

# Structural Studies of Geosmin Synthase, a Bifunctional Sesquiterpene Synthase with $\alpha\alpha$ Domain Architecture That Catalyzes a Unique Cyclization–Fragmentation Reaction Sequence

Golda G. Harris,<sup>†</sup> Patrick M. Lombardi,<sup>†,@</sup> Travis A. Pemberton,<sup>†</sup> Tsutomu Matsui,<sup>‡</sup> Thomas M. Weiss,<sup>‡</sup> Kathryn E. Cole,<sup>†,∇</sup> Mustafa Köksal,<sup>†,#</sup> Frank V. Murphy, IV,<sup>||</sup> L. Sangeetha Vedula,<sup>†</sup> Wayne K. W. Chou,<sup>‡</sup> David E. Cane,<sup>‡</sup> and David W. Christianson<sup>\*,†,§</sup>

<sup>†</sup>Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, United States

<sup>‡</sup>Department of Chemistry, Brown University, Box H, Providence, Rhode Island 02912-9108, United States

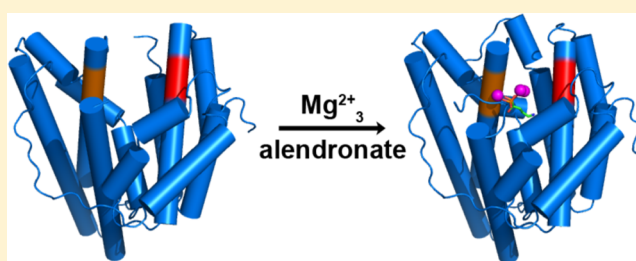
<sup>§</sup>Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Massachusetts 02138, United States

<sup>||</sup>Northeastern Collaborative Access Team/Cornell University, Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois 60439, United States

<sup>‡</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, P.O. Box 20450, Stanford, California 94309, United States

## Supporting Information

**ABSTRACT:** Geosmin synthase from *Streptomyces coelicolor* (ScGS) catalyzes an unusual, metal-dependent terpenoid cyclization and fragmentation reaction sequence. Two distinct active sites are required for catalysis: the N-terminal domain catalyzes the ionization and cyclization of farnesyl diphosphate to form germacradienol and inorganic pyrophosphate (PP<sub>i</sub>), and the C-terminal domain catalyzes the protonation, cyclization, and fragmentation of germacradienol to form geosmin and acetone through a retro-Prins reaction. A unique  $\alpha\alpha$  domain architecture is predicted for ScGS based on amino acid sequence: each domain contains the metal-binding motifs typical of a class I terpenoid cyclase, and each domain requires Mg<sup>2+</sup> for catalysis. Here, we report the X-ray crystal structure of the unliganded N-terminal domain of ScGS and the structure of its complex with three Mg<sup>2+</sup> ions and alendronate. These structures highlight conformational changes required for active site closure and catalysis. Although neither full-length ScGS nor constructs of the C-terminal domain could be crystallized, homology models of the C-terminal domain were constructed on the basis of ~36% sequence identity with the N-terminal domain. Small-angle X-ray scattering experiments yield low-resolution molecular envelopes into which the N-terminal domain crystal structure and the C-terminal domain homology model were fit, suggesting possible  $\alpha\alpha$  domain architectures as frameworks for bifunctional catalysis.



With more than 75000 members identified to date, terpenoids (also known as terpenes or isoprenoids) comprise the largest known family of natural products.<sup>1</sup> Their structural and stereochemical diversity allow myriad biological functions across all domains of life, such as defense against parasites and other predators, inter- and intraspecies communication, and intracellular signaling.<sup>2–4</sup> Many terpenoids are used as flavorings, such as menthol<sup>5</sup> and the cannabinoid  $\beta$ -caryophyllene;<sup>6</sup> fragrances, such as limonene<sup>7</sup> and sclareol;<sup>8</sup> advanced biofuels, such as farnesane<sup>9</sup> and bisabolane;<sup>10</sup> and pharmaceuticals, such as artemisinin<sup>11</sup> and Taxol.<sup>12</sup> Thus, the broad commercial importance of this family of natural products spans diverse chemical industries worldwide.<sup>13–16</sup>

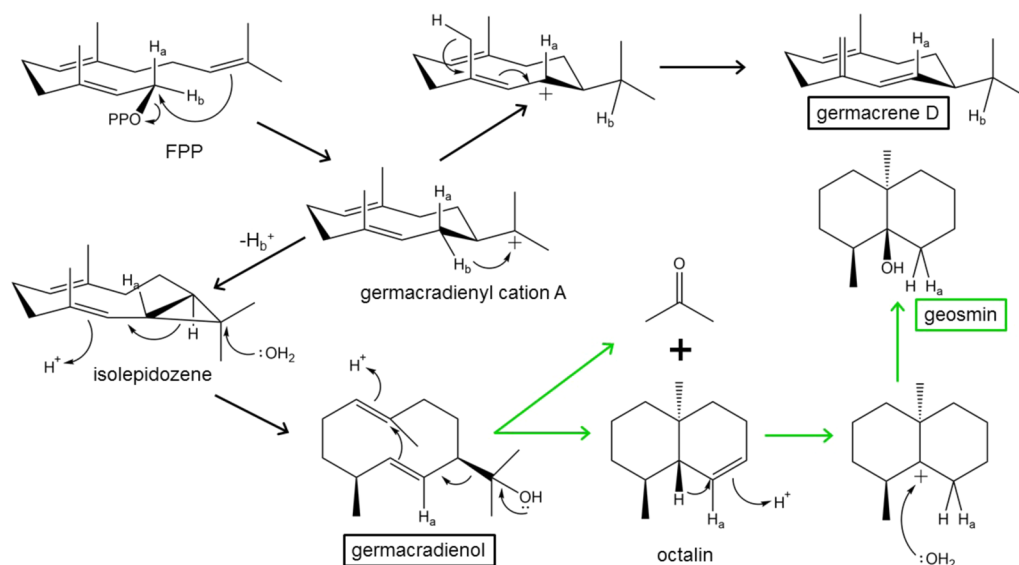
The vast chemodiversity of terpenoid natural products belies simple biosynthetic roots in the C<sub>5</sub> building blocks isopentenyl

diphosphate and dimethylallyl diphosphate, which are substrates for chain elongation reactions that produce acyclic isoprenoids such as the C<sub>15</sub> sesquiterpene farnesyl diphosphate (FPP) (Figure 1).<sup>17,18</sup> Isoprenoid diphosphates such as FPP then serve as substrates for terpenoid cyclases that generate diverse products containing multiple fused rings and stereocenters.<sup>19–24</sup>

Impressively, more than half of the carbon atoms of the isoprenoid diphosphate substrate typically undergo changes in bonding and/or hybridization during the course of a complex, multistep cyclization cascade. Given their biosynthetic roots in C<sub>5</sub> isoprenoid precursors, cyclic terpenoids usually contain 5*n* carbon atoms (*n* = 1, 2, 3, ...), unless the terpenoid has been

Received: October 20, 2015

Published: November 24, 2015



**Figure 1.** Cyclization sequence catalyzed by bifunctional geosmin synthase from *Streptomyces coelicolor* (ScGS). The FPP cyclization reaction (black arrows) is catalyzed by the N-terminal  $\alpha$  domain, and the retro-Prins fragmentation reaction (green arrows) is catalyzed by the C-terminal  $\alpha$  domain.

subject to additional biosynthetic modification preceding or following cyclization (e.g., methylation, acetylation, benzoylation, etc.).

The bicyclic natural product geosmin (Figure 1) is a noncanonical terpenoid, in that it contains 12 carbon atoms instead of  $5n$  carbon atoms.<sup>25,26</sup> Remarkably, the chemistry that yields this noncanonical terpenoid with  $C_{12} \neq C_{5n}$  is mediated by the terpenoid cyclase itself, rather than an upstream or downstream processing enzyme in the geosmin biosynthetic pathway.<sup>27,28</sup> Geosmin is a powerful odorant with an extremely low human detection threshold of <10 parts per trillion and is mainly responsible for the characteristic odor of freshly turned earth.<sup>29,30</sup> Although geosmin contributes to the pleasant, earthy flavor of beets,<sup>31</sup> it is also a commonly occurring contaminant of musty-tasting water, wine, and fish.<sup>32–35</sup> Geosmin is not known to cause human disease, but its detection and elimination from potable water sources are critical environmental and water quality issues.

Geosmin is produced by essentially all known species of the Gram-positive bacterial genus *Streptomyces*. Geosmin synthase from the soil bacterium *Streptomyces coelicolor* (ScGS) is a 726-residue, bifunctional sesquiterpene cyclase that catalyzes tandem metal ion-dependent cyclization and fragmentation reactions utilizing the  $C_{15}$  substrate FPP to yield  $C_{12}$  geosmin,  $C_3$  acetone, and inorganic pyrophosphate ( $PP_i$ ) (Figure 1).<sup>27,28,36,37</sup> The active site in the N-terminal domain of ScGS catalyzes the ionization-dependent cyclization of FPP to form  $PP_i$  and two cyclic products: germacradienol (major product, 85%) and germacrene D (minor product, 15%). After dissociation from the N-terminal domain, germacradienol is rebound to the active site of the C-terminal domain where it is converted to geosmin in a protonation-dependent cyclization reaction accompanied by the elimination of acetone through a retro-Prins reaction.<sup>28,38</sup> Analysis of the dependence of the ratio of geosmin to germacradienol on protein concentration has established that dissociation of the germacradienol intermediate is mandatory, suggesting that there is no direct channel for the transfer of the intermediate from the active site of the N-terminal domain to that of the C-terminal domain.<sup>28</sup> This diffusive transfer of the germacradienol intermediate is reminiscent of the established

mode of transfer of the copalyl diphosphate intermediate between the class II and class I terpenoid synthase domains of abietadiene synthase and other labdane synthases.<sup>39</sup> The observed fragmentation chemistry catalyzed by the C-terminal domain of ScGS is unprecedented in terpenoid cyclase reactions. Even more curious is the  $Mg^{2+}$  requirement for catalysis by the C-terminal domain, because the role for metal ion(s) is unclear in the absence of an isoprenoid diphosphate substrate; moreover, the fragmentation reaction does not require inorganic pyrophosphate.<sup>28</sup> These results are supported by experiments with the individual recombinant N-terminal and C-terminal domains as well as reconstituted mixtures of the two.<sup>28</sup>

Because the tandem cyclization–fragmentation reactions catalyzed by ScGS require two distinct active sites, a unique  $\alpha\alpha$  domain architecture is predicted for ScGS based on primary structure analysis.<sup>28</sup> Although bacterial terpenoid cyclases usually consist of a single  $\alpha$  domain, as first observed for the sesquiterpene cyclase pentalenene synthase,<sup>40</sup> plant cyclases often include multiple domains. For example,  $\alpha\beta$  domain architecture is found in 5-*epi*-aristolochene synthase from *Nicotiana tabacum*,<sup>41</sup> and  $\alpha\beta\gamma$  domain architecture is found in taxadiene synthase from *Taxus brevifolia*<sup>42</sup> (where  $\alpha$ ,  $\beta$ , and  $\gamma$  domains represent distinct folds, as classified by Oldfield<sup>43</sup>). In contrast, ScGS is believed to adopt  $\alpha\alpha$  domain architecture: the N-terminal domain and the C-terminal domain are separated by a 41-residue linker and share 28 and 29% amino acid sequence identity, respectively, with pentalenene synthase.<sup>28,36</sup> Each domain contains characteristic metal ion-binding motifs of class I terpenoid cyclases:<sup>44</sup> the aspartate-rich motif is found as D<sup>86</sup>DHFLE<sup>91</sup> and D<sup>455</sup>DYYP<sup>459</sup>, and the “NSE/DTE” motif is found as N<sup>229</sup>DLFSYQRE<sup>237</sup> and N<sup>598</sup>DVFSYQKE<sup>606</sup>.

Here, we report the X-ray crystal structure of the unliganded N-terminal cyclase domain of ScGS and the structure of its complex with three  $Mg^{2+}$  ions and alendronate. These structures show that this cyclase domain adopts the characteristic  $\alpha$  fold of a class I terpenoid synthase and also highlight the ligand-induced conformational changes required for complete active site closure and catalysis. Although neither full-length ScGS nor constructs of the C-terminal domain could be crystallized, the C-terminal domain is also predicted to adopt an  $\alpha$  fold homologous to that of

the N-terminal domain based on approximately 36% amino acid sequence identity between these domains. In the absence of crystal structure data for the full-length enzyme, small-angle X-ray scattering (SAXS) experiments with full-length ScGS and a nearly full-length construct containing residues 1–690 (ScGS<sub>690</sub>) suggest possible  $\alpha\alpha$  domain architectures that are consistent with and indeed may facilitate bifunctional catalysis.

## MATERIALS AND METHODS

**Expression and Purification of Full-Length ScGS and ScGS<sub>690</sub>.** The plasmid encoding full-length ScGS was prepared previously in the Cane laboratory, and this protein was expressed and purified as described previously.<sup>36</sup> However, we were unable to crystallize full-length ScGS. Because the last 36 residues were predicted to be disordered using the DISOPRED server<sup>45</sup> and could possibly hinder crystallization, we prepared a new construct in which these residues were deleted to yield a 690-residue protein. This truncated construct was prepared by polymerase chain reaction (PCR) mutagenesis using forward primer 5'-GGCATCTCAACTGGCACC GG TAGTACCCCGTTACAAGCCGAGTACC-3' and reverse primer 5'-AGGTACTCGGCCTTGTAACGGGGCTACTACCGGTGCCAGTTGAGGATGCC-3' (underlined nucleotides denote the mutagenic codons). PCR products were purified by gel electrophoresis and extracted according to the manufacturer's instructions (QIAGEN, QIAQuick Gel Extraction Kit). Purified PCR products were digested with *DpnI* at 37 °C for 1 h, transformed into XL1-Blue *Escherichia coli*, and plated on Lysogeny Broth (LB) agar medium typically containing 50  $\mu$ g/mL ampicillin. Single colonies were used to inoculate 5 mL cultures of LB medium with antibiotic (50  $\mu$ g/mL ampicillin). These 5 mL cultures were grown overnight (16–18 h) at 37 °C with 250 rpm shaking and pelleted by centrifugation (~4000g for 20 min at 4 °C), and plasmid DNA was extracted using a QIAGEN QIAprep Spin Miniprep Kit. The plasmid was confirmed to encode the desired 690-residue protein by sequencing at the DNA Sequencing Facility at the Perelman School of Medicine of the University of Pennsylvania.

A His<sub>6</sub> tag was added to the C-terminus of this protein to facilitate its purification. Thus, although 36 residues were deleted from the C-terminus of full-length ScGS, eight residues were added to the C-terminus of this construct (the His<sub>6</sub> motif and a two-residue linker) such that the ScGS<sub>690</sub> construct has a net loss of 28 residues from the C-terminus. PCR amplification (using *PfuUltra* DNA polymerase and 7% dimethyl sulfoxide) of an *NdeI*–*KpnI* fragment was conducted using forward primer 5'-GCAGCAGCACATATGACGCAACAGCCCTTCCAACCTCCGCAC-3' and reverse primer 5'-GCAGCAGCAGGTACC-CCGGTGCCAGTTGAGGATGCCGGC-3' (underlined nucleotides denote restriction sites for *NdeI* and *KpnI*, respectively). The PCR-amplified insert was purified by agarose gel electrophoresis and extracted. The insert was digested with *NdeI* and *KpnI* at 37 °C for 3 h, gel purified, and extracted to prepare the final insert fragment for ligation.

A variant of the pET22b vector (pET22bMV, Novagen) was used for ligation. A *KpnI* restriction site immediately upstream of the C-terminal His<sub>6</sub> tag was created using PCR with the following forward and reverse mutagenic primers: 5'-CGTCGACAAGC-TTGCGGCCGAGGTACCCACCACCACCACCACCTGAG-3' and 5'-CTCAGTGGTGGTGGTGGTGGTGGTGGTACCCTGCGGCCGCAAGCTTGTCGACG-3' (underlined nucleotides denote the restriction site for *KpnI*). The vector was digested with *NdeI*, *KpnI*, and calf intestinal alkaline phosphatase

at 37 °C for 3 h, gel purified, and extracted. The insert and vector were ligated using DNA ligase with 1:1, 2:1, and 5:1 insert:vector molar ratios at 15 °C overnight. Ligation products were transformed into XL1-Blue *E. coli* and plated on LB agar medium containing 50  $\mu$ g/mL ampicillin. Single colonies were used to inoculate 5 mL cultures of LB media with antibiotic (50  $\mu$ g/mL ampicillin), and plasmid DNA was extracted from the overnight cultures. DNA sequencing confirmed that the protein encoded by this plasmid, ScGS<sub>690</sub>, included residues 1–690 of ScGS and a C-terminal His<sub>6</sub> tag to facilitate purification.

This plasmid was transformed into BL21(DE3)pLysS *E. coli*, and cells were plated on LB agar medium containing 34  $\mu$ g/mL chloramphenicol and 100  $\mu$ g/mL ampicillin. Single colonies were used to inoculate 5 mL cultures of LB medium with antibiotic (34  $\mu$ g/mL chloramphenicol and 50  $\mu$ g/mL ampicillin). These 5 mL cultures were grown overnight (16–18 h) at 37 °C with 250 rpm shaking. Overnight cultures were used to inoculate 1 L cultures of LB medium with antibiotic (34  $\mu$ g/mL chloramphenicol and 50  $\mu$ g/mL ampicillin). These 1 L cultures were grown at 37 °C with 250 rpm shaking until the OD<sub>600</sub> reached 0.5–0.8. Cultures were cooled to 18–20 °C and induced with 1 mL of 1 M isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). Induction continued overnight at 18–20 °C with 250 rpm shaking.

Cells were harvested by centrifugation (4200g for 10 min at 4 °C) and resuspended in buffer A [50 mM Tris (pH 8.2), 50 mM NaCl, 5 mM MgCl<sub>2</sub>, 20% glycerol, and 5 mM  $\beta$ -mercaptoethanol (BME)] with protease inhibitor cocktail added. Cells were lysed by sonication, and the cell lysate was clarified by centrifugation (30000g for 1 h at 4 °C). The supernatant was loaded onto a Ni-IDA affinity column that had been equilibrated with buffer A. The column was washed with a 90% buffer A/10% buffer B [50 mM Tris (pH 8.2), 50 mM NaCl, 5 mM MgCl<sub>2</sub>, 250 mM imidazole, 20% glycerol, and 5 mM BME] mixture, and protein was eluted with a 20% buffer A/80% buffer B mixture. Protein purity was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). Unfortunately, we were unable to crystallize ScGS<sub>690</sub>, so we resorted to proteolytic treatment to generate a protein sample that would crystallize.

**Limited Proteolysis of ScGS<sub>690</sub>.** Fractions containing ScGS<sub>690</sub> were concentrated to 1 mg/mL and subjected to limited proteolysis by proteinase K using the following conditions: 1:1000 protease:substrate ratio in the presence of 5 mM CaCl<sub>2</sub> and incubation at room temperature for 30 min. The proteolysis reaction was quenched with 5 mM phenylmethanesulfonyl fluoride. The limited proteolysis mixture was concentrated to approximately 5 mL by centrifugation and loaded onto a Superdex-200 size exclusion column that had been equilibrated with size exclusion chromatography buffer [25 mM Tris (pH 8.2), 5 mM MgCl<sub>2</sub>, and 5 mM BME]. Two protein peaks were observed on the size exclusion chromatogram, corresponding to the molecular weight of a complex of two domains, and the molecular weight of a single domain. Fractions containing these two species were separated and concentrated to approximately 10 mg/mL. The protein purity was assessed by SDS–PAGE.

**Expression and Purification of the N-Terminal Domain, ScGS<sub>366</sub>.** Plasmid pRW22, containing ScGS residues 1–366 cloned into pET-21d, was prepared as previously described.<sup>28</sup> Plasmid pRW22 was transformed into BL21(DE3)pLysS *E. coli* cells, and the transformants were plated on LB agar medium containing 34  $\mu$ g/mL chloramphenicol and 100  $\mu$ g/mL ampicillin. Cell colonies were used to inoculate four 5 mL cultures of LB medium and antibiotic (34  $\mu$ g/mL chloramphenicol

nicol and 100  $\mu\text{g}/\text{mL}$  ampicillin) that were grown to saturation overnight at 37 °C with 250 rpm shaking. Overnight cultures were used to inoculate four 250 mL cultures of LB medium and antibiotic (34  $\mu\text{g}/\text{mL}$  chloramphenicol and 100  $\mu\text{g}/\text{mL}$  ampicillin) that were grown at 37 °C with 250 rpm shaking until the  $\text{OD}_{600}$  reached approximately 0.5. Protein expression was induced by the addition of 100  $\mu\text{L}$  of 1 M IPTG and allowed to continue overnight at 25 °C.

Cells were harvested by centrifugation at 6100g for 15 min and resuspended in lysis buffer [50 mM Bis-tris (pH 7.0), 20% glycerol, and 10 mM BME] with a protease inhibitor cocktail tablet added. The cells were lysed by sonication and then centrifuged at 57400g and 4 °C for 1 h. Following centrifugation, the supernatant was passed through a 0.22  $\mu\text{m}$  filter and then loaded onto a Q-Sepharose anion exchange column that had been equilibrated with buffer C [50 mM Bis-tris (pH 6.95), 20% glycerol, 5 mM  $\text{MgCl}_2$ , 10 mM BME, and 50 mM NaCl]. After being washed with 50 mL of buffer C, the protein was eluted with a 375 mL linear gradient of buffer C to buffer D [50 mM Bis-tris (pH 6.95), 20% glycerol, 5 mM  $\text{MgCl}_2$ , 10 mM BME, and 500 mM NaCl]. Fractions containing ScGS<sub>366</sub>, as determined by SDS–PAGE, were collected, and ammonium sulfate was added to a final concentration of 1.5 M. The solution was passed through a 0.22  $\mu\text{m}$  filter and loaded onto a methyl column that had been equilibrated in buffer E [50 mM Bis-tris (pH 6.95), 20% glycerol, 5 mM  $\text{MgCl}_2$ , 10 mM BME, and 1.5 M ammonium sulfate]. After being washed with 250 mL of buffer E, the protein was eluted using a 150 mL linear gradient from buffer E to buffer F [50 mM Bis-tris (pH 6.95), 20% glycerol, 5 mM  $\text{MgCl}_2$ , and 10 mM BME]. The purest fractions, as determined by SDS–PAGE, were pooled, concentrated to approximately 5 mL by ultrafiltration over a YM-10 membrane, and loaded onto a Superdex-200 size exclusion column that had been equilibrated in buffer G [25 mM Tris (pH 8.2), 5 mM  $\text{MgCl}_2$ , and 10 mM BME]. The purest fractions, as determined by SDS–PAGE, were collected and concentrated by ultrafiltration over a YM-10 membrane. A Bradford assay was conducted to determine the final protein concentration, and aliquots of 5–15 mg/mL ScGS<sub>366</sub> were prepared for crystallization trials.

**Expression Attempts with the C-Terminal Domain of ScGS.** Three constructs containing varying segments of the C-terminal domain of ScGS were prepared, each with a C-terminal His<sub>6</sub> tag: (1) residues 390–690, (2) residues 327–690, and (3) residues 327–726. Plasmids containing these constructs were transformed into BL21(DE3)pLysS *E. coli* and expressed as described above for ScGS<sub>690</sub>, except that 400  $\mu\text{L}$  of 1.0 M IPTG was used for induction. We attempted to purify these constructs by Ni-IDA affinity chromatography as with ScGS<sub>690</sub> but found that no expressed protein was in the soluble fraction of the lysate. It was determined that these C-terminal domain constructs were expressed primarily in inclusion bodies.

**X-ray Crystal Structure Determinations.** For ScGS<sub>690</sub> after limited proteolysis, fractions from the size exclusion column containing two-domain and single-domain proteins were isolated and used for separate crystallization trials using the sitting drop vapor diffusion method. Only single-domain protein samples yielded crystals. Typically, a 0.6  $\mu\text{L}$  drop of protein solution [7 mg/mL protein, 25 mM Tris (pH 8.2), 5 mM  $\text{MgCl}_2$ , 10 mM BME, and 1.5 mM sodium alendronate] was added to a 0.6  $\mu\text{L}$  drop of precipitant solution [0.2 M sodium acetate trihydrate (pH 7.0) and 20% (w/v) polyethylene glycol 3350] and equilibrated against a 100  $\mu\text{L}$  reservoir of precipitant solution at room temperature. Crystals formed from these conditions

contained only the N-terminal domain of ScGS, residues 1–338, as confirmed by mass spectrometry. Hence, this protein was designated ScGS<sub>338</sub>. Crystals of ScGS<sub>338</sub> belonged to space group  $P4_32_12$  ( $a = b = 67.2 \text{ \AA}$ ,  $c = 345.4 \text{ \AA}$ , and  $\alpha = \beta = \gamma = 90^\circ$ ) with two molecules in the asymmetric unit. These crystals diffracted X-rays to 2.11  $\text{Å}$  resolution at the National Synchrotron Light Source, beamline X-29, Brookhaven National Laboratory (Upton, NY).

For ScGS<sub>366</sub>, the hanging drop vapor diffusion method was used for crystallization. A 2.0  $\mu\text{L}$  drop of protein solution [5–15 mg/mL protein, 25 mM Tris (pH 8.2), 5 mM  $\text{MgCl}_2$ , and 10 mM BME] was added to a 2.0  $\mu\text{L}$  drop of precipitant solution [0.1 M Tris (pH 8.2), 0.5 M  $\text{MgCl}_2$ , and 28% PEG 400] and equilibrated against a 1 mL reservoir of precipitant solution at 4 °C. Showers of needles, approximately 0.005 mm  $\times$  0.005 mm  $\times$  0.040 mm, appeared in less than a week. X-ray diffraction data were collected from single needles using a 20  $\mu\text{m}$  wide beam at the Northeastern Collaborative Access Team beamline 24-ID-E, Advanced Photon Source, Argonne National Laboratory (Argonne, IL). Crystals belonged to space group  $P4_32_12$  ( $a = b = 76.8 \text{ \AA}$ ,  $c = 130.6 \text{ \AA}$ , and  $\alpha = \beta = \gamma = 90^\circ$ ) with one molecule in the asymmetric unit and initially diffracted X-rays to 2.4  $\text{Å}$  resolution.

Data collection from ScGS<sub>366</sub> needles was hindered by rapid crystal decay in the X-ray beam, even when crystals were frozen in liquid nitrogen. High-resolution X-ray diffraction disappeared after irradiation of the ScGS<sub>366</sub> needle crystals for just a few seconds. To produce the most complete data set of reflections achievable (88.7% complete overall, 77.6% complete in the highest-resolution shell), X-ray diffraction was recorded from several positions along the long axis of individual ScGS<sub>366</sub> needle crystals and then merged with diffraction data from crystals looped in different orientations. While the  $R_{\text{merge}}$  for the data is relatively high ( $R_{\text{merge}} = 0.201$  overall), the reflections exhibit a strong signal-to-noise ratio ( $I/\sigma = 4.00$  in the highest-resolution shell) and are internally consistent as judged by  $\text{CC}_{1/2}$  ( $\text{CC}_{1/2} = 0.896$  in the highest-resolution shell), a metric that is considered superior to  $R_{\text{merge}}$  in assessing data quality.<sup>46,47</sup>

X-ray diffraction data sets were indexed, integrated, and scaled using HKL2000.<sup>48</sup> The crystal structure of ScGS<sub>366</sub> was determined by molecular replacement using the program phenix.mr\_rosetta<sup>49</sup> with a model of pentalenene synthase<sup>40</sup> edited to match the ScGS<sub>366</sub> sequence and refined with Rosetta prior to rotation and translation function calculations. For the liganded ScGS<sub>338</sub> structure crystallized after limited proteolysis, the lower-resolution ScGS<sub>366</sub> structure was used as a search model. Manual model building and refinement were performed with COOT and PHENIX, respectively.<sup>50,51</sup> Structure validation of each final model was performed using MolProbity.<sup>52</sup> Data collection and refinement statistics for the unliganded and liganded structures of the N-terminal domain of ScGS are listed in Table 1. Disordered segments characterized by uninterpretable broken or missing electron density were excluded from the final models. In the structure of the ScGS<sub>338</sub>– $\text{Mg}^{2+}$ –alendronate complex, these segments included D116–A119; in the structure of unliganded ScGS<sub>366</sub>, these segments included D116–M121 and S167–A175.

**Homology Modeling.** Four online servers were used to create homology models of the C-terminal domain of ScGS: SWISS-MODEL,<sup>53</sup> I-TASSER,<sup>54</sup> Phyre2,<sup>55</sup> and HHPred/MODELLER.<sup>56,57</sup> For all models, residues G374–H698 (as expressed in the ScGS<sub>690</sub> construct containing a C-terminal His tag) or the full-length C-terminal domain (residues G374–H726) were provided as the sequences to be modeled. For I-TASSER, Phyre2, and HHPred/MODELLER, no template was

Table 1. Data Collection and Refinement Statistics

	unliganded ScGS <sub>366</sub>	liganded ScGS <sub>338</sub>
Data Collection		
resolution limits (Å)	34.3–3.20	48.2–2.11
no. of reflections measured (total/ unique)	22930/6094	567185/46408
space group	P4 <sub>3</sub> 2 <sub>1</sub> 2	P4 <sub>3</sub> 2 <sub>1</sub> 2
unit cell [ <i>a</i> , <i>b</i> , <i>c</i> (Å)]	76.8, 76.8, 130.6	67.2, 67.2, 345.4
<i>R</i> <sub>merge</sub> <sup>a,b</sup>	0.201 (0.326)	0.145 (0.918)
<i>I</i> / $\sigma$ ( <i>I</i> ) <sup>a</sup>	6.99 (4.00)	14.64 (1.68)
completeness (%) <sup>a</sup>	88.7 (77.6)	98.8 (98.6)
CC <sub>1/2</sub>	0.969 (0.896)	0.791 (0.391)
Refinement		
no. of reflections used in refinement/ test set	10938/534	43984/2282
<i>R</i> <sub>work</sub> <sup>a,c</sup>	0.240 (0.292)	0.207 (0.301)
<i>R</i> <sub>free</sub> <sup>a,c</sup>	0.288 (0.317)	0.250 (0.352)
no. of protein atoms <sup>d</sup>	2353	5099
no. of ligand atoms <sup>d</sup>	0	28
no. of Mg <sup>2+</sup> ions <sup>d</sup>	0	6
no. of solvent atoms <sup>d</sup>	6	145
root-mean-square deviation		
bonds (Å)	0.002	0.006
angles (deg)	0.7	0.9
average <i>B</i> factor (Å <sup>2</sup> )	43	44
main chain	44	43
side chain	43	45
ligand	–	37
solvent	14	44
Ramachandran plot (%) <sup>e</sup>		
allowed	94.7	92.7
additionally allowed	5.3	7.0
generously allowed	0.0	0.4
disallowed	0.0	0.0

<sup>a</sup>Numbers in parentheses refer to values for the highest-resolution shell of data. <sup>b</sup>*R*<sub>merge</sub> for replicate reflections.  $R = \sum |I_h - \langle I_h \rangle| / \sum \langle I_h \rangle$ , where *I*<sub>*h*</sub> is the intensity measure for reflection *h* and  $\langle I_h \rangle$  is the average intensity for reflection *h* calculated from replicate data. <sup>c</sup>*R*<sub>work</sub> =  $\sum ||F_o| - |F_c|| / \sum |F_o|$  for reflections contained in the working set. *R*<sub>free</sub> =  $\sum ||F_o| - |F_c|| / \sum |F_o|$  for reflections contained in the test set held aside during refinement (5% of total). *|F<sub>o</sub>|* and *|F<sub>c</sub>|* are the observed and calculated structure factor amplitudes, respectively. <sup>d</sup>Per asymmetric unit. <sup>e</sup>Evaluated with PROCHECK.

specified for modeling. Phyre2 created one ScGS C-terminal domain model using chain A of selinadiene synthase [Protein Data Bank (PDB) entry 4OKM]<sup>58</sup> as a structural template. I-TASSER created three different ScGS C-terminal models using threading templates identified by LOMETS<sup>59</sup> with the highest significance; the top 10 threading alignments included structures of *epi*-isozizaene synthase (PDB entries 3KB9 and 4LTV),<sup>60,61</sup> chain A of selinadiene synthase (PDB entry 4OKM),<sup>58</sup> chain A of hedycaryol synthase (PDB entry 4MC3),<sup>62</sup> and chain B of N219L pentalenene synthase (PDB entry 1HM7).<sup>63</sup> HHPred performed a multiple-sequence alignment, and MODELLER created one ScGS C-terminal domain model using chain A of hedycaryol synthase (PDB entry 4MC3)<sup>62</sup> as a template. When chain A of the ScGS<sub>338</sub> structure was specified as a template, I-TASSER and SWISS-MODEL created a sixth and seventh ScGS C-terminal domain model. For homology modeling of the full-length C-terminal domain (residues G374–H726), I-TASSER created five homology models based on chain A of the ScGS<sub>338</sub>

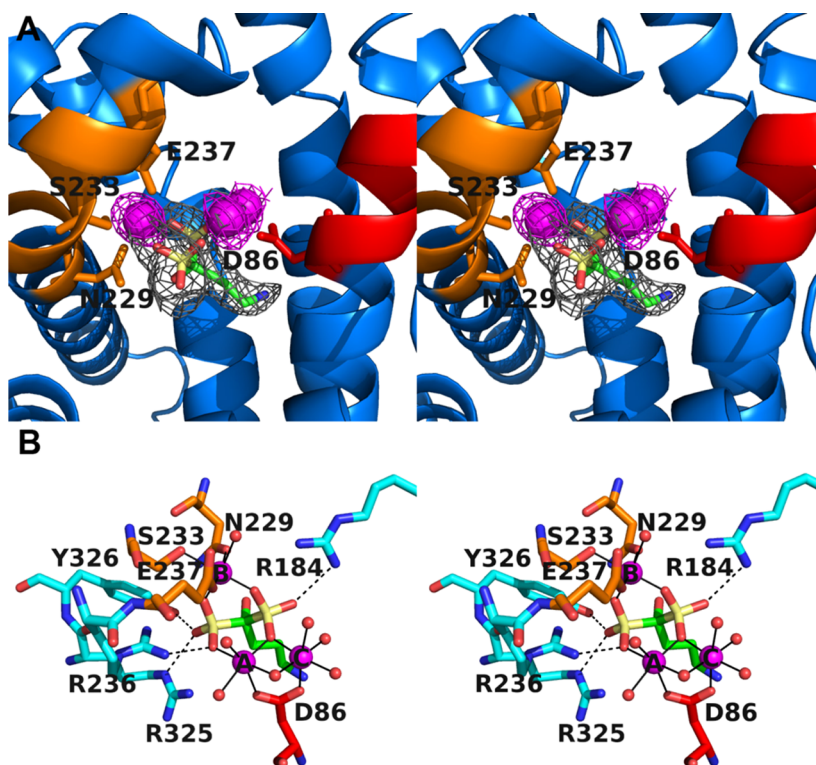
structure as a template. These models corresponded to the five largest clusters of predicted structures generated by I-TASSER.

For all C-terminal domain homology models, significant errors in geometry (e.g., clashes, poor rotamers, poor Ramachandran angles, and cis-peptide linkages) were detected by MolProbity. Phenix was used to normalize the geometry of these models using the program phenix.geometry\_minimization.<sup>51</sup> MolProbity was then used to assess the quality of each model. The side chains of certain asparagine, glutamine, and histidine residues were flipped as recommended to optimize hydrogen bond interactions. Final models were evaluated on the basis of their MolProbity statistics. We judged the C-terminal domain models generated by I-TASSER, with the N-terminal domain serving as a template, as the best models. The two C-terminal domain homology models, one based on the truncated construct ScGS<sub>690</sub> and the other based on full-length ScGS, were used in rigid body modeling with small-angle X-ray scattering data.

**Small-Angle X-ray Scattering (SAXS).** Samples of ScGS<sub>690</sub> after Ni-IDA purification were used for macromolecular SAXS experiments at the Stanford Synchrotron Radiation Lightsource (SSRL), beamline 4-2. Protein samples were concentrated to 21 mg/mL, flash-cooled in liquid nitrogen, and shipped to the beamline on dry ice. Aggregation was observed upon thawing at the beamline; samples were centrifuged to pellet the aggregated protein, and the supernatant was used for SAXS experiments. Protein samples were run on a Superdex 200 PC3.2 column (GE Healthcare) in SAXS buffer [25 mM Tris (pH 8.2), 5 mM MgCl<sub>2</sub>, and 5 mM dithiothreitol] in the FPLC-coupled solution SAXS system. X-ray exposure times of 1.0 s were used throughout the data collection. Scattering curves were corrected using scattering data from the size exclusion chromatography effluent as background. PRIMUS was used to analyze the scattering curve, Guinier plot, Kratky plot, and Porod–Debye plot and to calculate the radius of gyration (*R*<sub>g</sub>) and Porod volume for each data set.<sup>64</sup> GNOM was used to calculate the pair distribution function, *P*(*r*), and to estimate the maximal dimension (*D*<sub>max</sub>) for the scattering particle.<sup>65</sup> Molecular envelopes for the scattering particles were generated using the ShapeUp server.<sup>66</sup>

Samples of full-length ScGS were used for macromolecular SAXS experiments at SIBYLS beamline 12.3.1, Advanced Light Source, Lawrence Berkeley National Laboratory (Berkeley, CA). Samples of full-length ScGS were analyzed at three concentrations (2, 5, and 10 mg/mL) in a 96-well plate and shipped to the beamline at 4 °C. At the beamline, SAXS data for each sample were collected at exposure times of 0.5, 1.0, 2.0, and 4.0 s. The scattering curves for each sample were corrected for background by subtracting the scattering curve of the reference buffer. A merged data set for a sample of full-length ScGS at 10 mg/mL was created by combining the scattering at low *q* from the 0.5 s exposure scattering curve, and the scattering at high *q* from the 2.0 s exposure scattering curve. Data analysis was performed as described above for SAXS experiments with ScGS<sub>690</sub>. Data collection statistics for all SAXS experiments are listed in Table S1 of the Supporting Information.

**SAXS Rigid Body Modeling of the ScGS<sub>338</sub> Dimer.** Because the structure of liganded ScGS<sub>338</sub> contains a dimer in the asymmetric unit, this  $\alpha\alpha$  quaternary structure was used as a starting point for fitting the  $\alpha\alpha$  tertiary structure of ScGS<sub>690</sub> into the molecular envelope calculated from SAXS data. The scattering profile for the ScGS<sub>338</sub> dimer calculated with the FoXS server<sup>67</sup> was compared to the experimental scattering profile of ScGS<sub>690</sub> and the two-domain complex resulting from limited proteolysis using the FoXS  $\chi$  value. Manual rotation of



**Figure 2.** (A) Simulated annealing omit maps of the Mg<sup>2+</sup> ions (magenta map, contoured at 2.5 $\sigma$ ) and the bisphosphonate inhibitor alendronate (gray map, contoured at 2.5 $\sigma$ ) bound in the active site of the N-terminal domain construct ScGS<sub>338</sub> (structure determined at 2.11 Å resolution). Alendronate atoms are color-coded as follows: green for C, blue for N, red for O, and yellow for P. Metal ligands in the aspartate-rich segment (red) on helix D and the NSE segment (orange) on helix H are labeled. (B) Metal coordination interactions (solid lines) and hydrogen bond interactions (dashed lines) in the ScGS<sub>338</sub>-Mg<sup>2+</sup>-alendronate complex. Atoms are color-coded as follows: red for C in the aspartate-rich metal-binding motif, orange for C in the NSE metal-binding motif, cyan for C in the diphosphate recognition motif, and green for C in alendronate, blue for N, red for O, and yellow for P. Mg<sup>2+</sup> ions are shown as large magenta spheres and water molecules as small red spheres.

the two molecules of ScGS<sub>338</sub> was guided by visual inspection and fit within the molecular envelope calculated by the ShapeUp server;<sup>66</sup> the FoXS  $\chi$  value was used to score the fits of the different poses of the subunits obtained by manual rotation to the experimental scattering profile.

#### SAXS Rigid Body Modeling of the ScGS<sub>338</sub> Monomer with Homology Models of the ScGS C-Terminal Domain.

The unliganded and liganded structures of the N-terminal domain of ScGS were each paired with the seven homology models of the C-terminal domain of ScGS, resulting in 14 combinations representing possible models for the association of the N-terminal domain and the C-terminal domain in the full-length protein. Each of the 14 domain pairs was modeled against the small-angle X-ray scattering curve for ScGS<sub>690</sub> using the web service CORAL;<sup>68</sup> this process created 14 models for the association of the N- and C-terminal domains, with a linker of the appropriate length modeled as “dummy” residues connecting the two domains. This linker was manually deleted in the coordinate file, and the resulting coordinate file was compared to the scattering curve of ScGS<sub>690</sub> with the ShapeUp server<sup>66</sup> and the FoXS server<sup>67</sup> in the same manner as described above for the ScGS<sub>338</sub> dimer. The FoXS  $\chi$  value was used to score each model against the SAXS data measured from ScGS<sub>690</sub>. The two-domain models containing the unliganded and liganded N-terminal domain, each paired with the ScGS<sub>690</sub> C-terminal domain homology model from I-TASSER, were used for further analysis.

To model ScGS from SAXS data collected from the full-length enzyme, the unliganded and liganded N-terminal domain structures were each paired with the full-length C-terminal

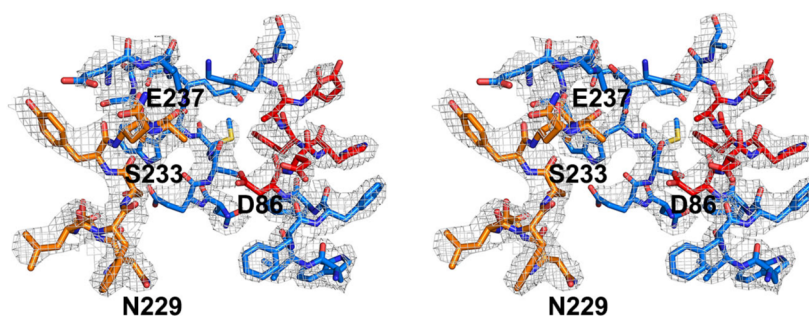
domain homology model from I-TASSER and modeled into the SAXS data using CORAL<sup>68</sup> in the same manner as described above. These two full-length ScGS models were similarly analyzed and evaluated using the FoXS  $\chi$  value.

## RESULTS

### Crystal Structure of the ScGS<sub>338</sub>-Mg<sup>2+</sup>-Alendronate Complex.

The N-terminal cyclase domain of ScGS resulting from proteolytic treatment of ScGS<sub>690</sub>, designated ScGS<sub>338</sub>, adopts the  $\alpha$ -helical class I terpenoid synthase fold as first observed in the crystal structure of avian farnesyl diphosphate synthase.<sup>69</sup> It is designated as the “ $\alpha$  fold” by Oldfield,<sup>43</sup> and its key features include characteristic metal ion-binding motifs D<sup>86</sup>DHFLE<sup>91</sup> and N<sup>229</sup>DLFSYQRE<sup>237</sup> in which underlined residues coordinate to a cluster of three Mg<sup>2+</sup> ions in complex with the bisphosphonate inhibitor alendronate (Figure 2; alendronate is formulated as the drug Fosamax used to treat osteoporosis<sup>70</sup>). Water molecules complete octahedral coordination polyhedra for all three Mg<sup>2+</sup> ions. In addition to metal ion coordination, the bisphosphonate moiety is also stabilized by hydrogen bonds with R184, R236, R325, and Y326. Thus, three metal ions, three basic residues, and the phenolic hydroxyl group of a tyrosine stabilize the anionic bisphosphonate moiety. These interactions are also expected to accompany the binding of the diphosphate group of substrate FPP.

Interestingly, in other cyclase structures, the three basic residues that donate hydrogen bonds to the substrate diphosphate group comprise a combination of arginine and lysine side chains.<sup>44</sup> However, in ScGS<sub>338</sub>, the three basic residues



**Figure 3.** Electron density map calculated with Fourier coefficients  $2|F_o| - |F_c|$  contoured at  $1.0\sigma$  showing the active site cleft of unliganded ScGS<sub>366</sub> at 3.2 Å resolution. Selected residues in the aspartate-rich and NSE metal-binding motifs are indicated (in the absence of bound Mg<sup>2+</sup> ions, the side chain of E237 is disordered and hence truncated at Cβ). Despite technical challenges with crystal decay, refinement yielded a reasonable structure illustrating the open active site conformation of the cyclase.

are all arginine side chains. Additionally, in some but not all cyclase active sites, there is a tyrosine that donates a hydrogen bond to a diphosphate oxygen of inorganic pyrophosphate (PP<sub>i</sub>) or a similar anionic ligand such as a bisphosphonate.<sup>44</sup> In addition to providing a sufficient driving force to trigger FPP ionization, these hydrogen bond interactions along with metal coordination interactions help maintain the active site in a closed conformation, inaccessible to bulk solvent, for the duration of the cyclization cascade. As a consequence, carbocation intermediates in catalysis are thereby protected from premature quenching by bulk solvent.

The pendant propylamino group of alendronate binds in the predominantly nonpolar region of the active site, along with eight solvent molecules. While the pK<sub>a</sub> of this amino group is approximately 10.9 in aqueous solution,<sup>70</sup> this pK<sub>a</sub> may be sufficiently perturbed by the hydrophobic environment so that the amino group is not protonated in the bound complex. The amino group donates a hydrogen bond to the backbone carbonyl oxygen of V82. It is interesting to note that four aromatic residues contribute significant surface area to the active site contour: F83, W192, F221, and W312. Although these residues may stabilize carbocation intermediates in catalysis through cation- $\pi$  interactions, they are not sufficiently well oriented to similarly stabilize the buried propylamino group of alendronate.

The ScGS<sub>338</sub>-Mg<sup>2+</sup><sub>3</sub>-alendronate complex crystallizes with an isologous dimer in the asymmetric unit. Accordingly, the active sites of the two protein molecules are oriented in an antiparallel fashion. Single- $\alpha$  domain terpenoid synthases generally crystallize as monomers or dimers; those that crystallize as dimers usually contain monomers oriented in an antiparallel fashion, as observed here. Occasionally, dimers comprised of parallel monomers are observed, for example, in the crystal structures of avian farnesyl diphosphate synthase<sup>69</sup> or methyl isoborneol synthase from *S. coelicolor*.<sup>71</sup> Assembly of the ScGS<sub>338</sub> dimer buries a total of 819 Å<sup>2</sup> of surface area, and the dimer interface is mainly comprised of helices G1, H1,  $\alpha$ -1, and I. Dimer assembly is most similar to that observed in pentalenene synthase.

In terms of primary structure, the N-terminal domain of ScGS is most similar to hedycaryol synthase (27% identical and 36% similar sequence).<sup>62</sup> Pentalenene synthase,<sup>40</sup> while having a slightly lower level of sequence identity (26%) with the N-terminal domain of ScGS, has a higher level of sequence similarity (41%). ScGS<sub>338</sub> has a helix topology nearly identical to that found in the structures of pentalenene synthase and *epi*-isozizaene synthase (EIZS).<sup>40,60</sup> Nonetheless, the tertiary structure of ScGS<sub>338</sub> most closely resembles that of selinadiene

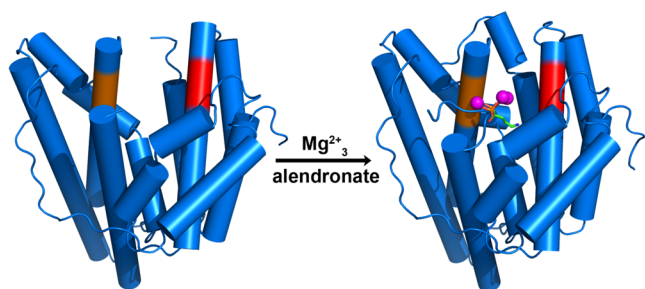
synthase<sup>58</sup> (25% identical and 38% similar; PDB entry 4OKZ) based on analysis with Dali.<sup>72</sup> The root-mean-square deviation (rmsd) of 275 C $\alpha$  atoms between ScGS<sub>338</sub> and selinadiene synthase is 1.6 Å (as calculated with Coot), indicating a closer structural match than ScGS<sub>338</sub> with 275 C $\alpha$  atoms of pentalenene synthase (2.1 Å), 293 C $\alpha$  atoms of EIZS (2.0 Å), or 267 C $\alpha$  atoms of hedycaryol synthase (2.6 Å), even though ScGS<sub>338</sub> has a slightly higher level of sequence identity with these three enzymes than with selinadiene synthase. It is noteworthy that the three-dimensional structures of selinadiene synthase and the N-terminal domain of ScGS, each of which catalyzes the 1,10-cyclization of FPP to yield a (*E,E*)-germacradienyl cation intermediate, are more similar to each other than to structures of enzymes that catalyze alternative cyclization reactions. For example, pentalenene synthase catalyzes an initial 1,11-ring closure reaction of FPP, while EIZS and hedycaryol synthase catalyze net 1,6- and 1,10-ring closure reactions, respectively, of the rearranged intermediate nerolidyl diphosphate. Evidently, similarities in tertiary structure rather than primary structure more accurately indicate similar template functions that determine the manner by which these enzymes enforce the initial cyclization of FPP.

There are also notable distinctions between the quaternary structures of these cyclases. Selinadiene synthase crystallizes as a tetramer, or a dimer of parallel dimers, in the asymmetric unit. EIZS forms crystallographic dimers that are oriented in an antiparallel fashion, but the crystallographic dimer interface differs from that of the ScGS<sub>338</sub> dimer. Pentalenene synthase appears to be most similar to ScGS<sub>338</sub>, as it crystallizes as a dimer in the asymmetric unit. The ScGS<sub>338</sub> dimer is reminiscent of the pentalenene synthase dimer, as both structures have a quasi-antiparallel orientation of subunits and share a similar dimer interface along helices H and I.

**Crystal Structure of Unliganded ScGS<sub>366</sub>.** Crystals of the unliganded N-terminal cyclase domain ScGS<sub>366</sub> were very poor in quality, as summarized in **Materials and Methods**. Crystals formed as very small needles, approximately 0.005 mm  $\times$  0.005 mm  $\times$  0.040 mm. X-ray diffraction data collected from these crystals were not highly complete, mainly because of rapid crystal decay in the X-ray beam. However, X-ray diffraction data collected from several crystals could be merged together with reasonably good agreement (CC<sub>1/2</sub> = 0.896 in the highest-resolution shell) to create a data set of reflections with 88.7% overall completeness. These data were sufficiently useful to determine a moderate 3.2 Å resolution structure of the unliganded N-terminal domain of ScGS to facilitate analysis of structural differences with the liganded enzyme. The ScGS<sub>366</sub>

model was refined to  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 0.240 and 0.288, respectively. A representative electron density map of the unliganded metal-binding motifs D<sup>86</sup>DHFLE<sup>91</sup> and N<sup>229</sup>DLFS-YQRE<sup>237</sup> is shown in Figure 3.

The overall structure of unliganded ScGS<sub>366</sub> is similar to that of ScGS<sub>338</sub> in its complex with three Mg<sup>2+</sup> ions and alendronate, with an rmsd of 1.2 Å for 301 C $\alpha$  atoms. However, in comparison with the liganded structure, notable conformational changes that reflect active site closure upon ligand binding are evident (Figure 4). Specifically, most of the helices shift 2–3 Å inward toward the



**Figure 4.** Comparison of unliganded and liganded N-terminal domain structures of ScGS reveals conformational changes that accompany active site closure triggered by the binding of three Mg<sup>2+</sup> ions and alendronate. These conformational changes are also expected to accompany the binding of three Mg<sup>2+</sup> ions and substrate FPP.

active site upon ligand binding; in particular, helix D, bearing the aspartate-rich motif, moves closer to helix H to optimize the geometry for coordination of Mg<sup>2+</sup><sub>A</sub> and Mg<sup>2+</sup><sub>C</sub> by D86. In contrast, metal-coordinating residues in the “NSE” motif on helix H adopt nearly identical conformations in the liganded and unliganded structures, so the Mg<sup>2+</sup><sub>B</sub>-binding site is essentially preformed for function. The most significant structural change upon ligand binding is the disorder–order transition for N168–A175 in helix F. This conformational transition contributes significant surface area to the enclosed active site contour and serves to fully close the active site during catalysis.

A significant conformational change in ligand-binding residues R325 and Y326 is also observed upon binding of the ligand to ScGS<sub>338</sub>. While these residues are oriented away from the active site in the unliganded state and make crystal contacts with E182 from a neighboring protein molecule, these residues both donate hydrogen bonds to the bisphosphonate moiety of alendronate in the liganded structure. These conformational changes are similar to those observed upon binding of the ligand to EIZS.<sup>60</sup> In the structure of EIZS complexed with Mg<sup>2+</sup><sub>3</sub>-PP<sub>i</sub> and the benzyltriethylammonium cation, conserved residues R338 and Y339 donate hydrogen bonds to the PP<sub>i</sub> anion. By contrast, in structures of unliganded EIZS, these residues either are disordered and not observed in the electron density or are flipped outward, away from the active site, similar to the unliganded structure of ScGS<sub>366</sub>. Other conformational changes generally observed upon binding of the ligand to class I terpenoid cyclases, such as the ordering of the H- $\alpha$ -1 loop,  $\alpha$ -1 helix, and the J–K loop (observed in EIZS), are not observed upon binding of the ligand to ScGS. Both the H- $\alpha$ -1 loop and the  $\alpha$ -1 helix are ordered and in very similar positions in the structures of unliganded ScGS<sub>366</sub> and liganded ScGS<sub>338</sub>. Electron density for the J–K loop is observed for almost identical portions of the sequence of ScGS in the unliganded and liganded structures (to K329 in the liganded structure and to N328 in the unliganded structure); in fact, helix J is slightly longer in the unliganded

structure, whereas the corresponding sequence uncoils to form the J–K loop in the liganded structure.

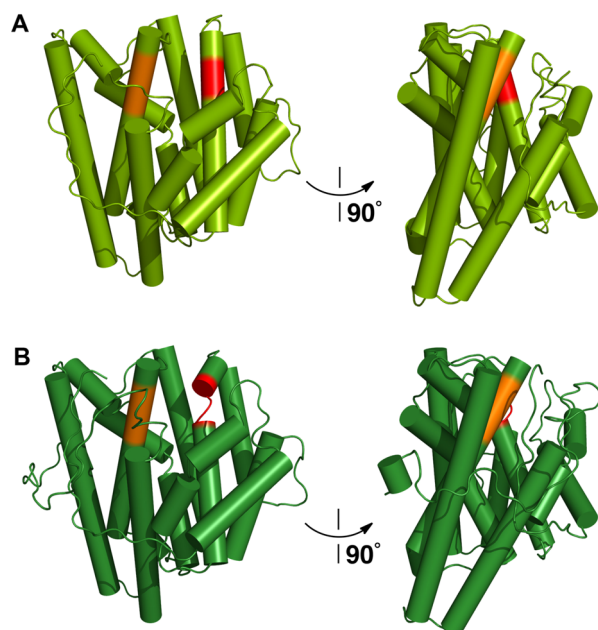
**Homology Model of the C-Terminal Fragmentation Domain.** Homology models of the C-terminal domain of ScGS based on the ScGS<sub>690</sub> truncation calculated by SWISS-MODEL, I-TASSER, Phyre2, and HHPred/MODELLER are similar to one another (Figure S1 of the Supporting Information), with pairwise rmsds ranging from 0.8 to 3.4 Å for 252–317 C $\alpha$  atoms. Homology models of the full-length C-terminal domain calculated by I-TASSER are also similar (Figure S2 of the Supporting Information), with pairwise rmsds ranging from 1.2 to 2.3 Å for 267–311 C $\alpha$  atoms.

The  $\alpha$ -helical topology of each model based on the ScGS<sub>690</sub> truncation is nearly identical to that of ScGS<sub>338</sub>, as expected on the basis of the approximately 36% amino acid sequence identity between the N-terminal and C-terminal domains. However, some minor differences are observed. Of the seven homology models calculated, additional, short  $\alpha$ -helices are predicted in the N-terminal segment preceding helix A (one model), loop segments that connect helices A and B (one model), helices D and E1 (two models), helices G2 and H (five models), helices I and J (two models), and the C-terminal segment following helix J (three models). Some helices present in ScGS<sub>338</sub> are absent in the SWISS-MODEL homology model (helix C), or shorter in the HHPred/MODELLER model (helix H). Helix D is shorter in several of the models (three models generated by I-TASSER without using a template, one model generated by HHPred/MODELLER, and one model generated by Phyre2); consequently, the aspartate-rich metal-binding motif is located at the beginning of a loop segment in these models. In the SWISS-MODEL homology model, helix J is broken into three shorter  $\alpha$ -helices connected by loop segments.

Following energy minimization using subroutines in PHE-NIX<sup>51</sup> and evaluation with MolProbity,<sup>52</sup> we judged the homology model generated by I-TASSER using the structure of ScGS<sub>338</sub> as a template as the highest-quality model based on molecular geometry statistics (Figure 5A). This model comprises a general reference point for understanding structure–activity relationships. Intact metal-binding motifs are located at the mouth of the active site: the aspartate-rich segment D<sup>455</sup>DYYP<sup>459</sup> on helix D and the NSE motif N<sup>598</sup>DVFSYQKE<sup>606</sup> on helix H. Although the aspartate-rich motif lacks the third aspartate residue that typically characterizes this motif, only the first aspartate coordinates to Mg<sup>2+</sup> ions in the crystal structure of ScGS<sub>338</sub>, so this could also be the case for D455 in the C-terminal domain. On the other hand, one or both aspartates of the D<sup>455</sup>DYYP<sup>459</sup> motif may well play a mechanistically distinct role as a general acid for proton-initiated cyclization of the rebound germacradienol intermediate, as discussed below.

Because I-TASSER yielded what we judged to be the best homology model of the C-terminal domain based on the truncated construct ScGS<sub>690</sub>, we also used I-TASSER exclusively to generate five homology models of the full-length C-terminal domain. These models correspond to the five largest clusters of structures predicted by I-TASSER, and these models were similarly evaluated and optimized using MolProbity and PHENIX. Each model exhibits an  $\alpha$ -helical topology similar to that of ScGS<sub>338</sub>, with minor differences: additional, short  $\alpha$ -helices are predicted in N-terminal segment before helix A (one model), loop segments that connect helices G2 and H (four models), and the C-terminal segment after helix J (three models). In one of these models (model 5), helix I is shorter than in ScGS<sub>338</sub> and an additional  $\alpha$ -helix is inserted between helices I





**Figure 5.** Homology models of the C-terminal domain of ScGS<sub>690</sub> (G374–R690) (A) and full-length ScGS (G374–H726) (B) generated by I-TASSER using the crystal structure of ScGS<sub>338</sub> as a template. For reference, the locations of the aspartate-rich motif (D<sup>455</sup>DYYP, helix D) and NSE motif (N<sup>598</sup>DVFS<sup>602</sup>YQKE<sup>606</sup>, helix H) are colored red and orange, respectively.

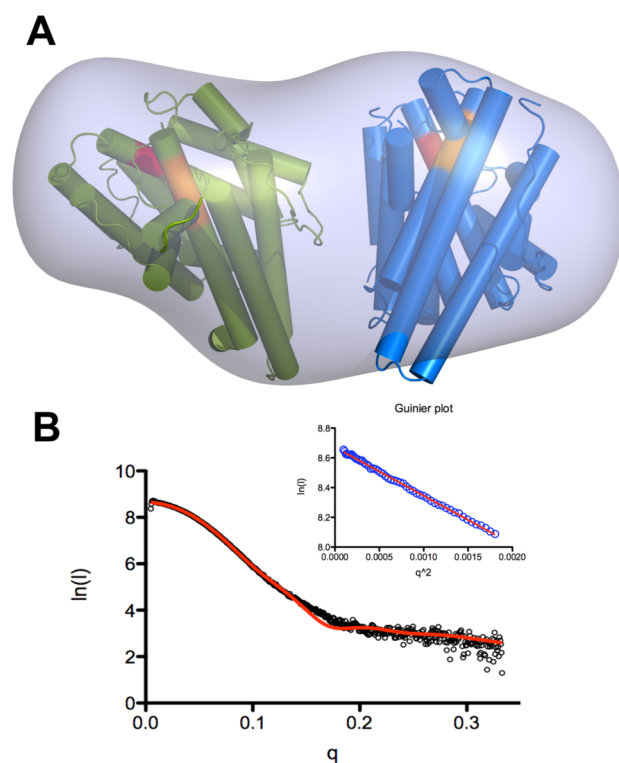
and J. Two of the  $\alpha$ -helices found in ScGS<sub>338</sub> are broken into two  $\alpha$ -helices in the full-length C-terminal domain homology models: helix J (model 3) and helix D (model 1 and model 4). Consequently, the aspartate-rich motif of model 1 is partially located on a loop segment at the break in helix D. Of the five models generated, we judged model 1 to be the best based on MolProbity statistics (Figure 5B). This model contains two additional short  $\alpha$ -helices when compared with ScGS<sub>338</sub>, lying in the loop segment between helices G2 and H and in the C-terminal segment after helix J, near the very end of the sequence (P722–T725). The break in helix D noted above causes the aspartate-rich motif of model 1 to lie partially on a loop segment at this break. The presence or absence of the 28-residue C-terminal segment does not appear to significantly impact the overall structure prediction, as might be expected for a disordered segment.

In the homology models of the C-terminal domain based on full-length ScGS or the truncated ScGS<sub>690</sub> construct, the chemical nature of the active site pocket is mainly nonpolar, as observed for ScGS<sub>338</sub>. The ring faces of two aromatic residues (F556 and W688) in part define the active site contour, and these residues could potentially stabilize carbocation intermediate(s) in the cyclization cascade through cation– $\pi$  interactions. Some partially polar residues are located in the active site pocket (C428, T452, and T561), but it is not clear that these residues could serve as general base/general acid residues in the retro-Prins fragmentation reaction catalyzed in this domain.

**Low-Resolution SAXS Structures of ScGS<sub>690</sub> and Full-Length ScGS.** Hypothesizing that the predicted disorder of the C-terminus of ScGS hindered crystallization of the full-length protein, we prepared the truncated construct ScGS<sub>690</sub> and used it for crystallization trials. Gas chromatographic–mass spectrometric analysis of cyclization products indicated that deletion of 36 residues from the C-terminus of ScGS and replacement with a

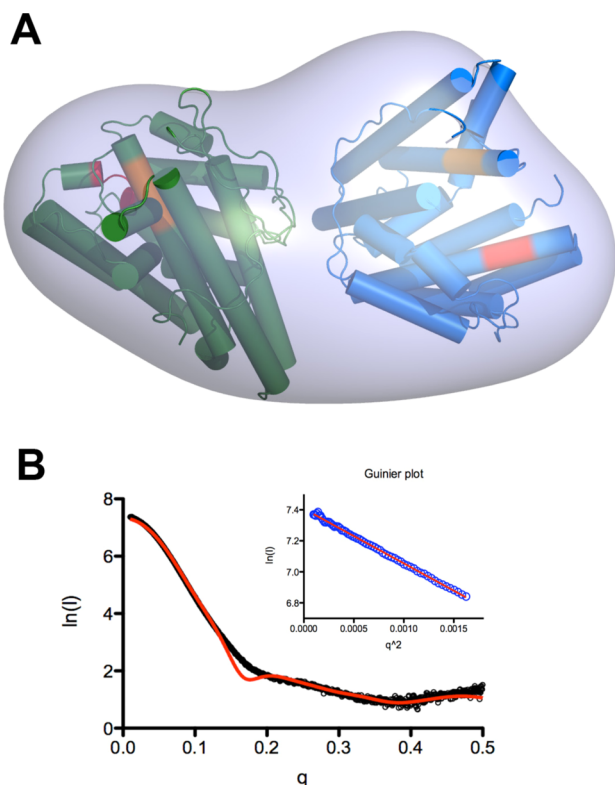
His<sub>6</sub> tag abolished catalytic activity in the C-terminal domain: the ScGS<sub>690</sub> construct generated increased concentrations of germacradienol (the cyclization product of the N-terminal domain) but no geosmin (data not shown). Thus, even though the C-terminus of ScGS is most likely disordered, it is absolutely required for catalytic activity in the C-terminal domain. Even so, both ScGS<sub>690</sub> and full-length ScGS proved to be excellent samples for small-angle X-ray scattering (SAXS) measurements.

SAXS data collected for ScGS<sub>690</sub> and full-length ScGS show that both proteins are well-folded and monodisperse in solution, and scattering in the Guinier region is linear (Figures 6 and 7;



**Figure 6.** (A) Best fit of the N-terminal domain (ScGS<sub>366</sub>, blue) and ScGS<sub>690</sub> C-terminal domain homology model (light green) to the molecular envelope calculated from SAXS data collected from ScGS<sub>690</sub>, generated by the ShapeUp server. (B) Experimental SAXS scattering profile for ScGS<sub>690</sub> (black circles) superimposed on the theoretical scattering profile for the  $\alpha\alpha$  domain model in panel A (red). For this model of  $\alpha\alpha$  domain assembly, the  $\chi$  value calculated by FoXS is 4.47. In the inset, the Guinier region is linear for SAXS data collected from ScGS<sub>690</sub>.

structural parameters are listed in Table S1 of the Supporting Information). The radius of gyration ( $R_g$ ) calculated by PRIMUS using the Guinier approximation for ScGS<sub>690</sub> is 31.3 Å, and the  $R_g$  for full-length ScGS is 32.5 Å. The slight increase in  $R_g$  for full-length ScGS is consistent with a slightly larger scattering particle because of 28 additional residues at the C-terminus. Because the FoXS-calculated  $R_g$  for ScGS<sub>338</sub> is 18.8 Å, the measured  $R_g$  values for ScGS<sub>690</sub> and full-length ScGS indicate a protein larger than a single  $\alpha$  domain, i.e., a monomeric protein containing two  $\alpha$  domains. Calculation of the pair-distance distribution function yields  $D_{\max}$  values of 110 and 117 Å for ScGS<sub>690</sub> and full-length ScGS, respectively. These  $D_{\max}$  values cannot distinguish between a monomer and a dimer. Inspection of the SAXS molecular envelopes for ScGS<sub>690</sub> and full-length ScGS, however,



**Figure 7.** (A) Best fit of the N-terminal domain (ScGS<sub>366</sub>, blue) and the full-length C-terminal domain homology model (forest green) to the molecular envelope calculated from SAXS data collected from full-length ScGS, generated by the ShapeUp server. (B) Experimental SAXS scattering profile for full-length ScGS (black circles) superimposed on the theoretical scattering profile for the  $\alpha$  domain model in panel A (red). For this model of  $\alpha$  domain assembly, the  $\chi$  value calculated by FoXS is 2.64. In the inset, the Guinier region is linear for SAXS data collected from full-length ScGS.

strongly suggests a monomeric protein containing two  $\alpha$  domains (Figures 6 and 7).

Using the experimentally determined crystal structures of the liganded and unliganded N-terminal domain of ScGS, and the best homology models of the C-terminal domain of ScGS<sub>690</sub> (G374–R690) and full-length ScGS (G374–H726) generated by I-TASSER, we performed rigid body modeling with CORAL<sup>68</sup> to generate four models illustrating potential domain assembly modes consistent with SAXS profiles. Using the FoXS server, the best models of ScGS<sub>690</sub> and full-length ScGS were judged to be those with the lowest  $\chi$  value (a lower  $\chi$  value indicates a closer match between the experimental scattering profile and the calculated scattering profile for a given structure), and these models are illustrated in Figures 6 ( $\chi = 4.47$ ) and 7 ( $\chi = 2.64$ ), respectively. The FoXS-calculated  $R_g$  values for the models in Figures 6 and 7 are 29.4 and 29.1 Å, respectively, consistent with the dimensions of a monomeric protein containing two  $\alpha$  domains. The orientation of one  $\alpha$  domain with respect to the other differs in these models. Thus, the active sites in the N-terminal and C-terminal domains are oriented in roughly parallel fashion in the model of ScGS<sub>690</sub> but are oriented approximately 90° away from each other in the model of full-length ScGS. It is nonetheless interesting to note that the same general faces of the N-terminal and C-terminal domains appear to mediate  $\alpha\alpha$  domain assembly. Regardless of the specific orientation of one  $\alpha$  domain with respect to the other, the molecular envelopes of

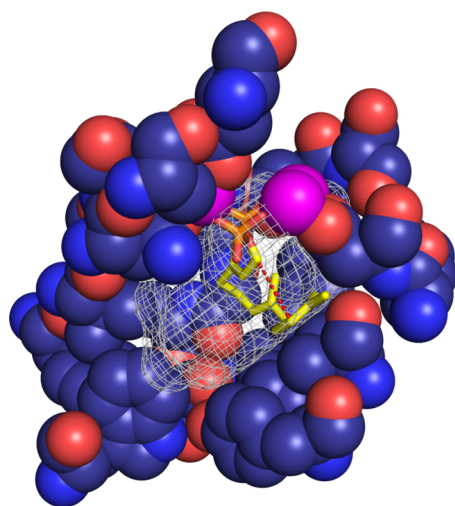
ScGS<sub>690</sub> and full-length ScGS are consistent with a model of  $\alpha\alpha$  domain assembly in which the active sites of the two domains are oriented away from each other rather than toward each other. Such a model would account for the lack of direct channeling of germacradienol between the two active sites. In fact, the observation of free germacradienol in solution is consistent with mandatory release of germacradienol from the N-terminal domain and diffusive rebinding to the C-terminal domain of the same or another geosmin synthase protein prior to conversion to geosmin.<sup>28</sup>

## DISCUSSION

Because the first step of the cyclization cascade catalyzed by a class I terpenoid cyclase is always the ionization of the diphosphate group of FPP, the three-dimensional contour of the active site encodes all the information required to direct a unique sequence of carbon–carbon bond-forming reactions. The most important step in such a sequence is the initial macrocyclization reaction, because this sets the stage for all subsequent steps. The initial binding conformation of FPP is enforced by the active site contour and determines whether a 1,6- or 1,7-cyclization reaction occurs via an ionization–isomerization–reionization sequence, whether a 1,10-cyclization reaction occurs with or without allylic isomerization, or whether a direct 1,11-cyclization reaction will take place. The active site contours of enzymes that direct initial 1,6-cyclizations, such as *epi*-isozizaene synthase<sup>60</sup> and trichodiene synthase,<sup>73</sup> tend to be somewhat narrow and deep. In contrast, enzymes that direct 1,10-cyclization reactions, such as selinadiene synthase<sup>58</sup> [via direct formation of an (*E,E*)-germacradienyl cation intermediate] and hedycaryol synthase<sup>62</sup> [via initial isomerization to nerolidyl diphosphate followed by formation of a (*Z,E*)-helminthogermacradienyl cation], as well as an enzyme that directs a 1,11-cyclization reaction, pentalene synthase,<sup>40</sup> all appear to have wider, shallower active sites (Figure S3 of the Supporting Information).

Intriguingly, the N-terminal domain of ScGS has an active site contour that is deeper than that of other enzymes that catalyze initial 1,10-cyclization reactions, and it also has a neck wider than that of enzymes that catalyze initial 1,6-isomerization–cyclization reactions (Figure S3 of the Supporting Information). Thus, the three-dimensional contour of the ScGS N-terminal domain active site appears to be a structural hybrid in that it exhibits features of both general active site shapes. Additionally, the active site contour of the N-terminal domain branches at its base, forming a shape somewhat like that of a “mitten”, a feature that is not observed in the 1,10-cyclization active sites of either selinadiene synthase (which generates a germacradienyl cation from FPP) or hedycaryol synthase (which generates a helminthogermacradienyl cation via nerolidyl diphosphate). Otherwise, the chemical nature of the N-terminal domain active site is similar to that of other terpenoid cyclases in that it is mainly hydrophobic. The active site contains a few aromatic residues, one of which in particular (F83) has its ring face oriented to permit stabilization of carbocation intermediates by cation– $\pi$  interactions.

In the active site of the N-terminal domain of ScGS, the diphosphate group of FPP is presumed to bind in a similar manner to the bisphosphonate group of alendronate as observed in its complex with ScGS<sub>338</sub>. With FPP locked in this orientation, there is sufficient active site volume for FPP to adopt the conformation required for the initial 1,10-cyclization reaction following ionization of the diphosphate group (Figure 8). The



**Figure 8.** Model of substrate FPP bound in the active site of the N-terminal domain of ScGS. The position of the FPP diphosphate group is based on the position of the bisphosphonate group of alendronate in the ScGS<sub>338</sub>–Mg<sup>2+</sup><sub>3</sub>–alendronate complex. Protein atoms are color-coded as follows: dark blue for C, blue for N, red for O, and magenta for Mg<sup>2+</sup>. Atoms are color-coded for the FPP stick figure as follows: yellow for C, orange for P, and red for O. The active site contour is indicated by light gray meshwork, and a red dashed line indicates the trajectory of formation of an initial carbon–carbon bond between C1 and C10 of FPP. The additional volume in the lower active site may accommodate a trapped solvent molecule that quenches the final carbocation intermediate. Active site contour in meshwork created with VOIDOO.<sup>80</sup>

initially formed germacradienyl cation intermediate must be deprotonated to generate the novel isolepidozene intermediate. Proton-initiated ring opening of the vinylcyclopropyl moiety and quenching of the resultant homoallylic cation by water then afford germacradienol, the major product of the cyclization reaction catalyzed by the N-terminal domain. Possible general acids and/or general bases that may participate in this reaction might include E161, which is located near Mg<sup>2+</sup><sub>C</sub> at the neck of the active site. Although E161 forms a salt link with R184, it appears to be ideally located to deprotonate H<sub>b</sub> from the C1 atom to form isolepidozene. Other potential catalytic groups include H226 and H320, which are situated more deeply in the active site. Finally, because water is a cosubstrate in the reaction, one or more of the active site water molecules observed in the ScGS<sub>338</sub>–Mg<sup>2+</sup><sub>3</sub>–alendronate complex may remain trapped in the active site upon FPP binding. There is sufficient extra volume in the active site of the modeled enzyme–substrate complex to accommodate trapped solvent (Figure 8).

The C-terminal domain of ScGS (Figure 5) is predicted to adopt the fold of a class I terpenoid cyclase based on its high level of amino acid sequence identity with the N-terminal domain. However, the C-terminal domain of ScGS catalyzes an unprecedented cyclization–fragmentation reaction involving a retro-Prins reaction, resulting in fragmentation of the C<sub>15</sub> substrate germacradienol to yield C<sub>12</sub> geosmin and C<sub>3</sub> acetone. The proton-initiated cyclization–fragmentation reaction may be initiated by one or both of the conserved aspartate residues of the D<sup>455</sup>DYYP<sup>459</sup> motif, consistent with the observation that the D455N/D456N double mutant of geosmin synthase no longer produces geosmin but accumulates only germacradienol and germacrene D, the characteristic products of the N-terminal domain. The aspartates of the DDYYP domain may therefore play a role in proton-initiated polyene cyclization similar to that

of the conserved DXDD motif (general acid residue underlined) of typical class II diterpene synthases such as *ent*-copalyl diphosphate synthase,<sup>74,75</sup> the triterpene cyclase squalene-hopene synthase,<sup>76</sup> and oxidosqualene-lanosterol synthase, which contains a closely related DCTA motif.<sup>77</sup>

The cyclization–fragmentation reaction catalyzed by the C-terminal domain of geosmin synthase does not involve the ionization of an isoprenoid diphosphate such as FPP to generate the initial carbocation, the hallmark of catalysis by a class I terpenoid cyclase, yet the cyclization–fragmentation reaction nevertheless has an absolute requirement for Mg<sup>2+</sup> ion(s).<sup>28</sup> Catalysis by a class I terpenoid cyclase typically involves coordination and activation of the substrate diphosphate group by three Mg<sup>2+</sup> ions to form a Mg<sup>2+</sup><sub>3</sub>–diphosphate/PP<sub>i</sub> complex that remains bound for the duration of the cyclization reaction; the closed conformation of the enzyme active site is exemplified by that of the Mg<sup>2+</sup><sub>3</sub>–alendronate complex with ScGS<sub>338</sub> (Figure 4). Because germacradienol, the substrate for the C-terminal domain of ScGS, lacks a diphosphate group altogether and the reaction has been shown not to require inorganic pyrophosphate, neither the mechanistic nor the structural basis for the requirement for Mg<sup>2+</sup> by the C-terminal domain is as yet apparent.

The crystal structure of the ScGS<sub>338</sub>–Mg<sup>2+</sup><sub>3</sub>–alendronate complex and the amino acid sequence identity between the N- and C-terminal domains of ScGS suggest a possible structural basis for the Mg<sup>2+</sup> requirement for catalysis by the C-terminal domain. Specifically, residues that coordinate to catalytic Mg<sup>2+</sup> ions in the N-terminal domain are conserved in the C-terminal domain as D455 in the aspartate-rich motif, and N598, S602, and E606 in the “NSE” motif. Additionally, the three basic residues and the tyrosine residue that hydrogen bond with the bisphosphonate anion in the N-terminal domain are conserved in the C-terminal domain as R552, K605, R694, and Y695. Metal-binding residues as well as residues that hydrogen bond with diphosphate/PP<sub>i</sub> are highly conserved in all class I terpenoid cyclases.<sup>44</sup> Conservation of these structural elements in the C-terminal domain of ScGS strongly suggests that this domain similarly binds one or more Mg<sup>2+</sup> ions. The binding of Mg<sup>2+</sup> might facilitate full active site closure of the C-terminal domain, just as it does for the N-terminal domain (Figure 4), but without interactions with a diphosphate group. As for all class I terpenoid cyclases, complete active site closure ensures the protection of reactive carbocation intermediates from premature quenching by bulk solvent. Notably, the more than 70 deduced or verified *Streptomyces* geosmin synthase sequences that have been reported to date in the protein databases exhibit an exceptionally high level of sequence conservation (60–85% sequence identity over more than 720 amino acids), with essentially 100% identity of the conserved aspartate-rich and NSE motifs in both the C-terminal and N-terminal domains.

The fragmentation reaction catalyzed in the C-terminal domain of ScGS utilizes germacradienol as a substrate. The first step is a proton-assisted cyclization, covalently linking C2 and C7 along with retro-Prins fragmentation to eliminate the 2-propanol side chain as acetone. The resulting bicyclic intermediate, octalin, has a *trans*-decalin-like configuration and subsequently undergoes protonation and a hydride shift to yield a tertiary carbocation, which is quenched by a water molecule to yield the final product alcohol, geosmin, which similarly adopts a *trans*-decalin-like configuration. It is highly unusual for a cyclization reaction to be initiated by protonation in a class I terpenoid cyclase; ordinarily, such a protonation is the chemical

strategy for cyclization adopted by a class II terpenoid cyclase, which adopts a completely unrelated protein fold.<sup>21</sup>

Both a general acid and a general base are required for the retro-Prins fragmentation reaction catalyzed by the C-terminal domain of ScGS (Figure 1), yet analysis of the homology model reveals a relative dearth of chemical functionality within the mainly nonpolar active site pocket beyond the conserved D455 and D456 residues. On the other hand, in the crystal structure of the ScGS<sub>338</sub>–Mg<sup>2+</sup><sub>3</sub>–alendronate complex with the N-terminal domain, several ordered solvent molecules are trapped in the active site. Corresponding solvent molecules might conceivably remain bound in the active site of the C-terminal domain upon the binding of the substrate germacradienol, with one of these waters ultimately quenching the final carbocation intermediate, as shown in Figure 1.

Finally, rigid body modeling with SAXS data provides a plausible model for the assembly of the N- and C-terminal  $\alpha$  domains of ScGS. Manual docking of the N-terminal domain structure and the homology model of the C-terminal domain yielded poor fits to the scattering profile. However, CORAL generated models with better fits, although not all of these models made good chemical sense. For example, domain interactions in some models appear to be mediated solely by loops rather than secondary structural elements (data not shown). Intriguingly, depending on the flexibility of the linker connecting the N-terminal domain and the C-terminal domain, multiple domain orientations could be feasible. While a single  $\alpha\alpha$  domain orientation cannot be definitively established, it is clear that the two  $\alpha$  domains interact with each other through a substantial interface based on the cross sections of the molecular envelopes shown in Figures 6 and 7. Although it could be argued that  $\alpha\alpha$  domain assembly in full-length ScGS ought to mimic  $\alpha\alpha$  dimer assembly as observed in the crystal structure of ScGS<sub>338</sub>, this particular  $\alpha\alpha$  assembly mode does not yield a satisfactory fit to the scattering data and yields a much higher  $\chi$  value of 9.96 (Figure S4 of the Supporting Information).

Regardless of the orientation of one domain to the other in the  $\alpha\alpha$  assembly, there is no channel specifically formed between the two active sites. Although bifunctional catalysis might be facilitated by a simple proximity or clustering effect,<sup>78,79</sup> it is not clear that germacradienol generated in the N-terminal domain is captured by the C-terminal domain of the same or a different geosmin synthase molecule. Future studies will continue to probe these aspects of catalysis and the role of domain architecture in bifunctional catalysis by ScGS.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b01143.

SAXS data collection statistics and structural parameters (Table S1), homology models of the C-terminal domain of ScGS<sub>690</sub> generated by various modeling programs (Figure S1), homology models of the C-terminal domain of full-length ScGS generated by I-TASSER (Figure S2), comparison of sesquiterpene cyclase active site contours (Figure S3), and best fit of the ScGS<sub>338</sub> dimer to the SAXS-derived molecular envelope of full-length ScGS (Figure S4) (PDF)

## ■ Accession Codes

The atomic coordinates and structure factors of the ScGS<sub>338</sub>–Mg<sup>2+</sup><sub>3</sub>–alendronate complex and unliganded ScGS<sub>366</sub> have been deposited in the Protein Data Bank as entries SDZ2 and 5DW7, respectively.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [chris@sas.upenn.edu](mailto:chris@sas.upenn.edu). Telephone: (215) 898-5714.

### Present Addresses

@P.M.L.: Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205.

∇K.E.C.: Department of Molecular Biology and Chemistry, Christopher Newport University, 1 Avenue of the Arts, Newport News, VA 23606.

#M.K.: Department of Molecular Biology and Genetics, İzmir Institute of Technology, Gülbahçe Köyü, Urla, 35430 İzmir, Turkey.

### Funding

Supported by National Institutes of Health (NIH) Grants GM56838 to D.W.C. and GM30301 to D.E.C., and an NIH Structural Biology and Molecular Biophysics Training Grant to G.G.H. D.W.C. thanks the Radcliffe Institute for Advanced Study for the Elizabeth S. and Richard M. Cashin Fellowship.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC02-06CH11357. We also thank the Stanford Synchrotron Radiation Lightsource for access to beamline 4-2 for SAXS experiments. Additionally, we thank the SIBYLS beamline at the Advanced Light Source for additional SAXS measurements. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract DE-AC02-05CH11231.

## ■ ABBREVIATIONS

BME,  $\beta$ -mercaptoethanol; EIZS, *epi*-isozizaene synthase; FPP, farnesyl diphosphate; IPTG, isopropyl  $\beta$ -D-1-thiogalactopyranoside; LB, lysogeny broth; PP<sub>i</sub>, inorganic pyrophosphate; SAXS, small-angle X-ray scattering; ScGS, full-length geosmin synthase from *S. coelicolor*; ScGS<sub>690</sub>, recombinant geosmin synthase construct containing residues 1–690; ScGS<sub>338</sub>, geosmin synthase N-terminal domain containing residues 1–338 resulting from proteolysis of ScGS<sub>690</sub>; ScGS<sub>366</sub>, recombinant geosmin synthase N-terminal domain construct containing residues 1–366.

## ■ REFERENCES

- (1) Dictionary of Natural Products (<http://dnp.chemnetbase.com>).
- (2) Trapp, S. C., and Croteau, R. B. (2001) Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158, 811–832.

- (3) Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* 9, 297–304.
- (4) Gershenzon, J., and Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3, 408–414.
- (5) Farco, J. A., and Grundmann, O. (2013) Menthol—pharmacology of an important naturally medicinal “cool”. *Mini-Rev. Med. Chem.* 13, 124–131.
- (6) Gertsch, J., Leonti, M., Raduner, S., Racz, I., Chen, J.-Z., Xie, X.-Q., Altmann, K.-H., Karsak, M., and Zimmer, A. (2008) Beta-caryophyllene is a dietary cannabinoid. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9099–9104.
- (7) Cometto-Muñiz, J. E., Cain, W. S., Abraham, M. H., and Kumarsingh, R. (1998) Trigeminal and olfactory chemosensory impact of selected terpenes. *Pharmacol., Biochem. Behav.* 60, 765–770.
- (8) Schalk, M., Pastore, L., Mirata, M. A., Khim, S., Schouwey, M., Deguerry, F., Pineda, V., Rocci, L., and Daviet, L. (2012) Toward a biosynthetic route to sclareol and amber odorants. *J. Am. Chem. Soc.* 134, 18900–18903.
- (9) Renninger, N., and McPhee, D. (2008) Fuel compositions comprising farnesane and farnesane derivatives and method of making and using same. U.S. Patent 7,399,323.
- (10) Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., and Lee, T. S. (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* 2, 483.
- (11) Miller, L. H., and Su, X. (2011) Artemisinin: Discovery from the Chinese herbal garden. *Cell* 146, 855–858.
- (12) Schiff, P. B., Fant, J., and Horwitz, S. B. (1979) Promotion of microtubule assembly *in vitro* by taxol. *Nature* 277, 665–667.
- (13) Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H., and Stephanopoulos, G. (2008) Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharmaceutics* 5, 167–190.
- (14) Bohlmann, J. (2011) Terpenoid synthases — from chemical ecology and forest fires to biofuels and bioproducts. *Structure* 19, 1730–1731.
- (15) Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., and Keasling, J. D. (2012) Microbial engineering for the production of advanced terpenes. *Nature* 488, 320–328.
- (16) Smanski, M. J., Peterson, R. M., Huang, S.-X., and Shen, B. (2012) Bacterial diterpene synthases: New opportunities for mechanistic enzymology and engineered biosynthesis. *Curr. Opin. Chem. Biol.* 16, 132–141.
- (17) Poulter, C. D., and Rilling, H. C. (1978) The prenyl transfer reaction. Enzymatic and mechanistic studies of the 1'-4 coupling reaction in the terpene biosynthetic pathway. *Acc. Chem. Res.* 11, 307–313.
- (18) Poulter, C. D. (2006) Farnesyl diphosphate synthase. A paradigm for understanding structure and function relationships in E-polyprenyl diphosphate synthases. *Phytochem. Rev.* 5, 17–26.
- (19) Cane, D. E. (1985) Isoprenoid biosynthesis. Stereochemistry of the cyclization of allylic pyrophosphates. *Acc. Chem. Res.* 18, 220–226.
- (20) Cane, D. E. (1990) Enzymatic formation of sesquiterpenes. *Chem. Rev.* 90, 1089–1103.
- (21) Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* 106, 3412–3442.
- (22) Christianson, D. W. (2008) Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* 12, 141–150.
- (23) Austin, M. B., O'Maille, P. E., and Noel, J. P. (2008) Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nat. Chem. Biol.* 4, 217–222.
- (24) Gao, Y., Honzatko, R. B., and Peters, R. J. (2012) Terpenoid synthase structures: A so far incomplete view of complex catalysis. *Nat. Prod. Rep.* 29, 1153–1175.
- (25) Gerber, N. N. (1968) Geosmin, from microorganisms, is *trans*-1,10-dimethyl-*trans*-9-decalol. *Tetrahedron Lett.* 9, 2971–2974.
- (26) Bentley, R., and Meganathan, R. (1981) Geosmin and methylisoborneol biosynthesis in streptomyces. Evidence for an isoprenoid pathway and its absence in non-differentiating isolates. *FEBS Lett.* 125, 220–222.
- (27) Jiang, J., He, X., and Cane, D. E. (2006) Geosmin biosynthesis. *Streptomyces coelicolor* germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. *J. Am. Chem. Soc.* 128, 8128–8129.
- (28) Jiang, J., He, X., and Cane, D. E. (2007) Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat. Chem. Biol.* 3, 711–715.
- (29) Gerber, N. N. (1979) Volatile substances from actinomycetes: Their role in the odor pollution of water. *Crit. Rev. Microbiol.* 7, 191–214.
- (30) Buttery, R. G., and Garibaldi, J. A. (1976) Geosmin and methylisoborneol in garden soil. *J. Agric. Food Chem.* 24, 1246–1247.
- (31) Tyler, L. D., Acree, T. E., Becker, R. F., Nelson, R. R., and Butts, R. M. (1978) Effect of maturity, cultivar, field history, and the operations of peeling and coring on the geosmin content of *Beta vulgaris*. *J. Agric. Food Chem.* 26, 1466–1469.
- (32) Heil, T. P., and Lindsay, R. C. (1988) Volatile compounds in flavor-tainted fish from the Upper Wisconsin River. *J. Environ. Sci. Health, Part B* 23, 489–512.
- (33) Jardine, C. G., Gibson, N., and Hrudey, S. E. (1999) Detection of odour and health risk perception of drinking water. *Water Sci. Technol.* 40, 91–98.
- (34) Schrader, K. K., and Dennis, M. E. (2005) Cyanobacteria and earthy/musty compounds found in commercial catfish (*Ictalurus punctatus*) ponds in the Mississippi Delta and Mississippi-Alabama Blackland Prairie. *Water Res.* 39, 2807–2814.
- (35) Darriet, P., Pons, M., Lamy, S., and Dubourdieu, D. (2000) Identification and quantification of geosmin, an earthy odorant contaminating wines. *J. Agric. Food Chem.* 48, 4835–4838.
- (36) Cane, D. E., and Watt, R. M. (2003) Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1547–1551.
- (37) He, X., and Cane, D. E. (2004) Mechanism and stereochemistry of the germacradienol/germacrene D synthase of *Streptomyces coelicolor* A3(2). *J. Am. Chem. Soc.* 126, 2678–2679.
- (38) Jiang, J., and Cane, D. E. (2008) Geosmin biosynthesis. Mechanism of the fragmentation-rearrangement in the conversion of germacradienol to geosmin. *J. Am. Chem. Soc.* 130, 428–429.
- (39) Peters, R. J., Ravn, M. M., Coates, R. M., and Croteau, R. B. (2001) Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.* 123, 8974–8978.
- (40) Lesburg, C. A., Zhai, G., Cane, D. E., and Christianson, D. W. (1997) Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology. *Science* 277, 1820–1824.
- (41) Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* 277, 1815–1820.
- (42) Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469, 116–120.
- (43) Oldfield, E., and Lin, F. Y. (2012) Terpene biosynthesis: Modularity rules. *Angew. Chem., Int. Ed.* 51, 1124–1137.
- (44) Aaron, J. A., and Christianson, D. W. (2010) Trinuclear metal clusters in catalysis by terpenoid synthases. *Pure Appl. Chem.* 82, 1585–1597.
- (45) Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20, 2138–2139.
- (46) Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* 336, 1030–1033.
- (47) Diederichs, K., and Karplus, P. A. (2013) Better models by discarding data? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* D69, 1215–1222.
- (48) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology. Macromolecular Crystallography (Part A)* (Carter, C. W., Jr., and Sweet, R. M., Eds.) Vol. 276, pp 307–326, Academic Press, New York.

- (49) Terwilliger, T. C., DiMaio, F., Read, R. J., Baker, D., Bunkóczi, G., Adams, P. D., Grosse-Kunstleve, R. W., Afonine, P. V., and Echols, N. (2012) *phenix.mr\_rosetta*: molecular replacement and model rebuilding with *Phenix* and *Rosetta*. *J. Struct. Funct. Genomics* 13, 81–90.
- (50) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* D66, 486–501.
- (51) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 213–221.
- (52) Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* D66, 12–21.
- (53) Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258.
- (54) Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738.
- (55) Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
- (56) Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248.
- (57) Šali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins: Struct., Funct., Genet.* 23, 318–326.
- (58) Baer, P., Rabe, P., Fischer, K., Citron, C. A., Klapschinski, T. A., Groll, M., and Dickschat, J. S. (2014) Induced-fit mechanism in class I terpene cyclases. *Angew. Chem., Int. Ed.* 53, 7652–7656.
- (59) Wu, S., and Zhang, Y. (2007) LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 35, 3375–3382.
- (60) Aaron, J. A., Lin, X., Cane, D. E., and Christianson, D. W. (2010) Structure of Epi-Isozoaene Synthase from *Streptomyces coelicolor* A3(2), a Platform for New Terpenoid Cyclization Templates. *Biochemistry* 49, 1787–1797.
- (61) Li, R., Chou, W. K. W., Himmelberger, J. A., Litwin, K. M., Harris, G. G., Cane, D. E., and Christianson, D. W. (2014) Reprogramming the Chemodiversity of Terpenoid Cyclization by Remolding the Active Site Contour of *epi*-Isozoaene Synthase. *Biochemistry* 53, 1155–1168.
- (62) Baer, P., Rabe, P., Citron, C. A., de Oliveira Mann, C. C., Kaufmann, N., Groll, M., and Dickschat, J. S. (2014) Hedycaryol synthase in complex with nerolidol reveals terpene cyclase mechanism. *ChemBioChem* 15, 213–216.
- (63) Seemann, M., Zhai, G., de Kraker, J. W., Paschall, C. M., Christianson, D. W., and Cane, D. E. (2002) Pentalenene Synthase. Analysis of Active Site Residues by Site-Directed Mutagenesis. *J. Am. Chem. Soc.* 124, 7681–7689.
- (64) Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277–1282.
- (65) Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495–503.
- (66) Liu, H., Hexemer, A., and Zwart, P. H. (2012) The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *J. Appl. Crystallogr.* 45, 587–593.
- (67) Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* 38, W540–W544.
- (68) Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* 45, 342–350.
- (69) Tarshis, L. C., Yan, M., Poulter, C. D., and Sacchettini, J. C. (1994) Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-Å resolution. *Biochemistry* 33, 10871–10877.
- (70) Ezra, A., and Golomb, G. (2000) Administration routes and delivery systems of bisphosphonates for the treatment of bone resorption. *Adv. Drug Delivery Rev.* 42, 175–195.
- (71) Köksal, M., Chou, W. K. W., Cane, D. E., and Christianson, D. W. (2012) Structure of 2-methylisoborneol synthase from *Streptomyces coelicolor* and implications for the cyclization of a noncanonical C-methylated monoterpene substrate. *Biochemistry* 51, 3011–3020.
- (72) Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–W549.
- (73) Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2001) Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13543–13548.
- (74) Pristic, S., Xu, J., Coates, R. M., and Peters, R. J. (2007) Probing the role of the DXDD motif in class II diterpene cyclases. *ChemBioChem* 8, 869–874.
- (75) Köksal, M., Hu, H., Coates, R. M., Peters, R. J., and Christianson, D. W. (2011) Structure and mechanism of the diterpene cyclase *ent*-copalyl diphosphate synthase. *Nat. Chem. Biol.* 7, 431–433.
- (76) Wendt, K. U., Poralla, K., and Schulz, G. E. (1997) Structure and function of a squalene cyclase. *Science* 277, 1811–1815.
- (77) Thoma, R., Schulz-Gasch, T., D’Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M., and Ruf, A. (2004) Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* 432, 118–122.
- (78) Brodelius, M., Lundgren, A., Mercke, P., and Brodelius, P. E. (2002) Fusion of farnesyldiphosphate synthase and *epi*-aristolochene synthase, a sesquiterpene cyclase involved in capsidiol biosynthesis *Nicotiana tabacum*. *Eur. J. Biochem.* 269, 3570–3577.
- (79) Castellana, M., Wilson, M. Z., Xu, Y., Joshi, P., Cristea, I. M., Rabinowitz, J. D., Gitai, Z., and Wingreen, N. S. (2014) Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.* 32, 1011–1018.
- (80) Kleywegt, G. J., and Jones, T. A. (1994) Detection, Delineation, Measurement and Display of Cavities in Macromolecular Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 50, 178–185.