

**---- SUPPORTING INFORMATION ----**

**Structure of Sesquisabinene Synthase 1, a Terpenoid Cyclase that  
Generates a Strained [3.1.0] Bridged-Bicyclic Product**

Patrick N. Blank, Stephen A. Shinsky, and David W. Christianson\*

Roy and Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, 231  
South 34<sup>th</sup> Street, Philadelphia, PA 19104-6323, United States

\*author to whom correspondence should be sent: E-mail [chris@sas.upenn.edu](mailto:chris@sas.upenn.edu)

## Methods

**Expression and Purification of SQS1.** A gene encoding full-length SQS1 (residues 1-566) was synthesized by Biomatik (Wilmington, DE) using codon optimization for expression in *Escherichia coli*. The gene was cloned into the pHis parallel expression vector<sup>1</sup> using the 5' BamHI site and the 3' XhoI site through standard molecular biology techniques. DNA sequencing was used to confirm the correct insertion and orientation of the gene within the vector. PCR amplification was achieved using forward primer 5'-AAAGGATCCGATGGATCTGTGCCAGA-3' and reverse primer 5'-AAACTCGAGTTATTCTTCATCCAGGGTAATCGG-3' and product purity assessed by agarose gel. The resulting plasmid was transformed into BL21(DE3) *E. coli* for expression.

For expression, 2  $\mu$ L of plasmid DNA was added to a 50- $\mu$ L aliquot of BL21(DE3) *E. coli* stored on ice and allowed to sit for 25 min. Cells were then heat-shocked at 42 °C for 45 s, after which they were put back on ice for 2 min. Lysogeny Broth (LB) media (100  $\mu$ L) was pipetted into the mixture and incubated with shaking at 250 rpm for 1 h at 37 °C. After incubation, the mixture was applied to a pre-warmed LB agar plate and incubated overnight for 14 h at 37 °C. The plate was stored at 4 °C.

Colonies that were uniform in shape with clean, defined edges were extracted via 1- $\mu$ L pipette tip suction and deposited into 6 individual Falcon tubes, each containing 5 mL of sterile LB media and 5  $\mu$ L of 100 mg/mL Ampicillin. The 6 mixtures were grown overnight for 13 h with shaking at 250 rpm at 37 °C. The resultant overnight cultures were then used to inoculate 6 x 1.0-L flasks containing sterilized LB media and 1 mL of 100 mg/mL Ampicillin. Flasks were shaken at 250 rpm for 3 h at 37 °C until the optical density at 600 nm ( $OD_{600}$ ) reached 0.8, after which flasks were equilibrated for 45 min without shaking at 16 °C. The cell mixture in each flask was then induced with 1.0 M IPTG (500  $\mu$ L). Flasks were shaken for 20 h at 250 rpm and then

centrifuged at 11700g for 12 min at 4 °C. The resultant cell pellets were combined and stored at -80 °C.

Pellets were thawed and resuspended in Buffer A [50 mM Na<sub>2</sub>PO<sub>4</sub> (pH = 7.7), 300 mM NaCl, 10 mM β-mercaptoethanol (BME), 20% v/v glycerol] by vortexing for 30 s on-30 s off (3 times) and chilled on ice. Protease inhibitor cocktail tablets (Roche) were added at a ratio of 1 tablet per 30 mL of cell suspension after 1 round of vortexing. The crude cell resuspension was transferred to a stainless steel beaker on ice and sonicated using a Q700 (Qsonica) with amplitude set at 30% for 10 min, cycling between 1 s on and 1 s off. Crude cell lysate was clarified by centrifugation at 41000g for 1 h at 4 °C. Supernatant was collected carefully so as to exclude cell debris, and then applied to a 5mL HisTrap™ HP column (GE Healthcare). The column was washed with 10% Buffer B [25 mM Na<sub>2</sub>PO<sub>4</sub> (pH = 7.6), 300 mM NaCl, 5 mM BME, 200 mM imidazole, 20% v/v glycerol] to remove contaminants with weak Ni<sup>2+</sup> affinity. The SQS1 protein was eluted at 100% Buffer B at 4 °C, concentrated to 5 mL, filtered, and then applied to a HiLoad™ Superdex™ 26/600 200 pg column (GE Healthcare). The buffer was exchanged on column with size-exclusion chromatography buffer [50 mM Tris (pH=7.5), 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol (BME), 10% v/v glycerol], concentrated to 11.4 mg/mL, and stored at -80 °C in 600 μL aliquots. The final protein construct consisted of residues 1-566 of SQS1 plus a 6-residue N-terminal His tag that is cleavable by TEV protease.

**Crystallization and Structure Determination of unliganded SQS1.** Crystallization of unliganded SQS1 was achieved in sitting drops consisting of 1 μL protein solution [5 mg/mL SQS1, 50 mM Tris (pH=7.5), 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 5 mM BME, 2 mM benzyltriethylammonium chloride (BTAC), 2 mM sodium pyrophosphate, 10% v/v glycerol] and 2 μL of precipitant solution [6% (v/v) Tacsimate, 0.1 M MES monohydrate (pH = 6.0), 25% (w/v) polyethylene glycol 4000] and equilibrated with a reservoir of 300 μL precipitant solution at 21

°C. Crystals appeared in 14 days and were harvested and cryoprotected in 25% (v/v) glycerol prior to flash-cooling in liquid nitrogen.

Crystals diffracted X-rays to 1.9 Å resolution at the Advanced Photon Source (APS), Northeastern Collaborative Access Team (NE-CAT) beamline 24-ID-E. Crystals belonged to space group  $I2_12_12_1$  with the following unit cell parameters:  $a = 63.42$  Å,  $b = 85.84$  Å,  $c = 220.31$  Å (one molecule in the asymmetric unit). The X-ray diffraction dataset was indexed and integrated using iMOSFLM.<sup>2</sup> Scaling was performed using AIMLESS.<sup>3</sup> Phasing was initiated using the atomic coordinates of (+)-limonene synthase from *Citrus sinensis* (PDB code: 5UV0)<sup>4</sup> as a search model for molecular replacement calculations with PHASER.<sup>5</sup> The model of SQS1 was built, refined, and validated using COOT, PHENIX, and MOLPROBITY, respectively.<sup>6-8</sup> Data collection and refinement statistics are recorded in Supplementary Table 1.

**Crystallization and Structure Determination of the SQS1-Mg<sup>2+</sup><sub>3</sub>-ibandronate complex.** Crystallization of SQS1 in the closed active site conformation, complexed with Mg<sup>2+</sup><sub>3</sub> and ibandronate, was achieved in sitting drops consisting of 600 nL protein solution [5 mg/mL SQS1, 50 mM Tris (pH=7.5), 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 5 mM BME, 2.0 mM ibandronate chloride, 10% v/v glycerol] and 600 nL of precipitant solution [0.2 M KCl, 20% w/v polyethylene glycol 3350 and equilibrated with a reservoir of 100 mL precipitant solution at 21 °C. Crystals appeared in 20 days and were harvested and cryoprotected in 25% (v/v) ethylene glycol prior to flash-cooling in liquid nitrogen.

Crystals diffracted X-rays to 2.1 Å resolution at the Stanford Synchrotron Radiation Lightsource (SSRL), beamline 9-2. Crystals belonged to space group C2 with unit cell parameters  $a = 181.55$  Å,  $b = 58.69$  Å,  $c = 53.59$  Å,  $\beta = 93.10^\circ$  (one molecule in the asymmetric unit). The X-ray diffraction dataset was indexed and integrated using iMOSFLM,<sup>2</sup> and scaling was performed using AIMLESS.<sup>3</sup> Phasing was achieved using the atomic coordinates of unliganded SQS1 as a search model in PHASER.<sup>5</sup> The model of the enzyme-inhibitor complex

was built, refined, and validated using COOT, PHENIX, and MOLPROBITY respectively.<sup>6-8</sup> Data collection and refinement statistics are recorded in Supplementary Table 1. All figures were generated with Pymol.<sup>9</sup>

**Molecular modeling.** The small molecules *epi*-isozizaene and sesquisabinene were drawn in a structure editor and converted to a SMILES format string. This string was input into the SMILES Translator and Structure File Generator (National Cancer Institute, <https://cactus.nci.nih.gov/translate/>), which output the file in .pdb format. The resultant .pdb file was double checked and optimized for proper spatial geometry using Phenix,<sup>7</sup> before being opened in Pymol<sup>9</sup> along with the atomic coordinates of EIZS and SQS1. To visualize the active site contour, a mesh was first generated using the DoGSiteScorer (<http://dogsite.zbh.uni-hamburg.de>). This program detects all pockets within the protein using a Difference of Gaussian method to generate a drug score for each pocket.<sup>10,11</sup> The resulting mesh pockets with the highest scores corresponded to the active sites of each enzyme. Each mesh was imported into Pymol and used to guide the generation of the active site contour with the surface cavity function in Pymol (probe radius = 1.6 Å). The surface cavity function was complemented by the NRGsuite subroutine implemented in Pymol,<sup>12</sup> which was used to delineate the surface adjacent to the Mg<sup>2+</sup><sub>3</sub>-bound pyrophosphate anion in the active site of EIZS and the Mg<sup>2+</sup><sub>3</sub>-bound bisphosphonate pyrophosphate mimic in the active site of SQS1. The enclosed volumes of the resultant active site contours were then measured in Å<sup>3</sup>. Conformations of *epi*-isozizaene, sesquisabinene, and the homobisabolyl cation were selected using the FlexAid function in NRGsuite and manually docked into active site clefts.

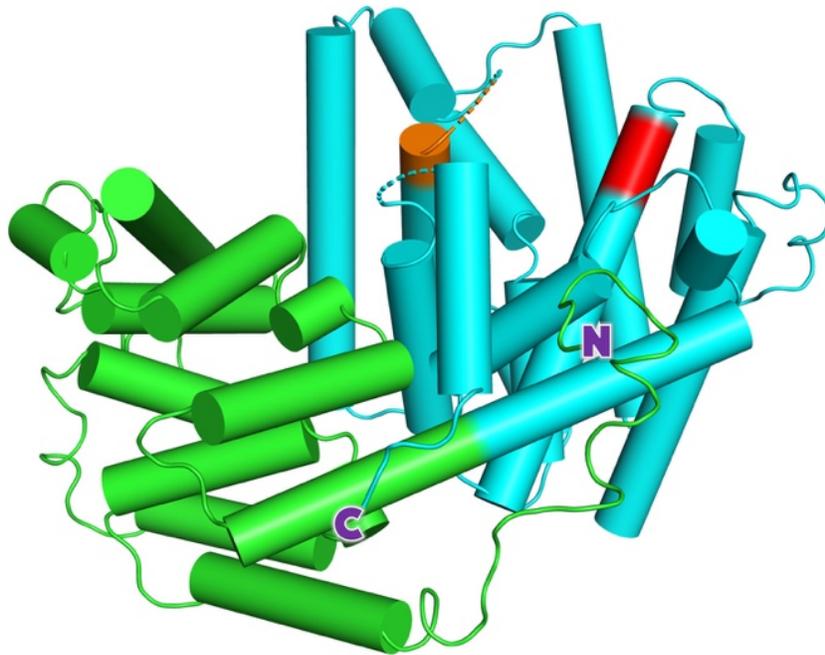
Models of F96M, F96Q, F96S, and F96N EIZS mutants were generated in Coot<sup>6</sup> using the simple mutate function with the closed conformation of wild-type EIZS as a starting point (PDB 3KB9).<sup>13</sup> Chemically reasonable side chain rotamers were chosen for each mutant that appeared to accommodate stabilizing interactions. The side chain rotamer of M96 was chosen

to optimize interactions with the spatially adjacent side chain of Y69, in view of the strong geometric preference for S- $\pi$  interactions at approximately 5 Å and 45° away from the ring normal.<sup>14</sup> The side chain rotamer of Q96 was chosen to optimize a hydrogen bond interaction with R338. The side chain rotamer of S96 was chosen to optimize hydrogen bond interactions with R338, Y69, and water 452. Furthermore, this S96 rotamer is consistent with one of the S96 rotamers observed in the crystal structure of unliganded F96S EIZS (PDB 6AXO). Finally, the side chain rotamer of N96 was chosen to optimize hydrogen bonds with R338 and Y69.

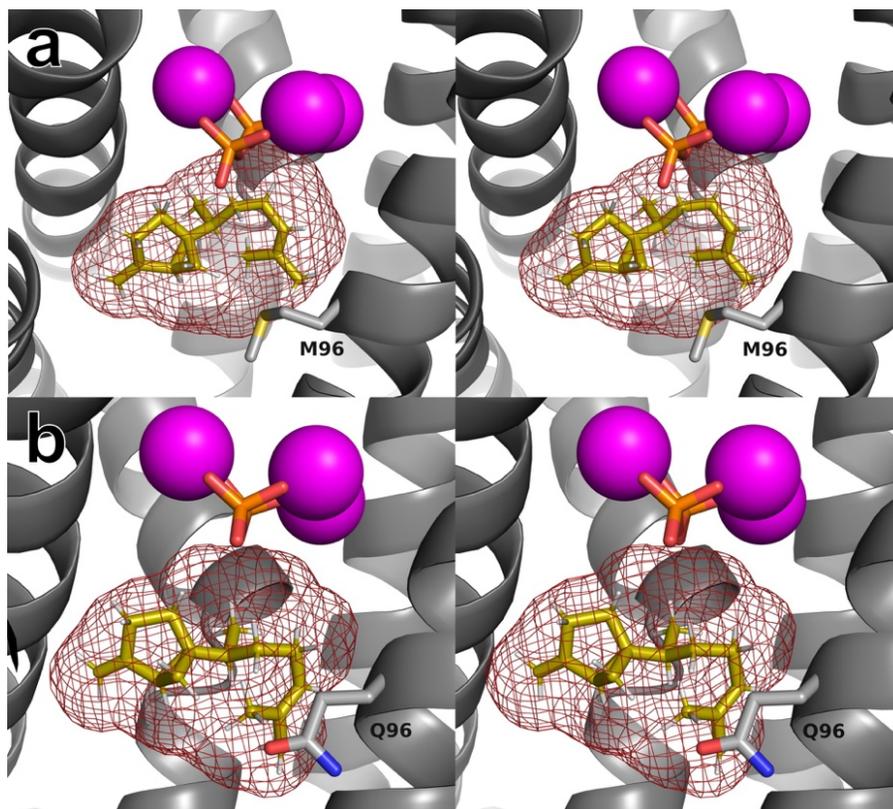
**Supplementary Table 1. Data Collection and Refinement Statistics**

<b>SQS1</b>	<b>Liganded</b>	<b>Unliganded</b>
Synchrotron	SSRL	APS
Beamline	9-2	24-ID-E
Detector	PILATUS 6M	EIGER 16M
Wavelength (Å)	0.979	0.979
Space group	C2	I2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
a, b, c (Å)	181.6, 58.7, 53.6	63.4, 85.8, 220.3
α, β, γ (°)	90.0, 93.1, 90.0	90.0, 90.0, 90.0
Resolution (Å) <sup>a</sup>	55.84-2.10 (2.16-2.10)	60.95-1.90 (1.94-1.90)
Total/unique no. of reflections	102065/31346 (7447/2469)	523533/47868 (33854/3065)
R <sub>merge</sub> <sup>a,b</sup>	0.087 (0.770)	0.075 (1.460)
R <sub>p.i.m.</sub> <sup>a,c</sup>	0.057 (0.524)	0.023 (0.457)
CC <sub>1/2</sub> <sup>a,d</sup>	0.997 (0.715)	0.999 (0.580)
I/σ(I) <sup>a</sup>	6.2 (1.3)	15.2 (1.8)
Redundancy <sup>a</sup>	3.3 (3.0)	10.9 (11.0)
Completeness (%) <sup>a</sup>	95.1 (91.8)	100.0 (100.0)
R <sub>work</sub> <sup>a,e</sup>	0.221 (0.352)	0.192 (0.321)
R <sub>free</sub> <sup>a,e</sup>	0.246 (0.404)	0.232 (0.376)
No. of non-hydrogen atoms:		
protein	4044	4125
water	190	295
other	46	20
<b>Average B-factors (Å<sup>2</sup>)</b>		
macromolecules	43	43
water	41	47
other	44	52
<b>Rms deviations from ideal geometry</b>		
bonds (Å)	0.003	0.005
angles (°)	0.6	1.2
<b>Ramachandran plot (%)<sup>f</sup></b>		
favored	97.20	96.65
allowed	2.80	3.15
outliers	0.00	0.20
Rotamer outliers (%)	0.24	0.23
Clashscore	5.75	4.16
Cβ outliers (%)	0	0
<b>PDB accession code</b>	6O9P	6O9Q

<sup>a</sup>Values in parentheses correspond to data in highest resolution shell. <sup>b</sup>R<sub>merge</sub> =  $\sum_{hkl} \sum_i |I_{i,hkl} - \langle I \rangle_{hkl}| / \sum_{hkl} \sum_i I_{i,hkl}$ ;  $\langle I \rangle_{hkl}$  = mean intensity of  $I_{hkl}$  calculated from replicate measurements. Note that R<sub>merge</sub> can sometimes be inordinately high for highly redundant data sets. <sup>c</sup>R<sub>p.i.m.</sub> =  $(\sum_{hkl} (1/(N-1))^{1/2} \sum_i |I_{i,hkl} - \langle I \rangle_{hkl}|) / \sum_{hkl} \sum_i I_{i,hkl}$ ;  $\langle I \rangle_{hkl}$  = mean intensity of  $I_{hkl}$  calculated from replicate measurements, N = number of reflections. <sup>d</sup>Pearson correlation coefficient between random half-datasets. <sup>e</sup>R<sub>work</sub> =  $\sum ||F_o| - |F_c|| / \sum |F_o|$  for reflections contained in the working set.  $|F_o|$  = observed structure factor amplitude,  $|F_c|$  = calculated structure factor amplitude. R<sub>free</sub> is calculated in the same manner using test set reflections held aside during refinement. <sup>f</sup>Calculated with MolProbity.



**Supplementary Figure 1.** Structure of unliganded SQS1 color coded as follows:  $\alpha$  domain, cyan;  $\beta$  domain, light green; DDXXD metal-binding motif, red; NSE/DTE metal-binding motif, orange. N- and C-termini are indicated; disordered polypeptide segments are indicated by dotted lines.



**Supplementary Figure 2.** (a) Stereoview of the enclosed active site contour in a model of the F96M *epi-isizizaene* synthase-Mg<sup>2+</sup><sub>3</sub>-PP<sub>i</sub> complex based on the structure of the *epi-isozizaene* synthase-Mg<sup>2+</sup><sub>3</sub>-PP<sub>i</sub>-BTAC complex, omitting the atomic coordinates of BTAC and instead fit with a model of product sesquisabinene. (b) Stereoview of the enclosed active site contour in a model of the F96Q *epi-isizizaene* synthase-Mg<sup>2+</sup><sub>3</sub>-PP<sub>i</sub> complex based on the structure of the *epi-isozizaene* synthase-Mg<sup>2+</sup><sub>3</sub>-PP<sub>i</sub>-BTAC complex, omitting the atomic coordinates of BTAC and instead fit with a model of product sesquisabinene.

## References

1. Sheffield, P., Garrard, S., and Derewenda, Z. (1999) Overcoming expression and purification problems of RhoGDI using a family of “parallel” expression vectors. *Protein Expr. Purif.* 15, 34-39.
2. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM *Acta Cryst. D67*, 271– 281.
3. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Cryst. D69*, 1204-1214.
4. Morehouse, B. R., Kumar, R. P., Matos, J. O., Olsen, S. N., Entova, S., Oprian, D. D. (2017) Functional and structural characterization of a (+)-limonene synthase from *Citrus sinensis*. *Biochemistry* 56, 1706-1715.
5. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software *J. Appl. Cryst.* 40, 658-674.
6. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Cryst. D66*, 486-501.
7. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L., 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 Cryst. D66*, 213-221.
8. 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 Cryst. D66*, 12-21.

9. The PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC.
10. Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., and Rarey, M. (2012) Combining global and local measures for structure-based druggability predictions. *J. Chem. Inf. Model.* 52, 360-372.
11. Volkamer, A., Griewel, A., Grombacher, T., and Rarey, M. (2010) Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J. Chem. Inf. Model.* 50, 2041-2052.
12. Gaudreault F., Morency L.-P., and Najmanovich, R. J. (2015) NRGsuite: a PyMOL plugin to perform docking simulations in real time using FlexAID. *Bioinformatics* 31, 3856–3858.
13. Aaron, J. A., Lin, X., Cane, D. E., and Christianson, D. W. (2010) Structure of *epi*-isozizaene synthase from *Streptomyces coelicolor* A3(2), a platform for new terpenoid cyclization templates. *Biochemistry* 49, 1787-1797.
14. Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J., and Sachs, J. N. (2012) The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem.* 287, 34979-34991.