

## Structure-Guided Functional Characterization of Enediye Self-Sacrifice Resistance Proteins, CalU16 and CalU19

Sherif I. Elshahawi,<sup>†,‡</sup> Theresa A. Ramelot,<sup>§</sup> Jayaraman Seetharaman,<sup>||</sup> Jing Chen,<sup>⊥</sup> Shanteri Singh,<sup>†,‡</sup> Yunhuang Yang,<sup>§</sup> Kari Pederson,<sup>#</sup> Madan K. Kharel,<sup>†,‡,□</sup> Rong Xiao,<sup>▽</sup> Scott Lew,<sup>||</sup> Ragothaman M. Yennamalli,<sup>○</sup> Mitchell D. Miller,<sup>○</sup> Fengbin Wang,<sup>○</sup> Liang Tong,<sup>||</sup> Gaetano T. Montelione,<sup>▽,◆</sup> Michael A. Kennedy,<sup>§</sup> Craig A. Bingman,<sup>¶</sup> Haining Zhu,<sup>⊥</sup> George N. Phillips, Jr.,<sup>\*,○</sup> and Jon S. Thorson<sup>\*,†,‡</sup>

<sup>†</sup>Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, Kentucky 40536, United States

<sup>‡</sup>Center for Pharmaceutical Research and Innovation (CPRI), College of Pharmacy, University of Kentucky, Lexington, Kentucky 40536, United States

<sup>§</sup>Department of Chemistry and Biochemistry, Northeast Structural Genomics Consortium, Miami University, Oxford, Ohio 45056, United States

<sup>||</sup>Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, New York 10027, United States

<sup>⊥</sup>Department of Molecular and Cellular Biochemistry & Center for Structural Biology, College of Medicine, University of Kentucky, Lexington, Kentucky 40536, United States

<sup>#</sup>Complex Carbohydrate Research Center, Northeast Structural Genomics Consortium, University of Georgia, Athens, Georgia 30602, United States

<sup>▽</sup>Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States

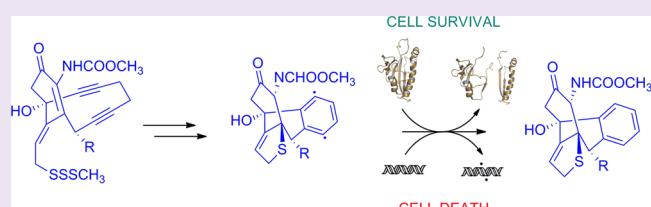
<sup>○</sup>Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005, United States

<sup>◆</sup>Department of Biochemistry, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States

<sup>¶</sup>Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

### Supporting Information

**ABSTRACT:** Calicheamicin  $\gamma_1$ <sup>1</sup> (**1**) is an enediye antitumor compound produced by *Micromonospora echinospora* spp. calicensis, and its biosynthetic gene cluster has been previously reported. Despite extensive analysis and biochemical study, several genes in the biosynthetic gene cluster of **1** remain functionally unassigned. Using a structural genomics approach and biochemical characterization, two proteins encoded by genes from the **1** biosynthetic gene cluster assigned as “unknowns”, CalU16 and CalU19, were characterized. Structure analysis revealed that they possess the STeroidogenic Acute Regulatory protein related lipid Transfer (START) domain known mainly to bind and transport lipids and previously identified as the structural signature of the enediye self-resistance protein CalC. Subsequent study revealed *calU16* and *calU19* to confer resistance to **1**, and reminiscent of the prototype CalC, both CalU16 and CalU19 were cleaved by **1** *in vitro*. Through site-directed mutagenesis and mass spectrometry, we identified the site of cleavage in each protein and characterized their function in conferring resistance against **1**. This report emphasizes the importance of structural genomics as a powerful tool for the functional annotation of unknown proteins.



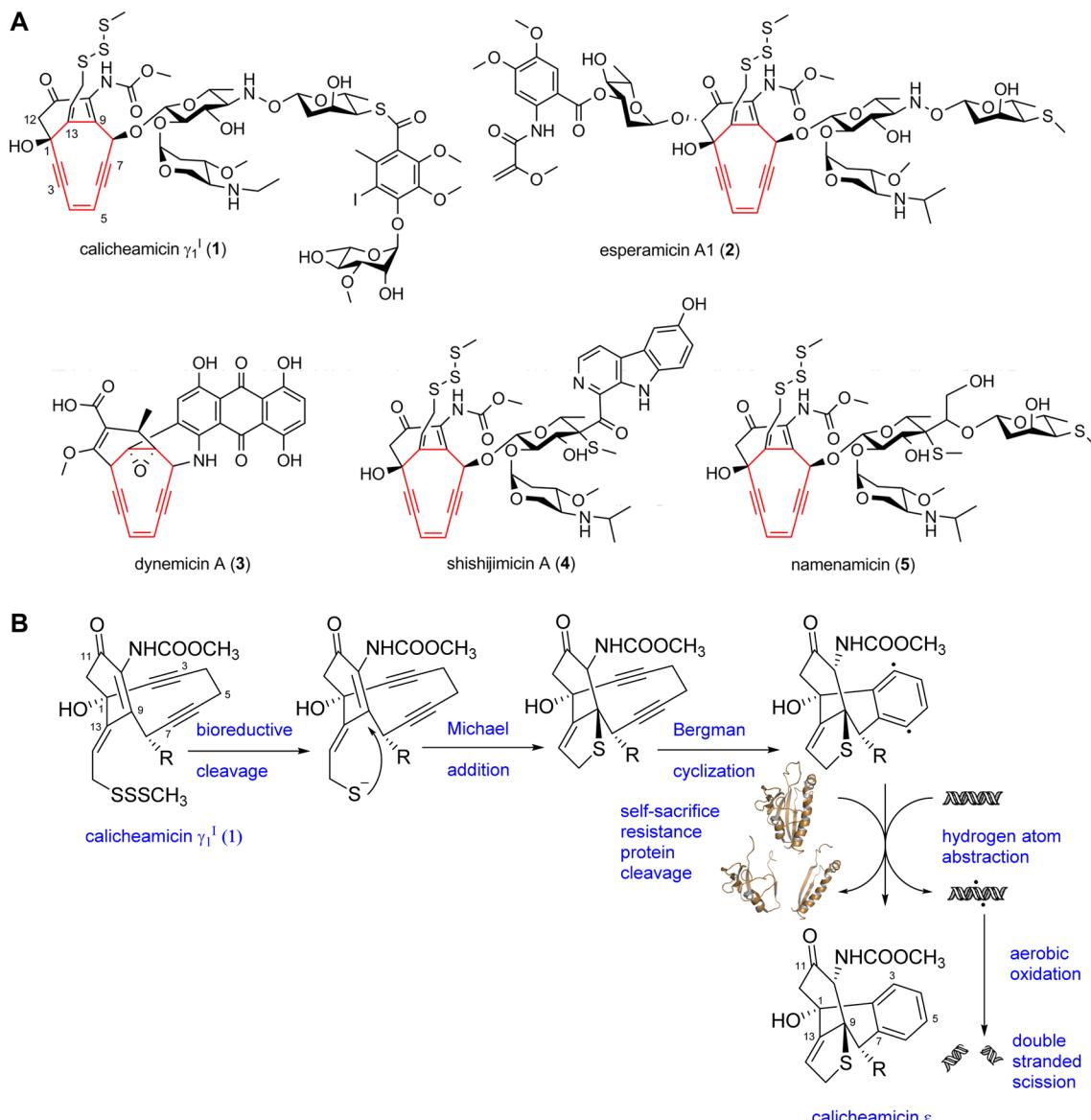
The calicheamicins are a prototype of the naturally occurring 10-membered enediye antitumor antibiotic family and were first reported in 1989 as metabolites of *Micromonospora echinospora* spp. calicensis.<sup>1–3</sup> Members of this family share a structurally conserved bicyclo[7.3.1]enediye core, also often referred to as the “warhead” as this structural unit is central to the fundamental enediye mechanism of action (Figure 1A).<sup>4–6</sup>

In all family members, the enediye core is strategically decorated with a bioorthogonal “triggering” system and specific appendages that enhance affinity to the metabolite’s target

**Received:** April 30, 2014

**Accepted:** July 31, 2014

**Published:** July 31, 2014



**Figure 1.** (A) Selected structures of naturally occurring 10-membered enediynes. The “warhead” is highlighted in red. (B) Proposed mechanism of cycloaromatization of **1** and its effect on DNA scission and CalC self-sacrifice mechanism.

(DNA/RNA). The calicheamicin trigger system (and that of esperamicins, shishijimicins, and namenamicins) is comprised of a unique trisulfide that, upon reduction by bioreductants such as glutathione, induces an intramolecular hetero-Michael addition at C-9 (Figure 1B).<sup>7–9</sup> The resulting fused enediyne core ring strain invokes a spontaneous cycloaromatization reaction which proceeds via a highly reactive diradical intermediate that is rapidly quenched by any suitable hydrogen source.<sup>7,10</sup> Calicheamicin’s high affinity for the minor groove of DNA, by virtue of its aryltetrasaccharide, insures that the diradical species is quenched via hydrogen abstraction from the backbone of opposing strands of dsDNA to form DNA radicals that, in the presence of oxygen, result in facile double-strand scission.<sup>11–13</sup> The study of calicheamicin’s many fascinating architectural and functional facets have led to numerous discoveries/advances in chemistry,<sup>14,15</sup> enzymology,<sup>16–26</sup> and anticancer drug development<sup>27,28</sup> over the past three decades. Yet, while the gene cluster encoding for calicheamicin biosynthesis was cloned from *M. echinospora* and sequenced nearly a decade ago,<sup>16</sup> there remain a number of genes ( $\sim 30\%$ )

within this locus annotated as “unknowns” due to a lack of homologues and/or biochemical characterization of corresponding gene products.

Herein, we describe the application of structural genomics as a basis for the functional characterization of two proteins encoded by such “unknowns”—CalU16 and CalU19. Specifically, structure elucidation (via both NMR and X-ray crystallography) revealed CalU16 to be a structural homologue of CalC, a protein previously characterized as among the first reported enediyne “self-sacrifice” resistance proteins.<sup>18,19</sup> Prompted by this structure-based revelation, subsequent biochemical characterization of CalU16 and its homologue CalU19 revealed both to serve in a similar capacity wherein CalU19 also displayed the unprecedented ability to trigger enediyne cycloaromatization in the absence of endogenous reducing agents. Cumulatively, this work extends the body of work focused upon understanding how bacteria construct and control highly reactive and lethal metabolites and expands the number of known “self-sacrifice” enediyne resistance proteins. This work also serves to illustrate the impact of structure

determination as a critical tool for the functional annotation of unassigned genes.

## RESULTS AND DISCUSSION

### CalU16 Structure and CalU19 Homology Model.

Protein Structure Initiative (PSI-Biology) employs a complementary combination of NMR spectroscopy and/or X-ray crystallography to determine 3D-structures of biologically relevant targets. The structure of CalU16 was solved both by X-ray crystallography (Protein Database Bank (PDB: 4FPW)) and NMR (PDB: 2LUZ). The crystal structure of CalU16 was determined at 2.50 Å resolution, refined to an  $R_{\text{cryst}}$  and  $R_{\text{free}}$  of 24.1% and 28.2% (Table 1), respectively, and is comprised of two subunits in an asymmetric unit belonging to space group

**Table 1. Summary of Crystal Parameters, Data Collection, and Refinement Statistics<sup>a</sup>**

property	value
Crystal Param.	
space group	$P6_1$
molecules per asymmetric unit	2
$V_M$ (Å <sup>3</sup> Da <sup>-1</sup> )	2.95
unit-cell param. (Å)	$a = b = 51.85, c = 305.70$
Data Collection Statistics	
wavelength (Å)	0.9790
resolution range (Å)	50.00–2.50 (2.59–2.50)
no. of reflections (measured/unique)	236417/16051
completeness (%)	99.9 (99.5)
$R_{\text{merge}}^b$	0.056 (0.169)
redundancy	14.8 (12.3)
mean $I/\sigma$ ( $I$ )	27.4 (10.4)
Refinement and Model Statistics	
resolution range (Å)	44.91–2.50
no. of reflections (work/test)	15160/794
$R_{\text{cryst}}^c$	0.241 (0.325)
$R_{\text{free}}^d$	0.282 (0.371)
twin operator	K, H, -L
twin fraction	0.50
RMSD bonds (Å)	0.009
RMSD angles (deg)	1.453
B factor (protein/solvent) (Å <sup>2</sup> )	53.8/50.9
no. of protein atoms	2497
no. of waters	128
Ramachandran Plot (%) <sup>e</sup>	
most favorable regions	91.3
allowed regions	5.8
disallowed regions	2.9
Global Quality Scores (Raw/Z-Score)	
Verify3D	0.4/−1.0
ProsaII	0.5/−0.6
Procheck G factor (phi-psi)	−0.6/−2.0
Procheck G factor (all)	−0.5/−2.7
Molprobity clash score	9.6/−1.8
PDB ID	4FPW

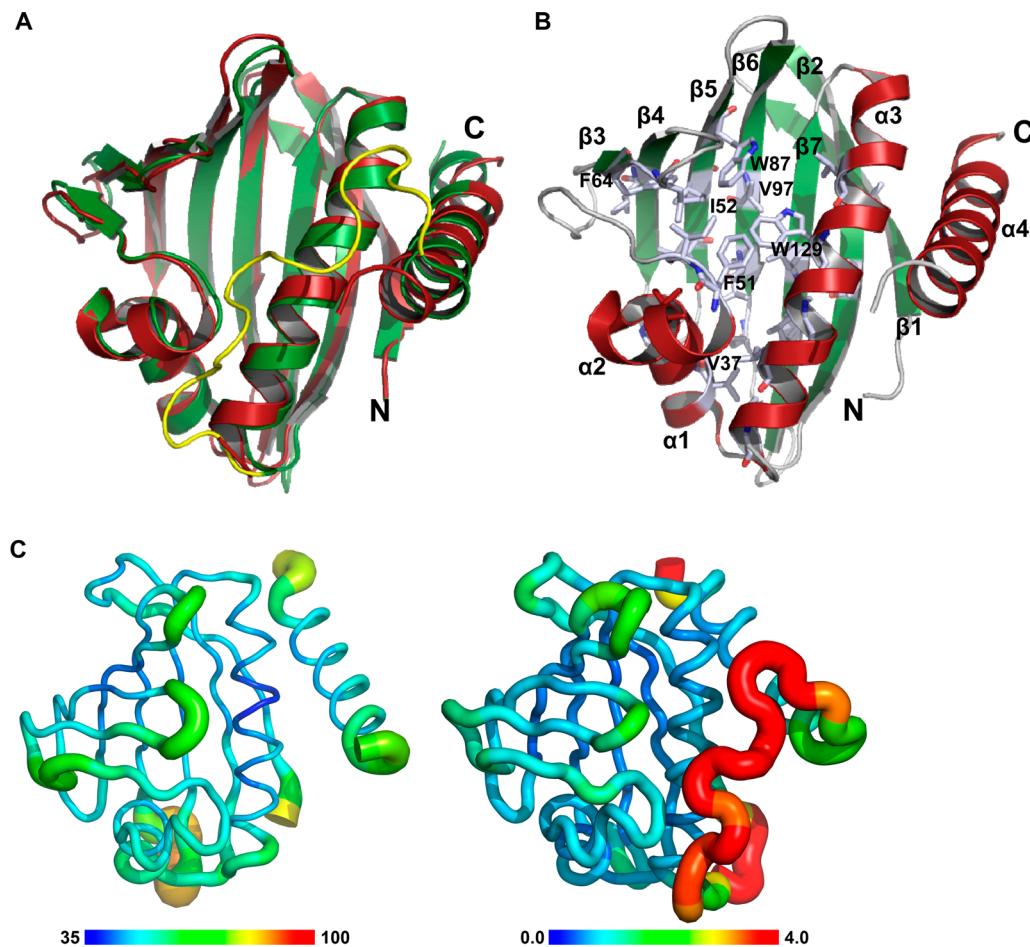
<sup>a</sup>Values in parentheses are for the highest resolution shell. <sup>b</sup> $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$ , where  $I_i(h)$  is the intensity of an individual measurement of the reflection and  $\langle I(h) \rangle$  is the mean intensity of the reflection. <sup>c</sup> $R_{\text{cryst}} = \sum_h \|F_{\text{obs}}\| - \|F_{\text{calc}}\| / \sum_h \|F_{\text{obs}}\|$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure-factor amplitudes, respectively. <sup>d</sup> $R_{\text{free}}$  was calculated as  $R_{\text{cryst}}$  except that it uses 10% of the reflection data omitted from refinement. <sup>e</sup>Calculated using Molprobity.<sup>71</sup>

$P6_1$  (Figure 2). Of the 182 residues of CalU16, the first 4 residues, the last 2 residues, and a dynamic loop containing 19 residues (a.a. 141–159) from subunit A and 15 residues (a.a. 144–158) from subunit B were not modeled due to insufficient electron density. Structure and quality statistics are given in Table 1. The corresponding solution structure of CalU16 was determined by multidimensional heteronuclear NMR spectroscopy (Supporting Information Figures S1–S3). Nearly complete <sup>1</sup>H, <sup>15</sup>N, and <sup>13</sup>C assignments were obtained and structures were calculated from residual dipolar couplings (RDCs), chemical shift-derived dihedral angle constraints, and Nuclear Overhauser Effect (NOE)-derived distance constraints via standard protocols. Structure and quality statistics are given in Table 2 for the ensemble of 20 lowest energy structures, indicating the high quality of the structure. Intrinsic disorder was observed for the first 6 residues, the last 3 residues, and 21 residues that includes a loop and part of the following helix (a.a. 143–163) and is corroborated by a lack of long-range NOEs and low (<0.6) <sup>15</sup>N{<sup>1</sup>H} heteronuclear values (Supporting Information Figure S3).

Superposition of the CalU16 solution structure with individual subunits of the crystal structure reveals a rmsd of 0.88 Å, indicating that the structures are quite similar. The major difference between the two structures is the absence of the previously mentioned dynamic loop region (residues 141–159) in the crystal structure (Figure 2). Interestingly, the crystal structure of CalU16 contains two subunits in the asymmetric unit related by 2-fold noncrystallographic symmetry, with a small contact interface (calculated to be ~530 Å<sup>2</sup> by PISA).<sup>29</sup> Based on analytical static light scattering in-line (Supporting Information Figure S4) with gel filtration chromatography and NMR correlation time estimates, CalU16 exists as a monomer in solution. The final difference map also reveals an additional long tubular feature in hydrophobic pocket, perhaps representing a bound unknown *E. coli* metabolite carried through the purification procedure (Supporting Information Figure S5).

The CalU16 monomer folds into a globular domain formed by four  $\alpha$ -helices and seven antiparallel  $\beta$ -strands in the order:  $\beta1-\beta2-\alpha1-\alpha2-\beta3-\beta4-\beta5-\beta6-\beta7-\alpha3-\text{loop}-\alpha4$  (Figure 2). The domain exhibits a common tertiary structure consisting of a helix-grip-fold, characteristic of STeroidogenic Acute Regulatory protein related lipid Transfer (START) domains implicated in possessing a hydrophobic cavity for ligand binding.<sup>30,31</sup> The hydrophobic cavity in CalU16 spans the entire length of CalU16, formed by seven  $\beta$ -strands and three helices,  $\alpha1$ ,  $\alpha2$ , and  $\alpha3$ , encompassing residues Ile24, Ile26, Val37, Trp38, Ile47, Trp50, Phe51, Ile52, Phe64, Leu66, Leu83, Ile74, Ile85, Trp87, Val97, Leu99, Leu101, Leu110, Leu112, Val125, Trp129, Leu133, Leu136, and Ile140 (Figure 2). In addition to the START domain core, the fourth helix ( $\alpha4$ ) of CalU16 contributes an additional structural unit. Such extra secondary structural elements are common upon START domain members where they contribute to a range of structural/functional roles.<sup>19,32</sup>

CalU19 is another protein of unknown function encoded by a gene in the calicheamicin gene cluster, which displays 42% sequence identity to CalU16 (Supporting Information Figure S6). A CalU16-based homology model of CalU19, generated using SWISS-MODEL also highlights the presence of a signature hydrophobic core consistent with START domain family members, implicating a similar function (Supporting Information Figure S7).



**Figure 2.** Structure of CalU16. (A) Overlay of NMR (green) and monomers of the crystal (brick-red) structures of CalU16. The N- and C-termini of the protein are labeled, while the dynamic loop is colored yellow. (B) Monomer with secondary structural elements labeled. The residues in the hydrophobic cavity are represented as stick models. (C) B-factors for the C $\alpha$  atoms in the crystal structure (left) and C $\alpha$  RMSD values from the NMR ensemble (right) are mapped to the color and tube diameter of “putty” traces showing the general agreement (correlation coefficient 0.559) between the structures.

**Structurally-Related Proteins.** A structure-based similarity search of CalU16 using DaliLite server<sup>33</sup> and PDBeFold<sup>34</sup> returned >100 hits that share close structural similarity. Figure 3 and Supporting Information Table S1 summarize a small selection of functionally characterized structural homologues of CalU16 (Figure 3A). Among these is CalC (PDB: 2L65) (Figure 3B), the prototype calicheamicin self-sacrifice resistance protein encoded by the gene *calC* from the same biosynthetic gene cluster as CalU16.<sup>19</sup> In addition, three structures (Figure 3C–E) associated with other natural product biosynthetic pathways were identified. Specifically, TcmN Aro/Cyc (PDB: 3TVQ) is a polyketide aromatase/cyclase (Figure 3C) involved in the regiospecific cyclization of tetracenomycin,<sup>35</sup> Hyp-1 (PDB: 3IE5) is involved in the formation of hypericin (Figure 3D),<sup>36</sup> while NCS (PDB: 2VQ5) is involved in the biosynthesis of (*S*)-noroclaurine (Figure 3E).<sup>37</sup> Other representative structural homologues identified include plant allergen Bet v1-J (PDB: 4A8U),<sup>38</sup> abscisic acid receptor (PDB: 4N0G),<sup>39</sup> cytokinin-specific binding protein (PDB: 2FLH),<sup>40</sup> and major celery allergen protein (PDB: 2BK0) (Supporting Information Table S1).<sup>41</sup> A structural alignment of CalU16 with these homologues highlights the striking conservation of the core structural fold, which contributes to the binding site for a structurally diverse array of ligands including steroids, lipids, hormones, and polyketides. However, a clear structural

difference in CalU16 is the presence of an additional helical region near the residues 160–180 ( $\alpha$ 4 helix as shown in Figure 2B). While this extra helix is not observed in the other homologues of known function, two functionally uncharacterized CalU16 structural homologues also have this helix (PDB: 2NN5 and 2K5G).<sup>42,43</sup> Despite the plethora of available sequences and structures,<sup>31,44</sup> START domain structure/sequence alone remains insufficient to assign function for newly discovered family members. Thus, additional structures and parallel biochemical characterization are needed to establish broader structure–activity relationships of START domain proteins.

**Heterologous Expression of *calU16* or *calU19* in *E. coli* Confers Resistance to Calicheamicin.** Our initial cloning of the 1 biosynthetic gene cluster was facilitated by employing a selection for 1 resistance using a *M. echinospora* cosmid library in *E. coli*.<sup>16</sup> From positive cosmids identified, an iterative subcloning and selection process ultimately led to the discovery of *calC* gene and subsequent elucidation of the CalC mechanism.<sup>18</sup> It is important to note that the initial screen for resistance genes relied upon the native *Micromonospora* promoters for heterologous expression in *E. coli*, and thus, *calU16* and *calU19* could have been missed simply due to poor heterologous expression in *E. coli*. However, the ability of *calC* (and presumably other self-sacrifice resistance protein encoding

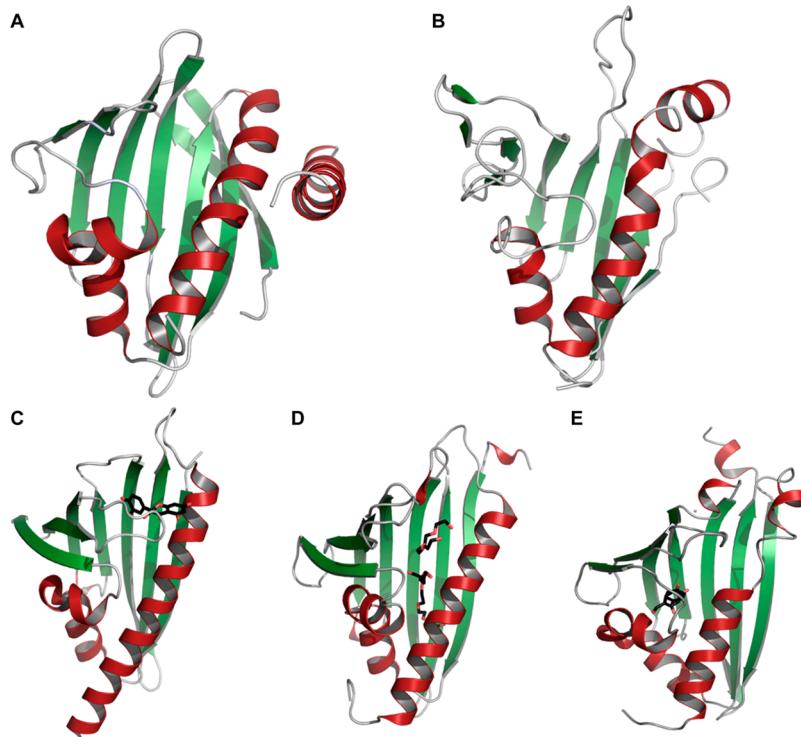
**Table 2.** Summary of CalU16 Solution Structural Statistics<sup>a</sup>

property	value
Completeness of Resonance Assignments <sup>b</sup>	
backbone (%)	98.3
side chain (%)	97.9
aromatic (%)	96.2
stereospecific methyl (%)	85.8
Conformationally Restricting Constraints <sup>c</sup>	
distance constraints	
total	2522
intraresidue ( $i = j$ )	471
sequential ( $ I - j  = 1$ )	614
medium range ( $1 <  i - j  < 5$ )	458
long-range ( $ i - j  \geq 5$ )	979
dihedral angle constraints	222
hydrogen bond constraints	192
NH RDC constraints (gel NH/NC')	75/59
no. of constraints per residue	16.5
no. of long-range constraints per residue	6.0
Residual Constraint Violations <sup>c</sup>	
avg. no. of distance violations per structure	
0.1–0.2 Å	24.8
0.2–0.5 Å	6.2
> 0.5 Å	0
avg no. of dihedral angle violations per structure	
1–10°	17.8
> 10°	0.2
RDC $Q_{\text{rmsd}}$ (gel NH/NC') <sup>d</sup>	0.4/0.3
Model Quality <sup>c</sup>	
RMSD backbone atoms (Å) <sup>e</sup>	0.7
RMSD heavy atoms (Å) <sup>e</sup>	1.2
RMSD bond lengths (Å)	0.018
RMSD bond angles (deg)	1.3
MolProbity Ramachandran statistics <sup>c,e</sup>	
most favored regions (%)	94.6
allowed regions (%)	5.4
disallowed regions (%)	0
global quality scores (Raw/Z-score) <sup>c</sup>	
Verify3D	0.4/–0.3
ProsaII	0.5/–0.6
Procheck G factor ( $\varphi$ – $\psi$ ) <sup>e</sup>	–0.4/ –1.3
Procheck G factor (all) <sup>e</sup>	–0.3/–2.0
MolProbity clash score	16.0/–1.2
RPF scores <sup>f</sup>	
recall/precision	0.98
F-measure/DP-score	0.94
Model Contents	
ordered residue range <sup>e</sup>	7–65, 71–89, 95–140, 161–179
BMRB accession no.	18547
PDB ID	2LUZ

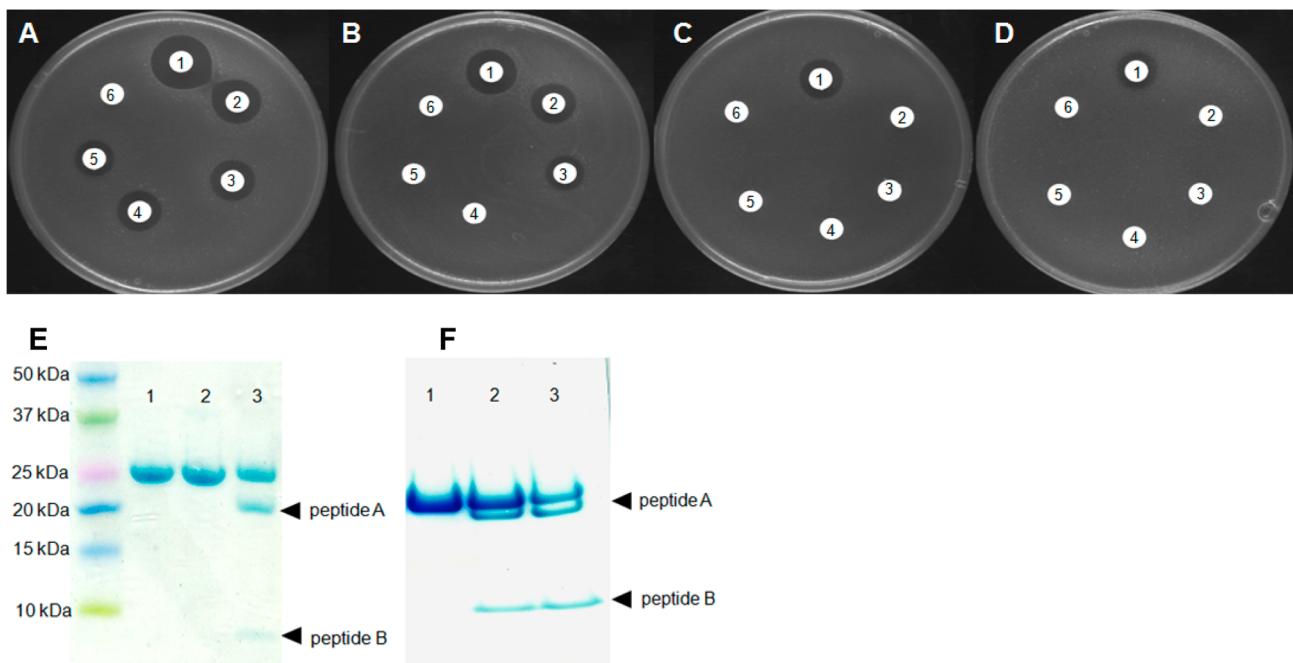
<sup>a</sup>Structural statistics computed for the ensemble of 20 deposited structures. <sup>b</sup>Computed using AVS software from the expected number of resonances, excluding highly exchangeable protons (N-terminal, Lys, and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, nonprotonated aromatic carbons, for residues 1–182. <sup>c</sup>Calculated using PSVS 1.4. Average distance violations were calculated using the sum over  $r^6$ . <sup>d</sup>RDC goodness-of-fit quality factor  $Q_{\text{rmsd}}$  determined using PALES. <sup>e</sup>Based on ordered residue ranges [ $S(\varphi) + S(\psi) > 1.8$ ]. <sup>f</sup>RPF scores reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NOESY data and resonance assignments.

genes) to confer resistance upon *E. coli* provides a convenient indicator for putative function. As a preliminary assessment of potential CalU16/19 function, the genes encoding each protein were expressed in *E. coli* and the corresponding recombinant strains tested for **1** resistance in a manner reminiscent to that previously reported for CalC.<sup>18</sup> To do so, *calU16* and *calU19*

were cloned into pET28a to provide pSECalU16-*E. coli* and pSECalU19-*E. coli*, respectively, and the corresponding heterologous production levels of both CalU16 and CalU19 in *E. coli* were confirmed via SDS-PAGE to be comparable (Supporting Information Figure S8). A subsequent disc diffusion assay was used to test for **1** resistance of



**Figure 3.** CalU16 structural homologues. (A) CalU16 (PDB: 4FPW); (B) CalC (PDB: 2L65), calicheamicin resistance protein; (C) TcmN Aro/Cyc (in complex with *trans*-dihydroquercetin; PDB: 3TVQ) involved in the biosynthesis of tetracenomycin; (D) Hyp-1 (in complex with ethylene glycol; PDB: 3IE5) involved in the biosynthesis of hypericin; (E) NCS (in complex with hydroxybenzaldehyde; PDB: 2VQS) involved in the biosynthesis of norcoclaurine.



**Figure 4.** CalU16 and CalU19 assays. Serial disc dilutions of **1** against (A) pSE28a-*E. coli* (control), (B) pSECalU16-*E. coli* (CalU16), (C) pSECalU19-*E. coli* (CalU19), (D) pJB2011-*E. coli* (CalC). Amount of **1** on discs 1–6 are 10  $\mu\text{g}$ , 1  $\mu\text{g}$ , 100 ng, 50 ng, 10 ng, and 1 ng, respectively. Coomassie-stained 8–12% SDS-PAGE gradient gel of (E) CalU16 and (F) CalU19 in the presence of DTT (lane 1), **1** (lane 2), and DTT and **1** (lane 3).

pSECalU16-*E. coli*, pSECalU19-*E. coli*, pET28a-*E. coli* (an empty vector negative control), and pJB2011-*E. coli* (a *calC*-expressing positive control).<sup>18</sup> Figure 4 illustrates the rank order of *in vivo* resistance to be CalC  $\approx$  CalU19 > CalU16 with

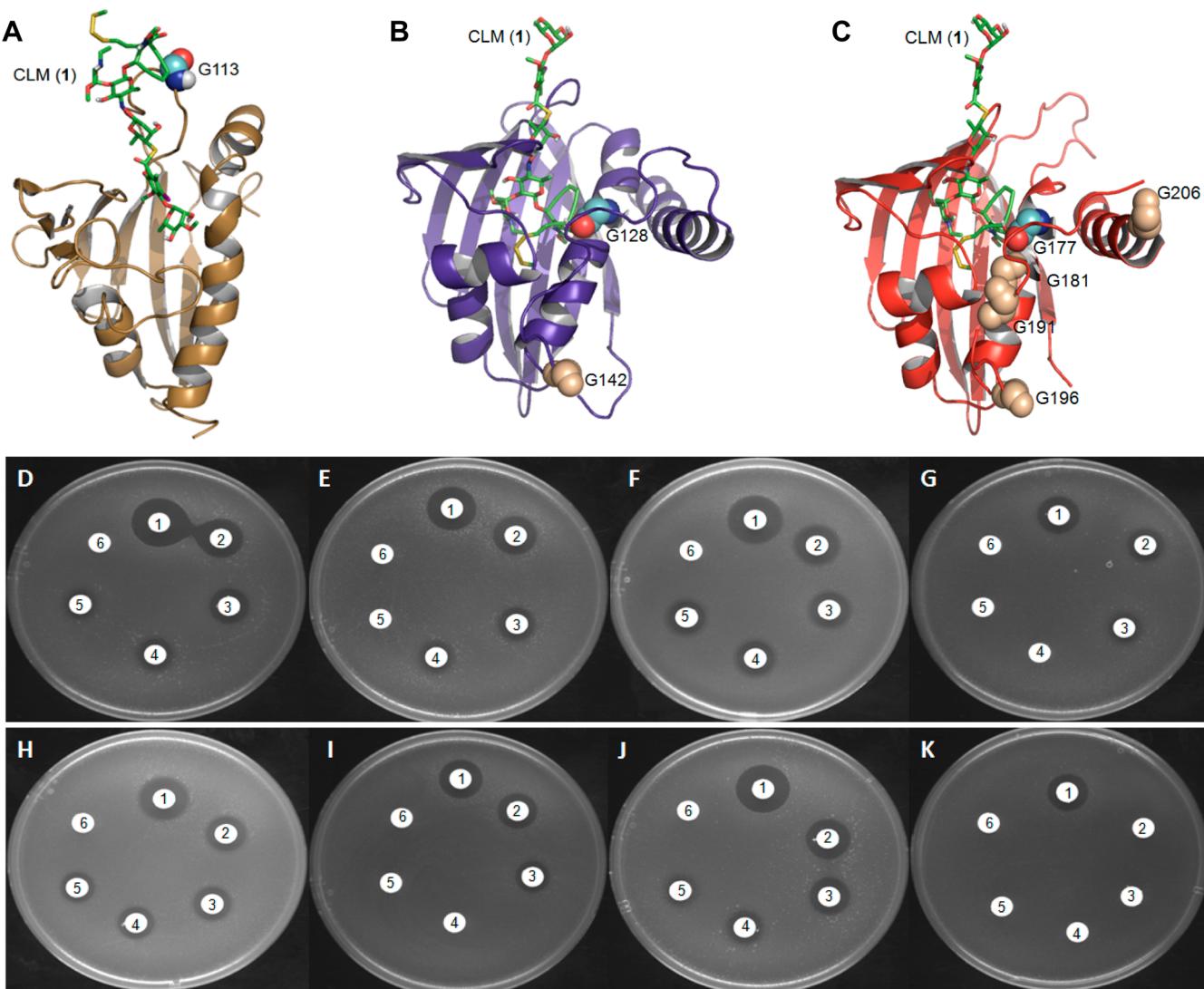
no resistance in the empty vector-containing negative control. Consistent with this qualitative assessment, the determination of minimum inhibitory concentration (MIC) (Table 3, Supporting Information Figure S9), revealed both pJB2011-*E.*

**Table 3. Minimum Inhibitory Concentration of Strains Used in This Study Against 1**

strain	MIC ( $\mu\text{M}$ )
pSE28a- <i>E. coli</i>	0.036
pSECalU16- <i>E. coli</i>	3
pSECalU19- <i>E. coli</i>	12
pJB2011- <i>E. coli</i>	12
pSEU16G128V- <i>E. coli</i>	0.36
pSEU16G128R- <i>E. coli</i>	0.18
pSEU16G142V- <i>E. coli</i>	3
pSEU16G142R- <i>E. coli</i>	3
pSEU19G177V- <i>E. coli</i>	6
pSEU19G177R- <i>E. coli</i>	0.36
pSEU19G181V- <i>E. coli</i>	12
pSEU19G191V- <i>E. coli</i>	12
pSEU19G196V- <i>E. coli</i>	12
pSEU19G206V- <i>E. coli</i>	12

*coli* and pSECalU19-*E. coli* to be 330-fold more tolerant to **1** than the pET28a-*E. coli* empty vector control while pSECalU16-*E. coli* was 80-fold more resistant than the control strain. Importantly, this supports the contention that CalU16/19 function in a manner similar to CalC. To further confirm that the **1**-resistance pattern of CalC, CalU16, and CalU19 is specific and not an artifact related to general proteins encoded by genes in the calicheamicin gene cluster, an identical study using the unknown-encoding gene *calU17* revealed no resistance toward **1** (Supporting Information Figure S10). It is also important to note that **1** does not indiscriminately cleave other proteins such as bovine serum albumin, illustrating the importance of the START domain hydrophobic core for enediyne binding and functional enediyne resistance.<sup>18</sup>

**CalU16 Biochemical Characterization.** The prototype self-sacrifice protein CalC functions via binding calicheamicin and providing an alternative hydrogen source for quenching the highly reactive diradical species formed upon calicheamicin cycloaromatization. In competition assays using the real-time



**Figure 5.** Site directed mutagenesis of CalU16 and CalU19. Docking models of (A) CalC (B) CalU16, and (C) CalU19 with mutated glycine residues represented as spheres where colored Gly residues indicate cleavage sites, wheat Gly residues indicate mutations that did not affect activity and calicheamicin (**1**) is represented as a stick model. Results of disc diffusion assay in CalU16 mutants (D) pSE28a-*E. coli* (control); (E) pSEU16G128V-*E. coli*; (F) pSEU16G128R-*E. coli*; (G) pSECalU16-*E. coli*; and CalU19 mutants (H) pSE28a-*E. coli* (control); (I) pSEU19G177V-*E. coli*; (J) pSEU19G177R-*E. coli*; (K) pSECalU19-*E. coli*. Amount of **1** on discs 1–6 are 10  $\mu\text{g}$ , 1  $\mu\text{g}$ , 100 ng, 50 ng, 10 ng, and 1 ng, respectively.

fluorescence-based molecular break light assay,<sup>13</sup> CalC was found to out-compete DNA for calicheamicin and thereby prevent calicheamicin-induced strand scission. In the presence of CalC, the calicheamicin diradical species abstracts the C<sub>α</sub> hydrogen of G113 to generate a protein radical that, in the presence of oxygen, leads to proteolysis into two distinct CalC fragments (Figure 1B).<sup>18</sup> Thus, a similar series of proteolysis experiments were conducted to determine whether CalU16 utilizes an analogous mechanism. Figure 4 highlights the outcome of this analysis using N-His<sub>6</sub>-CalU16. Importantly, incubation of CalU16 with **1** alone led to no reaction, incubation of CalU16 and **1** in the presence of reducing agent (dithiothreitol, DTT) led to the specific cleavage of CalU16 into two fragments of estimated size 16 kDa and 6 kDa based upon SDS-PAGE (Figure 4E). LC-MS/MS analysis of the tryptic digestion of the isolated fragments identified two nontryptic peptides flanking Gly128: <sup>101</sup>LSEEGDGTLLE-HATTSEQMLVEVG<sup>127</sup>V(-NH<sub>2</sub>) and <sup>129</sup>WEMALDFLGMFI<sup>141</sup>R (see MS/MS spectra of the two peptides in Supporting Information Figure S11). The first peptide ends with a nontryptic residue <sup>127</sup>V with an amine group (-NH<sub>2</sub>) from the <sup>128</sup>Gly residue attached. The second peptide starts with <sup>129</sup>W further supports that Gly128 is likely the cleavage site due to hydrogen abstraction. In addition, the LC-MS/MS data suggested <sup>142</sup>G as another potential cleavage site, based upon a weak peptide signal [<sup>129</sup>WEMALDFLGMFI<sup>141</sup>R(-NH<sub>2</sub>), ion score =24]. However, the corresponding peptide starting with <sup>143</sup>D was not identified. Nevertheless, we included both Gly128 and Gly142 in CalU16 as putative sites in the subsequent mutational study to further determine which glycine is the critical point of hydrogen abstraction.

To elucidate the key CalU16 glycine residue that serves as the putative hydrogen donor, four targeted CalU16 glycine mutants (G128V, G128R, G142V, and G142R) were created and tested both *in vivo* and *in vitro*. Strains expressing CalU16ΔG142 mutations (pSEU16G142V-*E. coli* and pSEU16G142R-*E. coli*) retained wtCalU16 (pSECALU16-*E. coli*) resistance levels to **1** while mutation of Gly128 (pSEU16G128V-*E. coli* and pSEU16G128R-*E. coli*) reduced or abolished tolerance to **1** (Figure 5, Supporting Information Figure S12, Table 3). Consistent with this, 1-based proteolysis of N-His<sub>6</sub>-CalU16 mutant proteins (Supporting Information Figure S13) revealed CalU16G142V and CalU16G142R to cleave in an identical manner to wtCalU16 in the presence of **1** and DTT while CalU16G128V and CalU16G128R were resistant to cleavage under identical conditions. Cumulatively, this data is consistent with CalU16 Gly128 as the key hydrogen donor in self-sacrifice mechanism, reminiscent of CalC Gly113, wherein cleavage of CalU16 (or CalC) inactivates **1** in a stoichiometric manner.

**CalU19 Biochemical Characterization.** To determine whether the CalU16 homologue CalU19 presents a **1** self-sacrifice resistance function analogous to CalC and CalU16, the expressed N-His<sub>6</sub>-CalU19 was purified and incubated with **1** alone or in the presence of a reducing agent (DTT). In this study, CalU19 was also cleaved into two fragments by **1** but, in stark contrast to CalC and CalU16, **1**-induced CalU19 cleavage occurred even in the absence of the reducing agent DTT (Figure 4F). The estimated sizes of the corresponding CalU19 fragments were ~21 and 6 kDa based upon SDS-PAGE (Supporting Information Table S2). LC-MS/MS analysis of the chymotrypsin digestion of the isolated fragments identified two nonchymotryptic peptides flanking Gly177: <sup>142</sup>RLTPSG-

DATVLELEHAPVPAAIPNAAPGAWGIG<sup>176</sup>A(-NH<sub>2</sub>) and <sup>178</sup>WEMGLVALDDYLAGTLPEGRAVD<sup>201</sup>W (see MS/MS spectra of the two peptides in Supporting Information Figure S14). The first peptide ends with a nonchymotryptic residue <sup>176</sup>A with an amine group (-NH<sub>2</sub>) from the <sup>177</sup>Gly residue attached. The second peptide starts with <sup>178</sup>W further supports that Gly177 is likely the cleavage site due to hydrogen abstraction.

To confirm the key CalU19 glycine residue, two targeted CalU19ΔGly177 mutants (G177V and G177R) were created and tested both *in vivo* and *in vitro*. To rule out any other possible glycine residues that might be involved in the protein cleavage, four additional glycine residues (Gly181, Gly191, Gly196 and Gly206) were mutated and the corresponding mutants tested in parallel. The *E. coli* strains expressing CalU19G181V, CalU19G191V, CalU19G196V, CalU19G206V (pSEU19G181V-*E. coli*, pSEU19G191V-*E. coli*, pSEU19G196V-*E. coli*, and pSEU19G206V-*E. coli*, respectively) retained wtCalU19 (pSECALU19-*E. coli*) resistance levels to **1** while mutation of Gly177 (pSEU19G177V-*E. coli* and pSEU19G177R-*E. coli*) reduced or abolished tolerance to **1** (Figure 5, Supporting Information Figure S15, and Table 3) as depicted by disc diffusion and MIC assays. Consistent with this, **1**-based proteolysis of N-His<sub>6</sub>-CalU19 mutant proteins revealed CalU19G177V and CalU19G177R to be resistant to cleavage when compared with wtCalU19 under identical conditions (Supporting Information Figure S16). In aggregate, this data is consistent with CalU19 Gly177 as the key hydrogen donor in the context of **1** inactivation and served as a basis for modeling the interactions of **1** with CalU16 and CalU19. Notably, the modeled orientation of **1** bound to CalU16 or CalU19 appears to be different from **1** with CalC (Figure 5A–C) consistent with prior observations that the late stage biosynthetic intermediates en route to **1** bind the glycosyltransferases involved in **1** maturation via distinct orientations.<sup>45</sup>

Distinct from CalC and CalU16, CalU19 functioned in the absence of a reducing agent—implicating a residue within CalU19 as a putative reductive activator of the process. CalU19 contains two cysteines (Cys25 and Cys120) where the CalU19 homology model suggests Cys25 to potentially be closer to the trisulfide of **1** in a ligand-bound model than Cys120 (estimated distance 14.80 Å, Supporting Information Figure S17). To test the putative role of Cys25 and Cys120, the corresponding alanine mutants were generated and tested *in vivo* and *in vitro*. Surprisingly, no differences in DTT-dependence were observed between wtCalU19, CalU19Cys25A, CalU19Cys120A and the double mutant CalU19Cys25A/Cys120A (Supporting Information Figure S18), suggesting **1** activation via CalU19 occurs via a nonredox mechanism. Notably, cycloaromatized calicheamicin *e* (Figure 1B) was detected in all CalU19 proteolysis experiments (including those with the double mutant CalU19Cys25A/Cys120A), indicating there is an alternative CalU19 contributor to trisulfide reductive initiation event.

**Enediyne Specificity of CalU16 and CalU19.** To test whether CalU16 and CalU19 confer resistance against other 10-membered enediynes, *in vivo* and *in vitro* studies similar to those described in the previous sections were performed where **1** was substituted with dynemicin (**2**) or esperamicin A<sub>1</sub> (**3**). CalU16 and CalU19 each failed to cause any cross resistance against **2** or **3** *in vivo* (Supporting Information Figure S19), and consistent with these observations, neither protein could be cleaved via **2** or **3** in the absence or presence of reducing agent

(Supporting Information Figure S20). Thus, CalU16 and CalU19 appear to display enediyne specificity similar to CalC.<sup>18</sup>

**Conclusions.** High throughput sequencing and genomics are powerful tools to implicate putative function but are still limited by three primary liabilities: (i) functional misannotation due to mistakes propagated throughout large sequence databases; (ii) functional misannotation due to the fact that even highly homologous proteins can present dramatically distinct mechanisms/functions; and (iii) a lack of suitably characterized homologues within existing large sequence databases.<sup>46–48</sup> Structural genomics can augment the use of sequence homology for functional annotation by presenting opportunities to identify structural homologues even where sequence homology may be too low to identify suitable homologues.<sup>49–51</sup> The current study highlights the power of structural genomics to inform putative function as the basis for subsequent biochemical characterization/confirmation and, in doing so, reveals two new enediyne self-sacrifice proteins (CalU16 and CalU19) encoded by genes in the calicheamicin biosynthetic gene cluster. The elucidation of these genes serve as potential new genetic markers which, in conjunction with the signature minimal enediyne PKS cassette, may facilitate future enediyne discovery via genome mining.<sup>17</sup>

While there exist many natural product biosynthetic loci that also encode for more than one resistance mechanism for the encoded natural product,<sup>52,53</sup> the encoded redundancy highlighted by the current study is uncommon and may suggest discrete self-sacrifice proteins to contribute to subtle distinctions in their localization and/or function. For example, the predicted isoelectric point of CalC (10.16) is dramatically different from CalU16 (4.24) or CalU19 (4.69) (Supporting Information Table S2) and is consistent with CalC's demonstrated ability to bind DNA (the target of enediynes) under physiological pH.<sup>19</sup> Under identical conditions, CalU16/U19 are predicted to possess an overall negative charge, which may contribute to distinct intracellular localization and/or unique protein–protein interactions possibly including those proteins/enzymes involved in 1 biosynthesis.

The enediyne self-sacrifice genes and corresponding encoded proteins have been validated in the context of conferring enediyne resistance, but it remains possible that such proteins or their proteolytic fragments could also serve alternative roles in *Micromonospora*. For example, many members of the START domain family function as molecular chaperones as exemplified by the Coq10 homologue CC1736, responsible for transporting ubiquinone to or within the respiratory chain complexes and contributes to dramatically higher biosynthetic efficiency.<sup>54</sup> Although the precedent for molecular chaperones in the context of the structurally related 9-membered enediynes is well established,<sup>5,55</sup> it remains to be determined whether such proteins play additional roles beyond stabilizing the highly reactive 9-membered enediyne chromophore. Preliminary evidence, based upon the location of the key glycines in CalC, CalU16, and CalU19, suggest that CalC binds 1 in a manner distinct from CalU16/CalU19 which, in the context of a molecular chaperone model, suggests different regions of the cargo would be accessible to putative associated partner proteins (Figure 5A–C). While CalC homologues lack apparent polyketide aromatase/cyclase TcmN Aro/Cyc catalytic residues based upon sequence (Supporting Information Figure S21) and structural (Supporting Information Figure S22) alignments, this relationship raises the intriguing possibility of CalC homologues as the long sought after

contributors to enediyne polyketide core folding/cyclization. Indeed, the surface area of the hydrophobic cavities of CalU16 (3986 Å<sup>2</sup>), CalC (3931 Å<sup>2</sup>), and TcmN ARO/CYC (3655 Å<sup>2</sup>) are similar (Supporting Information Figure S23), the latter of which accommodates large ( $\leq$ C20) linear polyketides.<sup>35</sup> Alternatively, certain proteolytic fragments from CalC, CalU16, and/or CalU19 could also potentially serve as intracellular signals of enediyne concentrations, in a manner conceptually similar to the vancomycin- or  $\beta$ -lactam-inducible three-component regulatory systems, where hydrolytic fragments of the bacterial cell wall serve as key inducers.<sup>56,57</sup> Studies are underway to further probe whether the functions of START domain congeners encoded by the calicheamicin gene cluster extend beyond simple protection of host against these highly reactive and toxic metabolites.

## METHODS

**Strains, Materials, and General Methods.** All primers (Integrated DNA Technology) and strains used in this study are summarized in Supporting Information Tables S3 and S4, respectively. *M. echinospora* strain LL6000 and calicheamicin  $\gamma_1^1$  were graciously provided by Pfizer. Dynemicin and esperamicin A<sub>1</sub> were generously provided by Bristol-Myers Squibb. All other reagents and chemicals were purchased from Sigma-Aldrich, unless otherwise stated. All protein structures were illustrated using PyMOL.<sup>58</sup> Gene alignments and analyses were performed using Geneious Pro 5.0.3.<sup>59</sup>

**Parental Plasmids for Protein Production.** For biochemical characterization, genomic DNA was extracted from *M. echinospora* LL6000 using the InstaGene Matrix Kit (BioRad) following manufacturer's protocol. Primers CalU16-NdeI-F/CalU16-HindIII-R, CalU17-NdeI-F/CalU17-HindIII-R, and CalU19-NdeI-F/CalU19-HindIII-R (Supporting Information Table S3) were used to amplify each of *calU16*, *calU17*, and *calU19*, respectively. Amplification was performed using the Advantage GC2 polymerase enzyme (Clontech) and the following PCR conditions: initial denaturation at 95 °C for 3 min; 25 cycles at 94 °C for 30 s, 65–62 °C for 30 s, 68 °C for 35 s; final extension temperature at 68 °C for 5 min. Each PCR product was gel purified using the QIAquick gel extraction kit (Qiagen). Restriction digestion with NdeI/HindIII (New Englands Biolabs, NEB) followed by ligation using T4 DNA ligase (NEB) into a linearized dephosphorylated pET28a (Novagen) yielded pSECALU16, pSECALU17, and pSECALU19 for *calU16*, *calU17*, and *calU19*, respectively. Each plasmid was transformed into NEB DH5 $\alpha$  chemically competent cells (NEB). Plasmids were purified using QIAprep Minispin kit (Qiagen). Each plasmid was verified by sequencing and subsequently transformed into the expression host BL21 (DE3) competent cells (NEB) to yield pSECALU16-*E. coli*, pSECALU17-*E. coli*, and pSECALU19-*E. coli*, respectively. For structural studies, plasmid preparation, overproduction, and purification were conducted following standard protocols of the Northeast Structural Genomics Consortium (NESG) to produce a uniformly <sup>15</sup>N/<sup>13</sup>C-enriched and 5% biosynthetically directed <sup>13</sup>C (NCS)-labeled protein samples for NMR spectroscopy and selenomethionine (Se-Met) labeled samples for X-ray crystallography and details can be found in Supporting Information Methods.<sup>60,61</sup> In brief, *calU16* was PCR amplified from genomic DNA and cloned into NdeI/XhoI-digested pET15 expression vector (NESG Clone ID MiR12-15.1; PSI:Biology Materials Repository clone ID MeCD000597015, see <http://psimr.asu.edu/>) to encode for the corresponding N-terminally tagged (MGHHHHHHSH) fusion protein. The plasmid was transformed into codon-enhanced *E. coli* BL21 (DE3) pMGK cells for protein production. Additional plasmids and protocols employed for protein overproduction and purification for biochemical studies are described in the Supporting Information Methods.

**Structure Determination of CalU16 by X-ray Crystallography.** Initial crystallization conditions for Se-Met labeled CalU16 were identified at the Hauptmann-Woodward Institute high-throughput screening facility<sup>62</sup> and further optimized manually by

microbatch methods at 18 °C. The protein was mixed in a 1:1 ratio with the precipitant containing 2.9 M sodium malonate pH 6.5. The crystals were cryo-protected by 10% (v/v) ethylene glycol and 2.6 M sodium malonate pH 6.5 prior to flash freezing in liquid nitrogen for data collection at 100 K. A single crystal of Se-Met labeled protein was used for data collection at a wavelength of 0.978 Å and diffracted to 2.5 Å resolution. The data was collected at beamline X4A at the National Synchrotron Light Source at Brookhaven National Laboratory, and further processed and scaled using HKL-2000 (Table 1).<sup>63</sup> Eight of the ten selenium sites in the asymmetric unit were successfully located by the SHELXD<sup>64</sup> program for initial phasing. RESOLVE<sup>65</sup> was utilized for automated model building and the model was further built by manual refitting with the program Coot.<sup>66</sup> The data were twinned, and a twin refinement was performed. The refinement involved iterations of manual model-building in Coot and Refmac.<sup>67</sup> The quality of the final structure was validated by PROCHECK.<sup>68</sup> The diffraction images illustrate diffuse scattering perpendicular to the  $6_1$  screw axis indicative of order/disorder (Supporting Information Figure S24). The effects of the disorder are also evident in the data anisotropy where the B-factor component along the  $6_1$  axis is less than half of the B-factor in the perpendicular direction. The unmodeled disorder leads to higher than expected R-factors and is likely contributing to the lower percentage of residues in the Ramachandran favored regions (Table 1). The atomic coordinates and structure factors have been deposited in the PDB (4FPW).

**NMR Spectroscopy and Structure Determination of CalU16.** NMR data were collected at 25 °C on [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]- and  $\text{U}^{15}\text{N}$ , 5% biosynthetically directed  $^{13}\text{C}$  (NC5)-labeled samples in 300 μL buffered solution (0.9 mM CalU16, 0.02%  $\text{NaN}_3$ , 10 mM DTT, 5 mM  $\text{CaCl}_2$ , 100 mM NaCl, 1× proteinase inhibitors, 20 mM MES pH 6.5) in 5 mm Shigemi NMR tubes on a 600 MHz Varian Inova spectrometer with a 5 mm HCN cold probe and a 850 MHz Bruker Avance III NMR spectrometer equipped with a conventional 5 mm HCN probe. A description of NMR experiments and methods for structure determination and refinement can be found in Supporting Information Methods. The protein was monomeric under the conditions used in the NMR experiments based on analytical static light scattering in-line with gel filtration chromatography and correlation time estimates based on one-dimensional  $^{15}\text{N}$   $T_1$  and  $T_2$  relaxation data (estimated  $\tau_c$  12.1 ns, Supporting Information Figure S1). The assigned  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum is provided as Supporting Information Figure S2. Chemical shifts, NOESY peak lists, and raw FIDS were deposited in the BioMagResDB (BMRB accession number 18547). The final ensemble of 20 models and NMR resonance assignments were deposited to the Protein Data Bank (PDB: 2LUZ).

**Resistance Assays.** *In vivo* resistance assays were performed using overnight cultures of the tested strain diluted to  $\text{OD}_{600} \sim 0.1$  in molten LB agar supplemented with final concentration of 30 μg mL<sup>-1</sup> kanamycin and 100 μM of isopropyl-β-D-thiogalactopyranoside (IPTG). The mixture was poured into Petri dishes and allowed to solidify. A stock solution of **1** was prepared in methanol at a concentration of 1 mg mL<sup>-1</sup> and serial dilutions were prepared such that the final concentrations tested ranged from 0.1 to 10 μg μL<sup>-1</sup>. These were applied to sterile discs such that the final quantities of **1** on the discs were 1 ng, 10 ng, 50 ng, 100 ng, 1 μg, and 10 μg. The disks were allowed to air-dry, and applied aseptically to the plates. All plates were incubated at 37 °C for 14 h.

**Protein Cleavage Assay.** Final concentration of 1 μM of each protein in 50 mM Tris pH 8.0 in the presence of 1 mol equiv of **1** with and without 0.2 mM DTT in a final volume of 500 μL was allowed to incubate at 30 °C for 30 min. Reactions were lyophilized and resuspended in 25 μL of 50 mM Tris, pH 8.0. Cleavage was analyzed by mixing with equal volume of 2× sample buffer. A volume of 30 μL was analyzed by NuPAGE 4–12% Bis-Tris Gels (Life technologies) using protein ladder (Bio-Rad) as size indicator.

**Trypsin Digestion and Mass Analysis.** A total of 4 μM of each of CalU16 and CalU19 were added to 2 mol equiv of **1** in the presence of 0.2 mM DTT. Reactions were incubated at 30 °C for 30 min. A fraction of each reaction was subjected to SDS-PAGE to confirm completion of the reaction. For proteolytic digestion, the protein

bands were excised from SDS-PAGE gel, washed with 50 mM ammonium bicarbonate, reduced with DTT, alkylated with iodoacetamide (IAA), and subjected to trypsin (for CalU16) or chymotrypsin (for CalU19) digestion at 37 °C for 16 h. Peptides resulting from the proteolytic digestion were extracted from gel pieces and analyzed by LC-MS/MS as previously described.<sup>69</sup> Briefly, the information-dependent acquisition of LC-MS/MS data were acquired using an LTQ Velos Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) coupled with a Nano-LC Ultra/cHiPLC-nanoflex HPLC system (Eksigent, Dublin, CA) through a nano-electrospray ionization source.<sup>70</sup> The results were subjected using ProteomeDiscoverer 1.3 software (Thermo Fisher Scientific, Waltham, MA) and MASCOT server. The cleavage sites were deduced from nonenzymatic peptides ending with an amine group ( $-\text{NH}_2$ ) from the Gly residue attached to a nontrypsin or nonchymotrypsin digestion site.

**Determination of Minimum Inhibitory Concentrations (MICs).** Fourteen strains created in this study (Supporting Information Table S4, strains 2–15) were grown overnight in LB supplemented with kanamycin 30 mg L<sup>-1</sup>. Each overnight culture was diluted to  $\text{OD}_{600} \sim 0.1$  using an Overnight Express Autoinduction System 1 (Novagen) supplemented with kanamycin 30 mg L<sup>-1</sup>. Enediyne **1** was serially diluted 1:2 in DMSO such that the final concentrations ranged from 2.4 μM to 4.5 nM for pSE28a-*E. coli* and from 12 μM to 22 nM for all other strains. Ampicillin and kanamycin were used as positive and negative controls, respectively. All plates were incubated for 18 h at 37 °C at 250 rpm. The residual metabolic activity was monitored based on the irreversible reduction of resazurin (7-hydroxy-3H-phenoxyazin-3-one-10-oxide, blue) to resorufin (7-hydroxy-3H-phenoxyazin-3-one, pink). For this assay, 10 μL of resazurin (final concentration 100 μM) was added to each well and allowed to incubate at 37 °C with shaking at 250 rpm for 1 h to allow viable cells to convert resazurin to resorufin. The well with the minimum concentration of **1** that did not display bacterial growth but did display a red color after incubation with resazurin was recorded as the MIC of that strain (Supporting Information Figure S9; Table 3).

## ■ ASSOCIATED CONTENT

### § Supporting Information

Additional protocols and methods, supplemental figures, and supplemental tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*Email: jsthorson@uky.edu.

\*Email: georgep@rice.edu.

### Present Address

□School of Pharmacy and Health Profession, University of Maryland Eastern Shore, Princess Anne, Maryland 21853, U.S.A.

### Notes

The authors declare the following competing financial interest(s): J.S.T. is a co-founder of Centrose (Madison, WI).

## ■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) grants CA84374 (to J.S.T.), U01GM098248 (to G.N.P.), the National Center for Advancing Translational Sciences (UL1TR000117), a NIH PSI grant (U54-GM094597 to M.A.K. and G.T.M.) and the BioXFEL Science and Technology Center under National Science Foundation Grant No. 1231306. Enediynes for this study were generously provided by Pfizer (calicheamicin) and Bristol-Myers-Squibb (esperamicin and dynemicin). Proteomics analyses were conducted by the University of Kentucky Proteomics Core

that is partially supported by grants from the National Institute of General Medical Sciences (P20GM103486), the National Cancer Institute (P30CA177558) and equipment acquired via a grant from National Center for Research Resources (1S10RR029127 to H.Z.). Protein NMR data collection was conducted at the Ohio Biomedicine Center of Excellence in Structural Biology and Metabolomics (Miami University). We thank Dr. J. Prestgard, Dr. T. Acton, and Dr. J. Everett at the Northeast Structural Genomics Consortium for technical assistance and J. Schwanof and R. Abramowitz for data collection assistance at beamline X4A, NSLS.

## ■ REFERENCES

- (1) Maiese, W. M., Lechevalier, M. P., Lechevalier, H. A., Korshalla, J., Kuck, N., Fantini, A., Wildey, M. J., Thomas, J., and Greenstein, M. (1989) Calicheamicins, a novel family of antitumor antibiotics: taxonomy, fermentation and biological properties. *J. Antibiot. (Tokyo)* **42**, 558–563.
- (2) Lee, M. D., Manning, J. K., Williams, D. R., Kuck, N. A., Testa, R. T., and Borders, D. B. (1989) Calicheamicins, a novel family of antitumor antibiotics. 3. Isolation, purification and characterization of calicheamicins  $\beta$  1Br,  $\gamma$  1Br,  $\alpha$  2I,  $\alpha$  3I,  $\beta$  1I,  $\gamma$  1I, and  $\delta$  1I. *J. Antibiot. (Tokyo)* **42**, 1070–1087.
- (3) Lee, M. D., Dunne, T. S., Chang, C. C., Siegel, M. M., Morton, G. O., Ellestad, G. A., McGahren, W. J., and Borders, D. B. (1992) Calicheamicins, a novel family of antitumor antibiotics. 4. Structure elucidation of calicheamicins  $\beta$  1Br,  $\gamma$  1Br,  $\alpha$  2I,  $\alpha$  3I,  $\beta$  1I,  $\gamma$  1I, and  $\delta$  1I. *J. Am. Chem. Soc.* **114**, 985–997.
- (4) Thorson, J. S., Sievers, E. L., Ahlert, J., Shepard, E., Whitwam, R. E., Onwueme, K. C., and Ruppen, M. (2000) Understanding and exploiting nature's chemical arsenal: The past, present, and future of calicheamicin research. *Curr. Pharm. Des.* **6**, 1841–1879.
- (5) Galm, U., Hager, M. H., Van Lanen, S. G., Ju, J., Thorson, J. S., and Shen, B. (2005) Antitumor antibiotics: Bleomycin, enediynes, and mitomycin. *Chem. Rev.* **105**, 739–758.
- (6) Liang, Z.-X. (2010) Complexity and simplicity in the biosynthesis of enediyne natural products. *Nat. Prod. Rep.* **27**, 499–528.
- (7) De Voss, J. J., Hangeland, J. J., and Townsend, C. A. (1990) Characterization of the *in vitro* cyclization chemistry of calicheamicin and its relation to DNA cleavage. *J. Am. Chem. Soc.* **112**, 4554–4556.
- (8) Chatterjee, M., Cramer, K. D., and Townsend, C. A. (1993) Kinetic nature of thiol activation in DNA cleavage by calicheamicin. *J. Am. Chem. Soc.* **115**, 3374–3375.
- (9) Myers, A. G., Cohen, S. B., and Kwon, B. M. (1994) A study of the reaction of calicheamicin  $\gamma$  1 with glutathione in the presence of double-stranded DNA. *J. Am. Chem. Soc.* **116**, 1255–1271.
- (10) Zein, N., Sinha, A. M., McGahren, W. J., and Ellestad, G. A. (1988) Calicheamicin  $\gamma$  1I: An antitumor antibiotic that cleaves double-stranded DNA site specifically. *Science* **240**, 1198–1201.
- (11) Zein, N., Poncin, M., Nilakantan, R., and Ellestad, G. A. (1989) Calicheamicin  $\gamma$  1I and DNA: Molecular recognition process responsible for site-specificity. *Science* **244**, 697–699.
- (12) Mah, S. C., Price, M. A., Townsend, C. A., and Tullius, T. D. (1994) Features of DNA recognition for oriented binding and cleavage by calicheamicin. *Tetrahedron* **50**, 1361–1378.
- (13) Biggins, J. B., Prudent, J. R., Marshall, D. J., and Thorson, J. S. (2006) A continuous assay for DNA cleavage using molecular break lights. *Methods Mol. Biol.* **335**, 83–92.
- (14) Danishefsky, S. J., and Shair, M. D. (1996) Observations in the chemistry and biology of cyclic enediyne antibiotics: Total syntheses of calicheamicin  $\gamma_1^1$  and dynemicin A. *J. Org. Chem.* **61**, 16–44.
- (15) Nicolaou, K. C., Hale, C. R. H., and Nilewski, C. (2012) A total synthesis trilogy: Calicheamicin  $\gamma_1$ (I), Taxol, and brevetoxin A. *Chem. Rec.* **12**, 407–441.
- (16) Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M., and Thorson, J. S. (2002) The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173–1176.
- (17) Liu, W., Ahlert, J., Gao, Q., Wendt-Pienkowski, E., Shen, B., and Thorson, J. S. (2003) Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11959–11963.
- (18) Biggins, J. B., Onwueme, K. C., and Thorson, J. S. (2003) Resistance to enediyne antitumor antibiotics by CalC self-sacrifice. *Science* **301**, 1537–1541.
- (19) Singh, S., Hager, M. H., Zhang, C., Griffith, B. R., Lee, M. S., Hallenga, K., Markley, J. L., and Thorson, J. S. (2006) Structural insight into the self-sacrifice mechanism of enediyne resistance. *ACS Chem. Biol.* **1**, 451–460.
- (20) Zhang, C., Griffith, B. R., Fu, Q., Albermann, C., Fu, X., Lee, I.-K., Li, L., and Thorson, J. S. (2006) Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **313**, 1291–1294.
- (21) Johnson, H. D., and Thorson, J. S. (2008) Characterization of CalE10, the *N*-oxidase involved in calicheamicin hydroxyaminosugar formation. *J. Am. Chem. Soc.* **130**, 17662–17663.
- (22) Belecki, K., Crawford, J. M., and Townsend, C. A. (2009) Production of octaketide polyenes by the calicheamicin polyketide synthase CalE8: Implications for the biosynthesis of enediyne core structures. *J. Am. Chem. Soc.* **131**, 12564–12566.
- (23) Horsman, G. P., Chen, Y., Thorson, J. S., and Shen, B. (2010) Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediynes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11331–11335.
- (24) Gantt, R. W., Peltier-Pain, P., Singh, S., Zhou, M., and Thorson, J. S. (2013) Broadening the scope of glycosyltransferase-catalyzed sugar nucleotide synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7648–7653.
- (25) Belecki, K., and Townsend, C. A. (2012) Environmental control of the calicheamicin polyketide synthase leads to detection of a programmed octaketide and a proposal for enediyne biosynthesis. *Angew. Chem., Int. Ed.* **51**, 11316–11319.
- (26) Belecki, K., and Townsend, C. A. (2013) Biochemical determination of enzyme-bound metabolites: Preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J. Am. Chem. Soc.* **135**, 14339–14348.
- (27) Ricart, A. D. (2011) Antibody-drug conjugates of calicheamicin derivative: Gemtuzumab Ozogamicin and Inotuzumab Ozogamicin. *Clin. Cancer Res.* **17**, 6417–6427.
- (28) Trail, P. A. (2013) Antibody drug conjugates as cancer therapeutics. *Antibodies* **2**, 113–129.
- (29) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
- (30) Iyer, L. M., Koonin, E. V., and Aravind, L. (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* **43**, 134–144.
- (31) Thorsell, A.-G., Lee, W. H., Persson, C., Siponen, M. I., Nilsson, M., Busam, R. D., Kotenyova, T., Schüller, H., and Lehtio, L. (2011) Comparative structural analysis of lipid binding START domains. *PLoS One* **6**, e19521.
- (32) Alpy, F., and Tomasetto, C. (2005) Give lipids a START: The StAR-related lipid transfer (START) domain in mammals. *J. Cell Sci.* **118**, 2791–2801.
- (33) Holm, L., and Rosenström, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549.
- (34) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2256–2268.
- (35) Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y., and Tsai, S.-C. (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5349–5354.
- (36) Michalska, K., Fernandes, H., Sikorski, M., and Jaskolski, M. (2010) Crystal structure of Hyp-1, a St. John's Wort protein implicated in the biosynthesis of hypericin. *J. Struct. Biol.* **169**, 161–171.

- (37) Ilari, A., Franceschini, S., Bonamore, A., Arenghi, F., Botta, B., Macone, A., Pasquo, A., Bellucci, L., and Boffi, A. (2009) Structural basis of enzymatic (*S*)-norcoclaurine biosynthesis. *J. Biol. Chem.* 284, 897–904.
- (38) Kofler, S., Asam, C., Eckhard, U., Wallner, M., Ferreira, F., and Brandstetter, H. (2012) Crystallographically mapped ligand binding differs in high and low IgE binding isoforms of birch pollen allergen bet v 1. *J. Mol. Biol.* 422, 109–123.
- (39) Li, W., Wang, L., Sheng, X., Yan, C., Zhou, R., Hang, J., Yin, P., and Yan, N. (2013) Molecular basis for the selective and ABA-independent inhibition of PP2CA by PYL13. *Cell Res.* 23, 1369–1379.
- (40) Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., Sikorski, M. M., and Jaskolski, M. (2006) Crystal structure of *Vigna radiata* cytokinin-specific binding protein in complex with zeatin. *Plant Cell* 18, 2622–2634.
- (41) Schirmer, T., Hoffmann-Sommergrube, K., Susani, M., Breiteneder, H., and Marković-Housley, Z. (2005) Crystal structure of the major celery allergen Api g 1: Molecular analysis of cross-reactivity. *J. Mol. Biol.* 351, 1101–1109.
- (42) Osipiuk, J., Wu, R., Moy, S., Joachimiak, A. (2006) X-ray crystal structure of conserved hypothetical protein EF\_2215 from *Enterococcus faecalis*. PDB ID: 2NN5. Unpublished.
- (43) Singarapu, K., Eletsky, A., Sathyamoorthy, B., Sukumaran, D., Wang, D., Jiang, M., Ciccosanti, C., Xiao, R., Liu, J., Baran, M. C., Swapna, G., Acton, T. B., Rost, B., Montelione, G. T., Szyperski, T. (2008) Solution NMR structure of protein encoded by gene BPP1335 from *Bordetella parapertussis*. PDB ID: 2K5G. Unpublished.
- (44) Clark, B. J. (2012) The mammalian START domain protein family in lipid transport in health and disease. *J. Endocrinol.* 212, 257–275.
- (45) Chang, A., Singh, S., Helmlich, K. E., Goff, R. D., Bingman, C. A., Thorson, J. S., and Phillips, G. N., Jr. (2011) Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17649–17654.
- (46) Gerlt, J. A., and Babbitt, P. C. (2000) Can sequence determine function? *Genome Biol.* 1, 1–10.
- (47) Whisstock, J. C., and Lesk, A. M. (2003) Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.* 36, 307–340.
- (48) Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: Misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* 5, e1000605.
- (49) Baker, D., and Sali, A. (2001) Protein structure prediction and structural genomics. *Science* 294, 93–96.
- (50) Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Rauschel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function. *Nature* 448, 775–779.
- (51) Guichou, J.-F., and Labesse, G. (2012) Fragment and conquer: From structure to complexes to function. *Structure* 20, 1617–1619.
- (52) Wright, G. D. (2007) The antibiotic resistome: The nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* 5, 175–186.
- (53) Cundliffe, E., and Demain, A. L. (2010) Avoidance of suicide in antibiotic-producing microbes. *J. Ind. Microbiol. Biotechnol.* 37, 643–672.
- (54) Allan, C. M., Hill, S., Morvaridi, S., Saiki, R., Johnson, J. S., Liau, W.-S., Hirano, K., Kawashima, T., Ji, Z., Loo, J. A., Shepherd, J. N., and Clarke, C. F. (2013) A conserved START domain coenzyme Q-binding polypeptide is required for efficient Q biosynthesis, respiratory electron transport, and antioxidant function in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 1831, 776–791.
- (55) Baker, J. R., Woolfson, D. N., Muskett, F. W., Stoneman, R. G., Urbaniak, M. D., and Caddick, S. (2007) Protein–small molecule interactions in neocarzinostatin, the prototypical enediyne chromoprotein antibiotic. *ChemBioChem* 8, 704–717.
- (56) Weigel, L. M., Clewell, D. B., Gill, S. R., Clark, N. C., McDougal, L. K., Flanagan, S. E., Kolonay, J. F., Shetty, J., Killgore, G. E., and Tenover, F. C. (2003) Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. *Science* 302, 1569–1571.
- (57) Qureshi, N. K., Yin, S., and Boyle-Vavra, S. (2014) The role of the Staphylococcal VraTSR regulatory system on vancomycin resistance and vanA operon expression in vancomycin-resistant *Staphylococcus aureus*. *PloS One* 9, e85873.
- (58) *The PyMOL Molecular Graphics System*, Version 1.3r1. (2010, August) Schrödinger, LLC., New York.
- (59) Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- (60) Acton, T. B., Xiao, R., Anderson, S., Aramini, J., Buchwald, W. A., Ciccosanti, C., Conover, K., Everett, J., Hamilton, K., Huang, Y. J., Janjua, H., Kornhaber, G., Lau, J., Lee, D. Y., Liu, G., Maglaqui, M., Ma, L., Mao, L., Patel, D., Rossi, P., Sahdev, S., Shastry, R., Swapna, G. V. T., Tang, Y., Tong, S., Wang, D., Wang, H., Zhao, L., and Montelione, G. T. (2011) Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* 493, 21–60.
- (61) Xiao, R., Anderson, S., Aramini, J., Belote, R., Buchwald, W. A., Ciccosanti, C., Conover, K., Everett, J. K., Hamilton, K., Huang, Y. J., Janjua, H., Jiang, M., Kornhaber, G. J., Lee, D. Y., Locke, J. Y., Ma, L.-C., Maglaqui, M., Mao, L., Mitra, S., Patel, D., Rossi, P., Sahdev, S., Sharma, S., Shastry, R., Swapna, G. V. T., Tong, S. N., Wang, D., Wang, H., Zhao, L., Montelione, G. T., and Acton, T. B. (2010) The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J. Struct. Biol.* 172, 21–33.
- (62) Luft, J. R., Wolfley, J. R., Said, M. I., Nagel, R. M., Lauricella, A. M., Smith, J. L., Thayer, M. H., Veatch, C. K., Snell, E. H., Malkowski, M. G., and Detitta, G. T. (2007) Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. *Protein Sci. Publ. Protein Soc.* 16, 715–722.
- (63) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- (64) Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 479–485.
- (65) Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 55, 849–861.
- (66) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 486–501.
- (67) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 53, 240–255.
- (68) Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- (69) Shi, J., Wang, Y., Zeng, L., Wu, Y., Deng, J., Zhang, Q., Lin, Y., Li, J., Kang, T., Tao, M., Rusinova, E., Zhang, G., Wang, C., Zhu, H., Yao, J., Zeng, Y.-X., Evers, B. M., Zhou, M.-M., and Zhou, B. P. (2014) Disrupting the interaction of BRD4 with diacetylated twist suppresses tumorigenesis in basal-like breast cancer. *Cancer Cell* 25, 210–225.
- (70) Zhai, J., Ström, A.-L., Kilty, R., Venkatakrishnan, P., White, J., Everson, W. V., Smart, E. J., and Zhu, H. (2009) Proteomic characterization of lipid raft proteins in amyotrophic lateral sclerosis mouse spinal cord. *FEBS J.* 276, 3308–3323.
- (71) Chen, V. B., Arendall, W. B., 3rd, 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.* 66, 12–21.