Crystal structure of SgcJ, an NTF2-like superfamily protein involved in biosynthesis of the nine-membered enediyne antitumor antibiotic C-1027

Tingting Huang1,9, Chin-Yuan Chang1, Jeremy R Lohman1,10, Jeffrey D Rudolf1, Youngchang Kim2,3,4, Changsoo Chang3,4, Dong Yang1, Ming Ma1, Xiaohui Yan1, Ivana Crnovcic1, Lance Bigelow4, Shonda Clancy4, Craig A Bingman3, Ragothaman M Yennamalli6,11, Gyorgy Babnigg2,4, Andrzej Joachimiak2,3,4, George N Phillips6 and Ben Shen1,7,8

Comparative analysis of the enediyne biosynthetic gene clusters revealed sets of conserved genes serving as outstanding candidates for the enediyne core. Here we report the crystal structures of SgcJ and its homologue NCS-Orf16, together with gene inactivation and site-directed mutagenesis studies, to gain insight into enediyne core biosynthesis. Gene inactivation in vivo establishes that SgcJ is required for C-1027 production in Streptomyces globisporus. SgcJ and NCS-Orf16 share a common structure with the nuclear transport factor 2-like superfamily of proteins, featuring a putative substrate binding or catalytic active site. Site-directed mutagenesis of the conserved residues lining this site allowed us to propose that SgcJ and its homologues may play a catalytic role in transforming the linear polyene intermediate, along with other enediyne polyketide synthase-associated enzymes, into an enzyme-sequestered enediyne core intermediate. These findings will help formulate hypotheses and design experiments to ascertain the function of SgcJ and its homologues in nine-membered enediyne core biosynthesis. The Journal of Antibiotics (2016) 69, 731–740; doi:10.1038/ja.2016.88; published online 13 July 2016

INTRODUCTION

The enediynes represent one of the most fascinating families of natural products for their unprecedented molecular architecture and extraordinary biological activities, and they have had profound impact on modern chemistry, biology and medicine.1–4 Since the structure of the neocarzinostatin (NCS) chromophore was first elucidated in 1985,5 the enediyne family of natural products has grown steadily with a total of 15 enediynes structurally characterized to date, of which four were isolated in the cycloaromatized form.4 The enediynes are classified into two subcategories according to the size of the enediyne core structures.1,2 Members of the 9-membered enediyne core subcategory included NCS, C-1027, kedarcidin, maduropeptin, N1999A2, the sporolides, the cyanosporasides (CYA, CYN) and the fijiolides (Figure 1a). Members of the 10-membered enediyne core subcategory included the calicheamicins (CAL), esperamincins (ESP), dynemicin (DYN), namenamicin, shishijimicin and uncialamycin (Supplementary Figure S1). The enediynes have provided an outstanding opportunity to decipher the genetic and biochemical basis for the biosynthesis of complex natural products,1–4 to explore ways to make novel analogues by manipulating genes governing their biosynthesis,5–10 and to discover new enediyne natural products by mining microbial genomes for the trademark enediyne biosynthetic machineries.3,4,11,12

The first set of biosynthetic gene clusters for the 9-membered enediyne C-102713 and the 10-membered enediyne CAL14 was cloned in 2002. Since then, a total of seven biosynthetic gene clusters for 9-membered enediynes (that is, C-1027,13 NCS,15 MDP,16 kedarcidin,17 sporolides,18 CYA19 and CYN19) and three biosynthetic gene clusters for 10-membered enediynes (that is, CAL,14 ESP (partial)11 and DYN20) have been reported. Comparative analysis of these gene clusters revealed a set of five genes common to both 9- and 10-membered enediynes (that is, the enediyne polyketide synthase (PKS) cassette consisting of E3/E4/E5/E/E10 (Figure 1b)),

1 Department of Chemistry, The Scripps Research Institute, Jupiter, FL, USA; 2 Center for Structural Genomics of Infectious Diseases, University of Chicago, Chicago, IL, USA; 3 Structural Biology Center, Argonne National Laboratory, Argonne, IL, USA; 4 Midwest Center for Structural Genomics, Biosciences Division, Argonne National Laboratory, Argonne, IL, USA; 5 Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA; 6 BioSciences at Rice and Department of Chemistry, Rice University, Houston, TX, USA; 7 Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, FL, USA and 8 Natural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, Jupiter, FL, USA.
9 Present address: State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.
10 Present address: Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA.
11 Present address: Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Himachal Pradesh 173234, India.
Correspondence: Prof B Shen, Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, #3A1, Jupiter, FL 33458, USA.
E-mail: shenb@scripps.edu
Dedicated to Professor David E Cane, Brown University, for his distinguished contributions to natural product biosynthesis and engineering.
Received 21 March 2016; revised 30 May 2016; accepted 15 June 2016; published online 13 July 2016
characterization of which has unambiguously established (i) the polypeptide origin for both 9- and 10-membered enediyne and (ii) a convergent model for enediyne biosynthesis. Although significant progress has been made toward elucidating the biosynthesis of the peripheral moieties present in enediyne, little is known about the enediyne core biosynthesis. In vivo and in vitro studies have established that the iterative type I PKS enzyme E initiates both 9- and 10-membered enediyne core biosynthesis via an acyl carrier protein-tethered linear polyene intermediate, which, in the absence of other enediyne PKS-associated enzymes, could be released by the thioesterase E10 to afford a heptaene. However, the enzymes and chemistry responsible for converting heptaene, or the nascent acyl carrier protein-tethered linear polyene intermediate, into the 9- and 10-membered enediyne cores remain elusive. Many of the candidate genes, predicted to be associated with enediyne core biosynthesis, are often annotated to encode proteins of unknown function. Inactivation of these candidate genes in vivo afforded mutant strains that often failed to accumulate any biosynthetic intermediate, revealing few clues for their function in enediyne core biosynthesis. Lack of functional prediction, together with the unavailability of suitable substrates, essentially forfeits any practical attempt to directly characterize these proteins biochemically in vitro.

Here we report the crystal structures of SgcJ and its homologue NCS-Orf16, together with gene inactivation and site-directed mutagenesis studies, to gain insight into enediyne core biosynthesis. We first closely examined the seven gene clusters that encode 9-membered enediyne biosynthesis and uncovered seven genes (E2, E7, E8, E9, E11, M, J) that are absolutely conserved but whose function could not be predicted on the basis of bioinformatics analysis alone. We next confirmed that SgcJ is absolutely required for C-1027 biosynthesis, inactivation of which in the C-1027 overproducer Streptomyces globisporus SB102210 completely abolished C-1027 production in the resultant ΔsgcJ mutant strain SB1027. We finally showed that SgcJ and NCS-Orf16 share a common structure with the nuclear transport factor 2 (NTF2)-like superfamily of proteins, featuring a hydrophobic pocket in the α/β barrel structure...
that could constitute as a putative substrate binding or catalytic active site. Site-directed mutagenesis of the conserved residues lining this site abolished C-1027 production, suggesting that SgcJ and its homologues may play a catalytic role in the 9-membered enediyne core biosynthesis.

RESULTS AND DISCUSSION
SgcJ and homologues are conserved among the 9-membered enediyne biosynthetic gene clusters but their function could not be predicted

Inspired by the enediyne PKS cassette, consisting of E3/E4/ES/E/E10, that is conserved among the seven 9-membered and three 10-membered enediyne biosynthetic gene clusters characterized to date, we recently completed a virtual survey of all bacterial genomes available in public databases using the enediyne PKS cassette as a probe.3,4 This effort resulted in the identification of an additional 77 putative enediyne biosynthetic gene clusters, implying that enediyynes are more common than currently appreciated on the basis of structurally characterized enediyne natural products.5-11,12 We subsequently constructed an enediyne genome neighborhood network, including both the 10 known and 77 putative enediyne gene clusters, to facilitate cluster annotation and predict 9- and 10-membered enediyne core biosynthesis. The enediyne PKS cassette is present in all 87 gene clusters, suggesting that they may be responsible for biosynthesis of a common intermediate for both 9- and 10-membered enediyne cores. Subsets of genes that are unique to either 9- or 10-membered enediyne gene clusters are also identified, as exemplified by the E2, E7, E8, E9, E11, M and J genes from the seven known 9-membered enediyne biosynthetic gene clusters (Figure 1b), and they may play roles in diversifying the common intermediate into the 9- or 10-membered enediyne cores, respectively.3,4

Among this set of genes is SgcJ (Figure 1b), and its homologues are present in the 34 putative 9-membered enediyne biosynthetic gene clusters (Supplementary Figure S2). SgcJ and homologues are comprised of 140–160 amino acids, with amino acid sequence identities ranging from 30 to 66%. According to the BLASTP search result, SgcJ homologues feature a domain of unknown function (DUF4440) and belong to the NTF2-like superfamily, a large group of related proteins that share a common protein fold. The NTF2-like superfamily proteins are widely found in both prokaryotic and eukaryotic organisms and possess versatile functions.26 Proteins in the NTF2-like superfamily are generally divided into two categories, enzymatically active and non-enzymatically active proteins. The former group includes enzymes with varying activities such as the ketosteroid isomerase,27 scytalone dehydrogenase28 and polyketide cyclase.29,30 The latter group includes proteins that could play roles as diverse as facilitating protein transport into the nucleus31 or mediating multimerization of calcium/calmodulin-dependent protein kinase II (CamKII),32 or may function as a receptor.33 The enediyne variants of SgcJ show less than 18% amino acid sequence identity to functionally characterized NTF2-like proteins. Owing to the diverse functions of the NTF2-like superfamily, bioinformatics analysis alone fell short of predicting the function of SgcJ and its homologues in the 9-membered enediyne core biosynthesis.

Gene inactivation reveals that sgcJ is necessary for enediyne biosynthesis
To establish a functional linkage of sgcJ and its homologues with enediyne biosynthesis, we inactivated sgcJ in the C-1027 overproducer S. globisporus SB1022,10 by replacing it with the kanamycin resistance cassette through λ-RED-mediated PCR targeting mutagenesis34 (Supplementary Figure S3a). The genotype of the resulting ΔsgcJ mutant strain SB1027 was confirmed by PCR and Southern analysis (Supplementary Figure S3b). SB1027 was fermented under the established conditions for C-1027 production with S. globisporus SB1022 as a positive control.9,10,13,35 Although C-1027 production by SB1022 was readily confirmed upon both bioassay against Micrococcus luteus and BIA, SB1027 completely abolished the production of C-1027, which was unambiguously verified by HPLC and ESI-MS analysis (Figure 2, panels I and II). The requirement for sgcJ in C-1027 biosynthesis was further supported by the fact that the ΔsgcJ mutation in SB1027 could be complemented by expressing a functional copy of sgcJ in trans, restoring C-1027 production in the
complementation strain SB1028 to the level comparable to that of SB1022 (Figure 2, panels I and III). Taken together, these data clearly established that SgcJ plays a necessary role in C-1027 biosynthesis and, by analogy, the essential role SgcJ homologues play in nine-membered enediyne core biosynthesis. However, SB1027 failed to accumulate any biosynthetic intermediate to sufficient levels for isolation and structural characterization, revealing no clues for its exact function. We therefore opted to solve the structures of SgcJ and its homologues in an attempt to elucidate their function in 9-membered enediyne core biosynthesis.

The overall structure of SgcJ and NCS-Orf16 reveals structural similarity to NTF2-like superfamily proteins

The crystals of SgcJ were obtained in the monoclinic space group C2 with unit cell parameters \(a = 72.7, b = 86.9, c = 55.3 \text{ Å} \) and \(\alpha = 90.0^\circ, \beta = 121.6^\circ\). The asymmetric unit contained two peptide chains, corresponding to a solvent content of 50.9%. The asymmetric unit also contained molecules of citric acid, glycerol, phosphate, pentaethylene glycol and tetraethylene glycol, which were present in the crystallization condition. The final model of SgcJ was refined to a resolution of 1.7 Å with an \(R\) factor of 16.9% and an \(R_{\text{free}}\) factor of 19.5%. Ramachandran analysis reveals that 99.6% of the residues were in the favored region with none in disallowed regions. Electron density map was well-defined for residues Ser3-Asp140 and Ala10-Asp140 for the two polypeptide chains in the asymmetric unit. Data collection and refinement statistics are summarized in Table 1.

The structures of SgcJ and NCS-Orf16 superimposed well with NCS-Orf16 with a root-mean-square deviation (rmsd) of 0.83 Å for the Cα atoms. The overall structures of SgcJ and NCS-Orf16 form a cone-like \(\alpha\beta\) barrel structure, which are both comprised of a long N-terminal \(\alpha\)-helix (\(\alpha_1\)-\(\alpha_2\)) passing through the curved six-stranded antiparallel \(\beta\)-sheet (\(\beta_1\)-\(\beta_6\)), with two additional shorter \(\alpha\)-helices (\(\alpha_3\) and \(\alpha_4\)) neighbor upon the \(\alpha_1\)-\(\alpha_2\) helix (Figure 3b). The \(\beta\)-sheet packs against the three \(\alpha\)-helices to form a hydrophobic core within the \(\alpha\beta\) barrel (Figure 3b). The crystal structures of SgcJ and NCS-Orf16 are packed as homodimers in an asymmetric unit, which are generated via non-crystallographic twofold axes (Figure 3c). The dimer interface is formed via a hydrogen-bonding network and salt bridges between the flat-face of the \(\beta\)-sheet from each monomer. Both SgcJ and NCS-Orf16 were indeed found to be homodimers in solution upon size exclusion chromatography (Supplementary Figure S4).

Consistent with the BLASTP search result, a search of the PDB databank using the DALI server revealed that SgcJ and NCS-Orf16 belong to the NTF2-like superfamily. This versatile superfamily is a classic example of divergent evolution wherein the proteins have similar overall structures but diverge greatly in their functions. Several crystal structures for NTF2-like superfamily proteins were reported, of which the functions have been characterized, including the association domain of CaMKII from mouse (PDB entry 1HKX), NTF2 from rat (PDB entry 1OUN), scytalone dehydratase (mgSD) from Magnaporthe grisea (PDB entry 1STD), and polyketide cyclases Snaol from Streptomyces avidinii (PDB entry 1SWJ) and Tcm ARO/CYC from Streptomyces glaucaescens (PDB entry 2RER). Despite low amino acid sequence identities, ranging from 9.8 to 17.5%, SgcJ was found to share similar folds with each of the NTF2-like superfamily proteins listed, with rmsds of 3.0, 2.6, 2.7, 2.7, 3.1 and 3.0 Å for the Cα atoms, respectively (Figure 4a). SgcJ, NTF2, KSI and Snaol form homodimers, while mgSD and CaMKII form a trimer and tetradecamer, respectively. Most importantly, all these NTF2-like superfamily proteins contain a hydrophobic pocket in the \(\alpha\beta\) barrel structure (Figure 4a), which forms a cavity that could be adapted to create an enzyme active site or a small molecule/peptide binding site, thereby serving the versatile functions.

Putative substrate binding cavity and catalytic residues of SgcJ and its homologues

KSI (isomerase), 27 mgSD (dehydratase), 28 Snaol (cyclase) 29 and Tcm ARO/CYC (cyclase) 30 are enzymatically active proteins within the NTF2-like superfamily. Although their functions are different, they share a common catalytic mechanism: (i) a general base abstracts a proton from Cα of a carbonyl group to form an enolate intermediate, which is stabilized by a general acid; (ii) the enolate intermediate...
tautomerizes back to the carbonyl group followed by double bond rearrangement or nucleophilic attack (Figure 4b). In KSI, mgSD, SnoaL and Tcm ARO/CYC, the general acid-base pairs that initiate the reactions are Asp40-Tyr16/Asp103, His85-a water bound by Tyr30 and Tyr50, Asp121-Gln105, and Tyr35-Arg69, respectively (Figure 4b). Interestingly, the crystal structures of SgcJ and NCS-Orf16 reveal conserved Asp111-Tyr72 and Asp115-Tyr76 pairs located at the entrance of the pocket, respectively (Figure 4a). Since these amino acids are known to act as the general acid-base pair in catalysis, it is tempting to speculate that SgcJ may play a similar catalytic role in transforming the nascent linear polyene intermediate along with other enediyne PKS-associated enzymes, into the 9-membered enediyne core.

Additionally, both SgcJ and NCS-Orf16 form a hydrophobic cavity within their pockets. Despite more than 50% amino acid sequence identity between SgcJ and NCS-Orf16, the amino acids lining the cavities are conserved: Trp29, Phe37, Tyr72, Trp118 and Tyr132 in SgcJ versus Trp32, Phe40, Tyr76, Trp122 and Tyr136 in NCS-Orf16 (Figures 3a and 5). Intriguingly, in the crystal structure of SgcJ, a molecule of pentaethylene glycol (1PE in chain A) and one of tetraethylene glycol (PG4 in chain B) were found bound in the cavity and surrounded by the conserved amino acid residues (Figure 5a). These polyethylene glycol molecules may mimic the binding of the linear polyene intermediate, which is sequestered and stabilized by the conserved aromatic residues lining the cavity during biosynthesis of the otherwise unstable 9-membered enediyne core intermediates.

The putative general acid-base catalytic pair and the amino acids lining the cavity are partially conserved among the SgcJ homologues in the seven known (Figure 3a) and 34 putative (Supplementary Figure S2) 9-membered enediyne biosynthetic gene clusters. To provide additional experimental data to support the catalytic role SgcJ and its homologues may play in enediyne core biosynthesis, we mutated each of the six conserved residues in SgcJ (that is, D111A and Y72A acting as the general acid-base pair, and W29A, F37A, W118A and Y132A lining the cavity) by site-directed mutagenesis. The expression constructs (pBS1148 to pBS1153) for the mutant variants...
of \textit{sgcJ} were identical to pBS1146, in which the expression of \textit{sgcJ} or its mutant variants was under the control of the constitutive \textit{ErnE}\* promoter. Introduction of pBS1148-pBS1153 individually into SB1027 afforded SB1028-SB1035, respectively, which were fermented, with SB1028 as a positive control, to examine if they could complement the Δ\textit{sgcJ} mutation in SB1027. Gratifyingly, none of six mutants restored C-1027 production (Figure 2, panels III and V-X), consistent with the proposal that these conserved residues are involved in substrate recognition, catalysis or both. Taken together, we now propose that \textit{SgcJ} plays a catalytic role in transforming the linear polyene intermediate, along with other enediyne PKS-associated enzymes, into an enzyme-sequestered 9-membered enediyne core intermediate.

\textbf{SgcJ and its homologues are pathway specific for enediyne biosynthesis}

We have previously demonstrated that PKSEs and thioesterases from different 9- and 10-membered enediyne machineries are freely interchangeable and 9- versus 10-membered enediyne core biosynthetic divergence occurs beyond the PKSE-thioesterase chemistry.\textsuperscript{21,22} Given the sequence homology among \textit{SgcJ} and its homologues, we now propose that these enzymes are specific for 9-membered enediyne core biosynthesis.

\textbf{Figure 4} Structure comparison between \textit{SgcJ} and selected homologues in the NTF2-like superfamily. (a) Each structure shows a curved antiparallel \(\beta\)-sheet wall with a group of \(\alpha\)-helices on one side of the wall to form the cone-like shapes. The putative general base and acid are shown as yellow and green sticks, respectively. The water molecule in the structure of mgSD is depicted by a red dot. Given in parentheses are PDB accession codes for each of the structures. (b) The proposed mechanisms for KSI, mgSD, SnoaL and Tcm ARO/CYC, featuring the conserved general acid-base pairs to catalyze the initial steps of reactions.
homologues (Figure 3a and Supplementary Figure S2) and the structural similarity as exemplified by SgcJ and NCS-Orf16 (Figures 3b, 4a and 5), as well as the common catalytic role proposed for SgcJ and homologues in 9-membered enediyne core biosynthesis, we finally asked if SgcJ and its homologues are pathway specific. An expression vector pBS1147 for ncs-orf16 was similarly constructed as pBS1146 for sgcJ, in which the expression of ncs-orf16 was under control of the constitutive promoter ErmE*. Introduction of pBS1147 into the ΔsgcJ mutant strain SB1027 afforded SB1029, which offered the opportunity to examine if ncs-orf16 could cross-complement the ΔsgcJ mutation in SB1027. SB1029 was fermented with SB1028 as a positive control. Cross-complementation was not observed as evidenced upon HPLC analysis that showed no C-1027 production in SB1029 (Figure 2, panels III and IV). This result would suggest that SgcJ and its homologues are pathway specific for 9-membered enediyne core biosynthesis. Close comparison of the SgcJ and NCS-Orf16 structures indeed showed subtle differences in protein surface electrostatics and the shape of the putative cavities (Figure 5), which may account for unique protein-protein interaction or accommodate varying enediyne core intermediates for different 9-membered enediyne biosynthetic machineries.

**CONCLUSIONS**

The enediyynes have served as an outstanding model to study the biosynthesis of complex natural products. Since cloning of the first set of enediyne biosynthetic gene clusters nearly 15 years ago, significant progress has been made toward elucidating the biosynthesis of the peripheral moieties present in enediyynes, but biosynthesis of the enediyne cores remains elusive. Comparative analysis of the enediyne gene clusters clearly revealed sets of genes that are highly conserved among the 9-membered, 10-membered or both enediyynes, serving as outstanding candidates to study enediyne core biosynthesis. Many of these candidate genes, however, are often annotated to encode proteins of unknown function, inactivation of which in vivo afforded mutant strains that often failed to accumulate any biosynthetic intermediates, thereby revealing few clues for their function in enediyne core biosynthesis. As a result, in spite of the great progress made in the past decade in characterizing the enediyne PKS enzyme E and its cognate thioesterase, cumulating to the discovery of a
linear heptane and its variants as the earliest possible intermediates or shunt metabolites for enediyne core biosynthesis,21–25 the exact nature of the nascent linear polyketide intermediates and their subsequent transformation to 9- and 10-membered enediyne cores remain unknown.

Recent technological advance in X-ray crystallography has made it possible to apply high-throughput structural biology as a practical tool to functionally characterize genes with deduced products that show little sequence homology to proteins of known function.39 While the current study fell short of establishing the exact function for SgcJ and its homologues may play a catalytic role, along with other enediyne function of SgcJ to C-1027 biosynthesis by site-directed mutagenesis of putative substrate binding or catalytic active site, (ii) correlate the function of SgcJ to C-1027 biosynthesis by site-directed mutagenesis of the conserved residues lining this site, and (iii) propose that SgcJ and its homologues may play a catalytic role, along with other enediyne PKS-associated enzymes, in transforming the linear polynye intermediate into an enzyme-sequestered 9-membered enediyne core intermediate. These findings will surely help formulate hypotheses and design experiments to ascertain the function of SgcJ and its homologues in 9-membered enediyne core biosynthesis in the future.

MATERIALS AND METHODS

Strains, plasmids and culture conditions

Bacterial strains, plasmids and primers used in this study are summarized in Supplementary Tables S1, S2, and S3, respectively. Escherichia coli strains and M. luteus ATCC 9431 were cultured in lysogeny broth or grown on lysogeny broth agar plates. S. globisporus wild-type and recombinant strains were cultivated at 28°C on ISP Medium 4 (Becton Dickenson, Franklin Lakes, NJ) for sporulation. Antibiotics for selection were used at the following concentrations: 25 μg ml⁻¹ for apramycin and thiostrepton, and 50 μg ml⁻¹ for chloramphenicol and kanamycin.

Construction of the ΔsgcJ mutant strain S. globisporus SB1027

The ΔsgcJ mutant strain SB1027 was constructed in the C-1027 overproducer S. globisporus SB1022 by gene replacement via homologous recombination. Briefly, the 1.3-kb kanamycin resistance cassette was amplified by PCR from pJTU4659 with primers sgcJtgtF and sgcJtgtR (Supplementary Table S3) and used to replace sgcJ in cosmids pBSI005 35 via XbaI-mediated PCR targeting mutagenesis44 to generate pBSI143. The ΔsgcJ gene was then excised from pBSI143 as a –21 kb XbaI-SphI fragment and inserted into the XbaI site of pSET151 to afford pBSI144. pBSI144 was finally introduced into S. globisporus SB1022 by E. coli-S. globisporus conjugation.40 Exconjugates resulting from the desired double-crossover homologous recombination were selected on the basis of kanamycin-resistant and thiostrepton-sensitive phenotype, and named SB1027, the genotype of which was confirmed by PCR and Southern analysis (Supplementary Figure S3).

Construction of ΔsgcJ complementation strains S. globisporus SB1028 and SB1029

A 0.8-kb fragment bearing oriT was amplified by PCR from plasmid pSET152 with primers oriT152F and oriT152R (Supplementary Table S3), digested with KpnI, and cloned into the same site of pWL201pwr to generate pBSI145. A 420-bp fragment of sgcJ and a 432-bp fragment of ncs-orf16 were amplified by PCR from cosmids pBSI00535 and pBSI00735, with primers sgcJ201NdeI and sgcJ201EcoRI, and ncs16NdeI and ncs16HindIII, respectively (Supplementary Table S3). The resultant products were digested with NdeI and EcoRI (for sgcJ), and NdeI and HindIII (for ncs-orf16), and cloned into the same sites of pBSI145 to afford pBSI146 (for sgcJ) and pBSI147 (for ncs-orf16), respectively. Both pBSI146 and pBSI147, in which the expressions of sgcJ and ncs-orf16 were under the control of the constitutive ErnE promoter,40 were finally introduced into the ΔsgcJ mutant strain S. globisporus SB1027 by E. coli-S. globisporus conjugation.40 Exconjugates were selected on the basis of thiostrepton-resistant phenotype as the desired complementation strains, and named SB1028 (that is, sgcJ expressing) and SB1029 (that is, ncs-orf16, expressing), respectively.

Site-directed mutagenesis of SgcJ

Plasmids of the sgcJ mutants, pBSI148 (W29A), pBSI149 (F37A), pBSI150 (Y72A), pBSI151 (D111A), pBSI152 (W118A) and pBSI153 (Y132A), were constructed by the QuikChange site-directed mutagenesis method, following the manufacturer’s protocol (Agilent Technologies, Santa Clara, CA) and using pBSI146 as a template. The primers used are listed in Supplementary Table S3. The mutations were verified by DNA sequencing. Each of the mutant constructs was then introduced into the ΔsgcJ mutant strain SB1027 by conjugation, yielding the complementation strains SB1030 (that is, SB1027/pBSI148), SB1031 (that is, SB1027/pBSI149), SB1032 (that is, SB1027/pBSI150), SB1033 (that is, SB1027/pBSI151), SB1034 (that is, SB1027/pBSI152) and SB1035 (that is, SB1027/pBSI153), respectively.

Production, isolation and analysis of C-1027

S. globisporus recombinant strains were cultured following a two-step fermentation procedure reported previously, and both stages utilized the same medium (1% glycerol, 2% dextrin, 1% fish meal, 0.5% peptone, 0.2% (NH₄)₂SO₄, 0.1% MgSO₄, 0.2% CaCO₃, pH 7.0).39,10,13,35 Briefly, fresh spores of the recombinant strains were inoculated into 250-ml baffled flasks containing 50 ml of medium and incubated at 28°C and 250 rpm for 48 h. The resultant seed cultures (2.5 ml) were then inoculated into 250-ml baffled flasks containing 50 ml of the same medium, and fermentation continued at 28°C and 250 rpm for 7 days. The C-1027 overproducer SB1022 and the ΔsgcJ mutant strain SB1027 were cultured in medium without any antibiotics. All other recombinant strains used in this study were cultured in medium supplemented with 5 μg ml⁻¹ thiostrepton to retain the introduced plasmids.

Isolation and HPLC analysis of the C-1027 chromophore were carried out by following published procedures.9,10,13,35 Briefly, fermentation broth (50 ml) was adjusted to pH 4.0 with 0.1 N HCl and centrifuged to remove any precipitate. To the supernatant, (NH₄)₂SO₄ was then added to 50% saturation, and the precipitated C-1027 chromoprotein was collected by centrifugation and re-dissolved in CH₃OH. HPLC was carried out on a Beckman ultrasphere-ODS dp analytical column (5 μm, 150 × 4.6 mm) (Beckman Coulter, Indianapolis, IN), eluted isocratically with 20 mM potassium phosphate (pH 6.8)/CH₃CN (50:50 v/v) at a flow rate of 1.0 ml min⁻¹ and UV detection at 350 nm on a Varian HPLC system with a Prostar 330 PDA detector (Agilent Technologies). LC-MS analysis of C-1027 was performed on an Agilent 6230 TOF LC-MS instrument (Agilent Technologies).

** Determination of C-1027 production by bioassay and biochemical induction assays

Determination of C-1027 production by bioassay against M. luteus ATCC 9431 was carried out as described previously.9,10,13,35 Alternatively, C-1027 production was also followed by the biochemical induction assay according to literature procedures,12 which uses the E. coli BR513 strain as an indicator and specifically detects agents with DNA damage activities. Briefly, 10 μl of fermentation supernatant or an agar plug were applied onto agar plates seeded with E. coli BR513 and incubated for 3–4 h at 37°C. The plates were then overlaid with soft agar containing 0.7 mg ml⁻¹ of X-gal and incubated at 37°C for additional 30–60 min to develop the characteristic blue color, indicative of DNA damage, and thus C-1027 production.

**Gene expression and protein purification

PCR amplification of sgcJ from S. globisporus genomic DNA13 and ncs-orf16 from S. carzinostaticus genomic DNA18 by KOD Hot Start DNA polymerase (EMD Millipore, Billerica, MA) followed the manufacturer’s protocols using primers sgcJ-F and sgcJ-R and primers orf16-F and orf16-R primers, respectively (Supplementary Table S3). The amplification buffer was supplemented with betaine to a final concentration of 2.5 M. The PCR products were
purified and cloned into pMCSG57, yielding pBS1154 (expressing sgcJ) and pBS1115 (expressing ncs-orf16), by the ligation-independent procedures.\(^1\) The expression plasmids were then transformed into E. coli BL21(DE3)-Gold strain (Stratagene, San Diego, CA) for protein production. Production and purification of SgcJ and NCS-Orf16 were performed according to standard protocol.\(^2\) Briefly, the cell were cultured at 37 °C in 1 L of enriched M9 medium\(^3\) until OD\(_{600}\) = 1.0. After air-cooling the culture down to 4 °C for 60 min, inhibitory amino acids (25 mg each per liter L-valine, L-isoleucine, L-leucine, L-lysine, L-threonine and L-phenylalanine), selenomethionine (SeMet) and isopropyl-β-D-thiogalactoside (IPTG) were added. The cells were incubated overnight at 18 °C, harvested and re-suspended in lysis buffer (500 mM NaCl, 5% (v/v) glycerol, 50 mM HEPES pH 8.0, 20 mM imidazole and 10 mM β-mercaptoethanol). The SeMet-labeled proteins were purified using Ni-NTA affinity chromatography by the AKTAdress system (GE Healthcare Life Sciences, Marlborough, MA) and digested with recombinant His6-tagged Tobacco etch virus (TEV) protease to remove the His6-tag. The final pure proteins were concentrated using Amicon Ultra-15 concentrators (Millipore, Bedford, MA) in 20 mM HEPES pH 8.0, 250 mM NaCl and 2 mM dithiothreitol. Protein concentrations were determined based on the absorbance at 280 nm using a molar absorption coefficient (ε\(_{280}\)) = 19,480 and 20,970 M\(^{-1}\) cm\(^{-1}\) for SgcJ and NCS-Orf16, respectively.\(^4\) The concentrations of SgcJ and NCS-Orf16 used for crystallization were both ~ 50 mg mL\(^{-1}\).

### Protein crystallization

Both SgcJ and NCS-Orf16 were screened for crystallization conditions using a Mosquito liquid dispenser (TTP Labtech, Melbourn, UK) and the sitting-drop vapor-diffusion technique in 96-well CrystalQuick plates (Greiner Bio-one, Monroe, NC). For each condition, 0.4 μL of protein (52.8 mg mL\(^{-1}\)) and 0.1 μL of the reservoir in the well. Commercially available crystallization screens (MCSG-1 (Microlytic Inc., Burlington, MA) at 24 °C, 16 °C and 4 °C. For SgcJ, crystals were obtained under several conditions, with the best result coming from 0.1 M Na\(_2\)HPO\(_4\) (adjust to pH 4.2 with citric acid) and 40% (v/v) PEG 300 at 16 °C. The crystals grew within 1 week and reached sizes of approximately 0.100 mm × 0.020 mm × 0.010 mm. For NCS-Orf16, suitable crystals for X-ray diffraction were grown from the condition containing 0.2 M sodium formate and 20% (w/v) PEG 3350 at 16 °C.

### Data collection, structure determination and refinement

Diffraction data were collected at 100 K at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory.\(^5\) A single data set was taken near the Se K-edge peak anomalous position (0.9792 Å) from a single protein crystal of SgcJ to a resolution of 1.7 Å. The crystals were indexed, integrated and scaled to 1.7 and 2.72 Å, respectively. For SgcJ, the initial model contains two protein chains consisting of at least 90% of the residues in each chain. For NCS-Orf16, the initial model contains 10 protein chains consisting of 67% of whole model with 20% assigned side-chain. Extensive manual model building with COOT\(^6\) and the subsequent refinement using phenix.refine\(^5\) were performed using R factors converged to final values of R\(_{\text{free}}\) = 0.168(0.195) and 0.217(0.256) for the structures of SgcJ and NCS-Orf16, respectively. The geometrical properties of the models were assessed using PROCHECK\(^7\) and Molprobity.\(^8\) The atomic coordinates and structure factors have been deposited in the Protein Data Bank with the accession code 4I4K for SgcJ and 4OVF for NCS-Orf16, respectively.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

This work is supported in part by a fellowship of Academia Sinica-The Scripps Research Institute Postdoctoral Talent Development Program (to C-YC), a German Research Foundation postdoctoral fellowship (to IC), US National Institute of General Medical Science Protein Structure Initiative Grants GM094585 (to AJ) and GM098248 (to GNP) and US National Institutes of Health Grants GM109456 (to GNP), CA078747 (to BS) and GM115575 (to BS). The use of Structural Biology Center beamlines at the Advanced Photon Source was supported in part by the US Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

### References

1. Van Laren, S. G. & Shen, B. Biosynthesis of edenediyne anticancer antibiotics. Curr. Top. Med. Chem. 8, 448–459 (2008).
2. Liang, Z. X. Complexity and simplicity in the biosynthesis of enediyne natural products. Bioorg. Med. Chem. Lett. 7, 3–0 (2017).
3. Kennedy, D. R., Ju, J., Shen, B. & Beerman, T. A. Designer enediynes generate DNA breaks, interstrand cross-links, or both, with concomitant changes in the regulation of ataxia-telangiectasia mutated in cellular responses to DNA double-strand breaks. Cancer Res. 67, 773–781 (2007).
4. Kennedy, D. R., Ju, J., Shen, B. & Beerman, T. A. Design of new enediyne anticancer drugs. J. Ind. Microbiol. Biotechnol. 43, 261–276 (2016).
5. Edo, K. et al. The structure of neocarzinostatin chromophore possessing a novel bicyclo[7,3,0]dodecadiyne system. Tetrahedron Lett. 46, 331–340 (2005).
6. Kennedy, D. R. et al. Single chemical modifications of the C-1027 enediyne core, a radiomimetic anticancer drug, affect both drug potency and the role of ataxia-telangiectasia mutated in cellular responses to DNA double-strand breaks. Cancer Res. 67, 773–781 (2007).
7. Kennedy, D. R., Ju, J., Shen, B. & Beerman, T. A. Designer enediynes generate DNA breaks, interstrand cross-links, or both, with concomitant changes in the regulation of DNA damage responses. Proc. Natl Acad. Sci. USA 107, 16176–16177 (2007).
8. Beerman, T. A., Gawon, L. S., Shin, S., Shen, B. & McHugh, M. M. C-1027, a radiomimetic enediyne anticancer drug, preferentially targets hypoxic cells. Cancer Res. 69, 593–598 (2009).
9. Chen, Y., Yin, M., Horsman, G. P. & Shen, B. Manipulation of pathway regulation in Streptomyces globisporus for overproduction of the enediyne antibiotic C-1027. J. Antibiot. 63, 482–486 (2010).
10. Chen, Y., Yin, M., Horsman, G. P. & Shen, B. Improvement of the enediyne antibiotic C-1027 production by manipulating its biosynthetic pathway regulation in Streptomyces globisporus. J. Nat. Prod. 74, 420–424 (2011).
11. Zazopoulos, E. et al. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat. Biotechnol. 21, 187–190 (2003).
12. Liu, W. et al. Rapid PCR amplification of minimal enediyne polycyclic synthase cassettes leads to a predictive familial classification model. Proc. Natl Acad. Sci. USA 100, 11999–11963 (2003).
13. Liu, W., Christensen, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antibiotic C-1027. Science 297, 1170–1173 (2002).
14. Ahlert, J. et al. The calicheamicin gene cluster and its iterative type I enediyne PKS. Science 297, 1173–1176 (2002).
15. Liu, W. et al. The neocarzinostatin biosynthetic gene cluster from Streptomyces caeziasomaticus ATCC 15944 involving two iterative type I polycyclic synthases. Chem. Biol. 12, 293–302 (2005).
16. Van Laren, S. G., Oh, T. J., Liu, W., Wende-Pienkowski, E. & Shen, B. Characterization of the maduropeptin biosynthetic gene cluster from Actinomadura madurae ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. J. Am. Chem. Soc. 129, 13082–13094 (2007).
17. Lohman, J. R. et al. Cloning and sequencing of the kerdacidin biosynthetic gene cluster from Streptos specialistus sp. ATCC 353650 revealing new insights into biosynthesis of the enediyne family of anticancer antibiotics. Mol. BioSyst. 9, 478–491 (2013).
18. McGlinchey, R. P. et al. Unraveling the biosynthesis of the sporolide cyclohexeneone building block. J. Am. Chem. Soc. 130, 2406–2407 (2008).
19. Lane, A. L. et al. Structures and comparative characterization of biosynthetic gene clusters for cyanosporides, enediyne-derived natural products from marine actinomycetes. J. Am. Chem. Soc. 135, 4171–4174 (2013).
20. Gao, O. & Thorsen, J. S. The biosynthetic genes encoding for the production of the dynemicin enediyne core in Micromonospora chersina ATCC53710. FEMS Microbiol. Lett. 282, 105–114 (2008).
21. Zhang, J. et al. A phosphopantetheinylation polyketide synthase producing a linear polypeptide to initiate enediyne antitumor antibiotic biosynthesis. *Proc. Natl Acad. Sci. USA* **105**, 1460–1465 (2008).

22. Horsman, G. P., Chen, Y., Thorson, J. S. & Shen, B. Polyketide synthase chemistry does not direct biosynthetic divergence between 9- and 10-membered enediyne. *Proc. Natl Acad. Sci. USA* **107**, 11331–11335 (2010).

23. Belecki, K., Crawford, J. M. & Townsend, C. A. 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 (2009).

24. Belecki, K. & Townsend, C. A. Biochemical determination of enzyme-nound metabo-

25. Belecki, K. & Townsend, C. A. Biochemical determination of enzyme-nound metabo-

26. Eberhardt, R. Y. et al. Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics* **14**, 327 (2013).

27. Cha, H. J. et al. Rescue of deltererious mutations by the compensatory Y30F mutation in ketosteroid isomerase. *Mol. Cells* **36**, 39–46 (2013).

28. Lundqvist, T. et al. Crystal structure of scytalone dehydratase—a disease determinant of the rice pathogen, *Magnaporthe grisea*. *Structure* **2**, 937–944 (1994).

29. Sultana, A. et al. Production of octaketide polyenes by the *Streptomyces globisporus* are clustered with the cagA gene that encodes the C-1027 apoprotein. *Antimicrob. Agents Chemother.* **48**, 382–392 (2000).

30. Ames, B. D. Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for the biosynthesis of enediyne antitumor antibiotic biosynthesis. *J. Mol. Biol.* **260**, 422–431 (1996).

31. Wang, J. W. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **28** (1999).

32. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **39**, 363–371 (2014).

33. Ott, M. Characterization of the polyketide cyclase SnoaL reveals a novel mechanism for isomerization of ketosteroid isomerase. *Mol. Cells* **37**, 382–392 (2000).

34. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted cloning of *Mba1*, a membrane-associated ribosome receptor in mitochondria. *J. Mol. Biol.* **431**, 431–441 (2013).

35. Liu, W. et al. Homology modeling of the calicheamicin polyketide synthase CalE8: implications for the biosynthesis of enediyne antitumor antibiotic biosynthesis. *J. Struct. Biol.* **181**, 319–326 (1989).

36. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Acta Cryst.* **62**, 859–866 (2006).

37. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).

38. Bulllock, T. L., Clarkson, W. D., Kent, H. M. & Stewart, M. The 1.6 angstroms resolution structure of nuclear transport factor 2 (NTF2). *J. Mol. Biol.* **260**, 422–431 (1996).

39. Jacobson, M. P., Fava, A., Zhao, S. & Tian, B. Leveraging structure for enzyme function prediction: methods, opportunities, and challenges. *Trends Biochem. Sci.* **39**, 363–371 (2014).

40. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical Strepptomycetes Genetics (The John Innes Foundation, Norwich, UK, 2000).

41. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074 (1990).

42. Kim, Y. et al. Automation of protein purification for structural genomics. *J. Struct. Funct. Genomics* **5**, 111–118 (2004).

43. Minor, W. et al. Phosphopantetheinylation polyketide synthase producing a linear polypeptide to initiate enediyne antitumor antibiotic biosynthesis. *Proc. Natl Acad. Sci. USA* **105**, 5349–5354 (2008).

44. Rosenbaum, G. et al. The 1.6 angstroms resolution structure of isomorphous replacement and anomalous scattering, proceedings of the CCP4 study meeting (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.), 80–86 (1991).

45. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Acta Cryst.* **D66**, 479–485 (2010).

46. Otwinowski, Z. in Isomorphous Replacement and Anomalous Scattering, Proceedings of the CCP4 Study Meeting (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.), 80–86 (1991).

47. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Acta Cryst.* **D69**, 31–38 (1994).

48. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).

49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126–2132 (2004).

50. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **D66**, 213–221 (2010).

51. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).

52. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **33**, W375–W387 (2005).

53. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **38**, 2138–2148 (2010).

54. Kim, Y. et al. Automation of protein purification for structural genomics. *J. Struct. Funct. Genomics* **5**, 111–118 (2004).

55. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **33**, W375–W387 (2005).

56. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **38**, 2138–2148 (2010).

57. Kim, Y. et al. Automation of protein purification for structural genomics. *J. Struct. Funct. Genomics* **5**, 111–118 (2004).

58. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **33**, W375–W387 (2005).

59. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **38**, 2138–2148 (2010).

60. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **33**, W375–W387 (2005).