Supporting information

Biosynthesis of chlorinated lactylates in *Sphaerospermopsis* sp. LEGE 00249

Kathleen Abt,[†,▽] Raquel Castelo-Branco[†] and Pedro N. Leão[†,*]

[†]*Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, Avenida General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal*

[▽]*Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Rua de Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal*

**author to whom correspondence should be addressed; E-Mail: pleao@ciimar.up.pt*

Contents

	page
Figure S1. In silico analysis of ClyE DH, KS and AT domains.	SI 3
Figure S2. Bioinformatic analysis of ClyB, ClyE and ClyF docking domains.	SI 4
Figure S3. Production of **1**-**4** by *Anabaena* sp. PCC 7108.	SI 5
Figure S4. Supplementation of *Sphaerospermopsis* sp. LEGE 00249 with deuterated fatty acids leads to incorporation of deuterium labels from d_{15}-octanoic, d_{19}-decanoic and d_{23}-dodecanoic acids into **2**/**3** and **4**.	SI 6
Figure S5. LC-HRESIMS-based detection of putative monochlorinated chlorosphaerolactylates analogues with C_{10} acyl chains in organic extracts of *Sphaerospermopsis* sp. LEGE 00249 cultures supplemented with decanoic acid or d_{19}-decanoic acid.	SI 7
Figure S6. [U-^{13}C]pyruvate isotopic incorporation experiment (50 hours exposure).	SI 8
Figure S7. Schematic representation of the proposed ^{13}C-label scrambling leading to different labeling patterns in the chlorosphaerolactylates upon supplementation of *Sphaerospermopsis* sp. LEGE 00249 with [U-^{13}C]pyruvate.	SI 9
Figure S8. Homology modeling of ClyF – model details	SI 10
Figure S9. Comparison of the homology model of ClyF and the crystal structure of StsB: key residues and pseudo A_{sub} domain.	SI 11
Figure S10. Output of the antiSMASH analysis for the KR domain of ClyF, highlighting the predicted configuration of the reduced product.	SI 12
References	SI 13

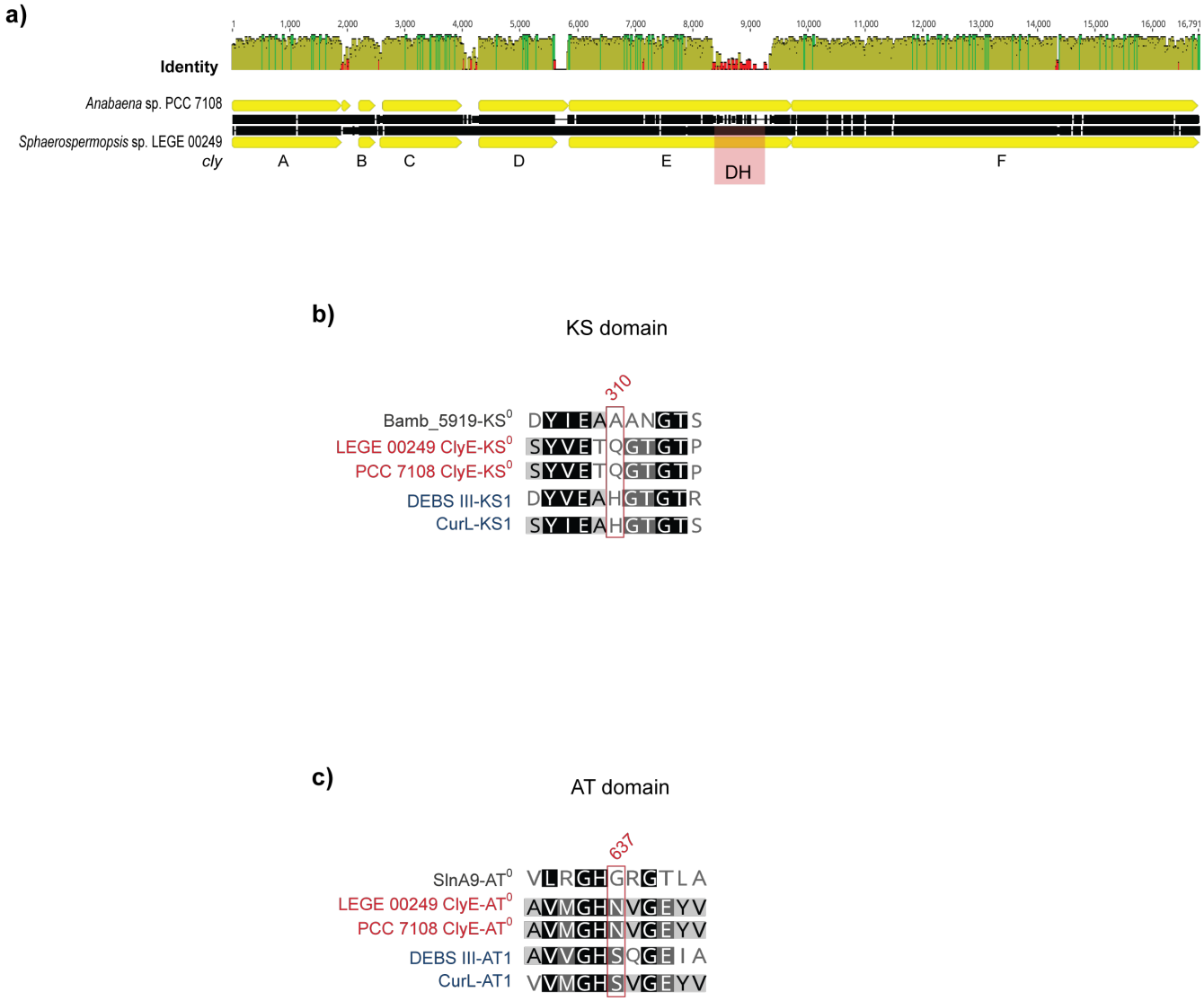

Figure S1. In silico analysis ClyE DH, KS and AT domains. Shown are a) Alignment of the *cly* locus from *Anabaena* sp. PCC 7108 and *Sphaerospermopsis* sp. LEGE 00249, highlighting the overall high identity and the lack of conservation in the region of *clyE* encoding the DH domain in the latter strain (a) and alignments with characterized KS/KS^0 (b) or AT/AT^0 (c) domains, highlighting the presence/absence of active site His (KS domains) and Ser (AT domains). Numbering in (b) and (c) corresponds to ClyE from *Sphaerospermopsis* sp. LEGE 00249.

a)

ClyB (ACP)	Top hits DDAP database C-terminal "heads"		
#identities	BGC linker ID	Corresponding proteins	Type of connection
41	BGC_27_LINKER_2_HC	JamC to JamE	(FAAL) ACP - PKS
20	BGC_84_LINKER_4_HC	OlmA4 to OlmA5	PKS - PKS
18	BGC_30_LINKER_4_HC	ChmGIV to ChmGV	PKS - PKS
18	BGC_82_LINKER_5_HC	HlsE to HlsF	PKS - PKS
17	BGC_37_LINKER_6_HC	SlnA6 to SlnA7	PKS - PKS
17	BGC_90_LINKER_1_HC	StmA to StmB	PKS - PKS
16	BGC_81_LINKER_2_HC	MakA2 to MakA3	PKS - PKS
16	BGC_28_LINKER_2_HC	EryAII to EryAIII	PKS - PKS
16	BGC_26_LINKER_4_HC	BorA4 to BorA5	PKS - PKS
16	BGC_84_LINKER_2_HC	OlmA2 to OlmA3	PKS - PKS

ClyE-N50 (PKS-KS)	Top hits DDAP database N-terminal "tails"		
#identities	BGC linker ID	Corresponding proteins	Type of connection
30	BGC_27_LINKER_2_TN	JamC to JamE	(FAAL) ACP - PKS
20	BGC_13_LINKER_2_TN	StiB to StiC	PKS - PKS
18	BGC_27_LINKER_7_TN	JamL to JamN	NRPS - PKS
17	BGC_49_LINKER_1_TN	RevC to RevA	PKS - PKS
16	BGC_27_LINKER_10_TN	JamO to JamP	NRPS - PKS
15	BGC_29_LINKER_6_TN	CurF to CurG	PKS - PKS
15	BGC_9_LINKER_2_TN	MxaE to MxaD	PKS - PKS
14	BGC_46_LINKER_3_TN	PieA3 to PieA4	PKS - PKS
14	BGC_3_LINKER_2_TN	EpoB to EpoC	NRPS - PKS
13	BGC_13_LINKER_4_TN	StiD to StiE	PKS - PKS

ClyE-C100 (PKS-T)	Top hits DDAP database C-terminal "heads"		
#identities	BGC linker ID	Corresponding proteins	Type of connection
28	BGC_27_LINKER_9_HC	JamN to JamO	PKS - NRPS
25	BGC_29_LINKER_12_HC	CurL to CurM	PKS - PKS
24	BGC_3_LINKER_1_HC	EpoA to EpoB	PKS - NRPS
23	BGC_81_LINKER_2_HC	MakA2 to MakA3	PKS - PKS
22	BGC_27_LINKER_5_HC	JamJ to JamK	PKS - PKS
21	BGC_9_LINKER_2_HC	MxaE to MxaD	PKS - PKS
21	BGC_27_LINKER_6_HC	JamK to JamL	PKS - PKS
21	BGC_29_LINKER_9_HC	CurI to CurJ	PKS - PKS
21	BGC_83_LINKER_2_HC	PldAII to PldAIII	PKS - PKS
20	BGC_2_LINKER_3_HC	MtaD to MtaE	PKS - PKS

ClyF-N50 (NRPS-C)	Top hits DDAP database N-terminal "tails"		
#identities	BGC linker ID	Corresponding proteins	Type of connection
19	BGC_3_LINKER_1_TN	EpoA to EpoB	PKS - NRPS
19	BGC_27_LINKER_9_TN	JamN to JamO	PKS - NRPS
18	BGC_2_LINKER_5_TN	MtaF to MtaG	PKS - NRPS
17	BGC_21_LINKER_5_TN	MelF to MelG	PKS - NRPS
16	BGC_32_LINKER_5_TN	CtaF to CtaG	PKS - NRPS
14	BGC_21_LINKER_1_TN	MelB to MelC	PKS - NRPS
13	BGC_54_LINKER_1_TN	LobR to LobA	PKS - PKS
13	BGC_32_LINKER_1_TN	CtaB to CtaC	PKS - NRPS
12	BGC_1_LINKER_3_TN	AveA3 to AveA4	PKS - PKS
12	BGC_87_LINKER_2_TN	RdmH to RdmI	PKS - PKS

b)

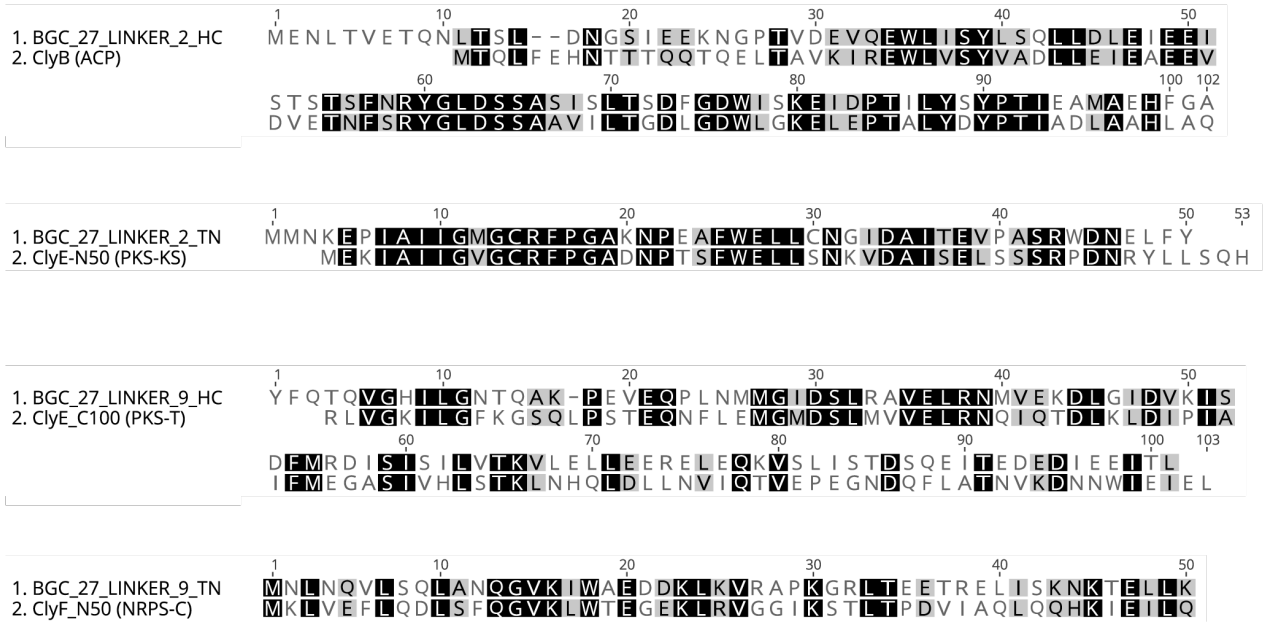

Figure S2. Bioinformatic analysis of ClyB, ClyE and ClyF docking domains. a) Distance table from MAFFT alignments of ClyB, the 50 N-terminal amino acids of ClyE, the 100 C-terminal amino acids of ClyE and the 50 N-terminal amino acids of ClyF with the respective (head or tail) 382 docking domain sequences currently deposited in the DDAP database[1] b) Sequence alignments of ClyB, ClyE and ClyF docking domains with the respective most similar docking domain from the DDAP database.

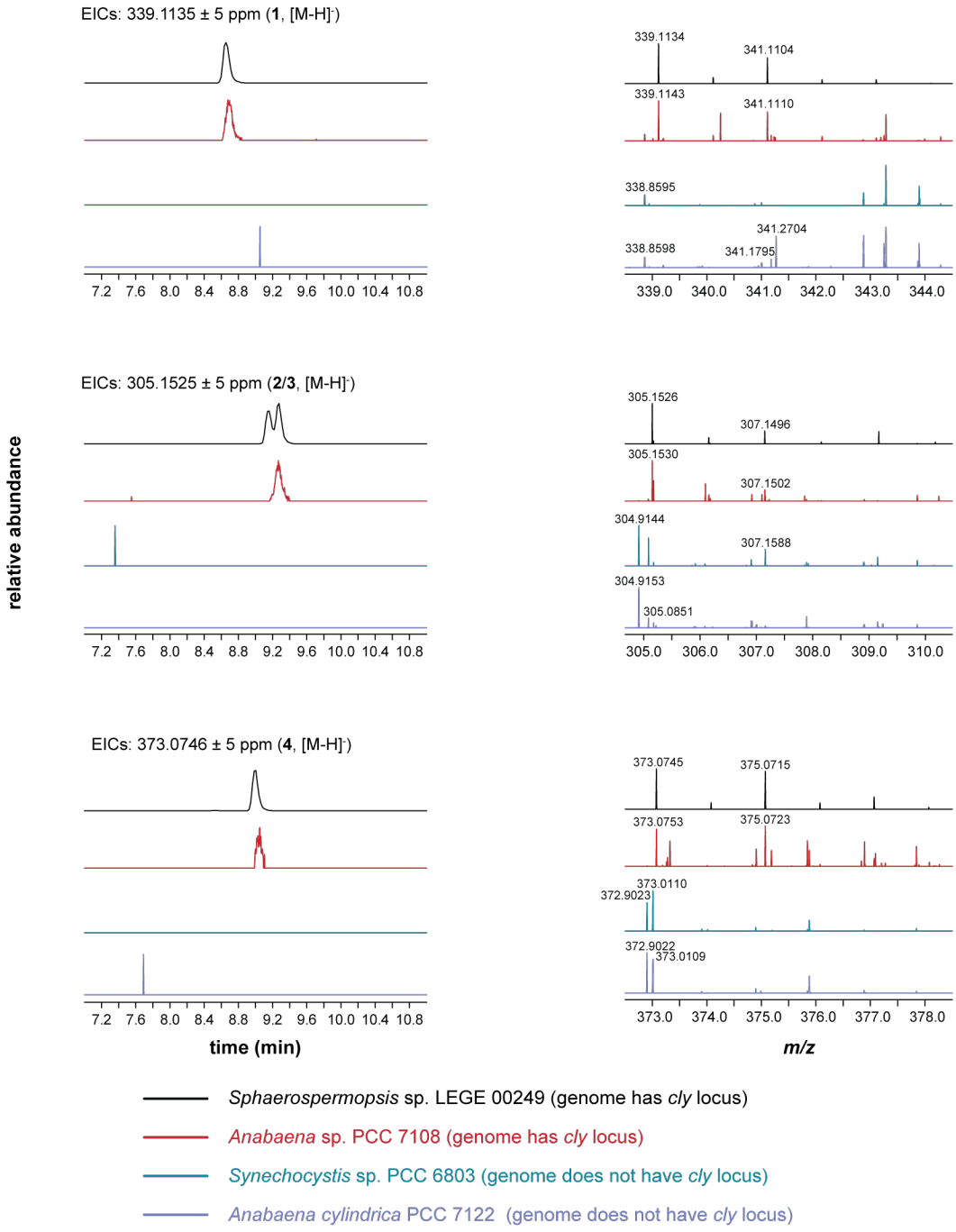

Figure S3. Production of **1-4** by *Anabaena* sp. PCC 7108. LC-HRESIMS analysis of **1-4** in organic (CH_2Cl_2/MeOH, 2:1 v/v) extracts of cyanobacterial strains that contain the *cly* locus in their genomes (*Sphaerospermopsis* sp. LEGE 00249 and *Anabaena* sp. PCC 7108) and strains that do not contain the *cly* locus (strains with full genome data available, *Synechocystis* sp. PCC 6803 and *Anabaena cylindrica* PCC 7122). Shown are extracted ion chromatograms (EICs) for the $[M-H]^-$ ions of the different compounds, as well as the corresponding region of the MS spectrum.

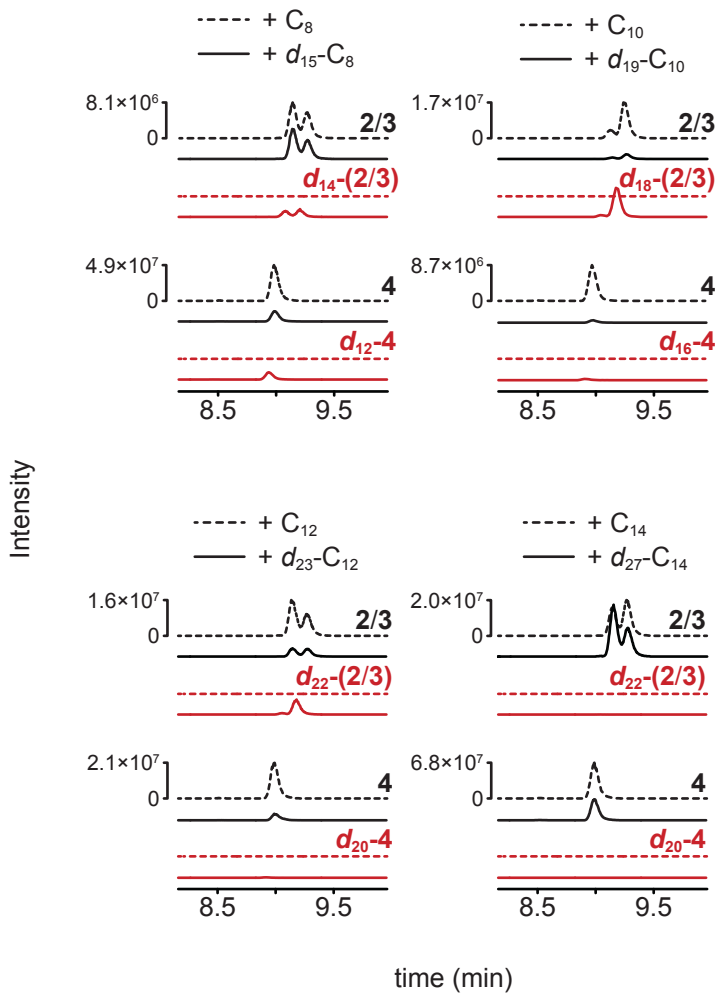

Figure S4. Supplementation of *Sphaerospermopsis* sp. LEGE 00249 with deuterated fatty acids leads to incorporation of deuterium labels from d_{15}-octanoic, d_{19}-decanoic and d_{23}-dodecanoic acids into **2**/**3** and **4**.

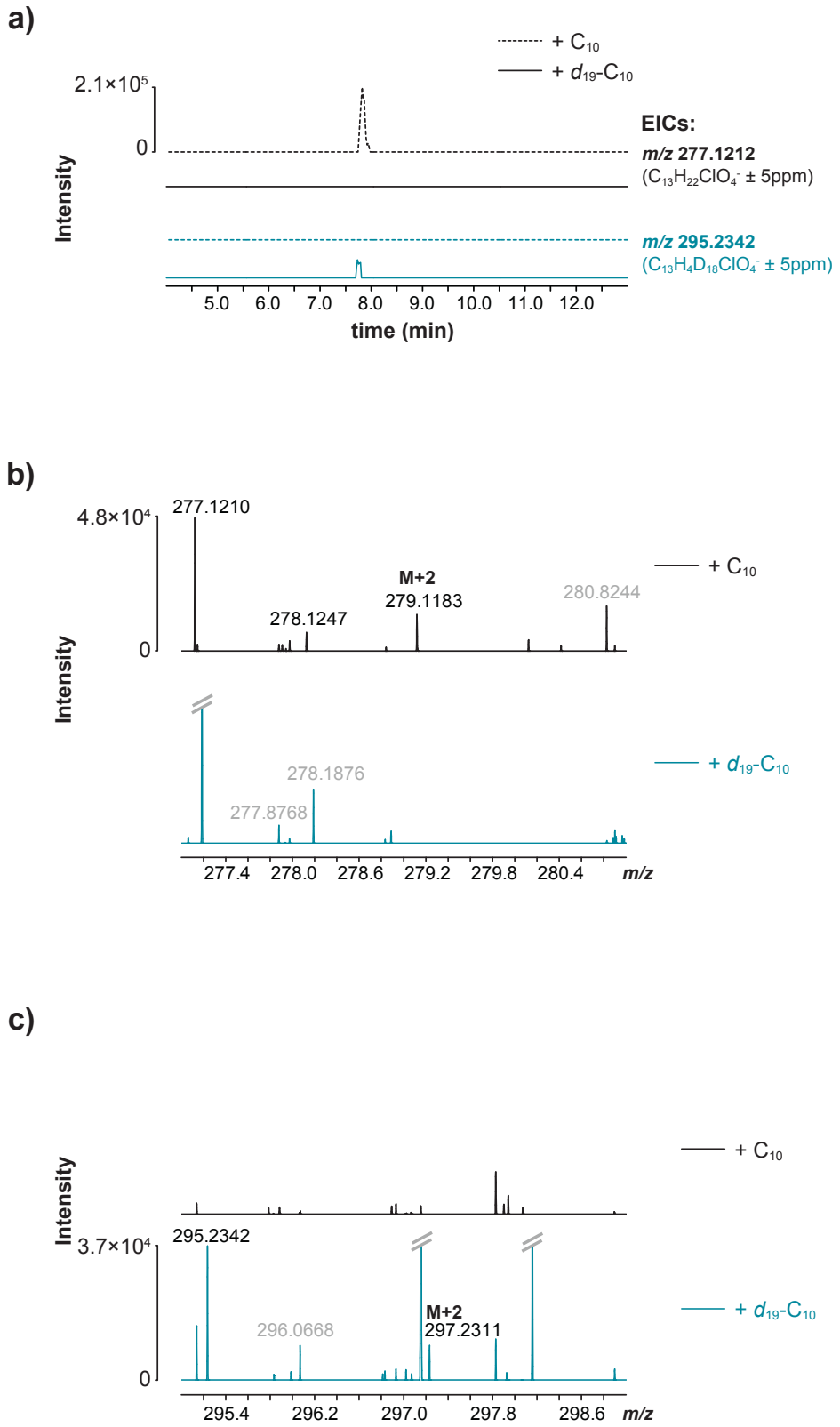

Figure S5. LC-HRESIMS-based detection of putative monochlorinated chlorosphaerolactylates analogues with C_{10} acyl chains in organic extracts of *Sphaerospermopsis* sp. LEGE 00249 cultures supplemented with decanoic acid or d_{19}-decanoic acid. Extracted ion chromatograms (EICs) for *m/z* values corresponding to analogues of compounds **2** or **3**, with a decanoic acid-derived monochlorinated acyl chain (a). Spectral regions containing the isotope clusters for the $[M-H]^-$ species of the shorter-chain derivative of **2** or **3** (b) or its d_{18} isotopologue (c).

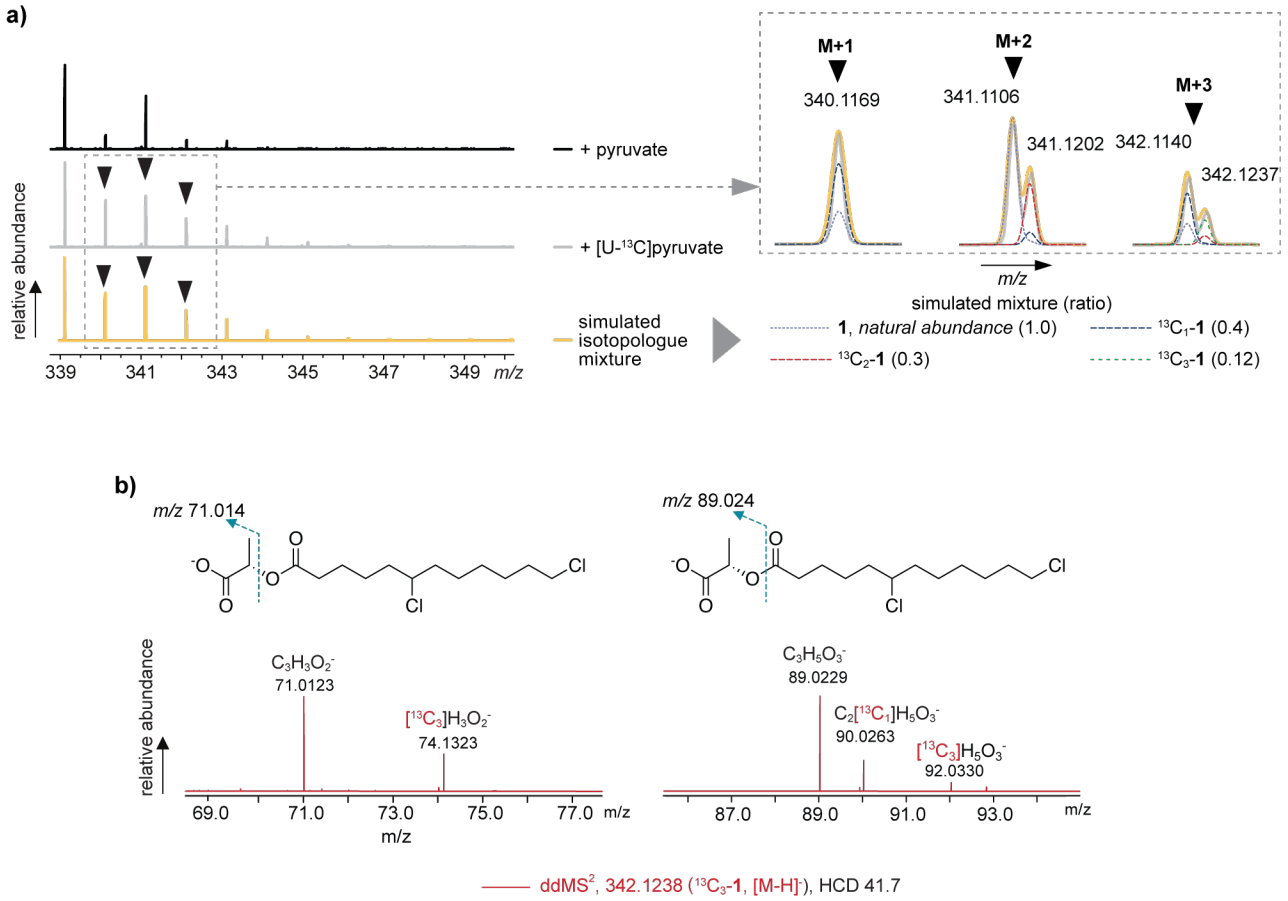

Figure S6. [U-^{13}C]pyruvate isotopic incorporation experiment (50 h exposure). a) Full MS (from LC-HRESIMS analysis) spectral regions for the isotopic cluster of the [M-H]$^-$ ion of **1**, and a spectrum simulation matching the M+1, M+2 and M+3 peak fine structures. b) LCHRESIMS/MS analysis of $^{13}C_3$-**1**, depicting key fragments which contain only the carbon atoms that are not derived from dodecanoic acid.

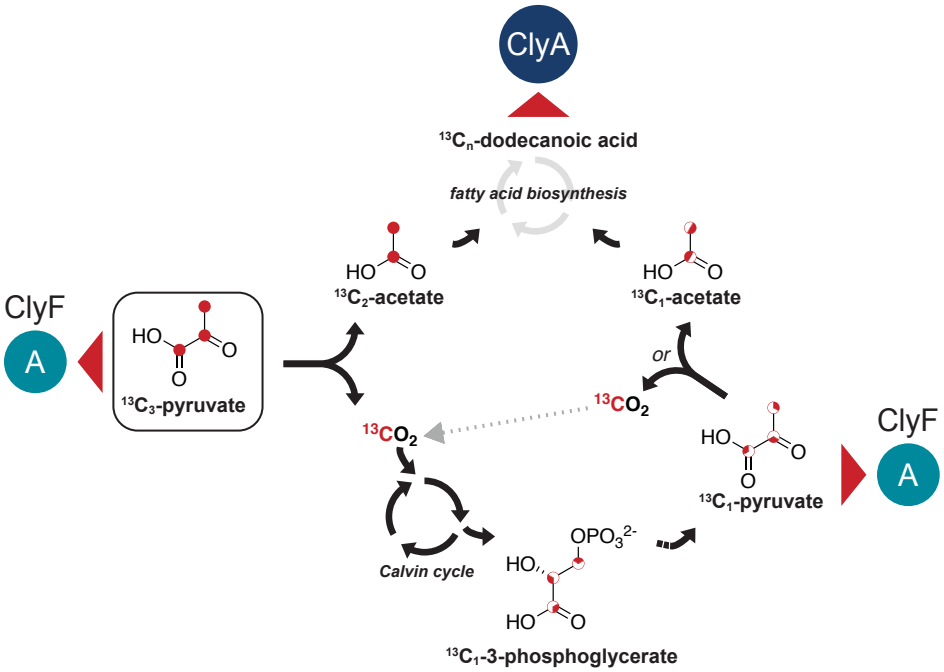

Figure S7. Schematic representation of the proposed ^{13}C-label scrambling leading to different labeling patterns in the chlorosphaerolactylates upon supplementation of *Sphaerospermopsis* sp. LEGE 00249 with [U-^{13}C]pyruvate.

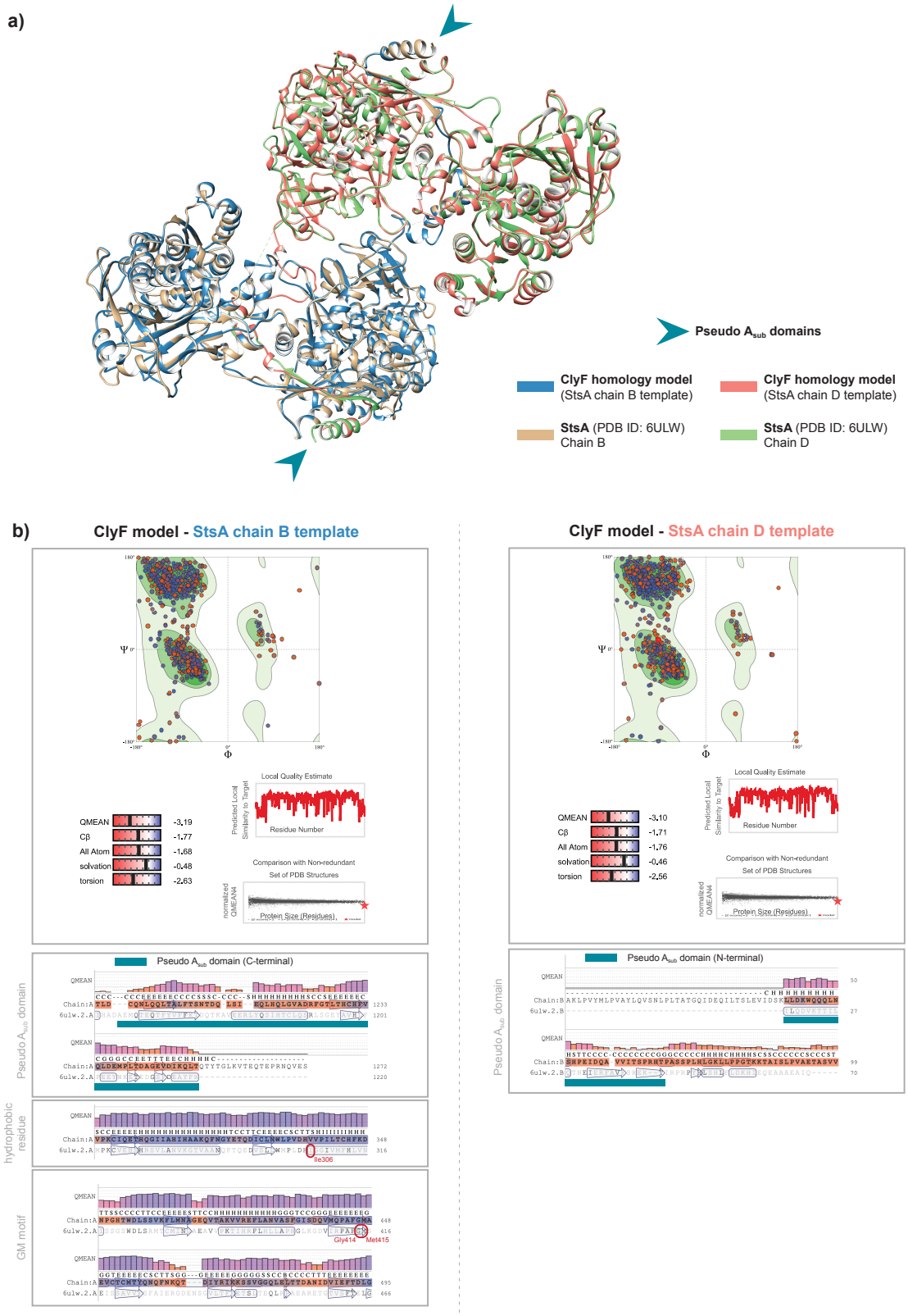

Figure S8. Homology modeling of ClyF. a) Superimposed view of the homology model of ClyF and the template used for the modeling (StsA, chains B and D), highlighting the positioning of the split pseudo A_{sub} domains. b) Global and local quality estimates for the models of each chain (Ramachandran plot, QMEAN estimates); highlighted are the regions associated specifically with depsipeptide synthetases (pseudo Asub domain, hydrophobic residue replacing amino acid-contacting Asp, GM motif)[2].

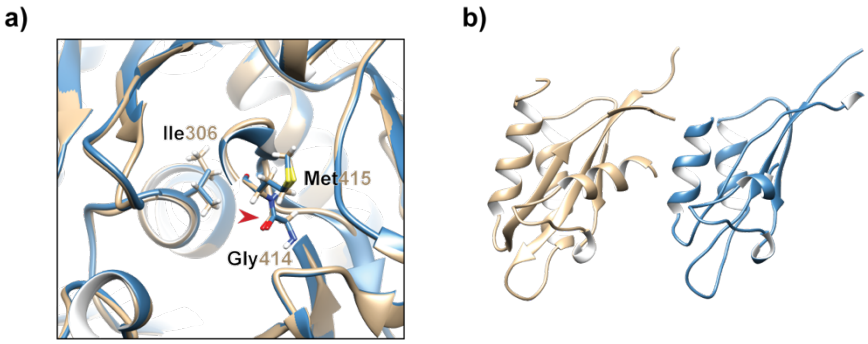

Figure S9. Comparison of the homology model of ClyF and the crystal structure of StsB: key residues and pseudo A_{sub} domain. a) Superimposed view of the crystallographic model of StsA (beige) and the homology model obtained for ClyF (blue), highlighting the position of the key residues outlined above. Residue numbering corresponds to StsA. The red arrow depicts the amide carbonyl previously proposed to interact with the alpha-keto group of the substrates. b) Comparison of the pseudo Asub domain, characteristic of depsipeptide synthetases, in the crystallographic model of StsA (beige) and the homology model of ClyF (blue).

Name: ctg1_166_KR1
Type: aSDomain
Length: 196
Interval: 1,608 -> 1,803
Residues: GVYLITGGLGGIGVKIAKYLL..
domain: PKS_KR
locus_tag: ctg1_166
detection: hmmscan
database: nrpspksdomains.hmm
evalue: 1.30 × 10-46
score: 150.4
Transferred Translation: GVYLITGGLGGIGVKIAKYLLENYQARLLIGRTPLPDESSGKNINAGEDQLTAKIRAYQELRQL
PGSVIYQDVDICNLDDMKKTLDLLSSEWQTQFDGVIHLAGVLQEQLITSASPESLIAGLQQKVMGTWVLHHLLQDQNHGFFIHFSSVNSF
FGGTGVSAYAAANSFQEAFSAYQRQHSYGNSYCLSWSMWDE
asDomain_id: nrpspksdomains_ctg1_166_KR1
specificity: KR activity: active; KR stereochemistry: ?
aSASF_scaffold: Scaffold coordinates: (149, 162, 166); Scaffold coordinates: (33,65,92,97,120,148); emission probability array (n.d.,n.d.,n.d.); emission probability array (n.d.,n.d.,n.d.,n.d.,n.d.,n.d.); expected: (R,D,G,A,K,S); expected: (S,Y,N); matchArray: (True,True,True); matchArray: (True,True,True,True,True,True); overall match: TRUE; scaffold residues: (R,D,G,A,K,S); scaffold residues: (S,Y,N)
aSASF_choice: Description: Active site S,Y,N catalytic triade, choice result: catalytic triade S,Y,N found, choice coordinates: (149,162,166); Description: KR domain putatively catalyzing D-configuration product formation, choice result: KR catalyzing D-configuration product, choice coordinates: (102); Description: KR domain putatively catalyzing L-configuration product formation, choice result: KR catalyzing L-configuration product, choice coordinates: (102); emission probability array (n.d.); emission probability array (n.d.,n.d.,n.d.); expected for choice: (D); expected for choice: (S,Y,N); expected for choice: ([^D]); matchArray: (False); matchArray: (True); matchArray: (True,True,True); overall match: FALSE; overall match: TRUE; residues: (E); residues: (S,Y,N)
aSASF_prediction: Full match for prediction: KR catalyzing L-configuration product; Full match for prediction: catalytic triade S,Y,N found
aSASF_note: ASF analyisis with definition ASP_PKSI-KR (type active_site); ASF analyisis with definition PKSI_KR_Stereo (type prediction)
NCBI Feature Key: aSDomain
Transferred From: NRPS translation
Transferred Similarity: 100.00%

Figure S10. antiSMASH analysis output for the KR domain of ClyF, highlighting the predicted configuration of the reduced product.

References

(1) Li, T.; Tripathi, A.; Yu, F.; Sherman, D. H.; Rao, A. DDAP: Docking Domain Affinity and Biosynthetic Pathway Prediction Tool for Type i Polyketide Synthases. *Bioinformatics* **2020**, *36* (3), 942–944. https://doi.org/10.1093/bioinformatics/btz677.

(2) Alonzo, D. A.; Chiche-Lapierre, C.; Tarry, M. J.; Wang, J.; Schmeing, T. M. Structural Basis of Keto Acid Utilization in Nonribosomal Depsipeptide Synthesis. *Nat. Chem. Biol.* **2020**, *16* (5), 493–496. https://doi.org/10.1038/s41589-020-0481-5.