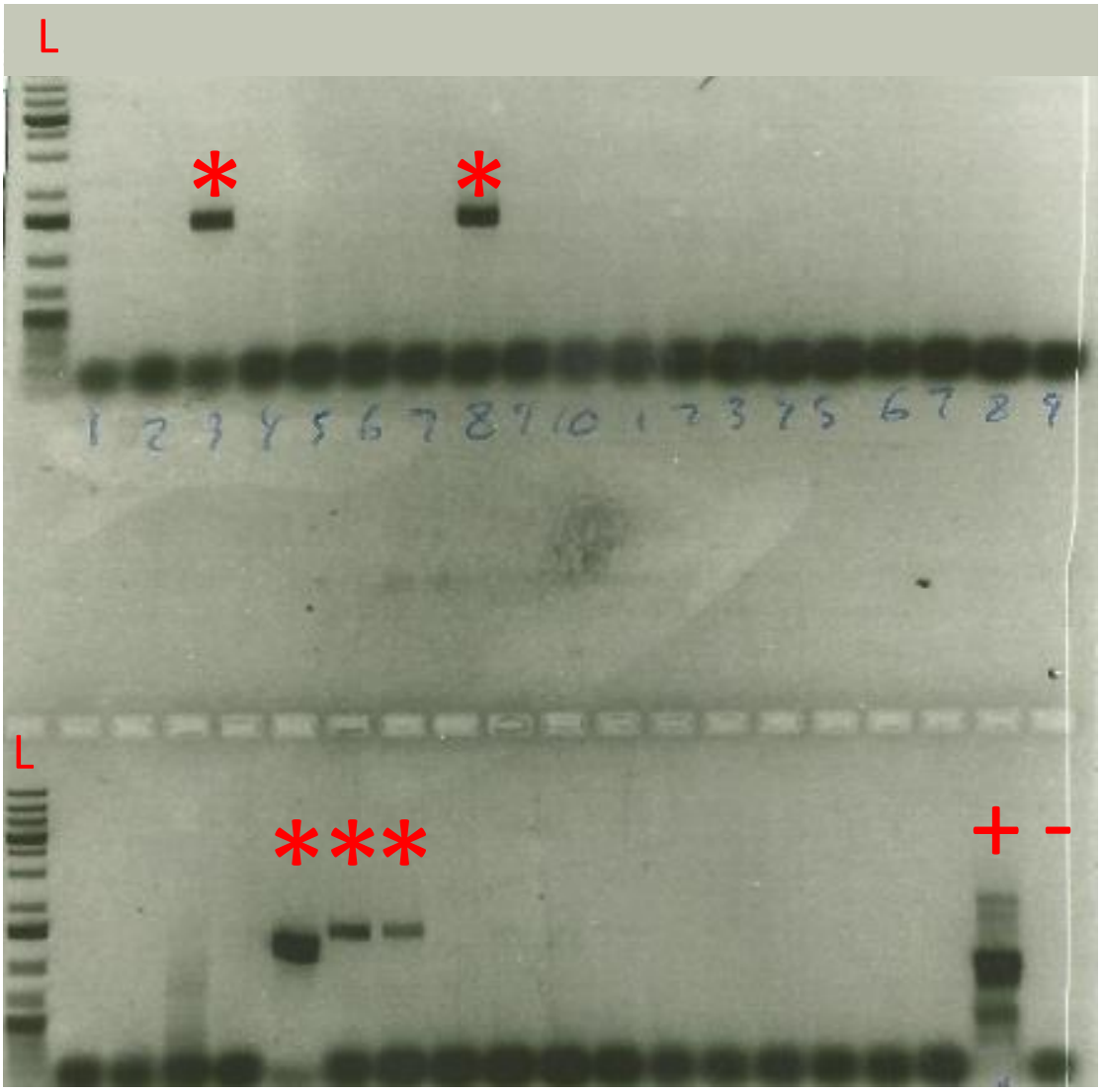
A PCR-independent Method of Transformation Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean Streptomycete

Charles B. Larson, Max Crüsemann, and Bradley S. Moore

Supporting Information

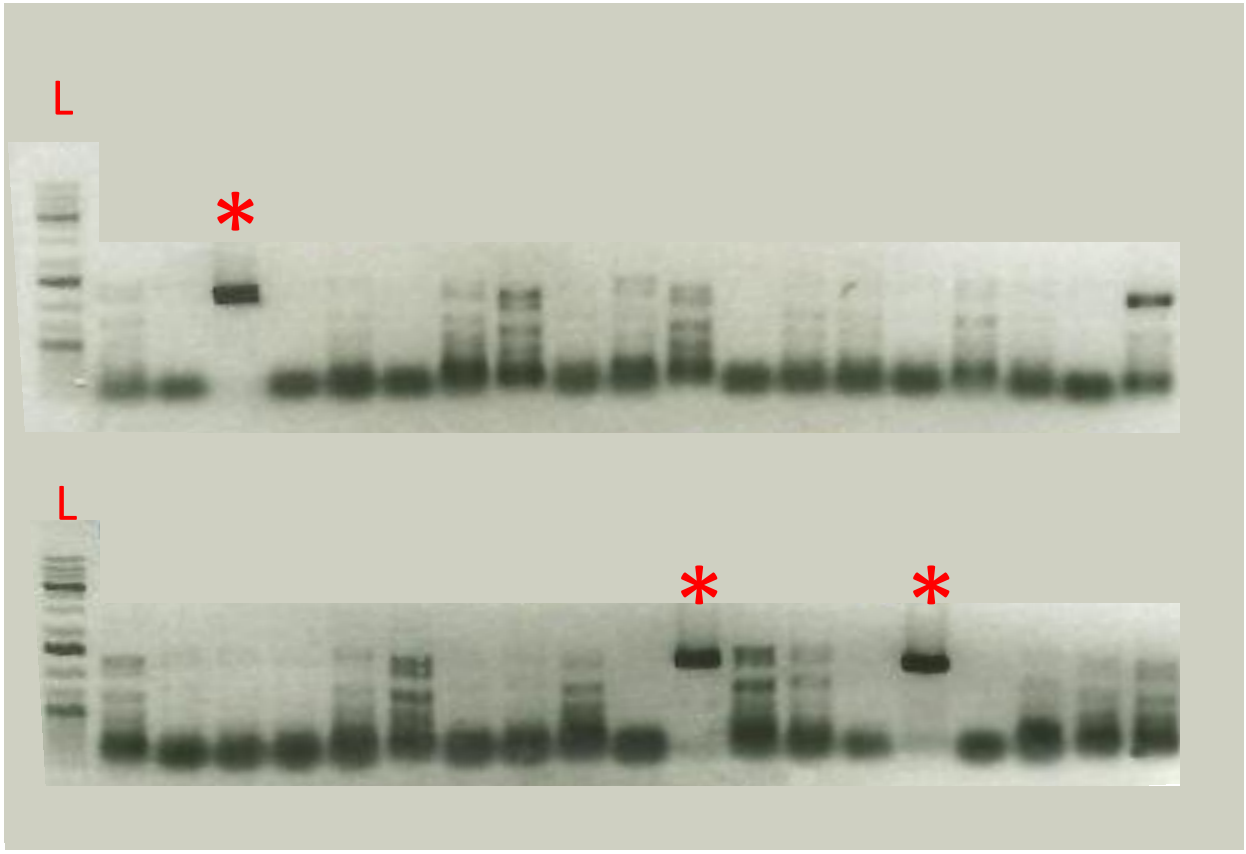
Table of Contents

Figure S1. PCR screening of multiple yeast colonies



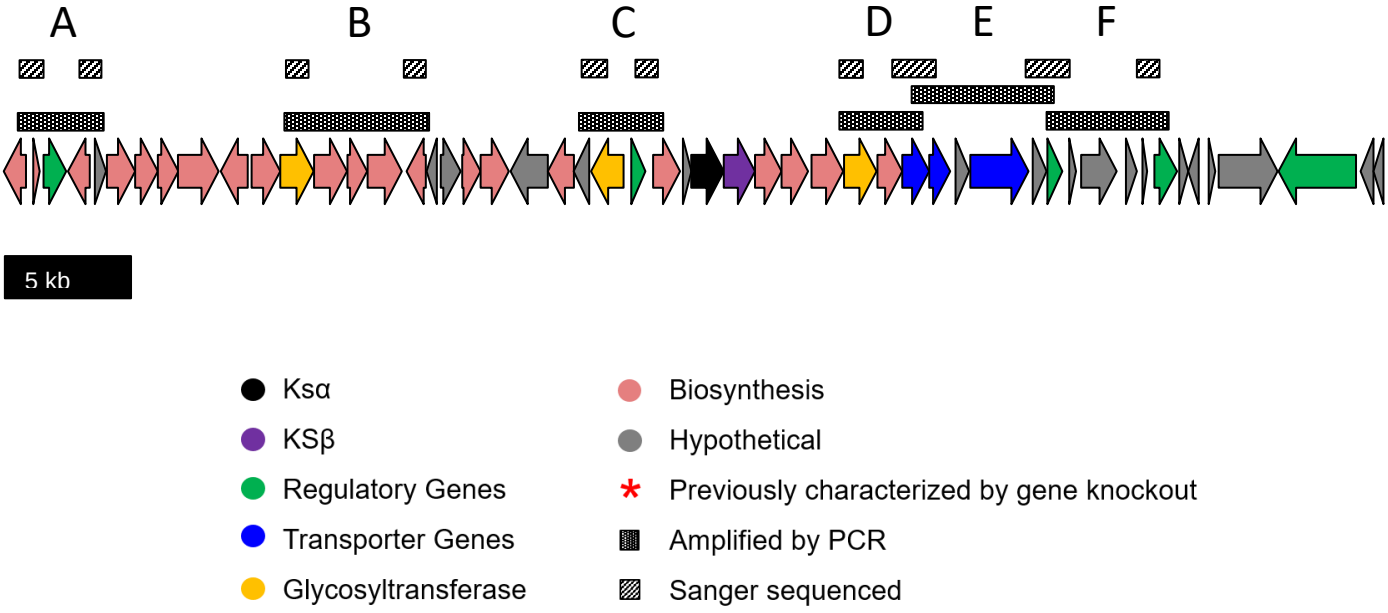
PCR amplification of a 1400 bp fragment of the target BGC performed on groups of 10 yeast clones. Positive amplification of target region indicated by "*". Positive and negative controls with and without CNT-302 gDNA shown as "+" and "–", respectively. Lanes loaded with 1 kb-plus DNA ladder (ThermoFisher) labeled with "L".

Figure S2. PCR screening of individual yeast colonies



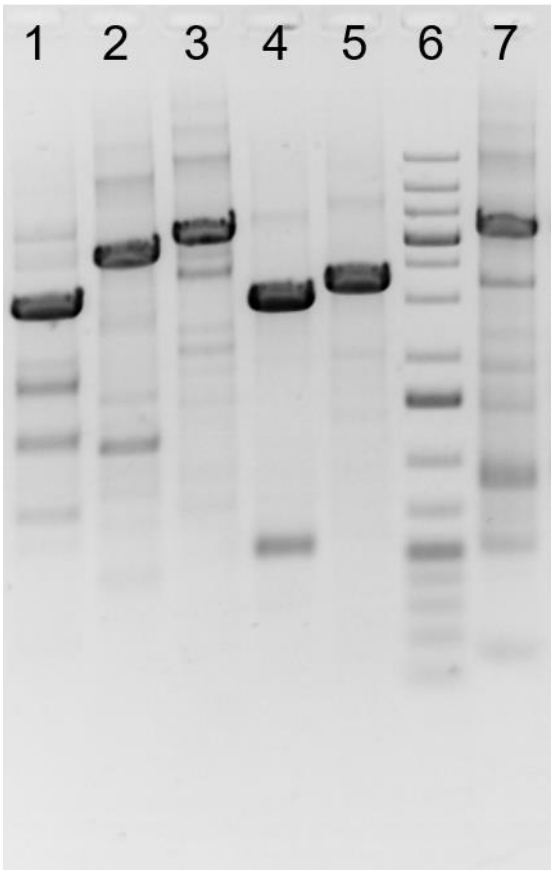
PCR amplification of a 1400 bp fragment of the target BGC performed on individual yeast clones. Lanes loaded with 1 kb-plus DNA ladder labeled with "L". Positive hits indicated by "*". A total of three clones showed clean amplification, with another clone showing an amplification band with lower molecular weight DNA also in the lane. A total of 200 colonies were initially investigated, calculated transformation efficiency was 1.5%.

Figure S4. Sequencing of integrated cosmomycin pathway



In order to further confirm proper assembly and integration of plasmid pCAP01-COS, regions A-F depicted above were amplified and Sanger sequenced. Results confirmed the sequence of the integrated pathway was correct for all amplicons. A total of 10 kb of the pathway was covered by the sequencing effort.

Figure S3. PCR amplification of integrated cosmomycin pathway

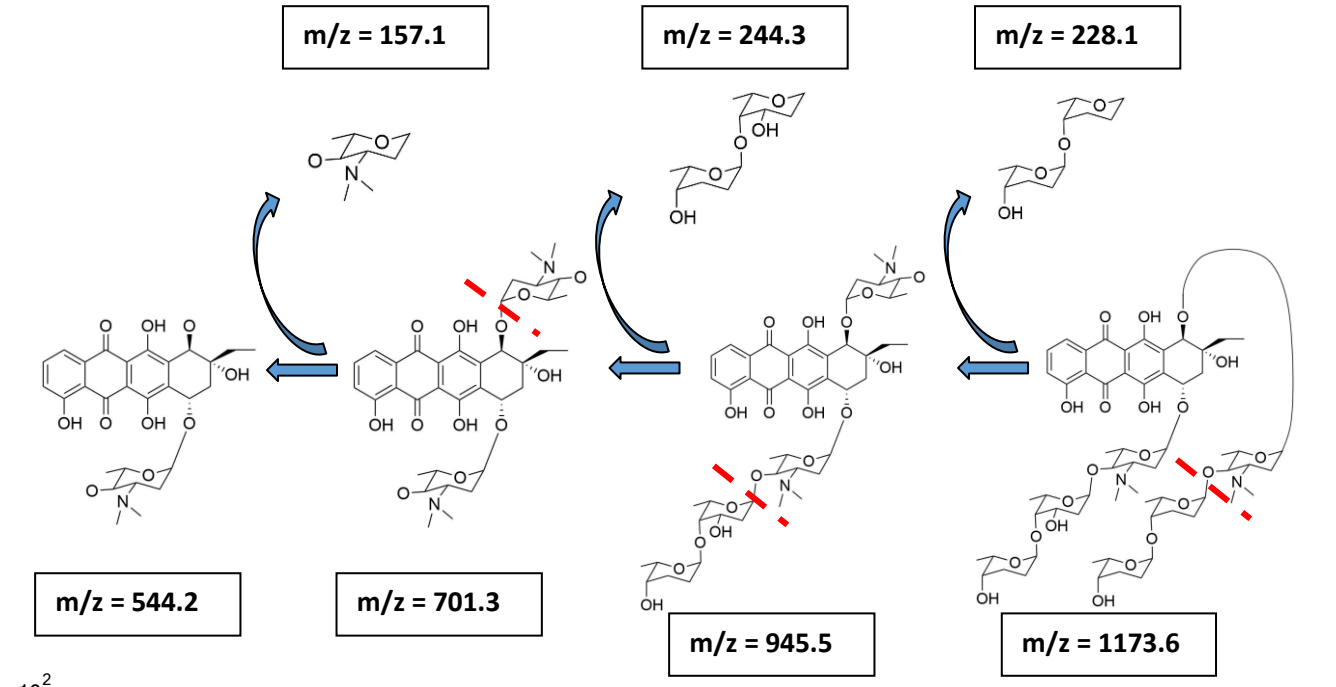


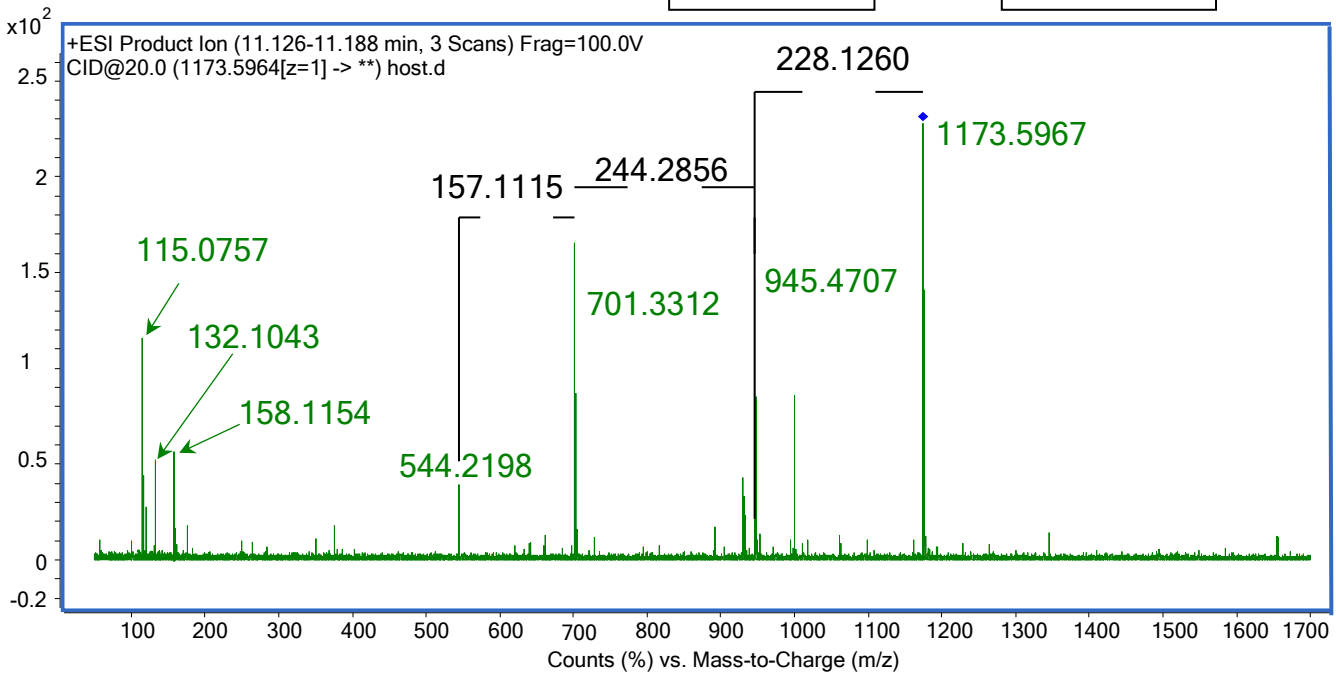
After integrating pCAP01-COS into the genome of *S. coelicolor* M512, various regions were amplified by PCR to confirm their size, proper assembly, and integration.

Expected Size (bp)

1: **3191** (Region D)
2: **4639** (Region F)
3: **5602** (Region E)
4: **3040** (Region C)
5: **3419** (Region A)
6: 1kb Plus Ladder: 20000, 10000, 7000, **5000**, 4000, 3000, 2000, **1500**, 1000, 700, **500**, 400, 300, 200
7: **5555** (Region B)

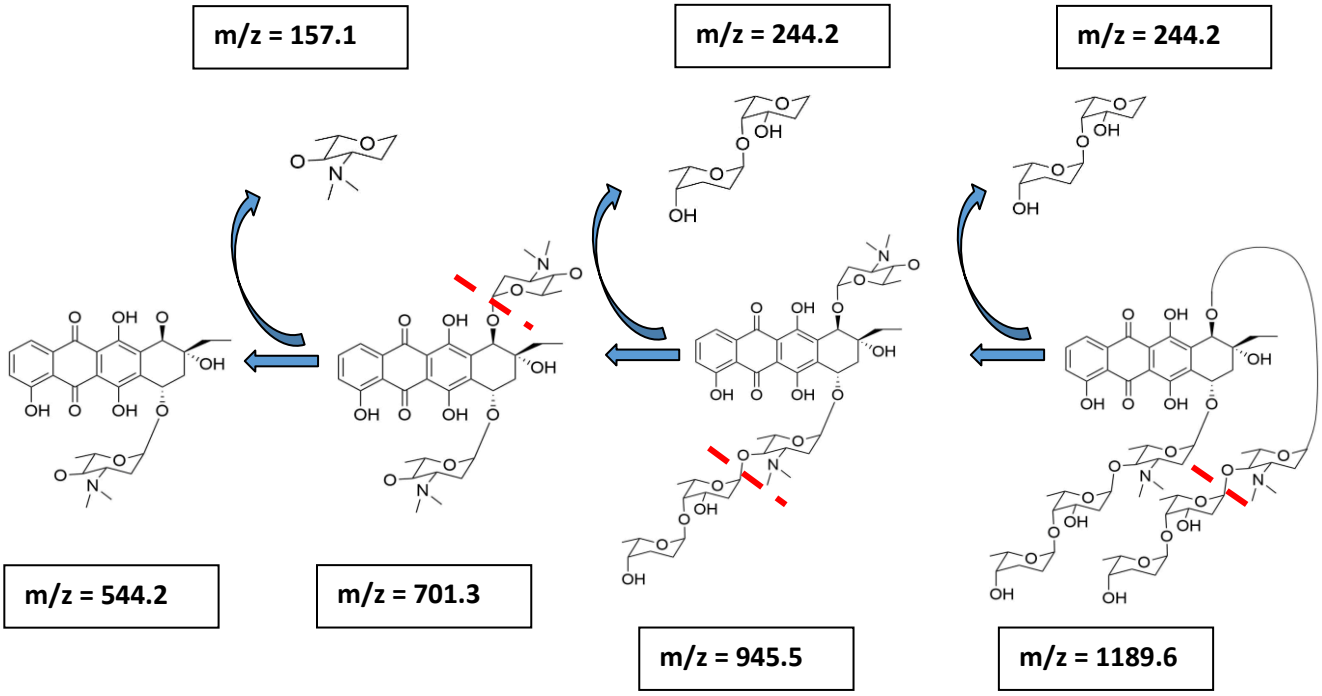
Figure S5. MS2 fragmentation analysis of cosmomycin C **(1)**

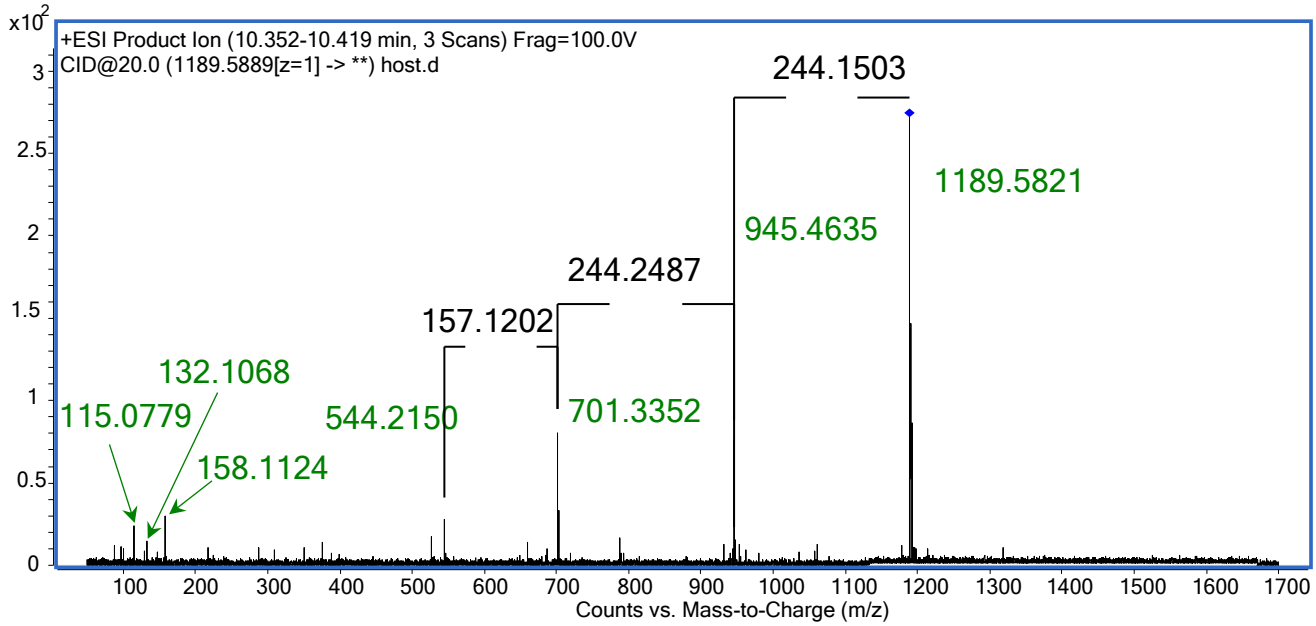




Fragmentation of parent ion 1173.59 shows losses corresponding to sugar moieties of cosmomycin C. The initial loss of 228 mass units correspond to the B ions of two rhodinose moieties, the two outermost sugars on one of the trisaccharide chains. The second loss of 244 corresponds to rhodinose and 2-deoxyfucose lost from the second trisaccharide chain. The last shift seen represents the loss of rhodosamine, with a B ion mass of 157. All three sugars can be seen in the lower mass range of the spectrum.

Figure S6. MS2 fragmentation analysis of cosmomycin D **(2)**





Fragmentation of parent ion 1189.589 shows losses corresponding to sugar moieties of cosmomycin D. The initial loss of 244 mass units correspond to the B ions of rhodinose and L-2-deoxy-fucose, the two outermost sugars on both the trisaccharide chains. The second loss of 244 corresponds to the same two sugars lost from the second trisaccharide chain. The last shift seen represents the loss of rhodosamine, with a B ion mass of 157. All three sugars can be seen in the lower mass range of the spectrum.

Figure S7. MS1 spectrum of cosmomycin analog **3**

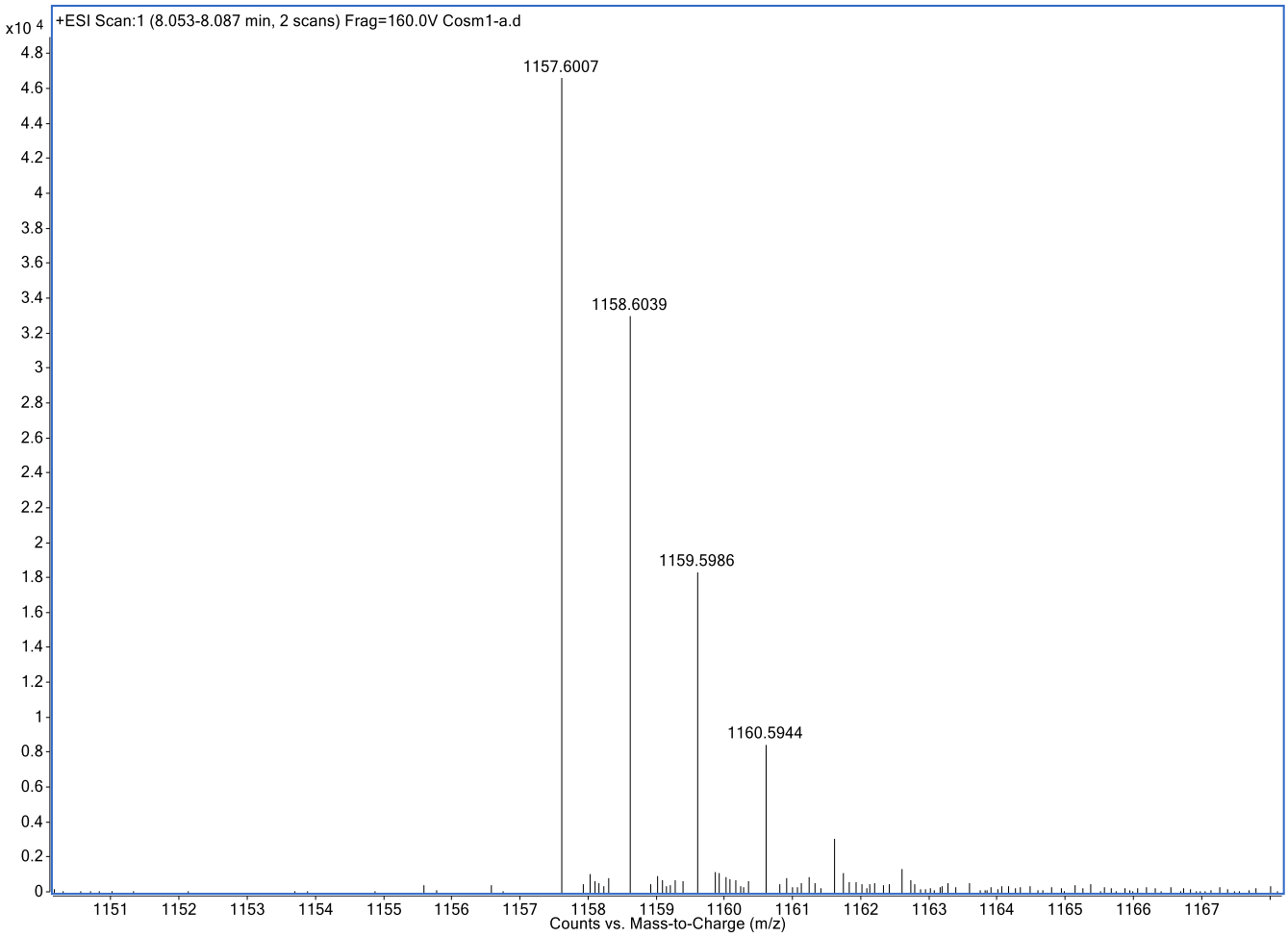
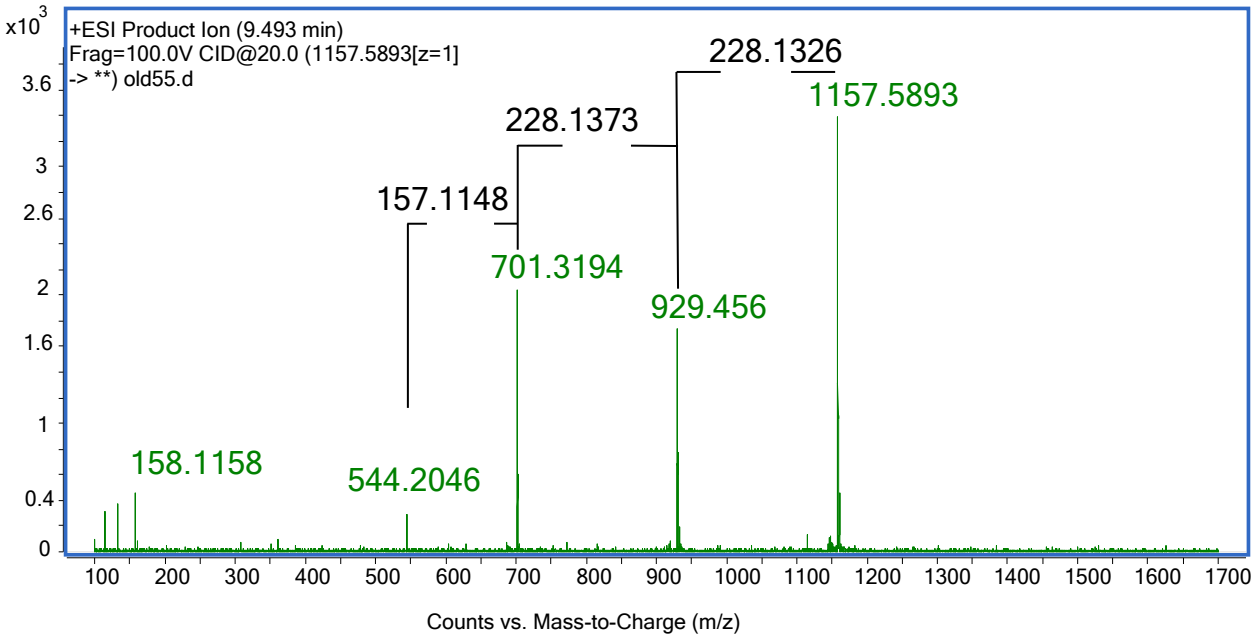


Figure S8. MS2 fragmentation analysis of cosmomycin analog **3**



Fragmentation of parent ion 1157.5893 shows losses corresponding to sugar moieties of the previously reported compound A447 C, also reported by Kelso et al[1]. The initial loss of 228 mass units corresponds to the B ions of two rhodinose moieties, the two outermost sugars on one of the trisaccharide chains. The second loss of 228 corresponds to the same moieties lost from the second trisaccharide chain. The final loss of 157 corresponds to a rhodosamine B ion.

Figure S9. MS1 spectrum of cosmomycin analog **A**

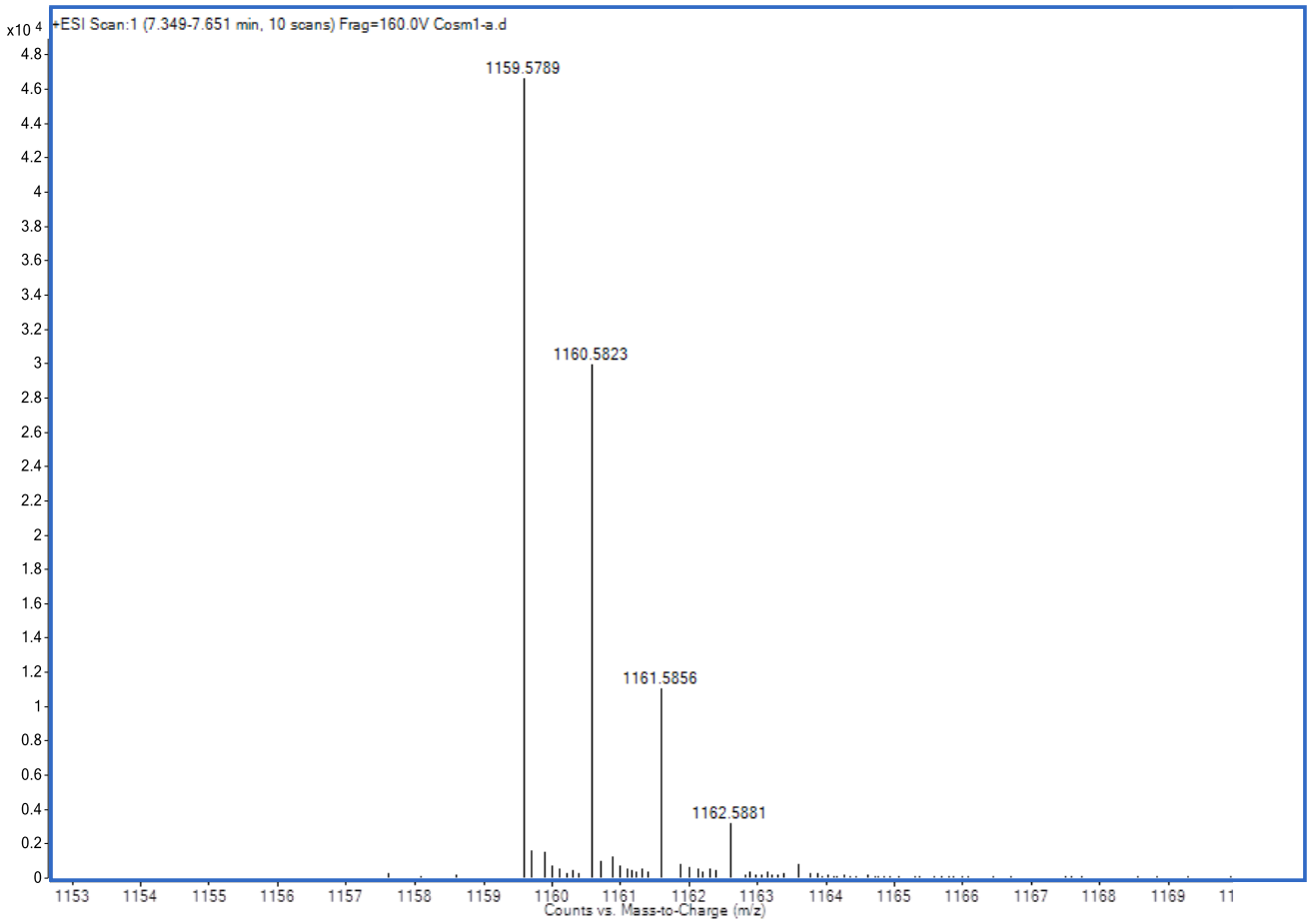
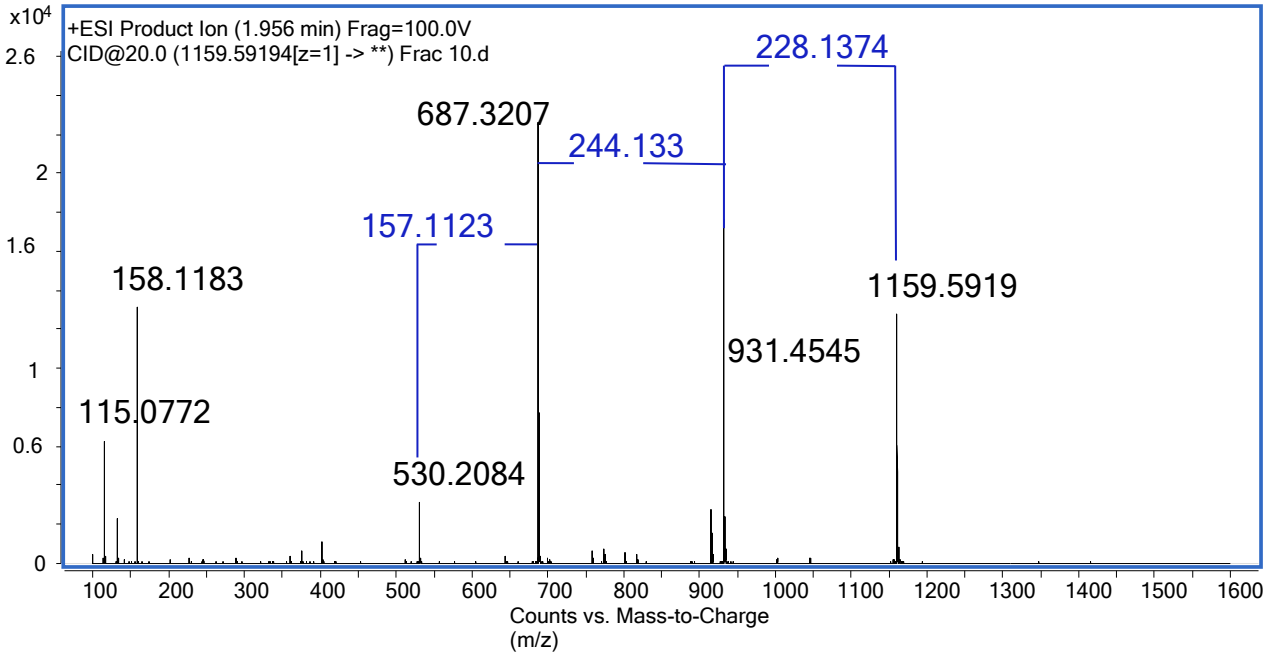
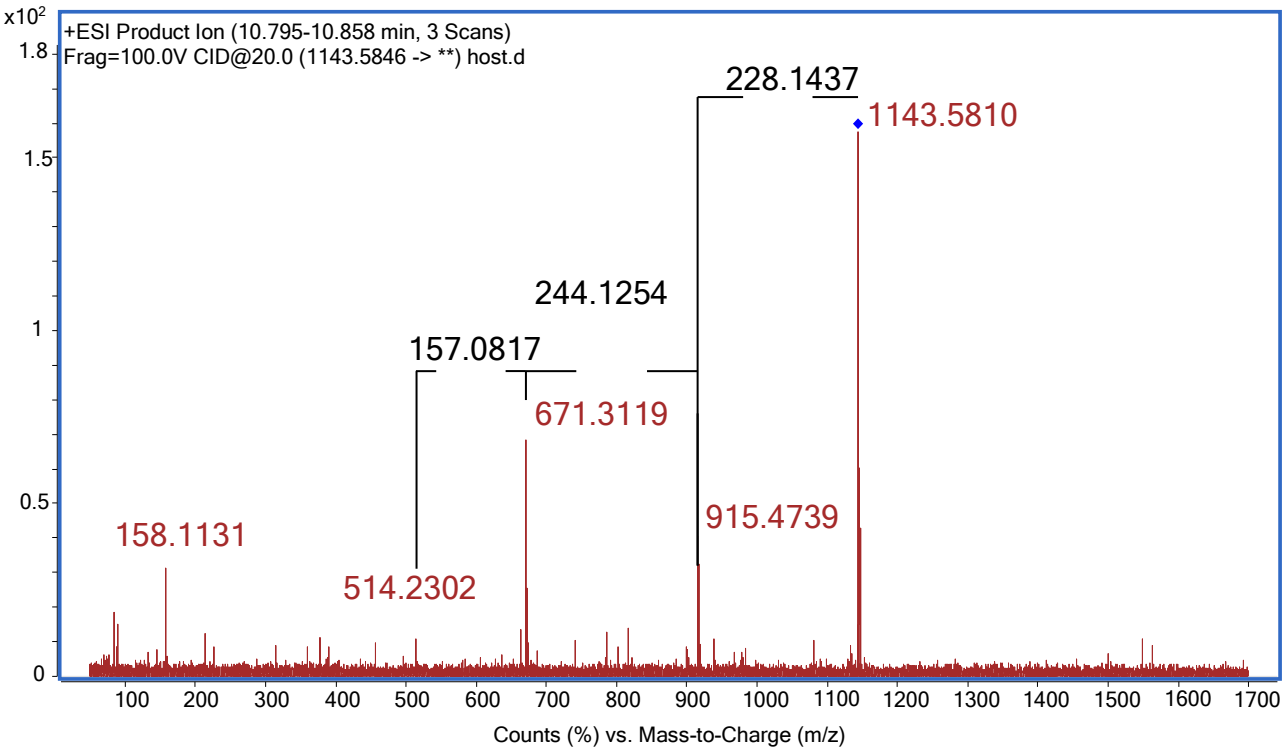


Figure S10. MS2 fragmentation analysis of cosmomycin analog **A**



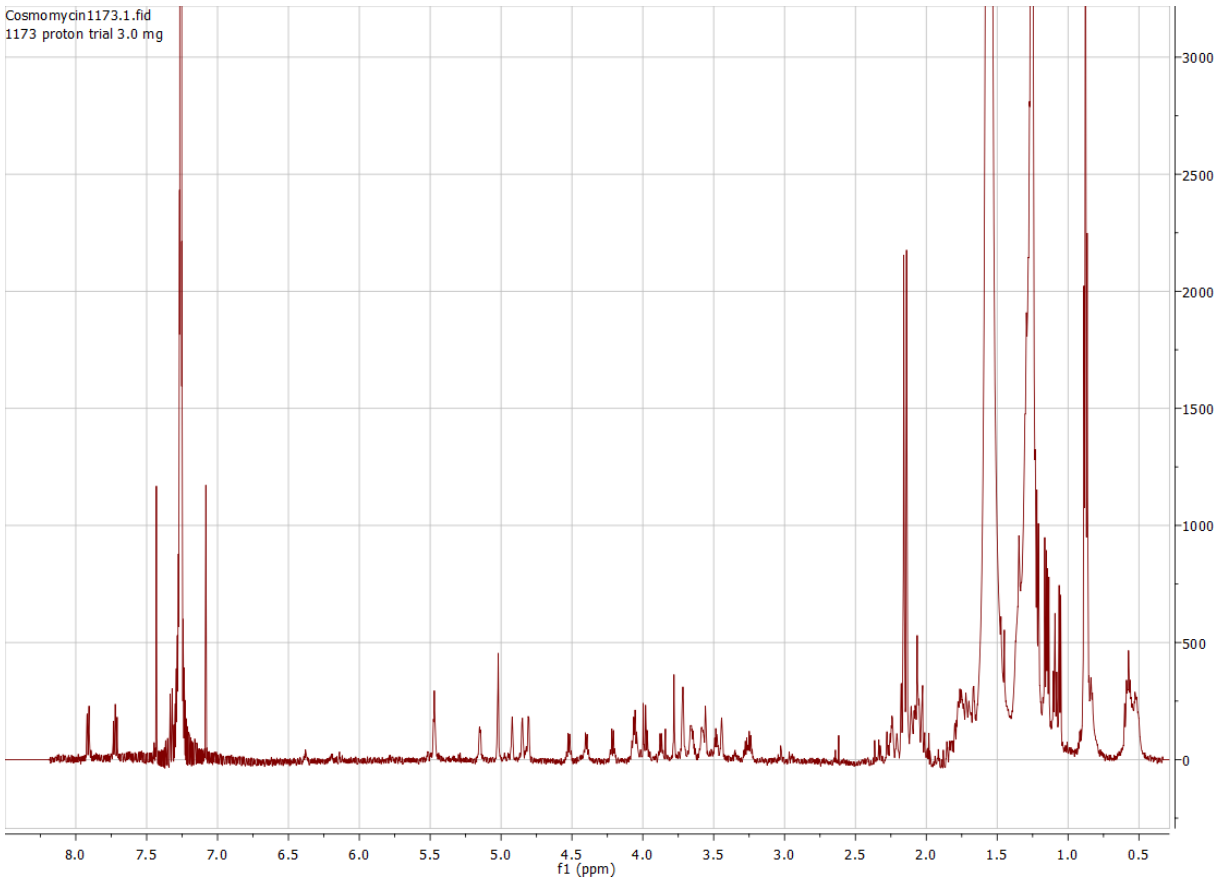
Fragmentation of parent ion 1159.5919 shows an initial loss of 228 mass units corresponding to the B ions of two rhodinose moieties, the two outermost sugars on one of the trisaccharide chains. The second loss of 244 corresponds to rhodinose and 2-deoxyfucose lost from the second trisaccharide chain. Although the sugar masses fragmented here are identical to the sugars lost from cosmomycin C, the mass of the rest of the molecule is 14 Daltons lighter than cosmomycin C at each stage of fragmentation. This suggests the lack of a methylene group the aglycone.

Figure S11. MS2 fragmentation analysis of cosmomycin analog **B**



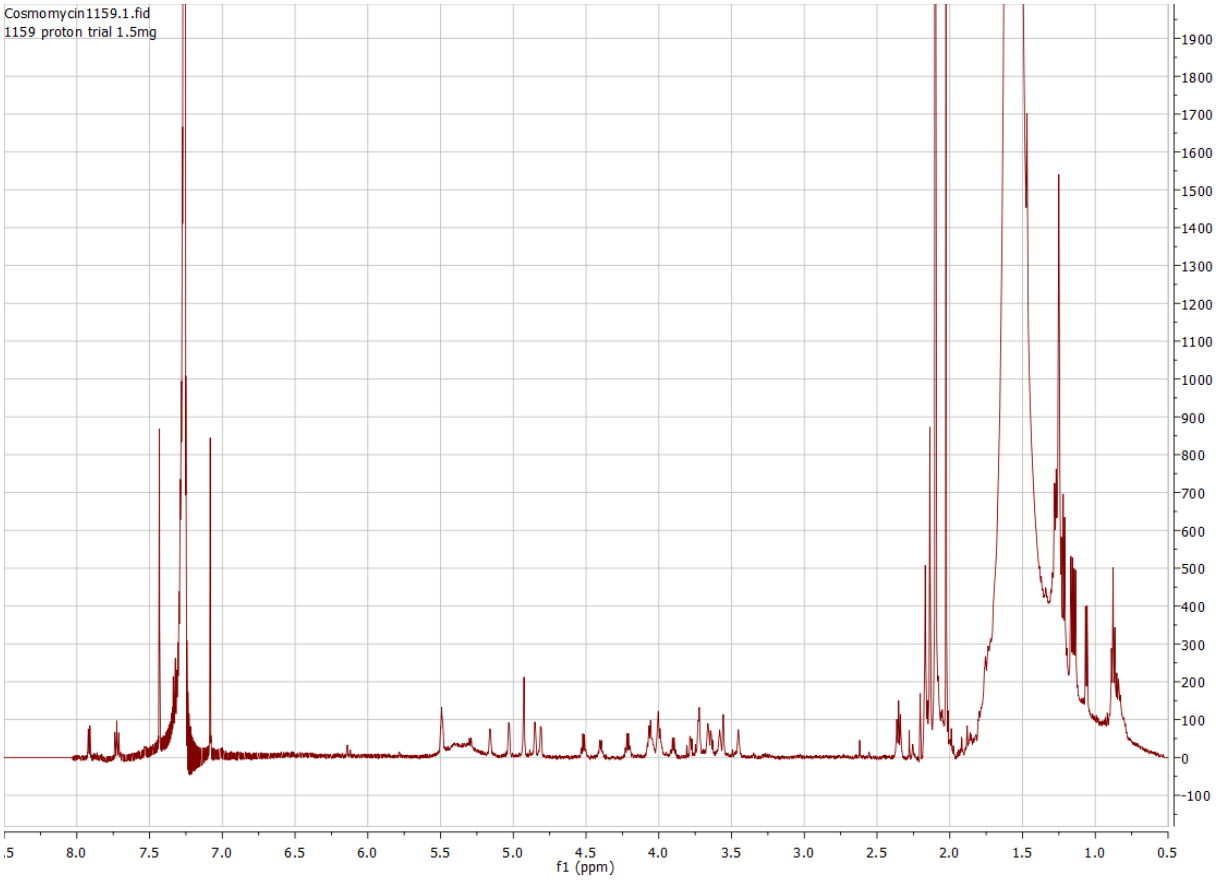
The initial loss of 228 mass units corresponds to the B ions of two rhodinose residues, the two outermost sugar units on one of the trisaccharide chains. The second loss of 244 corresponds to rhodinose and 2-deoxyfucose lost from the second trisaccharide chain. The mass shifts of the sugar moieties fragmented here match the first two fragments of cosmomycin C, indicating the same glycosylation pattern on the four outermost sugars. The mass of the unfragmented molecule, as well as the mass of the first three fragments, show a mass shift of -30 Daltons compared to cosmomycin C. This mass shift is consistent with the loss of an oxygen and a methylene group, suggesting this species is desoxy, desmethyl-cosmomycin C (compound **B**).

Figure S12. ^{1}H NMR spectrum of cosmomycin C (**1**)



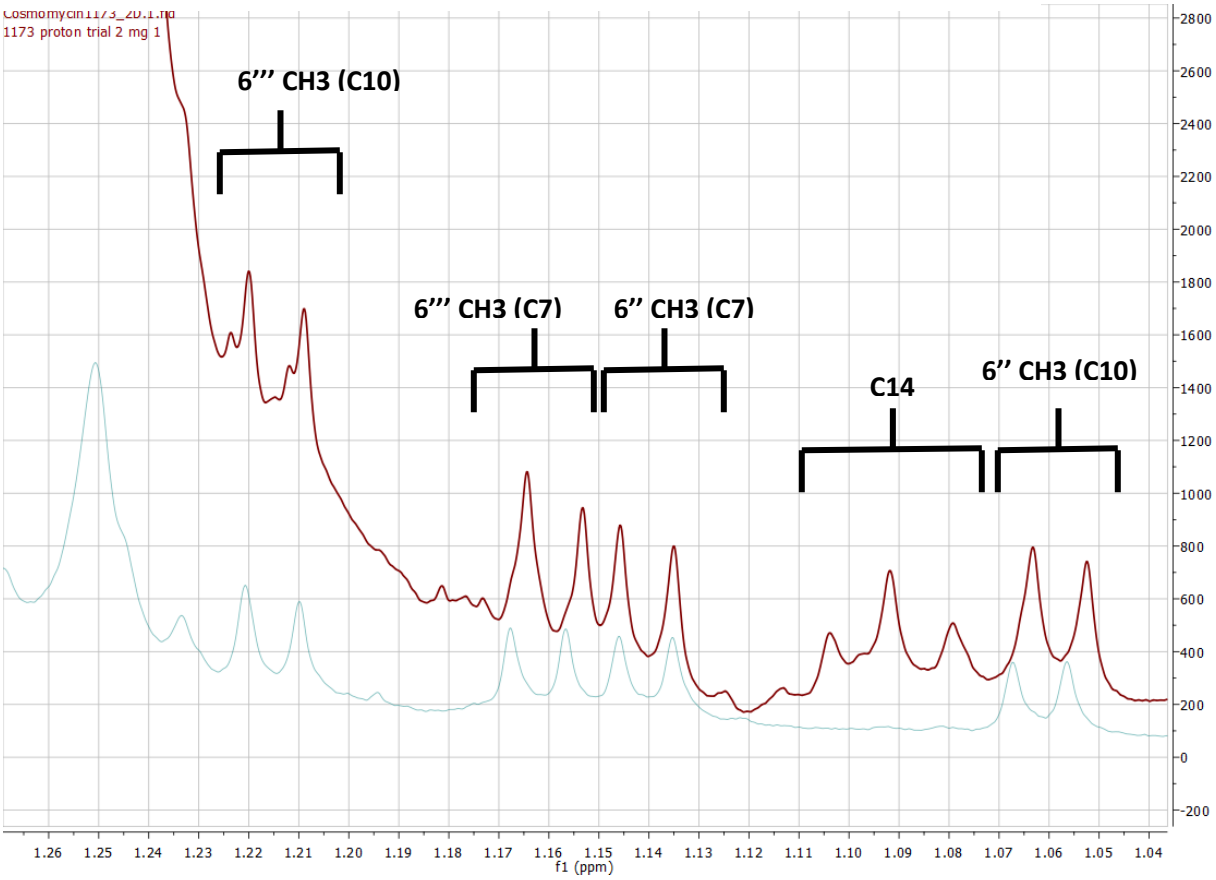
NMR Spectrum of cosmomycin C shows three distinct aromatic protons at 7.92 ppm, 7.72 ppm, and 7.33 ppm. Sugar protons can be seen from 3.5 to 5.5 ppm, and the N-methyl protons from the inner rhodosamines can be seen at 2.17 and 2.14.

Figure S13. ^{1}H NMR spectrum of **A**



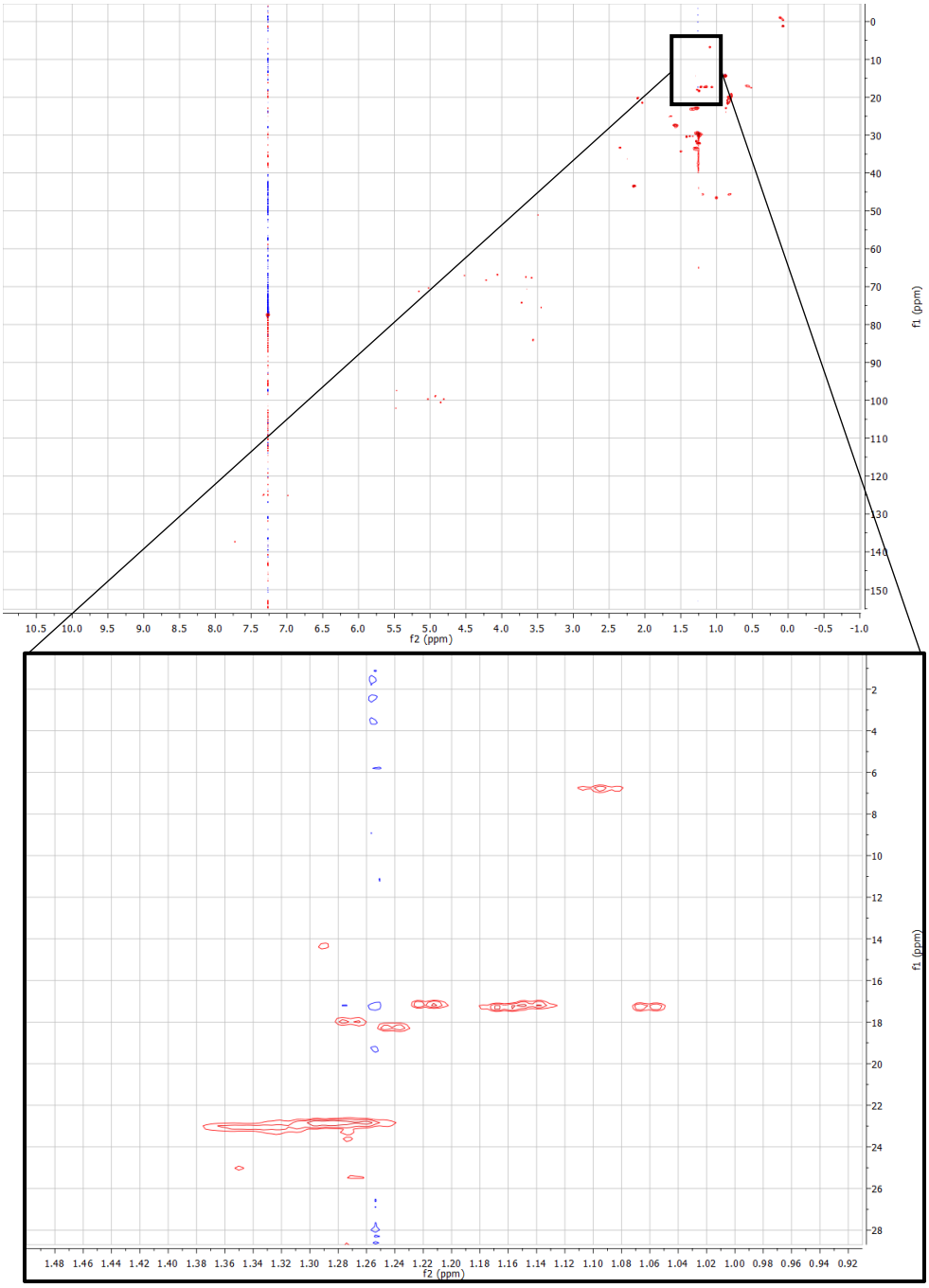
NMR Spectrum of compound **4** shows three distinct aromatic protons at 7.92 ppm , 7.72 ppm, and 7.33 ppm. Sugar protons can be seen from 3.5 to 5.5 ppm, and the N-methyl protons from the inner rhodosamines can be seen at 2.17 and 2.14.

Figure S14. Overlaid ^{1}H NMR spectrum of **1** and **A**



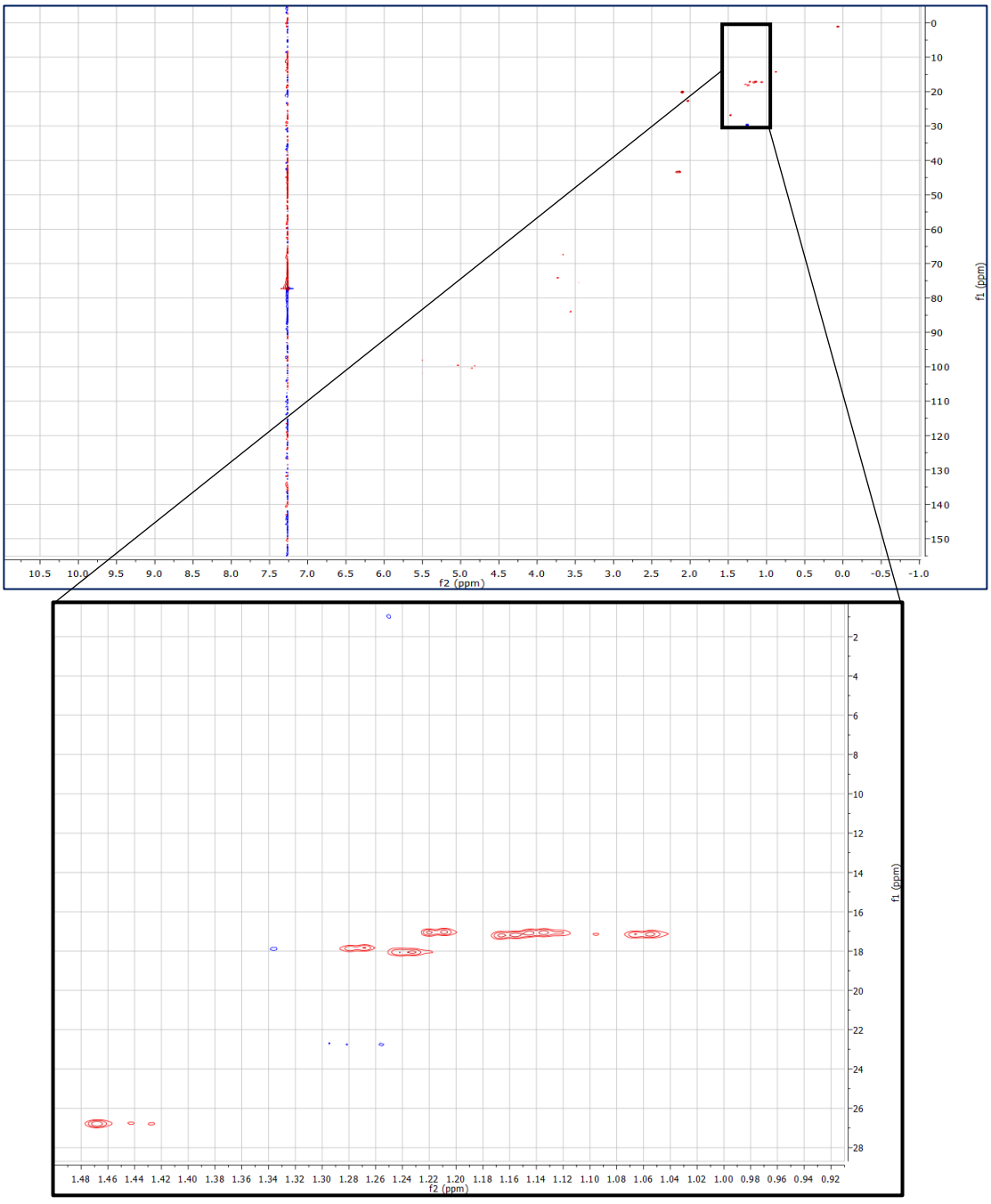
Overlaid ^{1}H NMR spectra of compounds **1** (red signal) and **A** (blue signal) show the loss of a triplet signal at 1.09 ppm in compound **A**. This signal corresponds to the C-14 methyl group of the aglycone.[4,5]

Figure S15. HSQC spectrum of **1**



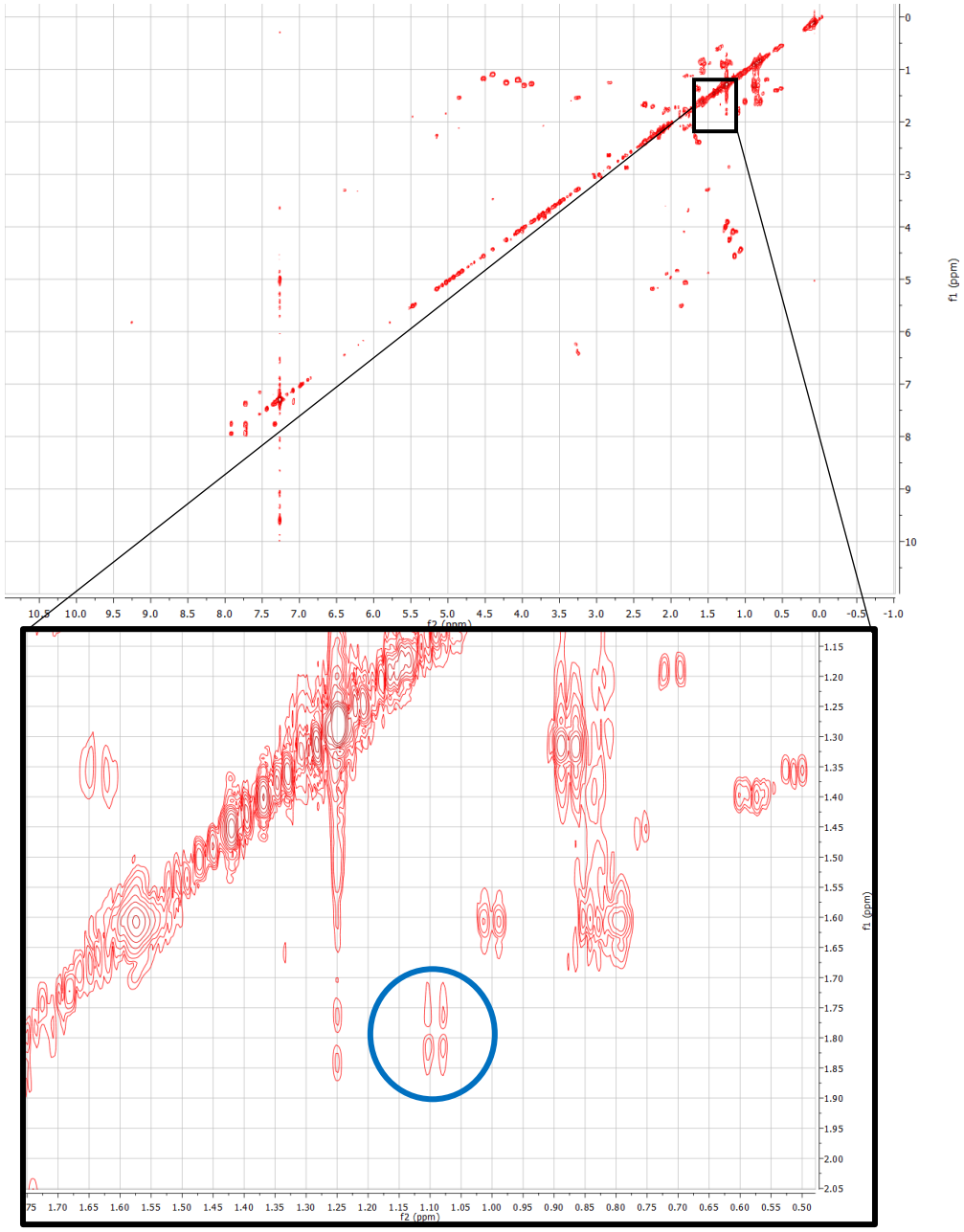
In the HSQC spectrum of compound **1** was acquired using 12.5% non-uniform sampling and reconstructed using SMILE[2] and NMRPipe[3] the triplet at 1.09 can clearly be seen correlating to carbon 14 of the aglycone which has a diagnostic chemical shift of 6.7 ppm. Also in the expanded region, the methyl protons attached to carbon six of all six sugars can be seen correlating to carbons with chemical shifts from ~17 to 18.3 ppm, matching literature values.[4-8]

Figure S16. HSQC NMR spectrum of **A**



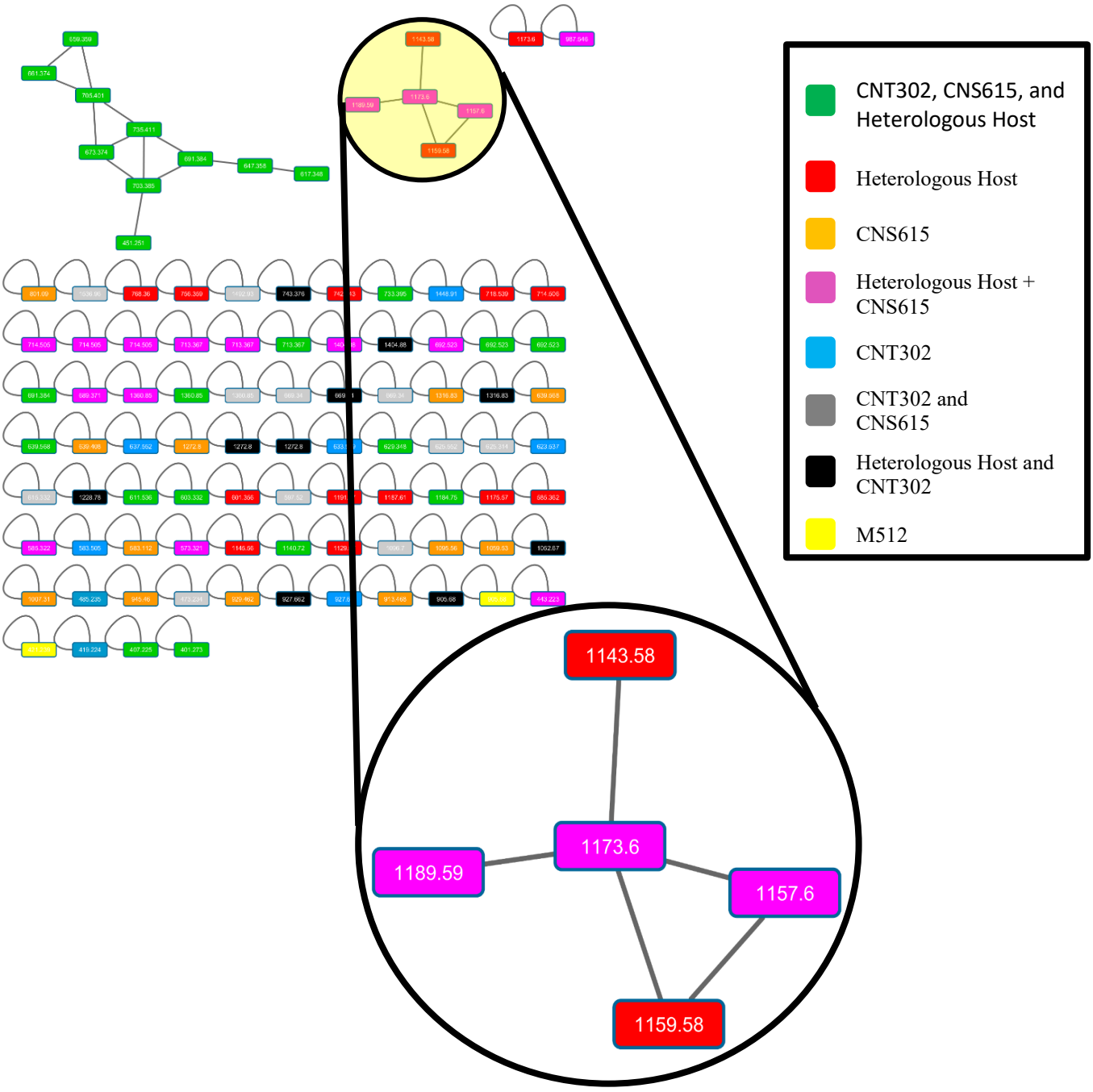
In the HSQC spectrum of compound **A**, the triplet at 1.09 ppm corresponding to carbon 14 is missing, and an additional singlet peak at 1.47 ppm which correlates to a carbon with a chemical shift of 26.8 ppm can be clearly seen. This data suggests the new peak represents carbon 13, which is now a terminal methyl group rather than a methylene. Such a structure is consistent with the canonical propionate PKS starter unit of cosmomycin being replaced by a malonate starter unit in the heterologous host.

Figure S17. COSY NMR spectrum of **1**



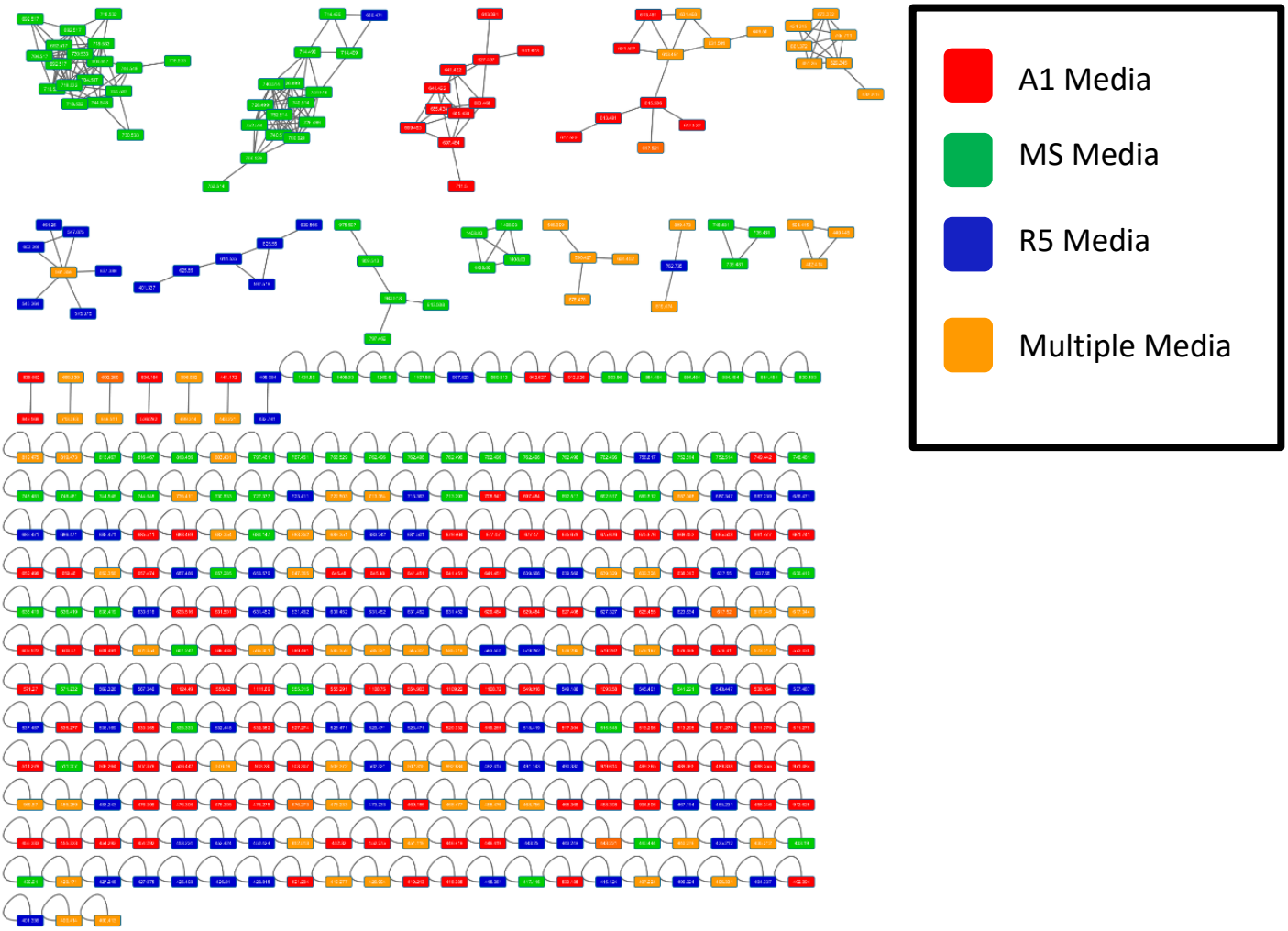
COSY correlation between carbon 14 protons observed at 1.09 ppm and two protons with chemical shifts of 1.76 and 1.82 (circled). These shifts are consistent with literature values for protons 13a and 13b.[6-8]

Figure S18. Molecular network of cosmomycin native producers and heterologous host



Molecular network generated using the GNPS algorithm.[9,10] Red nodes are found only in the heterologous host, while orange nodes are found only in CNS-615. Nodes colored purple are found in both strains. CNT-302 extracts were included in the networking, but no compounds related to cosmomycin were present. In addition to our own extracts, publicly available data on the strains was also included in the network (MassIVE accession: MSV000078836)[11] and visualization was performed by cytoscape.[12] The cosmomycin molecular family is highlighted and expanded.

Figure S19. *Streptomyces* sp. CNT-302 molecular network

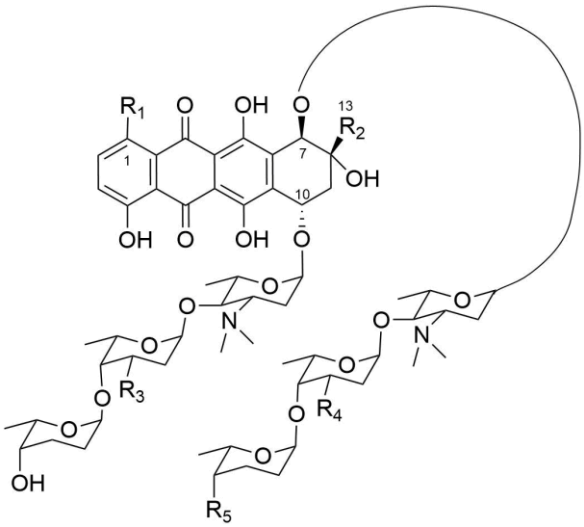


Networking results from CNT-302 butanol, methanol, and ethyl acetate extractions obtained from archived MassIVE data set MSV000078836. No masses matching cosmomycin C, cosmomycin D, or other cosmomycin family members were detected.

Table S1. [1]H NMR chemical shift assignments of compounds **1** and **4** and the related cosmomycin analogs obelmycin C[4] and A447C/D[5]

	Obelmycin C	A447 C	A447 D	Cosmomycin C (1)	Compound A
R1	OH	H	H	H	H
R2	CH2CH3	CH2CH3	CH2CH3	CH2CH3	CH3*
R3	OH	H	H	OH	OH
R4	OH	H	H	H	H
R5	OH	OH	=O	OH	OH

'*' denotes a putative assignment not confirmed by full NMR characterization.



	Obelmycin C		A447D		A447C		Cosmomycin C (**1**)		(**A**)	
Aglycone										
1 H	-	-	7.84	dd, J = 7.2, 1.1 Hz	7.89	d, J = 7.4 Hz	7.91	d, J = 7.2 Hz	7.91	d, J = 7.2 Hz
2H	7.31	d, J=10 Hz	7.65	dd, J = 7.4, 7.2 Hz	7.7	dd, J = 8.3, 7.4 Hz	7.72	t, J= 8.4, 7.8 Hz	7.72	t, J= 8.4, 7.8 Hz
3H	7.29	d, J=10 Hz	7.25	dd, J = 7.2, 1.1 Hz	7.3	dd, J = 8.3 Hz	7.33	d, J = 8.4 Hz	7.33	d, J = 8.4 Hz
7H	5.15	br d	5.15	m	5.14	m	5.15	br d 3 Hz	5.16	br d, J = 3 Hz
8 Ha	2.2		NR		NR		NR		NR	
8 Hb	2.2		NR		NR		NR		NR	
10 H	5.02	s	5.02	s	5.01	s	5.02	br s (2 protons)	4.92	s (2 protons)
13 Ha	NR		NR		NR		NR		NR	

13 Hb	NR		NR		NR		NR		NR	
14 CH3	1.1	t, J= 7 Hz	1.1	t, J = 7.4 Hz	1.1	t, J = 7.5 Hz	1.09	t, J=7.2 Hz		
C7 Trisaccharide										
1' H	5.48	d, J = 3.5 Hz	5.48	dd, J = 3.4 Hz	5.5	d, J = 3.6 Hz	5.47	m (2 protons)	5.49	m (2 protons)
2' H	NR		NR		NR		NR		NR	
3' H	NR		NR		NR		NR		NR	
4' H	3.73	br s	3.73	br s	3.78	br s	3.78	s	3.78	m
5' H	4.06	q, J = 7 Hz	3.99	q, J = 6.4 Hz	4	q, J = 6.8 Hz	4.06	q, J= 6.6 Hz	4.06	q, J = 6.6 Hz
6'-CH3	1.27	d, J= 7 Hz	1.28	dd, J = 6.4 Hz	1.27	d, J = 6.8 Hz	NR		NR	
3'-N(CH3)2	2.16	s	2.15	s	2.16	s	2.14	s (6 protons)	2.14	s (6 protons)
1" H	5.03		5.04	br s	4.95	br s	5.02	br s, 2 protons	5.03	br d, J = 3 Hz
2" H	NR		NR		NR		NR		NR	
3" H	3.98	br d	NR		NR		3.99	d, J= 10.2 Hz	4	br m
4" H	3.56	br s	4.06	m	3.47	br s	3.48	m	3.56	s
5" H	4.53	q, J = 7 Hz	4.52	q, J = 6.3 Hz	4.44	q, J = 6.7 Hz	4.52	q, J= 6.6 Hz	4.51	q, J = 6.6 Hz
6" CH3	1.14	q, J = 7 Hz	1.15	d, J = 6.3 Hz	1.07	d, J = 6.7 Hz	1.14	d, J=6.6 Hz	1.14	d, J = 6.6 Hz
1"' H	4.85	br s	4.85	br s	4.82	d, J = 3 Hz	4.85	s	4.85	br s
2"' H	NR		NR		NR		NR		NR	
3"' H	NR		NR		NR		NR		NR	
4"' H	3.66	br s	3.66	br s	3.58	br s	3.65	m	3.65	m
5"' H	4.06	q, J = 7 Hz	4.22	q, J = 6.7 Hz	4.06	q, J = 6.8 Hz	4.06	q, J= 6.6 Hz	4.06	q, J = 6.6 Hz
6"' CH3	1.16	q, J = 7 Hz	1.22	d, J = 6.7 Hz	1.15	d, J = 6.8 Hz	1.16	d, J=6.6 Hz	1.16	d, J = 6.6 Hz
C10 Trisaccharide										
1' H	5.44	d, J = 3.5 Hz	5.46	d, J = 3.5 Hz	5.46	d, J = 3.3 Hz	5.47	m (2 protons)	5.49	m (2 protons)
2' H	NR		NR		NR		NR		NR	
3' H	NR		NR		NR		NR		NR	
4' H	3.73	br s	3.73	br s	3.72	br s	3.72	br s (2 proton)	3.72	br s (2 protons)
5' H	3.88	q, J = 7 Hz	3.88	q, J = 6.7 Hz	3.87	q, J = 6.6 Hz	3.87	q, J= 6.6 Hz	3.9	q, J=6.6 Hz
6' CH3	1.24	d, J = 7 Hz	1.24	d, J = 6.7 Hz	1.23	d, J = 6.6 Hz	NR		NR	
3'-N(CH3)2	2.16	s	2.17	s	2.16	s	2.16	s (6 protons)	2.17	s (6 protons)
1" H	4.93	br s	4.92	br s	4.92	br s	4.92	s (2 protons)	4.92	s (2 protons)
2"	NR		NR		NR		NR		NR	
3" CH3	1.7, 2.0		NR		NR		NR		NR	
4" H	3.45	br s	3.55	br s	3.44	br s	3.45	s	3.45	s
5" H	4.4	q, J = 7 Hz	4.44	q, J = 6.3 Hz	4.39	q, J = 6.3 Hz	4.4	q, J= 6.6 Hz	4.4	q, J = 6.6 Hz

6" CH3	1.07	d, J = 7 Hz	1.09	d, J = 6.3 Hz	1.06	d, J = 6.3 Hz	1.06	d, J=6.6 Hz	1.06	d, J = 6.6 Hz
1‴ H	4.81	d, J = 3.5 Hz	5.03	t, J = 5.3 Hz	4.8	d, J = 3.2 Hz	4.8	doublet J= 3.6 Hz	4.81	d, J = 3.6 Hz
2‴ H	NR		NR		NR		NR		NR	
3‴ H	NR		2.51	ddd, J = 16, 6.7, 5.3 Hz			NR		NR	
			2.42	ddd, J = 16, 9.5, 5.3 Hz						
4‴ H	3.58	br s			3.58	br s	3.57	m	3.58	s
5‴ H	4.22	q, J = 7 Hz	4.33	q, J = 6.6 Hz	4.07	q, J = 6.8 Hz	4.21	q, J= 6.6 Hz	4.21	q, J = 6.6 Hz
6‴ CH3	1.22	d, J = 7 Hz	1.27	d, J = 6.6 Hz	1.17	d, J = 6.6 Hz	1.21	d, J=6.6 Hz	1.21	d, J = 6.6 Hz

All spectra were recorded in $CDCl_3$ and spectra for this study were referenced to residual chloroform (δH 7.26).

Table S2. Comparison of *Streptomyces* sp. CNT-302 and *S. olindensis* BGCs

Gene no. on CNT-302 ctg13	Gene product in CNT-302 (smcog)	Homologue in *S. olindensis*	Identity	Accession no.
91	glucose-1-phosphate thymidylyltransferase	glucose-1-phosphate thymidylyltransferase	97%	KDN80037.1
92	acyl carrier protein	actinorhodin polyketide synthase ACP	88%	WP_031119660.1
93	SARP family transcriptional regulator	transcriptional regulator	95%	KDN80039.1
94	methyltransferase	methyltransferase	92%	WP_037757689.1
95	unannotated	ester cyclase	95%	KDN80041.1
96	O-methyltransferase	methyltransferase	88%	KDN80042.1
97	alpha/beta hydrolase fold protein	short chain dehydrogenase	39%	WP_037760089.1
98	methyltransferase	SAM-dependent methyltransferase	89%	KDN80044.1
99	mono-oxygenase, FAD binding	FAD-dependent oxidoreductase	94%	WP_037757693.1
100	oxidoreductase	oxidoreductase	91%	WP_037757695.1
101	CYP450	hypothetical protein	74%	WP_037757696.1
102	glycosyl transferase	glycosyl transferase	94%	WP_037757699.1
103	aminotransferase	lipopolysaccharide biosynthesis protein Rfbh	95%	WP_037757701.1
104	short chain dehydrogenase	ketoacyl reductase	93%	WP_037757702.1
105	cyclase/dehydrase	actinorhodin polyketide cyclase	90%	WP_037757704.1
106	short chain dehydrogenase	ketoreductase	50%	WP_037758759.1
107	hypothetical	hydroxylacyl-CoA dehydrogenase	86%	WP_051648278.1

108	hypothetical	cyclase	94%	WP_037757865.1
109	methyltransferase	SAM-dependent methyltransferase	53%	KDN80044.1
110	aminotransferase	daunorubicin biosynthesis sensory transduction protein DnrJ	95%	WP_037757705.1
111	hypothetical	NDP-hexose 2,3-dehydratase	88%	WP_051648283.1
112	NAD-dependent epimerase/dehydratase	NAD-dependent epimerase	82%	KDN80058.1
113	hypothetical	dTDP-4-dehydrorhamnose 3,5-epimerase	89%	WP_037757707.1
114	glycosyl transferase	CosK	95%	ABC00725.1
115	PadR transcriptional regulator	CosS	95%	ABC00736.1
116	ornithine cyclodeaminase	hypothetical protein	81%	WP_037757710.1
117	hypothetical	CosX	91%	ABC00737.1
118	Beta-Ketoacyl Synthase (T2pks)	Beta-ACP synthase	94%	WP_037757712.1
119	Beta-Ketoacyl Synthase (clf)	Beta-ketoacyl synthase	90%	WP_037757715.1
120	3-oxoacyl ACP synthase	3-oxoacyl ACP synthase	91%	WP_037757720.1
121	malonyl-CoA ACP transacylase	CosF	83%	ABC00730.1
122	cytochrome CYP450	cytochrome CYP450	83%	WP_037757725.1
123	glycosyl transferase	glycosyl transferase family 28	94%	WP_037757728.1
124	NAD dependent epimerase/dehydratase	dTDP-glucose 4,6-dehydratase	94%	WP_037757730.1
125	ABC transporter ATP binding protein	CosI	94%	ABC00731.1
126	ABC-2 transporter	multidrug ABC transporter permease	97%	WP_037757733.1
127	hypothetical	glutathione peroxidase	87%	WP_037757871.1

128	ABC transporter ATP binding protein	daunorubicin resistance protein DrrC	94%	KDN80074.1
129	hypothetical	hypothetical protein DF19_01985	77%	KDN77779.1
130	PadR transcriptional regulator	transcriptional regulator	76%	KDN77778.1
131	hypothetical	membrane protein	51%	WP_031115629.1
132	hypothetical	ribonuclease BN	67%	KDN74392.1
133	hypothetical	membrane protein	86%	KDN73505.1
134	hypothetical	hypothetical protein	42%	WP_037764046.1
135	extracytoplasmic-function sigma-70 factor	RNA polymerase sigma factor	64%	KDN79303.1
136	hypothetical	hypothetical protein DF19_30470	47%	KDN79723.1
137	hypothetical	oxidoreductase	60%	KDN78332.1
138	hypothetical	hypothetical protein DF19_13800	31%	KDN77257.1
139	hypothetical	antitermination regulator	40%	WP_037761378.1
140	Sensor histidine kinase	histidine kinase	62%	WP_037761461.1
141	hypothetical	ATPase	43%	KDN79305.1
142	hypothetical	metal ABC transporter substrate-binding protein	52%	KDN79304.1

REFERENCES

(1) Kelso, C.; Rojas, J. D.; Furlan, R. L.; Padilla, G.; Beck, J. L. *Eur J Mass Spectrom* **2009,** *15*, 73-81.
(2) Johdo, O.; Watanabe, Y.; Ishikura, T.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T. *J Antibiot (Tokyo)* **1991,** *44*, 1130-1140
(3) Shimosaka, A.; Hayakawa, Y.; Nakagawa, M.; Furihata, K.; Seto, H.; Otake, N *Antibiot (Tokyo)* **1987,** *40*, 116-121.
(4) J. Ying, F. Delaglio, D.A. Torchia, and A. Bax. *J Biomol NMR*, in press **2016.**
(5) F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Baxz. *J. Biomol. NMR*, **1995**, *6*, 277-293.
(6) Johdo, O.; Watanabe, Y.; Ishikura, T.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T. *J Antibiot (Tokyo)* **1991**, *44*, 1121-1129.
(7) Johdo, O.; Tone, H.; Okamoto, R.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T., Anthracycline metabolites from Streptomyces violaceus A262. V. New anthracycline alldimycin A: a minor component isolated from obelmycin beer. *J Antibiot (Tokyo)* **1991**, *44*, 1160-1164.
(8) Johdo, O.; Tone, H.; Okamoto, R.; Yoshimoto, A.; Naganawa, H.; Sawa, T.; Takeuchi, T. *J Antibiot (Tokyo)* **1991**, *44*, 1155-1159.
(9) Bandeira, N.; Tsur, D.; Frank, A.; Pevzner, P. A. *Proc Natl Acad Sci U S A* **2007,** *104*, 6140-6145.
(10) Nguyen, D. D.; Wu, C. H.; Moree, W. J.; Lamsa, A.; Medema, M. H.; Zhao, X.; Gavilan, R. G.; Aparicio, M.; Atencio, L.; Jackson, C.; Ballesteros, J.; Sanchez, J.; Watrous, J. D.; Phelan, V. V.; van de Wiel, C.; Kersten, R. D.; Mehnaz, S.; De Mot, R.; Shank, E. A.; Charusanti, P.; Nagarajan, H.; Duggan, B. M.; Moore, B. S.; Bandeira, N.; Palsson, B. O.; Pogliano, K.; Gutierrez, M.; Dorrestein, P. C. *Proc Natl Acad Sci U S A* **2013,** *110*, E2611-2620.
(11) Crüsesmann, M.; O'Neill, E.; Larson, C.; Melnik, A.; Floros, D.; da Silva, R.; Jensen, P.; Dorrestein, P.; Moore, BS. *Journal of Natural Products* **2016**, epub November 11, 2016.
(12) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res* **2003**, *13*, 2498-2504.