



Supplementary Materials

Biosynthesis and Heterologous Expression of Cacaoidin, the First Member of The Lanthidin Family of Ripps

Fernando Román-Hurtado, Marina Sánchez-Hidalgo*, Jesús Martín, Francisco Javier Ortiz-López and Olga Genilloud

* Correspondence: marina.sanchez@medinaandalucia.es

Tables

|--|

Oligonucleotide	Sequence (5'-3')
Glyco1-sgRNA	TAATACGACTCACTATAGGACGACTCACGTGTCAAAGAGTTTTAGAGCTAGAAATAGCAA
Glyco2-sgRNA	TAATACGACTCACTATAGGGGCGAGATGCCATTCCAAGGTTTTAGAGCTAGAAATAGCAA
sgRNA-F	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC
sgRNA-R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT
pCAP01-Glyco-F	CGTGCGGTGGACCGCGCGTGACCCCCTTGTCGAGACTTGAGGTACCTGT
pCAP01-Glyco-R	GGGCCGGGTTCCAGCCGGTGATGCCGTCTTTCGAGGTTACTAGTCGATCT
cao4 check-F	CGCGCACTGACCGTCTG
cao4 check-R	CTGCTGCCGGACTGACC
cao8 check-F	CCAACATCGACCCGGACC
cao8 check-R	GTCACGCCGCCACACTC
cao15 check-F	TGGGGCATCTGATCAAGGAG
cao15 check-R	TCGCGCAGTACCTCGGTT

Fundación MEDINA, Avenida del Conocimiento, 34, 18016 Granada, Spain; fernando.roman@medinaandalucia.es (F.R-H), jesus.martin@medinaandalucia.es (J.M), javier.ortiz@medinaandalucia.es (F.J.O-L), olga.genilloud@medinaandalucia.es (O.G)

Table S2. Closest HHPred homolog for each ORF in cacaoidin BGC. HHPred probabilities give the most relevant rep										
resentation of sign	nificance with ≥90% usually being considered a true	positive.								
-	Closest HHpred Homolog	Probability	E-value							

	Closest HHpred Homolog	Probability	E-value
caoR1	Response regulator protein VraR; Enterococcus faecium, LiaR, response regulator	99.96	1.5e-25
caoR2	Methylphosphonate synthase	98.06	0.00017
cao3	dTDP-4-dehydrorhamnose reductase, RfbD ortholog	99.97	8e-28
cao4	O-methyltransferase family protein	100	6.2e-37
cao5	Adenosylhomocysteinase (E.C.3.3.1.1, 1.8.1.7	100	1.9e-37
саоб	Bifunctional protein Glm	100	7.8e-37
cao7	Type III effector HopA1	100	3.4e-39
cao8	Putative glycosyltransferase protein	99.94	2.4e-25
cao9	5-methylthioribose kinase (E.C.2.7.1.100)	99.9	2.3e-21
caoA	Myocyte-specific enhancer factor 2A	25.67	67
<i>ca</i> 011	Lipid A export ATP-binding/permease protein	100	1.3e-89
<i>ca</i> 012	Bacterial luciferase, 1,2-ethanediol; monooxygenase, flavoprotein	100	1e-41
caoD	MrsD protein	100	7.5e-32
cao14	Spectinomycin phosphotransferase; protein kinase, aminoglycoside phosphotransferase, antibiotic;	99.76	5.4e-17
cao15	Insulinase family protein; Protease, M16 Family, Processing Protease	100	3e-65
lcao16	Glycosyltransferase	99.95	3.5e-26
cao17	Light-harvesting protein B-875 alpha chain	54.44	10
<i>ca</i> 018	ABC transporter ATP-binding protein	100	2.5e-51
<i>ca</i> 019	ATP-binding cassette sub-family G member	99.93	7.2e-23
caoR3	Putative transposon-related DNA-binding protein	99.56	8.8e-14
cao21	Ubiquinol-cytochrome c reductase iron-sulfur subunit	59.39	56
caoR4	TRANSCRIPTIONAL REGULATORY PROTEIN (PROBABLY TETR- FAMILY); tetR family of transcriptional regulator	99.94	3.3e-24
cao23	Conserved protein; Structural genomics, unknown function, ligand;	99.89	7.2e-20
cao24	Glycosyl transferase (E.C.2.4.1.57)	100	5.8e-38
cao25	TarM; Glycosyltransferase, GT-A, Wall teichoic acid	97.73	0.0025
cao26	ABC transporter ATP-binding protein (E.C.3.6.3)	87.92	1.5
caoR5	Tetratricopeptide repeat sensor PH0952	100	1.2e-36

	Streptomy subsp. cac 12	ices cacaoi caoi NBRC 748	Strepton NRRL	nyces sp. F-5053	Strepto NRRI	myces sp. L S-1868	Streptomyces cacaoi subsp. cacaoi NRRL B-1220		Streptomyces ca- caoi subsp. ca- caoi OABC16	
	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity
caoR1	100	100	99.2	99.6	100	100	-	-	99.6	100
caoR2	97.9	99	98.1	98.1	34.4	40.6	-	-	97.9	99
cao3	97.8	98.1	99	99	96.6	97.1	99.6	100	98.1	98.3
cao4	99.1	99.7	99.7	100	99.1	99.7	99.1	99.7	100	100
cao5	97.9	98.8	97.5	98.2	98.2	98.8	97.9	98.8	98.2	98.8
cao6	100	100	99.7	100	100	100	100	100	99.2	99.7
cao7	99.7	99.7	98.9	99.4	99.1	99.4	99.4	99.7	99.7	100
cao8	98.7	99.5	99.5	99.5	99.5	99.7	98.7	99.5	99.2	99.5
cao9	99.7	99.7	99.7	100	100	100	99.7	99.7	99.5	100
caoA	100	100	100	100	100	100	100	100	100	100
cao11	99.1	100	99.7	100	99.1	100	99.1	100	99.5	100
cao12	99.7	100	100	100	99.7	100	99.7	100	99.7	100
caoD	99.6	100	92.1	92.1	93.3	93.3	99.6	100	94.6	95.7
cao14	99.3	99.3	99.3	99.3	98.7	99	-	-	99	99
cao15	98.6	99	98.6	98.8	99	99.3	-	-	98.4	98.8
cao16	100	100	99.7	99.7	99.4	99.4	-	-	100	100
cao17	-	-	-	-	100	100	-	-	100	100
cao18	99.1	99.1	100	100	99.7	99.7	-	-	99.4	99.4
cao19	99.7	99.7	99.7	99.7	99.7	99.7	-	-	99.3	99.7
caoR3	100	100	100	100	99	100	-	-	99.1	100
cao21	99.5	99.5	99.5	99.5	100	100	-	-	98.9	99.5
caoR4	100	100	100	100	100	100	-	-	100	100
cao23	99.4	99.4	100	100	98.8	98.8	-	-	99.4	100
cao24	100	100	100	100	100	100	-	-	100	100
cao25	99.1	99.1	96.9	97.8	98.8	99.1	-	-	98.4	98.8
cao26	-	-	-	-	-	-	-	-	-	-
caoR5	-	-	-	-	-	-		-		-

Table S3. BLAST homology between the ORFs in cacaoidin BGC and NCBI found clusters.

Figures

~	1	10	20	30	40	50	60	70 8	oe 90	100	110	122
a Consensus	MSXX:	******	*****	****	(XXXXXXXXXX	*******	*****	<u> XI</u>	*****	<u> </u>	*******	K
MDN-0207		MG	MAGE			GAPESS	APCIN	A5	SASI SAT	ASW	6 <mark>C</mark>	
Actagardine	MSAL	ATEX SUM DVD	LEDGAT SHEA			DRT YAAS	SGWYCT		CG		CA C	
Lichenicidin VK21 A1	MSRK	MILL SUK NPV	NRTES SHPA			HSLAGGTI	TISTCA				W-PSCN	
Moreacidin	NOR		CORNETONRA			GAG	AND A ACT		Besedicti			
Cytolysin Ls	1/10		NKI ELVG-PS			CSC DALL	PACET-				SA-MEC	
Haloduracin aloha	VT	EVEND	ERTHNNSNPA			ACUNCA	CAWWA				W-PSCN	
Haloduracin beta	1/10	SKDER-NP				AGSOBVEADT		GWSW			T-50C	
Lichenicidin VK21 A2	1/10	IMENSA-ARE	AREGAN HR	AG			PATTS SWICE TA	(WINS			T-SRC	
Mutacin 1140	1710					PETR			PGCA PT			
Mutacin A'	1/10		K-KOV-COLE			SECRETUS		CP	PD CT S			
Mutacin B Nu266	N III								PCCA NT			
Mutacin B-Ny200	N (2)								DDVI I			
Corocidin A1												
Corpolycip A1	191											
Cutolysin L								ATAAA				
EVA2 b								Real	33///C 000 30			
FIVAZ.0						AGAGE			G			
Ninchigatini A					SP SM					co in in		
Minabia aziria A2						S WINNE AGO	AGUSE					
Daanisidin A												
Paericium A					and the second sec	SSGC GSG	ANTOROU		DECTENCE		-	
Planosponcin		NC155-8	APONTAL		-	-Non Apr	ALL SUSA		Bacits au	GSG USH		
SAL-2242		INFO OF	P 11 Polys	QAN				G.				
Lacticin 3147 AZ		DIKEN					GGTTPA	(IF	A SILISAN IS	I NIC	M-TECTEAC	
FIVA La		MAC	GNW/VCLUNN	SUALIAIN	GAGIN		GINGU		a ag cc			
Nukacin A		MENS		ANLLEWU		GA	KKKSGV	P			Ca Ca	
Ginnamurin				AN								
Cirinamycin Eluad		MIA	SHELL SWELA	TERAALLEN	HAA	CASARATE			AATEA	MACINUSC SING	GNIK	
FIVAZ.d		MS										
Lacucin 3147 AT		INTERNET					SUNUES.					
Plantaricin w alpha		MIST	HADA		INSN	INC.	-C ROWA		GHICH			
SINDB		ME		AN					GIGICIN			
		ME	MININAGETP				WGIE		POICENS			
HVAZ.n		ME	GHEAGVER			GGTSSICVE	ASNTPE GT		RECPSAA-		MSCR	
Lactocin S		MR	I BERKYLDEL S	HASAMG	AROV	SSWINALS	IPVUASVA		- NELOPTASVINS	DVAGC EKY	SAN HHC	
BaccH91		M		DUDVOVKA		SOSAGO	RITSEIC	CT	RGCGKTG-	SE NSE	IC	
Epidermin			PAVKEKNOL	NEDVKVNA		SGARB	RIASKEI	CT	PCCAKTG-	SI N5M	C	-
Plantaricin w beta		M	TETSREENA	ANY EPVD		GAGEP		C	GAAMAAS	MAN .	PTTKCSKRCCKRK	K
Bovicin			MNNAT -	JIEV IMS		GG	A D GW			SICAGI ITA	M-N-CA	
Epicialn					MENN	AUL AQS	GFAIR	AI	ROM CE	KAI REV	WSCRNSECC	
Epilancin 15X				DENENKOI	AQK	SPENROS	ASIXX				TCGCHETGKK	
Epilancin K7			MNNSLE			SOUSROS	ASVLK	TS			TCGCNITGG	
Geobacillin II	-		MERCEOTE	VSKI SEEE		AGGMTEMS	POSTIN			CP	SEKWACEM	
Labyrinthopeptin A17A	3		MASIL			SSAADSN	ASMW	CC	<mark>STGSWW</mark>	PET		
Lacticin 481			MKEQNS			GA	KGGSCW		SHECMMNSW		[CS	
Macedovicin			MMNAT -	TWO TWS	DONI	GG	A			SICAGE TA	M-N-CA	
Mutacin II			MINKENSNA	VVSLINEVS		GGN		P	MSY CRANSA		1C	
Paenibaciilin				DUDURKSM	A	SCUSPO	ASIE		II KM SK			
Pep5			MENDIENCE		SQN1	Refe Pol	AGFAIR	A3	MRQCORTE			
Prochiorosin 1.1			MSHE CAL	ALLANU	ADIST		WA ARAAGES		HHUIESUDDLEGNA	GGEN LNG	JI-A-NEE ISINAC	
Ruminococcin A			MENUM			G	66-NGVI				10	
Salivaricin A			MENSKULL	NNATEENS	EXEL	MEMAGG	KRGSCW	AT		MEV.	1C	
Streptococcin A FF22				NSI QEVS		G	AGENCA		SHICHININ	A		
Inermophilin 1277			MMNAI -			GG	Alen Gw				M-N-CA	
Variacin			ML NZ	QALLEY		G	GGSGN				Call Call	
BritA2 Streptopollin						AGGS	TEACA			STAC		
Streptocollin				A-		GAS	SGICAC	-			G-SCA	
Catenunpeptin												
Coobacillin												
Nisin A						SGACD					ATCHCSHH	
Nisin II			STATE -			NNSGAED		<u></u>			ATCGCHIC	
Subtilin								(1	Bachton			
Ervthreanentin			Lallin a	VLELOELE	APNE	AGOPS	HGRGE		NESE ASCAN	S TH		
NAL 112			Lange and		SASA	TEONR		WS			NC	
Avermipentin			M4					-A	STWSLESC	544		
Griseonentin			MA		TPAR-			T G		55	W-MECTP	
SapB			6111					TG		558	-TTCN	
SR015-2212			6120	QANE	TPAR-			T G		55	W-MCTP	
Stackepentin A			6173		SEADE	SATE GG	<mark>55</mark> 65 GHS	CIP		I SG	- I	
Curvopentin				MED OGNT	PERM		GGG		SSASL LICE			
curvopeptin			1.00				666	-	55761	3 /10		
	1	10	on		80	40	50	60	70	RO	90 10	0 10-
Conconsus	K ADDEN	CANCENAN	20			TY YYYYYYYY	Ĩ	Ϋ́	YYYYYYY	T YYYYYYYYY		Men
D	MICH		MAMMAAAA/									FAIDING
MDN-0207			MC		ADVE		APESSAPCT	IAS	WSASISAT	ASWGC		
Cypemycin		MRS	MILTSTNS	ALALA	AQDE ANT-N	/LSAAAPG	HACCTPA-MA	ATPATPT	VACEVIC	-G <mark>STIC</mark>	EVC	
FrEUN1f DRAFT 0188		MSMRSE	PGSLRLSQL	ARIDALIT	AQSEC G	SDRERIHI	TEEQAAATPD/	AHHPLFOLSE H	DREILNQUIELT	GOLEHTTS		G
Grisemycin		MRL		TAPESMA	TODE ANS-	AGAVPG	HSDAETPA-MA	TPA	NAGEVIC	-G <mark>STIC</mark>	EVC	
Htur 3018				MS	SQTTEGWS	FTSGIVIL	VLK-ALPG			MELANG	LVLLYYR	
Legonaridin				MSM	ALL ANT-	LVDVEPGR	GSEATPT-M	TPLAT ATPL	ATPRGAATSAT	-AAAMN	MITHOVIEH	
Lxx19440				MSLOKT	AQTENDR-N	TTEIDPOA	TASTSVEH-R		GAAADVAA	-DVATGASA	AATALAL	
Lxx19470				MSLORT.	AQAL SOL-N	ATELNIOV	TAGSSVPL -H	VTCLGVIA		-DVASGAAA	AVITESV	
Mvan 2782			M	RVIVI I VTT.	AERERAA-A	AHPAATVS		VVSLLVGLGV	GWANATTAAA	PAVESGAG		TSRR
Nmul A1818			MVII			MTSDTSDK		GTAVAG		GUSUG		
SCAB 84151				MNDFLL	PSIVALG	GELIATR	AVSTPAVI GVA	VVIWG	ACSQARSTME	PRHTSSK	KRRQHP	
SCAB 84201		MSS		I PV GVVY	AARPTLG	PRIARIGR		AAAATAGVGV		ANAQNAAA	GGAPAS	
SGR 1511				MPEERQ	PONTROMAR		GGASPN-A	TPAWATAGA	WVALAACMAA	-HEANGG		
SGR 1512		MAS	PHRPPTSEE	KSENTAS		AL SE SDHGR		T		-NCETGS		
SGR 1513				MNKSSA	PAVLTATGA			TP-VALAA-T	MGVAFVAGYAAC	-RAATG	NMEEPM	
SGR 1514				MSMS PT	PAALRGAG	GESESDEGE	ALSSLAPY-TA		GVALVNALAAG	-MNHCGG	NVELPM	
SvirD4 22614	MPRE	SAPSERTA	GELALRIT	RKRTTMSV	ADE ANT-	ADVTPGR	GNDATPT-M	TPLAALATP	GNANTAAT	-ANALN	EVTNDEAG	

Figure S1. Alignment of the structural amino acid sequence (leader peptide + core) of cacaoidin (MDN-0207) with the sequences of other already known lanthipeptides **(a)** and linaridins **(b)**.



Figure S2. Schematic representation of the biosynthesis of dTDP-L-rhamnose from D-glucose.



Figure S3. Proposed pathway for the β -L-6-deoxy-gulose sugar biosynthesis for the BLM, TLM and ZMB compounds.





Figure S4. Neighbor-joining tree built with Mega X based on nearly complete 16S rRNA gene sequences of CA-170360, the 50 closest type strains of the genus *Streptomyces* and three strains of *Streptomyces cacaoi* subsp. *asoensis*. *Micromonospora auratinigra* DSM 44815 (T) was used as an outgroup. The numbers at the nodes indicate the bootstrap value (%) based on NJ analysis of 1000 replicates; only values higher than 50% are shown.