Supplementary Information

Functional exchangeability of oxidase and dehydrogenase reactions in the biosynthesis of hydroxyphenylglycine, a non-ribosomal peptide building block

Veronica Diez, Mark Loznik, Sandra Taylor, Michael Winn, Nicholas Rattray, Helen Podmore, Jason Micklefield, Royston Goodacre, Marnix H. Medema, Ulrike Müller, Roel Bovenberg, Dick B. Janssen and Eriko Takano*

Table SI. BlastP Hits obtained using the first 276 amino acid residues of Caci3265

Description[a]	Total Score[b]	Query cover[c]	E-value[d]	Ident[e]	Accession[f]
3-oxoacyl-[acyl-carrier-protein] synthase III *Clostridium perfringens* ATCC 13124	35.0	47	0.45	28	Q0TRH0.1
3-oxoacyl-[acyl-carrier-protein] synthase III *Arabidopsis thaliana*	34.7	60	0.64	24	P49243.2
3-oxoacyl-[acyl-carrier-protein] synthase III *Streptococcus agalactiae* 2603V/R	33.9	48	1.1	23	P64113.1
3-oxoacyl-[acyl-carrier-protein] synthase IIIA, chloroplastic. *Cuphea wrightii*	33.5	29	1.9	28	P49244.2
3-oxoacyl-[acyl-carrier-protein] synthase III *Polynucleobacter necessarius*	32.3	57	3.4	26	A4SW4.1
Probable pectin lyase A *Neosartorya fischeri* NRRL 181	32.3	15	3.6	32	A1CYC2.2
Protein translocase subunit SecY *Pyropia yezoensis*	32.3	24	3.8	30	Q1XDJ1.1
3-oxoacyl-[acyl-carrier-protein] synthase III *Prochlorococcus marinus* str. MIT 9313	32.3	39	3.9	28	Q7V4F6.1
3-oxoacyl-[acyl-carrier-protein] synthase III *Aromatoleum aromaticum* EbN1	32.3	55	4.1	28	Q5P0D6.1
3-oxoacyl-[acyl-carrier-protein] synthase III *Bradyrhizobium diazoefficiens* USDA 110	32.0	60	4.5	23	Q89K89.1
3-oxoacyl-[acyl-carrier-protein] synthase III *Azoarcus* sp. BH72	32.0	51	4.7	27	A1K5Y5.1
3-oxoacyl-[acyl-carrier-protein] synthase III *Microcystis aeruginosa* NIES-843	32.0	44	5.3	24	B0JXE2.1

BlastP against Swiss-Prot database.
a. Only the best 12 hits are shown.
b. Total score: the total alignment scores from all alignment segments
c. Query cover: the percentage of query covered by alignment to the database sequence (%)
d. E-value: the best (lowest) Expect value of all alignments from that database sequence
e. Ident: the highest percent identity of all query-subject alignments (%)
d. Accession number of the matched database sequence

Table SII. Primers used for sequencing

Primer name	Primer sequence 5'-3'	Template	Fragment size[a]
Amir4598 fw-out	TCGTCGAAGTCCAGGGTGTC	*A. mirum* DSM 43827	1644
Amir4598 rv-out	GATTTGCCGACCGGGAATGG		
Caci3265 fw-out	CCGTCAATGACAGAGCAGAG	*C. acidiphila* DSM 44928	2546
Caci3265 rv-out	GCGGAAGACCTGGAGATACC		
Csal1075 fw-out	TATGGCCGATGCTGTTGATG	*C. salexigens* DSM 3043	1505
Csal1075 fw-out	TGCCGAACTCTTCCGAAAGC		
Helo1144 fw-out	CGCCGTGTTCGGTGGAATAC	*H. elongata* DSM 3044	1493
Helo1144 rv-out	GCGCTTGTAGTGTCGTATTC		
Kfla6052 fw-out	TCCTCGATGGCGTAGAACTG	*K. flavida* DSM 17836	1653
Kfla6052 rv-out	TCTACCCGAACACGCTCTAC		
SCO3228 fw-out	CCTCGTTGAGGAAGTTCATC	*S. coelicolor* M145	1462
SCO3228 rv-out	GCTTCGGCAGTTCCAACATC		
Snov4068 fw-out	GCGGCCGGACTTCTACAATG	*S. novella* DSM 506	1834
Snov4068 rv-out	CACTCGACGGACCTTTATAC		

a. Size of PCR amplified product in bp.

Table SIII. Primers used for cloning

Primer name	Primer sequence 5'-3'[a]	Template	Fragment size[b]
Amir4598 rv XhoI	**CTCGAG**CCGGCCGTCGTCTTCC	*A. mirum* DSM 43827	1198
Amir4598 fw NdeI	**CATATG**GGCGTCCGCAACTCC		
Caci3265 fw NdeI	**CATATG**GCCGTGCATACAACTGTCCCG	*C. acidiphila* DSM 44928	2058
Caci3265 rv XhoI	**CTCGAG**GCACGTTCAGGGCTCCATC		
Csal1075 fw NdeI	**CATATG**AAGAGAAGCATGCCGCGCAGAC	*C. salexigens* DSM 3043	1232
Csal1075 rv XhoI	**CTCGAG**ACCGCCCATCCTACCCTG		
Helo1144 fw HindIII	**AAGCTT**TGAGGTGACGGGTCAAC	*H. elongata* DSM 3044	1206
Helo1144 rv NdeI	**CATATG**GTGAAACGCCGTCCCTATG		
Kfla6052 fw XhoI	**CTCGAG**ATCAGCCACAGGCTCTCC	*K. flavida* DSM 17836	1144
Kfla6052 rv NdeI	**CATATG**CCGATGGGTGACGGAG		
SCO3228 fw XhoI	**CTCGAG**TCATCCGTGGCTCCTGTC	*S. coelicolor* M145	1146
SCO3228 rv NdeI	**CATATG**GTGCGAGAGCCGCTCACG		
Snov4068 fw HindIII	**AAGCTT**CGCTTCGAGGCTTCTCAG	*S. novella* DSM 506	1214
Snov4068 rv NdeI	**CATATG**AAGATCGAGAGGATGATCAC		

a. Sequences in bold represent restriction sites added to the original sequence. The names of the restriction enzymes recognizing these sites are included in the names of the primers.

b. Size of PCR amplified products in bp.

Table SIV. Comparison of the CDA derivatives produced in M1144 and LW139.[a]

	Apex RT	Start RT	End RT	Area	%Area	Height	%Height
M1144 CDA-4B	9.78	9.74	9.85	5224461597	0.89	2087381264	2.02
M1144 CDA-4A	9.98	9.85	10.07	3361268084	0.57	958614880	0.93
LW139 CDA-4A	9.94	9.93	10.07	37768147.1	0.02	55246448.3	0.16

a. Data were extracted from continuum spectra generated from the Q-Exactive system. Each chromatogram was visualized within the Thermo Xcalibur software and deconvolved using the inbuilt Genesis algorithm. This processing generated the relevant area-under-curve information for each CDA peak. The data generated were cross-referenced via retention time and associated mass spectra and subsequently compared against the total-ion-count of each full spectrum yielding a % area contribution to each sample analyzed.

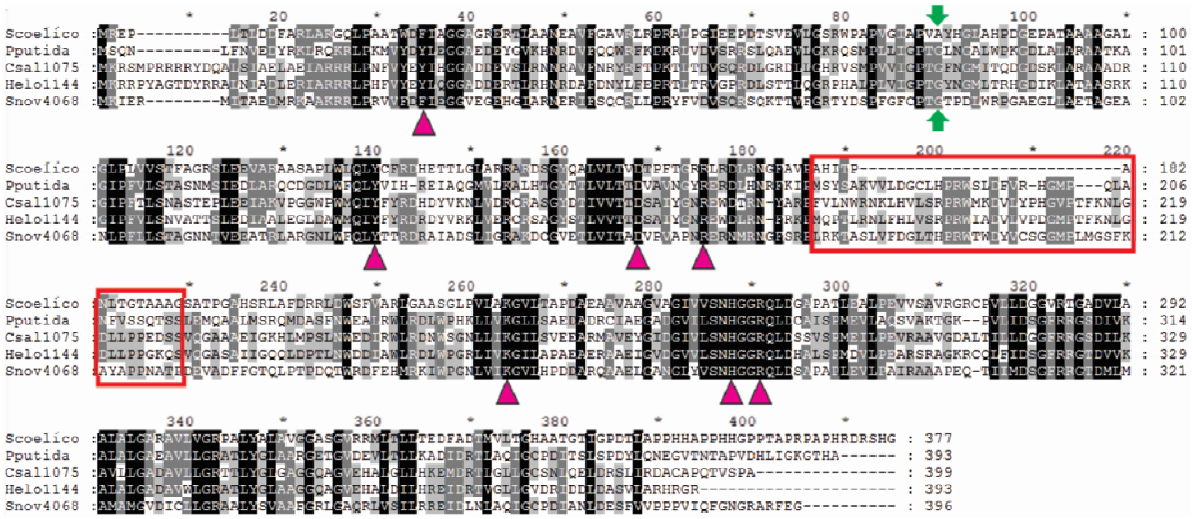
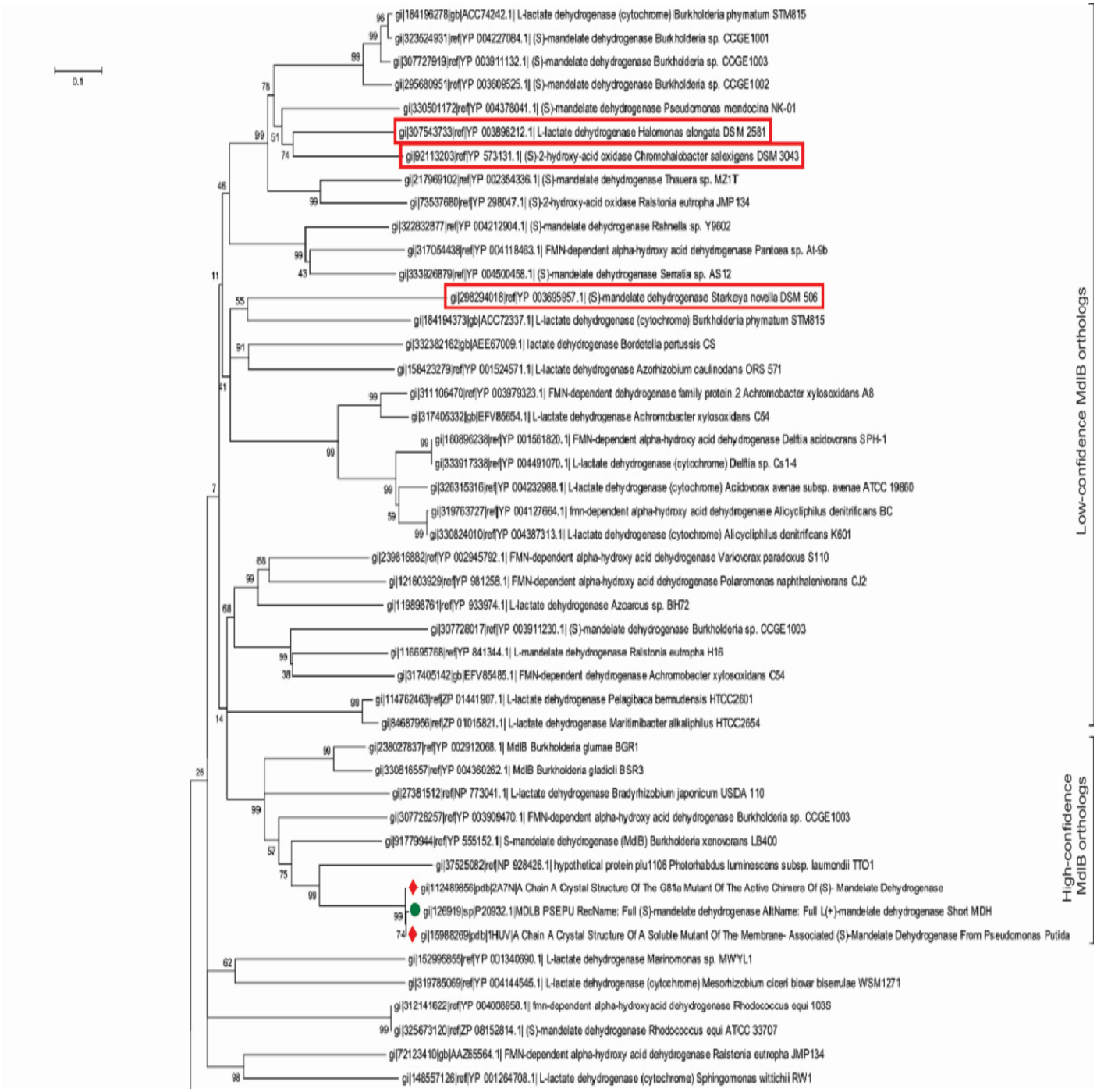
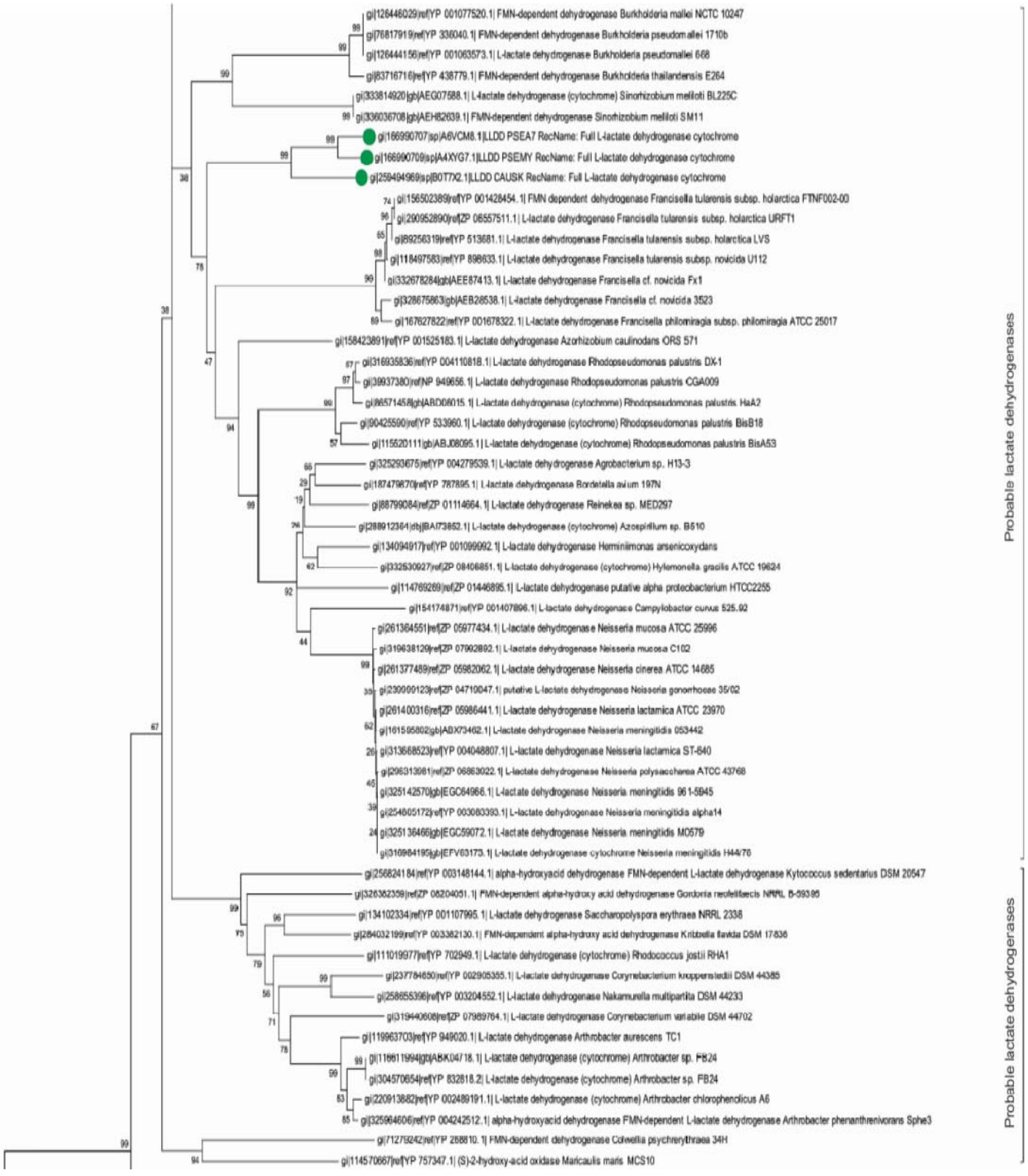


Figure S1. Alignment between the *S. coelicolor* HmO (Scoelico = Sco3228) and the L-mandelate dehydrogenase (MdlB) sequences from *P. putida* (Pputida), *C. salexigens* (Csal1075), *H. elongata* (Helo1144), and *S. novella* (Snov4068). The boxed segment of 39-aminoacids in *P. putida* MdlB (residues 177-215) is implicated in membrane association.[17] The pink triangles indicate the active site residues, Tyr^{26}, Tyr^{131}, Asp^{158}, Arg^{165}, Lys^{249}, His^{273}, and Arg^{276} that are highly conserved among the α-hydroxy acid oxidizing enzymes and are believed to be important for catalyzing the substrate oxidation half reaction.[22] The green arrows indicate the Gly81 residue in *P. putida* MdlB implicated in binding to the FMN that is conserved in the other dehydrogenases but changed to Ala in the oxidases.[34] Alignment was generated by using the CLUSTALW program.


0.1
gi|184196278|gb|ACC74242.1| L-lactate dehydrogenase (cytochrome) Burkholderia phymatum STM815
gi|323624931|ref|YP 004227084.1| (S)-mandelate dehydrogenase Burkholderia sp. CCGE1001
gi|307727919|ref|YP 003911132.1| (S)-mandelate dehydrogenase Burkholderia sp. CCGE1003
gi|295680951|ref|YP 003609525.1| (S)-mandelate dehydrogenase Burkholderia sp. CCGE1002
gi|330501172|ref|YP 004378041.1| (S)-mandelate dehydrogenase Pseudomonas mendocina NK-01
gi|307543733|ref|YP 003896212.1| L-lactate dehydrogenase Halomonas elongata DSM 2581
gi|92113203|ref|YP 573131.1| (S)-2-hydroxy-acid oxidase Chromohalobacter salexigens DSM 3043
gi|217969102|ref|YP 002354336.1| (S)-mandelate dehydrogenase Thauera sp. MZ1T
gi|73537680|ref|YP 298047.1| (S)-2-hydroxy-acid oxidase Ralstonia eutropha JMP134
gi|322832877|ref|YP 004212904.1| (S)-mandelate dehydrogenase Rahnella sp. Y9602
gi|317054438|ref|YP 004118463.1| FMN-dependent alpha-hydroxy acid dehydrogenase Pantoea sp. At-9b
gi|333926879|ref|YP 004500458.1| (S)-mandelate dehydrogenase Serratia sp. AS12
gi|298294018|ref|YP 003695957.1| (S)-mandelate dehydrogenase Starkeya novella DSM 506
gi|184194373|gb|ACC72337.1| L-lactate dehydrogenase (cytochrome) Burkholderia phymatum STM815
gi|332382162|gb|AEE67009.1| lactate dehydrogenase Bordetella pertussis CS
gi|158423279|ref|YP 001524571.1| L-lactate dehydrogenase Azorhizobium caulinodans ORS 571
gi|311106470|ref|YP 003979323.1| FMN-dependent dehydrogenase family protein 2 Achromobacter xylosoxidans A8
gi|317405332|gb|EFV85654.1| L-lactate dehydrogenase Achromobacter xylosoxidans C54
gi|160896238|ref|YP 001561820.1| FMN-dependent alpha-hydroxy acid dehydrogenase Delftia acidovorans SPH-1
gi|333917338|ref|YP 004491070.1| L-lactate dehydrogenase (cytochrome) Delftia sp. Cs1-4
gi|326315316|ref|YP 004232988.1| L-lactate dehydrogenase (cytochrome) Acidovorax avenae subsp. avenae ATCC 19860
gi|319763727|ref|YP 004127664.1| fmn-dependent alpha-hydroxy acid dehydrogenase Alicycliphilus denitrificans BC
gi|330824010|ref|YP 004387313.1| L-lactate dehydrogenase (cytochrome) Alicycliphilus denitrificans K601
gi|239816882|ref|YP 002945792.1| FMN-dependent alpha-hydroxy acid dehydrogenase Variovorax paradoxus S110
gi|121603929|ref|YP 981258.1| FMN-dependent alpha-hydroxy acid dehydrogenase Polaromonas naphthalenivorans CJ2
gi|119898761|ref|YP 933974.1| L-lactate dehydrogenase Azoarcus sp. BH72
gi|307728017|ref|YP 003911230.1| (S)-mandelate dehydrogenase Burkholderia sp. CCGE1003
gi|116695768|ref|YP 841344.1| L-mandelate dehydrogenase Ralstonia eutropha H16
gi|317405142|gb|EFV85485.1| FMN-dependent dehydrogenase Achromobacter xylosoxidans C54
gi|114762463|ref|ZP 01441907.1| L-lactate dehydrogenase Pelagibaca bermudensis HTCC2601
gi|84687956|ref|ZP 01015821.1| L-lactate dehydrogenase Maritimibacter alkaliphilus HTCC2654
gi|238027837|ref|YP 002912068.1| MdlB Burkholderia glumae BGR1
gi|330818557|ref|YP 004360262.1| MdlB Burkholderia gladioli BSR3
gi|27381512|ref|NP 773041.1| L-lactate dehydrogenase Bradyrhizobium japonicum USDA 110
gi|307726257|ref|YP 003909470.1| FMN-dependent alpha-hydroxy acid dehydrogenase Burkholderia sp. CCGE1003
gi|91779944|ref|YP 555152.1| S-mandelate dehydrogenase (MdlB) Burkholderia xenovorans LB400
gi|37525082|ref|NP 928426.1| hypothetical protein plu1106 Photorhabdus luminescens subsp. laumondii TTO1
gi|112480856|pdb|2A7N|A Chain A Crystal Structure Of The G81a Mutant Of The Active Chimera Of (S)- Mandelate Dehydrogenase
gi|126919|sp|P20932.1|MDLB PSEPU RecName: Full (S)-mandelate dehydrogenase AltName: Full L(+)-mandelate dehydrogenase Short MDH
gi|15988269|pdb|1HUV|A Chain A Crystal Structure Of A Soluble Mutant Of The Membrane- Associated (S)-Mandelate Dehydrogenase From Pseudomonas Putida
gi|152995855|ref|YP 001340690.1| L-lactate dehydrogenase Marinomonas sp. MWYL1
gi|319785069|ref|YP 004144545.1| L-lactate dehydrogenase (cytochrome) Mesorhizobium ciceri biovar biserrulae WSM1271
gi|312141622|ref|YP 004008958.1| fmn-dependent alpha-hydroxyacid dehydrogenase Rhodococcus equi 103S
gi|325673120|ref|ZP 08152814.1| (S)-mandelate dehydrogenase Rhodococcus equi ATCC 33707
gi|72123410|gb|AAZ65564.1| FMN-dependent alpha-hydroxy acid dehydrogenase Ralstonia eutropha JMP134
gi|148557126|ref|YP 001264708.1| L-lactate dehydrogenase (cytochrome) Sphingomonas wittichii RW1
Low-confidence MdlB orthologs
High-confidence MdlB orthologs


gi|126446029|ref|YP 001077520.1| FMN-dependent dehydrogenase Burkholderia mallei NCTC 10247
gi|76817919|ref|YP 336040.1| FMN-dependent dehydrogenase Burkholderia pseudomallei 1710b
gi|126444156|ref|YP 001063573.1| L-lactate dehydrogenase Burkholderia pseudomallei 668
gi|83716716|ref|YP 438779.1| FMN-dependent dehydrogenase Burkholderia thailandensis E264
gi|333814920|gb|AEG07588.1| L-lactate dehydrogenase (cytochrome) Sinorhizobium meliloti BL225C
gi|336036708|gb|AEH82639.1| FMN-dependent dehydrogenase Sinorhizobium meliloti SM11
gi|166990707|sp|A6VCM8.1|LLDD PSEA7 RecName: Full L-lactate dehydrogenase cytochrome
gi|166990709|sp|A4XYG7.1|LLDD PSEMY RecName: Full L-lactate dehydrogenase cytochrome
gi|259494969|sp|B0T7X2.1|LLDD CAUSK RecName: Full L-lactate dehydrogenase cytochrome
gi|156502389|ref|YP 001428454.1| FMN dependent dehydrogenase Francisella tularensis subsp. holarctica FTNF002-00
gi|290952890|ref|ZP 06557511.1| L-lactate dehydrogenase Francisella tularensis subsp. holarctica URFT1
gi|89256319|ref|YP 513681.1| L-lactate dehydrogenase Francisella tularensis subsp. holarctica LVS
gi|118497583|ref|YP 898633.1| L-lactate dehydrogenase Francisella tularensis subsp. novicida U112
gi|332678284|gb|AEE87413.1| L-lactate dehydrogenase Francisella cf. novicida Fx1
gi|328675863|gb|AEB28538.1| L-lactate dehydrogenase Francisella cf. novicida 3523
gi|167627822|ref|YP 001678322.1| L-lactate dehydrogenase Francisella philomiragia subsp. philomiragia ATCC 25017
gi|158423891|ref|YP 001525183.1| L-lactate dehydrogenase Azorhizobium caulinodans ORS 571
gi|316935836|ref|YP 004110818.1| L-lactate dehydrogenase Rhodopseudomonas palustris DX-1
gi|39937380|ref|NP 949656.1| L-lactate dehydrogenase Rhodopseudomonas palustris CGA009
gi|86571458|gb|ABD06015.1| L-lactate dehydrogenase (cytochrome) Rhodopseudomonas palustris HaA2
gi|90425590|ref|YP 533960.1| L-lactate dehydrogenase (cytochrome) Rhodopseudomonas palustris BisB18
gi|115520111|gb|ABJ08095.1| L-lactate dehydrogenase (cytochrome) Rhodopseudomonas palustris BisA53
gi|325293675|ref|YP 004279539.1| L-lactate dehydrogenase Agrobacterium sp. H13-3
gi|187479870|ref|YP 787895.1| L-lactate dehydrogenase Bordetella avium 197N
gi|88799084|ref|ZP 01114664.1| L-lactate dehydrogenase Reinekea sp. MED297
gi|288912364|dbj|BAI73852.1| L-lactate dehydrogenase (cytochrome) Azospirillum sp. B510
gi|134094917|ref|YP 001099992.1| L-lactate dehydrogenase Herminiimonas arsenicoxydans
gi|332530927|ref|ZP 08406851.1| L-lactate dehydrogenase (cytochrome) Hylemonella gracilis ATCC 19624
gi|114769269|ref|ZP 01446895.1| L-lactate dehydrogenase putative alpha proteobacterium HTCC2255
gi|154174871|ref|YP 001407896.1| L-lactate dehydrogenase Campylobacter curvus 525.92
gi|261364551|ref|ZP 05977434.1| L-lactate dehydrogenase Neisseria mucosa ATCC 25996
gi|319638129|ref|ZP 07992892.1| L-lactate dehydrogenase Neisseria mucosa C102
gi|261377489|ref|ZP 05982062.1| L-lactate dehydrogenase Neisseria cinerea ATCC 14685
gi|239099123|ref|ZP 04719047.1| putative L-lactate dehydrogenase Neisseria gonorrhoeae 35/02
gi|261400316|ref|ZP 05986441.1| L-lactate dehydrogenase Neisseria lactamica ATCC 23970
gi|161505802|gb|ABX73462.1| L-lactate dehydrogenase Neisseria meningitidis 053442
gi|313668523|ref|YP 004048807.1| L-lactate dehydrogenase Neisseria lactamica ST-640
gi|296313981|ref|ZP 06863022.1| L-lactate dehydrogenase Neisseria polysaccharea ATCC 43768
gi|325142570|gb|EGC64966.1| L-lactate dehydrogenase Neisseria meningitidis 961-5945
gi|254805172|ref|YP 003083393.1| L-lactate dehydrogenase Neisseria meningitidis alpha14
gi|325136466|gb|EGC59072.1| L-lactate dehydrogenase Neisseria meningitidis M0579
gi|319964195|gb|EFV63173.1| L-lactate dehydrogenase cytochrome Neisseria meningitidis H44/76
gi|256824184|ref|YP 003148144.1| alpha-hydroxyacid dehydrogenase FMN-dependent L-lactate dehydrogenase Kytococcus sedentarius DSM 20547
gi|326382359|ref|ZP 08204051.1| FMN-dependent alpha-hydroxy acid dehydrogenase Gordonia neofelifaecis NRRL B-59395
gi|134102334|ref|YP 001107995.1| L-lactate dehydrogenase Saccharopolyspora erythraea NRRL 2338
gi|284032199|ref|YP 003382130.1| FMN-dependent alpha-hydroxy acid dehydrogenase Kribbella flavida DSM 17836
gi|111019977|ref|YP 702949.1| L-lactate dehydrogenase (cytochrome) Rhodococcus jostii RHA1
gi|237784650|ref|YP 002905355.1| L-lactate dehydrogenase Corynebacterium kroppenstedtii DSM 44385
gi|258655396|ref|YP 003204552.1| L-lactate dehydrogenase Nakamurella multipartita DSM 44233
gi|319440608|ref|ZP 07989764.1| L-lactate dehydrogenase Corynebacterium variabile DSM 44702
gi|119963703|ref|YP 949020.1| L-lactate dehydrogenase Arthrobacter aurescens TC1
gi|116811994|gb|ABK04718.1| L-lactate dehydrogenase (cytochrome) Arthrobacter sp. FB24
gi|304570654|ref|YP 832818.2| L-lactate dehydrogenase (cytochrome) Arthrobacter sp. FB24
gi|220913682|ref|YP 002489191.1| L-lactate dehydrogenase (cytochrome) Arthrobacter chlorophenolicus A6
gi|325964606|ref|YP 004242512.1| alpha-hydroxyacid dehydrogenase FMN-dependent L-lactate dehydrogenase Arthrobacter phenanthrenivorans Sphe3
gi|71275242|ref|YP 268810.1| FMN-dependent dehydrogenase Colwellia psychrerythraea 34H
gi|114570667|ref|YP 757347.1| (S)-2-hydroxy-acid oxidase Maricaulis maris MCS10
Probable lactate dehydrogenases
Probable lactate dehydrogerases

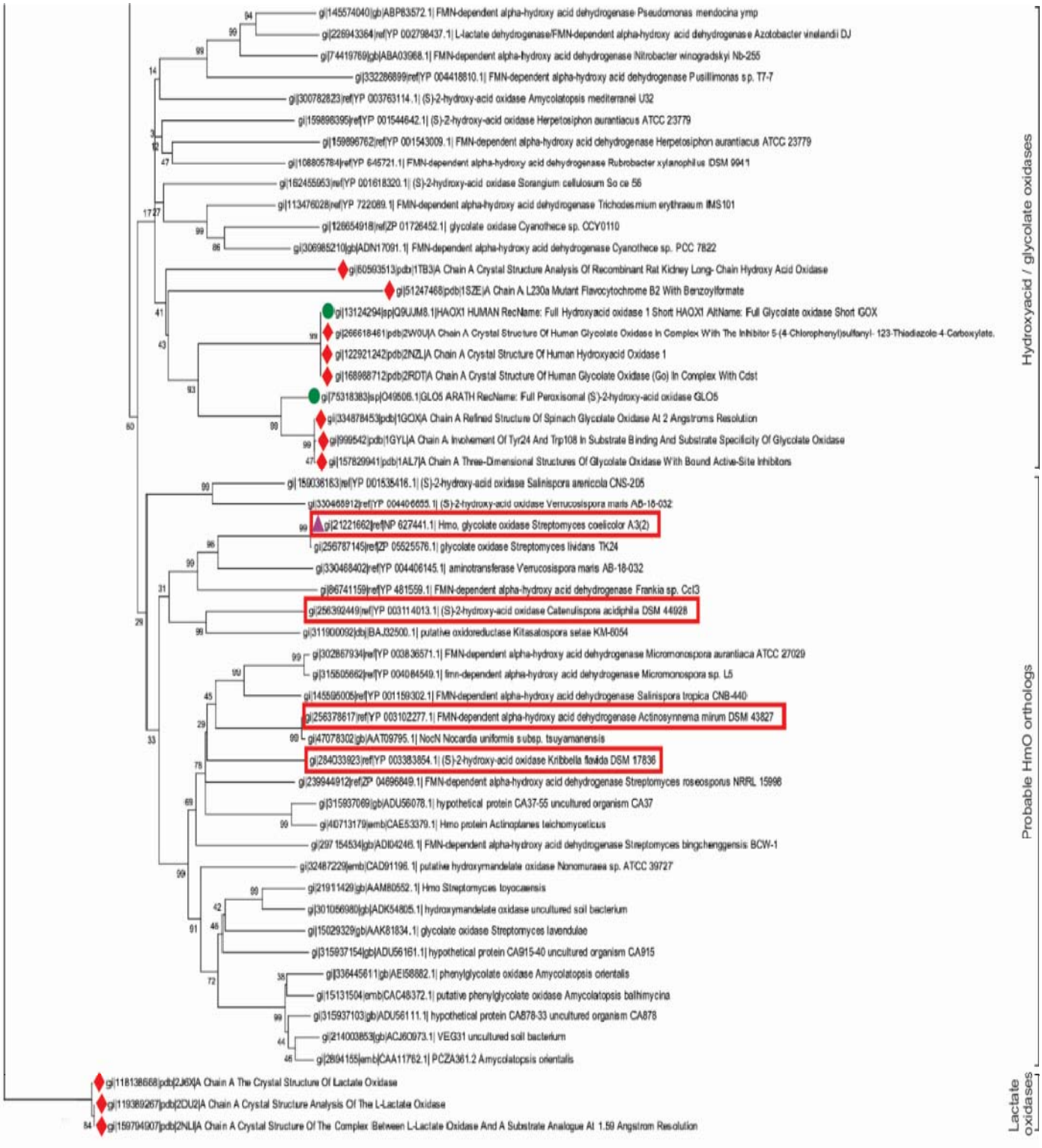
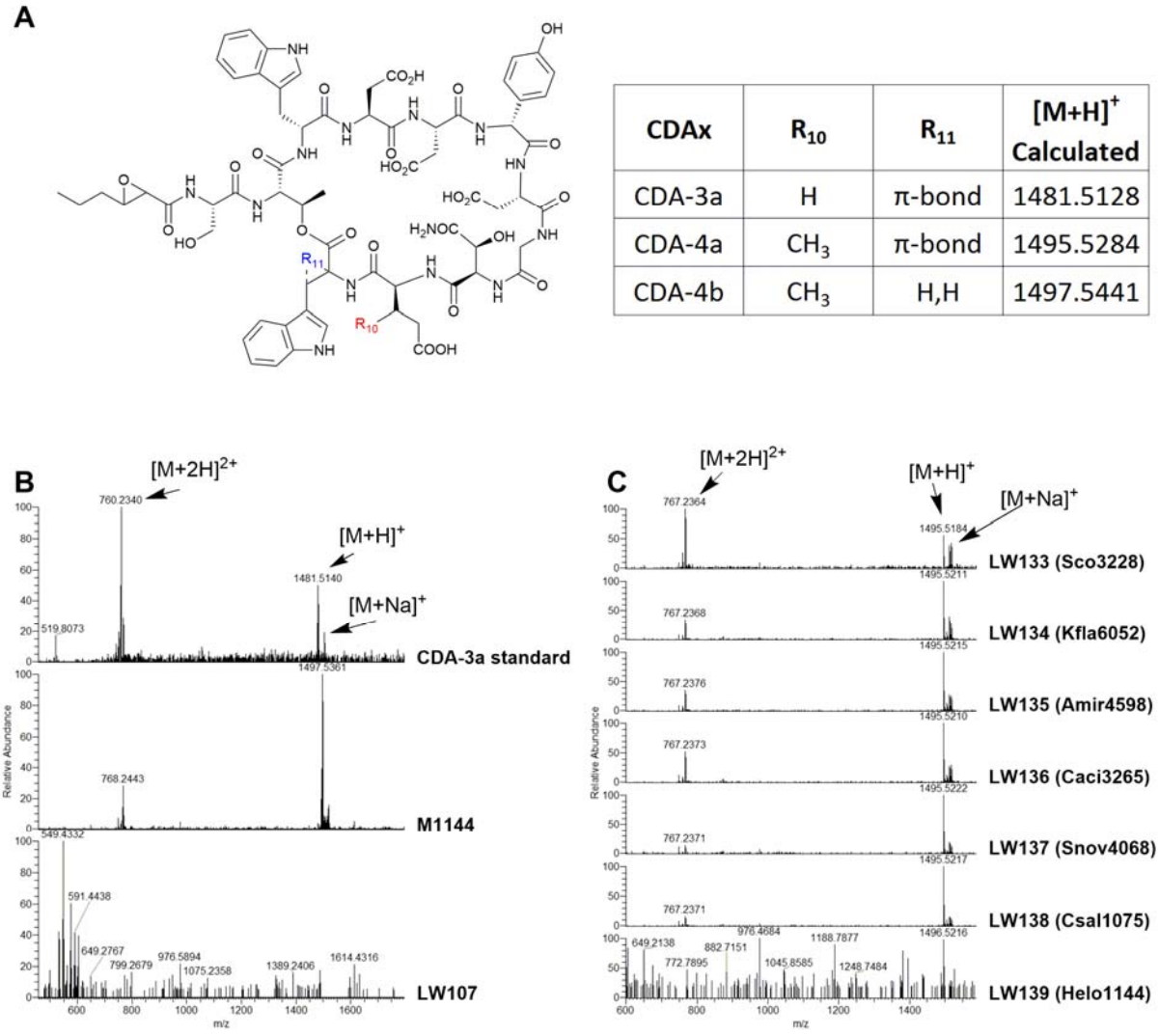


Figure S2. Phylogenetic Tree of the HmO and MdlB homologs. BLAST hits of HmO and MdlB proteins were combined in one alignment and in one phylogenetic tree. The evolutionary distances were calculated using the Poisson correction method[39] and a phylogenetic tree was created using the neighbour-joining method within MEGA5.[40] Red diamonds represent amino acid sequences with protein structures in the PDB database, green circles represent entries from the SwissProt database. Boxed in red, the selected proteins tested *in vivo* and/or *in vitro* in the present paper.



CDAx	R_{10}	R_{11}	$[M+H]^+$ Calculated
CDA-3a	H	π-bond	1481.5128
CDA-4a	CH_3	π-bond	1495.5284
CDA-4b	CH_3	H,H	1497.5441

Figure S3. MS analysis of CDA production in *S. coelicolor* M1144 (the parent), LW107 (*hmO*::Tn5) and its complemented strains. A sample of pure, previously fully characterized CDA-3a is provided for comparison.[53] **(A)** Structure and calculated theoretical masses of CDA variants identified in this study. **(B)** Masses detected in the peaks identified from extracted chromatograms (*m/z* 1480-1600) from figure 5A at 4.4 mins: CDA-3a pure standard (mass accurate to 0.8 ppm); M1144: *S. coelicolor* parent containing a wild type L-HPG operon (mass accurate to 5 ppm); LW107: M1144 *hmO::Tn5*. The WT strain M1144 can be seen to produce mainly CDA-4b. **(C)** Masses detected in the peaks identified from extracted chromatograms (*m/z* 1480-1600) from figure 5B at 4.4 mins: LW133: *hmO* mutant strain complemented with *S. coelicolor hmO*; LW134, LW135, and LW136: LW107 expressing Hmo homologs Kfla6052, Amir4598, Caci3265, respectively; LW137, LW138, LW139: LW107 complemented with MdlB orthologs, Snov4068, Csal1075 and Helo1144, respectively. No detectable CDA could be found in LW107 and only a trace in LW139. All other samples showed evidence of CDA-4a production, which represents a shift in production from CDA-4b produced by M1144. The reason for this shift is not speculated on. All masses are accurate to < 5 ppm.

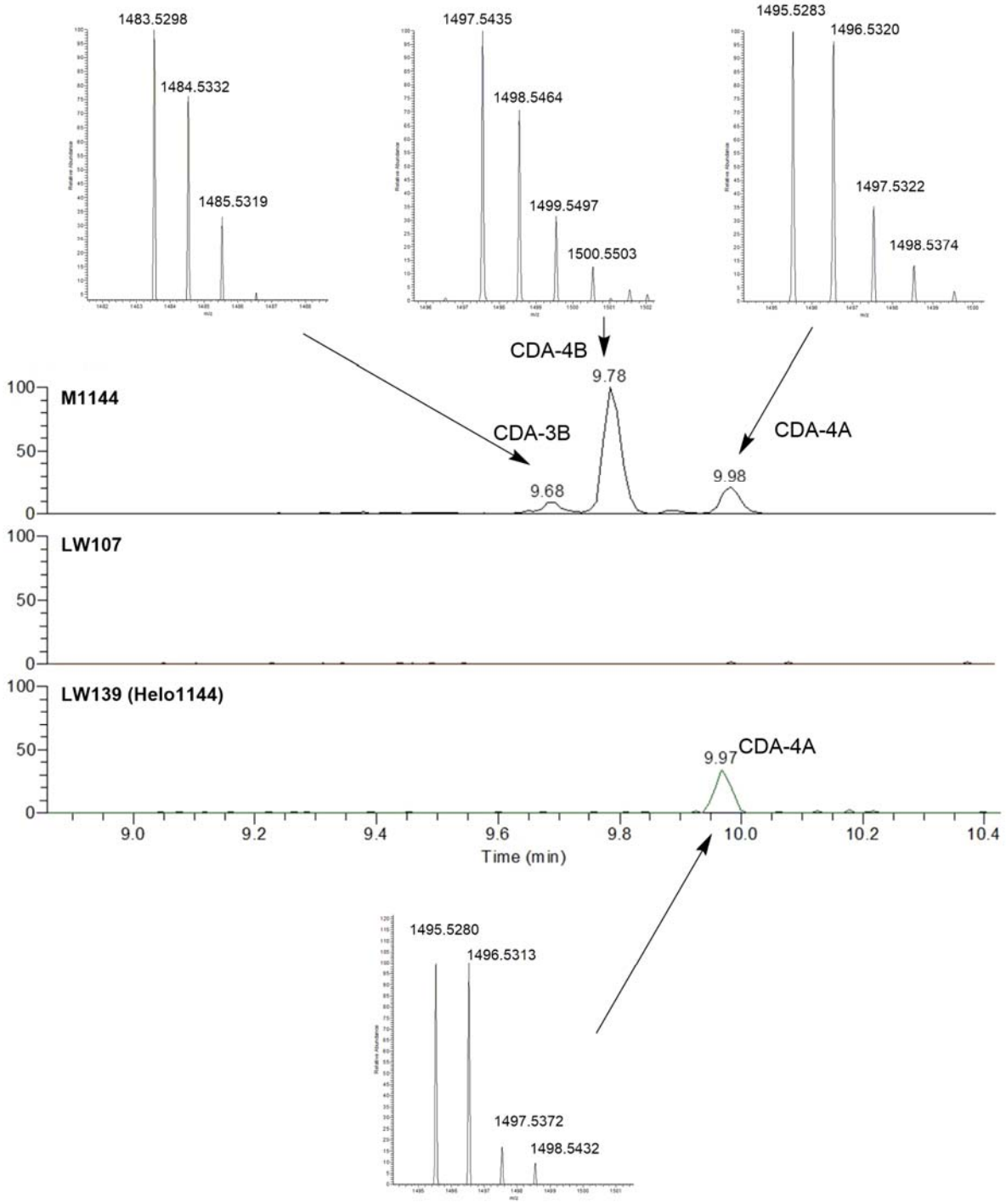


Figure S4. High resolution LC-MS analysis of CDA production in *S. coelicolor* M1144 (the parent), LW107 (*hmO*::Tn5), and LW139 complemented with MdlB ortholog Helo1144. Shown are extracted ion chromatograms for the mass range of known CDA variants (*m/z* = 1480-1600). Retention times of peaks are shown (min). Associated masses for major peaks are also shown. M1144 can be seen to produce mainly CDA-4b with smaller amounts of CDA-3b and CDA-4a. LW107 does not produce any detectable CDA as expected. LW139 can be seen to produce CDA-4a. All detected mass have an accuracy of less than 1 ppm.

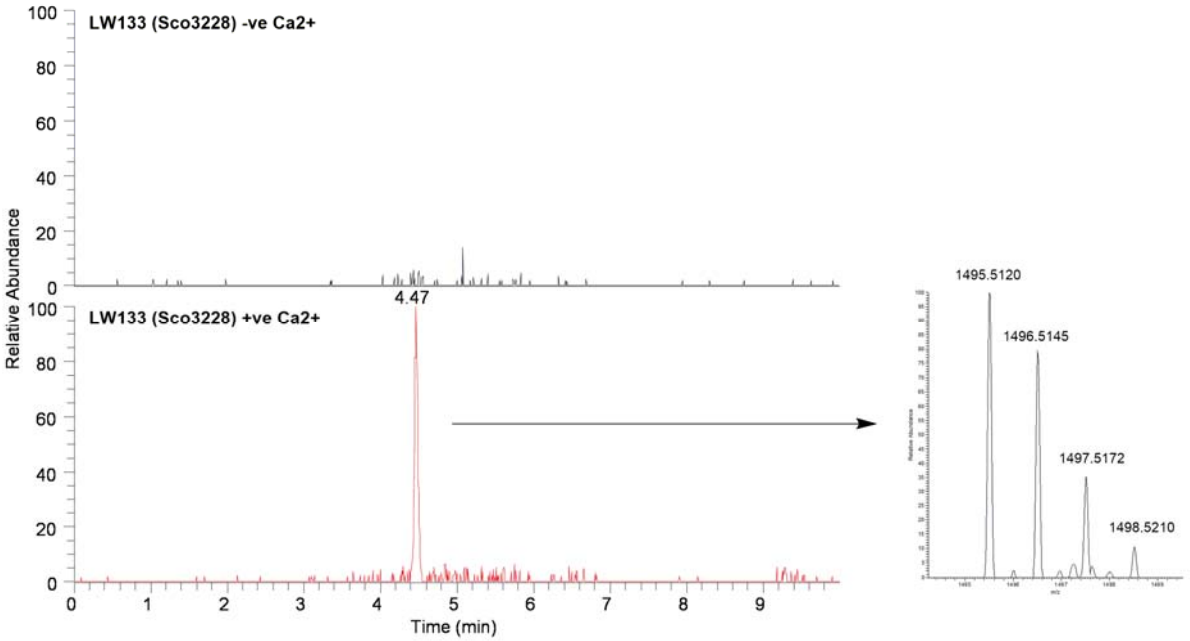


Figure S5. MS analysis of CDA production in *S. coelicolor* LW133 (*hmO* mutant strain complemented with *S. coelicolor hmO*) with or without the addition of calcium to the culture medium. Shown are extracted ion chromatograms for the mass range of known CDA variants (*m/z* = 1480-1600). Retention times of peaks are shown (minutes). CDA production is seen only in the presence of calcium (lower plot). The peak corresponds to a mass of 1495.5120 (theoretical mass of CDA-4a is 1495.5284 (5 ppm difference)).

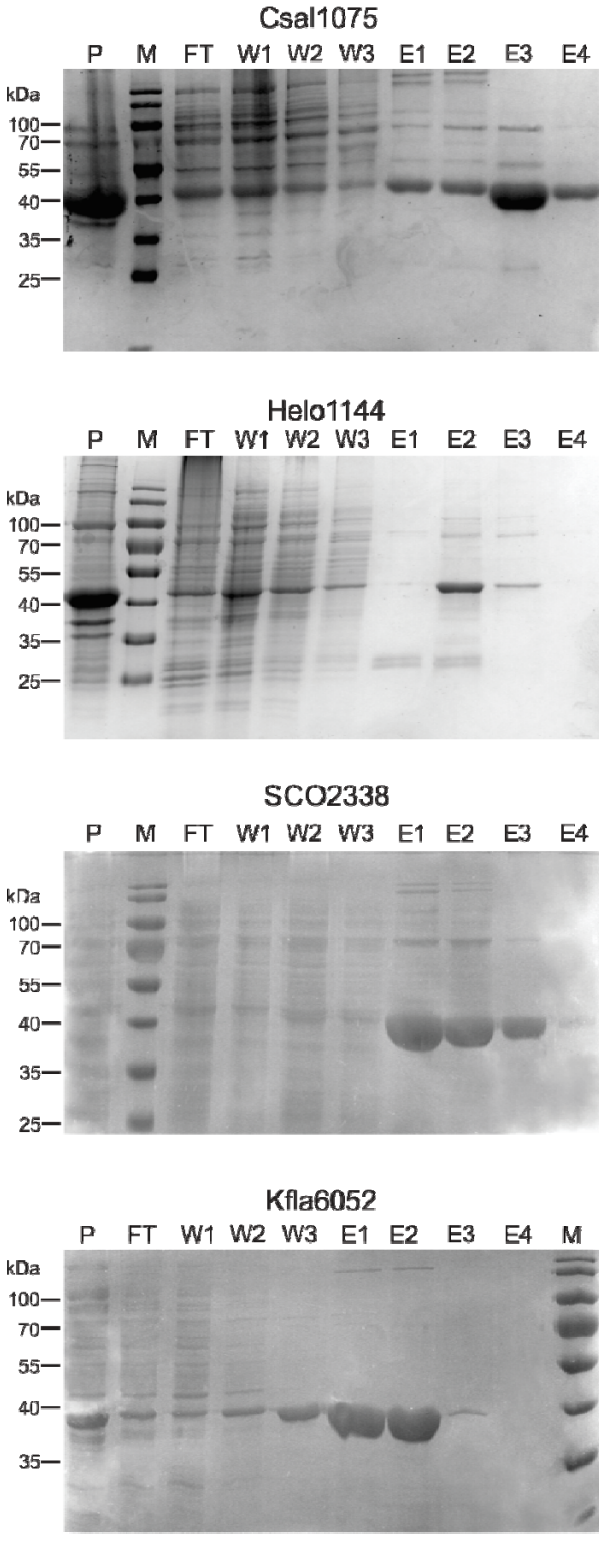


Figure S6. Purification of the HmO and MdlB homologs tested *in vitro*.
P = insoluble fraction corresponding to membranes and cell debris; FT: flow through; W1-3: wash fractions; E1-4: elution fractions. Buffer composition is detailed in Methods.