Electronic Supplementary Material (ESI) for Chemical Science.

Supplementary for

Structural basis of head to head polyketide fusion by CorB

Georg Zocher,[a,#] Joachim Vilstrup,[b,#] Daniel Heine,[c] Asis Hallab,[d] Emilie Goralski,[e] Christian Hertweck,[c,f] Mark Stahl,[g] Till F. Schäberle[e*] and Thilo Stehle[a,h*]

[a] Interfaculty Institute of Biochemistry, University Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany.
[b] Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10C, DK 8000 Aarhus C, Denmark.
[c] Leibniz Institute for Natural Product Research and Infection Biology (HKI), 07745 Jena, Germany.
[d] Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany.
[e] Institute for Pharmaceutical Biology, University of Bonn, Nussallee 6, 53115 Bonn, Germany.
[f] Chair for Natural Product Chemistry, Friedrich Schiller University, 07737 Jena, Germany.
[g] Center for Plant Molecular Biology, University Tübingen, Auf der Morgenstelle 1, 72076 Tübingen, Germany.
[h] Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.

[#] Authors contributed equally to this work

* To whom correspondence should be addressed:

For structural biology: Thilo Stehle, Interfaculty Institute of Biochemistry, University Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany, Tel.: +49-7071-2973043; Fax: +49-7071-295565; E-mail: thilo.stehle@uni-tuebingen.de, and for biochemistry Till Schäberle, Institute for Pharmaceutical Biology, University of Bonn, Nussallee 6, 53115 Bonn, Germany, Tel.: +49 228 73 2676; E-mail: till.schaeberle@uni-bonn.de

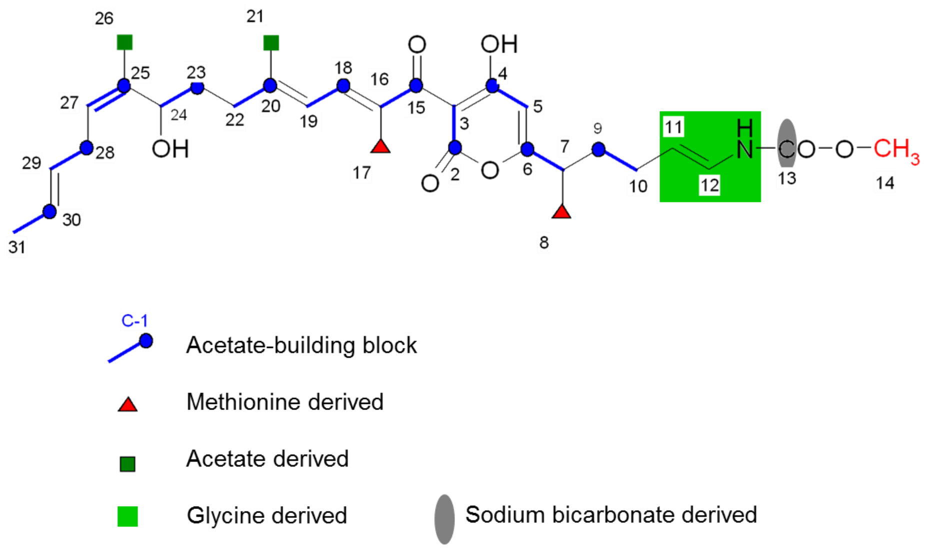


Figure S1. Result of the feeding experiments with labelled precursors

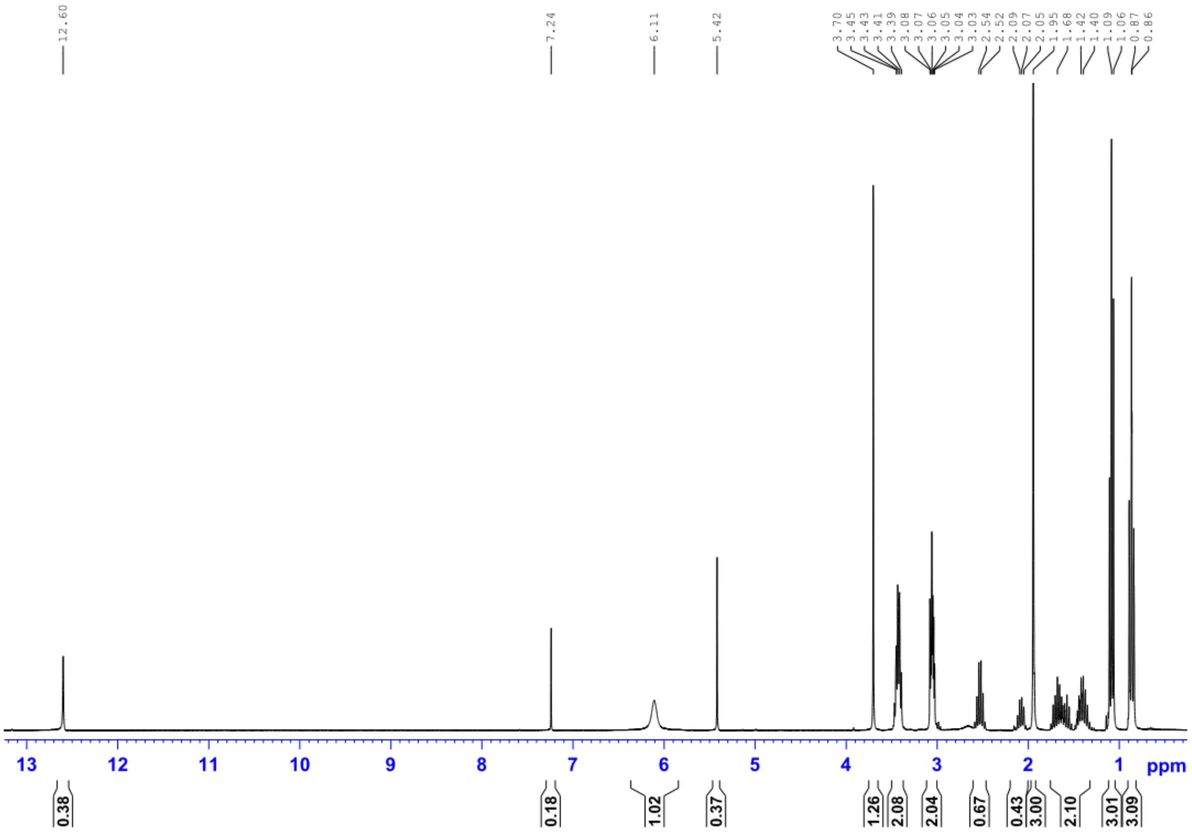


Figure S2 ^{1}H NMR (CDCl$_3$, 300 MHz) of substrate mimic **1**

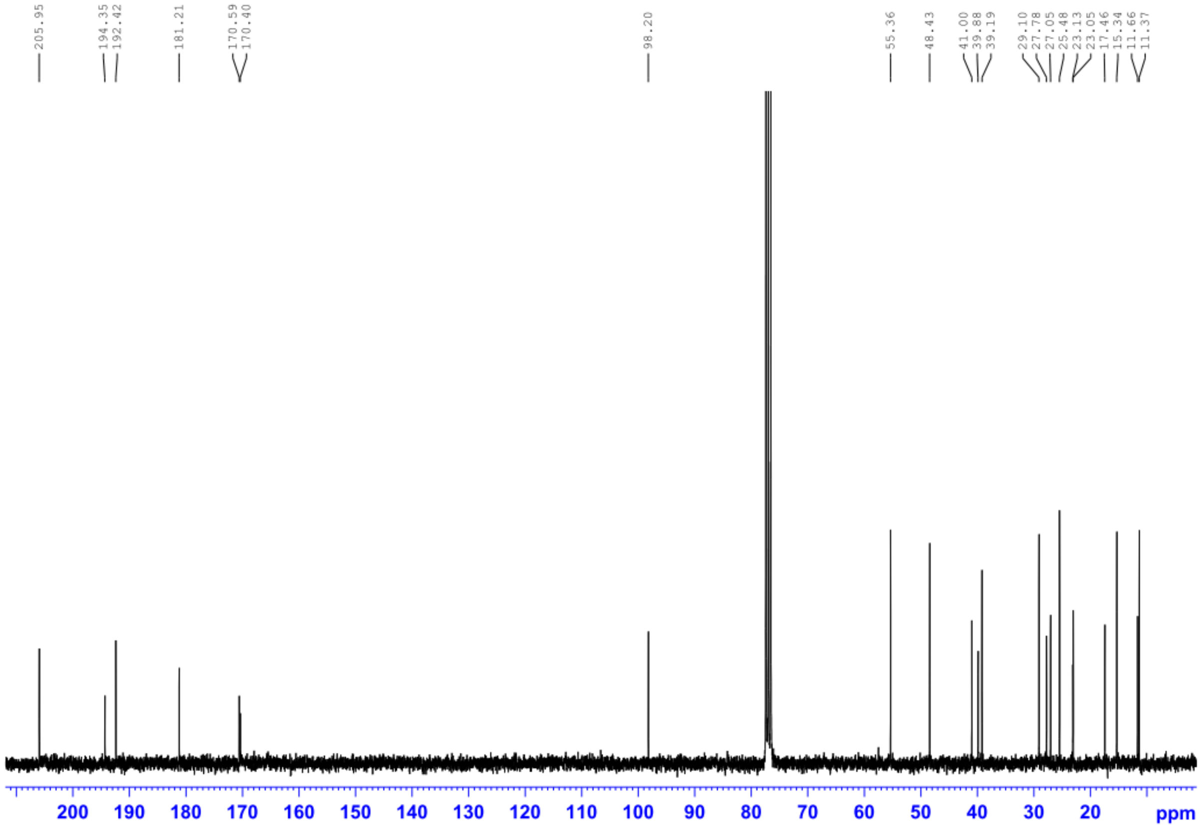


Figure S3 ^{13}C NMR ($CDCl_3$, 75 MHz) of substrate mimic **1**

KS	**Box1**	**Box2**	**Box3**	**Box4**
#1	CASSL	EAHGTG	NIGH	
#2	CSSSL	ETHGTG	NIGH	
#3	CSSSL	EAHGTG	NVGH	
#4	CSSSL	EAHGTG	NIGH	
#5	CSASL	ECAAAG	NVGH	
#6	CSSSL	ETHGTG	NIGH	
#7	CASSL	EAHGTG	NIGH	
#8	CSSSL	ECHGTG	NIGH	
#9	CSSSL	EAHGTG	NAGH	
#10	CSSSL	EAHGTG	NIGH	
#11	CSSSL	EAHGSG	NTGH	
#12	CSSSL	EAHGTG	NIGH	
#13	CSSSL	EAHGTG	NIGH	
CorB	CLSFF	HMNGAE	VIPH	A**N**C
CorD	**S**ASGN	NAHGSA	WTGH	

Figure S4. Conserved motifs of ketosynthase (KS) protein(domain)s encoded in the corallopyronin A gene cluster. The here analyzed CorB is boxed in red. CorD is involved in β-branching. KS domain #5 is inactive, due to the incomplete catalytic triad.

Table S1. Data collection and refinement statistics.

	CorB-ox	**CorB-red**	**CorB-1**
Wavelength [Å]	1.00000	1.5418	0.91840
Resolution [Å]	30-1.54 (1.58-1.54)	50-1.70 (1.80-1.70)	30-1.31 (1.34-1.31)
Space group	C222(1)	C222(1)	C222(1)
Unit cell dimension [Å]			
a	62.9	62.7	62.7
b	106.3	105.2	105.3
c	104.7	104.6	104.7
Measured reflections	659331 (29700)	146340 (22889)	840246 (132631)
Unique reflections	52020 (3695)	34975 (5725)	84188 (13261)
Redundancy	12.7 (8.0)	4.2 (4.0)	10.0 (10.0)
Completeness [%]	99.7 (97.3)	90.3 (92.5)	98.3 (99.7)
I/<σ**(I)**	22.1 (2.1)	11.1 (2.9)	16.9 (1.9)
CC1/2	100.0 (80.9)	99.6 (83.5)	99.9 (76.0)
Wilson B [Å^2]	24.3	20.4	19.6
No. of Atoms			
Chain A	2541	2594	2601
Ions/Water	255	317	341
Average B-Factors [Å^2]			
Chain A	33.8	18.8	18.3
Ions/Water	39.4	27.9	28.2
R_{work} [%]	17.7	19.7	17.5
R_{free} [%]	18.2	23.1	18.8
Rmsd bond angles [°]	1.33	0.010	1.16
Rmsd bond length [Å]	0.008	1.231	0.007
Ramachandran outliers [%]	0.7	0.6	0.6
Ramachandran favored [%]	96.1	95.3	96.2

Values in parentheses are for the highest resolution shell.

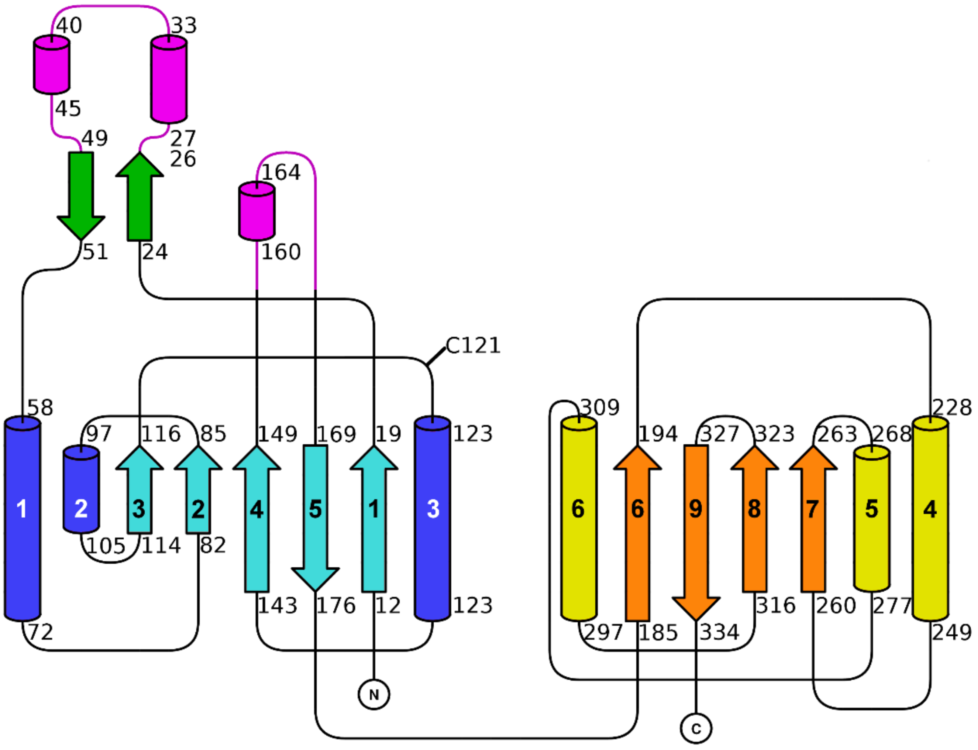


Figure S5 Topology plot of CorB.

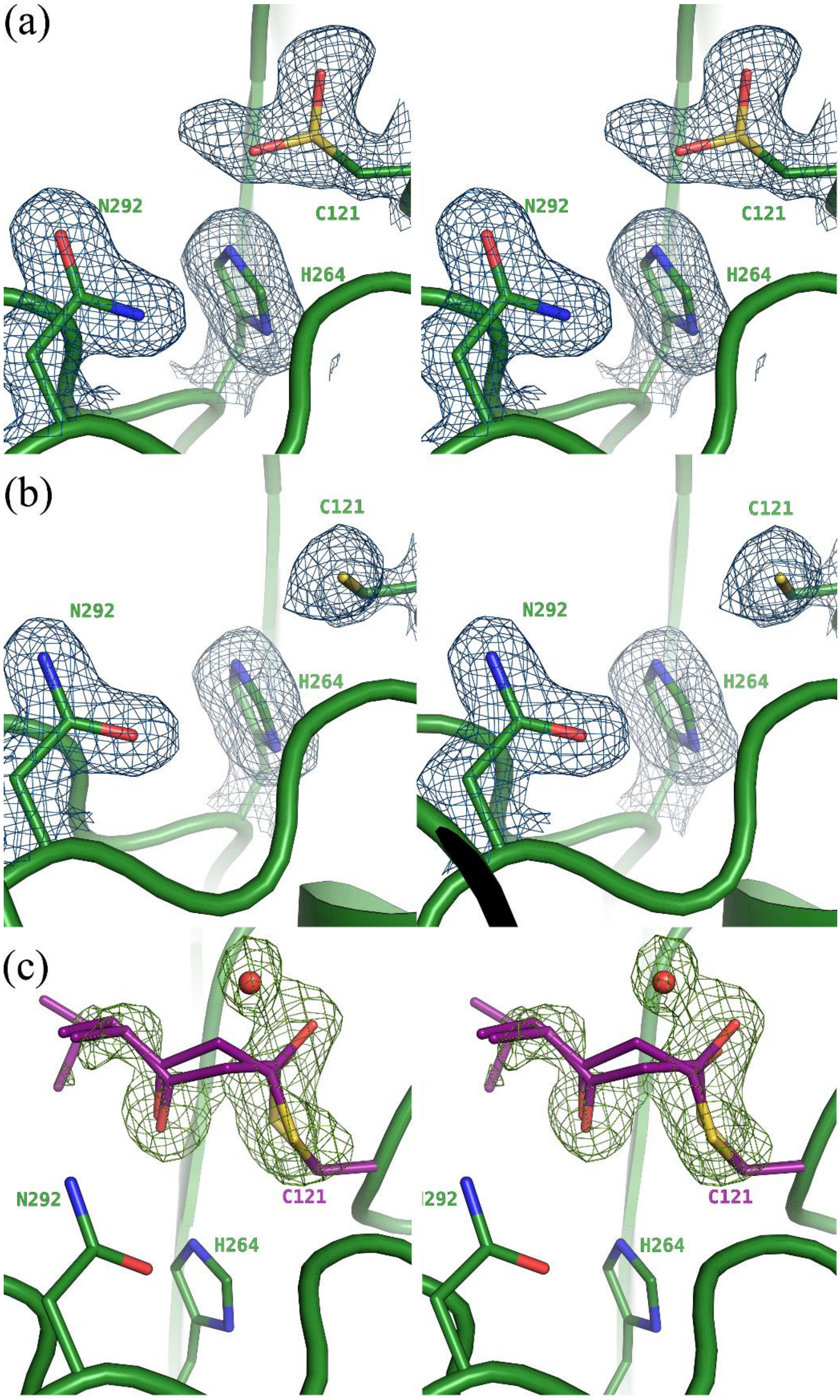
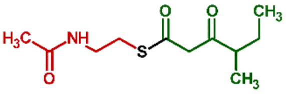
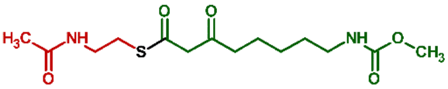


Figure S6 Stereoview into the active site of CorB for CorB-ox (a) for CorB-red (b) and for CorB-1 after incubation with **1** (c). The ($2F_o$-F_c)-electron density map (blue) for CorB-ox and CorB-red is shown at a contour level of 1.5. The substrate mimic **1** (violet) was modeled into the unbiased (F_o-F_c)-difference electron density map (green) depicted at a contour level of 2.9.



Substrate mimic, DH985, **1**



Substrate mimic, Eastern chain, JHS274, **2**

Molecular Formula: C_4H_8NOS

Formula Weight: 118.2 Da

Molecular Formula: $C_7H_{11}O_2$

Formula Weight: 127.2 Da

Molecular Formula: $C_{11}H_{19}NO_3S$

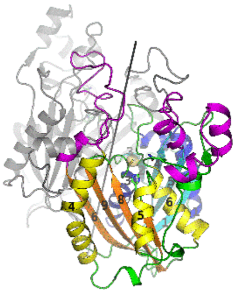
Formula Weight: 245.3 Da

Molecular Formula: $C_{10}H_{16}NO_4$

Formula Weight: 214.2 Da

Molecular Formula: $C_{14}H_{24}N_2O_5S$

Formula Weight: 332.4 Da



Formula Weight: 37233.6 Da

Figure S7 Theoretically expected molecular mass of all molecules used in the ESI-MS experiments.

Table S2. Structure comparison using DALI to FabH enzymes from different organisms

Protein and Source		PDB entry	Sequence identity [%]	C-rmsd [Å]	# of aligned residues
FabH	*Aquifex aeolicus vf5*	2EBD	34	2.1	304
FabH	*Burkholderia Xenovorans*	4DEF	25	2.4	322
FabH	*Vibrio Cholerae*	4NHD	30	2.5	313
FabH	*Escherichia coli*	1HND	31	2.4	313
FabH	*Staphylococcus aureus*	1ZOW	27	2.3	307
FabH	*Xanthomonas oryzae*	3FK5	22	2.3	309
FabH	*Haemophilus influenzae*	3IL3	29	2.7	312
FabH	*Xanthomonas Campestris Olea*	4KTI	22	2.4	307
FabH	*Thermus thermophilus hb8*	1UB7	29	2.4	310
FabH	*Mycobacterium tuberculosis*	2QNX	29	2.8	318

Table S3. Structure comparison using DALI to the closest 12 non-redundant enzymes

Protein and Source		PDB entry	Sequence identity [%]	C-rmsd [Å]	# of aligned residues
PqsD	*Pseudomonas aeruginosa pao1*	3H77	26	2.5	305
PKS11	*Mycobacterium tuberculosis*	4JAP	22	2.8	303
STEELY1	*Dictyostelium discoideum*	2H84	19	2.7	302
Pentaketid Synthase	*Neurospora crassa*	3EU0	17	2.9	300
HMG-CoA Synthase	*Brassica juncea*	2F9A	16	2.4	300
Chalcone Synthase 2	*Medicago sativa*	1U0W	22	3.0	305
Pentaketide Chromone Synthase	*Aloe arborescens*	2D3M	18	3.0	305
Curcumin Synthase	*Curcuma longa*	3OV3	21	3.0	304
PKS1	*Ectocarpus siliculosus*	4B0N	17	2.9	303
2-Pyrone Synthase	*Gerbera hybrid cultivar*	1EE0	20	3.0	306
Stilbene Synthase	*Vitis vinifera*	3TSY	20	3.0	305
Malonyl transferase	*Streptomyces tendae*	3S3L	17	2.9	298

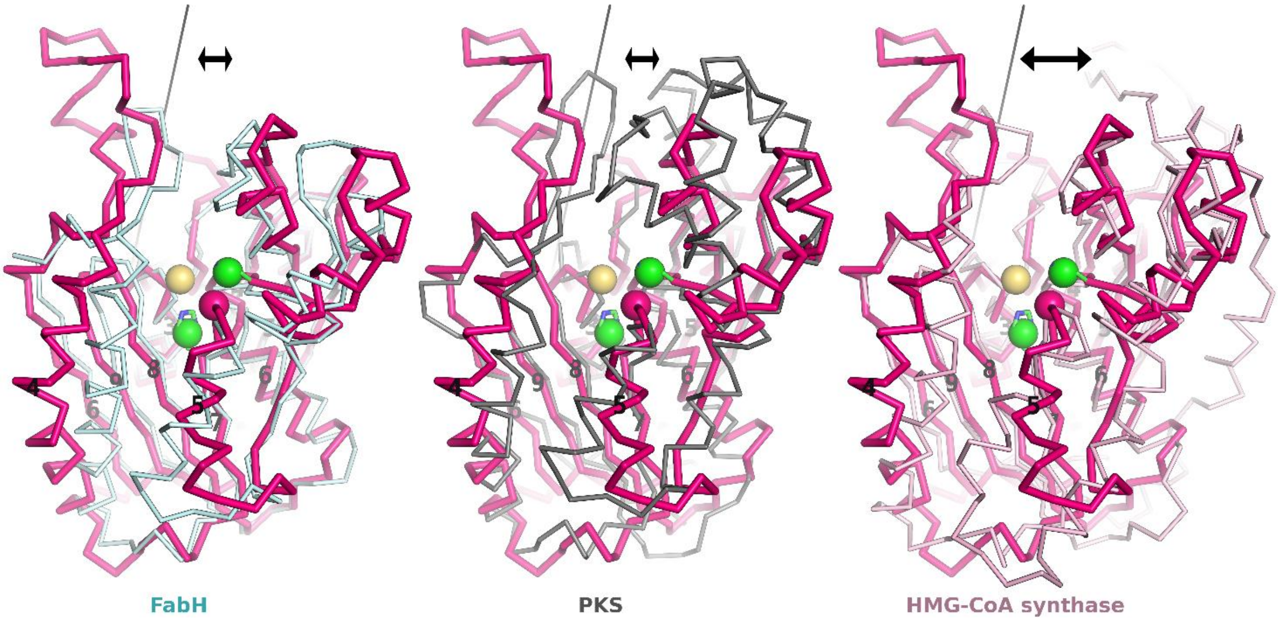


Figure S8 Structure comparison of CorB with FabH, PKS and HMG-CoA synthase. CorB is shown in pink. The catalytic triad is emphasized by colored spheres.

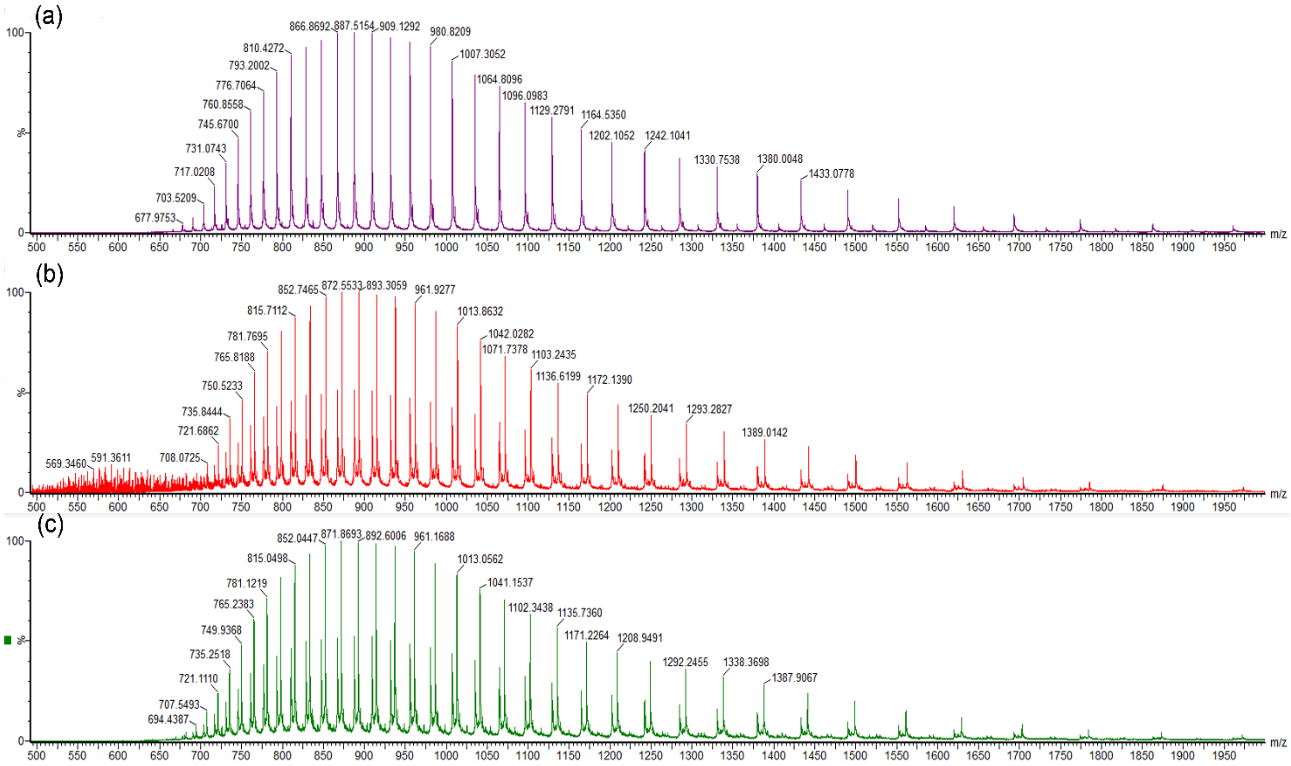


Figure S9 Complete ESI-MS spectra of (a) CorB, (b) CorB incubated with substrate **1**, and (c) CorB incubated with substrate **2**.

Table S4. Oligonucleotides used in this work.

Gene	**Primer**	**Sequence 5'-3'**
corB	CorB_TOPO_up	CACCATGAATCAGGGTGGTGTCTT
	CorB_TOPO_dn	TCAGTAGGTGAAGACCATTCC
*corB*_C121A	C121A for	ATCACCGTCAATGCCTCGGCCCTGAGCTTCTTCGTC
	C121A rev	GACGAAGAAGCTCAGGGCCGAGGCATTGACGGTGAT
*corB*_C121S	C121S for	TGCCGTCAATGCCTCGTCCCTGAGCTTCTTCGTC
	C121S rev	GACGAAGAAGCTCAGGGACGAGGCATTGAC
*corB*_H264A	H264A for	GGTACGTCATCCCGGCTCAGCCGAGCCGCGTGG
	H264A rev	CCACGCGGCTCGGCTGAGCCGGGATGACGTACC
*corB*_H264F	H264F neu f	GGTACGTCATCCCGTTTCAGCCGAGCCGCGTGG
	H264F neu r	CCACGCGGCTCGGCTGAAACGGGATGACGTACC
*corB*_N292A	N292A for	ATCGACCGGTTCGCCGCCTGCATCGGCGCCTCC
	N292A rev	GGAGGCGCCGATGCAGGCGGCGAACCGGTC

Table S5. Strains and plasmids used in this work.

Plasmid / Strain	**Genotype**	**Reference**
Corallococcus coralloides B035	Wild type	Erol et al., 2010
Escherichia coli One Shot®TOP10	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen
Escherichia coli BL21 (DE3) Star	F$^-$ *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm rne131* (DE3)	Invitrogen
pET151TOPO	N-terminal 6xHis tag, TEV protease cleavage site	Invitrogen
pET151TOPO+CorB	*corB*	this work
pET151TOPO+CorB_C121A	*corB*_C121A	this work
pET151TOPO+CorB_C121S	*corB*_C121S	this work
pET151TOPO+CorB_H264A	*corB*_H264A	this work
pET151TOPO+CorB_H264F	*corB*_H264F	this work
pET151TOPO+CorB_N292A	*corB*_N292A	this work
E. coli TFS-CorB	*corB*	this work
E. coli TFS-CorB_C121A	*corB*_C121A	this work
E. coli TFS-CorB_C121S	*corB*_C121S	this work
E. coli TFS-CorB_H264A	*corB*_H264A	this work
E. coli TFS-CorB_H264F	*corB*_H264F	this work
E. coli TFS-CorB_N292A	*corB*_N292A	this work