SUPPLEMENTARY INFORMATION

Analysis of the cryptic biosynthetic gene cluster encoding the RiPP curacozole reveals a phenylalanine-specific peptide hydroxylase

Samantha Hollands[1], Julia Tasch[2], David J. Simon[1], Dimah Wassouf[2], Isobel Barber[1], Arne Gessner[2,3], Andreas Bechthold[3], and David L. Zechel[1*]

1. Department of Chemistry, Queen's University, 90 Bader Lane, Kingston, Ontario, K7L 3N6, Canada
2. Pharmaceutical Biology and Biotechnology, Institute of Pharmaceutical Sciences, Albert-Ludwigs University, Freiburg, Germany.
3. Institute of Experimental and Clinical Pharmacology and Toxicology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

*Corresponding Author. DLZ: Email, dlzechel@chem.queensu.ca, Tel. +1 (613) 533-3259

TABLE OF CONTENTS

TABLES

Table S1. Chemical and spectroscopic properties of Curacozole

Appearance	Yellowish-white solid
Molecular Formula	$C_{36}H_{36}N_8O_6S_2$
HR-ESI-MS (*m/z*)	
Found	741.22699 $[M + H]^+$
Calcd.	741.22775
$[\alpha]^{25}_D$ (CH_2Cl_2)	+0.955
UV λ_{max} nm (CH_2Cl_2)	271, 303
IR ν_{max} (neat) cm^{-1}	3341, 3069, 2961, 2929, 2874, 1658, 1497, 1097, 1048

Table S2. ^{1}H NMR and ^{13}C NMR spectroscopic data for Curacozole (δ in ppm, J in Hz). **For atom numbering scheme and a summary of correlations, see Figure S2.**

Position	δ_C[a]	δ_H (mult., *J*)[b]	^{1}H-^{13}C HMBC	δ_N[c]	^{1}H-^{15}N HMBC
1	129.71, C	-			
2	153.11, C	-			
3	126.81, C	-			
4	128.21, CH	8.38 (d, 7.6)	C-2, C-5, C-6		
5	128.51, CH	7.45 (t, 7.4)	C-2(w), C-3, C-4		
6	130.25, CH	7.39 (t, 7.1)			
7	128.51, CH	7.45 (t, 7.4)	C-2(w), C-3, C-4		
8	128.21, CH	8.38 (d, 7.6)	C-2, C-5, C-6		
9	161.56, C				
10 (Ile1)	57.73, CH	4.97 (dd, 9.7, 5.0)	N-2, C-12, C-13, C-15		N-2, N-3
11 (Ile1)	37.54, CH	2.30 (m)	C-10, C-12, C-13, C-14, C-15		N-2
12 (Ile1)	15.08, CH_3	1.10 (d, 6.8)	C-10, C-11, C-13		
13 (Ile1)	26.49, CH_2	H_a, 1.27 (m) H_b, 1.46 (m)			
14 (Ile1)	11.70, CH_3	0.87 (t, 7.3)	C-11, C-13		
15 (Ile1)	173.36, C				
16 (Ile2)	59.15, CH	4.45 (t, 6.3)	N-2, C-12, C-13, C-21		N-3, N-4
17 (Ile2)	36.08, CH	2.08 (m)	C-16, C-18, C-19, C-20		
18 (Ile2)	16.15, CH_3	0.93 (d, 6.8)	C-16, C-17, C-19		
19 (Ile2)	25.39, CH_2	H_a, 1.24 (m) H_b, 1.51 (m)			
20 (Ile2)	11.60, CH_3	0.78 (t, 7.3)	C-17, C-19		
21 (Ile2)	171.82, sC				
22 (Gly)	36.85, CH_2	H_a, 4.25 (br. d, 17.5) H_b, 5.08 (dd, 17.8, 7.6)	C-23, C-21		N-4, N-5

23 (Gly)	162.32, C				
24	138.58, CH2	8.09 (s)	C-23, C-25		
25	129.66, C	-			
26	153.84, C	-			
27	148.12, C	-			
28	11.91, CH_3	2.62 (s)	C-29, C-30, C-31, C-32		N-6 (weak)
29	131.81, C	-			
30	158.08, C	-			
31	117.94, CH	7.75 (s)	C-29, C-30, C-32, C-33		N-7
32	148.18, C	-			
33	161.61, C	-			
34	119.27, CH	7.92 (s)	C-32, C-33, C-35, C-36		N-8
35	142.51, C	-			
36	153.94, C	-			
N-1	-	-		255.06, N	
N-2	-	8.52 (d, 9.7)	C-9, C-10, C-11	110.74, NH	N-1
N-3	-	7.75 (br. d, 6.9)	C-15, C-16, C-17	113.66, NH	
N-4	-	7.20 (br.)	C-21	99.93, NH	
N-5	-	-		245.13, N	
N-6	-	-		252.43, N	
N-7	-	-		313.78, N	
N-8	-	-		314.53, N	

[a] Recorded at 176 MHz in $CDCl_3$. [b] Recorded at 700 MHz in $CDCl_3$. [c] Recorded at 76 MHz in $CDCl_3$.

Table S3. Predicted functions of *czl* gene products.

Czl Proteins (# aa)	Annotation / Predicted Function	Accession	Homologs (% aa ID)	Accession
CzlI (322)	Fe/2OG dependent precursor peptide oxygenase*	KUM76188	YmI, (71) S. viridochromogenes hypothetical protein (96)	AFJ68073 ELS57320.1
CzlA (44)	Precursor peptide	WP_107116988	YmA, (86)	AFJ68074
CzlD (470)	YcaO family cyclodehydratase*	WP_159042661	YmD, (74) McmD1 (31) McmD2 (29) TlsQ (36)	AFJ68075 UBK24771 UBK24772.1 BBA31819
CzlE (469)	P450 monooxygenase	WP_062148882	YmE, (66) McmI, (30)	AFJ68076 UBK24777
CzlB1 (261)	putative cyclodehydratase	WP_062148878	YmB1, (85) McmE (37)	AFJ68086 UBK24773
CzlC1 (287)	unknown	WP_062148875	YmC1, (62) McmF (27)	AFJ68077 UBK24774
CzlF (362)	unknown	WP_062148872	YmF, (69)	AFJ68078
CzlBC (465)	E1-like / dehydrogenase fusion*	WP_062148869	YmBC, (67) McmC (40)	AFJ68085 UBK24770
CzlJ (166)	unknown	KQ947987 (Locus tag: AQI70_14560)	DUF6624 domain containing protein, (99)	WP_215158549
CzlK (409)	serine peptidase	WP_062148866	*Caulobacter vibrioides* beta-lactamase family protein, (36) *Bacillus simplex* D-alanyl-D-alanine carboxypeptidase (81)	PDB ID: 7UDA CEG31961
CzlR1 (138)	AraC family transcriptional regulator	WP_062148863	YmR1 (66)	AFJ68079

*Functions established in this work through in vitro reactions.

Table S4. Bacterial strains and plasmids

Strain or Plasmid	Description	Reference or Source
	Streptomyces	
S. curacoi / pTESa-*bldA*	*Streptomyces curacoi* DSM40107 carrying the integrative plasmid pTESa-*bldA,* Apr^R, curacozole producer.	Ref[1] and this work.
S. curacoi - bldA	*S. curacoi* with bldA sequence under control of ermE* promoter. Derived from *S. curacoi* / pTESa-*bldA* after excision of apramycin resistance gene, curacozole producer.	This work.
S. curacoi – bldA Δ*czlB1::apra*	Mutant strain with *czl*B1 sequence replaced with an apramycin resistance gene.	This work.
S. curacoi – bldA Δ*czlE::apra*	Mutant strain with *czl*E sequence replaced with an apramycin resistance gene.	This work.
S. coelicolor CH999	Heterologous host lacking actinorhodin BGC.	Ref[2]
S. lividans ΔYA9	Heterologous host lacking nine endogenous BGCs.	Ref[3]
	E. coli	
E. coli ET12567 / pUZ8002	Methylation deficient ET12567 strain (Cm^R) for intergeneric conjugation. Plasmid pUZ8002 carries conjugation functions (Kan^R)	Ref[4]
E. coli BW25113 / pIJ790	Strain for λ-RED based recombination	Ref[5]
	Plasmids	
pBSK	Subcloning plasmid derived from pBluescript II.	
pTESa	Integrative plasmid, pSET152 derivative, carrying an *ermE** promoter, φC31 *int* and *att* functions, and apramycin resistance gene for selection in *E. coli* and *Streptomyces*.	Ref[6]
pKCLP2	Replicative plasmid in *E. coli*, non-replicative in streptomyces, Hyg^R.	Ref[7]
pKGLP2	*gus*A containing derivative of pKCLP2, replicative plasmid in *E. coli,* non-replicative in streptomyces, Hyg^R	Ref[7]
pALCRE	Replicative plasmid in streptomyces, *creE* under control of *tipA* promoter, Hyg^R	Ref[8]
pLERE	Template for amplification of apramycin resistance gene	Ref[6]
pCDFDuet_1	Template for amplification of spectinomycin resistance gene	Novagen
pBSK-*czl*	Subcloning plasmid containing *czl* BGC	This work.
pTESa-czl	Integrative plasmid containing curacozole (czl) BGC from *czl*I to *czl*K	This work.
pTESa-czlΔI	Integrative plasmid containing *czl* BGC with *czl*I deleted.	This work.
pTESa-czlΔD	Integrative plasmid containing *czl* BGC with *czl*D deleted.	This work.
pTESa-czlΔC1	Integrative plasmid containing *czl* BGC with *czl*C1 deleted.	This work.
pTESa-czlΔF	Integrative plasmid containing *czl* BGC with *czl*F deleted.	This work.
pTESa-czlΔBC	Integrative plasmid containing *czl* BGC with *czl*BC deleted.	This work.
pTESa-czlΔJK	Integrative plasmid containing *czl* BGC with *czl*J and *czl*K deleted.	This work.
pKGLP2-czlE::apra	Suicide plasmid containing *czl* BGC with *czl*E replaced by apramycin resistance gene	This work.
pKCLP2-czlB1::apra	Suicide plasmid containing *czl* BGC with *czl*B1 replaced by apramycin resistance gene	This work.

Table S5. Primers for curacozole BGC cloning and mutagenesis

Primer	Sequence (5'-3')*	Notes
cur_xbalfw	ACG**TCTAGA**GGTGGGTTGCCTCTGGATG	pTESa-*czl* generation, XbaI site underlined
cur_EcoRIrv	GGC**GAATTC**TCAGTCGTCGATCGCCTGGTAG	pTESa-*czl* generation, EcoRI site underlined
czldell_fwxbal	AAAAA**TCTAGA**CCGGAGAAATCCCCCGGTCC	pTESa-*czl*ΔI generation, XbaI site underlined
cur2	TCAGTCGTCGATGCGCTGGTAG	pTESa-*czl*ΔI generation
cur_xbald_fw	ACG**TCTAGA**GGTGGGTTGCCTCTGGATG	pTESa-*czl*ΔJK generation, XbaI site underlined
czldelJ_v	GCACCTCGGCGAGCGGCGATAATG	pTESa-*czl*ΔJK generation
sect_czlD_fw	GGTAATTCTCCGGCACACCACCGCACCCATGGAATGGAGTCCCGC**ACTAGT**ttatttgccgactaccttg	*czl*D deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spect_czlD_rv	GGGCCGGGTCGTGGTCAGGGCAGGGGGTGCGGGCCTGGTACGTAG**ACTAGT**tttgtttatttttctaaataca	*czl*D deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spec_czlC1_fw	GGTGGCCCCCATGGGCGGACCGCGCGGAGCGACGGCATGAGC**ACTAGT**gaccgagtgagctagcta	*czl*C1 deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spec_czlC1_rv	CTCCAGGCCGCGCAGCTCGGCGAGGTGCCCCACGTCGGTCATCCC**ACTAGT**ttatttgccgactacctt	*czl*C1 deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spect_czlF_fw	CTGCTGGAGTCGGCGCGGCACTCGACGGCGGCGGCACGGGGATGA**ACTAGT**ttatttgccgactacctt	*czl*F deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spect_czlF_rv	CCAGCGGGTCGAAGCCGACGAGGGTGTACGAGGCGGTGGTCATA**ACTAGT**tttgtttatttttctaaatac	*czl*F deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spect_czlBC_fw	CGACGCGTACGCCCGGCACATCCACCGCACCGCCCGGGAGGCCGT**ACTAGT**ttatttgccgactaccttg	*czl*BC deletion, amplification of *aadA* for Spec^R, SpeI site underlined
spect_czlBC_rv	GGGGAACGGTGACAGGCACCTCGGCGAGCGGCGATAATGGAT**ACTAGT**tttgtttatttttctaaataca	*czl*BC deletion, amplification of *aadA* for Spec^R, SpeI site underlined
apra_czlB1_fw	GGCGACCCCCAGGCAACCGGAACCCGCCGGACACCGAAAGGAATCACGATGAGgcatatggatatctctagataccg	*czl*B1 deletion, amplification of *aac(3)IV* for Apra^R
apra_czlB1_rv	AGTGCAGCGCCGTGGCCGCCGCGTGCCGGTCGTGCCCGGCGGGTGCGCTCATGCcatatgaacaaaagctggagctc	*czl*B1 deletion, amplification of *aac(3)IV* for Apra^R
czlB1-fw	GTGTGCTTTCCCTACCTCCGGTC	*czl*B1 deletion, amplification of *czl*B1::apra sequence
czlB1-rv	GTCGAGTCGGAAGCCGTCCGG	*czl*B1 deletion, amplification of *czl*B1::apra sequence

apra_czlE_fw	ACGCCCGCCCCGACCGCGCCCTGCCGTG ACCGACGCCCCGAGGTGCCCGCGATgcatat ggatatctctagataccg	*czl*E deletion, amplification of *aac(3)IV* for Apra^R
apra_czlE_rv	ACTCCTCCTTGCGCATCAGCCCCGCGACC AGCGCGGTCGTCGCATGCCTCATCGcatatg aacaaaagctggagctc	*czl*E deletion, amplification of *aac(3)IV* for Apra^R
czlE-fw	GTGGTGGTCGTCCAGGTCAGCG	*czl*E deletion, amplification of czlE::apra sequence
czlE-rv	TGAAGCCGTTGAGACGCGCCAG	*czl*E deletion, amplification of *czl*E::apra sequence
FW_fw1	TGGATCATCGGGTCGACCTGCTGT	*czl*A F18W mutation, for amplification from pBSK-czlA
FW_rv1	TCCGACCTCGATGTCGACTTC	*czl*A F18W mutation, for amplification from pBSK-czlA
CzlA-F18Y-fw	TACATCATCGGGTCGACCTGCTGTTC	*czl*A F18Y mutation, for amplification from pBSK-czlA
CzlA-F18Y-rv	TCCGACCTCGATGTCGACTTCCTC	*czl*A F18Y mutation, for amplification from pBSK-czlA
corel_V_fw_b	GGTCGGATTCATCGTCGGGTCGAC	*czl*A_I20V mutation, for amplification from pBSK-czlA
corel_V_rv_b	TCGATGTCGACTTCCTCGATC	*czl*A_I20V mutation, for amplification from pBSK-czlA
corelL2fw	GGATTCATCCTCGGGTCGACCTGC	*czl*A_I20L mutation, for amplification from pBSK-czlA
corelL2rv	GACCTCGATGTCGACTTCCTC	*czl*A_I20L mutation, for amplification from pBSK-czlA
czlAT_Sfw_JT	ATCGGGTCGTCCTGCTGTTCCCTG	*czl*A_T23S mutation, for amplification from pBSK-czlA
coreC_S_fw_a	ATCGGGTCGTCCTCCTGTTCCCTG	*czl*A_T23S_C24S mutation, for amplification from pBSK-czlA_T23S
czlAT_Srv_JT	GATGAATCCGACCTCGATGTC	*czl*A_T23S and *czl*A_T23S_C24S mutations, for amplification from pBSK-czlA and pBSK-czlA_T23S
CzlA_check_fw	GTTCGCGCCCGCGGATCT	Primer to amplify *czl*A for DNA sequencing
CzlA_check_rv	CCGCCGCACTTGGTGTACCG	Primer to amplify *czl*A for DNA sequencing

*****Upper case corresponds to *czl* sequences (with exception of restriction enzyme sites), lower case to antibiotic resistance gene sequences. Restriction enzyme sites are underlined and in bold.

Table S6. Buffers used for purification of *czI* specified proteins.

Protein	Buffer A	Buffer B	Storage Buffer
CzIA and variants	pH 7.5 25 mM Bis-Tris 100 mM NaCl 1 mM DTT	pH 7.5 25 mM Bis-Tris 100 mM NaCl 1 mM DTT 500 mM imidazole	pH 7.5 25 mM Bis-Tris 50 mM NaCl 1 mM DTT 10 % glycerol
CzID CzIBC	pH 7.5 40 mM Tris 100 mM NaCl 1 mM DTT 10% glycerol	pH 7.5 40 mM Tris 100 mM NaCl 1 mM DTT 10 % glycerol 500 mM imidazole	pH 7.5 40 mM Tris 50 mM NaCl 1 mM DTT 10 % glycerol
CzII	pH 8.0 40 mM Tris 100 mM NaCl 10 % glycerol 1 mM DTT	pH 8.0 40 mM Tris 100 mM NaCl 10 % glycerol 1 mM DTT 500 mM imidazole	pH 8.0 40 mM Tris 50 mM NaCl 10 % glycerol

Table S7. Sequences of CzlA and variants used for in vitro assays

Variant	Sequence*	Av. Calculated Mass (Da)**
CzlA	GSSHHHHHHSSGLVPRGSQ**MFENTTAEIEEVDIEVG FIIGSTCCSLEMEEDDLDVAAEETEAV**	6851.36
CzlA C24S	GSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS**MFE NTTAEIEEVDIEVGFIIGSTSCSLEMEEDDLDVAAE ETEAV**	8224.89***
CzlA C25S	GSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS**MFE NTTAEIEEVDIEVGFIIGSTCSSLEMEEDDLDVAAE ETEAV**	8224.89***
CzlA C24S/C 25S	GSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS**MFE NTTAEIEEVDIEVGFIIGSTSSSLEMEEDDLDVAAE ETEAV**	8208.83***
CzlA F18G	GSSHHHHHHSSGLVPRGSH**MFENTTAEIEEVDIEVG GIIGSTCCSLEMEEDDLDVAAEETEAV**	6770.25

*Expressed sequences shown as missing *N*-terminal Met due to action of methionine aminopeptidase in *E. coli*. Native CzlA sequence shown in bold, core peptide in blue, substitutions in red. The underlined sequence in CzlA is predicted by AlphaFold to form β-sheets with the peptide binding domain of CzlBC.
**Calculations performed using Expasy PeptideMass.
***Masses of these CzlA variants are significantly larger due to differences in the N-terminal sequences arising from cloning of the corresponding genes.

Table S8. Calculated *m/z* values for CzlA and variants before and after heterocyclization by CzlD and CzlBC.

z	*Calculated**	*Observed*	*z*	*Calculated**	*Observed*
	CzlA MW_{calc} = 6851.36 Da			**CzlA-Ox** (-H_2O, -2H) MW_{calc} = 6831.36 Da	
4	1713.84	1714.30	4	1708.84	1709.1
5	1371.27	1371.53	5	1367.07	1367.53
6	1142.89	1142.92	6	1139.56	1139.71
7	979.77	980.02	7	976.94	977.01
	ClzA C24S MW_{calc} = 8224.89 Da			**ClzA C24S-Ox** (-H_2O, -2H) MW_{calc} = 8204.89 Da	
5	1645.98	1645.65	5	1641.98	1642.25
6	1371.82	1371.17	6	1368.48	1638.97
7	1175.98	1175.73	7	1173.13	1173.43
8	1029.11	1029.10	8	1026.6	1026.60
	CzlA C25S MW_{calc} = 8224.89 Da			**CzlA C25S-Ox** (-H_2O, -2H) **MW_{calc} = 8204.89**	
5	1645.98	1645.25	5	1641.98	1642.35
6	1371.82	1371.67	6	1368.48	1368.67
7	1175.98	1175.93	7	1173.13	-
8	1029.11	1028.7	8	1026.6	-
	CzlA C24S/C25S MW_{calc} = 8208.83			**CzlA C24S/C25S-Ox** (-H_2O, -2H) MW_{calc} = 8188.83	
5	1642.77	1642.35	5	1638.77	1638.25
6	1369.14	1368.67	6	1365.81	1366.27
7	1173.69	1173.43	7	1170.83	-
8	1027.10	1027.1	8	1024.60	1024.50
*Average calculated masses based on CzlA sequences missing *N*-terminal Met. Ox = oxazole.					

Table S9. Calculated *m/z* values for CzlA before and after hydroxylation by CzlI.

z	***Calculated****	***Observed***	*z*	***Calculated****	***Observed***
	CzlA MW_{calc} = 6851.36 Da			**CzlA-OH** MW_{calc} = 6867.36 Da	
4	1713.84	1713.25	4	1717.84	1717.89
5	1371.27	1370.65	5	1374.47	1373.89
6	1142.89	1142.61	6	1145.56	1145.34
7	979.77	979.52	7	982.05	981.72
*Average calculated masses based on CzlA sequences missing *N*-terminal Met					

FIGURES

Figure S1. Morphological characteristics of *S. curacoi* in response to constitutive expression of *bldA*.

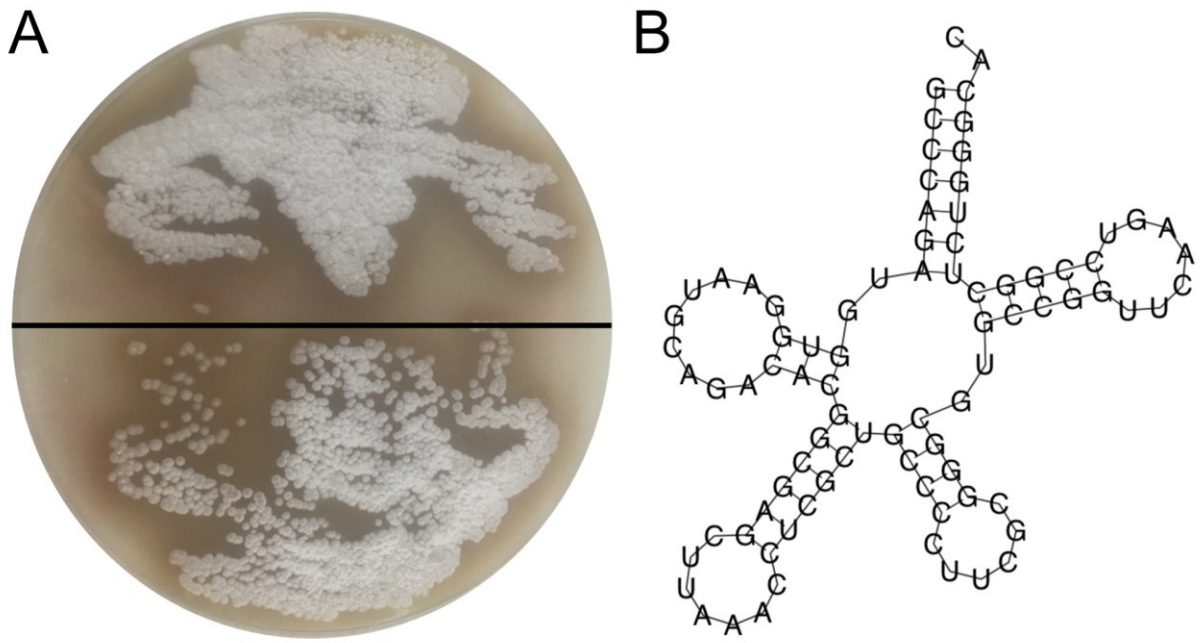

Figure S1. Morphological characteristics of *S. curacoi* in response to constitutive expression of *bldA*. (**A**) Growth of *S. curacoi* DSM40107 (top) and *S. curacoi* / pTESa-*bldA* (bottom) on mannitol-soya agar. (**B**) The predicted secondary structure of the native BldA sequence from *S. curacoi*. The secondary structure was predicted with the RNAfold server (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi).

Figure S2. Key NMR correlation data for Curacozole.

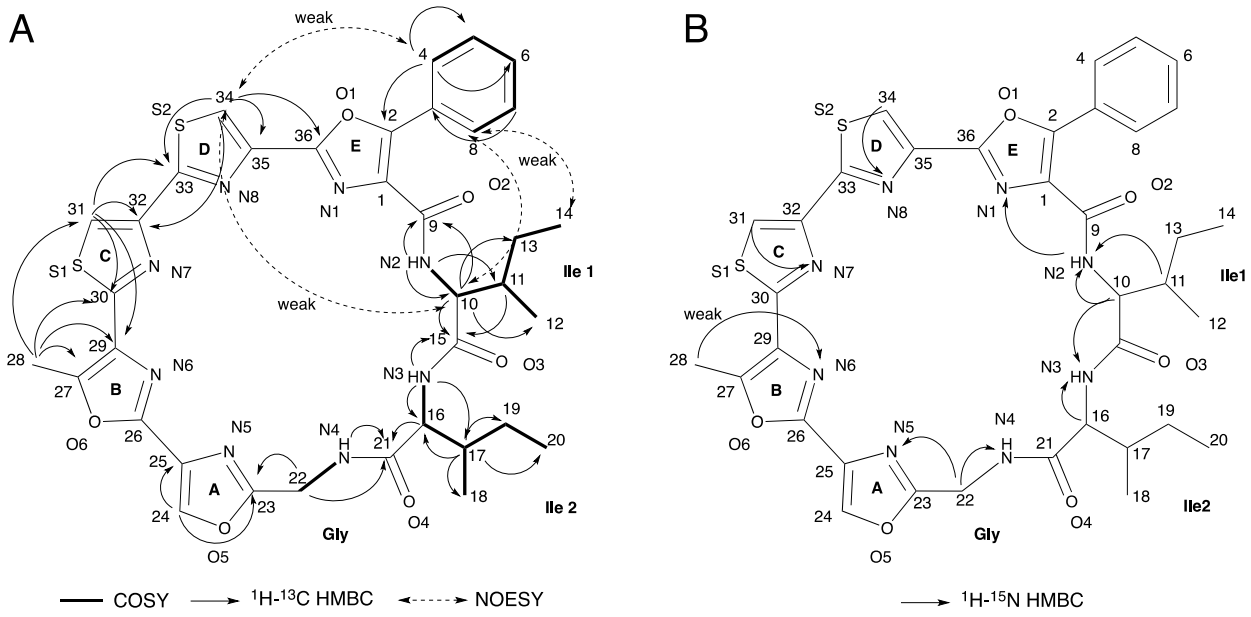

Figure S2. Key NMR correlation data for Curacozole. **(A)** COSY, ^{1}H-^{13}C HMBC, and NOESY correlations. **(B)** ^{1}H-^{15}N HMBC correlations.

Figure S3. A composite of MS^2 and MS^3 CID spectra for Curacozole.

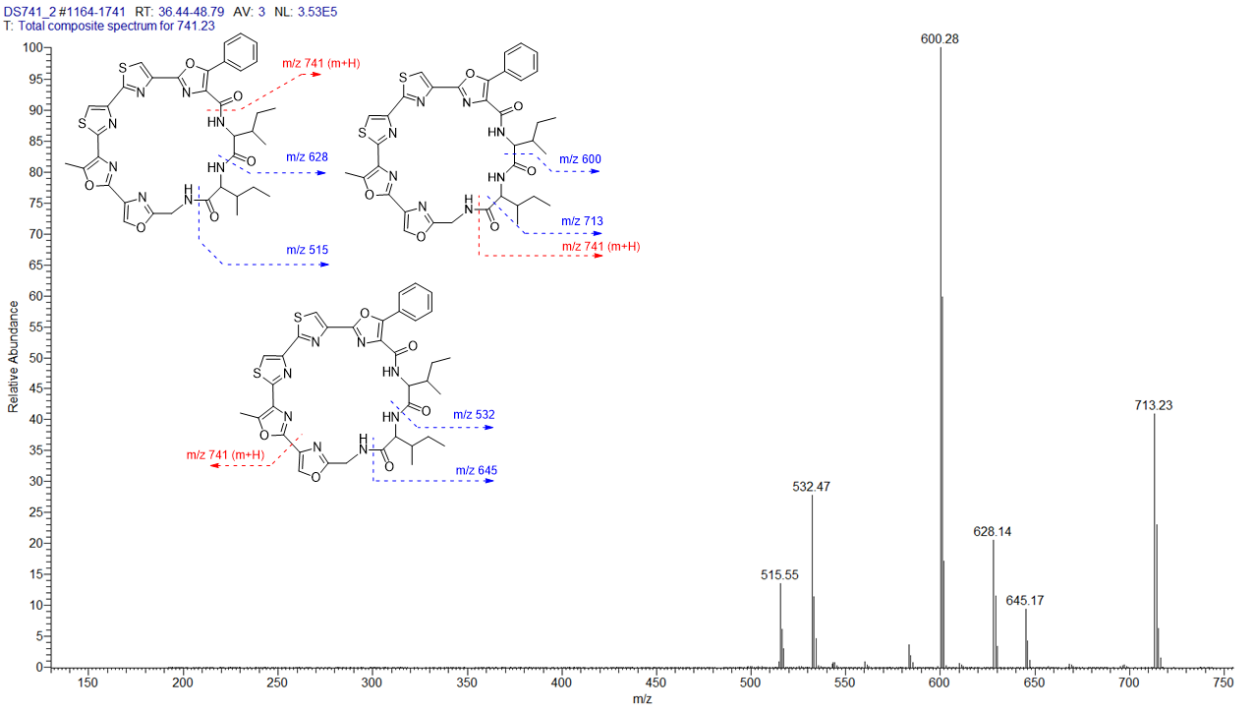

Figure S3. A composite of MS^2 and MS^3 CID spectra for Curacozole. The 741 *m/z* molecular ion (red) was subjected to CID fragmentation (MS^2) and the resulting fragments (blue) were selected for additional (MS^3) fragmentation.

Figure S4. ^{1}H-NMR spectrum of Curacozole (700 MHz, $CDCl_3$)

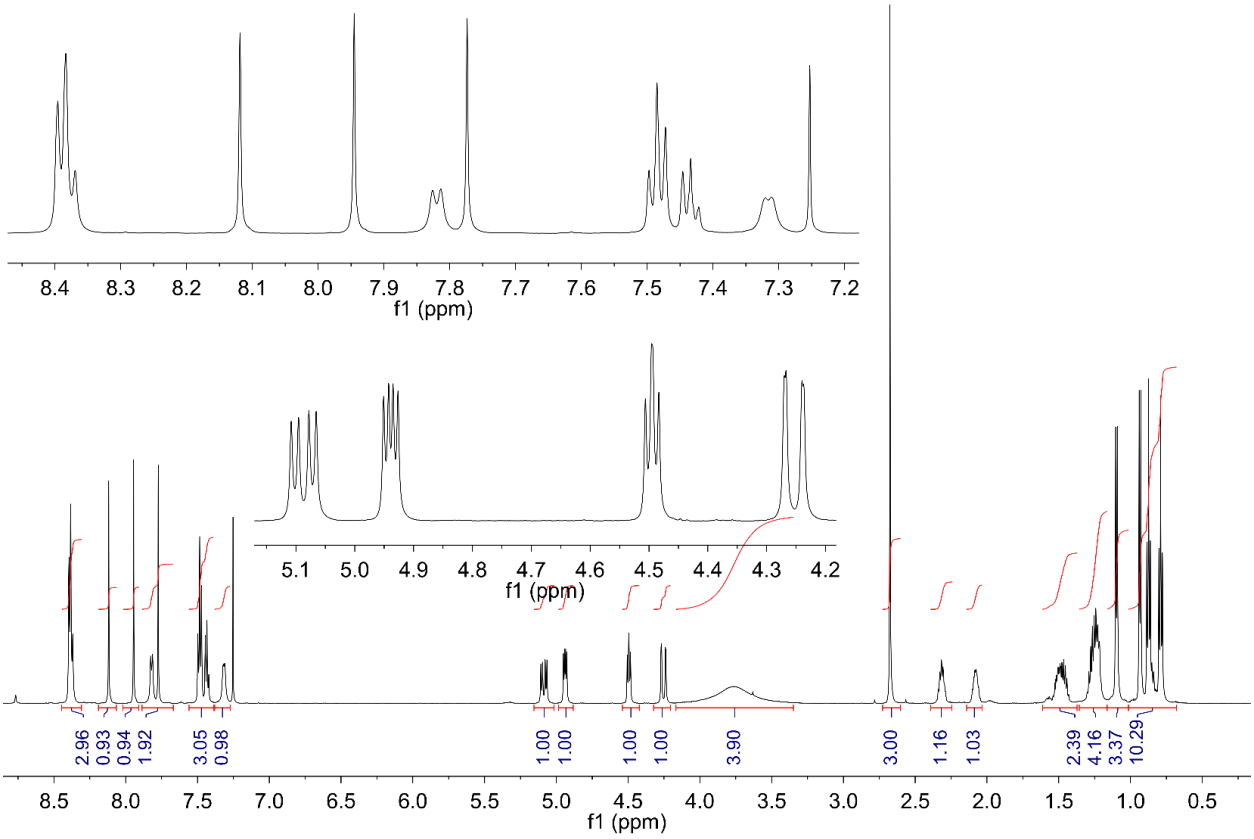

Figure S4. ^{1}H-NMR spectrum of Curacozole (700 MHz, $CDCl_3$). Spectrum continued next page.

Figure S4 (cont.)

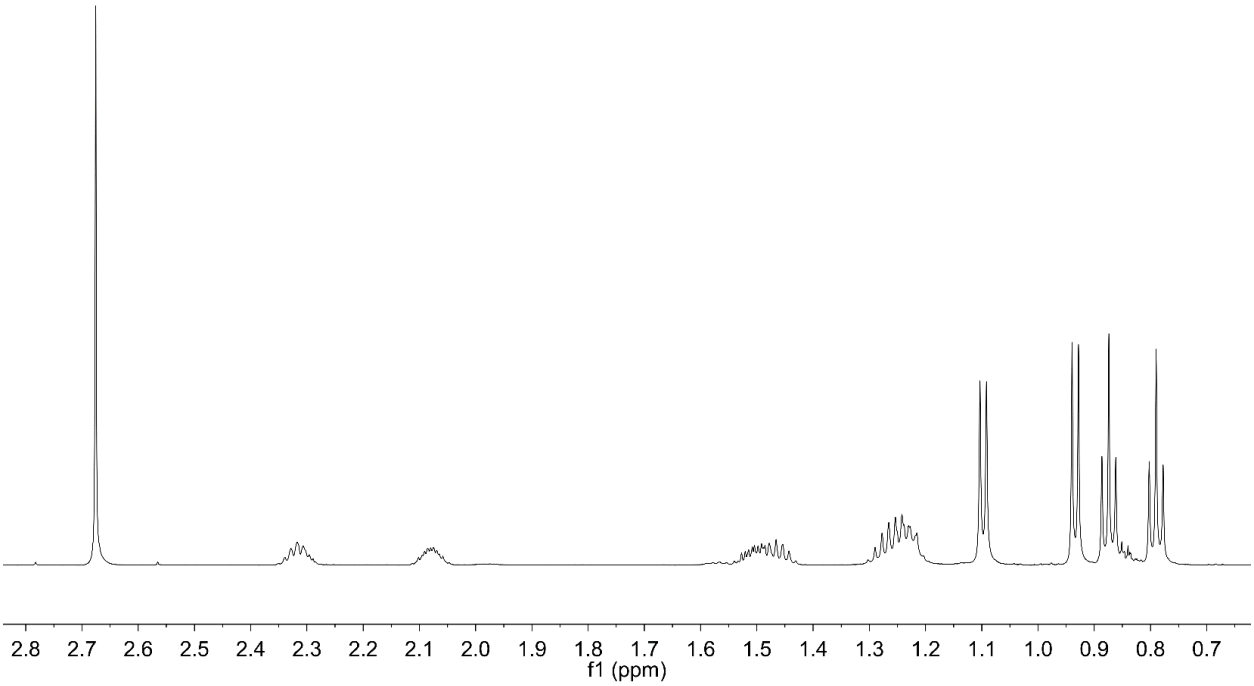

Figure S4 (cont.) Expansion of the upfield region of the ^{1}H-NMR spectrum of Curacozole (700 MHz, $CDCl_3$)

Figure S5. SDS-PAGE analysis of purified CzlA and variants.

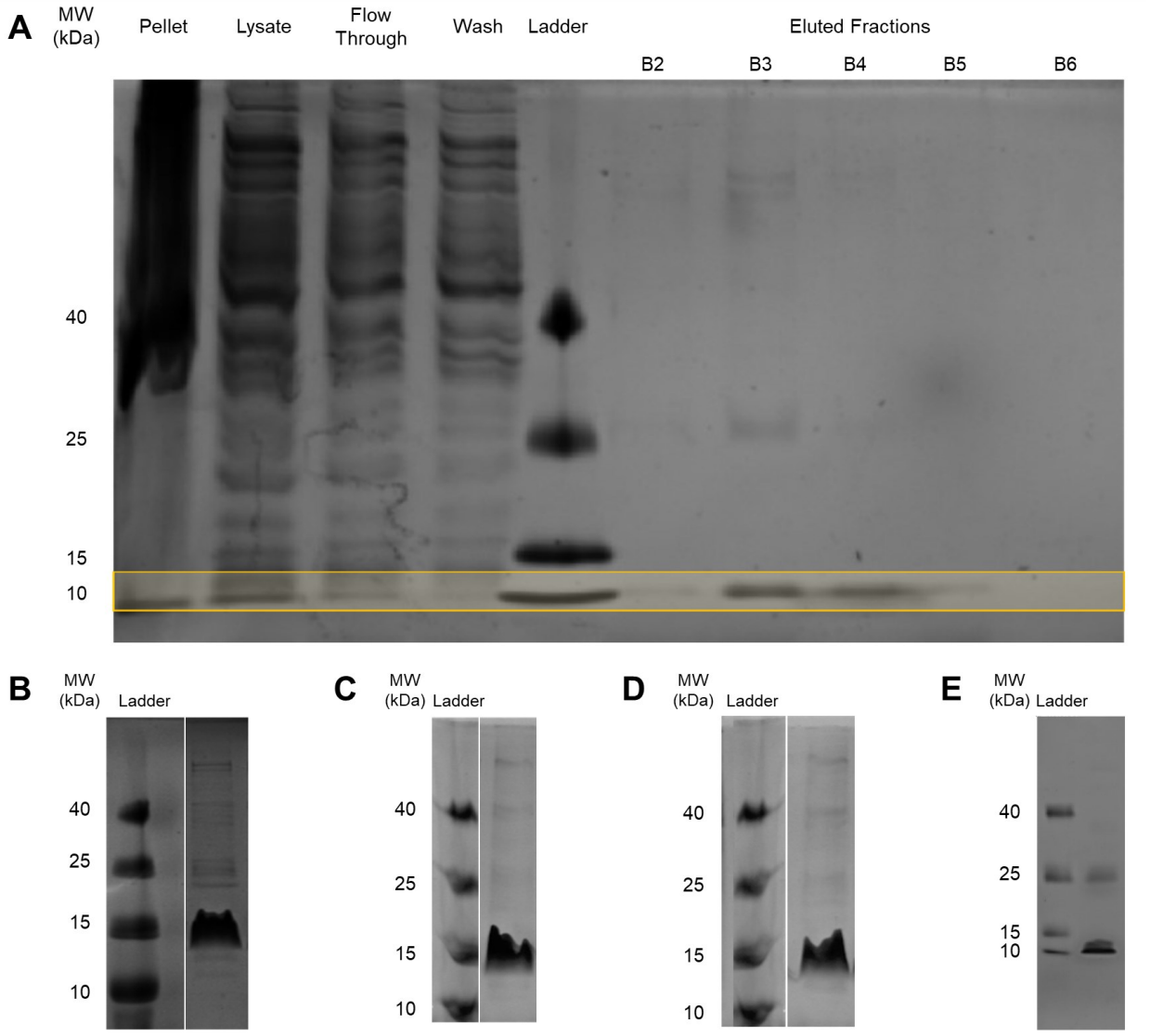

Figure S5. SDS-PAGE analysis of purified CzlA and variants. **A)** CzlA purification. Shown are samples of cell pellet and supernatant following cell lysis and centrifugation. Fractions eluted from the Ni-NTA column include flow-through during column loading, column wash, and fractions from the imidazole gradient (B2-B6). Fractions B3 and B4 were collected. **B)** Purified CzlA C24S. **C)** Purified CzlA C25S. **D)** Purified CzlA C24S/C25S. **E)** Purified CzlA F18G.

Figure S6. SDS-PAGE analysis of purified CzlD.

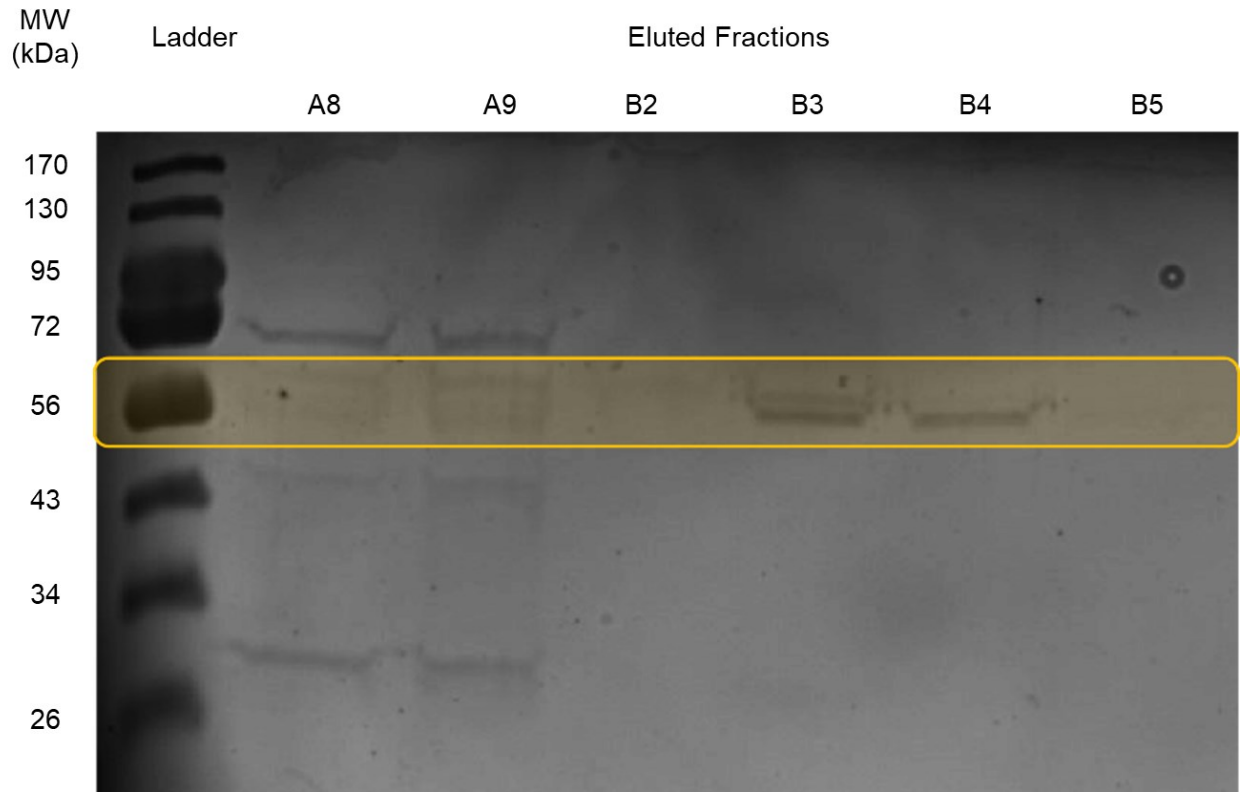

Figure S6. SDS-PAGE analysis of CzlD purification by Ni-NTA chromatography. Fractions B3 and B4 were collected. The predicted MW is 54.6 kDa.

Figure S7. SDS-PAGE analysis of purified CzlBC.

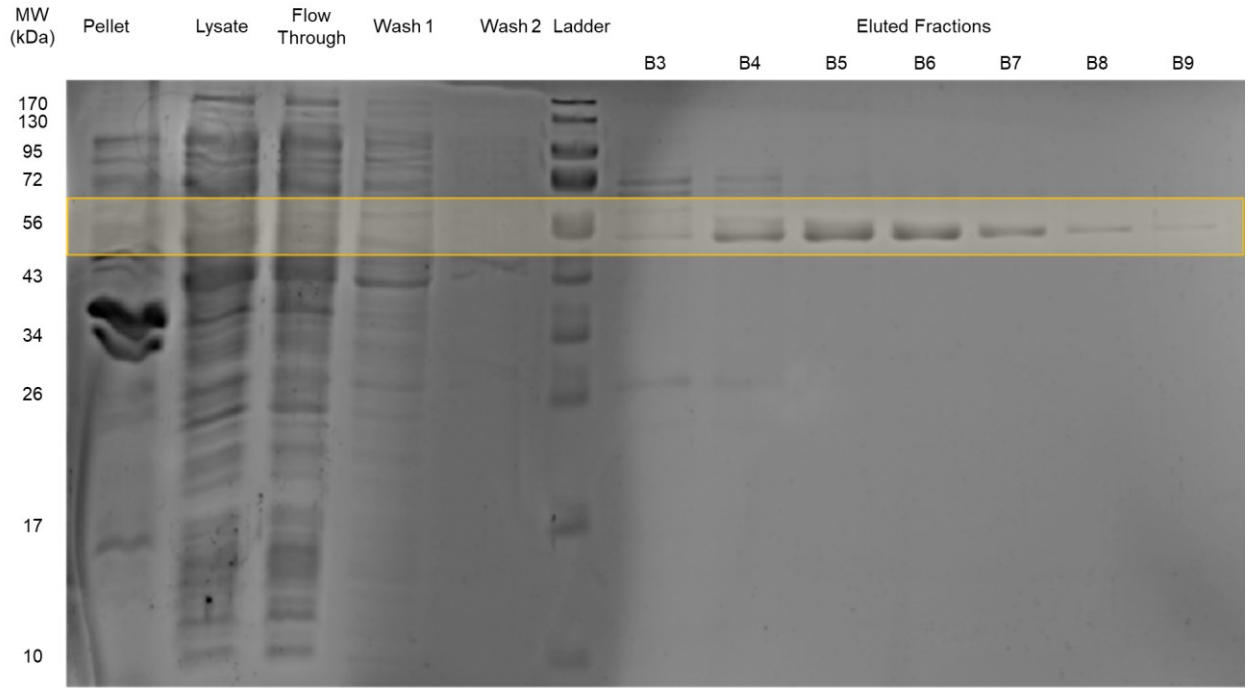

Figure S7. SDS-PAGE analysis of CzlBC purification by Ni-NTA chromatography. Fractions B4-B9 were collected. The predicted MW is 50.8 kDa.

Figure S8. SDS-PAGE analysis of purified CzlI.

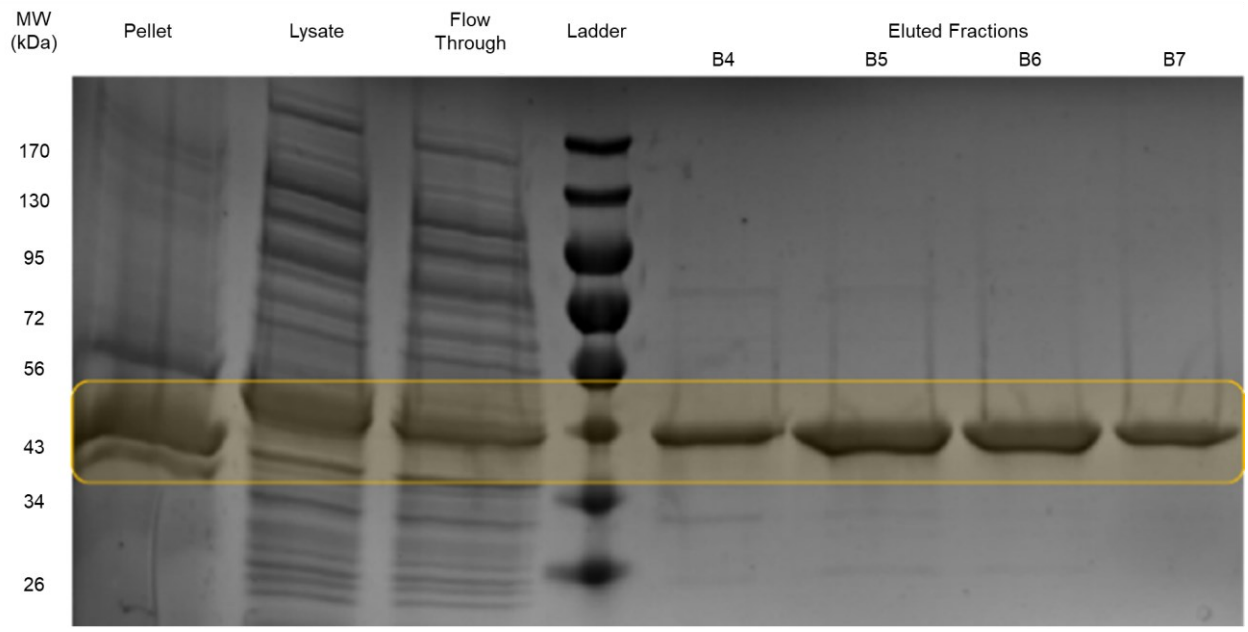

Figure S8. SDS-PAGE analysis of CzlI purification by Ni-NTA chromatography. Fractions B4-B7 were collected. The predicted MW is 40.5 kDa.

Figure S9. Supporting structural predictions for CzlD and CzlBC AlphaFold models.

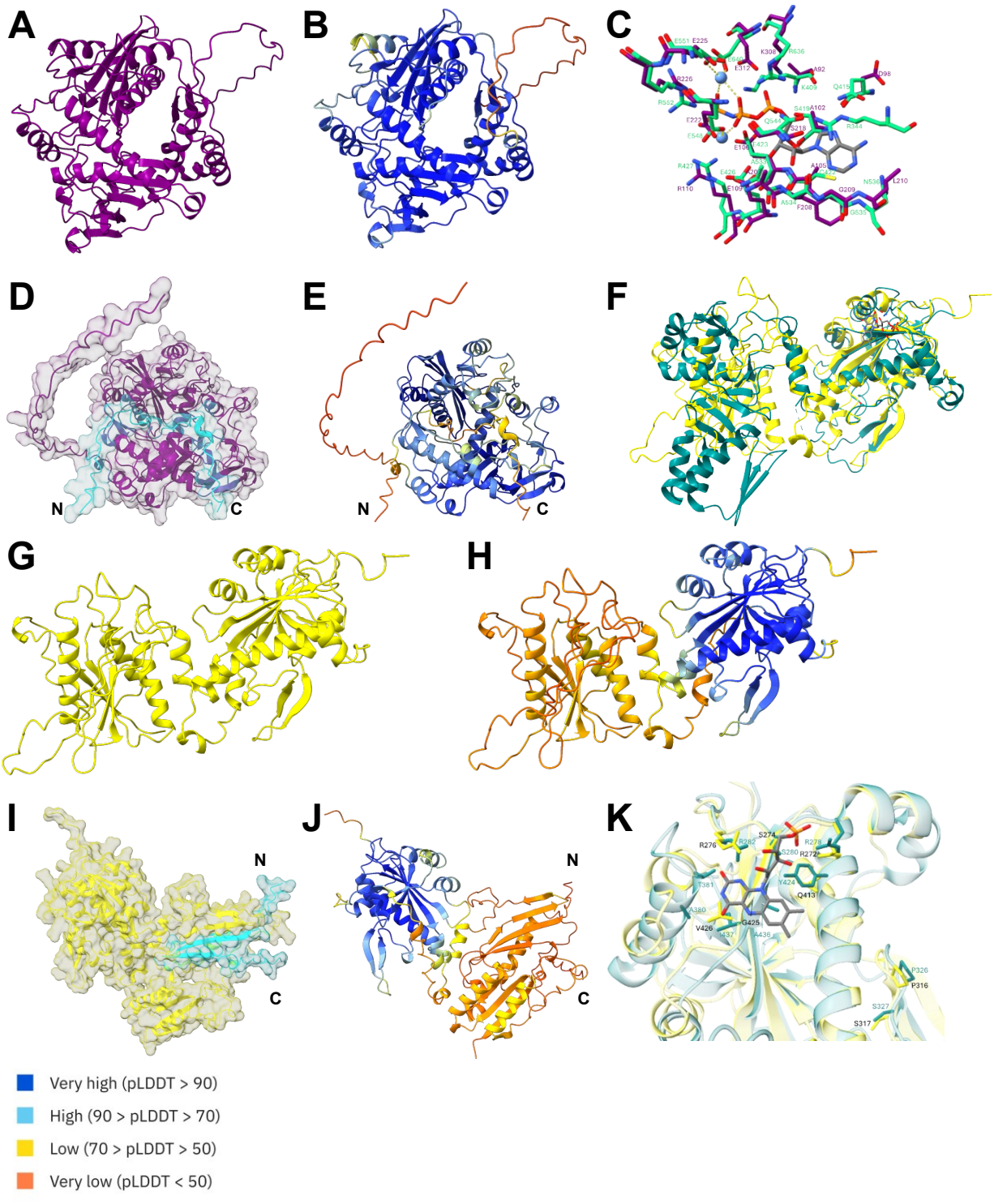

Figure S9. Supporting structural predictions for CzlD and CzlBC AlphaFold2 models. (**A**) CzlD. (**B**) CzlD coloured by pLDDT scoring. (**C**) Active site residues of CzlD (purple)

identified by overlaying the model with LynD (green). Mg^{2+} ions are shown as blue spheres and ANP is shown as grey sticks coloured by heteroatom. (**D**) CzlA (cyan) complexed with CzlD (purple). (**E**) CzlA complexed with CzlD coloured by pLDDT scoring. (**F**) CzlBC (yellow) aligned with ThcOx (PDB ID: 5LQ4; RMSD = 3.8 Å, Z = 28.4). ThcOx (teal) is shown bound to FMN, in grey sticks coloured by heteroatom. (**G**) CzlBC. (**H**) CzlBC coloured by pLDDT scoring. (**I**) CzlA (cyan) complexed with CzlBC (yellow). (**J**) CzlA complexed with CzlBC coloured by pLDDT scoring. (**K**) The FMN-binding residues of CzlBC (yellow) identified by overlaying the model with the known structure of ThcOx (teal). FMN is shown as grey sticks coloured by heteroatom.

Figure S10. Structure based sequence alignment of CzlD and LynD

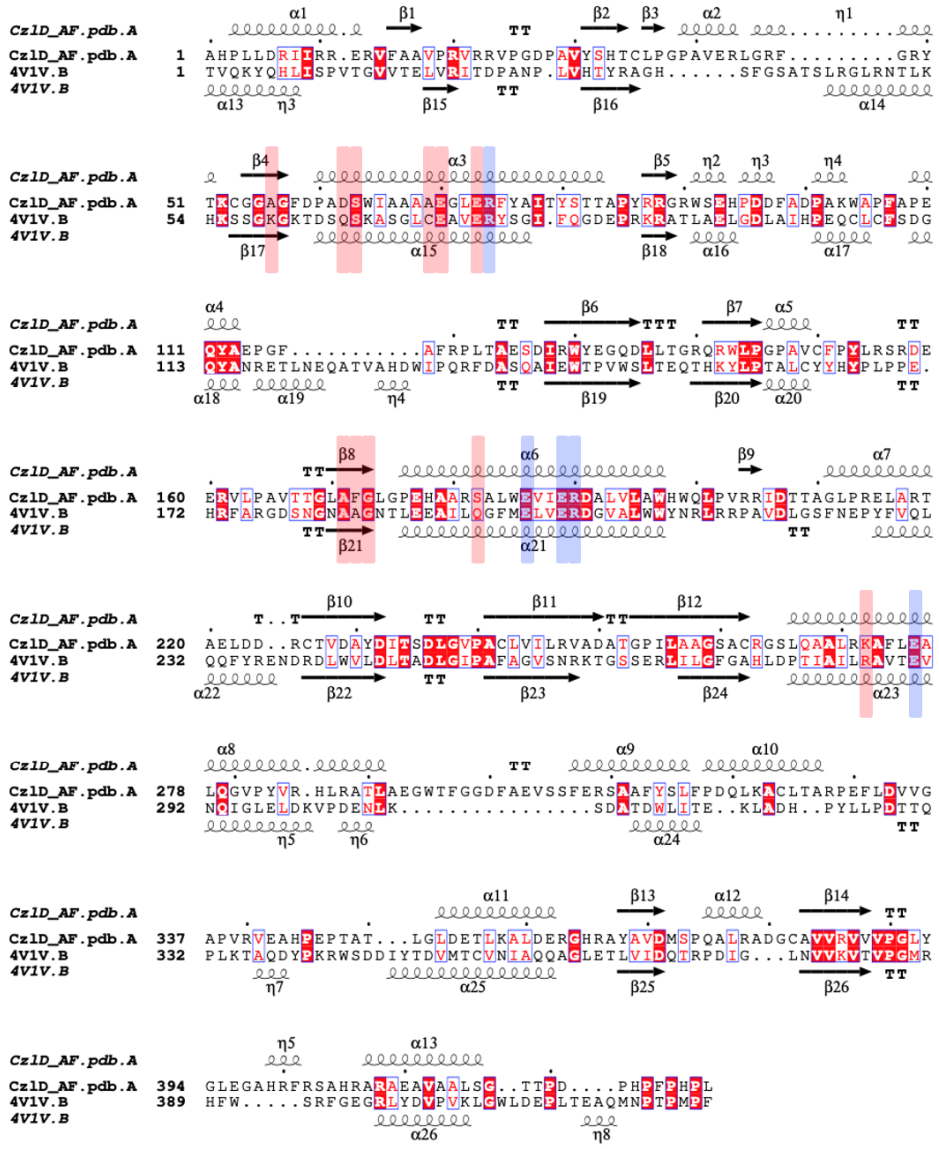

Figure S10. Structure based sequence alignment of CzlD and LynD. The AlphaFold model for CzlD was aligned to LynD (PDB ID: 4V1V-B) using the jFATCAT (flexible)[8] algorithm through the pairwise structural alignment tool of the PDB (https://www.rcsb.org/). The resulting sequence alignment was annotated using ESPript 3.0.[9] ATP and Mg^{2+} binding residues in LynD are highlighted with opaque red and blue bars, respectively.

Figure S11. Structural predictions for CzlD/BC/A heteromeric complexes.

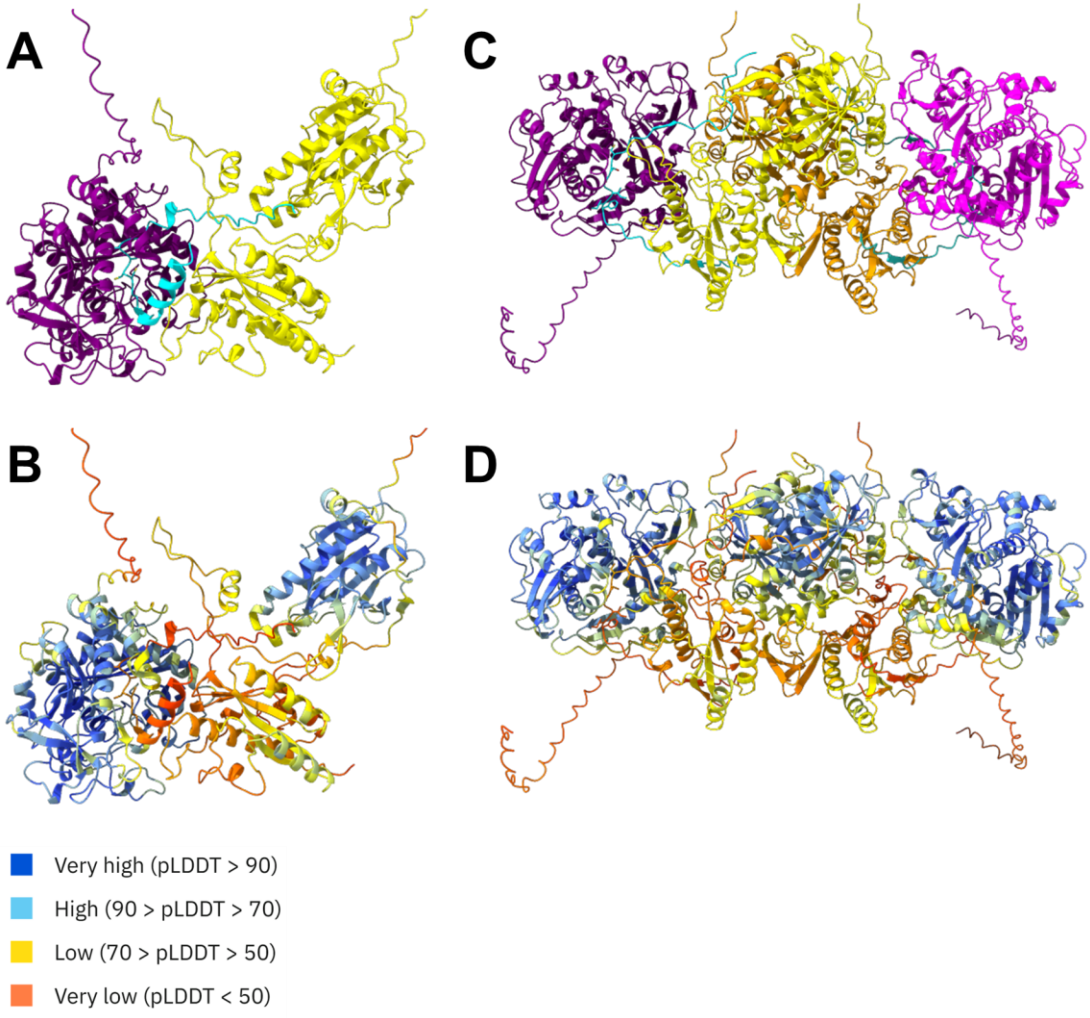

Figure S11. Structural predictions for CzlD/BC/A heteromers by AlphaFold3. (**A**) CzlD/BC/A heterotrimer including CzlD (purple), CzlBC (yellow), and CzlA (cyan) with $CzlA_{22\text{-}25}$ shown as sticks coloured by heteroatom. Confidence scores ipTM = 0.24 and pTM = 0.5. (**B**) CzlD/BC/A coloured by pLDDT. (**C**) $(CzlD/BC/A)_2$ heterohexamer model (ipTM = 0.31, pTM = 0.4) of CzlD (purple/magenta), CzlBC (yellow/orange), and CzlA (cyan/teal). CzlA S22, T23, C24, and C25 are shown as sticks coloured by heteroatom. (**D**) $(CzlD/BC/A)_2$ coloured by pLDDT.

Figure S12. Structure based sequence alignment of CzlBC and ThcOx (5LQ4-B).

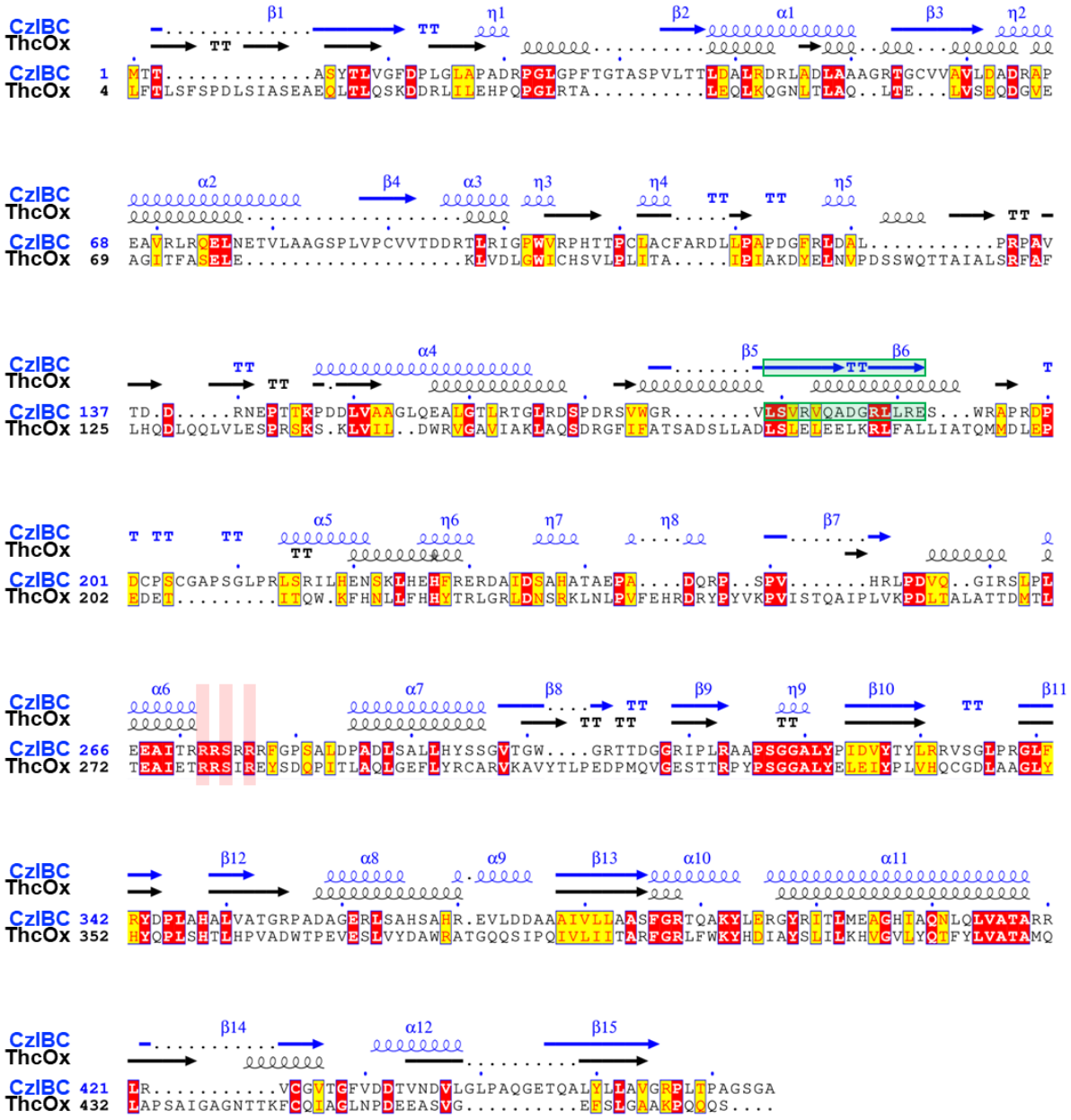

Figure S12. Structure based sequence alignment of CzlBC and ThcOx. The AlphaFold model for CzlBC (complexed with CzlA) was aligned to ThcOx (PDB ID: 5LQ4-B) using ENDscript 2.0.[9] The resulting sequence alignment was annotated using ESPript 3.0.[9] The predicted peptide-recruiting β-sheet residues are highlighted with green bars. FMN-binding residues are highlighted with red bars.

Figure S13. CzlA is not modified by CzlD without CzlBC.

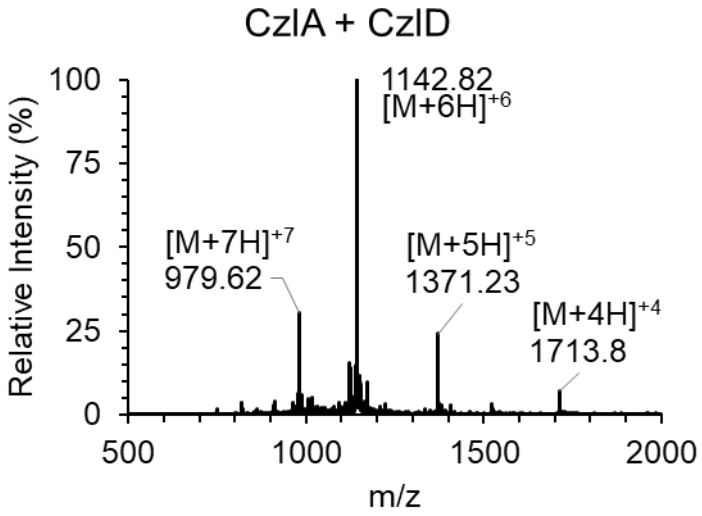

Figure S13. CzlD does not modify CzlA when CzlBC is not present. UPLC-MS analysis of in vitro reaction of CzlA with CzlD. CzlA MW_{calc} = 6851.36 Da.

Figure S14. CzlD and CzlBC form one heterocycle on CzlA and variants.

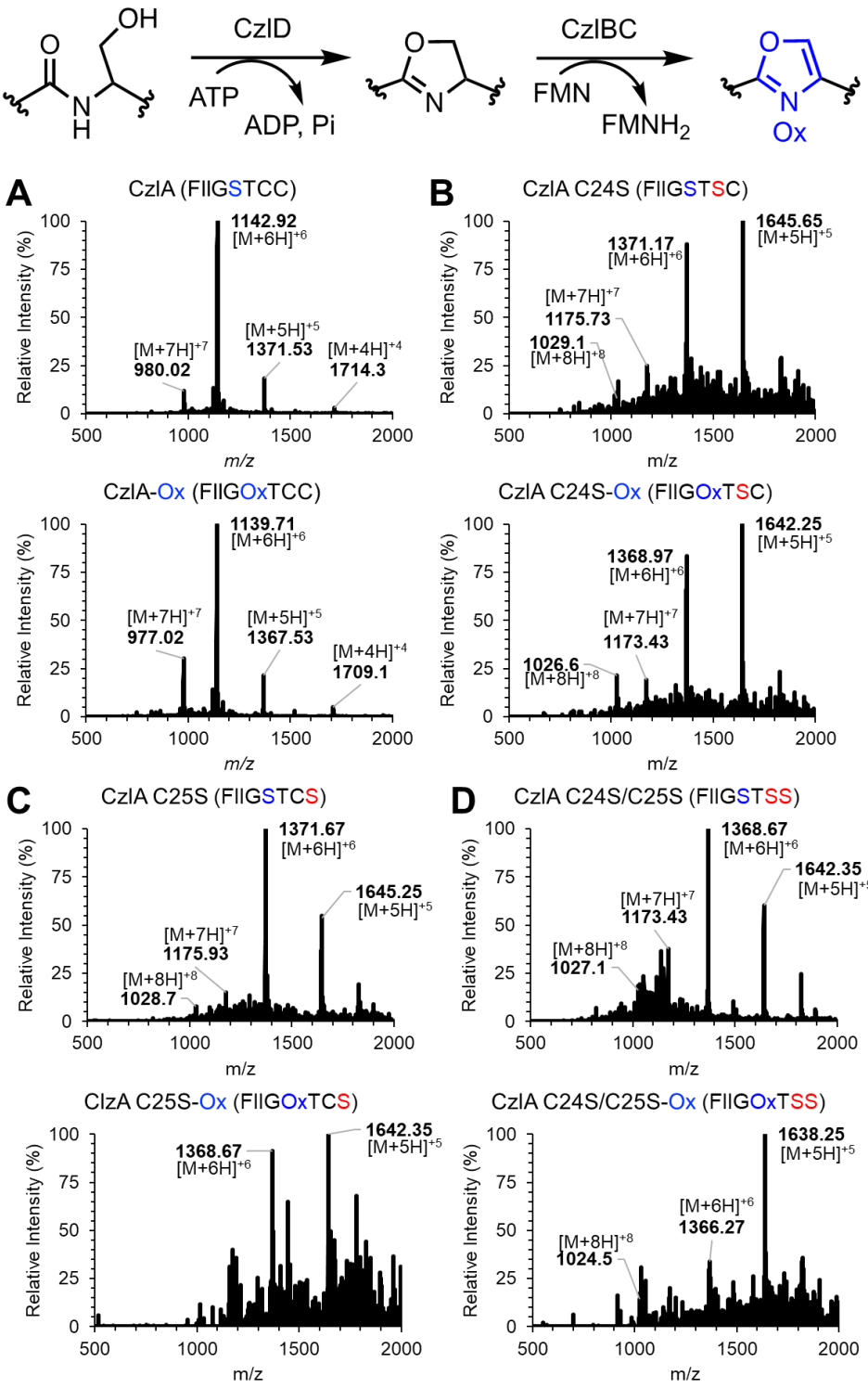

Figure S14. CzlD and CzlBC form one heterocycle on CzlA and variants. Shown are UPLC-MS spectra of reaction products with CzlA (**A**) and the variants: C24S (**B**), CzlA

C25S (**C**), and CzlA C24S/C25S (**D**). Upper panels and lower panels correspond MS of CzlA variants before and after the reaction, respectively. Core peptide sequences are shown in parentheses, blue letters correspond to modified Ser where Ox = oxazole, red letters indicate amino acid substitutions. See Tables S7 and S8 for calculated and observed masses.

Figure S15. MS/MS sequencing of oxazole-containing CzlA C24S/S25S.

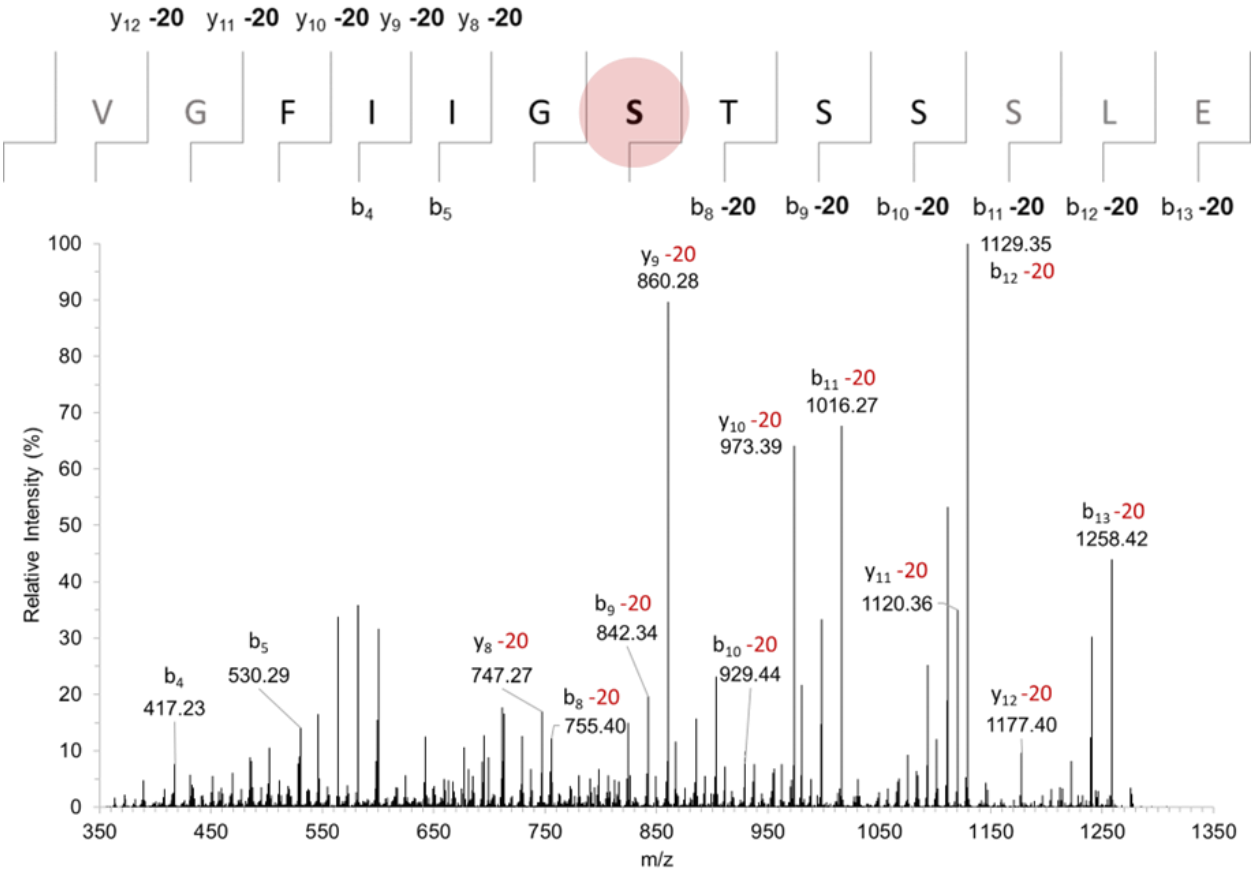

Figure S15. MS/MS sequencing of oxazole-containing CzlA C24S/S25S. The peptide was trimmed with GluC to release the residues 16 through 28 (shown above), which was then sequenced by fragmentation of the molecular ion (*m/z* = 1276.6, $[M+H]^{+1}$). The modified Ser22 residue is highlighted as a red sphere. Ions that deviate by 20 Da less than the expected value for the native sequence, and thus contain a single oxazole, are indicated.

Figure S16. Supporting structural predictions for CzlI AlphaFold models.

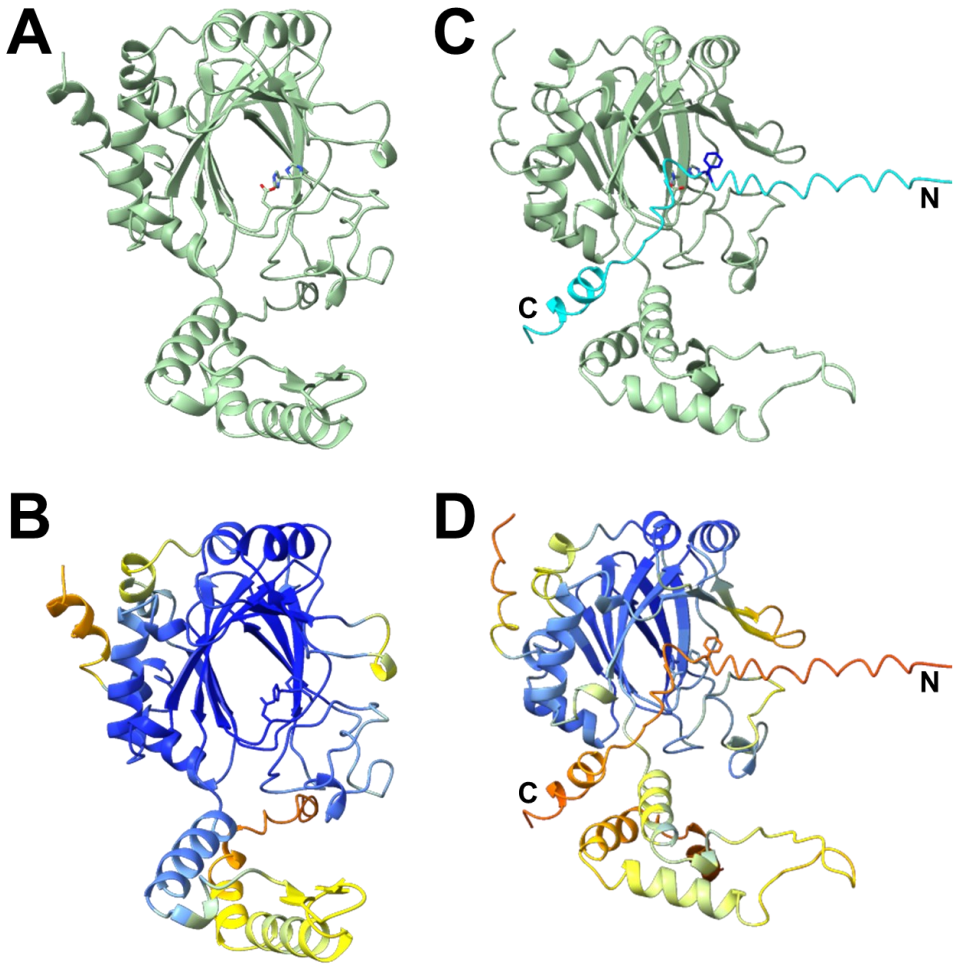

Figure S16. Supporting structural predictions for CzlI AlphaFold2 models. (**A**) CzlI. His 145, Asp 147, and His 224 are shown as sticks coloured by heteroatom. (**B**) CzlI coloured by pLDDT scoring. (**C**) CzlA (cyan) complexed with CzlI (green). Phe18 of CzlA is shown as blue sticks. (**D**) CzlA complexed with CzlI coloured by pLDDT scoring.

Figure S17. Structural alignment of CzlI with asparaginyl hydroxylase YxbC

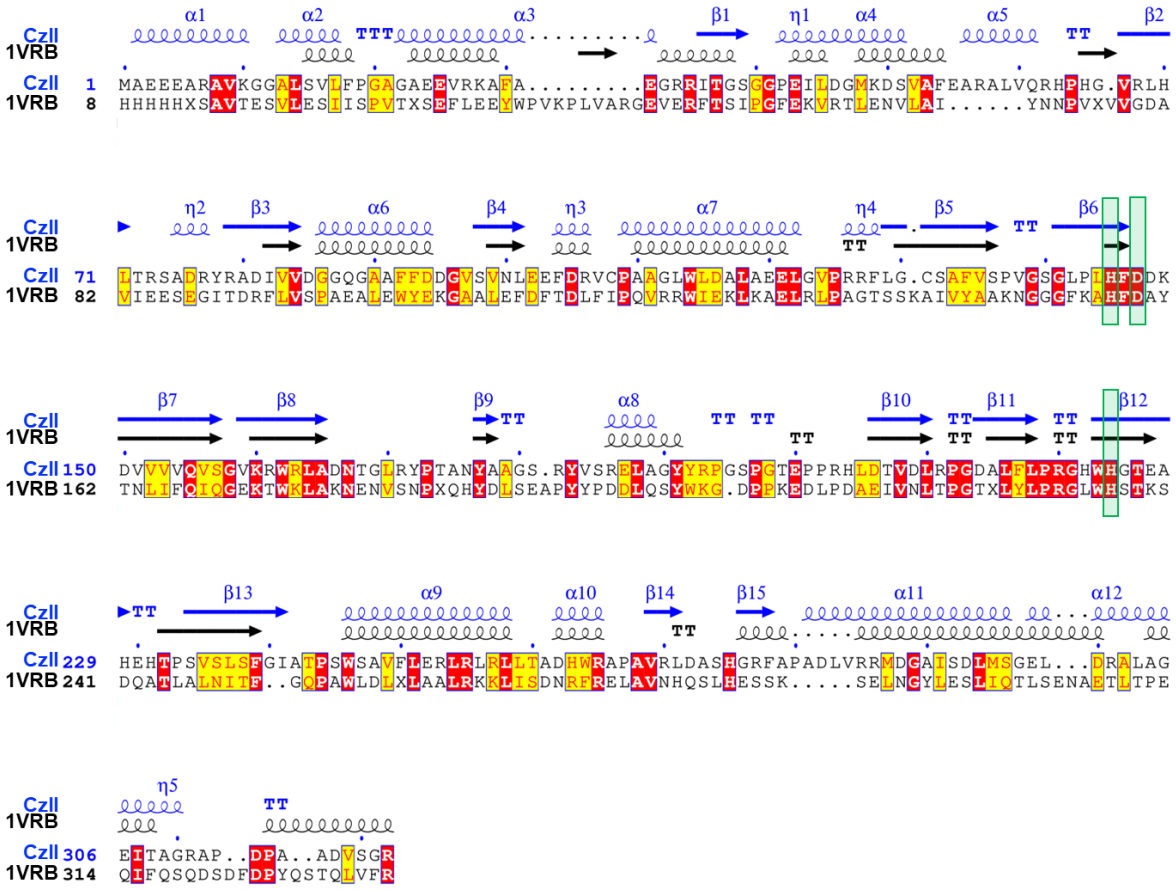

Figure S17. Structure based sequence alignment of CzlI and putative asparaginyl hydroxylase YxbC. The AlphaFold model for CzlI was aligned to the hydroxylase (PDB ID: 1VRB-A) using ENDscript 2.0.[9] The resulting sequence alignment was annotated using ESPript 3.0.[9] The predicted iron-coordinating residues are highlighted with green bars.

Figure S18. Hydroxylation of CzlA by the Fe/2OG dioxygenase CzlI

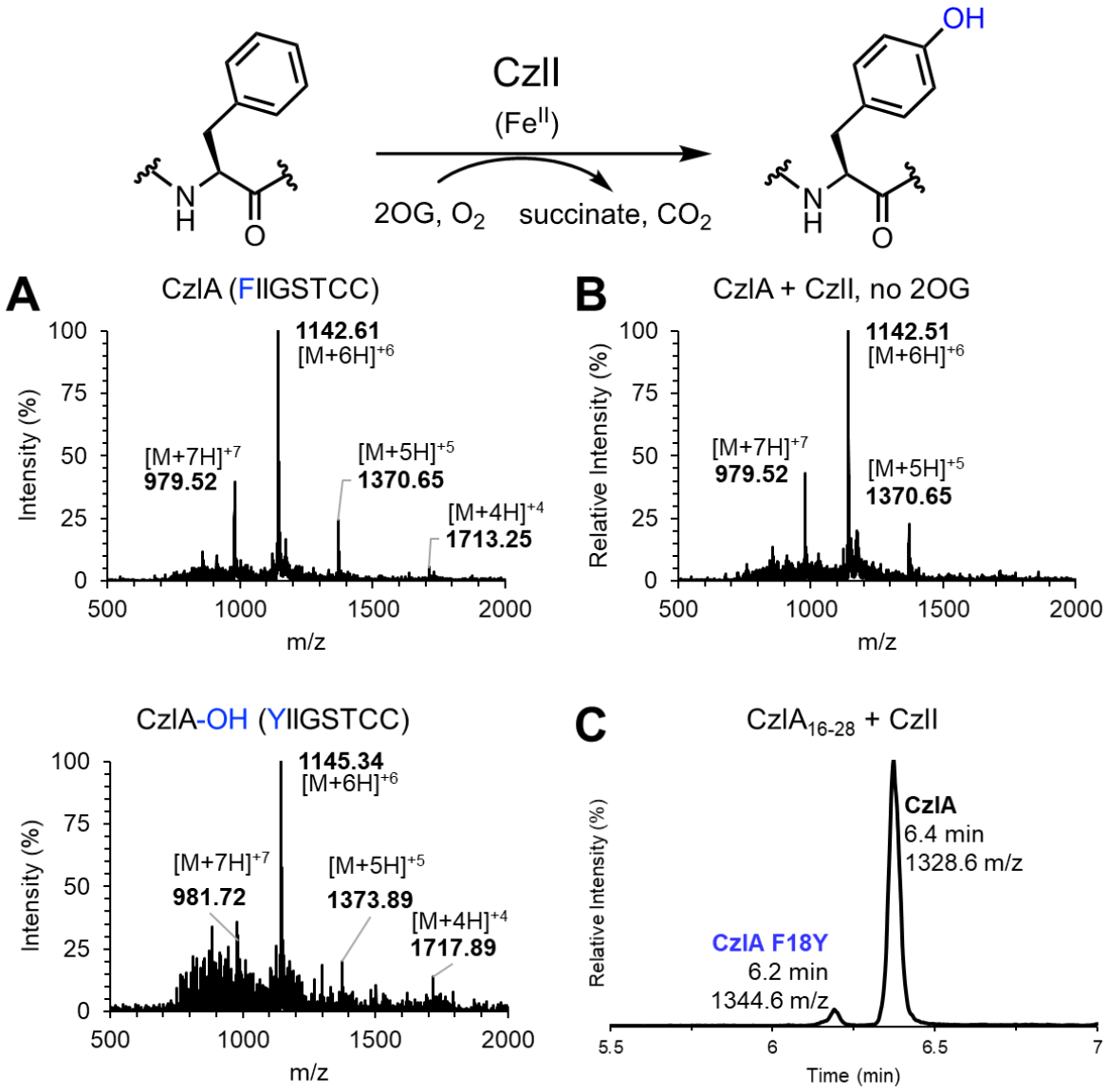

Figure S18. Hydroxylation of CzlA by the Fe/2OG dioxygenase CzlI. (A) UPLC-MS spectra of CzlA (upper panel) and CzlA following reaction with CzlI (lower panel). (**B**) Reaction of CzlA with CzlI in the absence of 2-oxoglutarate (2OG). See **Table S9** for calculated and observed masses. (**C**) UPLC-MS SIR chromatogram at 1328.6 and 1344.6 m/z for the reaction of $CzlA_{16-28}$, produced by proteolysis by GluC, with CzlI.

Figure S19. CzlA F18G is not modified by CzlI

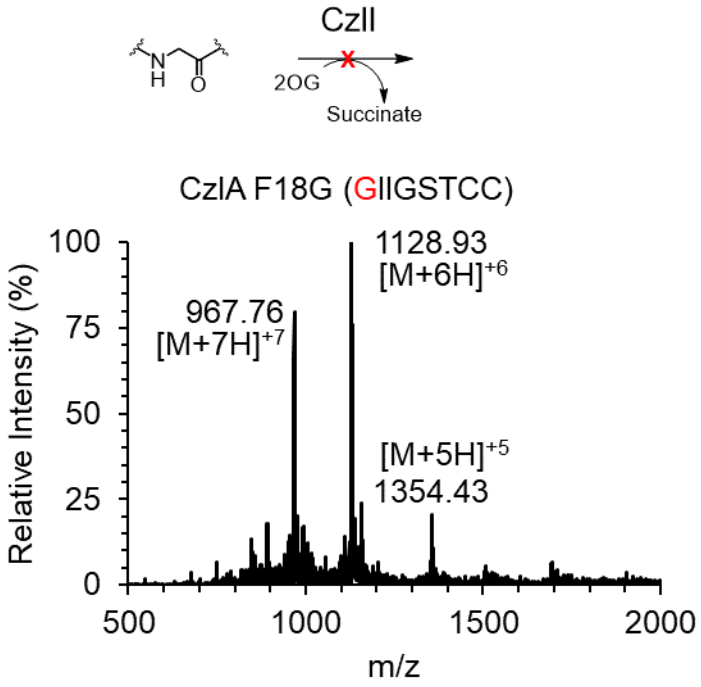

Figure S19. CzlI does not modify CzlA F18G. UPLC-MS analysis of in vitro reaction of CzlA F18G with CzlI. Unmodified CzlA F18G MW_{calc} = 6770.25 Da. The core peptide sequence is shown in parentheses.

Figure S20. The Fe/2OG dioxygenase CzlI creates a tyrosine residue on precursor peptide CzlA.

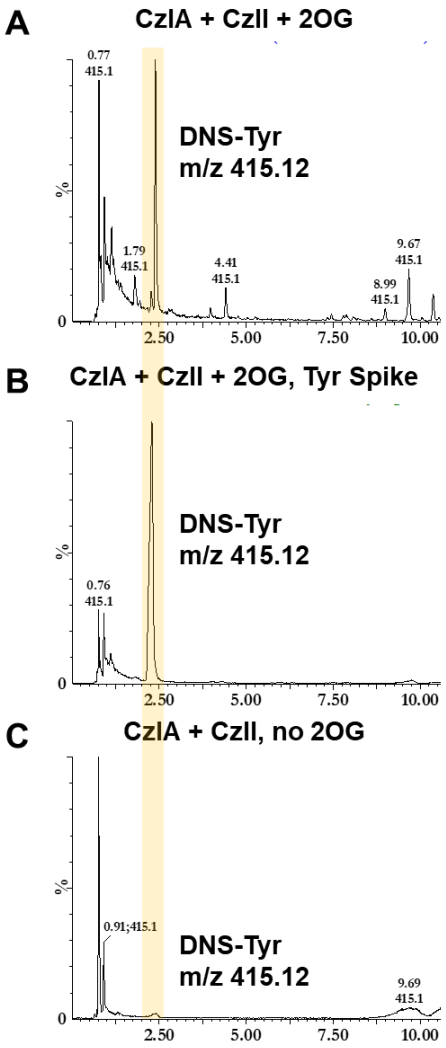

Figure S20. The Fe/2OG dioxygenase CzlI creates a tyrosine residue on precursor peptide CzlA. UPLC-MS analysis of CzlA following reaction with CzlI, total peptide hydrolysis, and dansylation of the resulting amino acids. The chromatograms show single ion recordings for *m/z* = 415.12 corresponding to dansylated L-tyrosine (DNS-Tyr), with corresponding peak highlighted in yellow. **(A)** CzlA after reaction with CzlI. (**B**) Addition of DNS-Tyr standard to the previous sample. **(C)** Analysis of CzlA following reaction with CzlI in the absence of the co-substrate 2OG.

Figure S21. UPLC-MS analysis of dansylated tyrosine structural isomers.

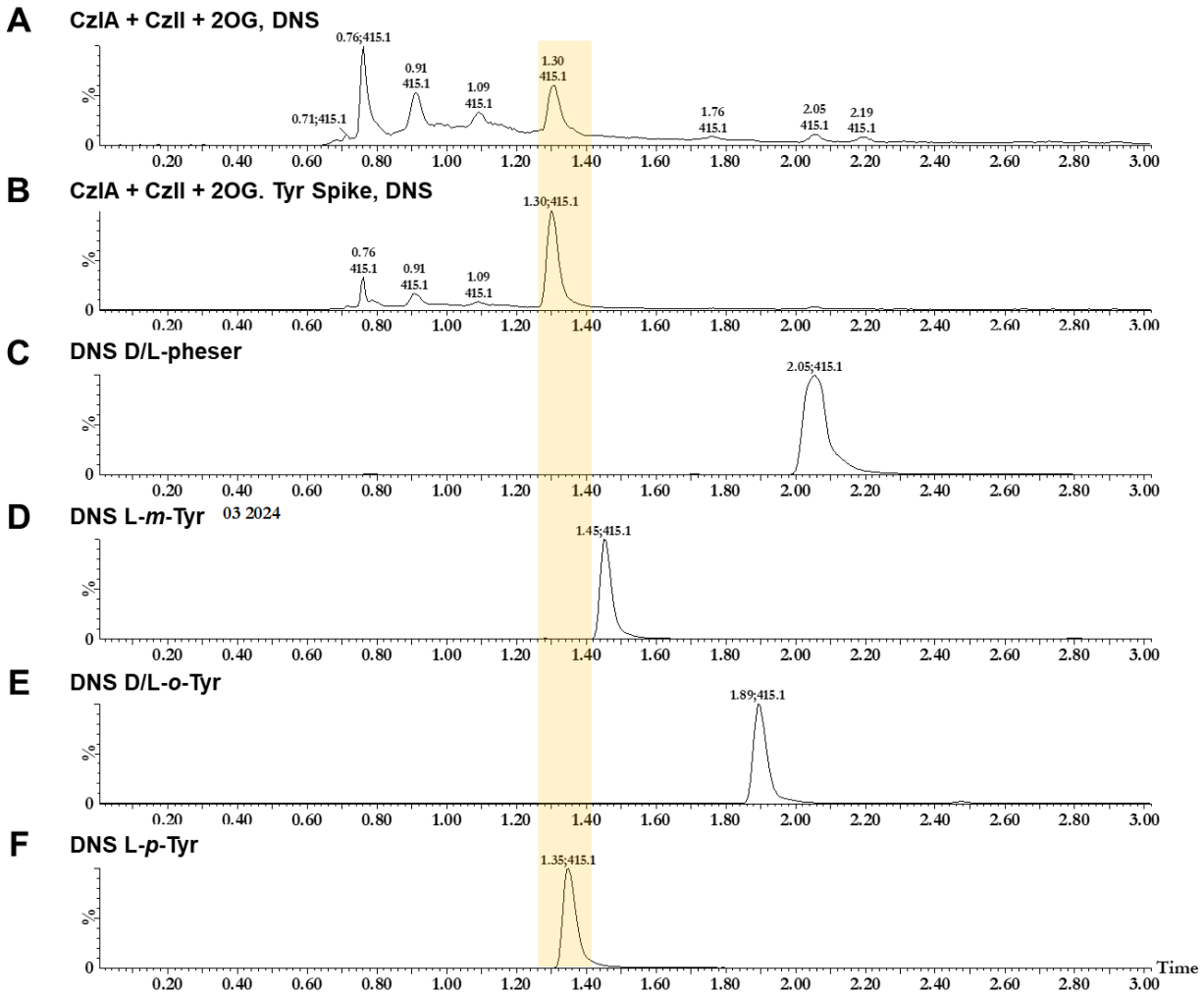

Figure S21. UPLC-MS analysis of dansylated tyrosine structural isomers. The chromatograms show single ion recordings for *m/z* 415.12 corresponding to dansylated tyrosine isomers. (**A**) CzlA following reaction with CzlI, total peptide hydrolysis, and dansylation of the resulting amino acids. (**B**) Addition of dansylated L-Tyr to the previous sample. (**C**) dansylated D/L-phenylserine. (**D**) dansylated L-meta-Tyr. (**E**) dansylated D/L-ortho-Tyr (**F**) dansylated L-Tyr.

Supplementary Experimental Methods

General

Reagents and solvents were obtained from Sigma-Aldrich, Fisher Scientific, Fluka, or Bioshop Life Science Products and used without further purification. Tryptic soy broth (TSB) was obtained from EMD Millipore. Solvents used in flash chromatography were of HPLC-grade. Silica gel flash chromatography was conducted on Silia-P Flash Silica Gel (40-63 μm, Silicycle). Flash chromatography was performed on a Biotage SP1 system. UV-visible spectra were recorded in CH_2Cl_2 on a Perkin-Elmer Lambda XLS+ UV-visible spectrophotometer. IR spectra were recorded on a Bruker Alpha Platinum ATR FT-IR spectrometer. Optical rotation measured with Anton-Parr MCP200 Polarimeter. NMR spectra were recorded on a Bruker UltraShield 600 MHz and a Bruker Ascend 700 MHz NMR spectrometer with $CDCl_3$ as a solvent. High resolution ESI-MS experiments were performed on a Thermo Scientific Orbitrap Velos Pro instrument.

The bacterial strains and plasmids used in this study are described in Table 1. Streptomyces strains were grown on solid or liquid media supplemented with antibiotics at the following concentrations where appropriate: apramycin, kanamycin, hygromycin: 50 μg mL^{-1}; ampicillin, 100 μg mL^{-1}; Fosfomycin, 200 μg mL^{-1}. Streptomyces strains were stored suspending mycelium in 50% glycerol or 25% sucrose and freezing at -80°C. For growth on solid media, glycerol stocks were streaked onto mannitol-soya (MS) agar (Per litre: 20 g D-mannitol, 20 g soy protein, 20 g agar, 10 mM $MgCl_2$, pH 7.2). Production medium consisted of (per litre): 20 g starch, 10 g glucose, 5 g peptone, 5 g yeast extract, 4 g NaCl, 0.5 g K_2HPO_4, 0.5 g $MgSO_4$, 2 g $CaCO_3$.

Large scale production and purification of curacozole

For large scale production of curacozole, a TSB culture of *S. curacoi* / pTESa-*bldA* was used to inoculate (2.5 % v/v) 1 L of production media in a 2.8 L baffled Fernbach flask containing a corrosion resistant metal spring (1 cm diameter) coiled into 12 cm diameter loop. The production cultures were incubated at 28°C, 180 rpm, for 7 days. *S. curacoi* / pTESa-*bldA* cells from the production cultures (6 L) were removed by centrifugation (4 °C, 2600 x *g*) and the resulting supernatant returned to the shake flasks. Amberlite XAD4 beads were added to the supernatant at 20 g L^{-1} and shaken at 28°C and 180 rpm for 2 hrs. The cell pellet and collected XAD4 beads were washed with acetone (4 L) and the solvent evaporated in vacuo to afford an aqueous concentrate (~500 mL). The aqueous extract was defatted by extracting with petroleum ether (3 x 500 mL) followed by extraction with n-butanol (3 x 500 mL). The solvent was removed in vacuo and the concentrate loaded onto a Biotage C18HS samplet and separated by linear gradient from 5% acetonitrile to 95 % acetonitrile over 90 minutes. Fractions containing the 741 *m/z* ion corresponding to curacozole were pooled, concentrated in vacuo, and subjected to flash silica gel chromatography using a stepwise gradient of 2%, 4%, and 6% MeOH in CH_2Cl_2. Evaporation of the curacozole containing fractions afforded 251 mg of an amorphous yellowish-white solid.

Cloning and Mutagenesis of the Curacozole BGC

Generation of pTESa-*czl*, pTESa-*czl*ΔI, and pTESa-*czl*ΔJK

Genomic DNA (gDNA) was isolated from *S. curcoi* DSM40107 using the salting-out method.[4] A ~11 kbp sequence containing the curacozole BGC (spanning from *czl*I to *czl*K) was amplified by PCR from *S. curacoi* gDNA using primers cur-xbaIfw and cur-EcoRIrv

(Table S5) and PrimeStar GXL DNA polymerase (Takara Bio). The resulting PCR product was treated with XbaI and EcoRI then ligated into the corresponding sites of pBSK. Subsequently the *czI* sequence was subcloned into the XbaI and EcoRI sites of the streptomyces integrative plasmid pTESa (Table S4), yielding pTESa-*czI*. The plasmids pTESa-*czI*ΔI and pTESa-*czI*ΔJK were generated similarly by PCR using the primer pairs czIdelI fwxbaI / cur2 for pTESa-*czI*ΔI and cur-xbaI fw / czIdelJ rv for pTESa-*czI*ΔJK, which amplify sequences that exclude the flanking genes *czI*I and *czI*J / K (Table S5). The resulting PCR products were digested with XbaI, then ligated into the XbaI (sticky) and EcoRV (blunt) sites of pTESa. The resulting pTESa-*czI*, pTESa-*czI*ΔI, and pTESa-*czI*ΔJK plasmids were transformed into *E. coli* ET12567 / pUZ8002 cells then introduced into *S. coelicolor* CH999 via intergeneric conjugation.

Generation of pTESa-*czI*ΔC1, pTESa-*czI*ΔD, pTESa-*czI*ΔF, pTESa-*czI*ΔBC

Single gene deletions of *czI*C1, *czI*D, *czI*F, and *czI*BC within pTESa-*czI* were performed using the λ-RED based recombination protocol described by Gust and co-workers[5]. Primer pairs were designed with the following elements: (1) 5'-sequences of approximately 40 bp that are complimentary to the regions that flank the gene to be deleted, (2) 3'-sequences to amplify the spectinomycin resistance gene from pCDFDuet_1, and (3) a SpeI restriction site between the 5'- and 3'-sequences (**Table S5**). PCR reactions with each primer pair and pCDFDuet_1 therefore generate a PCR product containing a spectinomycin resistance gene flanked by SpeI sites and 40 kb homology sequences that target the gene of interest. Each PCR product was purified by agarose gel electrophoresis then electroporated into *E. coli* BW25113 cells containing the plasmid pTESa-*czI* and the plasmid pIJ790 encoding the λ-phage Red proteins[5].

Transformants were selected on LB-agar containing apramycin (to select for pTESa-*czI*) and spectinomycin (to select for the integrated PCR product). The resulting *E. coli* clones contained pTESa-*czI* where the targeted gene on the plasmid had been replaced by the streptomycin resistance gene via λ-Red mediated homologous recombination (**Figure S5**). The desired mutations in pTESa-*czI* were verified by PCR. Next, the pTESa-*czI* derivative was treated with SpeI then re-ligated to remove the spectinomycin resistance gene, thereby leaving a short SpeI recognition sequence in place of the original *czI* gene that was deleted. The resulting pTESa-*czI* gene deletion derivatives were then transformed into *S. coelicolor* CH999 or *S. lividans* ΔYA9 by intergeneric conjugation from *E. coli*.

Deletion of *czI*B1 and *czI*E in *S. curacoi* / pTESa-*bldA*

To replace *czI*B1 and *czI*E in *S. curacoi* / pTESa-*bldA* using the apramycin resistance gene *aac(3')IV*, it was first necessary to remove the existing *aac(3')IV* gene from the integrated plasmid. The *attP* attachment site of pTESa-*bldA* is flanked by *loxP* recognition sites for Cre recombinase. This allows for removal of *aac(3')IV* from the chromosome after integration of a gene of interest (in this case *bldA*)[6]. To this end, *S. curacoi* / pTESa-*bldA* was transformed with the plasmid pALCRE encoding the Cre recombinase by intergeneric conjugation with *E. coli* ET12567/pUZ8002 cells[6]. The process of cultivating the exconjugants and inducing expression of the Cre recombinase was performed as described previously[6]. Exconjugants were selected on the basis of sensitivity to apramycin and hygromycin, indicating successful excision of *aac(3')IV* from the chromosome, and loss of the pALCRE plasmid. The final, marker free strain is denoted *S. curacoi-bldA*.

Single gene replacements of *czl*B1 and *czl*E within *S. curacoi-bldA* were performed through double cross-over utilizing the suicide plasmids pKGLP2 and pKCLP2 (Table S4). For *czl*B1, an apramycin resistance gene was amplified by PCR from the plasmid pLERE (Table S4) using the primers apra_czlB1_fw and apra_czlB1_rv (Table S5). These primers carry 5'-sequences of approximately 40 bp that are complimentary to the regions that flank *czlB1*. The PCR product was introduced by electroporation into *E. coli* BW25113 / pIJ790 cells containing the plasmid pBSK-*czl*, where λ-RED based recombination[5] replaced the *czlB1* sequence on pBSK-*czl* with the apramycin resistance gene, yielding pBSK-czl_*czl*B1::apra. PCR was subsequently used to amplify a *czl*B1::apra sequence from pBSK-czl_*czl*B1::apra using primers czlB1-fw and czlB1-rv (Table S5), yielding a PCR product containing the apramycin resistance gene flanked by ~2.5 kbp homology arms corresponding to the flanking regions of *czl*B1 within the *czl* BGC. The PCR product was then blunt-end ligated into the EcoRV site of the suicide vector pCKLP2. The resulting plasmid pCKLP2-*czl*B1::apra was introduced into *S. curacoi-bldA* via intergeneric conjugation from *E. coli* ET12567 / pUZ8002 cells. *S. curacoi-bldA* / Δ*czl*B1::apra exconjugants containing the double-crossover deletion of *czl*B1 were selected on MS-agar then TSB containing apramycin and fosfomycin. The mutation was confirmed by PCR analysis of gDNA isolated from the mutant strain.

The *czlE* gene was deleted in an analogous fashion. The primers apra_czlE_fw and apra_czlE_rv (Table S5) were used to generate a PCR product for replacement of *czl*E in pBSK-*czl* with an apramycin resistance gene via λ-RED based recombination.[5] The primers czlE-fw and czlE-rv were then used to PCR amplify a sequence from pBSK-czl_*czl*E::apra containing ~2.5 kbp homology arms, which was subsequently blunt-end

ligated into the EcoRV site of pKGLP2 (Table S4), yielding pKGLP2-*czl*E::apra. Transformation of the resulting plasmid into *S. curacoi-bldA* and selection for mutant exconjugants followed the same protocol as for *czl*B1 deletion.

Generation of point mutations in *czl*A and heterologous production of curacozole derivatives

Single point mutations in *czl*A were generated as follows. The plasmid pTESa-*czl* was digested with XbaI and NdeI and the resulting fragment containing *czl*A was ligated into the XbaI / NdeI sites of pBSK. The new plasmid pBSK-czlA has a size of approximately 7.3 kbp. Primer pairs that anneal back-to-back on pBSK-*czl*A, with one primer carrying the point mutation, were used to amplify the entire pBSK-*czl*A sequence to generate a PCR product carrying the desired mutation (Table S5). Accordingly, the primer pairs FW_fw1 and FW_rv1 generated *czl*A F18W; CzlA-F18Y-fw and CzlA-F18Y-rv generated *czl*A F18Y; coreI_V_fw_b and coreI_V_rv_b generated *czl*A I20V; coreIL2fw and coreIL2rv generated *czl*A_I20L; and czlAT_Sfw_JT and czlAT_Srv_JT generated czlA_T23S. Each PCR product was treated with DpnI to digest the methylated pBSK-*czl*A template. Subsequently, the PCR product was purified on an agarose gel, phosphorylated using T4 polynucleotide kinase, and then self-ligated using T4-ligase. The ligation mixture was transformed into *E. coli* DH10B cells, and single colonies were selected for plasmid purification. The desired mutation was confirmed by sequencing the PCR product that was generated from the mutant plasmid using the primers CzlA_check_fw and CzlA_check_rv.

Multiple mutations in *czl*A were generated in a sequential manner. To generate a PCR product carrying the czlA T23S_C24S double mutation, the pBSK-czlA_T23S plasmid was PCR amplified with the primers coreC_S_fw_a and czlAT_Srv_JT. This PCR

product was treated and cloned as described above, yielding pBSK-*czl*A_T23S_C24S. The mutant *czl*A* sequences were subsequently reintroduced into pTESa-*czl*. Following confirmation by sequencing, the mutant *czl*A* sequence was removed from pBSK by digestion with XbaI and NdeI, and the resulting fragment ligated into similarly digested pTESa-*czl*. The resulting pTESa-*czl** plasmid (where * denotes the *czl* BGC with a mutation in *czl*A) was confirmed once again by sequencing the PCR product generated with CzlA_check_fw and CzlA_check_rv.

For analysis of curacozole derivatives, each pTESa-*czl** plasmid specifying a CzlA variant was introduced into *S. coelicolor* CH999 or *S. lividans* ΔYA9 via intergeneric conjugation. *S. coelicolor* CH999 was used to express pTESa-*czl** plasmids specifying CzlA F18Y, I20V, T23S, and T23S_C24S variants as precursors to compounds **8**, **10**, **2**, and **11,** respectively. *S. lividans* ΔYA9 was used to express pTESa-*czl** plasmids specifying CzlA F18W and I20L variants as precursors to compounds **7** and **9**, respectively. A PCR reaction was performed on the resulting exconjugants using primers CzlA_check_fw and CzlA_check_rv, and the resulting PCR product confirmed by sequencing. The strains confirmed to carry pTESa-*czl** were cultivated and analyzed as described in **Cultivation of Streptomyces strains for analysis of curacozole and derivatives**, with the addition of a solid phase extraction step as follows. Following extraction of the culture with ethyl acetate and removing the solvent in vacuo, the sample was redissolved in 5 mL MeOH and then mixed with 5 mL water. The mixture was applied to an Oasis HLB 35cc LP Vac Cartridge (6 g, Waters Ltd.) which was then developed with a stepwise gradient of increasing MeOH in water, beginning with 50% MeOH:water and

increasing to 100% MeOH. The individual fractions were collected and analyzed as described in **LC-MS Analysis of curacozole and derivatives**.

DNA transformation into *E. coli* BW25113 / pIJ790 cells by electroporation

E. coli BW25113 / pIJ790 cells were grown overnight in LB medium at 37 °C in a shaker incubator. This pre-culture was used to inoculate (1% v/v) fresh LB medium and grown until the cell density reached an OD_{600} of 0.4 to 0.6. The cells were then placed on ice for approximately 10 min to stop growth. Subsequently, the cells were centrifuged, washed 2-3 times with ice-cold glycerol (10%), and centrifuged again. The pellet was resuspended in approximately 80 µL 10% glycerol, mixed with 3 µL gel purified PCR product and transferred to an ice-cold electroporation cuvette. Electroporation was performed at 1.8 kV, 25 µF, and 200Ω. The cells were transferred to 1 mL of fresh LB medium, incubated in a shaker incubator at 37°C for one hour, and then spread on an LB-agar plate containing ampicillin for pBSK-*czI*, or apramycin for pTESa-*czI*. This was then incubated at 37°C overnight.

Plasmid transformation into Streptomyces by intergeneric conjugation

The desired plasmid was introduced into *E. coli* ET12567 / pUZ8002 via heat-shock or electroporation. The cells were then cultured overnight in LB media supplemented with apramycin and kanamycin. 1% (v/v) of this culture was transferred to fresh medium and grown until the cell density reached an OD = 0.6. The cells were then centrifuged, washed approximately 2 times with LB medium, centrifuged again, and resuspended in approximately 500 µL of TSB. *Streptomyces* strains were prepared for conjugation either as mycelium or spores. The mycelial culture was also centrifuged, washed 2 times with

TSB medium, centrifuged again, and resuspended in approximately 500 µL of TSB medium. Subsequently, the *E. coli* ET12567 / pUZ8002 and *Streptomyces* cells were mixed and spread on a MS-agar plate containing 10 - 50 mM $CaCl_2$. Dilutions were performed if the cell concentration was too high. Spores, on the other hand, were first added to approximately 500 µL of TSB and reactivated at 50°C for 10 min. After cooling, the spores were mixed with *E. coli* ET12567 / pUZ8002 cells and added to an MS agar plate without antibiotics for 6 – 8 hours at 28°C to ensure DNA transmission. The plate was then flooded with a solution of the appropriate antibiotic (e.g. apramycin for pTESa) and fosfomycin (to eliminate the *E. coli* ET12567 / pUZ8002 cells) and allowed to dry. The plate was then incubated for 5 – 7 days at 28°C. Colonies corresponding to exconjugants were picked and regrown on MS-agar plates supplemented with the appropriate antibiotic.

Cultivation of *Streptomyces* strains for analysis of curacozole and derivatives

A cell freezer stock was used to inoculate 50 mL TSB supplemented with the appropriate antibiotic in a 300 mL baffled flask (containing a non-corroding spring), which was then incubated overnight at 28°C and 180 rpm. This pre-culture was used to inoculate (2% v/v) 200 mL production medium without any antibiotics in a baffled 1L flask (containing a non-corroding spring), which was then incubated for 5 days at 28°C, 180 rpm. The cells were removed by centrifugation. To the supernatant, an equal volume of ethyl acetate was added then mixed for 30 min. The organic phase was separated, and the solvent removed in vacuo. The resulting extract was redissolved in methanol for LC-MS analysis as described in **LC-MS Analysis of curacozole and derivatives**.

Cloning of *czl* specified proteins

CzlA. The *czlA* sequence was PCR amplified from *S. curacoi* gDNA using primers 5'-ggtgccgcgcggcagccaAATGTTCGAGAACACGACCG-3' and 5'-caccagtcatgctagccaTCAGACGGCCTCGGTCTC-3'. PCR was performed with Herculase II Fusion DNA polymerase (Agilent Technologies) in a reaction containing 6% DMSO, 0.25 μM dNTPs, 0.25 μM each primer, and Herculase II reaction buffer. PCR began with denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation (95 °C, 20s), annealing (54-62 °C, 20 s) and extension (72 °C, 30 s). A final extension was held at 72 °C for 3 min. The resulting PCR product was inserted into the NdeI site of pET28a by Gibson assembly. The resulting pET28a-*czlA* encodes CzlA with an N-terminal His_6 tag. The plasmid was verified by DNA sequencing.

CzlA Variants. The *E. coli* codon optimized genes encoding CzlA variants C24S, C25S, C24S/C25S, and F18G were synthesized and cloned into the NcoI and EcoRI sites of pET28a by Bio Basic (Canada). These sequences specify variants with N-terminal His_6 tags. The expressed sequences of CzlA and variants are listed in Table S7.

CzlD. The codon optimized *czlD* gene was synthesized by Bio Basic (Canada) and cloned into pET100 / D-TOPO using the Champion™ pET Directional TOPO® Expression Kit (Invitrogen). The sequence specifies CzlD with a N-terminal His_6 tag.

CzlBC. The codon optimized *czlBC* gene was synthesized and cloned into the NdeI and XhoI sites of pET21a by Bio Basic (Canada). The sequence specifies CzlBC with a C-terminal His_6 tag.

CzlI. The *czlI* sequence was PCR amplified from *S. curacoi* genomic DNA using primers 5'-gacagcaaatgggtcgcgATATGGCCGAGGAGGAAGCA-3' and 5'-cgacggagctcgaattcgTACAACGACCGCTGACATCG-3'. PCR was performed with Herculase II Fusion DNA polymerase (Agilent Technologies) in a reaction containing 6% DMSO, 0.25 μM dNTPs, 0.25 μM each primer, and Herculase II reaction buffer. PCR began with denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation (95 °C, 20s), annealing (54-62 °C, 20 s) and extension (72 °C, 30 s). Final extension was held at 72 °C for 3 min. The resulting PCR product was cloned into pET28a linearized with BamHI using Gibson assembly. The resulting plasmid pET28a-*czlI* specifies CzlI with with N- and C-teminal His_6 tags. The plasmid was verified by DNA sequencing.

Expression and purification of czl specified proteins.

Plasmids were transformed into the *E. coli* BL21 (DE3) heat shock competent cells and plated on LB agar with the appropriate antibiotic (50 μg/mL kanamycin for *czl* sequences in pET28a; 100 μg/mL ampicillin for *czl* sequences in pET100 and pET21a). The plates were grown overnight at 37 °C. A single colony was used to inoculate 25 mL LB broth with the appropriate antibiotic (either 50 μg/mL kanamycin or 100 μg/mL ampicillin) in a 125 mL flask. This starter culture was grown overnight at 37 °C, 200 rpm. Next, 1 L LB broth containing the appropriate antibiotic was inoculated with 20 mL (2% vol/vol) from the starter culture. The culture was grown at 37 °C, 180 rpm until the cell density reached OD_{600} ~0.6, then incubated on ice for 30 min. IPTG was added to a final concentration of 1 mM. Cultures expressing CzlA and variants were incubated at 30 °C and all others at 20 °C, for 20 hours, shaking at 200 rpm. The cultures were centrifuged at 5000 x*g* for 30

min at 4 °C, and the supernatant was discarded. The cell pellet was resuspended in ice cold Buffer A (Table S6) then lysed by 3 passes through an EmulsiFlex-C5 homogenizer (Avestin, Ottawa) at 15,000 psi. The lysate was cleared by centrifugation at 28 000 x*g* for 30 min at 4 °C using a Thermo Fisher Scientific Sorvall Lynx 6000 Superspeed Centrifuge. Ni-NTA Superflow (Qiagen) resin (5.0 mL) was packed into an XK-16 column (Cytiva) equilibrated with Buffer A for 10 column volumes (CV) using an AKTA FPLC system. The clarified lysate was filtered through a 0.45 µm polyethersulfone membrane and loaded onto the column. The column was developed at a flow rate of 5.0 mL/min as follows: 100% Buffer A for 5 CV; 4% Buffer B for 5 CV; a linear gradient to 100% Buffer B over 10 CV; 100% B for 2 CV. The eluted fractions were analyzed by SDS-PAGE Fractions containing pure protein were pooled for exchange into Buffer C by ultrafiltration using centrifugal filter units.

AlphaFold predictions and structural analysis

Native protein sequences were submitted to ColabFold: AlphaFold2 using MMseqs2 for structural predictions.[10] The top ranked models were selected for heuristic PDB searches using the DALI server.[11] To model CzlA-enzyme complexes, native protein sequences were submitted with specified inter-protein chain breaks, to ColabFold AlphaFold2 using MMseqs2 for structural predictions. Structural images were prepared in PyMOL.[12] Electrostatic surface potentials were calculated by submission of the calculated structures to the Adaptive Poisson-Boltzmann Solver (APBS) through the PyMOL plugin.[13] Pairwise structural alignments were performed using the jFATCAT algorithm[8] through the PDB server (https://www.rcsb.org/alignment). The corresponding sequence based amino acid sequence alignments were annotated using ESPript 3.0.[9]

CzlD and CzlBC Activity Assays

Reaction mixtures consisted of 5 uM CzlD, 5 uM CzlBC, 50 uM CzlA (or variants), 5 mM $MgCl_2$, 5 mM ATP, 50 uM FMN, and 5 mM TCEP (except with CzlA C24S/C25S which lacks Cys residues) in 25 mM Tris-HCl and 10 mM KCl (pH 7.5). The reactions were incubated at 25 °C for 20 hours prior to analysis. UPLC-MS analysis of modified CzlA peptides was performed as described below (see **UPLC-MS Analysis of CzlA**). The modified peptide was sequenced by MS/MS as described below (see **LC-MS/MS Peptide Sequencing of CzlA**).

CzlI Activity Assay

Assays were performed with 5 μM CzlI, 50 uM CzlA, 1 mM 2-oxoglutarate, 100 uM ferrous ammonium sulphate, and 200 uM ascorbate in 25 mM Tris-HCl and 10 mM KCl (pH 7.5). Reactions were incubated at 25 °C for 20 hours prior to analysis. Conversion of CzlA to the hydroxylated product was monitored by UPLC-MS (see **UPLC-MS Analysis of CzlA**). The modified peptide was sequenced by MS/MS as described below (see **LC-MS/MS Peptide Sequencing of CzlA**). Analysis of individual amino acids was achieved by heating 50 µL of the assay mixture at 60 °C for 10 min to precipitate enzyme and centrifuging at 16,000 rcf for 10 min. The supernatant was removed and loaded into a 1 mL vacuum hydrolysis tube (ThermoFisher Scientific, Catalog Number 29570) with 450 µL of 6N HCl. The tube was flushed with N_2 gas, placed under vacuum, sealed, then heated in an oven at 105 °C for 20 hours. The sample was allowed to cool at room temperature, dried by rotary evaporator, resuspended in 2 mL H_2O, then dried again by lyophilization. After drying, sample was dissolved in 50 µL H_2O then mixed with 10 µL of

10 µL of 1 M boric acid buffer (pH 8.5) and 10 mM dansyl chloride (in acetonitrile). Amino acid standards were derivatized following the same procedure. UPLC-MS analysis of derivatized amino acids was performed as described below (see **UPLC-MS Analysis of Dansylated Amino Acids Derived from CzlA**).

LC-MS analysis of curacozole and derivatives

Methanol extracts containing curacozole or derivatives were resolved on a Thermo Fisher LC-MS UltiMate 3000 instrument equipped with a Waters XBridge C18 column (100 mm x 4.6 mm, 3.5 µM). The column was maintained at 30°C and developed at a flow rate of 0.5 mL min^{-1} using a gradient comprised of 0.5% acetic acid in H_2O (solvent A) and 0.5% acetic acid in acetonitrile (solvent B). The gradient was as follows: 0 to 0.5 min: 95% A, 5% B; 0.5 to 18.5 min: linear gradient to 5% A, 95% B; 18.5 to 20.5 min: hold at 5% A, 95% B; 20.5 to 20.8 min: linear gradient to 95% A, 5% B; 20.8 to 25 min: hold at 95% A, 5% B. The MS settings in ESI positive ion mode were as follows: spray voltage, 2500 V; vaporizer temp., 450°C; sheath gas pressure, 30 arbitrary units; auxiliary gas pressure, 5 au; capillary temperature, 320°C.

UPLC-MS Analysis of CzlA

Reactions involving precursor peptide CzlA were treated with 1:1 v/v acetonitrile to precipitate enzymes then centrifuged at 16,000 rcf for 5 min. The supernatant was collected for CzlA analysis on a Waters Acquity™ UPLC-MS instrument using a Waters CORTECS T3 column (1.6 µM, 2.1mm X 100 mm) maintained at 40 °C. The column was developed with water (A), acetonitrile (B) and 1% formic acid in water (C) as follows: 10%

C maintained throughout the method; 5% B held for 2 min, followed by a linear gradient to 20% B over 1 min, then a linear gradient to 80% B over 7 min, then to 90% B over 1 min. The MS scan covered 50-2000 m/z with cone voltage of 35 V in ES+ ionization mode with source temperature of 150°C and desolvation temperature of 400 °C. Selective ion recording was used to detect dansylated amino acids.

LC-MS/MS Peptide Sequencing of CzlA

Prior to MS/MS, assay samples were heated at 60 °C for 5 min to precipitate enzymes. The supernatant was collected for proteolysis by GluC at 37 °C for 16 hours. After completion, proteolyzed assay samples were mixed 1:1 v/v with acetonitrile and centrifuged at 16,000 rcf for 5 min. The supernatant was collected for MS/MS analysis on a Thermo Scientific Orbitrap Velos Pro instrument coupled with an Agilent 1100 HPLC. The samples were resolved using a CORTECS C18 column (5 cm, 1.6 µm) maintained at 40 °C and a solvent system comprised of water with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B). Samples were eluted from the column with 90 % B over 10 minutes at a flow rate of 0.3 mL/min. MS was equipped with heated electrospray ion source running in positive ion mode with the following ion source parameters: high voltage 3.5 kV, evaporation temperature 350 °C, sheath gas 30 arbitrary units (arb), aux gas 10 arb, and ion transfer tube temperature 300 °C. The MS/MS method was programmed for each run to select the ion of interest, then fragmentation was performed using the ion trap with collision-induced dissociation set to 35 V.

UPLC-MS Analysis of Dansylated Amino Acids Derived from CzlA

A Waters Acquity™ Ultra High-Performance Liquid Chromatography Mass Spectrometry instrument was used to analyze the dansylated amino acids using a CORTECS T3 column (1.6 µM, 2.1mm X 100 mm) maintained at 30 °C. The column was developed with water (A), acetonitrile (B) and 1% formic acid in water (C) as follows: 10% C maintained throughout the method; 30 % B held for 1 min, followed by a linear gradient to 85% B over 25 min. MS scans were collected from 85 to 500 m/z with a cone voltage of 15 V in ES+ ionization mode, a source temperature of 150 °C, and a desolvation temperature of 400 °C. Selective ion recordings were used as indicated.

References

1 A. Gessner, T. Heitzler, S. Zhang, C. Klaus, R. Murillo, H. Zhao, S. Vanner, D. L. Zechel and A. Bechthold, *ChemBioChem*, 2015, **16**, 2244–2252.

2 R. McDaniel, S. Ebert-Khosla, D. A. Hopwood and C. Khosla, *Science*, 1993, **262**, 1546–1550.

3 Y. Ahmed, Y. Rebets, M. R. Estévez, J. Zapp, M. Myronovskyi and A. Luzhetskyy, *Microb. Cell Fact.*, 2020, **19**, 5.

4 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, United Kingdom, 2000.

5 B. Gust, G. L. Challis, K. Fowler, T. Kieser and K. F. Chater, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 1541–1546.

6 S. Herrmann, T. Siegl, M. Luzhetska, L. Petzke, C. Jilg, E. Welle, A. Erb, P. F. Leadlay, A. Bechthold and A. Luzhetskyy, *Appl. Environ. Microbiol.*, 2012, **78**, 1804–1812.

7 M. Myronovskyi, E. Welle, V. Fedorenko and A. Luzhetskyy, *Appl. Environ. Microbiol.*, 2011, **77**, 5370–5383.

8 Z. Li, L. Jaroszewski, M. Iyer, M. Sedova and A. Godzik, *Nucleic Acids Res.*, 2020, **48**, W60–W64.

9 X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, W320–W324.

10 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat. Methods*, 2022, **19**, 679–682.

11 L. Holm, *Nucleic Acids Res.*, 2022, **50**, W210–W215.

12 The PyMOL Molecular Graphics System (version 1.2r3pre) Schrödinger, LLC.

13 E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon and N. A. Baker, *Protein Science*, 2018, **27**, 112–128.