Supplementary Information

Flavoenzyme CrmK-Mediated Substrate Recycling in Caerulomycin Biosynthesis†

Yiguang Zhu,‡[a] Marie-Ève Picard,‡[b] Qingbo Zhang,[a] Julie Barma,[b] Xavier Murphy Després,[b] Xiangui Mei,[c] Liping Zhang,[a] Jean-Baptiste Duvignaud,[b] Manon Couture,[b] Weiming Zhu,[c] Rong Shi,[b]* and Changsheng Zhang[a]*

[a]CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Key Laboratory of Marine Materia Medica, RNAM Center for Marine Microbiology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, 164 West Xingang Road, Guangzhou 510301, P.R. China;

[b]Département de biochimie, de microbiologie et de bio-informatique, PROTEO, and Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, G1V 0A6, Canada;

[c]Key Laboratory of Marine Drugs, Chinese Ministry of Education, School of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China.

‡These authors contributed equally.

Equal Correspondence to C. Z. (czhang2006@gmail.com) for biochemistry or R. S. (rong.shi@bcm.ulaval.ca) for structure.

Content

Supplemental Materials and Methods

Bacterial strains, plasmids, and reagents. Bacterial strains and plasmids used and constructed in this study are listed in Table S1. Chemicals, enzymes, and other molecular biological reagents were purchased from standard commercial sources and used according to the manufacturers' recommendations.

General HPLC Analysis. HPLC analysis for metabolites and enzyme assays was carried out on a reversed phase column Luna C18 (Phenomenex, 150 × 4.6 mm, 5 μm) with UV detection at 313 nm under the following program: solvent system (solvent A, 0.15% TFA in water; solvent B, 100% CH_3CN); 5% B to 30% B (0 - 18 min), 30% B to 100% B (18 - 19 min), 100% B (19 - 24 min), 100% B to 5% B (24 - 25 min), 5% B (25 - 30 min); flow rate at 1 mL min^{-1}.

Cloning, Overexpression and Purification of CrmK. The *crmK* gene was PCR amplified from the genomic DNA of *A. cyanogriseus* WH1-2216-6 using primers CrmK-EF and CrmK-ER (Table S2). PCR products were digested with *Eco*RI/*Xho*I and subsequently inserted into pET28a linearized with *Eco*RI/*Xho*I, to yield the plasmid pCSG2218 after sequence confirmation. The plasmid pCSG2218 were introduced to *E. coli* BL21(DE3) to produce the *N*-$(His)_6$-tagged CrmK proteins. An overnight culture of transformed E. coli BL21(DE3)/pCSG2218 was used to inoculate 1 L TB medium and was grown at 37°C until the absorbance at 600 nm reached 0.6. Protein expression was induced with 100 μM isopropyl 1-thio-β-D-galactopyranoside (IPTG) followed by incubation at 18°C for 20 h. Cells were harvested and resuspended in lysis buffer (50 mM Tris-HCl, 200 mM NaCl, 5% glycerol, and 1 mM DTT, pH 8.0). An additional 5 mM imidazole was added to the lysis buffer and the overexpressed CrmK was purified by standard affinity chromatography using Ni^{2+}-NTA affinity resin (Qiagen). The CrmK protein was eluted with buffer containing 50 mM Tris-HCl (pH 8), 200 mM NaCl, 5% glycerol, 250 mM imidazole, and 1 mM DTT. Following the desalting step to remove the imidazole, CrmK was further purified by size-exclusion chromatography (SEC) on

a Superdex 200 column (GE Healthcare) equilibrated in a buffer containing 20 mM Tris-HCl pH 8, 150 mM NaCl, 2% (v/v) glycerol, and 5mM DTT. The eluted fractions were pulled, concentrated for enzyme assays and crystallization.

Site-directed Mutagenesis of *crmK* and Purification of CrmK Mutants. Site-directed mutagenesis of *crmK* was carried out according to manufacturer's instructions (TransGen). Thirteen single CrmK mutants (H64A, F123A, C124A, Y138F, I328A R340A, E376A, I378A, F403A, F403E, M405A, Y446F, Y449F) and three double mutants (H64A/C124A, Y449F/Y138F, Y449F/F403E) were constructed and were confirmed by sequencing. The plasmids carrying the *crmK* mutated genes were introduced to *E. coli* BL21(DE3) for the purification of CrmK mutants according to the aforementioned methods.

Biochemical assays of CrmK and Mutants. The concentration of purified proteins was determined by the Bradford method.[1] The CrmK assays for CRM P **7** were conducted in 50 μL reaction mixture in 50 mM Tris-HCl buffer (pH 8.0) containing 200 μM CRM P **7**, 240 nM CrmK at 28 °C, The CrmK assays for CRM F **8** were conducted in 50 μL reaction mixture in 50 mM Tris-HCl buffer (pH 8.0) containing 200 μM CRM F **8**, 2.4 μM CrmK at 28°C. For determining kinetic parameters, CRM P **7** was set as a variable substrate in concentrations of 5, 7.5, 10, 15, 20, 25, 50, 100, and 200 μM. Enzyme assay was performed in Tris-Cl buffer (50 mM, pH 8.0) containing 12 nM CrmK at 28°C for 5 min in triplicates. Similarly, CRM M **4** was set as a variable substrate in concentrations of 500, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 μM for the determination of kinetic parameters. K_m and V_{max} were calculated by nonlinear regression analysis using Origin 9.0 software. The relative enzymatic activities of CrmK mutants were obtained by comparison of initial velocities with that of the wild type CrmK. To determine the initial velocity on CRM P **7**, the assays were performed in Tris-HCl buffer (pH 8.0) containing 12 nM CrmK (or various mutants), 50 μM CRM P **7** at 28°C for 5 min in triplicates. To determine the initial velocity on CRM M **4**, the

assays were performed in 50 mM Tris-HCl buffer (pH 8.0) containing 240 nM CrmK (or various mutants), 5 mM CRM M **4** at 28°C for 5 min in triplicates. CrmK and mutants reactions were quenched by the addition of 50 µL MeOH, and denatured proteins removed by centrifugation. The assays were monitored by HPLC analysis.

Enzymatic Preparation and Isolation of CRM R 10. A CrmK reaction was conducted in 100 mL mixture consisting of 50 mM Tris-HCl (pH 8.0), 8 mg CRM F **8**, and 2.4 µM CrmK upon incubation at 28 °C for 48 h. MeOH extracts of the reaction mixture were subjected to purification by semi-preparative HPLC on a Varian Star Workstation, using a reversed phase column Luna 5µ Phenyl-Hexyl column (solvent A, 0.15% TFA in water; solvent B, 100% CH_3CN; eluted with constant 20% solvent B; flow rate at 2.5 mL min^{-1}) to obtain compound **10** (5 mg).

Crystallization, Data Collection and Refinement. Crystals of CrmK were obtained using the Classics II screen (Qiagen) through microbatch crystallization by mixing 0.7 µL of protein (12 mg mL^{-1}) with 0.5 µL reservoir solution (0.96 M Sodium citrate pH 7.0). The crystals of CrmK bound with its substrate (CRM P **7**) were obtained by soaking the CrmK crystals (~3 min) in the reservoir solution containing 5 mM CRM P **7**, which was initially dissolved in DMSO at a stock concentration of 100 mM. For data collection, the crystals were flash cooled in the N2 cold stream (Oxford Cryosystem, Oxford, UK) using 1.4 M Sodium Citrate pH 7.0 as cryoprotectant. Data for both CrmK-FAD and CrmK-FAD-CRM P complexes were collected to 1.84 Å and 2.15 Å, respectively, at a wavelength of 0.9795 Å at the CMCF1 beamline, Canadian Light Source. Both crystals belong to the space group P1 and are isomorphous with unit cell a=63.6, b=95.7, c=98.4 Å, α=95.2°, β=97.0°, γ=104.4°. Data processing and scaling were performed with iMosflm.[2] The structure solution for the CrmK-FAD complex was obtained by molecular replacement through MolRep[3] using the structure of TamL[4] (PDB code 2Y08) as the search model. Several cycles of refinement using REFMAC5[5] followed by

model rebuilding with Coot[6] were carried out. For the substrate-bound CrmK, guided by the Fourier difference map, the substrate molecule was placed in the model after the rebuilding of the protein part was finished. Except the first 2-3 residues at the N-terminal, all the residues in the sequence are clearly visible in the electron density maps. Both models have good stereochemistry as analyzed with PROCHECK.[7] Data collection and refinement statistics for both structures were shown in Table 1.

Table S1 Strains and plasmids used and generated in this study.

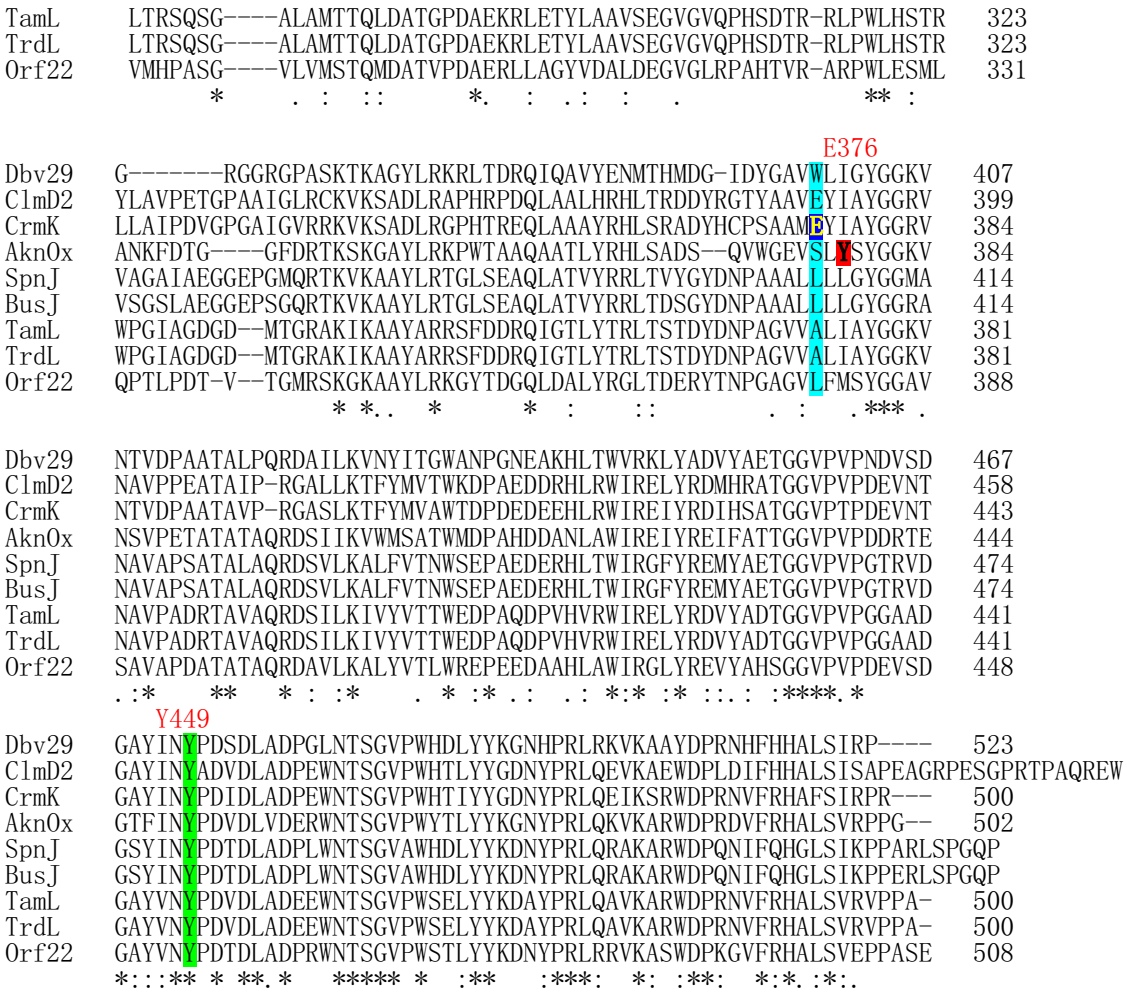
Strains/Plasmids	Characteristic(s)	Sources
Strains		
Escherichia coli		
DH5α	Host strain for cloning	Invitrogen
BL21(DE3)	Host strain for protein expression	Novagen
BW25113	Host strain for PCR targeting	[8]
ET12567	Donor strain for conjugation	[9]
Actinoalloteichus		
WH1-2216-6	Wild type, caerulomycin A producer	[10]
CRM01	The *crmG* gene disrupted mutant of WH1-2216-6	[11]
CRM11	The *crmK* gene disrupted mutant of WH1-2216-6	[12]
Plasmids		
pET28a	Km^r, expression vector	Stratagene
pIJ773	Aprr, source of *aac(3)IV*	[13]
pUZ8002	Kmr, including *tra* for conjugation	[14]
pCSG2016	Strain WH1-2216-6 genomic library cosmid	[12]
pCSG2102	pCSG2016 derivative where *crmG* was disrupted by *aac(3)IV*	This study
pCSG2201	1.58kb *crmG Nde*I/*Eco*RI PCR fragment from genomic DNA of strain WH1-2216-6 into pET28a	This study
pCSG2218	1.5kb *crmK Eco*RI /*Xho*I PCR fragment from genomic DNA of strain WH1-2216-6 into pET28a	This study
pCSG2229	pCSG2218 containing site-directed mutagenesis at H64A	This study
pCSG2230	pCSG2218 containing site-directed mutagenesis at C124A	This study
pCSG2231	pCSG2218 containing site-directed mutagenesis at F123A	This study
pCSG2236	pCSG2218 containing site-directed mutagenesis at Y138F	This study
pCSG2237	pCSG2218 containing site-directed mutagenesis at I328A	This study
pCSG2238	pCSG2218 containing site-directed mutagenesis at R340A	This study
pCSG2240	pCSG2218 containing site-directed mutagenesis at E376A	This study
pCSG2241	pCSG2218 containing site-directed mutagenesis at I378A	This study
pCSG2243	pCSG2218 containing site-directed mutagenesis at F403A	This study
pCSG2244	pCSG2218 containing site-directed mutagenesis at F403E	This study
pCSG2245	pCSG2218 containing site-directed mutagenesis at M405A	This study
pCSG2246	pCSG2218 containing site-directed mutagenesis at Y446F	This study
pCSG2247	pCSG2218 containing site-directed mutagenesis at Y449F	This study
pCSG2248	pCSG2218 containing site-directed mutagenesis at H64A/C124A	This study
pCSG2250	pCSG2218 containing site-directed mutagenesis at Y449F/Y138F	This study
pCSG2252	pCSG2218 containing site-directed mutagenesis at Y449F/F403E	This study

Table S2 Primers used in this study.

Primers	Sequences	Usage
For gene expression		
CrmK-EF	5' - GTGC**GAATTC**ATGCCAACCCGAGCG - 3'	*crmK* expression
CrmK-ER	5' - TCGT**CTCGAG**TTCACCGAGGACGGAT - 3'	
For *crmK* mutagenesis		
H64AF	5' - CCGTGCGCTCCGGTGGG**GCC**TGCGGGGAGGCGTT - 3'	*crmK* H64A mutagenesis
H64AR	5' - GCCCCACCGGAGCGCACGGTGAGGCGCTTGCC - 3'	
F123AF	5' - ACGTTCCCCGGCGGT**GCC**TGCATGGGGGTCG - 3'	*crmK* F123A mutagenesis
F123AR	5' - GCACCGCCGGGGAACGTCACCCCGTAGTT - 3'	
C124AF	5' - CG TTCCCCGGCGGTTTC**GCC**ATGGGGGTCGGG - 3'	*crmK* C124A mutagenesis
C124AR	5' - GCGAAACCGCCGGGGAACGTCACCCCGTAGTT - 3'	
Y138FF	5' - AC ATCTCCGGTGGGGGC**TTC**GGGCCGCTCTCG - 3'	*crmK* Y138F mutagenesis
Y138FR	5' - AA GCCCCCACCGGAGAT GTGGCCGCCCGCCCC - 3'	
I328AF	5' - CGACCAGCCAGCTGTTGGCG**GCC**CCCGACGTGGGG - 3'	*crmK* I328A mutagenesis
I328AR	5' - GCCGCCAACAGCTGGCTGGTCGCCAACCACGGCAAC - 3'	
R340AF	5' - CGGGTGCGATCGGGGTGCGG**GCG**AAGGTCAAGTCC - 3'	*crmK* R340A mutagenesis
R340AR	5' - GCCCGCACCCCGATCGCACCCGGCCCCACGTCGG - 3'	
E376AF	5' - ACTGCCCCAGCGCCGCGATG**GCG**TACATCGCCTAC - 3'	*crmK* E376A mutagenesis
E376AR	5' - GCCATCGCGGCGCTGGGGCAGTGGTAGTCGGCCC - 3'	
I378AF	5' - GCGCCGCGATGGAGTAC**GCC**GCCTACGGCGGG - 3'	*crmK* I378A mutagenesis
I378AR	5' - GCGTACTCCATCGCGGCGCTGGGGCAGTGGTA - 3'	
F403AF	5' - GCGGGGCGTCGTTGAAGACC**GCC**TACATGGTGGC - 3'G	*crmK* F403A mutagenesis
F403AR	5' - GCGGTCTTCAACGACGCCCCGC GGGGAACCGCCGT - 3'	
F403EF	5' - GCGGGGCGTCGTTGAAGACC**GAG**TACATGGTGGCG - 3'	*crmK* F403E mutagenesis
F403ER	5' - CTCGGTCTTCAACGACGCCCCGCGGGGAACCGCCGT - 3'	
M405AF	5' - CGTCGTTGAAGACCTTCTAC**GCG**GTGGCGTGGACC - 3'	*crmK* M405A mutagenesis
M405AR	5' - GCGTAGAAGGTCTTCAACGACGCCCCGCGGGGAA - 3'	
Y446FF	5' - AGGTCAACACGGGCGCC**TAC**ATCAACTACCCC - 3'	*crmK* Y446F mutagenesis
Y446FR	5' - AAGGCGCCCGTGTTGACCTCGTCGGGGGTGGG - 3'	
Y449FF	5' - ACGGGCGCCTACATCAAC**TTC**CCCGACATCGAC - 3'	*crmK* Y449F mutagenesis
Y449FR	5' - AAGTTGATGTAGGCGCCCGTGTTGACCTCGTC - 3'	

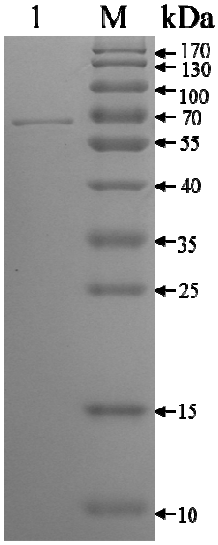
Fig. S1 Alignment of CrmK with bi-covalent flavinproteins from diverse pathways.

```
Dbv29     MTGGTG---ADAASAGASSTRPELRGERCLPPAGPVKVTPDDPRYLNLKLRGANSRFNGE   57
ClmD2     ----------MKRDDQDAEEESGPGSPTPPSSRVTPDDIRYENLRR-SYNYRFIAK   45
CrmK      ----------------------------MPTRAAVTVKPDDHRYDLLA-RADNYRFVAQ   30
AknOx     ------------------------MGGACGARTALVKVDRVDRRYQDLVTRGFNGRFRGR   36
SpnJ      MISAAGEQSGPVRKGGAVPEFHDPAPMNRRTPGTEITVEPDDPRYPDLV-VGHNPRFTGK   59
BusJ      MISAEGEQSGPVSKGGAVPDFHDPATMNRRTPGTEITVEPGDPRYPDLV-VGHNPRFTGK   59
TamL      ----------------------------MKHIDSVAPGDIRYEDLR-RGENLRFVGD   28
TrdL      ----------------------------MKHIDSVAPGDIRYEDLR-RGENLRFVGD   28
Orf22     ---------------------MTDGARHRPPLGPVRVTPADPRYRDLI-RGVNHRFTGS   37
                                           *   * **  *   . * ** .

                                          H64
Dbv29     PDYIHLVGSTQQVADAVEETVRTGKRVAVRSGGHCFEDFVDNPDVKVIIDMSLLTEIAYD   117
ClmD2     PDYFRLVHSPRQVEEAVREAVTAGKRITVRSRGHCGEDFVAAPDVEVILDLSPMSRVDYD   105
CrmK      PEYFRLPYSTAQVVEAVSEAVAAGKRLTVRSGGHCGEAFVASPDVDVIVDLSSMSHVGYD   90
AknOx     PDVVYVVHTADQVVDAVNQAMAAGQRIAVRSGGHCFEGFVDDPAVRAVIDMSQMRQVFYD   96
SpnJ      PERIHIASSAEDVVHAVADAVRTGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYYD   119
BusJ      PERIHIAGSTEDVVHAVAEAVRTGRRVGVRSGGHCFENLVADPAIRVLVDLSELNRVYFD   119
TamL      PEEIHLVGSAAEIEQVLSRAVRSGKRVAVRSGGHCYEDFVANSDVRVVMDMSRLSAVGFD   88
TrdL      PEEIHLVGSAAEIEQVLSRAVRSGKRVAVRSGGHCYEDFVANSDVRVVMDMSRLSAVGFD   88
Orf22     PGQIRIVGSAEQTVEAVQDAVDAGLRLAVRSGGHCLEDLVDGPDTAFLLDMSEMRQVYYD   97
          *  .  :  :: :  ..:  :: :* *: *** *** * :*       ::*:* :  : :*

                                          C124          Y138
Dbv29     PSMNAFLIEPGNTLSEVYEKLYLGWNVTIPGGVCGGVGVGGHICGGGYGPLSRQFGSVVD   177
ClmD2     RERNAFVIEAGAPVGKMLHTLFHNWGVTVPAGFCMGVGAGGHISGGGYGPLSRLLGLSVD   165
CrmK      EERGAFEVEAGATVGQIYRVLYKNYGVTFPGGFCMGVGAGGHISGGGYGPLSRLLGLTVD   150
AknOx     SGKRAFAVEPGATLGETYRALYLDWGVTIPAGVCPQVGVGGHVLGGGYGPLSRRDGVVAD   156
SpnJ      STRGAFAIEAGAALGQVYRTLFKNWGVTIPTGACPGVGAGGHILGGGYGPLSRRFGSVVD   179
BusJ      STRGAFAIEAGAALGQVYRTLFKNWGVTIPTGACPGVGAGGHIPGGGYGPLSRRFGSVVD   179
TamL      EERGAFAVEAGATLGAVYKTLFRVWGVTLPGGACPDVGAGGHILGGGYGPLSRMHGSIVD   148
TrdL      EERGAFAVEAGATLGAVYKTLFRVWGVTLPGGACPDVGAGGHILGGGYGPLSRMHGSIVD   148
Orf22     PGMRAFAVEPGAHLGDVYRTLYKGWGVTVPGGSCPTVAAGGHFAGGGYGPLTRLHGCVVD   157
              ** :* *  :.  *. *:  : **.* * *  *..***. *******:*  *  .*

Dbv29     YLYAVEVVVVNKQGKARVIVATRERDDPHHDLWWAHTGGGGGNFGVVTKYWMRVPEDV-G   236
ClmD2     HLYAVEVVVVDQDRNVSTVVATREKTDPNRDLWWAHTGGGGGNFGVITRYWMRSPEAS-G   224
CrmK      YLHAVEVVVVDAEGVVSTVVATREEDDPNRDLWWAHTGGGGGNFGVITRYWLRSPDAV-G   209
AknOx     HLYAVEVVVVDASGRARKVVATSAADDPNRELWWAHTGGGGGNFGIVTRYWFRTPGAT-G   215
SpnJ      YLQGVEVVVVDQAGEVHIVEADRNSTGAGHDLWWAHTGGGGGNFGIVTRFWLRTPDVV-S   238
BusJ      YLQGVEVVVVDRAGEVHIVEVDRNSIGAGHDLWWAHTGGGGGNFGVVTRFWLRAPDVV-S   238
TamL      YLHAVEVVVVDASGDARTVIATREPSDPNHDLWWAHTGGGGGNFGVVVRYWLRTAEADVP   208
TrdL      YLHAVEVVVVDASGDARTVIATREPSDPNHDLWWAHTGGGGGNFGVVVRYWLRTAEADVP   208
Orf22     HLYAVEVVVVDAAGRARLVVASRDDEGELGDLWWAHTGGGGGTFGVVTRYWLRTAGAT-G   216
          :* .******:    .  : .          :**********.**::.::*:*

Dbv29     RNPERLLPKPPATLLTSTVTFDWAGMTEAAFSRLLRNHGEWYERNSGPDSPYTGLWSQLM   296
ClmD2     AEPAGLLPRPPGALHIAEVSWPWDRLTGADFVRLVGNFMDWQIANSAVDSADADLYALLD   284
CrmK      DAPEEALPRPPASFHVARVSWSWAELTEADYVRLVSNFLDWQLRNCTVDSPNIGLYALLE   269
AknOx     TDPSQLLPKAPTSTLRHIVTWDWSALTEEAFTRIIDNHGAWHQSNSAAGTPYASMHSVFY   275
SpnJ      TDAAELLPRPPATVLLRSFHWPWHELTEQSFAVLLQNFGNWYEQHSAPESTQLGLFSTLV   298
BusJ      TDPSELLPRPPATVLLRSFHWPWCELTEQSFALLLRNFGTWYEQHSAPESTQLGLFSTLV   298
TamL      PEPGRLLPRPPAEVLLNTTVWPWEGLDEAAFARLVRNHGRWFEQNSGPDSPWCDLYSVLA   268
TrdL      PEPGRLLPRPPAEVLLNTTVWPWEGLDEAAFARLVRNHGRWFEQNSGPDSPWCDLYSVLA   268
Orf22     SDPAGLLPHPPSAVLDSLVTWSWEGMTAERFRRLMRNHAEWHERNGAADSPYASLFSVLG   276
                **: *:         : *  :    :  :: *.  *   .    :    : : :

Dbv29     IGNEVPGMGESGFMMPIQVDATRPDARRLLDAHIEAVIDGVPPAEVPEPI-EQRWLASTP   355
ClmD2     CPHRSAGD----ITLHAHLPEEAPRAHARMDAFLAALGAGVGIAPTVRRTS-LPWLAASQ   339
CrmK      CFHRSAG----HLAMHAQIPVDVPDAEERMSWFLAELNEGVAVAPSLTRR-RLPWLATSQ   324
AknOx     LNSRAAG----QILLDIQIDGGLDGAEALLNDFVAAVNEGTGVEPAVQRS-TEPWLRATL   330
SpnJ      CAHRQAG----YVTLNVHLDGTDPNAERTLAEHLSAINAQVGVTPAEGLRETLPWLRSTQ   354
BusJ      CAHRQAG----YVTLNIHLDGTDPNAERTLAEHLSAINDQVGVTPAEGLRETLPWLRSTQ   354
```



The letters in yellow background show the location of conserved FAD binding residues, involving a histidine and a cysteine. The letter in blue ground shows the position of a putative second catalytic residue (E376) in CrmK. The letters in cyan background shows CrmK E376 equivalents in other proteins. The letter Y in red ground shows the second catalytic residue in AknOx (Y378). TamL[4] (accession number ADC79636), TrdL[15] (accession number ADY38530), Dbv29[16] (accession number CAD91224), Orf22[17] (accession number ABL09969), AknOx[18] (accession number ABI15166), SpnJ[19] (accession number AAG23271), BusJ[20] (accession number AAY88927), CrmK[12] (accession number AFD30946), ClmD2 (accession number CCC55910).[21]

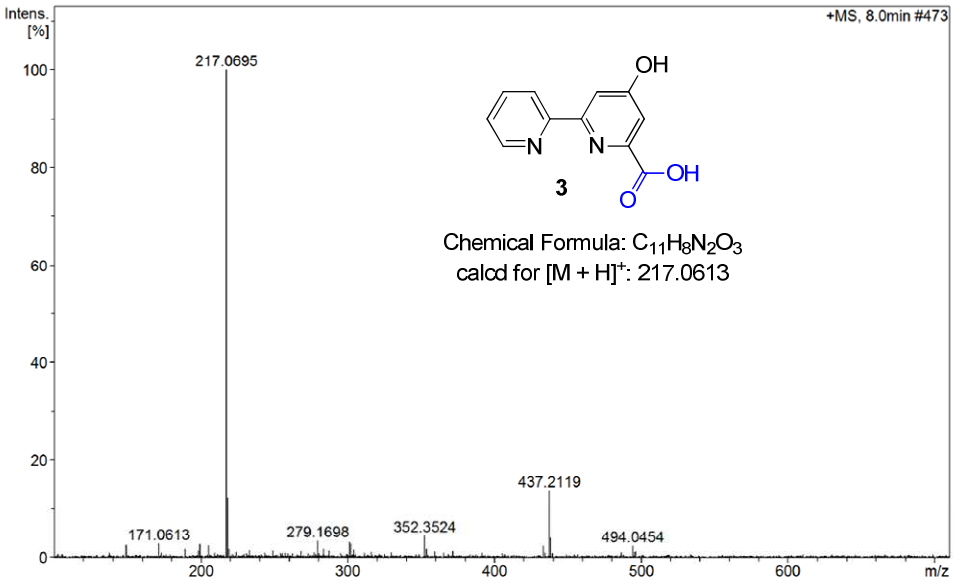
Fig. S2 SDS-PAGE analysis of the purification of CrmK.



The expression and purification of N-His_6-tagged CrmK in *E. coli* BL21(DE3)/pCSG2218. Lane 1, purified CrmK protein; Lane M, protein molecular weight marker. The acrylamide percentage of SDS-PAGE gels is 10 %.

Fig. S3 LC-MS characterization of products of CrmK reactions with CRM P **7**.

(A) LC-MS spectra of the product CRM O **3**



(B) LC-MS spectra of the product CRM M **4**

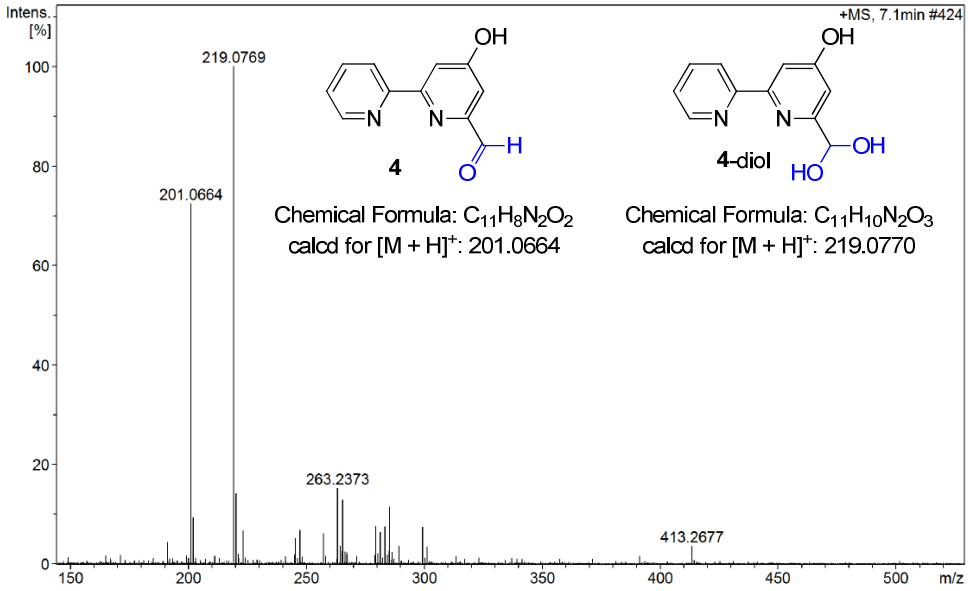
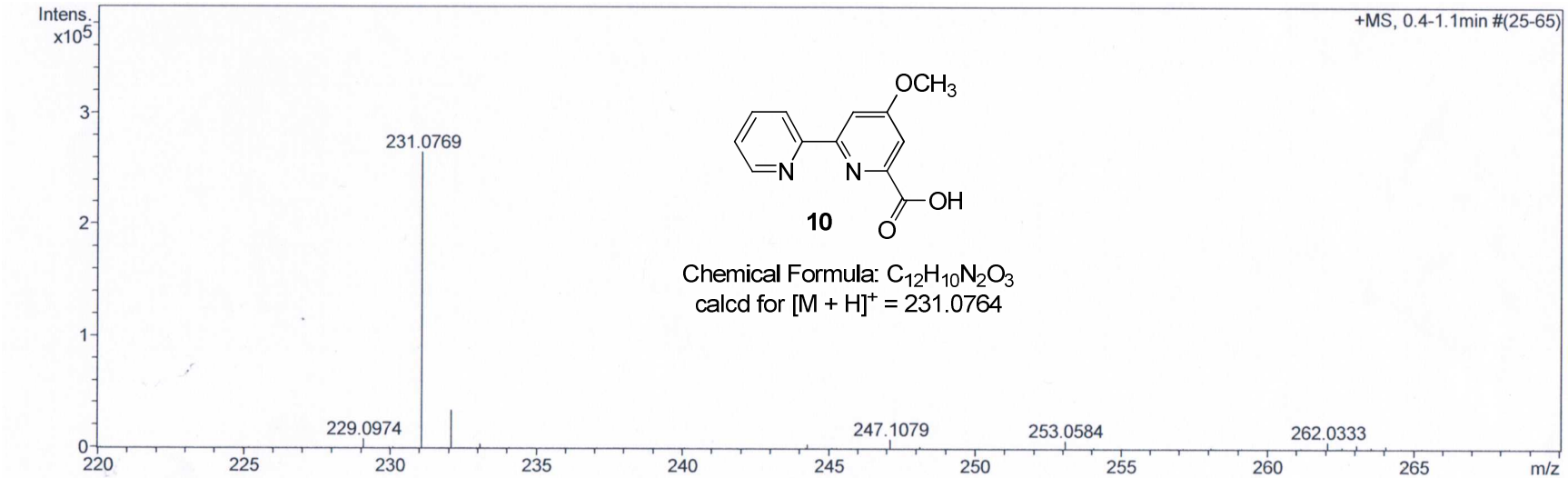


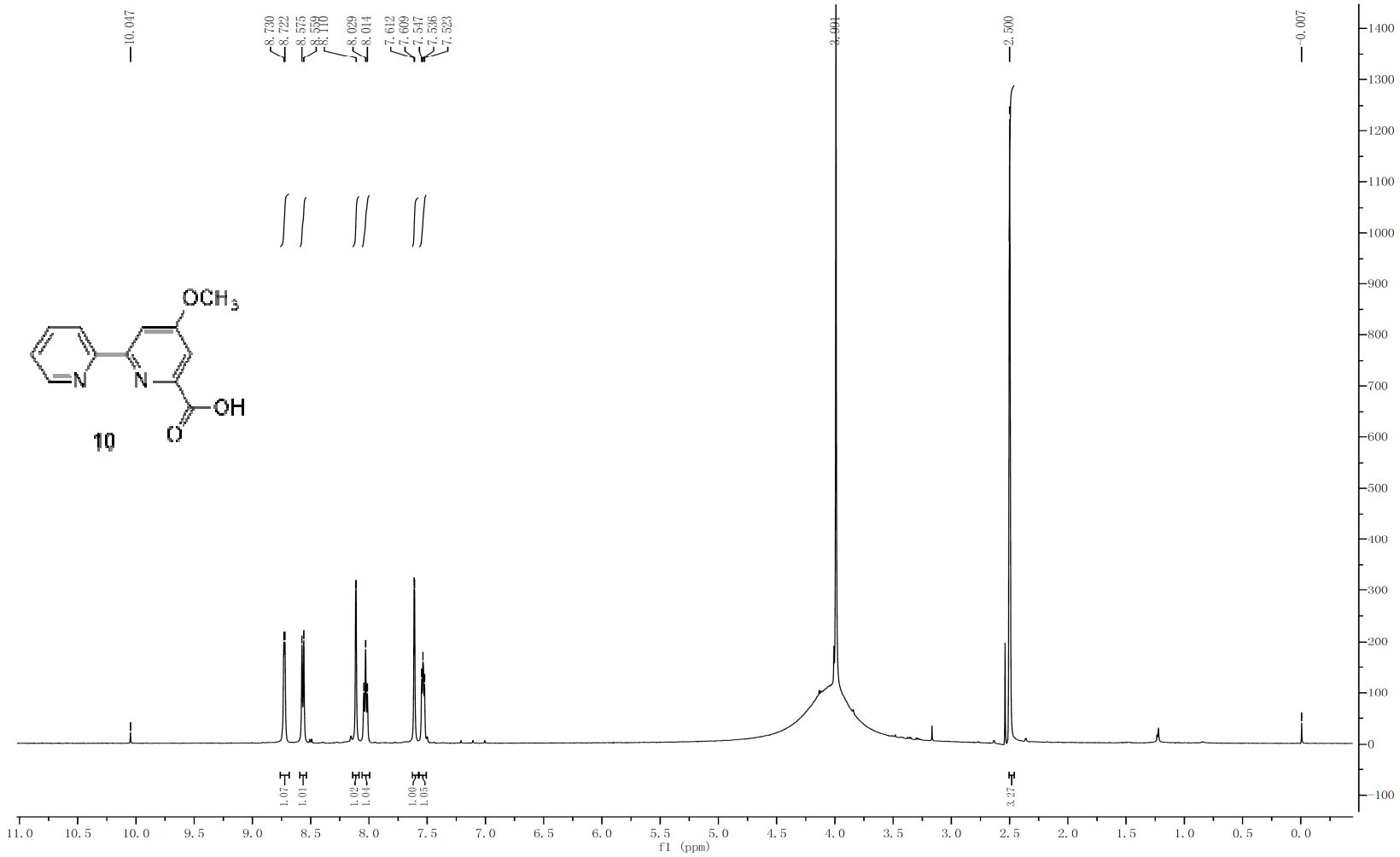
Fig. S4 ^{1}H and ^{13}C NMR spectroscopic data for CRM R **10**.

(A) The HRESIMS spectrum of CRM R **10**.



Meas. m/z	#	Formula	Score	m/z	err [mDa]	err [ppm]	mSigma	rdb	e⁻ Conf	N-Rule
231.0769	1	C 12 H 11 N 2 O 3	100.00	231.0764	-0.5	-2.0	6.4	8.5	even	ok
253.0584	1	C 12 H 10 N 2 Na O 3	100.00	253.0584	-0.0	-0.2	3.5	8.5	even	ok

(B) The ^{1}H NMR spectrum of CRM R **10** in DMSO-*d6*



(C) The ^{13}C NMR spectrum of CRM R **10** in DMSO-*d6*

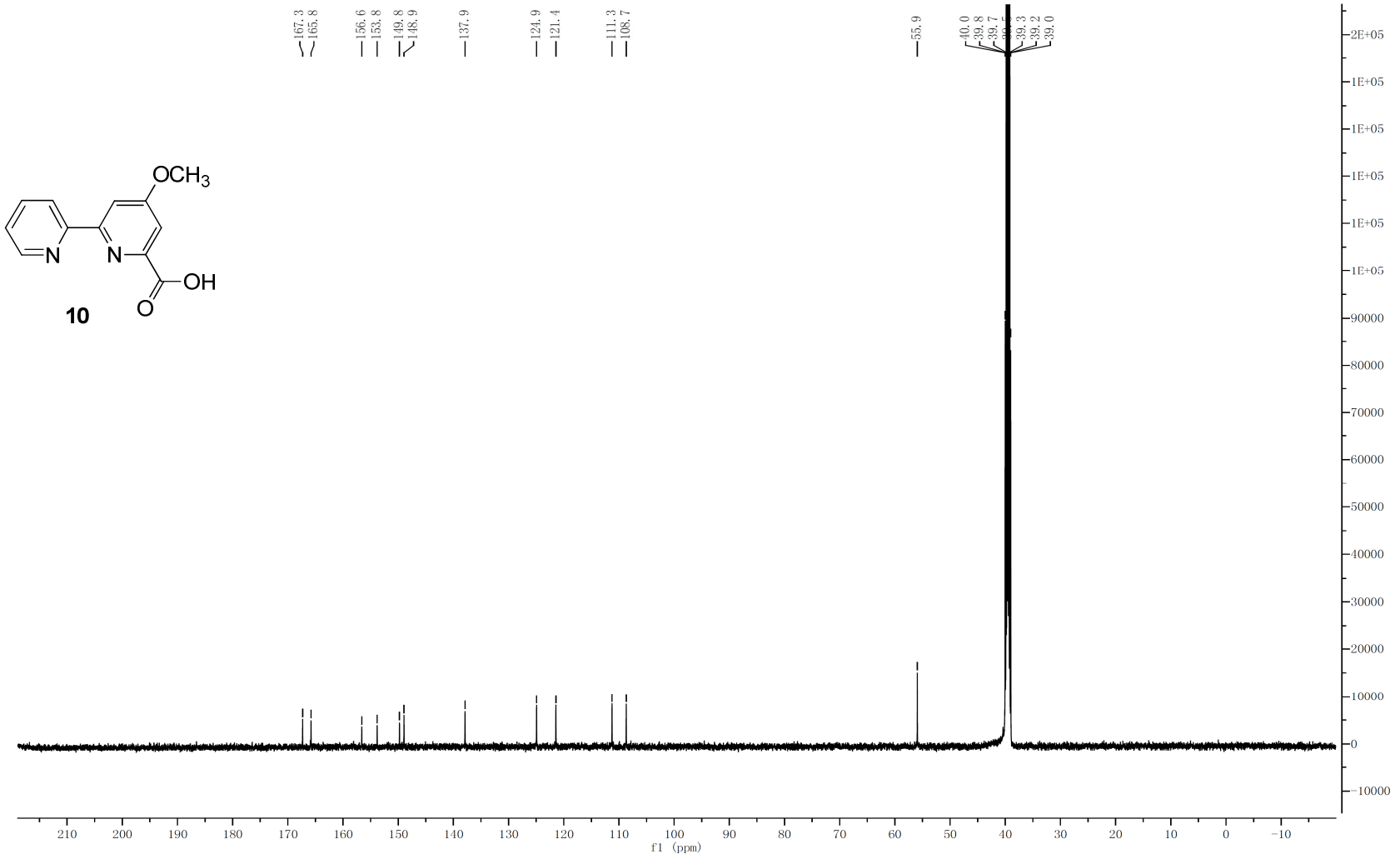
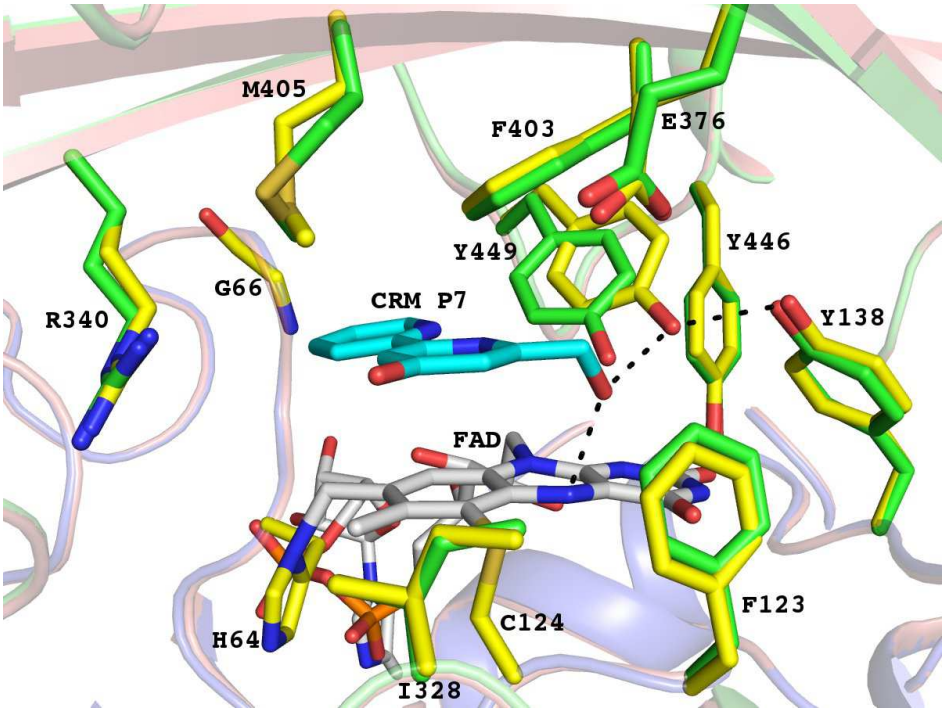


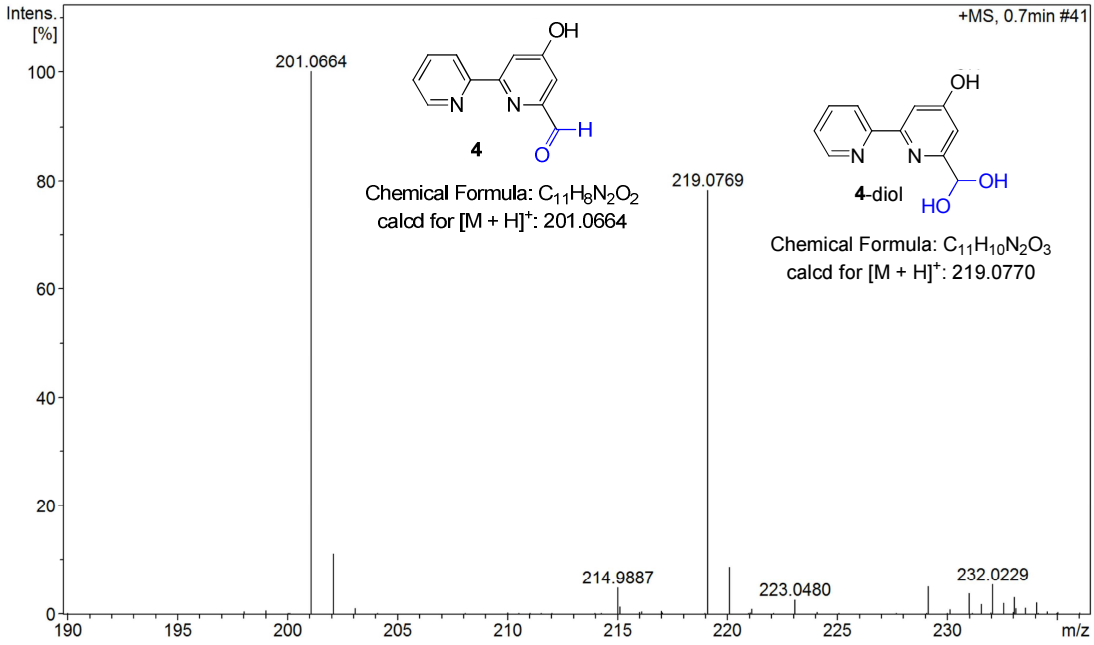
Fig. S5 Conformational change of Tyr449 in CrmK upon substrate binding.



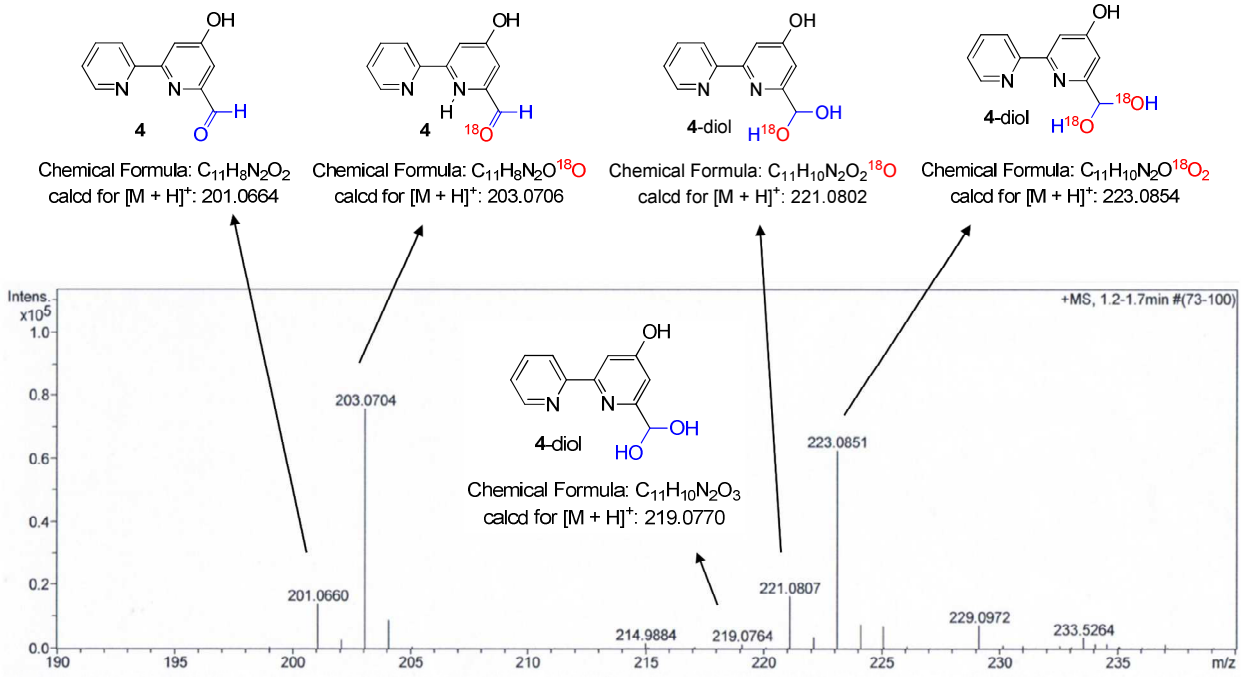
The carbon atoms of the residues lining the active site are colored in green and yellow, respectively, in the apo and substrate-bound form. The only major conformational change in CrmK upon the substrate binding is the swing of ~16° of the Y449 side chain to accommodate the substrate molecule, leading to a ~1.5 Å displacement of the Y449 hydroxyl group to enable an H-bond (2.6-3.0 Å) formation with the Y138 hydroxyl group, which is not found in the *apo* CrmK (~4 Å).

Fig. S6 MS data of CRM M 4 dissolved in H_2O or $H_2{}^{18}O$.

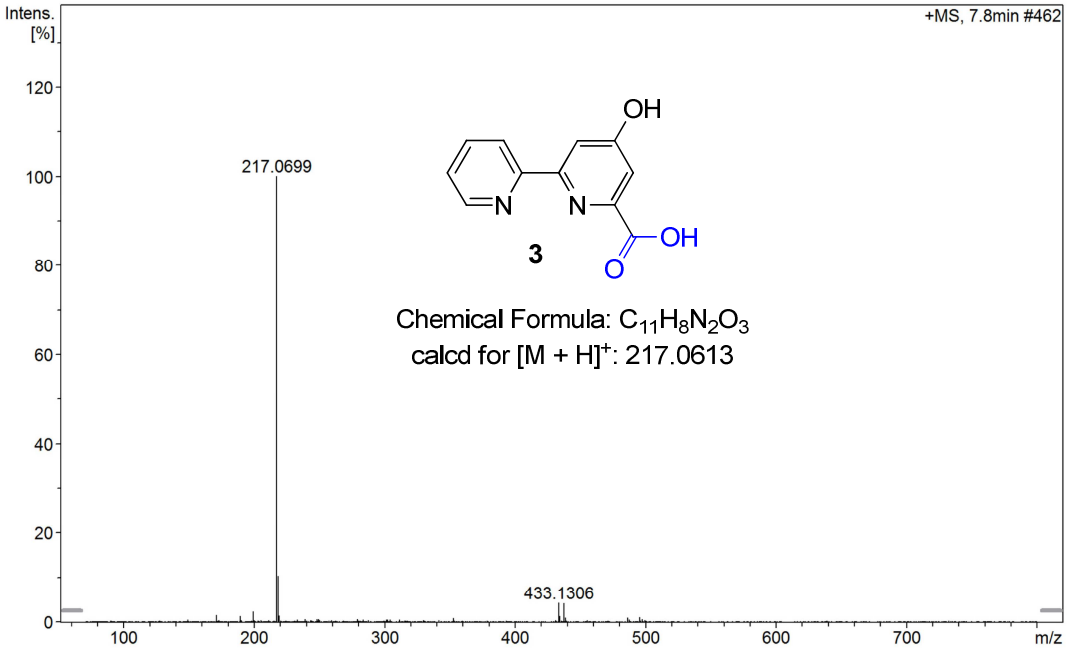
(A) Mass spectra of compound CRM M **4** in H_2O solution



(B) Mass spectra of compound CRM M **4** in $H_2{}^{18}O$ (97% purity) solution



(C) LC-MS spectra of CrmK enzymatic activity of compound CRM P **7** in H_2O solution



(D) LC-MS spectra of CrmK enzymatic activity of compound CRM P **7** in H_2O^{18} solution

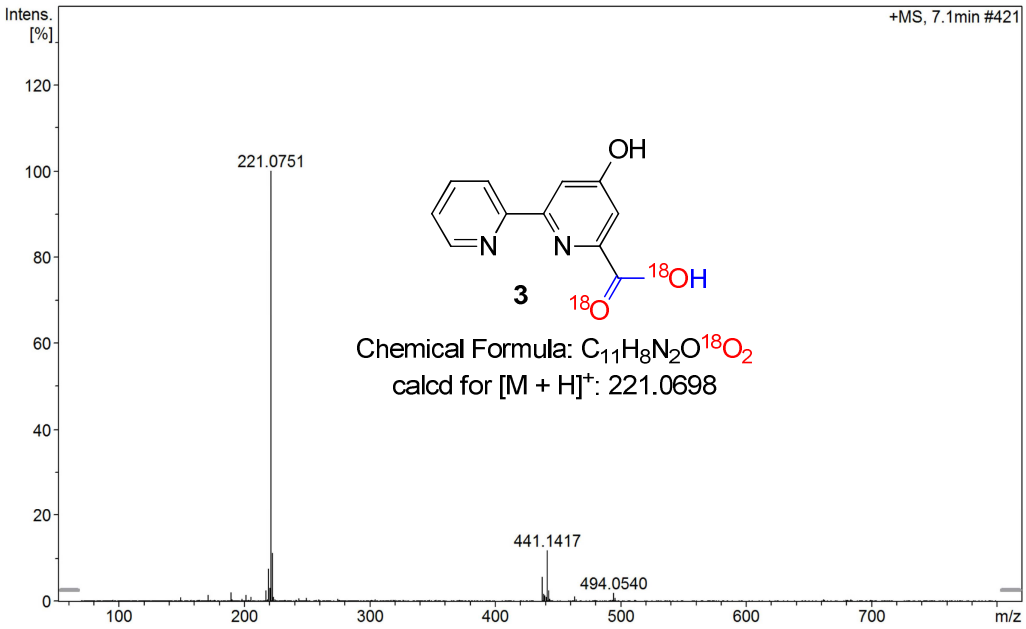
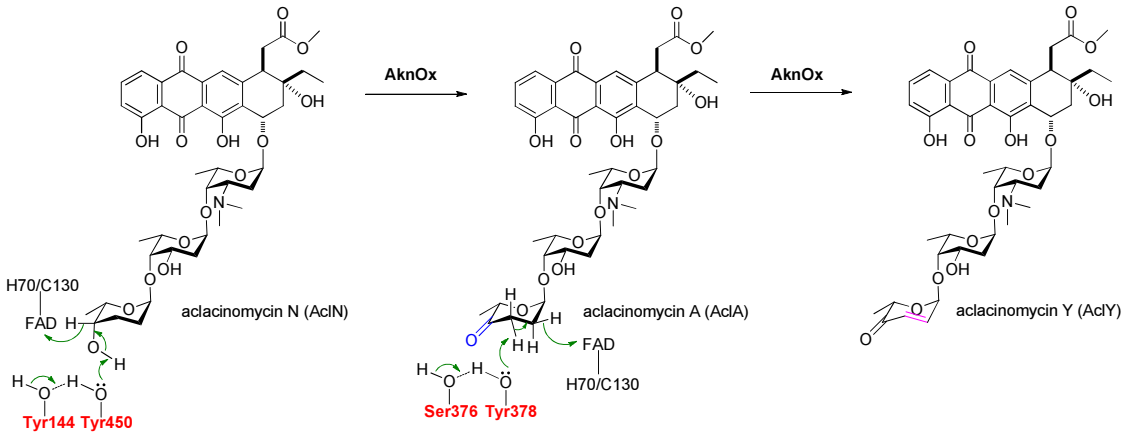
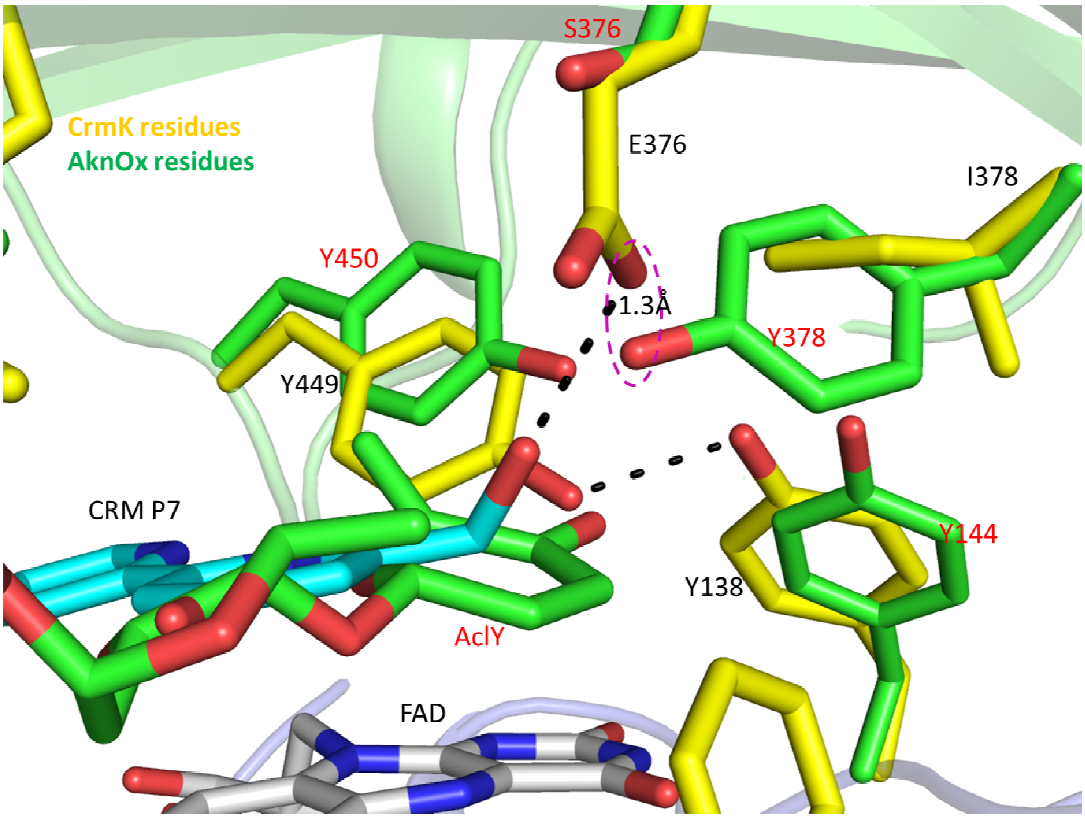


Fig. S7 Proposed catalytic mechanism with dual active site for AknOx.



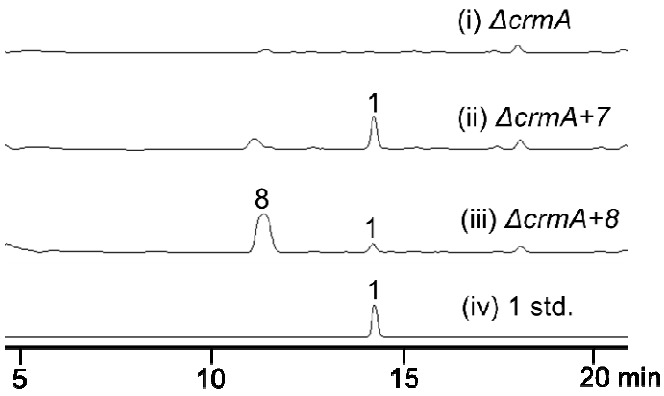
The dual active site (Tyr450/Tyr144, Tyr378/Ser376) was proposed to be responsible to catalyze two different dehydrogenation reactions in AknOx catalysis.[18]

Fig. S8 Superposition of active site residues in AknOx and CrmK.



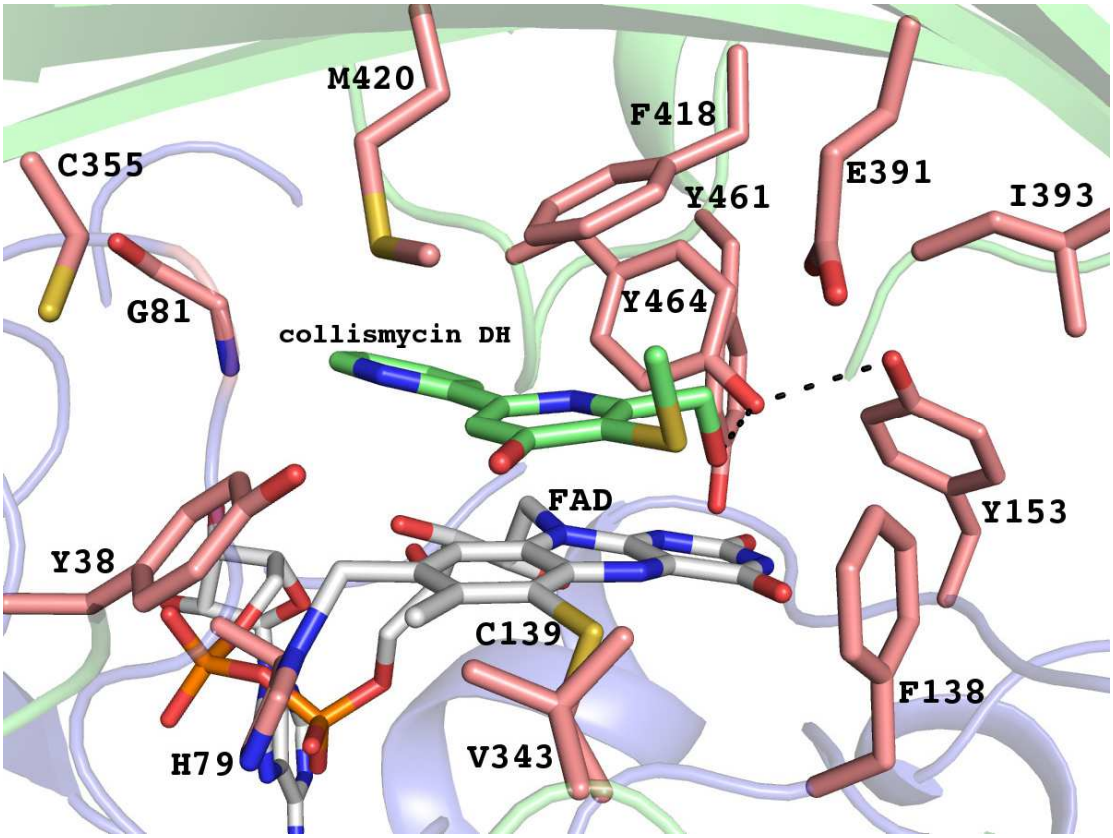
The carbon atoms of catalytic residues in AknOx (Y450/Y144 and Y378)[18] and CrmK (Y449/Y138 and E376) are colored in green and yellow, respectively. The carbon atoms of ligands are shown in cyan and green, respectively. Notably, upon superposition, the carboxylate group of E376 of CrmK is very close (~1.3 Å) to the hydroxyl group of Y378 in AknOx.

Fig. S9 HPLC analysis of feeding experiments in the *ΔcrmA* mutant with **7** and **8**.



(i) the *ΔcrmA* mutant (the CRMs-nonproducing mutant, in which the PKS/NRPS gene *crmA* was disrupted;[12] (ii) the *ΔcrmA* mutant was fed with CRM P **7** (30 μM); (iii) the *ΔcrmA* mutant was fed with CRM F **8** (30 μM); (iv) standard **1**.

Fig. S10 Active site architecture of ClmD2.



The homology model of ClmD2 was constructed based on the crystal structure of CrmK (68% sequence identity) and the substrate collismycin DH was docked in the active site of ClmD2 using the software MOE.[22] The orientation of active site shown in this figure is similar to that of CrmK presented in the main text. The carbon atoms of active site residues and the substrate are shown in salmon and pale green, respectively. The H-bonds are shown in black dash lines. Among all the residues lining the substrate binding pocket, three residues D23, R340, and I328 in CrmK have been replaced by Y38, C355, and V343, respectively in ClmD2. Notably, the replacement of I328 in CrmK by Val343 in ClmD2 would better accommodate the extra –SCH3 group on the collismycin DH substrate.[23]

Supplementary References.

1. M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248-254.
2. A. G. W. Leslie and H. R. Powell, *Evolving Methods for Macromolecular Crystallography*, 2007, **245**, 41-51.
3. A. Vagin and A. Teplyakov, *J. Appl. Crystallogr.*, 1997, **30**, 1022-1025.
4. J. C. Carlson, S. Y. Li, S. S. Gunatilleke, Y. Anzai, D. A. Burr, L. M. Podust and D. H. Sherman, *Nat. Chem.*, 2011, **3**, 628-633.
5. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Cryst. D*, 1997, **53**, 240-255.
6. P. Emsley and K. Cowtan, *Acta Cryst. D*, 2004, **60**, 2126-2132.
7. R. A. Laskowski, M. W. Macarthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283-291.
8. K. A. Datsenko and B. L. Wanner, *Proc. Natl. Acad. Sc.i U. S. A.*, 2000, **97**, 6640-6645.
9. D. J. MacNeil, K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons and T. MacNeil, *Gene*, 1992, **111**, 61-68.
10. P. Fu, S. Wang, K. Hong, X. Li, P. Liu, Y. Wang and W. Zhu, *J. Nat. Prod.*, 2011, **74**, 1751-1756.
11. Y. Zhu, J. Xu, X. Mei, Z. Feng, L. Zhang, Q. Zhang, G. Zhang, W. Zhu, J. Liu and C. Zhang, *ACS Chem. Biol.*, 2016, DOI: 10.1021/acschembio.1025b00984.
12. Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu and C. Zhang, *Org. Lett.*, 2012, **14**, 2666-2669.
13. B. Gust, G. L. Challis, K. Fowler, T. Kieser and K. F. Chater, *Proc. Natl. Acad. Sc.i U. S. A.*, 2003, **100**, 1541-1546.
14. M. S. B. Paget, L. Chamberlin, A. Atrih, S. J. Foster and M. J. Buttner, *J. Bacteriol.*, 1999, **181**, 204-211.
15. X. Mo, H. Huang, J. Ma, Z. Wang, B. Wang, S. Zhang, C. Zhang and J. Ju, *Org. Lett.*, 2011, **13**, 2212-2215.
16. Y. C. Liu, Y. S. Li, S. Y. Lyu, L. J. Hsu, Y. H. Chen, Y. T. Huang, H. C. Chan, C. J. Huang, G. H. Chen, C. C. Chou, M. D. Tsai and T. L. Li, *Nat. Chem. Biol.*, 2011, **7**, 304-309.
17. A. Luzhetskyy, A. Mayer, J. Hoffmann, S. Pelzer, M. Holzenkamper, B. Schmitt, S. E. Wohlert, A. Vente and A. Bechthold, *ChemBioChem*, 2007, **8**, 599-602.
18. I. Alexeev, A. Sultana, P. Mantsala, J. Niemi and G. Schneider, *Proc. Natl. Acad. Sc.i U. S. A.*, 2007, **104**, 6170-6175.
19. H. J. Kim, R. Pongdee, Q. Wu, L. Hong and H. W. Liu, *J. Am. Chem. Soc.*, 2007, **129**, 14582-14584.

20. D. R. Hahn, G. Gustafson, C. Waldron, B. Bullard, J. D. Jackson and J. Mitchell, *J. Ind. Microbiol. Biotechnol.*, 2006, **33**, 94-104.
21. I. Garcia, N. M. Vior, A. F. Brana, J. Gonzalez-Sabin, J. Rohr, F. Moris, C. Mendez and J. A. Salas, *Chem. Biol.*, 2012, **19**, 399-413.
22. *Molecular Operating Environment (MOE)*, 2013.08, Chemical Computing Group Inc., Montreal, Canada.
23. I. Garcia, N. M. Vior, J. Gonzalez-Sabin, A. F. Brana, J. Rohr, F. Moris, C. Mendez and J. A. Salas, *Chem. Biol.*, 2013, **20**, 1022-1032. .