Directed evolution of a cyclodipeptide synthase with new activities via label-free mass spectrometric screening

Songya Zhang, [‡a] Jing Zhu,[‡a] Shuai Fan,[b] Wenhao Xie,[a] Zhaoyong Yang,[*b] and Tong Si[*a]

1 Materials and chemicals.

PrimeSTAR DNA polymerase was purchased from Takara Bio, restriction enzyme was from New England Biolabs GmbH (NEB). Oligonucleotides and *albC* gene from *Streptomyces noursei* were synthesized from GENEWIZ Bio Inc. (Suzhou, China). The MALDI matrix α-cyano-4-hydroxycinnamic acid (CHCA) was obtained from Bruker. GeneMorph II Random Mutagenesis Kit was supplied from Agilent Technologies. DMSO, trifluoroacetic acid (TFA), high-performance liquid chromatography (HPLC)–grade acetonitrile, and HPLC-grade methanol were purchased from Merck. The chemical standards of cFL and cFY were purchased from GL Biochem (Shanghai, China). Deionized water was obtained from Watson's (Quchenshi, Shanghai, China).

2 General procedures

2.1 Construction of AlbC Mutant Libraries

Strains, plasmids, and primers are listed in Table S1 and S2. GeneMorph II Random Mutagenesis Kit (Agilent Technologies) was used to introduce mutations during the error-prone PCR (epPCR) process. The constructed pBluescript II SK (+)-AlbC plasmid was used as template. All procedures were performed according to the manufactures' protocol. Briefly, the library was constructed at a mutation rate of between 0 to 4.5 mutations per 1000 bp. EpPCR process was performed at a final volume of 50 µL containing 500 nM each primer, 10% DMSO, dNTPs (200 µM each), 552 ng template and 2.5 units of Mutazyme II DNA polymerase. PCR was performed under the following conditions: 95 °C for 2 min (1 cycle); 95 °C for 0.5 min, 62 °C for 0.5 min, and 72 °C for 1 min (25 cycles); and, 72 °C for 10 min (1 cycle). The primers AlbC-GA-F and AlbC-GA-R (Table S2) used for amplification with homologous overlap sequences to the vector pET28a. The PCR products and the linearized pET28a (linearized by PCR amplification using primers pET28a-GAalbC-F and pET28a-GAalbC-R) were purified and assembled using Gibson Assembly Master Mix (NEB). Subsequently, the assembled product was desalted by ethanol precipitation and electrotransformed into *E. coli* DH5α competent cells using the Gene Pulser Xcell System (Bio-Rad Laboratories, USA). All of the transformed cells were spread onto Luria Broth (LB) agar plates containing 50 µg/mL kanamycin. Overnight incubation at 37 °C yielded >10^6 independent transformants, from which plasmid DNA library was isolated and used to transform the production host *E. coli* Rosetta(DE3). Approximately 4500 mutants were screened at the first round. For the second round, the same procedure as the first round was applied to create the epPCR library using pBluescript II SK (+)-F186L-AlbC as the parent template, and around 2300 colonies were screened.

Site-saturation mutagenesies (SSM) plasmid libraries were created by Gibson Assembly following the "22c-trick" strategy[1]. All primers used are provided in Table S2, and pET28a-AlbC plasmid was used as template. The PCR reaction mixture (50 µL of final volume) contained 5.0 µL of DMSO (10%), 0.5 µL of template (1 ng/µL), 2.0 µL of primers (10 µM each mix), and 25 µL of 2x PrimeSTAR® HS DNA Polymerase (Sangon Biotech, Shanghai). The PCR reaction program consisted of an initial cycle at 98 °C for 3 min, followed by 28 cycles of denaturing at 98 °C for 15s, annealing at temperatures at 62 °C for 15s, and elongation at 72 °C for 1 min with a final extension at 72 °C for 5 min. PCR products were treated with 1 µL of DpnI (20 U/µL) to remove the methylated template before used as Gibson assembly substrates . Upon desalting by ethanol precipitation, Gibson assembly reaction mix was used to transform *E. coli* Rosetta (DE3) cells by electroporation. For each SSM library, 94 transformants obtained from LB agar plates containing 50 µg/mL kanamycin were screened.

2.2 Library Screening via automation

One of the robotic workcells of the Shenzhen Synthetic Biology Infrastructure was used in this study. It contains a spinnaker robotic arm on a 3.6-m track (Thermo Scientific, Waltham, MA), a QPix HT Colony Picker (Molecular Devices, San Jose, CA), a Fluent 780 liquid handling robot (Tecan, Männedorf, Switzerland) which was outfitted with a RGA robotic manipulation arm, a FCA 8-channel independent pipetter, a MCA384 with EVA adapter for 96-channel pipetting, a CPAC Ultraflat temperature-controlled block (Inheco, Munich, Germany), and two Te-Shake Silver shakers (Tecan, Männedorf, Switzerland), a Multidrop Combi reagent dispensers (Thermo Scientific, Waltham, MA), a Spark microplate reader (Tecan, Männedorf, Switzerland), a HiG 3 Robotic plate centrifuge (BioNex, San Jose, California), two WASP plate sealer (Kbiosystems, Essex, UK), and a Xpeel seal pealer (Brooks Life Sciences, Manchester, UK). Momentum (Thermo Scientific, Version 6.0.2) provided control and data integration for those devices. The liquid handling robot and program pipetting, temperature control and labware transportation were manipulated by Fluent control (Tecan, Version 3.0). Process for robotic screening of AlbC variant libraries in recombinant *E. coli* briefly described below: colonies were picked using QPix Colony Picker, and cultured overnight at 37 °C in 96-well microplates with 180 µL of LB medium (50 µg/mL kanamycin) in each well. Then, 20 µL of the overnight culture was transferred to a new 96-deep well plate with 400 µL of LB containing 50 µg/mL kanamycin and 0.1 mM IPTG, and resultant cultures were incubated at 30 °C for 17 h with constant shaking at 800 rpm. Separately, 60 µL of 50% sterile glycerol was added into the residue overnight culture to prepare glycerol stock. The induced culture was extracted with 400 µL of ethyl acetate at shaker for 0.5 h, and 0.6 µL of the upper layer of organic phase was used for high-throughput MS screening.

2.3 Metabolite analysis

The analytes from ethyl acetate extraction were analyzed using MALDI-Tof MS. In particular, 0.6 µL upper organic solution of each sample was spotted on 384 stainless steel MALDI target, and then overlaid with 0.6 µL of CHCA matrix solution (10 mg/mL in 70% acetonitrile with 0.1% TFA). Using air-dried samples, MALDI–ToF mass spectra (1000 laser shots) were recorded on an autoflex maX MALDI-ToF MS (Bruker Daltonic, USA) operating in reflection mode (m/z range: 100 to 400). The standard curves of MALDI-ToF MS analysis of two DKP products, cFL and cFY, were measured using chemically synthesized standards (Figure S7). The detection limits were 20 pmol and 10 pmol for cFL and cFY, respectively, when setting the signal-to-nois (S/N) threshold to >5. And the linear ranges of responses for cFL and cFY in our assay were 20-160 pmol and 10-160 pmol, respectively (Figure S7).

For LC-MS analysis, the culture was extracted with ethyl acetate and the organic phase was collected and dried with a speed vacuum concentrator. The dried extracts were dissolved in methanol and analyzed on an Agilent 1260 Infnity II liquid chromatograph coupled to an Agilent 6470 triple quadrupole mass spectrometer (Agilent, USA). Data were acquired and analyzed using Agilent MassHunter Workstation software. Chromatographic separation was performed on an ACE EXCEL 3 Super C18 UHPLC column (150 x 2.1mm, 2.0 µm). The mobile phase was composed of solvent A (deionized H_2O with 0.1% acetic acid, v/v) and solvent B (acetonitrile with 0.1% acetic acid, v/v). A gradient elution method (15% B over 3 min, 15−75% B from 3 to 9 min,15% B from 9.1 to 10 min, and a flow rate of 0.5 mL/min) was used with an injection volume of 1.0 µL. The following parameters for the spectrometer were utilized: electrospray positive-ion mode for ionization, a capillary voltage of 5.0 kV, and a collision energy of 0 or 10 eV. For the drying gas (350°C) and sheath gas (350°C), the flow rates were set to 10 and 15 L/min, respectively. The MRM parameters in analyzing the eleven compounds, including the quantitative and qualitative ions, collision energy, dwell time, and retention time are shown in Table S3. High-resolution mass spectrometry (HR-MS) data was obtained on a Vanquish Flex UHPLC (Thermo Fisher Scientific, USA) coupled online via electrospray ionization source with an Q ExactiveTM HF MS instrument (Thermo Fisher Scientific, USA).

2.4 Data processing and statistical analysis

The data files of MALDI-Tof mass spectra were imported into ClinProTools software (Bruker) and converted to csv format for feature extraction. Extracted features were then uploaded to the online MetaboAnalyst platform (http://www.metaboanalyst.ca/). The feature intensities were internal calibration by matrix fragment ion *m/z* 379 ($[2M+H]^+$). For data filtering, “None” option was selected; the features were log transformed and Pareto scaled before statistics analysis. Based

on the results of analysis of variance (ANOVA) in MetaboAnalyst, samples with m/z feature(s) exhibiting substantial differences among variants and WT were chosen for further verification using LC-MS/MS.

2.5 MD Analysis

2.5.1 Modeling the AlbC:Substrate Complex

The AlbC:Substrate model was prepared from the crystal structure with PDB ID: 4Q24[2] at 2.9 Å resolution. The UCSF Chimera program[3] was used to revert the mutated residues with the mutagenesis tool. As previously mentioned[4]. the wild-type Albc was built from the crystal structure by converting Cys37 into Ser37; the existed ligand was deleted the C91 atom and made a covalent bond between the O atom of Ser37 and the C92 atom for the ligand. And then different dipeptidyl ligands were built on this basis, such as cyclo(L-Phe-L-Leu) (cFL) and cyclo(L-Phe-L-Val) (cFV). The protonation state of the residues of AlbC was assigned after careful visual inspection of the environment of each residue and considering the pKa values obtained by PROPKA 3.0[5,6], except Glu182, which was in the deprotonated form. Finally, three models were built, WT-cFL, F186L-cFL and F186L-cFV.

2.5.2 Molecular Dynamics Simulation

The simulation were performed using the Gromacs 2019.6 package[7]. Parameterization of the protein was done with the CHARMM36 force field[8]. We then proceeded to parametrize the ligand obtained from CHARMM-GUI server[9]. The system thus obtained was solvated in a dodecahedron box with better simulation efficiency of TIP3P water[10] molecules extending 10 Å from any atoms of the protein in any directions. Three sodium ions were added to keep the system neutral. The final system had 28 608 atoms. In the latter, all bond lengths involving hydrogen atoms were restricted with the SHAKE algorithm[11]. The time step for the integration of the equations of motion was 2 fs. The particle mesh Ewald (PME) method[12] with a cutoff distance of 10 Å was used to calculate the Coulomb interactions. The steepest descent method was carried out to minimize the system with a tolerance value of 1000.0 kJ/mol/nm. Then all systems were minimized and then equilibrated in NVT and NPT ensembles. The system was then heated to 300 K using the v-rescale temperature coupling scheme[13] using the NVT ensemble in 1000 ps, followed by another 1000-ps NPT simulation using the Parrinello Rahman pressure coupling scheme[14]. The four system were performed for 100 ns each.

2.5.3 Binding Free Energy Calculations

We could further assess the free binding energies between Albc and ligands with the MD results. The MMPBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) method[15] implemented in the gmx_MMPBSA program[16] is adopted along with AmberTools21[17]. The PB model estimate only the polar component of the solvation. The nonpolar component is usually assumed to be proportional to the molecule's total solvent accessible surface area (SASA), with a proportionality constant derived from experimental solvation energies of small nonpolar molecules. This method has been proven to balance accuracy and computational efficiency, especially when dealing with large systems. In the 100ns simulation, the most stable 30ns trajectory was selected for free energy calculation. After MD, the activate site used for analysis using POcket Volume MEasurer (POVME)[18].

Table S1: Strains and plasmids used in this study

Strain/plasmid	**Description**	**Reference**
E. coli		
DH5α	General cloning host	Invitrogen
Rosetta(DE3)	BL21(DE3) pRARE (Cm^r)	Novagen
Plasmid		
pET28a	P_{T7}, pBR322 ori, Kan^r	Novagen
pBluescript II SK (+)	f1 ori, Amp^r	Addgene
pBluescript II SK (+)-AlbC	MSC::*albC*, f1 ori, Amp^r	This study
pBluescript II SK (+)-F186L-AlbC	MSC::F186L*albC*, f1 ori, Amp^r	This study
pET28a-AlbC	P_{T7}::*albC*, pBR322 ori, Kan^r	This study

Table S2: Primers used in this study

Primer name	Primer sequence (5'-3')
trick-AlbC-200L-R2	GACCCAGAGCGCTGCC
AlbC-trick-200L-F1	CGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGC
trick-AlbC-200L-F2	CACCGAGGAGGGGACGG
AlbC-trick-200L-R1	CCGTCCCCTCCTCGGTGXXXTGCTACCACATCGACACCC
trick-AlbC-33L-F2	GGCGTGCTCACCGCG
AlbC-trick-33L-R1	CGCGGTGAGCACGCCXXXATCGGAATCAGTGCGGGC
trick-AlbC-65V-F2	GACATCGGTGCGCTCGAAAC
AlbC-trick-65V-R1	GTTTCGAGCGCACCGATGTCXXXTATGTCGACACCCACATCGACG
trick-AlbC-119L-F2	GGACCGGACACGGAACC
AlbC-trick-119L-R1	GGTTCCGTGTCCGGTCCXXXTCCGAGCTCCAGGAGACC
trick-AlbC-185L-F2	CGGGGCCTCGGCCAG
AlbC-trick-185L-R1	TGCTGGCCGAGGCCCCGXXXTTCGCGGACTCGCCCG
trick-AlbC-152M-F2	GTCCTCGCAGGCGGT
AlbC-trick-152M-R1	ACCGCCTGCGAGGACXXXGTGCGGGCCGTGGTGA
trick-AlbC-159N-F2	CATCACCACGGCCCGC
AlbC-trick-159N-R1	GCGGGCCGTGGTGATGXXXCGGCCCGGTGACGGC
trick-AlbC-204I-F2	GTGGTAGCAGAGCACCGAGG
AlbC-trick-204I-R1	TCGGTGCTCTGCTACCACXXXGACACCCCGATCACGGCG
trick-AlbC-206T-F2	GTCGATGTGGTAGCAGAGCAC
AlbC-trick-206T-R1	GTGCTCTGCTACCACATCGACXXXCCGATCACGGCGTTCCTGT
trick-AlbC-207P-F2	GGTGTCGATGTGGTAGCAGAGC
AlbC-trick-207P-R1	CTGCTACCACATCGACACCXXXATCACGGCGTTCCTGTCC
trick-AlbC-205D-F2	GATGTGGTAGCAGAGCACCG
AlbC-trick-205D-R1	CGGTGCTCTGCTACCACATCXXXACCCCGATCACGGCG
trick-AlbC-98R-F2	CCGCAGATCCTTGAGCGTG
AlbC-trick-98R-R1	CACGCTCAAGGATCTGCGGXXXAGACTCCGGCGCTCGC
trick-AlbC-99R-F2	GCGCCGCAGATCCTTGA
AlbC-trick-99R-R1	TCAAGGATCTGCGGCGCXXXCTCCGGCGCTCGCTGG
trick-AlbC-101R-F2	GAGTCTGCGCCGCAGAT
AlbC-trick-101R-R1	ATCTGCGGCGCAGACTCXXXCGCTCGCTGGAGAGCGT
trick-AlbC-102R-F2	CCGGAGTCTGCGCCGC
AlbC-trick-102R-R1	GCGGCGCAGACTCCGGXXXTCGCTGGAGAGCGTGGG
pET28a-GAalbC-F	CATATGGCTGCCGCGCGGCA
pET28a-GAalbC-R	TAGCTCGAGCACCACCACCACCAC
AlbC-GA-F	TGCCGCGCGGCAGCCATATG
AlbC-GA-R	GTGGTGGTGGTGGTGCTCGAGCTA

Note: Codon sets XXX: NDT, VHG, TGG.

Table S3. MRM parameters and transitions of 11 cyclodipeptides

Compound Name	Formula	Precursor Ion	Product Ion	Fragmentor	Collision	Cell Accelerator Voltage	Polarity
cFL	$C_{15}H_{20}N_2O_2$	261	233	135	0	5	Positive
cFF	$C_{18}H_{18}N_2O_2$	295	267	135	0	5	Positive
cFY	$C_{18}H_{18}N_2O_3$	311	283	135	0	5	Positive
cFM	$C_{14}H_{18}N_2O_2S$	279	231	135	0	5	Positive
cYL	$C_{15}H_{20}N_2O_3$	277	249	135	0	5	Positive
cYM	$C_{14}H_{19}N_2O_3S$	295	247	135	0	5	Positive
cYY	$C_{18}H_{18}N_2O_4$	327	299	135	10	5	Positive
cLL	$C_{12}H_{22}N_2O_2$	227	199	135	0	5	Positive
cMM	$C_{10}H_{18}N_2O_2S_2$	263	215	135	0	5	Positive
cLM	$C_{11}H_{20}N_2O_2S$	245	197	135	10	5	Positive
cFV	$C_{14}H_{18}N_2O_2$	247	219	135	10	5	Positive

Table S4. The binding cavity volume of WT and mutants.

Site	**Volume size**
WT-cFL	197 $Å^3$
F186L-cFV	288 $Å^3$
F186L-cFL	215 $Å^3$

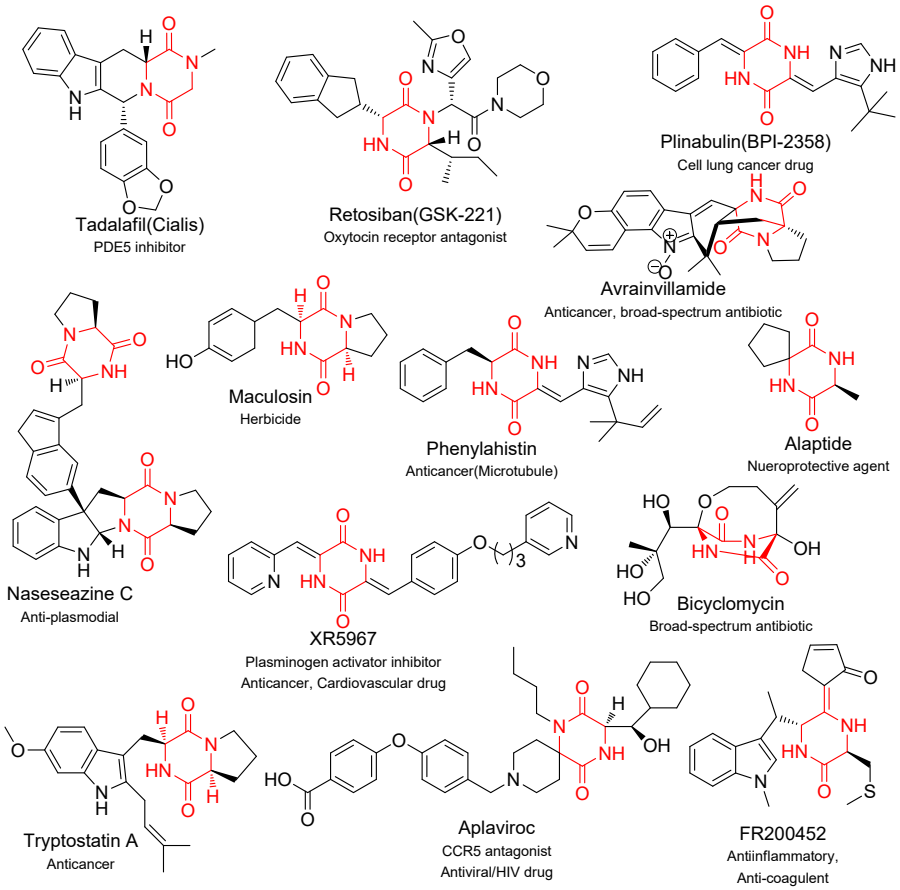


Figure S1. Representative molecule with DKPs scaffold.

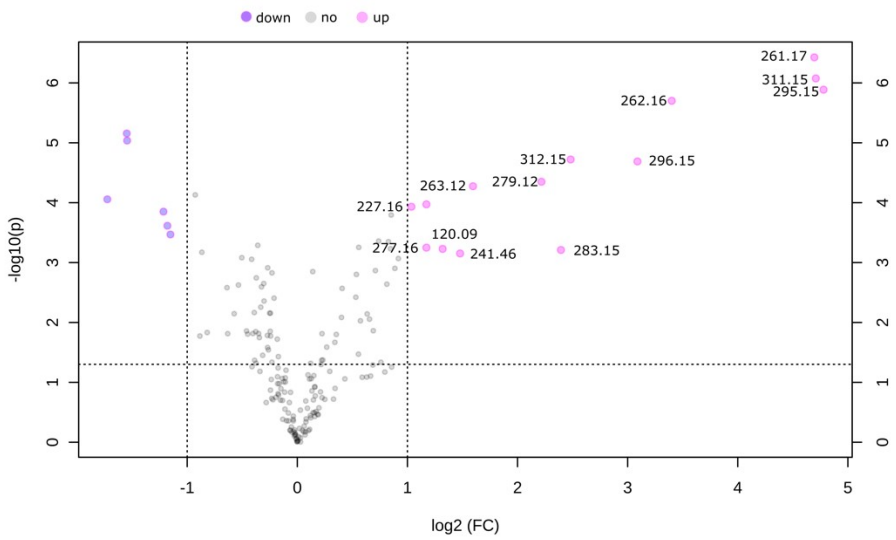
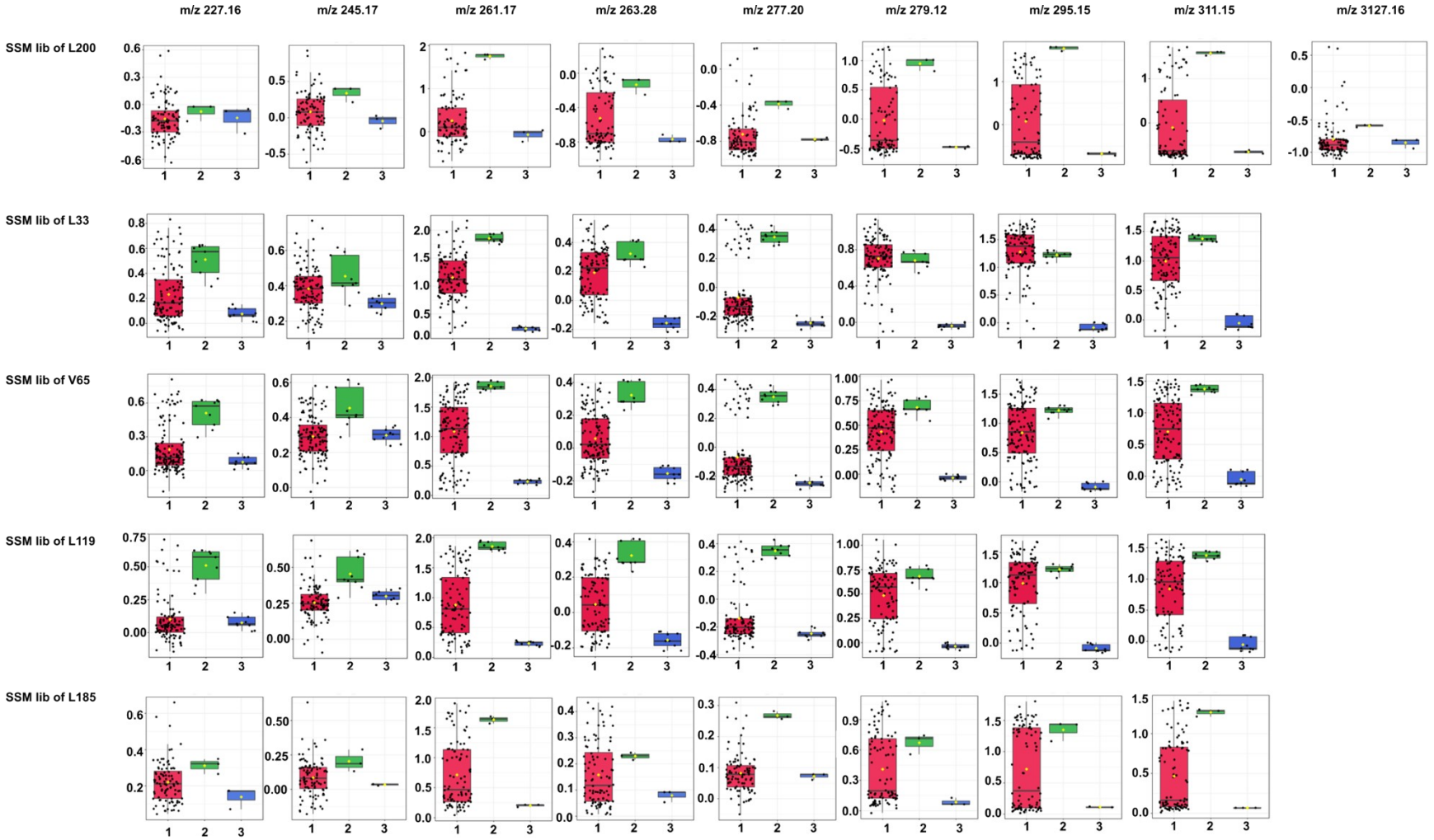
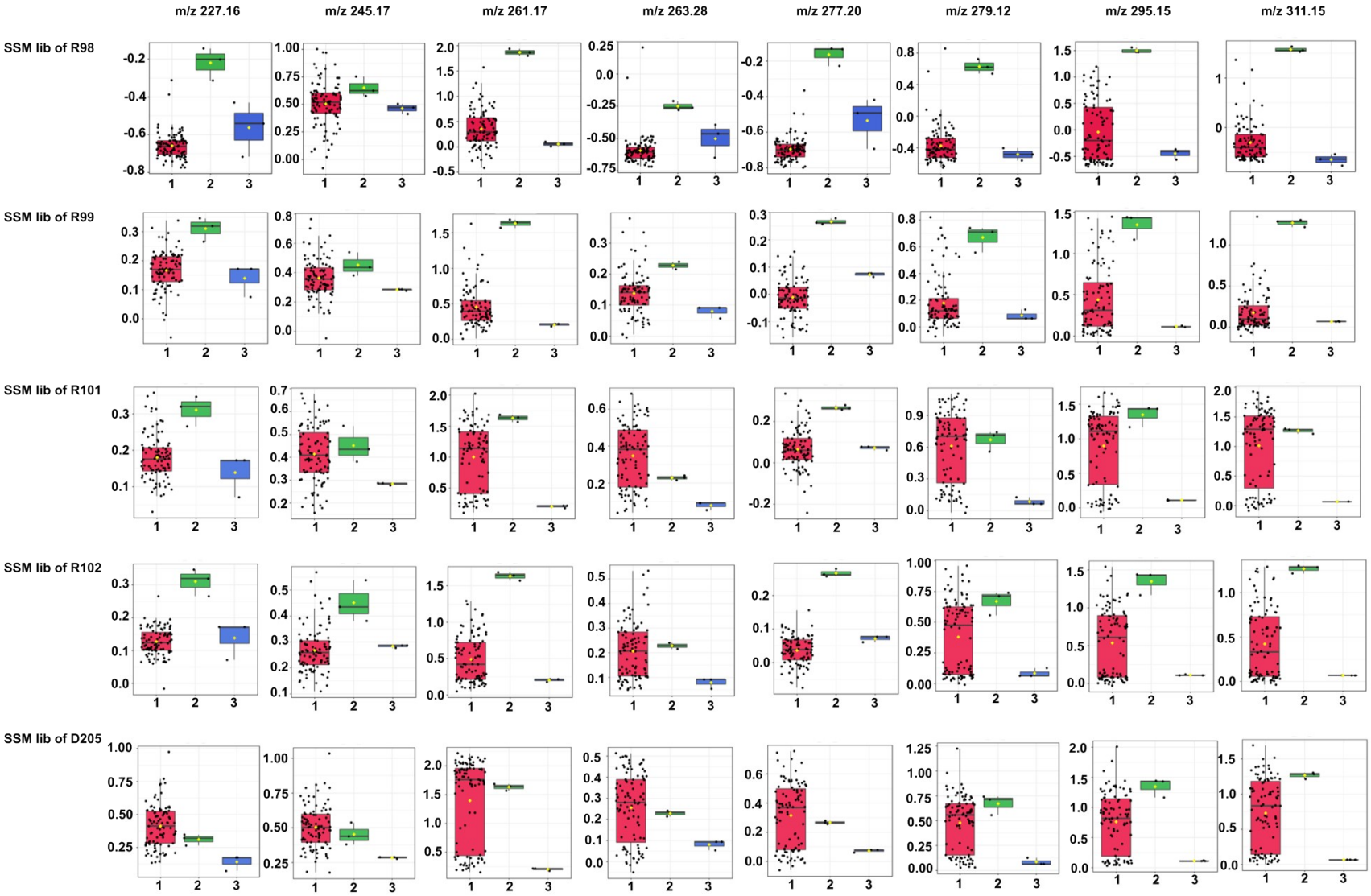


Figure S2. Volcano plots for the metabolite comparisons showing at least 2-fold differential expression between the wildtype and the control group (Fold change > 2.0 and $p < 0.05$).


m/z 227.16
m/z 245.17
m/z 261.17
m/z 263.28
m/z 277.20
m/z 279.12
m/z 295.15
m/z 311.15
m/z 3127.16
SSM lib of L200
SSM lib of L33
SSM lib of V65
SSM lib of L119
SSM lib of L185


m/z 227.16
m/z 245.17
m/z 261.17
m/z 263.28
m/z 277.20
m/z 279.12
m/z 295.15
m/z 311.15
SSM lib of R98
SSM lib of R99
SSM lib of R101
SSM lib of R102
SSM lib of D205

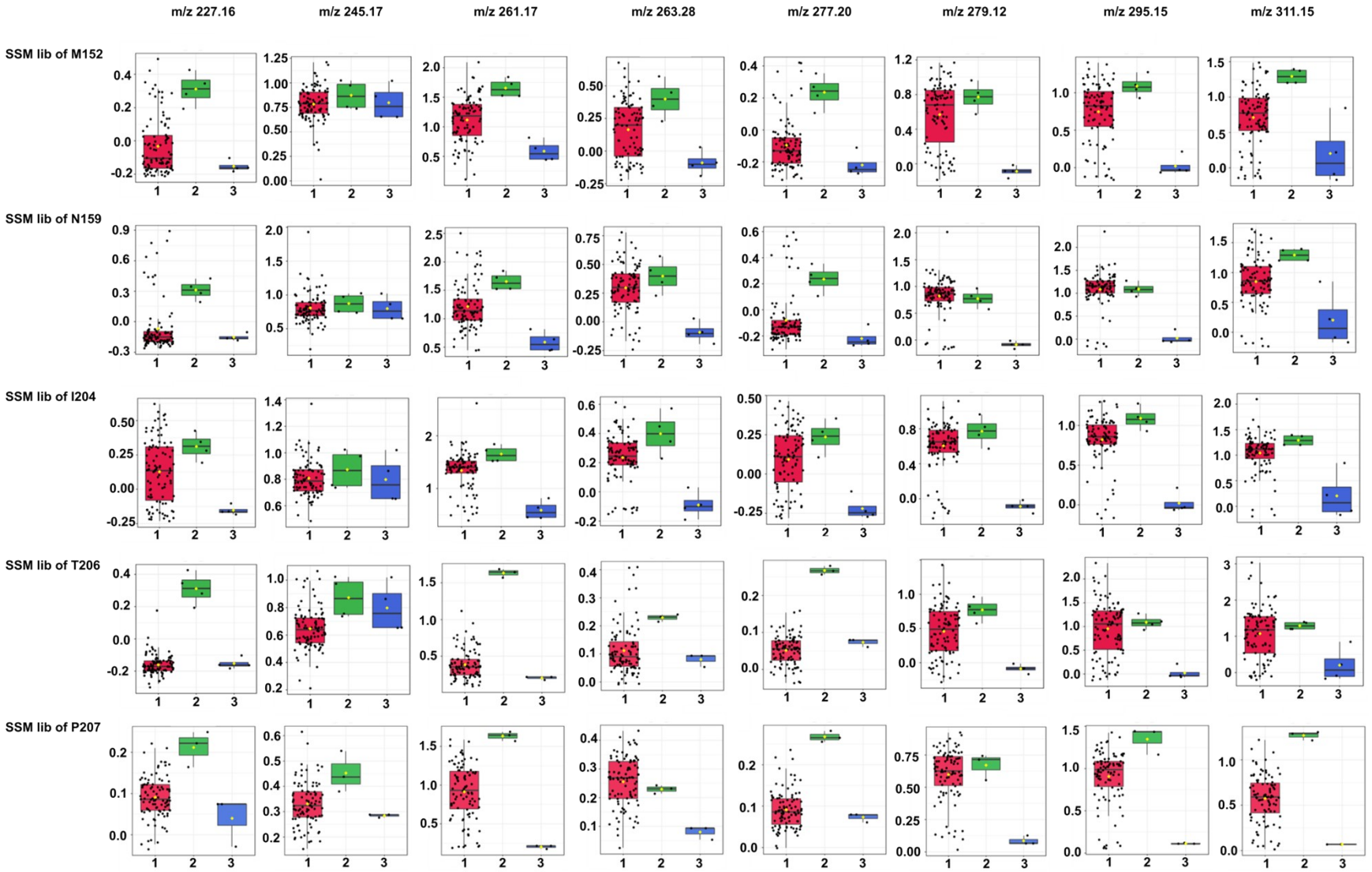


Figure S3. Box-whisker plots of selected peaks signals in each group based on ANOVA analysis. Note: Variants from SSM library (red); Rosetta(DE3)/pET28a-AlbC wildtype (green); Rosetta(DE3)/pET28a as control (blue). The selected peaks include *m/z* 227.16 (cLL), *m/z* 245.17 (cLM), *m/z* 261.17 (cFL), *m/z* 263.28 (cMM), *m/z* 277.16 (cYL), *m/z* 279.12 (cFM), *m/z* 295.15 (cFF or cYM) , *m/z* 311.15 (cFY) and *m/z* 327.16(cYY).

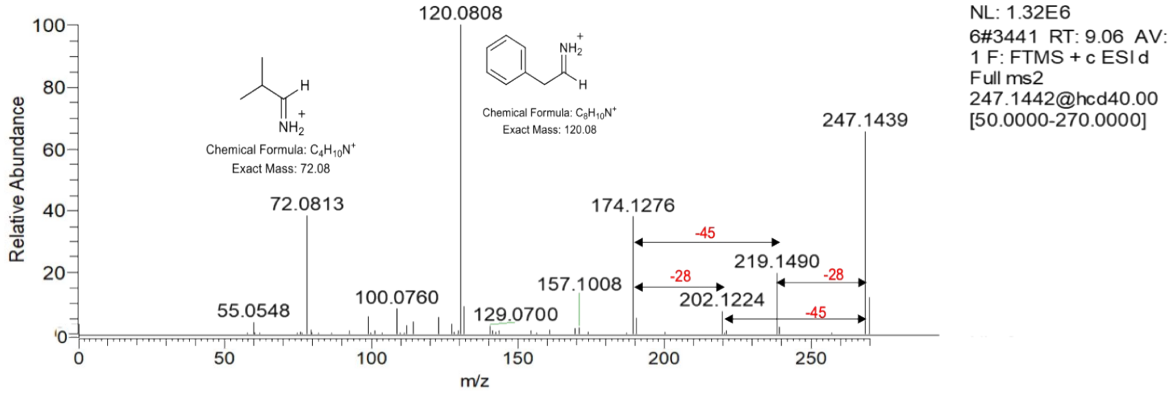


Figure S4. MS-fragmentation spectrum of the F186L-fermentations (cFV). Shown are the characteristic neutral losses of 28 and 45 Da, resulting in the detection of the immonium ions of the respective DKP constituents. Immonium ion structures are shown next to the detected fragment ion.

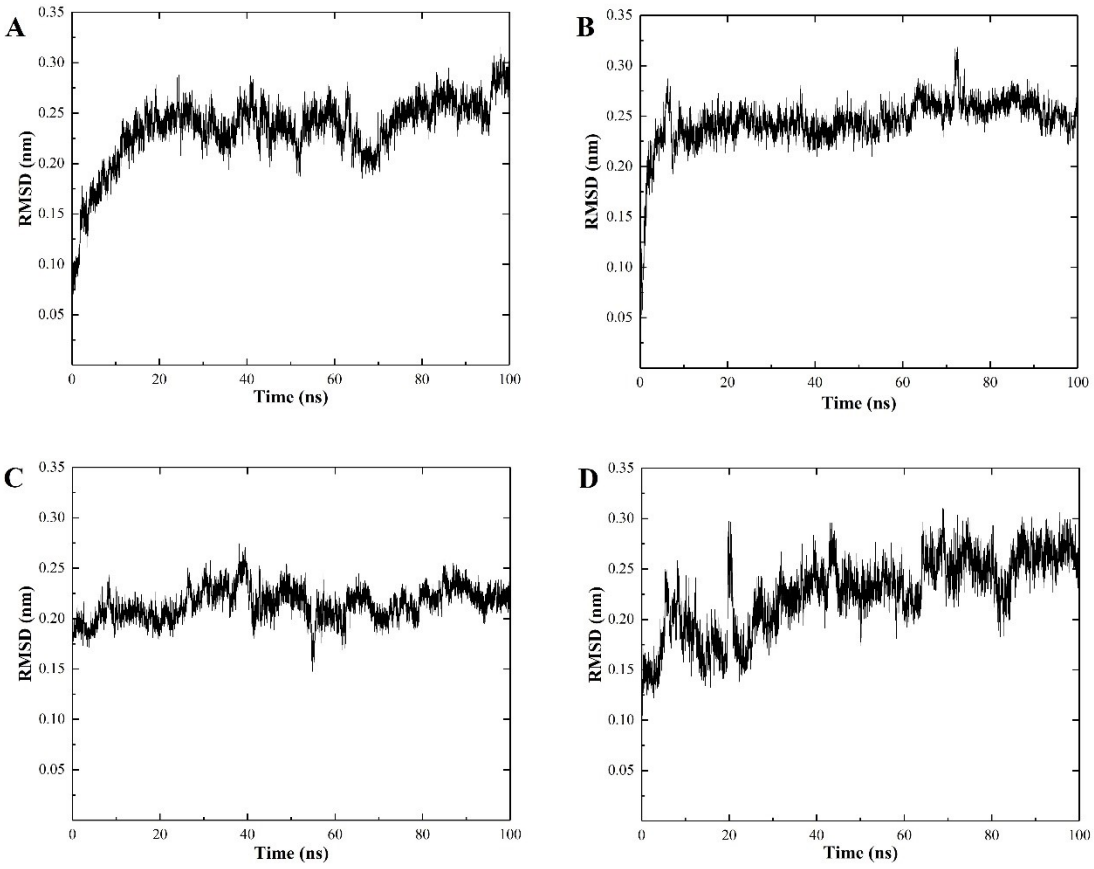


Figure S5. Time evolution of RMSD of WT-cFL (A), F186L-cFL (B), F186L-cFV (C) and T206F (D).

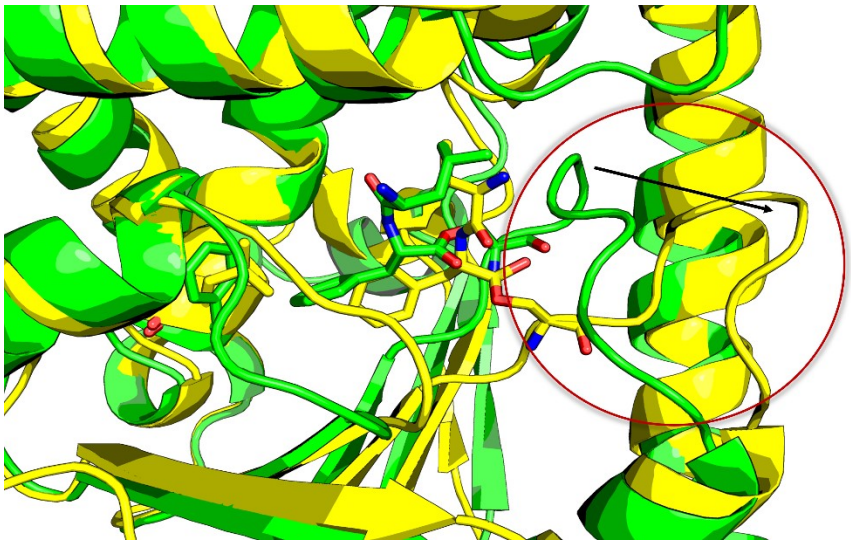


Figure S6. Loop movement in WT and F186L. The cFL and cFV are shown as stick model. WT and F186L are shown in green and yellow, respectively.

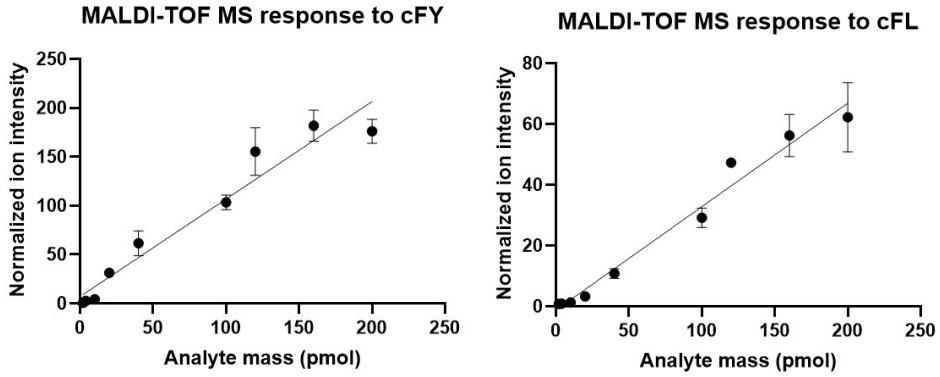


Figure S7. Standard curves of MALDI-ToF MS analyses of two main DKP products.

Supplementary References

1. S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz and J. P. Acevedo, *ACS Synth Biol*, 2013, **2**, 83-92.
2. M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam and M. Gondry, *Nat Commun*, 2014, **5**.
3. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605-1612.
4. E. Schmitt, G. Bourgeois, M. Gondry and A. Aleksandrov, *Sci Rep*, 2018, **8**, 7031.
5. C. R. Sondergaard, M. H. Olsson, M. Rostkowski and J. H. Jensen, *J Chem Theory Comput*, 2011, **7**, 2284-2295.
6. M. H. Olsson, C. R. Sondergaard, M. Rostkowski and J. H. Jensen, *J Chem Theory Comput*, 2011, **7**, 525-537.
7. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J Comput Chem*, 2005, **26**, 1701-1718.
8. J. Huang and A. D. MacKerell, Jr., *J Comput Chem*, 2013, **34**, 2135-2145.

9. S. Jo, X. Cheng, J. Lee, S. Kim, S. J. Park, D. S. Patel, A. H. Beaven, K. I. Lee, H. Rui, S. Park, H. S. Lee, B. Roux, A. D. MacKerell, Jr., J. B. Klauda, Y. Qi and W. Im, *J Comput Chem*, 2017, **38**, 1114-1124.
10. T. Yagasaki, M. Matsumoto and H. Tanaka, *J Chem Theory Comput*, 2020, **16**, 2460-2473.
11. G. C. Jean-Paul Ryckaert, Herman J.C Berendsen, *J Comput Phys*, 1977, **Volume 23, Issue 3**, 327-341.
12. T. Y. Darden, D.; Pedersen, L. , *J Comput Phys*, 1993, **98**, 10089−10092.
13. G. Bussi, D. Donadio and M. Parrinello, *J Chem Phys*, 2007, **126**.
14. M. Parrinello and A. Rahman, *J Appl Phys*, 1981, **52**, 7182-7190.
15. C. H. Wang, D. Greene, L. Xiao, R. X. Qi and R. Luo, *Front Mol Biosci*, 2018, **4**.
16. M. S. Valdes-Tresanco, M. E. Valdes-Tresanco, P. A. Valiente and E. Moreno, *J Chem Theory and Comput*, 2021, **17**, 6281-6291.
17. H. M. A. D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman, *University of California, San Francisco*, 2021.
18. J. R. Wagner, J. Sorensen, N. Hensley, C. Wong, C. Zhu, T. Perison and R. E. Amaro, *J Chem Theory and Comput*, 2017, **13**, 4584-4592.