Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

Tong Si[1,‡], Bin Li[2,3,4,‡], Troy J. Comi[2,3,‡], Yuwei Wu[2], Pingfan Hu[5], Yuying Wu[5], Yuhao Min[2], Douglas A. Mitchell[1,2,6], Huimin Zhao[*,1,2,5,7,8], Jonathan V. Sweedler[*,1,2,3]

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[2]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[3]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[4]State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing, 210009, China

[5]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[6]Department of Microbiology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[7]Department of Biochemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[8]Department of Bioengineering, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

Supporting Information

Experimental

Strains, media and cultivation conditions

All chemicals were purchased through Sigma-Aldrich (St. Louis, MO) or Fisher Scientific (Fisher Scientific, Pittsburgh, PA) unless noted otherwise. Zymo 5α Z-competent *Escherichia coli* (Zymo Research, Irvine, CA) and NEB 10β electrocompetent *E. coli* (New England Biolabs, Ipswich, MA) were used for general plasmid amplification and library construction, respectively. *E. coli* BL21(DE3) (Cell Media Facility, University of Illinois at Urbana-Champaign) was used as a host for the expression of multiple enzymes. For plasmid construction, *E. coli* strains were cultured at 37°C and 250 rpm in Luria broth (LB) liquid media, or at 37°C on LB plates solidified with 1.5% (w/v) agar. For plasmid maintenance using antibiotic selection, LB was supplemented with 100 μg ml^{-1} ampicillin and/or 50 μg ml^{-1} kanamycin. For inducible protein expression, BL21(DE3) cells were cultured at 30°C instead of 37°C, and isopropyl β-D-1-thiogalactopyranoside (IPTG) was supplemented at a final concentration of 1 mM. M9 minimal media supplemented with BME vitamin mix (Sigma-Aldrich, cat. #B6891), trace mineral solution (ATCC, Manassas, VA, cat. #MD-TMS), and 2 g L^{-1} acetate were used for plantazolicin (PZN) production.

DNA and strain construction

The list of primers used can be found in Table S1. All enzymes used for recombinant DNA cloning, including Q5 PCR polymerase and restriction digestion enzymes, were from New England Biolabs unless noted otherwise. Plasmid assembly was performed using a Gibson Assembly Cloning Kit (New England Biolabs) or T4 ligase following the manufacturer's instructions. QIAprep Spin Plasmid Mini-prep Kits (Qiagen, Valencia, CA) were utilized to isolate plasmid DNA from *E. coli*. PCR, digestion and ligation products were purified by QIAquick PCR Purification and Gel Extraction Kits (Qiagen). Error-prone PCR was performed using GeneMorph II Random Mutagenesis Kits (Agilent Technologies, Santa Clara, CA). The genomic DNA of *Pseudomonas aeruginosa* PAO1c was a kind gift from Prof. Joshua D. Shrout at the University of Notre Dame.

For PZN production, the partial operon containing the biosynthetic genes essential for PZN production (*ptnC*, *ptnD*, *ptnB*, *ptnE* and *ptnL*) were PCR-amplified in two pieces from a fosmid bearing the complete PZN pathway[1] with primer pairs NP5/NP6 and NP7/NP8. PCR products and the vector pRSFDuet-1 linearized by *Nde*I and *Mfe*I at the multiple cloning site II were assembled into pRSFDuet-T7-ptnJCDBEL(II), so that expression of the partial operon was under control of a T7 promoter. We previously discovered N-terminal fusion of the maltose-binding protein (MBP) was necessary for efficient production of the precursor peptide,[1] so pET28a-MBP-bamA was used for precursor expression in trans on a separate plasmid. The kanamycin resistance gene in pRSFDuet-T7-ptnJCDBEL(II) was replaced by the ampicillin resistance gene to allow co-selection with pET28a-MBP-bamA, which also harbors a kanamycin resistance cassette. For site-saturation mutagenesis, the NNK degenerative codons at I7 or I8 positions were introduced in primers NP127 and NP128, respectively. The plasmid libraries were assembled using two PCR products amplified from pET28a-MBP-bamA using primer pair #1 (NP119/NP124) and primer pair #2 (NP125/NP127 or NP125/NP128). Gibson assembly products were used to transform NEB 10β cells, which were selected on agar media to obtain $>10^4$ independent transformants for each library. Plasmid DNA was isolated and used to transform BL21(DE3) harboring pRSFDuet-T7-pntJCDBEL for strain library creation.

For monorhamnolipid (mono-RL) production, the *rhlB* gene was PCR-amplified using the primer pair NP21/NP22 from *P. aeruginosa* PAO1c genomic DNA. The PCR product was digested using *Eco*RI and *Hind*III, and ligated into multiple cloning site I of pRSFDuet-1 treated with the same set of enzymes to create pRSFDuet-T7-rhlB(I). The WT *rhlA* gene was PCR-amplified using the primer pair NP23/NP24 from *P. aeruginosa* PAO1c genomic DNA, digested using *Mfe*I and *Kpn*I, and ligated into multiple cloning site II of pRSFDuet-T7-rhlB(I). The resulting plasmid, pRSFDuet-T7-rhlB(I)-T7-rhlA(II), was

used to transform BL21(DE3) to create the 'wild-type' production strain. To introduce random mutagenesis to *rhlA*, error-prone PCR was performed at an average mutation rate of ~1.8 bp kb^{-1} (or 1.6 bp per the 888 bp *rhlA* gene) with the following PCR condition: 800 ng of pRSFDuet-T7-rhlB(I)-T7-rhlA(II) (WT or mutants) as a template, NP23/NP24 as primers, and 30 PCR cycles. The PCR product was digested using *Mfe*I and *Kpn*I before being ligated into pRSFDuet-T7-rhlB(I) treated with the same set of enzymes. Ligation products were used to transform electrocompetent NEB 10β cells, obtaining $>10^5$ independent transformants for each strain library on agar media. Plasmid DNA was isolated and transformed into BL21(DE3) cells for strain library construction.

Microscopy-guided MALDI MS with microMS

To prepare colonies for MALDI-ToF MS profiling, plasmid DNA libraries were used to transform BL21(DE3) cells, which were spread on Durapore™ PVDF membrane filters (0.22 μm pore size, 90 mm diameter, EMD Millipore, Kankakee, IL, cat. #GVWP08050) to allow growth using noninducing LB agar media. After cultivation at 30°C for 16–20 h, the filters were transferred to inducing plates containing 1 mM IPTG (M9+acetate media for PZN, and LB media for RL) for incubation at 30°C for 24 h. To transfer biomass onto MALDI targets, a colony-bearing filter was placed onto a clean, stainless steel substrate with colonies facing upwards. An indium-tin oxide (ITO)-coated glass slide (Delta Technologies, Loveland, CO) was delicately placed on the filter with the ITO coating facing the colonies. Colony patterns were imprinted onto the ITO slide by gently applying ~3.5 N force by hand for 10 s.

High-throughput MALDI-ToF MS screening of *E. coli* colonies was performed following previously reported single-cell profiling workflows[2-4] with modifications to the sample preparation steps. microMS[5], a custom-built image analysis software (available at http://neuroproteomics.scs.illinois.edu/microMS.htm), was employed for automatic MALDI-ToF MS profiling. Specifically, ITO-coated glass slides were etched with more than 16 fiducials surrounding the imprint region. Auto-fluorescence of *E. coli* colonies in the DAPI channel was used to aid colony finding. Whole-slide bright-field and fluorescence images were acquired on a Ziess Axio Imager M2 (Zeiss, Jena, Germany) using an Ab cam Icc5 camera, a HAL 100 halogen illuminator (Zeiss), and an X-CITE Series 120 Q mercury lamp (Lumen Dynamics, Mississauga, Canada). The 31000v2 DAPI filter set was used for auto-fluorescence excitation. The images were acquired as tiled mosaics using the 10× objective and 10% overlap. Images were processed and exported at 1/8 magnification as tiff files using ZEN software version 2 blue edition (Zeiss). The whole-slide tiff images were loaded into microMS, which performed automatic colony finding, target patterning around each colony, and correlation with a Bruker ultrafleXtreme MALDI-ToF/ToF mass spectrometer (Bruker Daltonics, Billerica, MA). Coordinate registration and correlation was performed by locating the etched fiducials in the mass spectrometer and recording their locations in microMS. The procedure is found to be accurate within ~20 μm when at least 16 fiducials are included in the training set.[5] Target patterning was positioned at most 10 targets around each colony, offset from the circumference by 25 pixels (110 μm), with a minimum shot-to-shot spacing of 10 pixels (44 μm). The custom geometry file was then ready to load into the mass spectrometer for automated acquisition.

Before spectra acquisition, the sample slide was coated with MALDI matrix using an artist's airbrush with a 0.2 mm nozzle (Paasche Airbrush Company, Chicago, IL). Several parameters were optimized to ensure even deposition of matrix on the imprinted colonies. The N_2 pressure was set at 40 psi and the imprint glass slides were spray-coated at a distance of 30–35 cm with 50 mg mL^{-1} 2,5-dihydroxybenzoic acid (DHB) (Sigma-Aldrich) dissolved in methanol:H_2O (1:1, v/v). After spraying 2 mL of the DHB solution with the airbrush, the sample was dried for 1 min to avoid over-wetting and analyte delocalization. A total of 10 mL of DHB solution was applied in 10 min per target plate.

Measurements were performed using a frequency tripled Nd:YAG solid state laser (λ=355 nm). The laser footprint was set to "Ultra" at a ~100 μm diameter. Mass spectrometer calibration was performed using Peptide Calibration Standard Kit II (Bruker Daltonics). Data acquisition was run in positive reflection

mode with pulsed ion extraction and a mass range of 440–700 Da for the detection of RLs and 905–2005 Da for PZN variants. Colony imprints were analyzed following the custom geometry file with 500 laser shots fired at 2000 Hz. Resulting spectra were analyzed as detailed below.

Multivariate data analysis

Mass spectra from this study have been deposited at OpenMSI (https://openmsi.nersc.gov/openmsi/client/) for open access. Spectra were read directly into MATLAB 2015b with the readbrukermaldi function (github.com/AlexHenderson/readbrukermaldi) and manually recalibrated with a third order polynomial to correct for mass shifts between analyses. Spectra were resampled with bin widths of 0.025 Da for RL and 0.1 Da for PZN.

For PZN, untargeted t-SNE[6] was performed utilizing each binned *m/z* value to evaluate population heterogeneity and variance in sample processing. It was determined that mutations resulted in mass shifts from the fully processed form of PZN while other clusters were due to experimental factors, including polymer contamination. A list of theoretical *m/z* values of monoisotopic masses were produced for each possible amino acid substitution by shifting the monoisotopic mass of PZN. Next, targeted t-SNE was performed utilizing just the intensity of the theoretical masses to locate colonies expressing each point mutation. t-SNE was performed with initial dimensionality reduction of 30 principal components and a perplexity of 30. The maximum intensity of each monoisotopic amino acid substitution was extracted from every spectrum with a tolerance of ±0.25 Da. The reduced dataset consisted of an array of 18 intensity values (20 amino acids, with L/I and Q/K unresolved), which was examined with t-SNE to cluster similar spectra. Apparent groups were examined manually to assign spectra to specific substitutions. Spectra without peptide signal were combined into the "N/A" cluster. These correspond to background, imaging artifacts, and spectra acquired on non-circular colonies. Additional filtering of putative colonies at the stage of optical image analysis could reduce the abundance of the N/A cluster. Next, the spectral classifications were mapped onto the optical image, leveraging the pixel positions encoded into the filename of each spectrum. To facilitate mutant recovery, the most common cluster is shown over the optical image, excluding the N/A cluster. Pseudocode is provided below as a general outline for the t-SNE analysis, and the MATLAB implementation is available upon request. The User's Guide for the tsne function can be found at https://lvdmaaten.github.io/tsne/User_guide.pdf.

```
mass spectra <- read in from data file, size n
mass list <- read in or enter m/z values of interest (PZN – (m/z of Ile) + (m/z of other AAs))
tolerance <- input (tolerance of extracted intensities, here we use ±0.25 m/z)
perform processing on mass spectra collection (recalibration/alignment, binning, normalization, etc)

initialize targeted data set
for each mass spectrum in mass spectra
        for each m/z value in mass list
                targeted data at mass spectrum and m/z value = maximum intensity of mass spectra in the
                interval of m/z value ± tolerance

tsne data = tsne of targeted data set, a 2 by n array of x,y coordinates in tsne "space".
plot tsne data
select or cluster points and examine average spectra
```

For RLs, a targeted analysis was performed to visualize the relative abundance of RL **5b**. The intensities of protonated, sodiated, and potassiated monoisotopic masses of RLs **5a**, **5b**, **5c**, and **5d** were extracted with a tolerance of ±0.2 Da. Spectra were filtered to remove colonies within 200 μm of each other and a total RL intensity of less than 500 (arbitrary counts). The intensities of each RL were summed for spectra

surrounding each colony. To visualize the molecular content of mutants on the target, the total RL intensity and relative content of RL **5b** were mapped onto the optical image of the colonies. The log base 10 of the total abundance of RLs determined the size of each data point overlaid on the whole slide image. The relative abundance of RL **5b** dictated the color of each point. Such visualization allows rapid assessment of desirable mutants in terms of total expression and detection of outliers.

***In situ* high-resolution and MS/MS analysis for PZN analogues**

In situ measurement of the accurate mass of peptides was conducted on colony imprints using a 7T solariX Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics) equipped with a dual ESI/MALDI source and a Smartbeam II laser. Mass calibrations were performed externally using DHB and Peptide Calibration Standard Kit II (Bruker Daltonics). An *m/z* range of 150–3000 was acquired at 4 Mword. Data were analyzed in Data Analysis version 4.0 software (Bruker Daltonics). *In situ* MS/MS was conducted on the colony imprints using the MALDI ToF/ToF LIFT mode of the mass spectrometer under manual control. Tandem mass spectra were smoothed, baseline-corrected, and analyzed in FlexAnalysis 3 (Bruker Daltonics).

Relative quantification of RL congeners using LC-MS/MS with multiple reaction monitoring (MRM) mode

To compare direct MALDI-ToF MS profiling results with LC-MS or MALDI-ToF MS quantification of RL congeners following organic solvent extraction (Figure S7), a filter bearing ~100 colonies with the WT mono-RL pathway subsequent to IPTG induction as described above was extracted using 500 µl of chloroform:ethanol (2:1, v/v) solution. The filter and cell debris were separated from supernatant using centrifugation (5 min, 10,000 rpm, 4°C), and the lower organic phase was dried under vacuum in a rotary flash evaporator (MiVac, GeneVac, UK). The samples were reconstituted in 50 µL of acetonitrile:H_2O (1:9, v/v) solution containing 2 mM ammonium acetate. For MALDI-ToF MS analyses, 1 µl of the reconstituted extracts was mixed with 1 µl of DHB solution (50 mg ml^{-1} in acetonitrile:H_2O (1:1, v/v)) and spotted onto an MTP 384 polished steel target (Bruker Daltonics). MALDI spectra were acquired in positive reflection mode as described above. LC-MS quantification was performed as detailed below. The same extraction and quantification procedure was also used to analyze the colonies of mutant strains subsequent to plasmid retransformation (Figure S7).

To characterize the ratios of RL congeners in liquid fermentation products, single colonies were obtained by streaking glycerol frozen stocks of the WT and mutant RL-producing strains on agar plates. Three colonies for each strain were inoculated into 3 mL of LB+kanamycin media. Following cultivation at 30°C and 250 rpm for 16 h, 60 µL of the cell cultures were added to 3 mL of fresh LB+kanamycin media to continue growth until cell densities reached OD_{600}=0.4~0.8. IPTG was added to a final concentration of 1 mM, and induction was performed at 30°C and 250 rpm for 24 h. Culture supernatants were obtained via centrifugation for cell separation (10 min, 4,000 rpm, 20°C), and were filtered through 0.22 µm-pore-size cellulose acetate membrane centrifuge tube filters (Sigma-Aldrich) before LC-MS analyses, as detailed below.

An MRM assay was performed in negative-ion mode using an ultrahigh performance liquid chromatography–triple quadrupole–electrospray ionization mass spectrometry (UHPLC-QqQ-ESI MS) system (Bruker Daltonics) consisting of an Advance UHPLC module and an EVOQ Elite triple quadrupole mass spectrometer. A Kinetex 1.7 µm C18 150 × 2.1 mm internal diameter column (Phenomenex, Torrance, CA) was used for LC analyte separation. Mobile phase A was H_2O containing 4 mM ammonium acetate and mobile phase B was acetonitrile. The gradient program was conducted as follows: 0–2 min, 5% B; 2–2.1 min, 5–50% B; 2.1–8 min, 50–90% B; 8–15 min, 90% B; 15–15.1 min, 90–5% B; 15.1–17 min, 5% B. Total run time was 17 min. The injection volume was 2 µL. The EVOQ source parameters were as follows: HESI, spray voltage (-) 4500 V; cone temperature, 250°C; cone gas

flow, 25; heated probe temperature, 450°C; probe gas flow, 45; nebulizer gas flow, 65; exhaust gas, Off. Monitored MRM transitions for RL **5a** were 475→305, 475→169; RL **5b** were 503→333, 503→169; RL **5c** were 531→333, 531→169, and RL **5d** were 559→395, 559→169. The *m/z* 169 product intensity was utilized as a quantitation transition while the other transitions provided confirmation of RL identities. EVOQ MRM chromatograms were analyzed using Data Review 8.2 (Bruker Daltonics). The peak area of quantitation transition of a specific RL congener was used to calculate its fraction relative to the sum of all RL peak areas.

Comparison with liquid-handling approaches

Here we provide a simple comparison of the estimated time and cost for screening between the optically guided MALDI-ToF method and the use of our robotic platform[7] containing an Evo200 liquid handling robot (https://www.youtube.com/watch?v=Hwb735qZ-IQ). As library construction and MALDI-ToF MS analysis are almost the same for both approaches, only sample preparation procedures are considered in this comparison.

Using a liquid-handling robot, the major consumables are disposable tips ($8 per 96) and 96-well plates ($7 per plate; in our experience, 96-well plates provide more consistent microbial cultivation than 384-well pates). The colony-picking add-on module can handle up to 800 colonies per hour, costing $0.15 and taking 4.5 s per colony. Re-inoculation and induction requires three liquid handling steps and one spectrometric measurement, adding an additional $0.31 and ~0.5 s per colony. Different methods will be used to extract target analytes from cells (i.e., plantazolicin) or liquid media (i.e., rhamnolipid), which is necessary to reduce salt concentrations, and each crude extract will be mixed with MALDI matrices and patterned on 384-well MALDI targets for MS analyses. These procedures require three liquid-handling steps and two centrifugation steps, costing $0.24 and taking ~9 s per colony. Additionally, the use of organic solvents in liquid-handling robots needs fine-tuning of pipetting parameters and results in long-term corrosion, and additional costs are associated with frequent cleaning (40 min) and replacement of expensive MALDI targets ($600 each for 3,840 total samples, or $0.16 per sample). In total, sample preparation on the liquid-handling platform costs $0.86 and takes ~14 s per colony.

Alternatively, for the optically guided method, a $3 filter can hold up ~2,000 colonies, which will be analyzed on two disposable ITO slides ($5 per slide). Transferring colonies to inducing plates takes negligible time, whereas imprinting, airbrush matrix coating, and whole-slide imaging take 2 min, 10 min and 1 h per slide, respectively. In total, sample preparation costs $0.0065 and takes 4.3 s per colony. This workflow also eliminates the need for cleaning and replacing 384-well MALDI targets. Moreover, as discussed in the main text, we are developing advanced machine vision approaches to replace whole-slide microscopy with an image acquired on a cell phone, potentially reducing the time of sample preparation to 0.9 s per colony. Notably, for certain applications, the optically guided method is used to accelerate, not to eliminate, the microtiter plate-based screening. For example, if the ultimate objective is large-scale fermentation, it is necessary to validate the improved phenotypes observed from colony screening in liquid cultures at different scales, including microtiter plates, flasks, and small-scale fermenters. In this way, the optically guided method serves as a rapid survey of the whole libraries to narrow down promising mutants for more expensive, intensive follow-up confirmation, to a level that is manageable for manual operations without liquid-handling robots. This practice was implemented in both the cases of PZN and RL for plasmid DNA isolation and liquid fermentation, respectively.

Table S1. DNA sequences in this study.

Primers		
NP5	T7-PZN-2 For	tataagaaggagatatacatgtgaaaattcactacatgggag
NP6	T7-PZN-2 Rev	tcgcgtggccggccgatatctcacgtatacctttttgtttttttataatcc
NP7	PZN-2-mid For	gatgtgaattcttctccgag
NP8	PZN-2-mid Rev	ctcggagaagaattcacatc
NP119	ptnA-I7 Rev	tgtggtacaggtacagcgtg
NP124	pET28-Mid For2	catcctgcgatgcagatccggaacataatggtg
NP125	pET28-Mid Rev2	ctgcatcgcaggatgctgc
NP127	ptnA-I7-NNK For	cacgctgtacctgtaccacannkatctctagttcatctacgttttaagcgg
NP128	ptnA-I8-NNK For	cacgctgtacctgtaccacaatcnnktctagttcatctacgttttaagcgg
RL021	pRSF-MCSI-RhlB For	atcaccacagccaggatccgaattcgatgcacgccatcctcatc
RL022	MCSI-RhlB Rev	ttaagcattatgcggccgcaagctttcaggacgcagccttcag
RL023	MCSII-RhlA For	atatacatatggcagatctcaattggatgcggcgcgaaagtctg
RL024	MCSII-RhlA Rev	tttaccagactcgagggtacctcaggcgtagccgatggc
The wild-type (yellow) and mutated (green) sequences of *rhlA*		
WT	atgcggcgcgaaagtctgttggtatcggtttgcaagggcctgcgggtacatgtcgagcgcgttgggcaggatcccgggcgcagca cggtgatgctggtcaacggcgcgatggcgaccaccgcctcgttcgcccggacctgcaagtgcctggccgaacatttcaacgtggt gctgttcgacctgcccttcgccgggcagtcgcgtcagcacaacccgcagcgcgggttgatcaccaaggacgacgaggtggaaat cctcctggcgctgatcgagcgcttcgaggtcaatcacctggtctccgcgtcctggggcggtatctccacgctgctggcgctgtcgc gcaatccgcgcggcatccgcagctcggtggtgatggcattcgcccctggactgaaccaggcgatgctcgactacgtcgggcggg cgcaggcgctgatcgagctggacgacaagtcggcgatcggccatctgctcaacgagaccgtcggcaaatacctgccgcagcgc ctgaaagccagcaaccatcagcacatggcttcgctggccaccggcgaatacgagcaggcgcgctttcacatcgaccaggtgctg gcgctcaacgatcggggctacttggcttgcctggagcggatccagagccacgtgcatttcatcaacggcagctgggacgaataca ccaccgccgaggacgcccgccagttccgcgactacctgccgcactgcagtttctcgcgggtggagggcaccgggcatttcctcg acctggagtccaagctggcagcggtacgcgtgcaccgcgccctgctcgagcacctgctgaagcaaccggagccgcagcgggc ggaacgcgcggcgggattccacgagatggccatcggctacgcctga
R1#15 A64V	atgcggcgcgaaagtctgttggtatcggtttgcaagggcctgcgggtacatgtcgagcgcgttgggcaggatcccgggcgcagca cggtgatgctggtcaacggcgcgatggcgaccaccgcctcgttcgcccggacctgcaagtgcctggccgaacatttcaacgtggt gctgttcgacctgcccttcgtcgggcagtcgcgtcagcacaacccgcagcgcgggttgatcaccaaggacgacgaggtggaaat cctcctggcgctgatcgagcgcttcgaggtcaatcacctggtctccgcgtcctggggcggtatctccacgctgctggcgctgtcgc gcaatccgcgcggcatccgcagctcggtggtgatggcattcgcccctggactgaaccaggcgatgctcgactacgtcgggcggg cgcaggcgctgatcgagctggacgacaagtcggcgatcggccatctgctcaacgagaccgtcggcaaatacctgccgcagcgc ctgaaagccagcaaccatcagcacatggcttcgctggccaccggcgaatacgagcaggcgcgctttcacatcgaccaggtgctg gcgctcaacgatcggggctacttggcttgcctggagcggatccagagccacgtgcatttcatcaacggcagctgggacgaataca ccaccgccgaggacgcccgccagttccgcgactacctgccgcactgcagtttctcgcgggtggagggcaccgggcatttcctcg acctggagtccaagctggcagcggtacgcgtgcaccgcgccctgctcgagcacctgctgaagcaaccggagccgcagcgggc ggaacgcgcggcgggattccacgagatggccatcggctacgcctga
R1#6 V10I	atgcggcgcgaaagtctgttggtatcgatttgcaagggcctgcgggtacatgtcgagcgcgttgggcaggatcccgggcgcagca cggtgatgctggtcaacggcgcgatggcgaccaccgcctcgttcgcccggacctgcaagtgcctggccgaacatttcaacgtggt gctgttcgacctgcccttcgtcgggcagtcgcgtcagcacaacccgcagcgcgggttgatcaccaaggacgacgaggtggaaat cctcctggcgctgatcgagcgcttcgaggtcaatcacctggtctccgcgtcctggggcggtatctccacgctgctggcgctgtcgc gcaatccgcgcggcatccgcagctcggtggtgatggcattcgcccctggactgaaccaggcgatgctcgactacgtcgggcggg cgcaggcgctgatcgagctggacgacaagtcggcgatcggccatctgctcaacgagaccgtcggcaaatacctgccgcagcgc

	ctgaaagccagcaaccatcagcacatggcttcgctggccaccggcgaatacgagcaggcgcgctttcacatcgaccaggtgctg gcgctcaacgatcggggctacttggcttgcctggagcggatccagagccacgtgcatttcatcaacggcagctgggacgaataca ccaccgccgaggacgcccgccagttccgcgactacctgccgcactgcagtttctcgcgggtggagggcaccgggcatttcctcg acctggagtccaagctggcagcggtacgcgtgcaccgcgccctgctcgagcacctgctgaagcaaccggagccgcagcgggc ggaacgcgcggcgggattccacgagatggccatcggctacgcctga
R2#71 V10I L269I	atgcggcgcgaaagtctgttggtatcgatttgcaagggcctgcgggtacatgtcgagcgcgttgggcaggatcccgggcgcagca cggtgatgctggtcaacggcgcgatggcgaccaccgcctcgttcgcccggacctgcaagtgcctggccgaacatttcaacgtggt gctgttcgacctgcccttcgtcgggcagtcgcgtcagcacaacccgcagcgcgggttgatcaccaaggacgacgaggtggaaat cctcctggcgctgatcgagcgcttcgaggtcaatcacctggtctccgcgtcctggggcggtatctccacgctgctggcgctgtcgc gcaatccgcgcggcatccgcagctcggtggtgatggcattcgcccctggactgaaccaggcgatgctcgactacgtcgggcgggg cgcaggcgctgatcgagctggacgacaagtcggcgatcggccatctgctcaacgagaccgtcggcaaatacctgccgcagcgc ctgaaagccagcaaccatcagcacatggcttcgctggccaccggcgaatacgagcaggcgcgctttcacatcgaccaggtgctg gcgctcaacgatcggggctacttggcttgcctggagcggatccagagccacgtgcatttcatcaacggcagctgggacgaataca ccaccgccgaggacgcccgccagttccgcgactacctgccgcactgcagtttctcgcgggtggagggcaccgggcatttcctcg acctggagtccaagctggcagcggtacgcgtgcaccgcgccctgatcgagcacctgctgaagcaaccggagccgcagcgggc ggaacgcgcggcgggattccacgagatggccatcggctacgcctga

Scheme S1. Main biosynthetic steps of PZN 1. The leader cleavage site in the precursor peptide is denoted by the asterisk. The main biosynthetic enzymes include a trimeric heterocycle synthetase consisting of a dehydrogenase (B) and two cyclodehydratase (C/D), a putative leader peptidase (E), and a methyltransferase (L). (Adapted from ref.[8] with permission. Copyright 2017 American Chemical Society).

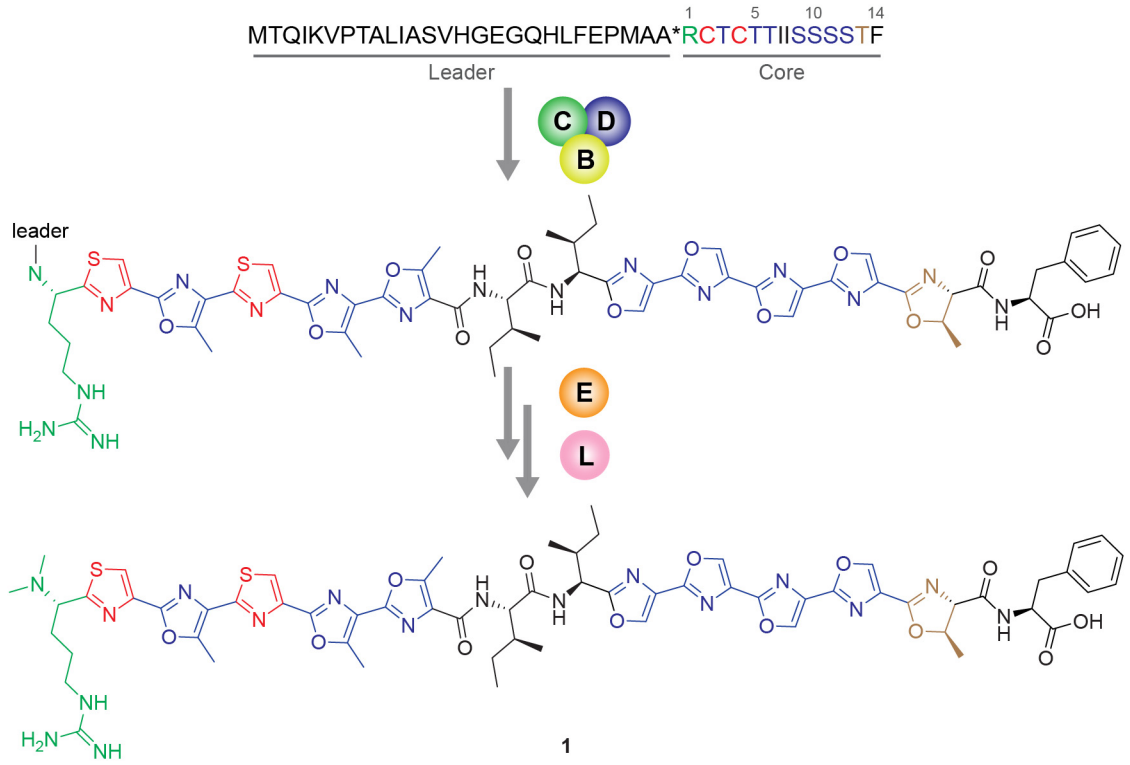
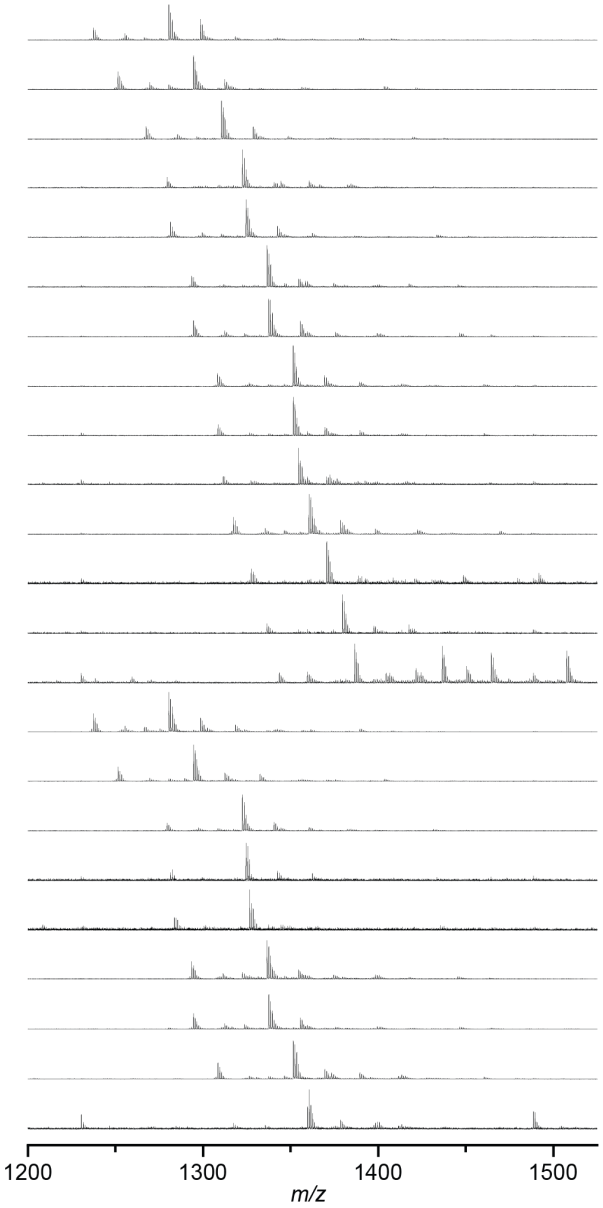


Figure S1. Summary of PZN 1 analogues observed in this study. For each PZN **1** analogue, a typical MALDI-ToF mass spectrum of the *E. coli* colony is included on the left. The table on the right summarizes the information of PZN **1** analogues including the amino acid (AA) mutation assignment based on mass spectra, theoretical and measured monoisotopic *m/z* value of the $[M+H]^+$ ion (by MALDI-ToF or MALDI-FT-ICR), and DNA sequencing results. High-resolution FT-ICR was performed on select colonies from the same sample target subsequent to MALDI MS screening, focusing on tentative Q and K mutants, as well as select **1** analogues (shown in red) not reported in a previous study.[1]



AA mutation by MS	Theoretical monoisotopic *m/z* $[M+H]^+$	TOF *m/z* $[M+H]^+$	FT-ICR *m/z* $[M+H]^+$	FT-ICR mass error (ppm)	DNA sequencing	
					Codon	AA
I7G	1280.41490	1280.452			GGT	I7G
I7A	1294.43055	1294.469			GCT	I7A
I7S	1310.42546	1310.541	1310.42424	-0.93	AGT	I7S
I7V	1322.46185	1322.528	1322.46206	0.16	GTT	I7V
I7T	1324.44111	1324.544	1324.44063	-0.37	ACG	I7T
I7L	1336.47750	1336.590	1336.47823	0.55	CTG	I7L
I7N	1337.43636	1337.512			AAT	I7N
I7Q	1351.45201	1351.389	1351.45219	0.13	CAG	I7Q
I7K	1351.48840	1351.571	1351.48848	0.06	AAG	I7K
I7M	1354.43392	1354.522	1354.43408	0.12	ATG	I7M
I7H	1360.45235	1360.441			CAT	I7H
I7F	1370.46185	1370.587			TTT	I7F
I7R	1379.49455	1379.576	1379.49482	0.20	AGG	I7R
I7Y	1386.45676	1386.520	1386.45743	0.48	TAT	I7Y
I8G	1280.41490	1280.521	1280.41506	0.13	GGG	I8G
I8A	1294.43055	1294.533			GCT	I8A
I8V	1322.46185	1322.497			GTT	I8V
I8T	1324.44111	1324.544	1324.44065	-0.35	ACT	I8T
I8C	1326.40262	1326.520			TGT	I8C
I8L	1336.47750	1336.536			CTT	I8L
I8N	1337.43636	1337.538			AAT	I8N
I8Q	1351.45201	1351.510	1351.45201	0.81	CAG	I8Q
I8H	1360.45235	1360.596			CAT	I8H

Figure S2. *In situ* MALDI-ToF/ToF tandem mass spectra of PZN-I7V, PZN-I7T and PZN-I8V. These analogues were reported previously[1,9] with tandem mass spectra (denoted as blue). Peaks with *m/z* values consistent with previous results[1] are labeled.

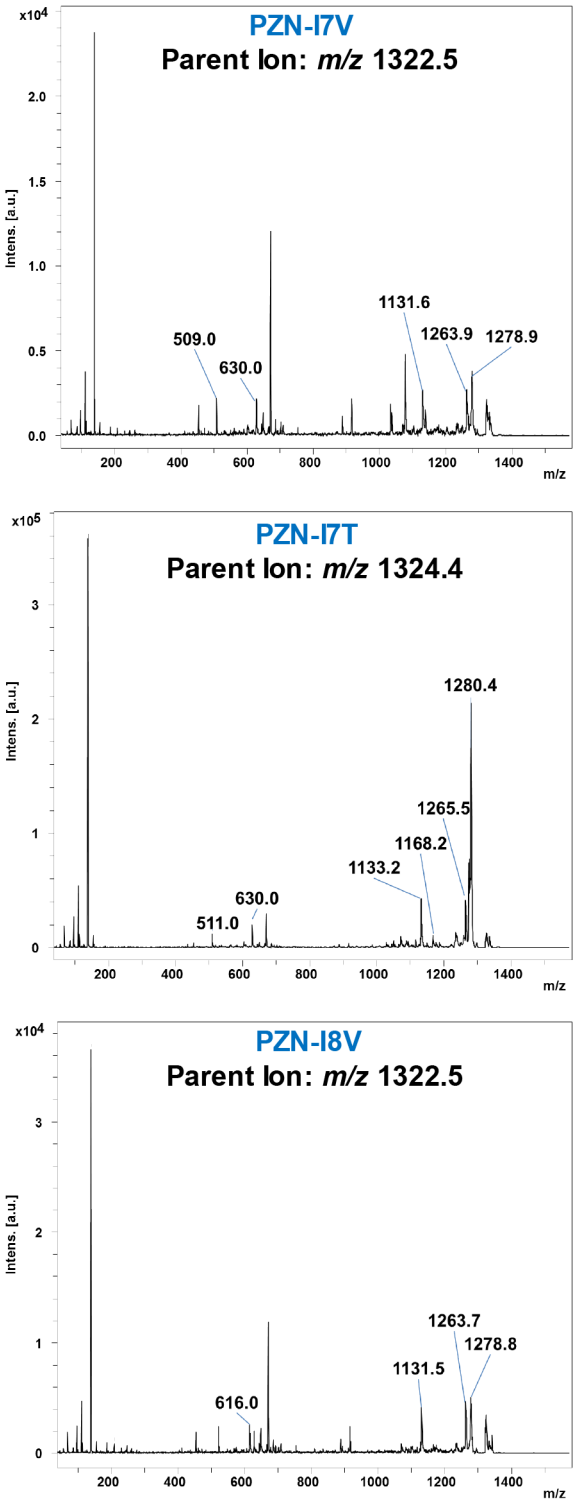


Figure S3. *In situ* MALDI-ToF/ToF tandem mass spectra of (A) PZN-I7L and (B) PZN-I7S. These analogues were reported previously[1] without tandem mass spectra (denoted as orange). Fragments derived from multiple bond cleavages are denoted by asterisks.

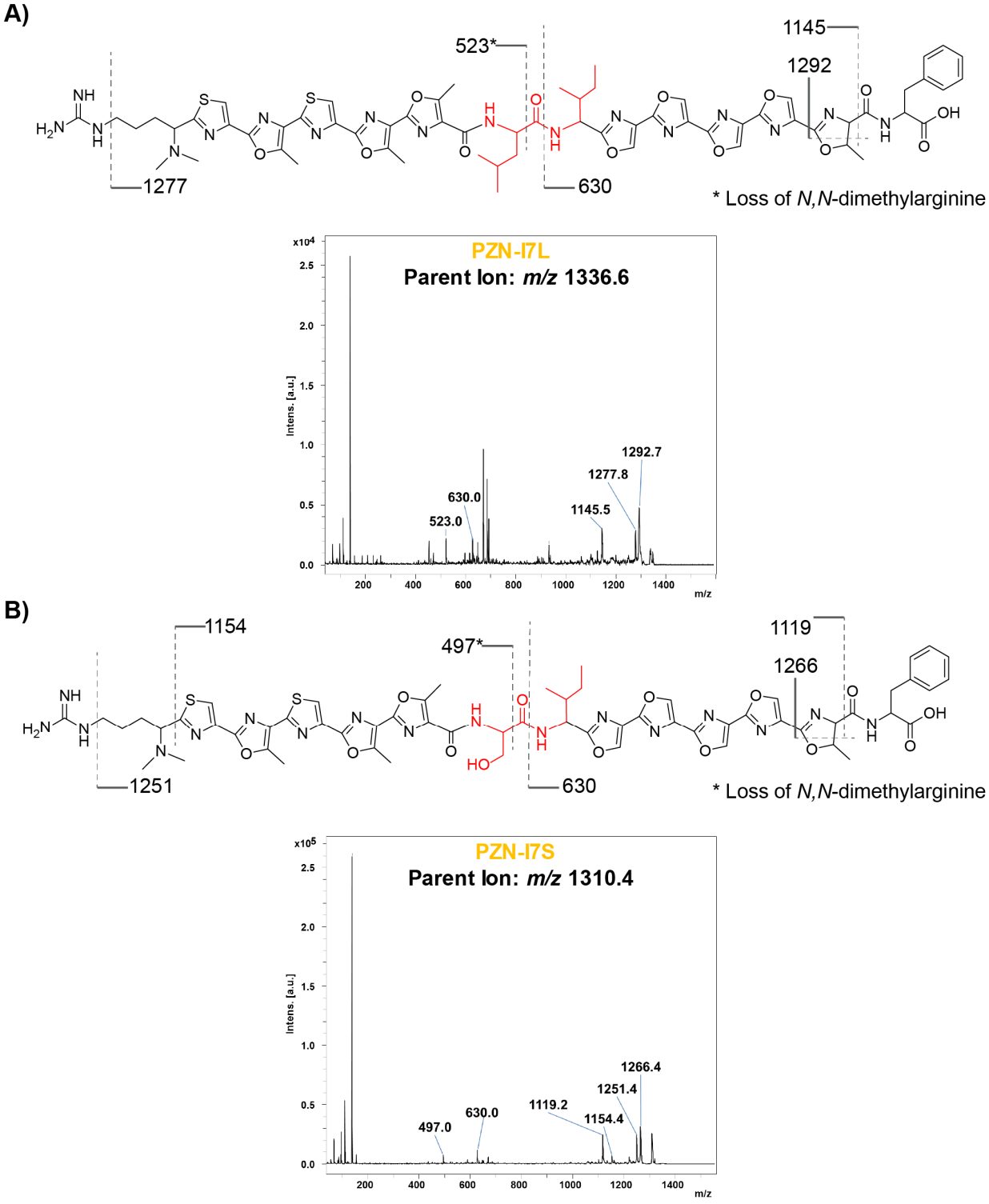


Figure S4. *In situ* MALDI-ToF/ToF tandem mass spectra of (A) PZN-I7K and (B) PZN-I7Q. These analogues were not reported previously[1] (denoted as red).

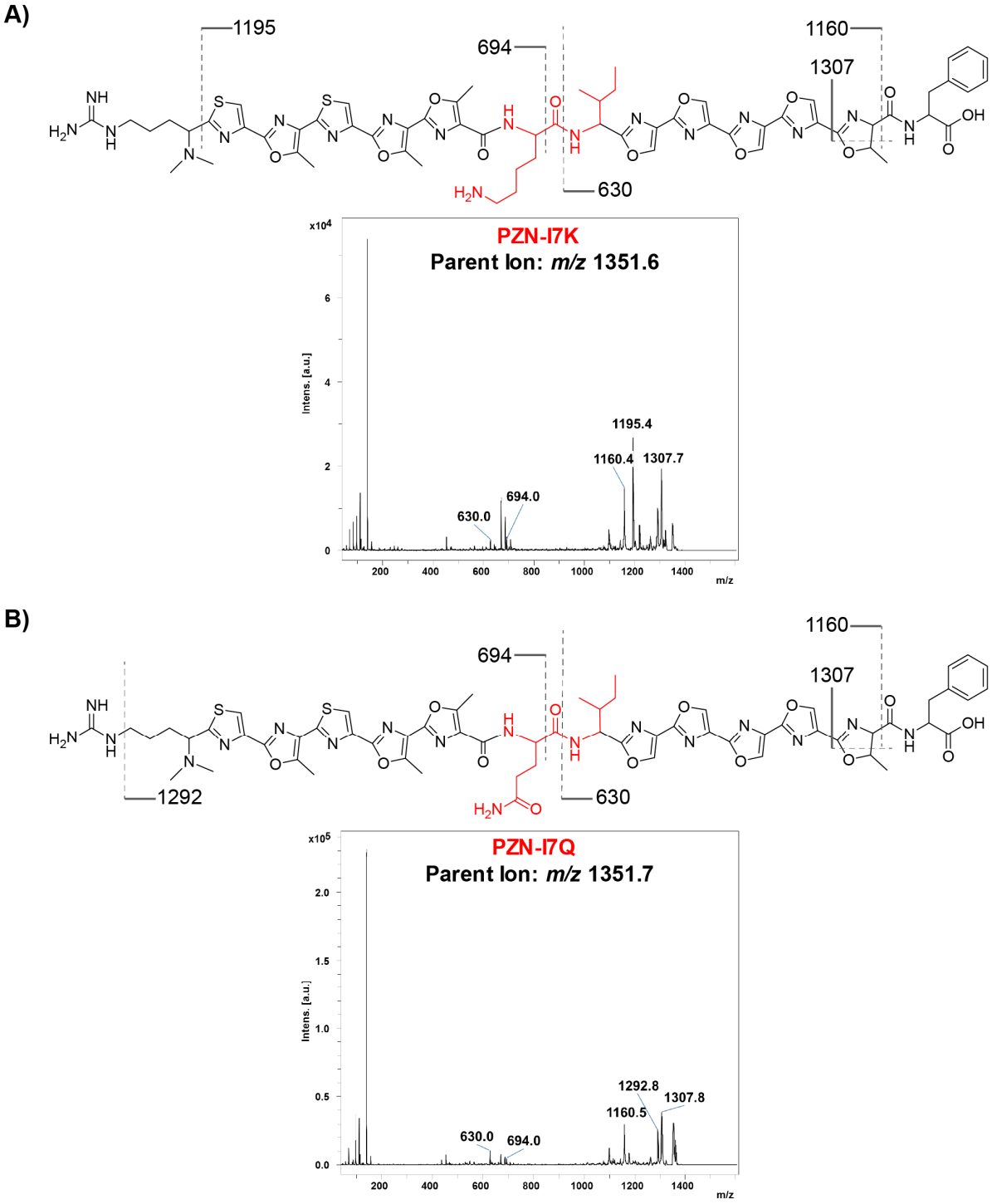


Figure S5. *In situ* MALDI-ToF/ToF tandem mass spectra of (A) PZN-I7Y and (B) PZN-I8T. These analogues were not reported previously[1] (denoted as red). Fragments derived from multiple bond cleavages are denoted by asterisks.

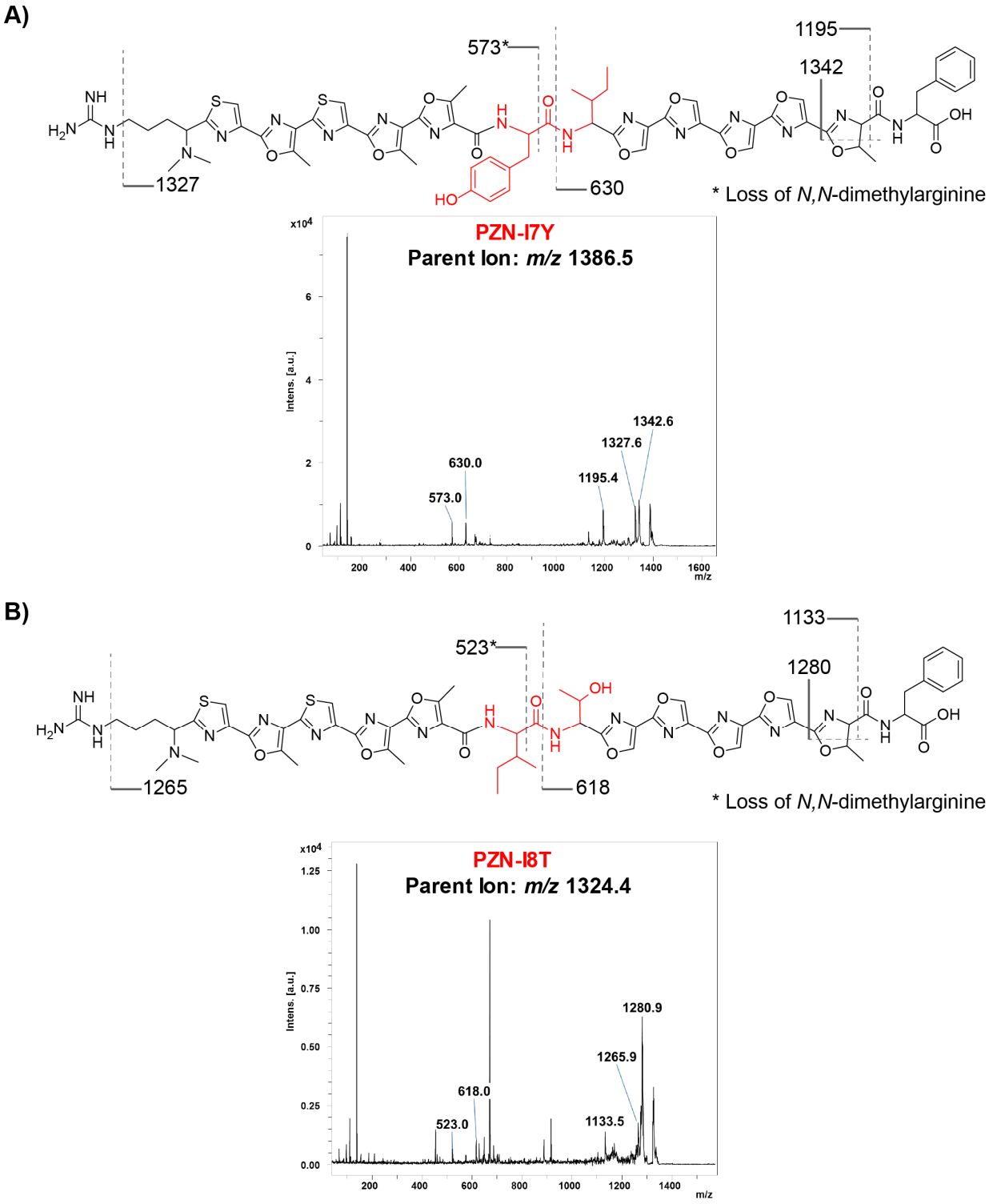


Figure S6. MALDI mass spectra of mono-RLs produced from *E. coli* colonies. (A) Typical colony-MALDI-ToF mass spectra from WT and a mutant strain. The mutant produced greater portions of RL **5c** and RL **5d** relative to WT. Tentative peak assignments of RLs **5a–d** are labeled with dashed lines. (B) *In situ* LIFT ToF/ToF analysis on tentative sodiated ions of RLs **5a–d**. Fragment ions with *m/z* values consistent with previous reports[10,11] are labeled.

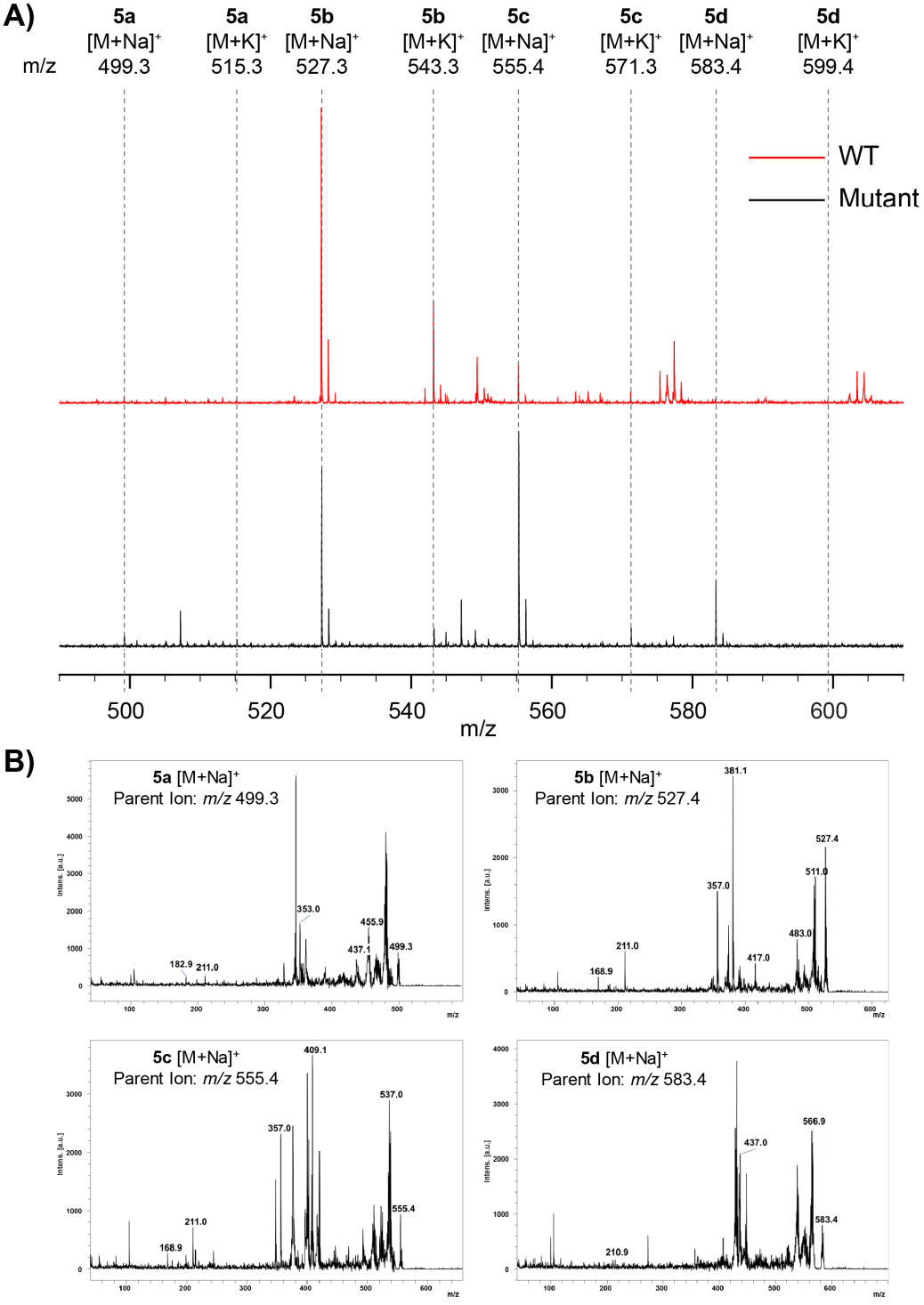


Figure S7. Relative abundance of mono-RL congeners produced from colonies following IPTG induction. Optically guided MALDI-ToF MS was performed on WT colonies. For WT and each mutant strain, ~100 colonies on a filter after IPTG induction were subjected to organic solvent extraction, and the extract was analyzed using either MALDI-ToF MS or LC-MS/MS. The fraction of a RL species relative to total RL amount was calculated as described in the Experimental. Error bars indicate standard deviations of biological replicates (n=81) for MALDI-ToF MS screening, or technical triplicates (n=3) for MALDI-ToF MS or LC-MS/MS measurement of the same extract.

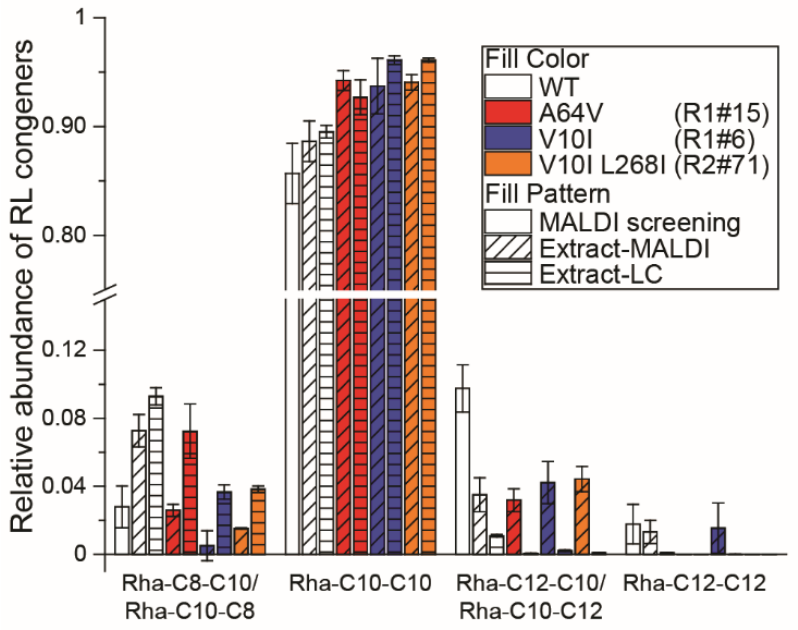
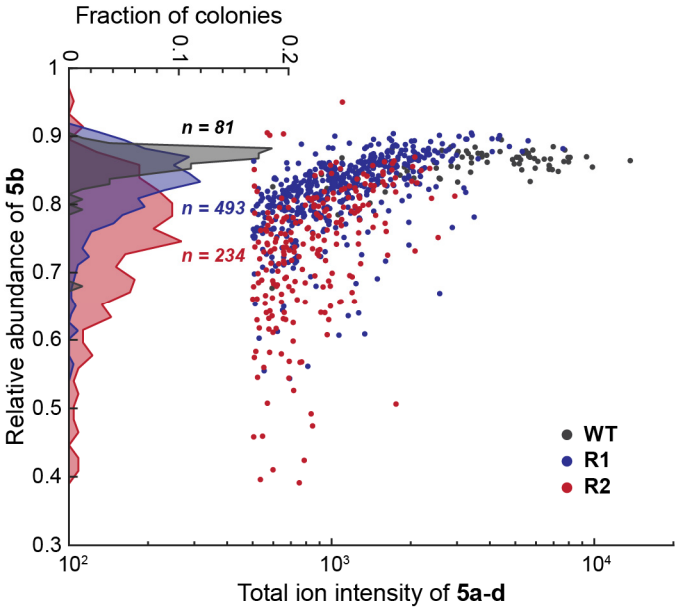


Figure S8. Comparison of relative RL 5b abundance among WT, R1, and R2. Each point on the scatter plot corresponds to the average spectrum of an individual colony. The overlaid histograms display the fractions of colonies with a given relative abundance of RL **5b**. Colonies with total RL intensities below 500 are excluded.



Supporting References

(1) Deane, C. D.; Melby, J. O.; Molohon, K. J.; Susarrey, A. R.; Mitchell, D. A. *ACS Chem. Biol.* **2013**, *8*, 1998-2008.
(2) Ong, T. H.; Kissick, D. J.; Jansson, E. T.; Comi, T. J.; Romanova, E. V.; Rubakhin, S. S.; Sweedler, J. V. *Anal. Chem.* **2015**, *87*, 7036-7042.
(3) Jansson, E. T.; Comi, T. J.; Rubakhin, S. S.; Sweedler, J. V. *ACS Chem. Biol.* **2016**, *11*, 2588-2595.
(4) Do, T. D.; Comi, T. J.; Dunham, S. J.; Rubakhin, S. S.; Sweedler, J. V. *Anal. Chem.* **2017**, *89*, 3078-3086.
(5) Comi, T. J.; Neumann, E. K.; Do, T. D.; Sweedler, J. V. *J. Am. Soc. Mass Spectrom.* **2017**, in press, DOI:10.1007/s13361-13017-11704-13361.
(6) van der Maaten, L.; Hinton, G. *J. Mach. Learn. Res.* **2008**, *9*, 2579-2605.
(7) Chao, R.; Liang, J.; Tasan, I.; Si, T.; Ju, L.; Zhao, H. *ACS Synth Biol* **2017**, *6*, 678-685.
(8) Hao, Y.; Blair, P. M.; Sharma, A.; Mitchell, D. A.; Nair, S. K. *ACS Chem. Biol.* **2015**, *10*, 1209-1216.
(9) Northen, T. R.; Lee, J.-C.; Hoang, L.; Raymond, J.; Hwang, D.-R.; Yannone, S. M.; Wong, C.-H.; Siuzdak, G. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3678-3683.
(10) Masyuko, R. N.; Lanni, E. J.; Driscoll, C. M.; Shrout, J. D.; Sweedler, J. V.; Bohn, P. W. *Analyst* **2014**, *139*, 5700-5708.
(11) Lanni, E. J.; Masyuko, R. N.; Driscoll, C. M.; Aerts, J. T.; Shrout, J. D.; Bohn, P. W.; Sweedler, J. V. *Anal. Chem.* **2014**, *86*, 9139-9145.