

<https://doi.org/10.1038/s42003-025-09405-w>

Multiple active-sites editing of adenylation domains of *srfA* operon for constructing diverse surfactin lipopeptide

Check for updates

Zhaoyang Wang^{1,3}, Li Liu^{1,3}, Jiaxun Wei¹, Ning Liu¹, Ming Ying¹ , Yan Jiang² & Lei Huang¹

The *srfA* operon, which encodes the surfactin synthase SrfA in *Bacillus* rhizobacteria, is essential for surfactin biosynthesis. As a typical non-ribosomal peptide synthetase (NRPS), the adenylation domain of SrfA determines the structure of the synthesized peptide. A major challenge in the field lies in the precise and efficient engineering of NRPS adenylation domains to generate diverse surfactin variants with enhanced bioactivities. Herein, we applied a non-specific guide RNA system (UgRNA/Cas9) for multi-site and diversified editing of the *srfA* operon in the high-secreting strain *Bacillus pumilus* LG3145. UgRNAs targeting four conserved motifs around the SrfA active pocket were designed to direct Cas9-mediated replacement of Stachelhaus codes with codons for 28 different amino acids. This strategy generated two engineered strains: WA1, which carries a Val→Phe mutation in module 2 and produces [Ser²]surfactin, and WA2, which harbors three mutations (Thr→Ser, Glu→Lys, Asn→Gly) across modules 3 and 6, yielding [Lys^{3,6}]surfactin. The two variants displayed distinct antifungal profiles: WA1 exhibited strong activity against *Ustilaginoidea virens*, whereas WA2 inhibited both *Pyricularia oryzae* and *U. virens*. In contrast, the parent strain LG3145 was only effective against *Rhizoctonia solani*. This work demonstrates the potential of UgRNA/Cas9-driven editing of NRPS domains to expand structural and functional diversity of surfactins.

Bacilli are known for their ability to secrete secondary metabolites that exhibit antibacterial, antifungal, and plant-growth-promoting properties, making them highly valuable for applications in industry, agriculture, and medicine¹. These compounds are assembled by large multi-enzyme complexes known as non-ribosomal peptide synthetases (NRPSs)^{2–6}. Among the most studied is surfactin, a potent lipopeptide produced by various *Bacillus* species, whose synthesis is directed by the *srfA* operon^{7–10}. The functional core of each NRPS module is the adenylation (A) domain, which is responsible for selecting, activating, and incorporating specific amino acid substrates into the growing peptide chain^{11–13}. The specificity of each A domain is largely determined by a set of ten critical residues within its active site, commonly referred to as the Stachelhaus codes¹⁴. These codes have provided a foundational blueprint for efforts to reprogram NRPS machinery and generate novel peptides with tailored biological activities¹⁵.

However, despite considerable advances in our understanding of NRPS biochemistry and the availability of genomic editing technologies, the engineering of these systems remains remarkably challenging. A persistent and significant research gap exists in our ability to effectively alter substrate specificity beyond minor, conservative substitutions¹⁶. Most successful

attempts to date have been limited to swapping amino acids with similar physicochemical properties, for instance, replacing glutamine with a methylated derivative¹⁷. These conservative changes often lead to only incremental modifications in the final peptide structure, thereby limiting the scope of biological diversity that can be achieved. This constraint underscores a critical shortcoming in the field is the lack of a robust and generalizable strategy for introducing non-conservative mutations that could potentially yield entirely new peptide scaffolds with novel functions. All these findings suggest the inherent complexity of A domain specificity may involve not only the Stachelhaus codes but also broader structural elements surrounding the binding pocket¹⁸.

In response to this limitation, our study aimed to develop and implement an innovative editing strategy designed to move beyond conservative mutagenesis. We hypothesized that regions flanking the canonical Stachelhaus codes, which exhibit a certain degree of conservation and are less characterized, might play a crucial role in defining substrate specificity. To test this, we targeted the A domains of all seven modules in the surfactin synthetase (SrfA) system of our previously engineered *Bacillus pumilus* strain LG3145. This strain possesses a mutated *cis*-regulatory element that

¹School of Chemistry and Chemical Engineering, Tianjin University of Technology, Tianjin, PR China. ²Department of Environmental Science, Atlantic Technological University, Ash Lane, Sligo, Ireland. ³These authors contributed equally: Zhaoyang Wang, Li Liu. e-mail: ym@tjut.edu.cn; yan.jiang@atu.ie; huanglei@tjut.edu.cn

alleviates carbon catabolite repression, resulting in a dramatically enhanced (~12-fold) expression of the *urfA* operon and providing an ideal expression background for detecting new surfactin variants¹⁹.

Our approach utilized a multiplexed, non-specific UgrNA/Cas9 system to introduce pools of degenerate repair templates across the targeted regions simultaneously. This high-throughput strategy was designed to saturate the potential mutational space and identify non-obvious mutations that confer altered specificity. Contrary to our initial plan of focusing on a single codon position (e.g., codon 278), this unbiased method led to the discovery of two novel mutant strains, WA1 and WA2. These strains harbored unexpected mutations located in previously unexplored core sequences between conserved motifs 1 and 3 within modules 2, 3, and 6 of the SrfA system. Importantly, these mutations resulted in successful alterations to the substrate specificity of their respective A domains. Strain WA1 was found to produce a novel surfactin variant, [Ser²]surfactin, indicating a change in the second module. Strain WA2 produced [Lys^{3,6}] surfactin, demonstrating dual mutations in the third and sixth modules. The incorporation of serine and lysine represents a clear move beyond conservative change, altering the charge and hydrophilicity of the resulting lipopeptides. Subsequent bioassays confirmed that these engineered strains exhibited distinct and diversified antifungal profiles, demonstrating that our engineering strategy not only altered the chemical output but also enhanced the functional capacity of the strain.

Results

Putative anchor motifs of A domain in SrfA of *B. pumilus* LG3145

The A domain of NRPS plays a critical role in substrate recognition and activation, with the 100-residue region surrounding the Stachelhaus codes serving as a key signature sequence²⁰. Based on the published sequences of 458 diverse A domains in the SBSPKv2 database (<http://www.nii.ac.jp/~pkfdb/sbspks/three.html>), we utilized the open-source software NRSPredictor2 (<https://bitbucket.org/chevrn/sandpuma>) to predict the substrate-binding pocket of the A domain (Supplementary Fig. 1)²¹. We generated an alignment map of the region from 32 diverse A domains, revealing several conserved sequences, which we designate as “anchors” located in close proximity to the Stachelhaus codes. These anchors, herein termed core sequence motifs 1–4, encompass one or more Stachelhaus codes (Supplementary Fig. 2). Among these motifs, motif 1 and 4 exhibit the highest conservation (average identity > 80%), while motif 3 is moderately conserved. In contrast, motif 2 displays the lowest conservation, with the Stachelhaus codon at position 278 showing only 13% identity (Fig. 1a, Supplementary Table 1 and Supplementary Data 1). Consequently, the region between motifs 1 to 4 likely represents a flexible segment of the A domain, potentially facilitating the recognition of diverse amino acid substrates.

In this study, we investigated the surfactin biosynthesis operon (*urfA*) in *B. pumilus* LG3145, as predicted by antiSMASH. The operon encodes three NRPSs and a type II thioesterase through the genes *urfA-A*, *urfA-B*, *urfA-C*, and *urfA-D*²². The synthetases SrfA-A, SrfA-B, and SrfA-C consist of three, three, and one module, respectively, each capable of specifically recognizing and adenylating their corresponding amino acid substrates via their A domains. Using the whole-genome dataset of *B. pumilus* LG3145 (GSA: CRA011959), we extracted the primary amino acid sequences of approximately 100 residues surrounding substrate-binding pockets of the A domains of the seven modules. These sequences were then subjected to detailed sequence alignment using Clustal Omega (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>), generating a detailed alignment map. Consistent with the earlier analysis, this map identified conserved “anchor” sequences (yellow boxes) adjacent to the Stachelhaus codes (red boxes), corresponding to the core sequence motifs 1–4. Notably, identical sequences were observed between modules 2 and 7, as well as between modules 3 and 6, which recognize the same substrates, L- and D-leucine, respectively (Fig. 1b, c). Based on these findings, the regions between motifs 1 and 4 may serve as ideal targets for editing, offering opportunities to modify substrate specificity and enable the production of novel surfactin variants.

Multiple editing protocols of active-sites in A domains

UgrNA/Cas9-mediated multi-sites editing has enabled the targeted mutation of conserved sequences, particularly the L/D-leucine modules with more than 90% identity in *urfA* operon. However, the distance from motif 1 to 4 exceeds 300 bp, making it impractical to design a single target fragment to edit the entire region. To address this, two pairs of non-specific UgrNAs, along with a series of repair templates, were designed to span the regions between motifs 1 and 4, aiming to simultaneously edit all seven modules. We utilized the web-based tool CRISPOR to select gRNA sequences to target specific regions of each module²³. These sequences served as target fragments for the Cas9 protein, using NNG as PAM sites. The corresponding UgrNAs were subsequently redesigned through sequence alignment, achieving at least 55% identity. However, due to the low conservation of motif 2 sequences, degenerate bases were incorporated into the design of UgrNA-Motif 2-F2 to enhance similarity (Fig. 2).

Subsequently, four multiple sites editing plasmids were designed, named pCas9-Motif1-F, pCas9-Motif1-R, pCas9-Motif2-F1 and pCas9-Motif2-F2. The insertion of UgrNA-Motif1-F and UgrNA-Motif1-R into plasmids pCas9-Motif1-F and pCas9-Motif1-R demonstrates higher similarity to modules 2, 3 and 6 (>70%) than to other modules. The UgrNA-Motif2-F1 sequence in the pCas9-Motif2-F1 plasmid closely matches the target fragments of modules 3 and 6 (>90% similarity), but shows only 55% similarity with module 7. To target the remaining modules with less conserved sequences, we engineered UgrNA-Motif2-F2 containing degenerate bases and inserted it into pCas9-Motif2-F2, thereby improving sequence similarity to modules 1, 2, 4, and 5. In order to disrupt the core sequences between motif 1 and 4, multiple single-stranded repair templates are used to replace the 278 codon. But, the single-stranded repair templates generally yield fewer transformants than double-stranded repair templates, thereby affording superior access to sequences with specificity²⁴. The repair templates used in this study contain asymmetric “homology arms” for targeting the fragments of motif 2 with the length of around 50 bp, which can enhance the repair efficiency. To avoid undesired cleavage of donor templates by Cas9, point mutations were introduced, including a PAM sequence alteration (NGG → NGC) to prevent Cas9 binding. First, we co-transformed the host strain LG3145 with Leu_{2,7}-template-A, B and C and pCas9-Motif2-F1/F2 plasmids to simultaneously modify codon 278 across all seven modules. Three degenerate codons (TGB, VAN, and SSN) were introduced at position 278 of motif 2 in the repair templates, enabling random incorporation of E/D/K/N/Q/H/R/W/C/A residues. To edit the core sequences located between motif 1 and 4, we performed co-transformation with three repair templates, D-Leu_{3,6}-template-A, B, and C, which are homologous to the sequence of motif 2 at module 3 and 6, with four CRISPR plasmids pCas9-Motif1-F, pCas9-Motif1-R, pCas9-Motif2-F1, and pCas9-Motif2-F2. This enabled simultaneous cleavage of both motif 1 and 2. The repair templates also contain degenerate codons SSN, VAN and SSN at position of 278, designed to introduce E/D/K/N/Q/H/R/W/C/A residues randomly (Fig. 3).

Repair templates disruption of *urfA* operon with UgrNA/Cas9 editing system

Many previous studies have proved that the alteration of Stachelhaus codes definitely led to substrate selectivity change²⁵. According to the editing protocols, the plasmids of pCas9-Motif1-F, pCas9-Motif1-R, and pCas9-Motif2-F1 were constructed and verified with agarose gel electrophoresis and DNA sequencing (Supplementary Fig. 3). The plasmids were divided into two pairs, pCas9-Motif1-F and pCas9-Motif1-R cleaving around motif 1, pCas9-Motif2-F1 and a degenerate plasmid pCas9-Motif 2-F2 cleaving around motif 2. To verify the position 278 is the most variable residue in changing substrate, we used the plasmids of pCas9-Motif2-F1 and pCas9-Motif2-F2 with the repair template Leu_{3,6}-template A, B and C to transform the protoplast of *B. pumilus* LG3145, we successfully obtained a transformant termed WA1²⁶. Additionally, the two pairs of plasmids pCas9-Motif1-F, pCas9-Motif1-R and pCas9-

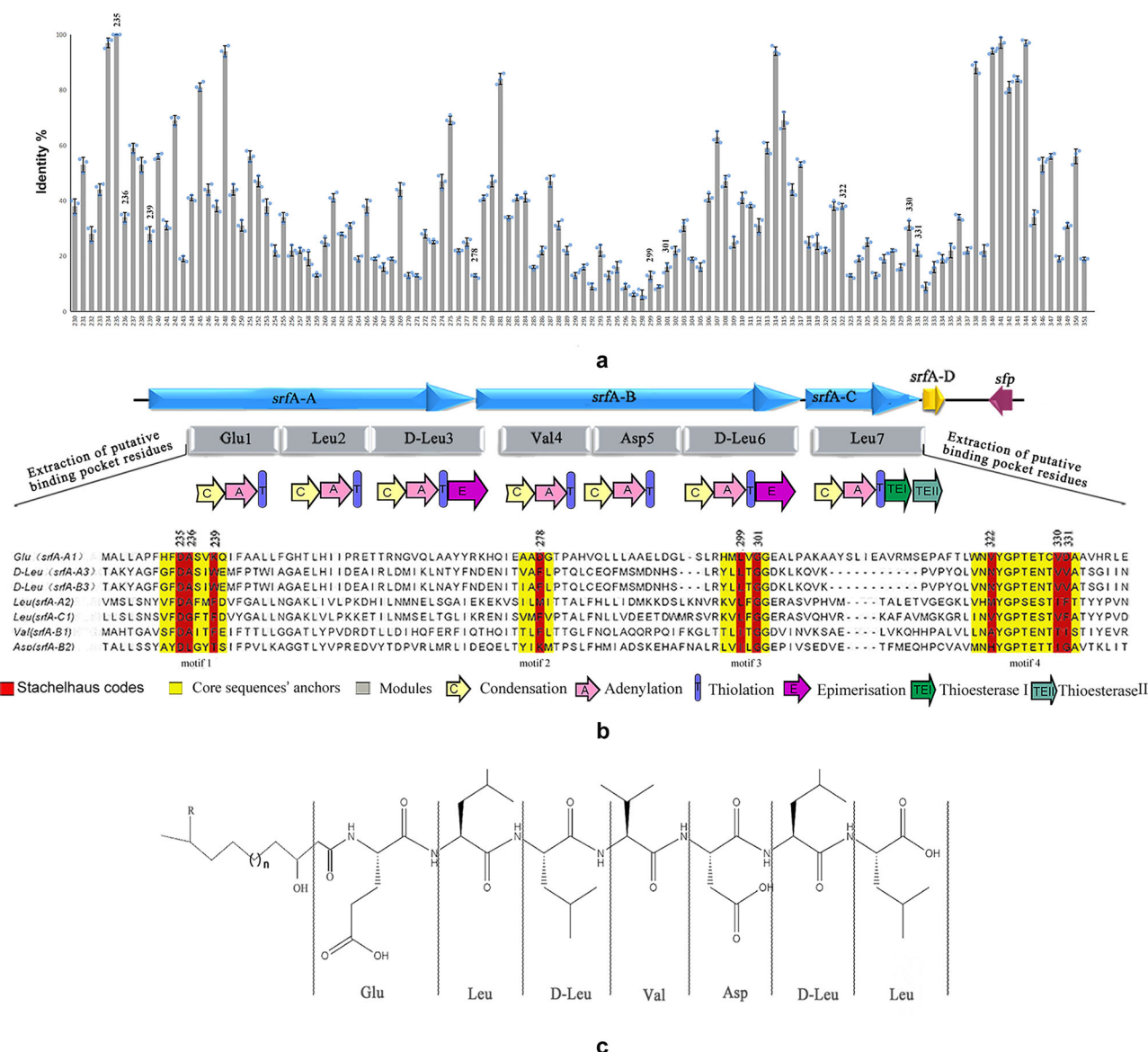


Fig. 1 | Active-site analysis of A domains in *srfA* operon. **a.** Assessment of sequence identity for the active pocket residues among 458 diverse adenylation (A) domain. Data in the graph are means±s.d. of ten replicates (n = 10) **b.** Using the amino acid

sequences of A domain binding pockets to determine core sequence anchors (yellow) and Stachelhaus codes (red) for individual modules (gray) in the *srfA* operon. **c.** Linearized surfactin molecular structure.

Motif2-F1, pCas9-Motif2-F2 were transformed into LG3145 with the repair templates D-Leu_{2,7}-template A, B and C, another transformant was emerged and termed as WA2. The PCR product sequencing results showed no mutation at position 278 but detected changes in the nearby region (close to motif 2), regardless of the protocols employed. However, in the strain WA1, the types of mutation are visually distinct from the strain WA2. The four sets of mutations occurred around 278, including one transversion (G/T) and three transitions (a G/A and two T/C), in which a missense mutation at the transversion site of 274 from GTC to TTC, resulting in a change of residue from Val to Phe. In contrast, the mutant strain WA2 produced by the second strategy, involving same mutant pattern both in the modules 3 and 6 simultaneously, happened many more mutations (about 36) than WA1. Therefore, all of changes distributing across the region from motif 1 to 3, including six transversions and thirty transitions, in which three missense mutations were observed at 244 (ACA → /TCA), 272 (GAA → AAA) and 273 (AAT → GGG), resulting the residues alternation from Thr to Ser, Glu to Lys, and Asn to Gly, respectively (Fig. 4 and Supplementary Fig. 4).

Versatile lipopeptide surfactin produced by editing the core sequence

In previous study, we found the strain LG3145 can produce heptapeptide surfactin formed as Glu-Leu-D-Leu-Val-Asp-D-Leu-Leu and linked with β-hydroxy fatty acids (C12) at the end of the exponential phase of growth²⁷. If the molecular structure of surfactin alters, its biological activity will change subsequently²⁸. In order to test the biological activity of the mutant strains, we selected three rice pathogens, *Pyricularia oryzae* (rice blast), *Ustilaginoidea virens* (rice false smut), and *Rhizoctonia solani* (rice sheath blight), to investigate the resistance properties of WA1, and WA2. In this experiment, we observed that all the strains LG3145, WA1, and WA2 exhibit antifungal activity, and show a certain degree of specificity. As shown in Fig. 5, strain LG3145 exhibited the strongest inhibitory effect against *Rhizoctonia solani*, with an inhibition zone diameter of 16.7 ± 0.14 mm after 5 days of cultivation on PDA plates. The corresponding EC₅₀ value was approximately (4.6 ± 0.23) × 10⁶ cells/mL. In contrast, the mutant strain WA2 showed significant antifungal activity against all three pathogens, particularly against *Rhizoctonia oryzae* and *Ustilaginoidea virens*, with inhibition zone diameters of 20.3 ± 0.18 mm and 10.8 ± 0.11 mm, respectively. The

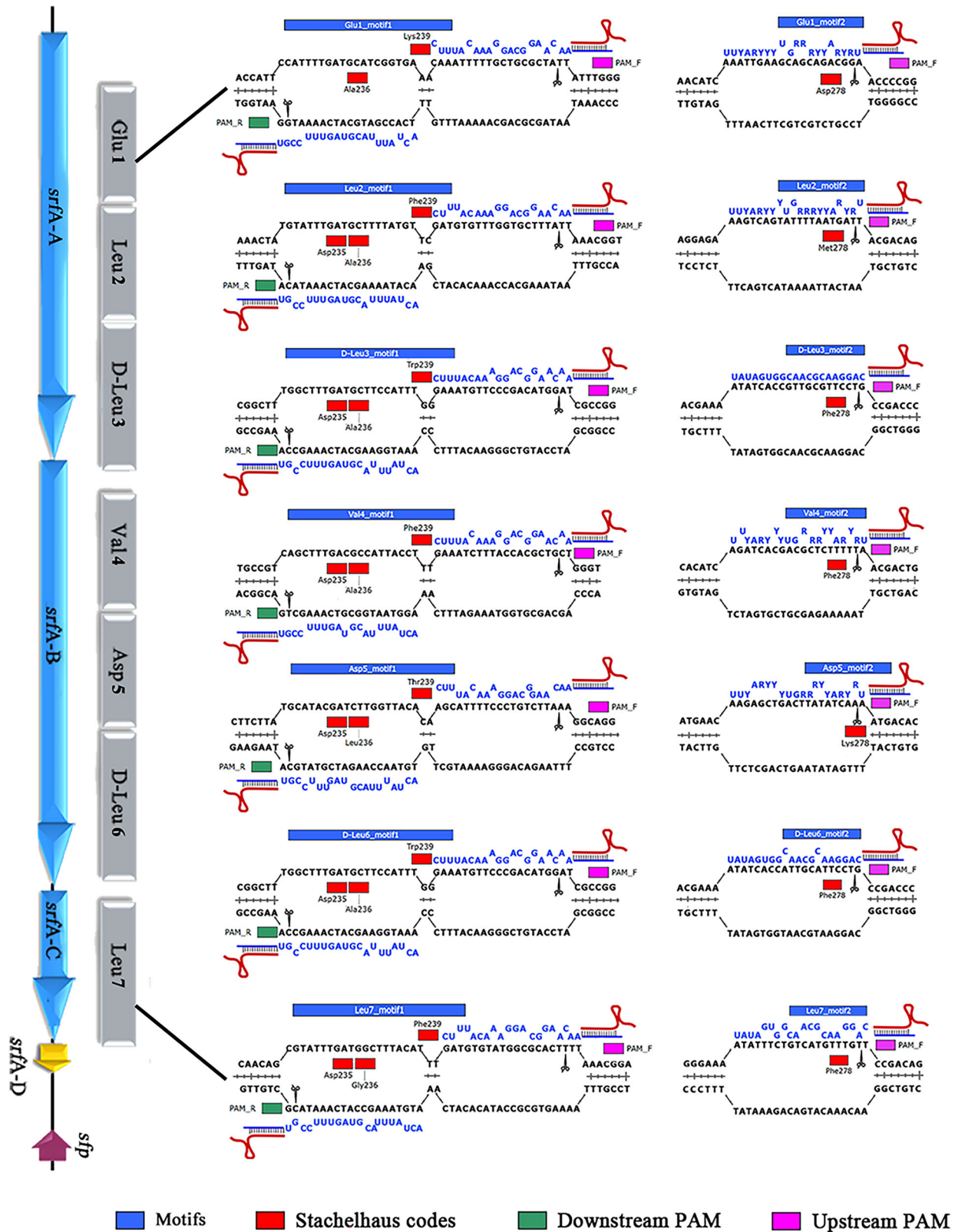


Fig. 2 | Schematic diagram of UgrNAs binding to target fragments in each module. The UgrNA sequences for motif 1 and motif 2 of all seven modules are highlighted in blue at their binding sites, with their corresponding PAM sites marked in pink. Non-complementary bases are displaced for emphasis.

corresponding EC_{50} values for WA2 against these pathogens were approximately $(4.5 \pm 0.23) \times 10^6$ cells/mL and $(7.9 \pm 0.40) \times 10^5$ cells/mL, respectively. Additionally, the inhibitory effect of WA1 on *Ustilagoindovirens* was more pronounced than that of LG3145 but weaker than that of WA2, with an inhibition zone diameter of approximately 65.7 ± 0.47 mm

and an EC_{50} value of $(1.3 \pm 0.07) \times 10^{12}$ cells/mL. The antifungal activity assay was conducted in triplicate, and all three strains demonstrated diverse bioactivities (Supplementary Table 2, Table 3 and Supplementary Data 2).

UPLC-MS was employed to characterize the molecular structure of surfactin produced by mutant strains WA1 and WA2, with the aim of

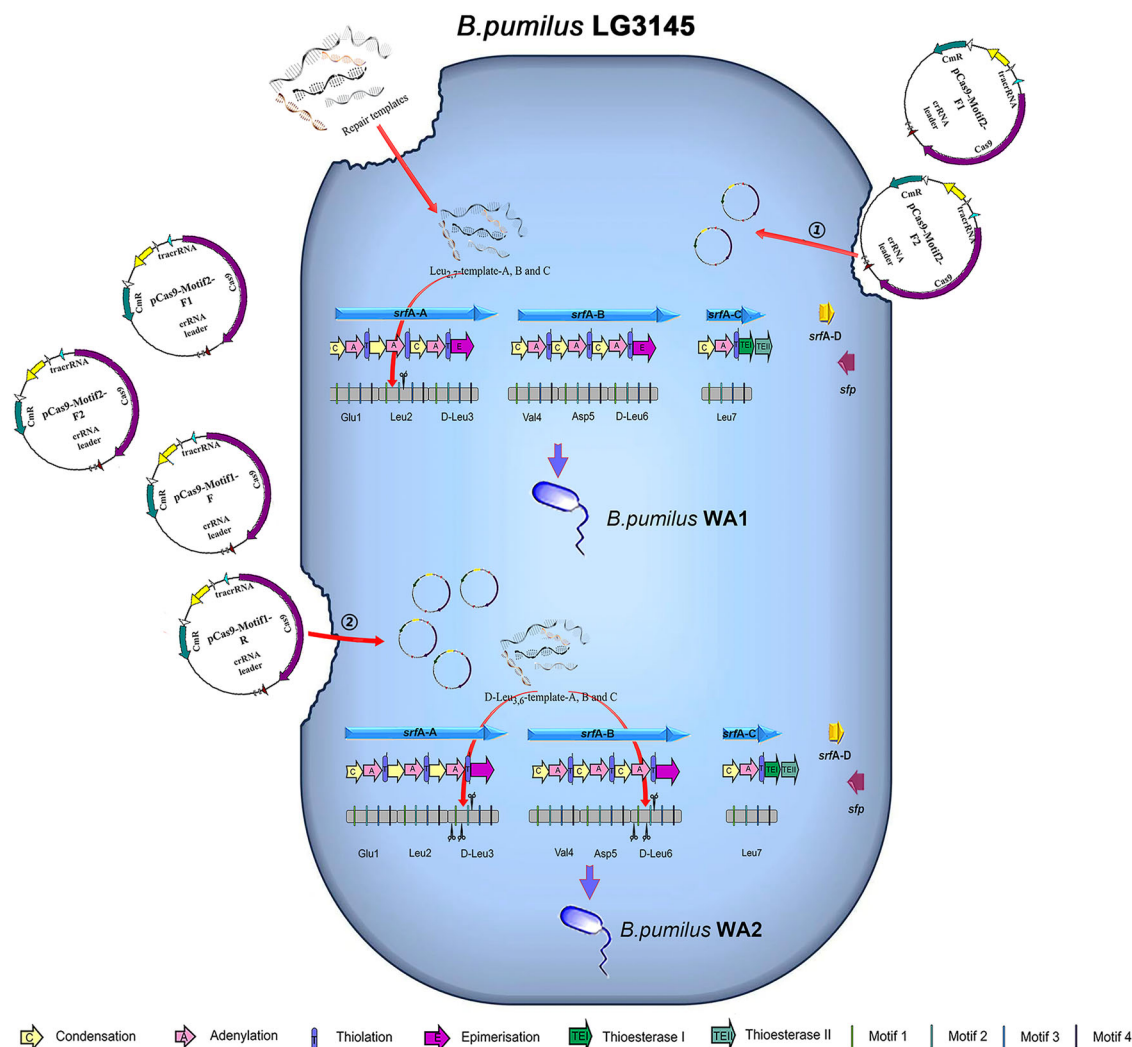


Fig. 3 | Active-sites editing protocols of multiple A domains in *srfA* operon. ① The protocol for constructing WA1 involved the use of two CRISPR editing plasmids, pCas9-Motif2-F1 and pCas9-Motif2-F2, to target and cleave motif 2 sequences. The degenerate nucleotides R and Y in pCas9-Motif2-F2 represent A/G and C/T. ② The protocol for constructing WA2 involved four plasmids: pCas9-Motif1-F, pCas9-

Motif1-R, pCas9-Motif2-F1, and pCas9-Motif2-F2, which were used to simultaneously edit motif 1 and motif 2. Subsequently, a series of repair templates are introduced to facilitate the genome editing process. Plasmid schematics for pCas9-Motif1-F, pCas9-Motif1-R, pCas9-Motif2-F1, and pCas9-Motif2-F2 were derived from Addgene plasmid # 42876.

elucidating the mechanisms underlying these changes in the mutants. Using a modified method described by Yan Zhi (2017) to purify the surfactin homologs when they were cultivated in glycerol minimal medium (GYMM), we obtained a series of data of UPLC and MS as shown in Fig. 6. Comparing with the MS data of LG3145, the peaks at *m/z* 1036.29 moving to 1008.600 and 1064.458 in the MS pattern of WA1 and WA2, which correspond to surfactin isomers with different saturated fatty acid chain lengths (C13 and C17) and consistent with the standard surfactin isomers reported by Yan Zhi et al. in 2017²⁹. The MS pattern of the surfactin obtained in LG3145 was consistent with the standard surfactin spectrogram in mzCloud dataset (<https://www.mzCloud.org/>) (Fig. 6a, b). However, in the MS pattern of WA1, the peaks observed around *m/z* 348.163 and 512.278 showed slight deviations from the standard, likely due to a missing hydrogen atom. Comparison with the standard serine (Ser) spectrum in mzCloud (<https://www.mzcloud.org/>) revealed that the peak at *m/z* 106.0499 corresponds to Ser incorporation in the novel molecule, while the peaks at *m/z* 348.163 and 512.278 represent the characteristic fragments of Glu-Ser-D-Leu and Val-Asp-D-Leu-Leu, respectively. These findings suggest that the novel molecule may be a [Ser²]surfactin variant (Fig. 6c, d). In the MS pattern of WA2, the peaks at 148.088 *m/z* is the representative structures of Lys, 389.696 *m/z*

and 529.745 *m/z* are speculated to be Glu-Leu-Lys and Val-Asp-Lys-Leu, which may be a novel [Lys^{3,6}]surfactin molecular (Fig. 6e, f).

Discussion

The large multimodular NRPS, involved in synthesizing bioactive molecules, were investigated more than 30 years to create novel bioactive peptides³⁰. The traditional DNA recombination techniques, widely used in previous studies, were difficult to meet the demands of large-scale drug screening for the low efficiency and limited diversity³¹. Although the NRPS enzymes have larger size, the regions of the active sites in the adenylation domains, only spanning around 100 residues, are the specificity-selecting units in each module which play the key role in determining the structure and function of the corresponding products³².

In this study, we conducted an in-depth analysis of Stachelhaus codes within the adenylation domains of NRPSs. Using the *Nrpspredictor2* software and based on the SBSPKSv2 database, we successfully predicted the core sequence region of the A domain and the Stachelhaus codes of 28 amino acids. The results indicate that the site of 278 is a highly variable codon, which may be the best effect mutation target for changing the substrate recognition properties of the A domain. So, our initial editing protocol

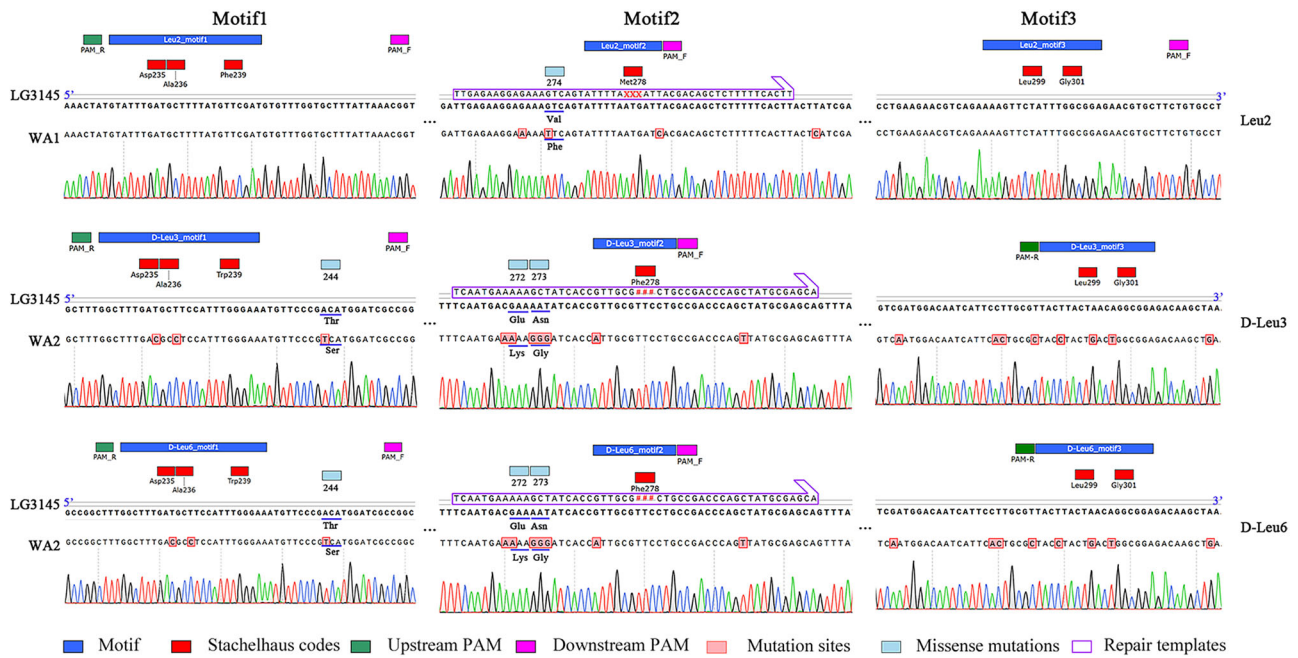
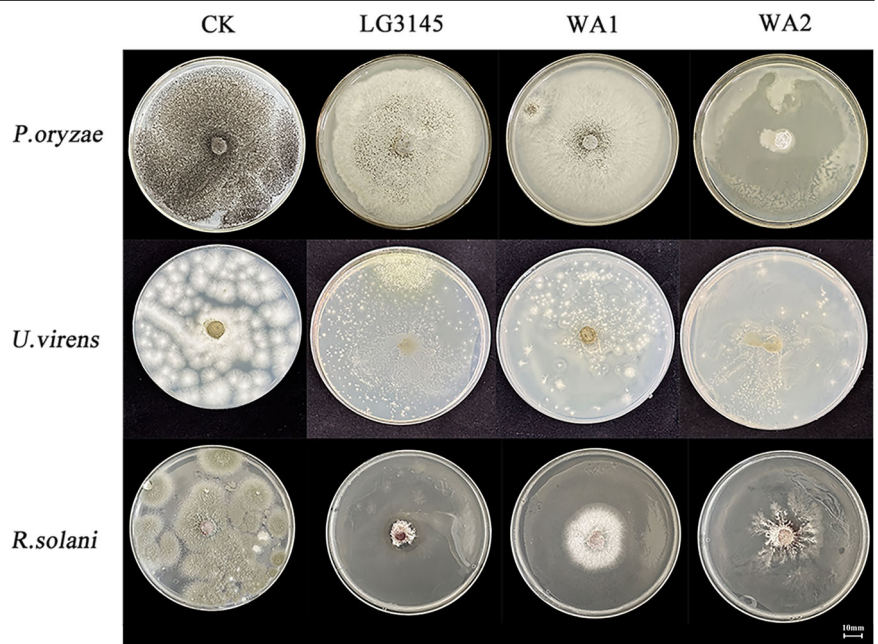


Fig. 4 | Chromatograms of the DNA sequencing of edited A domains active-sites in *srfA* operon *B. pumilus* WA1 and WA2 and coverage of the repair templates. All the mutations (pink box) were identified at three functional domains: motif 2 within module 2 (Leu2) of the *srfA* operon in strain WA1; motifs 1-3 spanning modules 3 and 6 (D-Leu3/D-Leu6) in strain WA2, in which the missense mutations were highlighted with blue underlines. XXX: TYB/VAN/SSN, ###: TYB/VAN/SSN.

Fig. 5 | Strain WA1 and WA2 phytopathogen inhibition assay on the fifth day of growth. Phytopathogenic fungi were inoculated onto potato dextrose agar (PDA) plates containing either strain WA1 or WA2, and their growth was compared with that of fungi inoculated onto PDA plates containing LG3145. Fungi inoculated onto PDA plates without any treatment served as the control (CK). Phytopathogenic fungi include *Pyricularia oryzae*, *Ustilagoidea vires* and *Rhizoctonia solani*.



was to alter the 278 residues of the seven modules with a series of degenerate repair templates simultaneously through UgrNA/Cas9 multi-site editing technology. Degenerate base-driven stochastic mutagenesis may lead to the generation of novel structural variants of surfactin molecules. However, mutagenesis events were restricted to module 2, 3, and 6 under the guidance of UgrNA-Motif1-F, UgrNA-Motif1-R, UgrNA-Motif2-F1 and UgrNA-Motif2-F2 with > 75% sequence similarity, rather than occurring at the designated site (278). Moreover, in the mutant strain WA1, all of mutations were exclusively localized to motif 2 which may be triggered by homologous degenerate template with this region. There is a transversion-induced missense mutation (Val→Phe), which is obviously repaired by the template of Leu_{3,6}-template-A. The same phenomenon is observed in strain WA2,

with same mutations exclusively distributed within the identical modules (3 and 6). There are three missense mutation (Thr→Ser), (Glu→Lys) and (Asn→Gly) which may be induced by the degenerate templates of D-Leu_{2,7}-template-A, B and C respectively. We propose that genome DNA strand distorts during HDR-mediated repair, driven by degenerate template-derived mutations symmetrically flanking the Cas9-mediated cleavage site at the DNA gap. This facilitated the bases at vicinity of 278 repaired according to the degeneracy templates. However, the template-mediated interference exhibits spatial restriction, occurring only in the target's proximal regions, as no changes were observed in motifs 1 and 3 of WA1. Moreover, the interference range expands with increasing cleavage sites when four UgrNA/Cas9 plasmids are co-expressed in WA2. More

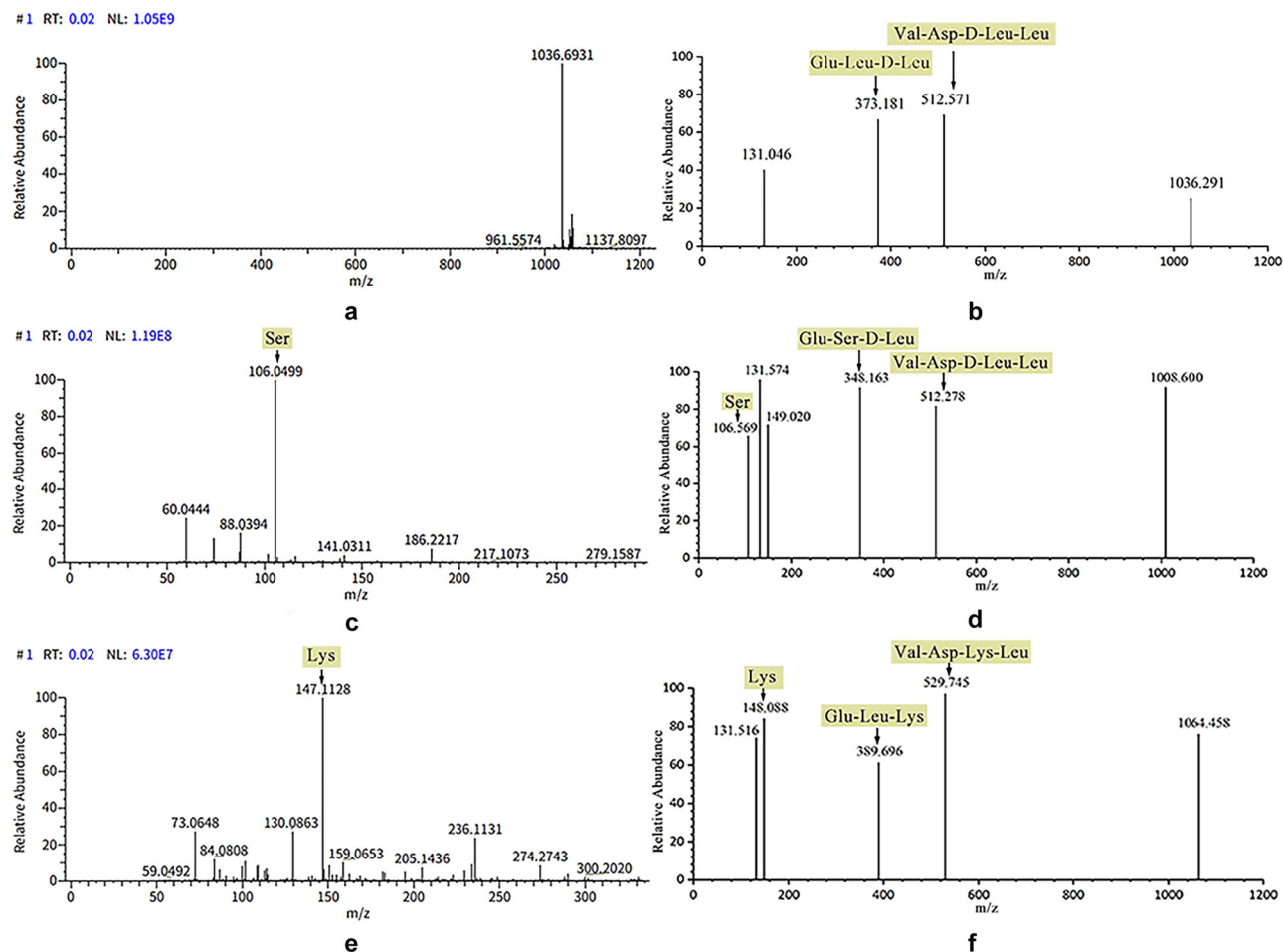


Fig. 6 | Detection of surfactin homologs produced by strain WA1 and WA2 compared with that from the strain LG3145 using ultra performance liquid chromatography-mass spectrometry. a Standards mass profile of surfactin in mzCloud database. **b** Peaks in the surfactin mass spectrometry profile of strain LG3145. **c** Standards mass profile of serine in mzCloud database. **d** The mass

spectrometry investigation focused on the extraction of surfactin homologs from the metabolites of strain WA1. **e** Standards mass profile of lysine in mzCloud database. **f** The mass spectrometry investigation focused on the extraction of surfactin from the metabolites of strain WA2. The bacteria were grown using glycerol as the carbon source (GYMM).

mutations symmetrically flanked the target site (278) between motifs 1 and 3, accompanied by an increase in degenerate template-derived missense mutations, likely due to the expansion of the DNA strand distortion region. Compared to WA1, WA2 received two-fold of degenerate repair templates which generated consistent editing pattern in the same modules (3 and 6). But these templates, which were homologous to the upstream sequence of the cleavage gap, failed to mediated repair at the downstream motif 3, which was likely repaired through NHEJ.

Interestingly, although all missense mutations have deviated from the predetermined target site, mutant strains WA1 and WA2 still produce distinct structural variants of surfactin with differential bioactivities. Specifically, strain WA1 produced [Ser²]surfactin, which showed specific activity against *U. virens*, while strain WA2 generated [Lys³⁶]surfactin that exhibited strong activity against both *P. oryzae* and *U. virens*. In contrast, strain LG3145 only demonstrated activity against *R. solani*. These phenomena further demonstrate that the ~100 residues region encompassing Stachelhaus codes from motif 1 to 4 also plays a key role in determining the specificity of the A domain reported by Eppelmann (2002).

We have successfully found a novel DNA self-repair mechanism in CRISPR/Cas9 system with the degenerate single templates, which exhibits wobble editing and diverse repair pathways. This approach will facilitate the biosynthesis of non-natural NRP molecule with novel bioactivities, addressing current limitations in editing efficiency and structural diversity. The UgrNA/Cas9 multiplex gene editing technology

with degenerate template-based random repair may be a potential approach to alleviate drug-resistance crisis in the medical and agriculture field. However, whether the template perturbation repair process can be controlled is unclear and its repair mechanisms need to be further explored.

Methods

Bacterial, fungal and reagents

Bacillus pumilus LG3145 used in this study was derived from the *Bacillus pumilus* wild-type strain SH-B9 (NCBI Accession no: NZ_CP011007.1), as previously described by Wang et al.³³ *Escherichia coli* DH5α (the host for *pCas9*) was purchased from Thermo Fisher Scientific (Waltham, MA, United States), were used for plasmid construction and amplification, and cultured in Luria-Bertani (LB) medium (10 g/L peptone, 5 g/L yeast powder, 10 g/L NaCl and pH 7.3–7.5), containing 25 μg/mL chloramphenicol. The plant fungal pathogens, *Pyricularia oryzae*, *Ustilago noidea virens*, and *Rhizoctonia solani*, were obtained from the Plant Protection Institute of Tianjin Academy of Agricultural Sciences, Tianjin, China (Supplementary Table 4). All chemicals, a DNA gel purification kit, and a plasmid extraction kit were purchased from TIANGEN Biotech (Beijing, China). Restriction enzymes and DNA ligase were purchased from New England Biolabs (Ipswich, MA, United States). Taq DNA polymerase and oligonucleotides used for PCR experiments were purchased from Ruibio Biotech (Beijing, China).

Plasmids and primers

The *pCas9* plasmid and sequences used in this study are available from the Addgene nonprofit plasmid repository (Cat: 42,876). The *pCas9* plasmid is 9326 bp based on the low-copy backbone of pACYC184 (ATCC 37033), which endows resistance to chloramphenicol (25 µg/mL). The other elements of this plasmid include tracrRNA, Cas9, and an array of crRNAs with two *BsaI* sites for spacer's insertion. The plasmids pCas9-Motif1-F, pCas9-Motif1-R, pCas9-Motif2-F1, and pCas9-Motif2-F2 were constructed for editing seven target modules. The primers pCas9-F and pCas9-R used for verifying recombinant plasmids (Supplementary Table 5).

Multisite editing strategies and sequence design of UgrNAs and repair templates for the *srfA* operon

To enable simultaneous editing of multiple active sites within the adenylation domains of the *srfA* operon, we adopted a modified UgrNA/Cas9-mediated multisite editing strategy based on our previous work³³. In this approach, the gRNA oligonucleotides must share at least 50% identity with the target sites and are therefore referred to as non-specific gRNAs (UgrNAs). This design allows the Cas9 protein to recognize and bind to multiple target sites concurrently. Importantly, these target sites are required to share conserved regions to facilitate such multiplex recognition. Analysis of the A domain protein residues revealed that the A domain motif in SrfA of *B. pumilus* LG3145 is highly conserved. A detailed discussion of these findings is provided in the "Results" section.

Firstly, we chose NNG for the PAM and the partial sequences of motif 1 and 2 designed for target fragments using the CRISPOR online tool (<http://crispor.tefor.net/>). For enhancing editing efficiency, we designed two directional UgrNAs, UgrNA-Motif 1-F (forward) and UgrNA-Motif 1-R (reverse), to specifically target the longer region of motif 1, while one UgrNA, UgrNA-Motif2-F1, to edit the shorter region of motif 2 (12 bp) in module 3, 6 and 7. Because of the less conservation, another one, UgrNA-Motif2-F2, with degenerate bases was designed to guide Cas9 for targeting module 1, 2, 4, and 5. The UgrNA oligonucleotides are listed in Supplementary Table 6, and the binding sites for UgrNA/Cas9 are presented in Fig. 2. Secondly, the four non-specific UgrNAs (UgrNA-Motif1-F, UgrNA-Motif1-R, UgrNA-Motif2-F1 and UgrNA-Motif2-F2) were inserted into the *BsaI* sites of *pCas9* as the CRISPR spacers after annealing at 95 °C for 5 min and 37 °C for 30 min using T4 ligase according to the *pCas9* protocol. *E. coli* DH5a cells were used to construct and amplify plasmids. Ligations of UgrNA-annealed oligos and *BsaI*-digested *pCas9* were gently mixed with competent *E. coli* DH5a cells at a ratio of 1:200 and incubated on ice for 30 min, then heated at 42 °C in a water bath for 90 s, and chilled for 1–2 min on ice. The cells were then incubated in preheated LB medium (150 µL) at 37 °C for 45 min, then plated onto 25 µg/mL chloramphenicol selective media and incubated overnight at 37 °C. The following morning, transformants were picked and incubated in LB medium. After cell growth, plasmids were extracted using TIANprep Mini Plasmid kit purchased from TIANGEN Biotech (Beijing, China) and the pCas9-Motif 1-F, pCas9-Motif 1-R, and pCas9-Motif 2-F1 constructs were verified by DNA sequencing. The degenerate plasmid pCas9-Motif2-F2 was constructed in a 20 µL reaction system containing 1 µL of annealed UgrNA-Motif2-F2 oligos, 10 µL of *BsaI*-digested *pCas9*, 2 µL of 10× T4 ligase buffer, 1 µL of T4 ligase, and 6 µL of ddH₂O. The mixture was incubated at 16 °C overnight in a biometra Tgradient PCR instrument (Jena, Germany), followed by gradual cooling to 4 °C at a rate of 5 °C per minute, and then used for transformation.

In this study, our objective is to modify the 100-residue region encompassing the Stachelhaus codes from motif 1 to 4 and substitute the codon at position 278 within motif 2. We designed six homologous repair templates: Leu_{2,7}-template-A, B, and C, containing the degenerate codons TYB, VAN, and SSN (representing S/F/L, E/D/K/N/Q/H, and G/A/R/P, respectively) for repairing modules 2 and 7; D-Leu_{3,6}-template-A, B, and C, carrying the same codons for homology-directed repair of modules 3 and 6 (Supplementary Table 6). Two editing strategies were used with these repair templates to simultaneously replace the codon 278. The first strategy

targeted motif 2 with plasmids pCas9-Motif2-F1 and pCas9-Motif2-F2, whereas the second targeted a 100-residue region from motif 1 to motif 4 using pCas9-Motif1-F, pCas9-Motif1-R, pCas9-Motif2-F1, and pCas9-Motif2-F2. These repair templates were mixed with the plasmids at equal molar ratios and then used to transform LG3145 using the same method as described below.

Protoplasts of *B. pumilus* LG3145 were prepared for transformation using a modified method based on that of Chang and Cohen (1979). A single colony of *B. pumilus* LG3145 was inoculated into 5 mL of LB medium and cultured overnight at 37 °C in a shaker (200 rpm). The overnight culture was diluted 50-fold into 5 mL of fresh LB medium and grown for 8 h at 37 °C (200 rpm). Cells were harvested by centrifugation (9000 × g, 10 min, 4 °C). The supernatant was removed and cells were resuspended in 1 ml of SMMP solution (equal volumes of 2× SMM buffer and 4× Penassay broth). Lysozyme powder was added to 0.4 mg/mL, mixed, and incubated at 37 °C for 45 min (100 rpm) to prepare protoplasts (globular appearance). Protoplasts were harvested by centrifugation (5000 × g, 10 min, RT), the supernatant was poured off, and the cell pellet was resuspended gently in 0.05 ml SMMP. Lastly, a 10 µL mixture of plasmids and repair templates at equal molar ratios was added to 100 µL of protoplast suspension, and the resulting mixture was maintained on ice for 10 min. The mixture was then transferred to a pre-cooled 0.2 cm Bio-Rad electroporation cuvette (Hercules, California, USA) and subjected to four pulses of 0.7 kV with a 4 ms interval using a Bio-Rad electroporator (Hercules, California, USA). Immediately afterwards, 1 mL of SMMP was added, and the suspension was incubated at 37 °C and 100 rpm for 12 h. Next, 200 µL of the culture was spread onto a CMR selective agar plate (10 g/L glucose, 10 g/L peptone, 10 g/L yeast extract, 5 g/L beef paste, 5 g/L NaCl, 20 g/L agar, 342.3 g/L sucrose, 8.13 g/L MgCl₂, and 8 µg/mL chloramphenicol) and incubated overnight to select transformants.

DNA sequencing

The recombinant plasmids were extracted from *E. coli* DH5a with TIANprep Mini Plasmid kit purchased from TIANGEN Biotech (Beijing, China). The region of spacer (240 bp) inserted in *pCas9* was amplified with the primer pCas9-F and pCas9-R. The genome of strain WA1 and WA2 was extracted using a TIANamp bacteria DNA kit purchased from TIANGEN Biotech (Beijing, China) and used as the PCR templates for amplifying the target sequences of module 1–7 with the primers in Supplementary Table 5. All of PCR products were detected using agarose gel electrophoresis, and subsequently sent to Ruibio Biotech (Beijing, China) for DNA sequencing on an Illumina HiSeq 4000 platform (Illumina, San Diego, CA, USA). The results were aligned with the sequences of UgrNAs or the target fragments of LG3145 with Snapgene software (<https://www.snapgene.com>). The oligonucleotides, Taq DNA polymerase, and dNTP utilized for the PCR experiments were purchased from Ruibio Biotech (Beijing, China).

Antifungal pathogens assessment

To further evaluate the inhibitory effect of the engineered strains on the fungal pathogens, an experiment was conducted using mycelial growth inhibition method and assessed by the half-maximal effective concentration value (EC₅₀)³⁴. The evaluated strains included *B. pumilus* LG3145, WA1 and WA2. These strains were inoculated individually in LB medium at 37 °C on a shaker at 125 rpm to conform their initial cell densities reaching approximately 1.32 × 10⁹ cells/mL (OD_{600nm} ≈ 2), and then, preparing a series of gradient dilutions from these bacterial suspensions and spreading 100 µL of each uniformly onto PDA plates (20 g/L glucose, 200 g/L potato, 15–20 g/L agar, and pH 7.3–7.5). At the same time, a 7-mm diameter fungal inoculum was then inoculated onto the center of each plate. Mycelial growth was assessed using the cross-measurement method at the fifth day. The average inhibition rate (μ) for each sample was calculated according to the equation $\mu = \frac{R_{CK} - R_x}{R_{CK} - R_0} \times 100\%$, in which R_{CK} and R_x denote the mycelial radii of the control and the experimental group, R_0 denotes the radius of the original fungal inoculum. The regression analysis was performed for each strain using μ versus log (C) for calculating EC₅₀ values, where C represents cell density.

Surfactin homologs assay

A modified method by Yan Zhi et al. in 2017 was used for the purification and assay of the surface homologs. Two *B. pumilus* strains WA1 and WA2 were cultivated in glycerol minimal medium (GYMM) containing 3.48 g/L KH_2PO_4 , 1.5 g/L $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$, 3.96 g/L $(\text{NH}_4)_2\text{SO}_4$, 0.7 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ with 60% (w/v) glycerol and 0.01 g/L yeast, at 35 °C and 125 rpm for 72 h, respectively. Thereafter, 60 mL of cell culture was collected and the pH was adjusted to 2.0 using 6 M HCl. The mixture was then subjected to centrifugation at $10,000 \times g$ for 30 min, in order to remove the upper layer of the cell culture medium. The precipitate was then dissolved in 1 mL of 100% CH_3OH and subjected to a second centrifugation step ($10,000 \times g$, 30 min) to ensure the removal of cellular residues. All samples (5 μL) were loaded onto a Waters ACQUITY UPLC BEH C18 column (5 μm , 150×4.6 mm; Waters Corp, Milford, MA, USA) and analyzed using a Waters UPLC I-Class/Xevo G2 XS-Quadrupole Time-of-flight system (Waters Corp, Milford, MA, USA) at 4 °C with a flow rate of 0.3 mL/min under isocratic elution with water (0.1% formic acid) and acetonitrile (0.1% formic acid). Mass spectrometry detection was performed in the positive ESI-MS mode with the following conditions: capillary voltage: 0.5 kV; cone voltage: 35 V; extractor voltage: 4.0 V; source temperature: 11 °C; desolvation temperature: 550 °C. UPLC-MS analysis data were collected and processed using MassLynx NT 4.1 software with QuanLynx program (Waters Corp). *B. pumilus* strain LG3145 was cultivated in the same condition as the control.

Statistics and reproducibility

Quantitative data are presented as mean \pm s.d. The experiments of pathogens resistance in this study were performed in triplicate. Assessment data of sequence identity for the active pocket residues are means \pm s.d. of ten replicates.

Data availability

The *B. pumilus* LG3145 raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China (GSA: CRA011959), which are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The plasmids generated in this study, pCas9-Motif1-F (249571), pCas9-Motif1-R (249572), pCas9-Motif2-F1 (249573), and pCas9-Motif2-F2 (249574) are deposited in Addgene (<https://www.addgene.org/>). All source data underlying the graphs in Fig. 1a is provided in Supplementary Data 1, and the source of Supplementary Table 3 is provided in Supplementary Data 2. All other data supporting the conclusions are included within the article and its Supplementary files.

Code availability

The substrate-binding pockets of these A domains were predicted using the open-source software NRPSpredictor2, available at <https://bitbucket.org/chevrn/sandpuma>. The antiSMASH tool, which was used to predict the *srfA* operon in *Bacillus pumilus* LG3145, is publicly available at <https://antismash.secondarymetabolites.org/>. Sequence alignment was performed using Clustal Omega, which is publicly accessible at <https://www.ebi.ac.uk/Tools/msa/clustalo>. For gRNA design, we used the web-based tool CRISPOR, available at <http://crispor.tefor.net/>.

Received: 21 May 2025; Accepted: 11 December 2025;

Published online: 06 January 2026

References

- Santos, V. S. V., Silveira, E. & Pereira, B. B. Toxicity and applications of surfactin for health and environmental biotechnology. *J. Toxicol. Env. Health-Pt B-Crit. Rev.* **21**, 382–399 (2018).
- Bozhuyuk, K. A. J. et al. Evolution-inspired engineering of nonribosomal peptide synthetases. *Science* **383**, eadg4320 (2024).
- Ishikawa, F. & Tanabe, G. Chemical strategies for visualizing and analyzing endogenous nonribosomal peptide synthetase (NRPS) megasynthetases. *Chembiochem.* **20**, 2032–2040 (2019).
- Ishikawa, F., Tanabe, G. & Kakeya, H. Activity-based protein profiling of non-ribosomal peptide synthetases. *Curr. Top. Microbiol. Immunol.* **420**, 321–349 (2019).
- Song, Y. et al. Development and application of CRISPR-based genetic tools in *Bacillus* species and *Bacillus* phages. *J. Appl. Microbiol.* **133**, 2280–2298 (2022).
- Koglin, A. & Walsh, C. T. Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep.* **26**, 987–1000 (2009).
- Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140 (2016).
- Iacovelli, R., Bovenberg, R. A. L. & Driessen, A. J. M. Correction to: nonribosomal peptide synthetases and their biotechnological potential in *penicillium rubens*. *J. Ind. Microbiol. Biotechnol.* **49**, (2022).
- Sussmuth, R. D. & Mainz, A. Nonribosomal peptide synthesis—principles and prospects. *Angew. Chem. Int. Ed. Engl.* **56**, 3770–3821 (2017).
- Qi, X., Liu, W., He, X. & Du, C. A review on surfactin: molecular regulation of biosynthesis. *Arch. Microbiol.* **205**, 313 (2023).
- Watzel, J., Duchardt-Ferner, E., Sarawi, S., Bode, H. B. & Wohnert, J. Cooperation between a T domain and a minimal C-terminal docking domain to enable specific assembly in a multiprotein NRPS. *Angew. Chem. Int. Ed. Engl.* **60**, 14171–14178 (2021).
- Gao, L. et al. Module and individual domain deletions of NRPS to produce plipastatin derivatives in *Bacillus subtilis*. *Microb. Cell. Fact.* **17**, 84 (2018).
- He, R. et al. Knowledge-guided data mining on the standardized architecture of NRPS: subtypes, novel motifs, and sequence entanglements. *Plos Comput. Biol.* **19**, e1011100 (2023).
- Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**, 9718–9726 (2002).
- Maruyama, C. & Hamano, Y. The Assembly-line enzymology of nonribosomal peptide biosynthesis. *Methods Mol Biol.* **2670**, 3–16 (2023).
- Alistair, S. B. et al. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228 (2018).
- Thirlway, J. et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chem. Int. Ed. Engl.* **51**, 7181–7184 (2012).
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
- Bi, M. et al. Genome-scale cis-acting catabolite-responsive element editing confers *Bacillus pumilus* LG3145 plant-beneficial functions. *iScience* **27**, 108983 (2024).
- Heard, S. C. & Winter, J. M. Structural, biochemical and bioinformatic analyses of nonribosomal peptide synthetase adenylation domains. *Nat. Prod. Rep.* **41**, 1180–1205 (2024).
- Aguero-Chapin, G., Perez-Machado, G., Sanchez-Rodriguez, A., Santos, M. M. & Antunes, A. Alignment-free methods for the detection and specificity prediction of adenylation domains. *Methods Mol. Biol.* **1401**, 253–272 (2016).
- Xu, Y., Wu, J., Liu, Q. & Xue, J. Genome-wide identification and evolutionary analyses of *srfA* operon genes in *Bacillus*. *Genes*. **14**, (2023).
- Huang, Z., Peng, Z., Zhang, M., Li, X. & Qiu, X. Structure, function and engineering of the nonribosomal peptide synthetase condensation domain. *Int. J. Mol. Sci.* **25**, (2024).
- Haber, J. E. DNA repair: the search for homology. *Bioessays* **40**, e1700229 (2018).
- Zhang, T. & Zhou, Q. Using large-scale multi-module NRPS to heterologously prepare highly efficient lipopeptide biosurfactants in

- recombinant *Escherichia coli*. *Enzyme. Microb. Technol.* **159**, 110068 (2022).
26. Chang, S. & Cohen, S. N. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol Gen Genet.* **168**, 111–115 (1979).
 27. Fira, D., Dimkic, I., Beric, T., Lozo, J. & Stankovic, S. Biological control of plant pathogens by *Bacillus* species. *J. Biotechnol.* **285**, 44–55 (2018).
 28. Yuan, L., Zhang, S., Peng, J., Li, Y. & Yang, Q. Synthetic surfactin analogues have improved anti-PEDV properties. *Plos One* **14**, e215227 (2019).
 29. Zhi, Y., Wu, Q. & Xu, Y. Genome and transcriptome analysis of surfactin biosynthesis in *Bacillus amyloliquefaciens* MT45. *Sci. Rep.* **7**, 40976 (2017).
 30. Crawford, J. M., Portmann, C., Kontrik, R., Walsh, C. T. & Clardy, J. NRPS substrate promiscuity diversifies the xenematides. *Org. Lett.* **13**, 5144–5147 (2011).
 31. Yan, H. et al. A rational multi-target combination strategy for synergistic improvement of non-ribosomal peptide production. *Nat. Commun.* **16**, 1883 (2025).
 32. Prave, L. et al. Investigation of the odilorhabdin biosynthetic gene cluster using NRPS engineering. *Angew. Chem. Int. Ed. Engl.* **63**, e202406389 (2024).
 33. Wang, Y. et al. Wobble editing of *cre*-box by unspecific CRISPR/Cas9 causes CCR release and phenotypic changes in *Bacillus pumilus*. *Front. Chem.* **9**, 717609 (2021).
 34. Li, J. et al. A Comparison of Different Estimation Methods for Fungicide EC₅₀ and EC₉₅ Values. *J. Phytopathol.* **163**, 239–244 (2015).

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant number 42077212]. We greatly appreciate Tianjin Academy of Agricultural Sciences of China for providing *Pyricularia oryzae*, *Ustilagoidea virens* and *Rhizoctonia solani* strains.

Author contributions

Zhaoyang Wang and Li Liu contributed equally. They designed and performed experiments, and wrote the first draft of the manuscript. Jiaxun Wei and Ning Liu participated in the experimental design and the analysis of the experiment results. Ming Ying, Yan Jiang and Lei Huang designed and supervised the project, and guided the experimental design, data analysis,

manuscript writing and revision. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-09405-w>.

Correspondence and requests for materials should be addressed to Ming Ying, Yan Jiang or Lei Huang.

Peer review information *Communications Biology* thanks Xiaoting Qiu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ophelia Bu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026