Published in final edited form as:
Nat Chem Biol. 2018 April ; 14(4): 375–380. doi:10.1038/s41589-018-0008-5.

A lanthipeptide library used to identify a protein–protein interaction inhibitor

Xiao Yang[#1], **Katherine R. Lennard**[#2], **Chang He**[1], **Mark C. Walker**[1], **Andrew T. Ball**[2], **Cyrielle Doigneaux**[2], **Ali Tavassoli**[2,*], and **Wilfred A. van der Donk**[1,*]

[1]Department of Chemistry and Howard Hughes Medical Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

[2]Department of Chemistry and Institute for Life Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom

[#] These authors contributed equally to this work.

Abstract

We describe the production and screening of a genetically encoded library of 10^6 lanthipeptides in *Escherichia coli* using the substrate-tolerant lanthipeptide synthetase ProcM. This plasmid-encoded library was combined with a bacterial reverse two-hybrid system for the interaction of the HIV p6 protein with the UEV domain of the human TSG101 protein, a critical protein–protein interaction for HIV budding from infected cells. Using this approach, we identified an inhibitor of this interaction from the lanthipeptide library, whose activity was verified *in vitro* and in cell-based virus-like particle budding assays. Given the variety of lanthipeptide backbone scaffolds that may be produced with ProcM, this method may be used for the generation of genetically encoded libraries of natural product-like lanthipeptides containing substantial structural diversity. Such libraries may be combined with any cell-based assay for the identification of lanthipeptides with new biological activities.

Macrocycles have proven to be an excellent scaffold for therapeutic agents; these molecules occupy unique chemical space, bridging the gap between small molecules and biologics, and have been successfully deployed against challenging drug targets such as protein–protein interactions (PPIs)1. While natural products and their synthetic derivatives have historically been the main source for therapeutic macrocycles2, cyclic peptides are increasingly viewed

[*]To whom correspondence should be addressed: Wilfred A. van der Donk, 600 S. Mathews Avenue, Urbana, Illinois 61801, United States; vddonk@illinois.edu; phone: (217) 244-5360; fax: (217) 244-8533.; Ali Tavassoli, Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom. a.tavassoli@soton.ac.uk; phone: +44 2380 592395.

Author Contributions

X.Y., A.T. and W.A.v.d.D. designed the study. X.Y., K.R.L, C.H., A.T.B. and C.D. performed all experiments. M.C.W. provided bioinformatics analysis of the plasmid library. All authors analyzed the data. X.Y., A.T., and W.A.v.d.D. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files, or are available from the corresponding author upon reasonable request.

Europe PMC Funders Author Manuscripts

as holding much potential in this area3–5. Macrocyclic peptides share many of the physical properties of natural product macrocycles, such as extended binding sites and limited conformational freedom, with the added advantage of being amenable to the creation of large and diverse genetically encoded libraries5–11.

Ribosomally-synthesized and post-translationally modified peptides (RiPPs) are a major group of natural products that are biosynthesized from a genetically encoded precursor peptide generally containing an N-terminal leader sequence and a C-terminal core peptide12, 13. After various post-translational modifications on the core peptide catalyzed by the biosynthetic enzymes and leader peptide removal, the mature RiPP is produced (Fig. 1a). The biosynthetic enzymes of many RiPPs have been shown to display high substrate tolerance14–25, and thus have the potential to be used for production of large libraries of non-natural macrocyclic peptides.

Lanthipeptides are a major group of RiPPs characterized by intramolecular thioether bridges (termed lanthionine or methyllanthionine) generated via two post-translational modification reactions26: dehydration of Ser or Thr residues followed by cyclization of Cys residues onto the dehydrated amino acids (Fig. 1b). Recent genome mining efforts have led to the discovery of a substrate-tolerant synthetase (ProcM) in *Prochlorococcus* that dehydrates and cyclizes up to 30 different linear precursor peptides encoded in the genome (designated as ProcAs; Supplementary Fig. 1). Remarkably, these peptides are transformed into single polycyclic lanthipeptide products (prochlorosins) with highly diverse ring topologies27–29.

Macrocyclic peptides hold much promise for recognition of protein surfaces and inhibition of PPIs because of their ability to act as structural mimics of native ligands1, 30. In addition to restricted conformational freedom, cyclization also offers increased stability against cellular catabolism compared to linear peptides. Among the multitude of methods for generating cyclic peptides31, genetic approaches offer advantages of large library size and the possibility to couple peptide generation with cell-based screening. Bicyclic peptides have also shown much promise, but other than disulfide-crosslinked peptides, they have thus far been limited to peptides generated by chemical methods6, 9, 32.

In this study, we used enzymatic methods to generate a bicyclic peptide library in *Escherichia coli*. The potential utility of this approach is demonstrated by coupling lanthipeptide library generation to a bacterial reverse two-hybrid system (RTHS) for the PPI between the HIV p6 protein and the ubiquitin E2 variant (UEV) domain of the human TSG101 protein33. This PPI is a potential target for antiviral therapy, as it has been shown to be essential for the budding of HIV from infected cells34–36.

Results

Generation of a lanthipeptide library in *E. coli*

Seeking to take advantage of the remarkable substrate tolerance of ProcM, as well as technology that allows successful production of lanthipeptides in *E. coli*37, we tested the use of ProcM for the generation of non-natural lanthipeptide libraries. Since the highly conserved prochlorosin leader peptide serves as a recognition motif for ProcM to carry out

dehydration and cyclization on diverse core peptides27, the library of substrates all contained the same leader peptide. The core peptides attached to the leader peptide were diversified using a degenerate oligonucleotide with nucleotide diversity introduced in predetermined positions during oligonucleotide synthesis. For proof-of-concept studies, we selected one of the thirty naturally encoded substrates, ProcA2.8 (ref. 27). ProcM converts ProcA2.8 into a product containing two non-overlapping, seven-residue thioether-crosslinked rings (Fig. 1a)28. We first tested the robustness of the ProcM catalyst for faithful installation of the two-ring scaffold when the ten amino acids within the rings were randomized (Fig. 1c). We set several criteria with respect to the randomized sequence. First, we wanted to avoid incorporation of additional Ser, Thr, and Cys residues, as these might disturb the original ring scaffold by forming additional or alternative rings. Second, we wanted to avoid incorporation of stop codons. Third, we preferred minimal bias among the amino acid variants. And lastly, we felt that incorporation of negative charge would be more important than positive charge given the sequence of the p6 peptide that binds to UEV (PEPTAPPEE). The NWY codon usage fulfills these criteria as it introduces two codons each for the amino acid residues Asp, Phe, His, Ile, Leu, Asn, Val, and Tyr. For ProcA2.8, a library of 1.07×10^9 members is theoretically possible by randomizing ten positions using the NWY codons that can introduce one of eight amino acids at each position (i.e. 8^{10}).

In order to test ProcM's catalytic fidelity on these non-native substrates, the library was first constructed in a pRSFDuet vector for co-expression of ProcM with individual members of the randomized N-terminally hexahistidine-tagged substrate library. The library was constructed starting from PCR amplification using degenerate primers that cover the 1.07×10^9 amino acid diversity based on the NWY codons. After digestion and ligation, the plasmid library isolated from transformed *E. coli* cells had a size of 10^6 as determined by the transformation efficiency, which currently limits the size of the library that can be functionally analyzed with our method. Deep sequencing was used to assess the diversity in the pRSF-Duet-derived plasmid library. As anticipated (based on the observed transformation efficiency) the size of the library was around 10^6, as 1,075,925 unique peptide sequences were obtained out of 2,008,108 total reads. Unique peptides that were sequenced ten or fewer times account for 99.7% of the unique peptides and 97.3 % of the total library, with a relatively small number of sequences highly represented (Supplementary Fig. 2). As desired, all eight amino acids encoded by the NWY codons were nearly evenly encoded at all ten randomized positions (Supplementary Fig. 3). Thirty-three clones were randomly chosen and the encoded peptides overexpressed with ProcM in *E. coli.* In the absence of a leader peptide protease and transporter, these peptides remain in the cytoplasm37, and hence the peptides were purified by Ni^{2+} affinity chromatography from lysed cells. The extent of dehydration for each core peptide was determined by matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) after removing most of the leader peptide by proteolysis with endoproteinase Glu-C. All 33 randomly chosen peptides were dehydrated twice (Supplementary Fig. 4). To interrogate the cyclization state, the core peptides were then treated with the thiol-alkylating reagents iodoacetamide (IAA) or N-ethylmaleimide (NEM) that would report on any non-cyclized Cys residues, and the ring topology was determined by tandem MS. Remarkably, ProcM catalyzed full cyclization of all 33 samples (Supplementary Table 1, Supplementary Figs. 4

and 5). Hence, based on the results of these 33 randomly chosen clones, ProcM produces a high quality library of bicyclic peptides containing two non-overlapping lanthionine rings in *E. coli*.

Screening the library for inhibitors of a PPI

We next assessed the feasibility of coupling the above lanthipeptide library with a cell-based selection assay to enable the identification of PPI inhibitors. A bacterial RTHS was used that couples the successful disruption of the p6–UEV PPI (Fig. 2a)38 to the transcription and expression of reporter genes critical for cell survival on selection medium (Fig. 2b)33. The p6–UEV RTHS was previously used to screen a library of 3.2×10^6 monocyclic peptides, generated via split-intein circular ligation of peptides and proteins (SICLOPPS); *cyclo*-SGWIYWNV was identified in this screen and shown to inhibit the formation of virus-like particles with an IC_{50} of 7 μM in a cell-based viral budding assay33. Given that the RTHS uses a PTac promoter to induce expression of the p6 and UEV proteins upon IPTG treatment, a pACYCDuet vector was engineered to contain two arabinose promoters to drive the expression of the library of substrates and ProcM (pARDuet; Supplementary Fig. 6) and the co-expression conditions were optimized (Supplementary Fig. 7). A pARDuet library encoding lanthipeptides based on ProcA2.8 was generated as before but without the His_6-tag, and used to transform the p6–UEV RTHS. In bacteria containing a plasmid encoding a p6–UEV PPI inhibitor, disruption of the p6–UEV PPI would also disrupt the P22–434 repressor complex to enable transcription of the kanamycin resistance and *HIS3* genes (Fig. 2b and Supplementary Fig. 8), leading to a growth advantage on selective media containing kanamycin and lacking histidine. In contrast, cells that contain a lanthipeptide that does not disrupt the p6–UEV interaction would not survive or would grow with a strong growth defect on selective media (Supplementary Fig. 8a). The p6–UEV RTHS cells transformed with the plasmid library were plated onto selective media containing IPTG and arabinose, and incubated at 33 °C until colonies were observed. Eighty colonies exhibited a growth advantage and were picked for further assessment. Individual colonies were subjected to secondary screening by drop spotting for IPTG-dependent growth inhibition and restoration of growth with arabinose. Colonies with the expected phenotype (Fig. 2c) were subjected to a tertiary screen by isolating the plasmid and re-transforming the p6–UEV RTHS cells with the plasmid. This procedure removes any false positives with chromosomal mutations that conferred the growth advantage. The substrate specificity of the hits was also assessed using another RTHS strain that was identical to the p6–UEV strain, except that p6 and UEV were replaced with an unrelated PPI (human CMG2 interacting with *Bacillus anthracis* protective antigen)39; peptides showing inhibition of this unrelated PPI were discarded as they were either non-specific inhibitors, or targeted other components of the RTHS (Supplementary Fig. 8b). One peptide, XY3-3 (**1**; Fig. 2d), remained after secondary and tertiary screening of the 80 initial hits. Analysis of our deep-sequencing data of the plasmid library used in the p6–UEV RTHS cells revealed that nucleotide sequences encoding XY3-3 were present fifteen times in the sequenced library that contained 8,673,291 unique peptides (Supplementary Fig. 9). Thus, the representation of XY3-3 in the pARDuet screening library is about the same as in the pRSFDuet-based library used to test the fidelity of ProcM discussed above.

Assessing the activity of XY3-3 *in vitro*

XY3-3 was prepared in *E. coli* on a larger scale and purified to assess its activity *in vitro*. We initially established a sandwich enzyme-linked immunosorbent assay (ELISA) to monitor the p6–UEV PPI. We first titrated p6 onto plate-bound UEV, and found an EC_{50} of 948 ± 21 nM for the p6–UEV interaction in this assay (Supplementary Fig. 10a). We next titrated XY3-3 onto plate-bound UEV, followed by p6, and observed disruption of the p6–UEV interaction by XY3-3 with an IC_{50} of 3.6 ± 0.3 μM (Fig. 3a). For comparison, a previously reported cyclic peptide inhibitor of this PPI (CP11, *cyclo*-SGWIYWNV)33 was also used in this assay, and disrupted the p6–UEV interaction with an IC_{50} of 30.2 ± 2.2 μM (Supplementary Fig. 10b). The contribution of the leader peptide to the activity of XY3-3 was assessed via two independent approaches. First, we determined the activity of the parent ProcA2.8 scaffold in the above assay; as this molecule contains the same leader peptide as XY3-3, ProcA2.8 would also disrupt the interaction if the observed p6–UEV inhibitory activity was a consequence of the leader peptide. However, neither the cyclic nor linear forms of ProcA2.8 inhibited the p6–UEV PPI (Supplementary Fig. 11). Second, we used the Lys-C and Glu-C endoproteases to remove portions of the leader peptide from XY3-3; Lys-C cleaves ~50% of the leader sequence to generate peptide **2**, whereas Glu-C removes all but five residues of the leader sequence to afford peptide **3** (Supplementary Fig. 12). Cleavage of the leader sequence increased the potency of the lanthipeptide against the p6–UEV PPI; Peptide **2** disrupted the p6–UEV interaction with an IC_{50} of 2.7 ± 0.6 μM (Fig. 3b), while peptide **3** disrupted this PPI with an IC_{50} of 1.9 ± 0.8 μM (Fig. 3c). The necessity of XY3-3's lanthipeptide backbone for its activity was assessed using linear variants of XY3-3 (full-length and Glu-C-cleaved) (Supplementary Fig. 12). Neither molecule was able to disrupt the p6–UEV PPI by ELISA (Supplementary Fig. 13). It may therefore be concluded that the lanthipeptide core of XY3-3 (rather than its leader sequence) is responsible for disruption of the p6–UEV PPI, and that the cyclic thioether backbone is required for its activity.

In order to determine whether XY3-3 binds p6 or UEV, we generated an N-terminal fluorescein-labeled derivative of XY3-3 (Supplementary Fig. 14). The ability of p6 or UEV to pull down this molecule was assessed by immobilizing either protein onto Ni^{2+}-coated 96-well plates, and washing the fluorescent derivative of XY3-3 over these wells. We observed a 4-fold increase in fluorescence in wells containing UEV, with no change over background in wells containing p6 (Fig. 3d), indicating that XY3-3 binds to UEV. The binding affinity for UEV was determined using microscale thermophoresis (MST). XY3-3 bound to UEV with a K_d of 16.0 ± 1.1 μM (Fig. 3e), while peptide **3** bound UEV with a K_d of 4.0 μM ± 3.3 μM (Fig. 3f). Thus, as with the ELISA data, removal of the leader sequence of XY3-3 resulted in a more potent molecule, indicating that the lanthipeptide core of this molecule is responsible for its activity. The binding constants obtained by MST were corroborated using a fluorescence polarization (FP) assay; a K_d of 5.5 ± 0.5 μM was measured for the binding of N-terminal FITC-labeled peptide **3** to UEV (Fig. 3g). For comparison, we also determined the affinity of UEV for the nonapeptide from p6 (PEPTAPPEE) that contains the PTAP motif that is essential for the binding of p6 to UEV (Fig. 2a)40. A fluorescently-labeled PTAP nonapeptide bound to UEV with a K_d of 16.6 ± 0.2 μM (Fig. 3h). This value is within the range of previously reported values for this

interaction, measured by various methods (3 μM41, 27 μM42, and 50 μM40). Thus, peptide **3** binds to UEV with ~3-fold better affinity than the PTAP nonapeptide when measured with the same assay. We also conducted competition FP experiments to assess whether FITC-labeled peptide **3** displaces the PTAP peptide; we did not observe displacement from UEV of PTAP by peptide **3**, nor peptide **3** by PTAP (Supplementary Fig. 15). These results indicated that PTAP and peptide **3** bind to non-overlapping regions of UEV. In contrast, peptide **3** was displaced from UEV by increasing concentrations of full length p6 peptide in the FP assay (Fig. 3i). Taken together, the FP data indicates that while peptide **3** does not bind to the PTAP binding pocket, it competes with p6 for binding to UEV.

The contribution from each of the two lanthionine rings to the binding of XY3-3 to UEV was next assessed. Two variants of peptide **3** were generated, with the amino acid sequence of each ring inverted one at a time. Inverting the amino acid sequence of ring A of peptide **3** (from LHFFL to LFFHL) caused the loss of UEV binding activity (Supplementary Fig. 16). Similarly, inverting the amino acid sequence of ring B (from HVLDI to IDLVH) in peptide **3** resulted in a lanthipeptide that did not bind UEV (Supplementary Fig. 16). These observations demonstrate the requirement of both rings A and B of peptide **3** in binding to UEV.

Assessing the activity of XY3-3 in cells

We next evaluated the potential of peptide **3** to disrupt viral budding in a cell-based assay33. This peptide did not affect the GAG-mediated budding of virus-like particles, which we attributed to lack of cell permeability. We therefore designed an analog of XY3-3 in which a Cys replaced what becomes the N-terminal Gly in peptide **3** (Supplementary Fig. 12). Co-expression of this full-length peptide with ProcM in *E. coli* resulted in a fully dehydrated and cyclized product and tandem ESI-MS showed that the same ring topology was formed as the parent peptide (Supplementary Fig. 17). Leader peptide removal with endoproteinase Glu-C produced an analogue of peptide **3** with an N-terminal Cys, that was conjugated to a Cys-containing variant of the cell-permeable Tat peptide via a disulfide bond (XY3-3-Tat, **4**; Supplementary Fig. 12)33. Previous studies have shown that loss of TSG101 function is lethal in cells and *in vivo*43, 44. We therefore initially assessed the cytotoxicity of the PPI inhibitor and found XY3-3-Tat to be toxic to HEK293T cells at doses of 1 μM and above (Fig. 4a). However, the compound was not toxic at 100 nM or 500 nM, and therefore these doses were used in the subsequent experiments. The effect of XY3-3-Tat on GAG-mediated viral budding was assessed in an established cell-based assay that utilizes GFP-tagged GAG polyprotein, and measures the formation of viral-like particles (VLPs) in the supernatant33, 45. XY3-3-Tat (100 nM) showed ~65% inhibition of viral budding from HEK293T cells in this assay, whereas no effect was observed when using 25 μM Tat alone (Fig. 4b). The UEV domain of TSG101 also mediates degradation of Epidermal Growth Factor Receptor (EGFR). The binding of Epidermal Growth Factor (EGF) to its receptor (EGFR) triggers EGFR degradation, a process mediated by the interaction between TSG101 and hepatocyte growth factor-regulated tyrosine kinase substrate (HRS)46. This interaction involves a PSAP motif on HRS that engages TSG101 via the same binding site as the p6 PTAP motif40. As such, a compound that disrupts the binding of p6 to UEV is likely to also interfere with HRS binding to TSG101 and therefore inhibit EGFR degradation. HeLa cells were treated with

500 nM Tat as a control and cultured either in the presence or absence of EGF. As expected, we observed a reduction in EGFR levels in cells cultured with EGF (Fig. 4c). We next assessed the effect of XY3-3-Tat on this process by repeating the above experiment with cells treated with 500 nM XY3-3-Tat and cultured either in the presence or absence of EGF. Cells treated with both XY3-3-Tat and EGF contained higher levels of EGFR than those treated with Tat and EGF, with EGFR levels approaching that of cells treated with Tat alone (Fig. 4c). Cells treated with XY3-3-Tat only (no EGF) showed substantially elevated levels of EGFR compared to Tat-treated equivalents (Fig. 4c). Together, these data demonstrate inhibition of EGFR degradation by XY3-3-Tat, in line with the proposed mechanism of action of XY3-3 binding to UEV and inhibiting its interaction with P(T/S)AP containing proteins.

Discussion

Our results demonstrate the feasibility of using a promiscuous lanthipeptide synthetase (ProcM) for the intracellular generation of lanthipeptide libraries of about a million members. While the tolerance of lanthipeptide synthetases has been previously demonstrated through alanine scanning and saturation mutagenesis at single sites15, 17, 21, this study represents the first example of a library of variants of a natural lanthipeptide by simultaneous randomization of multiple positions. We illustrate this possibility by generating a library with a theoretical size of 1.07×10^9 members by randomizing 10 amino acid positions within a bicyclic lanthipeptide scaffold. The library was screened for inhibitors of the p6–UEV PPI by transforming a previously reported RTHS for this PPI33 with the library-encoding plasmids. At present, the transformation efficiency of the selection strain and the relatively large size of the plasmid limited the screening for inhibitors to ~10^6 unique peptides. In future efforts, it may be possible to engineer a chromosomal copy of ProcM to decrease the plasmid size and increase the transformation efficiency. After secondary screening, a single inhibitor (XY3-3) was identified that selectively prevented the targeted PPI in the RTHS. The activity of XY3-3 was verified in a series of *in vitro* experiments. Removal of the leader sequence from XY3-3 improved its activity, demonstrating that the leader sequence plays no part in target engagement, and instead appears to hinder optimal binding. It should be noted that the sequence of XY3-3 is very different from that of the PTAP motif of p6 that is critical for its binding to UEV40. In fact, none of the amino acids in this nonapeptide (Pro, Glu, Thr, and Ala) are encoded by the NWY codons used in the randomized regions in our library. Given the central role played by this nonapeptide in the p6–UEV PPI40, 41, and considering the data from our competition FP assays, XY3-3 likely binds at a different position than the PTAP peptide, the “hotspot” for the p6–UEV PPI. We were unable to determine the precise site of binding by crystallography or NMR spectroscopy, in part because of the relatively low solubility of the peptide. Nonetheless, our *in vitro* and cell-based data suggest that XY3-3 is a functional inhibitor of the p6–UEV PPI. These findings have implications for inhibitor discovery approaches that focus on a single, known hotspot on a PPI, and illustrate the advantages of unbiased, functional screening. The results with the inverted sequences of the A and B rings suggest that both rings are making contact with UEV. Compared to a previously identified monocyclic inhibitor33, the bicyclic XY3-3 has about a 10-fold improved affinity, which is

modest compared to the expected increased affinity upon having two binding moieties47, suggesting that the sequence of XY3-3 is not optimal. Typical affinity maturation methods are not readily applied to RiPPs, but are under development to identify variants that take full advantage of two high affinity interactions.

Not unexpectedly, XY3-3 was not cell permeable, but conjugation of a Tat-tag enabled the activity of this molecule to be assessed in human cells. In line with the mode of action displayed *in vitro,* XY3-3-Tat inhibited virus-like particle formation in an HIV GAG-mediated, cell-based viral budding assay. The compound also prevented TSG101-mediated degradation of EGFR in HeLa cells, in line with our *in vitro* data showing XY3-3 acts by binding the UEV domain of TSG101. TSG101 function is critical for cell survival43, 44 providing a likely explanation for the observed toxicity of XY3-3-Tat to cells at higher doses. However, toxicity resulting from off-target effects cannot be ruled out. Regardless, this study illustrates the successful generation and screening of a lanthipeptide library in a cell-based assay, with the activity of the identified hit being verified *in vitro* and in cells. Given the range of other scaffolds that can be generated from the 30 different peptide substrates of ProcM (Supplementary Fig. 1) and the high quality of the library generated from ProcA2.8 (including XY3-3 and the two inverted ring sequences, 36 out of 36 examined peptides were fully dehydrated and cyclized), our studies suggest the possibility of generating lanthipeptide libraries with a wide range of structural and chemical diversity. In addition to its use with a RTHS, this lanthipeptide library may be interfaced with a variety of other cell-based assays for the identification of biologically active lanthipeptides. Whereas previous studies have used epitope grafting to bestow new biological activities onto existing RiPPs48, 49 or installation of RiPP-type crosslinks in known biologically active peptides50, this study utilizes the RiPP biosynthetic machinery for the discovery of entirely new biological activities, further underscoring the enormous possibilities of the enzymes that produce this class of natural products.

Online Methods

General methods

For lanthipeptide residue labelling, positive residue numbers are used for amino acids in the core peptide counting forwards from the (putative) leader peptide cleavage site. Negative residue numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site. For instance, in the ProcA2.8 precursor peptide, the C-terminal Gly residue of the leader sequence is −1, and when substituted by Lys the mutant is denoted as G–1K. Polymerase chain reaction (PCR) amplifications were carried out using an automated thermocycler (C1000, BioRAD). DNA sequencing was performed using appropriate primers by ACGT Inc. LC-ESI-Q/TOF MS analyses were conducted using a Synapt MS system equipped with an Acquity Ultra Performance Liquid Chromatography (UPLC) system (Waters). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analyses were conducted at the Mass Spectrometry Facility at UIUC using an UltrafleXtreme spectrometer (Bruker Daltonics). For MALDI-TOF MS analysis, samples were desalted using ZipTipC18 (Millipore), and spotted onto a MALDI target plate with a matrix solution containing 35 mg/mL 2,5-dihydroxybenzoic acid

(DHB) in 3:2 MeCN/H_2O with 0.1% TFA. Peptides were desalted by a C4 solid-phase extraction (SPE) column and further purified by preparative reversed-phase high performance liquid chromatography (RP-HPLC) on a Delta 600 instrument (Waters) equipped with a Phenomenex C18 column at a flow rate of 8 mL/min. For RP-HPLC, solvent A was 0.1% TFA in H_2O and solvent B was 4:1 MeCN/H_2O containing 0.086% TFA. An elution gradient from 0% solvent B to 74% solvent B over 40 min was used unless specified otherwise.

Materials

All chemicals used were purchased from Sigma Aldrich or Fisher Scientific unless noted otherwise. Endoproteinases Glu-C and Lys-C were purchased from Roche Applied Science. Oligonucleotide primers used for molecular cloning were purchased from Integrated DNA Technologies. The pBAD/HisA vector was purchased from Invitrogen. Electrocompetent *E. coli* DH5α cells, Phusion High-Fidelity DNA polymerase, Taq ligase, dNTP solutions, T4 DNA ligase and all restriction endonucleases were purchased from New England Biolabs. Gel extraction, plasmid miniprep, and PCR purification kits were purchased from QIAGEN. Protein Calibration Standard I and Peptide Calibration Standard II for MALDI-TOF MS were purchased from Bruker. *E. coli* DH5α was used as host for cloning and plasmid propagation, and *E. coli* BL21 (DE3) was used as a host for overexpression.

Statistical analysis

Data acquired from enzyme-linked immunosorbent assay (ELISA), Microscale Thermophoresis (MST) and Fluorescence Polarization (FP) experiments were processed by Prism 6 (Graphpad) unless otherwise indicated. EC_{50}, IC_{50} and dissociation constants were calculated using a non-linear regression model from three independent replicates. Errors are the standard error of the mean (S.E.M.) given by regression analysis. Data shown are mean ± S.E.M. from the three replicates.

Construction of lanthipeptide library in pRSFDuet-1 vector

Library primers were designed that added an EcoRI restriction site to the 5' end (Lib_EcoRI_FP, Supplementary Table 2) and a NotI restriction site to the 3' end of the construct. The reverse primer (Lib2.8NWY_NotI_RP, Supplementary Table 2) containing degenerate codons was synthesized by IDT using hand-mixed bases and purified by PAGE. The PCR template was constructed by inserting the procA2.8 leader sequence (amplified with primer Lib_EcoRI_FP and 2.8L_NotI_RP, Supplementary Table 2) into MCS-1 of pRSFDuet-1 plasmid between EcoRI and NotI restriction sites. This plasmid was used as template for PCR generation and amplification of the library. PCR was carried out using Phusion polymerase by 24 cycles of denaturing (94 °C for 30 s), annealing (67 °C for 30 s) and extending (72 °C for 20 s). The PCR products (appearing as a strong band between 250-300 bp) were gel-purified, digested with EcoRI and NotI restriction enzymes and purified using a Qiagen PCR purification kit.

The pRSFDuet-1 vector with *procM* inserted into MCS2 between NdeI and KpnI37 was used for library vector construction, the vector was digested with the same restriction enzymes as the PCR products and an additional enzyme SbfI to minimize vector self-

ligation, dephosphorylated with alkaline phosphatase and gel-purified. The digested vector (50 ng) was ligated with the insert DNA by T4 ligase at a molecular ratio of 1/6 (vector/insert), and the reactions were incubated at 15 °C overnight. The DNA product was desalted by butanol precipitation and used to transform electro-competent *E. coli* DH5α cells (New England Biolabs). After recovery in 1 mL of SOC medium (for each transformation) for one hour, 5 μL of the combined culture was plated as serial dilutions on LB-agar containing kanamycin (50 mg/L) to calculate the library size. The rest of the recovery culture was diluted in 10-fold volume of LB containing kanamycin (50 mg/L) and the library plasmids were isolated using a Qiagen Miniprep Kit after overnight incubation at 37 °C. Five ligation reactions described above were combined in order to scale up the library size. After transformation, the library size was approximately 10^6, calculated by dilution plating.

Construction of a library vector compatible with the RTHS strain

The dual-arabinose promoter vector pARDuet was constructed based on the pACYCDuet-1 backbone. pBAD/HisA plasmid was used as template for PCR amplification of the *araC* and P_{ara} region (primer araC_pBAD_NarI_FP and araC_pBAD_NcoI_RP, Supplementary Table 2). The PCR products were digested with NarI and NcoI and ligated into digested pACYCDuet-1 vector, which replaced the *lacI* with *araC* and the *T7* promoter in MCS-1 with P_{ara}. Next, the P_{ara} sequence was PCR-amplified from pBAD/HisA (primer Para_pBADHis_BsrGI-FP and Para_pBADHis_NdeI-FP, Supplementary Table 2), the PCR product was digested with BsrGI and NdeI, and cloned into the vector obtained in the first step, which replaced the *T7* promoter in MCS-2 with P_{ara}. Finally, the *rrnB* terminator sequence from pBAD/HisA was PCR-amplified (primer araTer_XhoI_FP and araTer_Bsu36I_RP, Supplementary Table 2), digested with XhoI and Bsu36I and cloned into the resulting vector, giving the empty library vector named pARDuet (Supplementary Fig. 6, Supplementary Table 3).

Construction of lanthipeptide libraries using the pARDuet vector

The libraries for genetic selection do not require the N-terminal His_6-tag. Therefore, a new forward library primer was designed that added an NcoI restriction site (closest to the start codon to eliminate unnecessary N-terminal residues) at the 5’ end (primer Lib_NcoI_FP, Supplementary Table 2). The reverse library primer remained the same as that used for the library in pRSFDuet. *procM* was cloned into the MCS-2 of pARDuet between NdeI and KpnI. The two NcoI cut sites inside the *procM* gene were deleted by silent mutations via site-directed mutagenesis (primer pMQC_1t2577c_FP, pMQC_1t2577c_RP, pMQC_2c2772t_FP and pMQC_2c2772t_RP, Supplementary Table 2), resulting in library vector ProcM(t2577c/t2775c)_pARDuet. The plasmid library with ProcA2.8 containing randomized core peptides was constructed as described above, except that NcoI and NotI were used as the restriction sites and chloramphenicol was used instead of kanamycin as selective marker. After PCR, digestion with restriction endonucleases, ligation, and transformation of *E. coli* DH5α cells (NEB), the plasmid library was isolated and plasmids were eluted in water during plasmid preparation for maximum electroporation efficiency in the next step.

Construction of pARDuet derivative for co-expression of His_6-tagged peptides

To test if the pARDuet vector could also produce a modified peptide library like the pRSFDuet vector, the gene encoding His_6-tagged ProcA2.837 was amplified and the PCR product digested with NcoI and NotI and inserted into MCS-1 of the ProcM(t2577c/t2775c)_pARDuet vector between the same restriction sites. This plasmid was tested using the expression conditions described below.

Over-expression of modified library peptides in *E. coli*

In order to test the modification efficiency of the library peptides, *E. coli* BL21 (DE3) cells were transformed with the previously-mentioned pRSFDuet derivatives encoding His_6-tagged substrates and ProcM, and colonies were randomly chosen from the transformation plate and inoculated into fresh medium for over-expression. The cell culture (50-100 mL LB) was induced with 0.25 mM IPTG at an OD_{600} between 0.6-0.8, and incubated at 18 °C for 18-20 h. MALDI-TOF MS analysis demonstrated that fully modified peptides were produced (Supplementary Fig. 4 and 5). For determining the quality of ProcM catalysis, in total 36 non-native ProcA derivatives were analyzed and all 36 were dehydrated and cyclized. With this sample size, at the 95% confidence level, the Clopper-Pearson exact lower confidence limit regarding the library quality is 90%51. In order to test the expression conditions for genetic selection, the co-expression plasmid containing genes encoding WT ProcA2.8 and ProcM37 was then used at different temperatures in 0.5-2 L minimal medium A, which showed that expressed ProcA2.8 could be fully dehydrated at temperatures up to 30 °C (Supplementary Fig. 7a).

To test the overexpression and modification using the pARDuet vector, after induction with 6.5 mM arabinose (note this concentration is higher than that used for selection, see below, in order to obtain sufficient amounts of peptide for analysis), the cells were cultured in LB or minimal medium A containing 25 mg/L chloramphenicol at 18 °C in the *E. coli* RTHS selection strain or in *E. coli* BL21 (DE3) cells. MALDI-TOF MS analysis demonstrated that fully modified peptides were produced (Supplementary Fig. 7b and c).

Sequencing of the plasmid library

Sequencing primers were designed to amplify the randomized portion of the pARDuet-derived plasmid library with Illumina adapter overhang nucleotide sequences per Illumina's protocol (Lib2.8NWYseq_FP and Lib2.8NWYseq_RP, Supplementary Table 2). PCR was carried out using Phusion polymerase by 15 cycles of denaturing (94 °C for 30 s), annealing (66 °C for 1 min) and extending (72 °C for 15 s). The PCR products were purified by Qiagen PCR purification kit and subjected to the following Index PCR. Nextera XT Index N701 and S517 (Nextera XT index kit FC-131-1001) were used to amplify the products of the first PCR using Phusion polymerase with 8 cycles of denaturing (94 °C for 30 s), annealing (66 °C for 1 min) and extending (72 °C for 30 s). The PCR products (appearing as a strong band at 200-250 bp) were gel-purified with Qiagen gel purification kit. DNA concentration was quantified with a Qubit® 2.0 fluorimeter (Invitrogen) using a Qubit HS dsDNA HS assay kit. Deep sequencing was performed in the Roy J. Carver Biotechnology Center (University of Illinois at Champaign-Urbana) on an Illumina HiSeq 2500.

The Illumina reads obtained were processed with in-house Python scripts (freely available directly from the authors upon request). Reads containing the sequence GGAGCG, which encodes the last residue of the leader and the first residue of the core, were identified and the quality scores for those bases and the following 57 (the rest of the core peptide including the stop codon) were analysed. Any read not containing the sequence or with a base with a quality score less than 30 in the examined region were discarded. For paired end sequencing, the reverse reads were searched for the reverse complement of the above sequence and the same quality score cut off was applied to the preceding 57 bases. If any base in the examined region of the forward or reverse reads dropped below the quality cut off, or if the sequence of examined region was not identical in the forward and reverse reads, both reads were discarded. The remaining reads were translated and the number of times a given peptide sequence was represented in the library were tabulated to determine the diversity of the library.

Library selection in the UEV-P6 RTHS strain

Each selection plate (150 mm diameter × 25 mm) contained 40 mL of M9 minimal media supplemented with 25 mg/L kanamycin, 2.5 mM 3-AT (3-amino-1,2,4-triazole), 30 mg/L carbenicillin, 25 mg/L spectinomycin, 35 mg/L chloramphenicol, 1 mM $MgSO_4$, 65 μM arabinose, 30 μM IPTG, 2% glycerol and 1.5% agar. A negative control plate was also made that omitted IPTG. UEV-p6 RTHS cells were cultured overnight in LB medium containing 30 mg/L carbenicillin and 25 mg/L spectinomycin, an aliquot of 1 mL cell culture was inoculated into 100 mL of the same medium and the cells were grown until the OD_{600} reached 0.5. The cells were quickly chilled in an ice-water bath and incubated on ice for 40 min before harvesting (3000 g, 15 min, 2 °C). The supernatant was discarded and the cell pellets were gently resuspended in 20 mL of ice-cold 10% glycerol and harvested again. The washing step was repeated two more times and the resulting electro-competent cell pellets were resuspended with a final total volume of 200-250 μL and split into four tubes. These freshly made competent cells were transformed (1 mm cuvette, 1.8 kV) with the library plasmids (4 × 25 ng), and 0.5 mL of SOC medium pre-warmed at 37 °C was immediately added into each transformation. After recovery for 1 h at 37 °C, the culture was centrifuged and the supernatant discarded. An aliquot of 0.5 mL of minimal medium A was added into each tube and the cell pellets were thoroughly resuspended by gentle pipetting. An aliquot of 5 μL of the combined culture was plated as serial dilutions on LB-agar containing chloramphenicol (35 mg/L) to calculate the library size. The rest of the culture was evenly spread onto the selection plates until all liquid was completely absorbed into the plates.

The selection plates were sealed with parafilm and incubated at 30-33 °C for 2-4 days. The growth was compared with cells on the control plate without IPTG, and a growth delay was observed on plates with IPTG during this time window. Visible colonies started to show up starting on the second day, and normally around the fourth day the background colonies started to grow due to IPTG degradation, antibiotic resistance, or mutations.

Eighty identified colonies that exhibited faster growth than the background colonies were inoculated into individual wells of a 96-well culture plate with 1 mL LB medium containing 30 mg/L carbenicillin, 25 mg/L spectinomycin and 35 mg/L chloramphenicol and the plate

was incubated with constant shaking at 37 °C overnight. Then, using a multi-channel pipette, each culture was sequentially diluted (10-fold each time) into M9 minimal media.

Four sets of plates were made from identical selection medium as described previously, but with or without IPTG or arabinose. Then 2.5 μL of each dilution culture (containing roughly 10^6, 10^5, 10^4 and 10^3 cells) from each sample were drop-spotted onto each plate. The plates were sealed with parafilm and incubated at 30-33 °C for 2-3 days. Growth difference was monitored by comparing the samples on the four plates at each dilution level, and two colonies that exhibited both IPTG-dependent inhibition and arabinose-dependent growth advantage were selected. The UEV-p6 strain and another heterodimeric RTHS strain (in which UEV and p6 were replaced with an unrelated protein-protein interaction of CMG2 with anthrax toxin protective antigen, PA) were transformed with these plasmids to confirm the specificity of the observed arabinose-dependent growth advantage. Each culture was sequentially diluted and drop-spotted onto the same four sets of selection plates with or without IPTG or arabinose as described above. A single plasmid was active in the UEV-p6 RTHS with no effect in the CMG2/PA RTHS. Sequencing of the entire plasmid revealed the identity of the isolated hit (XY3-3, Supplementary Table 4) without any other mutations in the plasmid that could have conferred the growth advantage.

Overexpression of the XY3-3 peptide

The pRSFDuet vector system was chosen for over-expression and characterization of the purified hit peptide. The gene was amplified by PCR from the pARDuet vector isolated from the hit colony described above and cloned into the MSC-1 of procM_pRSFDuet37 between the EcoRI and NotI sites. A thrombin cleavage site was also introduced between the His_6-tag and the peptide sequence in order to remove the tag after overexpression (primer 2.8Thrombin_EcoRI_FP and Lib_NotI_RP, Supplementary Table 2).

Purification of modified His_6-tagged ProcA2.8 derived peptides

After harvesting, the cell pellet was resuspended at 0.1 g/mL in LanA Start Buffer (20 mM NaH_2PO_4, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol, pH=7.5). The cell paste was subjected to sonication to lyse the cells. Insoluble cell debris was removed by centrifugation at 16,500 × g for 45 min. The supernatant was purified by immobilized metal affinity chromatography (IMAC) using 2-4 mL of His60 Ni Superflow Resin (Clontech). Following 1 h incubation at room temperature, the resin was washed with wash buffer (Start Buffer + 30 mM imidazole), and the peptide was eluted from the resin using elution buffer (start buffer + 4 M guanidine hydrochloride + 0.5 M imidazole). The elution fractions were desalted by C4 SPE column and purified on a C18 column using the solvent gradient described in the General Methods.

Expression and Purification of His_6-UEV and GST-p6

The gene encoding human UEV (ubiquitin E2 variant) was cloned into a pET28a vector using the NheI and XhoI restriction sites. The viral p6 gene was cloned into PGEX-2TK using the EcoRI and BamHI restriction sites. Both recombinant proteins were expressed in *Escherichia coli* BL21 (DE3) cells. Bacterial cultures were incubated at 37 °C with shaking until an OD_{600} of between 0.5 and 0.7 was reached. Protein expression was induced with 0.1

mM IPTG overnight at 16 °C. After harvesting, the cell pellets were suspended at 1 g/mL in Buffer A (20 mM Tris-HCl, 100 mM NaCl, 10% glycerol, pH 7.4) supplemented with 1% (v/v) Triton X-100, Pierce Protease Inhibitor tablets (ThermoFisher Scientific, 1 tablet per 20 mL) and lysozyme (100 μg/mL), and the cells were incubated on ice for 30 min. The lysate was sonicated (12 cycles of 10 s on followed by 10 s off) on ice to ensure complete lysis. Insoluble cell debris was pelleted via centrifugation at 22,700 × g for 45 min at 4 °C. The supernatant was filtered using 0.45 μm syringe filters. For His_6-UEV, the supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare) equilibrated in Buffer A coupled to an ÄKTA Prime fast protein liquid chromatography (FPLC) system (GE Healthcare). The protein was loaded onto the column at a flow rate of 1 mL/min. The column was washed with 30 mL of Buffer A and then with a gradient to 8% Buffer B (500 mM imidazole, 20 mM Tris-HCl, 100 mM NaCl, pH 7.4) in Buffer A at a flow rate of 1 mL/min, then the protein was eluted in 100% Buffer B. UV absorbance at 280 nm was monitored and fractions were collected and analysed by SDS-PAGE.

For GST-p6, the filtered supernatant after cell lysis was loaded onto a 1 mL GSTrap column (GE Healthcare) coupled to an ÄKTA Prime FPLC system (GE Healthcare) pre-equilibrated in Buffer A at a flow rate of 1 mL/min. The protein was loaded onto the column at a flow rate of 0.5 mL/min. The column was washed with 30 mL of Buffer A, and the GST-p6 was eluted with 10 mL GST Elution buffer (50 mM Tris-HCl, 150 mM NaCl, 15 mM reduced glutathione, pH 7.4) at a flow rate of 1 mL/min.

For both proteins, the fractions containing the desired protein were combined and dialysed overnight at 4 °C using SnakeSkin™ Dialysis Tubing (10K MWCO 22 mm; ThermoFisher Scientific) into Buffer A (1 L/2 mL protein). The buffer was exchanged for fresh buffer and allowed to dialyse for another 3 h. His_6-UEV was obtained with a yield of 20 mg/L cell culture, and GST-p6 was obtained with a yield of 10 mg/L cell culture. GST-p6 was subjected to thrombin cleavage at 4 °C overnight and purified with an FPLC system with a Superdex 200 10/300 GL size exclusion column (GE Healthcare) using buffer A. Protein aliquots were flash-frozen in liquid nitrogen and stored at −80 °C.

Endoproteinase digestion and purification of unmodified or modified peptides

For peptides prepared by *E. coli* co-expression, the majority of the leader peptide needs to be removed in order to accurately characterize the dehydrations in the core peptide by MS. Library samples were digested with Glu-C or Lys-C. XY3-3 (peptide **1)** was digested with thrombin to remove the His_6 tag, or with Lys-C, or Glu-C in order to obtain the cyclized core peptides with different leader length (peptide **2** and peptide **3**, respectively). The proteolysis reaction also contained 0.5 mM TCEP to reduce potential disulfide bonds in the core peptide. In contrast, when trying to remove the leader sequence of unmodified ProcA2.8 and linear unmodified XY3-3 by Glu-C, oxidative disulfide bond formation was performed in order to protect the Glu and Asp residues in the core peptide sequence from Glu-C cleavage. The linear peptide (150 μM) was treated with 5 mM each of reduced and oxidized glutathione in 50 mM HEPES buffer for 1 h at room temperature, before Glu-C was added. Purification of the digested sequence was carried out by RP-HPLC on an analytical Phenomenex C18 column, using the gradient described in the General Methods.

N-Ethylmaleimide (NEM) or iodoacetamide (IAA) alkylation assay to test for cyclization

An aliquot of the protease-digested peptide or full-length peptide solution was diluted into twice the volume of NEM alkylation buffer containing 500 mM HEPES, 3 mM NEM, 0.3 mM TCEP, pH 6.5, or IAA alkylation buffer containing 500 mM HEPES, 3 mM IAA, 0.3 mM TCEP, pH 8.0. The reaction was incubated at 37 °C for 30 min (for NEM alkylation) or at room temperature for 1-2 h (for IAA alkylation) in the dark, and analysed by MALDI-TOF MS. Species containing uncyclized free Cys residues that were alkylated were identified by a mass increase of 125 Da (for NEM) or 57 Da (for IAA) for each adduct.

ELISA assay characterizing the disruption of UEV-p6 interaction by lanthipeptides

To each well in a clear, nickel coated 96-well plate (Thermo Scientific), purified His_6-UEV protein diluted in PBS (Thermo Scientific; 1 tablet dissolved in 100 mL deionized water; at a final concentration of 1 μM of His_6-UEV; 100 μL per well), was added and incubated for 1 h at room temperature with rocking. The wells were then washed with PBS containing 0.05% TWEEN 20 (Sigma; 3 × 200 μL, 5 min incubation per wash). The wells were blocked for 1 h with 2% milk solution in deionized water (150 μL per well) at room temperature with rocking. Without washing the wells, descending concentrations of lanthipeptide (serially diluted in 100% DMSO, 10 μL) with 0.5 μM of purified GST-tagged p6 protein in PBS (90 μL) were added and the plate was incubated for 1 h at room temperature with rocking (10% final concentration of DMSO). The wells were then washed as before, and mouse anti-GST Ab-1 antibody (Neomarkers; 1:1000 made up in 2% milk in deionized water; 100 μL per well) was added and incubated for 1 h at room temperature with rocking. The wells were washed, then sheep anti-mouse IgG, peroxidase-linked whole antibody (GE Healthcare; 1:3000 made up in 2% milk in deionized water; 100 μL per well) was added and the plate was incubated for 1 h at room temperature with rocking. The wells were washed and 1-Step™ Ultra TMB ELISA solution (Thermo Scientific; 100 μL) was added to each well and the plate was incubated for 15 min at room temperature with rocking. To quench the reaction, 20 mM sulphuric acid (100 μL) was added to each well and the absorbance at 450 nm measured using a Tecan Infinite® M200 PRO plate reader. Data was processed with Prism 6 (GraphPad); EC_{50} and IC_{50} values were determined using a non-linear regression model from three replicates.

Fluorescein labelling of XY3-3

For fluorescein labelling of XY3-3 using fluorescein-5-maleimide (FM), the pRSFDuet plasmid with *procA* in MCS1 and *procM* in MCS2 was altered such that a Cys residue was inserted on the N-terminus of the leader peptide after a thrombin cleavage site (LVP**R**GC, thrombin cleaves after the Arg residue (bolded) in the underlined recognition sequence). The peptide sequence was PCR-amplified using primers 2.8ThrombinCys_EcoRI_FP and Lib_NotI_RP (Supplementary Table 2), and cloned between the EcoRI and NotI restriction sites in MCS1 of the pRSFDuet vector containing *procM* in MCS2. The peptide was overexpressed in *E. coli* and purified as described previously. The additional Cys residue did not interrupt the modification of the core peptide, as two non-overlapping rings still formed (Supplementary Fig. 14). The peptide was then subjected to thrombin cleavage. Thrombin was used to remove the N-terminal His_6-tag after pro-longed incubation time, and the

resulting untagged Cys-XY3-3 peptide was purified by HPLC on a C4 column. The fluorescein labelling reaction contained 20 μM purified peptide, 50 μM TCEP, 0.5 mM fluorescein-5-maleimide and 50 mM HEPES, pH 6.9. The reaction was incubated at room temperature for 3-4 h in the dark and monitored by MALDI-TOF MS (Supplementary Fig. 14). After C4 SPE desalting, the labelled peptide was further purified by HPLC on a C4 column as described under General Methods.

Fluorescent peptide pull-down

His-tagged p6 and UEV (5 μM protein in 45μL buffer A) were immobilised into separate wells of a Ni^{2+}-coated 96-well plate. Fluorescein-tagged XY3-3 (5 μL of a 100 μM solution) was added to each well and incubated with shaking at room temperature for 10 minutes. The wells were washed with buffer a (3 × 200 μL). The fluorescence of each well was measured using a BMG Labtech ClarioStar plate reader.

FITC labelling of peptide 3 and PTAP

HPLC-purified peptide **3** was dissolved in PBS buffer (pH=8.0) to a final concentration of 50 μM. FITC (Thermo Fisher) was dissolved in DMSO (10 mg/mL) and added slowly to the peptide solution to a final concentration of 250 μM. The reaction was incubated for 18 h at 25 °C. After centrifugation at 16,100 × g for 10 min, the supernatant was purified on an Agilent 1260 Infinity HPLC system with a Phenomenex Luna C18 column (250 mm × 4.6 mm, 10 μm particle size, 100 Å pore size). Solvent A was 0.1% TFA in H_2O and solvent B was 80% MeCN/H_2O containing 0.086% TFA. An elution gradient from 0% solvent B to 100% solvent B over 45 min at 1 mL/min was used and FITC labeled XY3-3 core peptide eluted at 66-67% solvent B. PTAP (GGPEPTAPPEE) peptide containing two additional N-terminal Gly residues (synthesized by United Biosynthesis) was labelled with FITC and purified as described above. The FITC labelled PTAP peptide eluted at 48-49% solvent B.

Microscale Thermophoresis (MST)

The purified recombinant His_6-UEV protein was labelled with the red NT-647-Maleimide dye using a protein-labelling kit (NanoTemper Technologies) according to the manufacturer's instructions. The labelled protein was then purified using a gravity flow column into MST assay buffer (20 mM HEPES, pH 7.4, 100 mM NaCl, 5% glycerol, 0.01% Tween-20). The inhibitor peptide (XY3-3, peptide **3**) was dissolved in 100% DMSO and then diluted 5-fold into MST buffer. The inhibitor was serially diluted in MST assay buffer supplemented with 20% DMSO (5 μL) and then mixed with the labelled protein 1:1 (10 μL final volume, 0.01 μM final concentration of protein, final DMSO concentration 10%). The samples were incubated at room temperature for 5 min and then loaded into MST premium coated capillaries (NanoTemper Technologies). The samples were analysed on a Monolith NT.115 MST Instrument (NanoTemper Technologies, Excitation Power 20%, MST Power 40%). Data points were excluded if the initial fluorescence was outside of the 10% error range as recommended by the manufacturer. Data was processed with MO.Affinity.Analysis (NanoTemper) and Prism 6 (GraphPad). Binding constants were determined using a non-linear regression model from three replicates.

Fluorescence polarization of UEV binding to peptide 3 and PTAP peptide

Purified recombinant His_6-UEV protein was diluted into binding buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 20% glycerol, 0.5 mM TCEP) and mixed with 150 nM FITC-labeled peptide **3** or FTIC-PTAP. Binding assays were incubated at 25 °C for 4 h and analyzed using a Synergy H4 Hybrid plate reader with λ_{ex} = 485 nm and λ_{em} = 538 nm using in triplicate. Then each experiment was repeated two more times for a total of three independent experiments. Dissociation constants were calculated from the 50% saturation point using a dose-response curve fit in Prism 6 (GraphPad) with three replicates. A blank was set up with protein with the FTIC fluorophore only.

Fluorescence polarization competition binding assay

Recombinant His_6-UEV protein 15 μM/4 μM was mixed with 1.5 μM FITC-PTAP/400 nM FITC-XY3-3 in binding buffer and incubated at 25 °C for 4 h to reach equilibration. Unlabelled peptide 3/PTAP peptide and thrombin cleaved GST-p6 were serially diluted in binding buffer and added to the pre-incubated mix of UEV with FITC-labelled PTAP/ peptide **3**. Binding assays were analysed on the same instruments and analysed with the same software as described above with three replicates.

Generation of XY3-3-Tat

For Tat labelling of peptide **3**, the Gly residue at the −5 position of XY3-3 was mutated to Cys using primers XY33G-5C_FP and XY33G-5C_RP (Supplementary Table 2, Supplementary Figure 12) in a pRSFDuet plasmid with the gene encoding XY3-3 in MCS1 and *procM* in MCS2. The peptide was overexpressed in *E. coli* and purified as described previously. The resulting peptide contained two non-overlapping rings, as desired, suggesting the mutation did not interrupt the modification process (Supplementary Fig. 17). The peptide was then digested by endoproteinase GluC and purified as described previously. The Tat-peptide (CGRKKRRQRRRPPQ) was synthesised using Fmoc solid-phase peptide synthesis on a Glutamine Wang Resin (Matrix Innovation). The cysteine side chain was activated by reacting with aldrithiol-4 (2 equivalents, 10 mL DMF, 16 h, Sigma). The Glu-C-cleaved peptide was dissolved in 3:1 DMF:H_2O (2 mL) with aldrithiol-activated Tat (3.5 equivalents) and stirred under argon for 18 h. The resulting peptide (named XY3-3-Tat) was precipitated in ice-cold diethyl ether and dried under argon.

Mammalian cell culture and transfection

HEK293T cells were maintained in DMEM Glutamax supplemented with antibiotics and 10% FBS (Life Technologies). HeLa cells were maintained in DMEM supplemented with 10% FBS. Plasmid DNA transfections were performed using Lipofectamine 2000 (Invitrogen).

Antibodies and Western Immunoblotting

To visualise expression of transfected proteins, the cells were harvested in cold PBS (ThermoFisher) and then centrifuged (1000 rpm, 4 min, r.t.). The cells were resuspended in RIPA Buffer with protease inhibitors (ThermoFisher) and incubated on ice for 5 min. The cell lysates were then sonicated in an ice water bath (12 cycles of 30 s on, 30 s off) before

being centrifuged (10,000 rpm, 20 min, 4 °C) and then the protein concentration in the supernatant was quantified by Bradford Assay. The proteins were separated on a 12% SDS-PAGE gel (v/v) under denaturing conditions (160 V, 50-70 min), then transferred to a nitrocellulose membrane (GE Healthcare, 250 mA, 2 h). The membrane was blocked with 5% (w/v) non-fat powdered milk with 0.1% Tween-20 (v/v) in PBS and then subjected to immunoblot analysis. The primary antibodies, either anti-eGFP (MA1-952, 1:1000, ThermoFisher), or anti-EGFR (ab252894, 1:1000, Abcam), were diluted in 5% (w/v) non-fat powdered milk with 0.1% Tween-20 (v/v) in PBS and incubated with the membrane overnight at 4 °C. Horseradish peroxidase conjugated anti-mouse (NA931, 1:100,000, GE Healthcare) or anti-rabbit (7074, 1:50,000, Cell Signalling) antibodies were used as the secondary antibodies for anti-eGFP and anti-EGFR respectively. Monoclonal anti-βactin-peroxidase antibody (A3854, 1:100,000, Sigma) was used as a loading control. Bound immunocomplexes were detected using ECL prime Western blot detection reagent (RPN2232, GE Healthcare) and analyzed using a ChemiDoc Imaging System (Bio-Rad) and Image Lab 4.0 Software (Bio-Rad).

Virus-like particle assay

HEK293T cells were seeded at 25,000 cells/well in a 24-well plate and incubated overnight such that the cells reached 60-70% confluency prior to transfection. The cells were transfected with Lipofectamine 2000 and pMET7-GAG-EGFP (Addgene) in a 4:1 ratio in Opti-MEM (Life Technologies). After 6 h, the cells were either mock-treated with control or Tat-peptide, or treated with increasing concentrations of XY3-3-Tat. After 18 h, the supernatants were collected and the cells washed and lysed prior to immunoblotting analysis. The supernatants were centrifuged (1000 rpm, 4 min), and then filtered through a 20% (w/v) sucrose cushion (12,000 rpm, 90 min, 4 °C). The filtered VLPs (final 15 μL) were resuspended with Laemmli Buffer (5 μL) and then subjected to SDS-PAGE and immunoblotting analysis.

Cytotoxicity assay

The cytotoxicity was investigated using the CytoTox96® Non-Radioactive Cytotoxicity Assay Kit (Promega) used according to the manufacturer's guidelines and the absorbance measured using a BMG CLARIOstar Plate reader.

EGFR down-regulation assay

HeLa cells were seeded at 150,000 cells/well in a 6-well plate and incubated overnight such that the cells reached 60-70% confluency. Cells were either treated with Tat peptide (0.5 μM) or XY3-3-Tat (0.5 μM) in DMEM Glutamax supplemented with 10% FBS and incubated for 18 h. The cells were then starved in serum-free DMEM Glutamax. After 1 h, the cells were either mock-treated or treated with EGF (150 ng/mL, Sigma) for 90 min. The cells were washed with cold PBS and immediately lysed and subjected to SDS-PAGE and immunoblotting analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Prof. Sven Eyckerman for pMET7-GAG-EGFP (via Addgene, plasmid # 80605), and Dr. Diego Gomez-Nicola for HEK239T cells. This work was supported by the National Institutes of Health (R37 GM 058822 to W.A.V.; F32 1 F32 GM0112284 to M.C.W.), Cancer Research UK (A20185 to A.T.), and the Engineering and Physical Sciences Research Council and C4X Drug Discovery (EP/L505067/1, Ph.D. studentship for K.R.L. to A.T.). Data underlying these experiments are available from the University of Southampton data repository at http://doi.org/10.5258/SOTON/D0133.

References

1. Marsault E, Peterson ML. Macrocycles are great cycles: applications, opportunities, and challenges of synthetic macrocycles in drug discovery. J Med Chem. 2011; 54:1961–2004. [PubMed: 21381769]
2. Driggers EM, Hale SP, Lee J, Terrett NK. The exploration of macrocycles for drug discovery--an underexploited structural class. Nat Rev Drug Discov. 2008; 7:608–624. [PubMed: 18591981]
3. Cardote TA, Ciulli A. Cyclic and macrocyclic peptides as chemical tools to recognise protein surfaces and probe protein-protein interactions. ChemMedChem. 2015; 11:787–794. [PubMed: 26563831]
4. Gao M, Cheng K, Yin H. Targeting protein-protein interfaces using macrocyclic peptides. Biopolymers. 2015; 104:310–316. [PubMed: 25664609]
5. Lennard KR, Tavassoli A. Peptides come round: using SICLOPPS libraries for early stage drug discovery. Chemistry. 2014; 20:10608–10614. [PubMed: 25043886]
6. Heinis C, Rutherford T, Freund S, Winter G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. Nat Chem Biol. 2009; 5:502–507. [PubMed: 19483697]
7. Miranda E, et al. A cyclic peptide inhibitor of HIF-1 heterodimerization that inhibits hypoxia signaling in cancer cells. J Am Chem Soc. 2013; 135:10418–10425. [PubMed: 23796364]
8. Passioura T, Katoh T, Goto Y, Suga H. Selection-based discovery of druglike macrocyclic peptides. Annu Rev Biochem. 2014; 83:727–752. [PubMed: 24580641]
9. Heinis C, Winter G. Encoded libraries of chemically modified peptides. Curr Opin Chem Biol. 2015; 26:89–98. [PubMed: 25768886]
10. Birts CN, et al. A cyclic peptide inhibitor of C-terminal binding protein dimerization links metabolism with mitotic fidelity in breast cancer cells. Chem Sci. 2013; 4:3046–3057.
11. Tavassoli A. SICLOPPS cyclic peptide libraries in drug discovery. Curr Opin Chem Biol. 2017; 38:30–35. [PubMed: 28258013]
12. Arnison PG, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013; 30:108–160. [PubMed: 23165928]
13. Ortega MA, van der Donk WA. New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. Cell Chem Biol. 2016; 23:31–44. [PubMed: 26933734]
14. Kuipers OP, et al. Protein engineering of lantibiotics. Antonie van Leeuwenhoek. 1996; 69:161–169. [PubMed: 8775976]
15. Cotter PD, et al. Complete alanine scanning of the two-component lantibiotic lacticin 3147: generating a blueprint for rational drug design. Mol Microbiol. 2006; 62:735–747. [PubMed: 17076667]
16. Pavlova O, Mukhopadhyay J, Sineva E, Ebright RH, Severinov K. Systematic structure-activity analysis of microcin J25. J Biol Chem. 2008; 283:25589–25595. [PubMed: 18632663]
17. Islam MR, et al. Evaluation of essential and variable residues of nukacin ISK-1 by NNK scanning. Mol Microbiol. 2009; 72:1438–1447. [PubMed: 19432794]
18. Pan SJ, Link AJ. Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions. J Am Chem Soc. 2011; 133:5016–5023. [PubMed: 21391585]

19. Young TS, Dorrestein PC, Walsh CT. Codon randomization for rapid exploration of chemical space in thiopeptide antibiotic variants. Chem Biol. 2012; 19:1600–1610. [PubMed: 23261603]
20. Zhang F, Kelly WL. In vivo production of thiopeptide variants. Methods Enzymol. 2012; 516:3–24. [PubMed: 23034221]
21. Boakes S, et al. Generation of an actagardine A variant library through saturation mutagenesis. Appl Microbiol Biotechnol. 2012; 15:1509–1517.
22. Deane CD, Melby JO, Molohon KJ, Susarrey AR, Mitchell DA. Engineering unnatural variants of plantazolicin through codon reprogramming. ACS Chem Biol. 2013; 8:1998–2008. [PubMed: 23823732]
23. Weiz AR, et al. Harnessing the evolvability of tricyclic microviridins to dissect protease-inhibitor interactions. Angew Chem Int Ed. 2014; 53:3735–3738.
24. Houssen WE, et al. An efficient method for the in vitro production of azol(in)e-based cyclic peptides. Angew Chem Int Ed. 2014; 53:14171–14174.
25. Ruffner DE, Schmidt EW, Heemstra JR. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. ACS Synth Biol. 2015; 4:482–492. [PubMed: 25140729]
26. Repka LM, Chekan JR, Nair SK, van der Donk WA. Mechanistic understanding of lanthipeptide biosynthetic enzymes. Chem Rev. 2017; 117:5457–5520. [PubMed: 28135077]
27. Li B, et al. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc Natl Acad Sci USA. 2010; 107:10430–10435. [PubMed: 20479271]
28. Tang W, van der Donk WA. Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria. Biochemistry. 2012; 51:4271–4279. [PubMed: 22574919]
29. Yu Y, Mukherjee S, van der Donk WA. Product formation by the promiscuous lanthipeptide synthetase ProcM is under kinetic control. J Am Chem Soc. 2015; 137:5140–5148. [PubMed: 25803126]
30. Giordanetto F, Kihlberg J. Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties? J Med Chem. 2014; 57:278–295. [PubMed: 24044773]
31. Foster AD, et al. Methods for the creation of cyclic peptide libraries for use in lead discovery. J Biomol Screen. 2015; 20:563–576. [PubMed: 25586497]
32. Quartararo JS, et al. A bicyclic peptide scaffold promotes phosphotyrosine mimicry and cellular uptake. Bioorg Med Chem. 2014; 22:6387–6391. [PubMed: 25438762]
33. Tavassoli A, et al. Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction. ACS Chem Biol. 2008; 3:757–764. [PubMed: 19053244]
34. Garrus JE, et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell. 2001; 107:55–65. [PubMed: 11595185]
35. Martin-Serrano J, Zang T, Bieniasz PD. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. Nat Med. 2001; 7:1313–1319. [PubMed: 11726971]
36. Demirov DG, Ono A, Orenstein JM, Freed EO. Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. Proc Natl Acad Sci USA. 2002; 99:955–960. [PubMed: 11805336]
37. Shi Y, Yang X, Garg N, van der Donk WA. Production of lantipeptides in Escherichia coli. J Am Chem Soc. 2011; 133:2338–2341. [PubMed: 21114289]
38. VerPlank L, et al. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). Proc Natl Acad Sci USA. 2001; 98:7724–7729. [PubMed: 11427703]
39. Male AL, et al. Targeting Bacillus anthracis toxicity with a genetically selected inhibitor of the PA/CMG2 protein-protein interaction. Sci Rep. 2017; 7:3104. [PubMed: 28596569]
40. Im YJ, et al. Crystallographic and functional analysis of the ESCRT-I /HIV-1 Gag PTAP interaction. Structure. 2010; 18:1536–1547. [PubMed: 21070952]

41. Pornillos O, Alam SL, Davis DR, Sundquist WI. Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. Nat Struct Biol. 2002; 9:812–817. [PubMed: 12379843]
42. Pornillos O, et al. Structure and functional interactions of the Tsg101 UEV domain. EMBO J. 2002; 21:2397–2406. [PubMed: 12006492]
43. Morris CR, Stanton MJ, Manthey KC, Oh KB, Wagner KU. A knockout of the Tsg101 gene leads to decreased expression of ErbB receptor tyrosine kinases and induction of autophagy prior to cell death. PLoS One. 2012; 7:e34308. [PubMed: 22479596]
44. Ruland J, et al. p53 accumulation, defective cell proliferation, and early embryonic lethality in mice lacking tsg101. Proc Natl Acad Sci USA. 2001; 98:1859–1864. [PubMed: 11172041]
45. Pornillos O, et al. HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. J Cell Biol. 2003; 162:425–434. [PubMed: 12900394]
46. Lu Q, Hope LW, Brasch M, Reinhard C, Cohen SN. TSG101 interaction with HRS mediates endosomal trafficking and receptor down-regulation. Proc Natl Acad Sci USA. 2003; 100:7626–7631. [PubMed: 12802020]
47. Jencks WP. On the attribution and additivity of binding energies. Proc Natl Acad Sci USA. 1981; 78:4046–4050. [PubMed: 16593049]
48. Knappe TA, et al. Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. Angew Chem Int Ed. 2011; 50:8714–8717.
49. Conibear AC, et al. Approaches to the stabilization of bioactive epitopes by grafting and peptide cyclization. Biopolymers. 2016; 106:89–100. [PubMed: 26566734]
50. Rink R, et al. To protect peptide pharmaceuticals against peptidases. J Pharmacol Toxicol Methods. 2010; 61:210–218. [PubMed: 20176117]
51. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med. 1998; 17:857–872. [PubMed: 9595616]

Europe PMC Funders Author Manuscripts

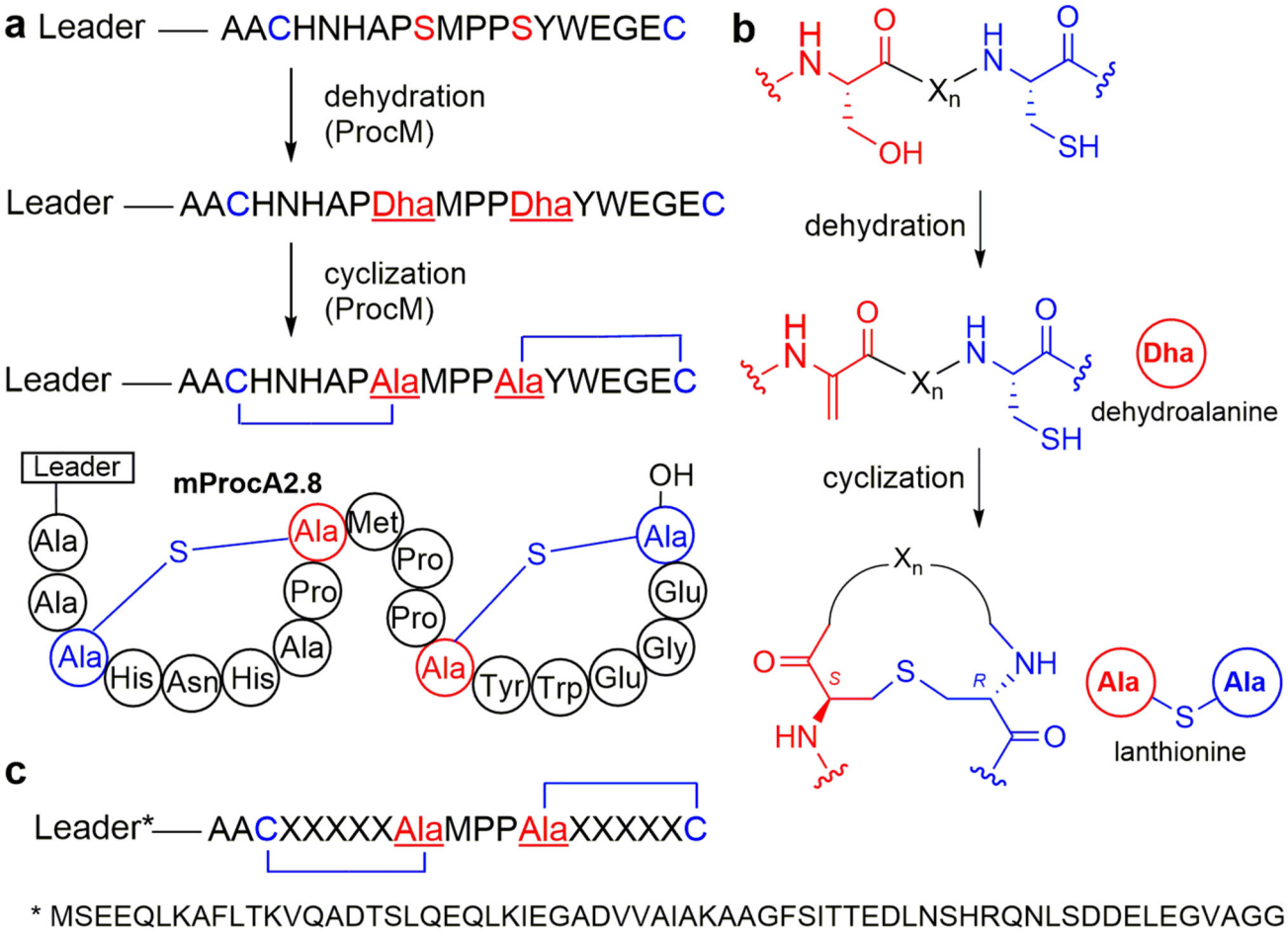

Figure 1. Representative illustration of lanthipeptide biosynthesis.

(**a**) Two Ser residues in the ribosomally synthesized linear precursor peptide ProcA2.8 are dehydrated by ProcM to generate two dehydroalanine (Dha) residues. The cyclization domain of ProcM then catalyzes the regioselective addition of two thiols of Cys residues to the Dha residues to generate modified ProcA2.8 (mProcA2.8). (**b**) Chemical structures showing the products of the dehydration and cyclization processes. (**c**) Generic structure demonstrating the randomization of the residues within the two rings of mProcA2.8 (X = D, F, H, I, L, N, V, or Y). The sequence of the leader peptide is also depicted. In all panels, structures derived from Ser are in red and from Cys in blue.

Europe PMC Funders Author Manuscripts

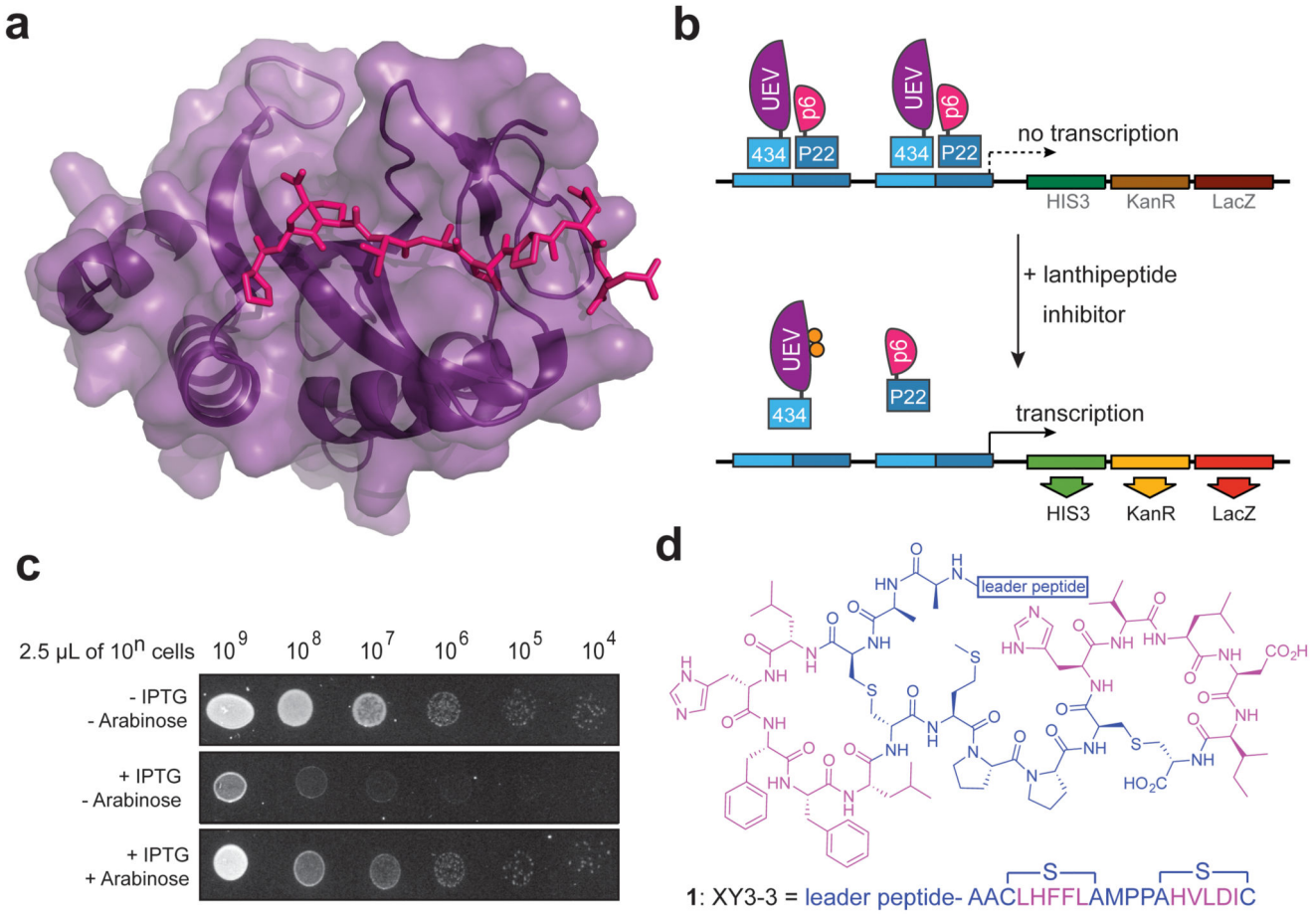

Figure 2. Identifying a lanthipeptide inhibitor of the p6–UEV PPI.

(**a**) Crystal structure of the PTAP peptide (PEPTAPPEE; pink) bound to UEV (purple); PDB ID = 3OBU40. (**b**) Top, IPTG-induced expression of the 434-UEV and P22-p6 fusion proteins results in a functional repressor that inhibits expression of the reporter genes; HIS3 and KanR are required for survival on selective media. Bottom, a lanthipeptide inhibitor of the p6–UEV interaction relieves the repression of the reporter genes and confers a growth advantage on selective media. (**c**) Drop spotting 10-fold serial dilutions of the p6–UEV RTHS transformed with the plasmid encoding XY3-3 in the absence of IPTG and arabinose (top row); expression of neither the repressors nor XY3-3 is induced, and full growth is observed. Upon adding IPTG (30 µM; middle row), a functional repressor prevents expression of His3/KanR, resulting in inhibited cell growth. Upon addition of IPTG (30 µM) and arabinose (65 µM; bottom row), the repressor and XY3-3 are produced; XY3-3 disrupts the p6–UEV interaction, preventing formation of a functional repressor and providing a growth advantage. All plates were incubated at 37 °C for 58 h. (**d**) Structure of XY3-3; residues in set positions are in blue, while randomized positions are in pink.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

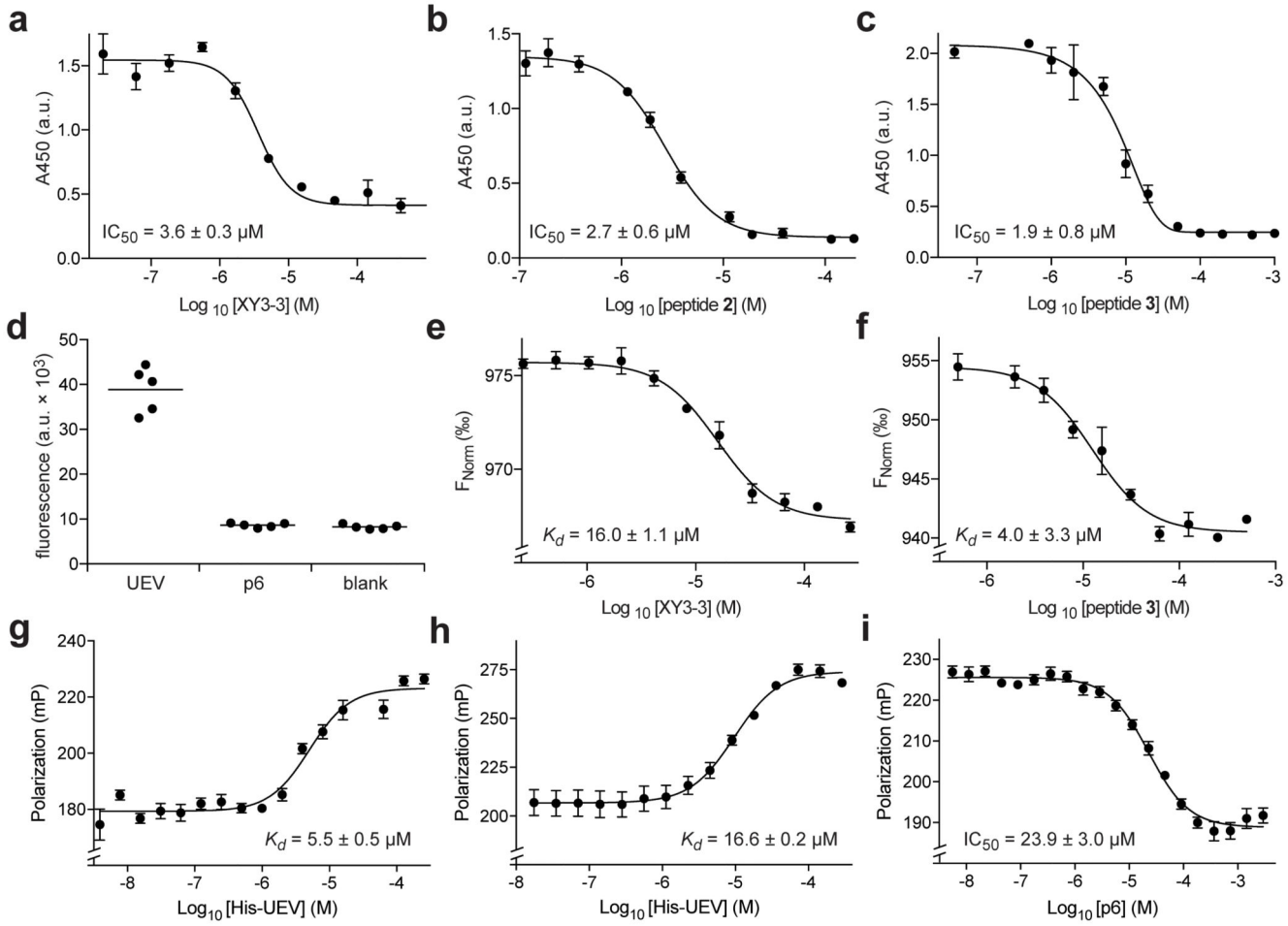

Figure 3. Assessing the activity of XY3-3 *in vitro*.

(**a–c**) ELISA data showing that XY3-3 disrupts the p6–UEV interaction with an IC_{50} of 3.6 ± 0.3 μM (**a**), peptide **2** disrupts the p6–UEV interaction with an IC_{50} of 2.7 ± 0.6 μM (**b**), and peptide **3** disrupts the p6–UEV interaction with an IC_{50} of 1.9 ± 0.8 μM (**c**). (**d**) UEV selectively pulls down a fluorescent-labeled XY3-3 whereas p6 does not, indicating that XY3-3 binds to UEV. Lines indicate the mean, n=5. (**e**,**f**) MST analysis reveals that XY3-3 binds to UEV with a K_d of 16.0 ± 1.1 μM (**e**) and peptide **3** binds to UEV with a K_d of 4.0 ± 3.3 μM (**f**). (**g–i**) FP analysis reveals that a FITC-tagged derivative of peptide **3** binds to UEV with a K_d of 5.5 ± 0.5 μM (**g**), a fluorescein-tagged PTAP nonapeptide binds to UEV with a K_d of 16.6 ± 0.2 μM (**h**), and a competition FP assay demonstrates that full length p6 displaces peptide **3** from UEV with an IC_{50} of 23.9 μM (**i**). Errors on K_d and IC_{50} values are the standard error of the mean (S.E.M.) given by regression analysis from three independent experiments.

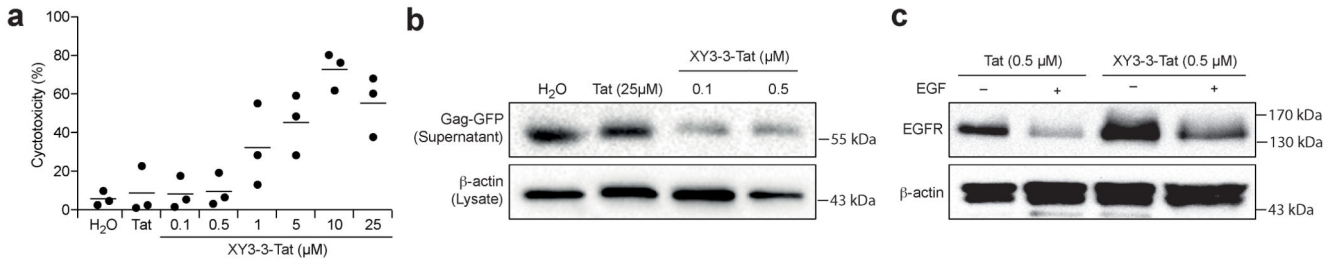

Figure 4. Cellular activity of XY3-3-Tat.

(**a**) Cytotoxicity assay of XY3-3-Tat in HEK293T cells shows that the compound is toxic to cells as doses above 1 μM. Lines indicate the mean, n=3. (**b**) Virus-like particle budding assay in HEK293T cells shows that a 100 nM dose of XY3-3-Tat inhibits virus-like particle production by ~65%. For uncropped blots see Supplementary Figure 18. (**c**) EGFR downregulation assay in HeLa cells shows XY3-3-Tat inhibits degradation of EGFR, in line with its mode of action of binding to UEV. For uncropped blots see Supplementary Figure 19.

Europe PMC Funders Author Manuscripts