scientific reports

OPEN

Discovery of antimicrobials by massively parallelized growth assays (Me^x)

Philipp Koch[1], Steven Schmitt[1], Mathias Cardner[2,3], Niko Beerenwinkel[2,3], Sven Panke[1] & Martin Held[1]✉

The number of newly approved antimicrobial compounds has been steadily decreasing over the past 50 years emphasizing the need for novel antimicrobial substances. Here we present Me^x, a method for the high-throughput discovery of novel antimicrobials, that relies on *E. coli* self-screening to determine the bioactivity of more than ten thousand naturally occurring peptides. Analysis of thousands of *E. coli* growth curves using next-generation sequencing enables the identification of more than 1000 previously unknown antimicrobial peptides. Additionally, by incorporating the kinetics of growth inhibition, a first indication of the mode of action is obtained, which has implications for the ultimate usefulness of the peptides in question. The most promising peptides of the screen are chemically synthesized and their activity is determined in standardized susceptibility assays. Ten out of 15 investigated peptides efficiently eradicate bacteria at a minimal inhibitory concentration in the lower μM or upper nM range. This work represents a step-change in the high-throughput discovery of functionally diverse antimicrobials.

Natural compounds are fundamental for drug discovery as they provide the biological relevance and structural diversity required to identify drug-like pharmacophores[1]. Owing to their high structural complexity and their ability to penetrate tissues and membranes, peptides are becoming increasingly important for many therapeutic areas[2]. Especially antimicrobial therapies have a very strong demand for novel compounds due to rising antimicrobial resistance[3]. Although about 3000 antimicrobial peptides have already been discovered, advances in genome sequencing and mining provide an ever-increasing number of peptides with elusive functions[4,5].

Large peptide libraries can be screened for antimicrobial activity using bacteria self-screening protocols. Here, peptides are expressed from their encoding DNA template and then accumulate either in the cytosol, the periplasm or at the bacterial surface[6]. If antimicrobial, their expression negatively impacts the proliferation rate or survival of the expressing cell. Sequencing of the peptide-encoding DNA of such impaired cells allows for the identification of antimicrobials in large pools of uncharacterized peptides. However, previous self-screening approaches failed to deliver large fractions of highly active peptides, or were unsuited for the screening of big libraries[7–9]. Thus, novel high-throughput screening methods are urgently needed.

We gathered naturally-encoded peptides from peptide and genomic sequence databases and assayed them for antimicrobial activity using massively parallelized growth assays (Me^x). Combined, the method delivered a rich collection of functionally diverse and highly active antimicrobial peptides.

Results

We first designed a library of naturally-encoded peptides. For this, we collected the amino acid sequences of 3063 peptides with already experimentally validated activity ("Parents" from here on) from the antimicrobial peptide database (APD) (Fig. 1a)[4]. Notably, Parents differed considerably with respect to the host from which they were derived, length, physiochemical properties, chemical modifications, 3D-structure, and sequence (Fig. 1a/b). Next, we applied *tblastn* on the translated nucleotide databases accessible through the NCBI using the amino acid sequence of the Parents as queries[10]. This search yielded 36,898 amino acid sequences with a similarity of ≥ 21.1% to the Parents ("Similars" from here on). Unlike the Parents, only very few of the Similars have been synthesized or experimentally tested. However, owing to their natural origin and similarity to the Parents, a fraction of the Similars is likely to display antimicrobial properties[11,12]. For technical reasons (Methods), we applied a cut-off of 42 amino acids in peptide chain length and selected Similars with at least 62.2% sequence

[1]Bioprocess Laboratory, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [2]Computational Biology, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland. [3]SIB Swiss Institute of Bioinformatics, 4058 Basel, Switzerland. ✉email: martin.held@bsse.ethz.ch
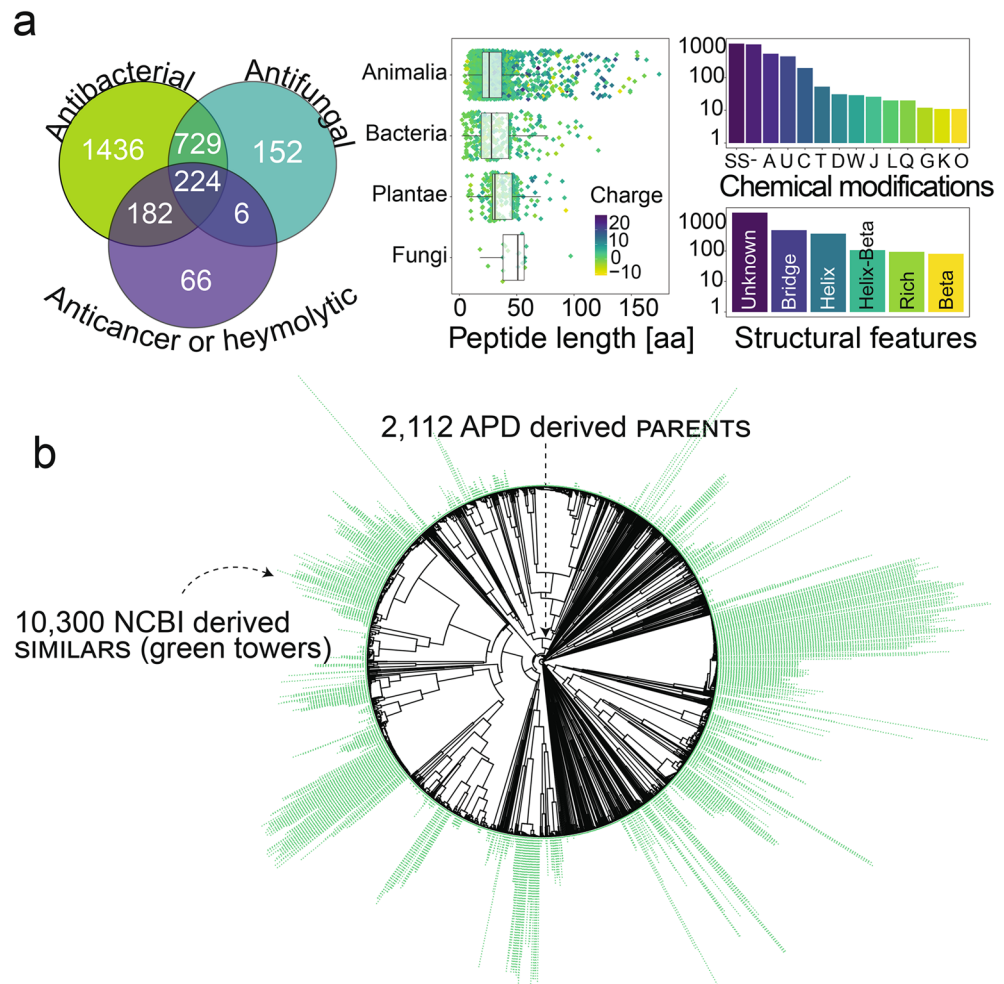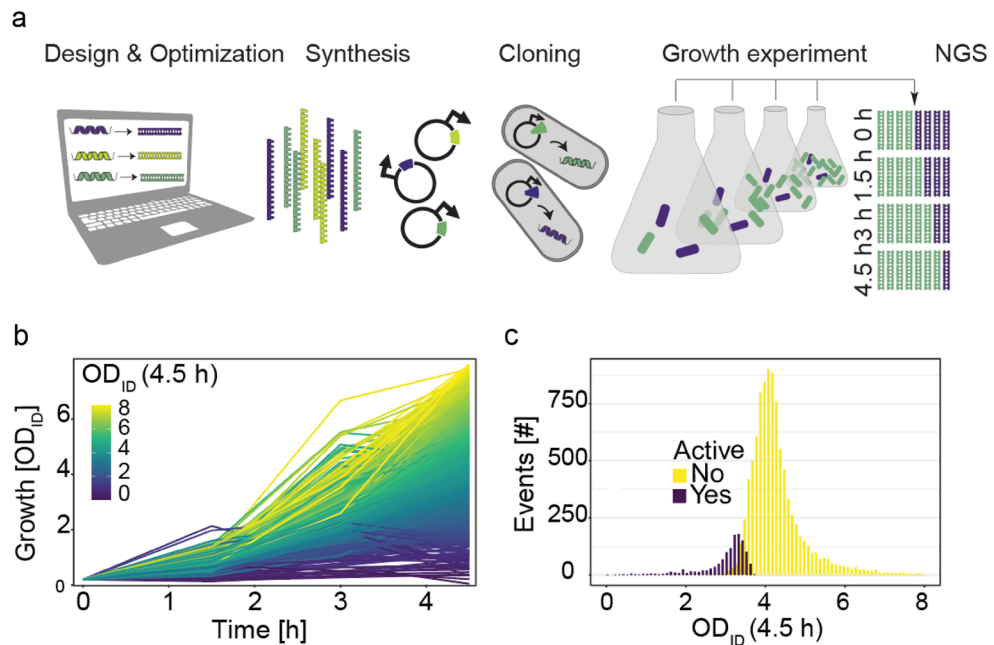
Figure 1. Generation of antimicrobial peptide library. (**a**) Biological diversity of Parents. Parents are derived from the APD. They have experimentally proven biological activity, e.g. antibacterial (Gram-negative and/or Gram-positive bacteria), antifungal, or anticancer or hemolytic, originate from species of various kingdoms of life, and differ considerably by length, charge, chemical modification (among others: *SS* disulfide bridges, *A* amidation, *U* terminal Rana box (disulfide bridge at C-terminus), *C* backbone cyclization, *T* thioether bridges, *D*-amino acids, *W* dehydration, *J* sidechain cyclization, *L* lipidation, *Q* terminal glutamate, *E* acetylation, *G* glycosylation, *K* hydroxylation, – no modification reported), and 3D-structure (*Beta* beta-sheet, *Bridge* disulfide bond, *Helix* alpha-helix, *Helix-Beta* alpha-helix and beta-sheet, *Rich* rich in unusual amino acids, *Unknown* no reported structure). (**b**) Sequence distances of the complete peptide library. Pairwise sequence distance between 2112 Parents (BLOSUM62) as a basis for hierarchical clustering. Similars found using *tblastn* for each Parents' search query are stacked as towers on the tips of the dendrogram.

similarity. In this way, a library of 2122 Parents and 10,300 Similars (Fig. 1b) was obtained. Examination of the final library indicated net charges from − 10 to + 15 and hydrophobicity of − 3.5 to 2.9 (GRAVY scale; Fig. S1). Additionally, we were able to allocate the origin of 7497 of these peptides to the kingdom animalia, 74 to fungi, 678 to bacteria, and 2485 to plantae (Fig. S2).

For Me^x, we converted the peptides into corresponding oligonucleotides (Fig. S3), retrieved the latter as a pool after chemical synthesis on a microarray, and ligated the sequences into a plasmid on which their expression was controlled by the tightly regulated P_{BAD} promotor (Fig. 2a). We then transformed the model organism, *E. coli* TOP10, with the peptide-encoding DNA library. Using next-generation sequencing (NGS), we only identified 10,663 different peptide-encoding DNA sequences (listed by ID in File S1) in *E. coli* indicating sequence bias in the initial oligonucleotide pool (Fig. S4).

To assess the antimicrobial activity of the DNA-encoded peptides, we performed Me^x and generated growth curves for each of the 10,663 peptide-expressing *E. coli* strains. To do so, we inoculated three liquid cultures each with 500 million transformed cells, and induced peptide synthesis after four cell doublings (Fig. S5). Because the synthesis of an antimicrobial peptide should inhibit the growth of the expressing host, the propagation rate of the peptide-encoding DNA will also be reduced. Hence, we harvested bacteria at the time of induction as well as 1.5 h, 3.0 h, and 4.5 h post-induction and used NGS to count reads for each peptide-encoding DNA. To derive growth curves (Fig. 2b), we calculated the abundance of each strain (ID) using the respective NGS read counts

Figure 2. Screening of antimicrobial peptides using Me^x. (**a**) Me^x workflow: Design and Optimization: Peptide sequences are reverse translated into *E. coli* codon-optimized nucleotide sequences. Synthesis: All peptide-encoding sequences are synthesized as oligonucleotides. Cloning: The sequences are inserted into plasmids. *E. coli* TOP10 is transformed with the generated peptide-encoding DNA library. Growth: Strains are incubated in shaking flasks, peptide expression is induced and plasmids are isolated. NGS: peptide-encoding DNA sequences are counted at four time points using NGS. (**b**) Growth curves of all 10,663 peptide-expressing strains, expressed as OD for a specific peptide-expressing strain (OD$_{ID}$; average of n = 3). Coloring from yellow to dark blue indicates higher growth inhibitory effects based on OD$_{ID}$ of last sampling point. Curves reaching a higher OD$_{ID}$ than eight (0.7%) are omitted for clarity. (**c**) Growth inhibition at 4.5 h recorded for all 10,663 peptide-expressing strains. OD$_{ID}$-values are recorded for each of the peptide-expressing strains and are averaged from three replicates. Me^x-active peptides (purple) significantly (Wald's test, adj. $p < 0.05$) reduce the OD$_{ID}$ of their expressing strain after 4.5 h while Me^x-inactive peptides (yellow) fail to do so. Note that some candidates may also fail to reach statistical significance in the performed Me^x-activity test due to low NGS read counts or high variance between replicates (Supplementary Table 2).

and multiplied these with the measured cell concentration of the entire liquid cultures (OD) thereby obtaining an approximation of the strain-specific concentrations (OD$_{ID}$) at each sampling point. Comparing OD$_{ID}$ of all peptide-expressing strains after 4.5 h, we found that intracellular expression of 1240 peptides (11.6%) significantly inhibited the growth of their host ("Me^x-actives" from here on; Wald's test, p-value (p) < 0.05, adjusted for multiple testing (adj.); Fig. 2c). The remaining peptides did not show growth inhibition in Me^x, likely because they are not antimicrobial at all or require chemical modifications not introduced in *E. coli*, could not access their (e.g. extracellular) target, or did not reach inhibitory concentrations due to limited mRNA or peptide stability.

Next, we confirmed that the intracellularly synthesized peptides also inhibited growth if the strains were grown individually. For this, we selected 110 peptide-expressing strains experiencing different levels of growth inhibition in Me^x and measured their growth in microtiter plate wells (Fig. S6a/b). As the growth curves recorded in Me^x and microtiter plates were comparable (Fig. S6c), we concluded that the complex dynamic of the Me^x-culture did not bias the results.

Screening 10,663 peptides at once allowed us to address several research questions. Firstly, we sought to confirm that our approach of exploiting sequence similarities to known antimicrobial peptides indeed allowed us to identify antimicrobials. In fact, 1035 out of 1240 Me^x-actives (83%) were SIMILARS, i.e. peptides whose functions were not reported on the APD. A closer look revealed that for 310 inactive PARENTS we found at least one active SIMILAR. As an example, PARENT Apo5 APOC1$_{667\ APD}$ (nomenclature: name of PARENT on APD $_{ID\ Origin}$), itself inactive, yielded 27 SIMILARS of which one showed eight amino acid differences to the PARENT and displayed antimicrobial activity (Fig. S7). We argue that the amino acids by which the inactive PARENT and the active SIMILAR differed were of high importance for activity and necessary for evading the abovementioned reasons for failed growth inhibition in Me^x. Furthermore, 47 PARENTS spawned an overrepresentation of active SIMILARS (Fisher's exact test, adj. $p < 0.05$; Fig. S8). Examples include Myticin-B (21/31), which yielded 31 SIMILARS, of which 21 were active, and PepG1 (11/11). This indicates that the respective peptide sequences have considerable plasticity and can accommodate multiple amino acid exchanges without losing activity. We argue that these peptides might be well suited for additional modifications performed for instance in the course of lead optimization programs[13].
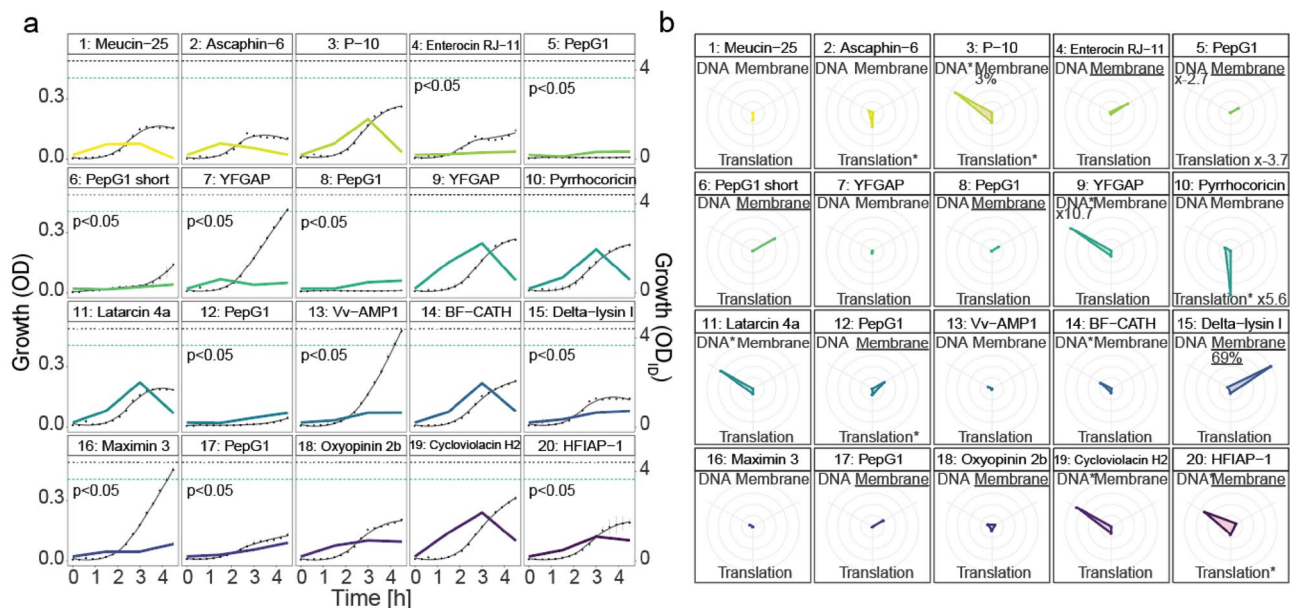
Figure 3. Characterization of the 20 most active peptides in Me^x. (**a**) Growth curves of *E. coli* TOP10 cells each expressing on of the 20 most active (by OD_{ID} at 4.5 h) peptides. Colored lines are Me^x-recorded growth curves (average of n = 3) determined via OD_{ID} approximation (header: 'rank: PARENT name'). Black lines are growth curves (n = 3, error bars: 2σ) determined via OD measurement in microtiter plates of individually grown strains. Horizontal dashed lines, in black (OD) or colored in green (OD_{ID}), show final values measured 4.5 h post-induction of a strain synthesizing the inactive control peptide HNP-1$_{3425\,APD}$ (obtained from Fig. S13). In each facet, we state if we obtain a $p < 0.05$ (Wald's test) for significant growth inhibition after 1.5 h in Me^x. (**b**) Potential mechanisms of action. Each radar plot shows the mean SOS-response (DNA; activation of the *recA* promoter; n = 3), translation inhibition (Translation; activation of the *cspA* promoter; n = 3), and membrane-damage (Membrane; PI stained cells in percent; n = 2) obtained after peptide expression in *E. coli* TOP10. Only the maximum and minimum values are reported in digits. The center represents values measured for the negative control peptide HNP-1$_{3425\,APD}$. Lower values are scaled to the center. Membrane damage is attributed if more than 10% of cells were PI-positive (underlined). For SOS and Translation, signals are reported relative to the signal obtained for the inactive control peptide HNP-1$_{3425\,APD}$. A significant increase (one-sided *t* test, adj. $p < 0.05$) compared to the inactive control is indicated by an asterisk (*).

Secondly, we evaluated the phylogeny of the hosts from which the inhibitory peptides were derived. For this, all peptides of the library were grouped taxonomically based on their natural host. We then calculated the fraction (%) of Me^x-actives within the ranks Kingdom and Class (Fisher's exact test; Fig. S9). Me^x-actives were significantly underrepresented ($p < 0.05$) among bacteria (8.5%), amphibians (7.7%), and mammals (10.3%) but overrepresented ($p < 0.05$) in insects (13.4%), birds (25%), ray-finned fishes (15.6%) and bivalves (31.8%). Since insects contain by far the most species in the animal kingdom, this indicates a huge and so far undiscovered pool of antimicrobials in insects.

Thirdly, as cationic and hydrophobic peptides frequently display antimicrobial activity, we wondered whether growth inhibition in Me^x was biased by the physiochemical properties of peptides[14]. However, linear regression analysis indicated no correlation of growth inhibition with hydrophobicity (correlation = 0.04) and charge (correlation = − 0.01; Fig. S10a). Furthermore, among the 47 PARENTS with overrepresented active SIMILARS, there was no clear relationship between charge or hydrophobicity and growth inhibition (Fig. S10b). We thus conclude that growth inhibition in Me^x-actives is driven by the specific antimicrobial activity of a peptide either damaging the cytoplasmic membrane or binding and inhibiting other cellular components.

To investigate peptides further, we characterized the 50 most growth inhibitory peptides as indicated by their rank in the Me^x screening (rank 1–50; 38 SIMILARS, 12 PARENTS) (Fig. S11a–c). Initial tests were performed with two biosensor constructs, containing the *cspA* and *recA* promoters, which upon activation are indicative of translation impairment and DNA damage, respectively[15]. The results indicated translational impairment for 11 and DNA damage for 12 peptide-expressing strains (one-sided *t* test, adj. $p < 0.05$; Fig. 3b; Fig. S12), which suggests that these peptides target intracellular macromolecules. In fact, many peptides traverse the membrane(s) of bacteria without permeabilization and kill cells by binding or blocking intracellular macromolecules[16,17]. For example, Metalnikowin IIA$_{8984\,APD}$, Metalnikowin III$_{9011\,APD}$, known ribosomal inhibitors, and Pyrrhocoricin$_{7122\,NCBI}$, whose PARENT is also a ribosomal inhibitor, caused the strongest indication for translational impairment in our assay[18]. Next, we measured membrane damage by quantifying propidium iodide (PI) uptake. Expression of 11 peptides resulted in membrane damage, with the strongest damages observed for Delta Lysin I SIMILARS whose PARENT is a membrane pore inducing bacteriocin from *Staphylococcus* (Fig. S12, Table S1). Interestingly, 11 out of 12 peptides that caused membrane damage significantly inhibited growth already after 1.5 h in Me^x (Wald's

Rank	PARENT name	Peptide sequence	Origin	ID	Charge	GRAVY	Intracellular characterization	MIC [μM] 100%MHB	MIC [μM] 25%MHB	Membrane damage	Hemolysis at MIC | 16MIC
1	Meucin-25	VKLIQIRIWIQYVTV-LQMFSMKTKQ	APD	11,598	+4	0.4	–	>60	>60		
2	Ascaphin-6	GFKDWIKGAAKKLIK-TVASSIANE	APD	9286	+3	−0.1	Translation	0.8	1	$\frac{1}{2}$ MIC	0% | 1.4%
3	P-10	VSKIKKYLKYKDRI	APD	8942	+5	−1	DNA; Translation	>60	13	No damage	0.5% | 2.2%
4	Enterocin RJ-11	AIAKLVAKFGWPIVK-KYYKQIMQFIGEG-WAINKIIEWIKK	NCBI	3780	+6	0.2	Membrane	0.5	1	$\frac{1}{4}$ MIC	22.7% | 103.3%
5	PepG1	MITISTMLQFGLFLIAL-IGLVIKLIELSIKK	NCBI	11,834	+2	1.6	Membrane	>60	>60		
7	YFGAP	VKVGINGFGRIGR-LVTRAAFQSKKVEI-VAIND	NCBI	8112	+4	0.3	–	>60	8	2MIC	0% | 3.3%
9	YFGAP	VKVGVNGFGRIGR-LVTRAAFNSGKVEI-VAIND	NCBI	8135	+3	0.4	DNA	>60	5	2MIC	0% | 0.4%
10	Pyrrhocoricin	VDKGGYLPRPTP-PRPVYR	NCBI	7122	+3	−1.2	Translation	20	8	No damage	1.1% | 2.3%
11	Latarcin 4a	LKD-KVKSMGEKLKQYIQT-WKAKF	APD	5147	+5	−1	DNA	8	5	2MIC	0.2% | 2.8%
13	Vv-AMP1	RACESQSHRFK-GTCVRQSNCAAVCQTE	NCBI	8053	+2	−0.7	–	>60	>60		
14	BF-CATH	KRFKKFFKKLKKSVK-KRAKKFFKKPRVIG-VSIPF	APD	9639	+16	−0.7	DNA	1	1	4MIC	2.1% | 4.8%
15	Delta lysin I	MAADIISTIGDLVKWI-IDTVNKFK	NCBI	3458	0	0.6	Membrane	>60	>60		
16	Maximin 3	TALKGAAKELASTYQH	NCBI	5468	+1	−0.4	–	>60	>60		
18	Oxyopinin 2b	GKFSGFAKILK-SIAKFFKGVGKVRKG-FKEASDLDKNQ	APD	9690	+7	−0.5	Membrane	0.5	0.5	$\frac{1}{2}$ MIC	0.7% | 63.4%
20	HFIAP-1	GWFKKAWRKVKHA-GRRVLDTAKGVG-RHYLNNWLNRYR	NCBI	4545	+10	−1.1	Membrane, DNA; translation	0.5	0.5	$\frac{1}{8}$ MIC	4.1% | 18.9%

Table 1. Summary of antimicrobial activity assays of the 20 most active peptides in Me^x. Peptides for which a MIC could be determined are highlighted in bold. PepG1 SIMILARS (rank 6, 8, 12, and 17), and Cycloviolacin H2 (rank 19) were removed from this table, as the purification failed. Intracellular characterization is derived from the experiments summarized in Fig. 3b. Mean MIC-values are recorded (n = 3) in microtiter plate assays using chemically synthesized peptides against the screening strain *E. coli* TOP10. Membrane damage is reported at the peptide concentration, relative to the MIC, at which at least 25% of cells acquired externally added PI. (n = 2; Fig. S14a). Hemolysis as a percentage is related to full lysis after treatment with 2.5% Triton-X100 (n = 4; Fig. S15). MIC of ampicillin (control) against *E. coli* TOP10: 34 μm (100% MHB); 17 μm (25% MHB).

test, $p < 0.05$) (Fig. 3a; Fig. S13) while for 25 peptides, and especially for those with putative intracellular targets, growth inhibition started only after 4.5 h (Fig. 3a; Fig. S13). Noteworthy, delay of the growth inhibition onset has been reported to be indicative of peptides that interact with an intracellular target[19]. We concluded that this effect could be observed in Me^x too, and hence reanalyzed all growth curves recorded for the Me^x-actives. Growth was significantly inhibited after 1.5 h by 806 peptides (65%) pointing towards membrane damage (File S1) but only after 4.5 h in the case of the remaining 434 peptides, suggesting a high likelihood for reduced membrane damage, or the interaction with an intracellular target.

Next, we chemically synthesized 15 out of the 20 peptides that were found to be most growth inhibitory in Me^x and determined their minimal inhibitory concentrations (MIC) in both 100% Mueller–Hinton Broth (MHB) and 25% MHB, their membrane damaging capabilities in 100% MHB, and hemolytic activity when added to cells as a synthesized chemical (Table 1; Fig. S11).

Firstly, no MIC was obtained for five peptides; however, as four of these five were either PARENTS or derived from PARENTS known to be inactive against *E. coli* (Table S1), we believe that these peptides exerted activity in the cytosol if synthesized intracellularly but could not reach their target (e.g. the cytoplasmic membrane) when added as a chemical substance to the growth medium. Remarkably, 10 of the 15 peptides for which MICs were recorded, very efficiently inhibited the growth of *E. coli* (MICs 0.4–20 μm; mean = 3.7 μm; median = 1 μm), in a concentration range that could qualify as a starting point for drug development[20]. In four cases (P-10$_{8942\ APD}$, YFGAP$_{8112\ NCBI}$, YFGAP$_{8135\ NCBI}$ and Pyrrhocoricin$_{7122\ NCBI}$) the activity increased in diluted MHB, while for the remaining six peptides, a similar MIC was recorded in both 100%, and 25% MHB. We selected the most active SIMILAR, HFIAP-1$_{4545\ NCBI,}$ and measured the activity against other clinically relevant Gram-negative and positive

bacteria. SIMILAR HFIAP-1$_{4545\,NCBI}$ inhibited growth of these strains (MICs: 0.4–5.6 μm; Table S2), which suggests a broad activity spectrum even though Me^x screening was performed in another host. These results indicated that although we screened the peptide library synthesized cytosolically, Me^x-active peptides also strongly inhibited growth when added to cells externally and that the employed *E. coli* lab strain was suited well for the identification of peptides also active against other bacterial species.

Secondly, to evaluate the degree rely on damaging of membranes damage for the exertion of antimicrobial, we measured the uptake of PI, when adding the peptides four log$_2$ concentration steps above and below their MIC. As complete lysis of both outer and inner membrane resulted in false negatives in PI measurement (see the decrease in PI signal when increasing the concentration of Melittin in Fig. S14), we also quantified the point at which both membranes were lysed by measuring the release of intracellularly expressed GFP. Only Ascaphin-6$_{9286\,APD}$, Enterocin RJ-11$_{3780\,APD}$, Oxyopinine 2b$_{9690\,APD}$, and HFIAP-1$_{4545\,NCBI}$, showed strong membrane damage (>25% PI-positive cells below MIC concentration) in a range of the membrane damaging peptide Melittin (Table 1; Fig. S14). For the remaining peptides, and especially for those with reported intracellular targets (Fig. 2b) and a delayed growth inhibitory effect in Me^x (Fig. 2a), no membrane damage could be detected at all, or occurred considerably above the MIC (Fig. S14). Hence, other mechanisms, such as blocking of protein translation as reported for the parent of Pyrrhocoricin$_{7122\,NCBI}$, must play a role in bacterial killing[18]. These results confirmed that the previously recorded intracellular characterization of the peptides can be a good indication for the activity of chemically produces peptides discovered in the Me^x assay.

Lastly, as hemolysis is a reliable and sensitive indicator for cytotoxicity assessment, we measured the hemolytic activity of all MIC active peptides[21,22]. Toxicity towards human erythrocytes (>5% hemolysis compared to the Triton-X100 control) at the MIC was only observed for the membrane damaging peptide Enterocin RJ-11$_{3780\,NCBI}$, a known hemolytic staphylococcal toxin, and for the control Melittin (Table 1)[23]. Additionally, the membrane damaging peptides Oxyopinine 2b$_{9690\,APD}$, a known hemolytic spider toxin, and HFIAP-1$_{4545\,NCBI}$ displayed hemolytic activity at higher concentrations (Fig. S15; Table 1)[24]. All other peptides did not damage erythrocytes at the tested concentration. This suggests that most of the active peptides found in the Me^x screen were not cytotoxic thereby further corroborating the potential of the isolated hits for drug development.

Discussion

We applied Me^x for the highly parallelized discovery and characterization of more than 10,000 structurally diverse, and naturally-encoded peptides (Fig. 1a/b; Fig. S1; File S1). Chemical production and microtiter plate based screening of a library of such high diversity and size, and containing peptides longer than 40 amino acids would have been very cost- and time-consuming. However, Me^x takes a shortcut by using in silico optimized and pooled oligonucleotides as templates for ribosomal peptide synthesis (Fig. S3) and intracellular activity assessment via monitoring the growth inhibitory effects (Fig. 2a/b).

Growth curves recorded by Me^x via NGS were comparable to those obtained for a few tested strains if grown compartmentalized in microtiter plates (Fig. S6). This indicated that pooling of the peptide-expressing strains did not bias the experimental outcome. Moreover, the high hit rate (10 out of 20 peptides) obtained for Me^x-active peptides if synthesized as chemicals and tested in MIC assays under stringent CSLI- assay conditions (Table 1) corroborates the robustness of Me^x. In addition, screening of naturally-encoded sequences delivered a large fraction of highly active peptides, by far exceeding the performance of other approaches selecting randomly or semi-randomly designed peptide libraries[7–9]. As naturally occurring peptides are preselected for biological activity, including antimicrobial activity, our results confirmed the advantage of screening sets derived from genomic databases[1,12,25].

When extrapolating from our hit-rates (50% of Me^x-actives were active in MIC assays using chemically synthesized peptides) to the entire library (1035 active SIMILARS and 205 active PARENTS), we found more than 500 previously unreported, active antimicrobial peptides, derived from organisms of various biological classes (Fig. 2c; Fig. S9). Moreover, by analyzing growth curves of the peptide-expressing strains (displayed as OD$_{ID}$, Fig. 3a; Fig. S13), including data indicative for intracellular stress response mechanisms (Fig. 3b; Fig. S12), and the status of the bacterial membrane integrity (Table 1, Fig. S14), we estimated in *E. coli* TOP10 about one-third of the active peptides most likely exert antimicrobial activity via the interaction with intracellular targets (see all in File S1). We want to point out that all data was generated using the weakened laboratory strain *E. coli* TOP10, and thus cannot be simply adapted to others strains of *E. coli* (e.g. clinical isolates), other Gram-negative, or even Gram-positive bacteria.

These results hence cast a fresh look on the field of antimicrobial peptides as only very few examples (< 50) of intracellularly active peptides are known[26]. We hypothesize that nature designed the peptides such that passage of bacterial membranes and binding to macromolecules is a frequently built-in feature. As the transition from the discovery pipeline to the patient is often hampered by the poor specificity of membrane damaging peptides in vivo, Me^x can be a valuable tool for the high-throughput discovery of peptides that do not rely on membrane damage for bacterial killing[27,28].

Our indicated mechanism (peptides of rank 1–20 summarized in Table 1) correlate well with those reported for the PARENT peptides by others (Table S1). Membrane damage was reported for PARENTS (of) Ascaphine-6, Enterocin RJ-11, Oxyopinin 2b, and HFIAP-1. For all of these peptides (mostly SIMILAR), besides for Ascaphine-6$_{9286\,APD}$, we also identified membrane damaging properties. On the other, no antibacterial mechanism of action was reported for PARENTS (of) P-10, or BF-CATH, a non-membrane damaging mechanism of action was reported for the YFGAP PARENTS, and protein translation was reported for the PARENT of Pyrrhocoricin. We indicated different non-membrane damaging mechanisms, including protein translation inhibition for Pyrrhocoricin$_{7122\,NCBI}$, and delayed growth inhibition in Me^x for most of them. Nonetheless, a direct comparison

between the data proposed by us and those for the PARENTS in literature might be difficult, as small changes in the amino acids sequence might result in an alteration of the antimicrobial function[29].

A limitation of the Me^x approach can be that the identified peptides are active intracellularly, but remain inactive when tested in MIC assays. For example, all 11 peptides derived from the PARENT PepG1 were among the most actives in Me^x (Figs. S8, S10, S12, S13), likely due to strong damage to the cytoplasmic membrane (Fig. 3a/b; Table 1). However, the most active candidate PepG1$_{11834 NCBI}$, remained inactive at the tested concentration (60 μm) when added extracellularly. In fact, the PepG1 PARENT has shown very weak activity against Gram-negative bacteria (MIC of 25–100 μm) by others previously[30]. To overcome this limitation, uptake for such peptides could be enhanced by linking them to cell-penetrating peptides[31].

Taken together, Me^x enables rapid discovery and classification of naturally-occurring and functionally diverse antimicrobial peptides. However, we argue that Me^x can also be used for de novo design or optimization of natural peptides by directed evolution approaches and that, the principal technology can eventually also be used for screening in drug-resistant (e.g. *Pseudomonas aeruginosa* or *Acinetobacter baumannii*). Ultimately, Me^x will hence allow paving the way towards the discovery of next-generation antibiotics.

Methods

Chemicals and reagents.
Unless otherwise stated, all chemicals, reagents, and primers were obtained from Sigma Aldrich (Buchs, CH). Restriction enzymes and their buffers were obtained from New England Biolabs (Ipswich, USA). Synthetic genes were obtained from Integrated DNA Technologies (Leuven, BE) or Twist Bioscience (San Francisco, USA). Kits for plasmid isolation and DNA purification were obtained from Zymo Research (Irvine, USA). Peptides in either purified (> 90%) or crude format were obtained from Pepscan (Lelystad, NL). Sanger-sequencing was done at Microsynth (Balgach, CH).

Bacterial strains and cultivations.
Unless otherwise stated, all experiments were performed using *Escherichia coli* TOP10 (F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *end*A1 *nup*G; Thermo Fisher Scientific, Waltham, USA). In this study, all cultivations were performed either in 14 ml polypropylene tubes (Greiner, Kremsmuenster, AT), filled with 5 ml of lysogeny broth (LB) medium (Difco, Becton Dickinson, Franklin Lakes, USA), or in 96-deep-well polypropylene plates (Greiner, Kremsmuenster, AT) filled with 500 μl of LB-medium. All samples were incubated at 37 °C with agitation on a shaker (Kuhner, Birsfelden, CH) operated at 200 r.p.m. and 25 mm amplitude. All media were supplemented with the appropriate antibiotic for plasmid maintenance (50 μg ml⁻¹ kanamycin; 100 μg ml⁻¹ carbenicillin) and 1% (w/v) D-glucose for repression of gene expression from catabolite-repression sensitive promoters such as P$_{BAD}$. In the case of peptide expression experiments, cultures were incubated without D-glucose and 0.3% (w/v) of the inducer L-arabinose was used for induction. For all cultivations on solid medium, 15 mg ml⁻¹ agar (Difco) was added to the broth, and incubation was performed without shaking in an incubator (Kuhner) at 37 °C. If not indicated differently, the optical densities (OD) of bacterial cultures were determined by measuring light scattering at 600 nm using a UV/VIS spectrophotometer (Eppendorf, Hamburg, DE).

In silico generation of peptide library.
We collected all peptide sequences (called "PARENTS") available on the APD in May 2017 (https://aps.unmc.edu/)[4]. These sequences were used as input queries to find sequence-similar peptide sequences in the NCBI non-redundant nucleotide collection (nr/nt), a collection that holds sequences from GenBank, European Molecular Biology Laboratory (EMBL), DNA Databank of Japan (DDBJ), and Reference Sequence database (RefSeq), as well as translated protein information from the protein database (PDB)[10]. By applying *tblastn*, 170,300 additional peptide sequences (called SIMILARS) were found[32]. Because we were limited to 12,412 different peptides with a maximum length of 42 amino acids (the chosen platform for the synthesis of the peptide-encoding oligonucleotides allowed 12,412 different sequences with a maximal length of 170 bases), we discarded SIMILARS with sequence similarity to the respective parent of less than 62.2%. The following parameters were used for the *tblastn* search: maximum sequences = 100; matrix = BLOSUM62; gap cost = 11.1; word size = 6; active low complexity filter; adjustment = conditional compositional score matrix adjustment.

Sequence distance among PARENTS and SIMILAR.
To visualize sequence diversity among PARENTS, we created a sequence-based phylogenetic tree. We performed pairwise global alignment of all PARENT sequences using the Needleman–Wunsch algorithm, as implemented in the R Bioconductor package 'Biostrings' (https://bioconductor.org/packages/release/bioc/html/Biostrings.html). The BLOSUM62 substitution matrix was used to compute the alignment scores, which were converted into pairwise distances following the method Scoredist[33]. Based on the pairwise distances between PARENTS, we used hierarchical clustering with average linkage to compute a dendrogram of sequences reflecting their similarities. PARENTS and their *tblastn*-derived SIMILARS were consolidated into groups, which were named after the PARENT from the APD (https://aps.unmc.edu/). In the sequence-based phylogenetic tree, each SIMILAR was stacked on top of its PARENT at the tip of the dendrogram. A SIMILAR may appear multiple times if it was found multiple times in the *tblastn* search using different PARENTS.

Peptide-encoding DNA architecture.
The corresponding oligonucleotide sequences of the peptide library were synthesized using microarray technology supplied from CustomArray Inc. (now GeneString, Piscataway, USA). The chosen platform allowed 12,412 different oligonucleotides with a maximal length of 170 bases. A generic oligonucleotide design employing four functional units was created (Fig. S3): A coding unit, a filler unit, and two universal units for amplification. This process was automated for each sequence by using an in-house written script in R. The coding unit contained the reverse translation of the peptide amino acid sequence

into a codon-optimized DNA for *E. coli*. We always chose the most abundant codon for each amino acid. In cases in which restriction sites had been introduced that could potentially interfere with subsequent manipulations, the crucial codon was replaced by the second most abundant one for this amino acid. The filler sequence was added to compensate for the various lengths of peptide genes (shortest coding sequence = 15 nucleotides, longest coding sequence = 126 nucleotides) and adjust the total of filler and coding unit to 129 nucleotides for all members of the library. To do so, we first added a UAA stop codon to the end of the coding sequence and then added downstream a semi-random sequence, ensuring a GC content of 40% for the filler sequence and limiting the number of identical nucleotides following each other to three. By adding this filler sequence we maximized sequence disparity at the DNA level (many coding sequences are homologs) thereby potentially increasing both synthesis and, later, sequencing quality. Two amplification units, of 23 and 18 bases, respectively, were appended upstream and downstream of the coding sequence and filler unit and contained the ribosomal binding site and restriction sites for the enzymes PstI and HindIII. Two amplify the peptide-encoding DNA, primer 1: CTGCAC AAAGCTTACGTG, complementary to the upstream amplification unit, and primer 2: CACGTAAGCTTTGTG CAG, reverse complementary to the downstream amplification unit were used. The final 170 bases long oligonucleotide sequences as synthesized are listed by ID in File S2 (erroneous sequences were discarded).

Peptide-encoding DNA cloning.   The chemically synthesized and single-stranded oligonucleotides were separated from their array and we received them as a pool. This pool was aliquoted in 10 mM Tris–Cl, 1 mM EDTA, pH 8 and deep-frozen at − 80 °C. The pool was amplified by polymerase chain reaction (PCR) in a 50 µl reaction using 5 ng of the template and 10 µm HPLC-purified primer 1 and primer 2, complementary to the amplification sites, and 25 µl of Phusion High-Fidelity PCR Master Mix with HF buffer. The amplification was performed using 25 cycles of 98 °C for 15 s, 55 °C for 20 s, and 72 °C for 5 s. The now double-stranded peptide-encoding DNA sequences were purified using a DNA purification kit. DNA concentration was measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) and 500 ng of the purified product was used for a restriction digest using enzymes HindIII-HF and PstI-HF in Cutsmart buffer. The digested product was again purified using a DNA purification kit and ligated to plasmid pBAD (Thermo Fisher Scientific) digested with the same enzymes[34]. This plasmid harbored the tightly controllable P_{BAD} promoter for peptide gene expression, a pBR322 replication of origin, and a resistance gene encoding for beta-lactamase. For ligation, pBAD was purified using a 1% agarose gel and a DNA gel recovery kit after digestion. Next, T4 ligase (800 units) was used to ligate 100 ng of cut pBAD vector and 10 ng peptide-encoding DNA sequences in T4 ligase buffer (molar ratio of 7:1 insert:vector). The ligation mix was incubated for 14 h at 16 °C. The ligation product was dialyzed in deionized water using filters (MilliporeSigma, Burlington, USA) and 1 µl of the mix was used to transform 20 µl of CloneCatcher Gold DH5G Electrocompetent *E. coli* (Genlantis, Burlington, USA) cells using electroporation. Recovered cells were plated and incubated overnight on LB agar plates supplemented with carbenicillin. Afterward, ~ 500,000 colonies were washed off the plates using LB medium, and the plasmids containing the peptide-encoding DNA sequences were extracted from 2.5×10^9 cells using a plasmid isolation kit. An aliquot of 5 ng of these plasmids was used to transform *E. coli* TOP10 cells using the protocol from the transformation above. A total of 1,000,000 colonies were recovered from the plates after overnight incubation by washing with LB medium, the suspension was diluted to OD = 1 with LB-medium, glycerol was added to a final concentration of 20% (v/v), and aliquots of 500 million cells were stored at − 80 °C.

Growth experiment.   Three aliquots of *E. coli* TOP10 harboring the peptide-encoding DNA sequences on the pBAD plasmid were thawed and added to three 1 l baffled shake flasks containing 100 ml of LB medium + 100 µg ml^{-1} carbenicillin. The cultures were grown for roughly 7.5 h at 37 °C. When the OD reached 0.2, the cultures were supplemented with ʟ-arabinose to a final concentration of 0.3% (w/v) to induce peptide expression. Cell samples were taken from each biological replicate at the point of induction and 1.5 h, 3 h, and 4.5 h post-induction. The plasmids were extracted from all samples using a plasmid isolation kit.

NGS.   For the generation of Me^x growth curves, peptide-encoding DNA sequences on plasmids, collected from the three replicates across four time points during the growth experiment, were sequenced by NGS. Additionally, the abundance of peptide-encoding DNA sequences in the original oligonucleotide pool and after transformation of the assay strain *E. coli* TOP10 was assessed by NGS as well. Peptide-encoding DNA sequences were amplified by primer 1 and primer 2 using 100 ng of plasmid and the PCR-amplification protocol mentioned before, but only for 10 cycles to avoid amplification bias. The amplification product was purified using an agarose gel. Single Index PentAdapters from Pentabase were used to prepare PCR-free libraries with the KAPA Hyper-Prep Kit (now Roche, Basel, CH) according to the manufacturer's specifications. Libraries were quantified using the qPCR KAPA Library Quantification Kit, pooled and sequenced PE 2 × 151 with an Illumina HiSeq 2500 using v4 SBS chemistry. Roughly 10% genomic PhiX library as spike-in to increase sequence diversity. Basecalling was done with bcl2fastq v2.20.0.422. The resulting fastq files were processed using in-house software written in R and C. This software aligns each sequence to our reference table of 12,412 sequences linking peptide-encoding DNA sequences and peptide sequence, identifies mismatches and sequencing errors, and counts how often each peptide-encoding DNA sequence was sequenced in each sample. NGS read counts for each sequence analyzed in Me^x were listed with a unique identifier (ID) in File S2.

Generation of Me^x growth curves.   We used the standard workflow of DESeq2 (NGS read count normalization, dispersion estimates, and Wald's tests) to analyze NGS read counts[35]. Only sequences that passed independent filtering were included in further analyses (= 10,633). To draw growth curves for each peptide-expressing strain, we calculated the log2-fold changes of NGS read counts (listed for each ID in File S2) between

the time of induction and all other time points (1.5 h, 3.0 h, and 4.5 h post-induction). A Bayesian shrinkage estimator was employed to shrink the log2 fold-change for each ID (lfcShrink$_{ID}$) between all time points using the R/Bioconductor package 'apeglm'[36]. To draw the Me^x growth curves, we calculated a strain-specific OD$_{ID}$ at each time point according to Eq. (1). OD values at the specific time points were averaged values from all three biological replicates (Fig. S5). The OD$_{ID}$ (0 h) for each peptide-expressing strain was set to 0.2 at the time of induction as lfcShrink$_{ID}$ (0 h) = 0 and OD = 0.2. This enabled us to compare peptide-expressing strains of different abundancies (see Fig. S6). OD$_{ID}$ values can be interpreted as the OD values that would have been measured when incubating the respective strain individually in the same experiment, i.e. in this case in LB medium in a 100 ml shake flasks.

$$OD_{ID}(t) = OD(t) \times 2^{lfcShrinkID(t)} \tag{1}$$

To find Me^x-active peptides, we also performed a one-sided Wald's test, with the alternative hypothesis that the expression of a given peptide leads to a reduced OD$_{ID}$ 1.5 h and 4.5 h post-induction. We rejected the null hypothesis at significance level alpha = 0.05. Peptides with a $p < 0.05$ (after adjustment for multiple testing using the Benjamini–Hochberg method) after 4.5 h are considered Me^x-active peptides. Peptides with $p < 0.05$ after 1.5 h do significantly inhibit growth already after 1.5 h. All values and results are reported in File S1.

Monoseptic growth experiments. Taking the OD$_{ID}$ (4.5 h) of each peptide-expressing strain, we could rank all peptides by their growth inhibitory effect. We selected 110 peptides (Ranks 1–50, 100–119, 1000–1019, and 10,000–10,019) and then generated an identical copy of the strain previously used in Me^x for its expression. First, the corresponding peptide-encoding DNA-sequences were synthesized as gene fragments. An aliquot of 400 ng of each gene fragment was directly used for a restriction digest using enzymes HindIII-HF and Pst-HF in Cutsmart buffer. The product was purified using a DNA purification kit. Next, T4 ligase (800 units) was used to ligate 50 ng of identically digested pBAD vector and 10 ng of purified gene fragment in T4 ligase buffer for 14 h at 16 °C. The ligation product was purified using a DNA purification kit. An aliquot of 5 μl of the purified ligation product was then used to transform chemically competent E. coli TOP10 cells. From the resulting colonies, we isolated one strain, sequence-verified the correct assembly of the expression plasmid, and stored it after overnight growth in glycerol at −80 °C. For the growth experiment, we first re-isolated single colonies on solid media and then picked three clones, incubated them separately overnight, and inoculated them into 200 μl fresh LB medium containing 0.3% (w/v) L-arabinose to a final OD of 0.01 into 96-well microtiter plates (Greiner). Growth was recorded by measuring OD in a Tecan Infinite 200 PRO (Tecan, Männedorf, CH) for 4.5 h (37 °C, 1.5 mm orbital shaking).

Enrichment analyses. We used Fisher's exact test to assess the over- or underrepresentation of Me^x-actives in various groups. This amounts to a hypergeometric test to assess the significance of drawing n active peptides in a group of k, from a population of size N containing K active peptides. We rejected the null hypothesis at significance level alpha = 0.05. Groups with a $p < 0.05$ had a significantly different representation of active peptides compared with the overall population. When adjusting for multiple testing, we used the Benjamini–Hochberg method.

Peptide classifications. The physicochemical parameters of the peptides were calculated at pH 7 using the R package 'Peptides' (https://cran.r-project.org/package=Peptides). For charge, we used the method by Lehninger[37]. For hydrophobicity (or GRAVY index), we used the calculations by KyteDoolittle[38]. The information for each PARENT such as the name, chemical modification, activity, 3D-structure, was extracted from the APD website (https://aps.unmc.edu/) using an in-house R script. The information on the species from which a specific peptide sequence originated, was extracted from the *tblastn* search and the APD website. The entire taxonomic classifications (kingdom, phylum, class) for each species were extracted, if available, from the Global Biodiversity Information Facility Data Portal (https://gbif.org) using the R package 'taxize' (https://cran.r-project.org/package=taxize). The results are summarized in File S1.

Membrane damage assay using intracellularly synthesized peptides. We selected the peptide-expressing strains of rank 1–50 in Me^x that we had previously constructed for the monoseptic growth assay. Additionally, we selected the strain expressing the inactive control peptide HNP-1$_{3425\ APD}$, a peptide known to be inactive if expressed in E. coli[8]. Each strain was re-isolated on solid media from frozen stock and incubated overnight. Then, two colonies were picked and incubated overnight in 96-deep-well polypropylene plates. These cultures were used to inoculate fresh media containing 0.3% (w/v) L-arabinose to a final OD of 0.01 into 96-well microtiter plates. The plates were then incubated on for 4.5 h (37 °C, 1.5 mm orbital shaking). After 4.5 h, an aliquot of 50 μl of cell suspension a Tecan Infinite 200 PRO plate reader was added to 150 μl of phosphate-buffered saline into a fresh 96-well microtiter plate. Propidium iodide (PI) was added to a final concentration of 1 μg ml^{-1}. PI is a DNA-intercalating dye that cannot pass an intact cytoplasmic membrane[39]. For each sample, PI fluorescence ($\lambda_{Ex} = 579$ nm/$\lambda_{Em} = 616$ nm) of ~ 10,000 cells were analyzed using a flow cytometer LSR Fortessa (BD Biosciences, Allschwil, CH). To determine the membrane damaging properties of each of the expressed peptides, we calculated the fraction of cells in percent for which a PI uptake was measured using the software FlowJo V10 (BD Biosciences).

Stress response assay using intracellularly synthesized peptides. We selected peptide-expressing strains of rank 1–50, previously generated for the monoseptic growth assay. Additionally, we selected the strain

expressing the inactive control peptide HNP-1$_{3425\,APD}$. Moreover, two plasmids (cloning vector: puA66) containing either the promoter of the gene for recombinase A (P$_{recA}$) or for the gene for cold shock protein A (P$_{cspA}$) were purified from the *E. coli* Alon collection[15]. Both plasmids contained a transcriptional fusion of their promoter with a downstream gene for green fluorescent protein (*gfp*), an additional kanamycin resistance cassette, and a pSC101 origin of replication. We transformed each of the 51 peptide-expressing *E. coli* strains with each of the two plasmids to generate 102 different strains and incubated them overnight on solid media. Then, three colonies were picked and incubated overnight. These cultures were used to inoculate fresh media containing 0.3% (w/v) L-arabinose to a final OD of 0.05 into 96-well microtiter plates. We recorded OD and GFP expression (λ_{Ex} 488 nm/λ_{Em} 530 nm) after 1.5 h and 4.5 h using a Tecan Infinite 200 PRO (37 °C, 1.5 mm orbital shaking). For each strain, we calculated the specific fluorescence change between the two time points [GFP/OD (4.5 h)– GFP/OD (1.5 h)]. Statistical significance was calculated by one-sided *t* tests, adjusted for multiple testing by Benjamini–Hochberg, using the signal of HNP-1$_{3425\,APD}$ as null distribution. We rejected the null hypothesis at significance level alpha = 0.05.

Purification of chemically synthesized peptides. Peptides were obtained from Pepscan (Lelystad, NL) in > 90% purity or in crude format and subsequently purified to > 90% purity in-house. For the latter, crude peptides were dissolved in 5 ml DMSO and 15 ml 0.1% aqueous trifluoroacetic acid, TFA. HPLC-purification of the dissolved crude peptides was performed on an ÄKTAexplorer chromatography system (GE Heathcare, SE). The entire peptide sample was loaded onto a RP C18 column (PRONTOSIL 120 C18 10 μm, 250 × 20 mm, 50 × 20 mm precolumn, Bischoff, Leonberg, DE), heated to 30 °C and operated at a flow rate of 10 ml min^{-1} using 0.1% aqueous TFA as solvent A and acetonitrile supplemented with 0.1% TFA as solvent B. The ratios of A to B were adapted for each peptide and typical values are given below. The column was equilibrated with the peptide-specific mixture of solvent A and solvent B (0–20%) prior to injection. After injection and an initial wash step of 6 min a gradient was imposed with the same mixture, and then a gradient was applied, in the course of which the amount of solvent B was increased to 50–90% in 40 min. The column was washed with 95% solvent B for 8 min and equilibrated with the specific solvent A/solvent B mixture for the next run for 13 min. Peptide elution was monitored spectrophotometrically at 205 nm, and generally the main peptide peak was collected. The sample was frozen at − 80 °C for > 2 h and lyophilized (approx. 18 h) using a freeze-dryer (Alpha 2–4 LDplus, Christ, DE), connected to a vacuum pump (RC6, Vacuubrand, DE). The lyophilized peptides were dissolved in 1 ml DMSO and stored at − 20 °C. The concentration of the peptide stocks was determined via HPLC using an Agilent 1200 series HPLC system. Each peptide stock was analyzed as a 1:100 dilution in water. An aliquot of 10 μl of the peptide stock was injected onto an RP-C18 column (ReproSil-Pur Basic C18, 50 × 3 mm, Dr. Maisch, Germany) operated with water supplemented with 0.1% TFA as solvent A and acetonitrile supplemented with 0.1% TFA as solvent B. Separation was performed using the same concentration profile previously used for purification. The concentration was measured using the integrated peak area at 205 nm and then calculated using peptide-specific absorption properties[40,41].

Measurement of the MIC using chemically synthesized peptides. On the same day at which MIC assays were executed, purified peptides were thawed and the concentration was determined by HPLC as described before. *E. coli* TOP10 cells were grown in Mueller Hinton Broth (MHB) or diluted MHB (25% of the original strength) overnight to stationary phase. Diluted MHB has been frequently used to assay antimicrobial peptides[42]. The cultures were then supplemented with 20% glycerol, aliquoted, and frozen at − 80 °C. For MIC measurements, a frozen stock of the cells was thawed, resuspended in MHB or 25% MHB to adjust to a density of 5×10^5 cells ml^{-1} in the experiment, and distributed to microtiter plate wells by an automated liquid handling system (Hamilton, Bonaduz, CH). Then the peptides were added by the liquid handling system in twofold dilutions using minimum of 100 μg ml^{-1} as the highest concentration. MICs were determined as broth microdilution assay in 384-well flat bottom polypropylene plates (Falcon 96-Well Flat-Bottom Microplate) adapted from the protocol of Wiegand et al.[43]. The plates were sealed airtight and incubated for 18 h without shaking at 37 °C before reading the OD using a Tecan Infinite 200 PRO plate reader. The MIC value corresponded to the concentration at which no growth of the bacterial strain was observed (< 5% of the OD value of the growth control). MIC experiments were performed at least in triplicate.

Membrane damage assay using chemically synthesized peptides. To measure extracellular membrane damage, *E. coli* TOP10 [pSEVA271-GFP] and the peptide dilutions were prepared as described for the MIC measurements (using 100% MHB as medium) but covering a concentration range of 16 × MIC to MIC/16 in twofold dilutions steps with a final assay volume of 200 μl. The bacterial strain suspension was furthermore supplemented with a final concentration of 1 μg ml^{-1} propidium iodide just before pipetting the assay. After 1 h incubation at room temperature membrane damage (= release of intracellularly expressed GFP and/or uptake of extracellularly added PI) was assessed by flow cytometry using a Fortessa Analyzer (BD Biosciences; 488 nm laser with 530/30 nm bandpass filter and 579 nm laser with 610/20 nm bandpass filter). The fractions of PI-positive and GFP-positive were determined with the same gate for all populations using the FlowJo V10 software (BD Biosciences). The extracellular membrane integrity assay was performed in biological duplicates analyzing at least 10,000 cells in each experiment.

Hemolysis assay using chemically synthesized peptides. Two samples of human blood were obtained from a blood bank (Blutspendezentrum SRK at the University Hospital Basel). The samples were pooled and erythrocytes were isolated by repeated centrifugation at 500×*g* for 10 min, removal of the blood plasma and resuspending the remaining cells in an equal volume of DPBS. Following last resuspension, erythro-

cytes were diluted 1:50 in DPBS. For the hemolysis assay, a log2 serial dilution of each peptide was prepared as described for the MIC but using DPBS and a 96-well plate (U-bottom, PP, 650201, Greiner) with a final volume of 200 µl. As lysis control, 2.5% Triton-X100 in DPBS was used in well 10, well 11 served as non-treated control (no peptide added), and well 12 as blank. To each well of the dilution plate, 100 µl of the red blood cells suspension was added. The plate was incubated for 1 h at 37 °C. After the incubation, the plate was centrifuged at 500×g for 10 min and 100 µl of the supernatant was transferred to a clean 96-well plate (F-bottom, PS, 655101, Greiner). The absorbance was measured at 540 nm using an Infinite M1000 PRO plate reader (Tecan) and corrected by the measurements from the untreated wells. The lysis of each peptide concentration was expressed relative to the lysis control (set as 100% lysis). The hemolysis assay was performed in triplicate.

References

1. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111–129 (2015).
2. Henninot, A., Collins, J. C. & Nuss, J. M. The current state of peptide drug discovery: Back to the future? *J. Med. Chem.* **61**, 1382–1414 (2018).
3. Lewis, K. The science of antibiotic discovery. *Cell* **181**, 29–45 (2020).
4. Wang, G., Li, X. & Wang, Z. APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44**, D1087–D1093 (2016).
5. Hao, Y. *et al.* SmProt: A database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief. Bioinform.* **19**, 636–643 (2017).
6. Raventos, D. *et al.* Improving on natures defenses: Optimization and high throughput screening of antimicrobial peptides. *Combin. Chem. High Throughput Screen.* **8**, 219–233 (2005).
7. Jensen, T. D. *et al.* The PepSeq pipeline. in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* 139–144 (ACM, 2018). https://doi.org/10.1145/3233547.3233599
8. Tucker, A. T. *et al.* Discovery of next-generation antimicrobials through bacterial self-screening of surface-displayed peptide libraries. *Cell* **172**, 1–11 (2018).
9. Guralp, S. A., Murgha, Y. E., Rouillard, J.-M. & Gulari, E. From design to screening: A new antimicrobial peptide discovery pipeline. *PLoS One* **8**, 1–7 (2013).
10. NCBI Resource Coordinators. Database resources of the National Center for Niotechnology information. *Nucleic Acids Res.* **45**, D12–D17 (2017).
11. Pearson, W. R. An introduction to sequence similarity ('homology') searching. *Curr. Protoc. Bioinform.* **42**, 1–8 (2013).
12. Watt, P. M. Screening for peptide drugs from the natural repertoire of biodiverse protein folds. *Nat. Biotechnol.* **24**, 177–183 (2006).
13. Torres, M. D. T., Sothiselvam, S., Lu, T. K. & de la Fuente-Nunez, C. Peptide design principles for antimicrobial applications. *J. Mol. Biol.* **431**, 3547–3567 (2019).
14. Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* **24**, 1551–1557 (2006).
15. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **3**, 623–628 (2006).
16. Elad, T., Seo, H. B., Belkin, S. & Gu, M. B. High-throughput prescreening of pharmaceuticals using a genome-wide bacterial bioreporter array. *Biosens. Bioelectron.* **68**, 699–704 (2015).
17. Bianchi, A. A. & Baneyx, F. Stress responses as a tool to detect and characterize the mode of action of antibacterial agents. *Appl. Environ. Microbiol.* **65**, 5023–5027 (1999).
18. Graf, M. *et al.* Proline-rich antimicrobial peptides targeting protein synthesis. *Nat. Prod. Rep.* **34**, 702–711 (2017).
19. Yang, S. T., Shin, S. Y., Hahm, K. S. & Kim, J. I. Different modes in antibiotic action of tritrpticin analogs, cathelicidin-derived Trp-rich and Pro/Arg-rich peptides. *Biochim. Biophys. Acta Biomembr.* **1758**, 1580–1586 (2006).
20. Luther, A. *et al.* Chimeric peptidomimetic antibiotics against Gram-negative bacteria. *Nature* **576**, 452–458 (2019).
21. Greco, I. *et al.* Correlation between hemolytic activity, cytotoxicity and systemic in vivo toxicity of synthetic antimicrobial peptides. *Sci. Rep.* **10**, 13206 (2020).
22. Pagano, M. & Faggio, C. The use of erythrocyte fragility to assess xenobiotic cytotoxicity. *Cell Biochem. Funct.* **33**, 351–355 (2015).
23. Cintas, L. M. *et al.* Enterocins L50A and L50B, two novel bacteriocins from *Enterococcus faecium* L50, are related to staphylococcal hemolysins. *J. Bacteriol.* **180**, 1988–1994 (1998).
24. Corzo, G. *et al.* Oxyopinins, large amphipathic peptides isolated from the venom of the wolf spider *Oxyopes kitabensis* with cytolytic properties and positive insecticidal cooperativity with spider neurotoxins. *J. Biol. Chem.* **277**, 23627–23637 (2002).
25. Atanasov, A. G. *et al.* Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
26. Le, C.-F., Fang, C.-M. & Sekaran, S. D. Intracellular targeting mechanisms by antimicrobial peptides. *Antimicrob. Agents Chemother.* **61**, 1–16 (2017).
27. Magana, M. *et al.* The value of antimicrobial peptides in the age of resistance. *Lancet Infect. Dis.* **20**, e216–e230 (2020).
28. Haney, E. F., Straus, S. K. & Hancock, R. E. W. Reassessing the host defense peptide landscape. *Front. Chem.* **7**, 1–22 (2019).
29. Fjell, C. D., Hiss, J. A., Hancock, R. E. W. & Schneider, G. Designing antimicrobial peptides: Form follows function. *Nat. Rev. Drug Discov.* **11**, 37–51 (2012).
30. Pinel-Marie, M. L., Brielle, R. & Felden, B. Dual toxic-peptide-coding *Staphylococcus aureus* RNA under antisense regulation targets host cells and bacterial rivals unequally. *Cell Rep.* **7**, 424–435 (2014).
31. Milletti, F. Cell-penetrating peptides: Classes, origin, and current landscape. *Drug Discov. Today* **17**, 850–860 (2012).
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
33. Sonnhammer, E. L. L. & Hollich, V. Scoredist: A simple and robust protein sequence distance estimator. *BMC Bioinform.* **6**, 1–8 (2005).
34. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
35. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
36. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

37. Butterworth, P. J. Lehninger: Principles of biochemistry (4th edn) D. L. Nelson and M. C. Cox, W. H. Freeman & Co., New York, 1119 pp (plus 17 pp glossary), ISBN 0-7167-4339-6 (2004). *Cell Biochem. Funct.* **23**, 293–294 (2005).
38. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
39. Mortimer, F. C., Mason, D. J. & Gant, V. A. Flow cytometric monitoring of antibiotic-induced injury in *Escherichia coli* using cell-impermeant fluorescent probes. *Antimicrob. Agents Chemother.* **44**, 676–681 (2000).
40. Anthis, N. J. & Clore, G. M. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci.* **22**, 851–858 (2013).
41. Pelillo, C. *et al.* Cellular internalization and cytotoxicity of the antimicrobial proline-rich peptide Bac7(1–35) in monocytes/macrophages, and its activity against phagocytosed *Salmonella typhimurium*. *Protein Pept. Lett.* **21**, 382–390 (2014).
42. Holfeld, L., Knappe, D. & Hoffmann, R. Proline-rich antimicrobial peptides show a long-lasting post-antibiotic effect on Enterobacteriaceae and *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **73**, 933–941 (2018).
43. Wiegand, I., Hilpert, K. & Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3**, 163–175 (2008).

Acknowledgements

Author contributions

M.H., P.K. and S.S. conceived the project. M.H. and S.P. supervised experimental work. N.B. supervised computational work. P.K. performed experiments. P.K. and S.S. purified peptides and developed software for designing the oligonucleotide architecture and processing of NGS data. M.C. developed computational methods to analyze sequence disparity. P.K., S.S. and M.C. analyzed data. P.K. and M.C. performed statistical analyses. P.K. and M.H. wrote the manuscript with input from all authors.

Competing interests

A patent application covering this technology has been filed, in which ETH Zurich is the owner, and M.H., S.S., P.K and S.P. are the inventors. Other authors do not have competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-07755-7.

Correspondence and requests for materials should be addressed to M.H.

Reprints and permissions information is available at www.nature.com/reprints.

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