Cell-free biosynthesis combined with deep learning accelerates de novo-development of antimicrobial peptides

Amir Pandi*[1], David Adam[1,2], Amir Zare[1], Van Tuan Trinh[3], Stefan L. Schaefer[4], Marie Wiegand[5], Björn Klabunde[5], Elizaveta Bobkova[1], Manish Kushwaha[6], Yeganeh Foroughijabbari[1], Peter Braun[2,7], Christoph Spahn[8], Christian Preußer[9,10], Elke Pogge von Strandmann[9,10], Helge B. Bode[8,11,12,13,14], Heiner von Buttlar[2,7], Wilhelm Bertrams[5], Anna Lena Jung[5,15], Frank Abendroth[3], Bernd Schmeck[5,14,15,16,17,18], Gerhard Hummer[4,19], Olalla Vázquez[3,14], and Tobias J. Erb*[1,14]

[1] Department of Biochemistry and Synthetic Metabolism, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany
[2] Bundeswehr Institute of Microbiology, Munich, Germany
[3] Department of Chemistry, Philipps-University Marburg, Germany
[4] Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt am Main, Germany
[5] Institute for Lung Research, Universities of Giessen and Marburg Lung Center, Philipps-University Marburg, German Center for Lung Research (DZL), Marburg, Germany
[6] Université Paris-Saclay, INRAe, AgroParisTech, Micalis Institute, Jouy-en-Josas, France
[7] German Center for Infection Research (DZIF), Munich, Germany
[8] Department of Natural Products in Organismic Interactions, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany
[9] Institute for Tumor Immunology, Center for Tumor Biology and Immunology, Philipps-University Marburg, Marburg, Germany
[10] Core Facility Extracellular Vesicles, Center for Tumor Biology and Immunology, Philipps University of Marburg, Marburg, German
[11] Molecular Biotechnology, Department of Biosciences, Goethe University Frankfurt, Frankfurt am Main, Germany
[12] Department of Chemistry, Chemical Biology, Phillips-University Marburg, Germany
[13] Senckenberg Gesellschaft für Naturforschung, Frankfurt, Germany
[14] SYNMIKRO Center of Synthetic Microbiology, Marburg, Germany
[15] Core Facility Flow Cytometry – Bacterial Vesicles, Philipps-University Marburg, Marburg, Germany
[16] Department of Medicine, Pulmonary and Critical Care Medicine, University Medical Center Marburg, Philipps-University Marburg, Marburg, Germany
[17] Institute for Lung Health (ILH), Giessen, Germany
[18] Member of the German Center for Infectious Disease Research (DZIF), Marburg, Germany
[19] Institute for Biophysics, Goethe University Frankfurt, Frankfurt am Main, Germany

* To whom correspondence should be addressed: toerb@mpi-marburg.mpg.de and amir.pandi@mpi-marburg.mpg.de

Abstract

Bioactive peptides are key molecules in health and medicine. Deep learning holds a big promise for the discovery and design of bioactive peptides. Yet, suitable experimental approaches are required to validate candidates in high throughput and at low cost. Here, we established a cell-free protein synthesis (CFPS) pipeline for the rapid and inexpensive production of antimicrobial peptides (AMPs) directly from DNA templates. To validate our platform, we used deep learning to design thousands of AMPs de novo. Using computational methods, we prioritized 500 candidates that we produced and screened with our CFPS pipeline. We identified 30 functional AMPs, which we characterized further through molecular dynamics simulations, antimicrobial activity and toxicity. Notably, six de novo-AMPs feature broad-spectrum activity against multidrug-resistant pathogens and do not develop bacterial resistance. Our work demonstrates the potential of CFPS for production and testing of bioactive peptides within less than 24 hours and <10$ per screen.

42   **Main**

43   According to the world health organization, antimicrobial resistance (AMR) is among the top 10
44   global health threats[1]. In 2019 alone, multidrug-resistant bacteria including pathogenic
45   *Escherichia coli*, ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella
46   pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Enterobacter* spp.),
47   *Streptococcus pneumoniae*, and *Mycobacterium tuberculosis* caused 1.27 million deaths[2]. This
48   number is predicted to reach 10 million annually by 2050[2]. Despite this looming threat, the
49   development of new antimicrobials is lagging behind. While more than 4,000 immuno-oncology
50   compounds were in clinical trials in 2021, only 40 antimicrobials (of which none is active against
51   multi-drug resistant Gram-negative bacteria) were subjected to clinical studies[3], highlighting the
52   urgent need to increase the development of novel antimicrobial compounds.

53   One promising class of antimicrobial compounds are antimicrobial peptides (AMPs)[4–8]. AMPs are
54   peptides of 12-50 amino acids (AA), which have evolved as part of nature's antimicrobial arsenal
55   of bacteria as well as the innate immune system of multicellular organisms, millions of years
56   before humans started to use antibiotics[4,6,8]. Compared to classical antibiotics, AMPs show
57   decreased resistance development mainly because (i) most AMPs act directly at the cell
58   membrane, (ii) show a relatively high killing rate, and (iii), resistance against AMPs is conferred
59   by rather non-specific mechanisms, which reduces the chances of mutational and/or horizontal
60   gene transfer events[4]. Overall, this makes AMPs interesting candidates for next-generation
61   antimicrobials.

62   About 5,000 AMPs have been characterized to date, most of which are of natural origin. However,
63   these 5,000 AMPs span only a tiny fraction of the possible solution space that nature could have
64   explored (~20^{30} for a 30 AA AMP). Yet, our ability to discover or develop new AMPs from this
65   *terra incognita* are still limited. One possibility to overcome this challenge is the use of deep
66   learning models, which are increasingly employed in protein and peptide design[9–13]. In these
67   approaches, known as generative deep learning, models "learn" the natural protein sequence
68   landscape in training sets with unlabeled data to propose new-to-nature protein sequences[14]. This
69   approach is distinct from predictive modeling that uses labeled data to predict specific properties
70   (labels) of proteins from their sequence[15]. Generative and predictive deep learning have been
71   recently used for the discovery of novel AMP sequences, which have been subsequently created
72   and validated through chemical synthesis of the individual candidates[6,16–19]. While this proof-of-
73   principle showcased the potential of deep learning in AMP discovery, a broader application of this
74   approach has been limited due to the lack of convenient methods for the production and screening
75   of more AMP candidates in medium to high-throughput.

76   One possibility to increase the throughput in AMP production is to switch from chemical synthesis
77   to DNA-based bioproduction methods. However, heterologous expression of AMPs in
78   microorganisms, such as *E. coli*, features several disadvantages: (i) it is time- and labor-intensive,
79   (ii) it requires the cloning, production and purification of AMPs from cell cultures, and most
80   importantly, (iii) many (potent) AMP candidates might not be available, as they potentially kill the
81   producer strain upon induction. Cell-free protein synthesis (CFPS) offers a promising solution to
82   these challenges. CFPS systems are in vitro transcription translation (TX-TL) systems that directly
83   use DNA templates for protein biosynthesis[20–22], which allows the production of peptides outside
84   of living cells. Thus these systems can help overcome potential cellular toxicity effects, and open
85   up the way for the rapid, small-scale production of several hundreds of peptides from linear DNA
86   in parallel.

87   Here, we combined deep learning and CFPS for de novo-design, rapid production and screening
88   of AMPs at small scale within 24 hours, and less than 10$ per individual AMP production assay

89   (excluding cost for the DNA fragment). Having explored ~500,000 theoretical sequences, we
90   screened 500 AMP candidates to identify 30 functional AMPs, which are completely unrelated to
91   any natural sequences. Notably, six of these AMPs exhibited high antimicrobial activity against
92   multidrug-resistant pathogens, showed no emergence of resistance and only minimal toxicity on
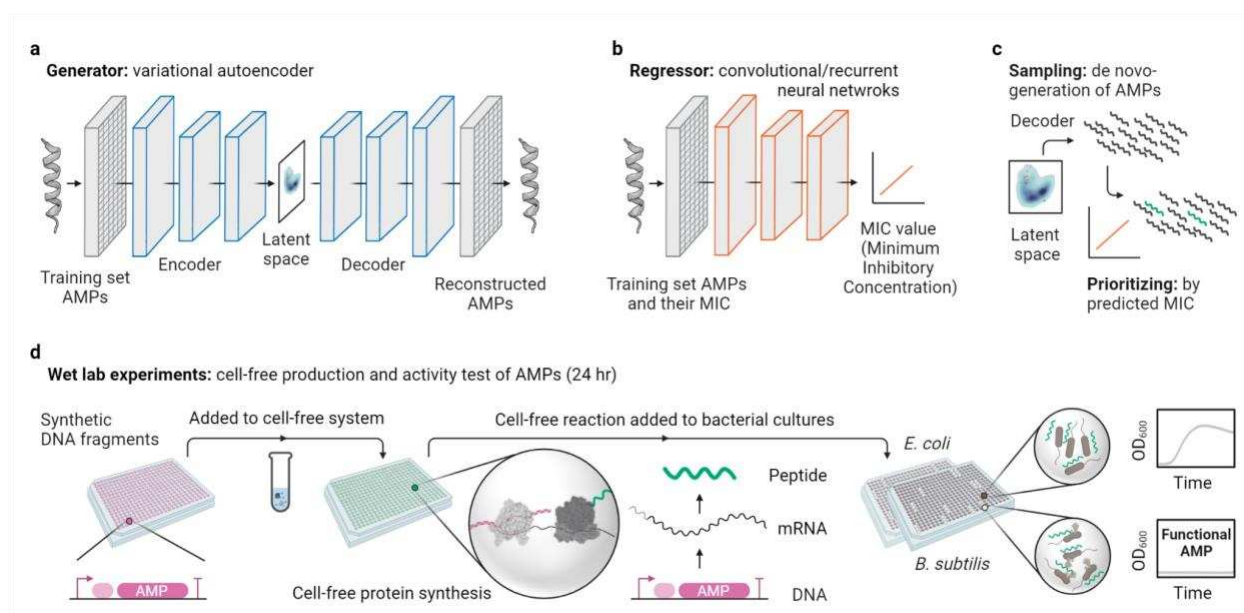93   human cells.

94   **Results**

95   De novo AMP design using deep learning

96   For de novo-design of AMPs, we adapted two versions of deep generative variational
97   autoencoders (VAE) from a previous study[23]. Generative VAE are unsupervised learning models,
98   which take as input only AMP sequences, and comprise an encoder, a latent space, and a
99   decoder. During model training, the encoder compresses the input sequences into a low-
100 dimensional space ("latent space"), while the decoder aims at reconstructing sequences from this
101 latent space (**Fig. 1a**). We first pretrained the VAEs using ~1.5 million peptide sequences from
102 UniProt as a generic dataset. Second, performed transfer learning on the pretrained VAEs using
103 a dataset of ~5,000 known AMPs to set up the latent space to be used for de novo AMP generation
104 (**Methods** and **Supplementary Table 1**).

105 To reduce the number of AMPs for experimental testing, we set up a method to select potential
106 candidates according to their predicted bioactivity, i.e., their Minimum Inhibitory Concentration
107 (MIC). To that end, we established predictive deep learning models that we trained with the
108 sequence and respective experimental MIC values of ~5,000 known AMPs (sequence-MIC
109 relationship, **Fig. 1b**). As regressors, we used convolutional neural networks (CNN) and recurrent
110 neural networks (RNN) (**Methods** and **Supplementary Table 1**).

111 To identify interesting AMP candidates, we first generated new AMPs by sampling points from
112 the latent space and subsequently feeding them into the decoder, which yielded peptide
113 sequences that share the same properties but are novel compared to the training dataset. These
114 de novo AMPs were then prioritized by the regressors according to their predicted MICs. In five
115 rounds, using different versions of models, we generated ~500,000 new peptides by sampling
116 from the latent space. We filtered these peptides by length and viability to ~50,000 candidates
117 and prioritized 500 AMP candidates for wet lab bioactivity test (**Fig. 1c** and **Supplementary Table
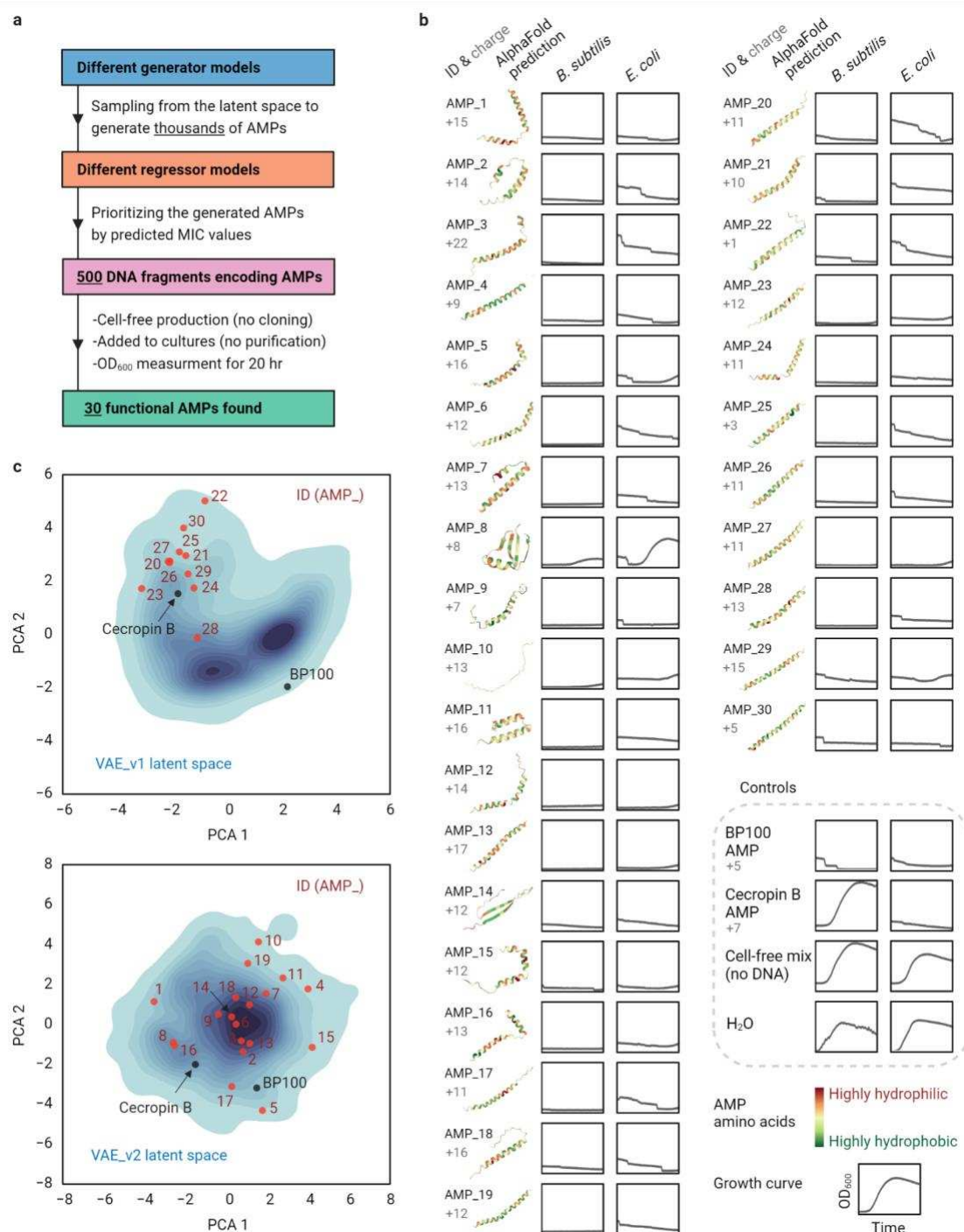118 2**).

119
120 **Fig. 1: The workflow for de novo-development of AMPs via deep learning and cell-free biosynthesis. a,**
121 Generative variational autoencoders (VAE) for de novo-design of AMPs after being trained on known AMP sequences.
122 **b,** Predictive convolutional or recurrent neural networks as regressors for the MIC prediction after being trained on
123 known AMPs and their MIC. **c,** Trained generative and predictive models are used for sampling from the latent space
124 (de novo-design of AMPs) and prioritization of AMPs (predicting their MIC), respectively. **d,** Experimental pipeline for
125 rapid cell-free biosynthesis of the designed AMPs from synthetic DNA fragments and direct testing of produced AMPs
126 in the cell-fee mix to bacterial cultures followed by overnight continuous growth assay.

127 Cell-free biosynthesis enables rapid screening for functional AMPs

128 To establish CFPS-based screening of AMPs, we designed an experimental pipeline for the high-
129 throughput synthesis and testing of AMPs in 384-well format (**Fig. 1d**). The system is based on
130 linear DNA templates, which comprise a T7 promoter and a ribosome binding site (RBS), to initiate
131 transcription (TX) and translation (TL), followed by the AMP coding region, and a T7 terminator.
132 After adding the DNA template (10 nM) directly into 10 µL of a cell-free TX-TL system, AMPs were
133 produced within 4 hours (**Methods**). To test the antimicrobial activity of the in vitro-produced
134 peptides, 4 µL of the cell-free mix was added into a final volume of 20 µL cultures of *E. coli* (Gram-
135 negative) and *Bacillus subtilis* (Gram-positive). Following the OD_{600} measurement for 20 hours
136 allowed identification of those peptides that show antimicrobial activity by suppressing growth.
137 Overall, the entire process of CFPS with subsequent bioactivity tests takes ~24 hours, as the
138 system works with linear DNA and does not require any extensive cloning or peptide purification
139 steps.

140 First, we validated the screening pipeline with two known AMPs, BP100[24] and Cecropin B[25], and
141 then screened 500 AMP candidates (see above) in five subsequent design-predict-build-test
142 cycles to identify 30 functional de novo AMPs (**Fig. 2**). During these five rounds, the success of
143 functional AMP discovery increased from 0% to 12.7% from the first to the fifth round, respectively
144 (**Supplementary Table 2**). Because translation initiation rates (TIRs) strongly affect protein
145 yield[26] we tested whether our screen had biased against candidates with low TIR. We calculated
146 the TIR for all sequences tested[27], but could not find a significant difference between the 500
147 candidates tested and the 30 functional AMPs identified (**Supplementary Note 1**,
148 **Supplementary Fig. 3**). Functional AMPs were re-validated in biological triplicates with 10 µL of
149 cell-free mix added to the final volume of 20 µL cultures (**Fig. 2b** and **Supplementary Fig. 1**) and
150 production of AMPs was analyzed through SDS-PAGE (**Supplementary Fig. 2**)

3

151
Fig. 2: Cell-free production of de novo-generated and prioritized AMPs and activity screening against _B. subtilis_ and _E. coli_. a, We used different generative and regressor models (**Supplementary Table 2**) to design and prioritize AMPs in five rounds, produced and screened a total number of 500 AMPs from synthetic DNA fragments and found 30 functional candidates. **b,** The charge and AlphaFold-predicted structure of the functional AMPs with associated slowed/stopped growth curves for _B. subtilis_ and _E. coli_. All (including control) AMPs were produced using CFPS and no peptide purification was carried out prior to the activity test. Growth curves (OD_{600} 0-0.45 over time 4-20 h for all) are the average of n = 3 independent experiments. Growth curves with error bars as standard deviation are provided in **Supplementary Fig. 1**. **c,** 2D-projections of the 50-dimensional latent space were obtained by principal component analysis (PCA) for the two generative variational autoencoders (VAE) that were used for de novo-design of AMPs. Blue
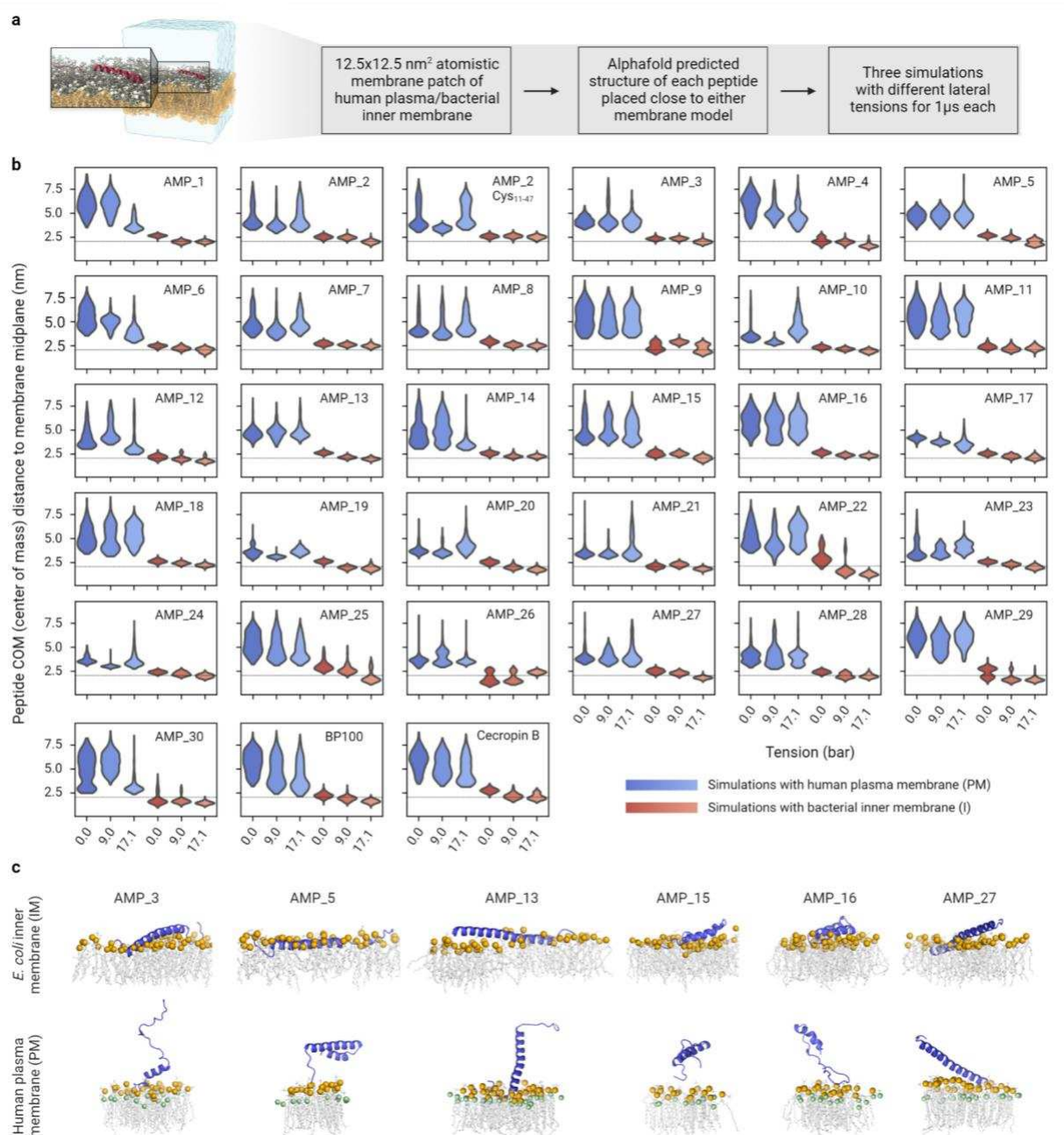
161 color intensity represents the frequency of training AMPs in the latent space. Functional AMPs (red), BP100 and
162 Cecropin B (black) annotated back into the latent space. Source data for b are provided as a **Source Data** file.

163 <u>Functional AMPs are unique but share properties with natural counterparts</u>

164 We next analyzed our de novo-designed AMPs in more detail. AlphaFold[28] predicted that 27 out
165 of the 30 sequences form a helical structure (**Fig. 2b**), which is a common feature of many
166 AMPs[4,5]. Interestingly, AMPs #1-19, generated by VAE 1, showed more structural diversity
167 compared to AMPs #20-30, generated by VAE 2, which is in line with the fact that the two VAEs
168 create two different latent spaces (**Fig. 2c**), thus generating AMPs with distinct structural,
169 physicochemical, and sequence features (**Supplementary Fig. 4**). One of the main
170 characteristics of AMPs is an amphiphilic character that results from alternating cationic and
171 hydrophobic amino acids in the AMP core[4,5]. In contrast to natural AMPs, which mainly feature
172 aliphatic amino acids, the hydrophobic core of our de novo AMPs was mostly aromatic
173 (**Supplementary Fig. 5a**). Phenylalanine was particularly overrepresented at the cost of leucine,
174 which was underrepresented (**Supplementary Fig. 5b**, **Supplementary Table 3**). BLAST
175 searching showed that our de novo AMPs were unique in their sequence. No significant similarity
176 was observed against the UniProt database, encompassing ~240 million entries, nor an AMP
177 sequence from the training data set. (**Supplementary Tables 4-6**, **Supplementary Note 2** for
178 detailed BLAST sequence similarity analyses). Altogether, these results demonstrated that our
179 de novo AMPs shared the physico-chemical building principles with their natural counterparts, but
180 were distinct from them in their amino acid sequences.

181 <u>De novo AMPs prefer bacterial over human membranes</u>

182 Structural and sequence analysis suggested that our de novo AMPs act as amphipathic helices
183 that insert into membranes. We used molecular dynamics (MD) simulations to study the
184 interaction of our AMPs with models of a negatively charged inner membrane of bacteria (IM) and
185 the human plasma membrane (PM) (**Fig. 3a**, **Supplementary Note 3**). According to our
186 simulations, all AMPs bind much stronger to the IM interface than to the PM (**Fig. 3b**). Binding of
187 the AMPs at the IM progressed rapidly, taking at most 200 ns to fully insert into the membrane
188 interface (**Supplementary Fig. 6a**). Once bound, AMPs stayed tightly bound to the IM for the
189 remainder of the simulations. In some cases, we observed a reorientation of the AMP after a few
190 hundred nanoseconds from a shallowly bound state to a binding mode that resided deeper in the
191 membrane (e.g., AMP #29). Several AMPs also partially bind the PM. However, in most cases,
192 this binding is transient with frequent un- and rebinding and without penetrating deeper into the
193 PM, as seen for the IM (**Supplementary Fig. 6b**). Furthermore, all AMPs show a higher number
194 of (mainly electrostatic) interactions with the IM and remain in an ordered structure, compared to
195 the PM system, where fewer interactions were observed and AMPs tended to unfold over time
196 (**Supplementary Fig. 6c-d**). The MD simulations suggested that all our AMPs generally target
197 bacterial membranes over the human plasma membrane, naturally however, the degree of
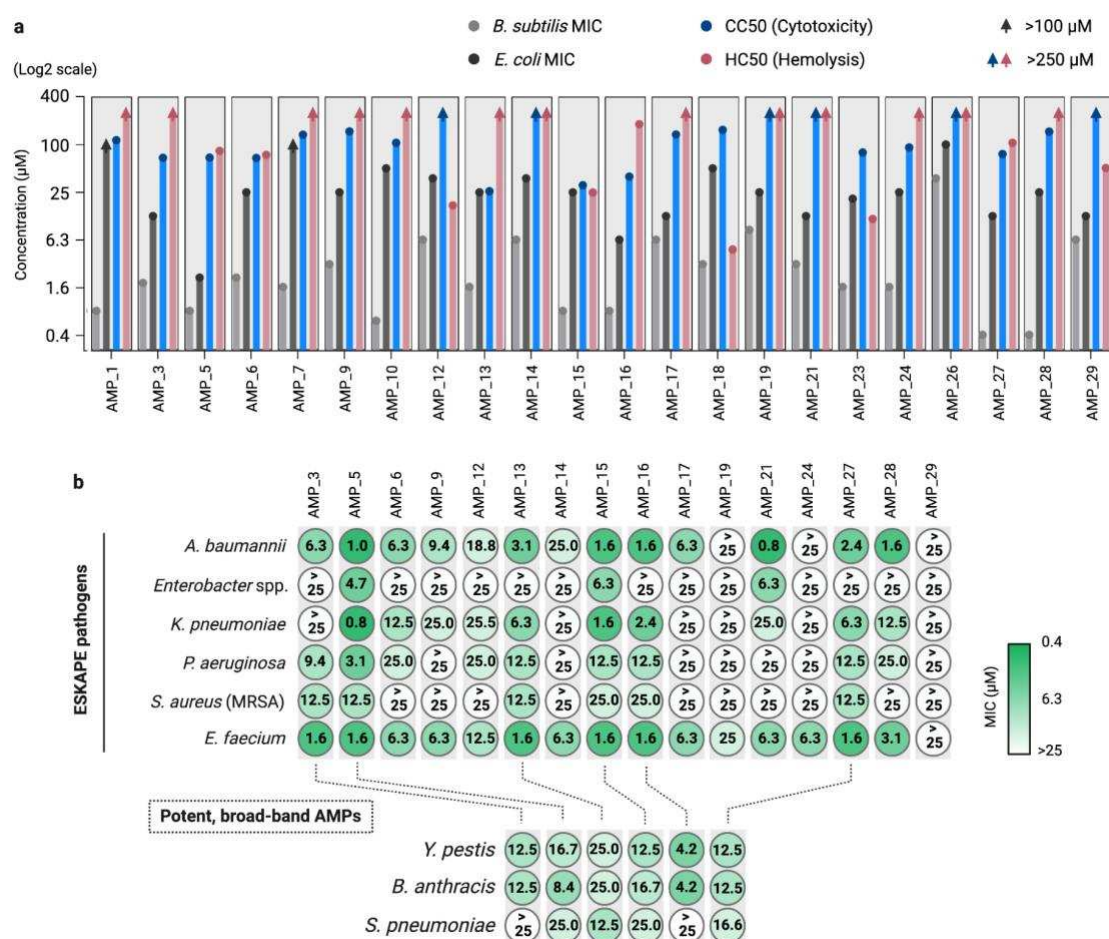198 preference is dependent on the individual AMP.

Fig. 3: Molecular dynamics (MD) simulations of AMP interactions with membranes. a, Overview of simulation setup and exemplary snapshot of AMP #5 after 1 µs of simulated time on the IM. The peptide is shown in red cartoon representation, solvent (water + ions) is shown as transparent blue surface and the membrane is shown as gray licorice representation with lipid headgroup phosphates shown as spheres. **b,** Distributions of distances between the centers of mass of the AMPs and the membrane midplane along the direction of the membrane normal (y axis). Distributions are calculated from the last 940 ns of 1 µs long replicates, run with different lateral membrane tensions (x axis) and with different membranes (blue: PM; red: IM). The thin dotted line indicates the headgroup phosphate positions. AMP #2 was simulated with and without disulfide bond (**Methods**). **c,** Rendered simulation snapshots of potent AMPs (**Fig. 4**) on IM (top) and PM (bottom). The interacting leaflet of the membrane is depicted with gray lipid tails and orange (phosphates) and green (cholesterol oxygen) spheres.

De novo AMPs show favorable MIC to toxicity ratios

211 To obtain pure compounds for cellular assays, we chemically synthesized the functional AMPs
212 and characterized their bioactivity, in particular the minimum inhibitory concentration (MIC)[29], as
213 well as hemolysis (HC50) and cytotoxicity (CC50), both expressed as 50% toxic concentration.
214 Of the 30 candidates, 22 peptides were successfully produced by chemical synthesis. Twenty
215 AMPs showed a MIC of ≤6 µM on *B. subtilis*, and fifteen AMPs showed a MIC of ≤25 µM on *E.*
216 *coli* (**Fig. 4a** and **Supplementary Table 7**). HC50 and CC50 were significantly higher in most
217 cases, with fifteen AMPs showing HC50 >100 µM (thirteen >250 µM) against fresh human red
218 blood cells, and thirteen AMPs showing CC50 >100 µM (six >250 µM) against HCT116 human
219 colon cells (**Fig. 4a** and **Supplementary Table 7**). This indicates that the bioactivity versus toxicity
220 relationship was very favorable for several of our de novo AMPs. We decided to continue with
221 sixteen AMPs that showed a favorable bioactivity to toxicity ratio and excluded six AMPs because
222 of high MIC and/or low HC50/CC50 values (AMP #1, #7, #10, #18, #23, and #26).

223
224 **Fig. 4: Bioactivity characterization of chemically synthesized functional AMPs. a,** Minimum inhibitory
225 concentration (MIC) of the AMPs against *E. coli* and *B. subtilis* (average of n = 3 independent experiments) and HC50
226 (hemolysis) and CC50 (cytotoxicity) values of the AMPs on human red blood cells and HCT116 human colon cells,
227 respectively (average of n = 2 independent experiments). MIC, HC50 and CC50 values are provided in **Supplementary**
228 **Table 7**. **b,** MIC values of the AMPs tested against ESKAPE pathogens including *E. faecium*, Methicillin-resistant *S.*
229 *aureus* (MRSA), *K. pneumoniae*, *A. baumannii*, *P. aeruginosa*, *Enterobacter* spp measured in duplicates (n = 2
230 independent experiments) and MIC of the six potent, broad-band AMPs on *Y. pestis* and *B. anthracis* and *S.*
231 *pneumoniae* as the average of triplicates (n = 3 independent experiments).

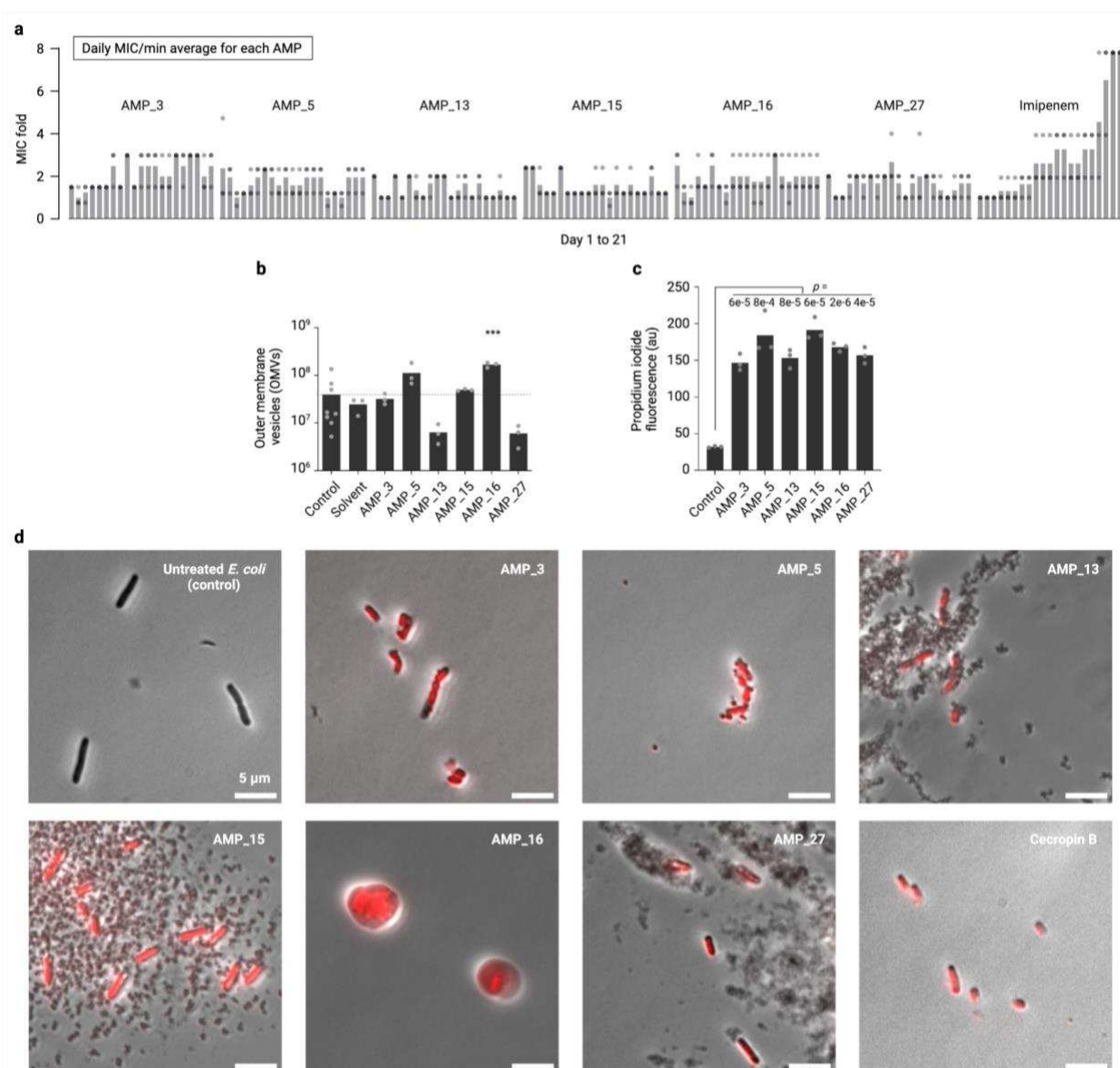232 De novo AMPs show broad-band activities

7

233   In the following, we tested the sixteen remaining de novo AMPs against clinically relevant strains,
234   and in particular multidrug-resistant ESKAPE pathogens (i.e., *Enterococcus faecium*,
235   *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas*
236   *aeruginosa*, and *Enterobacter* spp.; **Fig. 4b**). Our AMPs were most potent against *E. faecium* and
237   *A. baumannii*, resulting in fifteen AMPs with MIC ≤25 μM (thirteen ≤6.3 μM), and thirteen AMPs
238   with MIC ≤25 μM (ten ≤6.3 μM), respectively (**Fig. 4b**). For the rest of the ESKAPE pathogens, *K.*
239   *pneumoniae*, *P. aeruginosa*, *S. aureus*, and *Enterobacter* spp., ten, nine, eight, and three AMPs
240   showed MIC ≤25 μM, respectively (**Fig. 4b**). While some AMPs showed distinct activity profiles
241   against individual strains, six AMPs classified as "broad-spectrum" antimicrobials that showed
242   favorable therapeutic window i.e., antimicrobial activity at relevantly low hemolysis and
243   cytotoxicity (AMP #3, #5, #13, #15, #16, and #27). Notably, these AMPs were also active against
244   the notorious biothreat agents *Yersinia pestis* and *Bacillus anthracis* (**Fig. 4b**).

245   <u>No resistance was developed against de novo AMPs</u>

246   Next, we tested the emergence of resistance against the six broad-band AMPs identified in this
247   study (AMP #3, #5, #13, #15, #16, and #27). To that end, we performed *E. coli* serial passaging
248   experiments with the peptides. As control, we added imipenem, a broad-spectrum antibiotic that
249   is generally considered a "last resort" against multidrug-resistant pathogens[30]. During 21 days of
250   serial passage, we did not observe a significant increase in MIC of our AMPs, while the imipenem
251   MIC gradually increased up to 8-fold and exceeded the susceptibility breakpoint of clinically
252   relevant resistance defined by EUCAST (**Fig. 5a** and **Supplementary Fig. 7**). The fact that we
253   did not observe resistance is in line with the fact that our AMPs do not act on a specific cellular
254   target, but rather globally (i.e., at the membrane, **Fig. 3c**), which makes them less likely to cause
255   resistance development. This mode of action was further confirmed by propidium iodide staining
256   and microscopy (**Fig. 5c-d**), which clearly indicated membrane disruption upon AMP treatment.

257   Finally, we also investigated the effect of AMP treatment on the release of outer membrane
258   vesicles (OMVs) in *E. coli*. OMVs are naturally released by Gram-negative bacteria and can be a
259   reaction against surface attacking agents neutralizing the effect of membrane targeting
260   antibiotics[31–33]. Notably, none of the six broad-band AMPs did significantly increase the release
261   of OMV compared to untreated E. coli, except for AMP #16 that showed a 4-fold increase in OMV
262   production (**Fig. 5b**). Altogether, these experiments suggested that our six broad-band AMPs are
263   able to escape resistance development and self-defense reactions of bacteria.

Fig. 5: Resistance and mode of action test of AMPs on *E. coli* and the mode of action assays. a, 21 days MIC measurement of daily culture passage each day from the cells grown in half-MIC concentrations of AMPs. The MIC values were normalized by the minimum of 21 daily averages for each AMP. Bars are the average of n = 3 independent experiments. See **Supplementary Fig. 7** for raw MIC data. **b,** Impact of different AMPs at the concentration of ¼ MIC on the number of *E.coli* outer membrane vesicles (OMVs) for untreated *E. coli* (control), (AMP) solvent, and six potent AMPs. Bars are the average of n=8 and n = 3 independent experiments for the control and the rest, respectively. Statistics: one-way ANOVA; ***p <0.001, compared to control. **c,** Plate reader measured fluorescence of propidium iodide (DNA staining after membrane disruption) of *E. coli* cells untreated (control) or treated with AMPs. Bars are the average of n = 3 independent experiments. *P* values are for paired Student's t-test of each AMP with untreated *E. coli* cells. **d,** Phase-contrast microscopy of *E. coli* cells untreated or treated with six potent AMPs and Cecropin B (a membrane disruptive AMP) and stained with propidium iodide (red). Source data for **a**-**c** and six images (n = 3 independent experiments) of which one was used in **d** are provided as a **Source Data** file.

Discussion

In this work, we describe the design and validation of 30 de novo AMPs, of which six show broad-band activity. Recent work has impressively demonstrated the power of deep learning methods in discovering and/or developing novel AMPs[6,19]. However, these efforts still suffer from the limited

9

281 number of peptides that can be synthesized and tested, and the relatively long time it takes from
282 design to validation. Here we used CFPS to dramatically advance the design-build-test cycle in
283 AMP development. Compared to recent approaches[19], our success rate is in the same range (6%
284 versus 10%). However, the total number of functional AMPs discovered is an order of magnitude
285 higher (30 AMPs versus 2) and at massively increased rate (24 hours versus 28 days) and
286 reasonably low cost (<10$ for production of one AMP for screening on two strains in parallel).

287 While being unique and diverse, our de novo AMPs share common properties with known AMPs.
288 They are predicted to be mostly α-helical peptides rich in cationic and hydrophobic amino acids
289 and preferably act on negatively-charged IMs, showing that our pipeline was able to design new-
290 to-nature sequences that follow the general building principles of AMPs. While we put our focus
291 onto AMPs with broad-band activity, we note that our pipeline is in principle also suited to develop
292 and/or iteratively optimize peptides with more specific activities or desired functionalities (e.g.,
293 narrow spectrum, increased stability, etc.). A unique advantage of CFPS is the possibility to
294 incorporate modified and unnatural amino acids more rapidly and conveniently than by cell-based
295 methods.

296 Overall, our work provides a proof-of-principle, how CFPS can be used to leverage the full
297 potential of machine learning approaches in the future. Especially in the light of ever-decreasing
298 DNA synthesis costs, our combined approach of deep learning and CFPS provides a time-, cost-
299 , and labor-effective approach for peptide production and screening. Thus, our work holds the
300 potential to explore the design-function space of AMPs at increased rate and depth. This will
301 hopefully lead to the increased discovery and development of peptide-based drug candidates in
302 the future.

303 **Methods**

304 <u>Pretraining and training datasets</u>

305 Pretraining data. To gather a large corpus of protein sequences representing a general protein
306 grammar, we downloaded all protein sequences shorter than 49 amino acids from UniProt[34] (as
307 of July 2021). After removing duplicate sequences and entries with unknown amino acid
308 characters, 3,104,952 unique sequences remained of which a random subset of half of them was
309 used for pretraining.

310 AMP data. For experimentally validated AMP sequences, we used the Giant Repository of AMP
311 Activities (GRAMPA)[35] which has combined sequence and activity data from several public AMP
312 databases; APD[36], DADP[37], DBAASP[38], DRAMP[39], and YADAMP[40]. This database consists of
313 6,760 unique sequences and 51,345 total MIC measurements, spanning several bacterial and
314 nonbacterial target species. We filtered the MIC measurements to the ten most abundant bacterial
315 species (*E. coli, P. aeruginosa, Salmonella typhimurium, K. pneumoniae, A. baumannii, S.*
316 *aureus, B. subtilis, S. epidermidis, Micrococcus luteus and E. faecalis*) and omitted all peptides
317 containing any chemical modification other than C-terminal amidation, for feasible in-laboratory
318 expression. After removing duplicate sequences and entries longer than 48 amino acids, 5,319
319 unique AMP sequences were left.

320 Non-AMP data. We searched the UniProtKB[34,41] for proteins labeled as "NOT antimicrobial,
321 antibiotic, antiviral or antifungal" (downloaded as of July 2021), removed entries containing
322 ambiguous amino acids, and kept only unique sequences shorter than 49 amino acids. This
323 resulted in a dataset containing 10,612 unique non-AMP sequences.

324 <u>Generator variational autoencoder (VAE)</u>

325   We used VAEs because they have previously been used for de novo AMP design[16,17,19,42]. The
326   generative VAE consists of an encoder, a latent vector, and a decoder. The encoder feeds the
327   input data (one-hot encoded amino acid letter of peptides) into a latent vector that is an information
328   bottleneck, and the decoder aims to reconstruct the input data from the latent vector. Training the
329   VAE and minimizing the difference between the input data and the reconstructed data acts
330   twofold; the encoder learns to map the training dataset into a lower-dimensional space and the
331   decoder learns to generate samples similar to the training data from any vector in the latent space.
332   Thus, each peptide in the training dataset lands on a point in the multi-dimensional latent space.
333   Picking vectors from the empty regions in this space and feeding them into the decoder yield
334   peptide sequences that share the same grammar but are novel and not seen in the training
335   dataset. After pretraining and training (transfer learning), we generated new AMPs by sampling
336   from the latent space using different strategies in particular by exploring the neighborhood of a
337   control functional AMP, gradient descent, or random sampling (**Supplementary Tables 1, 2**).

338   Generator VAE models. We adapted the neural network architecture of the CNN-RNN hybrid VAE
339   model from Hawkins-Hooker et al.[23]. The encoder consists of 5 consecutive one-dimensional
340   convolution layers fed into a dense layer of size 50, which is the latent vector. The decoder is
341   made from 4 deconvolution layers that samples the latent vector and a GRU layer of 512 cells
342   outputs a sequence in the same dimension as the input. The model loss is the weighted sum of
343   a reconstruction loss and the Kullback-Leibner (KL) loss. The total loss function can be
344   dynamically changed in the training process for KL-term annealing[43]; as standard practice when
345   working with discrete data such as language. Our two final models were trained without and with
346   the KL-term annealing (VAE_v1 and VAE_v2, respectively). The models were compiled using the
347   Adam optimizer.

348   Pretraining and training. Our training dataset of ~5,000 AMP sequences is not sufficient for
349   learning what makes a protein sequence distinct from a random string of amino acid characters
350   and what makes a protein sequence an AMP. In such cases, pretraining with a much bigger
351   generic dataset is needed to enhance the model performance. We pretrained the generator
352   models with ~1.5 million protein sequences from UniProtKB for 600 epochs. We then trained the
353   models on the AMP data for 400 epochs. Model training metrics are provided in **Supplementary**
354   **Table 1**.

355   <u>Regressor convolutional and recurrent neural networks</u>

356   Regressor neural networks. We adapted a regressor model previously reported[35]. First, a CNN
357   regressor was built with two consecutive one-dimensional convolutional layers, a max pooling
358   layer, a flattening layer, a dropout layer (0.5), and three dense layers. Second, we built a simple
359   RNN regressor with an LSTM layer and two dense layers. The models were compiled with mean
360   squared error loss and the Adam optimizer.

361   Gram-specific regressors. Due to structural differences in the cell membrane, we assumed there
362   are differences between Gram-positive and Gram-negative bacteria in their response to AMPs.
363   To capture these differences, we trained Gram-specific regressor models. We trained the Gram-
364   negative model on 4619 unique AMP sequences and corresponding MIC measurements on *E.*
365   *coli;* and the Gram-positive model on 4089 AMP sequences and corresponding MIC
366   measurements on *B. subtilis*. This approach improved the accuracy and efficiency of the
367   regressor models (**Supplementary Tables 1, 2**).

368   Training. We trained these models on pairs of data containing AMP sequences and their
369   corresponding MIC measurements (in log 10). Based on a previous study[35], the nonAMP
370   sequences were labeled to have a log MIC of 4. We interpreted the predicted MIC value as

371 follows; below 3.5 as AMP, between 3.5-3.9 as potential AMP and above 3.9 as non-AMP. The
372 regressor models were trained by the pooled AMP and non-AMP data for 200 epochs. Model
373 training metrics are provided in **Supplementary Table 1**.

374 <u>Sampling and prioritizing of AMPs</u>

375 In each round of peptide synthesis, we selected thousands of random points from the VAEs latent
376 space and from each reconstructed a peptide sequence. We omitted non-viable sequences and
377 kept AMPs with 36-48 amino acids. This left us with viable peptide sequences (**Supplementary**
378 **Table 2**). We then fed these sequences into the regressors and prioritized them based on the
379 MIC prediction. 50-150 AMPs were chosen for wet-lab experiments in five rounds making a list of
380 over 500 AMPs tested in this work (**Supplementary Table 2**).

381 <u>Molecular dynamics (MD) simulations</u>

382 All MD simulations were performed using Gromacs 2020.3[44] and the CHARMM36m forcefield[45]
383 using an integration time step of 2 fs (partly 1 fs in the membrane equilibration; see
384 **Supplementary Table 9**). Bonds including hydrogens were constrained using the LINCS
385 algorithm[46]. Electrostatic interactions were computed using the Particle-Mesh Ewald (PME)
386 algorithm[47] with a real-space cut-off for pairs further apart than 1.2 nm. Lennard-Jones
387 interactions were smoothly switched to zero between 1 and 1.2 nm using the force-switch
388 algorithm.

389 The protein, the membrane and the solvent (water and ions) were individually coupled to thermal
390 baths set to 37 °C (310.15 K) using the v-rescale algorithm[48] with a time-constant of 1 ps. During
391 the equilibration runs pressure coupling was handled by the Berendsen barostat[49], which was
392 switched to the Parrinello-Rahman barostat[50] for all production runs. The barostat time-constant
393 and compressibility factor were consistently set to 5 ps and 4.5×10^{-5} bar^{-1}, respectively. Pressure
394 coupling was applied semiisotropically (with the x and y dimensions coupled together) for systems
395 with membrane and isotropically otherwise. The reference pressure was set to 1 bar. To
396 counteract an energetic penalty for peptide insertion into the membrane due to finite size effects,
397 we simulated each membrane system with three different lateral membrane tensions (0, 9, 17.1
398 bar). Therefore, the diagonal elements of the pressure tensor (P_{XX}, P_{YY}, P_{ZZ}) were set to

399 $$P_{XX} = P_{YY} = P - \frac{\Delta P}{3} \qquad\qquad P_{ZZ} = P + 2\frac{\Delta P}{3}$$

400 with P as the reference pressure (1 bar) and ΔP as the desired lateral membrane tension. To
401 ensure uncorrelated runs for the different tensions, the starting velocities of the atoms were
402 randomly initialized according to the Maxwell-Boltzmann distribution. The MD simulations of the
403 AMPs on membranes were performed for 1 μs each.

404 Visual analysis and renders used the VMD[51], PyMOL[52] and ChimeraX[53] software.

405 Membrane setup. Following earlier work and results from lipidomics experiments, we modeled
406 membranes resembling the outer leaflet of the human plasma membrane (PM)[54,55] and the *E. coli*
407 inner membrane (IM)[56]. The detailed compositions are summarized in **Supplementary Tables**
408 **10, 11**.

409 Using the CHARMM-GUI membrane builder[57,58], we generated 12.5 x 12.5 nm^2 patches of these
410 model membranes, energy minimized them with a steepest descent algorithm until the largest
411 force acting on any atom was below 1000 kJ mol^{-1} and subsequently equilibrated them following
412 the CHARMM-GUI equilibration scheme (summarized in **Supplementary Table 9**).

12

AMP system setup. Structures of the selected 30 AMPs and of Cecropin B were predicted using AlphaFold[28] while the short BP100 was modeled as a coil with initial angles φ=-60° and ψ=30°, using the Molefacture Protein Builder plugin for VMD[51]. Since several of our AMPs have more than one cysteine (AMPs #2, #4, #6, #8, #10, #11, #12, #14, #18, #19), we next performed 1-2 µs long MD simulations of them surrounded by only water and ions (150 mM NaCl plus counter ions for overall neutralization) and with no disulfide bonds imposed. Based on the frequency and distance with which two cysteines in the respective structure interacted with each other in these simulations, we assigned a disulfide bond for AMP #2 (Cys11, Cys47) and disulfide bonds connecting the predicted β-sheets of AMP #8 (Cys29, Cys35) and AMP #14 (Cys10, Cys32). Due to the particularly high abundance of cysteine in AMP #2 and the potential structural bias from imposing one specific disulfide bond, we additionally simulated it without any imposed disulfide bonds.

These structures were then orientated so that their first principal axis was orthogonal to the z-axis and placed in proximity, but not yet bound to the equilibrated PM and IM. The box was subsequently solvated with TIP3P water[59] and NaCl ions were added to a concentration of 150 mM, ensuring overall neutrality by adding additional counter ions to the systems. The systems were energy minimized in the same way as the pure membrane systems and were then equilibrated for 5 ns. During minimization and equilibration, we applied position restraints on the peptide heavy atom with a force constant of 1000 kJ mol^{-1} nm^{-2}.

Cell-free production and activity test of AMPs

DNA fragments encoding AMPs were designed with T7 promoter (GAATTTAATACGACTCACTATAGGGAGA), RBS (TCTAGAGATTAAAGAGGAGAATACTAG) sequences upstream of the AMP coding region, and a T7 terminator (TACTCGAACCCCTAGCCCGCTCTTATCGGGCGGCTAGGGGTTTTTTGT) downstream. 500 DNA fragments were purchased from Twist Bioscience. A final concentration of 10 nM of each fragment was used for cell-free transcription and translation of AMPs using PUREfrex®2.0 kits (GeneFrontier #PF201-0.25-5-EX). In 384-well plates (BRAND, #781687), 30 µL volume of the cell-free reaction was made for each AMP and incubated for 4 hours at 37 °C. The cell-free mix was directly used for the activity test on *E. coli* and *B. subtilis* or for the SDS-PAGEl of the functional AMPs.

E. coli MG1655 and *B. subtilis* PY79 were used as representatives of Gram-negative and Gram-positive bacteria. From LB agar plates into LB medium, three overnight cultures for each strain were made from three different colonies and grown while shaking at 37 °C. The next day, each was subcultured in LB (1:1000) and grown while shaking at 37 °C to OD ≈ 1. Cells were diluted in LB to 10^4 cfu mL^{-1}, and 16 µL of diluted cells were added to wells of a 384-well plate (Greiner Bio-One, #781185) in which 4 µL of the cell-free reaction mix (with AMPs produced) had been added beforehand. Cultures were mixed and the plate was sealed by a gas-permeable film (Carl Roth, #T093.1). OD$_{600}$ was measured every 10 min in a plate reader (Tecan Infinite® 200 PRO) shaking at 37 °C for 20 hours. Growth curves were analyzed for AMPs impairing bacterial growth.

SDS-PAGE of AMPs produced in the cell-free system

SDS-PAGE was used to detect produced functional AMPs in the cell-free reaction as previously described[22]. Cell-free reactions were boiled for 3 min in 2x Tricine buffer (Bio-Rad, #1610739), loaded in 16.5% Mini-PROTEAN polyacrylamide Tris/Tricine gels (Bio-Rad, #4563065), and run for 5 hours at 200 mA in the running buffer (10 mM Tris, 10 mM Tricine, and 0.01% SDS). Gels were then fixed for one hour in 12% trichloroacetic acid and one hour in 40% EtOH, 10% acetic

458   acid, followed by overnight staining in QC Colloidal Coomassie (Bio-Rad, #161-0803), 24 hours
459   of de-staining in water, and imaged using Intas GelStick Touch Imager.

460   <u>Measurement of minimum inhibitory concentration (MIC) and resistance test</u>

461   Strains used for MIC measurements are *E. coli* MG1655, *B. subtilis* PY79, *E. faecium* (isolate
462   from the gut of the cow), *S. aureus* DSM 11729, *K. pneumoniae* DSM 30104, *A. baumannii* (isolate
463   from the human abdominal wall), *P. aeruginosa* DSM 1117, *Enterobacter* spp., *Y. pestis* EV76,
464   *B. anthracis* Sterne, and *S. pneumoniae* D39. A commonly used standard protocol for
465   determination of MIC for antimicrobials was used to measure the MIC of the AMPs taking into
466   account all suggestions for cationic AMPs in the protocol[29]. Chemically synthesized peptides were
467   dissolved in BSA (0.2% w/v) acetic acid (0.01% v/v) solution to have 10x of the highest
468   concentration to be tested. In 96-well PCR plates (Axygen, #PCR-96-SG-C) two-fold serial
469   dilutions were made from columns 1-10 in each row specified to each peptide and BSA (0.2%
470   w/v) acetic acid (0.01% v/v) solution was pipetted into columns 11 and 12. Triplicates of 7.5 µL of
471   each dilution were pipetted into polypropylene 96-well plates (Corning, #3359). Triplicates of
472   bacterial overnight cultures in Mueller-Hinton broth 2 (MHB 2, Sigma-Aldrich, #90922) were
473   prepared from three different colonies the day before and grown shaking at 37 °C, subcultured in
474   the morning by diluting 1000x in MHB 2 grown shaking at 37 °C to OD ≈ 1. Bacterial cultures were
475   then diluted with MHB 2 to 10^5 cfu mL^{-1} and 67.5 µL of each triplicate was added on top of peptides
476   in columns 1-11 of 96-well plates. MHB 2 was added to column 12. The plates were sealed by
477   adhesive films (VWR, #391-1262) and incubated at 37 °C for 20 hours. MIC values were reported
478   as the highest concentration of each AMP in which no visible growth was observed. For *S.*
479   *pneumoniae* THY medium was used instead of MHB 2.

480   The same procedure was used for the resistance test except for cultures that from the second
481   day on, each of the triplicate cells grown in the highest AMP concentration (half MIC) was diluted
482   10,000x in MHB 2 and added to newly prepared peptides dilutions.

483   <u>Measurement of cytotoxicity (CC50)</u>

484   Cytotoxicity assay was performed on HCT116 human colon cells (ATCC, #CCL-247™) using
485   CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega, #G3580) which is a
486   colorimetric method based on MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-
487   (4-sulfophenyl)-2H-tetrazolium) for determining cell viability. The MTS tetrazolium compound
488   (Owen's reagent) is reduced into a colored formazan product by NADPH or NADH produced by
489   dehydrogenase enzymes in metabolically active cells.

490   On the first day, when cells reached the density of 50-80% of the covered surface, gently washed
491   twice using 10 mL of DPBS (Gibco, #14190367). 1 mL trypsin (Capricorn, #TRY-1B) was added,
492   5 min incubated at 37 °C and 9 mL medium was added including DMEM high glucose (Capricorn,
493   #DMEM-HPA), 10 % v/v fetal bovine serum (FBS, Capricorn, #FBS-11A), and antibiotic mix for
494   cell culture (Capricorn, #PS-B). The culture was transferred into a 15 mL falcon, and spun down
495   at 1000 rpm for 3 min. The supernatant was sucked out and 10 mL of fresh media was added and
496   transferred. Cells were diluted by the medium to have 5000 cells in 36 µL. 36 µL of the cell culture
497   was pipetted into wells of a 384-well plate (Greiner, #781185). The last well received only media.
498   Cells were incubated at 37 °C, 5% CO_2, for 24 hours. On the second day, peptides were prepared
499   in two-fold serial dilutions starting from a final concentration of 250 µM. Columns 11 and 12
500   received only water. 4 µL of each peptide dilution was added to wells of the cell culture plate
501   prepared on the first day and the plate was put at 37 °C, 5% CO_2, for 24 hours. On the third day,
502   8 µL of CellTiter 96® AQueous One Solution and 10 µL SDS 10% were added to each well and
503   after 90 min incubation at 37 °C, absorbance at 490 nm was measured and corrected by the value

14

504 of wells with only medium. CC50 values, the concentration of each AMP killing 50% of cells, were
505 calculated using Graphpad Prism 9.

Measurement of hemolytic activity (HC50)

507 Human blood from a healthy donor was washed three times with PBS and afterwards
508 resuspended in 2V PBS. AMPs with an initial concentration of 250 µM were titrated in 96-well
509 polypropylene plates (V-bottom, Greiner Bio-One GmbH). 5 µL AMP dilutions were overlaid with
510 45 µL of washed and concentrated human erythrocytes, and the plates were sealed and incubated
511 at 37°C for 1 h. 40 µL of supernatant were transferred after final centrifugation at 1000 g for 5 min
512 at room temperature to ELISA plates and absorbance was measured at 405 nm. Triton X 100
513 treated erythrocytes served as positive control. HC50 values, the concentration of each AMP
514 lysing 50% of RBCs, were calculated using Graphpad Prism.

Mode of action assay and microscopy using propidium iodide (PI)

516 Plate reader assay. Three colonies were picked to culture *E. coli* MG1655 cells in LB at 37 °C to
517 the exponential phase. Cells were harvested by centrifugation at 4000 g and washed three times
518 in 10 mM PBS (pH=7.0), and adjusted to OD_{600}=1 with 10 mM PBS. 10 µL of the cells in PBS
519 were mixed with 10 µL of AMPs to a final concentration of 4×MIC and incubated at 37 °C for 1
520 hour. 20 µM final concentration of PI[60] was added to each of the AMP-treated and untreated
521 samples and incubated at 37 °C for 30 min in the dark. Fluorescence was recorded at an excitation
522 of 535 nm and emission of 615 nm was measured using a Tecan Infinite® 200 PRO plate reader.

523 Sample preparation for microscopy. *E. coli* was grown in LB at 37 °C to the exponential phase
524 and diluted to 10^8 cfu mL^{-1} in fresh LB. 50 µL of diluted cells were pipetted in a 1.5 mL tube and
525 50 µL of AMPs was added at 4×MIC final concentration together with 20 µM PI. The mixture was
526 incubated at 37 °C for 1 h while shaking. 1 µL of samples was loaded onto agarose pads (2%
527 agarose in PBS) and imaged using a Zeiss AxioPlan 2 upright widefield microscope equipped
528 with a 100x NeoFluor phase contrast objective and a FluoArc HBO lamp. Fluorescence of PI was
529 recorded using TxRed HC Filter set (AHF, F36-504).

530 Image processing. The phase contrast and fluorescence images were overlaid. The dynamic
531 range of the fluorescence channel was set to a minimum of 125AU to remove background
532 fluorescence, while the contrast of the phase contrast channel was manually adjusted for better
533 visualization. For comparison of the different phenotypes, 232 pixels x 232 pixels regions of
534 interest were cropped. Raw image files are provided in this study.

Outer membrane vesicle (OMV) release of *E.coli* after AMP treatment

536 *Escherichia coli* MG1655 was grown on MacConkey agar (Carl Roth, Karlsruhe, Germany) plates
537 overnight. For overnight culture, a single colony was used for inoculation of 2 mL LB media at 37
538 °C, 160 rpm (MaxQ 6000, Thermo Fisher Scientific, Karlsruhe, Germany). Bacterial culture was
539 transferred to 10 mL fresh LB media and incubated (1 h, 37 °C, 160 rpm). The required amount
540 of bacteria was treated with ¼ MIC of the AMPs #3, #5, #10, #13, #15, #16, and #27, left untreated
541 for control or was treated with acetic acid and BSA as solvent control (90 min, 37°C, 160 rpm).
542 Samples were centrifuged (4,500 g, 15 min, 4 °C; Multifuge X3R, Thermo Fisher Scientific), the
543 supernatant was sterile filtered (0.22 µm) and afterwards concentrated by the factor of 20 using
544 100 kDa molecular weight cut-off filters (Merck KGaA, Darmstadt, Germany). To determine the
545 number of released vesicles, samples were measured by nano-flow cytometry (nFCM) using a
546 NanoAnalyzer (NanoFCM Co., Ltd, Nottingham, UK) as previously described[61].

15

547   <u>Chemical peptide synthesis of AMPs</u>

548   Materials. All commercially available reagents were purchased from the following companies, and
549   used without further purification: thioanisol, 1,2-ethandithiol (EDT) from Sigma Aldrich (USA);
550   piperidine, 2 mL polypropylene reactors with plunger and frit pore size 25 µm (#7926.1) from Carl
551   Roth (Germany); 2,6-lutidine, palladium acetate, phenylsilane, trifluoracetic acid (TFA) from Acros
552   (USA); microscale columns from Intavis (#35.091) (Germany); Fmoc-Gly-OH, Fmoc-L-Asn(Trt)-
553   OH, Fmoc-L-Asp(tBu)-OH, Fmoc-L-His(Trt)-OH, Fmoc-L-Ile-OH, Fmoc-L-Met-OH, Fmoc-L-Phe-
554   OH, Fmoc-L-Pro-OH*H20, Fmoc-L-Thr(tBu)-OH, Fmoc-L-Tyr(tBu)-OH, Fmoc-L-Val-OH, peptide
555   grade dimethylformamide (DMF), TentaGel S RAM resin from Iris Biotech (Germany); Fmoc-L-
556   Ala-OH-OH, Fmoc-L-Arg(Pbf)-OH, Fmoc-L-Cys(Trt)-OH, Fmoc-L-Gln(Trt)-OH, Fmoc-L-
557   Glu(OtBu)-OH, Fmoc-L-Leu-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Ser(tBu)-OH, Fmoc-L-Trp(Boc)-
558   OH, N,N′-diisopropylcarbodiimid (DIC), and Oxyma from Carbolution (Germany); acetic anhydride
559   (Ac$_2$O) from Grüssing GmbH (Germany); HPLC grade acetonitrile from Merck (Germany);
560   Ultrapure water of type 1 was obtained with a MicroPure Water Purification System from TKA
561   (Germany).

562   Solid-phase peptide synthesis. All peptides were synthesized via the Fmoc-solid phase strategy.
563   The synthesis was carried out by an automated peptide synthesizer using INTAVIS ResPep SLi
564   instrument for a 5 µmol scale. Higher amounts of peptides were achieved by running multiple 5
565   µmol syntheses in parallel. For all peptides, the TentaGel S RAM (0.22 mmol/g) resin was used.

566   Automated solid-phase peptide synthesis (INTAVIS ResPep SLi):

567   -   The conditions, reagents, and corresponding volumes of this synthesis protocol
568   correspond to a 5 µmol scale synthesis. No mixing was performed during incubation or
569   reaction time. In the coupling step, the temperature was set to 40°C.

570   -   Swelling: The appropriate amount of resin (23 mg) was swelling in 200 µL DMF for 30 min.

571   -   Deprotection of temporal Fmoc protecting groups: Piperidine (150 µL, 20% in DMF) was
572   added to the resin and incubated for 5 min. This step was repeated, and the resin was
573   filtered off, and washed with DMF (1 x 300 µL, 3 x 225 µL).

574   -   Coupling of amino acids: In a mixing vial the machine automatically added 53 µL of the
575   corresponding Fmoc-amino acid (4 eq, 0.5 M), 15 µL Oxyma (4 eq, 2 M), 13 µL DIC (4 eq,
576   2 M) each in DMF and 29 µL NMP. The resulting solution was activated by waiting for 1
577   min before addition to the resin. This suspension was incubated for 15 min. Next, the resin
578   was filtered off, and the coupling was repeated. No washing was performed after the
579   coupling.

580   -   Capping: 150 µL of a lutidine/Ac$_2$O/DMF 6:5:89 solution was added to the resin and
581   incubated for 8 min. The resin was filtered off and washed with DMF (3 x 225 µL).

582   -   After the last coupling, the resin was washed with DMF (1 x 300 µL, 3x 225 µL), ethanol
583   (4 x 150 µL) and CH$_2$Cl$_2$ (5 x 150 µL). The resin was finally dried under continuous air flow
584   for 5 min.

585   Final Cleavage, Purification, and Characterization:

586   -   Cleavage and deprotection of the amino acid side chains: All peptides have been cleaved
587   from a dry resin previously washed with 5x DMF and 10x CH$_2$Cl$_2$ following the last step of

16

588   the synthesis protocol. Depending on the total number of Cys, Met, or Trp, one of the
589   cleavage cocktails in **Supplementary Table 12** was utilized. Cleavage cocktail A was
590   used as the initial test cleavage after complete synthesis[62]. In the cases where oxidation
591   was observed after cleavage, cocktail B was applied[63,64]. For 5 μmol of resin, 2 mL of
592   cleavage cocktail was prepared. The total volume was increased by 1.5-fold for peptides
593   containing more than eight arginine residues. The dry resin was loaded into 2 mL reactors
594   with a plunger and the frit was treated with the corresponding mixture and shaken for 2.5
595   h and filtered off. The resin was washed with 1 mL of TFA and the filtrates were combined.
596   The TFA content of the filtrate was reduced via a gentle nitrogen flow. Next, ice-cold diethyl
597   ether (DEE) (1.00 mL of DEE for 100 μL cocktail) was added to precipitate the final
598   peptide. The precipitated peptide was centrifuged (8000 rpm, 4°C, 5 min), the supernatant
599   was discarded and the pellet washed once more; i.e. redissolved, and precipitated with
600   cold DEE. Afterward, the peptide pellet was dissolved in ultrapure water/MeCN (70:30)
601   with 0.1% of TFA (more MeCN was added when insoluble, not exceeding 1:1) to be
602   purified.

603   -   Purification: The peptides were purified by reverse-phase (RP)-HPLC using a preparative
604   Agilent 1260 Infinity II Series HPLC-system (Agilent Technologies) with column 1
605   (**Supplementary Table 13**). An isocratic regime during the first five minutes for column
606   equilibration, followed by the respectively stated linear gradient in 25 min (gradient is
607   specified at the respective peptide). The detection was carried out by measuring
608   absorption at the wavelengths: 220 nm and 260 nm. Ultrapure water (A) and MeCN (B)
609   were employed as eluents with an addition of 0.1% of TFA in both solvents.

610   -   Characterization: The freeze-dried products were identified via analytical HPLC-MS on an
611   Agilent 1260 Infinity II Series HPLC-system (Agilent Technologies) using column 2
612   (**Supplementary Table 13**). The detection was carried out by measuring absorption at the
613   wavelengths: 220 nm and 260 nm. Ultrapure water (A) with an addition of 0.05% of TFA
614   and MeCN (B) addition of 0.03% of TFA were employed as eluents. HR-ESI-MS was
615   performed for identification on an LTQ-FT Ultra device (Thermo Fischer Scientific). HPLC
616   chromatogram of purified peptides are provided in **Supplementary Figs. 8-34**.

617

Code availability

619 All deep learning models were built, trained, and tested using Keras 1.0 with TensorFlow 2.0
620 backend using Python 3.9 in the Google Colab pro environment. The deep learning codes and
621 models developed in this study as well as training data, 500 tested AMPs and 30 functional AMPs
622 can be found at https://github.com/amirpandi/Deep_cAMP.

Acknowledgments

Contributions

642 T.J.E and A.P. conceived the study and wrote the manuscript with contributions from other
643 authors, as indicated in the following. A.Z. developed deep learning codes and simulations and
644 wrote the corresponding method section and performed BLAST analysis. A.P. and Y.F. performed
645 cell-free protein synthesis and the initial bioactivity tests. M.K. performed the RBS calculator and
646 kinfold calculations. S.L.S. performed and, together with G.H analyzed molecular dynamics
647 simulation and wrote the corresponding methods and results section. A.P., D.A., B.K., P.B. and
648 H.v.B. performed and analyzed MIC, hemolysis, and cytotoxicity tests. V.T.T. performed chemical
649 peptide synthesis and analysis via automated solid-phase and wrote the corresponding methods
650 section. A.P., E.B, and C.S. performed and analyzed the mode of action assays. M.W. performed
651 and analyzed the outer membrane vesicle experiments and wrote the corresponding results and
652 methods sections. C.P. established the OMV quantification and performed the OMV quantification
653 in the Core Facility for Extracellular Vesicles. E.P.v.S., H.B.B., H.v.B., W.B., A.L.J., F.A., B.S,
654 G.H., O.V., and T.J.E supervised the work. All authors provided input on the manuscript and
655 confirmed the final draft.

Competing interests

657 The authors declare no competing interests.

References

659 1.   10 global health issues to track in 2021. https://www.who.int/news-room/spotlight/10-global-

660   health-issues-to-track-in-2021.
661   2.   Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance
662   in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
663   3.   Miethke, M. *et al.* Towards the sustainable discovery and development of new antibiotics.
664   *Nat Rev Chem* **5**, 726–749 (2021).
665   4.   Lazzaro, B. P., Zasloff, M. & Rolff, J. Antimicrobial peptides: Application informed by
666   evolution. *Science* **368**, (2020).
667   5.   Jaumaux, F., P Gómez de Cadiñanos, L. & Gabant, P. In the Age of Synthetic Biology, Will
668   Antimicrobial Peptides be the Next Generation of Antibiotics? *Antibiotics (Basel)* **9**, (2020).
669   6.   Ma, Y. *et al.* Identification of antimicrobial peptides from the human gut microbiome using
670   deep learning. *Nat. Biotechnol.* **40**, 921–931 (2022).
671   7.   Fjell, C. D., Hiss, J. A., Hancock, R. E. W. & Schneider, G. Designing antimicrobial
672   peptides: form follows function. *Nat. Rev. Drug Discov.* **11**, 37–51 (2011).
673   8.   Dijksteel, G. S., Ulrich, M. M. W., Middelkoop, E. & Boekema, B. K. H. L. Review: Lessons
674   Learned From Clinical Trials Using Antimicrobial Peptides (AMPs). *Front. Microbiol.* **12**,
675   616979 (2021).
676   9.   Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design.
677   *Nature* **537**, 320–327 (2016).
678   10.   Pan, X. & Kortemme, T. Recent advances in de novo protein design: Principles, methods,
679   and applications. *J. Biol. Chem.* **296**, 100558 (2021).
680   11.   Dauparas, J. *et al.* Robust deep learning-based protein sequence design using
681   ProteinMPNN. *Science* eadd2187 (2022).
682   12.   Ferruz, N., Schmidt, S. & Höcker, B. ProtGPT2 is a deep unsupervised language model for
683   protein design. *Nat. Commun.* **13**, 4348 (2022).
684   13.   Nijkamp, E., Ruffolo, J., Weinstein, E. N., Naik, N. & Madani, A. ProGen2: Exploring the
685   boundaries of protein language models. (2022) doi:10.48550/ARXIV.2206.13517.
686   14.   Strokach, A. & Kim, P. M. Deep generative modeling for protein design. *Curr. Opin. Struct.*
687   *Biol.* **72**, 226–236 (2022).
688   15.   Li, F. *et al.* Deep learning-based kcat prediction enables improved enzyme-constrained
689   model reconstruction. *Nat. Catal.* **5**, 662–672 (2022).
690   16.   Dean, S. N. & Walper, S. A. Variational Autoencoder for Generation of Antimicrobial
691   Peptides. *ACS Omega* **5**, 20746–20754 (2020).
692   17.   Dean, S. N., Alvarez, J. A. E., Zabetakis, D., Walper, S. A. & Malanoski, A. P. PepVAE:
693   Variational Autoencoder Framework for Antimicrobial Peptide Generation and Activity
694   Prediction. *Front. Microbiol.* **12**, 725727 (2021).
695   18.   Wang, C., Garlick, S. & Zloh, M. Deep Learning for Novel Antimicrobial Peptide Design.
696   *Biomolecules* **11**, (2021).
697   19.   Das, P. *et al.* Accelerated antimicrobial discovery via deep generative models and
698   molecular dynamics simulations. *Nature Biomedical Engineering* **5**, 613–623 (2021).
699   20.   Garenne, D. *et al.* Cell-free gene expression. *Nat Rev Methods Primers* **1**, (2021).
700   21.   Silverman, A. D., Karim, A. S. & Jewett, M. C. Cell-free gene expression: an expanded
701   repertoire of applications. *Nat. Rev. Genet.* **21**, 151–170 (2020).
702   22.   Pardee, K. *et al.* Portable, On-Demand Biomolecular Manufacturing. *Cell* **167**, 248–259.e12
703   (2016).
704   23.   Hawkins-Hooker, A. *et al.* Generating functional protein variants with variational
705   autoencoders. *PLoS Comput. Biol.* **17**, e1008736 (2021).
706   24.   Badosa, E. *et al.* A library of linear undecapeptides with bactericidal activity against
707   phytopathogenic bacteria. *Peptides* **28**, 2276–2285 (2007).
708   25.   Jarvis, D. L., Summers, M. D., Garcia, A., Jr & Bohlmeyer, D. A. Influence of different signal
709   peptides and prosequences on expression and secretion of human tissue plasminogen
710   activator in the baculovirus system. *J. Biol. Chem.* **268**, 16754–16762 (1993).

711 26. Espah Borujeni, A. & Salis, H. M. Translation initiation is controlled by RNA folding kinetics
712   via a ribosome drafting mechanism. *J. Am. Chem. Soc.* **138**, 7016–7023 (2016).
713 27. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding
714   sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
715 28. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
716   583–589 (2021).
717 29. Konur, O. *Handbook of Algal Science, Technology and Medicine*. (Academic Press, 2020).
718 30. Rodloff, A. C., Goldstein, E. J. C. & Torres, A. Two decades of imipenem therapy. *J.*
719   *Antimicrob. Chemother.* **58**, 916–929 (2006).
720 31. Kulp, A. & Kuehn, M. J. Biological functions and biogenesis of secreted bacterial outer
721   membrane vesicles. *Annu. Rev. Microbiol.* **64**, 163–184 (2010).
722 32. Bauwens, A., Kunsmann, L., Karch, H., Mellmann, A. & Bielaszewska, M. Antibiotic-
723   Mediated Modulations of Outer Membrane Vesicles in Enterohemorrhagic Escherichia coli
724   O104:H4 and O157:H7. *Antimicrob. Agents Chemother.* **61**, (2017).
725 33. Manning, A. J. & Kuehn, M. J. Contribution of bacterial outer membrane vesicles to innate
726   bacterial defense. *BMC Microbiol.* **11**, 258 (2011).
727 34. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids*
728   *Res.* **49**, D480–D489 (2021).
729 35. Witten, J. & Witten, Z. Deep learning regression model for antimicrobial peptide design.
730   *bioRxiv* 692681 (2019) doi:10.1101/692681.
731 36. Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for
732   research and education. *Nucleic Acids Res.* **44**, D1087–93 (2016).
733 37. Novković, M., Simunić, J., Bojović, V., Tossi, A. & Juretić, D. DADP: the database of anuran
734   defense peptides. *Bioinformatics* **28**, 1406–1407 (2012).
735 38. Pirtskhalava, M. *et al.* DBAASP v.2: an enhanced database of structure and
736   antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res.* **44**,
737   6503 (2016).
738 39. Fan, L. *et al.* DRAMP: a comprehensive data repository of antimicrobial peptides. *Sci. Rep.*
739   **6**, 1–7 (2016).
740 40. Piotto, S. P., Sessa, L., Concilio, S. & Iannelli, P. YADAMP: yet another database of
741   antimicrobial peptides. *Int. J. Antimicrob. Agents* **39**, 346–351 (2012).
742 41. UniProt. https://www.uniprot.org/help/uniprotkb.
743 42. Ghorbani, M., Prasad, S., Brooks, B. R. & Klauda, J. B. Deep attention based variational
744   autoencoder for antimicrobial peptide discovery. Preprint at
745   https://doi.org/10.1101/2022.07.08.499340.
746 43. Bowman, S. R. *et al.* Generating sentences from a continuous space. (2015)
747   doi:10.48550/ARXIV.1511.06349.
748 44. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-
749   level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).
750 45. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically
751   disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
752 46. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint
753   solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
754 47. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593
755   (1995).
756 48. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J.*
757   *Chem. Phys.* **126**, 014101 (2007).
758 49. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R.
759   Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684 (1984).
760 50. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular
761   dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).

762   51. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.*
763   **14**, 33–8, 27–8 (1996).
764   52. Schrödinger, L., & DeLano, W. PyMOL, Available at: http://www.pymol.org/pymol, (2020).
765   53. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators,
766   and developers. *Protein Sci.* **30**, 70–82 (2021).
767   54. Schaefer, S. L., Jung, H. & Hummer, G. Binding of SARS-CoV-2 Fusion Peptide to Host
768   Endosome and Plasma Membrane. *J. Phys. Chem. B* **125**, 7732–7741 (2021).
769   55. Lorent, J. H. *et al.* Plasma membranes are asymmetric in lipid unsaturation, packing and
770   protein shape. *Nat. Chem. Biol.* **16**, 644–652 (2020).
771   56. Berglund, N. A. *et al.* Interaction of the antimicrobial peptide polymyxin B1 with both
772   membranes of E. coli: a molecular dynamics study. *PLoS Comput. Biol.* **11**, e1004180
773   (2015).
774   57. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface
775   for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
776   58. Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes
777   for molecular dynamics simulations. *PLoS One* **2**, e880 (2007).
778   59. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L.
779   Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926
780   (1983).
781   60. Blaskovich, M. A. T. *et al.* Protein-inspired antibiotics active against vancomycin- and
782   daptomycin-resistant bacteria. *Nat. Commun.* **9**, 22 (2018).
783   61. Ponath, V. *et al.* Secreted ligands of the NK cell receptor NKp30: B7-H6 is in contrast to
784   BAG6 only marginally released via extracellular vesicles. *Int. J. Mol. Sci.* **22**, 2189 (2021).
785   62. King, D. S., Fields, C. G. & Fields, G. B. A cleavage method which minimizes side reactions
786   following Fmoc solid phase peptide synthesis. *Int. J. Pept. Protein Res.* **36**, 255–266
787   (1990).
788   63. Fujii, N., Otaka, A., Sugiyama, N., Hatano, M. & Yajima, H. Studies on Peptides. CLV.
789   Evaluation of Trimethylsilyl Bromide as a Hard-Acid Deprotecting Reagent in Peptide
790   Synthesis. *Chem. Pharm. Bull.* **35**, 3880–3883 (1987).
791   64. Pearson, D. A., Blanchette, M., Baker, M. L. & Guindon, C. A. Trialkylsilanes as scavengers
792   for the trifluoroacetic acid deblocking of protecting groups in peptide synthesis. *Tetrahedron*
793   *Lett.* **30**, 2739–2742 (1989).