Designing β-hairpin peptide macrocycles for antibiotic potential

Justin R. Randall[1], Cory D. DuPai[1,2], T. Jeffrey Cole[1,2], Gillian Davidson[1], Kyra E. Groover[1], Claus O. Wilke[2], and Bryan W. Davies*[,1]

Affiliations:
[1]Department of Molecular Biosciences, University of Texas at Austin; Austin, TX, USA
[2]Department of Integrative Biology, University of Texas at Austin; Austin, Texas, USA
*Corresponding author. Email: bwdavies@austin.utexas.edu

Abstract: Peptide macrocycles are a rapidly emerging new class of therapeutic, yet the design of their structure and activity remains challenging. This is especially true for those with β-hairpin structure due to weak folding properties and a propensity for aggregation. Here we use proteomic analysis and common antimicrobial features to design a large peptide library with macrocyclic β-hairpin structure. Using an activity-driven high-throughput screen we identify dozens of peptides killing bacteria through selective membrane disruption and analyze their biochemical features via machine learning. Active peptides contain a unique constrained structure and are highly enriched for cationic charge with arginine in their turn region. Our results provide a synthetic strategy for structured macrocyclic peptide design and discovery, while also elucidating characteristics important for β-hairpin antimicrobial peptide activity.

Brief Summary: We design, screen, and computationally analyze a synthetic macrocyclic β-hairpin peptide library for antibiotic potential.

Introduction: The stability and broad functionality of macrocyclic peptides makes them a promising area for drug development; however, there are few well-characterized strategies for their identification. Current *de novo* design remains challenging, especially for those with β-hairpin structure (*1–4*). They often lack sufficient inter-strand interactions to form stable conformations and their β-strands promote association which can lead to aggregation in solution (*3–5*).

Antibiotics provide an excellent example of macrocyclic peptide drug value and new discovery strategies are urgently needed (*6*). Macrocyclic β-hairpin antimicrobial peptides (β-AMPs) have recently gained popularity, with two such antibiotics having proceeded into clinical evaluation (*7–9*). These β-AMPs act primarily through bacterial membrane permeabilization, allowing them to overcome most mechanisms of bacterial drug resistance and also access and inhibit essential processes within the Gram-negative cell envelope (*10*, *11*); however, known examples are exceedingly rare. This makes it is difficult to understand how β-AMP sequence determines structure and function (*12–14*). Established strategies for *de novo* β-hairpin peptide design also limit the use of amino acids important for antibacterial activity and include residues that increase mammalian cell toxicity, which is detrimental for therapeutic development (*1*, *12*).

Here we describe the design, screening, and analysis of a synthetic macrocyclic β-hairpin peptide library using new strategies and technologies. Our results expand our understanding of β-AMP sequence-activity relationships and provides a new route for their design and discovery.

Results:

Design of a synthetic macrocyclic β-hairpin peptide library

Our design scheme leveraged our recently completed systematic analysis of over 49,000 β-hairpin motifs in the Protein Data Bank. This analysis identified position specific amino acid preferences in the strand and turn regions (*15*). Using this information, we designed two ribosomally encoded twenty amino acid cyclic β-hairpin peptide libraries (**Fig. 1AB** and **fig. S1**).

These libraries were intentionally designed to facilitate the development of stable antiparallel β-strands with amphipathic faces generated via the periodic alternation of aliphatic and charged/polar residues. We identified and applied a preference for glycine, asparagine, and aspartic acid in or near the turn regions (*15–17*) and excluded proline in contrast to canonical β-turn design (*18–21*). We also excluded aromatic residues because they promote mammalian toxicity despite their positive effects on β-hairpin structure (*22*). Cysteine was encoded at the amino and carboxyl termini to potentiate cyclization via a disulfide bridge and prevent fraying. Finally, we allowed select polar positions and the loop region to encode for positive residues to promote solubility and to mimic a trend we observed in natural β-AMPs. Using codon variation, we created and combined two synthetic, macrocyclic β-hairpin (SynCH) libraries based on our design scheme, one beginning with a charged residue and the other with an aliphatic residue. This allows different amphipathic characters to occupy different faces relative to the loop region. In total, this pooled library encompassed 196,608 unique peptide sequences possessing a variety of physiochemical properties. (**Fig. 1AB** and **fig. S1**).

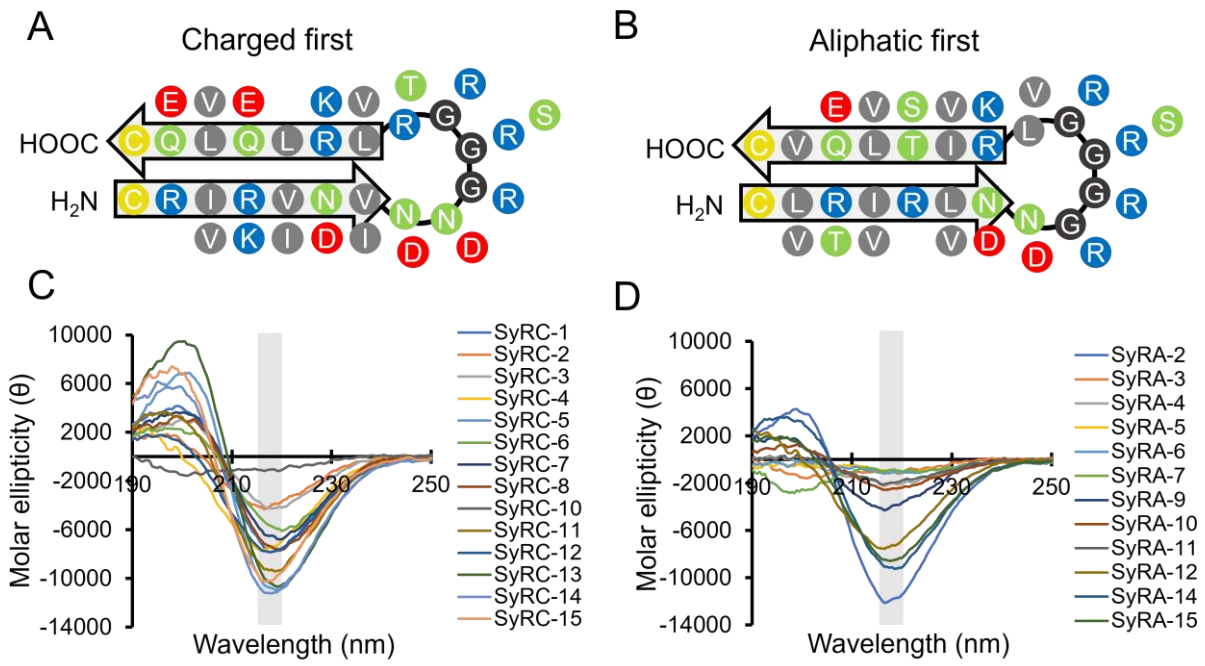


Figure 1. SynCH peptides show β-hairpin secondary structure. *Diagrams showing the designed structure and residue position of the charged first (**A**) and aliphatic first (**B**) SynCH libraries. Residues are color coded by side chain (yellow = cysteine, green = polar, gray = aliphatic, blue = positive, red = negative). (**C**, **D**) Circular dichroism spectra of randomly selected charged first (SyRC) and aliphatic first (SyRA) SynCH peptides. A single minimum between 215 and 220 nm (gray box) is characteristic of β-hairpin secondary structure. Each spectrum is the mean of three technical replicates with background subtracted.*

SynCH peptides spontaneously fold into macrocyclic β-hairpins

We examined the structure of thirty SynCH peptides at random from the charged first (SyRC) and aliphatic first (SyRA) libraries spanning a range of charge (-0.12 to 4.87) and grand average of hydropathicity (GRAVY) score (-0.72 to 0.41) (**Table 1**). One SyRC peptide and three SyRA peptides could not be synthesized. We first used circular dichroism spectroscopy (CD) to determine the secondary structure of these chemically synthesized peptides (**Fig. 1CD, Table 1**). Remarkably, 84.6% of all peptides had a CD spectrum with a single ellipticity minimum between 215 and 220 nm, indicative of antiparallel β-sheet secondary structure (*23*).

Table 1. Properties of randomly selected peptides from the SynCH library

Name	Sequence	Charge	GRAVY	β-hairpin CD	%S-S	MIC (μg/ml)
SyRC-1	CRVRINVDDRRGRLKLQVEC	2.88	-0.70	Yes	90.8 ± 2.7	>256
SyRC-2	CRVKINIDNGSRRVRLEVEC	1.88	-0.49	Yes	96.7 ± 0.7	>256
SyRC-3	CRIRVNVNDGSGRVRLQVQC	2.87	-0.33	Yes	21.9 ± 7.8	>256
SyRC-4	CRIRINVNDRGGTVRLEVEC	0.88	-0.31	Yes	100 ± 0.0	>256
SyRC-5	CRVRIDINNGRRRLKLQLQC	4.87	-0.70	Yes	89.7 ± 1.4	>256
SyRC-6	CRIKVNINNRRRTVKLELEC	3.88	-0.67	Yes	80.0 ± 1.7	>256
SyRC-7	CRIKVDIDDRRGRVRLQLEC	1.88	-0.68	Yes	68.7 ± 3.6	>256
SyRC-8	CRIKVNIDDRSRRLRLELEC	1.88	-0.72	Yes	69.2 ± 5.9	>256
SyRC-9	CRVRINVNDGSGTVRLQVQC*	-0.12	-0.31	-	-	-
SyRC-10	CRVRINVNDGSGTVRLQVQC	1.87	-0.15	No	95.7 ± 2.2	>256
SyRC-11	CRVRINVDNRGGTLKLQVQC	2.87	-0.32	Yes	89.5 ± 1.0	>256
SyRC-12	CRIRINVDNGGGTVRLQVEC	0.88	-0.12	Yes	nd	>256
SyRC-13	CRIRIDIDDRSRTLKLQLQC	1.88	-0.52	Yes	8.2 ± 3.6	>256
SyRC-14	CRVRINVNNRSGRLKLQVEC	3.88	-0.52	Yes	100 ± 0.0	>256
SyRC-15	CRVKVNVDDRGRRLKLELQC	2.88	-0.70	Yes	79.1 ± 3.0	>256
SyRA-1	CLRVRLNNGGRRVKVSVQVC*	1.88	-0.04	-	-	-
SyRA-2	CLRIRVNNRRGGVRITVQVC	4.87	0.00	Yes	100 ± 0.0	>256
SyRA-3	CLRVRLNNGGRRVKVSVQVC	4.87	-0.02	No	77.9 ± 2.0	256
SyRA-4	CLTIRVDNRRSRVKITLEVC	2.88	-0.02	Yes	85.2 ± 2.2	>256
SyRA-5	CLTIRLDNGGRGVRISLQVC	1.87	0.34	No	nd	>256
SyRA-6	CVRVRLNDGGSGVRISVQVC	1.87	0.36	Yes	78.4 ± 6.3	>256
SyRA-7	CVTIRLNDRRRGVKISVEVC	2.88	0.01	No	66.3 ± 8.2	>256
SyRA-8	CLRIRVNNGGSRVRISLEVC*	0.88	0.41	-	-	-
SyRA-9	CVTIRLDDRRRGLRVTVQVC	2.87	-0.04	Yes	94.3 ± 1.2	>256
SyRA-10	CLRIRVNNGGSRVRISLEVC	2.88	0.16	Yes	66.2 ± 11.9	>256
SyRA-11	CVTIRVDNRGSRVRVTLEVC	1.88	0.15	Yes	0.0 ± 0.0	>256
SyRA-12	CVRVRVDDRRGGLKISLQVC	2.87	0.00	Yes	100 ± 0.0	>256
SyRA-13	CVTVRLDNGRSGVRISVQVC*	1.87	0.34	-	-	-
SyRA-14	CVTVRVNNRGGRLKITLQVC	3.87	0.18	Yes	100 ± 0.0	>256
SyRA-15	CVTVRLNDRRSRLRITLEVC	2.88	-0.08	Yes	89.3 ± 2.3	>256

** was not synthesized, GRAVY: grand average of hydropathicity, β-hairpin CD was determined by having an ellipticity minimum between 215 and 220 nm, %S-S: percentage of peptide with a disulfide bond represented by the mean ± one standard deviation of three technical replicates, MIC: minimum inhibitory concentration, nd: not determined*

This is highly uncommon for peptides in aqueous solution. Most require a target interaction to form a β-hairpin or other secondary structure (*24*, *25*).

Next, we performed high resolution LC/MS on each peptide to determine whether an intramolecular disulfide bond was present (**Table 1, fig. S2**). All but three of the randomly selected SynCH peptides examined had a disulfide bond present in the majority of their molecular population. While sequences from the aliphatic first library were less likely to show β-sheet secondary structure (**Fig. 1CD**), they were slightly more likely to have a majority of their molecular population cyclized (**Table 1**). This data together suggests that ~72% of our SynCH peptide library forms a stable β-hairpin secondary structure and are cyclized through an intramolecular disulfide bond in solution.

Identification of SynCH peptides with antibiotic potential using SLAY

Our randomly selected SynCH peptides were not inherently antibacterial (**Table 1**), so we decided to use a high-throughput genetic platform developed in our lab called surface localized antimicrobial display (SLAY) (*26*, *27*) to screen for antibacterial activity. SLAY functions through the inducible display of a plasmid encoded peptide library on the Gram-negative bacterial cell surface. Next-generation sequencing is then used to generate a $\log_2$-fold change in peptide sequence read counts between induced and uninduced bacterial cultures (**Fig. 2A**).

We performed SLAY on our SynCH library at low (15 μM IPTG) and high (100 μM IPTG) induction concentrations to mimic treating bacteria with two different concentrations of peptide. The low induction (15 μM IPTG) condition identified 1,165 such peptides, referred to as "Hits" (**Fig. 2B, left**). The high induction (100 μM IPTG) condition identified significantly more "Hits" (5,697) (**Fig 2B, right**). Next, we plotted the $\log_2$-fold change of the hits from both induction concentrations against one another (**Fig. 2C**). 1,026 peptides, or 2.49% of the total number of peptides screened (~41,000), were identified as a "Hit" under both conditions. This group of 1,026 peptides are referred to as the SynCH SLAY Active group, or "SySA" going forward (**Fig. 2C, green**). The most promising 88 hits are also highlighted (**Fig. 2C**, cyan).

To see if there was enrichment of SynCH peptides with a certain charge or hydrophobicity in the SySA group, we plotted the distribution of charge versus GRAVY score for an equal number of randomly selected SynCH peptides and compared them to the SySA group (**Fig. 2D**). SySA distribution was highly enriched at charges greater than 5.5 relative to the SynCH library, suggesting that cationic charge is important for their antibacterial potential. The top 88 SySA peptides were further enriched toward cationic charge (**Fig. 2E, cyan**). Interestingly, 18/36 SySA peptides with verified *in vitro* activity (detailed below) were within a single distribution grouping (charge ~5.87, GRAVY -0.2 to -0.3) (**Fig. 2E, red, arrow**).

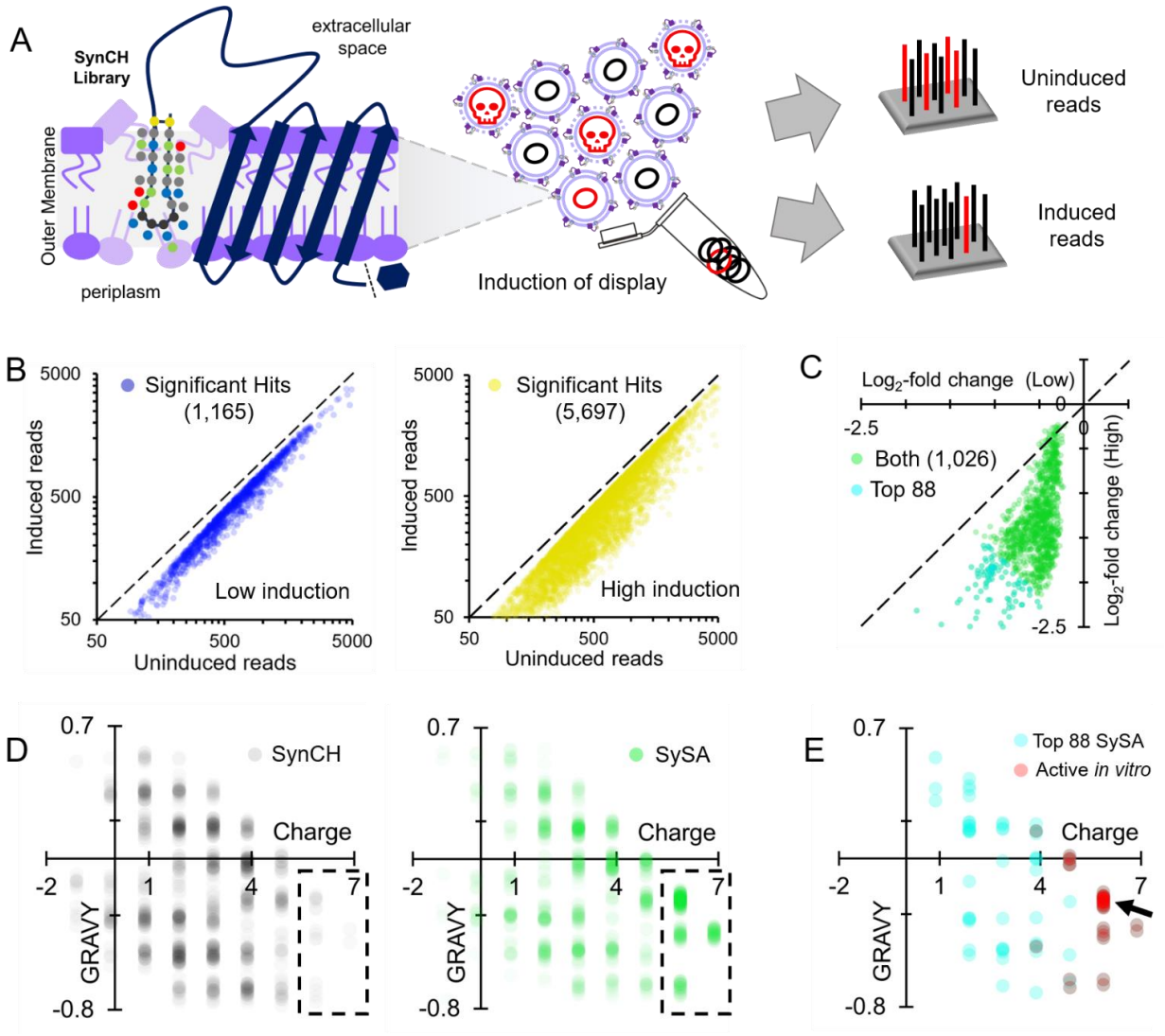


Figure 2. SLAY identifies SynCH peptides with antibacterial potential. *(**A**) Workflow of a surface localized antimicrobial display (SLAY) screen. (**B**) Density scatter plot for SynCH peptides with a significant, negative log_2-fold change ($p < 0.05$) from SLAY induced at low (left) and high (right) concentration. (**C**) Density scatter plot of each SynCH peptide with a significant, negative log_2-fold change in reads ($p < 0.05$) under both induction conditions (SySA). (**D**) Density scatter plots for a randomized subset of the SynCH library (left) and SySA (right). (**E**) Same plot but with the top 88 SySA peptides based on average log_2-fold change. Those which were verified to have in vitro activity are in red, 18/36 are from a single grouping (arrow).*

SySA peptides selectively disrupt bacterial membranes

We chose to chemically synthesize the top 88 SySA peptides as determined by lowest average log_2-fold change for further biochemical characterization (**Fig. 2CE cyan**) along with two natural β-AMPs for comparison: Protegrin-1 and Thanatin. We began by determining the minimum inhibitory concentration (MIC) for each SySA peptide against our screening strain, *E. coli* W3110 in both Mueller-Hinton broth (MH) and the tissue culture media RPMI 1640 (RPMI). MH broth is a standard for antibacterial testing, while RPMI medium better represents salt and buffer conditions found in the body. 44.4% (36/81) of the SySA peptides examined were active *in vitro* in both media with MICs ranging from 4 to 256 μg/ml (**Table 2**). Interestingly, the

vast majority of active peptides were from the aliphatic first library. All active SySA peptides had a charge greater than 3.8 with 78% having a charge greater than 5.8. Interestingly, 18 of the 36 peptides had a charge of 5.87 and a GRAVY score between -0.2 and -0.3 (**Fig. 2E,** arrow; **Table 2,** bolded). This grouping contained many of the most potent SySA peptides, so we chose five (MICs 4-32 μg/ml) to investigate further (**Fig. 3A**). Overall, this set of peptides was less potent than Protegrin-1 and Thanatin, whose MICs ranged from 0.5-4 μg/ml (**Fig. 3A**). This is not surprising considering our SySA peptides have not been optimized for antibacterial activity, while natural β-AMPs have evolved their activity over millennia.

We were curious whether SySA peptides functioned via membrane disruption like Protegrin-1 or by a periplasmic mode of action like Thanatin (*11*). To investigate, we measured the amount of propidium iodide (PI) uptake caused by our select SySA peptides and compared them to Protegrin-1 and Thanatin (**Fig. 3B**). PI fluorescence occurs upon DNA binding indicating that both the outer and inner membrane have been permeabilized. We treated *E. coli* W3110 cells with two-fold serial dilutions of peptide and observed PI uptake by relative fluorescence units (RFUs). All five or our SySA peptides showed high levels of PI uptake, similar to Protegrin-1, though requiring higher concentrations (**Fig. 3B**). As expected, Thanatin showed no PI uptake indicating it is not inner membrane disruptive.

Most natural β-AMPs like Protegrin-1 are also highly toxic to mammalian cells because they lack membrane specificity. For this reason, we examined each SySA peptide for its toxicity using a hemolysis assay measured as a percentage of red blood cells lysed by 300 μg/ml peptide compared to full lysis with a detergent (**Table 2**). Remarkably, our select SySA showed negligible hemolysis (< 3.73%) with some, like SySA-49, containing virtually no hemolysis (0.29 ± 0.49%) (**table S2, Fig. 3C**). For comparison, Protegrin-1 lysed 41.8 ± 5.03% of red blood cells and Thanatin, which does not function via membrane disruption, lysed only 0.3 ± 0.87% of red blood cells (**table S2, Fig. 3C**). This result suggests that active SySA peptides function through membrane disruption but demonstrate bacterial membrane selectively. This is an uncommon characteristic for members of the β-AMP class.

Lastly, we questioned if the naïve peptide sequences discovered through our screen could be optimized to improve their activity. Results from our biochemical characterization suggest high charge is important. Previous data from optimization of another SLAY identified β-AMP found that shorter peptide length and additional disulfide bonds increased potency (*25*). For these reasons we generated a 27-peptide optimization library around our most potent peptide (SySA-5) by shortening its length, increasing its charge, and adding the potential for a second intramolecular disulfide bond while also maintaining alternating residue side chain properties in the antiparallel β-sheets (**table S1**).

We had this library commercially synthesized and tested its ability to inhibit the growth of *Acinetobacter baumannii* AB5075, a clinically isolated Carbapenem Resistant Acinetobacter (CRA) pathogen (*28*). MICs were performed in MH as well as 100% fetal bovine serum (FBS), which is likely more representative of the *in vivo* environment. We found that 23 or our 27 variants increased potency as much as 8-fold in MH. Additionally, eleven gained the ability to kill *A. baumanii* AB5075 in FBS (**table S1**). This was surprising as degradation in serum is a common limitation of peptide use. Curious whether increased potency correlated with increased toxicity we also measured each variants hemolysis at 128 μg/ml and reported it as a fold change relative to SySA-5 (**table S1**). Most had less than a two-fold increase relative to SySA-5. The most potent of the variants (SySA 5.17) had the same MIC against *A. baumanii* AB5075 in FBS as Protegrin-1 (32 μg/ml), but was greater than ten-fold less hemolytic at 128 μg/ml. This data

suggests the SynCH peptides can be easily optimized to have greater therapeutic potential than naturally occurring β-AMPs.

Table 2. Properties of SynCH peptides verified as antibacterial.

		W3110 MIC (μg/ml)			
Name	**Sequence**	**MH**	**RPMI**	**Hemolysis (%)**	**β-hairpin CD**
SySA-1	CRIKVDVNNRRRRVRLQVQC	128	64	0.57 ± 0.23	Yes
SySA-4	CRIKVNVNNRGRRLRLEVQC	256	128	0.81 ± 0.56	Yes
SySA-5	CLRVRLNNRRGRVKVSLQVC	16	4	3.73 ± 1.28	Yes
SySA-7	CRVRININNGRRRVKLQLQC	256	32	0.50 ± 0.26	Yes
SySA-8	CRIKINVNDGSRRLKLQVQC	256	256	0.23 ± 0.19	Yes
SySA-9	CVRVRLNNRRRGLKISLQVC	128	8	1.13 ± 0.80	Yes
SySA-11	CVTVRVNNRRRRVRISVQVC	64	16	0.66 ± 0.12	Yes
SySA-16	CVRVRLNNRRRRLKVSLEVC	32	32	3.24 ± 1.16	Yes
SySA-17	CLRVRLNNRRRGLKVSVQVC	64	32	0.84 ± 0.18	Yes
SySA-25	CLRVRLNNRRRGVKVSVQVC	64	8	0.77 ± 0.53	Yes
SySA-26	CVRVRLNNGRRRVKVSLQVC	16	16	1.05 ± 0.31	Yes
SySA-27	CVRVRLNNRGGGLRITLQVC	256	64	3.40 ± 0.87	Yes
SySA-29	CVRVRLNNRRRGLRVSLQVC	16	8	1.22 ± 0.41	Yes
SySA-30	CLRIRLNNRRRGLKVSLQVC	64	8	0.61 ± 0.23	Yes
SySA-34	CRIKVNINNGRRRVKLQLQC	128	32	0.23 ± 0.07	Yes
SySA-37	CVRIRLNNRRRRVKVSLEVC	32	8	2.84 ± 0.49	Yes
SySA-38	CVRIRVNNRRSGVKVSLQVC	256	8	0.30 ± 0.41	Yes
SySA-40	CVRIRLNNRRRGLKVSLQVC	32	8	0.61 ± 0.36	Yes
SySA-45	CVRVRLNNRRRGLKVTLQVC	256	32	7.74 ± 1.14	No
SySA-49	CVRVRLNNRRRGVKISLQVC	32	8	0.29 ± 0.49	Yes
SySA-53	CVRIRLNNGGRRVKVTLQVC	32	32	0.71 ± 0.55	Yes
SySA-61	CLRVRLNNRRRGLKISLQVC	32	8	1.55 ± 0.74	Yes
SySA-62	CVRIRVNNRRRRVKVSVQVC	32	16	1.36 ± 0.74	Yes
SySA-63	CVTIRVNNRRSRVRISLQVC	256	64	0.79 ± 0.73	Yes
SySA-65	CVRVRLNNRRRRVKISVEVC	64	16	5.19 ± 1.35	Yes
SySA-66	CVRIRLNNRRRGVKVSVQVC	128	32	0.52 ± 0.53	Yes
SySA-71	CLRVRLNNGRRRLKVSLQVC	16	32	1.74 ± 0.57	Yes
SySA-72	CLRIRLNNRRRGLRISLQVC	32	8	3.17 ± 0.40	Yes
SySA-73	CLRVRLNNRRRGLRISLQVC	32	8	3.02 ± 0.82	Yes
SySA-78	CVRVRLNNRRRRLKISLEVC	32	16	6.59 ± 2.26	Yes
SySA-79	CVRVRVNNGRRRLKVSLQVC	64	32	1.19 ± 0.79	Yes
SySA-80	CVRIRLNNRRRRVKVTLQVC	16	16	2.93 ± 1.04	Yes
SySA-81	CLTVRLNNRRRGVKVSVQVC	64	32	0.76 ± 0.24	Yes
SySA-82	CVRIRLNNRRRGLKISLQVC	64	8	1.11 ± 0.61	Yes
SySA-83	CVRVRVNNRGGRLKISVQVC	128	16	0.59 ± 0.32	Yes
SySA-88	CVRVRLNNRRRGLKITLQVC	32	8	3.43 ± 0.72	Yes
Thanatin	GSKKPVPIIYCNRRTGKCQRM	2	4	0.30 ± 0.87	-
Protegrin-1	RGGRLCYCRRRFCVCVGR	0.5	0.5	41.84 ± 5.03	-

MIC: minimum inhibitory concentration, MH: Mueller-Hinton, RPMI: RPMI 11640; β-hairpin spectrum was determined by a circular dichroism ellipticity minimum between 215 and 220 nm. ***Bold:*** *part of the same charge-GRAVY score grouping.*

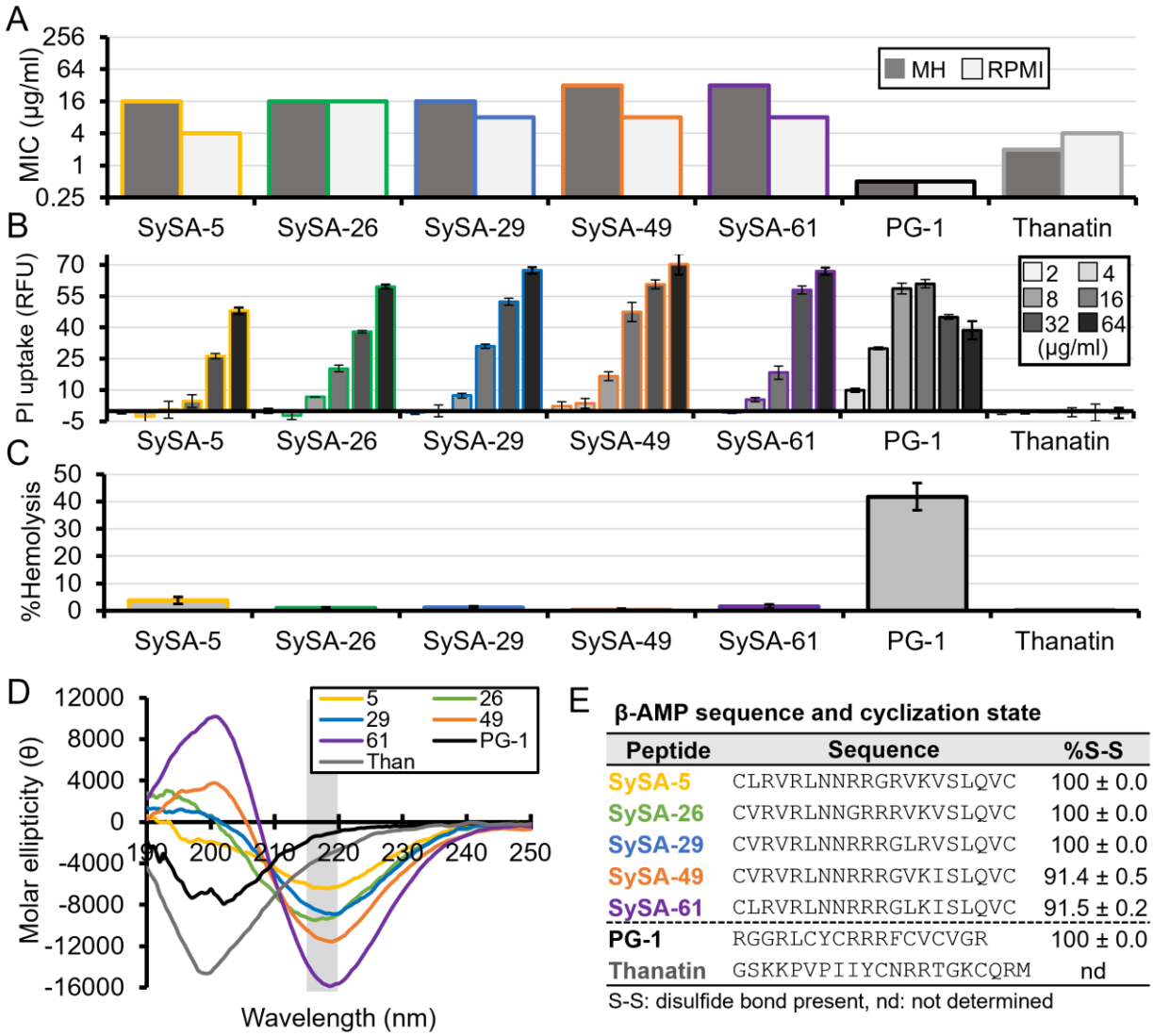


Figure 3. SySA peptide antibiotics demonstrate membrane specificity, have cyclic β-hairpin structure. *(**A**) Minimum inhibitory concentration of select SySA peptides, Protegrin-1 (PG-1) and Thanatin against E. coli W3110 in Mueller-Hinton (MH) and RPMI 1640 (RPMI) media. Each bar represents the median MIC (n=3) (**B**) The %hemolysis relative treatment with 1% Triton-X100. Bars represent the mean; error bars are one standard deviation (n=3) (**C**) Propidium iodide (PI) uptake of E. coli W3110 cells treated with peptide. Bars represent the median and error bars one standard deviation (n=3) (**D**) Circular dichroism spectra of peptides. A single minimum between 215 and 220 nm (gray box) is characteristic of β-hairpin structure. Each spectrum is the mean of three technical replicates with background subtracted. (**E**) Table of each peptides sequence and the percentage of the molecular population with an intramolecular disulfide bond. Error is one standard deviation of technical replicates (n=3).*

SySA peptides have a constrained cyclic β-hairpin structure

Natural antibacterial peptides, including β-AMPs, generally require membrane or membrane mimics like lipopolysaccharide (LPS) to form alpha helical or β-hairpin secondary structures (*24*, *25*). For example, our natural β-hairpin peptide controls have a single molar ellipticity minimum at 200 nm in phosphate buffer alone, consistent with a random coil secondary structure **(Fig. 3D)** (*23*); however, the CD spectra of all but one of the 36 active SySA peptides (SySA-45) have a single minimum between 215 and 220 nm consistent with a β-hairpin

secondary structure (**Table 2, Fig. 3D, and fig. S3**). This suggests that SySA peptides are more conformationally constrained than many natural β-AMPs. However, this alone does not dictate antibacterial activity because most of our inactive SySA peptides are also conformationally constrained (**fig. S3**). Conformational changes in β-AMP structure upon membrane interaction have been implicated in causing mammalian cytotoxicity (*29*). This could help explain the low hemolysis observed with our SynCH peptides.

Next, we looked to confirm that our most active SySA peptides were also cyclized via a disulfide bond by using LC/MS analysis. All five of the most potent SySA peptides also had a majority of their molecular population participating in a disulfide bond (**Fig. 3E**). Three of these peptides (SySA-5, 26, and 29) were 100% cyclized in solution. This data, combined with our CD analysis, strongly suggests that active SySA peptides have a macrocyclic β-hairpin structure which is conformationally constricted, a unique feature for members of the β-AMP class.

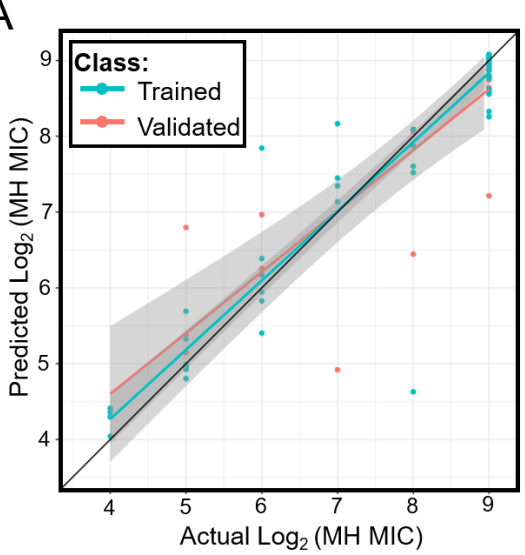
Machine learning identifies features important for antibacterial activity

The design of the SynCH library reliably produced cyclized β-hairpins with low/no hemolytic activity, but identifying sequences features responsible for antibacterial activity was more challenging. To overcome this hurdle, we trained a machine learning algorithm to predict peptide potency with 80% of our SySA biochemical data and used the remaining 20% for validation. For a detailed description of our model please see the supplemental text. The final trained model was able to explain 90% of the variation ($R^2 = 0.90$, $p < 2.2 \times 10^{-16}$) in the training dataset and 78% of the variation ($R^2 = 0.78$, $p = 2.5 \times 10^{-8}$) in the validation dataset. This implies the trained model accurately predicted the activities of SySA sequences set aside for validation (**Fig. 4A**). A predictive potency score was then generated using this algorithm for all 196,608 peptide sequences in the SynCH library.

To assess the accuracy of our predictive modeling we randomly selected 20,000 SynCH peptides excluded from our SLAY analysis and chose the ten with the lowest predicted potency score (SynCH Predicted Active, or "SyPA"). We had these ten peptides commercially synthesized and measured their MIC in MH against *E. coli* W3110. Remarkably, 8/10 SyPA peptides were active with MIC values ranging from 16 to 128 μg/ml (median MIC 64 μg/ml, **Fig. 4B**). This was a vast improvement to our top ten SLAY screening results where only 6/10 peptides were active and less potent overall (median MIC 256 μg/ml, **Fig. 4B**). One predicted peptide (SyPA-10) was as potent in MH media as any of the top 81 SySA peptides we examined. Furthermore, our modelling also identified active peptides with charges less than 5.8 (SyPA-2, 4, 5, and 9) and these peptides were more potent than peptides of similar charge identified using SLAY.

To explore the sequence features most important for predicting antibacterial potency, we plotted the residue frequency at each position for 100 random SynCH peptides and compared them to the top 100 peptides selected by SLAY (SySA) and predicted by machine learning (SyPA) (**Fig. 4C**) The SySA peptides had few obvious differences from SynCH other than a slight preference for the aliphatic first library. However, several sequence features were clearly enriched in the SyPA group (**Fig. 4C,** bottom). SyPA peptides were almost exclusively from the aliphatic first library possibly because of a greater potential for arginine in their loop region, which was also clearly enriched. Arginine was also highly enriched at position three. Additionally, leucine was highly preferred over valine at positions six and thirteen. Lastly, serine and glutamine were vastly favored over threonine and glutamate at positions sixteen and eighteen, respectively.

A



B

Comparison of antibacterial selection methods

SLAY	MIC (μg/ml)	Charge	Machine Learning	MIC (μg/ml)	Charge
SySA-1	128	5.87	SyPA-1	64	5.87
SySA-2	>256	1.88	SyPA-2	64	4.87
SySA-3	>256	1.88	SyPA-3	>256	3.87
SySA-4	256	4.88	SyPA-4	64	4.87
SySA-5	16	5.87	SyPA-5	128	4.87
SySA-6	>256	4.87	SyPA-6	32	6.87
SySA-7	256	5.87	SyPA-7	>256	3.87
SySA-8	256	3.87	SyPA-8	32	5.87
SySA-9	128	5.87	SyPA-9	64	3.87
SySA-10	>256	3.88	SyPA-10	16	6.87
Median	256	4.88	Median	64	4.88

MIC = *E. coli* W3110 minimum inhibitory concentration in Mueller-Hinton media; blue = positive charge; green = antibacterial potency

C

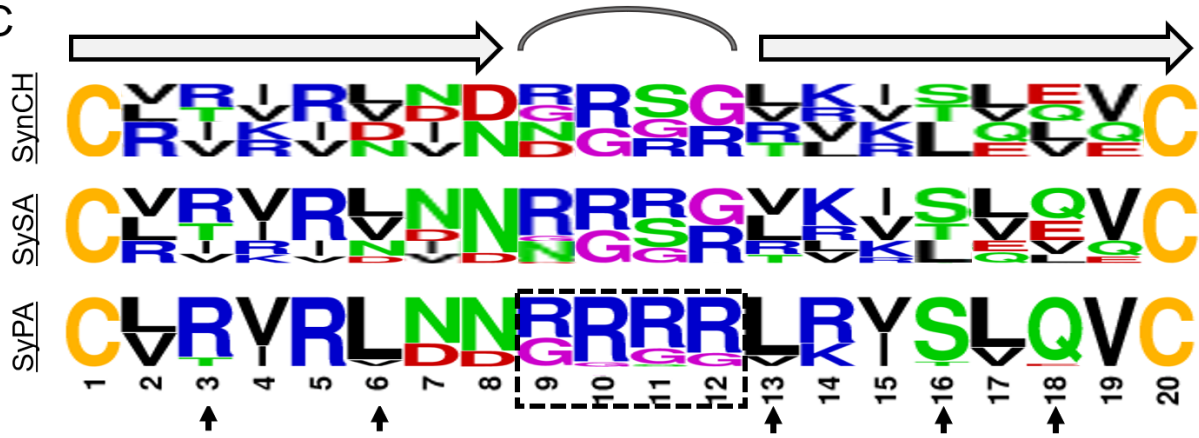


Figure 4. Machine leaning identifies features important for activity. *(**A**) Actual versus predicted log₂ MIC models produced via training and validation of the machine learning algorithm. (**B**) Table listing the MIC and charge of peptides identified with SLAY (left panel), or the machine learning (ML) (right panel). (**C**) residue frequencies at each position of 100 SynCH peptides (top), the top 100 SySA peptides (middle) and the top 100 SyPA peptides (bottom). Residues are color coded based on side chain properties: blue = basic, red = acidic, green = polar, black = aliphatic, purple = no side chain, yellow = sulfur containing.*

Discussion:

The goal of peptide design is to develop *de novo* sequences with and intended structure and activity. Here, we successfully designed and produced a large peptide library with predominantly macrocyclic β-hairpin structure and identified those with antibiotic activity. Our design lacked several features commonly used in β-hairpin peptide design including the use of proline in the loop region and aromatic residues like tryptophan in the β-sheet regions (*1*). In contrast, arginine was highly tolerated in both the loop region and β-sheets. We found this remarkable, as the loop sequence is considered critical for final β-hairpin stability (*30*), yet arginine is not often considered in loop design.

The β-AMPs identified from our library killed through membrane disruption but were not hemolytic. This membrane selectivity is highly unusual for β-AMPs (*13*) and warrants further study. It is possible that the constrained structure of SynCH peptides helps contribute to this selectivity. Conformational change has been shown to increase cytotoxicity in variants of

Protegrin-1 (*29*). However, the lack of aromatic residues in SynCH peptides could also contribute.

Our machine learning algorithm accurately predicted active SynCH peptides with greater accuracy than our cell-based approach and helped identify several sequence features associated with antibacterial potency. Most prominent was the enrichment of arginine, especially within the loop region which could also explain a preference for the aliphatic first library. This enrichment increases the overall charge of the peptide and is likely important for interaction with the negatively charged outer membrane. The reasons for the other enriched residues are less clear, but these preferences could be related to preferred inter-strand contacts between the antiparallel β-sheets.

We are excited about the future possibilities of pairing functional cell-based peptide screening technology with machine learning strategies, especially for antibiotic discovery. We believe as more antibacterial peptide data becomes available through synthetic screening technology, machine learning may be able to predict antibacterial activity *de novo*, bypassing the need for human design and cell-based screens entirely.

References:

1. P. Morales, M. A. Jiménez, Design and structural characterisation of monomeric water-soluble α-helix and β-hairpin peptides: State-of-the-art. *Arch Biochem Biophys*. **661**, 149–167 (2019).
2. M. Angeles Jiménez, Design of monomeric water-soluble β-hairpin and β-sheet peptides. *Methods in Molecular Biology*. **1216**, 15–52 (2014).
3. F. J. Blanco, G. Rivas, L. Serrano, A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat Struct Biol*. **1**, 584–590 (1994).
4. J. S. Richardson, D. C. Richardson, Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci U S A*. **99**, 2754–2759 (2002).
5. M. Angeles Jiménez, Design of monomeric water-soluble β-hairpin and β-sheet peptides. *Methods Mol Biol*. **1216**, 15–52 (2014).
6. J. R. Randall, B. W. Davies, Mining for novel antibiotics. *Curr Opin Microbiol*. **63**, 66–69 (2021).
7. A. Luther, C. Bisang, D. Obrecht, Advances in macrocyclic peptide-based antibiotics. *Bioorg Med Chem*. **26**, 2850–2858 (2018).
8. I. Martin-Loeches, G. E. Dale, A. Torres, Murepavadin: a new antibiotic class in the pipeline. *Expert Rev Anti Infect Ther*. **16**, 259–268 (2018).
9. P. v. Panteleev, S. v. Balandin, V. T. Ivanov, T. v. Ovchinnikova, A Therapeutic Potential of Animal β-hairpin Antimicrobial Peptides. *Curr Med Chem*. **24** (2017), doi:10.2174/0929867324666170424124416.
10. K. L. H. Lam, Y. Ishitsuka, Y. Cheng, K. Chien, A. J. Waring, R. I. Lehrer, K. Y. C. Lee, Mechanism of supported membrane disruption by antimicrobial peptide protegrin-1. *J Phys Chem B*. **110**, 21282–21286 (2006).
11. S. U. Vetterli, K. Zerbe, M. Müller, M. Urfer, M. Mondal, S.-Y. Wang, K. Moehle, O. Zerbe, A. Vitale, G. Pessi, L. Eberl, B. Wollscheid, J. A. Robinson, Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in Escherichia coli. *Sci Adv*. **4**, eaau2634 (2018).
12. P. v. Panteleev, I. A. Bolosov, S. v. Balandin, T. v. Ovchinnikova, Structure and Biological Functions of β-Hairpin Antimicrobial Peptides. *Acta Naturae*. **7**, 37–47 (2015).
13. I. A. Edwards, A. G. Elliott, A. M. Kavanagh, J. Zuegg, M. A. T. Blaskovich, M. A. Cooper, Contribution of Amphipathicity and Hydrophobicity to the Antimicrobial Activity and Cytotoxicity of β-Hairpin Peptides. *ACS Infect Dis*. **2**, 442–450 (2016).
14. I. A. Edwards, A. G. Elliott, A. M. Kavanagh, M. A. T. Blaskovich, M. A. Cooper, Structure-Activity and -Toxicity Relationships of the Antimicrobial Peptide Tachyplesin-1. *ACS Infect Dis*. **3**, 917–926 (2017).
15. C. D. DuPai, B. W. Davies, C. O. Wilke, A systematic analysis of the beta hairpin motif in the Protein Data Bank. *Protein Sci*. **30**, 613–623 (2021).

16. K. Fujiwara, H. Toda, M. Ikeguchi, Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct Biol*. **12** (2012), doi:10.1186/1472-6807-12-18.
17. J. M. Otaki, M. Tsutsumi, T. Gotoh, H. Yamamoto, Secondary structure characterization based on amino acid composition and availability in proteins. *J Chem Inf Model*. **50**, 690–700 (2010).
18. A. M. C. Marcelino, L. M. Gierasch, Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers*. **89**, 380–391 (2008).
19. S. R. Trevino, S. Schaefer, J. M. Scholtz, C. N. Pace, Increasing protein conformational stability by optimizing beta-turn sequence. *J Mol Biol*. **373**, 211–218 (2007).
20. S. N. Malkov, M. v. Živković, M. v. Beljanski, M. B. Hall, S. D. Zarić, A reexamination of the propensities of amino acids towards a particular secondary structure: classification of amino acids based on their chemical structure. *J Mol Model*. **14**, 769–775 (2008).
21. S. Costantini, G. Colonna, A. M. Facchiano, Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochem Biophys Res Commun*. **342**, 441–451 (2006).
22. C. M. Santiveri, M. A. Jiménez, Tryptophan residues: scarce in proteins but strong stabilizers of β-hairpin peptides. *Biopolymers*. **94**, 779–790 (2010).
23. N. Greenfield, G. D. Fasman, Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry*. **8**, 4108–4116 (1969).
24. C. Avitabile, L. D. D'Andrea, A. Romanelli, Circular Dichroism studies on the interactions of antimicrobial peptides with bacterial cells. *Sci Rep*. **4** (2014), doi:10.1038/SREP04293.
25. J. R. Randall, G. Davidson, R. M. Fleeman, S. A. Acosta, I. M. Riddington, T. J. Cole, C. D. DuPai, B. W. Davies, Synthetic antibacterial discovery of symbah-1, a macrocyclic β-hairpin peptide antibiotic. *iScience*. **25**, 103611 (2022).
26. A. T. Tucker, S. P. Leonard, C. D. DuBois, G. A. Knauf, A. L. Cunningham, C. O. Wilke, M. S. Trent, B. W. Davies, Discovery of Next-Generation Antimicrobials through Bacterial Self-Screening of Surface-Displayed Peptide Libraries. *Cell*. **172**, 618–628.e13 (2018).
27. J. R. Randall, C. D. DuPai, B. W. Davies, Discovery of Antimicrobial Peptide Macrocycles Through Bacterial Display. *Methods Mol Biol*. **2371**, 287–298 (2022).
28. L. A. Gallagher, Methods for Tn-Seq Analysis in Acinetobacter baumannii. *Methods Mol Biol*. **1946**, 115–134 (2019).
29. N. Soundrarajan, S. Park, Q. le Van Chanh, H. sun Cho, G. Raghunathan, B. Ahn, H. Song, J. H. Kim, C. Park, Protegrin-1 cytotoxicity towards mammalian cells positively correlates with the magnitude of conformational changes of the unfolded form upon cell interaction. *Sci Rep*. **9** (2019), doi:10.1038/S41598-019-47955-2.
30. J. M. Anderson, B. Jurban, K. N. L. Huggins, A. A. Shcherbakov, I. Shu, B. Kier, N. H. Andersen, Nascent Hairpins in Proteins: Identifying Turn Loci and Quantitating Turn Contributions to Hairpin Stability. *Biochemistry*. **55**, 5537–5553 (2016).
31. J. T. Roehr, C. Dieterich, K. Reinert, Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics*. **33**, 2941–2942 (2017).
32. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data, (available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
33. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. **34**, 525–527 (2016).
34. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. **15**, 1–21 (2014).
35. A. Tristan Bepler, B. Berger, T. Bepler, Synthesis Learning the protein language: Evolution, structure, and function In brief ll Synthesis Learning the protein language: Evolution, structure, and function. *Cell Systems*. **12**, 654–669 (2021).
36. G. Luo, B. L. Stone, M. D. Johnson, P. Tarczy-Hornoch, A. B. Wilcox, S. D. Mooney, X. Sheng, P. J. Haug, F. L. Nkoy, Automating Construction of Machine Learning Models With Clinical Big Data: Proposal Rationale and Methods. *JMIR Res Protoc*. **6** (2017), doi:10.2196/RESPROT.7757.
37. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, doi:10.5555/3294996.
38. C. Dallago, K. Schütze, M. Heinzinger, T. Olenyi, M. Littmann, A. X. Lu, K. K. Yang, S. Min, S. Yoon, J. T. Morton, B. Rost, Learned Embeddings from Deep Learning to Visualize and Predict Protein Sets. *Current Protocols*. **1**, e113 (2021).

Acknowledgments: We would like to thank the Targeted Therapeutic Drug Discovery and Development Program for access to circular dichroism training and equipment. Also, Dr. Kristin Blake at the Mass Spectrometry Facility at The University of Texas at Austin for her help with high-resolution LC/MS.

Funding: This work was funded by the National Institutes of Health (AI125337, AI148419, AI159203), The Welch Foundation (F-1870), The Defense Threat Reduction Agency (HDTRA1-17-C0008), and Tito's Handmade Vodka.

Author contributions:
Conceptualization: JR, CD, BW, TC, CW
Methodology: JR, CD, TC,
Investigation: JR, CD, GD, TC, KG
Visualization: JR, CW
Funding acquisition: BW, CW
Project administration: JR, BW
Supervision: JR, BD, CW
Writing – original draft: JR
Writing – review & editing: JR, BW, CW

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: Any data not available in the manuscript or supplementary material can be made available upon request to the corresponding author.

Supplementary Materials: Contains the Materials and Methods, Supplemental Text, Figures S1-S3, and Table S1-S2