

UCSF

UC San Francisco Previously Published Works

Title

Cross-molecular active learning for the discovery of antimicrobial polyacrylamides

Permalink

<https://escholarship.org/uc/item/49m4f2zn>

Journal

Matter

ISSN

2590-2393

Authors

Williams, Shoshana C
Makar-Limanov, Anna
Greenstein, Gabriel
[et al.](#)

Publication Date

2026-09-01

DOI

10.1016/j.matt.2026.103026

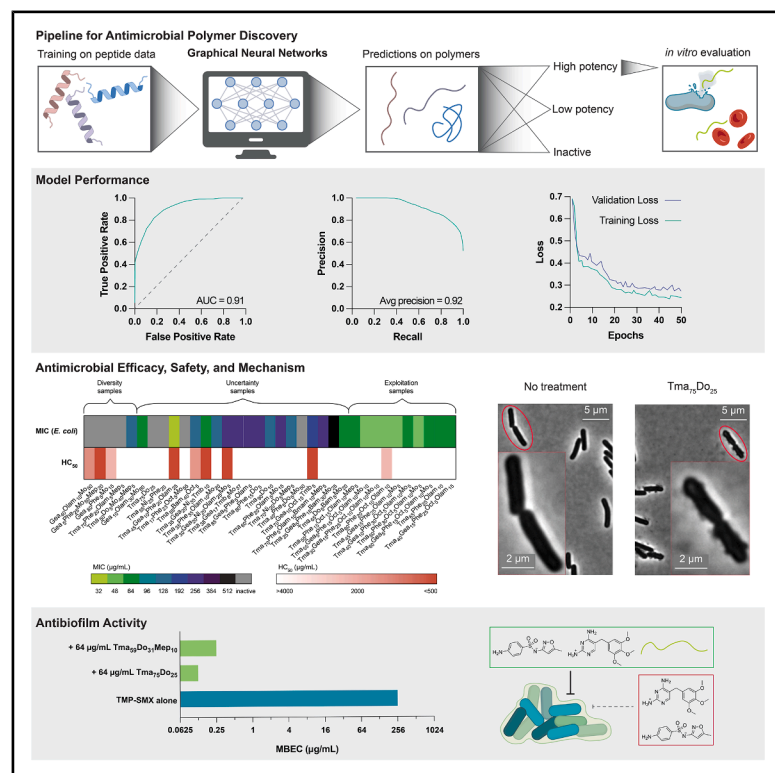
Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Cross-molecular active learning for the discovery of antimicrobial polyacrylamides

Graphical abstract



Authors

Shoshana C. Williams,
Anna Makar-Limanov,
Gabriel Greenstein, ..., Lynette Gegelski,
Hector Lopez Hernandez, Eric A. Appel

Correspondence

hector@conduciveml.com (H.L.H.),
eappel@stanford.edu (E.A.A.)

In brief

A cross-molecular machine learning pipeline was developed by using abundant antimicrobial peptide data for training before making predictions on the antimicrobial activity of synthetic copolymers. The predicted copolymers were synthesized and characterized, demonstrating favorable safety profiles and antimicrobial efficacy. They were shown to work through a membrane-disruption mechanism that helps overcome traditional resistance mechanisms. Further, they exhibit efficacy against difficult-to-target biofilm-associated bacteria.

Highlights

- Cross-molecular machine learning enables identification of antimicrobial polymers
- Predicted antimicrobial polymers show favorable safety profiles
- Predicted antimicrobial polymers show broad-spectrum antimicrobial efficacy
- One candidate shows antibiofilm efficacy and synergizes with existing antibiotics



Demonstrate

Proof-of-concept of performance with intended application/response

Williams et al., 2026, Matter 9, 103026
December 2, 2026 © 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
<https://doi.org/10.1016/j.matt.2026.103026>

Article

Cross-molecular active learning for the discovery of antimicrobial polyacrylamides

Shoshana C. Williams,¹ Anna Makar-Limanov,¹ Gabriel Greenstein,² Xinyu Liu,¹ Alessio Fragasso,^{3,4,5} Priya Ganesh,² Noah Eckman,⁶ Alexander N. Prossnitz,² Changxin Dong,² Christine Jacobs-Wagner,^{3,4,5,7} Lynette Cegelski,¹ Hector Lopez Hernandez,^{2,*} and Eric A. Appel^{2,4,8,9,10,11,*}

¹Department of Chemistry, Stanford University, Stanford, CA 94305, USA

²Department of Materials Science & Engineering, Stanford University, Stanford, CA 94305, USA

³Department of Biology, Stanford University, Stanford, CA 94305, USA

⁴Sarafan ChEM-H, Stanford University, Stanford, CA 94305, USA

⁵Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

⁶Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA

⁷Department of Microbiology and Immunology, Stanford University, School of Medicine, Stanford, CA 94305, USA

⁸Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

⁹Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA

¹⁰Department of Pediatrics (Endocrinology), Stanford University, Stanford, CA 94305, USA

¹¹Lead contact

*Correspondence: hector@conduciveml.com (H.L.H.), eappel@stanford.edu (E.A.A.)

<https://doi.org/10.1016/j.matt.2026.103026>

PROGRESS AND POTENTIAL Bacteria are becoming resistant to our existing arsenal of antibiotics at an alarming pace, posing a serious threat to global health. For decades, natural and synthetic antimicrobial peptides have been widely investigated for their broad-spectrum efficacy and propensity to overcome resistance. Their development has been complicated, however, by their off-target toxicity, high cost of synthesis, and poor pharmacokinetics. By contrast, peptide-inspired synthetic antimicrobial polymers can overcome these pitfalls, while maintaining the favorable properties of antimicrobial peptides. Such polymers, therefore, represent an attractive avenue for future antibiotic development.

Although machine learning pipelines have been implemented to successfully identify potent antimicrobial peptides, their application to synthetic copolymers has been limited due to insufficient training data. In this work, we leveraged abundant data on antimicrobial peptides to train an active learning pipeline to identify antimicrobial polymers. The predicted polymers were synthesized and evaluated for non-toxicity, antibacterial efficacy, and mechanism of activity. They demonstrated favorable safety profiles and broad-spectrum activity. One candidate copolymer was further characterized to show its antibiofilm efficacy, suggesting that it may be useful against hard-to-treat biofilm-associated infections. Further, this candidate copolymer works by permeabilizing bacterial membranes, a mechanism which has been shown to reduce antibiotic susceptibility to resistance, suggesting that such polymers may help overcome traditional resistance mechanisms. This work demonstrates the potential of cross-molecular machine learning by training a pipeline using peptides to offer high-accuracy predictions on copolymers, and it contributes to the development of a more robust antibiotic arsenal to combat the growing crisis of antimicrobial resistance.

SUMMARY

Antimicrobial resistance poses an urgent threat to global health. The development of new antimicrobials is crucial. Synthetic copolymers are attractive as a potential tool because they can be produced at scale and designed to mimic antimicrobial peptides and act as broad-spectrum antimicrobials capable of evading resistance mechanisms. This work leverages a cross-molecular machine learning pipeline, trained on antimicrobial peptides, to develop potent antimicrobial polymers—which were then synthesized and validated experimentally—to combat *Escherichia coli*. Crucially, all polymers predicted by the active learning pipeline during exploitation sampling exhibited impressive antimicrobial activity. One candidate copolymer was further characterized and shown to permeabilize the bacterial membrane, which is associated with

decreased resistance. Furthermore, this copolymer demonstrated remarkable synergy in eradicating biofilm-associated *E. coli* when combined with a clinical drug regimen. These results demonstrate promise for potentiating antibacterial activity of currently available antibiotics, treating serious and complicated infections, and combatting antimicrobial resistance.

INTRODUCTION

Antimicrobial resistance (AMR) is a global emergency as antimicrobial resistant infections contribute to 4.7 million deaths annually.¹ In the United States alone, AMR costs \$4.6 billion in direct healthcare spending.² The toll is even heavier in sub-Saharan Africa and South Asia.³ The burden of AMR is expected to soar in the coming decades, as antimicrobial resistant strains become more prominent and existing antibiotics become increasingly ineffective. It has been estimated that by 2050, AMR will contribute to 8.2 million deaths annually.¹

To combat this growing crisis, it is imperative to develop novel antimicrobial agents that help mitigate or overcome the onset of resistance. AMR arises through several mechanisms, which most commonly involve drug target modification, drug inactivation, drug uptake inhibition, and/or enhanced drug efflux.^{4–6} Host-defense peptides, or antimicrobial peptides, have gained attention in recent decades as potent, naturally occurring antimicrobials. These peptides are typically amphiphilic and contain cationic amino acids, which enhance interactions with the anionic bacterial cell surface, as well as hydrophobic amino acids, which promote interaction with and disruption of the hydrophobic compartment of the bacterial cell membrane, leading to bacterial cell death.^{7–12} Notably, this mechanism simultaneously enables broad-spectrum activity and limits the development of resistance, as it neither acts on a singular protein target nor depends on intracellular access.^{13,14}

Despite their promise, the clinical translation of host-defense peptides has been limited by synthetic challenges, off-target toxicity, susceptibility to proteolytic degradation, and poor pharmacokinetics.^{15–17} To address these challenges, researchers have explored the development of synthetic polymers that mimic the properties of host-defense peptides, several of which exhibit broad-spectrum activity, enhanced safety profiles, and resistance-evading properties.^{18–26} They present an affordable, effective alternative to host-defense peptides.

The further development of antimicrobial polymers is of great interest in combatting AMR, and the potential chemical space available to explore is nearly infinite. This opportunity also presents a challenge as it is impossible to comprehensively, experimentally evaluate candidates across the realm of polymers to find the most potent candidates. Machine learning pipelines have demonstrated great utility in evaluating vast chemical space to identify functional molecules, including for polymer applications.^{27–33} Researchers are curating databases of antimicrobial polymers to aid in such active learning approaches; however, the utility of these databases is constrained by their relatively small size.^{34,35} In contrast, vast databases detailing the activity of host-defense peptides are available, and these have been successfully implemented to identify new, potent antimicrobial peptide candidates.^{36–40} As antimicrobial polymers are inspired by host-defense peptides and can be characterized by the same key

chemical characteristics that influence peptide activity (charge density, hydrophobicity, etc.),⁴¹ we implemented an active learning pipeline leveraging peptide training data to predict antimicrobial polymers (Figure 1). We then evaluated the polymers' efficacy against *Escherichia coli* and their hemolytic activity. For one top-performing candidate, we investigated its membrane-permeabilizing activity, efficacy against biofilms, and synergy with a clinically used antibiotic regimen and compared these results with those for a previously reported antimicrobial polymer.

RESULTS

Design parameters for antimicrobial polymers

Due to the global burden of AMR, it is critical to design candidate polymers with favorable chemical features that would enable global distribution; that is, they must be affordable, simple to synthesize, and chemically stable. To meet these requirements, we chose to use statistical polyacrylamide derivatives. We used a library of 12 candidate monomers: 2 hydrophilic, 8 hydrophobic, and 2 cationic (Figure 1). These monomers are either commercially available or easily synthesized, and their controlled polymerization is scalable. The resultant copolymers feature a carbon backbone, which is highly chemically stable. Polymers of this class have been previously shown to demonstrate promising antimicrobial activity.²⁶ Although these synthetic polymers feature distinct backbones and side groups compared with naturally occurring amino acids (Figure 1), they nonetheless provide access to a chemically diverse space that enables investigation of variations in the cationic, hydrophobic, and hydrophilic properties of the resulting copolymers, each of which are known to influence antimicrobial efficacy and safety.^{42–49} The hydrophobic monomers span a wide range of logP values, as do hydrophobic amino acids. Our library did not include any negatively charged monomers, as cationic (rather than anionic) charge density is thought to be a key contributor to antimicrobial efficacy.⁴² We targeted polymers with a degree of polymerization (DP) of 70 monomers. This is significantly larger than the average antimicrobial peptide used for training (23 amino acids), but the polymer length was important to reduce the impact of compositional dispersity. We considered synthetic copolymers synthesized with 2–6 distinct monomers in increments of 5 wt %. This selection creates a combinatorial space of over 1.7 million possible synthetic polymers—far too many to screen experimentally without computational guidance. This computational library was screened using our cross-molecular machine learning to predict the antimicrobial efficacy of each entry.

Cross-molecular learning from peptides to polymers

We demonstrate that graph neural networks (GNNs) trained on antimicrobial peptide data can accurately predict the antimicrobial activity of structurally distinct synthetic polymers (Figure 2), as described in detail below. The promise of the system is

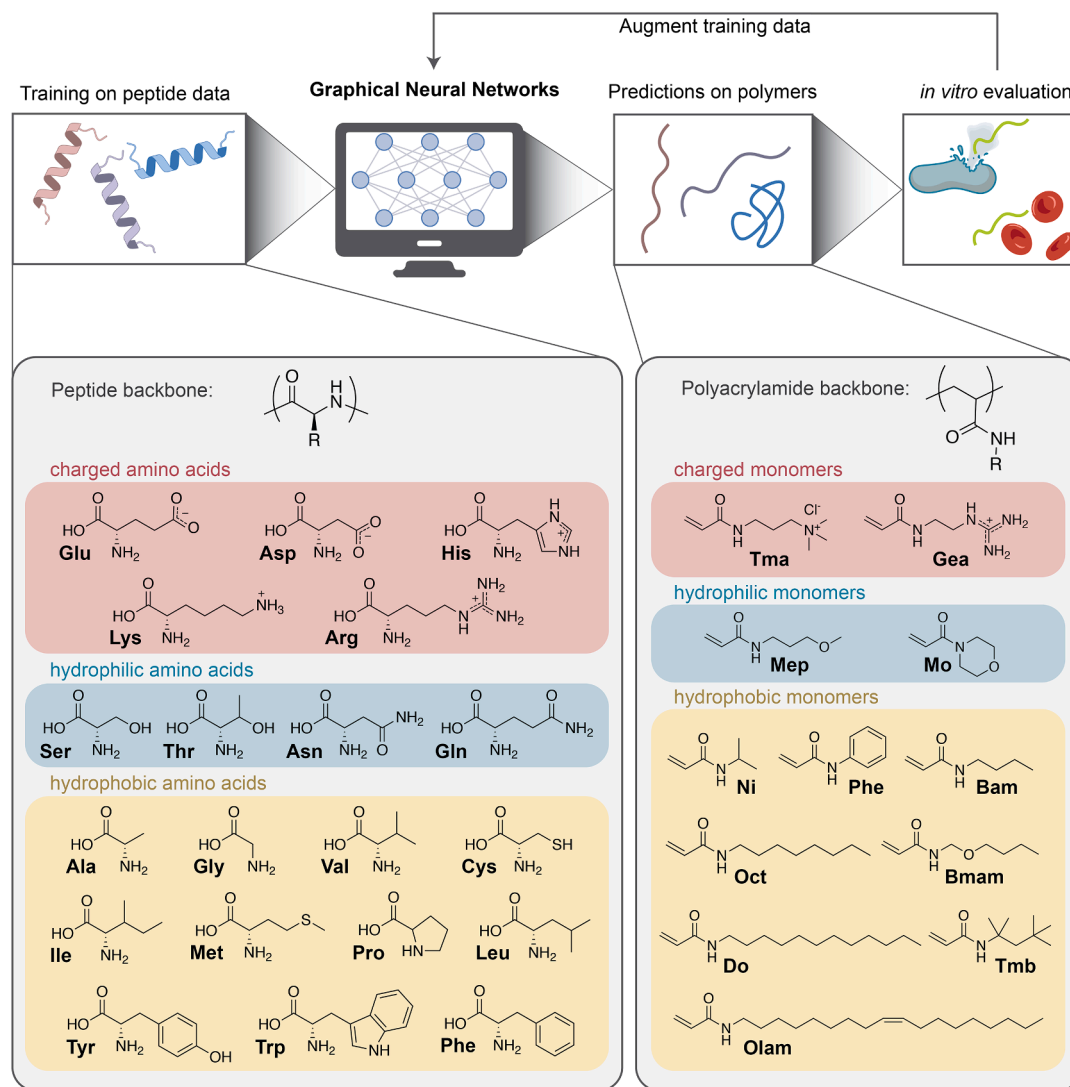


Figure 1. Schematic indicating the implementation of an active learning pipeline to mine for potent antimicrobial polymers

The active learning pipeline was trained on peptide data and leveraged to generate predictions on a large pool of synthetically accessible polyacrylamide candidates generated from commercially available or easily synthesized acrylamide derivatives.

underscored by the receiver operating characteristic area under the curve (ROC-AUC) of 0.91 and the precision-recall AUC (PR-AUC) of 0.92 when evaluated on a test set of known antimicrobial polymers (Figure 2). This cross-molecular learning capability represents a significant advance over traditional approaches that require extensive training data within the target material class. By leveraging shared physicochemical properties at the monomer level, our approach enables prediction of polymer antimicrobial activity using peptide training data, overcoming the fundamental limitation of insufficient polymer datasets in this domain.

Model architecture and cross-molecular implementation

The development of a unified GNN model capable of predicting antimicrobial efficacy across both peptides and synthetic poly-

mers required careful consideration of the molecular representation strategies. We adopted a graph-based approach where molecules are represented as networks of interconnected monomeric units, with nodes corresponding to individual monomers and edges representing chemical bonds (peptide bonds for peptides, carbon-carbon bonds for polymers). Each node is represented by a feature vector derived from RDKit molecular descriptors, with 104 chemically relevant descriptors selected from the original 217 available descriptors based on their relevance to antimicrobial activity. This representation preserves critical structural information, while enabling cross-molecular learning between chemically distinct compound classes.

A fundamental challenge in this approach stems from the inherent differences between sequence-defined peptide materials and stochastically defined synthetic polymers. While

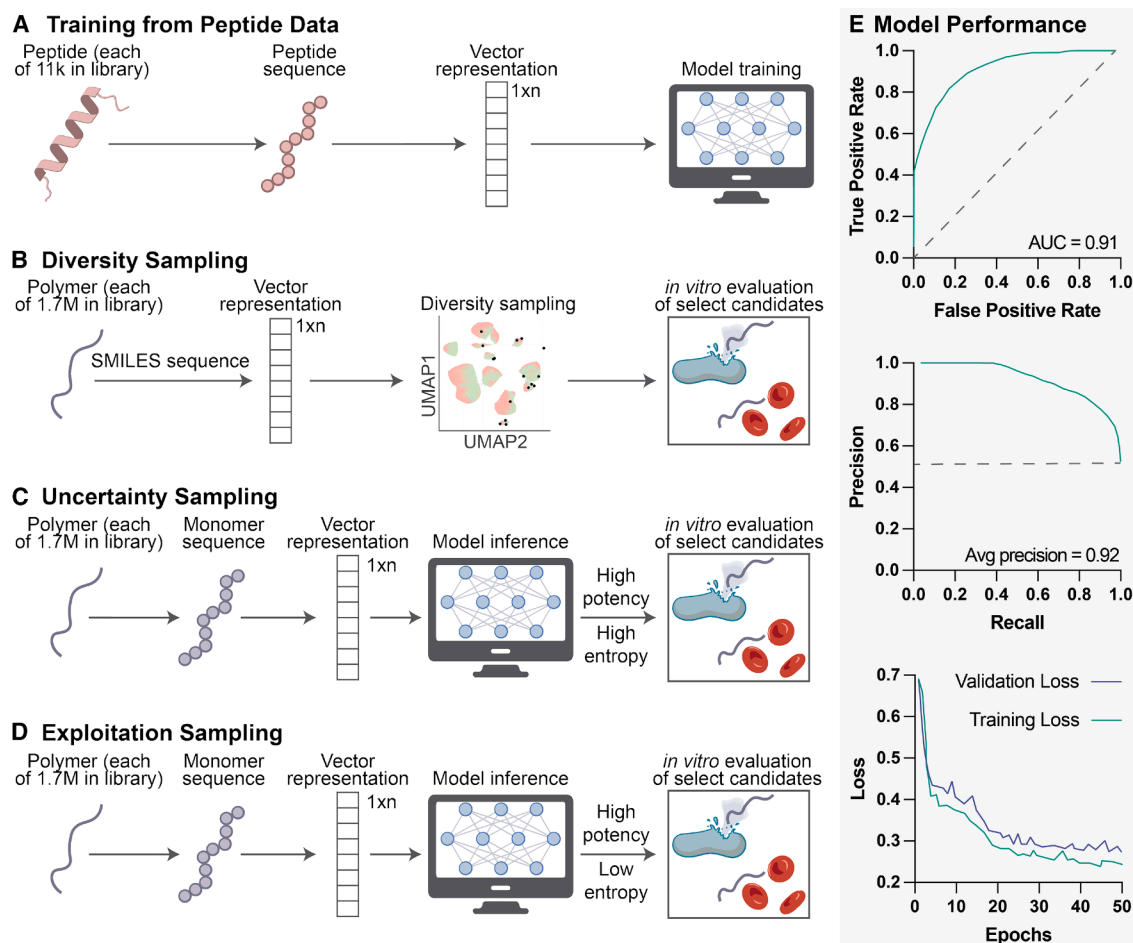


Figure 2. Schematic indicating the active learning workflow

(A) Model training.

(B–D) Diversity sampling (B), uncertainty sampling (C), and exploitation sampling (D) from a set of 1.7M synthetically accessible polymers.

(E) Model performance after training, evaluated on a test set of known antimicrobial polymers.

peptides have well-defined sequences, copolymers exhibit compositional dispersity. To address this challenge, we implemented a stochastic sampling strategy that generates multiple possible sequences for each polymer composition. Our peptide dataset comprised antimicrobial peptide data from the GRAMPA database⁵⁰ supplemented with non-antimicrobial peptides from UniProt⁵¹ and was split into train/validation/test sets. Each entry was characterized using its amino acid sequence and its reported minimum inhibitory concentration (MIC). The amino acid sequences were mapped to a representation of molecular descriptors. A separate holdout test set of 27 synthetic copolymers, previously synthesized and characterized in our laboratory,²⁶ served as the primary evaluation benchmark for cross-molecular transfer learning performance. Similarly, these polymer entries consisted of stochastic monomer sequences and their associated MIC values, where the monomer sequence was mapped to a representation of molecular descriptors.

We evaluated five GNN architectures: (1) graph convolutional networks (GCNs), (2) message-passing neural networks

(MPNN), (3) graph attention networks (GATs), (4) GraphSAGE, and (5) AttentiveFP. All models were implemented using the Deep Graph Library (DGL-LifeSci) framework. For each architecture, we trained 100 models using Bayesian hyperparameter optimization and selected the best-performing model from each architecture for comparison (Figure 2A; Figure S1). On the basis of this evaluation, we selected MPNN for the active learning pipeline, which achieved the ROC-AUC of 0.91 and PR-AUC of 0.92 on the polymer holdout set (Figure 2E). The high precision-recall performance (PR-AUC: 0.92) is particularly important for experimental efficiency, as it minimizes false positives that would waste laboratory resources. The training and validation loss curves demonstrated robust convergence without overfitting, validating our model architecture and training protocol (Figure 2E, bottom).

Active learning pipeline for polymer discovery

Leveraging this cross-molecular learning capability, we implemented an active learning pipeline to systematically explore

the polymer design space and identify promising antimicrobial candidates. Because our initial library of antimicrobial polymers was small and our model's performance was validated on only a subset of the chemical space, we expected model predictions to provide noisy labels. Building on this assumption, we used the trained GNN model to power an active learning loop, whereby model predictions guided the selection of screening candidates.

Our active learning strategy comprised three distinct sampling phases that progressed systematically from exploration to exploitation (Figures 2B–2D): (1) diversity sampling to ensure broad coverage of the chemical space, (2) uncertainty sampling to train our models with samples in which it had the least confidence, and (3) exploitation sampling to identify high-efficacy candidates. This systematic progression from exploration to exploitation ensures a comprehensive coverage of the design space, while efficiently identifying high-value targets.

We conducted three rounds of diversity sampling, including two initial rounds using terpolymers (a total of six polymers, four of which were previously published,²⁶ and the other two are reported here) and one expanded round incorporating 4-monomer copolymers (three additional polymers). We employed a MaxMin diversity (RDKit MaxMin Sampler) sampling algorithm that maximizes the minimum pairwise distance between selected polymers in the chemical feature space (Figure 2B).

We conducted the two rounds of uncertainty sampling (ten polymers each), using an ensemble-based approach that identifies regions where multiple trained models disagree in their predictions (Figure 2C). We trained multiple independent neural network models and measured the disagreement between their predictions using Kullback-Leibler (KL) divergence from the ensemble mean. Polymers with high inter-model disagreement were prioritized for experimental validation, as these represent regions where the models are most uncertain and additional data would be most valuable for improving the predictive performance. This strategy successfully identified polymers in chemically ambiguous regions of the design space, enabling targeted refinement of model decision boundaries.

We then conducted one round of exploitation sampling (ten polymers) designed to identify high-efficacy candidates by classifying polymers with an MIC of less than 100 $\mu\text{g/mL}$ (Figure 2D). We selected candidates with the highest model confidence and the lowest ensemble disagreement (low KL divergence).

The complete active learning pipeline synthesized and characterized 39 polymers across six sampling rounds. Our final polymer library included an additional 23 rationally designed and previously screened polymers—for a total of 62 polymers. Critically, all sampling operations (diversity, uncertainty, and exploitation) enabled the selection of candidates directly from this pre-enumerated library of synthetically accessible compositions, rather than through optimization in the continuous feature space and subsequently mapping back to monomer formulations.

Efficacy and safety of candidate copolymers

The copolymers identified through diversity sampling, uncertainty sampling, and exploitation sampling were synthesized using reversible addition-fragmentation chain-transfer (RAFT) polymerization from commercially available or readily synthesized monomers (Figure 1 lists monomers and their abbreviations).

Each candidate copolymer was prepared by statistical copolymerization of mixtures of 2–6 monomers and named according to those monomers and their weight percent in the feed ratio (Table 1). The targeted molar ratios of the monomers are also listed (Table S2). The copolymers were analyzed by NMR and gel permeation chromatography (GPC) (Tables 1, S1, and S2; Figures S4–S9). GPC traces for cationic polymers are known to exhibit tailing,⁵² causing the dispersity to appear high. Further, the common convention of size determination through comparison with standards such as poly(methyl methacrylate) (PMMA) or polyethylene glycol (PEG) can yield deviation from the true polymer mass for any novel copolymer.⁵³ As such, the GPC analysis is best interpreted qualitatively, showing that the vast majority of synthesized copolymers exhibits a monomodal size distribution, consistent with controlled radical polymerization.

These copolymer materials were then evaluated for their potential antimicrobial efficacy through determination of the MIC against *E. coli* grown in Mueller-Hinton broth (MHB) at a cell density of 5×10^5 colony-forming units (CFUs)/mL (Figure 3; Table S3). In total, fourteen copolymers were identified with experimentally confirmed MIC values below 100 $\mu\text{g/mL}$ (range: 32–64 $\mu\text{g/mL}$, corresponding to 1.7–15.5 μM). These represent strong antimicrobial agents. We compared this activity with those of previously reported antimicrobial polymers available in a published database.³⁵ Among 107 polymers with reported activity against *E. coli* and reported molecular weights, the average MIC was 163 $\mu\text{g/mL}$ or 58 μM . The top-performing candidates reported here compare favorably to the body of previously published antimicrobial polymers. This result demonstrates the utility of ensemble-based approaches for both exploration and exploitation in active learning pipelines.

To evaluate the potential for undesirable mammalian toxicity, hemolysis was measured for each copolymer (Figure 3; Table S4). Twenty-six polymers demonstrated very low hemolysis, with the hemolytic concentration at which 50% of red blood cells are lysed (HC_{50}) > 4,000 $\mu\text{g/mL}$. These HC_{50} values are among the highest reported for antimicrobial polymers.^{42,49,54–56} The therapeutic index was calculated as the ratio between the HC_{50} and the MIC against *E. coli* (Table S4) to quantify the relationship between the potency and safety of each polymer. Twenty-two polymers demonstrated a therapeutic index of greater than 10 (range: from 11.7 to >83.3), suggesting the potential safety and efficacy of these candidates. The lethal concentration at which cell viability decreased by 50% (LC_{50}) was also calculated for each polymer after an 8-h incubation with 3T3 cells (Table S4). Although the LC_{50} values were lower than the HC_{50} values, they were comparable to the LC_{50} of Tween 20 (polysorbate 20), a commonly used commercial excipient that is approved for use at concentrations that are orders of magnitude higher than its measured LC_{50} ,⁵⁷ suggesting that LC_{50} may underestimate actual *in vivo* tolerance.

For each of the twenty-two polymers with a therapeutic index greater than 10, their MICs were also evaluated against *Staphylococcus aureus* ATCC 29213 and *Klebsiella pneumoniae* ATCC 13884, each of which were originally sourced from human clinical samples (Table S3; Figure S10). Nineteen polymers showed activity against *S. aureus* (MICs ranging from 16 to 96 $\mu\text{g/mL}$), and all twenty-two showed activity against *K. pneumoniae*.

Table 1. Characterization of synthesized polyacrylamides

	Target DP ^a	Target M _w (kDa) ^a	Measured M _w (kDa) ^b	Dispersity ^b	Conversion (%) ^c
Diversity sampling					
Gea ₄₀ Olam ₁₀ Mo ₅₀	70	11.8	24.3	3.8	92.5
Gea ₈₀ Phe ₅ Mo ₁₅	70	12.6	12.3	1.9	84.3
Gea ₅ Phe ₂₅ Mo ₅₀ Mep ₂₀	70	10.1	ND	ND	95.9
Tma ₅₀ Do ₅ Mo ₄₀ Mep ₅	70	12.0	16.2	1.5	99.3
Tma ₇₀ Phe ₂₀ Olam ₅ Mep ₅	70	13.0	26.4	1.9	78.2
Uncertainty sampling					
Tma ₇₅ Do ₂₅	70	15.0	23.4	1.2	99.9
Gea ₁₀ Olam ₃₀ Mo ₆₀	70	12.3	11.7	1.8	98.9
Tma ₅₅ Ni ₂₅ Phe ₂₀	70	11.2	20.3	1.5	99.5
Tma ₅₅ Bam ₄₀ Oct ₅	70	11.5	18.4	1.2	97.8
Tma ₁₇ Phe ₂₃ Oct ₅ Mo ₅₅	70	10.7	13.3	1.3	99.9
Tma ₄₅ Gea ₁₀ Phe ₂₀ Olam ₂₅	70	14.5	51.2	3.2	74.0
Tma ₃₅ Gea ₂₀ Ni ₃₅ Tmb ₁₀	70	11.0	20.2	1.6	96.8
Tma ₃₅ Phe ₃₅ Olam ₁₀ Mo ₂₀	70	12.1	34.1	2.3	92.5
Tma ₃₈ Gea ₁₇ Tmb ₈ Mo ₃₇	70	12.1	14.0	1.7	99.2
Tma ₃₀ Gea ₂₀ Ni ₂₀ Olam ₂₅ Mo ₅	70	13.0	14.3	2.2	94.5
Tma ₉₀ Do ₁₀	70	14.7	ND	ND	56.8
Tma ₈₀ Phe ₁₅ Do ₅	70	13.7	16.0	1.5	89.7
Tma ₈₅ Gea ₅ Phe ₅ Olam ₅	70	14.4	15.4	2.0	84.0
Tma ₇₀ Ni ₂₀ Do ₅ Mep ₅	70	12.3	17.8	1.9	93.3
Tma ₆₀ Phe ₂₀ Olam ₅ Mo ₁₅	70	12.8	19.2	1.8	96.9
Tma ₄₀ Phe ₅ Do ₅ Mo ₅₀	70	11.6	15.4	1.5	98.9
Tma ₇₀ Gea ₁₀ Oct ₁₅ Tmb ₅	70	14.0	14.1	1.7	77.9
Tma ₆₅ Do ₅ Bam ₅ Mo ₂₀ Mep ₅	70	12.7	17.3	1.6	97.6
Tma ₂₅ Gea ₅ Phe ₁₅ Bam ₃₀ Mo ₂₅	70	10.6	15.0	1.5	94.3
Tma ₇₀ Phe ₅ Olam ₁₀ Bmam ₁₀ Mep ₅	70	13.9	25.9	1.9	80.2
Exploitation sampling					
Tma ₅₅ Phe ₁₅ Oct ₁₅ Olam ₁₀ Mep ₅	70	13.8	8.8	2.6	ND
Tma ₆₀ Phe ₂₀ Oct ₁₀ Olam ₁₀	70	13.9	11.8	1.9	ND
Tma ₅₅ Gea ₅ Phe ₁₅ Oct ₅ Olam ₁₀ Mo ₁₀	70	13.6	7.1	1.8	97.7
Tma ₄₅ Gea ₁₅ Phe ₂₅ Olam ₁₀ Mo ₅	70	13.2	3.1	3.0	94.3
Tma ₃₀ Gea ₁₅ Phe ₂₅ Oct ₁₀ Olam ₁₀ Mo ₁₀	70	17.2	8.4	2.2	94.1
Tma ₆₅ Phe ₂₅ Olam ₁₀	70	13.6	8.8	2.3	ND
Tma ₅₀ Phe ₃₀ Oct ₅ Olam ₁₀ Mo ₅	70	13.0	9.7	1.8	ND
Tma ₄₀ Gea ₁₀ Phe ₃₀ Oct ₅ Olam ₁₀ Mo ₅	70	13.2	7.7	2.4	93.0
Tma ₆₀ Gea ₅ Phe ₁₅ Oct ₅ Olam ₁₀ Mo ₅	70	14.7	7.6	2.5	97.4
Tma ₄₀ Gea ₁₅ Phe ₂₅ Oct ₅ Olam ₁₅	70	16.9	5.6	4.3	76.3

^aTarget DP and target M_w are theoretical values.

^bWeight-average molecular weight (M_w) and dispersity were determined by GPC analysis, using refractive index (RI) for diversity and uncertainty samples and UV for exploitation samples.

^cPercent conversion was determined by ¹H NMR analysis.

(MICs ranging from 48 to 192 μg/mL). Although the polymers were selected exclusively for their predicted activity on *E. coli*, this pipeline was successful in identifying broad-spectrum antimicrobial agents, as hypothesized.

Eleven polymers demonstrated impressive efficacies, with *E. coli* MICs equal to or lower than 64 μg/mL, while also exhibiting HC₅₀ values greater than 4,000 μg/mL. These eleven polymers include compositions spanning 2–6 monomers, represent-

ing the range used in designing the library. This finding is interesting as it demonstrates that greater complexity is not necessarily associated with greater performance, although the incorporation of more monomers is associated with higher compositional dispersity (Figure S11).

One copolymer candidate (Tma₇₅Do₂₅) was selected for further evaluation due to its favorable antimicrobial efficacy paired with its low hemolytic activity as well as relatively low

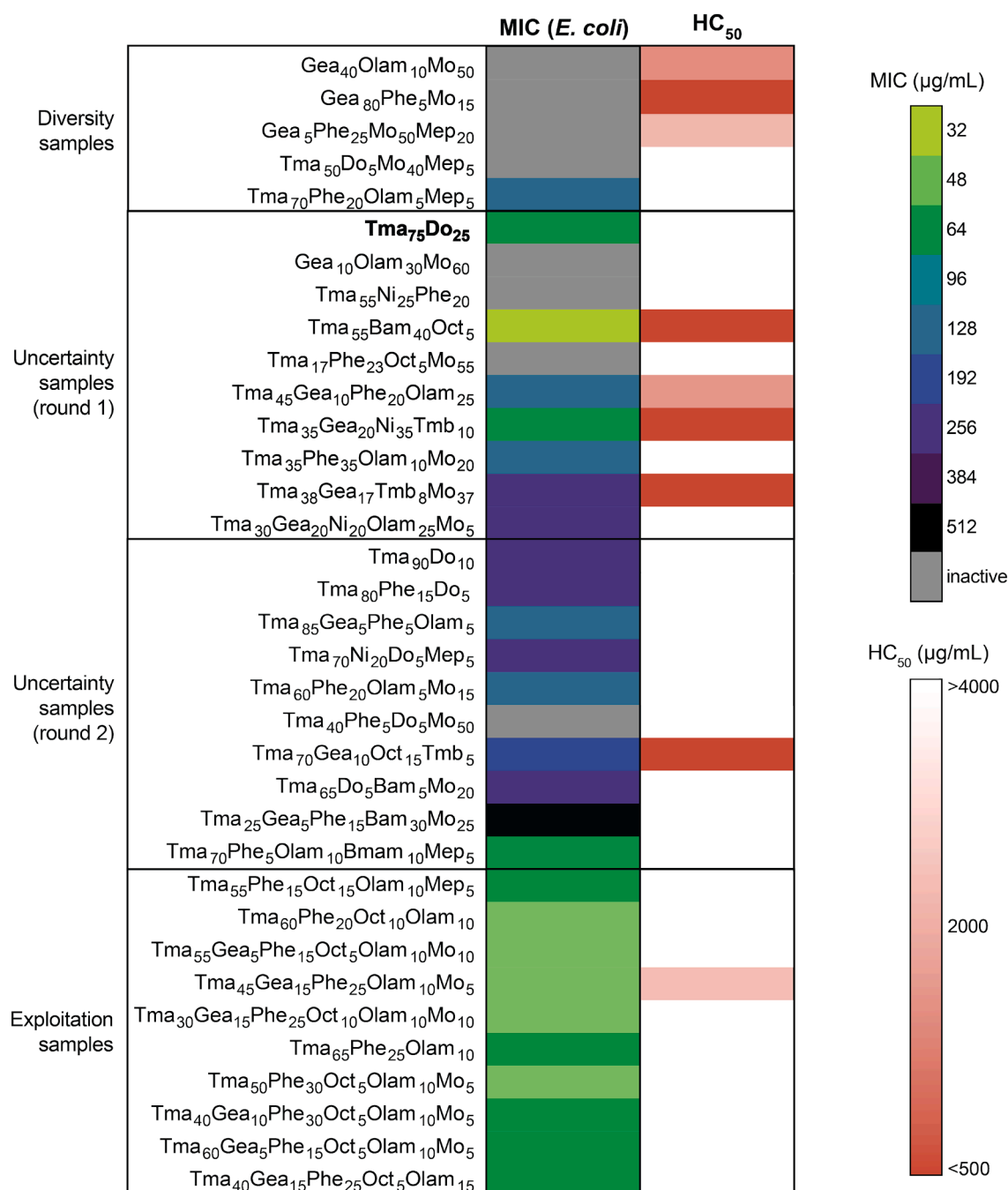


Figure 3. Safety and efficacy of polyacrylamide derivatives

(Left) Heatmap indicating the MIC of each copolymer against *E. coli*. (Right) Heatmap indicating their hemolytic activity. Bold is used to highlight Tma₇₅Do₂₅, a top-performing candidate that was selected for further characterization.

expected compositional dispersity. This top-performing candidate was evaluated *in vivo* in mice through both intraperitoneal (i.p.) and intranasal (i.n.) dosing at high (400 μg) and low (200 μg) doses, alongside a vehicle control (*n* = 5 animals per group). The animals were weighed daily, and their blood chemistry was analyzed 5 days after dosing (Figure S12). Animal

weights were stable following treatment, and blood chemistry values remained within the reference ranges, aside from a slight elevation in blood urea nitrogen for the animals receiving the highest i.n. dose (Figure S12). These observations indicate that our leading copolymer is well tolerated *in vivo* following either systemic i.p. dosing or locoregional i.n. dosing.

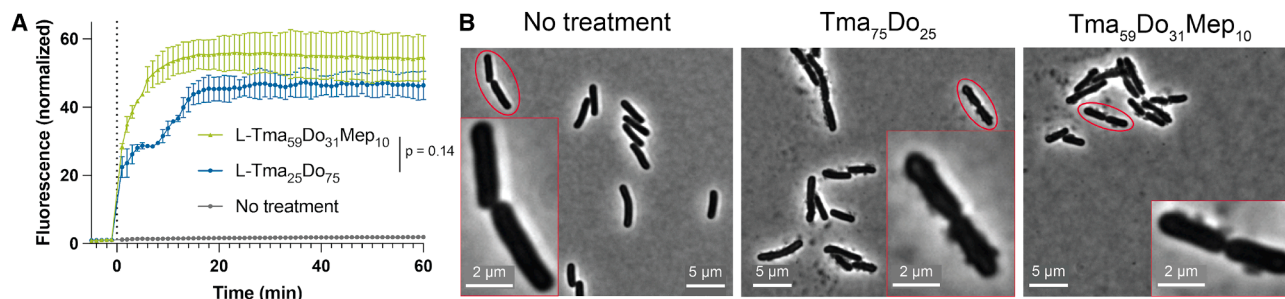


Figure 4. Mechanism of antimicrobial copolymers

(A) Membrane permeabilization assay for *E. coli* treated with each polymer at $2\times$ MIC, using the fluorescent probe Sytox Green, monitored continuously by plate reader. Baseline fluorescence was normalized to a value of 1. The samples were run in duplicate. Error bars represent one standard deviation. (B) Optical microscopy of *E. coli* after treatment with each polymer at $2\times$ MIC.

Copolymer activity against the bacterial membrane

The top-performing candidate Tma₇₅Do₂₅ was compared with a previously reported copolymer identified through rational design (Tma₅₉Do₃₁Mep₁₀; previously labeled L-Do₃₁Mep₁₀).²⁶ *E. coli* were incubated with the DNA-intercalating probe Sytox Green in the presence of each antimicrobial copolymer. A rapid and dramatic increase in Sytox Green fluorescence was observed upon the addition of the antimicrobial copolymer (Figure 4A), indicating that the copolymers induced damage to the inner membrane of the bacteria. The damaged inner membrane, which is impermeable to Sytox Green when intact, enabled the Sytox Green's intracellular access and subsequent intercalation into DNA. This result is general across both copolymers, and no statistically significant difference between the polymers was observed in the Sytox Green permeability ($p = 0.14$).

Additionally, copolymer-treated *E. coli* were examined by phase-contrast light microscopy (Figure 4B). Ruffling was observed on the surface of the bacterial cells after treatment with each polymer relative to untreated *E. coli*. Together, these results demonstrate that these antimicrobial copolymers disrupt the surface of the bacteria and compromise the cell membranes.

Antimicrobial copolymers eradicate biofilm-associated bacteria and potentiate clinically used antibiotics

Biofilm infections are notoriously difficult to eradicate with many clinically available antibiotics. This is, in part, because bacteria in a biofilm are protected by an extracellular polymeric matrix, and they can be less metabolically active and, therefore, less susceptible to antibiotics that target cell growth.⁵⁸ As such, biofilms can cause persistent or chronic infections, with limited treatment options.⁵⁹ New antimicrobials capable of combating biofilms, are therefore, highly desirable.

To test the efficacy of our leading copolymers against biofilms, we determined their minimum biofilm eradication concentration (MBEC) against *E. coli* grown in M63 medium with glucose. For each copolymer, the MBEC was 512 $\mu\text{g/mL}$ (Table 2; Figure 5; Figure S13), which falls far below the HC₅₀ (above 4,000 $\mu\text{g/mL}$ for each). By contrast, an antibiotic regimen commonly used as a first-line treatment for urinary tract infections—trimethoprim-sulfamethoxazole⁶⁰ (abbreviated as TMP-SMX)—demonstrated an MBEC of 256 $\mu\text{g/mL}$ (Table 2; Figures 5 and S13),

which may approach the limit of its therapeutic window, as its LD₅₀ for injection in mice is 700 mg/kg.⁶¹

The copolymers were then explored as adjuvants to investigate their potential to synergize with TMP-SMX. Each copolymer exhibited remarkable synergy in combination with TMP-SMX (Table 2; Figures 5 and S13). For the combination of Tma₇₅Do₂₅ and TMP-SMX, the polymer MBEC was 64 $\mu\text{g/mL}$ (corresponding to 1/8th of its MBEC value alone), and the TMP-SMX MBEC was 0.125 $\mu\text{g/mL}$ (corresponding to 1/2,048th of its MBEC value alone). For the combination of Tma₅₉Do₃₁Mep₁₀ and TMP-SMX, the polymer MBEC was 64 $\mu\text{g/mL}$ (corresponding to 1/8th of its MBEC value alone), and the TMP-SMX MBEC was 0.25 $\mu\text{g/mL}$ (corresponding to 1/1,024th of its MBEC value alone). The fractional inhibitory concentration indices (FICIs) were 0.1255 and 0.1260 for the combinations with Tma₇₅Do₂₅ and Tma₅₉Do₃₁Mep₁₀, respectively, indicating exceptionally strong synergy. This remarkable finding demonstrates the ability of the copolymers to dramatically improve the efficacy of a first-line antibiotic treatment, bringing it back to its therapeutically relevant dosing. The copolymers show potential to expand the scope of existing antibiotics in clinical applications.

DISCUSSION

New antimicrobial agents are urgently needed to address the growing crisis of AMR. Toward that goal, this work introduces a cross-molecular active learning pipeline to identify

Table 2. Efficacy of copolymers and TMP-SMX against biofilm-associated bacteria, alone and in combination

Treatment	TMP-SMX	Tma ₇₅ Do ₂₅	Tma ₅₉ Do ₃₁ Mep ₁₀
MBEC alone, $\mu\text{g/mL}$ (μM)	256 (984.9)	512 (21.9)	512 (42.2)
MBEC for TMP-SMX-Tma₇₅Do₂₅ combo, $\mu\text{g/mL}$ (μM)	0.125 (0.5)	64 (2.7)	N/A
MBEC for TMP-SMX-Tma₅₉Do₃₁Mep₁₀ combo, $\mu\text{g/mL}$ (μM)	0.25 (1.0)	N/A	64 (5.3)

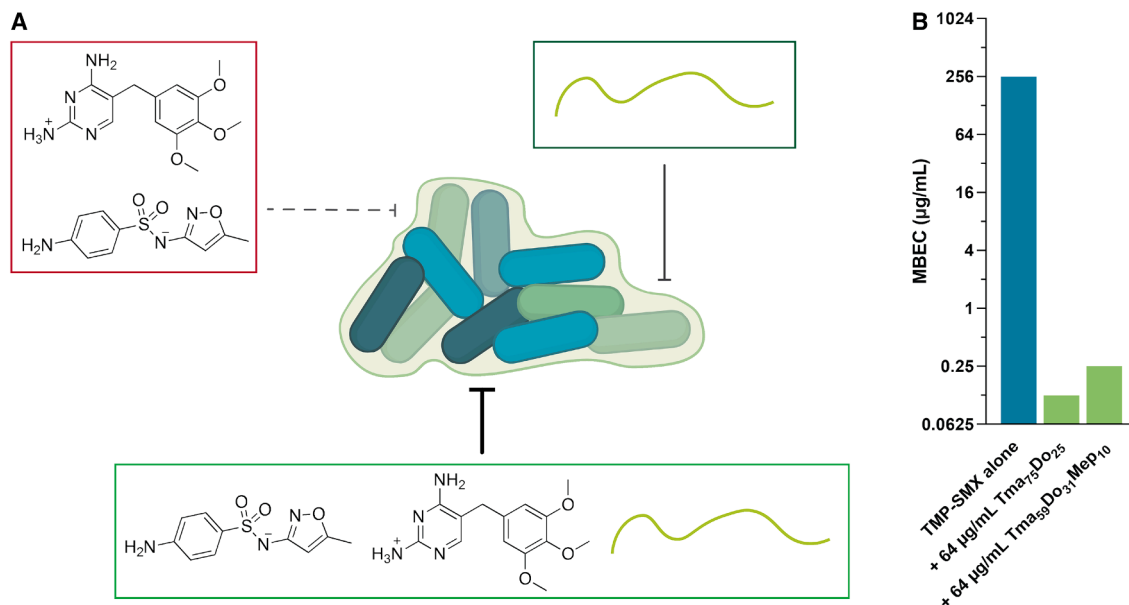


Figure 5. Anti-biofilm efficacy of copolymers

(A) Schematic for the anti-biofilm and antibiotic-potentiating activity of copolymers.

(B) Comparison of MBEC for TMP-SMX antibiotic treatment alone and in combination with each polymer.

polyacrylamide-based copolymers with antimicrobial activity against *E. coli*. We implemented GNNs trained on antimicrobial peptides to predict the properties of copolymer compositions, demonstrating the potential for cross-molecular training. We screened the predicted copolymers to identify compositions that are both effective and non-toxic, demonstrating that all copolymer compositions predicted by the active learning process during exploitation sampling exhibited efficacious antimicrobial activity. This result confirms that the machine learning pipeline can make predictions on simplified (i.e., monodisperse) model polymers, which can be applied to real-world polymers with outstanding accuracy, thereby supporting the strength of leveraging such models. We selected one top candidate, Tma₇₅Do₂₅, for further investigation, alongside a previously reported polymer, Tma₅₉Do₃₁Mep₁₀. Fluorescence nucleic acid staining and microscopy imaging demonstrated that these copolymers damage the bacterial cell surface and permeabilize the inner membrane. Furthermore, our results revealed tremendous synergy and potentiation of a first-line antibiotic regimen, TMP-SMX, in combination with each copolymer. Remarkably, each combination tested reduced the TMP-SMX concentration needed to eradicate bacteria in biofilms by three orders of magnitude, positioning it in the highly therapeutically attractive sub-micromolar range.

In summary, this work leverages cross-molecular machine learning to identify antimicrobial polyacrylamide derivatives that are more potent than those previously reported, without sacrificing hemolytic non-toxicity. As hypothesized, the newly characterized polymer attacks bacteria via a mechanism similar to a previously reported polymer as well as antimicrobial peptides, enabling broad-spectrum efficacy. This similar mechanism demonstrates the utility of cross-molecular machine learning, through which a

distinct class of molecules (peptides) can be used to predict the properties of another class of molecule (polyacrylamide derivatives) due to their shared chemical properties (i.e., charge density, hydrophobicity, etc.) and mechanism of action. The improved potency demonstrates that the computational screen was successful in identifying candidates that match or exceed those that have been previously reported, enabling a systematic evaluation of a much larger library than could be experimentally probed.

Future directions can expand on this promising strategy and should focus on incorporating multi-objective optimization into the machine learning pipeline to allow for simultaneous optimization on multiple goals, such as potency against multiple strains of bacteria and favorable safety profiles. In addition, while the copolymer design space described here is combinatorially large at 1.7 million compositions, it is constrained to 12 monomers, representing a narrow scope of possible copolymer chemistries. There is enormous potential to expand the chemistry and activities of the materials screened. The incorporation of diverse backbone architectures, alternative polymerization mechanisms, and broader monomer libraries will be essential for developing the most powerful, truly generalizable predictive models. Future work will also benefit from the implementation of sophisticated methods for modeling molecular weight dispersity⁶² and compositional dispersity⁶³ to enable more rigorous modeling of *in silico* polymerizations in a machine learning pipeline. Additional work should further explore the efficacy of the reported copolymer candidates against more bacterial species, both *in vitro* and *in vivo*. This work has demonstrated that these copolymers are effective antimicrobials, with low hemolysis and the potential to rescue existing antibiotics to combat dangerous biofilm-associated infections. Further efforts toward their translation could significantly broaden our arsenal of effective agents to combat AMR.

METHODS

Materials

The following solvents and reagents were used as received: *N,N*-dimethylformamide (DMF; Sigma-Aldrich), hexane (Sigma-Aldrich), acetone (Aldon Corporation), diethyl ether (Fisher Scientific), ethyl acetate (EtOAc; Sigma-Aldrich), dichloromethane (DCM; Sigma-Aldrich), methanol (MeOH; Sigma-Aldrich), deuterium oxide (D₂O; Sigma-Aldrich), chloroform-*d* (CDCl₃; Sigma-Aldrich), dimethyl sulfoxide-*d*₆ (DMSO-*d*₆; Sigma-Aldrich), acryloyl chloride (Sigma-Aldrich), triethylamine (TEA; Sigma-Aldrich), *N*-octylamine (TCI Chemicals), oleylamine (Sigma-Aldrich), 1,3-bis(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (Sigma-Aldrich), ethylenediamine (Sigma-Aldrich), and trifluoroacetic acid (TFA; Sigma-Aldrich or TCI Chemicals).

Similarly, the monomers *N*-phenylacrylamide (Phe; Sigma-Aldrich), *N*-dodecylacrylamide (Do; TCI Chemicals), *N*-butylacrylamide (Bam; TCI Chemicals), *N*-(butoxymethyl)acrylamide (Bmam; TCI Chemicals), and *N*-(1,1,3,3-tetramethylbutyl)acrylamide (Tmb; TCI Chemicals) were used as received. *N*-Isopropylacrylamide (Ni) was recrystallized from the solvent hexane. 4-Acryloylmorpholine (Mo; Sigma-Aldrich) and *N*-(3-methoxypropyl)acrylamide (Mep; Sigma-Aldrich) were filtered through basic alumina before use. (3-Acrylamidopropyl)trimethylammonium chloride (Tma; 75 wt % in H₂O, Sigma-Aldrich) was washed thrice with an equal volume of EtOAc and placed under vacuum for 2 min to remove residual EtOAc.

Azobisisobutyronitrile (AIBN; Sigma-Aldrich) was recrystallized from MeOH and dried under vacuum before use. 4-(((2-Carboxyethyl)thio)carbonothioyl)thio)-4-cyanopentanoic acid (CTA; Sigma-Aldrich) was used as received.

Bacterial strains were purchased from Microbiologics. MHB (BD Difco), tryptic soy media (Sigma-Aldrich), and agar (Sigma-Aldrich) were autoclaved before use.

Fluorescence images were acquired using a Nikon Ti2-E inverted microscope equipped with a Photometrics Prime BSI back-illuminated sCMOS camera (2,048 × 2,048 pixels, 6.5 μm pixel size), a 100× Plan Apo 1.45NA Ph3 oil objective (with type N immersion oil by Nikon), a Lumencore Spectra III LED (light-emitting diode) engine, a perfect focus system (PFS), temperature-controlled Okolab enclosure, and a polychroic mirror (FF-409/493/596-Di02 for by Shemrock) combined with a triple-pass emitter (FF-1-432/523/702-25 by Shemrock).

Dataset preparation and molecular representation

Antimicrobial peptide dataset

Our training dataset comprised antimicrobial peptide sequences from the GRAMPA database, which aggregates 51,345 peptides from multiple databases (APD, DADP, DBAASP, DRAMP, and YADAMP).⁵⁰ We retained only peptides reported to have the MICs against *E. coli*, excluded sequences with unusual modifications, and collapsed duplicate entries by taking the geometric mean of their MIC values. This yielded 5,154 unique positive AMPs, with lengths ranging from 2 to 190 amino acids (mean 22.6 ± 12.6 aa) and MIC values spanning 0.023–3,200 μM (mean 60.2 ± 213.3 μM). It should be noted that this database is imperfect in its aggregation, as it fails to account for inequalities. For instance, if a peptide was recorded in another database

as having an MIC >125 μM (i.e., no activity observed), GRAMPA indicates it as having an MIC = 125 μM. Although the database lacks precision on individual peptides, machine learning algorithms are generally tolerant to errors in individual points of a large library. Non-antimicrobial peptides were obtained from UniProt,⁵¹ selecting reviewed (Swiss-Prot) entries tagged as generic peptides with lengths between 1 and 40 amino acids, while excluding those annotated with “antimicrobial” keyword. After removing duplicates and non-canonical residues, we obtained 5,917 negative samples (mean length: 24.9 ± 10.8 aa) and assigned placeholder MIC values of 9,999 μM. The final peptide corpus contained 11,003 distinct sequences for model training and validation. The dataset was partitioned into training (70%), validation (15%), and test (15%) splits, using stratified sampling, to maintain class balance across the splits.

Synthetic polymer dataset

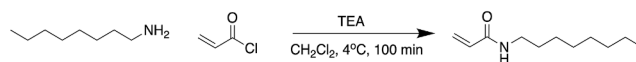
A holdout test set of 27 synthetic copolymers, previously synthesized and characterized in our laboratory, served as the primary evaluation benchmark for cross-molecular transfer learning performance. These polymers represent a subset of our compositional library containing 12 monomers representing diverse chemical functionalities: 2 hydrophilic, 8 hydrophobic, and 2 cationic monomers and spanning compositions from 2 to 6 monomers. The 27 synthesized polymers had experimentally determined antimicrobial activities against *E. coli*.

Graph representation

Both peptides and polymers were represented as molecular graphs where nodes correspond to monomeric units (amino acids for peptides, and chemical monomers for polymers) and edges represent covalent bonds. Node features were derived from RDKit molecular descriptors computed using rdkit.Chem.Descriptors, generating 217 descriptors per SMILES string. From these, we selected 104 descriptors deemed chemically relevant to the antimicrobial activity by excluding descriptors that primarily scale with chain length (SIZE), are redundant with retained descriptors (REDUN), represent simple atom counts with little variance (COUNT_dup), encode max/min versions of properties already captured by richer descriptors (EXTREMA), or count functional groups rarely present in peptides and antimicrobial polymers (FRAG_rare). Edge features encoded bond types and connectivity patterns, with 79 edge features retained after removing zero-variance descriptors across the combined peptide and polymer datasets.

Polymer sequence generation

To address the inherent polydispersity of synthetic copolymers, we implemented a stochastic sampling strategy that generates multiple possible sequences for each polymer composition. Each polymer is defined by its monomeric composition (monomer identities and molar ratios) and target DP. To ensure compatibility with the sequence-based format of peptide data, random sequences were sampled from each polymer distribution, constraining the sample means to match the corresponding theoretical composition, where each simulated copolymer exhibited the target DP. Our approach of randomized sampling from a computational monomer pool is comparable to previously published work describing Monte Carlo simulations to predict polymer sequences.⁶⁴ This approach captures the intrinsic variability of linear polymer synthesis. During training, we sampled



Scheme 1. Synthesis of *N*-octylacrylamide

50 sequences per polymer composition. For evaluation on our test set, we generated 200 sequences per polymer and determined final predictions through majority voting across all realizations. For evaluation of our computational library of 1.7M potential polymers, we used one random sequence per polymer.

GNN architecture and training

Model architecture

We evaluated five GNN architectures: GCN, MPNN, GATs, GraphSAGE, and AttentiveFP. GCN applies convolution directly over node features, treating each neighbor equally, while MPNN updates nodes by aggregating messages from neighboring nodes and their connecting edge features. GAT introduces self-attention to learn which neighbors deserve more weight, and AttentiveFP combines local message propagation with a global attention mechanism to capture both short- and long-range interactions. Notably, GCN and GAT use only the node features, whereas MPNN and AttentiveFP also incorporate edge information. All models were implemented using the DGL-LifeSci framework with consistent hyperparameter optimization protocols.

Hyperparameter optimization

For each architecture, we trained 100 models using Bayesian hyperparameter optimization with tree-structured Parzen estimator (TPE) sampling. Optimization targeted validation set performance over 50 iterations, with early stopping based on validation loss plateauing.

Model selection

The best-performing model from each architecture was selected based on the validation set ROC-AUC (Figure S1). MPNN was chosen for the active learning pipeline based on superior cross-molecular transfer performance (ROC-AUC: 0.91 and PR-AUC: 0.92 on the polymer holdout set).

Active learning pipeline

Design space definition

The polymer design space encompassed copolymer combinations synthesized from 12 candidate monomers (structures are shown in Figure 1). Each copolymer in the design space contained 2–6 monomers at 5 wt % compositional intervals. Some rational design was employed in further defining the design space; each polymer was at least 10 wt % cationic and 10 wt % hydrophobic. For the uncertainty and exploitation sampling, the total concentration of di-Boc-guanidinium ethyl acrylamide (Gea) was limited to minimize unintended crosslinking, and Gea and Bmam were not copolymerized due to the chemical incompatibility of Bmam with the Gea deprotection reaction. In total, over 1.7 million unique compositions were utilized.

Diversity sampling

We employed MaxMin diversity sampling to ensure broad coverage of the chemical space. The algorithm maximizes the minimum pairwise distance between selected polymers in the molecular descriptor space derived from RDKit molecular fingerprints. Distance calculations used Tanimoto similarity coefficients,

with selection prioritizing compositions that maximize the minimum distance to previously selected samples.

Query-by-committee uncertainty sampling

We implemented a query-by-committee (QBC) approach using ensemble disagreement to quantify prediction uncertainty. Three independent MPNN models were trained with different random initializations and hyperparameter configurations to create a diverse ensemble.

For each polymer composition, model disagreement was quantified using KL divergence between individual model predictions and the ensemble mean. The ensemble mean positive and negative probabilities were calculated across all models, with a tolerance factor of 1×10^{-6} applied to prevent numerical instabilities from $\log(0)$ calculations.

The KL divergence for each model's prediction p relative to the ensemble mean was computed as:

$$KL(p||q) = p \cdot \log_2(p / q_{\text{pos}}) + (1 - p) \cdot \log_2((1 - p) / q_{\text{neg}}),$$

where q_{pos} and q_{neg} represent the ensemble mean positive and negative probabilities, respectively. The final uncertainty score was the mean KL divergence across all ensemble models.

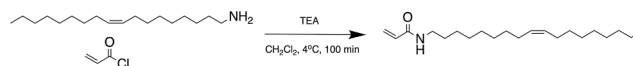
To focus on genuinely uncertain predictions while avoiding statistical outliers, we applied quartile-based filtering. Polymers were ranked by the KL divergence scores, and only those from the fourth quartile (Q4) were considered for selection. Outliers were removed using the 1.5 times the interquartile range ($1.5 \times \text{IQR}$) method, excluding samples beyond $Q3 + 1.5 \times \text{IQR}$. From this filtered set of high-uncertainty candidates, diversity sampling was applied using molecular descriptors to ensure broad chemical space coverage within the uncertain region.

Exploitation sampling

For exploitation sampling, we employed the same ensemble approach but focused on extracting the probability of polymers belonging to the lowest MIC bin ($<100 \mu\text{g/mL}$) from multi-class predictions. We selected polymers with high predicted probability for the lowest MIC class (>0.8) and low KL divergence scores (<0.1), indicating strong ensemble consensus for high-efficacy predictions.

Synthesis of *N*-octylacrylamide

N-octylacrylamide (Oct) was prepared by an adaptation of previously described methods (Scheme 1).^{26,65} Briefly, *N*-octylamine (992 μL , 6 mmol) and TEA (920 μL , 6.6 mmol) were dissolved in anhydrous DCM (60 mL) and cooled in an ice bath. The mixture was placed under a nitrogen atmosphere and stirred as acryloyl chloride (536 μL , 6.6 mmol) was added dropwise over 1 h. The mixture was stirred for an additional 40 min. It was then washed with saturated ammonium chloride (100 mL), followed by washing with saturated sodium bicarbonate (100 mL) and brine (100 mL). The product was dried over anhydrous magnesium sulfate, and the solvent was removed using a rotary evaporator. The product (Oct) was isolated as a waxy white solid with 97% yield. NMR spectra were recorded on a Bruker Avance Neo 500 MHz



Scheme 2. Synthesis of oyleylacrylamide

NMR spectrometer by collecting 32 scans with a 2 s relaxation delay.

Synthesis of oyleylacrylamide

Oyleylacrylamide (Olam) was prepared using previously reported methods (Scheme 2).²⁶ Briefly, oleylamine (1.974 mL, 6 mmol) and TEA (920 μ L, 6.6 mmol) were dissolved in anhydrous DCM (60 mL) and cooled in an ice bath. The mixture was placed under a nitrogen atmosphere and stirred as acryloyl chloride (536 μ L, 6.6 mmol) was added dropwise over 1 h. The mixture was stirred for an additional 40 min. It was then washed with saturated ammonium chloride (100 mL), followed by washing with saturated sodium bicarbonate (100 mL) and brine (100 mL). The product was dried over anhydrous magnesium sulfate, and the solvent was removed by rotary evaporator. The product (Olam) was isolated as a pale yellow oil with 87% yield. NMR spectra were recorded on a Bruker Avance Neo 500 MHz spectrometer by collecting 32 scans with a 2 s relaxation delay.

Synthesis of Gea

Di-Boc-guanidinium ethyl acrylamide (Gea) was prepared by an adaptation of previously reported methods (Scheme 3).⁶⁶ Briefly, 1,3-bis(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (7.260 g, 25 mmol, 1 eq.) in DCM (50 mL) was added dropwise over 45 min to a solution of ethylenediamine (5.0 mL, 75 mmol, 3 eq.) in DCM (50 mL) in an oven-dried 250 mL round-bottom flask, with stirring at 500 rpm. Following complete addition, the reaction was left to stir at room temperature (RT) for 4 h. The solution was washed with H₂O (2 $\times$ 50 mL) and brine (1 $\times$ 50 mL), the organic layer was dried over Na₂SO₄, and the DCM was removed under vacuum to yield crude 2-[1,3-bis(*tert*-butoxycarbonyl)guanidine]ethylamine as a white solid (7.07 g, 93% yield). This intermediate was characterized using ¹H NMR (Figure S2).

2-[1,3-Bis(*tert*-butoxycarbonyl)guanidine]ethylamine (1.5 g, 4.96 mmol, 1 eq.) was transferred to an oven-dried 100 mL round-bottom flask (RBF) equipped with a stir bar. TEA (761 μ L, 5.46 mmol, 1.1 eq.) was added, and the solution was cooled in an ice bath for 15 min while stirring. The reaction mixture was equipped with an N₂ balloon. Acryloyl chloride (403 μ L, 4.96 mmol, 1 eq.) was diluted in 1 mL DCM, and this solution was added dropwise by syringe. The balloon was removed. The reaction was covered with aluminum foil and allowed to stir and warm slowly to RT overnight with stirring at 400 rpm. The mixture was washed with saturated NaHCO₃. The aqueous layer was extracted with 50 mL DCM. The organic

layers were combined and dried over Na₂SO₄, and the solvent was removed under vacuum, yielding Gea as an off-white solid (1.29 g, 73% yield). The product was characterized using ¹H NMR (Figure S3).

Polymerization conditions

Statistical copolymerizations were carried out by RAFT polymerization. For each polymer, [CTA]/[monomers] = 70. [AIBN]/[CTA] was 0.2 or 0.5 for each entry. Generally, [AIBN]/[CTA] = 0.5 for entries containing Morph, and [AIBN]/[CTA] = 0.2 for other copolymers. The monomers (total mass of 0.5 g), CTA, and initiator were dissolved in a mixture of DMF and water to a final volume of 4 mL in a 20 mL scintillation vial equipped with a polytetrafluoroethylene/silicone septum. The mixture was sparged with nitrogen for 10 min and heated at 65°C overnight with stirring. The mixture was cooled to RT and exposed to air.

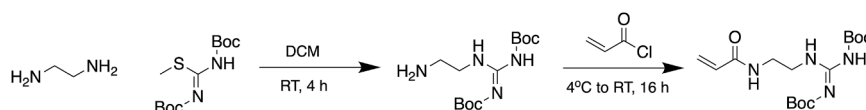
Polymers containing Gea were deprotected to remove the Boc groups by mixing the crude polymerization mixture with TFA (750 μ L). The mixture was equipped with a vent needle and allowed to stir at 300 rpm at RT for 1–2 h.

Polymers were precipitated two to three times in acetone or in a combination of ether and hexane. The precipitated polymers were flash-frozen and dried by lyophilization.

Polymer characterization

NMR spectra were recorded on a Bruker Avance Neo 500 MHz spectrometer by collecting 128 scans with a 5 s relaxation delay. *M_w* and dispersity values were determined via GPC after passing the samples through a size exclusion column (Resolve Mixed Bed Low divinylbenzene [DVB] [Jordi Labs]) in a mobile phase of DMF with 1 wt % NaBF₄ at 50°C and a flow rate of 1.0 mL/min (Thermo Fisher Scientific Dionex UltiMate 3000 pump, degasser, and auto-sampler). PEG standards were implemented for each polymer, except for Gea₁₀Olam₃₀Mo₆₀ and Tma₁₅Gea₁₀Ni₅Do₂₅Tmb₁₀Mep₃₅, for which a PMMA standard was applied to account for their relatively low cationic density and smaller radius of gyration in the ionic solvent.

To computationally probe the compositional dispersity of top candidate polymers, a previously described program, Compositional Drift, was utilized.⁶⁴ Briefly, this program predicted composition for a sample of 1,500 simulated polymer chains from a pool of 200,000 monomers, using the Mayo-Lewis model of monomer addition. Reactivity ratios of each monomer were set to 1. Although the actual reactivity ratios may deviate from 1, variation would impact the polymer sequence (i.e., forming a gradient rather than a random copolymer) rather than the overall



Scheme 3. Synthesis of di-Boc-guanidinium ethyl acrylamide

compositional dispersity of the polymer, particularly for high conversion polymers. Here, we used the program to predict overall compositional dispersity (comparing one polymer chain with another), rather than sequence (comparing the initiation end with the termination end within a single polymer chain); so, the precise reactivity ratios are less critical. The predicted set of polymers was analyzed to determine the coefficient of variance in the molar ratio of each monomer.

MIC assay

The MIC of each polymer was determined using the broth micro-dilution method, adapted from the Clinical and Laboratory Standards Institute (CLSI) guidelines.⁶⁷ Briefly, *E. coli* ATCC 25922, *S. aureus* ATCC 29213, and *K. pneumoniae* ATCC 13884 were each cultured in MHB overnight at 37°C with shaking at 200 rpm. Overnight cultures were diluted based on optical density at 600 nm (OD₆₀₀) measurements and corresponding CFUs/mL for each strain, which yielded cell densities of approximately 1×10^6 CFUs/mL. A 2-fold dilution series of 50 μ L of each polymer solution at desired concentrations as indicated in MHB was prepared from concentrated 20.48 mg/mL stocks of each polymer in a 96-well, U-bottom polypropylene plate. Subsequently, 50 μ L of each density-normalized bacterial culture was added to each well to achieve a final concentration of $\sim 5 \times 10^5$ CFUs/mL. All MIC determinations were performed in duplicate. Each plate contained nutrient medium-only controls and growth controls (without polymer). Assay plates were incubated statically at 37°C for 18–24 h. After incubation, each well was mixed by pipetting to resuspend bacteria, and 75 μ L from each well was transferred to a clear, flat-bottom 96-well polystyrene plate. Cell growth was quantified by OD₆₀₀ values measured on a BioTek Synergy H1 microplate reader. The MIC was defined as the lowest concentration of polymer before OD₆₀₀ increased by at least 0.1 units.

Hemolysis assay

Red blood cells (RBCs) were collected by centrifugation of whole blood at $500 \times g$ for 5 min. The plasma was removed, and the RBCs were washed twice with 150 mM sodium chloride and once with PBS. They were pelleted again and resuspended in PBS. The sample was diluted 1:50 in PBS. To each well of a 96-well plate was added 180 μ L of diluted RBCs and 20 μ L of antimicrobial agent stock solution at 10 $\times$ the desired concentration. The samples were run in duplicate. Triton X-100 at a final concentration of 1% was used as a positive control, and water was used as a negative control. The plate was incubated statically at 37°C for 1 h. Intact RBCs were pelleted by centrifugation at $1,000 \times g$ for 5 min, using a Thermo Scientific Sorvall Legend XTR centrifuge. 100 μ L of the supernatant was transferred from each well to a clear, flat-bottom 96-well plate, and the absorbance was measured at 540 nm. The percentage of hemolysis was calculated as follows:

$$\text{Hemolysis (\%)} = \frac{[(\text{Abs}_{450 \text{ nm}} \text{ of the treated sample} - \text{Abs}_{450 \text{ nm}} \text{ of the negative control}) / (\text{Abs}_{450 \text{ nm}} \text{ of positive control} - \text{Abs}_{450 \text{ nm}} \text{ of negative control})] \times 100\%.$$

The HC₅₀ was determined through interpolation after fitting the percentage of hemolysis to a five-parameter asymmetric sigmoidal function using GraphPad Prism software.

Fluorescence-based membrane integrity assay

The integrity of the cell membranes of *E. coli* in the presence of antimicrobial agents was evaluated through an adaptation of previously reported methods.^{68,69} Briefly, a stationary phase culture of *E. coli* ATCC 25922 was diluted 1:25 in MHB and grown at 37°C with shaking at 200 rpm until achieving the OD₆₀₀ $\approx$ 0.6. The bacteria were subsequently pelleted by centrifugation at 10,000 $\times g$ for 5 min at room temperature, using a Thermo Scientific Sorvall Legend Micro 21R centrifuge. They were washed once with HEPES/glucose buffer (5 mM HEPES and 5 mM glucose). They were pelleted again and resuspended to a final optical density of 0.1 in HEPES/glucose buffer. The cells were aliquoted into a black-walled 96-well plate, and the fluorescent probe Sytox Green was added to a final concentration of 1 μ M. The fluorescence was measured in a microplate for 5 min (excitation 505 nm; emission 523 nm). The plate was removed, and the antimicrobial agents were added at the desired concentrations. The fluorescence was measured for 1 h at room temperature. Statistical analysis comparing the fluorescence curve for each polymer was conducted with GraphPad Prism software using a two-way ANOVA.

Light microscopy

Phase-contrast light microscopy was performed as previously reported,⁷⁰ using the *E. coli* strain ATCC 25922. Briefly, the cells were grown to stationary phase in MHB medium at 37°C. To obtain exponentially growing cells, the culture was then diluted by at least 10,000-fold in fresh MHB and grown at 37°C until it reached an OD₆₀₀ of 0.2–0.4.

Polymer solutions were prepared by diluting a 20 mg/mL stock solution in fresh MHB and mixed 1:1 with the cell solution by pipetting, yielding a final polymer concentration of $2 \times \text{MIC}$ and OD₆₀₀ = 0.1. As a control, the same step was performed with fresh MHB instead of the polymer solution. The samples were incubated for 30 min at 37°C in a thermomixer (500 rpm), and 0.5 μ L was spotted onto a 1% (w/v) agarose pad. This pad was prepared by adding agarose powder in fresh MHB. The suspension was then gently warmed to dissolve agarose. Exposure time for phase-contrast images was 200 ms with 100% diascopic light intensity, and images were acquired at RT.

MBEC assay

E. coli ATCC 25922 biofilms were formed by overnight static incubation at 37°C of 100 μ L of inoculum (1×10^7 CFUs/mL in M63 minimal media with 0.8% w/v glucose) in the inner wells of a sterile polyvinyl chloride (PVC) plate. To limit evaporation and ensure biofilm formation consistency, 150 μ L of PBS was added to the outer wells of each plate. After 24 h of incubation, the nutrient medium was carefully removed from the inner wells by using a pipette, and the wells were rinsed three times with 120 μ L of sterile PBS to remove planktonic cells. After washing, 150 μ L treatment medium containing 2-fold serial dilutions of the antimicrobial compound in M63 + 0.8% glucose was added to the plate. The treatment medium was prepared by spiking M63 + 0.8%

glucose with a concentrated stock solution of respective agents prepared as follows: Tma₇₅Do₂₅ was prepared at 10.24 mg/mL in water, Tma₅₉Do₃₁Mep₁₀ at 10.24 mg/mL in 50:50 water/DMSO, and TMP-SMX in a 1:5 ratio at 30 mg/mL in DMSO. Each plate was incubated statically for an additional 24 h at 37°C. The treatment medium was then removed, and the wells were rinsed three times with 160 µL PBS. Fresh recovery medium (150 µL M63 with 0.8% glucose) was added, and the plate was incubated statically for an additional 24 h at 37°C. Then, 15 µL of 0.02% resazurin solution was added to each well and incubated for 30 min. The MBEC value was defined as the lowest treatment concentration where no bacterial growth occurred, as determined visually according to the color change of the resazurin dye.

Cytotoxicity assay

NIH3T3 cells were purchased from Sigma-Aldrich and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (Pen-Strep). The cells were maintained at 37°C with 5% CO₂ and handled under sterile conditions. A CCK-8 colorimetric kit was used to quantify the viability of cells following exposure to antimicrobial polymers. The cells were plated at a concentration of 10,000 cells per well in a clear, flat-bottom, 96-well plate and allowed to adhere for 24 h. Stock solutions of each tested compound were prepared in media and added to the cells to achieve a total volume of 100 µL per well. After 5 h of incubation, 10 µL of CCK-8 solution was added to each well. After 3 h, the absorbance was read at 450 nm, and background was read at 650 nm. The baseline-corrected absorbance (Abs_{450nm-650nm}) was calculated by subtracting the background absorbance at 650 nm from the absorbance at 450 nm. The percent viability was determined as follows:

$$\text{Viability (\%)} = \left[\frac{(\text{Abs}_{450\text{nm}-650\text{nm}} \text{ of sample} - \text{Abs}_{450\text{nm}-650\text{nm}} \text{ of media})}{(\text{Abs}_{450\text{nm}-650\text{nm}} \text{ of untreated control} - \text{Abs}_{450\text{nm}-650\text{nm}} \text{ of media})} \right] \times 100\%$$

LC₅₀ values were determined through interpolation after fitting the data to a four-parameter agonist vs. response function using GraphPad Prism.

In vivo toxicity assay

C57BL/6 mice (10 weeks old, female) were injected intraperitoneally or dosed intranasally with 400, 100, or 0 µg Tma₇₅Do₂₅ in 100 µL sterile PBS (i.p.) or 20 µL PBS (i.n.) (*n* = 5). The mice were weighed daily, and cardiac punctures were performed 5 days after injection to obtain blood for chemistry analysis, which was performed by the Stanford Animal Diagnostic Lab.

RESOURCE AVAILABILITY

Lead contact

Requests for further information, resources, and materials should be directed to and will be fulfilled by the lead contact, Eric A. Appel (eappel@stanford.edu).

Materials availability

All unique, stable reagents generated in this study are available from the lead contact, with a completed material(s) transfer agreement.

Data and code availability

Data generated to evaluate the conclusions of this article are provided within the article and [supplemental information](#). Raw data will be made available from the [lead contact](#) upon request. Original code is available at https://github.com/GabrielGreenstein01/polymer_gnn.

ACKNOWLEDGMENTS

This work was funded by the Stanford Bio-X Interdisciplinary Initiative Seed Grants Program (IIP). Research reported in this publication was also supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award number R35GM156332 (to L.C.). Part of this work was performed at the Stanford Nano Shared Facilities, supported by the National Science Foundation (NSF) under award ECCS-2026822. This work involved the use of the Stanford ChEM-H shared facility, enabled by the NIH High End Instrumentation grant (1 S10 OD028697-01). S.C.W. was supported by the Sarafan ChEM-H Chemistry/Biology Interface training program and the NSF GRFP. C.J.-W. is an investigator of the Howard Hughes Medical Institute. The authors gratefully acknowledge Dr. Emily Meany and Dr. Jerry Yan for providing murine blood for hemolysis assays. The authors also thank Dr. Carolyn Jons and Arjun Gupta for their advice regarding statistical characterization and implementation of the machine learning pipeline, respectively.

AUTHOR CONTRIBUTIONS

S.C.W., H.L.H., and E.A.A conceptualized the study, with contributions from L.C. and C.J.-W.; H.L.H, G.G., and N.E. developed the active learning pipeline; S.C.W., A.M.-L., A.F., X.L., P.G., C.D., and A.N.P. conducted the experiments; and S.C.W. wrote the original draft.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE MANUSCRIPT PREPARATION PROCESS

Claude Code was used to review code and provide an outline the methods sections pertaining to the ML-related work. The writing was then filled in and refined by the authors. The authors take full responsibility for the content of the publication.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.matt.2026.103026>.

Received: November 28, 2025

Revised: August 10, 2026

Accepted: August 25, 2026

REFERENCES

1. Naghavi, M., Vollset, S.E., Ikuta, K.S., Swetschinski, L.R., Gray, A.P., Wool, E.E., Robles Aguilar, G., Mestrovic, T., Smith, G., Han, C., et al. (2024). Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet* 404, 1199–1226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1).
2. Nelson, R.E., Hatfield, K.M., Wolford, H., Samore, M.H., Scott, R.D., II, Reddy, S.C., Olubajo, B., Paul, P., Jernigan, J.A., and Baggs, J. (2021). National Estimates of Healthcare Costs Associated With Multidrug-Resistant Bacterial Infections Among Hospitalized Patients in the United States. *Clin. Infect. Dis.* 72, S17–S26. <https://doi.org/10.1093/cid/ciaa1581>.

3. Murray, C.J.L., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., et al. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 399, 629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
4. Delcour, A.H. (2009). Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1794, 808–816. <https://doi.org/10.1016/j.bbapap.2008.11.005>.
5. Blair, J.M.A., Webber, M.A., Baylay, A.J., Ogbolu, D.O., and Piddock, L.J.V. (2015). Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* 13, 42–51. <https://doi.org/10.1038/nrmicro3380>.
6. Munita, J.M., and Arias, C.A. (2016). Mechanisms of Antibiotic Resistance. *Microbiol. Spectr.* 4, 4.2.15. <https://doi.org/10.1128/microbiolspec.VMBF-0016-2015>.
7. Zhang, Q.-Y., Yan, Z.-B., Meng, Y.-M., Hong, X.-Y., Shao, G., Ma, J.-J., Cheng, X.-R., Liu, J., Kang, J., and Fu, C.-Y. (2021). Antimicrobial peptides: mechanism of action, activity and clinical potential. *Milit. Med. Res.* 8, 48. <https://doi.org/10.1186/s40779-021-00343-2>.
8. Wu, M., Maier, E., Benz, R., and Hancock, R.E.W. (1999). Mechanism of Interaction of Different Classes of Cationic Antimicrobial Peptides with Planar Bilayers and with the Cytoplasmic Membrane of *Escherichia coli*. *Biochemistry* 38, 7235–7242. <https://doi.org/10.1021/bi9826299>.
9. Elliott, A.G., Huang, J.X., Neve, S., Zuegg, J., Edwards, I.A., Cain, A.K., Boinett, C.J., Barquist, L., Lundberg, C.V., Steen, J., et al. (2020). An amphipathic peptide with antibiotic activity against multidrug-resistant Gram-negative bacteria. *Nat. Commun.* 11, 3184. <https://doi.org/10.1038/s41467-020-16950-x>.
10. Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature* 415, 389–395. <https://doi.org/10.1038/415389a>.
11. Fjell, C.D., Hiss, J.A., Hancock, R.E.W., and Schneider, G. (2012). Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* 11, 37–51. <https://doi.org/10.1038/nrd3591>.
12. Paterson, D.J., Tassieri, M., Reboud, J., Wilson, R., and Cooper, J.M. (2017). Lipid topology and electrostatic interactions underpin lytic activity of linear cationic antimicrobial peptides in membranes. *Proc. Natl. Acad. Sci.* 114, E8324–E8332. <https://doi.org/10.1073/pnas.1704489114>.
13. Tavares, L.S., Silva, C., dos, S.F. da, Souza, V.C., Silva, V.L. da, Diniz, C.G., and Santos, M.D.O. (2013). Strategies and molecular tools to fight antimicrobial resistance: resistome, transcriptome, and antimicrobial peptides. *Front. Microbiol.* 4. <https://doi.org/10.3389/fmicb.2013.00412>.
14. Afacan, N.J., Yeung, A.T.Y., Pena, O.M., and Hancock, R.E.W. (2012). Therapeutic potential of host defense peptides in antibiotic-resistant infections. *Curr. Pharm. Des.* 18, 807–819. <https://doi.org/10.2174/138161212799277617>.
15. Hancock, R.E.W., and Sahl, H.-G. (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* 24, 1551–1557. <https://doi.org/10.1038/nbt1267>.
16. Marr, A.K., Gooderham, W.J., and Hancock, R.E. (2006). Antibacterial peptides for therapeutic use: obstacles and realistic outlook. *Curr. Opin. Pharmacol.* 6, 468–472. <https://doi.org/10.1016/j.coph.2006.04.006>.
17. Chen, C.H., and Lu, T.K. (2020). Development and Challenges of Antimicrobial Peptides for Therapeutic Applications. *Antibiotics* 9, 24. <https://doi.org/10.3390/antibiotics9010024>.
18. Palermo, E.F., and Kuroda, K. (2010). Structural determinants of antimicrobial activity in polymers which mimic host defense peptides. *Appl. Microbiol. Biotechnol.* 87, 1605–1615. <https://doi.org/10.1007/s00253-010-2687-z>.
19. Ilker, M.F., Nüsslein, K., Tew, G.N., and Coughlin, E.B. (2004). Tuning the Hemolytic and Antibacterial Activities of Amphiphilic Polynorbornene Derivatives. *J. Am. Chem. Soc.* 126, 15870–15875. <https://doi.org/10.1021/ja045664d>.
20. Tew, G.N., Liu, D., Chen, B., Doerksen, R.J., Kaplan, J., Carroll, P.J., Klein, M.L., and DeGrado, W.F. (2002). De novo design of biomimetic antimicrobial polymers. *Proc. Natl. Acad. Sci.* 99, 5110–5114. <https://doi.org/10.1073/pnas.082046199>.
21. Scott, R.W., DeGrado, W.F., and Tew, G.N. (2008). De novo designed synthetic mimics of antimicrobial peptides. *Curr. Opin. Biotechnol.* 19, 620–627. <https://doi.org/10.1016/j.copbio.2008.10.013>.
22. Lienkamp, K., and Tew, G.N. (2009). Synthetic Mimics of Antimicrobial Peptides—A Versatile Ring-Opening Metathesis Polymerization Based Platform for the Synthesis of Selective Antibacterial and Cell-Penetrating Polymers. *Chem. – Eur. J.* 15, 11784–11800. <https://doi.org/10.1002/chem.200900049>.
23. Som, A., Vemparala, S., Ivanov, I., and Tew, G.N. (2008). Synthetic mimics of antimicrobial peptides. *Pept. Sci.* 90, 83–93. <https://doi.org/10.1002/bip.20970>.
24. Thaker, H.D., Sgolastra, F., Clements, D., Scott, R.W., and Tew, G.N. (2011). Synthetic Mimics of Antimicrobial Peptides from Triaryl Scaffolds. *J. Med. Chem.* 54, 2241–2254. <https://doi.org/10.1021/jm101410t>.
25. Judzewitsch, P.R., Zhao, L., Wong, E.H.H., and Boyer, C. (2019). High-Throughput Synthesis of Antimicrobial Copolymers and Rapid Evaluation of Their Bioactivity. *Macromolecules* 52, 3975–3986. <https://doi.org/10.1021/acs.macromol.9b00290>.
26. Williams, S.C., Chosy, M.B., Jons, C.K., Dong, C., Prossnitz, A.N., Liu, X., Hernandez, H.L., Cegelski, L., and Appel, E.A. (2025). Polyacrylamide-Based Antimicrobial Copolymers to Replace or Rescue Antibiotics. *ACS Cent. Sci.* 11, 486–496. <https://doi.org/10.1021/acscentsci.4c01973>.
27. Yang, J., Tao, L., He, J., McCutcheon, J.R., and Li, Y. (2022). Machine learning enables interpretable discovery of innovative polymers for gas separation membranes. *Sci. Adv.* 8, eabn9545. <https://doi.org/10.1126/sciadv.abn9545>.
28. Yu, J., Yao, D., Wang, L., and Xu, M. (2025). Machine Learning in Predicting and Optimizing Polymer Printability for 3D Bioprinting. *Polymers* 17, 1873. <https://doi.org/10.3390/polym17131873>.
29. Bradford, G., Lopez, J., Ruza, J., Stolberg, M.A., Osterude, R., Johnson, J.A., Gomez-Bombarelli, R., and Shao-Horn, Y. (2023). Chemistry-Informed Machine Learning for Polymer Electrolyte Discovery. *ACS Cent. Sci.* 9, 206–216. <https://doi.org/10.1021/acscentsci.2c01123>.
30. Ting, J.M., Tamayo-Mendoza, T., Petersen, S.R., Van Reet, J., Ahmed, U.A., Snell, N.J., Fisher, J.D., Stern, M., and Oviedo, F. (2023). Frontiers in nonviral delivery of small molecule and genetic drugs, driven by polymer chemistry and machine learning for materials informatics. *Chem. Commun.* 59, 14197–14209. <https://doi.org/10.1039/D3CC04705A>.
31. Zhao, Y., Mulder, R.J., Houshyar, S., and Le, T.C. (2023). A review on the application of molecular descriptors and machine learning in polymer design. *Polym. Chem.* 14, 3325–3346. <https://doi.org/10.1039/D3PY00395G>.
32. Xu, C., Wang, Y., and Barati Farmani, A. (2023). TransPolymer: a Transformer-based language model for polymer property predictions. *npj Comput. Mater.* 9, 64. <https://doi.org/10.1038/s41524-023-01016-5>.
33. Martin, T.B., and Audus, D.J. (2023). Emerging Trends in Machine Learning: A Polymer Perspective. *ACS Polym. Au* 3, 239–258. <https://doi.org/10.1021/acspolymersau.2c00053>.
34. Kundi, V., Jin, Y., Chandrasekaran, A., Muralidhar, A., Masood, H., Boyer, C., and Kumar, P. (2024). Machine Learning Prediction of Antibacterial Activity of Block Copolymers. *ACS Appl. Nano Mater.* 7, 8939–8948. <https://doi.org/10.1021/acsnanm.4c00430>.
35. Zhao, Y., Mulder, R.J., Eyckens, D.J., Houshyar, S., and Le, T.C. (2024). Advancing antimicrobial polymer development: a novel database and accelerated design via machine learning. *Polym. Chem.* 15, 4063–4076. <https://doi.org/10.1039/D4PY00736K>.
36. Wang, J., Feng, J., Kang, Y., Pan, P., Ge, J., Wang, Y., Wang, M., Wu, Z., Zhang, X., Yu, J., et al. (2025). Discovery of antimicrobial peptides with notable antibacterial potency by an LLM-based foundation model. *Sci. Adv.* 11, eads8932. <https://doi.org/10.1126/sciadv.ads8932>.
37. Santos-Júnior, C.D., Torres, M.D.T., Duan, Y., Rodríguez del Río, Á., Schmidt, T.S.B., Chong, H., Fullam, A., Kuhn, M., Zhu, C., Houseman,

- A., et al. (2024). Discovery of antimicrobial peptides in the global microbiome with machine learning. *Cell* 187, 3761–3778.e16. <https://doi.org/10.1016/j.cell.2024.05.013>.
38. Wan, F., Torres, M.D.T., Peng, J., and de la Fuente-Nunez, C. (2024). Deep-learning-enabled antibiotic discovery through molecular de-extinction. *Nat. Biomed. Eng.* 8, 854–871. <https://doi.org/10.1038/s41551-024-01201-x>.
39. Wan, F., Wong, F., Collins, J.J., and de la Fuente-Nunez, C. (2024). Machine learning for antimicrobial peptide identification and design. *Nat. Rev. Bioeng.* 2, 392–407. <https://doi.org/10.1038/s44222-024-00152-x>.
40. de la Lastra, J.M.P., Wardell, S.J.T., Pal, T., de la Fuente-Nunez, C., and Pletzer, D. (2024). From Data to Decisions: Leveraging Artificial Intelligence and Machine Learning in Combating Antimicrobial Resistance – a Comprehensive Review. *J. Med. Syst.* 48, 71. <https://doi.org/10.1007/s10916-024-02089-5>.
41. Pane, K., Durante, L., Crescenzi, O., Cafaro, V., Pizzo, E., Varcamonti, M., Zanfardino, A., Izzo, V., Di Donato, A., and Notomista, E. (2017). Antimicrobial potency of cationic antimicrobial peptides can be predicted from their amino acid composition: Application to the detection of “cryptic” antimicrobial peptides. *J. Theor. Biol.* 419, 254–265. <https://doi.org/10.1016/j.jtbi.2017.02.012>.
42. Pham, P., Oliver, S., and Boyer, C. (2023). Design of Antimicrobial Polymers. *Macromol. Chem. Phys.* 224, 2200226. <https://doi.org/10.1002/macp.202200226>.
43. Colak, S., Nelson, C.F., Nüsslein, K., and Tew, G.N. (2009). Hydrophilic Modifications of an Amphiphilic Polynorbornene and the Effects on its Hemolytic and Antibacterial Activity. *Biomacromolecules* 10, 353–359. <https://doi.org/10.1021/bm801129y>.
44. Pham, P., Oliver, S., Wong, E.H.H., and Boyer, C. (2021). Effect of hydrophilic groups on the bioactivity of antimicrobial polymers. *Polym. Chem.* 12, 5689–5703. <https://doi.org/10.1039/D1PY01075A>.
45. Pham, P., Oliver, S., Nguyen, D.T., and Boyer, C. (2022). Effect of Cationic Groups on the Selectivity of Ternary Antimicrobial Polymers. *Macromol. Rapid Commun.* 43, 2200377. <https://doi.org/10.1002/marc.202200377>.
46. Yang, Y., Cai, Z., Huang, Z., Tang, X., and Zhang, X. (2018). Antimicrobial cationic polymers: from structural design to functional control. *Polym. J.* 50, 33–44. <https://doi.org/10.1038/pj.2017.72>.
47. Yin, L.-H., Ran, B., Hu, T.-J., Yang, C., Fei, J.-J., and Li, Y.-H. (2017). Preparation of highly efficient antibacterial polymeric films via the modulation of charge density and hydrophobicity. *RSC Adv.* 7, 6006–6012. <https://doi.org/10.1039/C6RA26071C>.
48. Engler, A.C., Tan, J.P.K., Ong, Z.Y., Coady, D.J., Ng, V.W.L., Yang, Y.Y., and Hedrick, J.L. (2013). Antimicrobial Polycarbonates: Investigating the Impact of Balancing Charge and Hydrophobicity Using a Same-Centered Polymer Approach. *Biomacromolecules* 14, 4331–4339. <https://doi.org/10.1021/bm401248t>.
49. Phuong, P.T., Oliver, S., He, J., Wong, E.H.H., Mathers, R.T., and Boyer, C. (2020). Effect of Hydrophobic Groups on Antimicrobial and Hemolytic Activity: Developing a Predictive Tool for Ternary Antimicrobial Polymers. *Biomacromolecules* 21, 5241–5255. <https://doi.org/10.1021/acs.biomac.0c01320>.
50. Witten, Z. (2025). Zswitten/Antimicrobial-Peptides.
51. UniProt UniProt. <https://www.uniprot.org>.
52. Patrizi, M.L., Diociaiuti, M., Capitani, D., and Masci, G. (2009). Synthesis and association properties of thermoresponsive and permanently cationic charged block copolymers. *Polymer* 50, 467–474. <https://doi.org/10.1016/j.polymer.2008.11.023>.
53. Schaefer, S., Pham, T.T.P., Brunke, S., Hube, B., Jung, K., Lenardon, M.D., and Boyer, C. (2021). Rational Design of an Antifungal Polyacrylamide Library with Reduced Host-Cell Toxicity. *ACS Appl. Mater. Interfaces* 13, 27430–27444. <https://doi.org/10.1021/acsami.1c05020>.
54. Nguyen, T.-K., Lam, S.J., Ho, K.K.K., Kumar, N., Qiao, G.G., Egan, S., Boyer, C., and Wong, E.H.H. (2017). Rational Design of Single-Chain Polymeric Nanoparticles That Kill Planktonic and Biofilm Bacteria. *ACS Infect. Dis.* 3, 237–248. <https://doi.org/10.1021/acsinfecdis.6b00203>.
55. Chin, W., Zhong, G., Pu, Q., Yang, C., Lou, W., De Sessions, P.F., Periaswamy, B., Lee, A., Liang, Z.C., Ding, X., et al. (2018). A macromolecular approach to eradicate multidrug resistant bacterial infections while mitigating drug resistance onset. *Nat. Commun.* 9, 917. <https://doi.org/10.1038/s41467-018-03325-6>.
56. Namivandi-Zangeneh, R., Kwan, R.J., Nguyen, T.-K., Yeow, J., Byrne, F.L., Oehlers, S.H., Wong, E.H.H., and Boyer, C. (2018). The effects of polymer topology and chain length on the antimicrobial activity and hemocompatibility of amphiphilic ternary copolymers. *Polym. Chem.* 9, 1735–1744. <https://doi.org/10.1039/C7PY01069A>.
57. Inactive Ingredient Search for Approved Drug Products <https://www.accessdata.fda.gov/scripts/cder/ig/index.cfm>.
58. Lewis, K. (2001). Riddle of Biofilm Resistance. *Antimicrob. Agents Chemother.* 45, 999–1007. <https://doi.org/10.1128/AAC.45.4.999-1007.2001>.
59. Mendhe, S., Badge, A., Ugemuge, S., and Chandi, D. (2023). Impact of Biofilms on Chronic Infections and Medical Challenges. *Cureus* 15, e48204. <https://doi.org/10.7759/cureus.48204>.
60. Nelson, Z., Aslan, A.T., Beahm, N.P., Blyth, M., Cappiello, M., Casaus, D., Dominguez, F., Egbert, S., Hanretty, A., Khadem, T., et al. (2024). Guidelines for the Prevention, Diagnosis, and Management of Urinary Tract Infections in Pediatrics and Adults: A WikiGuidelines Group Consensus Statement. *JAMA Netw. Open* 7, e2444495. <https://doi.org/10.1001/jama-networkopen.2024.44495>.
61. Sulfamethoxazole and Trimethoprim Injection USP.
62. Prossnitz, A.N., Nguyen, L.T., Eckman, N., Borkar, S., Tetef, S., Autzen, A.A.A., Fuller, G.G., and Appel, E.A. (2025). Defining Structure-Function Relationships of Amphiphilic Excipients Enables Rational Design of Ultra-Stable Biopharmaceuticals. *Adv. Sci.* 12, 2409604. <https://doi.org/10.1002/advs.202409604>.
63. Smith, A.A.A., Maikawa, C.L., Hernandez, H.L., and Appel, E.A. (2021). Controlling properties of thermogels by tuning critical solution behaviour of ternary copolymers. *Polym. Chem.* 12, 1918–1923. <https://doi.org/10.1039/D0PY01696A>.
64. Smith, A.A.A., Hall, A., Wu, V., and Xu, T. (2019). Practical Prediction of Heteropolymer Composition and Drift. *ACS Macro Lett.* 8, 36–40. <https://doi.org/10.1021/acsmacrolett.8b00813>.
65. Chou, B., Mirau, P., Jiang, T., Wang, S.-W., and Shea, K.J. (2016). Tuning Hydrophobicity in Abiotic Affinity Reagents: Polymer Hydrogel Affinity Reagents for Molecules with Lipid-like Domains. *Biomacromolecules* 17, 1860–1868. <https://doi.org/10.1021/acs.biomac.6b00296>.
66. Martin, L., Peltier, R., Kuroki, A., Town, J.S., and Perrier, S. (2018). Investigating Cell Uptake of Guanidinium-Rich RAFT Polymers: Impact of Comonomer and Monomer Distribution. *Biomacromolecules* 19, 3190–3200. <https://doi.org/10.1021/acs.biomac.8b00146>.
67. Lewis, J.S., Mathers, A.J., Bobenchik, A.M., Bryson, A.L., Campeau, S., Cullen, S.K., Dingle, T., Galas, M.F., Humphries, R.M., Kim, T.J., et al. (2024). CLSI M100: Performance Standards for Antimicrobial Susceptibility Testing (Clinical and Laboratory Standards Institute).
68. Antonoplis, A., Zang, X., Huttner, M.A., Chong, K.K.L., Lee, Y.B., Co, J.Y., Amieva, M.R., Kline, K.A., Wender, P.A., and Cegelski, L. (2018). A Dual-Function Antibiotic-Transporter Conjugate Exhibits Superior Activity in Sterilizing MRSA Biofilms and Killing Persister Cells. *J. Am. Chem. Soc.* 140, 16140–16151. <https://doi.org/10.1021/jacs.8b08711>.
69. Antonoplis, A., Zang, X., Wegner, T., Wender, P.A., and Cegelski, L. (2019). Vancomycin-Arginine Conjugate Inhibits Growth of Carbapenem-Resistant *E. coli* and Targets Cell-Wall Synthesis. *ACS Chem. Biol.* 14, 2065–2070. <https://doi.org/10.1021/acschembio.9b00565>.
70. Fragasso, A., Schlechtweg, T., Lin, W.-H., Barron, A.E., and Jacobs-Wagner, C. (2025). Time-resolved phenotyping at subcellular resolution reveals shared principles and key trade-offs across antimicrobial peptide activities. Preprint at bioRxiv. <https://doi.org/10.1101/2025.04.10.648262>.