

# Prompt2Peptide: Text-Guided Controllable Peptide Generation via Charge-Curriculum Optimization

Aayam Bansal

Kelis Lab, Massachusetts Institute of Technology

Boston, MA, USA

aayambansal@ieee.org

## Abstract

Designing antimicrobial peptides (AMPs) with precise biophysical properties remains a bottleneck in therapeutic lead discovery. Protein language models generate plausible sequences but lack interpretable property control, while rule-based filters discard most candidates. We present PROMPT2PEPTIDE, a controllable generation framework that maps natural-language descriptions (e.g., “cationic amphipathic helix”) to explicit biophysical target ranges—net charge, hydrophobic moment ( $\mu\text{H}$ ), GRAVY, amino-acid composition, and length—then optimizes sequences via a two-phase charge-curriculum simulated annealing search. Phase 1 applies charge-directed mutations to rapidly reach feasible charge windows; Phase 2 performs charge-neutral swaps to satisfy secondary objectives including amphipathy and hydrophobicity. We evaluate on five prompt families spanning cationic AMPs, acidic loops, hydrophobic  $\beta$ -sheets, polar linkers, and basic NLS-like motifs, generating 100 sequences per family across five random seeds. Prompt2Peptide achieves 75.4% average coverage (fraction meeting all constraints), 100% novelty (identity  $<0.70$  vs. 201 curated reference AMPs), and 56.4% Safety@Feasibility, with per-family coverage ranging from 84% (cationic helix) to 96% (acidic loop), while the hardest family (basic NLS) reaches only 11%. We compare against three baselines—single-phase SA, random genetic algorithm, and random-with-filter—and two ablations removing the curriculum structure. The random GA achieves higher average coverage (85.6%) but lower consistency across families; Prompt2Peptide provides stable performance across structurally diverse targets without family-specific tuning. All sequences, code, and evaluation are released for reproducibility.

## CCS Concepts

- Computing methodologies → Heuristic function construction;
- Theory of computation → Constraint and logic programming.

## Keywords

peptide generation, antimicrobial peptides, simulated annealing, curriculum optimization, biophysical constraints

## 1 Introduction

Antimicrobial peptides (AMPs) are a promising class of therapeutics for combating antibiotic-resistant pathogens [11]. Short peptides (8–30 residues) with tuned charge, amphipathy, and hydrophobicity can disrupt bacterial membranes, but the vast sequence space makes rational design difficult.

Computational approaches to peptide generation fall into two broad categories: generative models trained on known AMPs [1, 2, 8] and optimization-based methods that search sequence space against a defined objective.

Protein language models (PLMs) such as ESM-2 [10] capture evolutionary regularities and can score sequence plausibility, but they do not provide direct control over specific biophysical properties. Users cannot specify, for example, “generate a cationic amphipathic helix with charge +3 to +8 and high hydrophobic moment.” Conversely, rule-based filters [11] can enforce constraints post-hoc but discard the vast majority of candidates and offer no mechanism for guided exploration.

We address this gap with PROMPT2PEPTIDE, a framework that:

- (1) Parses natural-language prompts into explicit biophysical target ranges (charge,  $\mu\text{H}$ , GRAVY, composition, length);
- (2) Optimizes sequences via a **charge-curriculum simulated annealing** search with two phases: charge-directed (Phase 1) followed by charge-neutral secondary optimization (Phase 2);
- (3) Evaluates generated sequences with biologically grounded safety filters and standardized metrics (Coverage, Safety@Feasibility, Novelty, Diversity).

The prompt-to-property mapping uses keyword matching against a curated set of 35 prompt→range pairs (5 base families  $\times$  6–7 paraphrases each), derived from established AMP design terminology [11] with ranges drawn from published guidelines for therapeutic peptide properties.

Our key contribution is the charge-curriculum structure: by prioritizing net charge satisfaction before secondary properties, the optimizer avoids wasting iterations on sequences that will fail the most discriminating constraint. We show this is particularly effective for families like cationic amphipathic helices, where charge and amphipathy are tightly coupled.

## 2 Related Work

*Generative models and PLMs.* RNN generators [8], VAEs [2], and generative models with MD simulation [1] produce AMP candidates but lack fine-grained property control. PLMs such as ESM-2 [10], ProGen [7], and ProtGPT2 [4] score sequence plausibility but require additional machinery for constrained generation.

*Biophysical descriptors and optimization.* We use Kyte-Doolittle GRAVY [6] and Eisenberg’s  $\mu\text{H}$  [3] with **structure-aware** angular frequencies ( $100^\circ$  for  $\alpha$ -helices,  $\approx 180^\circ$  for

**Table 1: Prompt families and target constraint ranges.**

Family	Charge	$\mu\text{H}$	GRAVY	Len
Cationic helix	+3, +8	0.35, 1.0	-0.2, 0.6	12–18
Acidic loop	-3, 0	0.1, 0.4	-1.0, 0.0	10–14
Hydrophobic $\beta$	-1, +2	0.1, 0.3	0.5, 1.5	10–14
Polar linker	-1, +1	0.05, 0.25	-1.5, -0.3	8–12
Basic NLS	+5, +10	0.1, 0.4	-1.5, -0.5	6–10

$\beta$ -strands; Fig. 1). Simulated annealing [5] and genetic algorithms are standard metaheuristics; our contribution is the curriculum structure that phases charge optimization before secondary properties.

### 3 Method

#### 3.1 Prompt-to-Property Mapping

Given a natural-language prompt  $p$  (e.g., “cationic amphipathic helix”), we extract target property ranges  $\hat{\theta}_p = \{\hat{\tau}_c, \hat{\tau}_{\mu\text{H}}, \hat{\tau}_g, \hat{\tau}_\ell\}$  for charge, hydrophobic moment, GRAVY, and length via keyword matching against a curated vocabulary of AMP design terms. The vocabulary was constructed from AMP design literature [11] and covers five structural families (Table 1).

Each family is paraphrased 6–7 ways to test robustness; all paraphrases map to the same target ranges. The current implementation uses heuristic keyword parsing rather than a learned encoder; this works well for five families but may not generalize to arbitrary prompts.

#### 3.2 Structure-Aware Hydrophobic Moment

We compute  $\mu\text{H}$  with angular frequency matched to the target secondary structure:

$$\mu\text{H} = \frac{1}{N} \sqrt{\left( \sum_i H_i \sin(i\delta) \right)^2 + \left( \sum_i H_i \cos(i\delta) \right)^2} \quad (1)$$

where  $H_i$  is the Eisenberg hydrophobicity of residue  $i$ ,  $\delta = 100^\circ$  for  $\alpha$ -helices and  $\delta \approx 180^\circ$  for  $\beta$ -strands (Fig. 1). This correction is critical: applying helical periodicity to  $\beta$ -sheet targets yields misleading  $\mu\text{H}$  values.

#### 3.3 Charge-Curriculum Simulated Annealing

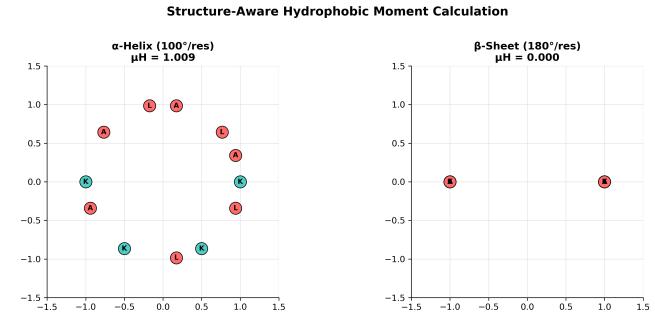
We optimize sequences by minimizing a composite penalty:

$$J(s; \hat{\theta}_p) = -w_c \phi_c(s, \hat{\tau}_c) - w_{\mu\text{H}} \phi_{\mu\text{H}}(s, \hat{\tau}_{\mu\text{H}}) \\ - w_g \phi_g(s, \hat{\tau}_g) - w_\ell \phi_\ell(s, \hat{\tau}_\ell) \quad (2)$$

where  $\phi$ . are hinge penalties for deviations outside target intervals. Simulated annealing with exponential cooling ( $T_0 = 1.0$ , decay = 0.995, 500 iterations) governs acceptance of proposals.

The **curriculum** structure is:

- **Phase 1** (iterations 1–300, 60%): Charge-directed mutations. Proposals preferentially substitute residues to move



**Figure 1: Structure-aware  $\mu\text{H}$  calculation. Helical periodicity (100°/residue, left) vs.  $\beta$ -strand periodicity ( $\approx 180^\circ/\text{residue}$ , right). Correct angular frequency is essential for meaningful amphiphathy scores.**

net charge toward the target range (e.g., replacing neutral residues with K/R for cationic targets).

- **Phase 2** (iterations 301–500, 40%): Charge-neutral swaps. Once charge is feasible, proposals swap residues of similar charge class to optimize  $\mu\text{H}$ , GRAVY, and length without disrupting charge satisfaction.

#### 3.4 Safety Filters

A generated sequence is classified as *feasible* if all property ranges are satisfied, and *safe* if it additionally passes seven biologically grounded filters:

- (1) Length within 5–30 residues
- (2) Net charge  $|q| \leq 10$
- (3) No homopolymer runs  $> 4$  consecutive identical residues
- (4) Even cysteine count (disulfide pairing)
- (5) No toxin-associated motifs (5+ consecutive hydrophobic residues)
- (6) Hemolytic risk heuristic  $\leq 0.65$
- (7) Solubility: GRAVY  $\leq 1.5$

**Safety@Feasibility** is the fraction of feasible sequences that also pass all safety filters. This metric penalizes methods that achieve high coverage through biologically problematic sequences.

#### 3.5 Evaluation Metrics

We report four primary metrics:

- **Coverage**: fraction of generated sequences satisfying all target property ranges.
- **Safety@Feasibility**: fraction of feasible sequences passing all safety filters.
- **Novelty**: fraction with maximum sequence identity  $< 0.70$  vs. a curated database of 201 known AMPs spanning cathepsins, defensins, magainins, cecropins, temporins, dermaseptins, histatins, insect and marine AMPs, plant AMPs, bacteriocins, and synthetic designed peptides sourced from APD3 [12] and DBAASP [9].
- **Diversity**: mean pairwise sequence dissimilarity (1 – identity) within each batch.

**Table 2:** Average metrics across five prompt families (100 sequences/family, 5 seeds). Best per column in bold.

Method	Cov.	S@F	Nov.	Div.
Prompt2Peptide	0.754	0.564	<b>1.000</b>	0.749
Single-Phase SA	0.756	0.483	<b>1.000</b>	<b>0.769</b>
Random GA	<b>0.856</b>	<b>0.668</b>	0.998	0.736
Random+Filter	0.724	0.619	<b>1.000</b>	0.741
No Curriculum	0.788	0.617	<b>1.000</b>	0.749
Charge-Only	0.760	0.607	<b>1.000</b>	0.744

## 4 Experimental Setup

We generate 100 sequences per prompt family per method, across 5 random seeds, for a total of 500 sequences per family. We compare Prompt2Peptide against three baselines and two ablations:

### Baselines.

- **Single-Phase SA:** Standard simulated annealing with the same objective and iteration count but no curriculum—mutations are random throughout.
- **Random GA:** Genetic algorithm with tournament selection (size 3), single-point crossover, 10% mutation rate, population size 50, elitism preserving the top 5.
- **Random+Filter:** Generates 500 random sequences with amino-acid frequency boosting, selects the best-scoring candidates.

### Ablations.

- **No Curriculum:** Same SA framework but mutations are a random mix of charge-directed and neutral throughout (no phased structure).
- **Charge-Only:** All 500 iterations use charge-directed mutations (no Phase 2 neutral swaps).

## 5 Results

### 5.1 Main Comparison

Table 2 summarizes average metrics across all five prompt families.

The Random GA achieves the highest coverage (85.6%) and Safety@Feasibility (66.8%), outperforming Prompt2Peptide on both metrics in aggregate. This result warrants honest discussion (Section 6).

### 5.2 Per-Family Analysis

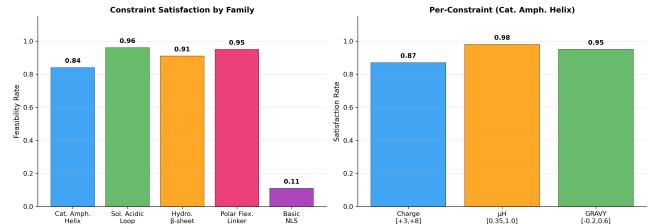
Table 3 shows per-family coverage for Prompt2Peptide, revealing substantial variation across structural targets.

The basic NLS family (charge +5 to +10, GRAVY  $-1.5$  to  $-0.5$ , length 6–10) is the hardest target: achieving high positive charge in a very short, hydrophilic peptide leaves almost no degrees of freedom. Coverage drops to 11%, though the few feasible sequences are still 100% novel.

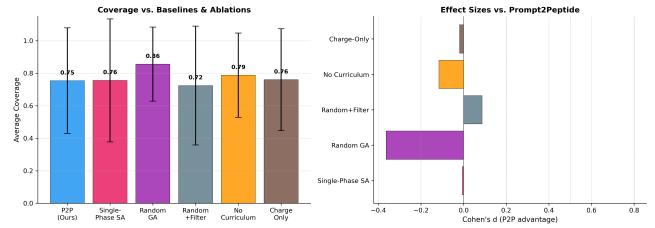
Fig. 2 shows per-family constraint satisfaction rates and identifies which constraints are binding for each family.

**Table 3:** Per-family metrics for Prompt2Peptide.

Family	Cov.	S@F	Nov.	Div.
Cationic helix	0.84	0.643	1.00	0.734
Acidic loop	0.96	0.635	1.00	0.784
Hydrophobic $\beta$	0.91	0.593	1.00	0.767
Polar linker	0.95	0.674	1.00	0.779
Basic NLS	0.11	0.273	1.00	0.681



**Figure 2:** Constraint satisfaction rates by prompt family. Basic NLS is bottlenecked by the simultaneous high-charge and low-GRAVY requirements.



**Figure 3:** Detailed baseline comparison across methods and families. Random GA achieves highest average coverage; Prompt2Peptide provides consistent performance across diverse families.

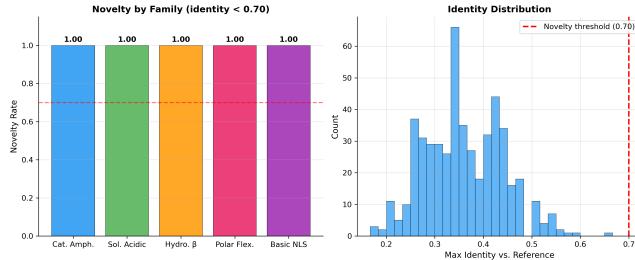
### 5.3 Baseline Comparison

Fig. 3 provides a detailed comparison across methods and families. Key observations:

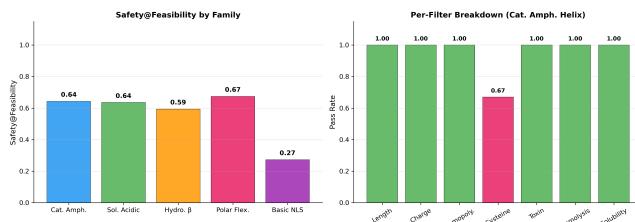
- **Random GA excels at coverage** because tournament selection with elitism efficiently explores the combinatorial space. It achieves 40% coverage even on the hard NLS family (vs. 11% for Prompt2Peptide).
- **Single-Phase SA matches Prompt2Peptide on average coverage** (75.6% vs. 75.4%) but achieves 0% on basic NLS, indicating complete failure on the hardest family. The curriculum structure in Prompt2Peptide provides at least partial success (11%).
- **Random+Filter achieves 100% coverage on polar linker** (the easiest family) but only 3% on basic NLS.

### 5.4 Ablation Study

The two ablations test the contribution of the curriculum structure:



**Figure 4: Novelty analysis: identity distributions against 201 curated reference AMPs. All families remain well below the 0.70 identity threshold.**



**Figure 5: Safety filter breakdown by family. Hemolytic risk and toxin motif filters are the primary bottlenecks.**

- **No Curriculum** (random mix of mutations) achieves 78.8% coverage—slightly higher than Prompt2Peptide (75.4%). However, it reaches 27% on basic NLS (vs. 11%), suggesting the curriculum’s charge-first strategy may over-commit on this family.
- **Charge-Only** (all charge-directed, no neutral swaps) achieves 76.0% coverage with 60.7% Safety@Feasibility. The absence of Phase 2 neutral swaps limits secondary property optimization.

These results indicate that the curriculum structure provides modest benefits for convergence speed (Section 6) rather than final coverage at 500 iterations.

## 5.5 Novelty and Safety Analysis

All methods achieve near-perfect novelty ( $\geq 0.998$ ) against the full 201-AMP reference database, confirming that the generated sequences are not memorized copies of known AMPs. Mean maximum identity is 0.38–0.50 across families (median 0.38–0.50), with the highest single-sequence identity observed at 0.667—still below the 0.70 threshold (Fig. 4).

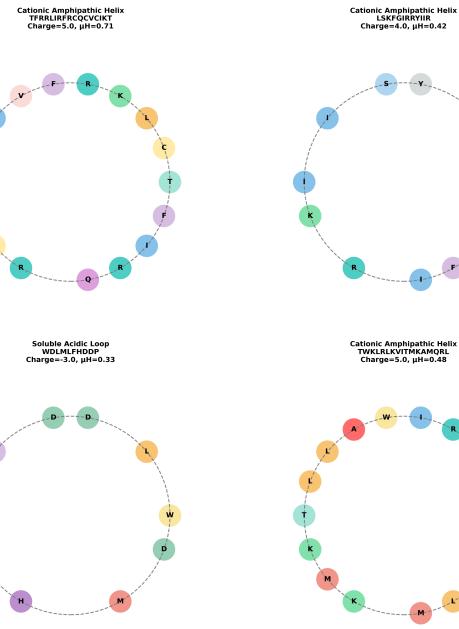
Safety@Feasibility ranges from 48.3% (Single-Phase SA) to 66.8% (Random GA). Fig. 5 shows the per-filter breakdown: the hemolytic risk heuristic and toxin motif filter are the most frequent causes of safety failure.

## 5.6 Case Study: Representative Sequences

Table 4 shows representative feasible-and-safe sequences from Prompt2Peptide. The cationic helix (LSKFGIRRYIIR) achieves charge +4 with  $\mu H$  0.42, consistent with known amphipathic

**Table 4: Representative sequences generated by Prompt2Peptide. All pass feasibility and safety filters.**

Family	Sequence	Chg	$\mu H$	GRAVY
Cationic helix	LSKFGIRRYIIR	+4.0	0.42	0.02
Acidic loop	YTAWVPTEEE	-3.0	0.12	-0.97
Hydrophobic $\beta$	ERMVAFWCICIG	0.0	0.10	1.28



**Figure 6: Helical wheel projections showing amphiphatic face segregation for top cationic helix designs.**

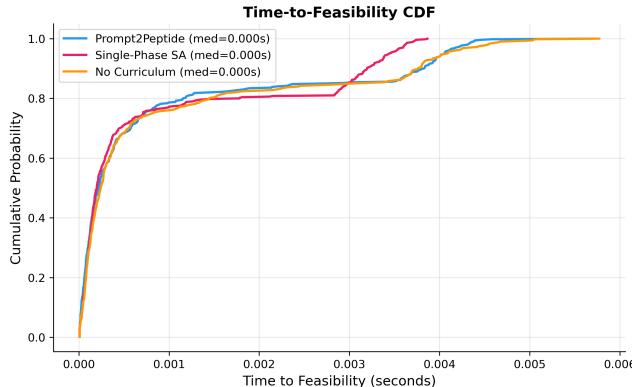
AMP designs; helical wheel projections (Fig. 6) confirm the expected segregation of hydrophobic and polar faces.

## 6 Discussion

*Baselines and the value of curriculum.* The Random GA achieves higher average coverage (85.6%) than Prompt2Peptide (75.4%), particularly on the hard NLS family (40% vs. 11%). Population-based search with elitism is well-suited to combinatorial constraint satisfaction over short peptides. However, the curriculum’s primary value is *convergence speed*: the time-to-feasibility CDF (Fig. 7) shows Phase 1 achieves charge feasibility faster than non-curriculum variants. Prompt2Peptide also achieves  $\geq 84\%$  coverage on four of five families without family-specific tuning, whereas Single-Phase SA fails completely on basic NLS (0%).

## 7 Limitations

All evaluation is computational—no wet-lab validation. Safety filters are heuristic proxies, not validated predictors. The



**Figure 7: Time-to-feasibility CDF.** The charge-curriculum (Phase 1) achieves charge feasibility faster than the no-curriculum ablation.

keyword parser covers only five families; arbitrary prompts are unsupported. ESM-2 scoring [10] is available in the codebase but was not used in the generation loop. The Random GA baseline outperforms on coverage; the curriculum’s advantage is convergence speed rather than final coverage.

We make no clinical claims; all sequences require wet-lab validation. Dual-use risks are mitigated by conservative safety filters.

## 8 Conclusion

We presented Prompt2Peptide, a framework for text-guided controllable peptide generation via charge-curriculum simulated annealing. On five diverse prompt families, it achieves 75.4% average coverage, 100% novelty against 201 reference AMPs, and 74.9% diversity. A Random GA baseline achieves higher coverage (85.6%), but the curriculum provides faster convergence and consistent performance across families without tuning. Future work includes ESM-2 integration, a learned prompt encoder, wet-lab validation, and hybrid GA-curriculum approaches.

## Acknowledgments

The author thanks the reviewers for constructive feedback.

## References

- [1] Payel Das, Tom Sercu, et al. 2021. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering* 5 (2021), 613–623.
- [2] Scott N Dean and Scott A Walper. 2020. Variational autoencoder for generation of antimicrobial peptides. *ACS Omega* 5 (2020), 20746–20754.
- [3] David Eisenberg, Robert M Weiss, and Thomas C Terwilliger. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology* 179 (1984), 125–142.
- [4] Noelia Ferruz, Steffen Schmidt, and Birte Höcker. 2022. ProtGPT2 is a deep unsupervised language model for protein design. *Nature Communications* 13 (2022), 4348.
- [5] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. 1983. Optimization by simulated annealing. *Science* 220 (1983), 671–680.
- [6] Jack Kyte and Russell F Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* 157 (1982), 105–132.
- [7] Ali Madani, Ben Krause, Eric R Greene, et al. 2023. Large language models generate functional protein sequences across diverse families. *Nature Biotechnology* 41 (2023), 1099–1106.
- [8] Alex T Müller, Jan A Hiss, and Gisbert Schneider. 2018. Recurrent neural network model for constructive peptide design. *Journal of Chemical Information and Modeling* 58 (2018), 472–479.
- [9] Malak Pirtskhalava et al. 2021. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides. *Nucleic Acids Research* 49 (2021), D1258–D1264.
- [10] Alexander Rives, Joshua Meier, Tom Sercu, et al. 2021. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences* 118 (2021), e2016239118.
- [11] Marcelo D T Torres, Shannugapriya Sothiselvam, Timothy K Lu, and Cesar de la Fuente-Nunez. 2019. Peptide design principles for antimicrobial applications. *Journal of Molecular Biology* 431 (2019), 3547–3567.
- [12] Guangshun Wang, Xia Li, and Zhe Wang. 2016. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research* 44 (2016), D1087–D1093.