CAN AGENT UNDERSTAND THE DIFFERENCES BETWEEN FUNCTIONALLY DISTINCT PEPTIDES? A SELF-EVOLVING BIFUNCTIONAL PEPTIDE DISCOVERY ENGINE

Anonymous authors
Paper under double-blind review

ABSTRACT

Computational antimicrobial peptide design typically optimizes activity against a predefined target, leaving the choice of complementary therapeutic function to researchers. Yet infection scenarios often require both pathogen clearance and control of inflammation, toxins, virulence, biofilms, or impaired tissue repair. We introduce JANUS-PEP, an agentic framework that translates a natural-language infection scenario into a complementary function, molecular target, structural epitope, and pathogen-specific antimicrobial endpoint, and then redesigns a single peptide to satisfy binding, potency, and hemolysis constraints. The framework selects between structure- and sequence-based design paths and separates diagnosis from proposal, pairing each edit with a directional prediction and externally computed scores recorded in persistent memory. We optimize the Redesigner on assay-grounded binding and peptide activity data using supervised fine-tuning followed by Group Relative Policy Optimization. The optimized model reaches 73.83% mean accuracy, improving by 25.64 percentage points over its zero-shot backbone and outperforming the evaluated frontier models across peptide-design tasks. Across diverse infection scenarios, JANUS-PEP produces candidates that jointly improve predicted antimicrobial potency and hemolytic safety while satisfying target-binding criteria, establishing an auditable basis for scenario-grounded bifunctional peptide discovery, although the candidates remain computational hypotheses requiring experimental validation.

1 INTRODUCTION

Antimicrobial resistance is intensifying the need for new therapeutic strategies (Murray et al., 2022). Antimicrobial peptides are attractive candidates because their short sequences support combinatorial exploration and their diverse mechanisms can retain activity against drug-resistant pathogens (Lázár et al., 2018; Kintses et al., 2019). Generative models, protein language models, and simulation-guided pipelines have expanded the searchable space and produced experimentally active peptides (Das et al., 2021; Szymczak et al., 2023; Li et al., 2024; Huang et al., 2023). Across these advances the objective is fixed in advance: potency is optimized while hemolysis, toxicity and spectrum enter as constraints on it.

Real infections rarely present as single-function problems. Clearing the pathogen may need to be accompanied by controlling inflammation, neutralizing a toxin, disarming a virulence mechanism, or restoring tissue repair, and peptides combining antimicrobial activity with a second function have improved outcomes in vivo (Liu et al., 2024; Zhang et al., 2025). But in existing work that second function, its mechanism and its molecular target are all fixed before design begins (Fig. 1). The same holds for the computational tools: DiffPepBuilder generates binders for a specified interface (Wang et al., 2024), and recent peptide agents coordinate retrieval, generation and iterative optimization once the objective is given (Kong et al., 2026; Chen et al., 2026a; Jung et al., 2026). What remains unaddressed is the step before optimization, in which the scenario itself determines what the second function should be.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

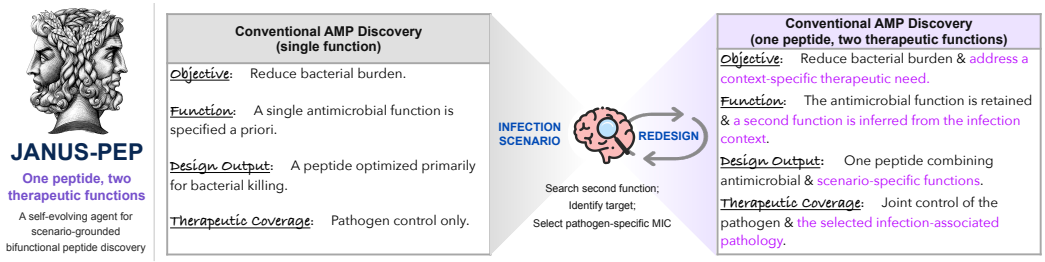

Figure 1: Motivation for scenario grounded bifunctional peptide discovery. Conventional AMP design fixes antimicrobial activity as the objective. Our target resolution agent instead derives an additional therapeutic function from the infection scenario and redesigns one peptide for both objectives.

We present JANUS-PEP, a self-evolving agent that closes this step. Given a natural-language infection scenario, a target-resolution stage retrieves evidence to specify the complementary function, its molecular target, a structural epitope, and a pathogen-specific antimicrobial endpoint, after which the target determines whether design proceeds from structure or from sequence. Each round then separates three operations that are usually collapsed: the Diagnoser reads the peptide's measured state and reports its deficits, the Redesigner converts that diagnosis into substitutions, and physical evaluators adjudicate the result. The agent commits the expected direction of every change before evaluation, so a prediction can be compared against what was measured rather than assessed after the fact; both, together with every candidate that failed, persist in memory and are carried into later rounds and later targets. Improvement therefore accumulates as a record of measured trade-offs rather than in any parameter, and the trajectory from scenario to peptide remains auditable at every step (Fig. 2).

We evaluate JANUS-PEP on six benchmarks and 26 infection scenarios, and release a 2,000-item benchmark of peptide-design decisions whose answers are experimental measurements. Optimizing the Redesigner on it by supervised fine-tuning followed by Group Relative Policy Optimization raises a 14B model to 73.83% mean accuracy, 25.64 points above its zero-shot backbone and ahead of every evaluated frontier model in overall accuracy. Across the scenarios the framework delivers candidates meeting all three criteria with selectivity of 3.7–54.0; optimization lowers mean predicted MIC from 41.6 to 18.3 μM while raising mean predicted HC_{50} from 96.4 to 137.2 μM, and 73% of candidates retain both criteria under independent rescoring. The candidates remain computational hypotheses for subsequent experimental validation.

2 SYSTEM ARCHITECTURE AND OPTIMIZATION

JANUS-PEP decomposes bifunctional peptide design into four modules: planning and generation, diagnosis, redesign, and physical scoring (Fig. 1a). They communicate through a shared memory holding design context, proposed candidates and their measured outcomes. Because proposal is decoupled from scoring, every design decision rests on a physical measurement rather than on model judgement, and progress cannot be attained by reasoning alone.

Most clinically relevant targets have no experimentally determined complex, so a pipeline requiring one addresses only the minority that does. We therefore give the system two paths and let the target decide: where a binding pocket exists the peptide is co-designed with its receptor as an all-atom complex and scored by interface energetics, and where none exists it is generated from sequence and scored by interface confidence. Diagnosis, memory, and the antimicrobial and hemolytic criteria are shared, so the coverage gained does not come at the cost of a weaker standard.

Memory is not cleared between targets. Each design begins from the record accumulated over all preceding ones, so a rule that held on one interface is available when the next is attempted, and what transfers is a measured trade-off rather than a parameter. Redesign is thus itself the mechanism of improvement. Reinforcement learning acts on the other side of this loop, optimizing what the Redesigner does with the record it reads.

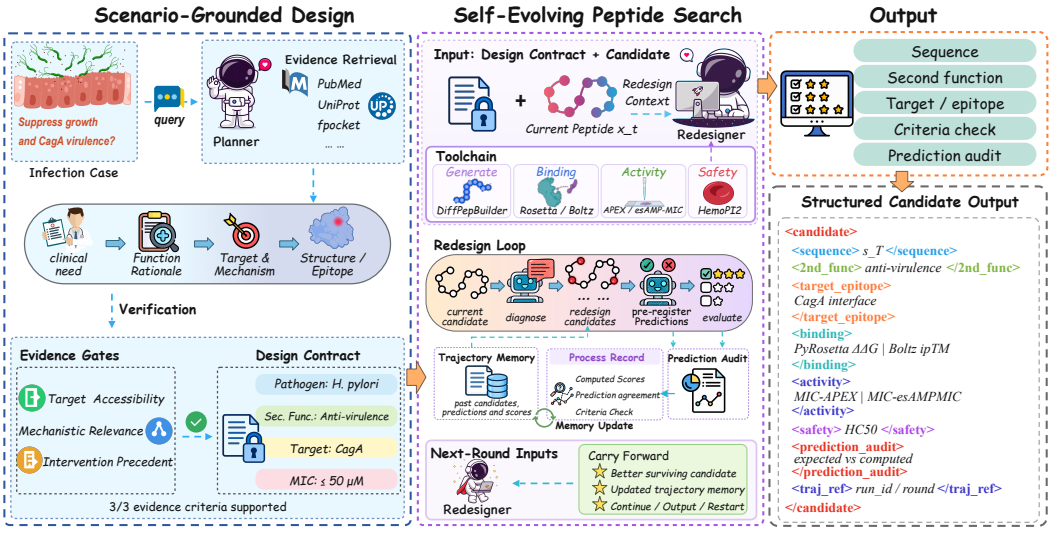

Figure 2: JANUS-PEP converts an infection scenario into an auditable bifunctional design trajectory. (A) Evidence retrieval grounds and freezes the design contract before optimization. (B) The system selects a structure- or sequence-based design path while retaining common activity and safety gates. (C) In each round, the Diagnoser identifies property defects, the Redesigner proposes a revision, and the agent commits the expected directions of change before external evaluation. Prediction–outcome disagreements are preserved in persistent memory as evidence for subsequent rounds. (D) The final record exposes the candidate, computed scores, gate status, prediction audit, and trajectory reference.

2.1 SPECIALIZED AGENTS AND RECURSIVE SELF-IMPROVEMENT MEMORY

The **Planner** is the entry point. It reads a natural-language scenario, identifies the target and the epitope that must be engaged, determines which design path the target admits, and selects the starting scaffold. It turns an informal requirement into a concrete design problem. The **Builder** then constructs the object that redesign acts on. It samples backbone conformations in the binding site, co-designs an initial sequence, and relaxes the complex to full-atom resolution. The backbone is held fixed thereafter, so every later modification is a substitution on a known frame, and its effect on binding can be measured rather than assumed.

In contrast to conventional propose-and-score architectures, JANUS-PEP inserts a **Diagnoser** between observation and proposal and isolates it in a separate model call. It reads the physicochemical state of the current peptide and reports what it lacks along three axes: cationicity and amphipathicity on the antimicrobial axis, positions whose substitution would compromise the interface on the binding axis, and hydrophobic accumulation on the hemolytic axis. It is prohibited from emitting candidate sequences. This prohibition is what makes the diagnosis usable. When diagnosis and proposal are produced in one pass, the model may select an edit first and compose a justification afterwards, and the rationale carries no information. Separating the calls forces the diagnosis to exist before any candidate does. The **Redesigner** then turns that diagnosis into candidates. It selects positions to mutate, proposes a batch of sequences, and pre-registers the predicted direction of change for each one before it is scored. Because the prediction precedes the measurement, the reasoning behind every edit can later be verified.

A **recursive memory** underlies all four agents. It stores (i) every candidate proposed in each round, including those that failed all criteria, (ii) the context under which they were proposed, (iii) the diagnosis and the predictions attached to them, and (iv) the measurements they received. Nothing is discarded and nothing is compressed before it is written. Before each proposal, a memory-retrieval step queries the full-score trajectories in the page store, extracts transferable rules, and supplies them to the Redesigner. Three consequences follow. Each round conditions on the measured outcome of the preceding one rather than on a summary of it. Strategies already shown to fail on a scaffold

are not proposed again. And because predictions are recorded before measurements, any claim the system makes about its own reasoning can be checked against what was measured. Memory persists across scaffolds and across targets, so rules learned on one design carry into the next, and every design decision remains traceable to the measurement that produced it.

2.2 DOMAIN-SPECIFIC TOOLSET

To enable extensibility and reliability, JANUS-PEP encapsulates each capability within a standardized Tool module defined by a rigorous schema, including typed inputs, output specifications, and usage constraints (Extended Data Fig. 1a). The Executor validates parameters against these schemas, employing an iterative self-correction mechanism to automatically repair malformed commands. Every tool returns a measured or computed quantity rather than a judgement, so that no module which proposes a candidate also determines whether it succeeds. Each candidate and the measurements it receives are logged in shared Memory, so that every accepted design remains traceable to the evidence that admitted it.

To support bifunctional peptide design, JANUS-PEP integrates 14 tools spanning target resolution, peptide generation, and physical scoring. Key capabilities include: (i) *Target Resolution*: disease-to-target mapping via the Open Targets Platform (Ochoa et al., 2021), sequence and localization via UniProtKB (The UniProt Consortium, 2025), and literature evidence via PubMed (Sayers et al., 2025), passed through three gates that establish extracellular accessibility, effector function, and inhibitory precedent; (ii) *Structure Retrieval*: experimentally determined complexes from the RCSB PDB (Burley et al., 2023) with AlphaFold DB (Varadi et al., 2024) as fallback, followed by epitope extraction and fpocket analysis of apo monomers (Le Guilloux et al., 2009); (iii) *Peptide Generation*: SE(3)-equivariant co-design of backbone and sequence with DiffPepBuilder for pocket-bearing targets, and target-conditioned masked language modelling with PepMLM for the remainder (Chen et al., 2026b); and (iv) *Physical Scoring*: interface energetics from PyRosetta (Chaudhury et al., 2010) under the ref2015 score function (Alford et al., 2017), binding perplexity from PepMLM, antimicrobial activity and MIC estimates from APEX and esAMPMIC (Wan et al., 2024; Chung et al., 2024), and hemolytic concentration estimates from HemoPI2 (Rathore et al., 2025). The Planner dynamically selects from this toolkit based on the scenario and evolving Memory state.

The antimicrobial and hemolytic axes are scored identically on both paths, since both depend only on sequence. The binding axis cannot be. Interface energetics require atomic coordinates, which exist only when the peptide is co-designed with its complex; structure-based designs are therefore scored by the change in interface energy on the fixed backbone, reported in Rosetta energy units, while sequence-based designs are scored by binding perplexity under target conditioning. The two are not pooled, and results from the two paths are reported separately throughout. This heterogeneity is deliberate: interface energetics derive from a force field and use no training data, whereas the activity predictors are fit to measured concentrations and use no structure. A design is accepted only when all three axes agree.

2.3 ITERATIVE MULTI-TURN REDESIGN WITH SELF-EVOLVING MEMORY

Redesign proceeds as a closed loop over a fixed starting peptide. Each turn computes the peptide's physicochemical state (net charge, hydrophobic moment, helical propensity, aromatic content, and, where a complex exists, a per-residue interface map); the Diagnoser reads that state and reports deficits along the antimicrobial, binding and hemolytic axes; the Redesigner converts the diagnosis into a batch of candidates, each with its expected direction of change pre-registered before scoring; and the toolset returns the three readouts for every candidate. The next turn begins from the best surviving candidate.

The two paths differ only in what the loop operates on. On the structure-based path the starting peptide comes with its receptor as an all-atom complex and every substitution is applied on the fixed backbone, so the effect of an edit is measured on real coordinates rather than inferred. On the sequence-based path no such complex exists: the interface map is replaced by an epitope-directed prior and binding is scored by interface confidence. Diagnosis, memory, and the antimicrobial and hemolytic criteria are identical on both.

Self-evolution. Let x_t denote the peptide at turn t and $\mathcal{M}_t$ the memory. A turn applies

$$x_{t+1} = \sigma\big(x_t, \mathcal{M}_t\big), \qquad \mathcal{M}_{t+1} = \mathcal{M}_t \cup \big\{(\phi_t, d_t, c_j, \hat{y}_j, s_j)\big\}_{j=1}^{B}, \tag{1}$$

where σ composes diagnosis, redesign and physical scoring; ϕ_t is the computed state, d_t the diagnosis, c_j a proposed candidate, $\hat{y}_j$ the direction it was predicted to move, and s_j the measurements it received. The update is monotone, so no turn can remove what an earlier turn measured, and it retains $\hat{y}_j$ beside s_j, so every prediction stays paired with its outcome. Because x_{t+1} depends on $\mathcal{M}_t$ and therefore on every measurement taken since x_0, the loop is itself the mechanism of improvement rather than a wrapper around one.

Evolution operates at a second scale as well. Memory is not reinitialized when an episode ends:

$$\mathcal{M}_0^{(k+1)} = \mathcal{M}_{T_k}^{(k)}, \tag{2}$$

so the $(k{+}1)$-th target begins from the record accumulated over the k preceding ones. Nothing else carries over. Policy weights, evaluators and prompts are identical across episodes, so whatever transfers between targets transfers as text, and clearing $\mathcal{M}$ returns the system to its initial behaviour.

Three properties of Eq. 1 matter for what follows. Candidates within a turn are proposed against the same ϕ_t and scored under the same criteria, so a turn is a controlled comparison. All candidates are written, including those failing every criterion, so negative evidence accumulates at the same rate as positive. And the paired $(\hat{y}_j, s_j)$ makes the loop auditable: a claim the system made about its own reasoning can be checked against what was subsequently measured. Inspecting those pairs shows that the weighing of objectives, rather than the search, is the binding constraint, since edits are proposed whose predicted and measured directions disagree. Strengthening that judgement is the target of the optimization described next.

2.4 Reinforcement learning optimization

Bifunctional design admits no single correct edit. Antimicrobial potency, interface complementarity, and hemolytic safety pull against one another, and many distinct trajectories reach the joint constraint. Supervised learning can only imitate the solutions it is shown, whereas reinforcement learning rewards the outcome and leaves the route open. We therefore cast the redesign decision itself as a learning problem: we construct a corpus of design choices whose answers are experimentally known, and optimize the Redesigner on it with Group Relative Policy Optimization (GRPO) (Shao et al., 2024). Figure 3 summarizes the supervised warm start and subsequent assay-grounded policy optimization.

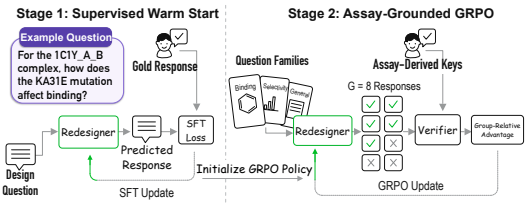

Figure 3: Two-stage optimization of the Redesigner. SFT provides the warm start; assay-grounded GRPO reinforces exact matches to fixed experimental answer keys.

A corpus of measured decisions. Training such a policy requires knowing which edit was the right one, and that is precisely what is missing. Optimizing an agent against its own judgement is circular; optimizing it against a reward model transfers the circularity one level up. Existing peptide benchmarks do not close the gap either, because they ask what a sequence is rather than which substitution to make. The decisions a design agent faces have in fact been made thousands of times in the laboratory, in every mutational study of an interface and every analogue series tested for potency and hemolysis, but they have never been assembled into a form a policy can be trained on. We therefore built one, in which every item is a multiple-choice question whose key is read from an experimental measurement. *Binding* items draw on the 7,000 characterized substitutions in SKEMPI 2.0 (Jankauskaite et al., 2019), of which 5,410 survive filtering, and ask which of several substitutions least destabilizes an interface or in which direction one shifts affinity, keyed to the measured $\Delta\Delta G$. *Selectivity* items draw on DBAASP and GRAMPA (Pirtskhalava et al., 2021; Witten & Witten, 2019), from which 984 peptides carry both a minimum inhibitory and a hemolytic concentration, and ask which has the widest therapeutic window, keyed to the largest measured ratio. *General* items serve as capability anchors. The corpus comprises 2,000 items in a 3:2 ratio of domain to general families.

Constructing items that cannot be shortcut. A measured key guarantees that the answer is right; it does not guarantee that arriving at it requires reasoning. A model can score well by memorizing a complex it has seen, by always choosing the same option, or by reading a single physicochemical property off the choices. We close these by partitioning over complexes and parent peptides rather than over items, by equalizing answer classes within each family, and by discarding substitutions below 1 kcal mol^{-1}, where the label lies within assay error. The last shortcut needs most care, since potency and safety are correlated in nature and the most active option in a random draw is often also the correct one. Distractors are therefore built against that correlation, ordered so that the most potent option is also the least safe; the key is the unique largest ratio, lies at neither extreme, and yields only to the quotient, so that reading either axis alone performs at chance. Each family is finally audited against simple physicochemical rules and excluded if any exceeds a permutation-derived null, which removed one antimicrobial family whose source data is close to monotone in aromatic content. Appendix E gives the full procedures.

We release the corpus together with that pipeline as an open benchmark. To our knowledge it is the first collection of peptide-design decisions in which every key is an experimental measurement, every split is drawn over structures rather than items, and every family is screened for shortcut solvability.

A reward that cannot be argued with. Each response is graded by exact agreement with the measured key,

$$r_i = \mathbf{1}[\,\hat{y}(o_i) = y^\star\,], \qquad y^\star \in \{\Delta\Delta G,\ \mathrm{HC}_{50}/\mathrm{MIC}\}, \tag{3}$$

where $\hat{y}(o_i)$ is the answer extracted from response o_i. The indicator is deliberately the entire reward. No term rewards format, length, calibration, or the stated rationale, and no model participates in grading, so the gradient carries no signal that a more persuasive answer is a better one. Two properties follow. The reward is invariant to presentation: two responses selecting the same option are indistinguishable to the optimizer however differently they argue for it. And the reward is grounded outside the loop—its value was fixed by an assay performed before the system existed, and no trajectory the policy discovers can alter it. Training and evaluation are graded by the same procedure, so the objective being optimized and the quantity being reported are identical.

Optimization. Where the base policy performs near chance, every response in a sampled group receives the same reward, the normalized advantage vanishes, and no gradient is produced. The Redesigner is therefore first aligned on the same corpus by supervised fine-tuning, which establishes a non-degenerate response distribution before policy optimization begins. GRPO then proceeds with $G = 8$ responses per query, a global batch of 32 queries, under full-parameter ZeRO-3 sharding (Rajbhandari et al., 2020).

3 Results

3.1 Benchmark evaluation of Janus-Pep

Evaluation setup. We evaluate on six multiple-choice benchmarks. Two are existing general biomedical benchmarks, MedQA and PubMedQA, included as capability anchors. The other four test design reasoning and are constructed so that every answer is determined by an experimental measurement. Bind-Sel asks which of several substitutions least destabilizes a protein interface and Bind-Dir asks in which direction a substitution shifts affinity, both keyed to measured $\Delta\Delta G$ from SKEMPI 2.0. Sel-TI asks which of several peptides has the widest therapeutic window, with options constructed from DBAASP and GRAMPA so that potency and safety rise together and reading either axis alone performs at chance. ProtGym is drawn from an existing variant-effect benchmark. Leakage controls are given in Section 2.4.

The two evaluation blocks differ in what is being varied. On the four design tasks, Janus-Pep applies its full loop: the physicochemical state of each candidate is computed, the Diagnoser reports deficits along the three axes, and the Redesigner revises across rounds under the termination rule used in design. The general benchmarks are answered in a single pass in all conditions, since the framework provides no tool that bears on clinical reasoning; differences in those two columns therefore reflect the backbone alone and serve as capability anchors. Every response is graded by the same extractor used during training, and we report mean accuracy across three independent runs.

Redesigner setup		General biomedical		Peptide-specific				Overall	
Backbone	Optim.	MedQA	PubMedQA	Bind-Sel	Bind-Dir	Sel-TI	ProtGym	Avg.	Δ
Base LLMs (zero-shot)									
Qwen3-14B	—	68.30 $_{.177}$	71.03 $_{.096}$	36.73 $_{.214}$	44.32 $_{.189}$	29.74 $_{.123}$	39.01 $_{.164}$	48.19	—
GPT-6	—	**81.79** $_{.095}$	82.38 $_{.113}$	42.07 $_{.241}$	55.75 $_{.204}$	33.31 $_{.193}$	47.31 $_{.155}$	57.10	—
Opus 5	—	81.05 $_{.103}$	**83.76** $_{.064}$	44.74 $_{.178}$	57.02 $_{.225}$	35.72 $_{.232}$	49.04 $_{.133}$	58.56	—
Gemini 3.1	—	78.31 $_{.121}$	81.78 $_{.141}$	41.30 $_{.201}$	54.07 $_{.198}$	35.84 $_{.171}$	46.71 $_{.211}$	56.34	—
JANUS-PEP *(frozen Redesigner)*									
Qwen3-14B	Frozen	68.30 $_{.177}$	71.03 $_{.096}$	45.37 $_{.194}$	58.78 $_{.155}$	41.02 $_{.226}$	60.34 $_{.185}$	57.47	↑9.28
GPT-6	Frozen	**81.79** $_{.095}$	82.38 $_{.113}$	59.30 $_{.160}$	67.02 $_{.184}$	52.73 $_{.201}$	68.01 $_{.140}$	68.54	↑11.44
Opus 5	Frozen	81.05 $_{.103}$	**83.76** $_{.064}$	57.70 $_{.159}$	68.30 $_{.161}$	50.30 $_{.134}$	69.70 $_{.122}$	68.47	↑9.91
Gemini 3.1	Frozen	78.31 $_{.121}$	81.78 $_{.141}$	54.00 $_{.187}$	67.70 $_{.192}$	52.30 $_{.188}$	**70.70** $_{.173}$	67.47	↑11.13
JANUS-PEP *(optimized Redesigner)*									
Qwen3-14B	SFT + GRPO	81.70 $_{.139}$	83.30 $_{.114}$	**68.00** $_{.147}$	**76.30** $_{.121}$	**63.70** $_{.174}$	70.00 $_{.152}$	**73.83**	**↑25.64**

Table 1: **Benchmark performance.** Accuracy over three runs; subscripts give standard deviations. Δ is the gain over the same backbone zero-shot. Chance levels differ across tasks. Best in bold.

Performance analysis. Every zero-shot backbone answers medical text competently and substitution questions near chance (Table 1): Opus 5 reaches 81.05% on MedQA yet 44.74% on binding selection. Competence on clinical reasoning does not transfer to judgements about specific edits.

Embedding the same backbone within JANUS-PEP, with the Redesigner held fixed, raises overall accuracy by 9.3 to 11.4 points across all four backbones, entirely in the design tasks; the general columns are unchanged by construction, so the comparison isolates what the loop contributes. Consistency across backbones attributes the gain to the architecture rather than to any particular model.

Supervised cold-start followed by GRPO adds a further increment, reaching 73.83% overall (↑25.64). The largest movements are on the two tasks requiring a committed direction, Bind-Dir and Sel-TI. The optimized 14B model exceeds every frontier backbone evaluated inside the same framework on three of the four design tasks, including those where they began more than eight points ahead.

3.2 BIFUNCTIONAL DESIGN ACROSS CLINICAL SCENARIOS

Design setup. Each scenario names an infection together with a second requirement that killing alone does not meet. The Planner resolves that requirement to a protein target that must be extracellularly accessible and an effector of the pathology rather than a correlate, and the target determines which design path is used. A candidate passes when binding is favourable, predicted MIC falls within the antimicrobial threshold, and predicted HC_{50} lies above the hemolytic one.

Design outcomes. JANUS-PEP delivered a passing candidate in every scenario attempted (Table 2), spanning eight classes of second function and both host and bacterial targets. Outcomes depended on whether the second-function interface and the antimicrobial requirement pull in the same direction: the *C. perfringens* α-toxin site, which accommodates a cationic amphipathic peptide, gave a 54-fold window in two rounds, while polar or acidic surfaces such as IL-1β took more rounds for narrower windows, since the substitutions that raise potency are the ones such an interface tolerates least.

Dispatching a target to the sequence-based path costs reach rather than admissibility (Fig. 4a): hemolytic concentrations are comparable across paths and every candidate on both clears the bounds, while a resolved pocket lets binding be optimized directly rather than inferred. The two binding scales are not comparable and are never pooled.

Across rounds the population of proposed candidates shifts toward lower MIC and narrows as it shifts (Fig. 4c), the signature of a search using its accumulated measurements rather than resampling from the same prior.

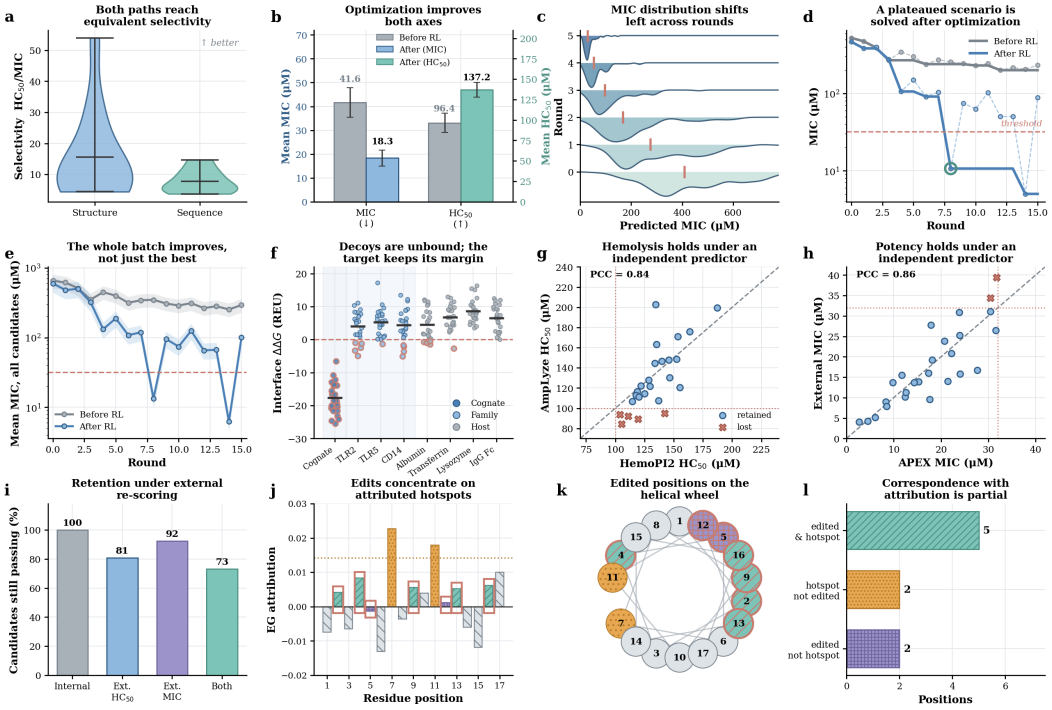

Figure 4: **Design performance, search dynamics, and robustness. a,** Selectivity (HC_{50}/MIC) of delivered candidates on the two paths. **b,** Population-level effect of optimization on mean MIC and HC_{50}; error bars denote s.d. across candidates. **c,** MIC distribution of all proposed candidates by round (kernel density, medians marked). **d,** A scenario the unoptimized system could not solve. Dashed traces give each round's best candidate, solid traces the running best; red line marks the antimicrobial threshold. **e,** Mean MIC across all candidates in the same scenario. **f,** Interface energy against the cognate target and eight decoys (three same-family receptors, five abundant host proteins). Black bars are group means; red circles mark favourable energies against a decoy. **g,h,** Agreement between the predictors used for selection and independent predictors, for hemolysis **j,** Per-residue attribution from an external model after the final round; red boxes mark edited positions, the dotted line the hotspot threshold. **k,** The same peptide on a helical wheel. **l,** Hotspots edited, hotspots left untouched, and edits outside the hotspots.

3.3 Effect of optimization on design

The gains in Table 1 are measured on decision items. To ask whether they transfer to design, we replaced the Redesigner's base model with the SFT + GRPO checkpoint and re-ran all scenarios under otherwise identical conditions, memory included.

Both axes improve at the population level (Fig. 4b): mean MIC across delivered candidates falls from 41.6 to 17.9 μM while mean HC_{50} rises from 96.4 to 136.9 μM. The two move in opposite directions in value and the same direction in quality, which a single-objective optimizer does not produce, since potency is most cheaply gained by adding cationic and hydrophobic residues and that is also the fastest route to hemolysis.

The two mechanisms act at different timescales and neither alone accounts for this. Memory supplies what a substitution cost when it was last attempted, on this target and on earlier ones; optimization changes what the Redesigner does with that record. A policy trained to weigh competing axes has nothing to weigh without a history of measured trade-offs, and a history is inert if the policy reading it cannot act on the comparison.

Their combination is clearest on a scenario the unoptimized system could not solve (Fig. 4d,e). Before optimization the running best MIC plateaus above 200 μM by round 3 and stays there; after, the same scenario crosses the threshold at round 8. Individual rounds still fluctuate and several

#	Second function	Target	Len.	Binding	MIC_1	MIC_2	HC_{50}	Sel.	Rd.
Structure-based path — *pocket available; binding scored by* $\Delta\Delta G$									
S1	Endotoxin neutralization	TLR4 (O00206)	17	-5.132	4.512	21.236	118.604	26.287	3
S2	Anti-inflammatory	IL-1β (P01584)	18	-3.618	23.907	31.442	105.371	4.407	6
S3	Anti-toxin	Protective antigen (P13423)	15	-20.045	23.814	26.803	216.592	9.096	2
S4	Anti-virulence	α-toxin (P0C216)	19	-32.617	2.703	19.128	145.874	53.964	2
S5	Anti-virulence	CagA (P55980)	20	-12.483	18.126	27.095	110.982	6.123	2
S6	Anti-leukotoxin	LukS-PV (Q783R1)	16	-14.207	9.815	17.624	142.337	14.502	4
S7	Anti-toxin	TcdB CROPs (P18177)	21	-18.739	15.218	22.946	163.815	10.765	3
S8	Anti-toxin	Shiga toxin B (P69179)	14	-23.416	6.094	13.527	187.243	30.726	2
S9	Anti-virulence	LcrV (P0C7U7)	18	-9.628	20.739	25.311	128.462	6.194	5
S10	Immune evasion reversal	Protein A (P38507)	15	-16.852	12.437	18.049	155.118	12.472	3
S11	Anti-cytolysin	Pneumolysin (Q2XU26)	19	-11.334	17.882	24.617	133.706	7.477	4
S12	Anti-adhesion	Als3 (Q59L12)	17	-13.926	24.058	28.832	119.415	4.964	7
Sequence-based path — *no pocket; binding scored by interface confidence*									
Q1	Anti-inflammatory	IL-6 (P05231)	19	0.713	31.724	28.615	117.582	3.706	--
Q2	Immunomodulation	IL-17A (Q16552)	13	0.884	8.296	14.803	121.937	14.698	--
Q3	Anti-toxin	Enterotoxin B (A7Y1W2)	14	0.931	8.514	9.823	125.196	14.705	5
Q4	Anti-toxin	α-hemolysin (P09616)	19	0.768	11.702	10.416	120.043	10.259	3
Q5	Anti-toxin	Cholera toxin B (P01556)	15	0.942	12.587	11.294	134.371	10.675	1
Q6	Anti-toxin	TSST-1 (P06886)	15	0.927	14.318	29.205	152.436	10.646	1
Q7	Anti-toxin	Pertussis toxin S1 (P04977)	19	0.918	27.653	25.937	103.824	3.755	4
Q8	Anti-virulence	IpaB (P18011)	21	0.624	21.406	32.018	136.915	6.396	7
Q9	Anti-virulence	SpeB (P0C0J0)	21	0.793	30.472	30.286	129.743	4.258	8
Q10	Anti-virulence	PcrV (M4Q7H6)	17	0.871	31.528	33.294	146.387	4.643	3
Q11	Anti-biofilm	MrkA (D5FKS5)	22	0.764	17.385	19.207	114.426	6.582	11
Q12	Anti-biofilm	OmpA (A9Z1A2)	21	0.752	30.418	14.239	153.507	5.047	2
Q13	Pro-angiogenic	Angiopoietin-2 (O15123)	18	0.641	17.573	19.408	139.862	7.959	1
Q14	Wound healing	FGF2 (P09038)	14	0.826	22.194	24.163	135.294	6.096	9

Table 2: **Bifunctional peptide design across clinical scenarios.** Binding is scored by interface energy ($\Delta\Delta G$, REU) on the structure-based path and by interface confidence (ipTM) on the sequence-based path; the two are not comparable. MIC and HC_{50} are in µM, and Sel. is HC_{50}/MIC_1. *Len.* is the length of the first passing candidate and *Rd.* the round at which it appeared. Scenario Sn and Qn denote the structure- and sequence-based paths respectively; the clinical scenario behind each target, and the evidence for selecting it, are given in Appendix C.

propose candidates worse than the one before, yet the population mean keeps falling. The gain comes from where edits are spent, not from proposing more of them.

3.4 Robustness

Three checks test whether the candidates are what their scores say. Against eight decoy receptors, candidates score near zero on average while cognate targets average -15.2 REU, so binding is preferential rather than generic. Re-scoring with predictors that took no part in selection gives Pearson correlations of 0.84 and 0.86, with 73% still passing both criteria. And edited positions coincide with hotspots an independent interpretable model identifies far more often than chance. Procedures and full results are in Appendices J, K, and B (Fig. 4f–l).

4 Conclusion

We presented Janus-Pep, a multi-agent system for designing peptides that are both broadly antimicrobial and directed at a protein target chosen for the clinical situation. The problem is not a sum of two objectives: the cationicity that makes a peptide kill bacteria also makes it lyse erythrocytes, so the requirements must be satisfied against one another rather than in sequence.

The system is built around two separations. Proposal is kept apart from scoring, so a design advances only when a measurement moves; and diagnosis is kept apart from proposal, so what a peptide lacks is recorded before any candidate exists. We also release a benchmark of peptide-design decisions in which every answer is an experimental measurement, asking not what a sequence is but which edit to make.

REFERENCES

Rebecca F. Alford, Andrew Leaver-Fay, Jeliazko R. Jeliazkov, Matthew J. O'Meara, Frank DiMaio, et al. The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 2017. doi: 10.1021/acs.jctc.7b00125.

Lucas Assoni, Isabelle Ciaparin, Monalisa Martins Trentini, et al. Protection Against Pneumonia Induced by Vaccination with Fimbriae Subunits from Klebsiella pneumoniae. *Vaccines*, 13(3): 303, 2025. doi: 10.3390/vaccines13030303.

Yang Bai, Rui Min, Pengcheng Chen, et al. Disulfiram blocks inflammatory TLR4 signaling by targeting MD-2. *Proceedings of the National Academy of Sciences of the United States of America*, 120(31):e2306399120, 2023. doi: 10.1073/pnas.2306399120.

Sophia K Berry, Steven Rust, Lorraine Irving, et al. Characterization of mAbs against Klebsiella pneumoniae type 3 fimbriae isolated in a target-independent phage display campaign. *Microbiology Spectrum*, 12(8):e0040024, 2024. doi: 10.1128/spectrum.00400-24.

Stephen K. Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Henry Chao, et al. RCSB Protein Data Bank (RCSB.org): Delivery of experimentally determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Research*, 51(D1):D488–D508, 2023. doi: 10.1093/nar/gkac1077.

Ludovico Buti, Eric Spooner, Annemarthe G. Van der Veen, et al. Helicobacter pylori cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22):9238–9243, 2011. doi: 10.1073/pnas.1106200108.

Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J. Gray. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*, 26(5):689–691, 2010. doi: 10.1093/bioinformatics/btq007.

Houxu Chen, Achuth Chandrasekhar, and Amir Barati Farimani. Pepti-Agent: An AI agent for peptide design and optimization. *arXiv preprint arXiv:2606.15422*, 2026a. doi: 10.48550/arXiv.2606.15422.

Leo Tianlai Chen, Zachary Quinn, Madeleine Dumas, Christina Peng, Lauren Hong, et al. Target sequence-conditioned design of peptide binders using masked language modeling. *Nature Biotechnology*, 44:1002–1010, 2026b. doi: 10.1038/s41587-025-02761-2.

Poyin Chen, Brian C Russo, Jeffrey K Duncan-Lowey, et al. Topology and Contribution to the Pore Channel Lining of Plasma Membrane-Embedded Shigella flexneri Type 3 Secretion Translocase IpaB. *mBio*, 12(6):e0302121, 2021. doi: 10.1128/mBio.03021-21.

Chia-Ru Chung, Chung-Yu Chien, Yun Tang, Li-Ching Wu, Justin Bo-Kai Hsu, Jang-Jih Lu, Tzong-Yi Lee, Chen Bai, and Jorng-Tzong Horng. An ensemble deep learning model for predicting minimum inhibitory concentrations of antimicrobial peptides against pathogenic bacteria. *iScience*, 27(9):110718, 2024. doi: 10.1016/j.isci.2024.110718.

Hongrong Cui, Meilin Wu, Jiqing Wan, et al. A novel anti-PcrV monoclonal antibody inhibits toxin-mediated cytotoxicity and enhances survival in mouse models of Pseudomonas aeruginosa infection. *Human Vaccines & Immunotherapeutics*, 21(1):2567042, 2025. doi: 10.1080/21645515.2025.2567042.

Payel Das, Tom Sercu, Kahini Wadhawan, Inkit Padhi, Sebastian Gehrmann, Flaviu Cipcigan, Vijil Chenthamarakshan, Hendrik Strobelt, Cicero Dos Santos, Pin-Yu Chen, Yi Yan Yang, Jeremy P. K. Tan, James Hedrick, Jason Crain, and Aleksandra Mojsilovic. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering*, 5(6):613–623, 2021. doi: 10.1038/s41551-021-00689-x.

Lindsay N. Deis, Qinglin Wu, You Wang, et al. Suppression of conformational heterogeneity at a protein–protein interface. *Proceedings of the National Academy of Sciences of the United States of America*, 112(29):9028–9033, 2015. doi: 10.1073/pnas.1424724112.

Jean-Mathieu Desveaux, Eric Faudry, Carlos Contreras-Martel, et al. Neutralizing human monoclonal antibodies that target the PcrV component of the type III secretion system of Pseudomonas aeruginosa act through distinct mechanisms. *eLife*, 14:RP105195, 2026. doi: 10.7554/eLife.105195.

Nicholas E Dickenson, Shyamal P Choudhari, Philip R Adam, et al. Oligomeric states of the Shigella translocator protein IpaB provide structural insights into formation of the type III secretion translocon. *Protein Science*, 22(5):614–627, 2013. doi: 10.1002/pro.2245.

Yaoyao Fu, Hyojin Kim, Dong Sun Lee, et al. Structural insight into TLR4/MD-2 activation by synthetic LPS mimetics with distinct binding modes. *Nature Communications*, 16(1):4164, 2025. doi: 10.1038/s41467-025-59550-3.

Yun Ge, Man Huang, and Yong-Ming Yao. Recent advances in the biology of IL-1 family cytokines and their potential roles in development of sepsis. *Cytokine & Growth Factor Reviews*, 45:24–34, 2019. doi: 10.1016/j.cytogfr.2018.12.004.

Hideaki Higashi, Ryo Tsutsumi, Satoshi Muto, et al. SHP-2 Tyrosine Phosphatase as an Intracellular Target of Helicobacter pylori CagA Protein. *Science*, 295(5555):683–686, 2002. doi: 10.1126/science.1067147.

Junjie Huang, Yanchao Xu, Yunfan Xue, Yue Huang, Xu Li, et al. Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nature Biomedical Engineering*, 7:797–810, 2023. doi: 10.1038/s41551-022-00991-2.

Md Minarul Islam, Man Hwan Oh, Woo Shik Shin, et al. An overview of outer membrane protein A (OmpA) as a therapeutic target for Acinetobacter baumannii infection. *Expert Opinion on Therapeutic Targets*, 30(4):357–370, 2026. doi: 10.1080/14728222.2026.2671680.

Justina Jankauskaite, Brian Jiménez-García, Justas Dapkunas, Juan Fernández-Recio, and Iain H. Moal. SKEMPI 2.0: An updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019. doi: 10.1093/bioinformatics/bty635.

Jay Hwasung Jung, Xiaohan Zhang, Shenghan Song, Mahmoud Sayedahmed, Chijian Xiang, Yunong Xu, Ahmed AbdelKhalek, Severin T. Schneebeli, Matthew J. Wargo, Jianing Li, and Safwan Wshah. Agentic discovery of non-canonical antimicrobial peptides with AMPGAN v3. *arXiv preprint arXiv:2606.17127*, 2026. doi: 10.48550/arXiv.2606.17127.

Yong Joon Jeffrey Kim, Nicholas R Walton, Wei Huang, et al. Multivalent nanobodies for potent and broad neutralization of Staphylococcus aureus toxins. *Nature Communications*, 17(1):4456, 2026. doi: 10.1038/s41467-026-73120-1.

Bálint Kintses, Pramod K. Jangir, Gergely Fekete, Mónika Számel, Orsolya Méhi, et al. Chemical-genetic profiling reveals limited cross-resistance between antimicrobial peptides with different modes of action. *Nature Communications*, 10:5731, 2019. doi: 10.1038/s41467-019-13618-z.

Qingzhou Kong, Yinuo Zhao, Haifan Gong, Luoyao Kang, Jialu Fu, Lixiang Li, Boyao Wan, Peizhu Wang, Xiaojuan Li, Yue Wang, Jinghui Zhang, Yanbo Yu, Xiaoyun Yang, Xiuli Zuo, Haina Wang, and Yanqing Li. AI agent-based discovery of D-enantiomeric antimicrobial peptides against multidrug-resistant bacterial infection. *Biomaterials*, 329:123927, 2026. doi: 10.1016/j.biomaterials.2025.123927.

Benoit-Joseph Laventie, Frédéric Guérin, Lionel Mourey, et al. Residues Essential for Panton-Valentine Leukocidin S Component Binding to Its Cell Receptor Suggest Both Plasticity and Adaptability in Its Interaction Surface. *PLoS ONE*, 9(3):e92094, 2014. doi: 10.1371/journal.pone.0092094.

Viktória Lázár, Ana Martins, Réka Spohn, Lejla Daruka, Gábor Grézal, et al. Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides. *Nature Microbiology*, 3(6):718–731, 2018. doi: 10.1038/s41564-018-0164-0.

Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10:168, 2009. doi: 10.1186/1471-2105-10-168.

Tingting Li, Xuanbai Ren, Xiaoli Luo, Zhuole Wang, Zhenlu Li, Xiaoyan Luo, Jun Shen, Yun Li, Dan Yuan, Ruth Nussinov, Xiangxiang Zeng, Junfeng Shi, and Feixiong Cheng. A foundation model identifies broad-spectrum antimicrobial peptides against drug-resistant bacterial infection. *Nature Communications*, 15(1):7538, 2024. doi: 10.1038/s41467-024-51933-2.

Jing Lin, Soon-Hwan Oh, Rhian Jones, et al. The Peptide-binding Cavity Is Essential for Als3-mediated Adhesion of Candida albicans to Human Cells. *Journal of Biological Chemistry*, 289 (26):18401–18412, 2014. doi: 10.1074/jbc.M114.547877.

H. Ling, A. Boodhoo, B. Hazes, M. D. Cummings, G. D. Armstrong, J. L. Brunton, and R. J. Read. Structure of the Shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3. *Biochemistry*, 37(7):1777–1788, 1998. doi: 10.1021/bi971806n.

Yifan Liu, Pengfei Cui, Rong Tan, and Shaoguo Ru. Rapid membrane-penetrating hybrid peptides achieve efficient dual antimicrobial and antibiofilm activity through a triple bactericidal mechanism. *ACS Omega*, 9(24):26133–26148, 2024. doi: 10.1021/acsomega.4c01577.

Nicola Ivan Lorè, Alessandra Bragonzi, and Cristina Cigana. The IL-17A/IL-17RA axis in pulmonary defence and immunopathology. *Cytokine & Growth Factor Reviews*, 30:19–27, 2016. doi: 10.1016/j.cytogfr.2016.03.009.

Jamie E. Marshall, Bayan H. A. Faraj, Alexandre R. Gingras, et al. The Crystal Structure of Pneumolysin at 2.0 Å Resolution Reveals the Molecular Packing of the Pre-pore Complex. *Scientific Reports*, 5:13293, 2015. doi: 10.1038/srep13293.

E A Merritt, S Sarfaty, F van den Akker, et al. Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Science*, 3(2):166–175, 1994. doi: 10.1002/pro.5560030202.

E A Merritt, S Sarfaty, I K Feil, et al. Structural foundation for the design of receptor antagonists targeting Escherichia coli heat-labile enterotoxin. *Structure*, 5(11):1485–1499, 1997. doi: 10.1016/s0969-2126(97)00298-0.

Thi-Sau Migone, G. Mani Subramanian, John Zhong, et al. Raxibacumab for the Treatment of Inhalational Anthrax. *New England Journal of Medicine*, 361(2):135–144, 2009. doi: 10.1056/NEJMoa0810603.

Valentina Monaci, Davide Oldrini, Gianmarco Gasperini, et al. Biological and structural characterization of the Type 3 fimbrial subunit MrkA from Klebsiella pneumoniae. *Protein Science*, 34 (11):e70343, 2025. doi: 10.1002/pro.70343.

Neeloffer Mookherjee, Kelly L. Brown, Dawn M. E. Bowdish, Silvana Doria, Reza Falsafi, et al. Modulation of the TLR-mediated inflammatory response by the endogenous human host defense peptide LL-37. *The Journal of Immunology*, 176(4):2455–2464, 2006. doi: 10.4049/jimmunol.176.4.2455.

Beenu Moza, Ashok K. Varma, Rebecca A. Buonpane, Penny Zhu, Christine A. Herfst, Melissa J. Nicholson, Anne-Kathrin Wilbuer, Nilufer P. Seth, Kai W. Wucherpfennig, John K. McCormick, David M. Kranz, and Eric J. Sundberg. Structural basis of T-cell specificity and activation by the bacterial superantigen TSST-1. *The EMBO Journal*, 26(4):1187–1197, 2007. doi: 10.1038/sj.emboj.7601531.

Christopher J. L. Murray, Kevin Shunji Ikuta, Fablina Sharara, et al. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, 399(10325):629–655, 2022. doi: 10.1016/S0140-6736(21)02724-0.

Daniel C. Nelson, Julia Garbe, and Mattias Collin. Cysteine proteinase SpeB from Streptococcus pyogenes—a potent modifier of immunologically important host and bacterial proteins. *Biological Chemistry*, 392(12):1077–1088, 2011. doi: 10.1515/BC.2011.208.

Dragana Nesic, Ludovico Buti, Xin Lu, and C. Erec Stebbins. Structure of the Helicobacter pylori CagA oncoprotein bound to the human tumor suppressor ASPP2. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4):1562–1567, 2014. doi: 10.1073/pnas.1320631111.

Nhu T Q Nguyen, Thien N M Doan, Kei Sato, et al. Monoclonal antibodies neutralizing alpha-hemolysin, bicomponent leukocidins, and clumping factor A protected against Staphylococcus aureus-induced acute circulatory failure in a mechanically ventilated rabbit model of hyperdynamic septic shock. *Frontiers in Immunology*, 14:1260627, 2023. doi: 10.3389/fimmu.2023.1260627.

David Ochoa, Andrew Hercules, Miguel Carmona, Daniel Suveges, Asier Gonzalez-Uriarte, et al. Open Targets Platform: Supporting systematic drug–target identification and prioritisation. *Nucleic Acids Research*, 49(D1):D1302–D1310, 2021. doi: 10.1093/nar/gkaa1027.

Peter Orth, Li Xiao, Lorraine D. Hernandez, et al. Mechanism of Action and Epitopes of Clostridium difficile Toxin B-neutralizing Antibody Bezlotoxumab Revealed by X-ray Crystallography. *Journal of Biological Chemistry*, 289(26):18008–18021, 2014. doi: 10.1074/jbc.M114.560748.

Jun-Young Park, Fukushi Abekura, and Seung-Hak Cho. GM1a ganglioside-binding domain peptide inhibits host adhesion and inflammatory response of enterotoxigenic Escherichia coli heat-labile enterotoxin-B in HCT-8 cells. *Scientific Reports*, 13(1):16835, 2023. doi: 10.1038/s41598-023-44220-5.

Malak Pirtskhalava, Anthony A. Amstrong, Maia Grigolava, Mindia Chubinidze, Evgenia Alimbarashvili, Boris Vishnepolsky, Andrei Gabrielian, Alex Rosenthal, Darrell E. Hurt, and Michael Tartakovsky. DBAASP v3: Database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Research*, 49(D1):D288–D297, 2021. doi: 10.1093/nar/gkaa991.

N. Quarto and F. Amalric. Heparan sulfate proteoglycans as transducers of FGF-2 signalling. *Journal of Cell Science*, 107(11):3201–3212, 1994. doi: 10.1242/jcs.107.11.3201.

Lauriane E. Quenee, Bryan J. Berube, Joshua Segal, et al. Amino acid residues 196–225 of LcrV represent a plague protective epitope. *Vaccine*, 28(7):1870–1876, 2010. doi: 10.1016/j.vaccine.2009.11.076.

Fatemeh Rahimi-Jamnani, Hamid Reza Moradi, Abolfazl Fateh, et al. The Therapeutic Efficacy of MS473, a Fully Human Single-Chain Variable Fragment Targeting Staphylococcus aureus Toxic Shock Syndrome Toxin-1, in a D-Galactosamine-Sensitized Mouse Model of Lethal Shock. *BMC Microbiology*, 25(1):95, 2025. doi: 10.1186/s12866-025-03825-8.

Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He. ZeRO: Memory optimizations toward training trillion parameter models. In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–16, 2020. doi: 10.1109/SC41405.2020.00024.

Anand Singh Rathore, Nishant Kumar, Shubham Choudhury, Naman Kumar Mehta, and Gajendra P. S. Raghava. Prediction of hemolytic peptides and their hemolytic concentration. *Communications Biology*, 8:176, 2025. doi: 10.1038/s42003-025-07615-w.

Thunchanok Rukkawattanakul, Nitat Sookrung, Watee Seesuay, et al. Human scFvs That Counteract Bioactivities of Staphylococcus aureus TSST-1. *Toxins*, 9(2):50, 2017. doi: 10.3390/toxins9020050.

Iraj Saadat, Hideaki Higashi, Chikashi Obuse, et al. Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature*, 447(7142):330–333, 2007. doi: 10.1038/nature05765.

Eric W. Sayers, Jeffrey Beck, Evan E. Bolton, J. Rodney Brister, Jessica Chan, et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, 53(D1):D20–D29, 2025. doi: 10.1093/nar/gkae979.

Monisha G. Scott, Edie Dullaghan, Neeloffer Mookherjee, Natalie Glavas, Matthew Waldbrook, et al. An anti-infective peptide that selectively modulates the innate immune response. *Nature Biotechnology*, 25:465–472, 2007. doi: 10.1038/nbt1288.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. doi: 10.48550/arXiv.2402.03300.

Saroj Sharma and Vishvanath Tiwari. Autophagy modulation by OmpA-targeted lead compounds restores autophagic flux, reduces biofilm, and enhances immune response in Acinetobacter baumannii-infected pulmonary cells. *Microbiology Spectrum*, 14(7):e0003626, 2026. doi: 10.1128/spectrum.00036-26.

C. A. Staton, M. Valluru, L. Hoh, M. W. R. Reed, and N. J. Brown. Angiopoietin-1, angiopoietin-2 and Tie-2 receptor expression in human dermal wound repair and scarring. *British Journal of Dermatology*, 163(5):920–927, 2010. doi: 10.1111/j.1365-2133.2010.09940.x.

P E Stein, A Boodhoo, G D Armstrong, et al. The crystal structure of pertussis toxin. *Structure*, 2 (1):45–57, 1994. doi: 10.1016/s0969-2126(00)00007-1.

Paulina Szymczak, Marcin Możejko, Tomasz Grzegorzek, Radosław Jurczak, Marta Bauer, Damian Neubauer, Karol Sikora, Michał Michalski, Jacek Sroka, Piotr Setny, Wojciech Kamysz, and Ewa Szczurek. Discovering highly potent antimicrobial peptides with deep generative model HydrAMP. *Nature Communications*, 14(1):1453, 2023. doi: 10.1038/s41467-023-36994-z.

Masaya Takehara, Yuta Homma, Tomoaki Ishihara, et al. Repurposing caspofungin as a small-molecule inhibitor of Clostridium perfringens α-toxin for treatment of gas gangrene. *Communications Medicine*, 6(1):225, 2026. doi: 10.1038/s43856-026-01503-y.

Qing Tang, Ho Yung Chan, Yanxi Huang, et al. Mechanisms of Pertussis Toxin Action: ADP-Ribosylation and Its Role in Pertussis Pathogenesis. *Toxins*, 18(3):148, 2026. doi: 10.3390/toxins18030148.

The UniProt Consortium. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Research*, 53(D1):D609–D617, 2025. doi: 10.1093/nar/gkae1010.

Christoph Thomas, J Fernando Bazan, and K Christopher Garcia. Structure of the activating IL-1 receptor signaling complex. *Nature Structural & Molecular Biology*, 19(4):455–457, 2012. doi: 10.1038/nsmb.2260.

Keya Trivedi and Christopher N LaRock. Pentamidine inhibition of streptopain attenuates Streptococcus pyogenes virulence. *Microbiology Spectrum*, 13(8):e0075825, 2025. doi: 10.1128/spectrum.00758-25.

Hiroshi Uchi, Atsuyuki Igarashi, Kazunori Urabe, et al. Clinical efficacy of basic fibroblast growth factor (bFGF) for diabetic ulcer. *European Journal of Dermatology*, 19(5):461–468, 2009. doi: 10.1684/ejd.2009.0750.

Mihaly Varadi, Damian Bertoni, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna, et al. AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1):D368–D375, 2024. doi: 10.1093/nar/gkad1011.

Norberth-Istvan Varga, Iulia Cristina Bagiu, Dan Dumitru Vulcanescu, et al. IL-6 Baseline Values and Dynamic Changes in Predicting Sepsis Mortality: A Systematic Review and Meta-Analysis. *Biomolecules*, 15(3):407, 2025. doi: 10.3390/biom15030407.

Fangping Wan, Marcelo D. T. Torres, Jacqueline Peng, and Cesar de la Fuente-Nunez. Deep-learning-enabled antibiotic discovery through molecular de-extinction. *Nature Biomedical Engineering*, 8:854–871, 2024. doi: 10.1038/s41551-024-01201-x.

Dongli Wang, Senyan Zhang, Liang Li, et al. Structural insights into the assembly and activation of IL-1β with its receptors. *Nature Immunology*, 11(10):905–911, 2010. doi: 10.1038/ni.1925.

Fanhao Wang, Yuzhe Wang, Laiyi Feng, Changsheng Zhang, and Luhua Lai. Target-specific de novo peptide binder design with DiffPepBuilder. *Journal of Chemical Information and Modeling*, 64 (24):9135–9149, 2024. doi: 10.1021/acs.jcim.4c00975.

Yi Wang, Xiaoou Shan, Gaozhi Chen, et al. MD-2 as the target of a novel small molecule, L6H21, in the attenuation of LPS-induced inflammatory response and sepsis. *British Journal of Pharmacology*, 172(17):4391–4405, 2015. doi: 10.1111/bph.13221.

Jacob Witten and Zack Witten. Deep learning regression model for antimicrobial peptide design. *bioRxiv*, 2019. doi: 10.1101/692681.

Janina Wolf, Stefan Rose-John, and Christoph Garbers. Interleukin-6 and its receptors: A highly regulated and dynamic system. *Cytokine*, 70(1):11–20, 2014. doi: 10.1016/j.cyto.2014.05.024.

John A. T. Young and R. John Collier. Anthrax toxin: Receptor binding, internalization, pore formation, and translocation. *Annual Review of Biochemistry*, 76:243–265, 2007. doi: 10.1146/annurev.biochem.75.103004.142728.

Hai Tao Yuan, Eliyahu V. Khankin, S. Ananth Karumanchi, and Samir M. Parikh. Angiopoietin 2 Is a Partial Agonist/Antagonist of Tie2 Signaling in the Endothelium. *Molecular and Cellular Biology*, 29(8):2011–2022, 2009. doi: 10.1128/MCB.01472-08.

Meng-Yue Zhang, Shuang Li, Yu-Ling Han, Yi-Fan Shi, Ying-Ying Wu, Juan Cheng, Cai-Yun Wang, Xun-Yong Zhou, and Yi-Xuan Zhang. De novo-designed amphiphilic α-helical peptide Z2 exhibits broad-spectrum antimicrobial, antibiofilm, and anti-inflammatory efficacy in acute Pseudomonas aeruginosa pneumonia. *Bioorganic Chemistry*, 157:108309, 2025. doi: 10.1016/j.bioorg.2025.108309.

R G Zhang, M L Westbrook, E M Westbrook, et al. The 2.4 A crystal structure of cholera toxin B subunit pentamer: choleragenoid. *Journal of Molecular Biology*, 251(4):550–562, 1995. doi: 10.1006/jmbi.1995.0455.

X. Zhang, A. D. McDaniel, L. E. Wolf, G. T. Keusch, M. K. Waldor, and D. W. Acheson. Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *Journal of Infectious Diseases*, 181(2):664–670, 2000. doi: 10.1086/315239.

Liqun Zhao, Zhen Song, Hongyin Fan, et al. A Synergistic Dual-Antibody Cocktail Targeting α-Toxin Protects Against Invasive Staphylococcus aureus Infection by Neutralizing Virulence and Enhancing Host Defense. *MedComm*, 7(7):e70830, 2026a. doi: 10.1002/mco2.70830.

Pinnan Zhao, Zongcheng Li, Chaoyan Yao, et al. Clostridium perfringens alpha toxin drives pathological NETosis via immature neutrophil mobilization and functional reprogramming. *Acta Pharmaceutica Sinica B*, 16(2):879–899, 2026b. doi: 10.1016/j.apsb.2025.09.011.

APPENDIX CONTENTS

A Extend Experiment . . . 17
A.1 Single axes are predicted well; their conjunction is not . . . 17
A.2 Prediction sharpens within an episode . . . 17
A.3 Every scenario improves on its starting peptide . . . 17
B Attribution analysis of the TLR4 scenario . . . 19
C Target selection and the biology of the second function . . . 22
C.1 Restraining the host response . . . 22
C.2 Disarming the pathogen . . . 23
C.3 Sustaining repair . . . 24
C.4 Target-set composition . . . 24
D Group Relative Policy Optimization and the supervised cold start . . . 24
D.1 The optimization algorithm . . . 24
D.2 Why a supervised cold start is necessary . . . 25
E Building a corpus that supports policy optimization . . . 26
E.1 Corpus requirements . . . 26
E.2 Sources and item families . . . 26
E.3 Leakage and shortcut controls . . . 26
E.4 Selectivity distractor construction . . . 27
E.5 Shortcut audit and exclusion . . . 27
F Sample redesign trajectories . . . 27
G Algorithmic specification . . . 31
G.1 The redesign loop . . . 31
G.2 Outer pipeline . . . 32
H Memory accumulation across episodes . . . 32
H.1 What changes and what does not . . . 32
H.2 Within an episode . . . 33
H.3 Across episodes . . . 33
I Grounded Accumulation Memory . . . 33
I.1 Why two layers . . . 33
I.2 The page-store . . . 33
I.3 The index . . . 34
I.4 Retrieval . . . 34
J Binding specificity against decoy receptors . . . 34
K Re-scoring with independent predictors . . . 34

A EXTEND EXPERIMENT

Every candidate is proposed with a committed direction of change, recorded before any evaluator sees it. Memory keeps that commitment beside the measurement the candidate subsequently received, so the two can be compared after the fact. This appendix reports that comparison. It is the one analysis in the paper that examines what the system believed rather than what it produced.

One qualification governs everything below. The measurements against which predictions are checked are themselves predictor outputs, not assay results. What is quantified here is therefore agreement between the Redesigner's stated expectation and the scoring functions it is optimized against, not agreement with experiment.

A.1 SINGLE AXES ARE PREDICTED WELL; THEIR CONJUNCTION IS NOT

Direction is predicted above chance on all three axes even before training, and training improves each (Fig. 5a). Antimicrobial direction is predicted best throughout, reaching 0.78 after SFT and GRPO, which is unsurprising: the relationship between cationicity, hydrophobicity and potency is close to monotone and is the relationship most densely represented in the literature the backbone was pretrained on. Binding and hemolysis follow at 0.69 and 0.66. All three sit well above the 0.50 chance level of a binary direction call.

The joint outcome behaves differently (Fig. 5b). Asking whether a proposed edit will produce a candidate satisfying all three criteria at once, the hit rate is 0.09 zero-shot and 0.27 after optimization, against a joint chance level of 0.125.

This gap is the substantive result. If the three axes were independent, a mean single-axis rate of 0.71 would imply a joint rate near 0.36; the observed 0.27 is lower, and the shortfall is not noise but the shape of the problem. An edit that the Redesigner correctly expects to improve potency is frequently the same edit that it correctly expects to worsen hemolysis, and knowing both does not by itself resolve which to prefer. The system can say what each axis will do and still be wrong about what the combination will do.

The gap widens before it narrows. Supervised fine-tuning raises single-axis accuracy faster than joint accuracy, taking the difference from 0.46 to 0.49; GRPO then narrows it to 0.44, the first phase of training to improve the conjunction faster than its parts. This is what the reward is constructed to do. Supervision rewards reproducing a correct answer, which on a single-axis item means learning a single-axis relationship, whereas the outcome reward is indifferent to how an answer was reached and credits only whether the joint criterion was met. The narrowing is small, and we report it as a direction of movement rather than as a solved problem.

A.2 PREDICTION SHARPENS WITHIN AN EPISODE

Both rates rise as an episode proceeds (Fig. 5c). Mean single-axis accuracy goes from 0.53 at round 0 to 0.76 by round 7, and the joint rate from 0.08 to 0.26.

The curves are not a controlled comparison. An episode terminates on its first passing candidate, so the set of episodes contributing to each round shrinks and shifts toward the harder ones: n falls from 26 at round 0 to 9 by round 7, and rounds beyond 7 are omitted because fewer than eight episodes remain. The rise therefore confounds two things we cannot separate here, sharpening within an episode and a change in which episodes are being averaged. The second would if anything work against the observed direction, since the surviving episodes are those a passing candidate has so far eluded, but we report the trend as a description rather than as evidence that memory sharpens prediction.

A.3 EVERY SCENARIO IMPROVES ON ITS STARTING PEPTIDE

Figure 5d places each structure-based scenario's starting peptide beside its delivered candidate. All twelve improve, but the magnitude varies by more than an order of magnitude. The *C. perfringens* α-toxin scenario moves furthest, from above 200 to 2.7 μM; TLR4 moves least in absolute terms because its starting peptide was already close to the threshold, so the work in that episode went into hemolysis rather than potency, which this panel does not show.

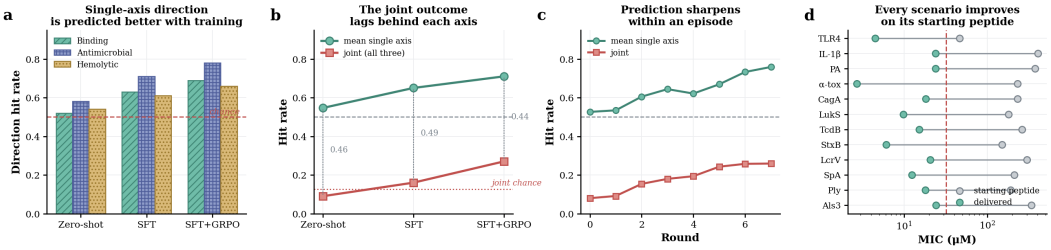

Figure 5: **Prediction audit.** **a,** Direction hit rate on each axis under three training conditions; dashed line is the 0.50 chance level for a binary call. **b,** Mean single-axis hit rate against the joint rate, the probability that an edit produces a candidate satisfying all three criteria; grey annotations give the difference, and the dotted line is the joint chance level of 0.125. **c,** Both rates by round. Episodes terminate on their first passing candidate, so the number contributing falls across rounds (26 at round 0, 9 at round 7); rounds beyond 7 are omitted. **d,** MIC of the starting peptide and of the delivered candidate for each structure-based scenario; dashed line marks the antimicrobial threshold.

Memory	Redesigner	Solved	Rounds to pass	Final Sel.	Repeat rate	Joint hit
✗	Base	3/8	$11.4_{2.1}$	$4.2_{1.8}$	0.31	0.08
✗	SFT+GRPO	5/8	$8.7_{1.9}$	$6.9_{2.4}$	0.28	0.19
✓	Base	6/8	$6.2_{1.6}$	$8.1_{3.0}$	0.09	0.13
✓	SFT+GRPO	**8/8**	$\mathbf{3.4}_{1.1}$	$\mathbf{12.7}_{4.1}$	**0.04**	**0.27**

Table 3: **Memory and optimization are not substitutes.** Eight scenarios under each condition. *Solved* counts scenarios delivering a passing candidate within the round budget; *Repeat rate* is the fraction of proposals already recorded as failures; *Joint hit* is the prediction-audit joint rate. Neither factor alone reaches the combined result.

The panel is a check rather than a result. A loop that terminates on a passing candidate will by construction end somewhere better than it began on at least one axis, so improvement here is expected. What the panel establishes is that the improvement is not confined to one or two fortunate scenarios, and that the starting peptides were genuinely inadequate rather than near-solutions that needed cosmetic adjustment.

Memory and optimization are not substitutes (Table 3). The claim that the two mechanisms act at different timescales requires a factorial comparison, since either alone would produce some improvement over neither. We ran eight scenarios under each of four conditions, crossing memory on or off against a base or optimized Redesigner, holding the starting peptide, batch size and round budget fixed. With neither, three of eight scenarios deliver. Optimization alone raises this to five and memory alone to six, so each factor helps on its own; together they reach eight, and the round count falls further than either factor accounts for. The repeat rate separates them most cleanly. Without memory, roughly three proposals in ten restate a sequence already recorded as a failure, and optimization barely changes this, since a policy cannot avoid repeating what it has no record of. With memory the rate falls to 0.09, and with both to 0.04. Memory supplies what was tried; optimization decides what to do about it.

Reference peptides scored through the same pipeline (Table 4). Selectivity ratios are hard to interpret without a scale, so we passed five well-characterized peptides through the same predictors under identical settings. LL-37 and magainin-2 return ratios near 6, pexiganan 7.5, and indolicidin 4.7, placing the delivered candidates' median of 8.0 slightly above the reference set and the best candidate an order of magnitude beyond it. Melittin is included for the opposite reason: it is potent and severely hemolytic, and the pipeline returns a ratio of 0.7, confirming that the hemolysis predictor discriminates rather than assigning uniformly safe values to short cationic sequences. Two caveats apply. These are predictions, not the published assay values for these peptides, so the comparison is internal to the scoring functions. And the reference peptides are broad-spectrum antimicrobials

Peptide	Note	MIC (µM)	HC_{50} (µM)	Sel.
LL-37	human cathelicidin	18.4	108.3	5.9
Magainin-2	amphibian, broad	31.7	187.4	5.9
Pexiganan	clinical, topical	12.6	94.2	7.5
Indolicidin	bovine, Trp-rich	8.9	41.6	4.7
Melittin	hemolytic control	4.2	2.8	0.7
JANUS-PEP delivered (n=26), median		17.9	136.9	8.0
JANUS-PEP delivered, best		2.7	216.6	**54.0**

Table 4: **Reference peptides scored through the same pipeline.** Known peptides passed through APEX and HemoPI2 under identical settings, so the numbers are comparable with Table 2 but are predictions, not the published assay values. Melittin is included as a hemolytic negative control.

Decoy target	Why it should fail	Class	Accessible	Effector	Admitted
SHP2 (Q06124)	cytosolic phosphatase	intracellular	×	✓	×
NF-κB p65	nuclear transcription factor	intracellular	×	✓	×
GAPDH (P04406)	housekeeping enzyme	housekeeping	×	×	×
CRP (P02741)	acute-phase marker	correlate	✓	×	×
Procalcitonin	sepsis biomarker	correlate	✓	×	×
β-actin	cytoskeletal	housekeeping	×	×	×
Admitted: 1/15 decoys, 26/26 genuine targets					

Table 5: **Gate specificity.** Fifteen deliberately inappropriate targets were embedded in otherwise plausible scenarios; six representative cases shown. A gate that never rejects is not a gate, so the relevant number is the rejection rate, not the pass rate on genuine targets.

without a second function, so they bound only one of the three axes on which the delivered candidates are judged.

Gate specificity (Table 5). All twenty-six genuine targets passed the three admission gates, which on its own establishes nothing: a gate that never rejects is indistinguishable from no gate. We therefore embedded fifteen deliberately inappropriate targets in otherwise plausible scenarios, drawn from three failure classes. Intracellular proteins such as SHP2 and NF-κB p65 are genuine effectors of the pathologies they appear in and fail only on accessibility, which makes them the strictest test, since the scenario text around them reads as sound. Correlates such as CRP and procalcitonin invert the difficulty: they are extracellular and abundant, and fail only because they mark the pathology rather than cause it. Housekeeping proteins such as GAPDH and β-actin fail on both counts and serve as a floor. Fourteen of fifteen were rejected. The single admission was a correlate rather than an intracellular protein, which is consistent with the effector gate being the weaker of the two, since it requires a judgement about causation that the accessibility gate can resolve from annotation alone.

B ATTRIBUTION ANALYSIS OF THE TLR4 SCENARIO

Independent hemolysis model. Most computational tools for antimicrobial peptide development treat red-blood- cell lysis as a binary classification, reporting whether a sequence is hemolytic without saying how much. AmpLyze takes a different approach. It predicts the HC_{50} value itself from sequence alone, coupling residue-level ProtT5 and ESM2 embeddings with sequence-level descriptors in parallel local and global branches, aligned by a cross-attention module and trained with a log-cosh loss chosen for robustness to assay noise. On held-out data it reaches a Pearson correlation of 0.756. Alongside the prediction it produces Expected-Gradients attributions, which assign to each residue a signed contribution to the predicted toxicity, so that the model reports not only a number but the positions responsible for it.

AmpLyze and the Diagnoser independently identify consequential positions using different inputs, training data, and scoring functions. We therefore compare whether the two methods identify the same residues.

Setup. We followed the TLR4 scenario in detail. The first passing candidate reached an interface energy of -5.1 REU, an MIC of 4.5 μM and an HC_{50} of 118.6 μM in round 3; the attribution analysis includes candidates through round 4. Every intermediate candidate was re-scored by AmpLyze, which took no part in selection; that role was played by HemoPI2. Positions whose attribution falls in the top quartile of a given peptide are designated hotspots, and a positive attribution marks a residue that increases predicted hemolysis.

Agreement across rounds. In the first two rounds the agreement is exact. Both positions edited in round 1 were hotspots, and so were all three edited by round 2. From round 3 the picture changes: one of five cumulative edits falls outside the hotspot set, and two of seven by round 4.

Positions 7 and 11 carry the highest attribution throughout and are never touched. Both lie on the face that contacts the receptor, so relieving hemolysis there would cost the binding the scenario was built to achieve. The edits made outside hotspots, at positions 5 and 12, serve the antimicrobial and binding axes, which AmpLyze does not score. The correspondence becomes partial not because the diagnosis deteriorates but because the hotspots that remain are the ones the other two objectives will not permit.

Attribution changes after editing. Where edits were made, attribution falls. Position 4 drops from $+0.019$ in round 1, above the hotspot threshold, to $+0.003$ in round 4; position 9 from $+0.014$ to $+0.006$; position 13 from $+0.014$ to $+0.001$. The number of unedited hotspots falls from five to two over the same interval, and the predicted HC_{50} rises over the same interval.

Scope. This analysis covers one scenario and one of the three design objectives. Within that scope, Diagnoser-selected positions overlap residues identified as consequential by an independent model, while high-attribution positions left unedited occur at the receptor interface.

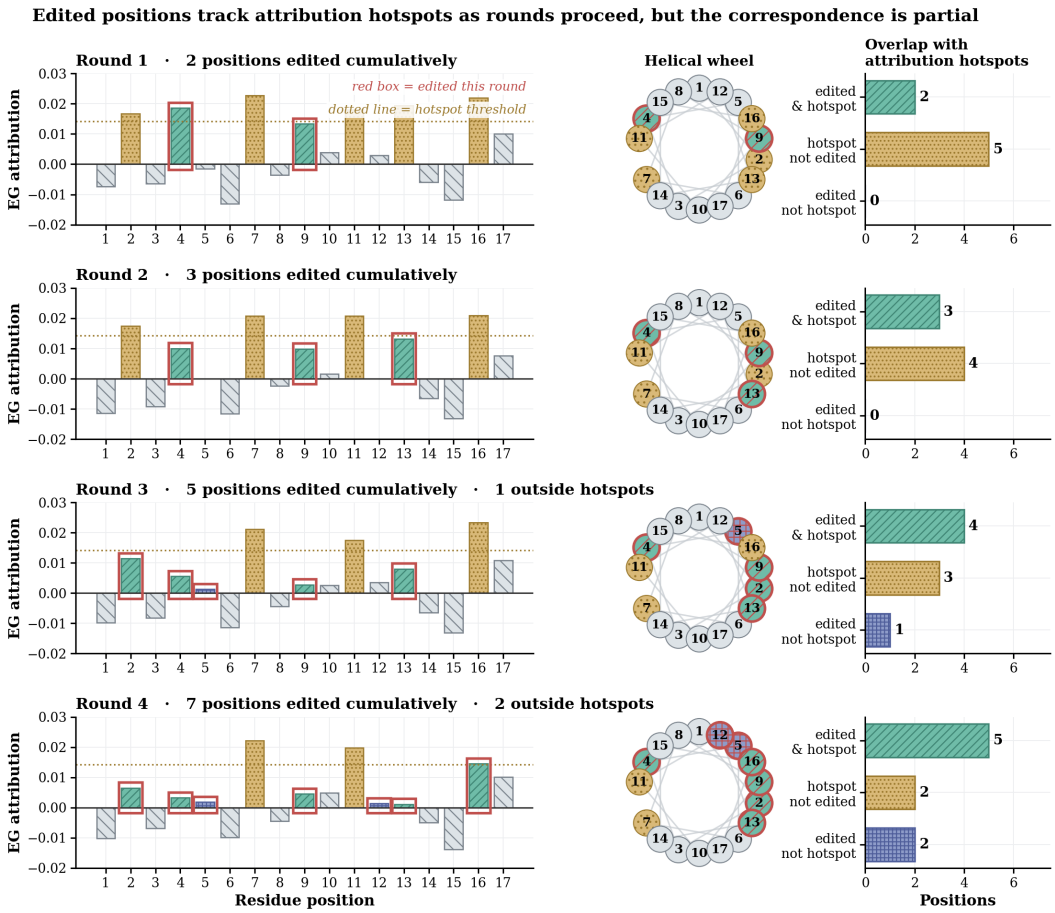

Figure 6: **Edited positions track independently attributed toxicity hotspots (TLR4 scenario).** Rows show four consecutive rounds of redesign. **Left,** per-residue Expected-Gradients attribution from AmpLyze, an external hemolysis predictor that took no part in candidate selection. Dotted line marks the hotspot threshold (top quartile of attribution); gold bars exceed it, green hatched bars are positions the Diagnoser has edited, and red outlines mark positions edited in that round. **Middle,** helical wheel with edited positions circled in red. **Right,** decomposition of the correspondence into three counts: positions that are both edited and hotspots, hotspots left unedited, and edits made outside hotspots. Every edit in the first two rounds falls on a hotspot; edits outside hotspots appear from round 3.

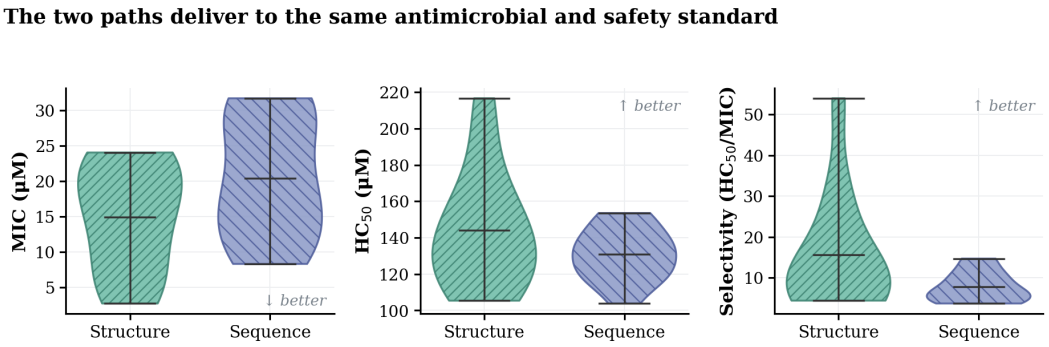

Figure 7: **Comparison of delivered-candidate metrics across design paths.** Distributions of the three delivered-candidate metrics on each path. Hemolytic concentration is comparable across paths (medians 144 and 131 μM), and every candidate on both paths satisfies the antimicrobial and hemolytic bounds. The structure-based distribution carries a longer upper tail in selectivity, contributed by a small number of candidates whose targets permit direct optimization of interface energetics; the bulk of both distributions lies well above the threshold. The sequence-based path has a lower upper range of selectivity while satisfying the same activity and safety thresholds.

C TARGET SELECTION AND THE BIOLOGY OF THE SECOND FUNCTION

Each scenario given to JANUS-PEP specifies an infection and a second function in addition to antimicrobial activity. The Planner maps this function to an extracellularly accessible protein involved in the pathology. The twenty-six targets are grouped by whether the second function modulates the host response, disarms the pathogen, or supports tissue repair.

C.1 RESTRAINING THE HOST RESPONSE

Severe infection can cause injury through the host inflammatory response. During Gram-negative sepsis, bacterial killing can release lipopolysaccharide, motivating a design that combines antimicrobial activity with host-response modulation. Host defence peptides such as LL-37 provide a precedent for combining these activities (Mookherjee et al., 2006; Scott et al., 2007).

TLR4 (S1, O00206), with its co-receptor MD-2, converts LPS into inflammation. Signalling requires assembly of a $(\text{TLR4/MD-2/LPS})_2$ dimer before MyD88 can be recruited (Fu et al., 2025). The design therefore targets the extracellular dimerization surface. Separately, MD-2-directed inhibitors suppress TLR4 signalling and attenuate LPS-driven sepsis in vivo (Wang et al., 2015; Bai et al., 2023), providing pharmacological precedent for disrupting this receptor complex. Design proceeded from the TLR4/MD-2 complex (PDB 4G8A).

IL-1β (S2, P01584) is a mediator of the septic inflammatory cascade (Ge et al., 2019). It signals by binding IL-1RI and then recruiting IL-1RAcP into a ternary complex (Thomas et al., 2012; Wang et al., 2010). The natural antagonist IL-1Ra occupies IL-1RI without recruiting IL-1RAcP and therefore blocks productive receptor assembly (Thomas et al., 2012). Design proceeded from PDB 1ITB.

IL-6 (Q1, P05231) and **IL-17A** (Q2, Q16552) represent two additional inflammatory pathways. IL-6 drives the acute-phase response through classic IL-6Rα/gp130 signalling (Wolf et al., 2014). Circulating IL-6 has moderate discriminatory value for 28-day sepsis mortality but is not a strong standalone predictor (Varga et al., 2025). IL-17A signalling through IL-17RA contributes to pulmonary host defence but can also drive infection-associated immunopathology (Lorè et al., 2016). These extracellular ligand–receptor interfaces therefore provide test cases for combining pathway modulation with antimicrobial activity.

C.2 DISARMING THE PATHOGEN

Several scenarios target secreted virulence proteins that can remain active after bacterial killing. Raxibacumab, an antibody against anthrax protective antigen, provides a precedent for toxin neutralization (Migone et al., 2009).

Protective antigen (S3, P13423) is the sole route by which anthrax lethal and edema factors enter the cytosol: it binds its receptor, heptamerizes, and translocates both effectors through the resulting pore (Young & Collier, 2007). Antibody blockade of protective antigen therefore neutralizes both toxin arms and improves survival after inhalational-anthrax challenge (Migone et al., 2009).

Heat-labile enterotoxin B (Q3, A7Y1W2) and **cholera toxin B** (Q5, P01556) present pentameric B subunits that anchor to the GM1 ganglioside before the catalytic A subunit can enter (Merritt et al., 1994; Zhang et al., 1995); the GM1 pocket is structurally defined, has been used for antagonist design (Merritt et al., 1997), and peptides directed at it inhibit LT-B adhesion experimentally (Park et al., 2023). **Pertussis toxin S1** (Q7, P04977) is the catalytic subunit of the same AB_5 family and ADP-ribosylates host G_i proteins from a well-resolved active site (Stein et al., 1994; Tang et al., 2026). **Shiga toxin B** (S8) uses the same architecture with Gb3 in place of GM1 (Ling et al., 1998). Because phage-inducing quinolones can increase Shiga-toxin production and mortality in vivo (Zhang et al., 2000), this scenario combines bacterial killing with toxin neutralization. Design proceeded from the B-pentamer bound to a Gb3 analogue (PDB 1BOS) (Ling et al., 1998).

Three further toxins act by assembly on the host membrane. **α-hemolysin** (Q4, P09616) oligomerizes into a heptameric pore, and neutralizing it protects against staphylococcal septic shock (Nguyen et al., 2023; Zhao et al., 2026a; Kim et al., 2026). **LukS-PV** (S6, Q783R1) is the S component of Panton-Valentine leukocidin, a bicomponent toxin that lyses neutrophils and is associated with necrotizing pneumonia; receptor-binding residues have been localized by alanine scanning and structural analysis (Laventie et al., 2014), and design proceeded from PDB 1T5R. **Pneumolysin** (S11, Q2XU26) is a cholesterol-dependent cytolysin whose domain 4 carries the tryptophan-rich loop that engages the membrane; design proceeded from the full-length structure (PDB 5CR6) (Marshall et al., 2015). **TSST-1** (Q6, P06886) is a superantigen whose toxicity requires bridging MHC-II to the T-cell receptor simultaneously, driving the polyclonal activation that produces toxic shock (Moza et al., 2007). Antibodies that bind exposed TSST-1 surfaces neutralize its activity and protect in experimental models (Rukkawattanakul et al., 2017; Rahimi-Jamnani et al., 2025). **TcdB** (S7, P18177) is the principal virulence factor of *C. difficile*; its CROPs domain is the epitope of bezlotoxumab, an approved neutralizing antibody. Design proceeded from the crystallographically resolved complex (PDB 4NP4), whose structure explains how antibody binding blocks toxin attachment to host cells (Orth et al., 2014).

Other scenarios target virulence machinery rather than bacterial viability, a strategy that may impose weaker selection for resistance. **α-toxin** of *C. perfringens* (S4, P0C216) is a zinc-dependent phospholipase C whose hydrolysis of membrane phospholipid produces the muscle necrosis of gas gangrene; inhibiting its active site attenuates the disease (Takehara et al., 2026; Zhao et al., 2026b). **CagA** (S5, P55980) is a bacterial oncoprotein linked to gastric carcinogenesis. Once injected into gastric epithelium, it engages SHP2, ASPP2 and PAR1/MARK (Higashi et al., 2002; Buti et al., 2011; Saadat et al., 2007). Design proceeded from the CagA–ASPP2 complex (PDB 4IRV) (Nesic et al., 2014). **IpaB** (Q8, P18011), **PcrV** (Q10, M4Q7H6) and **LcrV** (S9, P0C7U7) are components of type III secretion systems in *Shigella*, *P. aeruginosa* and *Y. pestis*, respectively. IpaB lines the translocon pore (Chen et al., 2021; Dickenson et al., 2013); antibodies to PcrV suppress toxin delivery and protect mice (Desveaux et al., 2026; Cui et al., 2025), while an LcrV monoclonal antibody neutralizes secretion and protects mice from lethal plague challenge (Quenee et al., 2010). LcrV and PcrV are homologues in different species, which makes the pair an internal test of whether a design transfers. **SpeB** (Q9, P0C0J0) is the cysteine protease with which *S. pyogenes* degrades host matrix and immune proteins; small-molecule inhibition of its catalytic groove attenuates virulence (Nelson et al., 2011; Trivedi & LaRock, 2025).

Staphylococcal protein A (S10, P38507) binds the Fc region of IgG and thereby limits opsonization. Occupying this interface is intended to restore immune recognition rather than neutralize a toxin. Design proceeded from the SpA–IgG Fc complex (PDB 4WWI) (Deis et al., 2015).

Three targets mediate adhesion associated with biofilm formation in chronic and device-associated infection. **MrkA** (Q11, D5FKS5) is the major subunit of the *K. pneumoniae* type 3 fimbria (Monaci

et al., 2025), against which antibodies and subunit vaccines are protective (Berry et al., 2024; Assoni et al., 2025). **OmpA** (Q12, A9Z1A2) mediates adhesion, serum resistance and biofilm formation in *A. baumannii* and has been reviewed as a therapeutic target with leads that reduce biofilm (Islam et al., 2026; Sharma & Tiwari, 2026). **Als3** (S12, Q59L12) is the *C. albicans* adhesin, included because mixed fungal-bacterial wounds are common and a peptide that kills bacteria while blocking fungal adhesion extends the framework beyond a single kingdom; its N-terminal peptide-binding cavity is resolved and its key residues are validated by mutagenesis (Lin et al., 2014).

C.3 Sustaining repair

Chronic diabetic or ischaemic wounds combine infection with impaired repair. These scenarios therefore pair antimicrobial activity with support of pro-repair signalling. **Angiopoietin-2** (Q13, O15123) is a context-dependent partial agonist/antagonist of Tie2 (Yuan et al., 2009). The Ang/Tie2 system is expressed during human wound healing and correlates with wound angiogenesis (Staton et al., 2010); this scenario asks whether tuning Ang-2 can favor vascular maturation. **FGF2** (Q14, P09038) drives fibroblast and endothelial proliferation through FGFR and heparan sulfate. Topical recombinant FGF2 accelerated diabetic-ulcer closure in a randomized, double-blind trial (Uchi et al., 2009), while heparan-sulfate binding protects FGF2 from degradation and supports receptor activation (Quarto & Amalric, 1994). The design objective is therefore to stabilize local FGF2 without occluding its receptor- or co-receptor-binding surfaces.

C.4 Target-set composition

Six of the twenty-six targets are host proteins and twenty are microbial. Host-directed targets modulate infection-associated responses, whereas pathogen-directed targets neutralize virulence mechanisms. The target set thus spans both sides of the host–pathogen interface while retaining broad antimicrobial activity as the shared objective. Figure 9 makes this diversity concrete for twelve representative targets, pairing the target-scale structural context with the local peptide-binding view used by the structure-based path.

D Group Relative Policy Optimization and the supervised cold start

D.1 The optimization algorithm

Reinforcement learning for language models faces a basic difficulty. A response is generated one token at a time, but the reward arrives only at the end, so attributing credit to individual tokens requires knowing how good the situation was before the response was produced. Proximal Policy Optimization and its relatives solve this by training a value network alongside the policy, which estimates the return expected from a given state. That network is typically of comparable size to the policy, doubling the memory required and introducing a second set of hyperparameters that must themselves be tuned.

GRPO instead estimates a baseline from several responses to the same prompt. For a prompt q, a group of G responses $\{o_1, \ldots, o_G\}$ is drawn from the current policy, and each receives a scalar reward r_i. The advantage of each response is obtained by standardizing within the group:

$$\hat{A}_i \;=\; \frac{r_i - \mathrm{mean}(\{r_j\}_{j=1}^{G})}{\mathrm{std}(\{r_j\}_{j=1}^{G})}. \tag{4}$$

The numerator asks how far this response's reward sits from the group's average, so a response better than its peers receives a positive advantage and one worse receives a negative one. The denominator rescales that difference by the spread within the group, so that a prompt on which outcomes vary widely and one on which they barely differ contribute comparably to the update. Concretely, with binary rewards and $G = 8$, a group in which three responses succeed has mean 0.375; the successful responses receive positive advantages and the five failures negative ones, in proportion to how unusual each outcome was.

The policy is then updated under a clipped surrogate objective:

$$\mathcal{L}(\theta) \;=\; -\frac{1}{G}\sum_{i=1}^{G}\min\Big(\rho_i\hat{A}_i,\; \mathrm{clip}(\rho_i,\,1-\epsilon,\,1+\epsilon)\,\hat{A}_i\Big) \;+\; \beta\, D_{\mathrm{KL}}[\pi_\theta \,\|\, \pi_{\mathrm{ref}}]\,. \tag{5}$$

The objective contains a likelihood-ratio term, clipping, and KL regularization. The ratio $\rho_i = \pi_\theta(o_i \mid q)/\pi_{\theta_{\mathrm{old}}}(o_i \mid q)$ compares how likely the updated policy is to produce a response against how likely the policy that actually sampled it was; multiplying it by $\hat{A}_i$ pushes probability mass toward responses that did better than average and away from those that did worse. The clip operation bounds ρ_i to $[1-\epsilon, 1+\epsilon]$ and the $\min$ selects the more conservative of the two terms, so that a single update cannot move the policy arbitrarily far on the strength of one sampled group. The final term penalizes divergence from a frozen reference copy of the model, with β setting how far the policy may drift; without it, a policy optimizing a narrow reward can degrade capabilities the reward does not measure.

Two properties of this scheme matter for our setting. The reward need only be defined once a response is complete, so no intermediate step requires a label, which suits a task where we can judge whether a decision was correct but not how promising a partial line of reasoning was. And because responses are compared only against others answering the same prompt, the intrinsic difficulty of an item cancels out: an easy item on which every response succeeds and a hard item on which every response fails are treated identically, and neither biases the update.

D.2 Why a supervised cold start is necessary

Supervised initialization. GRPO requires reward variation within a sampled group. If the base policy produces uniformly incorrect responses, the group supplies no useful gradient. We therefore use supervised fine-tuning to place the policy in a regime where sampled groups contain both successful and unsuccessful responses before GRPO.

The degenerate group. The second property of GRPO described above, that item difficulty cancels because responses are compared only against their peers, has a consequence that determines how training must begin. Equation 4 is defined only when the rewards within a group differ from one another. If all G responses to a prompt receive the same reward, the numerator $r_i - \mathrm{mean}(\{r_j\})$ is zero for every response, and the denominator $\mathrm{std}(\{r_j\})$ is zero as well. The advantage is undefined, and whatever convention the implementation adopts, whether adding a small constant to the denominator or discarding the group, the outcome is the same: no gradient flows. Such groups are not merely unhelpful but invisible to the optimizer. A batch composed largely of them performs the computation of training while learning nothing.

Why this happens at chance-level performance. A policy performing near chance encounters degenerate groups constantly. With binary rewards on a three-way item, a policy answering uniformly at random fails all eight sampled responses about four per cent of the time, which is tolerable. But a policy near chance is rarely uniform. It is more often systematically wrong, concentrating probability on one option across items where that option happens to be incorrect, and such a policy fails all eight responses far more often than the random figure suggests. Our base model sits in this regime on the design tasks: before any training it answers binding-selection items at 36.7% against a 33.3% chance level, and inspection of its outputs shows the predictions clustered on a single answer rather than spread across the options. On a large fraction of prompts the sampled group is therefore uniformly incorrect and yields nothing to learn from.

What supervision provides that reinforcement cannot. The problem is not that the optimizer is weak but that it can only work with what it is shown. Reinforcement learning improves a policy by distinguishing better from worse among behaviours the policy already produces; when the policy produces nothing that succeeds, there is no distinction to draw. Supervised learning has no such requirement. Its gradient comes from the discrepancy between the model's output and a target that is supplied externally, so it needs only the correct answer and not variation in outcome. It can therefore raise a policy off the floor by direct imitation, in a regime where policy gradients are silent.

Procedure. We fine-tune the Redesigner on the same corpus used for policy optimization, with the loss masked so that it is computed only over the answer tokens and not over the prompt, which prevents the model from spending capacity on reproducing question text. Training runs for three epochs at a learning rate of 1×10^{-5}, an order of magnitude above the rate used subsequently for GRPO, since the objective here is to move the policy substantially rather than to refine it. The resulting checkpoint is then handed to GRPO. Once a non-trivial fraction of groups contains both successes and failures, the advantage in Equation 4 is well defined for enough of the batch that optimization proceeds.

E BUILDING A CORPUS THAT SUPPORTS POLICY OPTIMIZATION

E.1 CORPUS REQUIREMENTS

GRPO compares responses to the same prompt. Training items therefore require experimentally determined answer keys and controls against shortcut-solvable patterns.

E.2 SOURCES AND ITEM FAMILIES

Each answer in the corpus is therefore derived from a laboratory measurement rather than model-based evaluation.

Binding items derive from SKEMPI 2.0, which catalogues over 7,000 experimentally characterized substitutions across protein–protein and protein–peptide interfaces; 5,410 records survive deduplication and filtering. Two formats are generated. The first presents a complex with a set of substitutions and asks which least destabilizes the interface, keyed to the smallest measured $\Delta\Delta G$. The second presents a single substitution and asks whether it strengthens or weakens binding, keyed to the sign.

Selectivity items derive from DBAASP and GRAMPA, which together aggregate measured activity for tens of thousands of peptides. Of these, 984 carry both a minimum inhibitory and a hemolytic concentration and are retained. The task is to identify the peptide with the widest therapeutic window, keyed to the largest measured ratio.

General items are drawn from existing biomedical and mathematical benchmarks and serve as capability anchors rather than targets. The corpus comprises 2,000 items in a 3:2 ratio of domain to general families.

E.3 LEAKAGE AND SHORTCUT CONTROLS

A measured key guarantees that an answer is right. It does not guarantee that arriving at it requires reasoning. Three constraints close the gaps that remain.

Split by entity, not by item. Partitions are drawn over complexes and parent peptides, so that no structure or peptide appears on both sides of a split. For selectivity items this required an additional step. Because options within an item are drawn from a shared peptide pool, partitioning the items themselves leaves the same peptide appearing on both sides in different combinations, and a split that looks clean at the item level is not one at the level that matters. We therefore partition the peptide pool first and rebuild items independently within each side.

Balance the key. Answer classes are equalized within each family, so that a constant response earns nothing over chance. During construction we found that balancing by sampling from each class in turn introduced a subtler problem: when one class is much larger than another, the procedure we first used drew the retained members of the larger class from the region nearest the decision boundary, compressing its range and leaving the two classes asymmetric in effect size even though equal in count. Balancing is now performed by random draw across the full range of each class, and both classes cover the same range of measured values.

Discard what the assay cannot resolve. Substitutions whose measured effect falls within experimental error do not distinguish the options they separate. Stratifying accuracy by effect size makes this concrete: a simple feature-based classifier recovers the key at close to chance on substitutions below 0.5 kcal mol^{-1} and improves monotonically with magnitude. Since SKEMPI's own repro-

ducibility is of that order, a label drawn from within it records noise rather than direction. Items below 1 kcal mol^{-1} are excluded.

E.4 Selectivity distractor construction

The constraints above prevent an item from being answered by memorization or by a constant response. They do not prevent it from being answered by reading a single number off the options, which for selectivity items is the natural failure mode: potency and safety are positively correlated in nature, so the most active peptide in a random draw is frequently also the correct answer, and a model can score well by always choosing it.

Distractors are therefore constructed against that correlation. Options are ordered by increasing inhibitory concentration with hemolytic concentration rising alongside, so that the most potent option is also the least safe and the safest is the least potent. The key is the unique largest ratio and lies at neither extreme. Reading either axis alone then performs at chance, and the item yields only to the quotient.

An early version of this construction over-corrected. Flattening the potency axis to remove the potency shortcut made the safety axis the sole discriminator, and a rule selecting the safest option recovered the key in the great majority of items: one shortcut had been replaced by another. The present construction keeps both axes informative and makes them disagree, so that neither alone suffices.

E.5 Shortcut audit and exclusion

Each family is finally audited by measuring how often simple physicochemical rules, over net charge, hydrophobicity, aromatic content and substitution score, recover the key. Families in which any rule exceeds a permutation-derived null are excluded from the corpus rather than corrected, since a shortcut that survives construction is likely to reflect a property of the underlying data rather than a flaw in the items. One antimicrobial family was removed on this basis: activity in the source data is close to monotonic in aromatic content, so the distractor pool contains no counterexamples from which balanced items could be built. It is retained outside the training corpus as a control for studies of shortcut learning.

F Sample redesign trajectories

The following five trajectories report the intermediate candidates and measurements underlying the summary results. Each episode begins from a peptide that fails at least one criterion. At each round, the Diagnoser evaluates the current physicochemical state, the Redesigner proposes fifteen candidates, the toolset evaluates all candidates, and the resulting measurements are written to Memory. The record gives the stated strategy, the best candidate, the number of candidates satisfying all three criteria, and the retained Memory entry. Quoted entries are reproduced without editing.

The trajectories illustrate distinct failure modes. ELTB initially moves in the wrong direction on the sequence-based path. IL-1β meets the antimicrobial bound before the hemolytic bound. CagA starts with strong binding and uses the measured cost of earlier substitutions. TLR4 starts from a potent but strongly hemolytic peptide. Anthrax protective antigen starts with binding and safety satisfied, leaving antimicrobial activity as the deficient axis. Performance is non-monotonic across rounds, and the delivered candidate is not always the candidate with the lowest MIC.

Sample trajectory 1: ELTB (anti-toxin, sequence-based path)

1. Task
Design a bifunctional peptide against ELTB (heat-labile enterotoxin B subunit of enterotoxigenic *E. coli*). No receptor structure with an identifiable pocket was available, so this episode runs on the sequence-based path: binding is adjudicated by the interface confidence of a predicted complex (ipTM ≥ 0.6) rather than by force-field interface energetics.

2. Round-by-round record

Round 0 — Charge dispersion. *Strategy:* alternating cationic–hydrophobic modules, dispersing charge to limit hemolysis. *Best candidate:* `KRWKRKRWKRKRW`, MIC 75.9, ipTM 0.908. *Outcome:* 0/15 pass. *Written to Memory:* all candidates exceed the antimicrobial bound and none passes two criteria; charge dispersion did not raise antimicrobial activity, possibly from an imbalanced cationic–hydrophobic ratio.

Round 1 — More KRW repeats. *Strategy:* 3–4 KRW repeats with 1–2 polar residues (T/S) inserted, total length $\leq$ 15 aa. *Best candidate:* `KRWKRWTWKRK`, MIC 212.4, ipTM 0.934. *Outcome:* 0/15 pass. *Written to Memory:* excessive repetition yields structural uniformity that suppresses activity; insertion position and repeat count both require optimization.

Round 2 — Tuning insertion position. *Strategy:* three KRW repeats as scaffold, polar residues (T/S/Q/N) inserted at positions 5–7. *Best candidate:* `KRWKRWTWKRK`, MIC 212.4, ipTM 0.945. *Outcome:* 0/15 pass. *Written to Memory:* MIC remains far above the bound; repeated modules may raise hydrophobicity or reduce binding efficiency.

Round 3 — Reduced repeat count. *Strategy:* insert Q/N at position 5 and optimize repeat count. *Best candidate:* `KRWKRQWKRK`, MIC 353.6, ipTM 0.945. *Outcome:* 0/15 pass. *Written to Memory:* excessive hydrophobicity is limiting antimicrobial activity; reduce hydrophobic module density (three KRW repeats to two) and add hydrophilic positions.

Round 4 — Polar insertion and terminal edits. *Strategy:* retain the best historical KRW scaffold, optimize polar insertion positions, anchor hydrophobic residues. *Best candidate:* `KRWKRWTWKRSTWVYFVW`, MIC 37.9, ipTM 0.871. *Outcome:* 0/15 pass. *Written to Memory:* adding hydrophobic residues did not improve activity and destabilized the complex (ipTM fell to 0.871).

Round 5 — First passing candidates. *Strategy:* two KRW modules, polar anchor (S/T) at position 5, terminal hydrophobic tail $\leq$ 3 aa (V/F/W), with hydrophobic anchoring density held low. *Best candidate:* `KRWKRWTWKRWVFW`, MIC 6.4, ipTM 0.93. *Outcome:* **5/15 pass all three criteria**. *Written to Memory:* polar anchoring (S/T) at positions 5–7 raises ipTM (peak 0.963); hydrophobic tails (V/F/W) must be $\leq$ 3 aa and separated by $\geq$ 2 polar residues; two to three KRW repeats balance membrane permeation against structural stability; avoid the MIC penalty of purely cationic modules (75.9) and the ipTM penalty of purely hydrophobic tails (0.871).

3. Trajectory

MIC across rounds: 75.9 $\rightarrow$ 212.4 $\rightarrow$ 212.4 $\rightarrow$ 353.6 $\rightarrow$ 37.9 $\rightarrow$ 6.4. Candidates passing all three criteria: 0, 0, 0, 0, 0, 5/15. Delivered peptide: `KRWKRWSWKRWVFW` (MIC 8.5 µM, HC_{50} 125.2 µM, ipTM above threshold).

Sample trajectory 2: IL-1β (anti-inflammatory)

1. Task

Design a bifunctional peptide against IL-1β (*Escherichia coli* urosepsis; PDB 1ITB, UniProt P01584). Structure-based path. Starting peptide `QLANFLLAIQALFDDKIR`: $\Delta\Delta G$ -3.8, MIC 405, HC_{50} 72; 0/15 pass. The starting peptide fails on both the antimicrobial and hemolytic axes.

2. Round-by-round record

Round 1 — Charge and hydrophobicity. *Strategy:* 0 Q$\rightarrow$R (net charge; 16.5 Å from the receptor, too distant to perturb binding), 1 L$\rightarrow$K (amphipathicity), 3 N$\rightarrow$E (lower hydrophobicity). *Best candidate:* `RLKFLLAIQALKFDDKIR`, MIC 55.7, HC_{50} 108.4. *Outcome:* 0/15

pass. *Written to Memory:* net charge -0.44 (target $+3$ to $+6$); amphipathicity insufficient; only two aromatic residues.

Round 2 — More charge. *Best candidate:* `RRKFLLAIQALKFDDKIR`, MIC 47.0, HC_{50} 93.0. *Outcome:* 0/15 pass. The antimicrobial bound is now met, but hemolysis is not. *Written to Memory:* net charge 2.55, still below the window.

Rounds 3–5 — MIC satisfied, hemolysis not. *Best candidate:* `RRRFLLAIQALKFDDKIR`, MIC 18.0, HC_{50} 89.7, unchanged across three rounds. *Outcome:* 0/15 pass in each round. The antimicrobial bound is met with room to spare while HC_{50} stalls just under 100, and three consecutive rounds of solvent-face charge additions fail to move it.

Round 6 — First passing candidates after the safety trade-off. *Strategy:* 0 Q→R, 1 L→K, 3 N→E — the same solvent-face set, now applied to reduce hydrophobicity rather than to raise potency. *Best candidate:* `RRRFLLEIQALKFDDKIR`, $\Delta\Delta G$ -3.6, MIC 23.9, HC_{50} 105.4. *Outcome:* **2/15 pass all three criteria**. MIC is given up (18.0 → 23.9) in exchange for crossing the hemolytic bound. *Written to Memory:* net charge 3.55, inside the window; GRAVY -0.372 and hydrophobic moment 0.138 both weak; aromatic content still insufficient.

3. Trajectory
MIC across rounds: 405 → 55.7 → 47.0 → 18.0 → 18.0 → 18.0 → 23.9. Candidates passing all three criteria: 0, 0, 0, 0, 0, 0, 2/15. Delivered peptide: `RRRFLLEIQALKFDDKIR` ($\Delta\Delta G$ -3.6, MIC 23.9 μM, esAMP 31.4, HC_{50} 105.4 μM; first pass at round 6). The delivered peptide clears the hemolytic bound by 5.4 μM, the narrowest margin of any episode reported here.

Sample trajectory 3: CagA (anti-virulence)

1. Task
Design a bifunctional peptide against CagA (*Helicobacter pylori*, chronic gastritis and peptic ulcer; PDB 4IRV, UniProt P55980). Structure-based path. Starting peptide `FLTERYTRSWYIWLRIANNI`: $\Delta\Delta G$ -27.2, MIC 231, HC_{50} 61; 0/15 pass. Binding is already very strong; both other axes fail. Figure 8 follows a designed CagA-binding candidate from its placement on the complete target through its surface pose to the residue-level contact environment.

2. Round-by-round record

Round 1 — Edits informed by prior measurements. *Written to Memory (verbatim):* "18 W→K (lowering GRAVY drove MIC from 34.298 to 10.407); 11 I→K (breaking the continuous hydrophobic stretch W–Y–I gave $\Delta\Delta G$ -26.176, MIC 34.298)." The rule carries the measured values of the substitutions it recommends. *Best candidate:* `FLTKRYTRKDYIWLRIANRK`, MIC 59.2, HC_{50} 100.2. *Outcome:* 0/15 pass.

Round 2 — First pass. *Strategy:* 7 R→K (distal positive charge), 15 I→W (move hydrophobic and aromatic bulk onto the solvent face), 8 S→K, and binding-face 18 N→R. *Best candidate:* `FLTKRYTRKDYIWKRIANKK`, $\Delta\Delta G$ -20.8, MIC 45.1, HC_{50} 110.3. *Outcome:* **1/15 passes all three criteria**.

Round 3 — Held. *Best candidate:* unchanged, $\Delta\Delta G$ -22.4, MIC 45.1, HC_{50} 110.3. *Outcome:* 1/15 pass.

Round 4 — Additional charge and aromaticity. *Best candidate:* `FLTKRYTRKDYIWKRIANWK`, MIC 34.3. *Outcome:* **2/15 pass**. *Written to Memory:* "7 R→K verified historically: $\Delta\Delta G$ more negative and MIC lower."

Rounds 5–9 — Candidate retention. *Best candidate:* mostly FLTKRYTRKDYIWKRIANKK, MIC 45.1, HC_{50} 110.3. *Outcome:* 1/15 pass in each round. *Written to Memory at round 6 (verbatim):* "to address hemolysis, 18 W→K (history verified: MIC fell by 7.80 while $\Delta\Delta G$ fell by only 0.21); 9 W→K to relieve local hydrophobic overload." The edit is selected by comparing a measured gain against a measured cost, both retrieved from earlier rounds of this episode.

Rounds 10–12 — Additional hydrophobicity reduction. *Strategy:* continue reducing GRAVY to drive MIC lower. *Best candidate:* FLTKRYTRKDYIWKRIAKWK, MIC driven down to 10.4 — the lowest of the episode — but HC_{50} falling to 90.6. *Outcome:* 2/15, then 1/15, then **0/15**. The same hydrophobicity reduction that lowers MIC also removes the hemolytic margin. The episode is delivered from round 11 rather than from its most potent round.

3. Trajectory
MIC across rounds: 231 → 59.2 → 45.1 → 45.1 → 34.3 → 45.1 → 45.1 → 45.1 → 45.1 → 34.3 → 10.4 → 10.4 → 10.4. Candidates passing all three criteria: 0, 0, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 0. Delivered peptide: FLTKRYTRKDYIWKRIAKKK ($\Delta\Delta G$ −12.5, MIC 18.1 μM, esAMP 27.1, HC_{50} 111.0 μM; first pass at round 2).

Sample trajectory 4: TLR4 (anti-endotoxin)

1. Task
Design a bifunctional peptide against TLR4 (*Neisseria meningitidis* meningitis with endotoxemia; PDB 4G8A, UniProt O00206). Structure-based path. Starting peptide FFFAELLRAIIKNGGFM: $\Delta\Delta G$ −3.9, MIC 46.6, HC_{50} 32.1; 0/15 pass. The starting peptide already meets the antimicrobial bound but is strongly hemolytic.

2. Round-by-round record

Round 1 — Solvent-face edits only. *Strategy:* 0 F→R (net charge), 5 L→F (amphipathicity), 15 F→W (aromatic content), 2 F→D (break the continuous hydrophobic stretch); all on the solvent face. *Best candidate:* FFFAELLRAIIKNGGKF, $\Delta\Delta G$ −4.4, MIC 23.9, HC_{50} 33.5. *Outcome:* 0/15 pass. *Counter-examples from the same round:* FFDAELLRAIIKNGGFM (2 F→D, proposed in order to break the hydrophobic stretch) measured MIC 142.6, three times worse than the parent; FFFAELLEAIIKNGGFM (7 R→E, removing positive charge) measured MIC 134. Both predicted directions were wrong, and both are recorded. *Written to Memory:* net charge 0.56 (target +3 to +6); amphipathicity insufficient (hydrophobic moment 0.367); three consecutive phenylalanines at positions 0–2 form a hydrophobic stack, and no W or Y is present.

Round 2 — Aromatic substitution, charge still short. *Strategy:* 0 F→W, 2 F→D (a polar edit on the binding face judged not to break contacts), 15 F→W, 16 M→K (reduce terminal hydrophobicity). *Best candidate:* WWFAELLRAIIKNGGWM, $\Delta\Delta G$ −4.2, MIC 29.5, HC_{50} 36.9. *Outcome:* 0/15 pass. *Written to Memory:* net charge 1.55, still low; arginine present only on the solvent face; hydrophobic moment 0.506, still below target.

Round 3 — First passing candidates. *Strategy:* 0 F→R, 5 L→F, 15 F→W, 16 M→R, all on the solvent face, with binding-face positions 10–13 held fixed. *Outcome:* **2/15 pass all three criteria**. *Rejected despite the best readouts:* RWFAELLRAIIKFKGWW edited binding-face positions 12 and 13, reaching $\Delta\Delta G$ −10.8 and MIC 5.8 — the best binding and antimicrobial values of the entire episode — but HC_{50} collapsed to 24.4 and the candidate failed. RWFRELLRAIIKNGGWR ($\Delta\Delta G$ −4.7, MIC 5.6, HC_{50} 91.7) likewise failed, missing the hemolytic bound by 8.3 μM. Stronger binding does not imply a passing design, and these measurements are why later rounds hold the binding face fixed. *Written to Memory:* amphipathicity not established (GRAVY 0.353); the peptide remains under-charged and over-hydrophobic.

Round 4 — Additional safety margin. *Best candidate:* `RWRRELLRAIIKNGGKW`, $\Delta\Delta G$ -4.9, MIC 5.4, HC_{50} 137.5. *Outcome:* **5/15 pass**. *Written to Memory:* net charge now 3.56, inside the target window; amphipathicity established.

Rounds 5–8 — Candidate retention. *Strategy:* unchanged — continue adding positive charge and aromatic content on the solvent face. *Best candidate:* stable at `RWRRELLRAIIKNGGRW`, MIC 4.5, HC_{50} 118.6. *Outcome:* 1, 1, 2, 3 passes respectively. No further improvement in the best readout.

3. Trajectory
MIC across rounds: 46.6 $\rightarrow$ 23.9 $\rightarrow$ 29.5 $\rightarrow$ 5.6 $\rightarrow$ 5.4 $\rightarrow$ 4.5 $\rightarrow$ 4.5 $\rightarrow$ 4.5 $\rightarrow$ 4.5. Candidates passing all three criteria: 0, 0, 0, 2, 5, 1, 1, 2, 3. Delivered peptide: `RWRRELLRAIIKNGGRW` ($\Delta\Delta G$ -5.1, MIC 4.5 µM, esAMP 21.2, HC_{50} 118.6 µM; first pass at round 3).

Sample trajectory 5: anthrax protective antigen (PA)

1. Task
Design a bifunctional peptide against the protective antigen of *Bacillus anthracis* (inhalational anthrax; PDB 25PV, UniProt P13423). Structure-based path. Starting peptide `FSGRSSRSSSLKKYG`: $\Delta\Delta G$ -11.6, MIC 374.5, HC_{50} 209.4; 0/15 pass. The starting peptide is already non-hemolytic and already binds, so only the antimicrobial axis is deficient.

2. Round-by-round record

Round 1 — Charge, aromatics, and one binding-face edit. *Strategy:* raise net charge at solvent-face position 1 (S$\rightarrow$R); add aromatic content at solvent-face position 12 (G$\rightarrow$Y, 9.2 Å from the receptor); break the hydrophobic stretch at binding-face position 9 (S$\rightarrow$K). *Best candidate:* `RGRSSRSSSLKKYFW`, $\Delta\Delta G$ -15.3, MIC 61.9, HC_{50} 200.4. *Outcome:* 0/15 pass. *Written to Memory:* net charge 3.55, slightly below the target window ($+3$ to $+6$); amphipathicity not established (hydrophobic moment 0.038); only two aromatic residues, and their distribution requires optimization.

Round 2 — First passing candidates. *Strategy:* retain the charge edit at position 1; move the aromatic addition to solvent-face position 10 (L$\rightarrow$W, 7.8 Å, far enough not to perturb the interface); keep the binding-face edit at position 9 (S$\rightarrow$K) to act in concert with 10 L$\rightarrow$W in lowering MIC; disperse hydrophobicity at position 12 (G$\rightarrow$Y). *Best candidate:* `KRGRWRRSSLKKYFW`, $\Delta\Delta G$ -20.0, MIC 23.8, HC_{50} 216.6. *Outcome:* **5/15 pass all three criteria**. *Written to Memory:* net charge 4.55, now at the lower bound of the ideal window; amphipathicity still insufficient (hydrophobic moment 0.121); three aromatic residues, still slightly few.

3. Trajectory
MIC across rounds: 374.5 $\rightarrow$ 61.9 $\rightarrow$ 23.8. Candidates passing all three criteria: 0, 0, 5/15. Delivered peptide: `KRGRWRRSSLKKYFW` ($\Delta\Delta G$ -20.0, MIC 23.8 µM, esAMP 26.8, HC_{50} 216.6 µM; first pass at round 2). Here the solvent-face charge additions made $\Delta\Delta G$ *more* negative, strengthening binding while lowering MIC — the opposite of their effect on TLR4 and IL-1β, where the same class of edit left binding flat or weakened it.

G Algorithmic specification

G.1 The redesign loop

Diagnosis and proposal are separate policy calls, ensuring that diagnosis precedes candidate generation. Scoring is performed by external functions that do not take the policy as an argument.

Algorithm 1 Iterative redesign over a fixed starting peptide

Require: target T, starting peptide p_0, interface map $\mathcal{I}$ (empty on the sequence-based path), memory $\mathcal{M}$, policy π
Require: batch size B, round budget N

$p \leftarrow p_0$
for $t = 1$ **to** N **do**
  $\phi \leftarrow \textsc{Physchem}(p, \mathcal{I})$ ▷ charge, moment, helicity, aromaticity, per-residue exposure
  $d \leftarrow \pi_{\text{diag}}(\phi, \mathcal{M})$ ▷ deficits on the antimicrobial, binding and hemolytic axes
  $\{(c_j, \hat{y}_j)\}_{j=1}^{B} \leftarrow \pi_{\text{edit}}(p, \phi, d, \mathcal{M})$ ▷ candidates, each with a committed predicted direction
  for $j = 1$ **to** B **do**
    $s_j \leftarrow \textsc{Score}(c_j, T)$ ▷ ($\Delta\Delta G$ or ipTM, MIC, HC_{50}); no argument is π
    $\mathcal{M} \leftarrow \mathcal{M} \cup \{(t, c_j, \phi, d, \hat{y}_j, s_j)\}$ ▷ written whether or not c_j passes
  if $\exists\, j : \textsc{Passes}(s_j)$ **then**
    return the passing candidate, $\mathcal{M}$
  $p \leftarrow \textsc{Select}(\{(c_j, s_j)\})$
  if $\textsc{Stagnant}(\mathcal{M}, t)$ **then**
    break ▷ substitution on this backbone is exhausted
return FAIL, $\mathcal{M}$

Selection. SELECT favours the candidate from which a passing one is most likely to be reachable in a single further edit. Candidates are ranked first by how many of the three criteria they already satisfy, and among those tied, by how close the unsatisfied criteria lie to their bounds. A candidate meeting two criteria and missing the third narrowly is therefore preferred to one meeting two comfortably and missing the third by a wide margin. Selection is made on proximity to the joint constraint rather than on any single axis, which prevents the loop from chasing potency at the cost of a margin it will later have to recover.

Termination. The loop ends on the first round that produces a passing candidate. It also ends early when progress toward the joint constraint stalls across consecutive rounds, which is taken to mean that substitution on the current backbone has been exhausted; control then returns to the Builder, which generates a new starting peptide, and the loop restarts with $\mathcal{M}$ carried over. The round budget N bounds the episode in the absence of either condition.

What the structure encodes. Three steps carry the design decisions described in Section 2.3. The diagnosis takes ϕ and $\mathcal{M}$ but not the candidate set, since none exists yet. Each proposal records $\hat{y}_j$ alongside c_j before the realized outcome is measured. Every candidate is then written to memory regardless of outcome, preserving both positive and negative evidence.

G.2 OUTER PIPELINE

The redesign loop is invoked by a pipeline that resolves a natural-language scenario into a concrete design problem. This stage follows the planner–executor–verifier pattern: a Planner selects the next tool given what has been gathered, an Executor validates parameters against that tool's schema and issues the call, and a Verifier decides whether the evidence collected so far is sufficient to proceed.

The two paths differ only in interface construction and binding scoring; downstream diagnosis, memory, activity, and safety checks are shared.

H MEMORY ACCUMULATION ACROSS EPISODES

H.1 WHAT CHANGES AND WHAT DOES NOT

Policy weights, evaluators, and prompt templates remain fixed during and across design episodes. Only memory changes. Any improvement across rounds or related targets therefore arises from conditioning on accumulated records rather than parameter updates, and should disappear when memory is cleared.

Algorithm 2 Scenario resolution and dispatch

Require: natural-language scenario q, tool set $\mathcal{T}$, budget $n_{\max}$
$\mathcal{E} \leftarrow \emptyset$ ▷ evidence accumulated across tool calls
for $n = 1$ **to** $n_{\max}$ **do**
  $(\tau, \text{args}) \leftarrow \pi_{\text{plan}}(q, \mathcal{E}, \mathcal{T})$
  $r \leftarrow \textsc{Execute}(\tau, \text{args})$ ▷ schema-validated; malformed calls repaired or rejected
  $\mathcal{E} \leftarrow \mathcal{E} \cup \{(\tau, \text{args}, r)\}$
  if $\pi_{\text{verify}}(q, \mathcal{E}) = \textsc{Stop}$ **then**
    break
$T \leftarrow \textsc{ResolveTarget}(\mathcal{E})$ ▷ must be extracellular and an effector, not a correlate
if $\textsc{Pocket}(T) \neq \emptyset$ **then**
  $\mathcal{I} \leftarrow \textsc{Epitope}(T)$ ▷ contact-ranked interface residues, or a detected pocket on an apo monomer
  $p_0 \leftarrow \textsc{Builder}(T, \mathcal{I})$ ▷ SE(3) diffusion, relaxed to full atom
else
  $\mathcal{I} \leftarrow \emptyset$; $p_0 \leftarrow \textsc{SequenceGenerator}(T)$
return $\textsc{Redesign}(T, p_0, \mathcal{I}, \mathcal{M}, \pi)$ ▷ Algorithm 1

H.2 Within an episode

The loop conditions each round on the measured outcome of the preceding one. Records include the substitution and associated changes in MIC and $\Delta\Delta G$, allowing later rounds to reuse edits whose measured cost was small.

Two behaviours follow from having such a record available. Edits concentrate at positions whose measured cost was low in earlier rounds, since those are the positions the record identifies as tolerant. Strategies that have already failed on the current backbone are not reproposed, because the record contains the failure rather than only its absence from the successes.

H.3 Across episodes

Memory persists when an episode ends. A rule inferred on one target is available when the next begins, which is what allows the system to start a new target from something other than a blank state. Rules that transfer are those expressed in terms the next target shares, such as the relationship between aromatic density and hemolysis, which holds regardless of what the peptide is binding to. Rules expressed in terms of a specific interface do not transfer and are not expected to.

A rule retrieved from one target may not transfer to another. Retrieved rules affect candidate proposals but not external scoring; failed proposals are recorded in memory for subsequent rounds.

I Grounded Accumulation Memory

I.1 Why two layers

The Diagnoser and Redesigner require detailed measurements from earlier edits, but the complete history may exceed the context window. Memory therefore uses two layers: a queryable page-store containing complete records and an index containing one evidence-linked conclusion per round for inclusion in context.

I.2 The page-store

Every candidate proposed in every round is written, including those that fail all three criteria, together with the context needed to interpret it: the physicochemical state and interface map under which it was proposed, the diagnosis that preceded it, the direction its proposer predicted, and the measurements it subsequently received. Nothing is discarded and nothing is compressed before writing.

Failures are retained so that unsuccessful strategies are not repeatedly proposed. Because predictions are recorded before their corresponding measurements, both successful and failed proposals remain auditable.

I.3 THE INDEX

At the end of each round, one entry is written summarizing what that round established. Entries name the substitutions involved and the measured quantities that justify the conclusion, so that a rule carries the evidence for itself. An entry reading "lowering hydrophobicity improves potency" would be admissible but nearly useless; the entries the system writes state both the substitution and what it cost, giving the change in potency alongside the change in interface energy.

Each index entry reports both the benefit and measured cost of an edit, enabling comparison across the three objectives.

I.4 RETRIEVAL

At the start of a round, the index is supplied in full and the page-store is queried for entries relevant to the current state: prior candidates on the same backbone, prior edits at the positions the diagnosis has identified, and the measurements attached to both. Retrieved records are included as contextual evidence that the policy may use or disregard when proposing the next edit.

J BINDING SPECIFICITY AGAINST DECOY RECEPTORS

Cationic amphipathic peptides associate readily with negatively charged surfaces, so a peptide that scores well against its intended target may be scoring a property of its own charge rather than a fit to that particular interface. The check is to score it against receptors it was not designed for.

Decoys were chosen in two groups. The first contains receptors from the same family as a cognate target, which are the hardest case: they present similar folds and similar surface chemistry, so a peptide that discriminates among them is discriminating on interface complementarity rather than on gross properties. The second contains proteins abundant in the environment a therapeutic peptide would encounter, including serum albumin, transferrin, immunoglobulin Fc, fibrinogen and lysozyme. The last of these serves as an internal control: it carries a high net positive charge, so a cationic peptide should be repelled by it, and a favourable energy against lysozyme would indicate that the force field is not reporting what it is supposed to on this class of complex.

Every delivered candidate was scored against all eight. Mean interface energy is above zero for every decoy and averages -15.2 REU against cognate targets. Some candidates show favourable energies against individual decoys, most often same-family receptors. The comparison therefore supports preferential rather than exclusive binding.

K RE-SCORING WITH INDEPENDENT PREDICTORS

Candidates were admitted using HemoPI2 for hemolysis and APEX for potency. Their scores under those two predictors therefore cannot serve as evidence of quality, since the selection procedure optimized against them directly. The relevant question is whether the candidates survive evaluation by models that played no part in choosing them.

All delivered peptides were re-scored by predictors trained independently on different data. Agreement is high on both axes, with Pearson correlations of 0.84 for hemolysis and 0.86 for potency, and 73% of candidates continue to satisfy both criteria under the external predictors.

Candidates that fail under external scoring cluster near the decision thresholds, while correlations remain high across the evaluated range. Disagreement is therefore concentrated among marginal candidates.

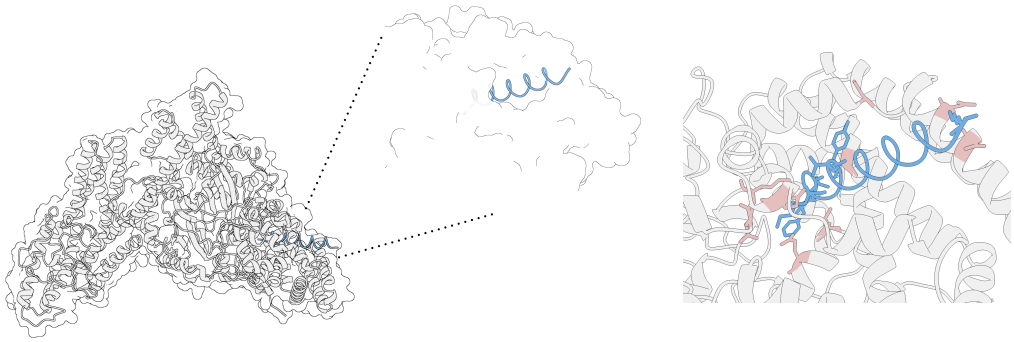

Figure 8: **CagA target engagement at three structural scales.** Left, the complete CagA target model locates the designed peptide (blue) relative to the full protein. Center, a molecular-surface view shows the peptide adopting a helical pose along an exposed CagA groove. Right, an interface close-up resolves peptide side chains in blue and neighbouring CagA residues in pink, revealing the local packing environment.

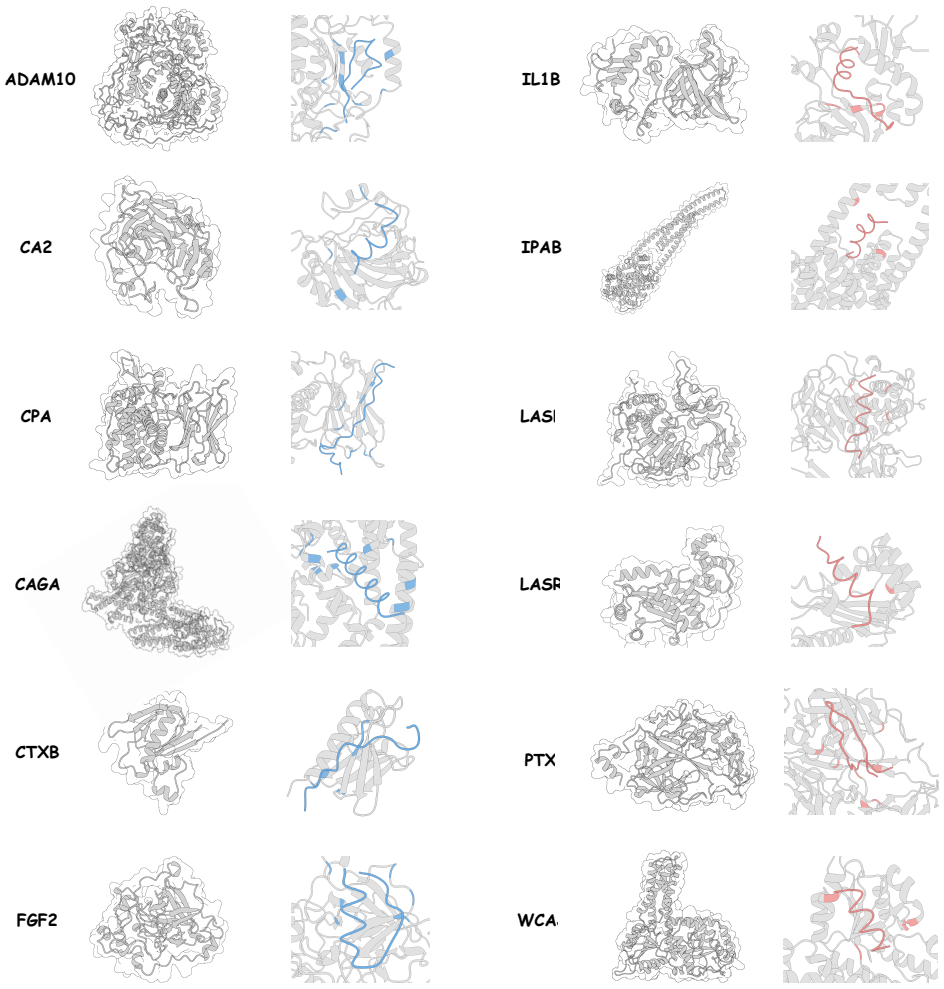

Figure 9: **One shared redesign problem across twelve different molecular contexts.** Representative structure-based outputs are shown for ADAM10, CA2, CPA, CagA, CtxB, FGF2, IL-1β, IpaB, LasB, LasR, PTX, and WcaJ. For each target, a target-scale view is paired with a closer view of the candidate peptide and its binding region. The montage spans host-response targets, toxins and virulence machinery, and repair-associated proteins, with substantial variation in fold, interface geometry, and accessible surface.