nature machine intelligence

Supplementary information

https://doi.org/10.1038/s42256-026-01237-5

A generative artificial intelligence approach for peptide antibiotic optimization

In the format provided by the authors and unedited

Supplementary Materials

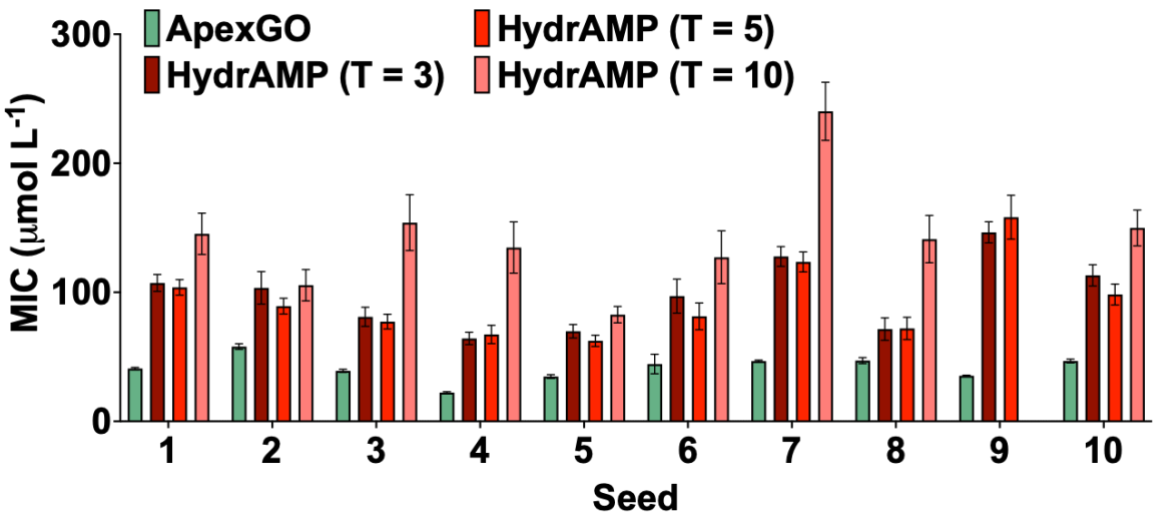
Supplementary Figs. 1 to 17

Supplementary Tables 1 to 2

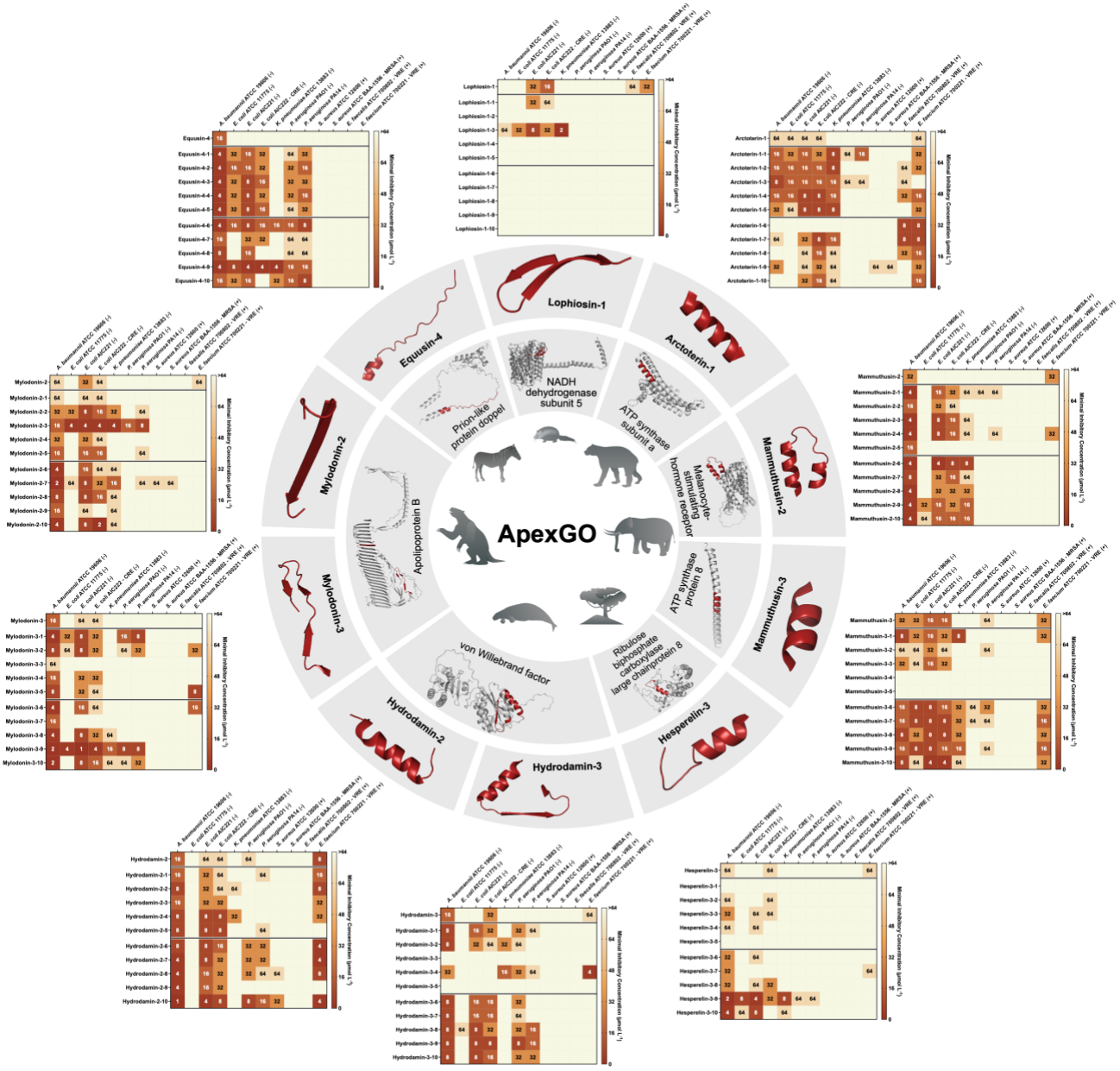
Supplementary Table 1. Template and derivatives sequences. In bold amino acid residues that are different from the original sequence.

Peptide	Sequence	Peptide	Sequence
Equusin-4	CVLLFSQLPAVKARGTKHRIKWNRK	Mammuthusin-3	KTLKIIRLLF
Equusin-4-1	C**H**L**E**F**R**QLPAVKAR**LL**K**L**RIKWNRK	Mammuthusin-3-1	KTLKIIRLLF**CK**
Equusin-4-2	C**H**ELF**R**QLPAVKAR**LL**K**L**RIKWNRK	Mammuthusin-3-2	KT**S**KIIRLLF**K**
Equusin-4-3	C**E**L**H**F**R**QLPAVKAR**LL**K**L**RIKWNRK	Mammuthusin-3-3	KTLKIIRLL**CK**
Equusin-4-4	C**E**L**R**F**H**QLPAVKAR**LL**K**L**RIKWNRK	Mammuthusin-3-4	KT**N**KIIRLLF**K**
Equusin-4-5	C**R**LLF**R**QLPAVK**E**R**LL**K**L**RIKWNRK	Mammuthusin-3-5	KTLKI**Q**RLLF**K**
Equusin-4-6	**IW**LLF**A**QL**I**AVKAR**L**TK**I**RIKWNRK	Mammuthusin-3-6	K**W**LKIIRL**R**LF
Equusin-4-7	**IW**LLASQL**I**AVKAR**LL**KHRIKWNRK	Mammuthusin-3-7	K**W**L**R**KIIRLLF
Equusin-4-8	**IW**LLASQL**I**AVKAR**LI**KHRIKWNRK	Mammuthusin-3-8	KT**WL**KIIRLL**K**F
Equusin-4-9	**IW**LLFSQL**A**AVKAR**AL**K**I**RIKWNRK	Mammuthusin-3-9	K**W**LK**R**IIRLLF
Equusin-4-10	**IW**LLF**A**QL**W**AVKAR**I**TKLRIKWNRK	Mammuthusin-3-10	K**W**TLKIIRLL**K**F
Arctoterin-1	GHLLIHLIGKATLAL	Mammuthusin-2	RACLHARSIARLHKRWRPVHQGLGLK
Arctoterin-1-1	GH**K**LIHLIGKA**K**L**K**L	Mammuthusin-2-1	**IQK**LHARSIARL**L**KRWRPVHQ**L**LE**L**K
Arctoterin-1-2	GHL**S**IHLIGKA**K**L**K**L	Mammuthusin-2-2	R**KE**LHAR**Q**IARLHK**L**WRPVHQ**E**L**W**LK
Arctoterin-1-3	GHL**K**IHLIGKA**K**L**N**L	Mammuthusin-2-3	R**KE**LHAR**Q**IARLHK**L**WRPVHQ**W**L**E**LK
Arctoterin-1-4	GHL**K**IHLIGKA**S**L**K**L	Mammuthusin-2-4	**IQK**LHARSIARLHK**L**WRPVHQ**F**L**E**LK
Arctoterin-1-5	GHLLI**K**LIGKATL**K**L**K**	Mammuthusin-2-5	**IQK**LHARSIARL**L**KRWRPVHQ**E**L**N**LK
Arctoterin-1-6	GHLLI**S**LIGK**L**TL**K**L	Mammuthusin-2-6	**I**A**S**LHARSIARL**L**K**I**WR**N**VHQGLGLK**W**
Arctoterin-1-7	GHLLI**R**LIGK**L**TL**K**L	Mammuthusin-2-7	**I**A**S**LHARSIARL**L**KRW**IN**VHQGLGLK**W**
Arctoterin-1-8	GHL**S**IHLIGK**L**TL**K**L	Mammuthusin-2-8	**I**A**W**LHARSIARL**L**KRW**IN**VHQGLGLK**W**
Arctoterin-1-9	GHLLI**G**LIGK**L**TL**K**L	Mammuthusin-2-9	**I**A**S**LHARSIARL**L**K**L**WR**N**VHQGLGLK**W**
Arctoterin-1-10	GHL**S**LIHLIGK**L**TL**K**L	Mammuthusin-2-10	**I**A**S**LHARSIARL**L**KRW**LN**VHQGLGLK**W**
Lophiosin-1	HWITINTIKLSISLKI	Hydrodamin-2	RMARNLVRYVQGLKKKKVI
Lophiosin-1-1	**KM**I**G**IN**K**IKLSISLKI	Hydrodamin-2-1	R**L**ARNLVRYVQ**L**LK**W**KKV**K**
Lophiosin-1-2	**EIK**INTIKLSI**K**LKI	Hydrodamin-2-2	R**P**ARNLVRYV**A**GLKK**L**KV**W**
Lophiosin-1-3	H**K**I**K**INTIKL**K**I**K**LKI	Hydrodamin-2-3	R**P**ARNLVRYVQGLK**L**KKV**WK**
Lophiosin-1-4	H**K**ITIN**IK**L**K**I**K**LKI	Hydrodamin-2-4	R**L**ARNLVRYVQ**A**LKK**L**KV**W**
Lophiosin-1-5	**EK**ITIN**G**IKLSI**K**LKI	Hydrodamin-2-5	R**L**ARNLVRYVQ**A**LK**W**KKV**W**
Lophiosin-1-6	HWI**R**IN**A**IKL**R**ISLKI**W**	Hydrodamin-2-6	R**I**AR**W**LVRYVQ**W**LKKKKV**W**
Lophiosin-1-7	HWI**R**INTIKL**R**I**A**LKI**W**	Hydrodamin-2-7	**IW**ARNLVRYV**W**GLKKKKVI**W**

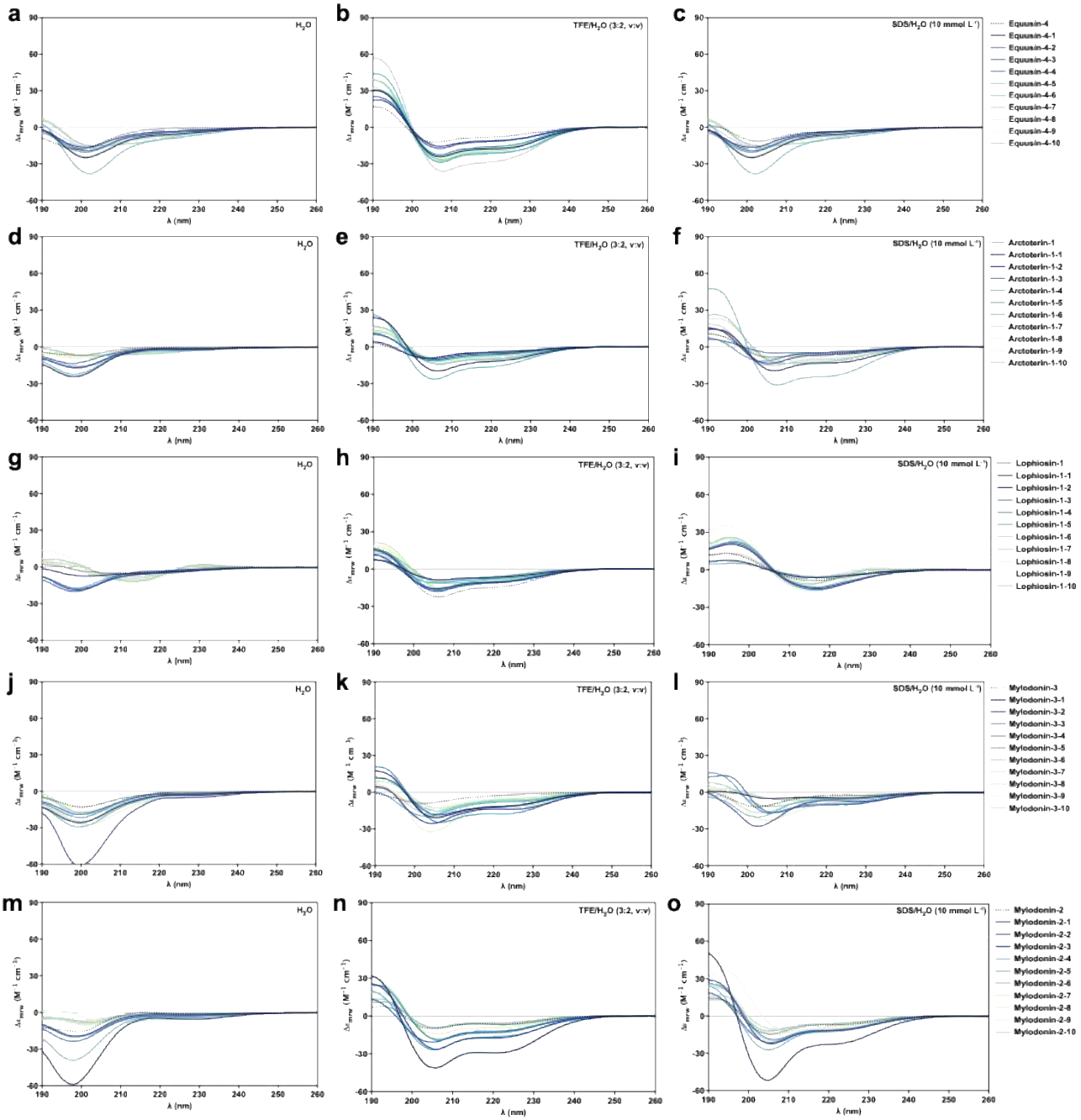
Lophiosin-1-8	HWI**A**IN**G**IKL**R**ISLKI**W**	Hydrodamin-2-8	**IW**ARNLVRYV**W**GLKKKKV**W**
Lophiosin-1-9	HWI**A**IN**A**IKL**R**ISLK**W**	Hydrodamin-2-9	**IW**ARNLVRYVQ**W**LKKKKVI**W**
Lophiosin-1-10	HWI**G**IN**A**IKL**R**ISLKI**W**	Hydrodamin-2-10	R**I**ARNLVR**W**V**L**GLKKKKV**W**
Mylodonin-3	KIYKKLSTPPFTLNIRTLPKVKFPK	Hydrodamin-3	RNLVRYVQGLKKKKVIVIPVGIGPHANIK
Mylodonin-3-1	KI**N**KKLS**I**PPF**K**L**K**IR**F**L**G**KVKFPK	Hydrodamin-3-1	**K**NL**S**R**I**VQGLKK**F**KK**K**I**N**IPVGIGP**F**ANIK
Mylodonin-3-2	KI**N**K**I**LS**L**PPF**K**L**K**IR**G**LPKVKFPK	Hydrodamin-3-2	**K**NL**S**R**I**VQGLKK**F**KK**Q**I**R**IPVGIGP**F**ANIK
Mylodonin-3-3	KI**G**KKLS**I**PPF**K**L**K**IR**N**L**F**KVKFPK	Hydrodamin-3-3	**K**NLI**R**P**V**QGLK**F**KK**K**I**K**IPVGIGP**F**ANIK
Mylodonin-3-4	KI**G**KKLS**L**PPF**K**L**K**IR**Q**L**M**KVKFPK	Hydrodamin-3-4	**K**NL**F**R**K**VQGLK**L**KKK**L**IV**R**PVGIGP**F**ANIK
Mylodonin-3-5	KI**N**KKLS**I**PPF**K**L**K**IR**G**LP**Q**VKFPK	Hydrodamin-3-5	**K**NLVR**K**VQGLKK**F**KK**P**I**K**IPVGIGP**N**AN**F**K
Mylodonin-3-6	**W**I**A**KKLS**LRW**F**R**LNIRTLPKVKFPK	Hydrodamin-3-6	RNL**A**RYV**W**GLK**L**KKVI**K**I**K**VGIG**RW**ANIK
Mylodonin-3-7	**W**IY**A**KLS**LRW**F**R**LNIRTLPKVKFPK	Hydrodamin-3-7	RNL**A**RYV**W**GLK**L**KKVI**K**I**V**GIG**RW**ANIK
Mylodonin-3-8	**F**IY**A**KLS**LRW**F**R**LNIRTLPKVKFPK	Hydrodamin-3-8	RNLVR**W**V**I**GLK**L**KKKVI**K**I**K**VGIG**RW**ANIK
Mylodonin-3-9	KI**G**KKLS**RWL**F**R**LNIR**G**LPKVKFPK	Hydrodamin-3-9	RNLVR**W**V**A**GLK**L**KKKVI**K**I**K**VGIG**RW**ANIK
Mylodonin-3-10	**F**I**A**KKLS**LRW**F**R**LNIRTLPKVKFPK	Hydrodamin-3-10	RNL**W**RYV**A**GLK**L**KKVI**K**I**K**VGIG**RW**ANIK
Mylodonin-2	KRKRGLKLATALSLNNKF	Hesperelin-3	RQKNHGIHFRVLAKALR
Mylodonin-2-1	K**I**K**I**GLKLA**K**ALSLNNK**L**	Hesperelin-3-1	RQKN**LP**I**K**FRVLAKALR**P**
Mylodonin-2-2	K**M**K**I**GLKLATALSL**K**NK**L**	Hesperelin-3-2	RQKN**LK**I**P**FRVLAKALR**P**
Mylodonin-2-3	K**I**K**I**GLKLAT**K**LSLNNK**L**	Hesperelin-3-3	RQKN**PL**I**KL**RVLAKALR
Mylodonin-2-4	K**V**KR**I**LKLATAL**K**LNNK**L**	Hesperelin-3-4	RQKN**PL**I**K**FRVLAKALR**P**
Mylodonin-2-5	K**I**KR**L**LKLAT**K**LSLNNK**L**	Hesperelin-3-5	RQKN**LP**I**P**FRVLAKAL**K**
Mylodonin-2-6	**I**RKR**I**L**W**LA**R**ALSLNNKF	Hesperelin-3-6	**WL**KNH**A**IHFRVLAKAL**W**
Mylodonin-2-7	**I**RKR**L**L**W**LA**R**ALSLNNKF	Hesperelin-3-7	**LA**KNH**W**IHFRVLAKAL**W**
Mylodonin-2-8	**I**RK**I**GLKLA**W**ALSL**K**NKF	Hesperelin-3-8	R**L**KNH**W**I**A**FRVLAKAL**W**
Mylodonin-2-9	KRK**I**GLKLA**W**ALSL**KL**KF	Hesperelin-3-9	**L**R**I**KNH**W**IHFRVLAKAL**W**
Mylodonin-2-10	K**II**RGLKLA**W**ALSL**K**NKF	Hesperelin-3-10	R**L**KN**WA**IHFRVLAKAL**W**



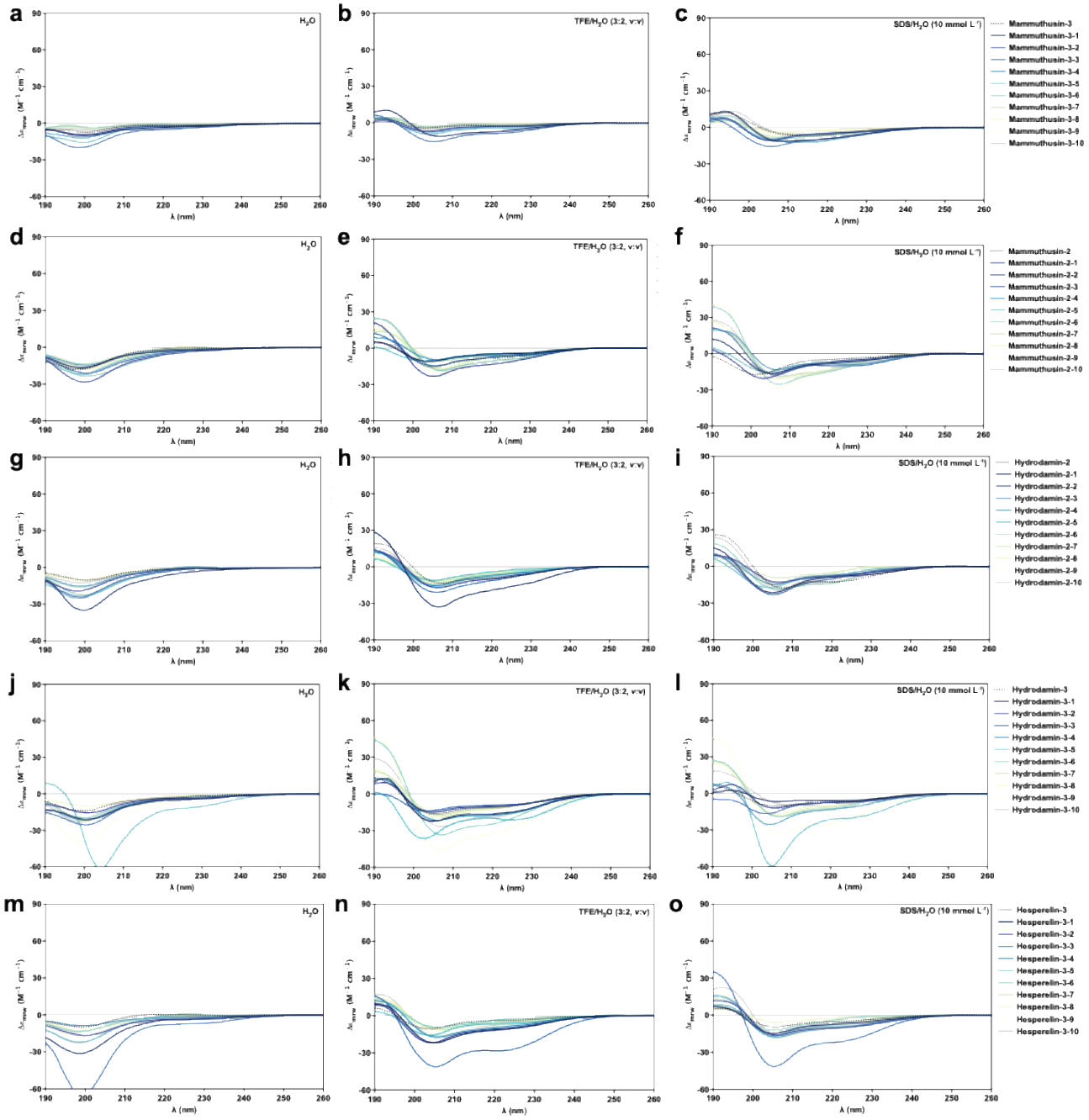
Supplementary Figure 1. Comparison of ApexGO and HydrAMP across 10 seeds on the Gram-negative objective under a ≥75% sequence-similarity constraint. Bars show the mean APEX-predicted MIC of the top-K feasible sequences per seed; error bars denote ±1 s.d. Colors denote methods: ApexGO, HydrAMP with decoder temperature T = 3, T = 5, and T = 10. Lower MIC is better. For seed 9, HydrAMP at T = 10 produced no feasible analogue within the 200k-unique / 1M-proposal budget, and thus it is not shown for that condition.



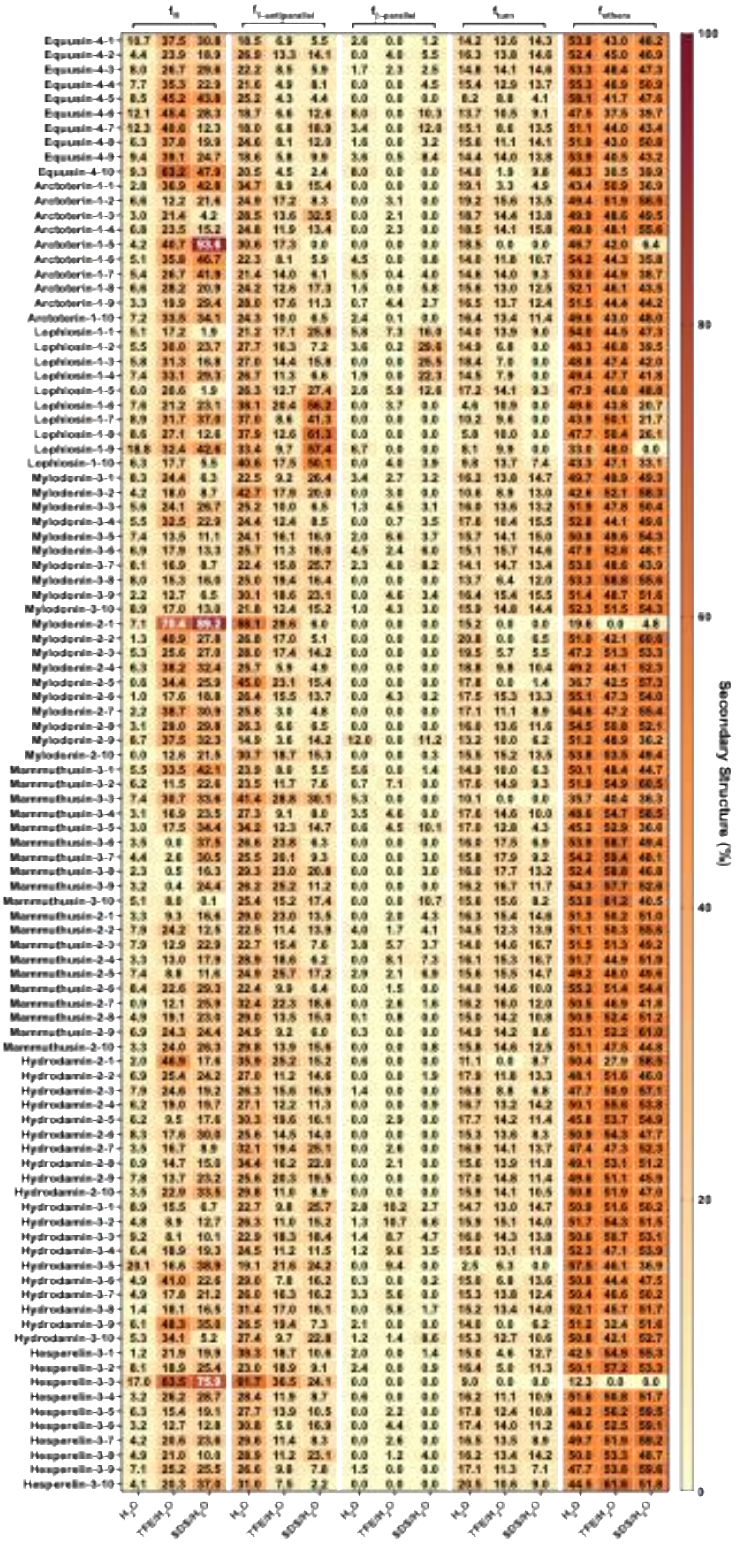
Supplementary Figure 2. Antimicrobial activity of peptides optimized by ApexGO. Heat map displaying the minimum inhibitory concentrations (MICs, in μmol L^{-1}) of peptides optimized by ApexGO against clinically relevant bacterial pathogens. The MICs were determined using the broth microdilution method, and the bar plots presents the mode MIC values from three independent biological replicates obtained for the best performing optimized peptide for each of the templates. The optimized peptides demonstrate significant improvements in antimicrobial efficacy compared to their parent templates, validating the effectiveness of ApexGO in generating peptides with enhanced activity against bacterial strains.



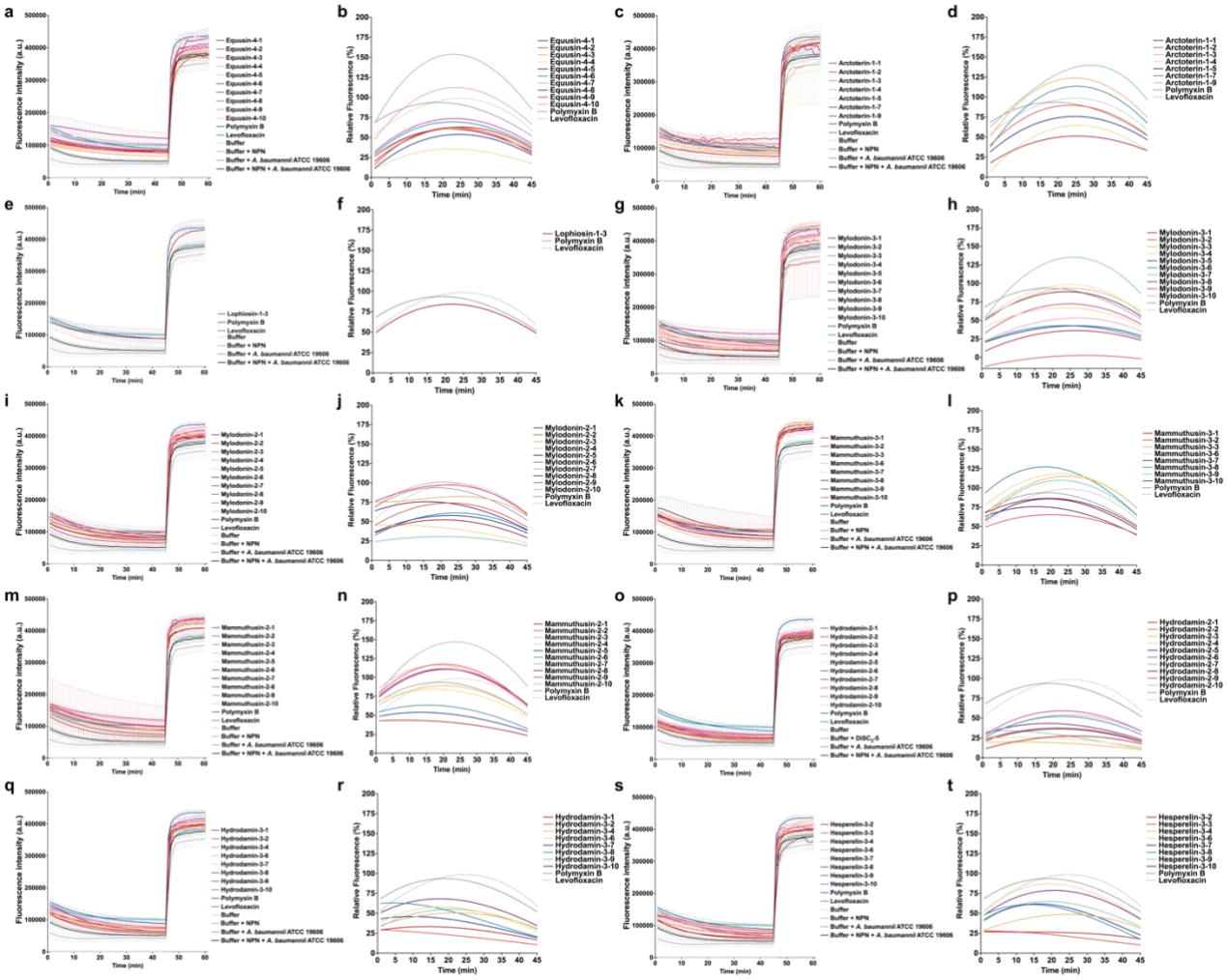
Supplementary Figure 3. Circular dichroism spectra of templates equusin-4, arctoterin-1, lophiosin-1, mylodonin-3, and mylodonin-2 and their respective ApexGO-optimized derivatives. Circular dichroism experiments were conducted in a J-1500 Jasco circular dichroism spectrophotometer. The spectra were recorded in three different media: water, 60% trifluoroethanol in water, and Sodium dodecyl sulfate (SDS) in water (10 mmol L^{-1}), after three accumulations at 25 ºC, using a 1mm path length quartz cell, between 260 and 190 nm at 50 nm min^{-1}, with a bandwidth of 0.5 nm. The concentration of all peptides tested was 50 µmol L^{-1}.



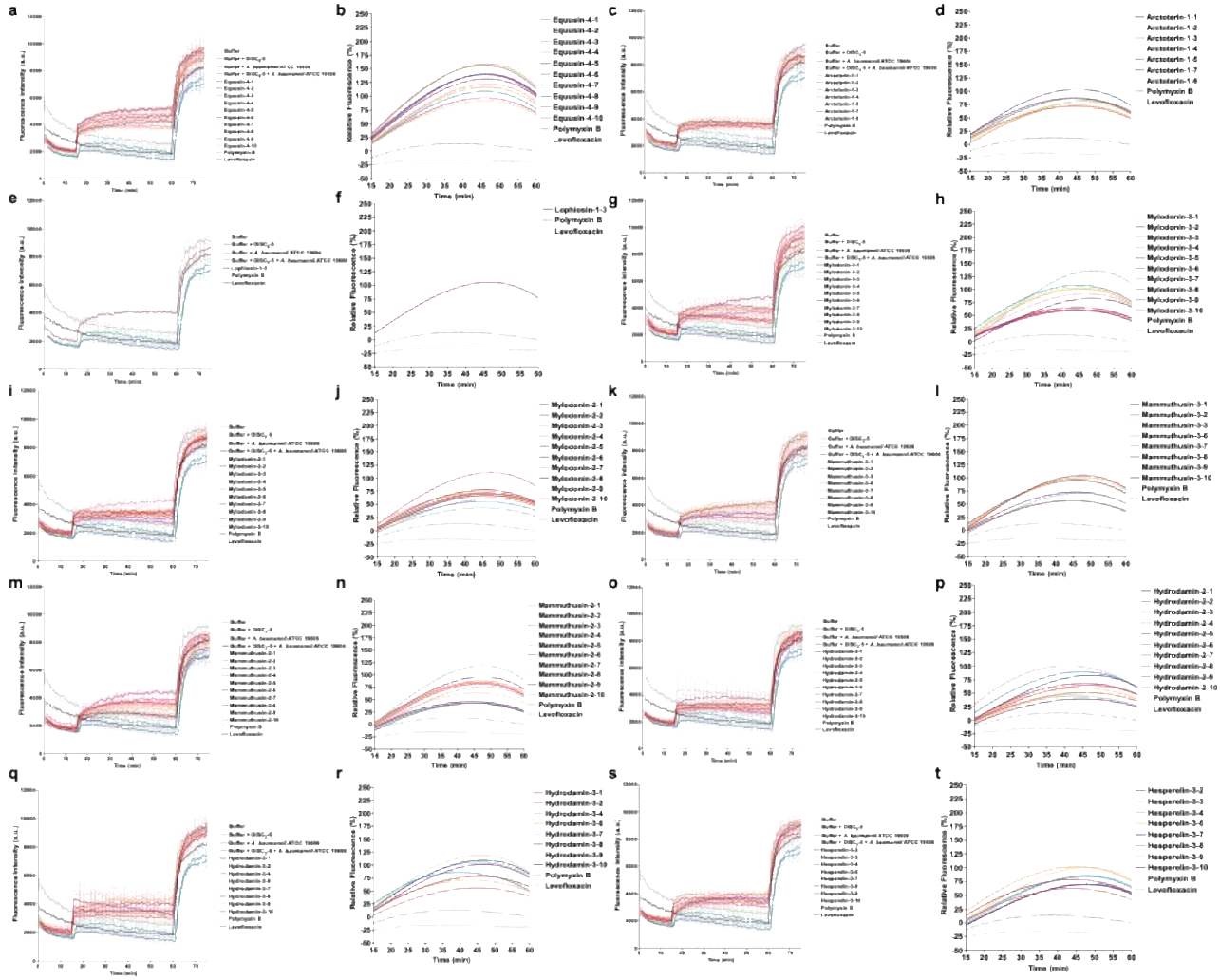
Supplementary Figure 4. Circular dichroism spectra of templates mammuthusin-3, mammuthusin-2, hydrodamin-2, hydrodamin-3, and hesperelin-3 and their respective ApexGO-optimized derivatives. Circular dichroism experiments were conducted in a J-1500 Jasco circular dichroism spectrophotometer. The spectra were recorded in three different media: water, 60% trifluoroethanol in water, and Sodium dodecyl sulfate (SDS) in water (10 mmol L^{-1}), after three accumulations at 25 °C, using a 1mm path length quartz cell, between 260 and 190 nm at 50 nm min^{-1}, with a bandwidth of 0.5 nm. The concentration of all peptides tested was 50 µmol L^{-1}.



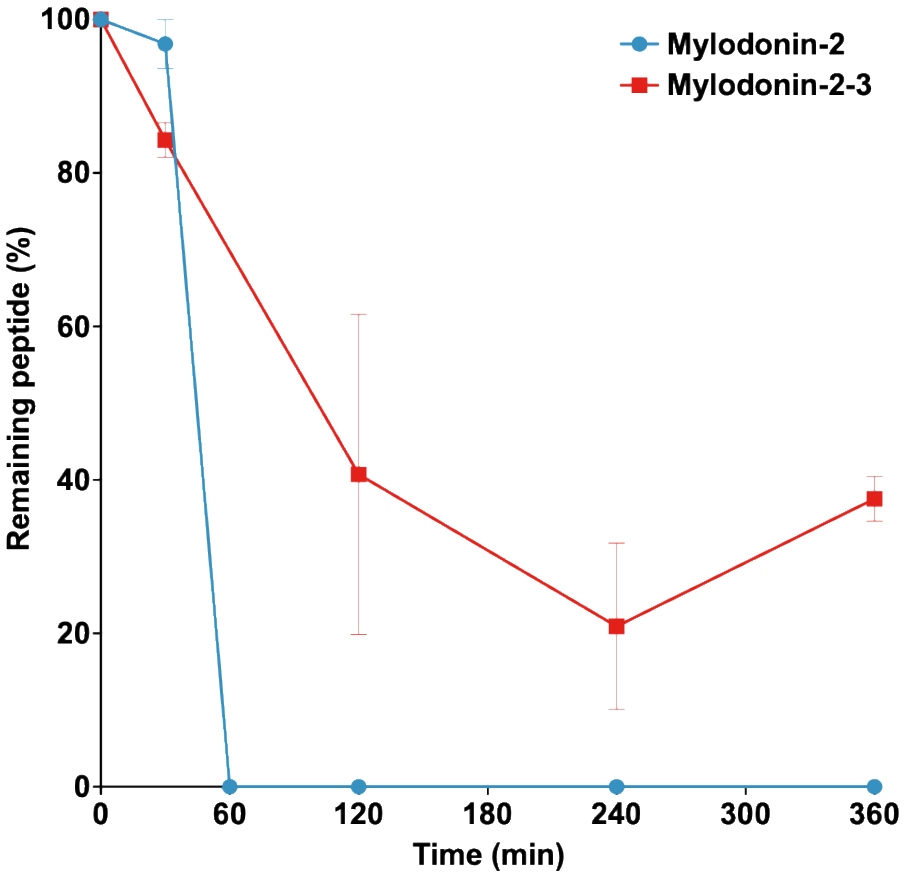
Supplementary Figure 5. Secondary structure fractions of ApexGO-optimized peptides. Heatmap with the percentage of secondary structure found for each peptide in three different solvents: water, 60% trifluoroethanol (TFE) in water, and SDS (10 mmol L^{-1}) in water. Secondary structure fraction was calculated using the BeStSel server[32].



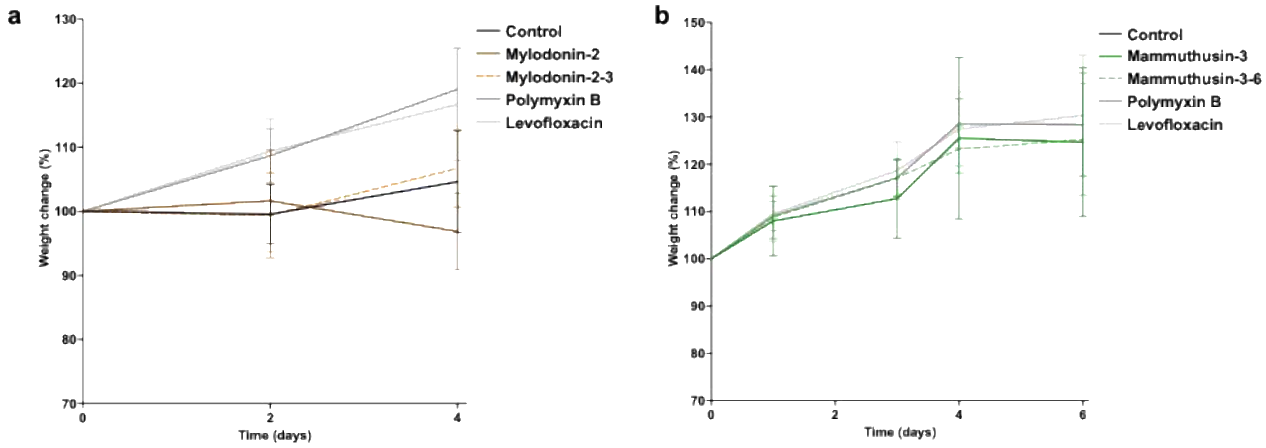
Supplementary Figure 6. Outer membrane permeabilization of *A. baumannii* ATCC 19606 induced by ApexGO-optimized peptides . Outer membrane permeabilization was assessed using the probe 1-(N-phenylamino)naphthalene (NPN), showing the permeabilization effects of peptides derived from each of the following 10 parent de-extinct EPs that showed antimicrobial activity against *A. baumannii* ATCC 19606: **(a)** equusin-4, **(c)** arctoterin-1, **(e)** lophiosin-1, **(g)** mylodonin-3, **(i)** mylodonin-2, **(k)** mammuthusin-3, **(m)** mammuthusin-2, **(o)** hydrodamin-2, **(q)** hydrodamin-3, and **(s)** hesperelin-3. The relative fluorescence values of the probe's fluorescence obtained for each of the experiments and represented by a non-linear fitting compared to the baseline of the untreated control (buffer + bacteria + fluorescent dye) and benchmarked against the antibiotics polymyxin B and levofloxacin are represented for: **(b)** equusin-4, **(d)** arctoterin-1, **(f)** lophiosin-1, **(h)** mylodonin-3, **(j)** mylodonin-2, **(l)** mammuthusin-3, **(n)** mammuthusin-2, **(p)** hydrodamin-2, **(r)** hydrodamin-3, and **(t)** hesperelin-3.



Supplementary Figure 7. Cytoplasmic membrane depolarization of *A. baumannii* ATCC 19606 induced by ApexGO-optimized peptides. Membrane depolarization assays were performed using the hydrophobic probe 3,3′-dipropylthiadicarbocyanine iodide [$DiSC_3$-(5)] showing the depolarization effects of ApexGO-optimized peptides derived from each of the following 10 parent de-extinct peptides that showed antimicrobial activity against *A. baumannii* ATCC 19606: **(a)** equssin-4, **(c)** arctoterin-1, **(e)** lophiosin-1, **(g)** mylodonin-3, **(i)** mylodonin-2, **(k)** mammuthusin-3, **(m)** mammuthusin-2, **(o)** hydrodamin-2, **(q)** hydrodamin-3, and **(s)** hesperelin-3. The relative fluorescence values of the probe's fluorescence obtained for each of the experiments and represented by a non-linear fitting compared to the baseline of the untreated control (buffer + bacteria + fluorescent dye) and benchmarked against the antibiotics polymyxin B and levofloxacin are represented for: **(b)** equssin-4, **(d)** arctoterin-1, **(f)** lophiosin-1, **(h)** mylodonin-3, **(j)** mylodonin-2, **(l)** mammuthusin-3, **(n)** mammuthusin-2, **(p)** hydrodamin-2, **(r)** hydrodamin-3, and **(t)** hesperelin-3.



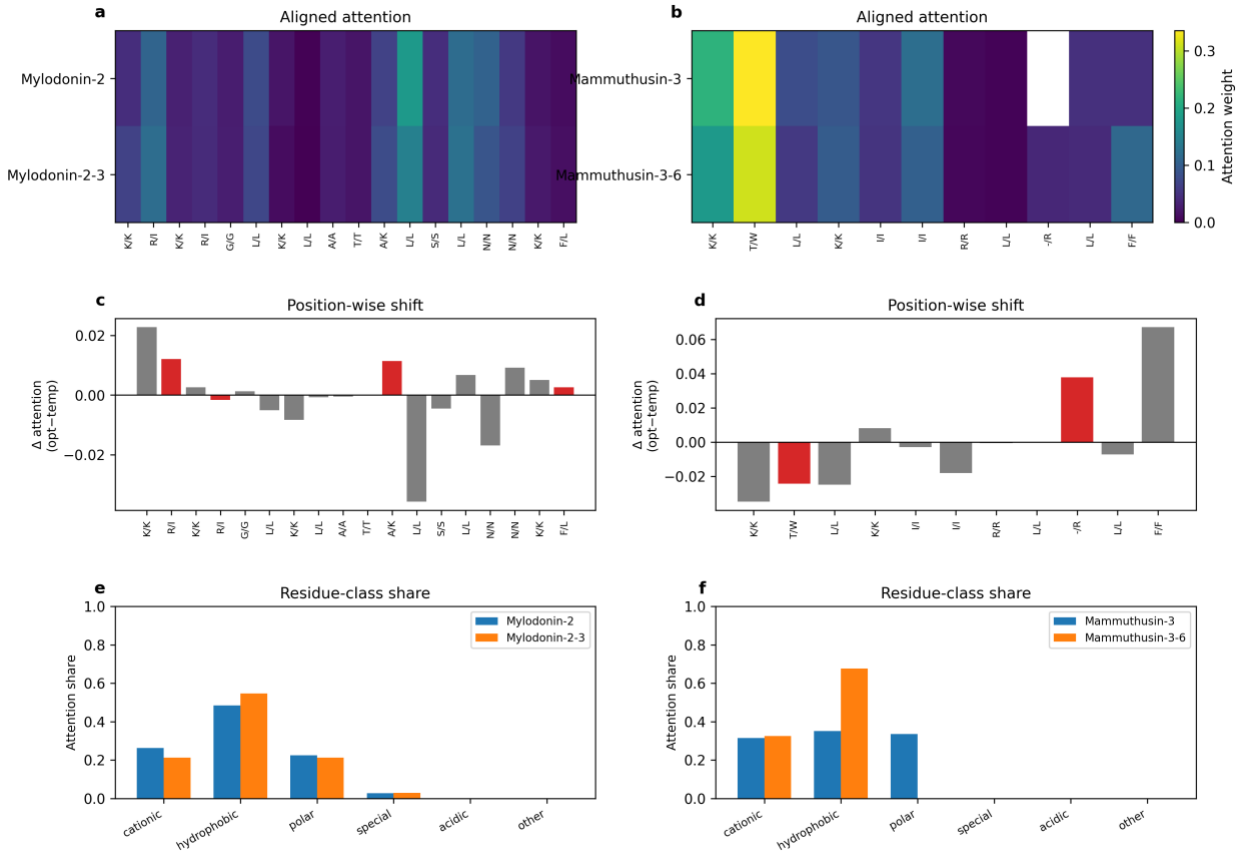
Supplementary Figure 8. Resistance to proteolytic degradation of de-extinct template and its lead optimized derivative. Shown are the stability profiles of mylodonin-2 and mylodonin-2-3 following 6 hours of exposure to human serum, illustrating each peptide's resistance to proteolytic degradation.



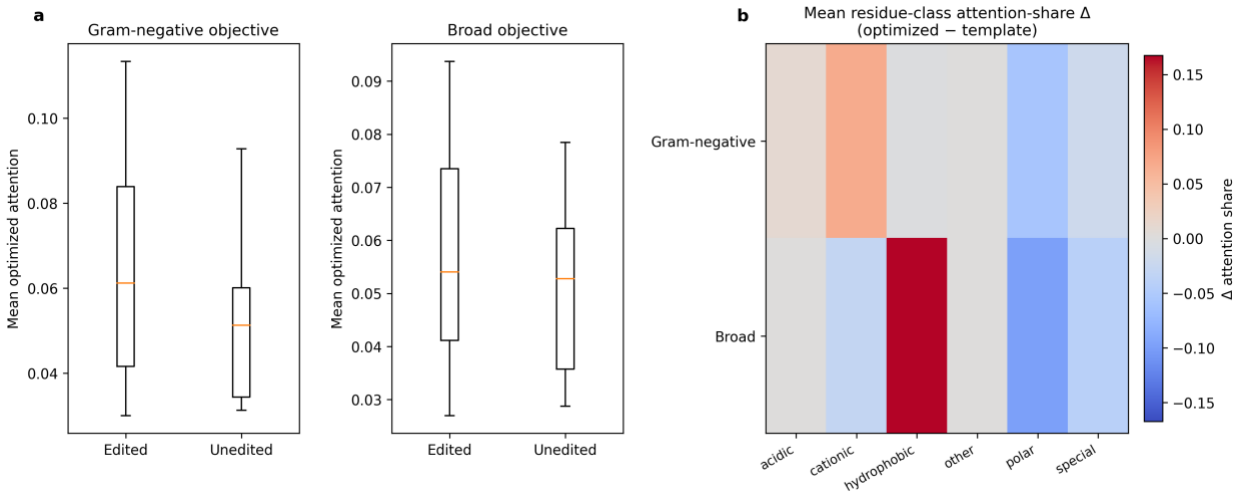
Supplementary Figure 9. Weight change monitoring in both skin abscess and deep thigh infection mouse models infected with *A. baumannii*. Mouse weight was monitored throughout the duration of the **(a)** skin abscess model (4 days total) and the **(b)** deep thigh infection model (6 days total) to assess potential toxic effects of both the bacterial load and the peptides.

Supplementary Table 2. Template-free peptides and their most similar known AMP.

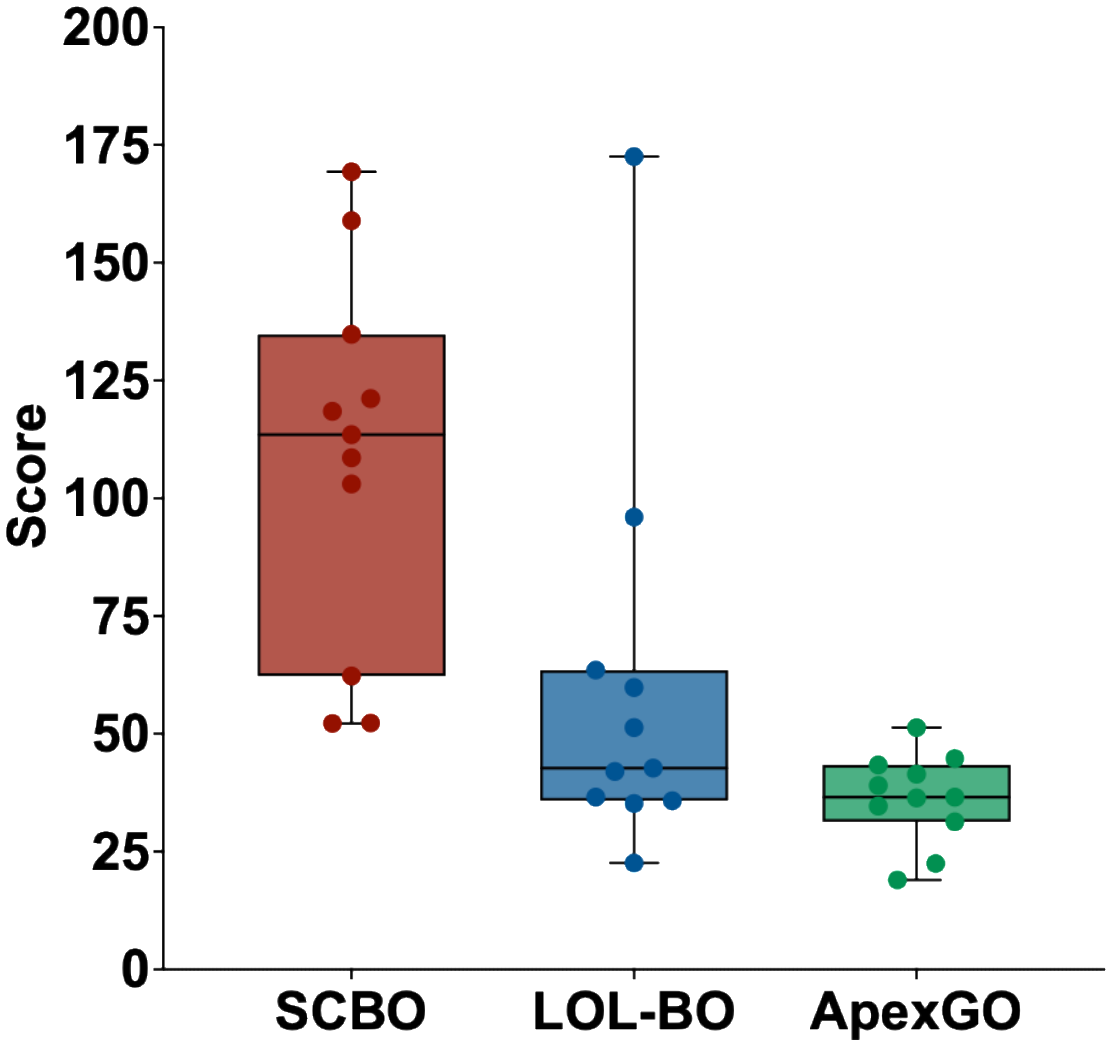
Peptide	Sequence	Most similar known AMP	SSW Similarity score
Free-1	KIIRTLKVKFCK	IRTLPKVKF	0.634
Free-2	KKIRLKFTCKFK	KKIRLKVTFKCK	0.679
Free-3	KIRKIFKIFKCK	KKIRLKVTFKCK	0.629
Free-4	KLRKISLTLKLKKLKLKLLALKLRLKTSIA	LKLKLKLKLKLKLKLKLKLKLK	0.591
Free-5	KLRKISLKSLLKKLKKLLKLRLKLLKLTIA	LKKLLKLLKLLKL	0.607
Broad-free-1	ILGGTLSKLAKLLSKLKLLASSK	ILGKLLSTAAKLLSKL	0.567
Broad-free-2	LIGAAKSSLSLLSAKLL	AAKKVLKLLKKLL	0.504
Broad-free-3	LIASALGKAKLLLSKLKTLL	FASLLGKALKALLAKLAKQ	0.505
Broad-free-4	IISGKLLASKLLASKLLLSALK	ILGKLLSTAAKLLSKL	0.518
Broad-free-5	IIGAKLAKNLKLLSKLLALLSK	LKLLKKLLKLLKKL	0.545
Free-75diff-1	KLRIIKFKVTCKK	LRIIKKILKKLI	0.527
Free-75diff-2	KKLSSLRIKLLKKLKKLLLRKLASKLLQSG	LKKLLKKLKKLL	0.551
Free-75diff-3	KIPKIRINLKVKVFK	KKIRLKVTFK	0.489
Free-75diff-4	KPIKKIKIKIGLSKFKNLNFPLK	KKIRLKVVKKF	0.431
Free-75diff-5	KFKFNVQFKKIKKLKKIRWDNEFFLPNNNS	KFKKFKKFKKFKKFKKFK	0.432
Broad-free-75diff-1	ILIAARIWLKRLLK	KWLKKLLKKLL	0.580
Broad-free-75diff-2	KLRWRIFKFKKIF	RWRFKWKK	0.636
Broad-free-75diff-3	KIWLKRIRIKIKFF	WIQRIRIWV	0.539
Broad-free-75diff-4	LIGLSGKLGLLLKKKLAL	KGLKKLGKLLKKLLKL	0.531
Broad-free-75diff-5	FIRWRLRFRLRIRFL	LRLRLRLRLRLRLR	0.593



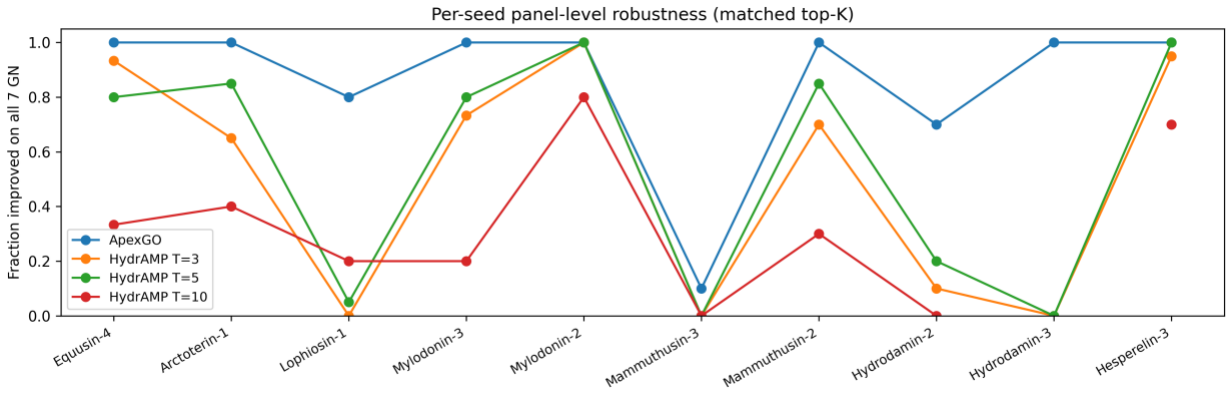
Supplementary Figure 10. Attention-based interpretation of APEX for representative ApexGO designs. We analyzed token-level attention weights from APEX for two template/optimized peptide pairs: Mylodonin-2/Mylodonin-2-3 (Gram-negative objective) and Mammuthusin-3/Mammuthusin-3-6 (broad objective). **(a,b)** Aligned attention heatmaps for template (top row) and optimized (bottom row) sequences; column labels denote aligned template/optimized residues at each position (including insertions/deletions). **(c,d)** Position-wise attention shifts (optimized minus template) across aligned positions; edited sites correspond to mismatched residue labels. **(e,f)** Attention mass aggregated by residue class (cationic, hydrophobic, polar, special, acidic, other) illustrates objective-dependent redistribution of attention across physicochemical residue types.



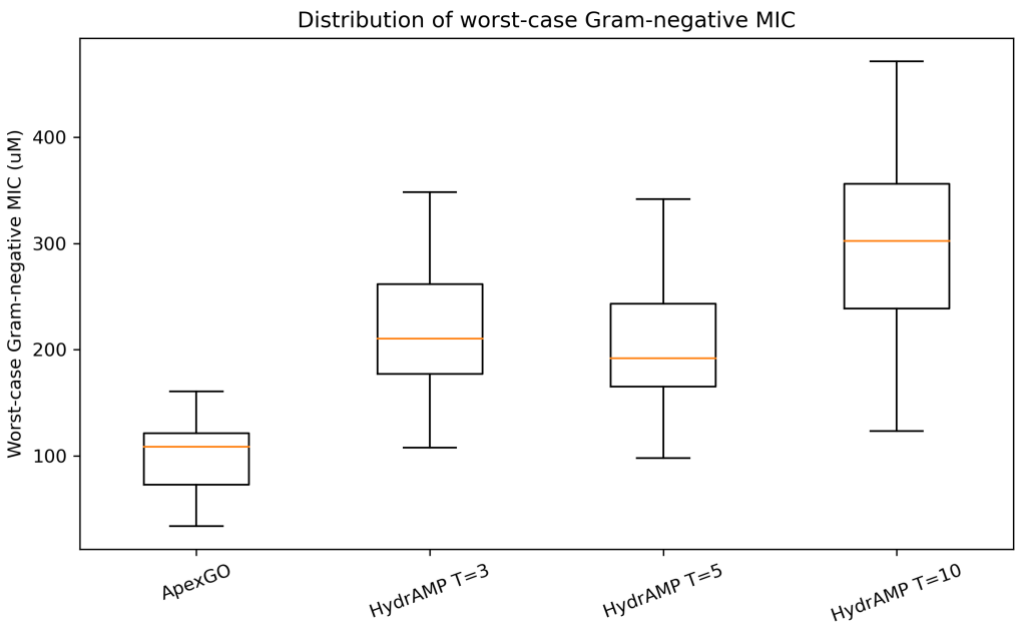
Supplementary Figure 11. Aggregate attention analysis across the top-2 ApexGO cohort. We analyzed 40 optimized peptides (top two designs per seed per objective; 20 Gram-negative objective and 20 broad objective). **(a)** For each objective, distributions of per-peptide mean optimized attention at edited vs unedited positions show that edited sites receive higher attention mass. **(b)** Mean residue-class attention-share shifts (optimized minus template) across the cohort highlight objective-dependent trends, including increased attention share on cationic residues for Gram-negative optimization and increased attention share on hydrophobic residues for broad-spectrum optimization.



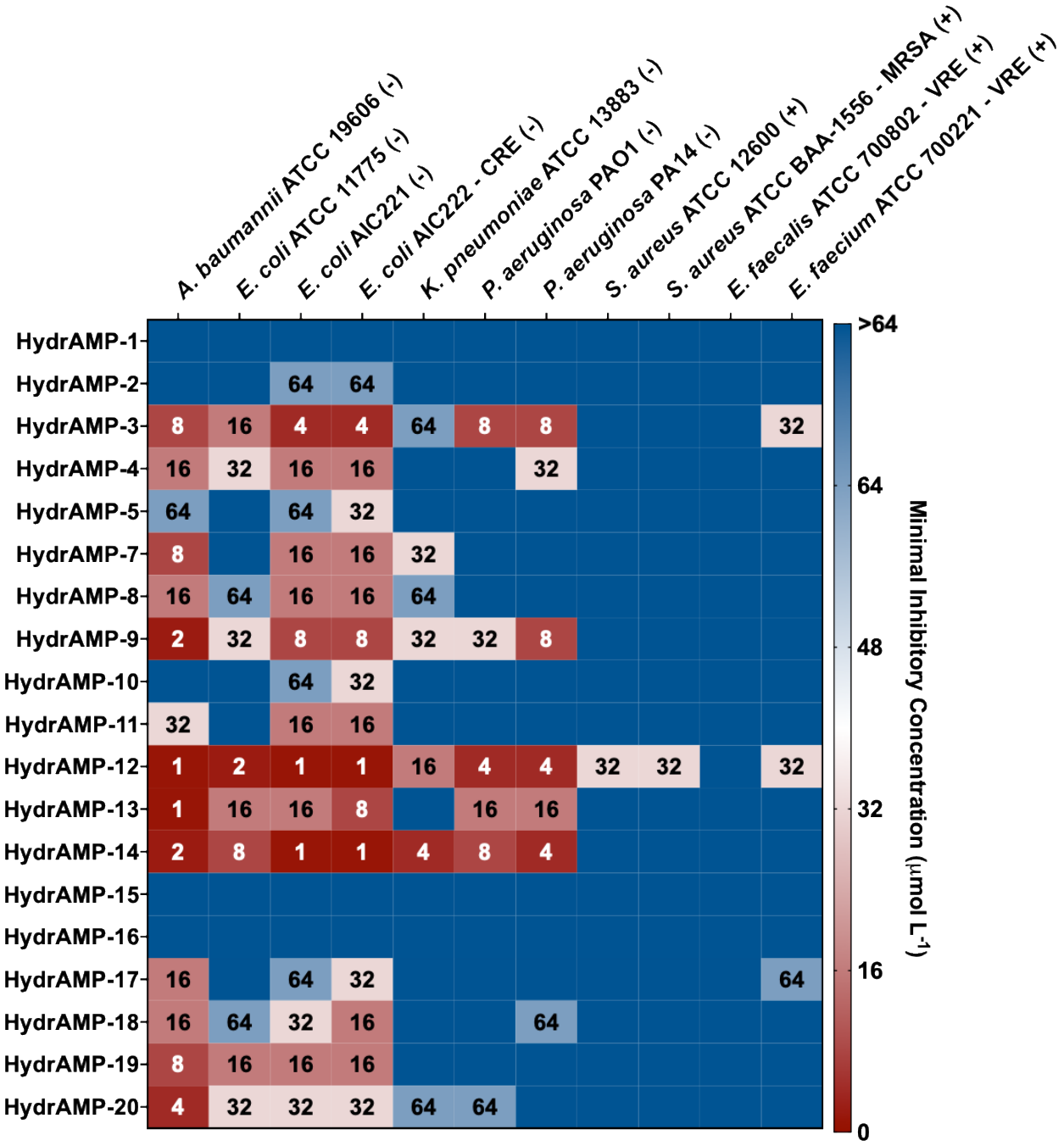
Supplementary Figure 12. Comparing ApexGO against LOL-BO (Maus et al., 2021), and standard latent-space BO (SCBO), modified using the approach of Eriksson et al., 2020 to high dimensional optimization. Each point represents one template-specific run (10 points per method). All methods were evaluated using the same peptide VAE to ensure fair comparison.



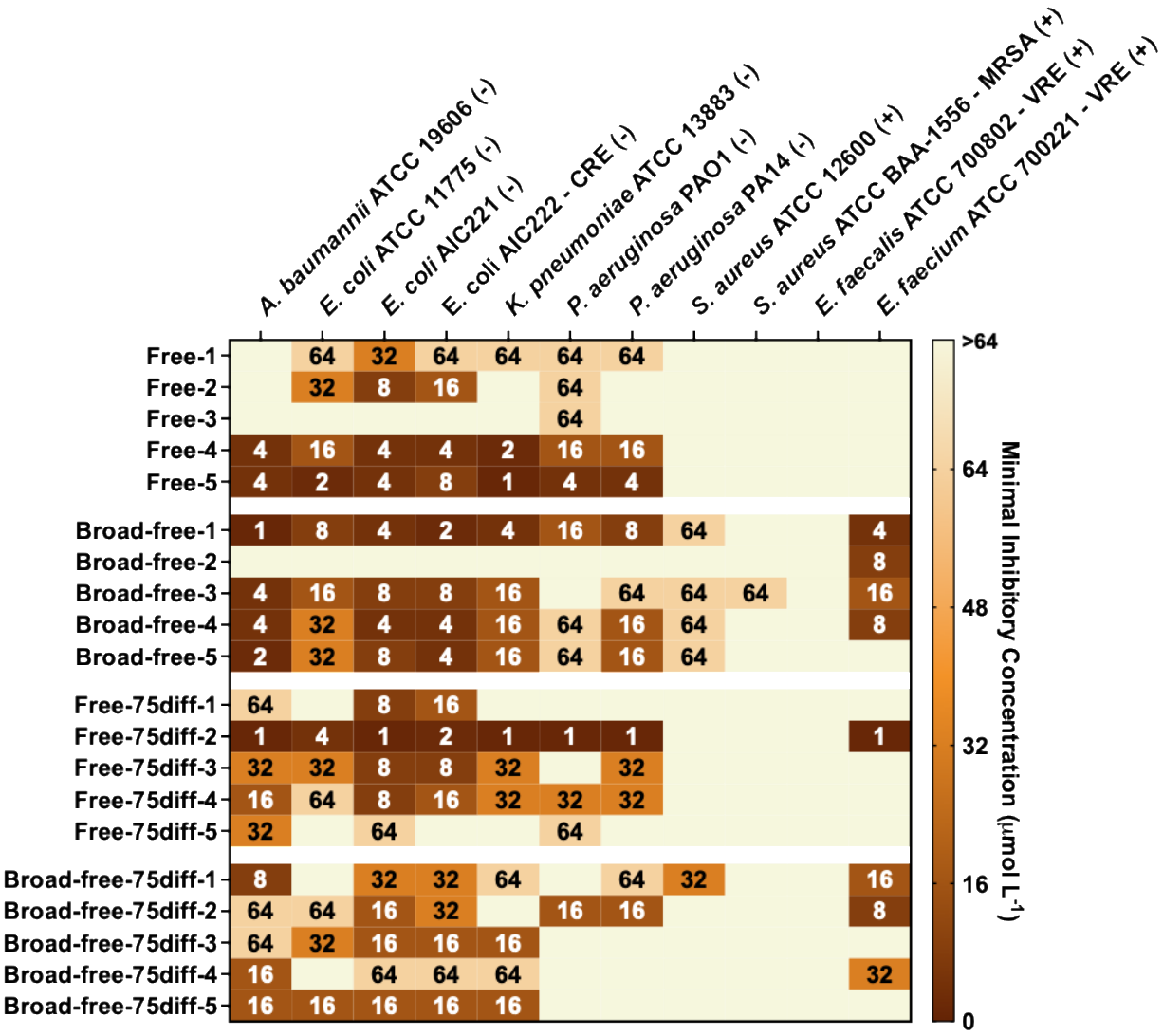
Supplementary Figure 13. Per-seed panel-level robustness for Gram-negative optimization (matched top-K). For each seed template and method, we report the fraction of feasible designs whose APEX-predicted MIC improved relative to the seed across all seven Gram-negative pathogens. To ensure a fair comparison across methods with different feasibility rates, K is matched per seed to the number of feasible ApexGO designs (required K).



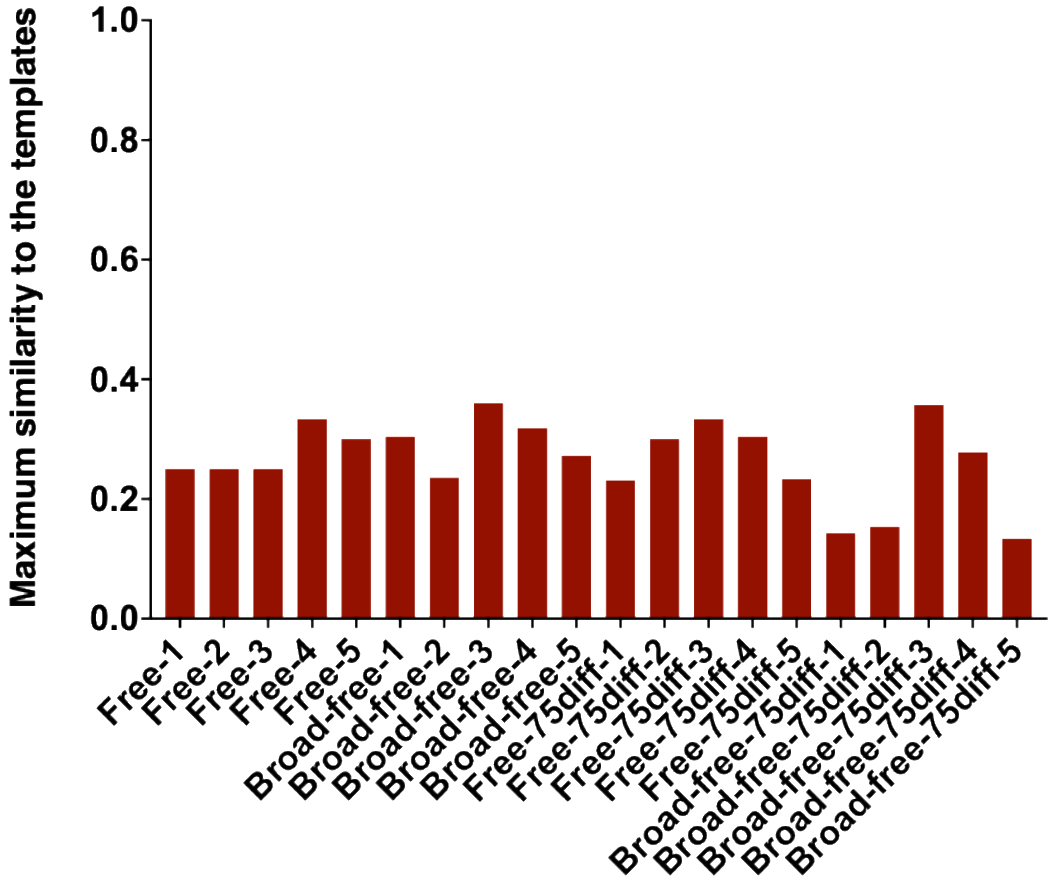
Supplementary Figure 14. Worst-case Gram-negative predicted MIC across feasible sets (matched top-K). Boxplots show, for each seed, the distribution of worst-case APEX-predicted MIC across the seven Gram-negative organisms for each method's top-K feasible designs. Lower values correspond to better predicted activity.



Supplementary Figure 15. Heatmap of minimal inhibitory concentration (MIC) values in μmol L^{-1} for the HydrAMP generated peptides. Red color indicates lower MIC values (higher antimicrobial activity), whereas blue color indicates lower MICs or inactive molecules. The MIC values are indicated within each cell of the heatmap, and the value is the mode of three independent replicates.



Supplementary Figure 16. Heatmap of minimal inhibitory concentration (MIC) values in μmol L^{-1} for the freely generated peptides. Darker colors indicate lower MIC values (higher antimicrobial activity), whereas lighter colors indicate lower MICs or inactive molecules. The MIC values are indicated within each cell of the heatmap, and the value is the mode of three independent replicates.



Supplementary Figure 17. Maximum similarity between freely generated peptides and the parent templates used in our main experiments. For each template, we measured $1 - \frac{1}{|s|}\min_t d(s,t)$ where $|s|$ is the length of the freely generated peptide s, and $\min_t d(s,t)$ is the minimum edit (Levenshtein) distance from the freely generated peptide to any of our templates. No freely generated peptide has more than 33% similarity to any peptides generated by our original experiments.