nature biomedical engineering

PEER-REVIEW INFORMATION

Deep mutational scanning and machine learning for the analysis of antimicrobial-peptide features driving membrane selectivity

Corresponding author: Bryan Davies

Editorial note

This document includes relevant written communications between the manuscript's corresponding author and the editor and reviewers of the manuscript during peer review. It includes decision letters relaying any editorial points and peer-review reports, and the authors' replies to these (under 'Rebuttal' headings). The editorial decisions are signed by the manuscript's handling editor, yet the editorial team and ultimately the journal's Chief Editor share responsibility for all decisions.

Any relevant documents attached to the decision letters are referred to as **Appendix #**, and can be found appended to this document. Any information deemed confidential has been redacted or removed. Earlier versions of the manuscript are not published, yet the originally submitted version may be available as a preprint. Because of editorial edits and changes during peer review, the published title of the paper and the title mentioned in below correspondence may differ.

Correspondence

Tue 19 Sep 2023
Decision on Article nBME-23-2228

Dear Dr Davies,

Thank you for submitting to *Nature Biomedical Engineering* your manuscript, "Deep mutational scanning and machine learning uncover antimicrobial peptide features driving membrane selectivity". We decided to recruit expert advice, and we have now received feedback from three experts. You will find their reports at the end of this message.

You will see that the reviewers appreciate the study. However, their criticisms suggest that the work would need substantial additional evidence to support the reproducibility, robustness and validity of the findings. We hope that you will be able to follow the reviewers' many useful suggestions and improve the manuscript accordingly. In particular, I am hoping that substantial effort will allow you to address the criticisms, particularly regarding the suitability of the method used for feature selection, the validation of the model in independent datasets, and the quantitative performance and robustness of the workflow for deep mutational and surface localized antimicrobial display. Also, please ensure that the methodology is thoroughly reported.

When you are ready to resubmit your manuscript, please upload the revised files, a point-by-point rebuttal to the comments from all reviewers, the reporting summary, and a cover letter that explains the main improvements included in the revision and responds to any points highlighted in this decision.

Please follow the following recommendations:

* Clearly highlight any amendments to the text and figures to help the reviewers and editors find and understand the changes (yet keep in mind that excessive marking can hinder readability).

* If you and your co-authors disagree with a criticism, provide the arguments to the reviewer (optionally,

indicate the relevant points in the cover letter).

* If a criticism or suggestion is not addressed, please indicate so in the rebuttal to the reviewer comments and explain the reason(s).

* Consider including responses to any criticisms raised by more than one reviewer at the beginning of the rebuttal, in a section addressed to all reviewers.

* The rebuttal should include the reviewer comments in point-by-point format (please note that we provide all reviewers will the reports as they appear at the end of this message).

* Provide the rebuttal to the reviewer comments and the cover letter as separate files.

We hope that you will be able to resubmit the manuscript within 16 weeks from the receipt of this message. If this is the case, you will be protected against potential scooping. Otherwise, we will be happy to consider a revised manuscript as long as the significance of the work is not compromised by work published elsewhere or accepted for publication at *Nature Biomedical Engineering*.

We hope that you will find the referee reports helpful when revising the work. Please do not hesitate to contact me should you have any questions.

Best wishes,

Pep

__
Pep Pàmies
Chief Editor, *Nature Biomedical Engineering*

_____
Reviewer #1 (Report for the authors (Required)):

In this manuscript, the authors have developed deep mutational surface localized antimicrobial display (dmSLAY) for identifying potent antimicrobial peptides from a library of Protegtrin-1 (PG-1) variants. Subsequently, these data were used for training machine learning models for predicting the antimicrobial potency, toxicity and bacterial membrane selectivity of the full spectrum of possible one, two and three residue PG-1 variants. While the study demonstrates 28 predictions exhibiting heightened bacterial membrane selectivity, alongside comparable antimicrobial activity and reduced toxicity when compared to PG-1 in vitro. However, there are concerns about the robustness and function of the machine learning models used in this work.
Here are my detailed concerns:

Major Concerns:
1. The authors trained SVM models with only 96 sequences to predict various properties of peptides. This small dataset could lead to overfitting, where the models perform well on the training data but fail to generalize to unseen data. The manuscript should provide details on how overfitting was addressed and how model generalization was ensured.

2. Fig. 5 shows mutational heat maps of machine learning predictions, but the manuscript does not provide the accuracy of these models for generalizing to unseen data. Without an assessment of model accuracy, it's challenging to interpret the significance of these predictions.

3. The manuscript mentions that 36 candidates were chosen for further wet-lab measurements from a pool of 17,348 potential variants. The criteria and methods for selecting these 36 candidates should be clearly explained, as this is a critical step in the process.

4. The manuscript mentions that the dmSLAY technique was integrated with SLAY, assessing antibacterial

potential of hundreds of thousands of sequences, and deep mutational scanning, assessing up to millions of mutations. Still, it's unclear what advantages dmSLAY offers over these techniques, especially since dmSLAY was applied to assess only thousands of peptides. The manuscript should clarify the benefits of dmSLAY in comparison to existing methods.

5. The authors have previously reported work combining wet-lab experiments and machine learning for mining low-toxicity AMPs (Sci. Adv., 2023, 9, eade0008). The manuscript should highlight the differences and unique contributions of the current work compared to this previous study.

6. The manuscript should consider providing more information about the novelty of the identified AMPs, their potential for drug resistance, and their therapeutic effects in vivo, as these are crucial aspects for assessing their practical utility.

Minor Issues:

1. Given the importance of machine learning in this work, the introduction should provide more information about the role of machine learning in the context of antimicrobial peptide discovery.

2. The manuscript should provide a clear legend explaining the meaning of color-coded cells (e.g., blue, red, gray, black, white) in Fig. 1d.

Reviewer #2 (Report for the authors (Required)):

Randall et al. evaluate the potency and selectivity of variants of the antimicrobial peptide protegrin-1. They implement their previous innovation – surface-localized antimicrobial display (SLAY) – to measure bacterial killing for >7,000 variants. In benchmarking with bacteriocidal assays with purified peptide, minimal correlation is observed although binary prediction of active/inactive was reasonable (86% accuracy). Secondary structure analysis of 40 variants suggested that potency benefited from maintained structure, which is consistent with the well-established structure-function dependence of proteins. Five variants were observed to have improved selectivity for bacterial lysis relative to hemolysis, and the mutational dependence of these variants was studied by mutational reversion. Finally, a machine learning model was used to explore a broader range of sequences, which suggested sequence elements that drive potency and selectivity. Testing of 36 designed variants yielded several peptides with improved selectivity and maintained potency relative to wild-type protegrin-1.

Overall, the manuscript has strong potential to be a good contribution to the field. Antimicrobial peptide engineering is an important technological goal to combat antibiotic resistance. Selectivity, to avoid human cell lysis, is an important design parameter. The combination of a high-throughput platform (dmSLAY), purified peptide characterization, and a mathematical model of the sequence-performance relationship demonstrates a reasonably effective approach to peptide engineering and identifies improved variants of protegrin-1. Yet the inability of dmSLAY (with or without supplementation of the machine learning model) to predict bacterial potency is a significant drawback. The ability of the study to assess (and improve upon) selectivity may be the larger impact. However, the selectivity analysis does not emerge from dmSLAY but rather from ‘classic’ experiments with individual peptides. As a result, the title and other writing is somewhat misleading in the driver of the advance. The work is generally thorough with assertions generally supported by presented data (see exceptions below). The writing and data presentation are generally clear. The manuscript will benefit from attention to several current shortcomings, detailed below.

1. There are three considerations for Figure 1C.
a. Figure 1C presents data from a highly valuable experiment to benchmark the dmSLAY platform. As such its presence is a key benefit to the rigor of the paper. Unfortunately, the results indicate a lack of quantitative utility to the dmSLAY platform as there is no appreciable correlation between read change and MIC. The authors appropriately acknowledge this in the text, but it nevertheless diminishes from the utility of the platform.

b. As for the focus on a binary binning of active/inactive, it could be useful to provide a receiver-operator characteristic curve generated by varying the MIC threshold and/or the read change threshold.

c. Also, Figure 1C could be improved by labeling the y-axis more clearly. The text indicates that 64 μg/mL is used as the inactive/active threshold, yet the figure places an active/inactive line halfway between tick marks labeled 64 and >64. Thus, the ticks seem to be 0, 32, 64, 96, >64. Please edit for clarity.

2. Data reproducibility of dmSLAY should be evaluated. While visual indication of variance would be too crowded in Figure 1B, the text could provide the average read variance; or a histogram of variances could be provided in a supporting figure. The particular format is not critical, but evaluation of reproducibility is critical.

3. Figure 1D is relatively sparse. While the increase in throughput relative to traditional methods is notable, sequence space (particularly for single mutants) remains relatively under-examined. Thus, the limitations in asserting the mutational flexibility at each site should be acknowledged.

4. There are numerous statements that would benefit from more rigorous, quantitative analysis. For example:

“Residues in the loop and tail regions of PG-20 were generally the most flexible while residues in the sheet regions were predicted to be more important for antibacterial activity.”

“Variants retaining activity mostly contained changes in the loop and tail regions, but some contained changes to sheet residues with similar side chain amino acids.”

5. Data reproducibility (e.g. standard deviation of replicate measurements) should be provided for the MIC values in Figure S1A as well as all values in Table S2.

6. For the five variants with improved selectivity scores (Figure 3A), it would be useful to know if the selectivity was improved by reduced hemolysis, decreased MIC, or a combination of both. Perhaps provide a supplemental table of the hemolysis, MIC, and selectivity for all Figure 3A variants or maybe simply note the hemolysis and MIC below the variant names in Figure 3A (for PG1.0 and the five variants).

7. How were the 36 experimental variants (Table S2) chosen from the 17,348 designed variants (Figure 5)?

8. The relative inability of the machine learning model to predict potency is interesting (and technologically disappointing), especially because the model was trained on dmSLAY potency. It would be useful for the authors to comment on this. Also, Figure S5 should show the model’s train/test performance on potency (akin to Figure S5B,C for hemolysis and selectivity).

9. The machine learning model should be described in more detail. The shared code is effective, but a concise description of key elements would be useful for readers who do not view the code and supplementary Github content. Also, it would be valuable to assess the impact of the data characteristics on predictive performance: how is predictive quality impacted by data depth, data breadth (range), and model parameters? This will provide useful insight for future practitioners.

10. Minor point: Figure 1A is mildly misleading. The cells with potent protegrin variants are killed, but the DNA is not eliminated (as suggested by the schematic). Rather those cells lack the ability to multiply.

Reviewer #3 (Report for the authors (Required)):

In this manuscript, the authors describe a method called deep mutational SLAY scanning (dmSLAY) to assess peptide sequence variants and their antimicrobial profiles. dmSLAY was applied to the relatively toxic antimicrobial peptide Protegrin-1, leading to variants that preserved their antibacterial efficacy while displaying improved membrane specificity towards bacteria as opposed to red blood cells. The authors further showed that machine learning tools could be built on the results derived from dmSLAY to computationally uncover additional antimicrobial Protegrin-1 variants with improved membrane selectivity. Overall, the proposed method could be useful to investigate the quantitative sequence/structure–activity relationships of peptides. Unfortunately, as it stands, the paper is too preliminary for Nature BME, as delineated below.

Major issues:
- Regarding the protein representations, the authors used the embeddings extracted from the ESM model and performed a feature selection to reduce the feature dimensionality. Since the features extracted by

neural networks without disentanglement are usually known as distributed representations (i.e., a single concept is expressed by multiple units in a vectorized representation), discrete feature selection may not be the best way to reduce the dimensionality. The authors should investigate other continuous dimensionality reduction approaches (e.g., PCA or other supervised alternatives) on ESM embeddings.

- NPN assays should be performed to assess whether the peptide variants permeabilize the outer membrane of Gram-negative bacteria such as E. coli as this will shed further insight into the mechanism of action of the peptides generated.

- Cytotoxicity assays using human cell lines should be performed as this is standard in the field.

- In vivo validation using a mouse model is needed to ensure the peptides generated target bacterial pathogens while not causing toxicity in a realistic and relevant setting. This would provide a robust validation of dmSLAY and is expected for a journal like Nature BME.

Tue 16 Jan 2024
Decision on Article nBME-23-2228A

Dear Dr Davies,

Thank you for your patience in waiting for the feedback on your revised manuscript, "Deep mutational scanning and machine learning uncover antimicrobial peptide features driving membrane selectivity", which has been seen by the original reviewers. In their reports, which you will find at the end of this message, you will see that the reviewers are mostly happy with the improvements to the work and that Reviewer #3 finds that in vivo evidence of the activity and toxicity of the selected membrane-specific PG-1 variants would be warranted. We agree with this point, and hope that you can provide in vivo data, at least in G. mellonella larvae.

As before, when you are ready to resubmit your manuscript, please upload the revised files, a point-by-point rebuttal to the comments from all reviewers, the reporting summary, and a cover letter that explains the main improvements included in the revision and responds to any points highlighted in this decision.

We look forward to receive a further revised version of the work. Please do not hesitate to contact me should you have any questions.

Best wishes,

Pep

__
Pep Pàmies
Chief Editor, *Nature Biomedical Engineering*

_____
Reviewer #1 (Report for the authors (Required)):

The authors revised the manuscripts well and addressed most question I have. And I think it is can be accepted by Nature BME.

Reviewer #2 (Report for the authors (Required)):

The authors have improved the manuscript by effectively addressing most reviewer critiques. One primary concern was not addressed (perhaps because it was in the main paragraph and not a numbered element):

The ability of the study to assess (and improve upon) selectivity may be the larger impact. However, the selectivity analysis does not emerge from dmSLAY but rather from ‘classic’ experiments with individual peptides. As a result, the title and other writing is somewhat misleading in the driver of the advance.

In addition, it remains true that dmSLAY does not provide a continuous quantitative evaluation of molecular potency, which is a technological limitation. Nevertheless, the rigor and clarity in regard to the binary approach have been dramatically improved.

Overall, the manuscript is now in a form that provides substantial value to the scientific community. Ideally, brief editing to address the primary concern above would be ideal, although this reviewer does not advocate for another round of formal review to address this matter.

Reviewer #3 (Report for the authors (Required)):

Though I commend the authors for slightly improving the paper through their comments and revisions, I still think it is too preliminary and does not meet the minimum bar for Nature BME.

The authors mention that most AMPs require extensive synthetic modification to be effective in animal models. However, this is simply not true as has been extensively demonstrated in the literature (using mouse infection models) and in the author's own recent work (e.g., Randall JR et al. Adapting antibacterial display to identify serum-active macrocyclic peptide antibiotics. PNAS Nexus. 2023 Aug 17;2(8):pgad270. doi: 10.1093/pnasnexus/pgad270. PMID: 37637199; PMCID: PMC10449418). Testing the safety and efficacy of AMPs in mouse models is standard in the field and most studies these days show data with more than one mouse model.

Wed 08 May 2024
Decision on Article nBME-23-2228B

Dear Dr Davies,

Thank you for your revised manuscript, "Deep mutational scanning and machine learning uncover antimicrobial peptide features driving membrane selectivity". Having consulted with Reviewer #3, I am pleased to write that we shall be happy to publish the manuscript in *Nature Biomedical Engineering*, provided that the points specified in the attached instructions file are addressed.

When you are ready to submit the final version of your manuscript, please upload the files specified in the instructions file.

We encourage authors to take up transparent peer review. If you are eligible and opt in to transparent peer review, we will publish, as a single supplementary file, all the reviewer comments for all the versions of the manuscript, your rebuttal letters, and the editorial decision letters. **If you opt in to transparent peer review, in the attached file please tick the box 'I wish to participate in transparent peer review'; if you prefer not to, please tick 'I do NOT wish to participate in transparent peer review'.** In the interest of confidentiality, we allow redactions to the rebuttal letters and to the reviewer comments. If you are concerned about the release of confidential data, please indicate what specific information you would like to have removed; we cannot incorporate redactions for any other reasons. More information on transparent peer review is available.

Please do not hesitate to contact me should you have any questions.

Best wishes,

Pep

__
Pep Pàmies
Chief Editor, *Nature Biomedical Engineering*

_____
Reviewer #3 (Report for the authors (Required)):

The authors have addressed my prior concerns.

_____
Nature Biomedical Engineering is a Transformative Journal. Authors may publish their research with us through the traditional subscription access route, or make their paper immediately open access through payment of an article-processing charge. More information about publication options is available.

You may need to take specific actions to comply with funder and institutional open-access mandates. If the work described in the accepted manuscript is supported by a funder that requires immediate open access (as outlined, for example, by Plan S) and your manuscript was originally submitted on or after January 1st 2021, then you will need to select the gold OA route. Authors selecting subscription publication will need to accept our standard licensing terms (including our self-archiving policies), and these will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Rebuttal 1

Response to Editor/Reviewer comments

Key
black = Editor/reviewer comments
blue = response to comment
bolded = manuscript change

From the Editor:

You will see that the reviewers appreciate the study. However, their criticisms suggest that the work would need substantial additional evidence to support the reproducibility, robustness and validity of the findings. We hope that you will be able to follow the reviewers' many useful suggestions and improve the manuscript accordingly. In particular, I am hoping that substantial effort will allow you to address the criticisms, particularly regarding the suitability of the method used for feature selection, the validation of the model in independent datasets, and the quantitative performance and robustness of the workflow for deep mutational and surface localized antimicrobial display. Also, please ensure that the methodology is thoroughly reported.

We thank the reviewers for their thoughtful comments and suggestions and feel they have served to greatly improve the manuscript. We have responded to these comments individually below and any changes made to the manuscript resulting from these comments are **bolded**. While working on the reviewer suggestions, we discovered three sequence duplications in the dataset used to train our original machine learning models. These duplications have been removed in the revised manuscript, we have rerun the machine learning and updated the figures where necessary. This small change in the dataset did not have any major impact on any conclusions drawn in the original manuscript.

Briefly, some additional modifications include running a continuous dimensionality reduction approach used for feature selection (PCA) in our modelling, showing no large differences in performance between the two methods (see Github). We also reemphasized the evaluation of the machine learning model performance using an independent dataset by moving a summary of this analysis to the main text (now Fig. 6A). We also added cytotoxicity assays with selective PG-1 variants to further validate reduction in mammalian cell toxicity. We have added raw sequencing files used for dmSLAY analysis to the SRA database (Accession: PRJNA1022479) and ensured that intermediate files used for this analysis are included in the Github link. This should now allow for full independent reproduction of all our dmSLAY and machine learning methodology and analysis.

Reviewer #1 (Report for the authors (Required)):

In this manuscript, the authors have developed deep mutational surface localized antimicrobial display (dmSLAY) for identifying potent antimicrobial peptides from a library of Protegtrin-1 (PG-1) variants. Subsequently, these data were used for training machine learning models for predicting the antimicrobial potency, toxicity and bacterial membrane selectivity of the full spectrum of possible one, two and three residue PG-1 variants. While the study demonstrates 28 predictions exhibiting heightened bacterial membrane selectivity, alongside comparable antimicrobial activity and reduced toxicity when compared to PG-1 in vitro. However, there are concerns about the robustness and function of the machine learning models used in this work.
Here are my detailed concerns:

Major Concerns:
1. The authors trained SVM models with only 96 sequences to predict various properties of peptides. This small dataset could lead to overfitting, where the models perform well on the training data but fail to generalize to unseen data. The manuscript should provide details on how overfitting was addressed and how model generalization was ensured.

We appreciate the reviewer's insightful observation regarding the potential overfitting risk associated with our limited dataset size.

Due to the nature of our dataset and the challenges of obtaining more data (given the cost and labor-intensive process of testing peptides in the lab), we recognize that our analysis might not be as quantitatively robust as with larger datasets. Nevertheless, our approach emphasizes qualitative measures and methodologies designed to mitigate overfitting. This included:

1. Dimensionality Reduction: To address the overfitting risks associated with high-dimensional data, we reduced our feature dimensions significantly, from 5,120 to approximately 150. This strategy aimed not only to streamline our models but also to minimize the risk of fitting noise present in the data. The 150 features gave us a good balance between performance while it tries to mitigate overfitting. If more features are used, signals of overfitting start to show, such as a good performance on train but a poor performance on test set.

2. Cross-Validation: To validate our model's performance and its capability to generalize, we adopted a cross-validation strategy. It's worth noting that we encountered some small values during this process. Such values could suggest sensitivity to specific data splits of the data, although the overall mean and when trained on whole train set, we obtain a good performance. Suggesting that if more data is add this could contribute with a better model.

3. Test vs. Train Performance: One of the primary methods to detect overfitting is by evaluating a model's performance on a training set. Overfitting is suspected when the model performs considerably better on the training set than on the unseen data (test set). Interestingly, in our study, the models, when trained on the entire training set, frequently exhibited similar or even superior performance on the test set compared to the training set.
See notebook 2 04-06 on results visualization:

4. Consensus Model Approach: As part of our strategy for robustness, we utilized three distinct models: one Lasso, one SVM regressor and a Gaussian Naive Bayes (GNB) classifier. Each model was trained to predict different features from the peptide dataset. For final selection, peptides were shortlisted only if they met our strict criteria across all three models, ensuring a higher level of confidence in our predictions.

In summary, while the limitations posed by our dataset size are acknowledged, our methodologies and qualitative measures were implemented with the intent of maximizing model reliability and preventing overfitting.

In response to this comment, we have added a more detailed description of our machine learning modeling to the supplemental information. We also created a version 2 of our models

which is included in the GitHub link and discussed in more detail below (Reviewer #3, first comment). These modifications did not substantially improve the accuracy of the original modelling.

2. Fig. 5 shows mutational heat maps of machine learning predictions, but the manuscript does not provide the accuracy of these models for generalizing to unseen data. Without an assessment of model accuracy, it's challenging to interpret the significance of these predictions.

The presentation of our data must not have been clear. We did evaluate the bacterial specific modeling with an independent dataset in Table S3 and reported on the accuracy of the three models used. To remedy this, **we have moved Table S3 to the main text as part of Figure 6.**

3. The manuscript mentions that 36 candidates were chosen for further wet-lab measurements from a pool of 17,348 potential variants. The criteria and methods for selecting these 36 candidates should be clearly explained, as this is a critical step in the process.

The candidates were randomly chosen from four different machine learning predicted log10 selectivity score bins. **We have added text explaining this process (page 7 line 1) and added a column delineating this to Table S2.** We removed three duplicate sequences found in the original training dataset so now only 32 of the 36 original peptides chosen still meet the original cut offs. The four which no longer met the predictive cut offs were dropped from the analysis in the revised manuscript.

4. The manuscript mentions that the dmSLAY technique was integrated with SLAY, assessing antibacterial potential of hundreds of thousands of sequences, and deep mutational scanning, assessing up to millions of mutations. Still, it's unclear what advantages dmSLAY offers over these techniques, especially since dmSLAY was applied to assess only thousands of peptides. The manuscript should clarify the benefits of dmSLAY in comparison to existing methods.

In the past, SLAY was used to identify antibacterial peptide sequences from hundreds of thousands of highly variable sequences. Here we apply the same SLAY procedure, but to a library of several thousand peptides representing a detailed mutational spectrum of Protegrin-1; hence the change in nomenclature to deep mutational SLAY scanning (dmSLAY). The advantage of dmSLAY over original SLAY is that it is used to identify both active and inactive variants of a known antibacterial peptide sequence. This allows us to determine positional residue flexibility and importance and select peptides with diverse activities to create better datasets to train machine learning models. **We have adjusted text in the introduction to better emphasize this difference (Introduction, paragraph 3).**

5. The authors have previously reported work combining wet-lab experiments and machine learning for mining low-toxicity AMPs (Sci. Adv., 2023, 9, eade0008). The manuscript should highlight the differences and unique contributions of the current work compared to this previous study.

Good point. The original study evaluated the antimicrobial activity of a portion of a synthetic peptide library with beta-hairpin structure using SLAY (~40,000 of ~200,000 total unique sequences). We then using machine learning to predict the antimicrobial activity for the remaining ~160,000 peptides in the library. Here we use dmSLAY to evaluate how thousands of mutational variations positively or negatively impact the antimicrobial activity of a known AMP sequence (PG-1). Using data obtained during the evaluation of dmSLAY leads, we applied machine learning to identify mutational profiles

impacting potency, toxicity, and bacterial or mammalian cell membrane selectivity. **We have adjusted text in the introduction to better emphasize this point (Introduction, paragraph 3)**

6. The manuscript should consider providing more information about the novelty of the identified AMPs, their potential for drug resistance, and their therapeutic effects in vivo, as these are crucial aspects for assessing their practical utility.

The selective PG-1 mutants identified here have not been previously described in any literature. PG-1 and our selective variants kill through inner and outer membrane disruption (Fig. 4A). Previous work (Steinberg et. al., Antimicrob Agents Chemother. 1997) has shown that resistance mechanisms to PG-1 activity do not readily develop. **Additionally, we have added experiments as Figure 6BC showing selective PG-1 variants are less toxic to tissue culture cells (HEK293) and that PG-1 and our selective variants are unaffected by a known mechanism of AMP resistance (mobile colistin resistance, a.k.a. mcr-1). We have also added some text describing these experiments and discussing that extensive synthetic modifications are usually required to have clinical antibiotic relevance. This goes beyond simple toxicity issues (page 7, paragraph 2).**

Minor Issues:

1. Given the importance of machine learning in this work, the introduction should provide more information about the role of machine learning in the context of antimicrobial peptide discovery.

Great point. **We have added text and cited articles regarding this topic to the Introduction (Introduction, paragraph 3).**

2. The manuscript should provide a clear legend explaining the meaning of color-coded cells (e.g., blue, red, gray, black, white) in Fig. 1d.

This has been added to the figure.

Reviewer #2 (Report for the authors (Required)):

Randall et al. evaluate the potency and selectivity of variants of the antimicrobial peptide protegrin-1. They implement their previous innovation – surface-localized antimicrobial display (SLAY) – to measure bacterial killing for >7,000 variants. In benchmarking with bactericidal assays with purified peptide, minimal correlation is observed although binary prediction of active/inactive was reasonable (86% accuracy). Secondary structure analysis of 40 variants suggested that potency benefited from maintained structure, which is consistent with the well-established structure-function dependence of proteins. Five variants were observed to have improved selectivity for bacterial lysis relative to hemolysis, and the mutational dependence of these variants was studied by mutational reversion. Finally, a machine learning model was used to explore a broader range of sequences, which suggested sequence elements that drive potency and selectivity. Testing of 36 designed variants yielded several peptides with improved selectivity and maintained potency relative to wild-type protegrin-1.

Overall, the manuscript has strong potential to be a good contribution to the field. Antimicrobial peptide engineering is an important technological goal to combat antibiotic resistance. Selectivity, to avoid human cell lysis, is an important design parameter. The combination of a high-throughput platform

(dmSLAY), purified peptide characterization, and a mathematical model of the sequence-performance relationship demonstrates a reasonably effective approach to peptide engineering and identifies improved variants of protegrin-1. Yet the inability of dmSLAY (with or without supplementation of the machine learning model) to predict bacterial potency is a significant drawback. The ability of the study to assess (and improve upon) selectivity may be the larger impact. However, the selectivity analysis does not emerge from dmSLAY but rather from 'classic' experiments with individual peptides. As a result, the title and other writing is somewhat misleading in the driver of the advance. The work is generally thorough with assertions generally supported by presented data (see exceptions below). The writing and data presentation are generally clear. The manuscript will benefit from attention to several current shortcomings, detailed below.

1. There are three considerations for Figure 1C.
a. Figure 1C presents data from a highly valuable experiment to benchmark the dmSLAY platform. As such its presence is a key benefit to the rigor of the paper. Unfortunately, the results indicate a lack of quantitative utility to the dmSLAY platform as there is no appreciable correlation between read change and MIC. The authors appropriately acknowledge this in the text, but it nevertheless diminishes from the utility of the platform.

This is true. We did not observe a correlation between potency and dmSLAY log_2-fold change in reads. However, the accuracy of the general activity predictions does allow for identification of positional importance and flexibility as well as prioritization of leads based on retention or loss of activity.

b. As for the focus on a binary binning of active/inactive, it could be useful to provide a receiver-operator characteristic curve generated by varying the MIC threshold and/or the read change threshold.

This is a great suggestion. **We have added this as Fig. S1D**. The binning of active and inactive was best optimized for a log2-fold change of 0 and active/inactive cut off of ≤ 64 ug/ml. **We added mention of this in the text and adjusted Fig. 1C slightly to better reflect this (page 3, paragraph 2).**

c. Also, Figure 1C could be improved by labeling the y-axis more clearly. The text indicates that 64 µg/mL is used as the inactive/active threshold, yet the figure places an active/inactive line halfway between tick marks labeled 64 and >64. Thus, the ticks seem to be 0, 32, 64, 96, >64. Please edit for clarity.

We have modified the axis to better reflect the data. Minimum inhibitory concentrations are performed in two-fold dilutions. Our active cut off was at 64 ug/ml so inactive peptides fall into ≥128. **The axis tick marks and labels in Fig. 3C have been modified to better reflect this.**

2. Data reproducibility of dmSLAY should be evaluated. While visual indication of variance would be too crowded in Figure 1B, the text could provide the average read variance; or a histogram of variances could be provided in a supporting figure. The particular format is not critical, but evaluation of reproducibility is critical.

This is a great point. **We conducted a principal component analysis (PCA) to assess the variance in the dmSLAY samples, both in their induced and uninduced states, taken in triplicate. This has been added as Figure S1C**. This shows clear grouping, distinct separation, and high variance captured. These results suggest that our data collection and analysis methods for dmSLAY samples

are both consistent and reproducible. Some d**escriptive text was also added (page 3, paragraph 1)**

3. Figure 1D is relatively sparse. While the increase in throughput relative to traditional methods is notable, sequence space (particularly for single mutants) remains relatively under-examined. Thus, the limitations in asserting the mutational flexibility at each site should be acknowledged.

The data for single mutants with a significant log_2-fold change greater than zero (blue) or less than -1 (red) is fairly sparse. However, data used to generate average log_2-fold change across a position includes data from single mutants with significant log_2-fold change from -1 to 0 (gray) which greatly increases the number of changes considered. From our *in vitro* validation (Fig. 1C) it appears these sequences are likely to retain activity. **We have adjusted the text and figure to better reflect this point while still also stating that not every possible residue change is included in the dataset (page 3, paragraph 1).**

4. There are numerous statements that would benefit from more rigorous, quantitative analysis. For example:

“Residues in the loop and tail regions of PG-20 were generally the most flexible while residues in the sheet regions were predicted to be more important for antibacterial activity.”

“Variants retaining activity mostly contained changes in the loop and tail regions, but some contained changes to sheet residues with similar side chain amino acids.”

After reviewing the data, we have slightly adjusted these statements. We feel the adjusted statements are strongly supported by dmSLAY log_2-fold change scores from both single and multiple variants data (Fig. 1D and Fig. S2), which have been shown to be 86% accurate. Residues located in the tail regions are generally more flexible (retaining activity in dmSLAY, negative log_2-fold change) while changes in the sheet regions were generally more likely to cause a loss of activity (high log_2-fold change). Some changes in the sheet regions are tolerated (i.e. V14I, V14L) which are changes of residues with hydrophobic side changes to other hydrophobic residues. **Some additional text adjustments have been made to these sections to help clarify these claims and which data supports them (page 3, paragraphs 3 and 4).**

5. Data reproducibility (e.g. standard deviation of replicate measurements) should be provided for the MIC values in Figure S1A as well as all values in Table S2.

MIC values are generally examined at two-fold dilutions. This is how all MICs were determined here. This would make means and standard deviations deceiving and is why we chose to report median values of triplicate reactions. **We have added to the MIC methods text that we never observed more than a two-fold change in MIC score for any peptide examined (page 16, paragraph 1). We have also described how MIC is reported in the Figure S1 legend and Table S1 and S2 legend.**

6. For the five variants with improved selectivity scores (Figure 3A), it would be useful to know if the selectivity was improved by reduced hemolysis, decreased MIC, or a combination of both. Perhaps provide a supplemental table of the hemolysis, MIC, and selectivity for all Figure 3A variants or maybe simply note the hemolysis and MIC below the variant names in Figure 3A (for PG1.0 and the five variants).

This is a great point. The individual MIC and hemolysis values for these peptides are listed in Table S1 **However, we have added some text and a panel to Fig. S3 showing the log_2-fold change in MIC versus hemolysis relative to PG-1.0 for this group which clearly shows differences in hemolysis are most strongly impacting selectivity score (page 4, paragraph 3).**

7. How were the 36 experimental variants (Table S2) chosen from the 17,348 designed variants (Figure 5)?

We have added text describing how these variants were selected (page 7, paragraph 1). The number of experiment variants was reduced to 32 after machine learning was rerun without the duplicate sequences because 4 of the original sequences no longer met the stated predictive cut offs.

8. The relative inability of the machine learning model to predict potency is interesting (and technologically disappointing), especially because the model was trained on dmSLAY potency. It would be useful for the authors to comment on this. Also, Figure S5 should show the model's train/test performance on potency (akin to Figure S5B,C for hemolysis and selectivity).

We may have been unclear regarding the specifics of our potency model. This is a binary model which only predicts the probability a variant has an MIC ≤ 8 µg/ml (potent) or > 8 µg/ml (not potent) it doesn't attempt to predict a range of potencies. Also, the model was trained solely on MIC data from collected from PG-1 dmSLAY variants not the log_2-fold change scores generated from dmSLAY. **We have made some subtle text changes to the abstract and the machine learning portion of the main text (page 6, last paragraph) and the detailed machine leaning methods section in the supplemental information to help better clarify these points.**

9. The machine learning model should be described in more detail. The shared code is effective, but a concise description of key elements would be useful for readers who do not view the code and supplementary Github content. Also, it would be valuable to assess the impact of the data characteristics on predictive performance: how is predictive quality impacted by data depth, data breadth (range), and model parameters? This will provide useful insight for future practitioners.

We have added a more detailed description of how our machine learning models set up and trained in the supplementary information.

10. Minor point: Figure 1A is mildly misleading. The cells with potent protegrin variants are killed, but the DNA is not eliminated (as suggested by the schematic). Rather those cells lack the ability to multiply.

This is true. We have adjusted Fig. 1A and the legend to help reflect this.

Reviewer #3 (Report for the authors (Required)):

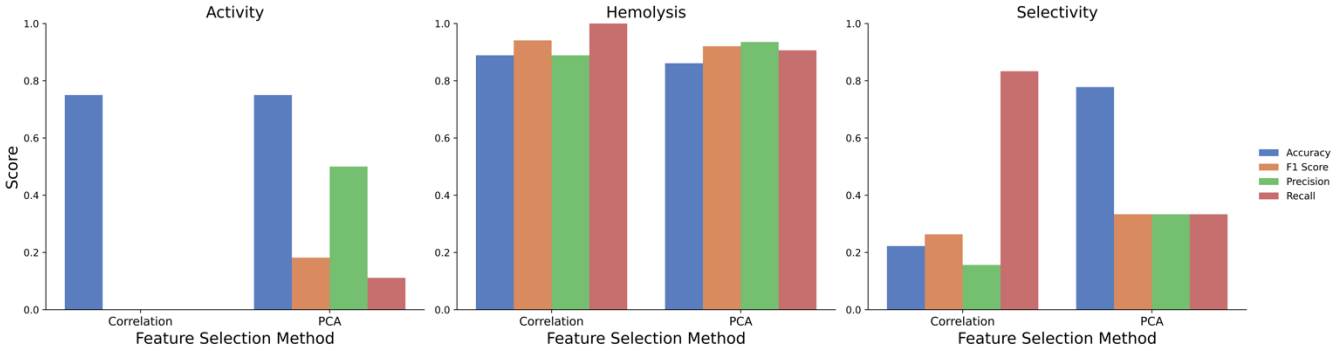
In this manuscript, the authors describe a method called deep mutational SLAY scanning (dmSLAY) to assess peptide sequence variants and their antimicrobial profiles. dmSLAY was applied to the relatively toxic antimicrobial peptide Protegrin-1, leading to variants that preserved their antibacterial efficacy while displaying improved membrane specificity towards bacteria as opposed to red blood cells. The

authors further showed that machine learning tools could be built on the results derived from dmSLAY to computationally uncover additional antimicrobial Protegrin-1 variants with improved membrane selectivity. Overall, the proposed method could be useful to investigate the quantitative sequence/structure–activity relationships of peptides. Unfortunately, as it stands, the paper is too preliminary for Nature BME, as delineated below.

Major issues:
- Regarding the protein representations, the authors used the embeddings extracted from the ESM model and performed a feature selection to reduce the feature dimensionality. Since the features extracted by neural networks without disentanglement are usually known as distributed representations (i.e., a single concept is expressed by multiple units in a vectorized representation), discrete feature selection may not be the best way to reduce the dimensionality. The authors should investigate other continuous dimensionality reduction approaches (e.g., PCA or other supervised alternatives) on ESM embeddings.

We agree, this could be a problem, so we compared the performance of two feature selection methods—correlation and PCA—on ESM embeddings, using the same bacterially selective PG-1 variants as in our original model evaluation. The metrics analyzed include accuracy, F1 score, precision, and recall. We find that the impact of the feature selection method differs widely for the three different quantities we predicted with our models. For predicting activity, the PCA-based method surpasses the correlation-based method. For predicting hemolysis, both methods exhibit comparable results. Finally, for predicting selectivity, the PCA-based method shows reduced performance compared to the correlation-based method. Together this suggests that our continuous dimensionality reduction approach was appropriate and similarly successful to other approaches. **A full description and implementation of the PCA-based method was added to the Github as a version 2 of the manuscript.**

Here is the graph:



- NPN assays should be performed to assess whether the peptide variants permeabilize the outer membrane of Gram-negative bacteria such as E. coli as this will shed further insight into the mechanism of action of the peptides generated.

Propidium iodide (PI) uptake assays assess both outer and inner membrane disruption simultaneously (Fig. 4A). This is because PI is membrane impermeable and can only fluoresce when bound to DNA forcing both membranes to be disrupted to have a positive fluorescence signal. This is mentioned in the text (page 5 line 21)

- Cytotoxicity assays using human cell lines should be performed as this is standard in the field.

We have added kidney cell line (HEK293) SYTOX green uptake upon treatment with PG-1.0 and the most selective dmSLAY and machine learning variants (PG1.37 and bsPG-1.2) as Figure 6B. SYTOX green is membrane impermeable and fluoresces upon DNA binding so it is a measurement of membrane integrity. We also assayed Colistin, a clinically used AMP, for comparison. Selective variants are fare less toxic to kidney cells than PG-1.0 and machine learning variants were comparable or less membrane disruptive than Colistin overall. **Text can be found on page paragraph 2.**

- In vivo validation using a mouse model is needed to ensure the peptides generated target bacterial pathogens while not causing toxicity in a realistic and relevant setting. This would provide a robust validation of dmSLAY and is expected for a journal like Nature BME.

We understand this concern but feel that this is outside the scope of this particular manuscript. We want to keep methods and data focused on helping to understand rules governing AMP membrane selectivity rather than direct therapeutic function. This is because most AMPs require extensive synthetic modification, like cyclization and use of non-canonical amino acids, in order to successfully treat infection in animal models. These modifications are not compatible with dmSLAY and therefore we feel testing in animal models is best reserved for additional studies using the methods and principles outlined here to decrease toxicity of therapeutically relevant AMPs. **We have added some conversation of this important point to the manuscript's discussion section.**

Rebuttal 2

Response to Reviewers #2:

We again thank the reviewers for their time and comments. We have added *in vivo* mouse maximum tolerated dose and IP/IP infection model data to help alleviate any remaining reviewer concerns. We hope the manuscript is now deemed adequate for publication in Nature Biomedical Engineering. For a more detailed response to specific comments please see below.

Reviewer #1 (Report for the authors (Required)):

The authors revised the manuscripts well and addressed most question I have. And I think it is can be accepted by Nature BME.

Reviewer #2 (Report for the authors (Required)):

The authors have improved the manuscript by effectively addressing most reviewer critiques. One primary concern was not addressed (perhaps because it was in the main paragraph and not a numbered element):

The ability of the study to assess (and improve upon) selectivity may be the larger impact. However, the selectivity analysis does not emerge from dmSLAY but rather from ‘classic’ experiments with individual peptides. As a result, the title and other writing is somewhat misleading in the driver of the advance.

In addition, it remains true that dmSLAY does not provide a continuous quantitative evaluation of molecular potency, which is a technological limitation. Nevertheless, the rigor and clarity in regard to the binary approach have been dramatically improved.

Overall, the manuscript is now in a form that provides substantial value to the scientific community. Ideally, brief editing to address the primary concern above would be ideal, although this reviewer does not advocate for another round of formal review to address this matter.

We understand the reviewer's concern but disagree with the idea that dmSLAY did not aid in the assessment of peptide selectivity. Nearly all selective PG-1 variants identified, and later assessed using 'classic' experiments, contained multiple simultaneous amino acid changes and would have never been identified to begin with without the use of high-throughput dmSLAY analysis. We have edited the text to highlight this use of dmSLAY and its important role in down selecting otherwise unmanageable numbers of sequences for classical study and machine learning. We have edited some text at the end of the manuscript to help highlight dmSLAY's important role in down selecting otherwise unmanageable numbers of sequences for classical study and machine learning.

Reviewer #3 (Report for the authors (Required)):

Though I commend the authors for slightly improving the paper through their comments and revisions, I still think it is too preliminary and does not meet the minimum bar for Nature BME.

The authors mention that most AMPs require extensive synthetic modification to be effective in animal models. However, this is simply not true as has been extensively demonstrated in the literature (using mouse infection models) and in the author's own recent work (e.g., Randall JR et al. Adapting antibacterial display to identify serum-active macrocyclic peptide antibiotics. PNAS Nexus. 2023 Aug 17;2(8):pgad270. doi: 10.1093/pnasnexus/pgad270. PMID: 37637199; PMCID: PMC10449418). Testing the safety and efficacy of AMPs in mouse models is standard in the field and most studies these days show data with more than one mouse model.

I think there may be a bit of confusion. Our work (Randall et al., PNAS Nexus 2023), cited by the reviewer above, shows *in vivo* activity for a highly synthetically modified AMP (SAP-26) in a *G. mellonella* infection model. In fact, this work specifically mentions that Protegrin-1, which is also terminally modified via amidation in its naturally occurring form, does not rescue infection in this model at doses as high as 100 mg/kg.

Still, two examples do exist where naturally occurring Protegrin-1 is effective *in vivo* (see citations below). One shows activity in an IP/IP and IV/IV mouse infection model; the other in just an IP/IP model. Both studies do not use does above 5 mg/kg, one specifically citing toxicity concerns.

Seeing this data, and with the reviewers comment in mind, **we ran maximum tolerated dose (MTD) assays with Protegrin-1 and an amidated and unamidated version our least toxic variant, bsPG-1.2 (see updated Figure 6CD). Both bsPG-1.2 variants showed an MTD greater than five-fold higher than Protegrin-1 (>25 mg/kg versus 5 mg/kg). bsPG-1.2 was also shown to drastically reduce bacterial organ infection in an IP/IP mouse model.** We hope the inclusion of this additional data satisfies any remaining reviewer concerns.

Bolosov, I. A., Panteleev, P. V., Sychev, S. V., Khokhlova, V. A., Safronova, V. N., Toropygin, I. Y., Kombarova, T. I., Korobova, O. V., Pereskokova, E. S., Borzilov, A. I., Ovchinnikova, T. V., & Balandin, S. V. (2023). Design of Protegrin-1 Analogs with Improved Antibacterial Selectivity. *Pharmaceutics*, *15*(8). https://doi.org/10.3390/PHARMACEUTICS15082047/S1

Steinberg, D. A., Hurst, M. A., Fujii, C. A., Kung, A. H. C., Ho, J. F., Cheng, F. C., Loury, D. J., & Fiddes, J. C. (1997). Protegrin-1: a broad-spectrum, rapidly microbicidal peptide with in vivo activity. *Antimicrobial Agents and Chemotherapy*, *41*(8), 1738–1742. https://doi.org/10.1128/AAC.41.8.1738