Pharmacophore-Inspired Virtual Receptor: Generating Realistic Binding Affinity Data for Machine Learning

Kilian Conde-Frieboes[1] **Carsten Stahlhut**[1]
[1]AI & Digital Innovation, Novo Nordisk A/S
{kcf,ctqs}@novonordisk.com

Abstract

VReceptor is a virtual receptor simulation framework that generates realistic, sequence-dependent binding affinity data for peptide-based therapeutics. It combines pharmacophore-inspired weighting schemes with amino acid similarity metrics to evaluate peptide design strategies in a controlled environment. Validation using the Prolactin-Releasing Peptide (PrRP) dataset confirms VReceptor's ability to approximate binding behaviors and optimize experimental planning. This framework empowers researchers in computational peptide design and active learning workflows while highlighting areas for further refinement.

1 Introduction

Peptide-based therapeutics have demonstrated significant potential over the past few decades for the treatment of various diseases with both specificity and biological safety [1]. The development of *in silico* tools for drug discovery has revolutionised early-stage pharmaceutical research, enabling rapid hypothesis testing and virtual screening of vast molecular libraries. However, a persistent challenge remains: how to validate these computational models effectively, especially when experimental resources are limited or prohibitively expensive. This is particularly true for peptide-based therapeutics, where the combinatorial explosion of possible sequences makes exhaustive experimental testing infeasible.

To address this, we introduce VReceptor, a virtual receptor simulation framework designed to generate realistic, sequence-dependent binding affinity data. Inspired by the pharmacophore concept originally proposed by Paul Ehrlich [2, 3], VReceptor provides a controlled and interpretable environment for evaluating peptide design strategies and active learning algorithms. It is particularly suited for benchmarking machine learning models in peptide optimisation tasks, where the goal is to predict or improve binding affinity based on sequence information.

Unlike purely random or black-box data generators, VReceptor is built on a transparent and biologically motivated model. It simulates ligand–receptor interactions by assigning importance to specific amino acid positions using a pharmacophore-like weighting scheme. These weights are then combined with amino acid distance metrics to produce a binding score that reflects both sequence composition and structural relevance. The result is a synthetic yet biologically plausible dataset that can be used to test hypotheses, compare design strategies, and train predictive models.

1.1 Related work

To address the challenge of molecular optimization, recent research has introduced the Poli library, which provides a suite of artificial fitness functions, or oracles, designed to serve as reliable proxies for molecular traits of interest [4, 5]. While this library facilitates systematic benchmarking of

European Conference on Neural Information Processing Systems (EurIPS) Workshop: SimBioChem 2025.

molecular optimization methods specifically for protein optimization, further efforts are needed to broaden the range of modalities to include peptides and to validate the consistency and reliability of these oracles across various relevant wet-lab characterizations.

Although numerous bioinformatics tools such as PeptideRanker [6] exist for targeting the modulation of bioactive peptides, [7] highlights a general deficiency in their effectiveness. A significant challenge is the weak correlation between predictions made by these tools and experimentally derived IC50 values. This discrepancy indicates that many models may not adequately capture the complex biological interactions governing peptide activity, making them less reliable for practical applications. This underscores the urgent need for more accurate computational methods to enhance the predictive validity of bioinformatics tools, facilitating their integration into peptide design and leading to more effective bioactive peptides. Similarly, recent benchmarking work, as exemplified in [8], provides benchmarking framework tailored for protein folding neural networks in protein-peptide complex prediction. While AlphaFold 3 [9] accordingly to [8] demonstrates strong performance in structure prediction, they note that the confidence metrics correlate poorly with experimental binding affinities.

In this context, VReceptor offers a complementary approach: a transparent and controllable virtual environment for generating realistic synthetic data to accelerate method development. Although we validate the realism of VReceptor by fitting it to experimentally measured affinities from the PrRP peptide family [10], this serves only as a demonstration of plausibility rather than the primary contribution of the work. This work makes the following contributions:

- **A realistic virtual receptor framework (VReceptor):** A transparent, pharmacophore-inspired simulation model that generates biologically plausible, sequence-dependent binding affinity data for peptides.
- **A synthetic, controllable environment for benchmarking and accelerating peptide-model development:** VReceptor enables systematic testing of peptide optimisation strategies, search algorithms, active learning methods, and machine learning models under realistic and interpretable conditions—supporting rapid iteration before experimental investment.
- **A biologically interpretable receptor model:** The framework combines position-specific weighting with amino acid distance metrics, offering mechanistic transparency absent in black-box oracle generators.
- **Empirical demonstration of realism:** We show that VReceptor can be calibrated to approximate real binding landscapes using data from the PrRP peptide family—serving as a validation step, not the main contribution.

2 Peptide-Receptor Binding Model

The foundation of the VReceptor framework is a generative model that simulates peptide–receptor binding affinities based on peptide sequence composition. The core idea is to construct a function that maps a peptide sequence to a predicted binding strength, mimicking the behaviour of a biological receptor. This function is designed to be deterministic and sequence-dependent, ensuring that similar sequences yield similar binding profiles, thereby enabling meaningful comparisons and learning.

The binding affinity is modelled as a function of the peptide sequence (eq. 1):

$$IC_{50} = f(\text{sequence}) \tag{1}$$

To achieve this, the model begins with a reference peptide of length n, where each position $i \in \{1, ..., n\}$ is assigned an importance score b_i. These scores reflect the contribution of each position to the overall binding interaction and are derived from a hypothetical pharmacophore model. This model is expressed as a mixture of Gaussian functions g_k (eq. 2, Fig. 1).

$$b_i = \sum_k a_k g_k(i, \mu_k, \sigma_k) \tag{2}$$

Here, each Gaussian g_k is parameterised by an amplitude a_k, a mean μ_k, and a width σ_k, allowing the model to flexibly encode position-specific relevance across the peptide. The selection of gaussian functions to model the importance, or weights, of different amino acid stretches in the peptides allows

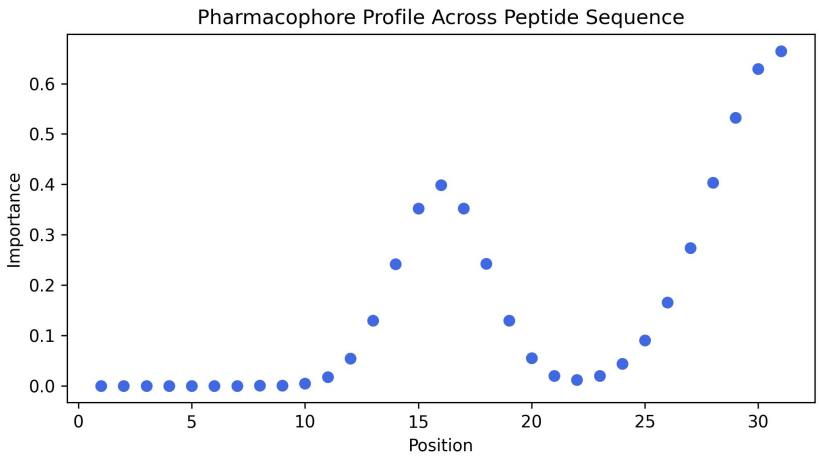

Figure 1: Hypothetical pharmacophore model for a peptide with 31 amino acids.

a smaller number of parameter as compared to individual weights, but also reflects ligand-receptor interactions, where small areas in the peptide interact with the receptor, but nearby amino acids have a diminishing effect.

In the VReceptor model, each amino acid is represented by a feature vector $\mathbf{v}_i^j \in \mathbb{R}^d$, where d denotes the number of physicochemical descriptors used—herein derived from z-scales as described in [11], i indicates the position in the peptide and j indicates the amino acid selected from the amino acid alphabet. These descriptors capture properties such as hydrophobicity, steric bulk, and electronic characteristics.

To calculate the score of a target peptide relative to the reference peptide, the model needs a measure of distance between the amino acids. The distance function $\text{dist}(\mathbf{v}_i^{ref}, \mathbf{v}_i^j)$ quantifies the dissimilarity between the amino acid at position i in the reference peptide and the amino acid in the same position of the target peptide based on their physicochemical properties (e.g., z-scale descriptors), and is elaborated in Section 2.1.

Next, a weight matrix $\mathbf{W} \in \mathbb{R}^{n \times m}$ is constructed, where m is the number of amino acids in the alphabet (typically 20). Each element W_{ij} represents the weighted contribution of substituting the amino acid at position i with amino acid j, and is computed as (eq. 3):

$$W_{ij} = \text{dist}(\mathbf{v}_i^{ref}, \mathbf{v}_i^j) * b_i \tag{3}$$

In the VReceptor model, each peptide sequence is represented using a one-hot encoding matrix $\mathbf{E}$, which captures the identity of amino acids at each position in the sequence. For a peptide of length n and an amino acid alphabet of size m (typically the 20 standard amino acids), the matrix $\mathbf{E} \in \{0, 1\}^{n \times m}$ is defined such that:

$$E_{ij} = \begin{cases} 1 & \text{if the amino acid at position } i \text{ is the } j\text{-th amino acid in the alphabet} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Each row of $\mathbf{E}$ corresponds to a position in the peptide, and each column corresponds to a specific amino acid. Only one entry per row is set to 1, indicating the presence of a particular amino acid at that position, while all other entries are 0.

This encoding allows the model to compute the *Hadamard product* $\mathbf{R} \in \mathbb{R}^{n \times m}$ (eq. 5), where $\mathbf{W}$ is the weight matrix derived from the pharmacophore model and distance scores. The resulting matrix $\mathbf{R}$ reflects the contribution of each amino acid to the overall binding affinity of the target peptide, which is then summed to produce a scalar prediction score.

$$\mathbf{R} = \mathbf{W} \circ \mathbf{E} \tag{5}$$

To simulate experimental variability, homoscedastic noise is added to the computed binding affinity. The final expression for the relative binding affinity is (eq. 6):

$$\log(\mathrm{IC}_{50,\mathrm{rel}}) = \sum_{i=1}^{n} \sum_{j=1}^{m} R_{ij} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{6}$$

Here, R_{ij} denotes the weighted contribution of amino acid j at position i, and ε is a normally distributed noise term with zero mean and variance σ^2, consistent with the homoscedasticity assumption commonly used in linear mixed models for assay variability estimation [12].

Together, these components form the basis of the VReceptor simulation engine, enabling the generation of realistic, interpretable binding data for downstream analysis and model training. The model does not account for interactions between amino acids within the peptide nor any structural or conformational considerations; it focuses solely on position-specific importance scores.

2.1 Distance Measures

Three different distance measures have been integrated into the VReceptor, with the possibility of further extensions. Similarly, extension of the VReceptor can be done with respect to the feature vector $\mathbf{v}$ by replacing z-scale derived physicochemical descriptors with alternative representative descriptors.

2.1.1 Cosine Distance

To assess the distance between amino acids based on their vector representations (e.g. z5 descriptors), we define a pairwise distance measure using the cosine of the angle between their feature vectors. Let us denote the vector of descriptors of amino acids g as $\mathbf{v}_g$ and the corresponding vector of amino acid h as $\mathbf{v}_h$. The cosine distance is then defined as (eq. 7):

$$\mathrm{dist}_{cosine}(\mathbf{v}_g, \mathbf{v}_h) = 1 - \frac{\mathbf{v}_g^T \mathbf{v}_h}{\|\mathbf{v}_g\| \|\mathbf{v}_h\|} \tag{7}$$

This formulation yields values in the interval $[0, 2]$, where higher values indicate greater distance. It captures the angular proximity between amino acid descriptors and is particularly useful when the direction of the feature vector is more informative than its magnitude.

2.1.2 Chebyshev-like Distance

To quantify distance based on the maximum absolute difference between features, we define a Chebyshev-like distance. Let $\mathbf{v}_g$ be the descriptor vector of amino acid g and $\mathbf{v}_h$ be the feature vector of amino acid h. The raw Chebyshev distance is the following with d being the dimension of the z-scale vectors $\mathbf{v}$ (eq. 8):

$$\mathrm{dist}_{\text{Chebyshev}}(\mathbf{v}_g, \mathbf{v}_h) = \max_{\alpha \in \{1,\ldots,d\}} |\mathbf{v}_{g\alpha} - \mathbf{v}_{h\alpha}| \tag{8}$$

To normalize this distance across the amino acid set, we divide by the maximum observed difference for each feature (eq. 9).

$$\mathrm{dist}_{\text{Clike}}(\mathbf{v}_g, \mathbf{v}_h) = -\frac{\max_{\alpha \in \{1,\ldots,d\}} |\mathbf{v}_{g\alpha} - \mathbf{v}_{h\alpha}|}{\max\limits_{\mathrm{AA}_m, \mathrm{AA}_n, \alpha^*} |\mathbf{v}_{\mathrm{AA}_m,\alpha^*} - \mathbf{v}_{\mathrm{AA}_n,\alpha^*}|} \tag{9}$$

This normalized form ensures comparability across features and amino acid pairs, making it suitable for models where dominant feature differences are most relevant.

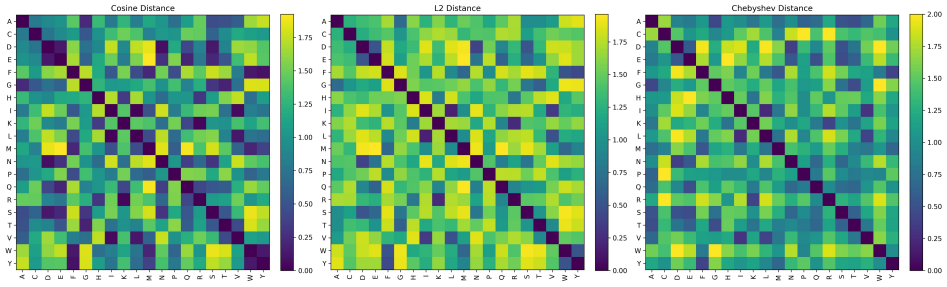

Figure 2: Visual comparison between the 3 discussed distance measures.

2.1.3 L2 Distance

To quantify the distance between amino acids based on their physicochemical properties, we define a pairwise distance measure derived from the L2 norm of normalized feature vectors. Let $\mathbf{v}_g$ denote the vector of properties of amino acid g, in this case the z-scale vector. Each vector is normalized to the unit length (eq. 10):

$$\hat{\mathbf{v}}_g = \frac{\mathbf{v}_g}{\|\mathbf{v}_g\|_2} \tag{10}$$

The distance between amino acids g and h is then defined as (eq. 11):

$$\text{dist}_{L2}(\mathbf{v}_g, \mathbf{v}_h) = \|\hat{\mathbf{v}}_g - \hat{\mathbf{v}}_h\|_2 * 2 \tag{11}$$

This formulation ensures that the distance values are in the interval $[0, 2]$, with higher values indicating greater distance. The approach captures relative proximity in the normalized feature space and is particularly suited for comparing amino acids in embedded or kernel-based models. Figure 2 shows a comparison of the three different distance measures.

3 Experiments

The intended purpose of this tool was to serve as a data generator for evaluating peptide library design strategies, tasks that would otherwise be prohibitively expensive and time-consuming to evaluate experimentally. A simple question is, for example, whether a full mutational scan of a given target peptide should be conducted or whether a random library of peptides with multiple mutations would yield a better model. All experiments were conducted on compute architecture featuring an x86_64 configuration with a single Intel Xeon Platinum 8360H CPU @ 3.00 GHz. In the supplied notebook (see GitHub) we address this problem for a peptide of length 31. We chose the Prolactin-Releasing Peptide (PrRP, [10]) sequence as a starting peptide and created an artificial pharmacophore with the use of two Gaussian functions as shown in figure 1. A complete mutational scan with the 20 canonical amino acids yields 589 peptides ($31 * 19$). We compared this design with a random design of equal size. Sequences from the 30 million possible triple mutations for this target peptide were selected. After generating simulated data with the primed VReceptor, two similar models are trained on these datasets and evaluated on a random test set of 200 peptides, again selected from the set of all triple mutated peptides (figure 3).

The models consist of a 2-layer, bidirectional LSTM [13] neural network as the regression model. The hidden state of the LSTM layer is fed into the output layer. After training the models are evaluated on the test data set. In this case the model trained on the random data set has a smaller mean-squared error as compared to the systematic generated scan data set (0.023 vs 0.140) (fig. 4).

Figure 4 shows the mean squared error (mse) comparing the two different models.

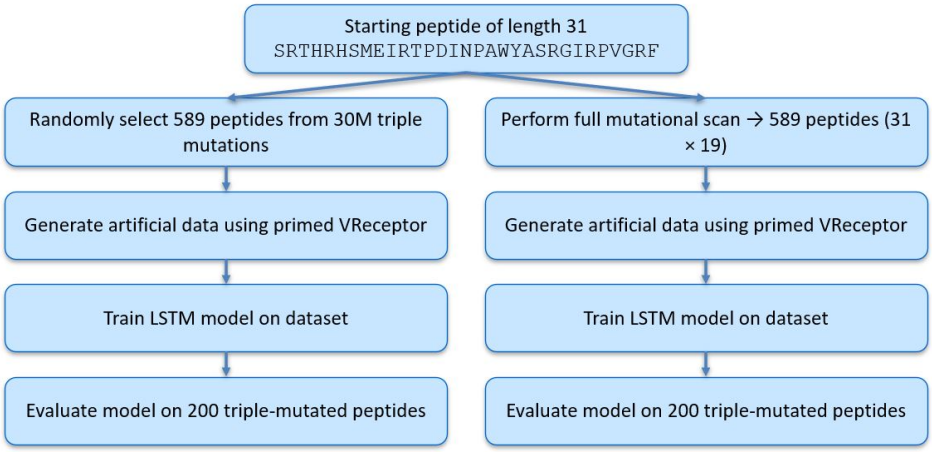

Figure 3: Flow chart to compare two different design strategies using the VReceptor.

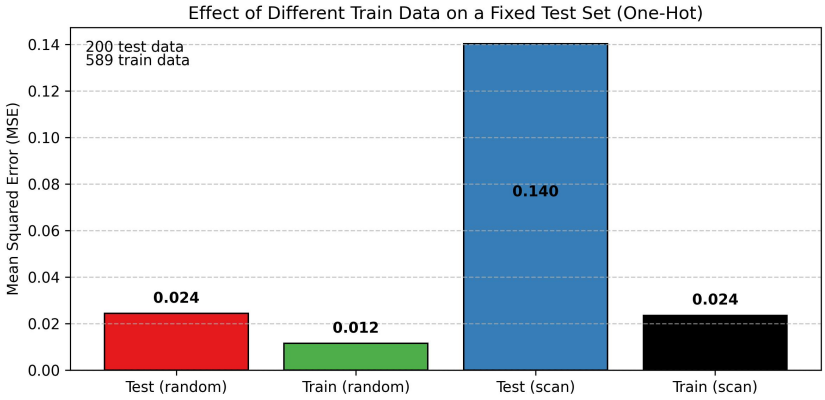

Figure 4: Train/Test error for the scan and random design strategy

3.1 Realism of VReceptor

During the development and testing of VReceptor, a central question emerged: How realistic is this model? Can peptide–receptor interactions truly be captured with such simplicity? VReceptor is intentionally designed with a limited number of parameters, determined by the number of Gaussian functions N_G specified during initialization. The only additional degree of freedom lies in the choice of distance measure $\mathrm{dist}(g, h)$, which governs how amino acid substitutions are evaluated. To assess the model's authenticity and practical value, we explored fitting these minimal parameters to real experimental data. This approach allows us to evaluate whether VReceptor can meaningfully replicate observed binding behaviours, thereby validating its usefulness as a simulation tool.

3.1.1 The PrRP dataset

Prolactin-Releasing Peptide (PrRP) is a neuropeptide involved in the regulation of food intake and energy homeostasis [10]. It exerts its physiological effects by binding to at least three G-protein coupled receptors: GPR10, NPFF receptor 1 (NPFFR1), and NPFF receptor 2 (NPFFR2). The dataset used to validate the VReceptor tool comprises experimentally measured binding affinities ($\log IC_{50}$ values) of PrRP-derived peptides against these three receptors. The listed data are relative to wildtype human PrRP. All peptides in the dataset are 31 amino acids long and C-terminally amidated, reflecting their biologically active form. A subset of peptides includes fatty acid modifications, denoted by an 'X' in the sequence; however, these were excluded from the present analysis to maintain consistency

in molecular structure and avoid confounding effects. The estimated standard deviation is 0.1 on the $\log_{10}$ scale.

3.1.2 Parameter Fitting Procedure for VReceptor Models

To evaluate the performance of the VReceptor model under different amino acid similarity metrics, we implemented a parameter optimisation workflow using three distinct similarity functions: cosine similarity, Chebyshev-like distance, and L2 norm-based similarity. For each similarity function, we trained a VReceptor model to predict the binding affinity ($\log IC_{50}$) of peptide sequences using a minimal set of parameters. The optimisation objective was to minimise the mean squared error (MSE) between predicted and experimentally measured $\log IC_{50}$ values, represented as $\hat{y}_n$ and y_n, respectively, as detailed in (eq. 12):

$$min \frac{1}{N} \sum_{n}^{N} \|\hat{y}_n - y_n\|^2 \tag{12}$$

We selected a subset of the PrRP dataset where the target response variable (relGPR10) was below a threshold of 2.0. The data was split into training and test sets using a 50:50 ratio with a fixed random seed for reproducibility. The model parameters were optimised using the scipy.optimize.minimize function. The parameter vector included the positions and amplitudes of two Gaussian functions used to define the pharmacophore model. A grid of initial parameter guesses was explored to avoid local minima, iterating over combinations of Gaussian centres and amplitudes.

For each candidate parameter set, a VReceptor instance was constructed and used to simulate binding affinities for the training sequences. The predicted values were compared to experimental data using MSE as the loss function. The best-performing parameter set was retained for each similarity function. The optimised models were then evaluated on both training and test sets. Predicted $\log IC_{50}$ values were compared to experimental values using Pearson and Spearman correlation coefficients, and mean squared error (MSE). The results were stored for each similarity function and visualised to compare model performance across metrics.

When comparing the VReceptor model metrics with a Ridge linear regression with the same data split and an empirical determined alpha = 10, the VReceptor performs similar good (figure 5, table 1). The larger difference between the VReceptor and Ridge regression on the train data is most likely a reflection of the vast difference in parameter, 6 vs 155, respectively, and a train set size of 250 compounds.

Method	Pearson r	Spearman r	MSE
Training Set			
Cosine	0.82	0.80	0.131
Chebyshev	0.85	0.83	0.109
L2_sim	0.85	0.83	0.108
Ridge	**0.95**	**0.92**	**0.036**
Test Set			
Cosine	0.77	0.75	0.155
Chebyshev	**0.83**	**0.80**	**0.118**
L2_sim	**0.83**	**0.80**	**0.118**
Ridge	0.80	0.73	0.140

Table 1: Performance metrics for different methods on training and test sets.

The fitted VRecptors can predict the test data with decent metrics. The model using the cosine distance is slightly worse than the other two, but still acceptable.

4 Discussion

The VReceptor framework presents a streamlined yet biologically grounded approach to simulating peptide–receptor interactions. By combining a pharmacophore-inspired weighting scheme with interpretable amino acid similarity metrics, it enables the generation of synthetic binding data that

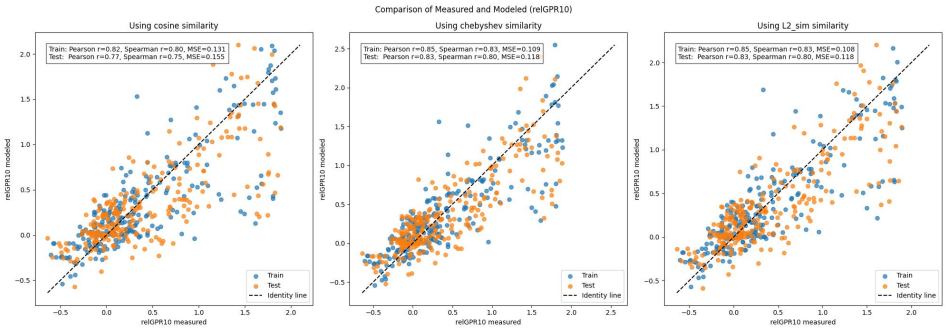

Figure 5: Performance of VReceptor models using three different similarity measures

reflects meaningful sequence dependencies. This makes it particularly valuable for benchmarking peptide design strategies and evaluating active learning workflows.

A key strength of VReceptor is its balance between simplicity and flexibility. The model is parameterised by a small number of Gaussian functions, which define the positional importance of amino acids in a peptide sequence. This allows for intuitive control over the simulated binding landscape while maintaining computational efficiency. The integration of multiple distance measures—cosine, Chebyshev-like, and L2—further enhances the model's adaptability, enabling users to explore how different assumptions about amino acid similarity influence predictive performance.

The application to the PrRP dataset demonstrates that VReceptor can approximate experimental binding affinities with reasonable accuracy, even when using a minimal parameter set. This suggests that the model can serve as a practical surrogate for experimental assays in early-stage research, where rapid iteration and hypothesis testing are critical. The comparison of design strategies—full mutational scans versus random triple-mutant libraries—also highlights the tool's utility in guiding experimental planning.

However, the model's simplicity also imposes limitations. It assumes additive contributions from individual amino acid positions and does not account for higher-order interactions or structural constraints. While this is a deliberate design choice to preserve interpretability, it may limit the model's applicability in more complex biological contexts. This can be seen when fitting a model on the other two provided receptor data (NPFFR1 and NPFFR2). These fittings have higher mean-squared error values and lower Spearman and Pearson correlations. Two possible explanations are the collected data are not diverse enough to provide a good data set and/or there are other effects like e.g. the influence of the helicity of the peptides on binding affinities which are not captured in this simple model. Also the fitting of the model to real data was done to evaluate the realism of VReceptor and the simple global minimum search strategy might not be thorough enough.

5 Conclusion & Future Directions

In summary, VReceptor offers a lightweight, interpretable, and extensible platform for simulating realistic peptide binding data. It is well-suited for use in computational peptide design, active learning evaluation, and methodological benchmarking. Its realism and applicability is validated through approximating the binding behavior of a real wetlab experimental dataset with enhanced generalization compared to simple ridge regression when facing limited experimental data.

Beyond its role as a synthetic data generator and benchmarking tool, VReceptor can serve as a robust baseline in predictive machine learning workflows. Its transparent, biologically motivated design allows researchers to systematically evaluate the performance of more complex models against a well-understood reference. By incorporating additional features—such as structural priors or higher-order sequence interactions—VReceptor may achieve predictive accuracy on par with state-of-the-art machine learning approaches, while retaining interpretability and computational efficiency.

Acknowledgments and Disclosure of Funding

We gratefully acknowledge Claus Bekker Jeppesen and Lone Honoré at Novo Nordisk for their invaluable contribution in collecting the PrRP dataset. Both authors, KCF and CS, are employees of Novo Nordisk A/S.

References

[1] Marian Vincenzi, Flavia Anna Mercurio, and Marilisa Leone. Virtual screening of peptide libraries: The search for peptide-based therapeutics using computational tools. *International Journal of Molecular Sciences*, 25(3):1798, 2024.

[2] P. Ehrlich. Über die constitution des diphtheriegiftes. *Deutsche Medizinische Wochenschrift*, pages 597–600, 1898.

[3] Osman F. Güner and J. Phillip Bowen. Setting the record straight: The origin of the pharmacophore concept. *Journal of Chemical Information and Modeling*, 54(5):1269–1283, 2014. PMID: 24745881.

[4] Miguel González-Duque, Richard Michael, Simon Bartels, Yevgen Zainchkovskyy, Søren Hauberg, and Wouter Boomsma. A survey and benchmark of high-dimensional bayesian optimization of discrete sequences. *Advances in Neural Information Processing Systems*, 37:140478–140508, 2024.

[5] M. Gonzalez-Duque, S. Bartels, R. Michael, Y. Zainchkovskyy, S. Hauberg, and W. Boomsma. Poli: A library for discrete objective functions. `https://github.com/MachineLearningLifeScience/poli?tab=readme-ov-file`, 2024. Accessed: October 14, 2025.

[6] Catherine Mooney, Niall J. Haslam, Gianluca Pollastri, and Denis C. Shields. Towards the improved discovery and design of functional peptides: Common features of diverse classes permit generalized prediction of bioactivity. *PLOS ONE*, 7(10):1–12, 10 2012.

[7] Fernando Rivero-Pino, Maria C Millan-Linares, and Sergio Montserrat-De-La-Paz. Strengths and limitations of in silico tools to assess physicochemical properties, bioactivity, and bioavailability of food-derived peptides. *Trends in Food Science & Technology*, 138:433–440, 2023.

[8] Silong Zhai, Huifeng Zhao, Jike Wang, Shaolong Lin, Tiantao Liu, Shukai Gu, Dejun Jiang, Huanxiang Liu, Yu Kang, Xiaojun Yao, et al. Peppcbench is a comprehensive benchmarking framework for protein–peptide complex structure prediction. *Journal of Chemical Information and Modeling*, 65(16):8497–8513, 2025.

[9] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

[10] Veronika Pražienková, Andrea Popelová, Jaroslav Kuneš, and Lenka Maletínská. Prolactin-releasing peptide: Physiological and pharmacological properties. *International Journal of Molecular Sciences*, 20(21), 2019.

[11] Maria Sandberg, Lennart Eriksson, Jörgen Jonsson, Michael Sjöström, and Svante Wold. New chemical descriptors relevant for the design of biologically active peptides. a multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry*, 41(14):2481–2491, 1998. PMID: 9651153.

[12] Hang Li, Tomasz M. Witkos, Scott Umlauf, and Christopher Thompson. Potency assay variability estimation in practice. *Pharmaceutical Statistics*, 24(1):e2408, 2025.

[13] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 11 1997.