BMC Biology

RESEARCH

Open Access

NeuroScale: evolutional scale-based protein language models enable prediction of neuropeptides

Hongqi Zhang[1], Shanghua Liu[1], Wei Su[1], Xueqin Xie[1], Junwen Yu[1], Fuying Dao[2], Mi Yang[1*], Hao Lyu[1*] and Hao Lin[1*]

Abstract

Background Neuropeptides (NPs) are critical signaling molecules involved in various physiological and behavioral processes, including development, metabolism, and memory. They function within both the nervous and endocrine systems and have emerged as promising therapeutic targets for a range of diseases. Despite their significance, the accurate identification of NPs remains a challenge, necessitating the development of more effective computational approaches.

Results In this study, we introduce NeuroScale, a multi-channel neural network model leveraging evolutionary scale modeling (ESM) for the precise prediction of NPs. By integrating the GoogLeNet framework, NeuroScale effectively captures multi-scale NP features, enabling robust and accurate classification. Extensive benchmarking demonstrates its superior performance, consistently achieving an area under the receiver operating characteristic curve (AUC) exceeding 0.97. Additionally, we systematically analyzed the impact of protein sequence similarity thresholds and multi-scale sequence lengths on model performance, further validating NeuroScale's robustness and generalizability.

Conclusions NeuroScale represents a significant advancement in neuropeptide prediction, offering both high accuracy and adaptability to diverse sequence characteristics. Its ability to generalize across different sequence similarity thresholds and lengths underscores its potential as a reliable tool for neuropeptide discovery and peptide-based drug development. By providing a scalable and efficient deep learning framework, NeuroScale paves the way for future research in neuropeptide function, disease mechanisms, and therapeutic applications.

Keywords Neuropeptide, Deep learning, ESM, GoogLeNet, Protein large language model

*Correspondence:
Mi Yang
Yangmi_Pro@126.com
Hao Lyu
hao.lyu@uestc.edu.cn
Hao Lin
hlin@uestc.edu.cn
[1] The Clinical Hospital of Chengdu Brain Science Institute, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China
[2] School of Biological Sciences, Nanyang Technological University, Singapore 639798, Singapore

Background

Neuropeptides (NPs) represent a critical class of signaling molecules essential for a multitude of biological functions across diverse physiological and behavioral processes, including development [1, 2], differentiation [3], muscle contraction [4], and digestion [5, 6], as well as learning [7], memory [8, 9], and aging [10]. Typically comprising no more than 100 amino acids [11, 12], these peptides are cleaved from larger precursor proteins, often characterized by distinct C-terminal or N-terminal sequence motifs [13–15]. Beyond their pivotal role

in the nervous system, NPs exert significant effects via the endocrine system, modulating an extensive array of physiological functions. These functions encompass regulation of food intake [16–18], metabolism [19, 20], reproduction [21–23], fluid and cardiovascular homeostasis, energy balance, stress management, pain perception, social behaviors [24], and circadian rhythms [25]. Consequently, NPs are increasingly recognized as potential therapeutic targets for a broad spectrum of disorders, including sleep disturbances, autism, depression, heart failure, obesity, diabetes, hypertension, and epilepsy [12, 26–32].

Although traditional experimental methods, such as mass spectrometry and liquid chromatography techniques, can accurately identify novel NPs, these approaches are costly and time-consuming. To fill these gaps, various computational methods have been developed to predict NPs [33–38]. PredNeuroP utilizes a two-layer stacking strategy with traditional machine learning algorithms to achieve high-precision identification of NPs, achieving an accuracy of 0.872 [39]. By adopting feature representation learning methods combined with a two-step feature selection framework, NeuroPred-FRL achieved an AUC of 0.960 in a meta-model of random forests on an independent dataset [40]. To obtain an interpretable prediction model, NeuroPpred-Fuse employs stacking algorithms and sequence-derived features to identify NPs. Results indicate that NeuroPpred-Fuse can produce reliable predictive performance with an AUC of 0.958 [41]. Recently, deep learning has been widely applied to bioinformatics problems due to its ability to avoid feature engineering and its efficiency. Inspired by this, NeuroPred-PLM employs a multi-scale convolutional neural network and a global multi-head attention network to identify NPs, achieving a prediction accuracy of 0.922 [42].

Although previous studies have achieved commendable results, they still fail to address several issues: (i) most prior research has relied on feature engineering combined with machine learning strategies, where the high demand for domain knowledge limits the scalability of the prediction model. Even for methods based on deep learning such as protein large language models have not been systematically compared in NP identification; (ii) there is a lack of comprehensive comparison of computational efficiency between models; (iii) protein sequence similarity, as a primary factor of model performance drift, requires analysis concerning the impact of varying similarity thresholds; (iv) the model needs to be customized to accommodate multiple scales of protein sequence lengths as input.

In this study, we introduce NeuroScale, a multi-channel neural network model based on evolutionary scale modeling, for identifying NPs. By leveraging the design principles of GoogLeNet, NeuroScale achieves significant performance improvements in F1 score (F1), accuracy (ACC), and AUC. Furthermore, we systematically evaluated the impact of changes in protein sequence similarity on model performance, demonstrating the robustness of NeuroScale with a five-fold cross-validation mean AUC value consistently above 0.97. We also proved the strong scalability of NeuroScale by assessing changes in model performance when using multi-scale protein sequence lengths as inputs. We believe NeuroScale can serve as a powerful tool for psychiatric treatment and peptide drug screening.

Results and discussion

The architecture of NeuroScale

NeuroScale is an NP identification algorithm that combines a protein large language model with an advanced neural network architecture. NeuroScale consists of four parts: (i) multi-scale protein sequences are input into the language model, which are then processed by the ESM2Tokenizer module to obtain attention masks; (ii) utilizing the core component esm2_t33_650M_UR50D to extract features. This component is trained on hundreds of millions of protein sequences, comprises 33 layers, and has 650 M parameters, making it exceptionally adept at capturing the sequence-structure–function relationships of proteins. (iii) Establishing a multi-channel fully connected neural network that uses GoogLeNet, which extracts and fuses multi-dimensional information from different channels. Inspired by GoogLeNet's multi-branch architecture, NeuroScale adopts a three-channel structure in its fully connected layers. Each channel processes input features independently and outputs a 64-dimensional vector. The final feature representation is obtained by summing the outputs element-wise, simplifying computation and avoiding overfitting due to excessive complexity. This aggregation strategy enables NeuroScale to efficiently integrate multi-scale features, improving both model stability and performance when handling long and complex sequences. NeuroScale's unique design enables the understanding of deep information hidden within large-scale protein sequences, thereby characterizing the proteins' evolutionary patterns and biological representations (Fig. 1).

Comparison of feature extraction effects of protein large language models

The continuous emergence of protein large language models has provided new possibilities for the accurate identification of NPs. However, there is still a lack of systematic comparison of the effectiveness of various model designs and parameters for NP prediction.

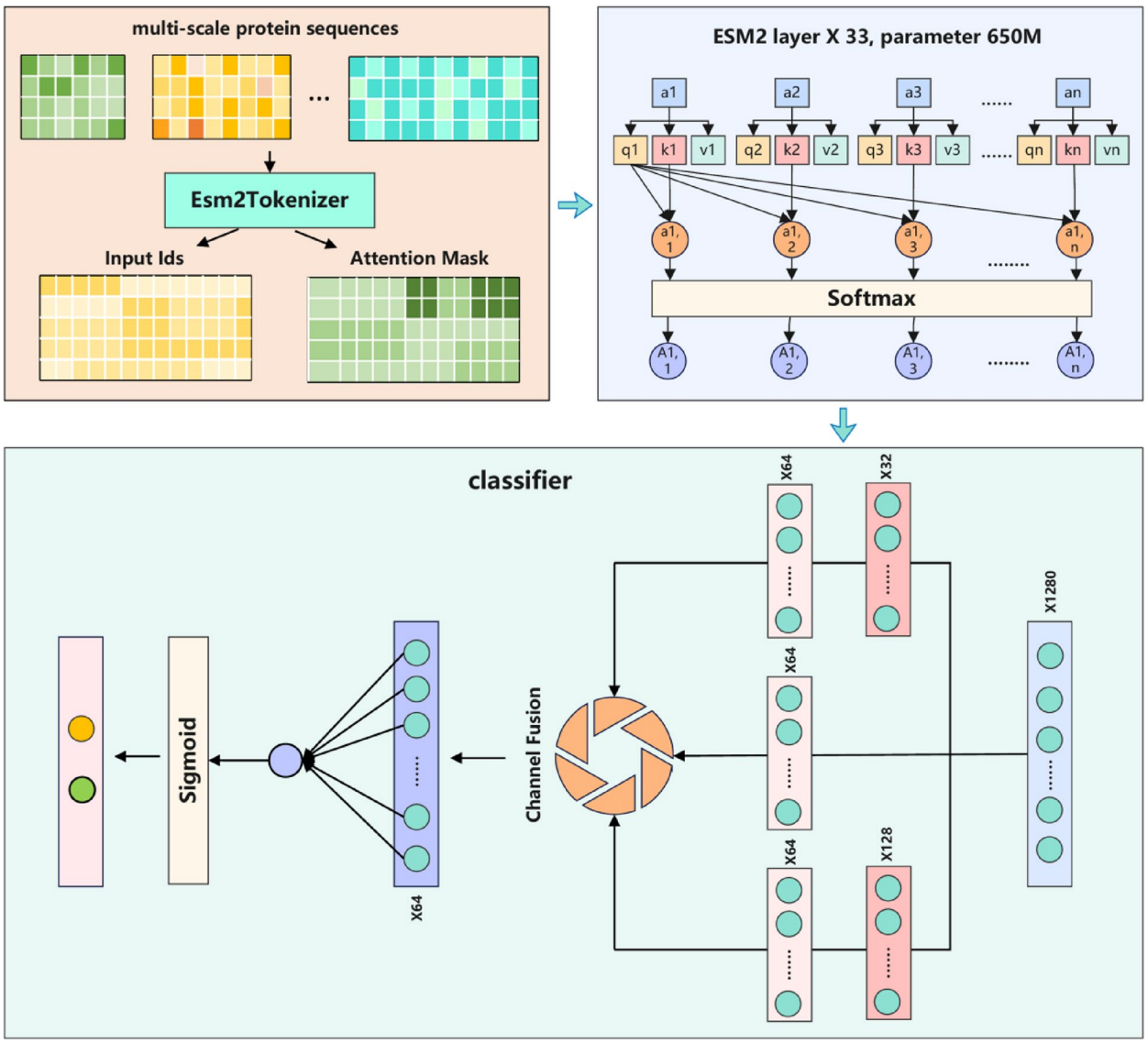

Fig. 1 Model framework of NeuroScale

To address this issue, we established and evaluated the effectiveness of several popular and efficient protein large language models as feature extraction layers in NP prediction models: different scale models of the ESM series esm2_t6_8M_UR50D (E1), esm2_t12_35M_UR50D (E2), esm2_t30_150M_UR50D (E3), esm2_t33_650M_UR50D (E4), prot_bert, DistilProtBert, RITA_s, RITA_m, and RITA_l [43–45]. As shown in Table S1 and Fig. 2A, the ESM series models performed exceptionally well, with the E4 model exhibiting outstanding performance across multiple evaluation metrics. Specifically, the E4 model achieved an ACC of 0.9513, recall (Rec) of 0.9704, precision (Pre) of 0.9349, Matthews correlation coefficient (MCC) of 0.9034, and F1 of 0.9523, confirming its efficacy in the NP prediction task. In addition, we compared the performance of two BERT-based models, prot_bert and DistilProtBert. As depicted in Fig. 2B, both yielded satisfactory performances, with DistilProtBert outperforming prot_bert. We speculate that during the distillation process, the smaller model DistilProtBert effectively learned the smooth outputs of the larger model prot_bert, which better represent inter-class relationships than hard labels, thereby aiding the smaller model in learning more robust feature representations and reducing the risk of overfitting. We also compared the performance of the RITA series models, with an approximate F1 of 0.91 indicating they are not the best practice for distinguishing NPs from non-NPs (Fig. 2C). Furthermore, intuitive

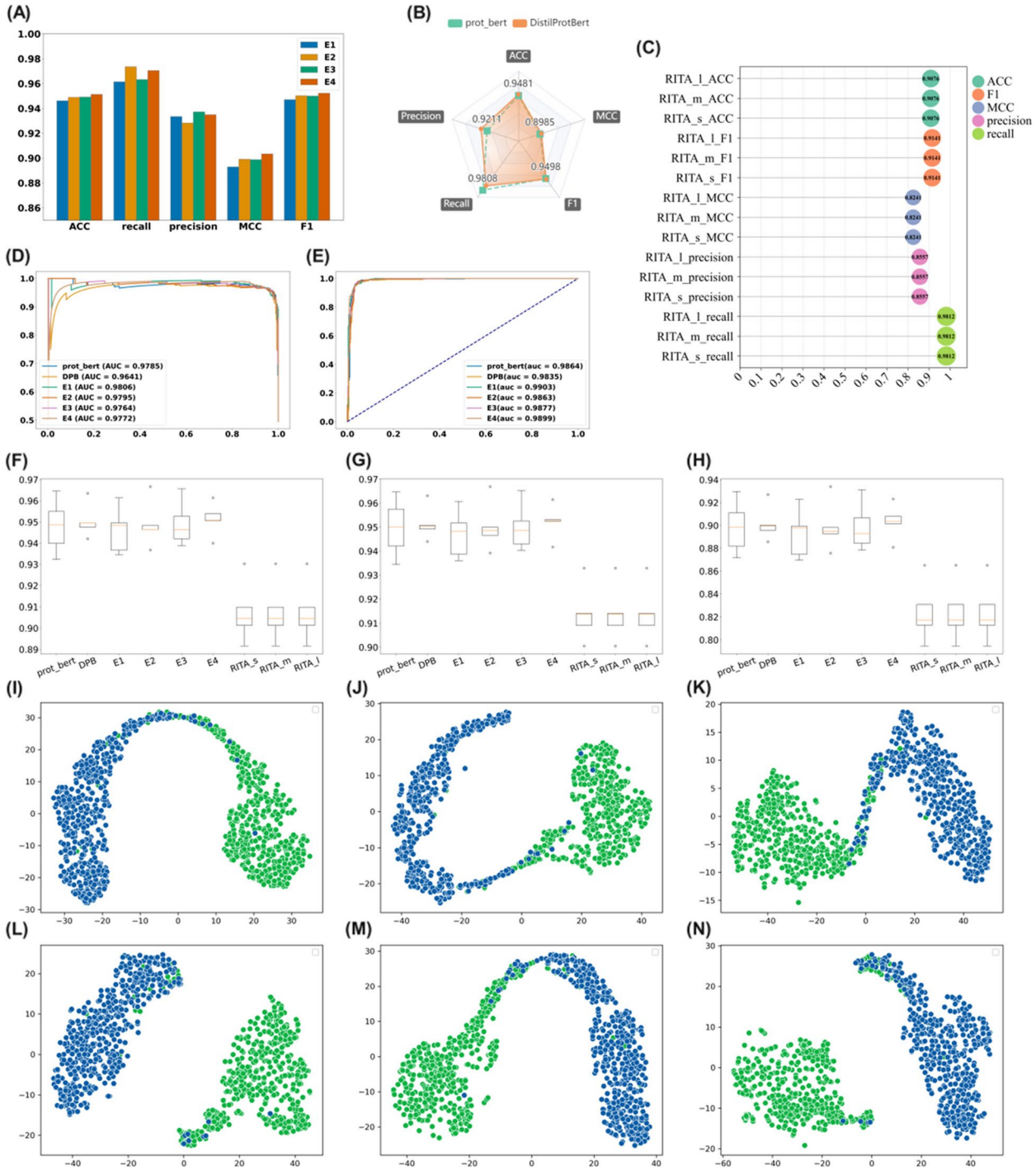

Fig. 2 Comparisons of protein large language models. **A**–**C** Comparing the average performance of ESM, BERT, and RITA series models under five-fold cross-validation, respectively. **D**–**H** Comparing PRC, ROC, ACC, F1, and MCC of E4 and other popular protein large language models under five-fold cross-validation, respectively. **I**–**N** t-SNE visualization of prot_bert, DistilProtBert, E1, E2, E3, and E4, respectively

precision-recall curve (PRC) and receiver operating characteristic curve (ROC) consistently demonstrate the reliability of the ESM series and BERT-based models (Fig. 2D–H). In order to obtain a visual representation of the features learned by the model, we used the test set of NPs and non-NPs and projected the attention layer based on the t-SNE algorithm. As shown in Fig. 2I–N, the ESM series and BERT-based models are able to generate clear

class intervals, indicating that such algorithms are capable of capturing high-order differences between NPs and non-NPs.

Comparing the running time of protein large language models

In practical scenarios, large language models are often hampered by their extensive parameter counts, leading to prolonged operation times that limit their applicability. However, existing studies frequently overlook the discussion of computational efficiency. To assess the real-world viability of the E4 model, we conducted an in-depth evaluation of its operational speed in an environment equipped with an NVIDIA GeForce RTX 3090 GPU, Intel(R) Xeon(R) Gold 6148 CPU, and 128 GB of memory. For the prot_bert, DistilProtBert, E1, E2, E3, and E4 models, we performed 500 random time evaluations. The results indicated that while the E4 model exhibited slightly longer running times compared to other models, it consistently completed analyses within 0.035 seconds, with most operations ranging between 0.032 and 0.034 seconds (Table S2, Fig. 3). This finding demonstrates that the E4 model not only delivers excellent performance but also maintains high computational efficiency. Overall, the E4 model excels in both performance and speed, providing an efficient and reliable solution for the prediction of NPs.

Independent external validation

In order to verify the generalization ability of NeuroScale, we constructed multiple datasets, including bitter peptides, natriuretic peptides A, hypocretin neuropeptide precursor, and progonadoliberin. These datasets were used for five-fold cross-validation training and testing. As shown in Fig. 4A–B, NeuroScale achieved MCC and ACC values greater than 0.98 in identifying the three types of peptides. Interestingly, NeuroScale also demonstrated superior performance in recognizing bitter peptides, yielding MCC and ACC values of 0.7687 and 0.8828, respectively. To further confirm the robustness of NeuroScale, we plotted the ROC curve and PRC curve. As expected, NeuroScale performed excellently in both neuropeptide and non-neuropeptide identification, showcasing its strong versatility (Table S3, Fig. 4C–D).

Compared to the latest methods of NP prediction

To further evaluate the prediction performance of NeuroScale, we compared it with the state-of-the-art methods NeuroPred-PLM [42], PredNeuroP [39], NeuroPred-FRL [40], and NeuroPpred-Fuse [41]. To eliminate the possible influence of test data on the results, we retrained PredNeuroP, NeuroPpred-Fuse, and NeuroPred-PLM using the same dataset and five-fold cross-validation. It should be noted that since the source code of the NeuroPred-FRL was not provided, we obtained the test results from the web server of NeuroPred-FRL. We found that the highest accuracy rate was achieved by the NeuroScale (0.9513), followed by the NeuroPred-PLM (0.8962), NeuroPpred-Fuse (0.8753), PredNeuroP (0.8665), and NeuroPred-FRL (0.8414). In addition, the scores for MCC, Rec, Pre, and F1 of the NeuroScale are approximately 13.89%, 6.74%, 5.46%, and 6.13% higher than the suboptimal NeuroPred-PLM, respectively (Table S4, Fig. 5).

Exploring the influence of protein similarity threshold on the model

The similarity threshold between protein sequences can significantly influence the efficacy of the model in classifying these sequences. Variations in this threshold alter the degree of similarity among protein sequences, potentially impacting the model's performance in classification tasks. To rigorously assess the model's generalization capabilities and robust performance across different

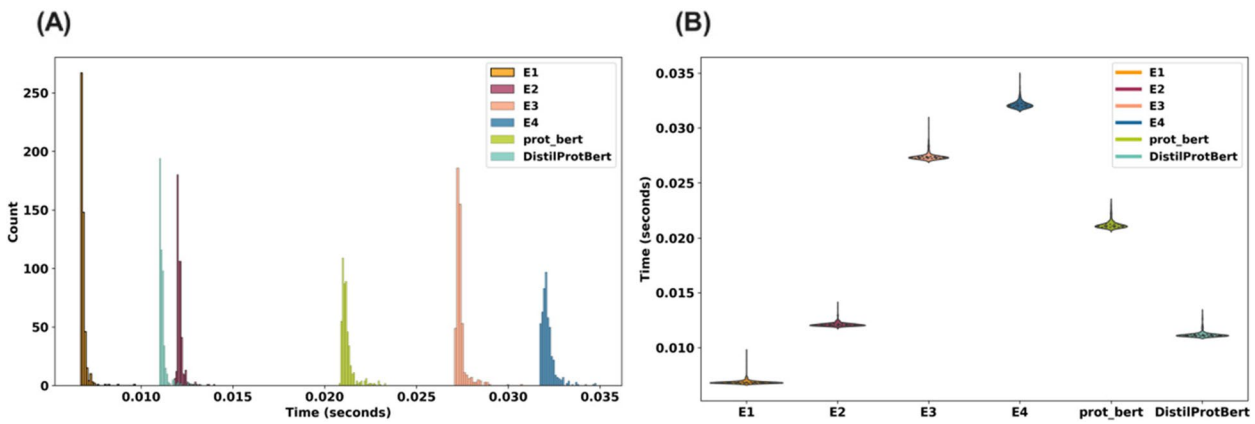

Fig. 3 The runtime statistics of protein large language models

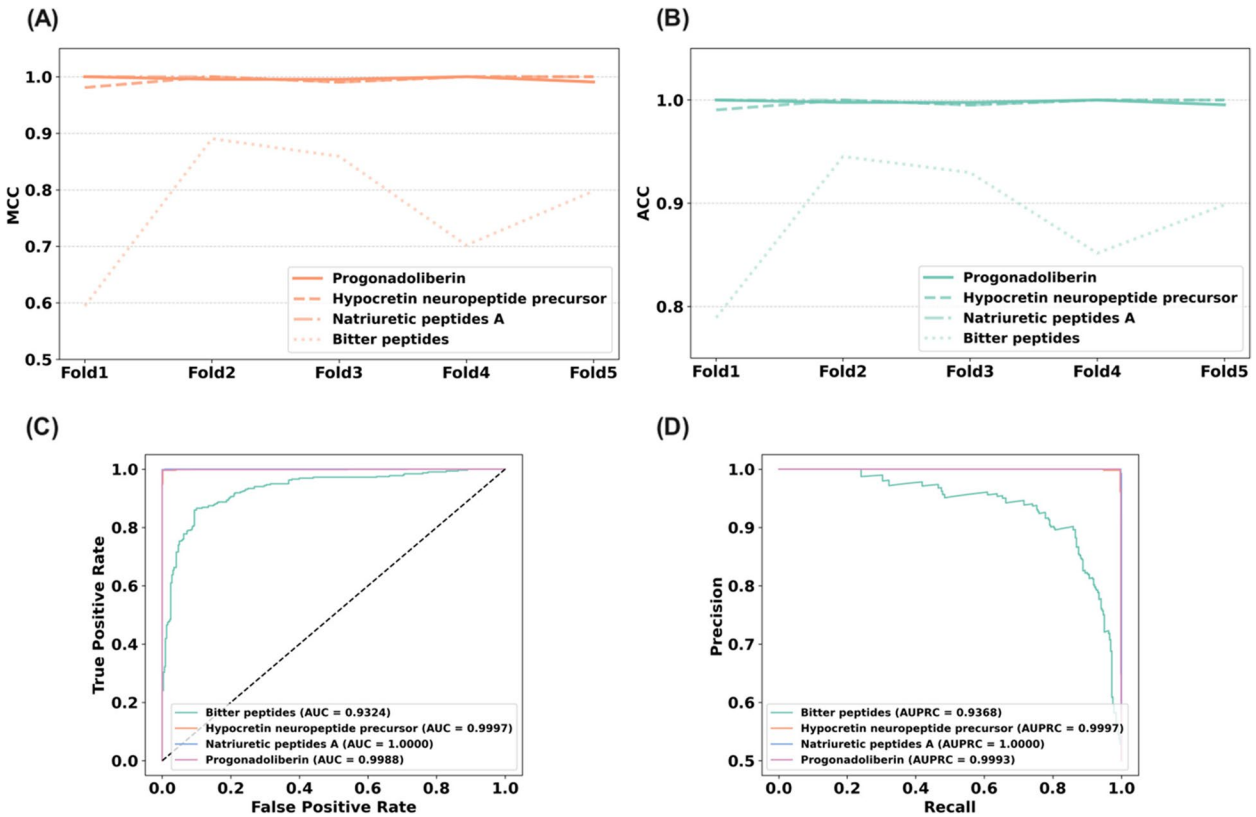

Fig. 4 Statistics of independent external validation results. **A** Five-fold MCC results for multiple species. **B** Five-fold ACC results for multiple species. **C** ROC curves for multiple species. **D** PRC curves for multiple species

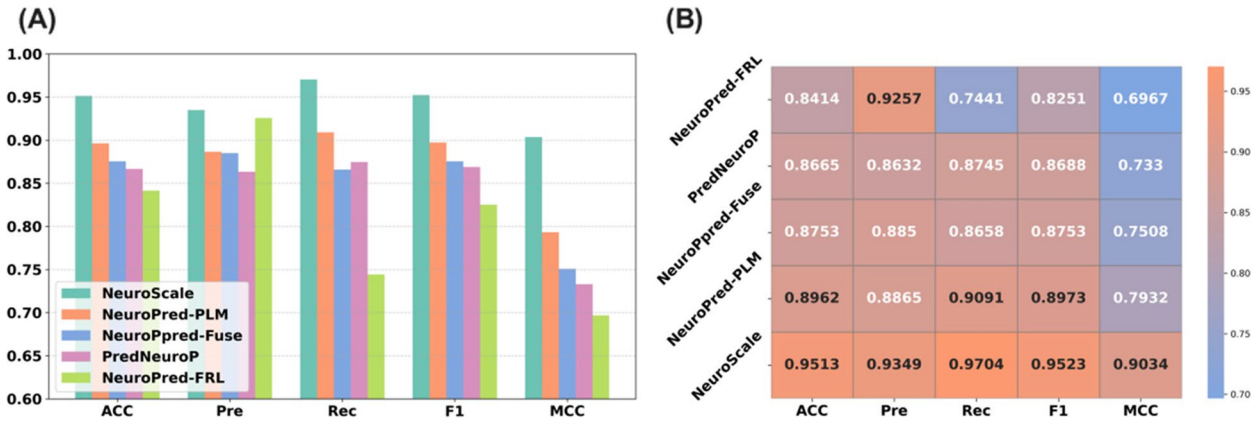

Fig. 5 Comparison results with the latest NP prediction methods

similarity thresholds, we constructed five datasets with thresholds ranging from 0.4 to 0.8. We then evaluated the model's performance using five-fold cross-validation. The experimental results showed that NeuroScale consistently performed well across different similarity thresholds. Specifically, we examined performance in terms of ACC, F1, and MCC. Although performance slightly decreased in the Fold 5 at a similarity threshold of 0.6, the overall predictive performance remained consistently excellent (Table S5, Fig. 6A–C). Moreover, NeuroScale's AUC values across different thresholds are 0.9855, 0.9826, 0.9780, 0.9785, and 0.9807, demonstrating its stability (Fig. 6D). In terms of AUPRC, we achieved prediction performance around 0.8 (Fig. 6E). Overall, NeuroScale exhibited stable

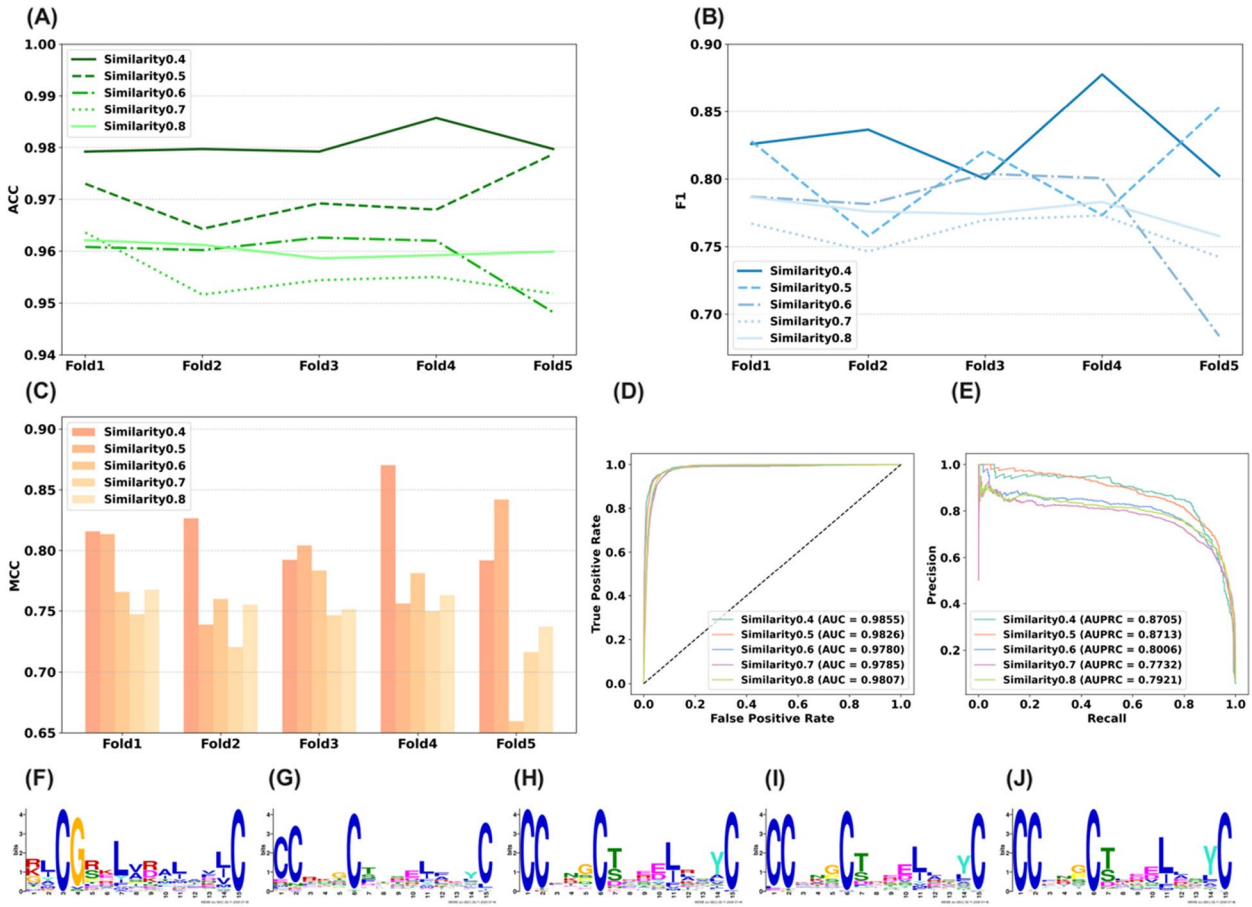

Fig. 6 Model performance results across different similarity thresholds. **A** Five-fold ACC results for different similarity thresholds. **B** Five-fold F1 results for different similarity thresholds. **C** Five-fold MCC results for different similarity thresholds. **D** ROC curves for different similarity thresholds. **E** PRC curves for different similarity thresholds. **F**–**J** MEME results for different similarity thresholds, with thresholds set at 0.4, 0.5, 0.6, 0.7, and 0.8

performance across varying similarity thresholds. To further confirm the top-level reasons behind our strong performance across different thresholds, we performed a MEME analysis [46]. The results showed that the identified potential conserved motifs at different thresholds exhibited highly similar distribution preferences, with only slight differences observed when the threshold was set to 0.4 (Fig. 6F–J). A lower similarity threshold reveals evolutionary differences between species, reflecting the functional diversity of neuropeptides, while a higher threshold focuses on conserved sequences [47–50]. NeuroScale demonstrates excellent predictive performance across different similarity thresholds, proving its ability to accurately capture both the diversity and conservancy of neuropeptides in diverse biological contexts.

Exploring the impact of protein sequence length on model performance

It is a recognized fact that as protein sequence length increases, the performance of a model may be disrupted. To verify whether NeuroScale can overcome the adverse effects associated with variations in protein sequence length, we constructed five datasets of varying scales. Five-fold cross-validation results showed that NeuroScale was able to maintain stable performance across MCC, ACC, and F1 for each fold (Table S6, Fig. 7A–C). Furthermore, evaluation metrics based on AUC and AUPRC exceeded 0.99, demonstrating NeuroScale's ability to handle protein sequences of varying lengths (Fig. 7D–E). Shorter sequences typically represent more specialized functional modules, while longer sequences may contain additional regulatory information or functional domains [51–53]. NeuroScale performed excellently on protein sequences of varying lengths, demonstrating its ability to effectively identify the diverse functional features embedded in sequences of different lengths.

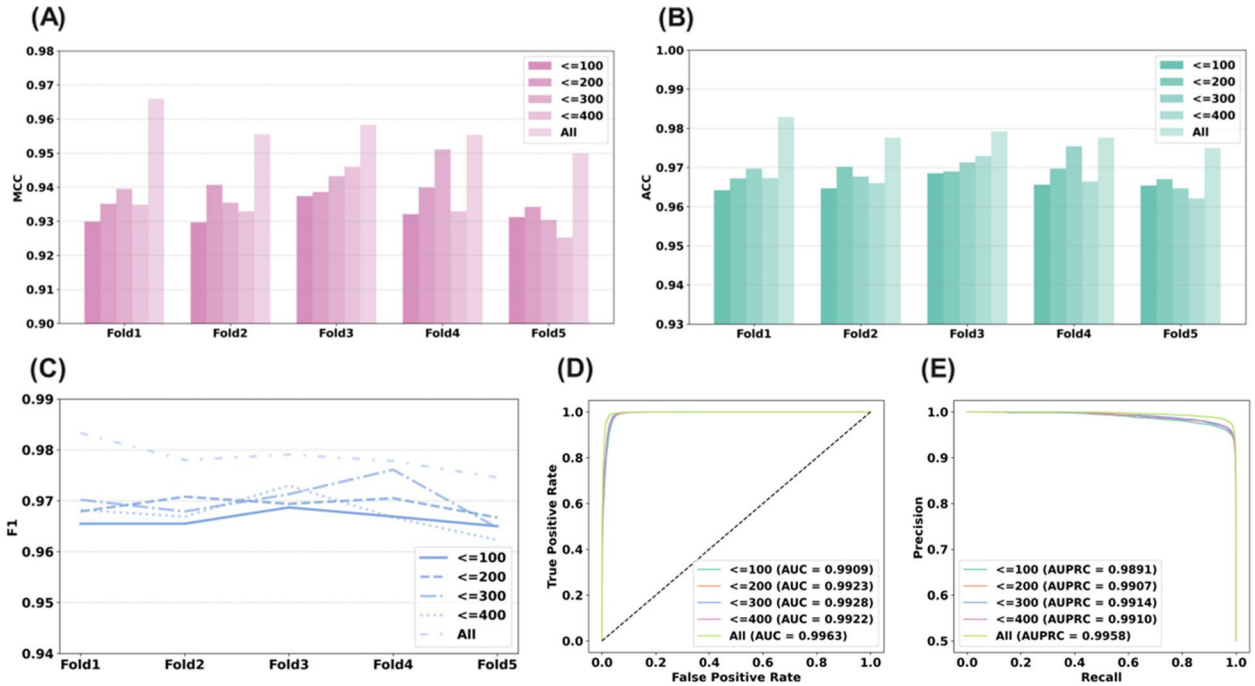

Fig. 7 Model performance results across different sequence lengths. **A** Five-fold MCC results for different sequence length ranges. **B** Five-fold ACC results for different sequence length ranges. **C** Five-fold F1 results for different sequence length ranges. **D** ROC curves for different sequence length ranges. **E** PRC curves for different sequence length ranges

Conclusions

In this study, we designed the NeuroScale, a multi-channel neural network model based on ESM (evolutionary scale modeling), specifically designed for the accurate prediction of NPs. The main contribution of this work was to use advanced deep learning technology and the powerful representation ability of the protein large language model to efficiently and accurately classify NPs.

Not only did NeuroScale surpass the existing NP prediction methods on standard dataset, but it also demonstrated its unique advantages in the comparison of feature extraction effects, running time comparison, external verification, and comparison with other NP prediction works. By combining the powerful feature extraction ability of the ESM2 model and the fine design of the multi-channel fully connected layer, our model can capture the complex features of NP sequences, which are crucial for accurate prediction of NP sequences.

In addition, when exploring the ability of NeuroScale models to process protein sequences, we not only focused on the length of protein sequences, but also investigated the impact of similarity thresholds on model performance. By constructing datasets with different sequence lengths and similarity thresholds, the adaptability of the model to deal with diverse sequence conditions is verified. The experimental results show that NeuroScale can maintain a high prediction accuracy regardless of the long sequence or various similarity thresholds. This not only emphasizes the reliability and robustness of the model in practical application, but also highlights the importance of setting appropriate similarity threshold in improving the adaptability and accuracy of the model. This discovery is of great significance for bioinformatics research and drug development related to NPs. In practical applications, the diversity of protein sequence lengths requires that the model must be capable of handling this diversity. At the same time, by setting the similarity threshold reasonably, the accuracy of the model for sequence feature recognition can be further improved, thus playing a key role in biomarker discovery, disease mechanism understanding, and new drug development.

NeuroScale, with its unique multi-channel neural network architecture and powerful feature extraction capabilities based on ESM, effectively captures multi-level sequence information when processing neuropeptide sequences. It extracts and integrates multi-dimensional features from different perspectives, enabling the identification of complex biological patterns. We believe that this architectural advantage allows NeuroScale to not only be applied to neuropeptide sequences but also to handle other types of biological sequences. Despite differences in structure and function across various biological sequences, NeuroScale's multi-scale feature extraction ability enables it to capture both the commonalities and

differences within these sequences, ensuring outstanding performance across other biological sequences and diverse biological contexts. Future research will focus on enhancing the model's interpretability, aiming to deepen our understanding of the mechanisms behind neuropeptide sequence prediction and provide more reliable support for biomarker discovery and the exploration of disease mechanisms.

Methods

Dataset construction

We collected 11,282 sequences of NPs from the NeuroPep 2.0 database [54]. After that, we took a series of screening steps. First, the NPs between 5 and 100 in length were preserved. Secondly, we used the CD-HIT (version: 4.8.1) tool and set a similarity threshold of 0.8 to eliminate samples with more than 80% similarity with other sequences. After the above screening steps, 2331 NPs were still retained. For the negative samples, we used the same method as the one used in the NeuroPred-FRL and the NeuroPred-PLM, which was to extract from the UniProt database and maintain a similar length distribution as the positive samples. In the end, we got 2331 negative samples. We randomly split the data into a training set and an independent test set in an 8:2 ratio. This dataset will serve as the benchmark for comparing subsequent models.

The datasets for the hypocretin neuropeptide precursor, natriuretic peptides A, and progonadoliberin were downloaded directly from the UniProt database, with sequence lengths between 1 and 200. The negative samples were also extracted from the UniProt database and maintained a length distribution similar to the positive samples. The size of the hypocretin neuropeptide precursor dataset was 520 positive samples and 520 negative samples, the size of the natriuretic peptides A dataset was 529 positive samples and 529 negative samples, and the size of the progonadoliberin dataset was 1079 positive samples and 1079 negative samples.

The bitter peptide-based dataset was obtained from http://pmlab.pythonanywhere.com/BERT4Bitter. The data contained 640 records, including 320 bitter and 320 non-bitter peptides [55].

To validate our model's ability to handle multi-scale protein sequences, specifically its capability to process protein sequences under different sequence length constraints and similarity threshold conditions, we further processed the data based on the NeuroPep 2.0 dataset. This dataset was filtered to retain sequences of length 5–100 and clustered using CD-HIT with similarity thresholds of 0.4, 0.5, 0.6, 0.7, and 0.8, resulting in datasets of 480, 943, 1501, 1937, and 2331 samples, respectively. Additionally, five datasets were generated with sequence length constraints of ≤100, ≤200, ≤300, ≤400, and no limit, yielding sample sizes of 10,470, 10,900, 10,982, 11,162, and 11,282, respectively (Table S7).

Feature representing

The model is divided into two parts: feature extraction and classifier. In the feature extraction section, we selected esm2_t33_650M_UR50D in the ESM2 model. With 33 layers and 650 M parameters, the pre-trained model has been trained on hundreds of millions of protein sequences, making it one of the most popular protein large language models. esm2_t33_650M_UR50D can capture the sequence-structure–function relationship in large-scale protein sequence database, revealing the evolutionary mode and characteristics of proteins. This deep learning approach is expressive and can extract highly abstract features from protein sequences, including relative positional information and amino acid interactions, which are critical for protein function and structure prediction. All protein large language models were fine-tuned during training without freezing any layers. For the final model used in NeuroScale (esm2_t33_650M_UR50D), full fine-tuning was applied to optimize all parameters for the neuropeptide prediction task.

Multi-channel residual neural network

In our fully connected neural network's classifier, we leverage the architectural principles of GoogLeNet to develop a multi-channel fully connected neural network model. This design involves three distinct channels, each employing specific strategies to manipulate the feature space.

The first channel utilizes a fully connected layer to directly map the 1280-dimensional features to 64 dimensions. The mathematical formulation for this channel is:

$$output_1 = \mathrm{ReLU}(W_1 \cdot input + b_1) \quad (1)$$

where W_1 and b_1 are the weights and biases for the first channel, respectively, and ReLU is the activation function used.

The second channel, drawing inspiration from GoogLeNet, focuses on reducing the dimensionality of the feature space before expanding it again. This channel effectively compresses the features from 1280 dimensions to 32, then expands them back to 64. The operations in this channel are expressed as:

$$output_2 = \mathrm{ReLU}(W_{2b} \cdot (\mathrm{ReLU}(W_{2a} \cdot input + b_{2a})) + b_{2b}) \quad (2)$$

here, W_{2a}, b_{2a}, W_{2b}, and b_{2b} are the parameters adjusting the feature dimensions.

Similarly, the third channel modifies the feature dimensions from 1280 to 128 and then to 64. This is mathematically represented as:

$$output_3 = \mathrm{ReLU}(W_{3b} \cdot (\mathrm{ReLU}(W_{3a} \cdot input + b_{3a})) + b_{3b}) \quad (3)$$

here, W_{3a}, b_{3a}, W_{3b}, and b_{3b} are the parameters adjusting the feature dimensions.

Finally, the outputs from these three channels are aggregated to compute the final feature extraction value. This aggregation is encapsulated in the formula:

$$output = output_1 + output_2 + output_3 \quad (4)$$

This comprehensive approach allows for more detailed information extraction from different angles.

Loss function

In the classifier section of NeuroScale, we use a loss function known as the binary cross-entropy loss (BCELoss), which is ideal for binary classification problems. The BCELoss function measures the performance of a classification model whose output is a probability value between 0 and 1. The loss increases as the predicted probability diverges from the actual label.

The mathematical expression for binary cross-entropy loss is:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(p_i) + \left(1 - y_i\right) \cdot \log(1 - p_i)\right] \quad (5)$$

where N is the number of observations in the batch, y_i is the actual label of instance i (0 or 1), and p_i is the predicted probability of instance i being in class 1.

Parameter setting

Our detailed model parameters are set as shown in Table S8. The learning rate is 0.001, the batch size is 1, the optimizer is SGD, and we have incorporated dropout in the model.

Evaluation index

In this study, we used several widely used machine learning evaluation indicators, including ACC, Pre, Rec, F1, and MCC. We also drew ROC and PRC and obtained the area under the curve (AUC, AUPRC) [56–61]. The specific formulas for these indicators are as follows:

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (6)$$

$$\mathrm{Rec} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (7)$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (8)$$

$$\mathrm{F1} = \frac{2\mathrm{PreRec}}{\mathrm{Pre} + \mathrm{Rec}} \quad (9)$$

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})}} \quad (10)$$

among them, TP, TN, FP, and FN represented the true positive, true negative, false positive, and false negative of the sample, respectively.

Abbreviations

NPs	Neuropeptides
ESM	Evolutionary scale modeling
F1	F1 score
ACC	Accuracy
E1	Esm2_t6_8M_UR50D
E2	Esm2_t12_35M_UR50D
E3	Esm2_t30_150M_UR50D
E4	Esm2_t33_650M_UR50D
Rec	Recall
Pre	Precision
MCC	Matthews correlation coefficient
PRC	Precision-recall curve
ROC	Receiver operating characteristic curve
BCELoss	Binary cross-entropy loss

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02243-6.

Additional file 1: Table S1 Performance of protein large language models. Table S2 Running time of different protein large language models. Table S3 The performance of NeuroScale based on the external test set. Table S4 Comparison table between NeuroScale and the latest NPs prediction methods. Table S5 Model performance evaluation results at different protein similarity thresholds. Table S6 Model performance evaluation results at different protein sequence lengths. Table S7 Data construction information. Table S8 Model parameter settings.

Acknowledgements

Not applicable.

Authors'contributions

H.L.1, H.L.2, and M.Y. conceived the idea and supervised the study. H.Q.Z. and S.H.L. analyzed the data. W.S. and X.Q.X. wrote the manuscript with input from all authors. J.W.Y. and F.Y.D. made contributions to the revision of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Scientific Foundation of China (82130112, 62402089, 62373079), Science and Technology Department of Sichuan Province (2025ZNSFSC1465, 2024ZYD0039), China Postdoctoral Science Foundation (2023 TQ0047, GZC20230380), Health Commission of Sichuan Province (24 CXTD11), Sichuan Medical Association (S23012), and Health Commission of Chengdu (2024141).

Data availability

https://github.com/ZhangHongqi215/NeuroScale; https://doi.org/10.5281/zenodo.15064998.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 22 February 2025 Accepted: 12 May 2025
Published online: 28 May 2025

References

1. Casini G. Neuropeptides and retinal development. Arch Ital Biol. 2005;143(3–4):191–8.
2. Escudero Castelan N, Semmens DC, Guerra LAY, Zandawala M, Dos Reis M, Slade SE, Scrivens JH, Zampronio CG, Jones AM, Mirabeau O, et al. Receptor deorphanization in an echinoderm reveals kisspeptin evolution and relationship with SALMFamide neuropeptides. BMC Biol. 2022;20(1):187.
3. Gozes I, Brenneman DE. Neuropeptides as growth and differentiation factors in general and VIP in particular. J Mol Neurosci. 1993;4(1):1–9.
4. Choi U, Hu M, Zhang Q, Sieburth D: The head mesodermal cell couples FMRFamide neuropeptide signaling with rhythmic muscle contraction in C. elegans. Nat Commun 2023, 14(1):4218.
5. Yoneda M, Watanobe H, Terano A. Central regulation of hepatic function by neuropeptides. J Gastroenterol. 2001;36(6):361–7.
6. Xu W, Qin X, Liu Y, Chen J, Wang Y: Advances in enzyme-responsive supramolecular in situ self-assembled peptide for drug delivery. Curr Drug Deliv 2023.
7. De Fruyt N, Yu AJ, Rankin CH, Beets I, Chew YL: The role of neuropeptides in learning: insights from C. elegans. Int J Biochem Cell Biol 2020, 125:105801.
8. Borbely E, Scheich B, Helyes Z. Neuropeptides in learning and memory. Neuropeptides. 2013;47(6):439–50.
9. Ogren SO, Kuteeva E, Elvander-Tottie E, Hokfelt T. Neuropeptides in learning and memory processes with focus on galanin. Eur J Pharmacol. 2010;626(1):9–17.
10. Botelho M, Cavadas C. Neuropeptide Y: an anti-aging player? Trends Neurosci. 2015;38(11):701–11.
11. Wang Y, Wang M, Yin S, Jang R, Wang J, Xue Z, Xu T: NeuroPep: a comprehensive resource of neuropeptides. Database (Oxford) 2015, 2015:bav038.
12. Hokfelt T, Bartfai T, Bloom F. Neuropeptides: opportunities for drug discovery. Lancet Neurol. 2003;2(8):463–72.
13. Tyburski AL, Cheng L, Assari S, Darvish K, Elliott MB. Frequent mild head injury promotes trigeminal sensitivity concomitant with microglial proliferation, astrocytosis, and increased neuropeptide levels in the trigeminal pain system. J Headache Pain. 2017;18(1):16.
14. Carniglia L, Ramírez D, Durand D, Saba J, Turati J, Caruso C, Scimonelli TN, Lasaga M. Neuropeptides and microglial activation in inflammation, pain, and neurodegenerative diseases. Mediators Inflamm. 2017;2017(1):5048616.
15. Cai W, Kim C-H, Go H-J, Egertová M, Zampronio CG, Jones AM, Park NG, Elphick MR. Biochemical, anatomical, and pharmacological characterization of calcitonin-type neuropeptides in starfish: discovery of an ancient role as muscle relaxants. Front Neurosci. 2018;12:382.
16. Sobrino Crespo C, Perianes Cachero A, Puebla Jimenez L, Barrios V, Arilla Ferreiro E. Peptides and food intake. Front Endocrinol. 2014;5:58.
17. Mishra P, Yang SE, Montgomery AB, Reed AR, Rodan AR, Rothenfluh A. The fly liquid-food electroshock assay (FLEA) suggests opposite roles for neuropeptide F in avoidance of bitterness and shock. BMC Biol. 2021;19(1):31.
18. Odekunle EA, Semmens DC, Martynyuk N, Tinoco AB, Garewal AK, Patel RR, Blowes LM, Zandawala M, Delroisse J, Slade SE, et al. Ancient role of vasopressin/oxytocin-type neuropeptides as regulators of feeding revealed in an echinoderm. BMC Biol. 2019;17(1):60.
19. Valente R, Cordeiro M, Pinto B, Machado A, Alves F, Sousa-Pinto I, Ruivo R, Castro LFC. Alterations of pleiotropic neuropeptide-receptor gene couples in Cetacea. BMC Biol. 2024;22(1):186.
20. Saini M, Trehan K, Thakur S, Modi A, Jain SK: Advances in iron deficiency anaemia management: exploring novel drug delivery systems and future perspectives. Curr Drug Deliv 2024.
21. Shahjahan M, Kitahashi T, Parhar IS. Central pathways integrating metabolism and reproduction in teleosts. Front Endocrinol. 2014;5:36.
22. Jonsson M, Morin M, Wang CK, Craik DJ, Degnan SM, Degnan BM. Sex-specific expression of pheromones and other signals in gravid starfish. BMC Biol. 2022;20(1):288.
23. Hu S, Wang Y, Han X, Dai M, Zhang Y, Ma Y, Weng S, Xiao L. Activation of oxytocin receptors in mouse GABAergic amacrine cells modulates retinal dopaminergic signaling. BMC Biol. 2022;20(1):205.
24. Kormos V, Gaszner B. Role of neuropeptides in anxiety, stress, and depression: from animals to humans. Neuropeptides. 2013;47(6):401–19.
25. Hökfelt T, Broberger C. Xu Z-QD, Sergeyev V, Ubink R, Diez M: Neuropeptides—an overview. Neuropharmacology. 2000;39(8):1337–56.
26. Jacobson LH, Hoyer D, de Lecea L. Hypocretins (orexins): the ultimate translational neuropeptides. J Intern Med. 2022;291(5):533–56.
27. Rana T, Behl T, Sehgal A, Singh S, Sharma N, Abdeen A, Ibrahim SF, Mani V, Iqbal MS, Bhatia S et al: Exploring the role of neuropeptides in depression and anxiety. Progress in Neuro-Psychopharmacology & Biological Psychiatry 2022, 114.
28. Shen YC, Sun X, Li L, Zhang HY, Huang ZL, Wang YQ: Roles of neuropeptides in sleep-wake regulation. Int J Mol Sci 2022, 23(9).
29. Wu Y, Berisha A, Borniger JC. Neuropeptides in cancer: friend and foe? Adv Biol (Weinh). 2022;6(9): e2200111.
30. Lasagni Vitar RM, Rama P, Ferrari G. The two-faced effects of nerves and neuropeptides in corneal diseases. Prog Retin Eye Res. 2022;86: 100974.
31. Xu J, Xu L, Sui P, Chen J, Moya EA, Hume P, Janssen WJ, Duran JM, Thistlethwaite P, Carlin A et al: Excess neuropeptides in lung signal through endothelial cells to impair gas exchange. Dev Cell 2022, 57(7):839–853 e836.
32. Wang J, Yin T, Xiao X, He D, Xue Z, Jiang X, Wang Y: StraPep: a structure database of bioactive peptides. Database (Oxford) 2018, 2018.
33. Polfer NC, Oomens J. Reaction products in mass spectrometry elucidated with infrared spectroscopy. Phys Chem Chem Phys. 2007;9(29):3804–17.
34. Patriarchi T. New technologies to investigate neuropeptides at scale. ACS Chem Neurosci. 2022;13(16):2353–5.
35. Wang H, Qian T, Zhao Y, Zhuo Y, Wu C, Osakada T, Chen P, Chen Z, Ren H, Yan Y et al: A tool kit of highly selective and sensitive genetically encoded neuropeptide sensors. Science 2023, 382(6672):eabq8173.
36. Phetsanthad A, Vu NQ, Li L: Multi-faceted mass spectrometric investigation of neuropeptides in Callinectes sapidus. J Vis Exp 2022(183).
37. Liu Y, Wang S, Li X, Liu Y, Zhu X. NeuroPpred-SVM: a new model for predicting neuropeptides based on embeddings of BERT. J Proteome Res. 2023;22(3):718–28.
38. Wang L, Zeng Z, Xue Z, Wang Y. DeepNeuropePred: a robust and universal tool to predict cleavage sites from neuropeptide precursors by protein language model. Comput Struct Biotechnol J. 2024;23:309–15.
39. Bin Y, Zhang W, Tang W, Dai R, Li M, Zhu Q, Xia J. Prediction of neuropeptides from sequence information using ensemble classifier and hybrid features. J Proteome Res. 2020;19(9):3732–40.
40. Hasan MM, Alam MA, Shoombuatong W, Deng HW, Manavalan B, Kurata H: NeuroPred-FRL: an interpretable prediction model for identifying neuropeptide using feature representation learning. Brief Bioinform 2021, 22(6).
41. Jiang M, Zhao B, Luo S, Wang Q, Chu Y, Chen T, Mao X, Liu Y, Wang Y, Jiang X et al: NeuroPpred-Fuse: an interpretable stacking model for prediction of neuropeptides by fusing sequence information and feature selection methods. Brief Bioinform 2021, 22(6).
42. Wang L, Huang C, Wang M, Xue Z, Wang Y: NeuroPred-PLM: an interpretable and robust model for neuropeptide prediction by protein language model. Brief Bioinform 2023, 24(2).
43. Geffen Y, Ofran Y, Unger R: DistilProtBert: a distilled protein language model used to distinguish between real proteins and their randomly shuffled counterparts. Bioinformatics 2022, 38(Suppl_2):ii95-ii98.
44. Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, et al. ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Trans Pattern Anal Mach Intell. 2022;44(10):7112–27.

45. Hesslow D, Zanichelli N, Notin P, Poli I, Marks D: RITA: a study on scaling up generative protein sequence models. 2022.
46. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28–36.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
48. Nepomnyachiy S, Ben-Tal N, Kolodny R. Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths. Proc Natl Acad Sci U S A. 2017;114(44):11703–8.
49. Shi Q, Chen W, Huang S, Wang Y, Xue Z. Deep learning for mining protein data. Brief Bioinform. 2021;22(1):194–218.
50. Wang L, Zhong H, Xue Z, Wang Y. Improving the topology prediction of alpha-helical transmembrane proteins with deep transfer learning. Comput Struct Biotechnol J. 2022;20:1993–2000.
51. Sandhya S, Rani SS, Pankaj B, Govind MK, Offmann B, Srinivasan N, Sowdhamini R. Length variations amongst protein domain superfamilies and consequences on structure and function. PLoS ONE. 2009;4(3): e4981.
52. Nevers Y, Glover NM, Dessimoz C, Lecompte O. Protein length distribution is remarkably uniform across the tree of life. Genome Biol. 2023;24(1):135.
53. Koehler Leman J, Szczerbiak P, Renfrew PD, Gligorijevic V, Berenberg D, Vatanen T, Taylor BC, Chandler C, Janssen S, Pataki A, et al. Sequence-structure-function relationships in the microbial protein universe. Nat Commun. 2023;14(1):2351.
54. Wang M, Wang L, Xu W, Chu Z, Wang H, Lu J, Xue Z, Wang Y: NeuroPep 2.0: an updated database dedicated to neuropeptide and its receptor annotations. J Mol Biol 2024, 436(4):168416.
55. Zhang YF, Wang YH, Gu ZF, Pan XR, Li J, Ding H, Zhang Y, Deng KJ. BitterRF: a random forest machine model for recognizing bitter peptides. Front Med (Lausanne). 2023;10:1052923.
56. Dao F, Xie X, Zhang H, Guan Z, Wu C, Su W, Wei Y, Hong F, Luo X, Xie S et al: PlantEMS: a comprehensive database of epigenetic modification sites across multiple plant species. Plant Commun 2024:101228.
57. Abdelkader GA, Kim JD. Advances in protein-ligand binding affinity prediction via deep learning: a comprehensive study of datasets, data preprocessing techniques, and model architectures. Curr Drug Targets. 2024;25(15):1041–65.
58. Lv H, Dao FY, Lin H. DeepKla: an attention mechanism-based deep neural network for protein lysine lactylation site prediction. Imeta. 2022;1(1): e11.
59. Ma M, Zhao R, Li X, Jing M, Song R, Fan J: Biological properties of arginine-rich peptides and their application in cargo delivery to cancer. Curr Drug Deliv 2023.
60. Mou Y, Cao M, Zhang D: The blood-prostate barrier: an obstacle to drug delivery into the prostate. Curr Drug Deliv 2023.
61. Fu Q, Liu Y, Peng C, Muluh TA, Anayyat U, Liang L. Recent advancement in inhaled nano-drug delivery for pulmonary, nasal, and nose-to-brain diseases. Curr Drug Deliv. 2025;22(1):3–14.

Publisher's Note