

Table S1. Five-fold cross-validation performance of BiToxNet.

Datasets	Metrics	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	95% confidence intervals
Protein	Accuracy	90.65	89.68	90.32	89.68	89.68	90.00 $\pm$ 0.51
	Sensitivity	87.10	86.45	89.68	89.03	88.39	88.13 $\pm$ 1.49
	Specificity	94.19	92.90	90.97	90.32	90.97	91.87 $\pm$ 1.80
	MCC	0.815	0.795	0.807	0.794	0.794	0.801 $\pm$ 0.011
	AUC	96.74	0.964	97.49	96.15	97.29	96.81 $\pm$ 0.64
Peptide	Accuracy	95.44	95.16	94.87	96.01	95.73	95.44 $\pm$ 0.50
	Sensitivity	94.32	94.89	0.943	96.59	95.45	95.11 $\pm$ 1.06
	Specificity	96.57	95.43	0.954	95.43	96.00	95.77 $\pm$ 0.57
	MCC	0.909	0.903	0.898	0.920	0.915	0.909 $\pm$ 0.010
	AUC	99.18	98.96	98.49	98.36	99.00	98.80 $\pm$ 0.39
Combined	Accuracy	92.13	91.68	92.28	91.38	92.28	91.95 $\pm$ 0.45
	Sensitivity	90.94	96.07	93.05	93.05	93.96	93.41 $\pm$ 2.06
	Specificity	93.33	87.27	91.52	89.70	90.61	90.48 $\pm$ 2.49
	MCC	0.843	0.837	0.846	0.828	0.846	0.840 $\pm$ 0.009
	AUC	98.04	98.24	98.13	97.71	97.97	98.02 $\pm$ 0.22

Table S2. Quantitative results of ablation studies for BiToxNet.

Datasets	Model	Accuracy	Sensitivity	Specificity	MCC	AUC
Peptide	w/o ESM2 Encoder	93.16	87.50	98.86	0.869	99.31
	w/o Protein Encoder	92.02	93.18	90.86	0.841	96.33
	w/o BAN	93.16	92.05	94.29	0.864	98.22
	BiToxNet	96.01	94.89	97.14	0.921	99.39
Protein	w/o ESM2 Encoder	91.29	89.03	93.55	0.827	97.94
	w/o Protein Encoder	88.06	86.45	89.68	0.762	94.30
	w/o BAN	90.97	88.39	93.55	0.820	97.63
	BiToxNet	92.26	89.68	94.84	0.846	98.19
Combined	w/o ESM2 Encoder	91.83	93.66	90.0	0.837	97.29
	w/o Protein Encoder	90.02	95.47	84.55	0.805	96.65
	w/o BAN	90.17	90.33	90.0	0.803	96.49
	BiToxNet	92.74	94.56	90.91	0.855	97.54

Table S3. Hyperparameter search space and optimal settings.

Hyperparameters	Search Space	Best hyperparameter
Learning Rate	{2e-4, 4e-4, 8e-4}	4e-4
WEIGHT_DECAY	{5e-5, 1e-4, 2e-4}	1e-4
BATCH_SIZE	{16, 32, 64}	32
HIDDEN_SIZE	{512, 768, 1024}	768
REDUCE_DIM	{128, 256, 512}	256
DROPOUT	{0.3, 0.5, 0.7}	0.5
BAN_HEADS	{2, 4, 6}	4
BAN_DROPOUT	{0.05, 0.1, 0.2}	0.1

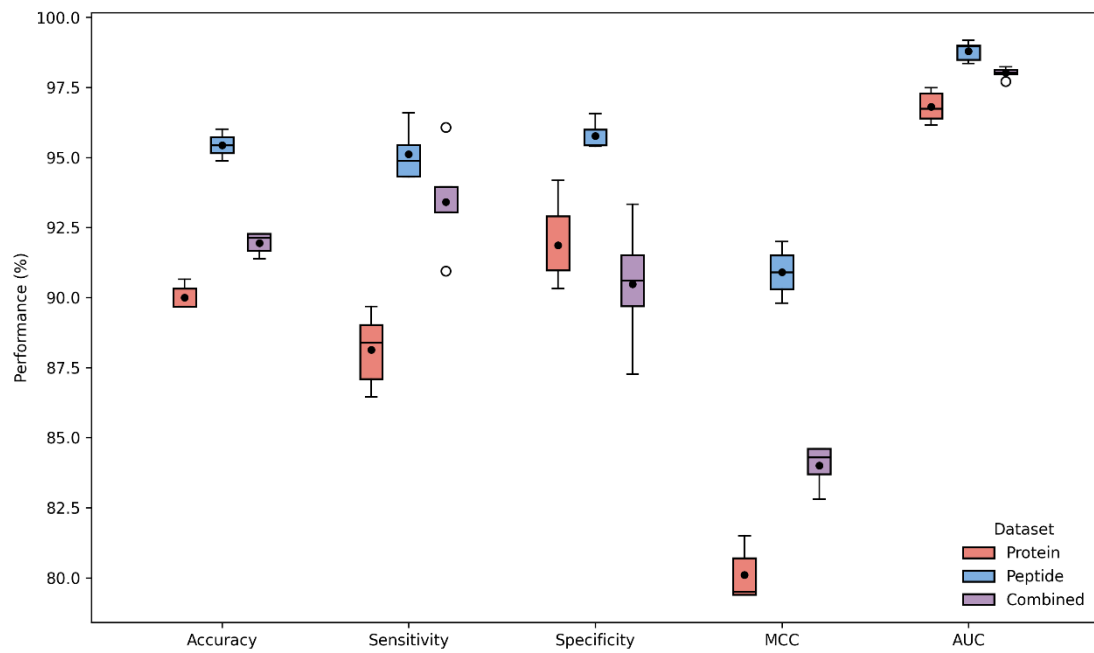


Figure S1. Boxplots illustrating the five-fold cross-validation results of BiToxNet across five performance metrics: Accuracy, Sensitivity, Specificity, MCC, and AUC. Red represents the protein dataset, blue signifies the peptide dataset, and purple denotes the combined dataset.

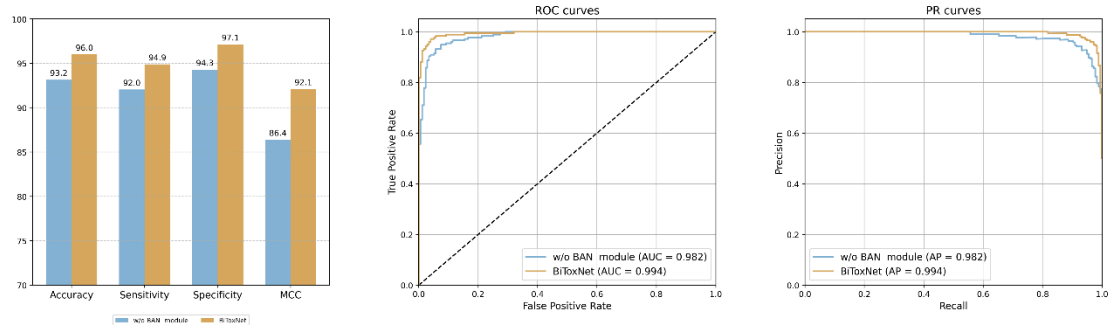


Figure S2. Performance comparison between models without BAN (blue) and full BiToxNet (orange) across four key metrics on the peptide dataset.

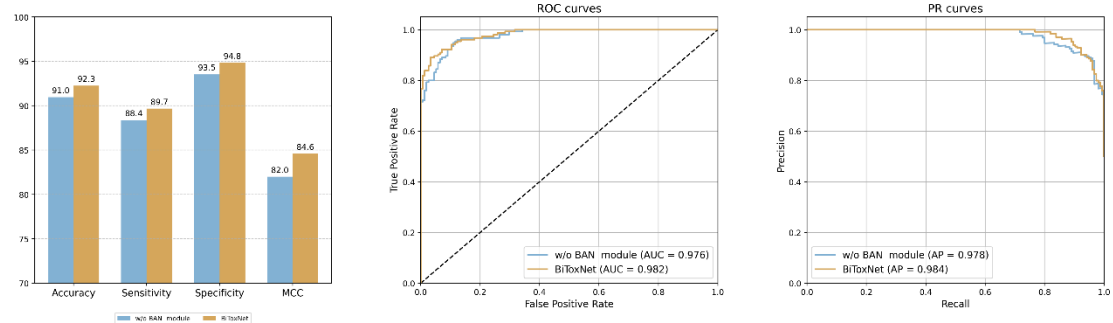


Figure S3. Performance comparison between models without BAN (blue) and full BiToxNet (orange) across four key metrics on the protein dataset.

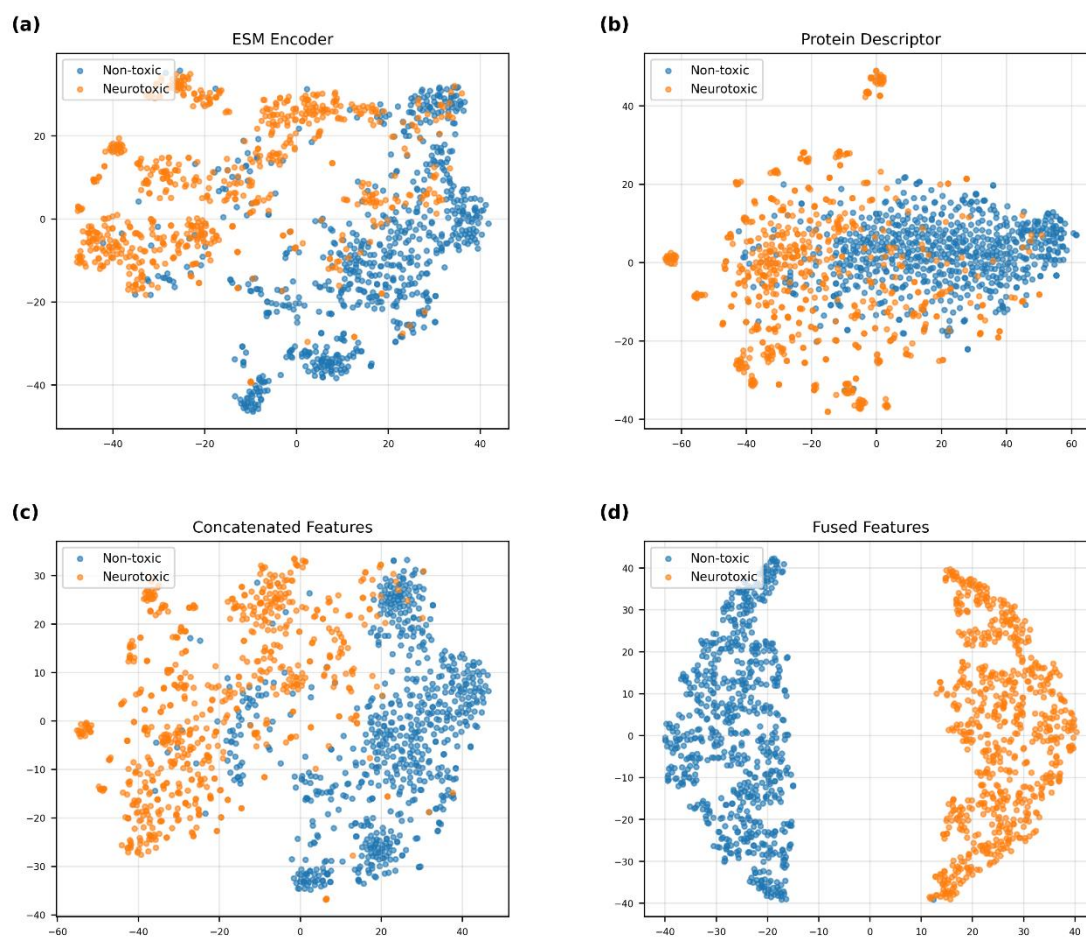


Figure S4. t-SNE visualization of feature evolution on the peptide training dataset. (a) ESM-2 embeddings show mixed distributions. (b) Handcrafted descriptors reveal clustered overlaps. (c) Concatenated features exhibit improved but adversarial separation. (d) BAN-fused features achieve optimal class isolation.

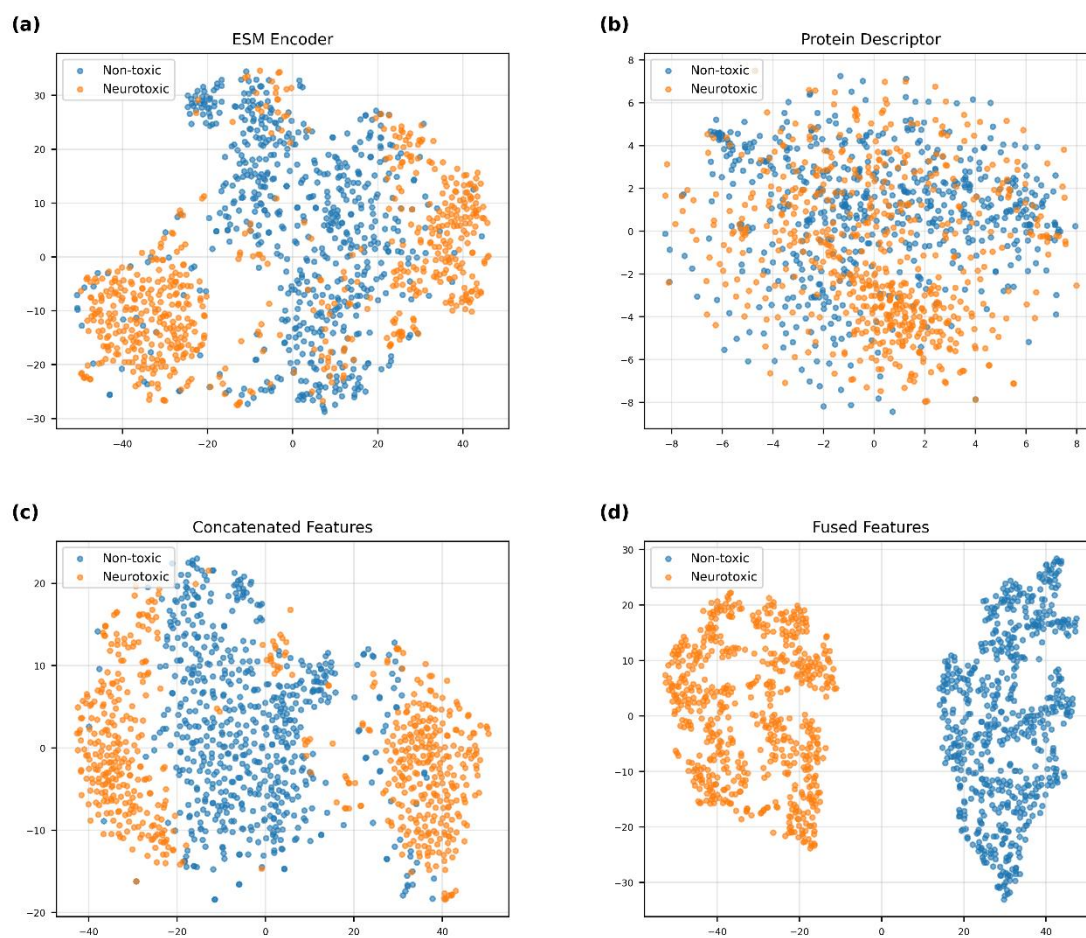


Figure S5. t-SNE visualization of feature evolution on the protein training dataset. (a) ESM-2 embeddings show mixed distributions. (b) Handcrafted descriptors reveal clustered overlaps. (c) Concatenated features exhibit improved but adversarial separation. (d) BAN-fused features achieve optimal class isolation.

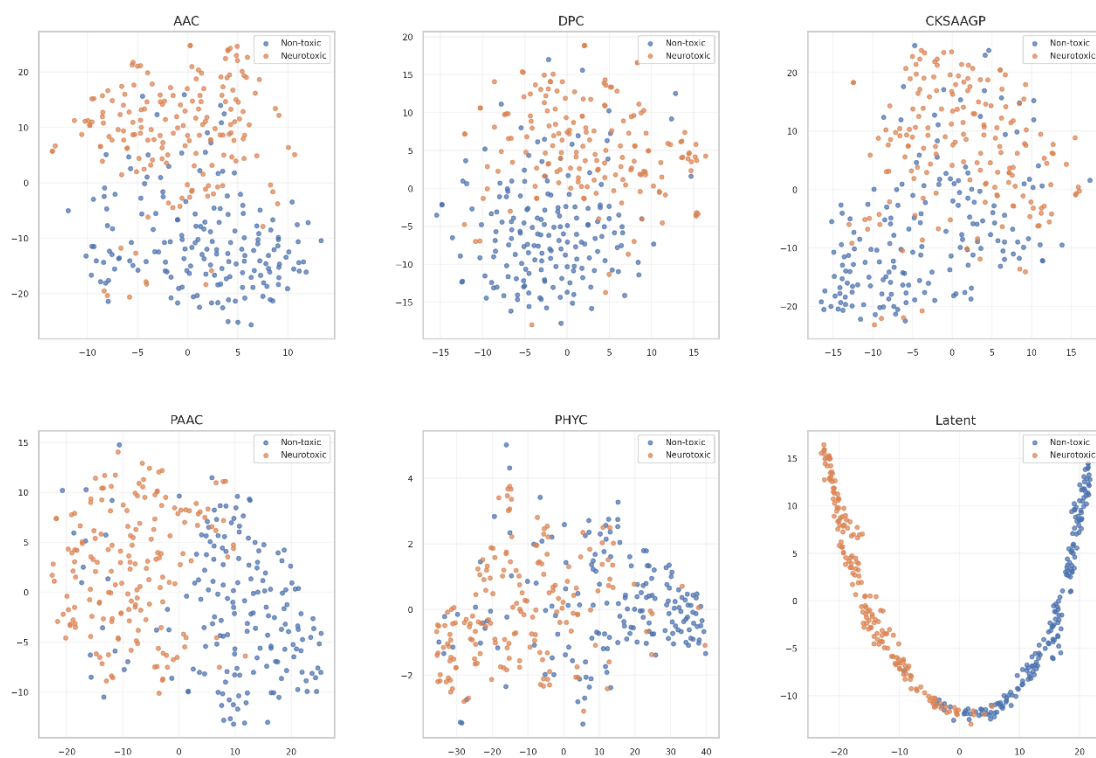


Figure S6. t-SNE visualization of feature discriminability on peptide testing dataset. (a-e) Traditional descriptors show chaotic class overlap. (f) BiToxNet's latent features yield distinct neurotoxic (orange) and non-toxic (blue) clusters.

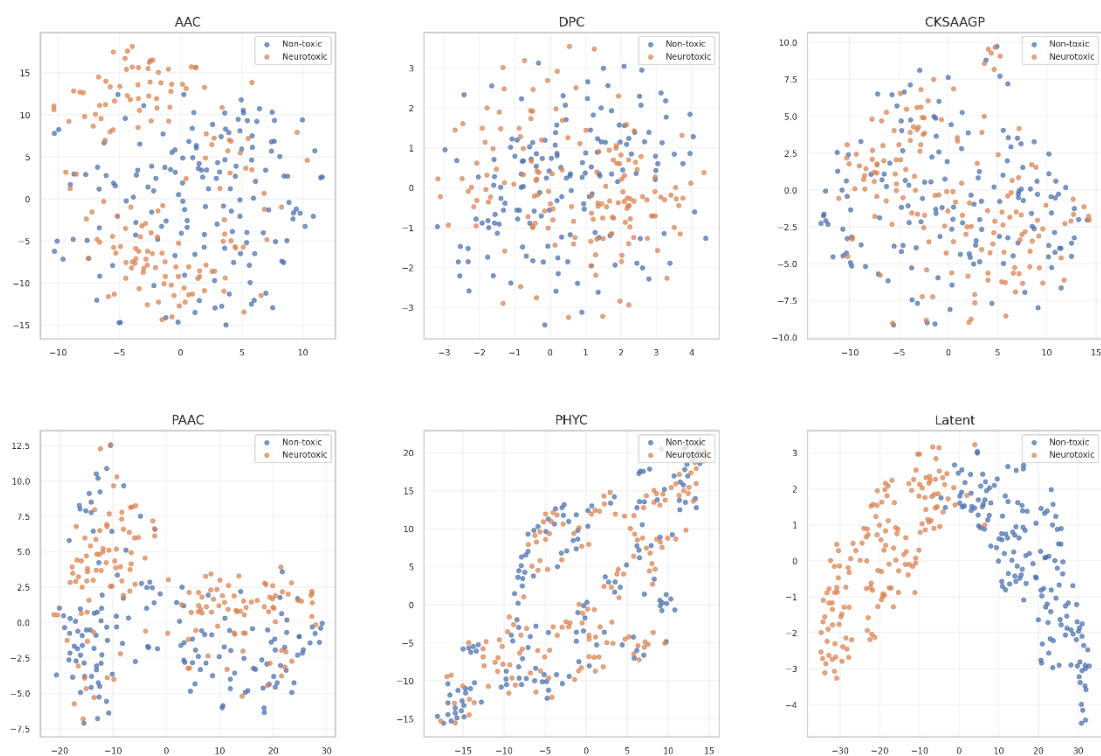


Figure S7. t-SNE visualization of feature discriminability on protein testing dataset. (a-e) Traditional descriptors show chaotic class overlap. (f) BiToxNet's latent features yield distinct neurotoxic (orange) and non-toxic (blue) clusters.

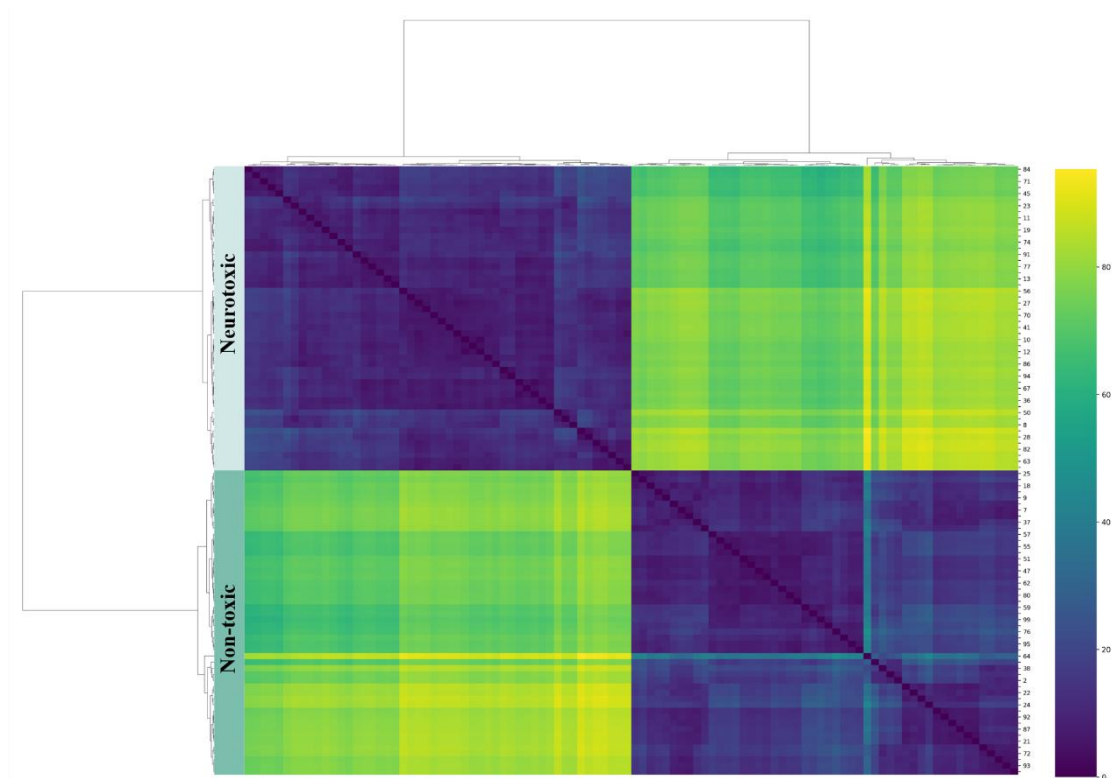


Figure S8. Hierarchical clustering of BiToxNet features on peptide testing dataset. Neurotoxic and non-toxic sequences form perfectly separated subtrees.

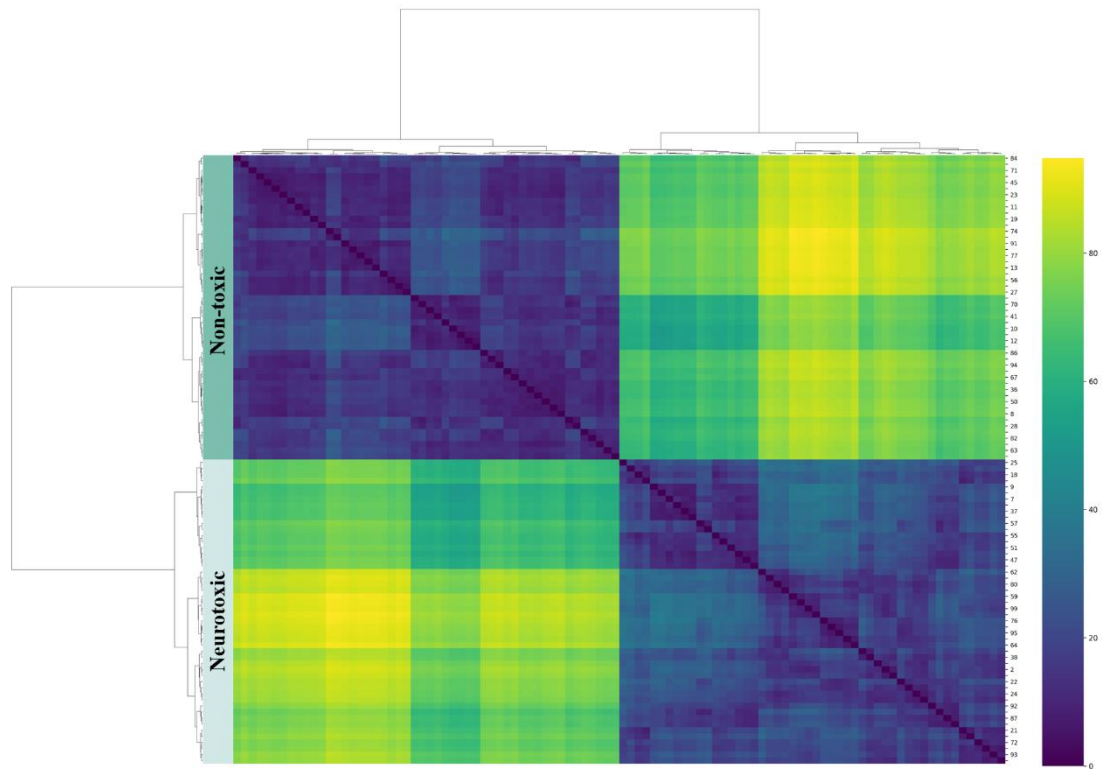


Figure S9. Hierarchical clustering of BiToxNet features on protein testing dataset. Neurotoxic and non-toxic sequences form perfectly separated subtrees.

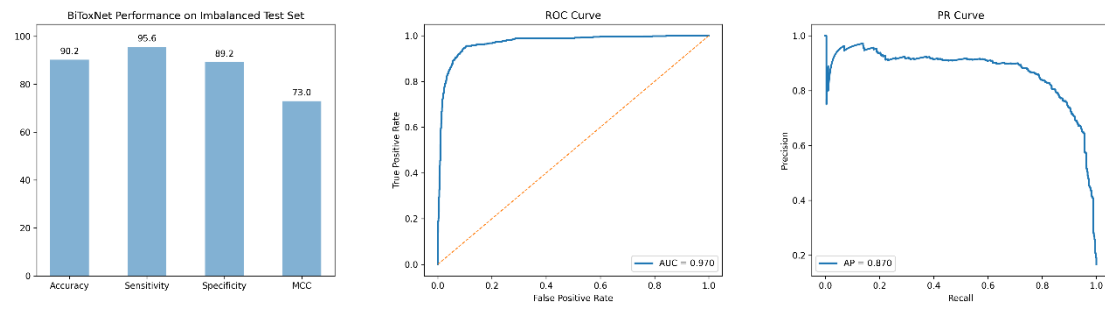


Figure S10. Performance of BiToxNet on the independent imbalanced dataset, including bar-plotted evaluation metrics (accuracy, sensitivity, specificity, and MCC), the ROC curve with AUC, and the PR curve with AP.

Details of ten protein descriptors:

Amino acid composition

The amino acid composition (AAC) quantifies the occurrence frequency of each standard amino acid within a peptide sequence. For a specific amino acid residue r among the 20 canonical amino acids, the AAC value is computed as:

$$AAC(r) = \frac{N(r)}{L}$$

where $N(r)$ denotes the total occurrences of residue r in the sequence, and L represents the sequence length.

Dipeptide Composition

The dipeptide composition (DPC) calculates the relative frequency of each dipeptide, defined as two consecutive amino acid residues, within a peptide sequence. This method generates a 400-dimensional feature vector (20×20) representing all possible ordered pairs of the 20 canonical amino acids. The proportion of a specific dipeptide formed by residue r followed by residue s is derived as:

$$DPC(r, s) = \frac{N_{rs}}{L-1}$$

where N_{rs} denotes the number of dipeptide rs occurrences in the sequence, L is the total sequence length.

Composition of k -Spaced Amino Acid Group Pairs

The Composition of k -Spaced Amino Acid Group Pairs (CKSAAGP) computes the relative frequency of amino acid group pairs separated by exactly k residues. This method categorizes the 20 canonical amino acids into five groups based on physicochemical properties: aliphatic ($g1$), aromatic ($g2$), positively charged ($g3$), negatively charged ($g4$), and uncharged ($g5$). For a specified spacing value k , CKSAAGP generates a 25-dimensional feature vector (5×5) representing all possible group pair combinations ($g1g1, g1g2, \dots, g5g5$). The frequency of a specific group pair (g_i, g_j) is calculated as:

$$CKSAAGP((g_i, g_j; k) = \frac{N_{g_i g_j}}{L - (k+1)}$$

where $N_{g_i g_j}$ denotes the number of occurrences where a residue from group g_i is followed by a residue from group g_j with exactly k residues between them, and L is the sequence length. The denominator $L - (k+1)$ represents the total number of valid positions for k -spaced residue pairs in the sequence.

Pseudo-amino Acid Composition

The Pseudo-amino Acid Composition (PAAC) is a feature extraction method that simultaneously captures standard AAC and sequence order information. This descriptor extends traditional AAC by incorporating sequence correlations through two key parameters: the sequence order correlation factor λ and a weighting factor ω .

The parameter λ determines the maximum separation distance between residues considered for sequence order correlations, while ω controls the relative contribution of sequence order features versus the standard AAC components. Higher ω values increase the influence of sequence order patterns while proportionally reducing the weight of AAC features in the final descriptor.

Physicochemical Properties

Physicochemical properties significantly influence peptide functionality and interactions. This study considers eight fundamental physicochemical properties (PHYC) derived from amino acid sequences: hydrophobicity, isoelectric point, net charge, aliphatic index, alpha-helical propensity, transmembrane propensity, hydrophobic moment, and Boman index.

BLOSUM62

The BLOSUM62 matrix provides evolutionary information by quantifying amino acid substitution patterns observed in conserved protein blocks with less than 62% sequence identity. This encoding scheme converts each amino acid residue in a peptide sequence into a 20-dimensional vector derived from its corresponding row in the BLOSUM62 substitution matrix. Consequently, a peptide sequence of length L is represented as an $L \times 20$ feature matrix, where each row corresponds to a residue in the sequence and each column represents one of the 20 standard amino acids. The matrix elements encode the log-odds scores for observed substitutions, with higher values indicating more evolutionarily favorable replacements. This transformation translates biological sequences into structured numerical data suitable for computational analysis while preserving critical evolutionary relationships between amino acids.

Z-Scale

The Z-Scale provides a quantitative representation of amino acid physicochemical properties through five-dimensional vectors. Each dimension corresponds to distinct biophysical characteristics: z_1 quantifies hydrophobicity and hydrophilicity, z_2 measures steric bulk and polarizability, z_3 reflects polarity, while z_4 and z_5 capture electronic effects. For a peptide sequence of length L , this encoding generates an $L \times 5$ matrix where each row corresponds to the Z-scale values of an amino acid in the sequence.

Quasi sequence Order

The quasi-sequence order (QSOrder) encoding utilizes a comprehensive approach that captures both global amino acid composition and local sequence order information. It simultaneously analyzes the overall distribution of amino acids while incorporating their positional relationships within the protein sequence. This method effectively integrates compositional patterns with sequential arrangement characteristics, creating a unified representation that reflects both local sequence motifs and global structural properties.

Dipeptide Deviation from Expected mean

The Dipeptide Deviation from Expected mean (DDE) encoding employs a statistical approach that captures both amino acid composition and dipeptide occurrence patterns. It calculates the normalized deviation between observed dipeptide frequencies and their expected values based on random amino acid distribution. This method effectively quantifies non-random dipeptide associations in protein sequences, revealing preferential residue combinations and compositional biases through standardized statistical measures.

The PseAAC of distance-pair and reduced alphabet

The PseAAC of distance-pair and reduced alphabet (DistancePair) characterizes protein sequences by quantifying pairwise distances between amino acids grouped via a reduced alphabet. It captures the distribution of residue-pair distances along the sequence, reflecting their relative spatial positions. This encoding implicitly models potential residue interaction patterns through distance-dependent frequency statistics.