Bioorganic Chemistry 165 (2025) 109027

Contents lists available at ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Multimodal profiling of Pepcan-CB1 receptor structure-activity relationships: integrating molecular dynamics simulations, biological profiling, and the deep learning model MuMoPepcan

Hongyang Man [a], Huiming Bao [a], Zhanyu Niu [a], Zhonghua Zhang [a], Jerine Peter Simon [a], Tong Yang [a], Pengtao Li [a], Shouliang Dong [a,b,*]

[a] Department of Animal and Biomedical Sciences, School of Life Sciences, Lanzhou University, 222 Tianshui South Road, Lanzhou 730000, China
[b] Key Laboratory of Preclinical Study for New Drugs of Gansu Province, Lanzhou University, 222 Tianshui South Road, Lanzhou 730000, China

ARTICLE INFO

Keywords:
Computer-aided drug design
Cannabinoid receptors type 1
Pepcans
Molecular dynamics simulation
Deep learning

ABSTRACT

In machine learning of drug discovery, the scale of accessible data is often strictly limited, while few-shot learning in wet-lab experimental data limits the accuracy of machine learning algorithms. Cannabinoid receptors are involved in various important physiological activities, and pepcans are key components of the endocannabinoid system. Herein, we proposed a combined dry-wet lab experimental framework that incorporated molecular dynamics simulation (MDS) data into peptide biological activity prediction. We validated our hypothesis on cannabinoid receptors type 1 (CB1) and pepcans: (1) In the study, we synthesized 45 pepcan peptides to establish a bioactivity dataset and identified RD-pepcan-11 as an lead analgesic compound by Bio-screening, with systematic characterization of its CB1 selectivity and pharmacodynamics.; (2) Millions of conformational data were generated by MDS and a CB1-pepcans conformation dataset was constructed; (3) Combining wet-lab data and MDS data, a deep learning model - MuMoPepcan was developed, reducing prediction errors to within the error range of wet-lab experiments. This study not only identified novel high-potential pepcans - RD-pepcan-11, but also demonstrated that MDS can serve as an effective data augmentation method to scale up drug-receptor datasets, thereby improving model generalizability and performance.

1. Introduction

Machine learning techniques, particularly deep learning methods, can often establish good mappings between two data distributions and are used for computer-aided drug design (CADD) [1]. In drug discovery, predictive workflows based on graph neural network models have demonstrated strong performance in quickly assessing the binding affinity between receptors and ligands [2,3]. However, current deep learning models generally rely heavily on large-scale general training datasets, which limits their adaptability to specific receptors [4]. Even more concerning is that, despite being state-of-the-art (SOTA), existing methods still exhibit prediction errors for ligand-receptor binding affinity within a range of 1.2 to 1.4 units (which corresponds to $-\log(Kd/Ki)$), indicating that the field is still some distance from practical applications [5].

Molecular docking, a computational method for predicting the optimal binding arrangement (pose) of a ligand within a receptor's binding pocket, leverages scoring functions to rank potential binding modes [6]. Starting from these static conformations, further integration with molecular dynamics simulation (MDS) technology allows for the investigation of dynamic interaction properties between ligands and receptors [7,8]. Numerous studies [9–11] have demonstrated that long-term MDS can effectively explore and reveal the interaction patterns and mechanisms between G-protein coupled receptors (GPCRs) and drug molecules. In modern CADD workflows, these techniques are often used synergistically: molecular docking provides high-throughput screening capability due to its computational efficiency, while MDS offers high-resolution mechanistic insights by accounting for full flexibility and dynamics, albeit at a greater computational cost.

In recent years, researchers have gradually shifted their focus to MDS, exploring its potential as a data augmentation tool. By leveraging MDS technology, it is possible to generate conformational datasets for

* Corresponding author at: Department of Animal and Biomedical Sciences School of Life Sciences, Lanzhou University, 222 Tianshui South Road, Lanzhou 730000, China.
E-mail address: dongsl@lzu.edu.cn (S. Dong).

https://doi.org/10.1016/j.bioorg.2025.109027
Received 25 July 2025; Received in revised form 31 August 2025; Accepted 22 September 2025
Available online 23 September 2025

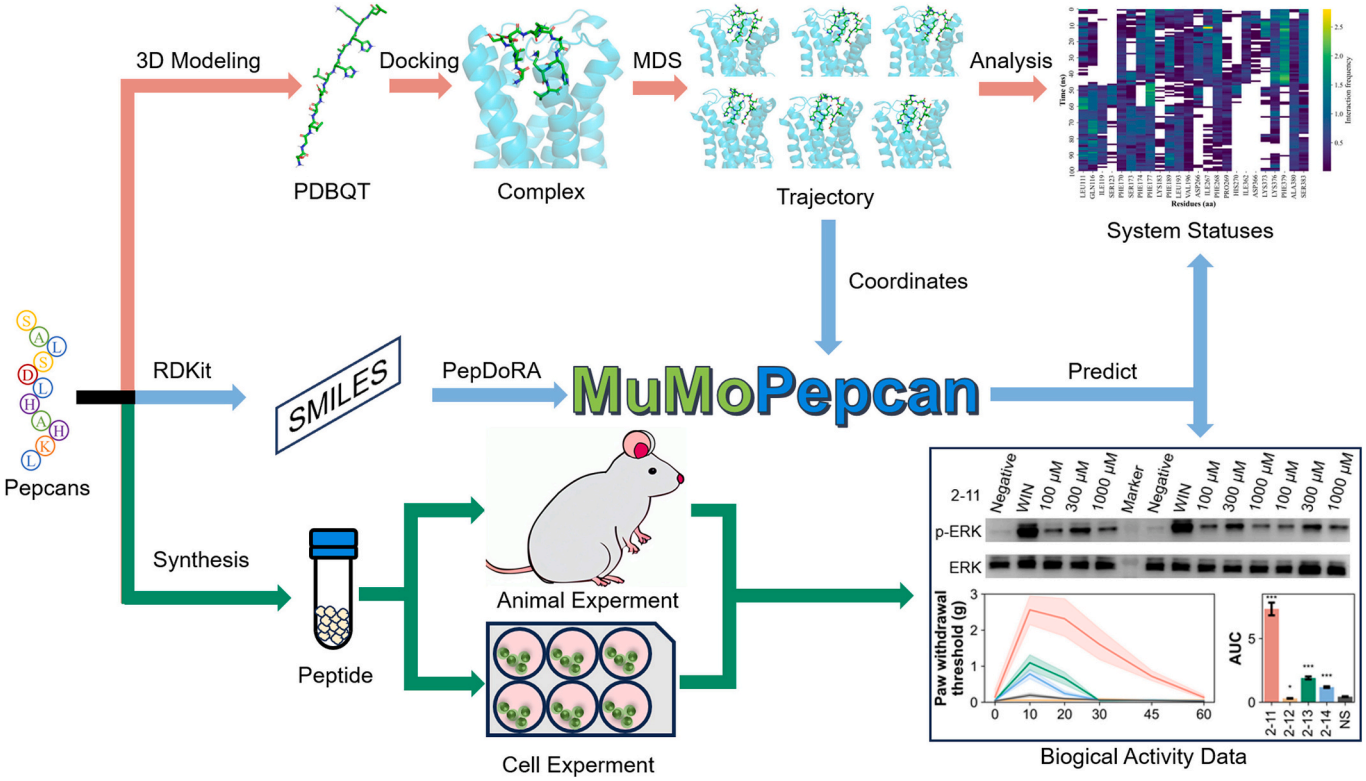

Fig. 1. MuMoPepcan's integrated dry-wet experimental framework. (1) Analysis of the interaction modes between pepcans and CB1 receptors using Molecular Docking and MDS. (2) Chemical synthesis of pepcan family members, followed by anti-inflammatory pain relief experiments and receptor activation activity assays on CB1-overexpressing cells to obtain wet-lab experimental data for biological activity evaluation of pepcans. (3) The chemical information of pepcans, dynamic interaction information with CB1 receptors, and ground truth data from wet-lab experiments are utilized for training MuMoPepcan, ultimately achieving a closed-loop pipeline that integrates both dry and wet experiments.

specific receptors in a relatively reasonable manner. On this basis, researchers further integrate machine learning methods to establish reliable mappings between receptor biological states and receptor conformational features or ligand-receptor binding modes [12,13]. These studies not only elucidate the mechanisms by which specific micro-structure influences overall receptor functionality but also provide new theoretical foundations and computational frameworks for drug development.

Cannabis has been widely used and increasingly legalized for medical and recreational purposes in many areas globally. Cannabis compounds, represented by Δ9-tetrahydrocannabinol, exert their neuroactive effects by acting on the cannabinoid receptors type 1 (CB1) and cannabinoid receptors type 2 (CB2) receptors of the endocannabinoid system. Both receptors belong to the GPCR family. These receptors are widely distributed in the nervous system, immune system, and various organs, including the liver, gastrointestinal tract, adipose tissue, and skin [14,15]. Among them, CB1 receptors play a crucial role in regulating neural functions and maintaining physiological homeostasis, influencing aspects such as pain perception, mood, memory, sleep, appetite, immune response, prenatal and postnatal development [16].

Pepcans are degradation products of the α chain of hemoglobin [17]. As an important component of the endogenous cannabis system, they exert analgesic effects by targeting the CB1 receptor. At present, more than 10 endogenous pepcan members have been identified [18]. Pepcan-9 (Hemopressin), the first discovered member of the pepcans, has been characterized as a CB1 inverse agonist [19]. It was reported to exert analgesic effects in a carrageenan-induced plantar pain sensitivity model; however, this analgesic property appears to be model-dependent. However, it has no analgesic effect in the spinal cord injury neuralgia model, the hot plate acute pain model, and the arthritis pain model caused by carrageenan, and even antagonizes the analgesic effect of 2-AG [20,21]. Pepcan-12 acts as a negative allosteric modulator of the CB1 receptor, reducing orthosteric agonist efficacy in cAMP and [^{35}S]GTPγS binding assays [17], while demonstrating agonist activity through stimulation of ERK1/2 phosphorylation and calcium release pathways [18], suggesting potential functional selectivity at CB1. It also exhibits positive allosteric modulation at the CB2 receptor, significantly potentiating the effects of CB2 agonists including 2-AG in [^{35}S]GTPγS binding and cAMP inhibition assays [22]. Research on the analgesic effects of pepcans is constrained by the complexity of its mechanisms, with the relationship between structural characteristics and functional outcomes remaining unclear, thereby hindering further development and practical applications.

In this study, we employed a sequential truncation strategy targeting pepcan-23 to construct a peptide library, while incorporating other minimal active fragments identified in previous studies [23,24]. Functional validations of pepcans were conducted both *in vivo* and *in vitro*. Specifically, a carrageenan-induced mouse inflammatory pain model was established to assess the antinociceptive effects of candidate pepcans *in vivo*. Additionally, their effects on the extracellular signal-regulated kinase 1/2 (ERK1/2) signaling pathway were examined in cell lines overexpressing the CB1 receptor. The results demonstrated that RD-pepcan-11 exhibited the most potent antinociceptive effects. To overcome the limitations of few-shot learning, this study conducted multiple 100 ns MDS with ligand-receptor interaction analysis, successfully constructing a conformation dataset, and expanding the scale from several dozen to millions. Following this, we leveraged deep learning technology to develop a novel deep learning model-MuMoPepcan, combining wet-dry lab data. In our pipeline (Fig. 1), (1) Through molecular docking and MDS, various CB1 receptor coordinates and interaction data were generated; (2) Pepcan library was synthesized. The biological effects were evaluated by *in vivo* and *in vitro*

methods; (3) Chemical information of pepcans, dynamic interaction information, and wet-lab experiment data as ground truth were used to train MuMoPepcan. Ultimately, our model can more accurately predict the biological activities of pepcans targeting the CB1 receptor and achieves comparable prediction error levels to wet-lab experiments.

2. Materials and methods

2.1. 3D modeling of protein and ligand

The 3D structure of the CB1 receptor was constructed using a multi-template comparative modeling strategy to integrate structural information from several high-resolution experimental structures and enhance model quality. Multiple experimental structures of the human CB1 receptor were retrieved from the RCSB database (https://www.rcsb.org), including 7WV9, 8IKG, 5 U09, 6KPG, 8WRZ, and 7FEE. These structures were selected to represent diverse conformational states and resolutions. They were manually preprocessed to ensure only the target protein sequence was retained and were then used as templates for integrated model building. Subsequently, multi-template homology modeling was performed using MODELLER software (version 10.5), followed by backbone refinement with ModRefiner [25]. To further improve model accuracy, we employed the Relax protocol of PyRosetta (version 2024.18) for fine-tuning the receptor conformation [26,27]. Notably, in the CB1 receptor structure, the N-terminal unstructured region obscures the conventional orthostatic pocket, posing technical challenges for peptide docking studies. Given this, and considering the intrinsically disordered nature and lack of resolvable density for the N-terminal region (residues 1–110) in available high-resolution structures, the sequence corresponding to amino acids 1–110 was excluded from the structure to ensure the validity of subsequent molecular interaction analyses focused on the orthosteric pocket. Regarding ligand preparation, we utilized PyRosetta to model and optimize the three-dimensional structures of peptide ligands. For small molecule ligands such as WIN55212–2, AEA, 2-AG, AM6538, and AM251, their precise three-dimensional structures were generated using the CORINA (https://demos.mn-am.com/corina_interactive.html) based on SMILES formats retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov).

2.2. Molecular docking

Molecular docking was performed using AutoDock Vina software (version 1.1.2), along with its Python scripts to prepare input files for the CB1 receptor and ligands [28]. Using AutoDock Tools (version 1.5.7), reasonable box dimensions were set in the orthosteric pocket region of the CB1 receptor to ensure coverage of potential binding sites [29]. In the molecular docking parameter settings, the exhaustiveness value was set to 8 to balance computational efficiency and result accuracy, while n_model was specified as 20 to guarantee sufficient sampling quantity. Each ligand underwent three independent molecular docking runs to enhance the reliability of the results. After completing all molecular docking runs, we conducted an integrated analysis of the results from multiple independent runs. First, PLIP (version 2.4.0) was employed for a systematic analysis of interaction patterns between ligands and receptors [30]. To further validate the accuracy of the results, PyMOL (version 3.0.3) was used to visually inspect and analyze interactions in the ligand-receptor complexes, ensuring the reliability of docking results and interaction analyses [31].

2.3. Molecular dynamics simulation

For structural preparation, PyMOL and Avogadro [32] (version 1.99.0) software were utilized. Peptide ligands were directly modeled as protein structures and engaged in docking with the CB1 receptor. In contrast, small molecule ligands were processed using the CGenFF web app (https://cgenff.com) to generate input files compatible with the GROMACS [33]. All simulations were conducted on the GROMACS software platform [34,35] (version 2023.3), utilizing LazyDock (https://github.com/BHM-Bob/LazyDock, version 0.13.0) for automated workflow construction. In the process of constructing the simulation system, the CB1 receptor and its ligand were first moved to the origin of a 3D coordinate system using Open Babel (version 3.1.0) [36]. The complex was then aligned along principal axes *via* LazyDock, and its minimum bounding box was calculated by compass (version 2.8.0) [37]. Based on this volume data, a safety padding of 1.2 nm was added to generate the final simulation box. For force field selection, the CHARMM36 force field (version jul2022) was uniformly employed to construct topology files for the CB1 receptor, peptide ligands, and small molecule ligands [38]. To fill the system box and ensure electro neutrality, a TIP3P water model was incorporated into the simulation system, with an appropriate amount of Cl^- ions added. During the energy minimization phase, the steepest descent method was applied to optimize the system for up to 6000 steps, with a maximum allowed energy set at 200 kJ/mol to maintain stability. Subsequently, the system underwent 100 ps of NVT isothermal equilibration at 300 K, followed by 100 ps of NPT isothermal-isobaric equilibration at 1 bar pressure. Finally, a total of 100 ns MDS were conducted on the completed system, with coordinate snapshots saved every 10 ps for subsequent analysis. To ensure the reliability and accuracy of simulation results, three independent MDS runs of 100 ns each are performed for every CB1 receptor or ligand-receptor system after constructing the simulation setup.

2.4. Chemicals

All peptides used in this study were synthesized by Solid Phase Peptide Synthesis with 2-chloro-triphenylmethyl chloride (2-CTC) resin. The crude peptides were isolated and purified by reversed-phase HPLC (LC2030, Shimadzu, Japan). Finally, the purity and molecular weight of the peptide were determined by reversed-phase high performance liquid chromatography and MS (LCMS-2020, Shimadzu, Japan). In short, the 2-CTC resin was first added into the reactor and oscillated with DCM for 30 min. The first amino acid (1.2 eq) was dissolved in DCM along with DIEA (1.2 eq), then added to the reactor and oscillated for 3 h. Subsequently, the system was conditioned with a mixture of MeOH and DIEA (3,1, *v*/v) for an additional 30 min. The resin was washed alternately with DCM and DMF 3 times, each wash lasting 3 min. Following this, the Fmoc protection groups were deprotected by adding a 20 % piperidine/DMF (*v*/v) solution, which was oscillated for 15 min and repeated twice. After washing with DMF 3 times (3 min per wash), a small resin sample was taken to monitor the deprotection reaction *via* Kaiser test. The remaining Fmoc-protected amino acids were sequentially coupled to the resin in the desired peptide sequence. Once all couplings were complete, the Fmoc protection groups were removed, and the peptidyl resin was washed three times with each of DMF, DCM, and MeOH for 3 min. The final purification step involved drying the peptidyl resin for 12 h and treating it with a cleavage mixture (TFA/EDT/TIS/water = 90:4:2:4, *v*/v) for 2 h. The crude peptide was purified using reversed-phase HPLC and analyzed by mass spectrometry. All peptides were obtained through lyophilization.

Pepcans and carrageenan used in animal experiments were dissolved in normal saline (NS), while those for cell experiments were dissolved in serum-free medium. WIN55212–2 was dissolved in DMSO and diluted with serum-free medium for cell experiments. For animal experiments, both WIN55212–2, AM251 and AM630 were dissolved in a mixture of DMSO/castor oil/NS (5/5/90). In the Western blot (WB) experiment, 5 % skimmed milk, antibodies including CB1 (cat: PA5–85080, Thermofisher, USA), H3 (cat: B1055, Biodragon, China), Phospho-p44/42 MAPK (cat:9101, Cell signaling, USA), p44/42 MAKP(cat:4695, Cell signaling, USA), and secondary antibodies (cat: BF03008, Biodragon, China; cat: BF03001, Biodragon, China) were all prepared with Tris-buffered saline (pH 7.4) with 0.05 % Tween-20.

2.5. Animals

Kunming mice (6–8 weeks old, weighing 25 ± 5 g, male) were used for behavioral experiments. All animals were purchased from the Experimental Animal Center of Lanzhou University. Each cage housed 3–5 mice, with free access to food and water. To ensure the reliability of the experiment, the environmental temperature was controlled at 22 ± 1 °C, maintaining a 12-h light-dark cycle. The animal ethics and experimental procedures strictly followed the National Institutes of Health's "guide for the care and use of laboratory animals" and were approved by the Lanzhou University Animal Experiment Ethics Committee (EAF2022012).

2.6. Intrathecal injection

The injection process was performed on awake mice according to the method described by Hvlden and Wilcox in 1980 [39,40]. A 26-gauge needle attached to a 10-μL Hamilton microliter syringe was inserted into the subdural space, and 5 μL of the drug was administered over 5 s. A light reflexive lateral tap on the mouse's tail or the formation of an "S" shape indicates a successful epidural puncture. After the injection was completed, the needle was left in place for 5 s and then gently rotated to remove it. AM251 and AM630 were injected intrathecally 5 min before the drug injection.

2.7. Carrageenan-induced inflammatory pain

The carrageenan-induced inflammatory pain model in mice was used to assess the anti-hyperalgesic effects of pepcans on chronic inflammatory pain [19]. The experimental procedure is as follows: 20 μL of 1 % carrageenan was injected into the plantar surface of the right hind paw of mice. Behavioral assessments were performed 6 h post-injection.

Mechanical sensitivity was assessed by von Frey filament, as previously reported [41]. Briefly, each mouse was placed in a Plexiglass chamber (dimensions: 10 × 10 × 15 cm) for 30 min to acclimate. After drug administration, paw withdrawal thresholds (PWT) were assessed at 0, 10, 20, 30, 45, and 60 min post-injection, and the area under the curve (AUC) was calculated to comprehensively evaluate the antinociceptive effects of the drugs. Followed by the measurement of baseline mechanical thresholds using von Frey filaments based on the classical "up-down method".

2.8. Construction of stable cell lines overexpressing CB1 receptor

To establish a stable cell line with overexpression of CB1 receptors, lentivirus infection was employed. HEK-293 T cells (cultured in the lab) were maintained in Dulbecco's modified eagle's medium (DMEM, cat: 12491023, Gibco, USA) supplemented with 10 % fetal bovine serum (FBS, cat: AB-FBS-00505, ABW, Uruguay) and 1 % penicillin-streptomycin (cat: 15140122, Gibco, USA). These cells were incubated at 37 °C under a humidified atmosphere containing 5 % CO_2. Similarly, CHO-K1 cells (also cultured in our lab) were maintained in Ham's F-12 K (Kaighn's) medium (F-12 K, cat: 21127022, Gibco, USA) supplemented with 10 % FBS and 1 % penicillin-streptomycin under identical incubation conditions.

The CB1 gene DNA fragment was amplified *via* PCR from the template plasmid (cat: P3163, MiaoLing Plasmid, China). The target fragment was then cloned into the vector plasmid (cat: VT8070, YouBio, China). Subsequently, the packaging plasmid psPAX2 and envelope plasmid pMD2.0 were transfected into HEK-293 T cells using transfection reagent (cat: T101, Vazyme, China). After 48 h, viral particles were harvested and used to infect CHO-K1 cells. To establish a stable cell line expressing CB1 receptors, antibiotic selection was performed using purinomycin (cat: CL13900, Ambole, USA) at a concentration of 8 μg/mL. Through limiting dilution, a monoclonal stable cell line expressing the CB1 receptor was successfully obtained. The sequences for each primer pair were as follows:

CB1: 5'-GCTCTAGAATGAAGTCGATCCTAGATGGCCTTGC-3'(Forward) /.

5'-AAGGAAAAAAGCGGCCGCTCACAGAGCCTCGGCAGACGTGTCT-3′ (Reverse).

2.9. Quantitative real-time PCR

When the monoclonal cell line reached approximately 80 % confluence, RNA was isolated using RANiso Plus Reagent (cat: 9108, Taraka, Japan). The extracted RNA was reverse-transcribed into cDNA following the protocols provided in the Takara Reverse Transcription Kit (cat: RR047A, Takara, Japan). Subsequently, the cDNA underwent fluorescent quantitative PCR Reactions according to the instructions outlined in the Takara Quantitative PCR Reagent Kit (cat: RR820A, Takara, Japan). The relative expression levels of CB1 mRNA were calculated using the $2^{-\Delta\Delta Ct}$ method. The sequences for each primer pair were as follows:

CB1: 5'-CTGTTCCTCACAGCCATCGACA-3'(Forward) /.

5'-GGCTATGGTCCACATCAGGCA-3′ (Reverse),

B2M: 5'-TGGCTCACACGGAGTTTACA-3'(Forward) /.

5'-CATGTCTCGTTCCCAGGTGA-3′ (Reverse),

2.10. Western blot analysis

After activating the CB1 receptors at the cellular level, which significantly promotes the phosphorylation of ERK1/2, WB analysis assay was employed to assess the agonistic activity of pepcans on CB1 receptors [19].

When the density of overexpressed CB1 cells reached approximately 50 %, they were incubated in a serum-free medium for 2 h. Subsequently, peptides, WIN55212–2, and negative control (serum-free medium) were added separately, followed by a 10-min incubation period. RIPA buffer (cat: R0010, Solarbio, China), Protease Inhibitor Cocktail (cat: M5293, Abmole, USA), and Phosphatase Inhibitor Cocktail (cat: M7528, Abmole, USA) were then added. The cells collected in tubes were left on ice for 20 min and centrifuged at 12,000 rpm for 10 min. The resulting total protein was separated by 10 % SDS-PAGE and transferred to PVDF membranes (cat: ipvh00010, Merck, Germany), which were then blocked by 5 % skimmed milk for 2 h. Subsequently, the membranes were incubated with primary antibodies at 4 °C overnight, followed by 1-h incubation with horseradish peroxidase (HRP)-conjugated secondary antibodies. The chemiluminescence signal was detected and recorded using ECL Western blotting Substrate (cat: PYT005–100, Yoche, China), and the grayscale values of protein bands were analyzed using Image-J (version 1.54f) and normalized (calculated using follow equation).

$$\text{Phosphorylation level} = \frac{A_1 - A_0}{A_{WIN} - A_0} \times \frac{B_{WIN} - B_0}{B_1 - B_0}$$

where: the character "A" represents p-ERK value, "B" represents ERK value; the subscript "1", "WIN", and "0" represent pepcans, WIN55212–2, negative control.

2.11. Data sets

This study constructed a multi-modal dataset that integrates peptide sequences with interaction data from MDS of complexes. To generate SMILES representations for peptides, we used the RDKit tool (version 2024.9.5) to convert amino acid sequences into chemical structure formats and parsed PDB files for small molecules to obtain their SMILES encodings. Subsequently, these SMILES strings were processed by a pretrained PepDoRA model [42], with the maximum length set to match the pepcan-23. This generated corresponding SMILES token sequences and their associated masks, serving as ligand representation features.

In terms of receptor feature extraction, we aligned the MDS trajectories of complexes based on the transmembrane helical region structure of the CB1 receptor and used the obtained 3D coordinate data as the receptor representation. For the selection of prediction target labels, this study adopted multi-level experimental data from wet-lab experiment: first, AUC values were used in a mouse analgesia model to evaluate analgesia efficacy; second, ERK and phospho-ERK (p-ERK) at three different concentrations were selected as wet-lab experiment validation indicators at the cellular level using normalized ratios. Additionally, we analyzed ligand-receptor interactions in complex MDS data using the PLIP software. We associated each of the 128 potential binding sites on the receptor with the presence or absence of interactions across frames, forming a complete MDS labeling system.

Since the peptides in C-pepcans were derived from the pepcan-9, their sequence mode shows significant differences from other pepcans. Therefore, all these samples were designated as a separate test set in this study. The remaining pepcans and small-molecule data were combined and divided into training and validation sets using a 5-fold cross-validation strategy. In terms of model evaluation: The validation set results served as the primary reference for assessing model prediction performance; The test set results provided the final reference for evaluating model generalization ability.

2.12. Single modal deep learning model

The advantage of single-modal single-task deep learning models lies in their ability to implement end-to-end prediction processes, thereby providing reliable tools for efficient and extensive candidate peptide screening. In this study, the numerical dataset comprised the multi-level experimental data described in the Data sets section (*in vivo* analgesia AUC and *in vitro* p-ERK/ERK ratios), which were used as target labels for model training. Based on the Transformer architecture, two innovative model structures were designed:

1. ML-Decoder-Based Architecture: ML-Decoder is a model proposed in 2021 that achieves SOTA performance in image multi-label tasks [43]. Considering the wet-lab experimental dataset in this study as essentially a numerical dataset, we developed an architecture by combining Transformer with ML-Decoder as the first candidate solution.

2. Prediction Token-Based Architecture: Based on the design concept of pre-trained feature inputs and leveraging ML-Decoder's characteristics as a tokenized predictor, this study further proposed a second candidate architecture that directly integrates prediction tokens with the Transformer for deep combination. This design not only simplifies the overall model structure but also retains the core advantages of the previous version.

For loss calculation, we used a below-noise strategy to augment the training labels and a gradient-scaled squared error loss to ensure stable convergence:

$$Y'_{true_{animal}} = Y_{true_{animal}} + 0.1 \times \varepsilon, \varepsilon \sim N(0,1)$$

$$Y'_{true_{cell}} = Y_{true_{cell}} + 0.01 \times \varepsilon, \varepsilon \sim N(0,1)$$

$$\text{Scaled MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(20 \times \left(y_{pred} - Y'_{true}\right)\right)^2$$

where n is the batch size, and the scale factor of 20 was chosen to amplify the gradient signal during backpropagation, which proved essential for effective optimization given the small numerical range of our target biological activity values.

2.13. Multimodal multitask deep learning model

Since the wet-lab experimental dataset used in this study only contains 45 peptide samples, a multimodal multitask prediction model was further developed to effectively control the risk of model overfitting. By incorporating MDS data into training, reasonable and simple data augmentation objectives can be achieved. In terms of processing and integrating ligand and receptor inputs, the following basic model architecture was designed:

1. Ligand Module: The Prediction Token-based architecture from the single-modal model was adopted.
2. Receptor Module: Inspired by the work of Pu, Xuemei, we desided to use Convolutional Neural Network (CNN) to process receptor coordinates. Considering that the CB1 receptor backbone contains 932 atoms, directly applying a Transformer model would result in significant computational costs. To address this, an architecture similar to CvT [44] was employed for processing receptor coordinate data, with optimizations made to key components: The standard CNN was replaced with multi-scale 1D convolution to more efficiently and rationally extract spatial features; The initial layers of the Transformer structure were removed to further reduce computational complexity.
3. Integration Module: Multi-head self-attention mechanisms were used to compute attention matrices for ligand and receptor information,

Table 1
Peptide information of pepcan library.

Series name	Name	Short name	Sequence
Pepcan	Pepcan-23	1–23	SALSDLHAHKLRVDPVNFKLLSH
	Pepcan-22	1–22	ALSDLHAHKLRVDPVNFKLLSH
	Pepcan-22	1–22	ALSDLHAHKLRVDPVNFKLLSH
	Pepcan-21	1–21	LSDLHAHKLRVDPVNFKLLSH
	Pepcan-20	1–20	SDLHAHKLRVDPVNFKLLSH
	Pepcan-19	1–19	DLHAHKLRVDPVNFKLLSH
	Pepcan-18	1–18	LHAHKLRVDPVNFKLLSH
	Pepcan-17	1–17	HAHKLRVDPVNFKLLSH
	Pepcan-16	1–16	AHKLRVDPVNFKLLSH
	Pepcan-15	1–15	HKLRVDPVNFKLLSH
	Pepcan-14	1–14	KLRVDPVNFKLLSH
	Pepcan-13	1–13	LRVDPVNFKLLSH
	Pepcan-12	1–12	RVDPVNFKLLSH
	Pepcan-11	1–11	VDPVNFKLLSH
	Pepcan-10	1–10	DPVNFKLLSH
	Pepcan-9	1–9	PVNFKLLSH
	Pepcan-8	1–8	VNFKLLSH
	Pepcan-7	1–7	NFKLLSH
	Pepcan-6	1–6	FKLLSH
	Pepcan-5	1–5	KLLSH
	Pepcan-4	1–4	LLSH
	Pepcan-3	1–3	LSH
RD-pepcan	RD-pepcan-22	2–22	SALSDLHAHKLRVDPVNFKLLS
	RD-pepcan-21	2–21	SALSDLHAHKLRVDPVNFKLL
	RD-pepcan-20	2–20	SALSDLHAHKLRVDPVNFKL
	RD-pepcan-19	2–19	SALSDLHAHKLRVDPVNFK
	RD-pepcan-18	2–18	SALSDLHAHKLRVDPVNF
	RD-pepcan-17	2–17	SALSDLHAHKLRVDPVN
	RD-pepcan-16	2–16	SALSDLHAHKLRVDPV
	RD-pepcan-15	2–15	SALSDLHAHKLRVDP
	RD-pepcan-14	2–14	SALSDLHAHKLRVD
	RD-pepcan-13	2–13	SALSDLHAHKLRV
	RD-pepcan-12	2–12	SALSDLHAHKLR
	RD-pepcan-11	2–11	SALSDLHAHKL
	RD-pepcan-10	2–10	SALSDLHAHK
	RD-pepcan-9	2–9	SALSDLHAH
	RD-pepcan-8	2–8	SALSDLHA
	RD-pepcan-7	2–7	SALSDLH
	RD-pepcan-6	2–6	SALSDL
	RD-pepcan-5	2–5	SALSD
	RD-pepcan-4	2–4	SALS
	RD-pepcan-3	2–3	SAL
C-pepcan	C-pepcan-7	3–7	PVNFKLL
	C-pepcan-6	3–6	PVNFKL
	C-pepcan-5	3–5	VNFKL
	C-pepcan-4	3–4	NFKL

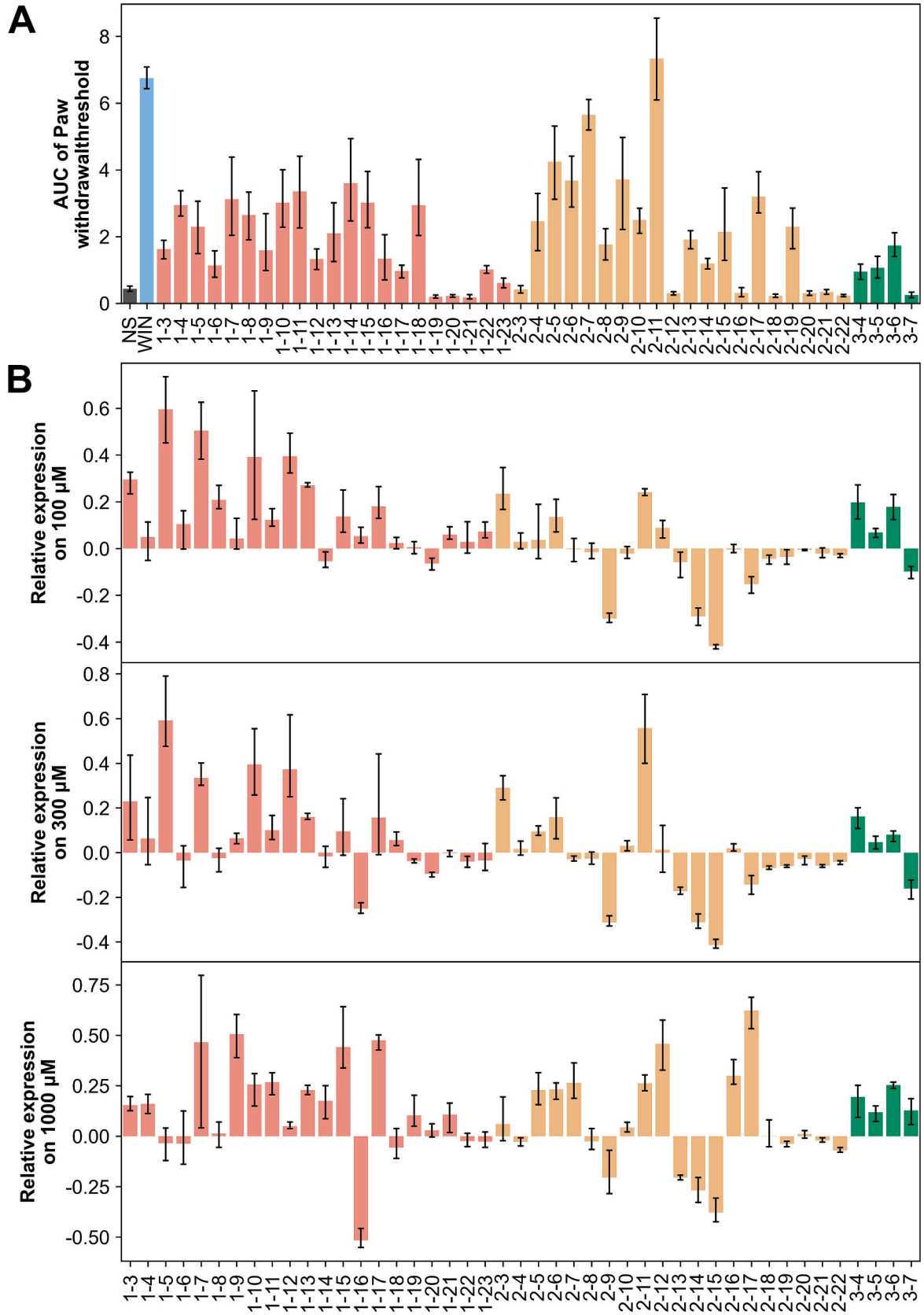

Fig. 2. Summary of wet-lab experiments (A) The area under the antinociceptive curves of chronic inflammatory pain models induced by carrageenan for each member of the pepcan family at the dose of 30 nmol per mouse. WIN55212–2 was administered at the dose of 7.5 nmol per mouse ($n = 6$). (B) WB detection of ERK1/2 phosphorylation levels in CB1-overexpressing cells treated with each pepcan family member at concentrations of 100 µM, 300 µM, and 1000 µM. Results were normalized based on negative and positive controls ($n = 3$). Data are presented as mean ± 1.96 × SE.

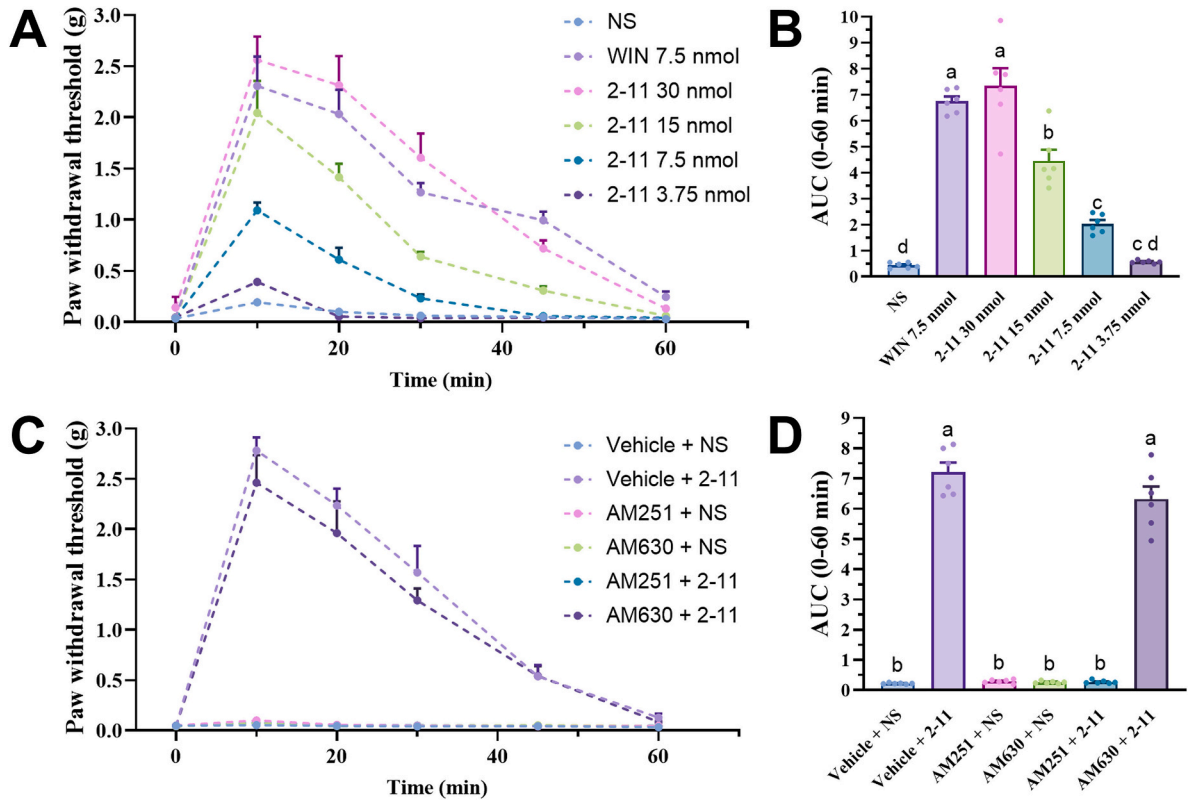

Fig. 3. Antinociceptive effects of RD-pepcan-11 at the spinal level in the carrageenan-induced inflammatory pain model and the impact of receptor antagonists. (A, B) Dose-dependent antinociceptive effects of RD-pepcan-11: (A) Dose-response curve; (B) Corresponding AUC analysis. (C, D) Blockade effects of the CB1 receptor antagonist AM251 (30 nmol per mouse, i.t.) and CB2 receptor antagonist AM630 (30 nmol per mouse, i.t.) on RD-pepcan-11 (30 nmol per mouse, i.t.): (C) Dose-response curve; (D) Corresponding AUC analysis. Data are presented as mean ± standard error of the mean (S.E.M.), $n = 6$ per group. Different lowercase letters indicate statistically significant differences ($P < 0.05$) as determined by one-way ANOVA with Tukey's test.

which were then applied to weight the features of ligand and receptor information *via* a sheared Transformer Decoder layers.

4. Wet Experiment Prediction Module: Four independent Feed-Forward Neural Networks were employed to predict the four wet-lab experiment values separately.
5. Interaction Prediction Module: A single Feed-Forward Neural Network was used to predict interaction labels.

For the loss calculation of wet-lab experimental data, we borrowed the same noise strategy and scaled MSE loss from the single modal. For loss calculation of interaction data, we use the bellow loss called Balanced Focal Loss to address the unbalanced label issue:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$w = \frac{n_{neg}}{n_{pos}}$$

$$BCE(x,y) = -(w \cdot y \log(\sigma(x)) + (1-y)\log(1-\sigma(x)))$$

$$p\_t = y \cdot \sigma(x) + (1-y)(1-\sigma(x))$$

$$Balanced\ Focal\ Loss = \alpha \cdot \frac{1}{n}\sum_{i=1}^{n} BCE(x_i, y_i) \cdot (1-p_t)^{\gamma}$$

where w is the weight of positive samples in training set, n is the batch size, x_i is the output of the model, y_i is the ground truth, α is the scale factor, γ is the factor of Focal Loss.

3. Results

3.1. Pepcan library

Starting with pepcan-23 as the parent peptide, this research progressively removed amino acids one by one from both the N- and C-terminal ends, down to tripeptides. Additionally, we collected different minimal active fragments identified in previously published studies [23,24] and integrated them into our experimental system. Ultimately, this process resulted in the construction of a peptide library comprising 45 peptides belonging to the pepcans (Table 1).

3.2. RD-pepcan-11 exhibits strongest Antinociceptive effects in the carrageenan model

Members of the pepcans initially garnered significant attention due to their ability to reverse plantar hyperalgesia caused by carrageenan through intrathecal administration [19]. Building on this research foundation, this study utilized the same animal experimental model to validate the biological activity of newly developed pepcans. Different concentrations of WIN55212–2 and the classical pepcan, pepcan-12, were initially tested in a carrageenan-induced plantar inflammatory pain relief assay (Fig. S1). The results indicated that the optimal intrathecal administration dose was 30 nmol per mouse. Subsequently, the AUC values for analgesia and mechanical pain threshold values at 10 min post-administration were statistically analyzed (Fig. 2A, Fig. S2). Based on AUC rankings, we selected the top 21 pepcans and WIN55212–2 and used the CB1 receptor-specific antagonist AM251 to verify that their targets are CB1 receptors (Fig. S3). The top five peptides demonstrating the most significant antinociceptive effects were: RD-

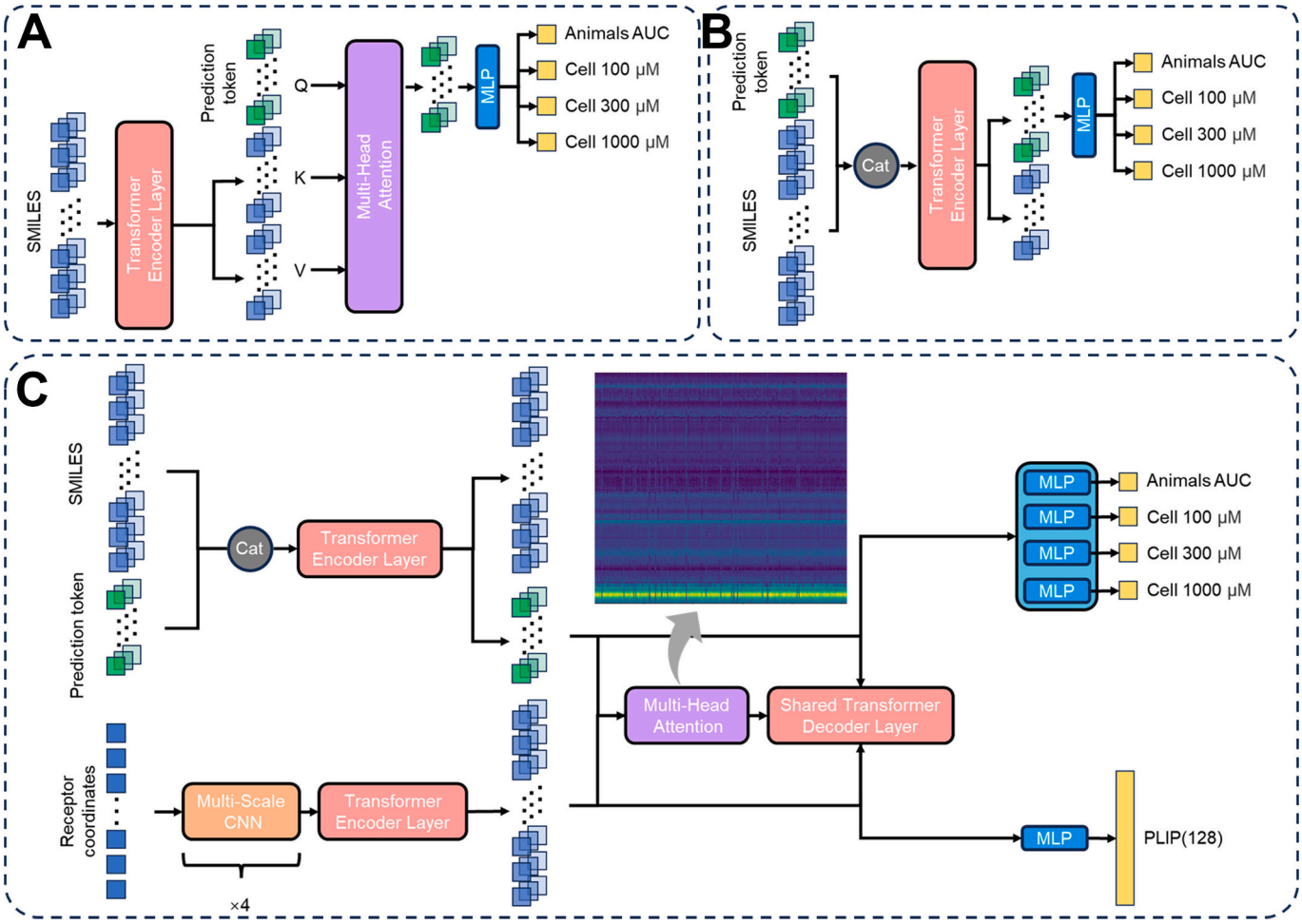

Fig. 4. Framework Diagram of MuMoPepcan. (A) Single modal candidate architecture 1 of MuMoPepcan. The prediction token serves as the query; SMILES features are processed through a single layer of Transformer encoder, serving as keys and values; Through multi-head attention, the query is projected *via* MLP to predict the four wet-lab experimental endpoints shown in Table 3. (B) Single modal candidate architecture 2 of MuMoPepcan. The prediction token and SMILES features are concatenated and processed through a Transformer encoder layer; The resulting prediction token is then projected *via* MLP to predict the four wet-lab experimental endpoints shown in Table 3. (C) Dual-Modal V3 Architecture of MuMoPepcan. Ligand Information Processing: The prediction token and SMILES features are concatenated and processed through a Transformer encoder layer; The resulting prediction token is used as the ligand information. Receptor Information Processing: Receptor coordinates undergo processing through four layers of multi-scale CNN, followed by a single layer of Transformer encoder, forming the receptor information; Attention mechanism and output projection: The ligand and receptor information are input into a multi-head attention module to compute the attention matrix. Both ligand and receptor data are then processed through the same layer of Transformer decoder with this attention matrix. Finally, *via* MLP projection, the model outputs predictions for the four wet-lab experimental endpoints and the interaction sites shown in Table 3.

pepcan-11, RD-pepcan-7, RD-pepcan-5, RD-pepcan-9, and RD-pepcan-6. In terms of mechanical pain threshold values, the best-performing peptides within 10 min post-administration were: RD-pepcan-11, pepcan-5, RD-pepcan-7, RD-pepcan-5, and pepcan-4.

Notably, RD-pepcan-11 ranked highest in both assessment metrics, showcasing outstanding antinociceptive effects and prolonged duration of action. This indicates that the peptide not only exhibits favorable effectiveness and robust initial activity but also demonstrates superior persistence. Furthermore, the RD-pepcan series-comprising RD-pepcan-7, RD-pepcan-5, RD-pepcan-9, and RD-pepcan-6-exhibited remarkable antinociceptive effects, combining high capability in increasing mechanical pain thresholds with prolonged duration of action. In contrast, while pepcan-5 and pepcan-4 also demonstrated significant increases in mechanical pain thresholds, their antinociceptive effects exhibited a more rapid decline over time. This observation underscores the importance of considering both efficacy strength and duration when developing therapeutic agents.

To evaluate the antinociceptive effects of RD-pepcan-11 in an inflammatory pain model, we administered four doses: 30 nmol, 15 nmol, 7.5 nmol, and 3.25 nmol per mouse (Fig. 3A, B). Results demonstrated that the AUC values for the 30 nmol, 15 nmol, and 7.5 nmol per mouse dose groups were significantly higher than those in the NS group. Notably, at the 30 nmol per mouse dose, the peak antinociceptive effects of RD-pepcan-11 and its corresponding AUC value slightly exceeded the levels observed with the positive control drug WIN55212–2 (7.5 nmol per mouse). Subsequently, to elucidate the receptor mechanisms underlying the *in vivo* antinociceptive effects mediated by RD-pepcan-11 at the spinal level, we conducted pharmacological blockade experiments using the CB1 receptor antagonist AM251 and the CB2 receptor antagonist AM630(Fig. 3C, D). Results indicated that the CB1 receptor antagonist AM251 almost completely blocked the antinociceptive effects of RD-pepcan-11. Conversely, the CB2 receptor antagonist AM630 did not significantly attenuate the antinociceptive effects of RD-pepcan-11. Therefore, the CB1 receptor plays a critical role in the antinociceptive effects exerted by RD-pepcan-11 at the spinal level, whereas the CB2 receptor does not appear to be significantly involved.

3.3. *Pepcan-CB1 receptor interaction exhibit high complexity*

Given the critical role of CB1 in pepcan-mediated antinociception at

Table 2
Comparison results of MuMoPepcan and other models on pepcan biology activity map data set. These results are all the best performance for the same model in the test set and validation set of 5-fold cross-validation. Performance values are presented as mean (std), $n = 5$.

Model	Modal	Validation set		Test set	
		Scaled MSE	R_p	MSE	R_p
SVM	SMILES	0.9050 (0.7536)	0.4548 (0.3754)	0.3922 (0.1478)	0.1404 (0.1794)
RF	SMILES	0.7393 (0.3991)	0.5387 (0.2182)	0.9224 (0.2312)	0.3315 (0.0989)
XGBoost	SMILES	0.7348 (0.5292)	0.5306 (0.2724)	0.8583 (0.5007)	−0.1475 (0.1987)
MuMoPepcan	SMILES(t)	0.4374 (0.2790)	0.5194 (0.2332)	1.0549 (0.2458)	−0.2119 (0.1281)
	SMILES(v)	1.4278 (0.8075)	0.3824 (0.1372)	**0.0710 (0.0114)**	−0.3294 (0.2182)
	SMILES+pose (t)	**0.0069 (0.0014)**	**0.9985 (0.0002)**	1.2783 (0.8353)	0.7761 (0.1258)
	SMILES+pose (v)	1.6090 (0.2113)	0.5726 (0.0922)	0.1085 (0.0136)	**0.8250 (0.0412)**

the spinal level, this study further investigated the interaction mechanisms between pepcans and CB1 receptors. It is established that following ligand binding, CB1 activate the downstream ERK1/2 signaling pathway, leading to increased phosphorylation levels of ERK1/2. Based on this mechanism, we employed WB to detect ERK1/2 phosphorylation in a CB1 receptor-overexpressing cell, aiming to evaluate the potential agonist activity of pepcan peptides on CB1 (Fig. S4). Experimental results revealed significant differences in the agonist activity of various peptides for the CB1 receptor under different dosages (Fig. 2B, Fig. S5). Specifically: at a concentration of 100 μM, peptides pepcan-5, pepcan-7, pepcan-12, pepcan-10, and pepcan-3 exhibited strong agonist activity toward the CB1 receptor; at 300 μM, peptides pepcan-5, RD-pepcan-11, pepcan-10, pepcan-12, and pepcan-7 showed higher agonist activity; when the dosage was increased to 1000 μM, RD-pepcan-17, pepcan-9, pepcan-17, pepcan-7, and RD-pepcan-12 demonstrated notable agonist effects on the CB1 receptor.

Notably, at the relatively lower concentrations tested in this assay (100 μM and 300 μM), pepcan-5 consistently demonstrated the highest agonist activity for the CB1 receptor. This finding aligns with animal experiments where intrathecal administration resulted in higher mechanical pain thresholds (indicating stronger antinociceptive effects) within 10 min post-administration. However, its performance in terms of antinociceptive effects AUC was relatively weaker, likely due to limitations in specificity and stability. Additionally, we found that RD-pepcan-9 consistently reduced the basal phosphorylation level of ERK1/2 under different concentrations, demonstrating inverse agonist activity toward the CB1 receptor. Despite its inverse agonistic properties, RD-pepcan-9 effectively alleviated pain symptoms in the carrageenan-induced plantar subcutaneous inflammatory model following intrathecal administration. At the concentration of 1000 μM, RD-pepcan-17, pepcan-9, pepcan-17, pepcan-7, and RD-pepcan-12 exhibited notable agonist effects on the CB1 receptor. These findings provide valuable references for future development of highly specific and stable CB1 receptor agonists.

3.4. *MuMoPepcan addresses few-shot learning challenge through multi-modal and tokenized architecture*

Given the lack of significant correlation between the antinociceptive effects of pepcans at the spinal level and there *in vitro* CB1 receptor agonist activity, this study employed deep learning approaches to further explore their structure-activity relationships (SAR). To this point, we developed a multi-level, multi-modal predictive model named MuMoPepcan. This model demonstrated predictive performance,

Table 3
Results of MuMoPepcan (single modal) with varying settings on the pepcan biology activity map data set. These results are all the best performance for the test set and validation set separately of 5-fold cross-validation. Performance values are presented as mean (std), $n = 5$.

Wet predictor	Test set		Validation set	
	Scaled MSE	R_p	Scaled MSE	R_p
ML-decoder	0.6996 (0.5003)	0.5326 (0.1782)	0.0783 (0.0359)	−0.2083 (0.2419)
PTA + MLP	**0.4374 (0.2790)**	**0.5194 (0.2332)**	0.0710 (0.0114)	−0.3294 (0.2182)
PTA + MCMLP	0.6278 (0.4348)	0.5756 (0.1856)	**0.0867 (0.0388)**	**0.1200 (0.2568)**

achieving accuracy comparable to or even surpassing *in silico* assays (intra-group error of 0.3, equivalent to a mean squared error of 0.09 after conversion), thereby providing a powerful computational tool for in-depth mechanistic analysis of pepcan's antinociceptive effects. The model can be adapted to different prediction tasks according to speed and accuracy requirements: The single-modal model (Fig. 4A, B) enabled rapid prediction screening requiring only ligand SMILES information; The dual-modal model (Fig. 4C), while requiring MDS data, significantly enhanced prediction accuracy (Table 2).

By incorporating MDS receptor conformational modalities, this study aims to leverage them for data augmentation effects and to constrain model learning of receptor conformation and ligand relationships through PLIP interaction site predictions. This design allows the dual-modal model to achieve excellent performance without requiring explicit alignment of modalities. Furthermore, traditional machine learning models based on SMILES (SVM, RF, and XGBoost) were trained, and their performance was evaluated using both validation and test sets. While these conventional methods achieved a good balance between the two datasets, their best performance remained significantly below that of MuMoPepcan.

3.5. *MuMoPepcan archived balanced performance on training and unseen dataset*

The peptides in C-pepcan series constructed in this study were derived from the sequence of pepcan-9 and exhibits significant differences compared to the pepcan-series and RD-pepcan-series. Therefore, we designated it as a separate validation set to evaluate the generalization performance of the MuMoPepcan model. To achieve an optimal balance between the test set and validation set, multiple optimization improvements were implemented for the base line model (Table 3). Experimental results indicate that while single-modal models achieved low prediction errors on the validation set, their correlation coefficient R_p values were relatively low or even negative, suggesting the limited applicability of this approach. Based on these findings, we further optimized the dual-modal model architecture, primarily adjusting the receptor atom selection criteria, MDS simulation time window, number of Transformer network layers, and predictor components (Table 4). Experimental results confirmed that reducing one layer of the Transformer structure significantly enhanced validation set performance while modifying the predictor architecture to a combination of Prediction Token and ML-Decoder achieved desirable effects on both the validation set and test set. Additionally, Table 1 lists the optimal model performance data for each fold during the 5-fold training process for versions v3 and v4, demonstrating the model's applicability.

3.6. *Interpretability research*

To explore the correlation between different ligands and receptor residue sites, this study extracted attention matrices from the MuMoPepcan model incorporating attention modules and averaged them

Table 4
Results of MuMoPepcan (dual modal) with varying settings on the pepcan biology activity map data set. These results are all the best performance for the test set and validation set separately in 5-fold training. Performance values are presented as mean (std), $n = 5$.

Name	Trajectory (ns)	Receptor	CNN	Wet predictor	Interaction predictor	Parameters	Test set		Validation set	
							Scaled MSE	R_p	Scaled MSE	R_p
Baseline	50–100	TM	1DCNN × 4 transformer × 2	MHMLP	MLP	9.34 M	0.0079 (0.0012)	0.9984 (0.0002)	0.6755 (0.1434)	0.7446 (0.1015)
v1	50–100	TM + LOOP	1DCNN × 4 transformer × 2	MHMLP	MLP	9.34 M	0.0100 (0.0034)	0.9979 (0.0008)	0.3928 (0.0831)	**0.8536 (0.0519)**
v2	0–50	TM	1DCNN × 4 transformer × 2	MHMLP	MLP	9.34 M	0.0088 (0.0035)	0.9983 (0.0006)	0.3327 (0.2887)	0.7838 (0.0926)
v3[aa]	50–100	TM	1DCNN × 4 transformer × 1	MHMLP	MLP	9.33 M	**0.0069 (0.0014)**	**0.9985 (0.0002)**	0.3783 (0.1763)	0.8010 (0.0692)
v4	50–100	TM	1DCNN × 4 transformer × 2	PTA	ML-decoder	9.80 M	0.1650 (0.0338)	0.9926 (0.0049)	**0.1085 (0.0136)**	0.8250 (0.0412)
v5	50–100	TM	2DCNN × 4 transformer × 2	MHMLP	MLP	10.56 M	0.0088 (0.0013)	0.9981 (0.0004)	0.3469 (0.1707)	0.7956 (0.1635)

[aa] v3 is the final version to be used in interpretability research.

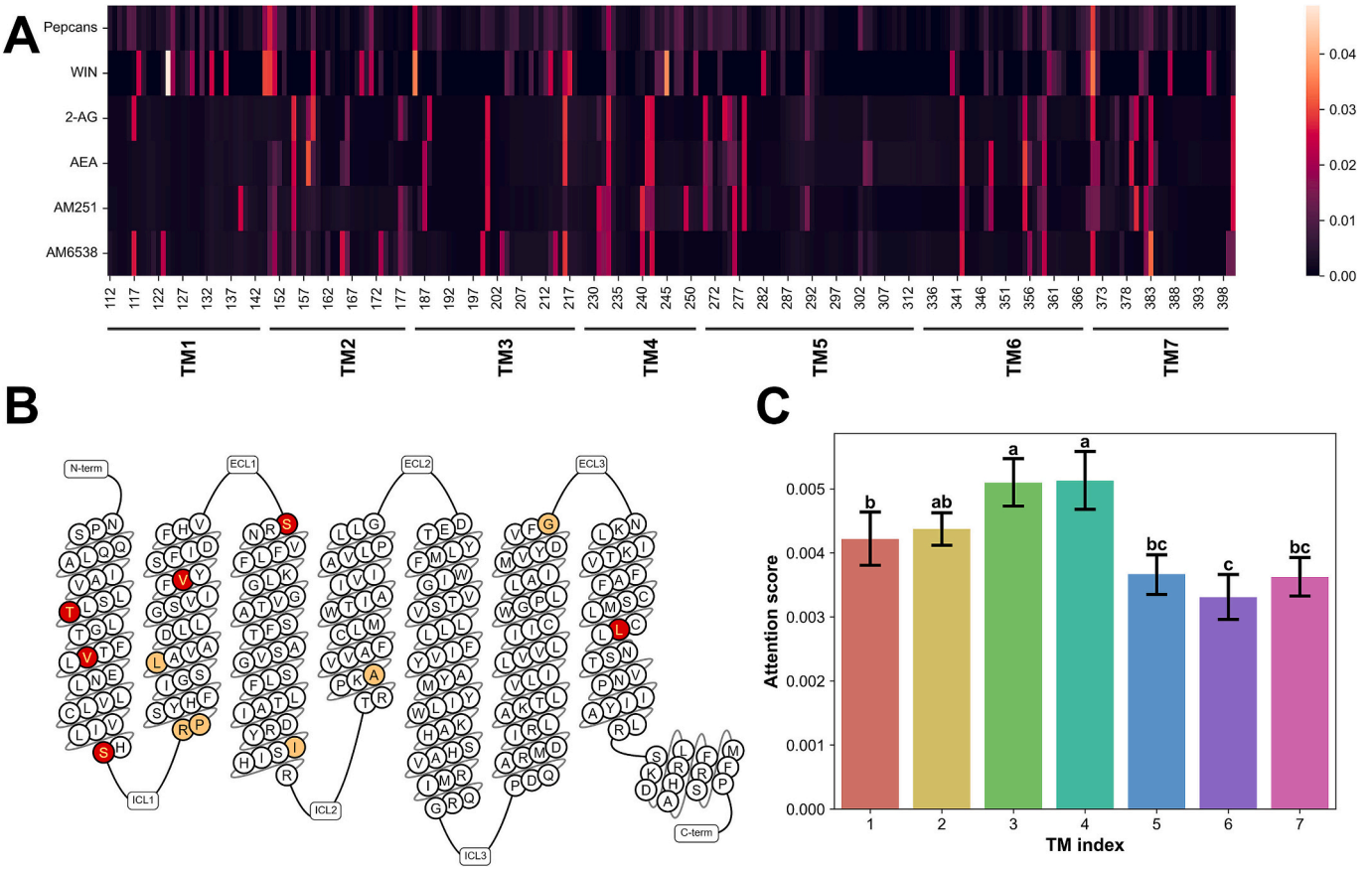

Fig. 5. Attention Visualization Analysis of the Dual-Modal MuMoPepcan Model. (A) Ligand Attention Heatmap: Each ligand's attention scores are averaged, followed by averaging across pepcan values. The x-axis represents residues in the CB1 receptor transmembrane helical region, labeled every five units. Horizontal lines on the TM index only mark annotated sites. (B) Distribution of Attention Scores in the Attention Matrix: Schematic illustrating site distributions where: Attention scores for pepcans and WIN55212–2 exceed 0.1 and attention scores for cannabinoids and antagonists are below 0.1. This Fig. provides a comprehensive visualization of attention mechanisms within the dual-modal MuMoPepcan model, highlighting key interactions and patterns across different ligand types. (C) Attention Scores Across Transmembrane Helices: Display of attention scores for all pepcans across various transmembrane helices. Different lowercase letters indicate statistically significant differences according to Tukey's test ($P < 0.05$).

across multiple samples for each ligand. This analysis generated heatmaps illustrating correlations between pepcans, their corresponding ligands, and individual amino acid residues on the receptor (Fig. 5A, B). The findings showed that WIN55212–2 and pepcans exhibited significant associations with residues such as THR125, VAL131, SER144, VAL171, SER185, and LEU387. Consequently, we propose that these sites may play a role in the unique activation mechanisms by which WIN55212–2 and pepcans exert their effects.

Furthermore, it was observed that neural networks paid significantly more attention to the N-terminal and C-terminal regions of transmembrane helices compared to their core regions (Fig. 5A). In GPCR receptors, these terminal regions are essential for ligand-specific recognition and regulation of downstream signaling pathways. They also serve as important spatial features in studying receptor

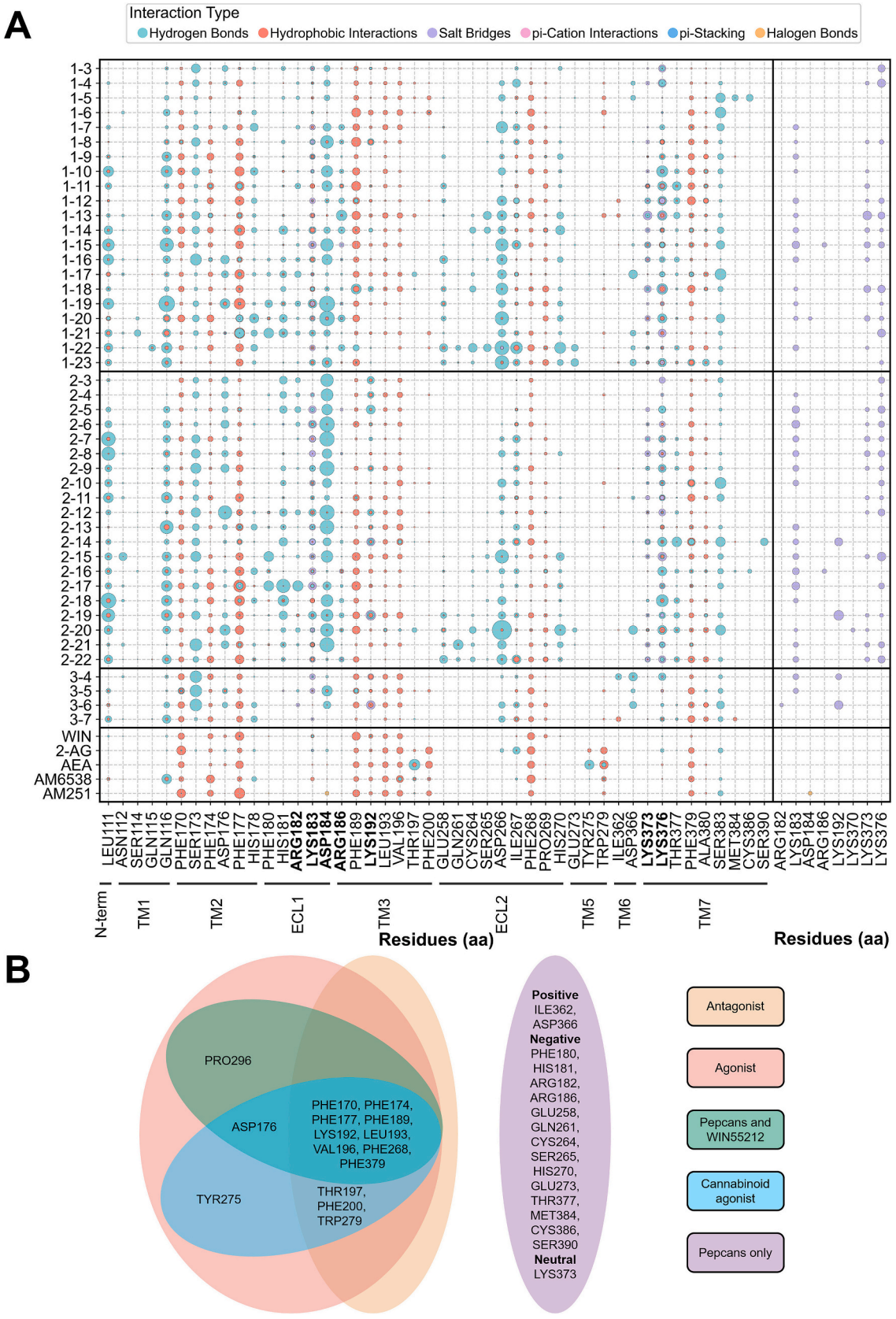

Fig. 6. Interaction sites in MDS. (A) Statistical diagram of critical binding sites for each ligand, where point colors represent different interaction types and point sizes indicate interaction strength (based on the frequency and bond length of receptor-ligand interactions within 100 ns, $n = 3$). (B) Venn diagram of receptor interaction site classifications, with different ellipses representing distinct categories of interaction sites.

conformational changes from active to inactive states. Notably, the networks displayed significant variations in attention levels across different transmembrane helices (Fig. 5C): higher attention was observed for TM3 and TM4, while lower attention was directed toward TM5, TM6, and TM7. This pattern likely reflects the distinct functional roles of these transmembrane helices in receptor activity and ligand binding mechanisms [45].

3.7. Unique binding sites were categorized through in silico and in vivo methods

To further validate whether the active residue sites identified in the MuMoPepcan interpretability experiments are relevant to the entire receptor and unique to the computational biology method, we conducted an in-depth analysis of the interactions between pepcans and CB1 receptors. Specifically, based on interaction patterns within the binding pockets, we systematically classified and investigated the binding sites of pepcans.

Based on the statistical analysis of interaction frequency, duration, and bond length between ligands and amino acid residues on the CB1 receptor during MDS, we conducted a classification study on key amino acid residues (Fig. 6). The first category sites (PHE170, PHE174, PHE177, PHE189, LYS192, LEU193, VAL196, PHE268, PHE379) displayed binding characteristics in all ligands, potentially playing a significant role in promoting stable binding between the receptor and ligand. Secondly, during the interaction analysis of agonists and antagonists, we identified the second category site (ASP176), which specifically binds to WIN55212–2, endocannabinoid agonists (such as AEA and 2-AG), and pepcans, but not to antagonists such as AM251 or AM6538. This finding suggests that ASP176 may play a critical role in receptor activation. Notably, the third category site PRO296 was only observed in WIN55212–2 and pepcan binding patterns, possibly serving as one of the key factors distinguishing these agonists' distinct biological effects. Furthermore, the fourth category sites (THR197, PHE200, TRP279) exhibited common interactions between cannabinoid agonists and antagonists, indicating their potential correlation with CB1 receptor activation mechanisms. Notably, the fifth category site TYR275 was uniquely identified in cannabinoid agonists, suggesting its unique role in the subtype-specific binding of this agonist class.

Finally, the last category is unique to pepcans. Based on the results of inflammatory pain analgesia experiments, these sites are categorized into three types: positive sites (ILE362, ASP366), negative sites (PHE180, HIS181, ARG182, ARG186, GLU258, GLN261, CYS264, SER265, HIS270, GLU273, THR377, MET384, CYS386, SER390), and neutral sites (LYS373).

4. Conclusion and future direction

For a long time, as an important component of the endogenous cannabinoid system, the interaction between pepcans and CB1 receptors has not been fully clarified, which has limited the development of peptide-based drugs targeting CB1 receptors. In this study, through *in vitro* and *in vivo* methods, we systematically elucidated for the first time the agonist activity profiles of CB1 receptor targets among members of the pepcans, as well as their inflammatory pain antinociceptive effects profiles, identifying a new pepcan member RD-pepcan-11 with the best inflammatory pain antinociceptive effects. Using molecular docking and MDS research methods, this study further discovered that the binding pockets of CB1 receptors by pepcans and WIN55212–2 exhibited significant differences from those of cannabinoids, which may lead to their unique modes of action. Furthermore, the distinct signaling profile observed, particularly ERK1/2 activation, suggests potential biased agonism properties that warrant further investigation in future studies to fully characterize their signaling selectivity over other pathways such as β-arrestin recruitment.

Deep learning, particularly multi-modal models, is increasingly becoming a powerful tool for integrating experimental assay data to advance biological predictions [46,47]. Building on this foundation, this study constructed a deep learning model that successfully learned SAR underlying the antinociceptive effects of pepcans, demonstrating robust predictive accuracy. To overcome limitations faced by small-scale bioactivity datasets, we further developed the MuMoPepcan model, which innovatively integrates biological assays with computational simulation data. By incorporating receptor conformational information obtained through MDS (data augmentation) and introducing interaction site prediction constraints, this model effectively enriched learning modalities and guided model optimization. As a result, MuMoPepcan achieved an error level identical to that within the wet-lab experimental group. This precision is sufficient for large-scale and efficient virtual screening of peptide drugs targeting CB1 receptors, making it a powerful tool in CADD. However, it must be acknowledged that the prediction errors for *in vivo* analgesia AUC increased significantly for unknown sequences, ranging from 0.3 to 1.2. Given that the AUC for high-efficacy compounds exceeds 6.0, this level of error remains acceptable for the screening of novel analgesic peptides. There are also opportunities for improvement with MuMoPepcan. Many studies have successfully employed pre-training, adversarial learning, and other techniques to enhance the performance of deep learning models on unseen data [48,49].

This study has systematically elucidated, for the first time, SAR underlying the inflammatory pain antinociceptive effects of the pepcans and their agonist activity profiles targeting CB1, identifying RD-pepcan-11 as a lead analgesic compound. By integrating biological assays with MDS data, the constructed multi-modal MuMoPepcan model achieved high-precision predictions. The integrated computational-experimental framework established in this study provides a transferable paradigm for innovative drug design targeting other therapeutic targets. The multi-modal strategy of MuMoPepcan offers a generalizable blueprint, potentially extendable to small molecule discovery and the prediction of detailed functional profiles like biased signaling, as demonstrated for kinases [50]. Furthermore, our approach could be adapted to identify novel binders from diverse starting points, such as perturbing protein-protein interactions [51], broadening its application to challenging therapeutic targets.

Associated Content

Supplementary Fig.s, characterization of synthesized pepcans by ESI-MS and analytical RP-HPLC, original images of WB (PDF).

CRediT authorship contribution statement

Hongyang Man: Writing – original draft, Methodology, Investigation, Conceptualization. **Huiming Bao:** Validation, Investigation, Formal analysis, Data curation. **Zhanyu Niu:** Methodology, Data curation. **Zhonghua Zhang:** Writing – review & editing, Visualization, Conceptualization. **Jerine Peter Simon:** Data curation. **Tong Yang:** Data curation. **Pengtao Li:** Visualization. **Shouliang Dong:** Writing – review & editing, Software, Resources, Project administration, Funding acquisition.

Funding

This work was supported by the National Natural Science Foundation of China (32271307), the Natural Science Foundation of Gansu Province (24JRRA411, 24JRRA312), and the Fundamental Research Funds for the Central Universities (lzujbky-2022-it23).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgment

The authors want to thank School of Life Sciences for the FUSION FX 6 chemiluminescence imaging system and the IX53 inverted fluorescence microscope.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bioorg.2025.109027.

Data availability

The data processing, model training, and attention extraction code are available at https://github.com/BHM-Bob/MuMoPepcan.

References

[1] X. Chen, K. Wang, J. Chen, C. Wu, J. Mao, Y. Song, Y. Liu, Z. Shao, X. Pu, Integrative residue-intuitive machine learning and MD approach to unveil allosteric site and mechanism for β2AR, Nat. Commun. 15 (1) (2024) 8130, https://doi.org/10.1038/s41467-024-52399-y.

[2] S. Xu, L. Shen, M. Zhang, C. Jiang, X. Zhang, Y. Xu, J. Liu, X. Liu, Surface-based multimodal protein–ligand binding affinity prediction, BIOINFORMATICS 40 (7) (2024) btae413, https://doi.org/10.1093/bioinformatics/btae413.

[3] Q. Hong, G. Zhou, Y. Qin, J. Shen, H. Li, Q. Hong, SadNet: a novel multimodal fusion network for protein–ligand binding affinity prediction, Phys. Chem. Chem. Phys. 26 (16) (2024) 12880–12891, https://doi.org/10.1039/d3cp05664c.

[4] P.Y. Libouban, C. Parisel, M. Song, S. Aci-Sèche, J.C. Gómez-Tamayo, G. Tresadern, P. Bonnet, Spatio-temporal learning from molecular dynamics simulations for protein-ligand binding affinity prediction, BIOINFORMATICS 41 (8) (2025) btaf429, https://doi.org/10.1093/bioinformatics/btaf429.

[5] Z. Yang, W. Zhong, Q. Lv, T. Dong, G. Chen, C.Y.-C. Chen, Interaction-based inductive Bias in graph neural networks: enhancing protein-ligand binding affinity predictions from 3D structures, IEEE Trans. Pattern Anal. Mach. Intell. (2024) 1–18, https://doi.org/10.1109/TPAMI.2024.3400515.

[6] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings, JCIM 61 (8) (2021) 3891–3898, https://doi.org/10.1021/acs.jcim.1c00203.

[7] H. Wang, F. Hetzer, W. Huang, Q. Qu, J. Meyerowitz, J. Kaindl, H. Hübner, G. Skiniotis, B.K. Kobilka, P. Gmeiner, Structure-based evolution of G protein-biased μ-opioid receptor agonists, Angew. Chem. Int. Ed. 61 (26) (2022) e202200269, https://doi.org/10.1002/ange.202200269.

[8] M. Masureel, Y. Zou, L.-P. Picard, E. van der Westhuizen, J.P. Mahoney, J.P.G.L. M. Rodrigues, T.J. Mildorf, R.O. Dror, D.E. Shaw, M. Bouvier, et al., Structural insights into binding specificity, efficacy and Bias of a β2AR partial agonist, Nat. Chem. Biol. 14 (11) (2018) 1059–1066, https://doi.org/10.1038/s41589-018-0145-x.

[9] F. Zhang, Y. Yuan, Y. Chen, J. Chen, Y. Guo, X. Pu, F. Zhang, Molecular insights into the allosteric coupling mechanism between an agonist and two different transducers for μ-opioid receptors, Phys. Chem. Chem. Phys. 24 (9) (2022) 5282–5293, https://doi.org/10.1039/d1cp05736g.

[10] W. Huang, A. Manglik, A.J. Venkatakrishnan, T. Laeremans, E.N. Feinberg, A. L. Sanborn, H.E. Kato, K.E. Livingston, T.S. Thorsen, R.C. Kling, et al., Structural insights into μ-opioid receptor activation, Nat 524 (7565) (2015) 315–321, https://doi.org/10.1038/nature14886.

[11] R.O. Dror, D.H. Arlow, P. Maragakis, T.J. Mildorf, A.C. Pan, H. Xu, D.W. Borhani, D.E. Shaw, Activation mechanism of the β2-adrenergic receptor, PNAS 108 (46) (2011) 18684–18689, https://doi.org/10.1073/pnas.1110499108.

[12] P. Mollaei, A. Barati Farimani, Activity map and transition pathways of G protein-coupled receptor revealed by machine learning, JCIM 63 (8) (2023) 2296–2304, https://doi.org/10.1021/acs.jcim.3c00032.

[13] I. Buyanov, P. Popov, Characterizing conformational states in GPCR structures using machine learning, Sci. Rep. 14 (1) (2024) 1098, https://doi.org/10.1038/s41598-023-47698-1.

[14] H.C. Lu, K. Mackie, Review of the endocannabinoid system, BIOL PSYCHIAT-COGN N 6 (6) (2021) 607–615, https://doi.org/10.1016/j.bpsc.2020.07.016.

[15] K.A. Sharkey, J.W. Wiley, The role of the endocannabinoid system in the brain-gut Axis, GASTROENTEROLOGY 151 (2) (2016) 252–266, https://doi.org/10.1053/j.gastro.2016.04.015.

[16] A. Kilaru, K.D. Chapman, The endocannabinoid system, Essays Biochem. 64 (3) (2020) 485–499, https://doi.org/10.1042/EBC20190086.

[17] M. Bauer, A. Chicca, M. Tamborrini, D. Eisen, R. Lerner, B. Lutz, O. Poetz, G. Pluschke, J. Gertsch, Identification and quantification of a new family of peptide endocannabinoids (Pepcans) showing negative allosteric modulation at CB1 receptors, J. Biol. Chem. 287 (44) (2012) 36944–36967, https://doi.org/10.1074/jbc.M112.382481.

[18] I. Gomes, J.S. Grushko, U. Golebiewska, S. Hoogendoorn, A. Gupta, A.S. Heimann, E.S. Ferro, S. Scarlata, L.D. Fricker, L.A. Devi, Novel endogenous peptide agonists of cannabinoid receptors, FASEB J. 23 (9) (2009) 3020–3029, https://doi.org/10.1096/fj.09-132142.

[19] A.S. Heimann, I. Gomes, C.S. Dale, R.L. Pagano, A. Gupta, L.L. de Souza, A. D. Luchessi, L.M. Castro, R. Giorgi, V. Rioli, et al., Hemopressin is an inverse agonist of CB1 cannabinoid receptors, PNAS 104 (51) (2007) 20588–20593, https://doi.org/10.1073/pnas.0706980105.

[20] A. Hama, J. Sagen, Centrally mediated Antinociceptive effects of cannabinoid receptor ligands in rat models of nociception, Pharmacol. Biochem. Be. 100 (2) (2011) 340–346, https://doi.org/10.1016/j.pbb.2011.09.004.

[21] Z. Petrovszki, G. Kovacs, C. Tömböly, G. Benedek, G. Horvath, The effects of peptide and lipid endocannabinoids on arthritic pain at the spinal level, Anesth. Analg. 114 (6) (2012) 1346–1352, https://doi.org/10.1213/ANE.0b013e31824c4eeb.

[22] V. Petrucci, A. Chicca, S. Glasmacher, J. Paloczi, Z. Cao, P. Pacher, J. Gertsch, Pepcan-12 (RVD-hemopressin) is a CB2 receptor positive allosteric modulator constitutively secreted by adrenals and in liver upon tissue damage, Sci. Rep. 7 (1) (2017) 9560–9567, https://doi.org/10.1038/s41598-017-09808-8.

[23] C.B. de Araujo, A.S. Heimann, R.A. Remer, L.C. Russo, A. Colquhoun, F.L. Forti, E. S. Ferro, Intracellular peptides in cell biology and pharmacology, BIOMOLECULES 9 (4) (2019) 150, https://doi.org/10.3390/biom9040150.

[24] C.S. Dale, L. Pagano Rde, V. Rioli, S. Hyslop, R. Giorgi, E.S. Ferro, Antinociceptive action of Hemopressin in experimental hyperalgesia, PEPTIDES 26 (3) (2005) 431–436, https://doi.org/10.1016/j.peptides.2004.10.026.

[25] D. Xu, Y. Zhang, Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization, BJ 101 (10) (2011) 2525–2534, https://doi.org/10.1016/j.bpj.2011.10.024.

[26] S. Chaudhury, S. Lyskov, J.J. Gray, PyRosetta: a script-based Interface for implementing molecular modeling algorithms using Rosetta, BIOINFORMATICS 26 (5) (2010) 689–691, https://doi.org/10.1093/bioinformatics/btq007.

[27] X. Kong, Y. Jia, W. Huang, Y. Liu, Full-atom Peptide Design with Geometric Latent Diffusion. arXiv:2402.13555, **2024**, https://doi.org/10.48550/arXiv.2402.13555.

[28] S. Forli, R. Huey, M.E. Pique, M.F. Sanner, D.S. Goodsell, A.J. Olson, Computational protein–ligand docking and virtual drug screening with the AutoDock suite, Nat. Protoc. 11 (5) (2016) 905–919, https://doi.org/10.1038/nprot.2016.051.

[29] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (16) (2009) 2785–2791, https://doi.org/10.1002/jcc.21256.

[30] M.F. Adasme, K.L. Linnemann, S.N. Bolz, F. Kaiser, S. Salentin, V.J. Haupt, M. Schroeder, PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA, Nucleic Acids Res. 49 (W1) (2021) W530–W534, https://doi.org/10.1093/nar/gkab294.

[31] L.L.C. Schrodinger, The PyMOL molecular graphics system, Version 3.1.4.1. https://pymol.org, 2015.

[32] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, J. Chemother. 4 (1) (2012) 17, https://doi.org/10.1186/1758-2946-4-17.

[33] K. Vanommeslaeghe, A.D. MacKerell, Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing, JCIM 52 (12) (2012) 3144–3154, https://doi.org/10.1021/ci300363c.

[34] M. Abraham, A. Alekseenko, B. Andrews, V. Basov, P. Bauer, H. Bird, E. Briand, A. Brown, M. Doijade, G. Fiorin, et al., GROMACS 2025.2 Manual, Zenodo (2025), https://doi.org/10.5281/zenodo.14025865.

[35] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1-2 (2015) 19–25, 10.1016j.softx.2015.06.001.

[36] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open babel: an open chemical toolbox, J. Chemother. 3 (2011) 33, https://doi.org/10.1186/1758-2946-3-33.

[37] T.V. Mele, many, o, COMPAS: A Framework for Computational Research in Architecture and Structures, 2017, https://doi.org/10.5281/zenodo.2594510.

[38] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (4) (2010) 671–690, https://doi.org/10.1002/jcc.21367.

[39] J.L.K. Hylden, G.L. Wilcox, Intrathecal morphine in mice: a new technique, Eur. J. Pharmacol. 67 (2–3) (1980) 313–316, https://doi.org/10.1016/0014-2999(80)90515-4.

[40] C.A. Fairbanks, Spinal delivery of analgesics in experimental models of pain and analgesia, Adv. Drug Deliv. Rev. 55 (8) (2003) 1007–1041, https://doi.org/10.1016/S0169-409X(03)00101-7.

[41] C. He, X. Wang, J. Zhang, H. Wang, Y. Zhao, J.N. Shah, C. Ma, H. Li, W. Su, Z. Zhang, et al., MCRT, a multifunctional ligand of opioid and neuropeptide FF receptors, attenuates neuropathic pain in spared nerve injury model, Basic Clin. Pharmacol. Toxicol. 128 (6) (2021) 731–740, https://doi.org/10.1111/bcpt.13566.

[42] L. Wang, R. Pulugurta, P. Vure, Y. Zhang, A. Pal, P. Chatterjee, PepDoRA: A Unified Peptide Language Model via Weight-decomposed Low-rank Adaptation. arXiv: 2410.20667, **2024**, https://doi.org/10.48550/arXiv.2410.20667.

[43] T. Ridnik, G. Sharir, A. Ben-Cohen, E. Ben-Baruch, A. Noy, ML-decoder: scalable and versatile classification head, IEEE (2021) 32–41, https://doi.org/10.48550/arXiv.2111.12933.
[44] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, L. Zhang, CvT: Introducing Convolutions to Vision Transformers. arXiv:*2103.15808*, **2021**, https://doi.org/10.48550/arXiv.2103.15808.
[45] A.S. Hauser, A.J. Kooistra, C. Munk, F.M. Heydenreich, D.B. Veprintsev, M. Bouvier, M.M. Babu, D.E. Gloriam, GPCR activation mechanisms across classes and macro/microscales, Nat. Struct. Mol. Biol. 28 (11) (2021) 879–888, https://doi.org/10.1038/s41594-021-00674-7.
[46] Y. Jiang, B. Zhao, X. Wang, B. Tang, H. Peng, Z. Luo, Y. Shen, Z. Wang, Z. Jiang, J. Wang, et al., UKB-MDRMF: a multi-disease risk and multimorbidity framework based on UK biobank data, Nat. Commun. 16 (1) (2025) 3767, https://doi.org/10.1038/s41467-025-58724-3.
[47] H. Cui, A. Tejada-Lapuerta, M. Brbić, J. Saez-Rodriguez, S. Cristea, H. Goodarzi, M. Lotfollahi, F.J. Theis, B. Wang, Towards multimodal foundation models in molecular cell biology, Nat 640 (8059) (2025) 623–633, https://doi.org/10.1038/s41586-025-08710-y.
[48] S. Jin, X. Liu, J. Xu, S. Yuan, H. Xiang, L. Shen, C. Li, Z. Niu, Y. Jiang, Adaptive symmetry-based adversarial perturbation augmentation for molecular graph representations with dual-fusion attention information, INFORM FUSION. 120 (2025) 103062, https://doi.org/10.1016/i.inffus.2025.103062.
[49] H. Cai, Z. Zhang, M. Wang, B. Zhong, Q. Li, Y. Zhong, Y. Wu, T. Ying, J. Tang, Pretrainable geometric graph neural network for antibody affinity maturation, Nat. Commun. 15 (1) (2024) 7785, https://doi.org/10.1038/s41467-024-51563-8.
[50] E. Frasnetti, I. Cucchi, S. Pavoni, F. Frigerio, F. Cinquini, S.A. Serapian, L. F. Pavarino, G. Colombo, Integrating molecular dynamics and machine learning algorithms to predict the functional profile of kinase ligands, JCTC 20 (20) (2024) 9209–9229, https://doi.org/10.1021/acs.jctc.4c01097.
[51] A. Paladino, M.R. Woodford, S.J. Backe, R.A. Sager, P. Kancherla, M.A. Daneshvar, V.Z. Chen, D. Bourboulia, E.F. Ahanin, C. Prodromou, et al., Chemical perturbation of oncogenic protein folding: from the prediction of locally unstable structures to the Design of Disruptors of Hsp90-client interactions, Chem. Eur. J. 26 (43) (2020) 9459–9465, https://doi.org/10.1002/chem.202000615.