Acta Pharmaceutica Sinica B 2026;16(7):4147−4165

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

Deep learning-driven discovery and mechanism of action study of a minimalist conopeptide targeting α7 nicotinic acetylcholine receptor

Jinghui Zhang[a,b,†], Zhengji Yin[a,b,†], Yue Li[c,†], Cheng Ge[a,b,†], Zixuan Zhang[a,b], Pu Yuan[a,b], Tao Jiang[a,b], David J. Craik[d], Yan Zhao[c,*], Rilei Yu[a,b,*]

[a]*Key Laboratory of Marine Drugs, Chinese Ministry of Education, School of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China*
[b]*Laboratory for Marine Drugs and Bioproducts, Qingdao Marine Science and Technology Center, Qingdao 266237, China*
[c]*National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China*
[d]*Institute for Molecular Bioscience, Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, The University of Queensland, Brisbane, Queensland 4072, Australia*

Received 25 May 2025; received in revised form 30 July 2025; accepted 27 August 2025

KEY WORDS

Deep learning;
Conopeptides;
α7 nicotinic acetylcholine receptor;
Structure−activity relationship studies;
Cryo-EM;
Computational modeling;
Structure optimization;
Molecular dynamic simulations

Abstract Despite extensive structural and functional characterization of the α7 nicotinic acetylcholine receptor, valuable structural insights into its interactions with conopeptides remain limited, thereby hindering the rational development of peptide-based modulators for this clinically important receptor subtype. Here, we present an integrated pipeline combining deep learning, structural biology, computational modeling and electrophysiology to accelerate the discovery and optimization of α7 nAChR-targeting conopeptides. To overcome data scarcity, we developed a deep learning model using the ESM-2 protein language framework, enabling efficient screening of 689 disulfide-poor conopeptides. This approach identified SS1, a novel antagonist of α7 nAChR, which was systematically optimized *via* structure−activity relationship studies to yield [ΔQP,S8R]SS1—a minimalist peptide with nanomolar potency (IC_{50} = 49.2 nmol/L), enhanced selectivity, and improved stability. Cryo-EM and computational modeling resolved the 3.3 Å resolution structure of α7 nAChR bound to [S8R]SS1, revealing a unique binding mode stabilized by hydrogen bonds, hydrophobic interactions, and glycan contacts, while hybrid

This article is part of special issue entitled Machine Learning in Drug Discovery.

*Corresponding authors.
E-mail addresses: zhaoy@ibp.ac.cn (Yan Zhao), ryu@ouc.edu.cn (Rilei Yu).
†These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.12.035

receptor conformations (closed/desensitized) elucidated its inhibitory mechanism. This work establishes a transformative deep learning-to-experiment framework for accelerating the discovery and optimization of nature-inspired peptide therapeutics.

1. Introduction

Marine cone snails produce a diverse array of bioactive conopeptides, comprising both disulfide-rich conotoxins and non-disulfide-rich variants, that represent valuable resources for drug development or the design of neurochemical tools[1,2]. While modern venomics approaches have dramatically expanded conopeptide discovery, yet their corresponding pharmacological characterization has significantly lagged behind[3-5]. This disparity highlights an urgent need for innovative strategies to bridge the gap between peptide identification and pharmacological validation.

Traditionally, the identification of functionally specific conopeptides has relied on random screening through labor-intensive electrophysiological experiments. Some existing high-throughput screening methods developed recent years, including comprehensive venomics and phage display technologies, have significantly advanced the pharmacological development of bioactive peptides[6-9]. Nevertheless, developing rapid, high-throughput screening methods without the need for initial wet lab experiments remains highly attractive for the rapid characterization of conopeptide functions. Recent studies highlight the potential of deep learning (DL) approaches for mining functional drugs from genomics and transcriptomics data[10,11]. Furthermore, some language models—including PepMLM, Contrastive Language Modeling, and SaLT&PepPr—have achieved significant breakthroughs in various peptide binder design[12-15]. Meanwhile, DL-based strategies have been employed for the prediction and design of conotoxins[16-18]. These findings motivated us to utilize DL and language models to discover peptides with desired functions.

In addition to DL and language models, particularly using AlphaFold3 (AF3), and several advanced techniques, such as cryo-electron microscopy (cryo-EM) and molecular dynamic (MD) simulations, have significantly advanced the determination of receptor−peptide complex structures for drug development. Cryo-EM is particularly valuable for analyzing large macromolecular proteins as it enables atomic-level determination of receptor−ligand interactions[19]. MD simulations provide insights into the movements of atoms within a molecular system over time, based on physical models of interatomic interactions, which are vital for understanding the interaction modes between ligands and receptors[20]. Moreover, AF3, the most advanced model with diffusion-based architecture, has been demonstrated to perform structure prediction of biomolecular interactions with unprecedented accuracy[21]. These methods, combined with chemical mutagenesis studies, provide robust approaches for the structural determination of receptor-ligand complexes.

In particular relevance to the current study, a series of conopeptides targeting α7 nAChR have been identified (Table 1)[22-37]. However, existing α7-targeting conopeptides are limited by relatively low potency, overly long sequences, and a lack of diversity. For instance, αD-VxXXB from *Conus vexillum*, existing as an 11 kDa homo-dimeric protein that is a potent α7 nAChR inhibitor with a half maximal inhibitory concentration (IC_{50}) of 0.4 nmol/L[37]. Czon1107, a short 10-amino acid conopeptide with a single disulfide, only weakly inhibits the α7 nAChR, with an IC_{50} of 77.1 μmol/L[22]. Furthermore, many of the conopeptides antagonizing α7 nAChR discovered so far are similar in that they contain four cysteine residues and share similar sequences and the same disulfide bond framework. For example, conotoxins ImI, ImII and RgIA share the same 4/3 disulfide framework with a 'globular' (*i.e.*, Cys^{I} to Cys^{III} and Cys^{II} to Cys^{IV}) disulfide connectivity pattern[24,25], while Mr1.1, RegIIA and CIB have a 4/7 disulfide framework[28,30,38]. Despite extensive SAR investigations and structural optimization efforts focused on these classic α-conotoxins in recent years, no α-conotoxin-based drugs have yet been approved for clinical use. Meanwhile, this structural similarity inevitably restricts the diversity of peptide ligands acting on α7 nAChRs, thereby limiting their potential as drug leads or neurochemical tools. Therefore, there is an urgent need to discover potent α7-targeting conopeptides with new disulfide frameworks and shorter sequences.

In this study we developed an integrated pipeline (Fig. 1) combining DL, AF3, cryo-EM, MD simulations, chemical modification and electrophysiology to discover and optimize conopeptides targeting the α7 nicotinic acetylcholine receptor (nAChR)—a therapeutically relevant target for neurodegenerative diseases and cancer[39-42]. Using our BERT-based ESM-2[43] screening model with dual classification/regression heads, we identified SS1 from 689 non-disulfide-rich conopeptides. Subsequent structure−activity relationship studies and rational optimization yielded [S8R]SS1, whose mechanism was elucidated through cryo-EM (3.3 Å resolution) and computational modeling. This workflow serves as a generalizable template for natural peptide drug discovery elucidating the mechanism of action of structurally flexible peptides.

2. Materials and methods

2.1. Dataset preparation

For the training set of the classification model, we collected a dataset of conopeptides from UniProt with length less than 100 amino acids. All peptides in UniProt contain functional descriptions, though not all include experimental IC_{50} values. In light of this situation, during the model construction process, we attempted to filter for $IC_{50} < 10$ μmol/L. However, this significantly reduced the number of positive samples, leading to suboptimal results. As an alternative, we extracted peptide sequences from the positive samples in UniProt that contained functional descriptions such as "bind" or "inhibit" and "nicotinic acetylcholine receptors". Non-nicotinic acetylcholine receptor binders

Table 1 Representative conopeptides targeting α7 nAChR.

Conopeptide	Sequence	Size (aa)[a]	No. Cys[b]	IC_{50} (nmol/L)[c]	Ref.
Czon1107	GFRSPCPPFC[d]	10	2	77100 (h)	22
Conorfamide-As1a	RIKKPIFAFPRF[d]	12	0	736.5 (h)	23
ImI	GCCSDPRCAWRC[d]	12	4	595.0 (h)	24
ImII	ACCSDRRCRWRC[d]	12	4	571.0 (h)	24
RgIA	GCCSDPRCRYRCR	13	4	4600.0 (r)	25
BuIA	GCCSTPPCAVLYC[d]	13	4	272.0 (r)	26
LvIB	QCCSNPPCAHEHCR	14	4	1760.0 (r)	27
Mr1.1	GCCSHPACSVNNPDIC[d]	16	4	259.1 (h)	28
EpI	GCCSDPRCNMNNPD(sTy)C[d]	16	4	30.0 (r)	29
CIB	GCCSNPVCHLEHPNAC[d]	16	4	1500.0 (r)	30
RegIIA	GCCSHPACNVNNPHIC[d]	17	4	210.0 (h)	31
AnIB	GGCCSHPACAANNQD(sTy)C[d]	17	4	76.0 (r)	32
ArIB	DECCSNPACRVNNPHVCRRR	20	4	1.8 (r)	33
SIIID	CCGEGSSCPKYFKNNFICGCC	21	6	880.7 (h)	34
GeXIVA	TCRSSGRYCRSPYDRRRRYCRRITDACV	28	4	660.0(r)	35
GeXXA	PCQSVRPGRVWGKCCLTRLCSTMCCARADCTCVY HTWRGHGCSCVM	46	10	210.0 (h)	36
VxXXB	DD(Gla)S(Gla)CIINTRDSPWGRCCRTRMCGSMCCPRNGCT CVYHWRRGHGCSCPG[e]	100	20	0.4 (r)	37

[a]The amino acids sequence length of conopeptides.
[b]The number of cysteine residues in the sequence.
[c]The half maximal inhibitory concentration of conopeptides targeting α7 nAChR, h and r represent human and rat origin, respectively.
[d]C-terminal amidation, sTy = sulfotyrosine, and Gla = gamma-carboxylic glutamic acid.
[e]Exists in dimer form through interchain disulfide bonds.

were manually excluded. Using the CD-HIT software (CD-HIT, UCSD, La Jolla, CA, USA), we removed redundant sequences and obtained a total of 166 peptides that target nAChRs, which served as our positive samples (Supporting Information File S1). Additionally, we randomly selected 166 non-redundant peptides that without reported activities against nAChRs, as negative samples (Supporting Information File S2). All peptides in the positive and negative samples are composed solely of natural amino acids. The sequence length distribution of the dataset is shown in Supporting Information Fig. S1. The training and test sets were randomly assigned in a ratio of 8:2.

For the training set of the regression model, we collected a total of 93 conopeptides targeting the α7 nAChR from the ConoServer database, all of which have specific potencies for the α7 nAChR (Supporting Information File S3). Due to the wide range of potencies, we applied a logarithmic transformation to these values. The training and test sets were randomly assigned in a ratio of 9:1.

For the disulfide-poor conopeptides dataset, we retrieved 117 relevant articles from PubMed using 'conus Transcriptome' as the keyword (Supporting Information File S4). From these articles, we identified numerous sequences and further curated using the ConoPrec tool, which extracts conotoxins from precursor proteins. In total, we obtained 689 disulfide-poor conopeptides containing either one or no disulfide bonds (Supporting Information File S5). The source of each peptide is detailed in Supporting Information File S6. This disulfide-poor dataset was used to identify conopeptides targeting the α7 nAChR.

2.2. *Evaluation metrics*

The classification model was evaluated with several metrics, including accuracy (Acc), sensitivity (Sens), precision (Pre), specificity (Spe), and Mathew correlation coefficient (MCC). The specific calculation process was shown in Eqs. (1)−(5):

$$\mathbf{Acc} = \frac{\mathbf{TN} + \mathbf{TP}}{\mathbf{TN} + \mathbf{TP} + \mathbf{FN} + \mathbf{FP}} \tag{1}$$

$$\mathbf{Sens} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}} \tag{2}$$

$$\mathbf{Spec} = \frac{\mathbf{TN}}{\mathbf{TN} + \mathbf{FP}} \tag{3}$$

$$\mathbf{Prec} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}} \tag{4}$$

$$\mathbf{MCC} = \frac{(\mathbf{TP} \times \mathbf{TN}) - (\mathbf{FP} \times \mathbf{FN})}{\sqrt{(\mathbf{TP} + \mathbf{FP})(\mathbf{TP} + \mathbf{FN})(\mathbf{TN} + \mathbf{FP})(\mathbf{TN} + \mathbf{FN})}} \tag{5}$$

The regression model utilizes Spearman's rank correlation coefficient (Spearman's ρ) and Root Mean Square Error (RMSE) as evaluation metrics, which were defined in Eqs. (6) and (7):

$$\boldsymbol{\rho} = \mathbf{1} - \frac{\mathbf{6}\sum \boldsymbol{d}_i^2}{\boldsymbol{n}(\boldsymbol{n}^2 - \mathbf{1})} \tag{6}$$

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{i=1}^{n}(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)^2} \tag{7}$$

Spearman's ρ is a non-parametric measure of rank correlation that assesses the strength and direction of association between two ranked variables. RMSE is a commonly used metric to measure the difference between values predicted by a model and the actual observed values.

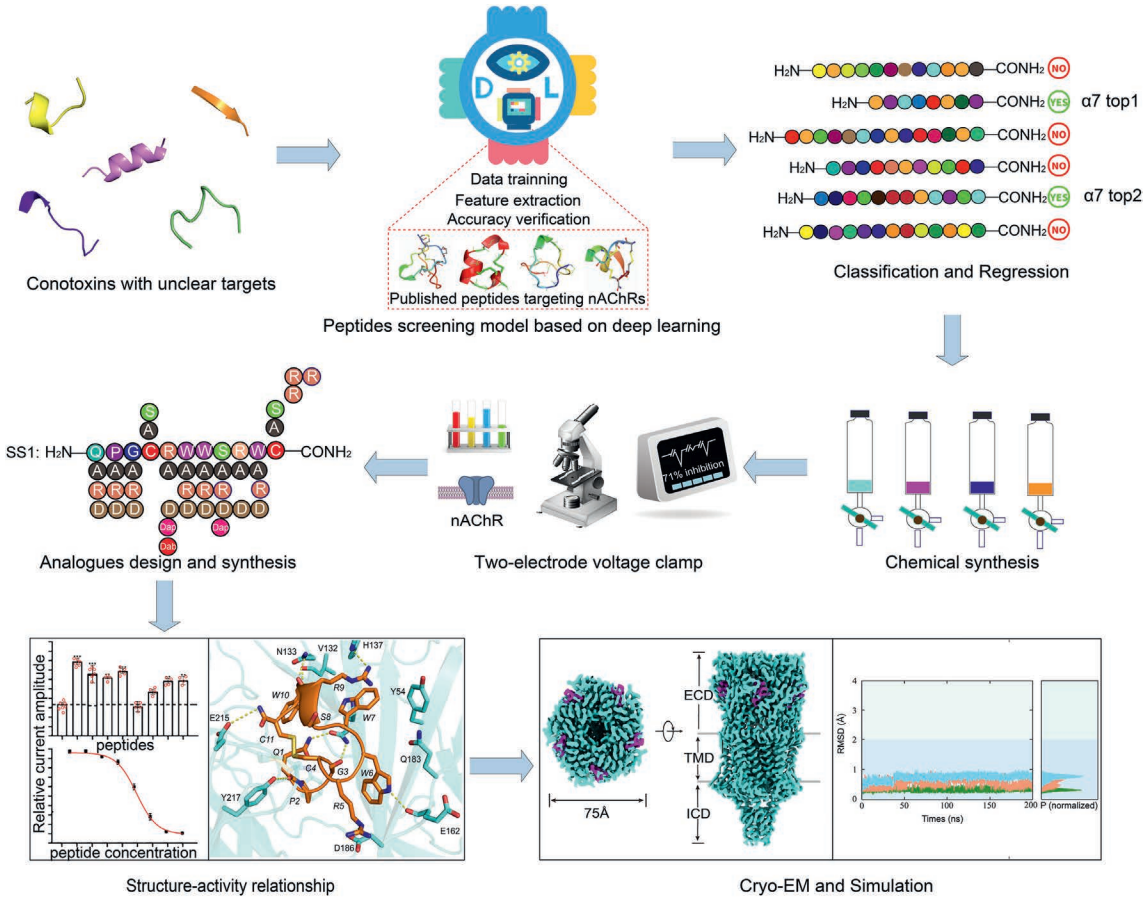

Figure 1 Research protocol. In this study, peptides targeting the nAChRs were collected and used to train a binary classification model, Additionally, a database from ConoServer was constructed and used to train a regression model targeting the α7 nAChR, which was then applied to screen disulfide-poor conopeptides with unclear targets. Then chemical synthesis and two-electrode voltage clamp experiments of candidate conopeptides were performed in the wet laboratory. Next, site-directed mutagenesis and MD simulations were performed to elucidate the structure-activity relationship and identify analogues with improved activity and structural stability. Furthermore, the binding mode of the most potent analogue with α7 nAChR was revealed through the combination of cryo-EM and computational modeling.

2.3. *Synthesis of candidate peptides and analogues of SS1*

All peptides were synthesized using manual solid-phase peptide synthesis (SPPS) with Fmoc chemistry, as detailed in previous studies[44]. The amino acids and resins used for peptide synthesis were purchased from Nanjing Peptide Biotech Co., Ltd. (Nanjing, China), and other chemicals were obtained from Sinopharm Group Co., Ltd. (Beijing, China). The peptides synthesized in this study comprised two types: those with C-terminal amidation and those with a C-terminal carboxylic acid. For C-terminal amidated peptides, Rink Amide (RAM) resin was used, while Wang resin preloaded with Fmoc-amino acid was employed for peptides with a C-terminal carboxylic acid (Nanjing Peptide Biotech Co., Ltd., Nanjing, China). SS1-1 was used as a representative peptide to illustrate the synthetic procedure. SS1-1 was synthesized with Wang resin (0.372 mmol/g loading) preloaded with Fmoc-Cys(Trt)-OH. The resin (0.1 mmol) was swollen in *N*,*N*′-dimethylformamide (DMF) for 60 min. Fmoc protecting groups were removed with 20% piperidine in DMF for 30 min. During coupling, Fmoc-protected amino acids (4 eq.) were activated by 4 eq. HCTU and 8 eq. *N*,*N*-diisopropylethylamine (DIPEA) in DMF for 60 min at room temperature. The resin was thoroughly washed with DMF and dichloromethane (DCM) and deprotected with 20% piperidine in DMF. This cycle was repeated until the final amino acid was deprotected. Peptides were cleaved from the resin using a TFA cleavage solution (TFA/H_2O/TIPS, 90:5:5) for 3 h at 25 °C. The product was filtered into cold diethyl ether, centrifuged at 5000 rpm for 3 min, then dissolved in 10% acetonitrile aqueous solution, adjusted to pH 8−10 with ammonium bicarbonate. The solution was stirred in air for 48 h for oxidation, then purified by reverse-phase high-pressure liquid chromatography (RP-HPLC, Hanbang Biotech Co., Ltd., Huanan, China). The products were combined and lyophilized. Purity and molecular weight were confirmed using HPLC (RP-HPLC, Waters Corporation, Milford, MA, USA) and ESI-MS (Waters Corporation, USA). All peptides used in this study were synthesized at least 95% purity.

2.4. *Xenopus oocytes preparation and microinjections*

Experimental procedures were approved by the Scientific Ethics Review Board of the Qingdao Marine Biomedical Research Institute (approval no. E-MBWLS-2024-3). All animal housing and experiments were carried out in strict accordance with the institutional guidelines for the care and use of laboratory animals. Oocytes were harvested from *Xenopus laevis* female clawed frogs

after anesthesia and washed twice in Ca^{2+}-free OR2 solution (82.5 mmol/L NaCl, 2.5 mmol/L KCl, 1 mmol/L $MgCl_2$, 5 mmol/L HEPES, pH 7.4). Oocytes were transferred to approximately 15 mL tubes and treated with 2 mg/mL collagenase (Sigma type II, Sigma−Aldrich Inc., St Louis, MO, USA) in OR2 solution for 20 min at 20−25 °C with gentle rotation. Stage V or VI oocytes were selected for microinjections. For *in vitro* cRNA synthesis, capped cRNAs were transcribed *in vitro* using the mMACHINE™ T3 Kit (Ambion, Austin, TX, USA) following the linearization of plasmids in pBluescript KSM vectors. Oocytes were injected with 25 nL of cRNA solution, containing approximately 20 ng of cRNA for human $\alpha7^{WT}$, $\alpha7^{EM}$, or WT α7 mutants nAChRs, and 5 ng of cRNA for h$\alpha3\beta2$, h$\alpha3\beta4$, h$\alpha4\beta2$, and h$\alpha4\beta4$ nAChRs. cRNA using a microinjector (Drummond Scientific, Broomall, PA, USA). Oocytes were kept at 16 °C in ND96 solution (96 mmol/L NaCl, 2 mmol/L KCl, 1.8 mmol/L $CaCl_2$, 1 mmol/L $MgCl_2$, 5 mmol/L HEPES, pH 7.4 adjusted with NaOH).

2.5. Two-electrode voltage clamp recording

Two-electrode voltage clamp recordings were made 2−5 days post-injection. Oocytes were impaled with two microelectrodes (0.5−1.0 MΩ) filled with 3 mol/L KCl in a 40-μL recording chamber. The membrane potential was held at −80 mV using standard voltage clamp procedures. Currents were recorded in ND96 solution at room temperature (22 ± 1 °C) using a HEKA-iTEV90 amplifier and PatchMaster software (HEKA Instruments, Germany). Initially, oocytes were briefly washed with ND96 solution, followed by three applications of ACh: 300 μmol/L for h$\alpha3\beta4$ and h$\alpha7$ nAChR, 3 μmol/L for h$\alpha4\beta2$, 6 μmol/L for h$\alpha3\beta2$ and h$\alpha4\beta4$. Washout with ND96 solution was done for 3 min between ACh applications. Oocytes were incubated with peptides for 5 min with the perfusion system turned off, followed by co-application of 300 μmol/L ACh and peptide with flowing ND96 solution. All peptide solutions were prepared in ND96 solution containing 0.1% bovine serum albumin. Peak current amplitudes before and after peptide incubation were measured using Igor Pro software (Wave-metrics), where the ratio of ACh + peptide-evoked current amplitude to ACh alone-evoked current amplitude was used to assess the potency of the peptides at the h$\alpha7$ nAChR.

2.6. Cryo-EM data acquisition and processing

Cryo-EM data were acquired on a 200-kV ThermoFisher Titan Krios microscope (ThermoFisher Scientific, Waltham, MA, USA) equipped with a Gatan K2 Summit direct electron detector (Gatan, Pleasanton, CA, USA), using EPU software (Thermo Fisher Scientific, Waltham, MA, USA) for automated collection. A total of 690 movie stacks were recorded at a magnification of 130,000×, yielding a pixel size of 0.96 Å. The nominal defocus range was set between −1.0 and −2.0 μm. Each movie received a total electron dose of 50 $e^-/Å^2$, delivered at a dose rate of approximately 9 $e^-/Å^2/s$, and was fractionated into 32 frames.

All subsequent processing was performed within the CryoSPARC v4.2.1 (Structural Biotechnology Inc., Canada). Movies were pre-processed with Patch Motion Correction, followed by contrast transfer function (CTF) estimation with Patch CTF Estimation. Particles were selected using the blob picker tool, with diameters ranging from 80 to 180 Å. These particles were then extracted into 256 × 256-pixel boxes and subjected to multiple rounds of 2D and 3D classification to eliminate suboptimal particles. The final map was generated through non-uniform refinement and CTF refinement, achieving a resolution of 3.3 Å as determined by the gold-standard Fourier shell correlation (GSFSC) criterion.

2.7. Peptide structure generation and complex modeling

Peptide structures were generated using AlphaFold with the amino acid sequences as input on locally deployed AF2 software. C-terminal amidation of peptides was achieved using PyMOL. The sequence for human α7 nAChR was obtained from UniProt (https://www.uniprot.org/) under entry Q05941. The human α7/SS1 and α7/[S8R]SS1 complexes were predicted using the locally deployed AF3 Program. For the human $\alpha7^{EM}$/[S8R]SS1 model, the structure was constructed by superimposing the ligand from the AF3-predicted human α7/[S8R]SS1 complex onto the cryo-EM receptor using PyMOL.

2.8. Molecular dynamic simulations

The 3D structure of the peptide or complex was solvated in a 10 Å TIP3P water box. Then Na^+ and Cl^- ions were introduced to maintain system neutrality, at a concentration of 0.15 mol/L. MD simulations were performed using AMBER 20 (AMBER, University of California, San Francisco, CA, USA) with ff19SB force field for protein, Lipid17 force field for lipid, and GLYCAM_06j-1 force field for carbohydrate. After setting up the system, optimization was conducted with a solute constraining force of 100 kcal/(mol·$Å^2$), initially applying a 2000-step steepest descent minimization and subsequently a 3000-step conjugate gradient minimization. Upon completion of the solute restrained optimization, the positional constraints were withdrawn, and the whole system was minimized using the same parameters as above again. The entire system underwent a 100 ps heating process under isobaric conditions to incrementally raise the temperature from 50 K to 300 K, with a solute constraints force set to 5 kcal/(mol·$Å^2$). For refinement of $\alpha7^{EM}$/[S8R]SS1 model, a harmonic restraining force of 1 kcal/(mol·$Å^2$) was imposed only on the receptor, while all restraints on the [S8R]SS1 peptide were released during the 200-ns production phase. The trajectory of the MD simulations was analyzed based on the root-mean-square deviation (RMSD) of the backbone using the Visual Molecular Dynamics software. The binding patterns of complexes were analyzed using PyMOL software.

2.9. Circular dichroism

A JASCO J-810 spectropolarimeter (JASCO Corporation, Easton, MD, USA) was employed to characterize the secondary structure of synthetic peptides. Peptides were dissolved in pure water to a final 0.2 mg/mL concentration. Molar ellipticity [θ] values at wavelengths between 180 and 250 nm were measured using a 1.0 mm path length cell, with a bandwidth of 1.0 nm and a response time of 2 s at room temperature under nitrogen protection conditions. The sample was scanned in triplicate. The spectra were presented in terms of corresponding molar ellipticity [θ], calculated as $[\theta] = 1000 \cdot mdeg/(l \cdot c)$. In this equation, *mdeg* represents the raw CD data, c is the peptide molar concentration, and l is the optical diameter.

2.10. Peptide stability

All experiments were conducted in accordance with the guidelines of the Scientific Ethics Review Board of Qingdao Marine Biomedical Research Institute (E-MBWM-2024-14). The stability

of the peptide in water was assessed at a 1 mmol/L aqueous solution. The peptide was prepared in water and maintained at a temperature of 25 °C. Samples were taken at various intervals over 14 days and analyzed using high-performance liquid chromatography (RP-HPLC, Waters Corporation, USA) to monitor the production of impurities. For serum stability, the peptide was first dissolved in saline to create a 10 mmol/L stock solution. It was then diluted at a 1:9 ratio into human serum and maintained at 37 °C. Samples were collected at different time points and analyzed by HPLC to assess the peak areas of the peptides. Each experiment was conducted in triplicate independently.

2.11. Statistical analysis

All electrophysiological recordings were performed at room temperature and all electrophysiological data were analyzed with Igor Pro (Wave-metrics), Origin 8 (OriginLab) and GraphPad Prism (GraphPad Software, La Jolla, CA, USA). All electrophysiological data were pooled ($n \geq 5$) and represented as means ± standard error of the mean (SEM). Data sets were compared using one-way ANOVA. Differences were regarded as statistically significant when $P < 0.05$. The IC_{50} was determined from concentration-response curves fitted to a non-linear regression function and reported with the error of the fit.

3. Results

3.1. Building the deep learning-aided peptides screening model

A major challenge in developing our model was the scarcity of peptides known to target the α7 nAChR. To compensate for this data limitation and solve the homology issue among nAChR subtypes, we developed a binary classification model predicting peptide targeting of nAChRs using a broader sequence dataset. Additionally, a regression model was created to estimate the potency of peptides targeting the α7 nAChR, enhancing model accuracy and subtype recognition. The workflow of the DL-assisted peptide screening model is depicted in Fig. 2. Our earlier research demonstrated that models pretrained on large-scale protein sequences and fine-tuned on specific tasks outperform traditional models trained with limited data[45]. We utilized the Transformers module and the ESM-2 protein pretrained model for sequence feature extraction. The ESM-2 model, employing the BERT architecture, was pretrained on the UniRef50 protein sequence database by Lin et al.[43].

For the classification task, we fixed the parameters of the ESM-2 model and integrated a multilayer perceptron (MLP) for fine-tuning on the conopeptide dataset. The MLP has four fully connected layers; the first three are each followed by a batch normalization layer and a ReLU activation function. The final layer is a fully connected output layer that maps features into the output space. The softmax function converts the output features into a probability distribution, providing predicted classification probabilities. We used t-distributed Stochastic Neighbor Embedding (t-SNE) to represent high-dimensional data on a two-dimensional map, effectively simplifying data complexity. Fig. 2A shows the t-SNE analysis, indicating that our model effectively distinguishes between two conopeptide groups (*i.e.*, those targeting nAChRs and those that do not).

For the regression, we utilized the transformer's AutoModel for sequence classification with num_labels set to 1. The model's final layer consists of a linear transformation, and its outputs are directly utilized to compute the relative potency ranking. The scatter plot in Fig. 2A shows that the validation data (blue points) are essentially distributed along the diagonal, indicating accurate potency predictions.

3.2. Optimization and evaluation of the DL-aided peptides screening model

The development environment and requirements for the DL-aided screening model are outlined in Supporting Information Table S1. We optimized model performance by adjusting the hyper-parameters, as detailed in Supporting Information Figs. S2−S8. The optimized hyper-parameter values are listed in Supporting Information Tables S2 and S3. We conducted 10-fold cross-validation on both binary classification and regression models using optimal hyperparameters. Performance results are displayed in Fig. 2B and C, with detailed outcomes in Supporting Information Table S4. The classification model scored above 0.8 across all evaluation metrics, demonstrating robust performance. The model achieved an average accuracy of 0.904, highlighting its exceptional capability to differentiate between nAChRs and non-nAChRs. Sensitivity and specificity metrics both surpassed 0.9, indicating the model's high efficacy in identifying conopeptides targeting nAChRs and minimizing false positives. Despite some instability in RMSE performance due to the small validation set, an average RMSE of 0.831 with a standard deviation of 0.266 was achieved, as depicted in the box plot in Fig. 2B. The average Spearman score of our model was 0.930, with all except the fifth fold demonstrating excellent performance. This result indicates that the model performs excellently in ranking relative potency values, meeting wet lab screening requirements for identifying active conopeptides from a large pool.

3.3. Peptide screening and inhibitory activity validation

In the next step, we employed our DL-aided peptide screening model to identify non-canonical conopeptides targeting the α7 nAChR from the disulfide-poor dataset (Supporting Information File S7). In the classification task, 81 conopeptides were identified as potentially targeting nAChRs, while 608 conopeptides were predicted not to interact with nAChRs. In the regression task, we predicted the relative potency ranking of these 81 peptides (Supporting Information File S8). Five conopeptide sequences with the highest ranking were selected as candidates for experimental investigation (Table 2). These peptides, renamed ZS1, ZS2, and SS1 to SS3, were synthesized with a purity of ≥95%. In electrophysiology experiments, SS1 demonstrated the strongest inhibition of ACh-evoked peak current amplitude in α7 nAChR-expressing oocytes and was selected for further study (Supporting Information Fig. S9).

In addition, we collected some α7 nAChR-targeting conopeptide sequences published in 2024−2025[46-48]. Although these sequences were not included in the training set (since the model was developed starting in 2023), they were accurately predicted by our model as potential nAChR binders with 100% accuracy (8/8) (Supporting Information Table S5). To further validate the classification effectiveness, we synthesized and tested the top five conopeptides predicted to be non-nAChR binders (Table 2 and Supporting Information Fig. S10). At concentrations of 1 and 10 μmol/L, none showed significant inhibitory activity, confirming that the classification process effectively minimizes

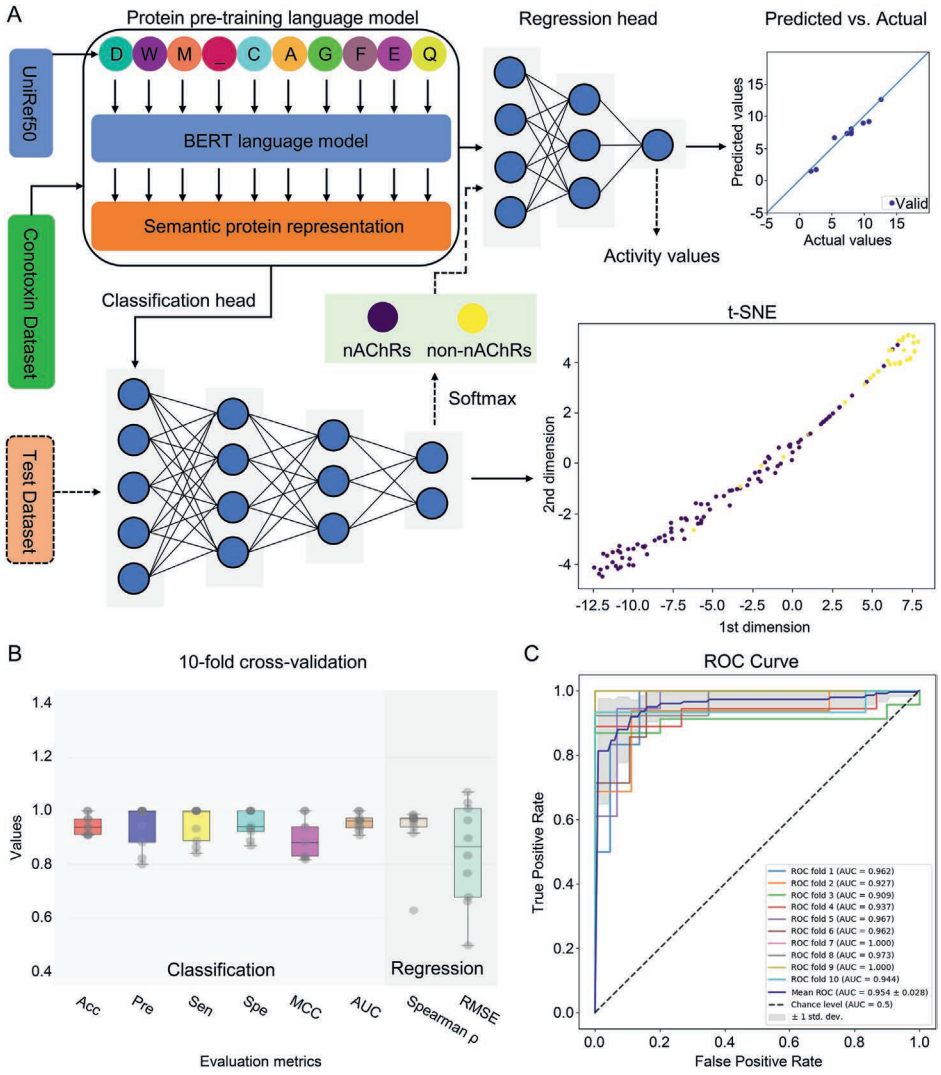

Figure 2 Structure and performance of deep-learning based peptide screening model. (A) The binary classification task and the regression task are each built using the BERT-based protein pre-trained language model ESM-2, combined with a classification head and a regression head, respectively. The binary classification task is designed to predict whether the input conopeptide targets are the nAChRs or not, while the regression task is used to predict the potency of conopeptides targeting the α7 nAChR. The data flow of test dataset is represented by a dashed line. (B) The performance of the model was evaluated using 10-fold cross-validation. (C) The ROC curves of the 10-fold cross-validation for the screening model.

false negatives and avoids the omission of active peptides. This result strongly suggests that our model has successfully acquired the ability to identify nAChR binders, which is essential for guiding downstream wet-lab experiments.

Despite high accuracy in the evaluation metrics, wet lab experiments yielded only 20% accuracy. This discrepancy may be due to four factors: the inclusion of ligands targeting various nAChR subtypes, which diluted model specificity; significant differences between peptide sequences in the training (disulfide-rich peptides) and input (disulfide-poor peptides) datasets, reducing accuracy; limited sampling, where increasing the sample size would conflict with the goal of minimizing wet-lab experiments; and variability in peptide potency data due to non-standardized testing criteria in different laboratories[24,49,50].

For DL and artificial intelligence-based approaches, the authenticity of the training set data is critical to the model's performance. Accordingly, we attempted to collect data exclusively from tests conducted on *Xenopus* oocytes or human SH-SY5Y neuroblastoma cells and standardized concentrations and dosages of agonists to improve data quality. Despite these efforts, the limited amount of data resulted in a suboptimal outcome. Subsequently, we fine-tuned the encoder layers of ESM-2 using a masked language modeling (MLM) task tailored for peptide sequences[51]. This approach aimed to enhance the model's ability to prioritize mutations and adjust weights to better capture conopeptide-specific features, resulting in slight performance improvements (Supporting Information Table S6 and Fig. S11). In future related studies, we advocate for the establishment of standardized protocols to enhance data reproducibility across different laboratories.

3.4. Probing the binding mode of SS1 by Ala, Arg and Asp scanning

Given its robust inhibition of ACh-evoked currents mediated by the hα7 nAChR, SS1 was selected for further investigation. Alanine

Table 2 The synthesis peptides and their corresponding sequence.

Peptide	Sequence	Peptide	Sequence	Peptide	Sequence
Candidate conopeptides		[S8R]SS1	QPGCRWWR**R**WC[a]	[ΔQP,R5Dap,S8Dap]SS1	GC**Dap**WW**Dap**RWC[a]
ZS1	DWMYIHAHPYSNTLWSLV[a]	[W10R]SS1	QPGCRWWSR**R**C[a]	[ΔQP,C4A,C11A,R5Dap,S8Dap]SS1	G**A****Dap**WWDap**R**W**A**[a]
ZS2	TRNYYVYSAPENSWWTRVV[a]	Aspartic scanning analogues		[ΔQP,C4A,C11A,R5Dap,S8R]SS1	G**A****Dap**WW**R**RW**A**[a]
SS1	QPGCRWWSRWC[a]	[Q1D]SS1	**D**PGCRWWSRWC[a]	Third-generation analogues	
SS2	DPPCPILHCP[a]	[P2D]SS1	Q**D**GCRWWSRWC[a]	[S8R]SS1-1	QPGCRWW**R**RWC
SS3	NGCNGNTCSNTPLP[a]	[G3D]SS1	QP**D**CRWWSRWC[a]	[Q1Pyr, S8R]SS1	**Pyr**PGCRWW**R**RWC[a]
Negative conopeptides		[R5D]SS1	QPGC**D**WWSRWC[a]	[ΔQP, S8R]SS1	GCRWWR**R**WC[a]
AI(-1)	GEEEYQKMLENLREDKKVKKIDKEKE[a]	[W6D]SS1	QPGCR**D**WSRWC[a]	[C4S, C11S, S8R]SS1	QPG**S**RWW**R**RW**S**[a]
AI(-2)	GYEEDWEIWENIRKLIEAGK[a]	[W7D]SS1	QPGCRW**D**SRWC[a]	[C4A, C11A, S8R]SS1	QPG**A**RWW**R**RW**A**[a]
AI(-3)	GEEEYQKMLENLREDKKLKKIDKEKE[a]	[S8D]SS1	QPGCRWW**D**RWC[a]	SD-S8R	GCCSDPRCWWRRWC[a]
AI(-4)	GEEEYQKMLENLREARKLKKIDKEKE[a]	[R9D]SS1	QPGCRWWS**D**WC[a]	SH-S8R	GCCSHPACWWRRWC[a]
AI(-5)	GNTARGTEEFIEFEEKMRELDIAEKLKKD[a]	[W10D]SS1	QPGCRWWSR**D**C[a]	Fourth-generation analogues	
Alanine scanning analogues		First-generation analogues		[ΔQP, C4A, C11A, S8R]SS1	G**A**RWW**R**RW**A**[a]
[Q1A]SS1	**A**PGCRWWSRWC[a]	Ca-73	QPGCRWWSRWC	[ΔQP, C4S, C11S, S8R]SS1	G**S**RWW**R**RW**S**[a]
[P2A]SS1	Q**A**GCRWWSRWC[a]	[ΔQP]SS1	GCRWWSRWC[a]	[ΔQP, C4Pen, S8R]SS1	G**Pen**RWW**R**RWC[a]
[G3A]SS1	QP**A**CRWWSRWC[a]	[12R]SS1	QPGCRWWSRWC**R**[a]	[ΔQP, S8R, C11Pen]SS1	GCRWW**R**RW**Pen**[a]
[R5A]SS1	QPGC**A**WWSRWC[a]	[12R,13R]SS1	QPGCRWWSRWC**RR**[a]	[ΔQP, G3Sar, S8R]SS1	**Sar**CRWW**R**RWC[a]
[W6A]SS1	QPGCR**A**WSRWC[a]	[C4A,C11A]SS1	QPG**A**RWWSRW**A**[a]	Others	
[W7A]SS1	QPGCRW**A**SRWC[a]	[R5Dap]SS1	QPGC**Dap**WWSRWC	[R5A, S8R]SS1	QPGC**A**WW**R**RWC[a]
[S8A]SS1	QPGCRWW**A**RWC[a]	[R5Dab]SS1	QPGC**Dab**WWSRWC[a]	[W6R, S8R]SS1	QPGCR**R**W**R**RWC[a]
[R9A]SS1	QPGCRWWS**A**WC[a]	[S8Dap]SS1	QPGCRWW**Dap**RWC[a]	[W6A, S8R]SS1	QPGCR**A**W**R**RWC[a]
[W10A]SS1	QPGCRWWSR**A**C[a]	[P2Hyp]SS1	Q**Hyp**GCRWWSRWC[a]	[S8R, R9A]SS1	QPGCRWW**RA**WC[a]
Arginine scanning analogues		Second-generation analogues		[P2Hyp, S8R]SS1	Q**Hyp**GCRWW**R**RWC[a]
[Q1R]SS1	**R**PGCRWWSRWC[a]	[R5Dap,S8Dap]SS1	QPGC**Dap**WW**Dap**RWC[a]	[Q1E, S8R]SS1	**E**PGCRWW**R**RWC[a]
[P2R]SS1	Q**R**GCRWWSRWC[a]	[R5Dap,S8R]SS1	QPGC**Dap**WW**R**RWC[a]	[Q1R, S8R]SS1	**R**PGCRWW**R**RWC[a]
[G3R]SS1	QP**R**CRWWSRWC[a]	[C4A,C11A,R5Dap,S8Dap]SS1	QPG**A****Dap**WW**Dap**RW**A**[a]	[Q1Pyr, S8R]SS1	**Pyr**PGCRWW**R**RWC[a]
[W6R]SS1	QPGCR**R**WSRWC[a]	[C4A,C11A,R5Dap,S8R]SS1	QPG**A****Dap**WW**R**RW**A**[a]		
[W7R]SS1	QPGCRW**R**SRWC[a]	[ΔQP,R5Dap,S8R]SS1	GC**Dap**WW**R**RWC[a]		

[a]=amidated C-terminal; Dap = 2,3-diaminopropionic acid; Dab = 2,4-diaminobutanoic acid; Pyr = Pyroglutamic acid; Sar = N-methyl alanine; Hyp = Hydroxyproline; Pen = Penicillamine; Disulfide bonds formed between cysteine residues.

(Ala) scanning identifies functional residues in bioactive peptides by eliminating residue side chains[52], while arginine (Arg) and aspartic (Asp) substitution investigates the impact of charge change on the potency of peptides at specific sites[52,53]. To elucidate the roles of individual amino acid residues in ligand-receptor interactions, Ala, Arg and Asp substitutions were made in SS1, excluding Cys residues (Table 2). The activities of these analogues were assessed at 1 μmol/L, near the IC_{50} of SS1 at the hα7 nAChR using electrophysiology assays (Supporting Information Fig. S12, Fig. 3A and B).

In general, the effectiveness of designing analogues based on structure-activity relationships hinges on an unchanged secondary structure of the set of analogues. We thus recorded the circular dichroism (CD) spectrum of SS1 and its scanning mutations. Notably, all Ala, Arg, and Asp substitutions—except at positions 6 and 7—showed two absorption peaks at 208 nm and 220 nm, consistent with the parent SS1 (Supporting Information Figs. S13−S17). However, mutations at the Trp residues at positions 6 and 7 led to the loss of the 220 nm peak, suggesting a conformational change in these analogues (Figs. S16 and S17). To determine if these secondary structure changes were due to alterations in the peptide backbone or side chains, we conducted site-directed mutagenesis on SS1 using PyMOL, followed by molecular dynamic (MD) simulations to analyze backbone stability. The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) values for the W7A, W7R, W7D, W6A, W6R, and W6D mutants remained stable (Supporting Information Figs. S18 and S19), indicating that the observed changes in the CD spectra at positions 6 and 7 originate from side chain modifications rather than backbone alterations.

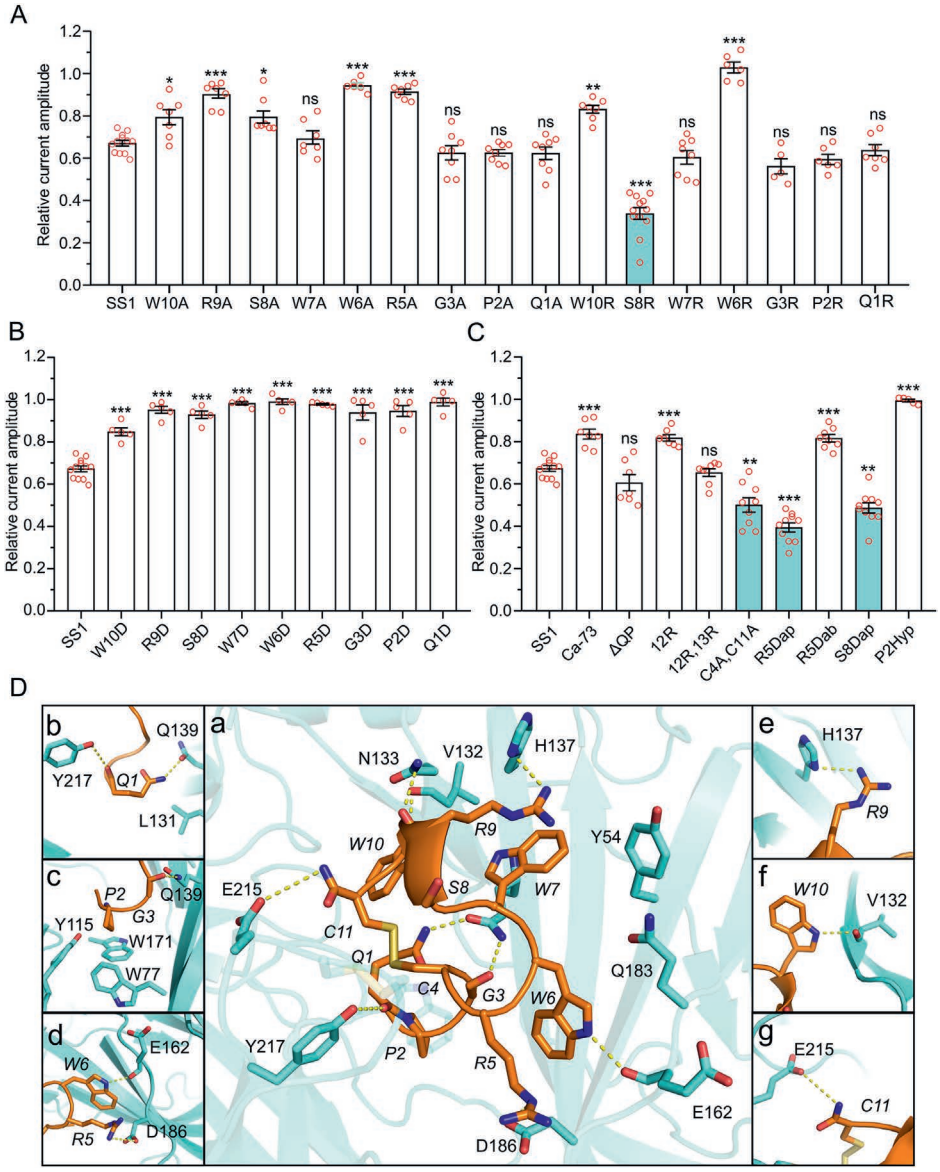

Figure 3 Binding mode and structure-activity relationship of SS1 and hα7 nAChR. (A) Bar graphs of relative inhibition of ACh-evoked peak current amplitude by alanine and arginine scanning analogues of SS1. (B, C) Percentage of ACh-evoked peak current amplitude inhibition by the aspartic scanning (B) and the first-generation (C) analogues of SS1. (D) Binding mode of SS1 at the neurotransmitter binding site of α7 nAChR. The α7 subunit is depicted in cyan, and SS1 is shown in orange. The peptide concentration was 1 μmol/L and whole-cell currents mediated by hα7 nAChR were activated using 300 μmol/L ACh (mean ± SEM, $n = 5 - 12$, $^{*}P < 0.01$, $^{**}P < 0.001$, $^{***}P < 0.0001$, one-way ANOVA).

For the Ala mutants of SS1, only the Q1A, P2A, G3A, and W7A analogues maintained SS1 inhibitory activity (31%−35% inhibition, $n = 7-8$), whereas the other analogues showed significantly reduced potency (5%−21% inhibition, $n = 6-8$). Consistent with Ala scanning, Q1R, P2R, G3R, and W7R showed no significant differences ($P > 0.05$, $n = 5-8$) in ACh-evoked currents compared to SS1. Additionally, the W10R analogue showed approximately 17% inhibition of ACh-evoked currents ($P < 0.001$, $n = 7$), while the W6R analogue completely lost potency ($P < 0.0001$, $n = 6$). Different from the results obtained from Ala and Arg scanning, all Asp mutant analogues exhibited a significant reduction in potency (1%−16% inhibition, $n = 5$). The S8R mutant exhibited the highest inhibitory activity (∼66% inhibition, $n = 12$) among all analogues (Fig. 3A and B).

Overall, residues at positions 5, 6 and 8−10 appear to interact with the α7 nAChR, as removing their side chains, significantly reduced potency. Though mutations to Ala and Arg on residues Q1, P2, G3, and W7 in SS1 had minimal impact on potency, replacing these positions with Asp markedly reduced potency. Notably, substituting Ser at position 8 with Arg significantly increased potency, likely due to the positive charge enhancing interactions with the α7 nAChR.

3.5. Mechanism of action and SAR of SS1 at α7 nAChR

To elucidate the binding mode and better understand the structure−activity relationships (SAR) of SS1, we constructed a model of the complex of SS1 with hα7 nAChR using computational modeling, validated by mutagenesis data. AlphaFold2 (AF2) generated the three-dimensional structure of SS1, followed by MD simulations to evaluate the peptide's stability. During the 500 ns of MD simulations, SS1 displayed substantial flexibility, alternating between random coil and partial α-helix configurations, likely attributed to its single disulfide bond and limited conformational constraints (Supporting Information Fig. S20). This flexibility rendered traditional docking methods unsuitable for accurately predicting the receptor−peptide complex. We used the AlphaFold3 (AF3) to create a model of the hα7/SS1 complex. After 200 ns of MD simulation to relax the hα7/SS1 complex, the final conformation was analyzed to explore the interactions between SS1 and the hα7 nAChR (Fig. 3D).

In the hα7/SS1 complex model, SS1 binds to the ACh binding site and forms extensive hydrogen bonds, salt-bridges and hydrophobic interactions with the C-loop of the principal subunit and the β-sheet of the complementary subunit (Fig. 3B and Table 3). To validate the complex model, we compared it with electrophysiological data from Ala, Arg and Asp scanning. In the complex model, the side-chain nitrogen and main-chain carbonyl of Q1 form hydrogen bonds with Q139 and Y217 in the receptor, respectively. When Q1 is mutated to Ala, the main-chain carbonyl hydrogen bond is preserved, and the unfavorable interaction with L131 is reduced, thereby retaining the activity. When Q1 was mutated to Arg, the original interactions remain intact, and the inhibitory activity is also maintained. However, mutation to Asp abolishes the interaction with Q139, resulting in reduced potency ($n = 5$, $P < 0.0001$). For P2, it engages in strong hydrophobic interactions with Y115, W77, and W171 of the receptor, and the introduction of a hydroxyl group to the Pro will be detrimental to the hydrophobic interactions with these aromatic residues. Indeed, substituting P2 with hydroxyproline (Hyp) resulted in a complete loss of activity (<1% inhibition, $P < 0.0001$), which is highly consistent with the model predictions. Replacing Pro with Ala or Arg can change the peptide flexibility, potentially leading to new interaction modes with the receptor and maintaining peptide potency. The guanidine group of Arg at position 5 forms salt bridges with the carboxyl group of D186. Replacing the side chain of R5 with a methyl group, significantly reduced the potency (<10% inhibition, $P < 0.0001$) due to the loss of these hydrogen bonds. Likewise, Arg at position 9 forms hydrogen bonds with H137, contributing to the significant loss of potency of R9A (<10% inhibition, $P < 0.0001$). For W6 and W10, the N−H on the indole rings forms hydrogen bonds with E162 and V132, respectively. Disruption of these hydrogen bonds by Ala and Arg led to decreased potency (5%−21% inhibition, $n = 6-8$). Residue W7 does not have significant interactions with the receptor, and the Ala and Arg mutants retain comparable potency as SS1. During the 200 ns MD simulation, a weak hydrogen bond forms between the Ser at position 8 and C212 on the receptor (Supporting Information Fig. S21). The S8A mutation disrupts this hydrogen bond between S8 and C212, resulting in slightly decreased potency (∼30% inhibition, $n = 7$, $P < 0.01$). Intriguingly, S8R significantly enhances potency against hα7 nAChR (66%

Table 3 Pairwise interactions between the SS1/[S8R]SS1 and hα7 nAChR.

Residue[a]		Q1	P2	G3	C4	R5	W6	W7	S8(R8)	R9	W10	C11
Interaction residue with SS1	α7(+)[b]	Y217	*Y115, W171*	/	/	/		/	/	/	/	E215
	α7(−)[c]	Q139	*W77*	Q193	/	**D186**	E162	/	/	H137	V132	/
Interaction residue with [S8R]SS1	α7(+)	D219,R208	/	/	/	/	Y210	/	/	/	/	/
	α7(−)	/	/	/	/	**D186**, S58, S188	*W77*	/	**D186**, E184, S56, Q79	Y54, Q183	S81, Q139, NAG	NAG
Potency change[d]	A	0	0	0	/	−	−	0	−	−	−	/
	R	0	0	0	/	/	−	0	+	/	−	/
	D	−	−	−	/	−	−	−	−	−	−	/

[a]All SS1/[S8R]SS1 residues are listed. Residues of the hα7 nAChR forming hydrogen bonds and with peptides are underlined. Salt bridge interactions are bolded, and pi−pi stacking and hydrophobic interactions are italicized.

[b]Residues on the principal side of the binding site.

[c]Residues on the complementary side of the binding site.

[d]Potency changes in alanine and arginine scanning analogues of SS1: "0" indicates no change, "+" signifies enhanced Potency, and "−" denotes reduced Potency. The "/" symbol means the test was not conducted. Residues of the human α7 nAChR that interact pairwise with SS1/[S8R]SS1 are shown in black and purple, respectively.

inhibition, $n = 12$, $P < 0.0001$), although the model does not indicate any interactions with the newly introduced Arg at position 8. We speculate that the bulky side chain of R8 might induce a binding orientation change of the peptide at the binding site. Surprisingly, introducing Asp at nearly any position of SS1 results in significantly decreased potency, suggesting that the hα7 nAChR binding site is intolerant to negative charges (Fig. 3B). This finding is consistent with the binding site's overall negative electrostatic potential, which likely prevents the incorporation of additional negative charges on the SS1 surface (Supporting Information Fig. S22).

3.6. Rational design and potency evaluation of four generations of SS1 analogues

Building on a comprehensive understanding of the action mechanism and SAR of SS1 at the α7 nAChR, we utilized site-directed mutagenesis, non-natural amino acid substitutions, and sequence truncation to generate a first generation of SS1 analogues (Table 2) and evaluated their potency (at 1 μmol/L) at the hα7 nAChR (Fig. 3C).

Disulfide bond formation and C-terminal amidation are common post-translational modifications (PTMs) in conopeptides and are crucial for their stability and potency[54]. Surprisingly, replacing Cys4 and Cys11 with Ala in SS1 slightly increased its activity (50% inhibition, $n = 9$, $P < 0.0001$), likely due to enhanced peptide flexibility without disulfide constraints, allowing the peptide to better adapt to the binding site. The analogue of SS1 with a free C-terminal, Ca-73[55], had reduced potency against the α7 nAChR ($\sim$16% inhibition, $n = 7$, $P < 0.0001$). This aligns with our model, which shows a crucial hydrogen bond between the C-terminal amide and E215. Consistent with previous findings[56], glutamic acid at the N-terminus rapidly converts to pyroglutamic acid at room temperature (Supporting Information Figs. S23 and S24). Given that the first amino acid in many conopeptides is Gly, this prompted us to truncate the two N-terminal amino acids of SS1 to obtain [ΔQP]SS1. The truncated peptides exhibited comparable potency to SS1 ($\sim$40% inhibition, $n = 7$), indicating that the N-terminal tail is not critical for inhibitory activity. In addition, the introduction of Hyp at position 2 of SS1 led to a significant reduction in potency (<5% inhibition, $P < 0.0001$), primarily due to the disruption of hydrophobic interactions between P2 and the receptor's W77, W171 and Y115.

A hydrophilic region near SS1's C-terminus led us to hypothesize that adding Arg residues might enhance interactions. However, [12R]SS1 and [12R,13R]SS1 showed no improvement in potency (18%−35% inhibition, $n = 7$−8), likely due to steric hindrance from the extended peptides. Dap and Dab, two non-natural lysine analogues with varied side chain lengths, were used to replace R5, exploring the optimal distance for forming a hydrogen bond with D186. R5Dap showed enhanced inhibitory activity (51% inhibition, $n = 10$, $P < 0.0001$) while R5Dab showed reduced potency (17% inhibition, $n = 10$, $P < 0.0001$), suggesting a one-carbon distance may be optimal. Our previous work indicates that mutating Ser to Dap in α-conotoxins significantly enhances inhibitory activity by forming stronger salt-bridges with the receptor[28]. As expected, mutating Ser at position 8 to Dap significantly improved potency (61% inhibition, $n = 10$, $P < 0.0001$), probably due to forming stronger hydrogen bonds with C212.

Through mutagenesis and design of first-generation analogues, we identified mutants such as [S8R/Dap]SS1, [R5Dap]SS1, and [C4A, C11A]SS1 that enhanced the potency of SS1 at hα7 nAChR, in the order of [S8R]SS1 > [R5Dap]SS1 > [S8Dab]SS1 > [C4A, C11A]SS1. Additionally, the N-terminal truncation analogue [ΔQP]SS1 maintained the potency of SS1, indicating that the N-terminal tail, although involved in receptor interactions, is not essential for inhibitory activity.

We designed a second generation of SS1 analogues using [S8R]SS1, [R5Dap]SS1 and [S8Dab]SS1 as starting points. We employed Cys-to-Ala mutations and N-terminal truncation to further optimize peptide structure and activity. Eight second-generation analogues were synthesized and their activities were evaluated using two-electrode voltage clamp electrophysiology (Table 1 and Fig. 4A). At 1 μmol/L, only [ΔQP, R5Dap, S8R]SS1 had comparable inhibition (69% inhibition, $n = 10$) to [S8R]SS1. All other mutants had reduced potency, with inhibition values ranging from 21% to 53% ($n = 7$−10). It seems likely that these mutations, especially the multi-site mutations, significantly altered the conformation of SS1, as confirmed by CD results (Supporting Information Figs. S25 and S26). Analogues with N-terminal truncation, such as [ΔQP, R5Dap, S8R]SS1 and [ΔQP, R5Dap, S8Dap]SS1, showed greater potency than those without truncation. Conversely, Cys-to-Ala mutations did not enhance potency compared to those without mutations. These results indicate that N-terminal truncation is more effective than Cys-to-Ala mutation in enhancing peptide inhibition potency.

For further optimization, we focused on [S8R]SS1, the most potent analogue. We drew on insights gained from the enhancements observed in the first two rounds of design, resulting in the design of five third-generation analogues (Table 2 and Fig. 4B). At 1 μmol/L concentration, replacing the C-terminal amide of [S8R]SS1 with a carboxyl group did not affect potency at the α7 nAChR (70% inhibition, $n = 10$). In contrast, substituting Cys with Ala or Ser significantly reduced potency (26%−47% inhibition, $n = 7$−8, $P < 0.0001$), contrasting with results on SS1. Given that many α-conotoxins targeting α7 nAChR share a 4/7 disulfide framework, we engineered two analogues, SH-S8R and SD-S8R, by replacing the loop II residues of Epi and Mr1.1 with the sequence from [S8R]SS1, aiming to enhance the potency. Unexpectedly, the potency of these mutants decreased significantly, likely due to differences in binding modes. Given N-terminal glutamine's tendency to convert to pyroglutamate (Pyr), we removed the two N-terminal amino acids or replaced Gln with Pyr. Consistent with earlier mutational results, [ΔQP, S8R]SS1 showed a slight increase in inhibitory activity (78% inhibition, $n = 6$, $P < 0.05$). [Q1Pyr, S8R]SS1 achieved 66% inhibition of α7 nAChR-mediated currents ($n = 7$, $P > 0.05$), comparable to that of [S8R]SS1.

In the fourth optimization round, five analogues were designed to optimize the structure based on [ΔQP, S8R]SS1 (Table 2 and Fig. 4C). The two cysteine residues were replaced with Ala and Ser, significantly reducing the inhibitory activity (12%−33% inhibition, $n = 6$−7, $P < 0.0001$). Furthermore, these cysteines at positions 4 and 11 were substituted with the non-natural amino acid Pen to potentially improve peptide serum stability as suggested in our previous study[57]. The mutation at position 4 led to a considerable decrease in inhibitory activity ($n = 5$, $P < 0.001$), while the potency was retained at position 11 ($n = 5$, $P > 0.01$). Gly at position 3 was also replaced with the non-natural amino acid N-methyl Gly, and we found that this modification could effectively preserve an inhibitory effect at a 1 μmol/L concentration ($\sim$75% inhibition, $n = 5$).

Concentration−response relationships for SS1, [S8R]SS1 and [ΔQP, S8R]SS1 showed that SS1 had an IC_{50} of 1.4 ± 0.3 μmol/L,

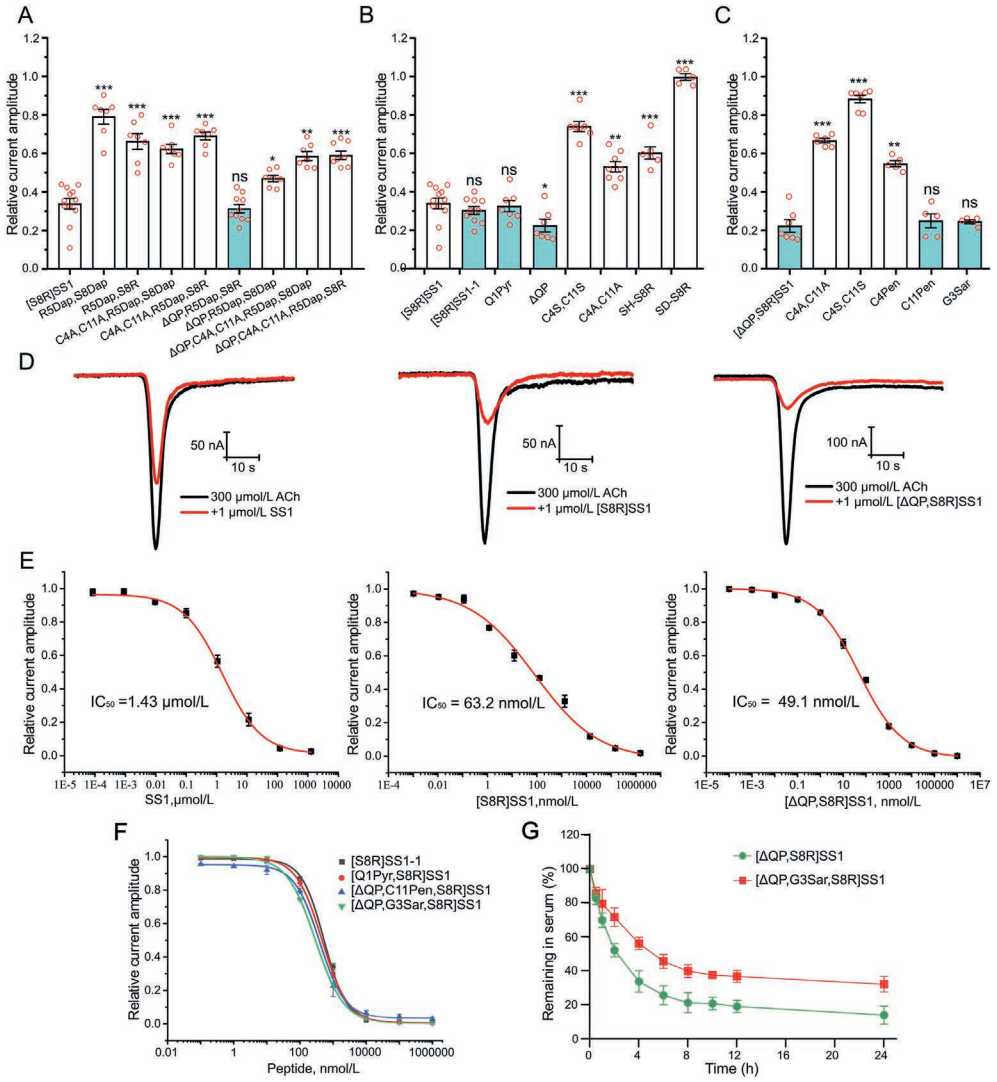

Figure 4 The inhibitory activity of conopeptide analogues and optimization results of SS1. Bar graphs of relative inhibition by 1 μmol/L SS1 second-generation (A), Third-generation (B) and Fourth-generation analogues (C) of ACh-evoked current amplitude mediated by hα7 nAChR. Data were acquired using 300 μmol/L ACh as the agonist. (mean ± SEM, $n = 5-12$, $^{*}P < 0.01$, $^{**}P < 0.001$, $^{***}P < 0.0001$, one-way ANOVA). (D) Representative currents mediated by hα7 nAChR, recorded in the absence (control, black trace) and presence (red lines) of SS1, [S8R]SS1, and [ΔQP, S8R]SS1. (E) Concentration-response of relative ACh-evoked current amplitude mediated by hα7 nAChR in the presence of SS1, [S8R]SS1 and [ΔQP, S8R]SS1 giving IC_{50} of 1.4 μmol/L (0.5 − 2.4; 95% CI), 63.2 nmol/L (30.1 − 96.5; 95% CI) and 49.1 nmol/L (34.1 − 64.2; 95% CI), respectively. (F) Concentration-response of relative ACh-evoked current amplitude mediated by hα7 nAChR in the presence of [S8R]SS1-1 (539.6 ± 52.6 nmol/L), [Q1Pyr,S8R]SS1(462.7 ± 49.8 nmol/L), [ΔQP,C11Pen,S8R]SS1 (389.6 ± 57.1 nmol/L), [ΔQP,G3Sar,S8R]SS1 (290.8 ± 12.6 nmol/L), respectively. (G) The stability of [ΔQP, S8R]SS1 and [ΔQP, G3Sar, S8R]SS1 in human serum over a 24-h period. The values are presented as means ± SEM from at least three independent experiments.

while [S8R]SS1 and [ΔQP, S8R]SS1 had values of 63.3 ± 13.6 and 49.2 ± 6.4 nmol/L, respectively (Fig. 4D and E). These represent nearly 23-fold and 29-fold increases in potency for [S8R]SS1 and [ΔQP, S8R]SS1 compared to SS1.

We further examined the concentration-response curves for four other analogues with comparable relative inhibition potency to [S8R]SS1 identified in the third and fourth rounds (Fig. 4F). Surprisingly, the results indicate that all analogues exhibited decreased potency compared with [S8R]SS1. Among these peptides, [ΔQP, G3Sar, S8R]SS1 showed the highest potency, with an IC_{50} of 290.8 ± 12.6 nmol/L, exhibiting a roughly 4.6-fold loss in potency compared to [S8R]SS1. This was followed by [ΔQP, C11Pen, S8R]SS1 with an IC_{50} of 389.6 ± 57.1 nmol/L. Next, [Q1Pyr, S8R]SS1 had an IC_{50} of 462.7 ± 49.8 nmol/L, whereas [S8R]SS1-1, at 539.6 ± 52.6 nmol/L, was the least potent.

We also investigated the selectivity and serum stability of the conopeptides (Supporting Information Fig. S27 and Fig. 4G). At concentrations of 1 and 10 μmol/L, SS1 showed broad inhibitory activity across several common acetylcholine receptor subtypes, particularly α3β2 and α4β4, with IC_{50} values of 5.6 and 2.1 μmol/L, respectively. By contrast, [S8R]SS1 and [ΔQP, S8R]SS1 showed only weak potency against other tested subtypes at 10 μmol/L, with the strongest effect observed on the α3β4 subtype, with only 35% inhibition at 10 μmol/L. This result indicates that the selectivity of [S8R]SS1 and [ΔQP, S8R]SS1 is significantly improved compared to SS1. Considering the high stability

of [ΔQP, S8R]SS1 in aqueous solution (Fig. S24), we examined its serum stability and found that 14% of the peptide remained after 24 h of degradation in serum. In comparison, the Pen-modified peptide [ΔQP, S8R, C11Pen]SS1 exhibited significantly higher stability, with 32% of the peptide remaining after 24 h.

Overall, although [ΔQP, S8R]SS1 has a slightly lower IC_{50} than [S8R]SS1, it is a truncated version of [S8R]SS1. We compared the CD spectra of SS1 and [ΔQP, S8R]SS1 in 50% TFE, 50% PBS, and pure water. Both peptides retained characteristic absorption peaks in all three solvent environments, suggesting that their three-dimensional structures are quite similar (Supporting Information Figs. S28 and S29). To more clearly illustrate the SARs of SS1 analogues, [S8R]SS1 was used in subsequent binding mode analyses.

3.7. Determination of the binding mode between [S8R]SS1 and α7 nAChR through the integration of cryo-EM and computational modeling

We co-expressed the modified wild-type hα7 ($\alpha7^{EM}$) gene and the chaperone NACHO in HEK293 cells (Supporting Information Fig. S30A). Electrophysiological experiments confirmed that our EM α7 construct retained the functional characteristics of wild-type α7 nAChR, responding similarly to ACh and [S8R]SS1 (Fig. S30B and S30C). A sharp SEC peak was collected that yielded receptor−ligand complex structure (Supporting Information Fig. S31A), and SDS-PAGE gel confirmed that the molecular weight matched the α7 subunit (Fig. S31B). We collected cryo-EM data using a Titan Krios microscope and performed single-particle analysis (Supporting Information Fig. S32). The three-dimensional density map of full-length α7 nAChR bound to [S8R]SS1 was determined at 3.3 Å resolution, revealing a closed pore configuration (Fig. 5). The highest resolution was in the extracellular domain (ECD), followed by the transmembrane domain (TMD), with the intracellular domain (ICD) exhibiting the lowest resolution, likely due to its dynamic nature (Supporting Information Fig. S33). The map is rich in structural features, including densities for sidechains, *N*-glycans, and lipid molecules, enabling us to reliably build atomic models of the α7 nAChR. However, despite clear density in the ACh binding site of the ECD, we were unable to confidently position each density to perfectly match every residue in [S8R]SS1.

Although the cryo-EM experiment did not yield a complete α7/[S8R]SS1 complex model, we successfully generated a precise structure of α7 co-incubated with [S8R]SS1, confirming its location at the neurotransmitter binding site of the ECD. The density of [S8R]SS1 is located between Loop C and Loop D, surrounded by R208, Y210, C212, C213, and Y217 from the principal subunit, and Q79, Q139, and D186 from the complementary subunit (Fig. 5A). In 2021, Noviello CM and colleagues elucidated the structure and gating mechanism of the α7 nAChR, identifying three key conformations in its gating cycle: a closed resting state, an open activated state, and a closed desensitized state[58]. We characterized the state of the $\alpha7^{EM}$ structure by measuring the pore radius and comparing it to the structure reported by Noviello CM, finding the ECD and TMD remarkably similar to the closed state bound to alpha-bungarotoxin (PDB ID: 7KOO) (Fig. 5B, C and F), whereas the intracellular domain (ICD) displayed similarities to both the closed state and the desensitized state typically observed with epibatidine binding (PDB ID: 7KOQ) (Fig. 5C). In our previous study on the hα4β2 nAChR, we characterized the C-loop opening by measuring the distance between the C_α of Cys199 in the C-loop of the α4 subunit and the C_α of Ser58 in the complementary β2 subunit, and defined the activation distance at the interface as the distance between the C_α atoms of K68 (at the β1−β2 linker) and P284 (at the M2−M3 linker)[59]. As illustrated in Fig. 5D and E, the C-loop opening distance and activation distance in our α7 receptor are 21.5 Å and 6.9 Å, respectively, indicative of a closed state. As a comparison, the cryo-EM structure of the α7/alpha-bungarotoxin complex, with a C-loop opening of 20.0 Å and an activation distance of 6.9 Å, further supports our model of the α7 nAChR in an overall closed state.

Subsequently, AF3 and MD simulations were used to construct structure of the hα7/[S8R]SS1 complex. Specifically, AF3 was used to predict the initial structure of hα7/[S8R]SS1, which was then superimposed onto orthosteric site of the cryo-EM α7 structure to obtain the structure of the cryo-EM hα7/[S8R]SS1 complex (Fig. 6A). The predicted local distance difference test (pLDDT) and Predicted Aligned Error (PAE) values for the AlphaFold-generated structure indicate high confidence in the receptor region and significant uncertainty in the peptide conformations (Supporting Information Figs. S33 and S34). The observed discrepancy is likely attributed to the substantial sequence divergence of [S8R]SS1 from known conopeptides, which hindered the AlphaFold model from identifying suitable homologous templates. To address this discrepancy, we conducted a 200 ns MD simulation to refine the modeled complex structure by imposing a soft-restraint merely on the receptor. This strategy can preserve the conformation of the $\alpha7^{EM}$ while enabling the peptide to fully explore its optimal bound conformation within the binding site. The simulation revealed that the ligand maintained remarkable stability at the binding site throughout the trajectory (Fig. 6B). The mean RMSD values of [S8R]SS1 in the complex during 200 ns simulations was less than 1.00 Å (Fig. 6C). RMSF analysis revealed that the fluctuations of each residue in [S8R]SS1 were also less than 1.00 Å (Fig. 6D). The stability of the ligand within the binding pocket, as observed during the MD simulations, suggests a high degree of reliability in the constructed complex.

The MD refined structures of hα7/[S8R]SS1 and hα7/SS1 show that both [S8R]SS1 and SS1 are deeply embedded in the binding pocket, yet they display different SARs (Figs. 3D and 6E). Specifically, Q1 of [S8R]SS1 is positioned near residues D219 and R208 of the receptor, forming a hydrogen bond with D219 and van der Waal contacts with R208. Additionally, Arg5 of [S8R]SS1 is proximate to the carboxyl group of D186, the carbonyl group on the main chain of S188 and the hydroxyl group of S58, forming salt-bridges and hydrogen bonds, respectively. In the [S8R]SS1 complex, W6 forms a hydrogen bond with the phenol hydroxyl group of Y210 and pi−pi stacking with the indole aromatic ring of W77 from the complementary subunit. At the same time, R8 forms multiple hydrogen bonds with Q79, S56 and E184, and a salt bridge with D186. Furthermore, R9 forms two hydrogen bonds with Y54 and Q183, respectively, while the side chain of W10 establishes hydrogen bonds with the hydroxyl group of S81 and side chain of Q139. Notably, the receptor's glycan, NAG, forms van der Waals interactions with the indole ring of W10 and the amide group at the C11 terminus, along with a weak hydrogen bond to the main-chain carbonyl of W10. In a previous study, we also identified one carbohydrate at the same position of the α3β4 nAChR involves direct interaction with the α-conotoxin RegIIA[31].

Results from mutagenesis studies on [S8R]SS1 validated the accuracy of our model (Fig. 7A and B and Supporting Information Fig. S35). Mutating residues at positions 5, 6, and 9 of [S8R]SS1

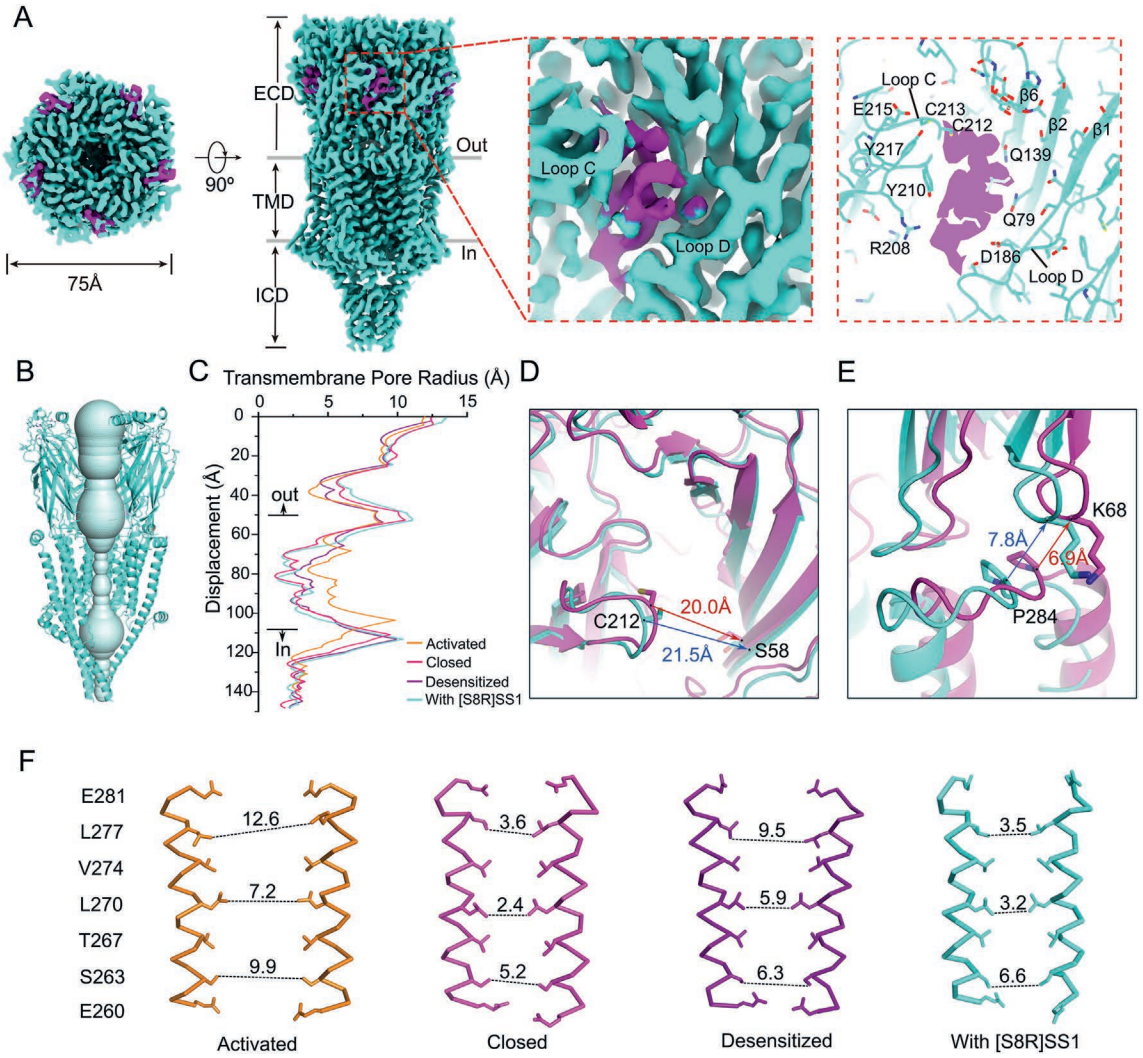

Figure 5 Cryo-EM results of α7/[S8R]SS1 complex. (A) The cryo-EM density map of the α7/[S8R]SS1 complex. Density beyond the receptor (magenta) was observed at the neurotransmitter binding site in the extracellular domain of the α7 receptor (cyan). (B) The structure of the $\alpha7^{EM}$ nAChR, with the pore calculated using Hole. (C) Structural comparison of $\alpha7^{EM}$ (cyan) with α7 nAChR in activated state (PDB ID: 7KOX, orange), resting state (PDB ID: 7KOO, magenta), and desensitized state (PDB ID: 7KOQ, purple). (D, E) Superimposition of α7 nAChR in resting state (magenta) and $\alpha7^{EM}$ nAChR (cyan). The distance used to characterize the C-loop opening in the α7 nAChR was measured between the Cα of Cys212 of the α4 subunit and the Cα of Ser58 of the β2 subunit (D). The activation distance was defined in the α4 subunit as the distance between the Cα atoms of K68 (at the β1−β2 linker) and P284 (at the M2−M3 linker) (E). (F) Two M2 helices are displayed for each conformational state, colored according to panel C. Side chains of pore-lining residues are shown, with pore diameters (in Å) indicated by dashed lines, as calculated by the HOLE.

to Ala, key sites for receptor interactions, resulted in significant decreases in potency, confirming the model's accuracy (Fig. S35A). Interestingly, mutation of W6 to Arg (W6R) maintained potency, likely due to its positively charged side chain forming a new cation−π interaction with the aromatic ring of W77 of the complementary subunit, which compensated for the lost interactions (Fig. 6E). Notably, the side chain of Q1 in [S8R]SS1 orients in different positions from that of SS1 in our model. To accurately validate these interactions, Q1 was mutated to E, R and Pyr, respectively, and the IC_{50} values for these analogues were tested at the $\alpha7^{EM}$ nAChR (Fig. 7A). The results suggest that [Q1Pyr] decreased potency by 2.1-fold, while the [Q1E] and [Q1R] increased the potency by 1.8 and 4.8-fold, respectively. The side chain of Q1 is oriented to R208 and D219, and thus introducing either a positively charged Arg or negatively charged Glu can increase potency by forming overall favorable electrostatic interactions with R208 and D219 nearby. By contrast the side chain of Pyr cannot form such interactions, and the potency is consequently decreased. P2 in SS1 forms strongly hydrophobic interactions with several aromatic residues (W77, Y115 and W171) and replacement of P2 with Hyp almost abolished inhibitory activity (Fig. 3). By contrast, P2 in [S8R]SS1 is oriented differently, forming only a weak hydrophobic interaction with Y217, and thus this position is more intolerant to a hydroxyl group. Indeed, replacement of P2 with Hyp only decreased the potency by 1.5-fold (Supporting Information Fig. S36). Furthermore, we compared four mutations of SS1 and [S8R]SS1 and observed opposite trends in the relative current amplitude

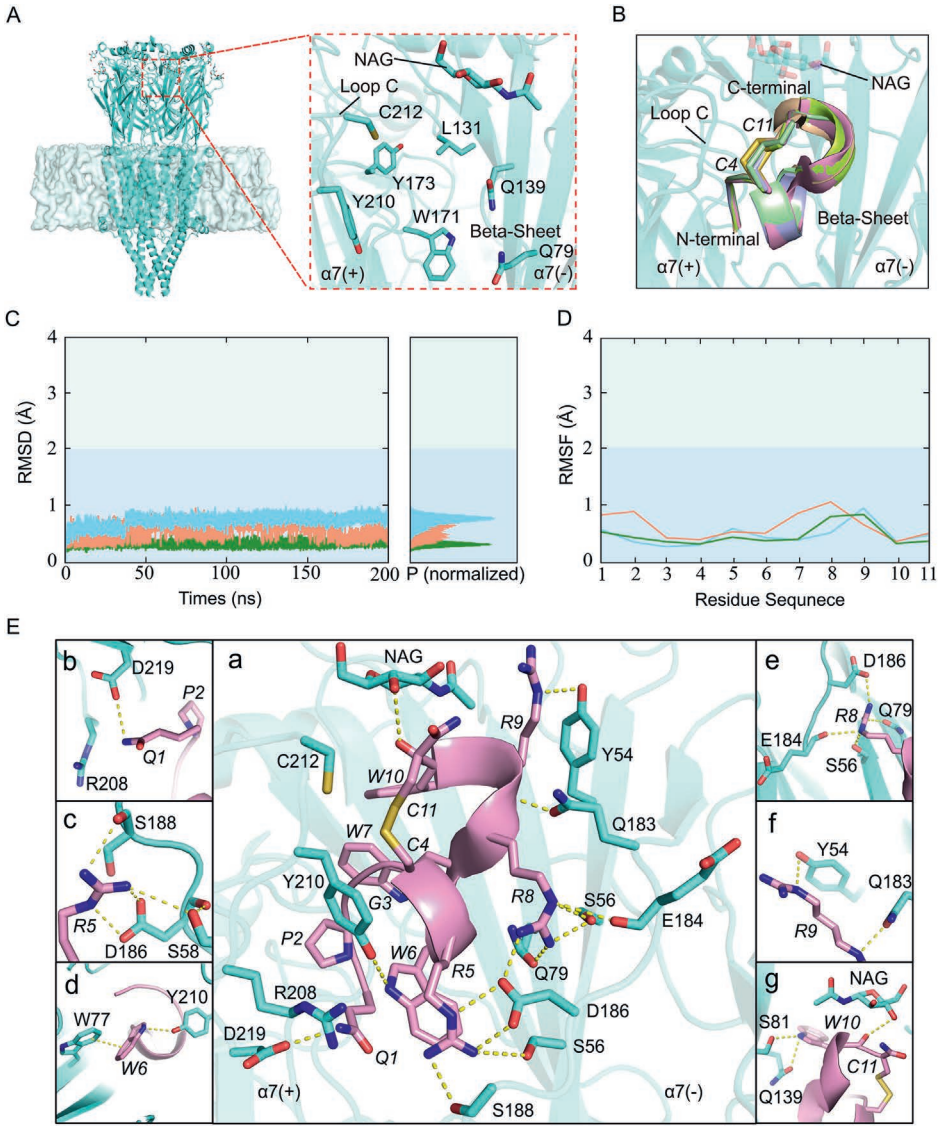

Figure 6 MD simulations of [S8R]SS1 binding to hα7 nAChR. (A) The orthosteric binding site of α7 nAChR was resolved in the cryo-EM structure. (B) Conformational variations of the [S8R]SS1 within the hα7/[S8R]SS1 complex were analyzed by capturing frames at 50 ns intervals in MD. (C) The left column shows RMSD of [S8R]SS1 in the hα7/[S8R]SS1 complex across 200 ns MD simulations, while the distribution of RMSD during the MD simulations are shown in the right column. (D) RMSF plot of [S8R]SS1 in the hα7/[S8R]SS1 complex across 200 ns MD simulations. Three distinct colors represent the independent triplicate MD simulation. (E) Binding mode of [S8R]SS1 with hα7 nAChR. The α7 subunit is shown in cyan and [S8R]SS1 is presented in pink.

differences (Fig. S35B). This finding supports distinct binding modes for SS1 and [S8R]SS1 with the hα7 nAChR.

We also conducted site-directed mutagenesis on the receptor to further validate the binding mode of [S8R]SS1 (Fig. 7B). To investigate the effects of receptor mutations, we measured the concentration−response curves of [S8R]SS1 on $\alpha7^{EM}$ nAChR mutants. The S56A mutation resulted in an IC_{50} of 233.5 nmol/L, significantly lower than the $\alpha7^{EM}$ receptor, which exhibited an IC_{50} of 659.1 nmol/L. This increase in potency can be attributed to the altered interactions at the binding site. While the S56A mutation led to the loss of a hydrogen bond between S56 and R8 of the [S8R]SS1, it also removed the hindrance that prevents closer contact between the guanidyl group of R8 and the side chain of Q79 on the receptor. Alternatively, the hydroxyl group of S56 forms a strong H-bond with the main chain of E184, and the S56A mutation can perturb local conformation change, and indirectly affect the potency of SS1 (Supporting Information Fig. S37). Q139 forms a hydrogen bond with the indole moiety of the W10 side chain on the [S8R]SS1, and this interaction is maintained when Q139 is mutated to Glu. Similarly, the backbone carbonyl group of E184 forms a hydrogen bond with R8 on the peptide, an interaction that remains unchanged after mutating E184 to Gln. As a result, the IC_{50} values for the Q139E and E184Q mutants are 852.1 and 555.5 nmol/L, respectively, both comparable to the $\alpha7^{EM}$ receptor. For the R208A mutant, the IC_{50} value was 1925.2 nmol/L, indicating a significant reduction in potency. This decrease can be primarily attributed to the loss of a hydrogen bond between R208 and the Q1 residue on the peptide.

The complex model generated by AF3 and MD simulations exhibits a high degree of similarity to the obtained cryo-EM map, further

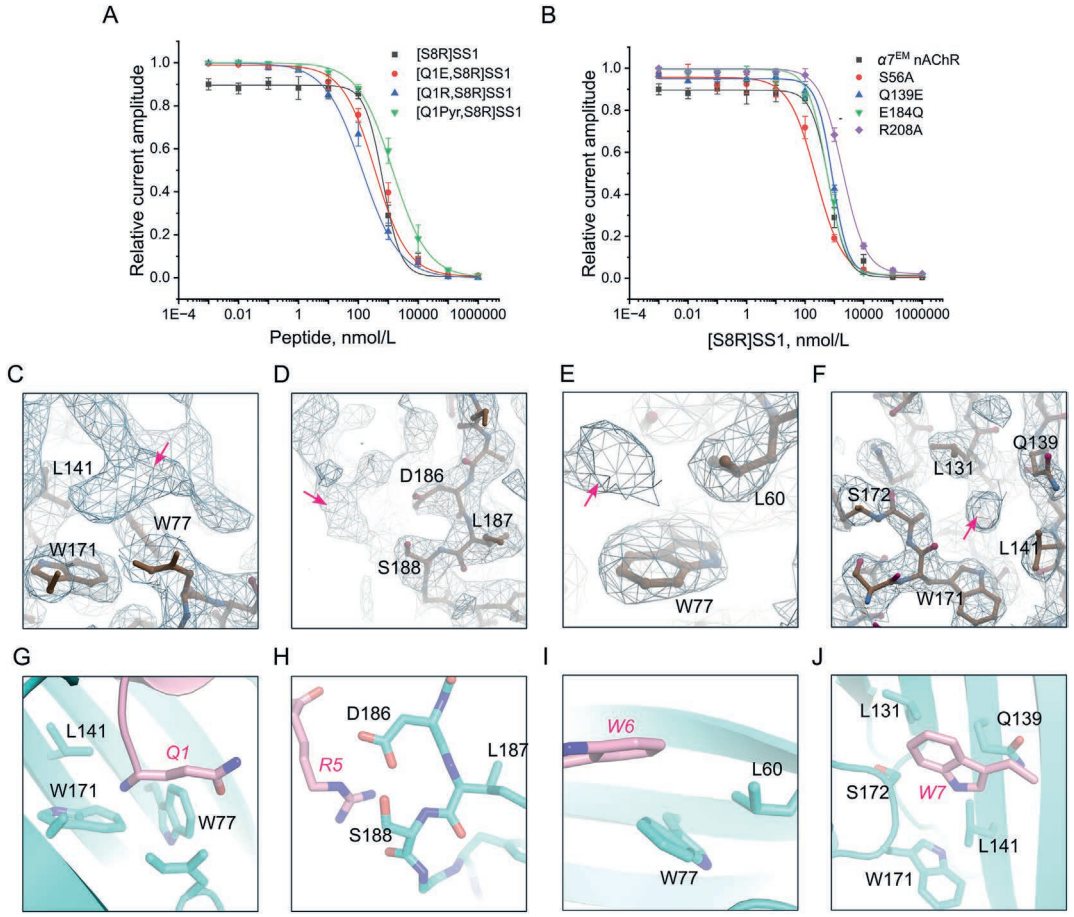

Figure 7 Validation of [S8R]SS1/hα7 complex structure. (A) Concentration−response of relative ACh-evoked current amplitude mediated by $\alpha7^{EM}$ nAChR in the presence of [S8R]SS1 (659.1 ± 92.9 nmol/L), [Q1E,S8R]SS1 (363.51 ± 102.6 nmol/L), [Q1R,S8R]SS1 (136.1 ± 27.8 nmol/L), [Q1Pyr,S8R]SS1 (1392.3 ± 259.9 nmol/L). (B) Concentration−response of relative ACh-evoked current amplitude mediated by mutant $\alpha7^{EM}$ nAChR in the presence of [S8R]SS1. The IC_{50} for $\alpha7^{EM}$, S56A, Q139E, E184Q, R208A are 659.1 ± 92.9, 233.5 ± 36.1, 852.1 ± 66.3, 555.5 ± 99.9, 1925.2 ± 196.4 nmol/L respectively. (C−F) The density features of the cryo-EM map. The ligand density is indicated by magenta arrows. (G−J) The amino acids corresponding to the density in panels C−F of the [S8R]SS1/α7 nAChR complex. The receptor is depicted in cyan, while the ligand is shown in pink. The values are presented as means ± SEM from at least three independent experiments.

supporting the reliability of our model (Fig. 7C−J). Specifically, in the cryo-EM map, a density near W171, L141, and W77 corresponds to the Q1 of [S8R]SS1 (Fig. 7C and G). In the complex model, the residue at position 5 is adjacent to D186 and S188, aligning with the density observed in the map (Fig. 7D and H). Likewise, the density near L60 and W77 corresponds to the indole ring of the tryptophan at position 6 (Fig. 7E and I). And the tryptophan at position 7 is surrounded by L131, Q139, L141, and W171, consistent with the density map (Fig. 7F and J). Overall, these data support the accuracy of our structural model of the [S8R]SS1/hα7 complex.

4. Discussion

Recent advances in transcriptomics and genomics have led to the discovery of a wealth of novel conopeptides[55,60,61], yet the pharmacological characterization of these peptides remains insufficient, underscoring the critical need for more efficient strategies to identify bioactive conopeptides with desired therapeutic properties from natural sources. This study establishes an innovative pipeline integrating deep learning, structural biology, computational modeling and electrophysiology to overcome key challenges in conopeptide drug discovery.

Our findings demonstrate that protein language models can effectively prioritize candidate peptides from limited datasets, while cryo-EM and computational modeling provide critical structural insights for rational optimization. The resulting [ΔQP, S8R]SS1 peptide represents a significant advance in conopeptide engineering, achieving nanomolar potency with minimalist structure while improving selectivity and stability through minimal structural modifications.

Several key implications emerge from this work. First, the success of transfer learning with ESM-2 suggests that existing protein language models can be adapted for specialized peptide discovery tasks, even with sparse target-specific data. Second, our integrative strategy, combining cryo-EM-derived receptor with computational modeling, effectively addresses the long-standing challenges of resolving flexible regions of complex structures and is rigorously validated through *in situ* mutagenesis and cryo-EM density fitting. Third, the structural resolution of a conopeptide-nAChR complex reveals previously unrecognized

interactions, including glycan-mediated contacts and hybrid receptor conformations, that expand our understanding of peptide-mediated channel inhibition. Finally, the pipeline's modular design—combining computational prediction with experimental validation—provides a generalizable framework that could be extended to other challenging membrane protein targets.

However, several limitations should be considered in this study. The predictive accuracy of our model was constrained by the heterogeneity of existing pharmacological data and the intrinsic promiscuity of many conopeptides across nAChR subtypes, highlighting the importance of establishing standardized experimental protocols to improve inter-laboratory reproducibility. While cryo-EM provided critical structural insights, the dynamic nature of both the peptide and receptor resulted in regions of unresolved density, necessitating complementary computational approaches. Additionally, the current study focused on *in vitro* characterization, and further work is needed to assess the therapeutic potential of these peptides *in vivo*.

Overall, the integrated methodology demonstrates how DL-driven approaches can transform natural product discovery, from overcoming data scarcity through transfer learning to enabling structure-based optimization. These advances provide a generalizable framework that bridges AI prediction with experimental validation, significantly accelerating the development of nature-inspired peptide therapeutics. Future applications could expand to other ion channel targets while incorporating emerging omics technologies for comprehensive peptide characterization.

Data availability statement

The mass spectrometric characterization and chromatographic profiles of the synthesized conopeptides (Supporting Information Figs. S38−S121), along with other Supporting Information, are provided in the Supporting Information. All code scripts and notebooks may be found at: https://github.com/gc-js/ConotoxinFinder. Users can submit their own sequences and get prediction results at HuggingFace Spaces: https://huggingface.co/spaces/oucgc1996/ConotoxinFinder. The cryo-EM model of α7 nAChR has been deposited in the Protein Data Bank (PDB) under accession code 9J37.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC, Nos. 82473832, 82122064, and U25A20178), Qingdao Science and Technology Demonstration Project (25-1-1-sfgc-2-nsh, China), the Qingdao Marine Science and Technology Center (No. 2022QNLM030003-4, China), the Fundamental Research Funds for the Central Universities (No. 202441004, China), and the Foundation of Shandong Province (No. 2022GJJLJRC02-046, China) to Rilei Yu; the National Laboratory of Biomacromolecules (No. 2023kf09, China) and the Chinese National Programs for Brain Science and Brain-like Intelligence Technology (No. 2022ZD0205800, China) to Yan Zhao; and the National Health and Medical Research Council (NHMRC) of Australia (GNT2009564) and the Australian Research Council (ARC) Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) to David J. Craik.

Author contributions

Rilei Yu, Yan Zhao conceived the project; Cheng Ge conducted conopeptides screening based on deep learning; Jinghui Zhang synthesized the conopeptides; Zhengji Yin performed Two-electrode voltage clamp experiments; Yue Li and Yan Zhao carried out cryo-electron microscopy experiments; Jinghui Zhang, Zhengji Yin, Zixuan Zhang and Pu Yuan processed the experimental data; Jinghui Zhang and Zixuan Zhang conducted molecular dynamic simulations, complex construction and analyzed the structures. Jinghui Zhang, Yan Zhao, Cheng Ge and Yue Li prepared the Figure. Jinghui Zhang, Rilei Yu, Zhengji Yin, Cheng Ge, Tao Jiang, and David J. Craik wrote and revised the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2025.12.035.

References

1. Zhan C, Li C, Wei X, Lu W, Lu W. Toxins and derivatives in molecular pharmaceutics: drug delivery and targeted therapy. *Adv Drug Deliv Rev* 2015;**90**:101−18.
2. Abraham N, Lewis RJ. Neuronal nicotinic acetylcholine receptor modulators from cone snails. *Mar Drugs* 2018;**16**:208.
3. Jin AH, Muttenthaler M, Dutertre S, Himaya SWA, Kaas Q, Craik DJ, et al. Conotoxins: chemistry and biology. *Chem Rev* 2019;**119**: 11510−49.
4. Oldrati V, Arrell M, Violette A, Perret F, Sprüngli X, Wolfender JL, et al. Advances in venomics. *Mol Biosyst* 2016;**12**:3530−43.
5. Calvete JJ. Venomics: integrative venom proteomics and beyond. *Biochem J* 2017;**474**:611−34.
6. Cao X, Liu T, Wang T, Wang X, Xu Z, Zhou L, et al. *De novo* screening and mirror image isomerization of linear peptides targeting α7 nicotinic acetylcholine receptor. *ACS Chem Biol* 2024;**19**: 592−8.
7. Xiang H, Bai L, Zhang X, Dan T, Cheng P, Yang X, et al. A facile strategy for the construction of a phage display cyclic peptide library for the selection of functional macrocycles. *Chem Sci* 2024;**15**: 11847−55.
8. Muttenthaler M, King GF, Adams DJ, Alewood PF. Trends in peptide drug discovery. *Nat Rev Drug Discov* 2021;**20**:309−25.
9. Kong XD, Tian C. *De novo* discovery of bicycles. *Nat Chem Biol* 2025;**21**:29−31.
10. Ma Y, Guo Z, Xia B, Zhang Y, Liu X, Yu Y, et al. Identification of antimicrobial peptides from the human gut microbiome using deep learning. *Nat Biotechnol* 2022;**40**:921−31.
11. Lee MH, Lee B, Park SE, Yang GE, Cheon S, Lee DH, et al. Transcriptome-based deep learning analysis identifies drug candidates targeting protein synthesis and autophagy for the treatment of muscle wasting disorder. *Exp Mol Med* 2024;**56**:904−21.
12. Bryant P, Elofsson A. Peptide binder design with inverse folding and protein structure prediction. *Commun Chem* 2023;**6**:229.
13. Brixi G, Ye T, Hong L, Wang T, Monticello C, Lopez-Barbosa N, et al. SaLT&PepPr is an interface-predicting language model for designing peptide-guided protein degraders. *Commun Biol* 2023;**6**: 1081.
14. Bhat S, Palepu K, Hong L, Mao J, Ye T, Iyer R, et al. *De novo* design of peptide binders to conformationally diverse targets with contrastive language modeling. *Sci Adv* 2025;**11**:eadr8638.
15. Chen T, Dumas M, Watson R, Vincoff S, Peng C, Zhao L, et al. PepMLM: target sequence-conditioned generation of therapeutic peptide binders *via* span masked language modeling. *ArXiv* 2024. Available from: https://doi.org/10.48550/arXiv.2310.03842.

16. Truong DP, Monroe LK, Williams RF, Nguyen HB. Machine learning framework for conotoxin class and molecular target prediction. *Toxins (Basel)* 2024;**16**:475.
17. Guo M, Li Z, Deng X, Luo D, Yang J, Chen Y, et al. ConoDL: a deep learning framework for rapid generation and prediction of conotoxins. *J Comput Aided Mol Des* 2024;**39**:4.
18. Ge C, Tae H-S, Zhang Z, Lu L, Huang Z, Wang Y, et al. CreoPep: a universal deep learning framework for target-specific peptide design and optimization. *ArXiv* 2025. Available from: https://arxiv.org/html/2505.02887v1. https://doi.org/10.48550/arXiv.2505.02887.
19. Zhu KF, Yuan C, Du YM, Sun KL, Zhang XK, Vogel H, et al. Applications and prospects of cryo-EM in drug discovery. *Mil Med Res* 2023;**10**:10.
20. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron* 2018;**99**:1129−43.
21. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493−500.
22. Mohan MK, Abraham N, R PR, Jayaseelan BF, Ragnarsson L, Lewis RJ, et al. Structure and allosteric activity of a single-disulfide conopeptide from Conus zonatus at human $\alpha 3\beta 4$ and $\alpha 7$ nicotinic acetylcholine receptors. *J Biol Chem* 2020;**295**:7096−112.
23. Jin AH, Cristofori-Armstrong B, Rash LD, Román-González SA, Espinosa RA, Lewis RJ, et al. Novel conorfamides from Conus austini venom modulate both nicotinic acetylcholine receptors and acid-sensing ion channels. *Biochem Pharmacol* 2019;**164**:342−8.
24. Ellison M, Gao F, Wang HL, Sine SM, McIntosh JM, Olivera BM. Alpha-conotoxins ImI and ImII target distinct regions of the human alpha7 nicotinic acetylcholine receptor and distinguish human nicotinic receptor subtypes. *Biochemistry* 2004;**43**:16019−26.
25. Ellison M, Feng ZP, Park AJ, Zhang X, Olivera BM, McIntosh JM, et al. Alpha-RgIA, a novel conotoxin that blocks the alpha9alpha10 nAChR: structure and identification of key receptor-binding residues. *J Mol Biol* 2008;**377**:1216−27.
26. Azam L, Dowell C, Watkins M, Stitzel JA, Olivera BM, McIntosh JM. Alpha-conotoxin BuIA, a novel peptide from *Conus bullatus*, distinguishes among neuronal nicotinic acetylcholine receptors. *J Biol Chem* 2005;**280**:80−7.
27. Wang S, Zhu X, Zhangsun M, Wu Y, Yu J, Harvey PJ, et al. Engineered conotoxin differentially blocks and discriminates rat and human $\alpha 7$ nicotinic acetylcholine receptors. *J Med Chem* 2021;**64**:5620−31.
28. Liang J, Tae HS, Zhao Z, Li X, Zhang J, Chen S, et al. Mechanism of action and structure−activity relationship of α-conotoxin Mr1.1 at the human $\alpha 9\alpha 10$ nicotinic acetylcholine receptor. *J Med Chem* 2022;**65**:16204−17.
29. Nicke A, Samochocki M, Loughnan ML, Bansal PS, Maelicke A, Lewis RJ. Alpha-conotoxins EpI and AuIB switch subtype selectivity and activity in native versus recombinant nicotinic acetylcholine receptors. *FEBS Lett* 2003;**554**:219−23.
30. Giribaldi J, Wilson D, Nicke A, El Hamdaoui Y, Laconde G, Faucherre A, et al. Synthesis, structure and biological activity of CIA and CIB, two α-conotoxins from the predation-evoked venom of *Conus catus. Toxins (Basel)* 2018;**10**:222.
31. Zheng M, Tae H-S, Xue L, Jiang T, Yu R. Mechanism of interactions between α-conotoxin RegIIA and carbohydrates at the human $\alpha 3\beta 4$ nicotinic acetylcholine receptor. *Mar Life Sci Technol* 2022;**4**:98−105.
32. Loughnan ML, Nicke A, Jones A, Adams DJ, Alewood PF, Lewis RJ. Chemical and functional identification and characterization of novel sulfated alpha-conotoxins from the cone snail *Conus anemone. J Med Chem* 2004;**47**:1234−41.
33. Whiteaker P, Christensen S, Yoshikami D, Dowell C, Watkins M, Gulyas J, et al. Discovery, synthesis, and structure activity of a highly selective alpha7 nicotinic acetylcholine receptor antagonist. *Biochemistry* 2007;**46**:6628−38.
34. Wang H, Li Y, Yang M, Zhou M. Synthesis and characterization of αM-conotoxin SIIID, a reversible human $\alpha 7$ nicotinic acetylcholine receptor antagonist. *Toxicon* 2022;**210**:141−7.
35. Luo S, Zhangsun D, Harvey PJ, Kaas Q, Wu Y, Zhu X, et al. Cloning, synthesis, and characterization of αO-conotoxin GeXIVA, a potent $\alpha 9\alpha 10$ nicotinic acetylcholine receptor antagonist. *Proc Natl Acad Sci U S A* 2015;**112**:E4026−35.
36. Xu S, Zhang T, Kompella SN, Yan M, Lu A, Wang Y, et al. Conotoxin αD-GeXXA utilizes a novel strategy to antagonize nicotinic acetylcholine receptors. *Sci Rep* 2015;**5**:14261.
37. Loughnan M, Nicke A, Jones A, Schroeder CI, Nevin ST, Adams DJ, et al. Identification of a novel class of nicotinic receptor antagonists: dimeric conotoxins VxXIIA, VxXIIB, and VxXIIC from. *Conus vexillum. J Biol Chem* 2006;**281**:24745−55.
38. Yu J, Zhu X, Zhang L, Kudryavtsev D, Kasheverov I, Lei Y, et al. Species specificity of rat and human $\alpha 7$ nicotinic acetylcholine receptors towards different classes of peptide and protein antagonists. *Neuropharmacology* 2018;**139**:226−37.
39. Quik M, Zhang D, McGregor M, Bordia T. Alpha7 nicotinic receptors as therapeutic targets for Parkinson's disease. *Biochem Pharmacol* 2015;**97**:399−407.
40. Wang S, Hu Y. $\alpha 7$ nicotinic acetylcholine receptors in lung cancer. *Oncol Lett* 2018;**16**:1375−82.
41. Waka Lin NHYS, Yasunari K. Role of $\alpha 7$-nicotinic acetylcholine receptor in normal and cancer stem cells. *Curr Drug Targets* 2012;**13**:656−65.
42. Yaniv D, Mattson B, Talbot S, Gleber-Netto FO, Amit M. Targeting the peripheral neural-tumour microenvironment for cancer therapy. *Nat Rev Drug Discov* 2024;**23**:780−96.
43. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123−30.
44. Zhang J, Xu H, Wang B, Zhang X, Fu L, Li Y, et al. *In silico* design of anti-tumor mini-protein targeting MDM2. *Chin Chem Lett* 2023;**34**:107871.
45. Cao Q, Ge C, Wang X, Harvey PJ, Zhang Z, Ma Y, et al. Designing antimicrobial peptides using deep learning and molecular dynamic simulations. *Brief Bioinform* 2023;**24**:bbad058.
46. Cao X, Liu T, Wang T, Wang X, Xu Z, Zhou L, et al. *De novo* screening and mirror image isomerization of linear peptides targeting $\alpha 7$ nicotinic acetylcholine receptor. *ACS Chem Biol* 2024;**19**:592−8.
47. Yu J, Xie J, Ma Y, Wei P, Zhang P, Tang Z, et al. Single amino acid substitution in Loop1 switches the selectivity of α-conotoxin RegIIA towards the $\alpha 7$ nicotinic acetylcholine receptor. *Mar Drugs* 2024;**22**:390.
48. Hone AJ, Santiago U, Harvey PJ, Tekarli B, Gajewiak J, Craik DJ, et al. Design, synthesis, and structure−activity relationships of novel peptide derivatives of the severe acute respiratory syndrome-coronavirus-2 spike-protein that potently inhibit nicotinic acetylcholine receptors. *J Med Chem* 2024;**67**:9587−98.
49. Jin AH, Vetter I, Dutertre S, Abraham N, Emidio NB, Inserra M, et al. MrIC, a novel alpha-conotoxin agonist in the presence of PNU at endogenous alpha7 nicotinic acetylcholine receptors. *Biochemistry* 2014;**53**:1−3.
50. Rogers JP, Luginbühl P, Pemberton K, Harty P, Wemmer DE, Stevens RC. Structure−activity relationships in a peptidic alpha7 nicotinic acetylcholine receptor antagonist. *J Mol Biol* 2000;**304**:911−26.
51. Atif HB, Alvi H, Naveed H. Masked language modeling for resource constrained biological natural language processing. *Annu Int Conf IEEE Eng Med Biol Soc* 2023;**2023**:1−5.
52. Halai R, Clark RJ, Nevin ST, Jensen JE, Adams DJ, Craik DJ. Scanning mutagenesis of α-conotoxin Vc1.1 reveals residues crucial for activity at the $\alpha 9\alpha 10$ nicotinic acetylcholine receptor. *J Biol Chem* 2009;**284**:20275−84.
53. Rojas L, Cabrera-Muñoz A, Gil Pradas D, González JB, Alonso-Del-Rivero M, González-González Y. Arginine substitution by alanine at the P1 position increases the selectivity of CmPI-II, a non-classical Kazal inhibitor. *Biochem Biophys Rep* 2021;**26**:101008.

54. Ho TNT, Lee HS, Swaminathan S, Goodwin L, Rai N, Ushay B, et al. Posttranslational modifications of α-conotoxins: sulfotyrosine and C-terminal amidation stabilise structures and increase acetylcholine receptor binding. *RSC Med Chem* 2021;**12**:1574−84.
55. Yao G, Peng C, Zhu Y, Fan C, Jiang H, Chen J, et al. High-throughput identification and analysis of novel conotoxins from three vermivorous cone snails by transcriptome sequencing. *Mar Drugs* 2019;**17**:193.
56. Chelius D, Jing K, Lueras A, Rehder DS, Dillon TM, Vizel A, et al. Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Anal Chem* 2006;**78**:2370−6.
57. Li T, Tae HS, Chen S, Li X, Liang J, Pan T, et al. Development of an intravenously stable disulfide-rich peptide for the treatment of chemotherapy-induced neuropathic pain. *J Med Chem* 2024;**67**: 18741−52.
58. Noviello CM, Gharpure A, Mukhtasimova N, Cabuco R, Baxter L, Borek D, et al. Structure and gating mechanism of the α7 nicotinic acetylcholine receptor. *Cell* 2021;**184**:2121−34.
59. Yu R, Tae HS, Xu Q, Craik DJ, Adams DJ, Jiang T, et al. Molecular dynamics simulations of dihydro-β-erythroidine bound to the human $\alpha4\beta2$ nicotinic acetylcholine receptor. *Br J Pharmacol* 2019;**176**: 2750−63.
60. Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. *BMC Genom* 2011;**12**:60.
61. Lavergne V, Dutertre S, Jin AH, Lewis RJ, Taft RJ, Alewood PF. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. *BMC Genom* 2013;**14**:708.