HiLL PUBLISHING

Advance in Biological Research, 2026, 7(1), 54-58
https://www.hillpublisher.com/journals/abr/
ISSN Online: 2635-0319

Computational Intelligence-driven DPP-4 Inhibitory Peptide Prediction and Bispecific Antibody Design: Integrating Deep Learning, Molecular Dynamics, and Nanodisc Validation

Taozhi Wang

The Alice Smith International School, No.3, Jalan Equine, Taman Equine 43300, Seri Kembangan, Selangor Darul Ehsan, Malaysia.

How to cite this paper: Taozhi Wang. (2026). Computational Intelligence-driven DPP-4 Inhibitory Peptide Prediction and Bispecific Antibody Design: Integrating Deep Learning, Molecular Dynamics, and Nanodisc Validation. *Advance in Biological Research*, 7(1), 54-58.
DOI: 10.26855/abr.2026.06.006

Received: April 27, 2026
Accepted: May 29, 2026
Published: June 30, 2026

Corresponding author: Taozhi Wang, The Alice Smith International School, No.3, Jalan Equine, Taman Equine 43300, Seri Kembangan, Selangor Darul Ehsan, Malaysia.

© 2026 by the author(s).
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution- NonCommercial-NoDerivatives (CC BY-NC-ND) license, which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and is not modified or adapted.
https://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract

Investigating peptides and antibodies targeting DPP-4 in vitro experiments requires the presence of the native lipid bilayer, as catalytic activity depends on it. We propose a methodology that involves molecular docking, 100-ns molecular dynamics simulations of DPP-4 in a POPC lipid bilayer mimicking nanodisc conditions, a novel dual-channel BiLSTM deep learning model for predicting DPP-4 inhibitor peptides, CDR affinity maturation based on Rosetta, and optimizing the structure of DPP-4 antibodies. For predicting peptides, we developed a novel dual-channel model based on the ESM-2 protein language model and docking data with expected R^2 and Spearman correlations of 0.80-0.85 and 0.70-0.80, respectively, based on the same peptide data used in the experiment within the nanodisc system. In optimizing the structure of DPP-4 antibodies, we use IgFold for predicting the structure of DPP-4 antibodies, ClusPro/HADDOCK for mapping binding sites, and virtual saturation mutagenesis for the optimization of CDRs, resulting in a 2-5 times increased binding affinity. A bispecific DPP-4 inhibitor antibody with GLP-1R agonist activity is designed following the principle of knob-into-hole and using AlphaFold-Multimer.

Keywords

DPP-4; deep learning; molecular dynamics; bispecific antibody; nanodisc

1. Introduction

Type 2 diabetes mellitus and obesity share an incretin-related pathogenic axis in which DPP-4 rapidly inactivates incretin hormones [1]. DPP-4 inhibition and GLP-1R agonism provide complementary approaches, suggesting a bispecific strategy that combines both mechanisms. Detergent solubilization can disturb membrane-protein conformation [2], whereas nanodiscs better preserve a native-like lipid environment but still require expensive and lengthy wet-lab screening. Molecular docking [3], membrane molecular dynamics [4], protein language models [5], and complex structure prediction can help prioritize peptide and antibody candidates. This work therefore develops a pipeline integrating virtual screening, membrane-embedded molecular dynamics, a dual-channel BiLSTM model, and Rosetta-based CDR affinity maturation [6].

2. Biological Foundation

2.1 DPP-4 Structure and Incretin Axis

DPP-4 (CD26) is a type II transmembrane protein with cytoplasmic, transmembrane, and extracellular catalytic regions. Its catalytic triad comprises Ser630, Asp708, and His740 and recognizes substrates with proline or alanine at the second residue from the N-terminus [10]. The functional enzyme exists as a homodimer, while the surrounding lipid environment can influence dimer formation, conformation, and catalytic behavior.

2.2 Nanodisc and Food-derived Peptides

Nanodiscs generated with membrane scaffold proteins (MSPs) create a soluble, detergent-free phospholipid environment that can maintain membrane-protein conformation. Food-derived peptides from cow milk, soy protein, and fish are known natural DPP-4 inhibitors with micromolar IC50 values [8]. Machine-learning models have successfully screened bovine whey hydrolysate peptides, while atom-level encoding has improved the classification of short inhibitory fragments [9]. These findings support combining food-derived peptide libraries with structural screening.

2.3 Rationale for Computational Approach

From a structural perspective, DPP-4 has a high-resolution crystal structure (PDB 1X70; 2.1 Å) suitable for docking and molecular dynamics studies. From a sequence perspective, public databases such as BIOPEP-UWM and published studies provide more than 1,000 peptide sequences with quantitative or reported inhibitory activities, supporting supervised deep-learning models [8-10].

3. Computational Methods

3.1 Data Mining and Peptide Library Construction

The positive sample set was assembled from BIOPEP-UWM entries annotated with experimentally determined DPP-4 inhibitory activity. After removing duplicate sequences and entries lacking quantitative IC50 values, approximately 1,200 peptides were retained. Each IC50 value was converted to pIC50 = −log10(IC50 [M]) to normalize the distribution and compress its dynamic range. The combined dataset was split 8:1:1 into training, validation, and test partitions using stratified sampling to preserve the pIC50 distribution. The candidate library was expanded to approximately 5,000 sequences by adding computationally generated fragments from simulated enzymatic hydrolysis of food proteins.

3.2 Molecular Docking

Virtual screening was performed with AutoDock Vina 1.2.0 using the prepared DPP-4 structure. A 20 × 20 × 20 Å grid was centered on the catalytic triad (Ser630, Asp708, and His740) and covered the two substrate-binding sites. Three-dimensional peptide conformers were prepared, and the top 100 candidates with docking scores below −7.5 kcal/mol were selected for the next stage. In addition to pose ranking, four structural descriptors were calculated: hydrogen-bond number (NHB), hydrophobic surface area (Ahydro), catalytic-triad contacts (Ncat), and buried surface area (BSA), as defined in Equation (1):

$$\mathbf{F}_{struct} = [N_{HB}, A_{hydro}, N_{cat}, BSA] \tag{1}$$

3.3 Molecular Dynamics Simulation

Molecular dynamics simulations were performed on the top 20 peptide-DPP-4 complexes using GROMACS 2023 and the CHARMM36m force field [4]. Each complex was placed in a 128 × 128 POPC phospholipid bilayer and solvated in TIP3P water containing 0.15 M NaCl. After minimization and equilibration, production simulations were run under NPT conditions at 310 K and 1 atm, using a 2 fs time step and PME electrostatics with a 1.2 nm cutoff. Complexes with RMSD drift above 3 Å were considered unstable. RMSF, intermolecular hydrogen-bond occupancy above 30%, and MM-PBSA per-residue binding-energy contributions were analyzed, as defined in Equation (2):

$$\Delta G_{bind} = \Delta E_{vdW} + \Delta E_{elec} + \Delta G_{polar} + \Delta G_{nonpolar} - T\Delta S \tag{2}$$

Here, ΔEvdW and ΔEelec represent van der Waals and electrostatic interaction energies, ΔGpolar and ΔGnonpolar denote polar and nonpolar solvation contributions, and TΔS is the entropic penalty estimated by normal-mode analysis.

3.4 Deep Learning Activity Prediction

The dual-channel network combines peptide sequence features and docking-derived structural information. Residue-level ESM-2 embeddings are processed by a two-layer bidirectional LSTM to capture sequential dependencies, whereas NHB, Ahydro, Ncat, and BSA are processed by a two-layer MLP with ReLU activation. The two representations are concatenated and passed through fully connected layers to predict pIC50. Model performance is evaluated using R^2, RMSE, Pearson r, and Hit Rate, defined as the overlap between the predicted and experimental top-50 peptides.

4. AI-assisted Bispecific Antibody Design

4.1 CDR Optimization

Fv structures were predicted with IgFold without templates [11]. The predicted antibody was docked to DPP-4 using ClusPro for rigid-body sampling, followed by HADDOCK refinement with epitope residues as ambiguous interaction restraints [12]. After interface identification, residues in CDR-H1, CDR-H2, CDR-H3, and CDR-L3 underwent virtual saturation mutagenesis with Rosetta InterfaceDesign. Mutations with $\Delta\Delta G < -1.0$ REU were retained for combinatorial analysis, and the top five variants underwent 50 ns all-atom molecular dynamics using the GROMACS/CHARMM36m protocol in Section 3.3. Variants with RMSD > 2.5 Å, hydrogen-bond retention < 80%, or interfacial instability were excluded.

4.2 Bispecific Assembly and Structure Prediction

For bispecific assembly, the optimized anti-DPP-4 Fab was connected to a GLP-1R agonist module. Heavy-chain heterodimerization used the knobs-into-holes approach, with T366W on the knob chain and T366S, L368A, and Y407V on the hole chain. The GLP-1R agonist module was attached to the second heavy chain through a flexible (G4S)3 linker. The complete bispecific construct was modeled from paired sequences using AlphaFold-Multimer [7, 13], and the spatial arrangement of the two functional ends was assessed.

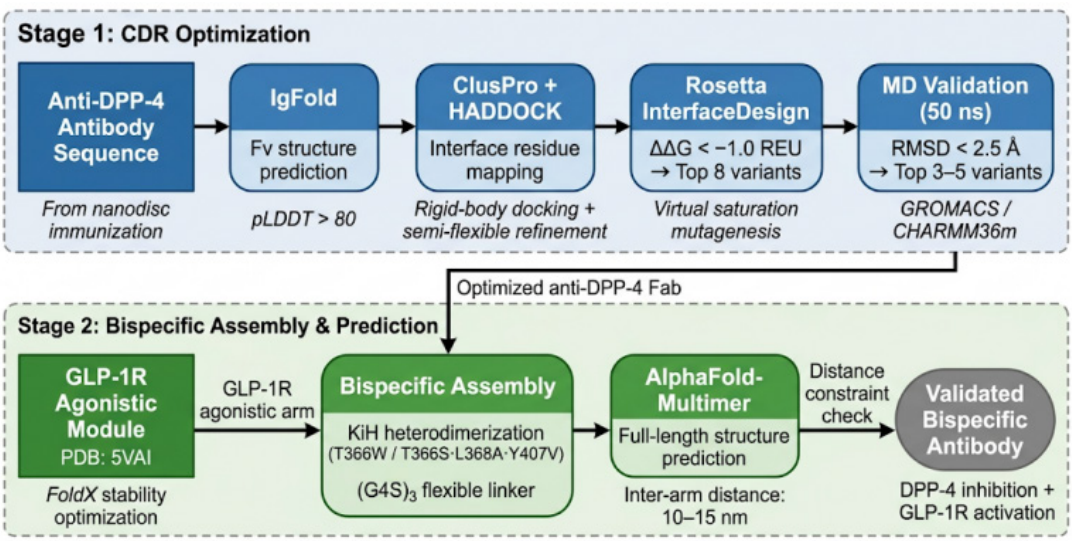

Figure 1. Workflow of the AI-assisted bispecific antibody design pipeline.

5. Expected Performance and Feasibility Analysis

5.1 Peptide Prediction

Docking is expected to reduce the library from approximately 5,000 peptides to a top-100 candidate set using a docking-score threshold of −7.5 kcal/mol. Anticipated performance across five modeling approaches is summarized in Table 1,

with the dual-channel BiLSTM expected to outperform sequence-only and conventional machine-learning models. The top 50 predictions will then be tested against nanodisc-embedded DPP-4, with an expected Spearman ρ of 0.70-0.80.

Table 1. Anticipated model performance comparison for DPP-4 inhibitory peptide activity prediction (five-fold cross-validation)

Model	Input Features	Expected R^2	Expected RMSE	Expected Pearson r
Dual-channel BiLSTM	ESM-2 + docking descriptors	0.80-0.85	0.32-0.38	0.89-0.92
Transformer (4-layer)	ESM-2 embedding only	0.76-0.81	0.36-0.42	0.86-0.90
1D-CNN	One-hot + docking descriptors	0.72-0.78	0.40-0.46	0.83-0.87
XGBoost	Physicochemical + docking	0.68-0.74	0.44-0.50	0.79-0.84
Random Forest	Physicochemical + docking	0.64-0.70	0.48-0.54	0.75-0.80

5.2 Antibody Optimization

Fifty-nanosecond molecular dynamics validation is expected to reduce the candidate mutants to 3-5 by excluding variants with CDR-H3 displacement, interfacial instability, or substantial hydrogen-bond loss. The best mutant is anticipated to improve binding by 2-5-fold and reduce the dissociation constant (KD) from approximately 45 nM for the parent antibody to 10-20 nM.

5.3 Integrated Pipeline Efficiency

The combined docking, molecular dynamics, and deep-learning cascade is expected to reduce approximately 5,000 peptides to fewer than 500 experimental candidates, lowering the screening burden by nearly 90%. Computational-experimental agreement may improve in nanodiscs because the simulations include a POPC membrane similar to the validation environment rather than a detergent-solubilized enzyme. Estimated computational costs include peptide-complex simulations, five-fold model training, and AlphaFold-Multimer prediction. Overall, the workflow is expected to require about two weeks on an 8-16-GPU HPC system, compared with three to six months for broad experimental screening.

6. Discussion

The proposed pipeline combines environmental context, sequence learning, and structural optimization in a single workflow. Membrane-embedded molecular dynamics should better approximate nanodisc reconstitution than detergent-based models, while the dual-channel model uses complementary sequence and docking information. A bispecific construct targeting DPP-4 while activating GLP-1R may preserve complementary therapeutic effects [10]. Several limitations remain. CHARMM36m may not fully represent short flexible peptides, and enhanced sampling such as REMD would increase computational cost. ESM-2 was trained mainly on full-length proteins, so embeddings for peptides shorter than 10 amino acids may be less reliable. The dataset of approximately 1,200 peptides may limit generalization, and AlphaFold-Multimer remains uncertain for flexible linkers with pLDDT < 50.

7. Conclusion

We propose an integrated computational framework for DPP-4 inhibitory peptide prediction and bispecific antibody design using docking, membrane-embedded molecular dynamics, a dual-channel BiLSTM model, and Rosetta-based CDR maturation. The workflow treats the POPC bilayer as part of the modeling context and combines docking-derived descriptors with ESM-2 sequence embeddings. The antibody-design stage extends peptide discovery toward therapeutic protein engineering. Future work will validate the pipeline experimentally in nanodiscs, compare peptide-focused language models with ESM-2, and expand the labeled training dataset.

References

[1] Holst JJ, Rosenkilde MM. GIP as a therapeutic target in diabetes and obesity: insight from incretin co-agonists. J Clin Endocrinol Metab. 2020;105(8):e2710-e2716.

[2] Pettersen JM, Yang Y, Robinson AS. Advances in nanodisc platforms for membrane protein purification. Trends Biotechnol. 2023;41(8):1041-1054.

[3] Eberhardt J, Santos-Martins D, Tillack AF, et al. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model. 2021;61(8):3891-3898.

[4] Páll S, Zhmurov A, Bauer P, et al. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. J Chem Phys. 2020;153(13).

[5] Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023; 379(6637):1123-1130.

[6] Boorla VS, Chowdhury R, Ramasubramanian R, et al. De novo design and Rosetta-based assessment of high-affinity antibody variable regions (Fv) against the SARS-CoV-2 spike receptor binding domain (RBD). Proteins. 2023;91(2):196-208.

[7] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

[8] Zhang Y, Zhu Y, Bao X, et al. Mining bovine milk proteins for DPP-4 inhibitory peptides using machine learning and virtual proteolysis. Research. 2024;7:0391.

[9] Wang D, Jin J, Li Z, et al. StructuralDPPIV: a novel deep learning model based on atom structure for predicting dipeptidyl peptidase-IV inhibitory peptides. Bioinformatics. 2024;40(2):btae057.

[10] Zhang M, Zhu L, Wu G, et al. Food-derived dipeptidyl peptidase IV inhibitory peptides: production, identification, structure-activity relationship, and their potential role in glycemic regulation. Crit Rev Food Sci Nutr. 2024;64(8):2053-2075.

[11] Ruffolo JA, Chu LS, Mahajan SP, et al. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. Nat Commun. 2023;14(1):2389.

[12] Vajda S, Yueh C, Beglov D, et al. New additions to the ClusPro server motivated by CAPRI. Proteins. 2017;85(3):435-444.

[13] Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2021. doi:10.1101/2021.10.04.463034