

Zhixing Zhang

**PhD Student in Pharmacy | Machine Learning for Therapeutic Peptides and Computational Protein Design
| Research Software Development**

School of Pharmacy, East China University of Science and Technology, Shanghai, China

Expected PhD completion: June 2027

Email: 2145573258@qq.com | Phone: +86 182 8721 1875

[GitHub](#) | [PepDiana](#) | [Bilingua-DPP-IV](#)

Research Profile

Research focuses on machine learning for therapeutic peptide discovery and computational protein design, with experience across small molecules, peptides, proteins and antibodies. Led the research and implementation of peptide prediction models, a literature evidence-mining agent and a peptide database, spanning data processing, method evaluation, manuscript preparation, full-stack development and deployment. Three years of experience managing laboratory computing infrastructure and more than 20 research websites and services. Collaborated with research and experimental teams on antibody design, protein structure analysis, fluorescent probe engineering and small-molecule screening.

Research Interests

- Sequence and structural representations of therapeutic peptides; prediction of activity and developability properties; computational methods across small molecules, peptides and proteins.
- Computational protein and antibody design, structural evaluation, and collaboration between computational and experimental researchers.
- Literature evidence extraction, scientific databases and reusable computational workflows for drug discovery.

Education

East China University of Science and Technology — School of Pharmacy

Integrated master's–doctoral programme in Pharmacy | September 2022–present

PhD in Pharmacy expected June 2027

Supervisor: Professor Yun Tang, School of Pharmacy, East China University of Science and Technology
(ytang234@ecust.edu.cn)

Research: AI-assisted drug design, focusing on therapeutic peptide data mining, databases and predictive methods.

Beijing University of Chemical Technology

Bachelor's degree in Pharmaceutical Engineering | September 2018–June 2022

Selected Research Experience

PepMinerva — Literature Evidence Mining for Therapeutic Peptides

Principal researcher and software developer; first author | East China University of Science and Technology

April–July 2026 | Manuscript under review at *Protein Science*

- Designed and implemented a literature extraction workflow integrating retrieval, full-text parsing, content routing, structured extraction, evidence linking and export to capture peptide entities and activity measurements.
- Built and annotated Gold50, a benchmark comprising 50 papers, 165 peptide entities and 746 activity records. Compared three model backends and two processing routes using record recovery, core-field agreement and numeric matching.

- Achieved a record-level F1 of 0.843 in the frozen Gold50 full-text workflow, including fallback processing; this measures record detection and matching rather than complete-field accuracy.
- Developed the software, command-line interface and Agent Skill interface; conducted evaluations and drafted the manuscript. Contributed extracted literature records to PepDiana.

[Code and documentation](#)

PepDiana — A Therapeutic Peptide Database with Source Traceability

Principal researcher and full-stack developer | East China University of Science and Technology

June 2025–June 2026: initial release; maintenance and revision ongoing

- Led integration of peptide data from existing databases and literature evidence extracted with PepMinerva, linking peptide records, activity observations and their sources.
- Built the June 2026 release, with 82,224 peptide records and 376,740 activity observations in the integrated data layer; maintained separate provenance for aggregated database records and primary literature evidence.
- Independently implemented data processing, backend services, APIs and frontend interfaces, and deployed and maintained the service for browser-based and programmatic access.
- Launched the initial website in June 2026; ongoing work addresses display and field-level issues and prepares the database manuscript.

[Database website](#)

Bilingua-DPP-IV — Peptide Prediction through Sequence and Molecular Representations

Principal researcher and software developer; first author | East China University of Science and Technology

November 2024–June 2025 | Paper accepted by *ChemMedChem* in October 2026

- Curated DPP-IV inhibitory peptide data and designed a dual-branch convolutional network with gated fusion, using representations from pretrained ESM-C and Uni-Mol2 models for activity classification.
- Performed data cleaning, model implementation, training, comparative evaluation and ablation studies. Obtained an MCC of 0.764 and ROC AUC of 0.963 on a stratified random holdout set of 300 records.
- Integrated the model into a web service, developing and deploying both frontend and backend components; drafted the manuscript and prepared the software copyright registration application.

[Web service](#) | [Code](#)

Research Computing and Software Engineering

Laboratory Cluster and Website Administration | School of Pharmacy, East China University of Science and Technology

Three years of experience, concurrent with doctoral research

- Administered Linux servers, computing resources and task scheduling across one control node and 12 compute nodes, including four GPU nodes with 11 GPUs in total, supporting CPU and GPU workloads.
- Deployed and managed more than 20 research websites and services, including laboratory admetSAR platforms; handled Nginx/Apache configuration, deployment and operational maintenance.
- Supported admetSAR platforms with cumulative visit counts exceeding one million across versions 1.0–3.0, including more than 170,000 for [admetSAR 3.0](#) as of October 2026. These are platform lifetime visit counts.
- Maintained node and resource-use documentation and supported laboratory users through shared guides on server access, remote connections, environment configuration, website deployment and maintenance.

Industry Collaboration and Internship

Computational Design of Anti-TSLP Antibodies — Biotechnology Startup

Doctoral Research Intern, flexible schedule

October 2025–October 2026

- Developed the computational workflow for antibody CDR redesign, integrating RFantibody/RFdiffusion backbone generation, ProteinMPNN sequence design and structure prediction for candidate generation and prioritisation.
- Evaluated structures with RoseTTAFold2 and AlphaFold3, combining multiple random seeds, local interface-contact analysis and molecular visualisation to assess confidence and interface features.
- Used HADDOCK, PRODIGY and complementary structure-prediction tools to examine ranking and consistency across modelling settings and inform experimental priorities.
- Independently handled workflow adaptation, input generation, Slurm batch jobs and result parsing; delivered candidate lists, structural visualisations and technical reports, and discussed experimental feedback and interpretation with the company's experimental team.

Protein Conformation and Antigen–Antibody Interface Analysis — Clover Biopharmaceuticals Collaboration

University–industry collaborative research | East China University of Science and Technology

September 2022–September 2026

- Performed whole-complex and chain-wise structural superposition, interface-residue comparisons and analysis of predicted complexes; prepared figures, editable molecular sessions and reports to support interpretation and follow-up validation.
- Prepared and delivered theoretical introductions and practical training on AlphaFold2 structure prediction, molecular dynamics and protein developability, using case studies to discuss method selection and prediction limitations.

Additional Computational Research

Computational Design of MatC-Based Fluorescent Malate Probes

Collaborative research | East China University of Science and Technology

September 2022–July 2025

- Contributed to Rosetta-based binding-pocket variant design, candidate selection and result organisation; incorporated collaborators' fluorescence measurements into iterative analysis, gaining experience in protein–small-molecule recognition and structure-guided protein engineering.

Computational Screening of NMDA Receptor-Targeting Small Molecules

Team project | Shanghai International Computational Biology Innovation Competition, 2023 edition

- Participated in screening small molecules targeting GluN1/GluN3A receptors, using structural and binding-site information for candidate prioritisation. The team workflow included initial deep-learning-based screening, molecular docking and MM/GBSA ranking.
- The team's experimental feedback summary reported three candidates with IC₅₀ values below 100 μ M, with a minimum of 2.45 μ M, providing preliminary signals for retesting and experience linking computational screening to experimental evaluation.

Publications and Manuscripts

Applicant's name in bold; † denotes equal contribution. First/co-first-author papers and other co-authored papers are grouped separately and listed from most recent to oldest. Accepted and under-review manuscripts are identified explicitly.

First and Co-First Author Papers

1. **Zhixing Zhang**, Qiule Yu, Chen Yang, Hao Duan, Jiaojiao Fang, Jiahao Xu, Changda Gong, Weihua Li, Guixia Liu, Yun Tang. Bilingua-DPP-IV: A bimodal fusion convolutional neural network model for prediction of dipeptidyl peptidase IV inhibitory peptides. *ChemMedChem*. **Accepted 11 October 2026; not yet published online.**
2. **Zhixing Zhang**, Hua Deng, Yun Tang. Applications and challenges of artificial intelligence in the development of anticancer peptides. *Journal of China Pharmaceutical University*. 2024;55(3):347–356. doi:10.11665/j.issn.1000-5048.2024040201. [Review; in Chinese].
3. Yang Tan†, **Zhixing Zhang**†, Mingchen Li†, Fei Pan, Hao Duan, Zijie Huang, Hua Deng, Zhuohang Yu, Chen Yang, Guoyang Shen, Peng Qi, Chengyuan Yue, Yuxian Liu, Liang Hong, Huiqun Yu, Guisheng Fan, Yun Tang. MedChatZH: A tuning LLM for traditional Chinese medicine consultations. *Computers in Biology and Medicine*. 2024;172:108290. doi:10.1016/j.compbiomed.2024.108290. Co-first author, listed second.

Other Co-Authored Papers

4. Qiule Yu, **Zhixing Zhang**, Weihua Li, Guixia Liu, Yun Tang. ToxiCompass: an evidence-tracked benchmark for short-peptide toxicity screening and confidence-aware assay prioritization. *Briefings in Bioinformatics*. 2026;27(5):bbag526. doi:10.1093/bib/bbag526.
5. Qiule Yu, **Zhixing Zhang**, Xiang Li, Weihua Li, Guixia Liu, Yun Tang. ncAA-RepDistill: a chemistry-native SMILES representation for permeability prediction and local edit ranking of non-canonical cyclic peptides. *Journal of Cheminformatics*. 2026. Published online. doi:10.1186/s13321-026-01311-5.
6. Zhili Pu, Lantian Zhang, Hao Duan, **Zhixing Zhang**, Keyun Zhu, Yongqi Fan, Ruihui Hou, Tong Ruan, Yun Tang. MCLE-Mol: Empowering LLM with Molecular Comprehension and Low-Cost Continual Evolution for Interpretable Property Prediction. In: *Findings of the Association for Computational Linguistics: ACL 2026*. 2026:5305–5333. doi:10.18653/v1/2026.findings-acl.262.
7. Jiaojiao Fang, Changda Gong, Keyun Zhu, Xiang Li, Chen Yang, **Zhixing Zhang**, Guixia Liu, Yun Tang, Weihua Li. Prediction of Intravenous Pharmacokinetic Parameters across Multiple Species by a Multifidelity Deep Learning Framework. *Journal of Chemical Information and Modeling*. 2026;66(2):1035–1049. doi:10.1021/acs.jcim.5c02546.
8. Chen Yang, Changda Gong, **Zhixing Zhang**, Jiaojiao Fang, Weihua Li, Guixia Liu, Yun Tang. In silico prediction of pKa values using explainable deep learning methods. *Journal of Pharmaceutical Analysis*. 2025;15(6):101174. doi:10.1016/j.jpha.2024.101174.
9. Qiule Yu, **Zhixing Zhang**, Guixia Liu, Weihua Li, Yun Tang. ToxGIN: an In silico prediction model for peptide toxicity via graph isomorphism networks integrating peptide sequence and structure information. *Briefings in Bioinformatics*. 2024;25(6):bbae583. doi:10.1093/bib/bbae583.
10. Yujia Tian, **Zhixing Zhang**, Aixia Yan. Discovering the Active Ingredients of Medicine and Food Homologous Substances for Inhibiting the Cyclooxygenase-2 Metabolic Pathway by Machine Learning Algorithms. *Molecules*. 2023;28(19):6782. doi:10.3390/molecules28196782.

Manuscript Under Review

1. **Zhixing Zhang**, Chen Yang, Qiule Yu, Hao Duan, Jiaojiao Fang, Jiahao Xu, Weihua Li, Guixia Liu, Yun Tang. PepMinerva: an evidence-grounded agent for extracting experimentally validated therapeutic peptide activities from full-text literature. **Under review at Protein Science**. First author.

Research Software and Data Resources

- **Bilingua-DPP-IV** — DPP-IV inhibitory peptide prediction and web service. Contributions: model research, full-stack development, deployment, manuscript preparation and software copyright registration. [Web service](#) | [Code](#).
- **PepMinerva** — Literature evidence extraction, the Gold50 benchmark and Agent Skill integration. Contributions: workflow and software development, annotation, evaluation and manuscript preparation. [Code and documentation](#).
- **PepDiana** — Peptide data integration, source traceability, website and API access. Contributions: data processing, full-stack development, deployment and maintenance. [Database](#).

Selected software copyright registration: Bilingua-DPP-IV V1.0, a deep-learning-based prediction system for dipeptidyl peptidase IV inhibitory peptides. Registration no. **2025SR2002825**, 16 October 2025. Copyright holder: East China University of Science and Technology. Responsible for software development and the registration application.

Technical Skills

- **Machine learning and scientific programming:** Python, PyTorch; data cleaning, feature and pretrained representation use, predictive modelling, training, evaluation and visualisation.
- **Protein and antibody design:** RFantibody/RFdiffusion, ProteinMPNN; AlphaFold2/3 and RoseTTAFold2 for structure prediction and candidate evaluation.
- **Molecular modelling across modalities:** small-molecule structure and pocket analysis and virtual screening; peptide sequence/molecular representations and activity/property prediction; protein/antibody structure analysis and design. PyMOL, Rosetta and Schrödinger; HADDOCK docking, PRODIGY scoring and team-based Glide/MM/GBSA screening workflows.
- **Literature mining and research agents:** LLM API integration, prompt and workflow design, structured extraction, evidence linking and evaluation at multiple levels.
- **Full-stack research software development:** connecting data processing and model inference to backend services, APIs and frontend interfaces; independent development and deployment of research websites, web servers and standalone tools.
- **Research computing infrastructure:** Linux server and CPU/GPU cluster administration, Slurm scripts and batch processing, Nginx/Apache configuration, website deployment and maintenance.

Selected Awards

- Second Prize, China Postgraduate Mathematical Contest in Modeling, 2025.
- Third Prize, China Postgraduate Mathematical Contest in Modeling, 2024 and 2023.
- First-Class Postgraduate Academic Scholarship, East China University of Science and Technology, 2023 and 2022.

Languages

Chinese: native. English: passed the College English Test Band 6 (CET-6).