

# Xibin Bayes Zhou (周禧彬)

*Protein foundation models for search, dialogue, and design*

*Ph.D. Candidate in Computer Science, Westlake University & Zhejiang University, Hangzhou*

zhoubay.github.io | zhouxibin@westlake.edu.cn | Google Scholar | ORCID | GitHub | LinkedIn

## Education

- Ph.D. in Computer Science, Joint Program of Westlake University and Zhejiang University, Hangzhou, China (2022–2027, expected)
- B.S. in Data Science and Big Data Technology, Tongji University, Shanghai, China (2018–2022)
- High School, Tianjin Nankai School, Tianjin, China (2015–2018)

## Research Interests

- **Structure-aware protein language modeling** — Integrating 3D structure tokens with sequence for general PLMs
- **Multimodal protein understanding** — Aligning sequence, structure, and language for protein search & annotation
- **AI-driven protein engineering** — Enzyme optimization, base editing, de novo design, and autonomous lab workflows
- **Democratizing PLM research** — No-code PLM training platforms and open collaboration for biologists

## Honors & Awards

- **Oct 2025** — Best Poster Award, *Nature Conference on AI Augmented Biology* (Nanjing; poster on Evolla)

## Community Impact

- Contributor to Hugging Face **Transformers** (Evolla integration) — commit
- **Evolla** post on X: 130K+ views — post
- **Evolla** on Hugging Face: ~100K downloads/month — model
- **Pinal** featured in *Nature News* (May 2025) — article
- Research featured in **Westlake University** press (ProTrek / trimodal PLM) — news

## Selected Publications

\* equal contribution

- **Xibin Zhou, et al.** Decoding the molecular language of proteins with Evolla. *bioRxiv* (under review at *Nature*), 2025.
- F. Dai, et al., **X. Zhou.** Toward de novo protein design from natural language. *bioRxiv* (under review at *Nature*), 2024.
- Y. He\*, **X. Zhou\***, et al. PLM-assisted optimization of uracil-N-glycosylase enables programmable T-to-G/C base editing. *Molecular Cell* 84(7), 1257–1270, 2024.
- H. Qian\*, Y. Wang\*, **X. Zhou\***, et al. ESM-Ezy: deep learning strategy for mining novel multicopper oxidases. *Nature Communications* 16, 3274, 2025.
- J. Su\*, Y. He\*, S. You\*, et al., **X. Zhou.** A trimodal protein language model enables advanced protein searches. *Nature Biotechnology*, 2025.
- J. Su, et al., **X. Zhou.** Democratizing protein language model training, sharing and collaboration. *Nature Biotechnology*, 2025.
- J. Su, et al., **X. Zhou.** SaProt: protein language modeling with structure-aware vocabulary. *ICLR*, 2024.

## Experience

- **Investment Intern**, ZhenFund Jul–Sep 2023 (remote)

## Extracurriculars

- Former **President**, Westlake Water Sports Club; **Rowing** (4 yrs, captain); **SUP** (3 yrs); sailing, dragon boat, and SUP polo

## Languages

- Mandarin (native; 普通话水平测试一级乙等), English (proficient), Japanese (JLPT N2)
- Spanish, Latin (learning)