Acta Pharmaceutica Sinica B 2025;15(5):2805−2807

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

Acta Pharmaceutica Sinica B

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

EDITORIAL

VenusMutHub—A benchmark for protein mutation effect prediction

KEY WORDS

Protein engineering;
Protein mutation;
Computational methods;
Benchmark dataset

Protein engineering has become a cornerstone in numerous fields, from biocatalysis to biological drug development, offering innovative solutions with enhanced or novel protein functions[1]. One of the critical challenges in this field is the prediction of protein mutation effects, a task of great importance for both drug development and precision medicine[2]. The large mutational sequence space poses a significant challenge for traditional experimental approaches, underscoring the importance of computational strategies in protein engineering. Recent advancements in zero-shot computational methods, such as physics-based approaches and machine learning models, have substantially improved the prediction of protein mutation effects[3-5]. However, while these models perform well on large-scale mutation datasets, they face limitations when applied to real-world scenarios where high-throughput screening is not feasible or when specific biochemical properties need to be considered. The practical challenge lies in the small-scale, experimentally validated datasets often used in protein engineering, particularly in directed evolution efforts[6]. These datasets are more reflective of real-world constraints, where experimental data is sparse and derived from a limited set of mutations[7]. Thus, predicting mutation effects in these small-scale datasets, which more closely resemble the conditions protein engineers face, becomes a crucial task.

Recently, Zhang et al.[8] present VenusMutHub, a benchmark study for protein mutation effect prediction. It represents a specific advancement in protein mutation prediction by providing a standardized, comprehensive framework for evaluating 23 computational models on 905 small-scale experimental datasets, which were curated from 527 unique proteins derived from published literature and public databases (Fig. 1). VenusMutHub covers four key functional properties: stability (59.7%), activity (19.3%), binding affinity (15.8%), and selectivity (5.2%). The evaluated models fall into three categories: sequence-only (*e.g.*, ESM[9]), evolution-informed (*e.g.*, GEMME[10] and VESPA[11]), and structure-aware (*e.g.*, MIF[12] and VenusREM[13]). Performance was assessed using robust metrics, including Spearman correlation, normalized discounted cumulative gain, accuracy, and F1 score, to measure ranking, classification, and prediction consistency.

They found that different models shine in different areas. For stability, structure-aware models (*e.g.*, MIF) performed best, achieving high accuracy (0.627). For activity, evolution-informed models (*e.g.*, VESPA) led with a strong correlation (0.338). Binding affinity predictions varied: multichain models excelled for protein−protein interactions, while various models demonstrated better average predictive capabilities for DTI (drug−target interaction) than PPI (protein−protein interaction). However, all models struggled with selectivity, showing very low correlations (0.099), due to the complexity of these predictions. The study also explored the impact of dataset size, finding that model performance improves significantly with datasets containing 8−13 mutations or more. Structure-aware models exhibited lower variance in stability predictions, making them more reliable for this property, while evolution-informed models were more consistent for activity predictions. These insights are critical for researchers selecting models for specific applications.

VenusMutHub is a transformative resource for protein engineering, offering a rigorous evaluation of computational models and practical guidance for their application. Its focus on small-scale, biochemically validated datasets bridges the gap between computational predictions and real-world needs, making it a valuable tool for biopharmaceutical development and precision medicine. The study's findings—that structure-aware models excel in stability predictions, evolution-informed models in activity, and multichain models in specific binding scenarios—provide actionable insights for optimizing protein design workflows.

However, the benchmark has limitations. The dataset's uneven distribution, with 59.7% of data related to stability and

https://doi.org/10.1016/j.apsb.2025.05.001

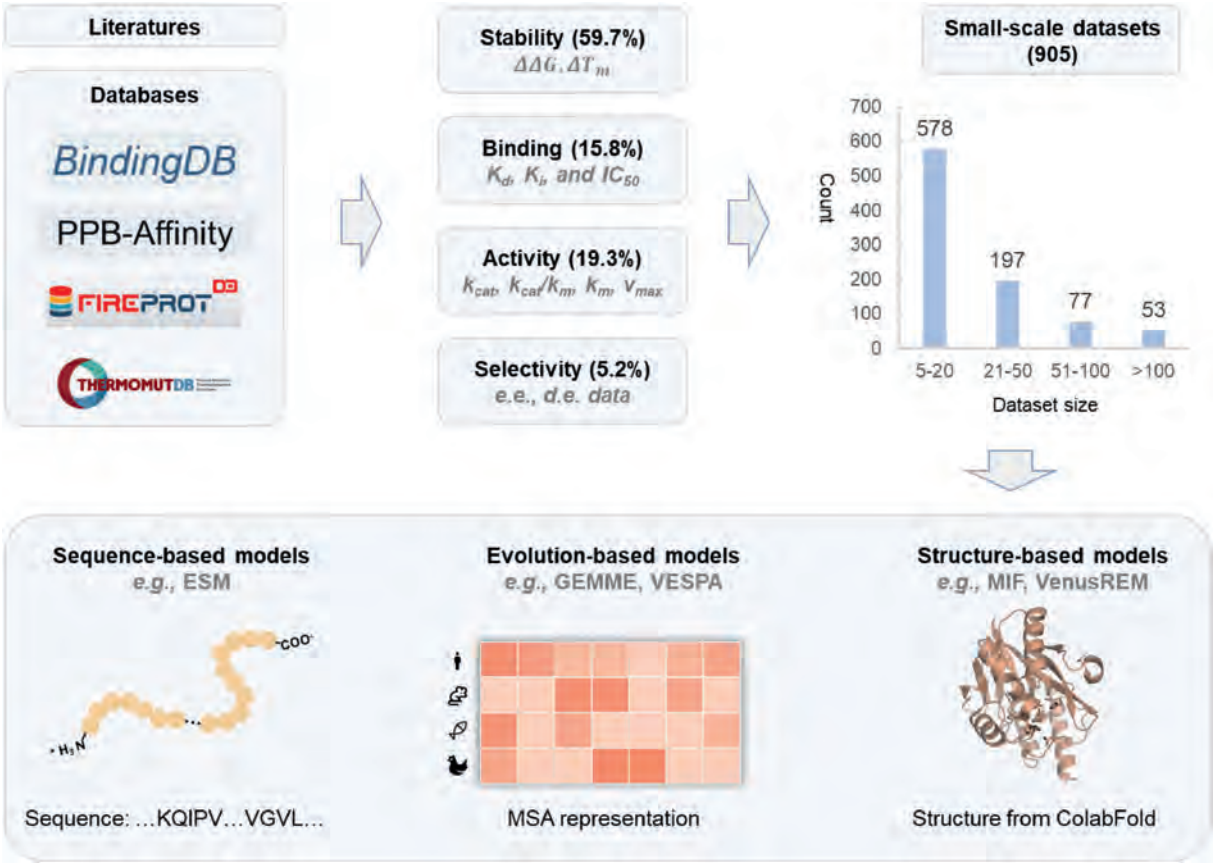

Figure 1 The benchmark establishment and evaluation for predicting protein mutation effects.

only 5.2% to selectivity, may limit its generalizability across all functional properties. Potential biases in the selection of protein families could also affect the applicability of the results. Moreover, the poor performance in selectivity predictions and the challenges in handling multiple mutations, which exhibit non-additive epistatic effects, highlight significant gaps in current modeling approaches. Future research should prioritize the development of hybrid models that integrate sequence, structure, and evolutionary data to improve prediction accuracy across all properties. Incorporating substrate-specific information through docking simulations could address the selectivity challenge, while uncertainty quantification would enhance model reliability by providing confidence measures alongside predictions. As the field progresses, VenusMutHub's open-access dataset will continue to drive innovation, encouraging the creation of next-generation models to tackle these persistent challenges and advance protein engineering.

References

1. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 2009;**10**:866−76.
2. Wittrup KD, Verdine GL. *Protein engineering for therapeutics, part A*. 1st ed. Academic Press; 2012.
3. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 2025;**381**:adg7492.
4. Cheng P, Mao C, Tang J, Yang S, Cheng Y, Wang W, et al. Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. *Cell Res* 2024;**34**:630−47.
5. Mansoor S, Baek M, Juergens D, Watson JL, Baker D. Language models enable zero-shot prediction of the effects of mutations on protein function. *NeurIPS* 2021;**34**:29287−303.
6. Hsu C, Nisonoff H, Fannjiang C, Listgarten J. Learning protein fitness models from evolutionary and assay-labeled data. *Nat Biotechnol* 2022;**40**:1114−22.
7. Zhou Z, Zhang L, Yu Y, Wu B, Li M, Hong L. Enhancing efficiency of protein language models with minimal wet-lab data through few-shot learning. *Nat Commun* 2024;**15**:5566.
8. Zhang L, Pang H, Zhang C, Li S, Tan Y, Jiang F, et al. VenusMutHub: a systematic evaluation of protein mutation effect predictors on small-scale experimental data. *Acta Pharm Sin B* 2025;**15**:2805−7.
9. 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.
10. Laine E, Karami Y, Carbone A. GEMME: a simple and fast global epistatic model predicting mutational effects. *Mol Biol Evol* 2019;**36**:2604−19.
11. Marquet C, Heinzinger M, Olenyi T, Dallago C, Erckert K, Bernhofer M, et al. Embeddings from protein language models predict conservation and variant effects. *Hum Genet* 2022;**141**:1629−47.
12. Yang KK, Zanichelli N, Yeh H. Masked inverse folding with sequence transfer for protein representation learning. *Protein Eng Des Sel* 2023;**36**:gzad015.
13. Tan Y, Wang R, Wu B, Hong L, Zhou B. Retrieval-enhanced mutation mastery: agmenting zero-shot prediction of protein language model. *arXiv* 2024. https://doi.org/10.48550/arXiv.2410.21127.

Junlin Yu[a], Guobo Li[a,b,*]
[a]*Key Laboratory of Drug Targeting and Drug Delivery System of Ministry of Education, Department of Medicinal Chemistry, West China School of Pharmacy, Sichuan University, Chengdu 610041, China*
[b]*Children's Medicine Key Laboratory of Sichuan Province, Chengdu 610041, China*

*Corresponding author.
E-mail address: liguobo@scu.edu.cn (Guobo Li)