**Multimodal Framework For Early Developability Assessment To Accelerate Protein/antibody Development**

**Jiayin Denga, 1,** Qiong Huangb, 1, Jiayi Lvb, Yiqi Yangc, Zhuyifan Yed, Yanyi Chue, Qi Zhaoc, \*, Wei-Jie Fangb, \*, Defang Ouyanga, f, \*

The State Key Laboratory of Mechanism and Quality of Chinese Medicine, Institute of Chinese Medical Sciences a, University of Macau, Taipa, Macau, China;

Institute of Drug Metabolism and Pharmaceutical Analysis, College of Pharmaceutical Sciences b, Zhejiang University, Hangzhou, China;

Cancer Centre, Faculty of Health Sciences, University of Macau c, Taipa, Macau, China;

Faculty of Applied Sciences, Macao Polytechnic University d, Macau, China;

Arc Institute, Palo Alto e, CA, USA;

DPM, Faculty of Health Sciences, University of Macau f, Taipa, Macau, China.

**Background and aims.** Protein/antibody therapeutics, distinguished by their exceptional activity, specificity, and precise biological functions, are crucial in modern pharmaceuticals. Their clinical success relies on developability—a set of properties that ensure stability and safety—determined by both the protein molecule and the formulation composition. However, formulation development of biopharmaceutical drugs remains hindered by inefficient and expensive trial-and-error assessments. This study aims to address these challenges by developing a multimodal deep learning framework for comprehensive developability prediction.

**Methods.** Four developability-related datasets were constructed, each containing both protein sequence and excipient information. A multimodal deep learning framework (FormulationProtein) was designed to predict these developability attributes by systematically integrating three-dimensional structural data, sequence, descriptors, and formulation composition. Transfer learning and conventional machine learning algorithms were employed to enable comprehensive feature representation and prediction. Experimental validation was conducted on 9 proteins across 145 formulations.

**Results.** The datasets cover conformational stability (986 entries), colloidal stability (919 entries), viscosity (900 entries), and solubility (749 entries). FormulationProtein demonstrated strong predictive performance, achieving average accuracies of 0.925 for conformational stability, 0.858 for colloidal stability, 0.917 for viscosity, and 0.742 for solubility on the test sets. These results highlight the framework's capability to systematically predict key developability attributes. Experimental validation further confirmed its reliability across diverse proteins and formulations.

**Conclusion.** This study presents FormulationProtein, a multimodal deep learning framework, for comprehensive developability assessment. By integrating structural, sequence, and formulation data, this approach offers a powerful tool to accelerate protein/antibody development in early-stage pharmaceutical research.

**Acknowledgements:** We thank the support provided by the University of Macau Multi-Year Research Grant (MYRG-GRG2024-00123-ICMS-UMDF) and Macau Science and Technology Development Fund (0071/2024/RIA1) We thank the funding provided by Macao Polytechnic University (RP/FCA-13/2023).



**Figure 1.** (A) Conventional approaches to protein formulation development. (B) Workflow for the development of FormulationProtein.