

PSF: A Web Application Tool for Protein Scaffold Filling

Kushal Badal*, Binhai Zhu†, Xiaowen Liu‡, Letu Qingge*

*Department of Computer Science, North Carolina Agricultural and Technical State University, Greensboro, NC, USA. Email: kbadal@aggies.ncat.edu, lqingge@ncat.edu

†Gianforte School of Computing, Montana State University, Bozeman, MT, USA.
Email: bhz@montana.edu

‡John W. Deming Dept. of Medicine, Tulane University, New Orleans, LA, USA. Email: xwliu@tulane.edu

Abstract—The protein scaffold filling problem remains a significant challenge in computational proteomics, which is critical for accurate protein function prediction and drug design. Despite recent advancements, current sequencing methods often yield incomplete protein sequences, referred to as scaffolds, which require precise filling for further analysis. This paper presents a web-based application, implemented using the Django framework, adopting our previously developed machine learning and deep learning techniques for protein scaffold filling. The platform allows users to try our pre-trained models or train models on their datasets for new scaffolds. This system provides a versatile tool for researchers in computational proteomics, enhancing the efficiency of protein sequence prediction. The developed web application can be accessed through <https://psf.ncat.edu/>.

Keyword— *Protein Scaffold Filling, Deep Learning, Generative AI, Django, Web Application*

Mr. Kushal Badal is currently a Master student at North Carolina A&T State University with a strong academic record and a passion for AI and data science. He gained practical experience in data analysis, machine learning, and research, complemented by leadership roles and a commitment to continuous learning. His goal is to leverage his technical skills and enthusiasm for problem-solving to make meaningful contributions in the field of Artificial Intelligence.

Dr. Binhai Zhu obtained his PhD in computer science at McGill University in 1994. Following a 2-year post-doc at Los Alamos National Laboratory, he has been teaching at City University of Hong Kong and Montana State University since Aug, 1996. He is currently a professor at Gianforte School of Computing at Montana State University. His research interests are Algorithms, Computational Biology, Computational Geometry, and Combinatorial Optimization. He has published over 220 papers in these areas. He is currently on the editorial board of Acta Informatica and Algorithms.

Dr. Xiaowen Liu received his bachelor's (1996) and master's (2004) degrees in computer science from Shandong University, China and obtained his Ph.D. degree (2008) in computer science from City University of Hong Kong, Hong Kong. After 4-year postdoc training at the University of Western Ontario, the University of Waterloo, and the University of California, San Diego, Dr. Liu took positions as an Assistant Professor and Associate Professor at the Department of BioHealth Informatics, Indiana University-Purdue University Indianapolis from 2012 to 2021. Dr. Liu is currently a professor of bioinformatics in the Division of Biomedical Informatics and Genomics, John W. Deming Department of Medicine, Tulane University School of Medicine. Dr. Liu's research focuses on computational proteomics, especially mass spectrometry-based top-down proteomics, which can identify various proteoforms with alterations in biological samples. Dr. Liu's lab is supported by several NIH grants, and his work has resulted in more than 80 peer-reviewed journal papers.

Dr. Letu Qingge PhD, is an Assistant Professor in the Department of Computer Science at North Carolina A&T State University, NC. He received his PhD degree in Computer Science from Montana State University, MT, USA. His research interests include algorithms, machine learning, deep learning, bioinformatics and computational biology, computer vision and neural network control. He has published over 50 peer-reviewed papers. His research has been funded by NSF and NIH.