

GIGA-RFP: Generative and Incorporated Geometric Architectures for *de novo* Red Fluorescent Protein design

Bo-Kai Wu, Ruei-En Hu, I-Son Ng*

Department of Chemical Engineering, National Cheng Kung University, Tainan, Taiwan

*yswu@mail.ncku.edu.tw



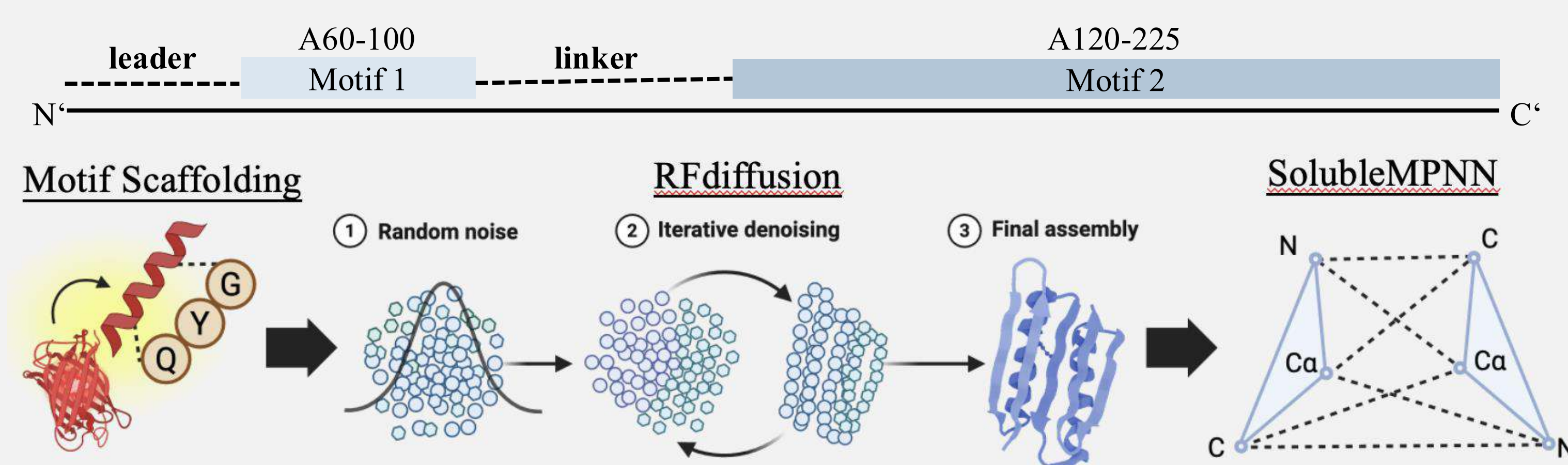
Abstract

With the advancement of artificial intelligence and deep learning, culminating in the **2024 Nobel prize** in chemistry and **ESM3-deisgned GFP**, it is now possible to perform *de novo* protein design. This project leverages **RFdiffusion** and **SolubleMPNN** to conduct *de novo* design, then screened based on structural confidence scores, solubility and fluorescence to select a small number of high-potential sequences for experimental validation. The accuracy of the gene sequences will be verified through *Escherichia coli* genetic engineering. This project aims to successfully generate novel RFP gene sequences, providing new strategic options for the development of unknown proteins.

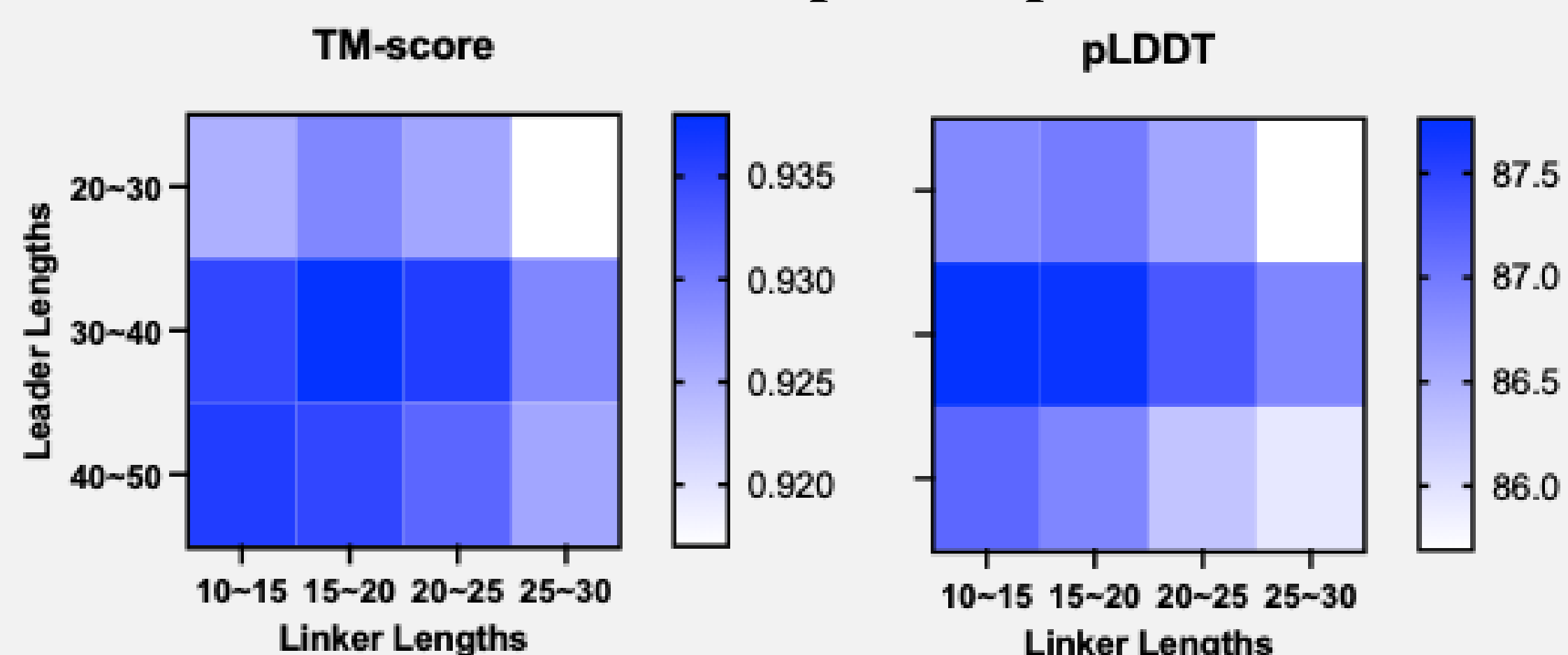
Keywords: protein design, deep learning, *E. coli*, genetic engineering, red fluorescent protein.

Core Architecture

1. Motif Selection & Sequences Generation

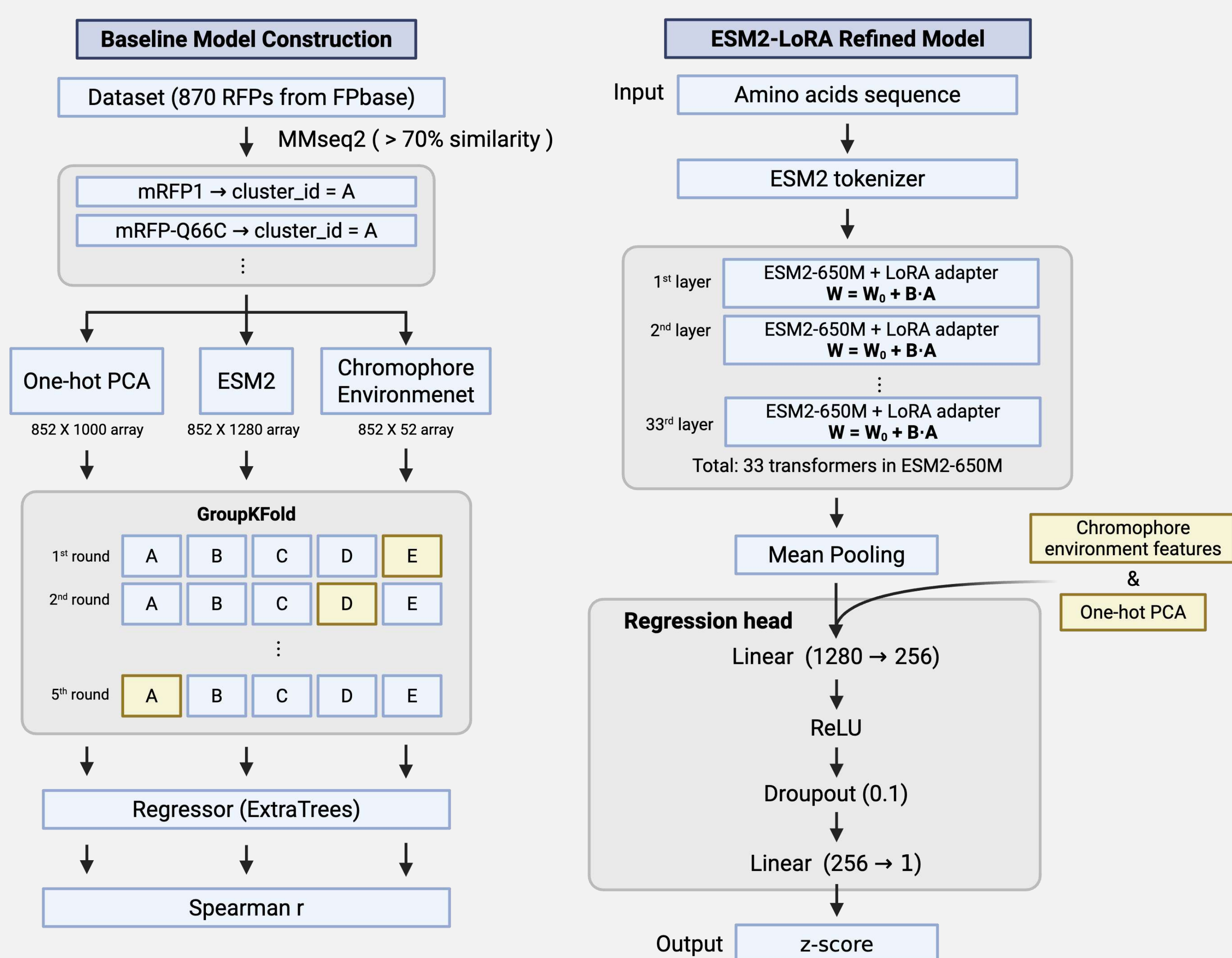


2. Leader & Linker Modulations Explore Optimistic Combinations



- **TM-score:** Global structural similarity relative to the template.
- **pLDDT:** Local confidence and conformational stability.
- **Final Candidate:** Leader (30~40), linker (15~20)

3. Fluorescence Prediction Model Construction

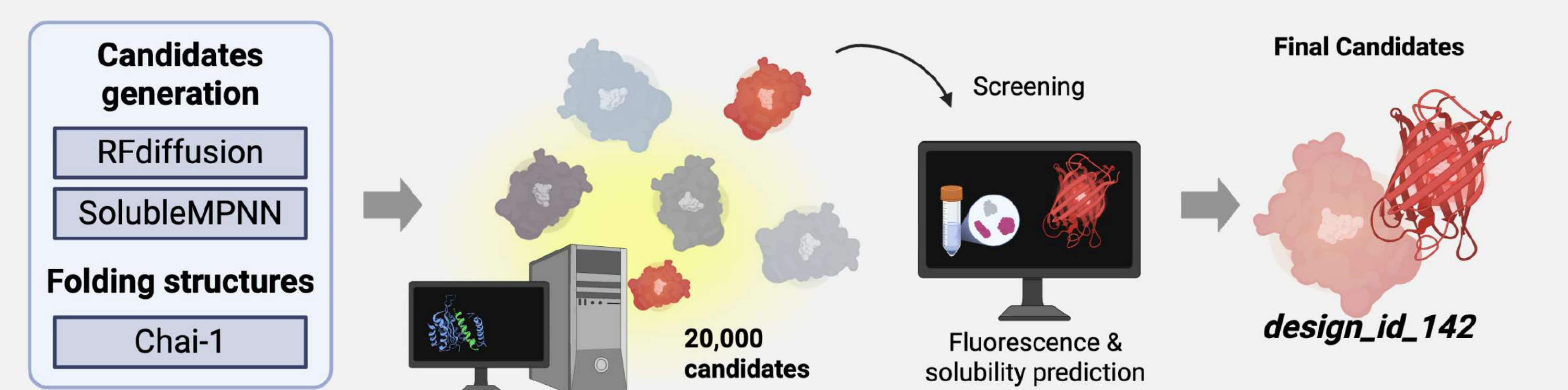


Performance Benchmarking:

Property	One-hot PCA	ESM2	Chromophore env.	ESM2-LoRA
Emission peak	+0.468	+0.362	-0.126	+0.494
Stokes shift	-0.022	+0.012	-0.135	+0.302
Extinction coeff.	+0.110	+0.239	-0.004	+0.192
Quantum yield	+0.118	+0.160	+0.040	+0.385
Brightness	+0.087	+0.148	+0.075	+0.354

All Spearman *r* are cross validated under 5 folds and regressed by ExtraTrees.

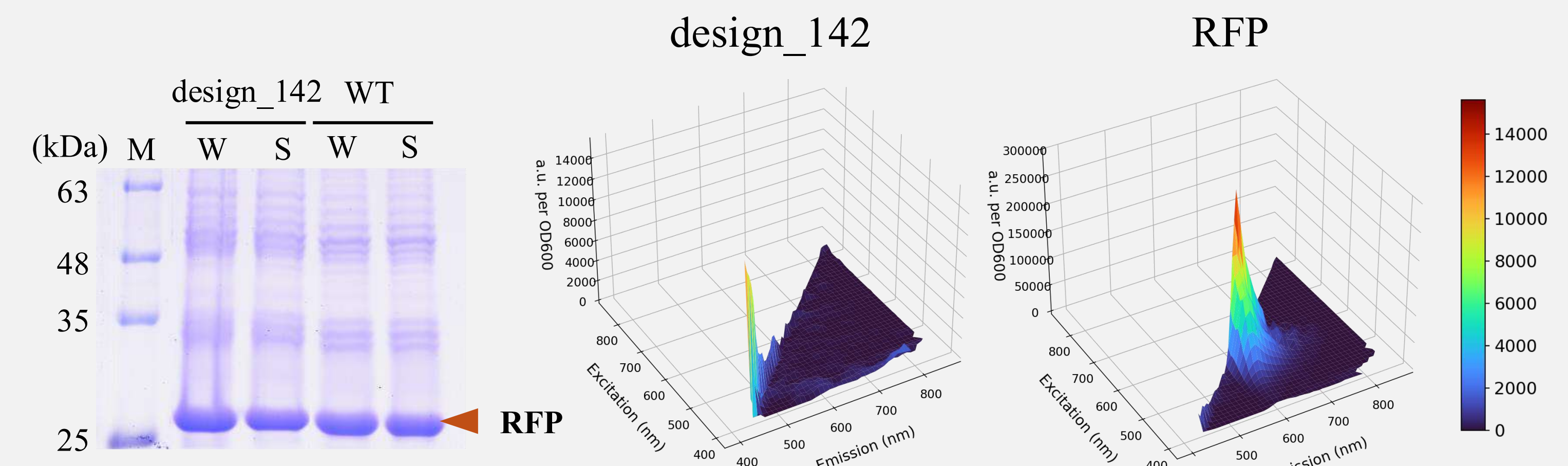
3. Fluorescence & Solubility Prediction



In this stage, we utilize three open-source protein solubility model (DeepSolNet, PLM_Sol and FGNN_Sol) to predict the solubility, setting the criteria of > 0.60 . Subsequently, we construct our own fluorescence prediction model to benchmark all the candidates against the original mRFP1. Lastly, only few candidates past through all the stages, then validated via wet-lab experiment.

Predicted Property	design_142	Template (mRFP1)
Emission peak	580.3 nm	607 nm
Stokes shift	24.8 nm	23 nm
Extinction Coefficient (EC)	89354 M ⁻¹ cm ⁻¹	50000 M ⁻¹ cm ⁻¹
Quantum Yield (QY)	0.32	0.25
Brightness (EC × QY)	28.8	12.5

4. Wet-lab Validation



The final candidate (design_id_142) was constructed into pSU-pLacI-sfGFP-B0015 backbone, expressed via *E. coli*. The fluorescence analysis was examined by excitation-emission metrics (EEM).

Conclusion & Future Perspectives

- ✓ **Computational Protein Design:** Integrated cutting-edged generative AI with deep-learning frameworks to assist the *de novo* design red fluorescent proteins (RFPs).
- ✓ **Experimental Findings:** The top-ranked design was successfully expressed in *E. coli* but exhibited no detectable fluorescence.
- ✓ **Future Work:** By incorporating robust geometric and structural validation tools, aiming to overcome maturation bottlenecks and generate functional red fluorescent proteins.

Acknowledgement

This work is financially supported by National Science and Technology Council, Taiwan. Project number: 115-2813-C-006-114-E

