Summary of Activities (Quarter 2)

Yuting Chen and Haoling Zhang

Overview

Our *PyMOL Fellowship* Project is to establish high-level interfaces based on *PyMOL* framework[1] called *PyMOL-advance*. Its repository is set out in https://github.com/BGI-SynBio/PyMOL-advance. These high-level interfaces effectively establish the channel from structure data to publication-standard figures. As shown in Figure 1, they can help bioinformaticians, algorithm scientists and computer engineers to create figures for their manuscripts in a simpler and more rapid way.

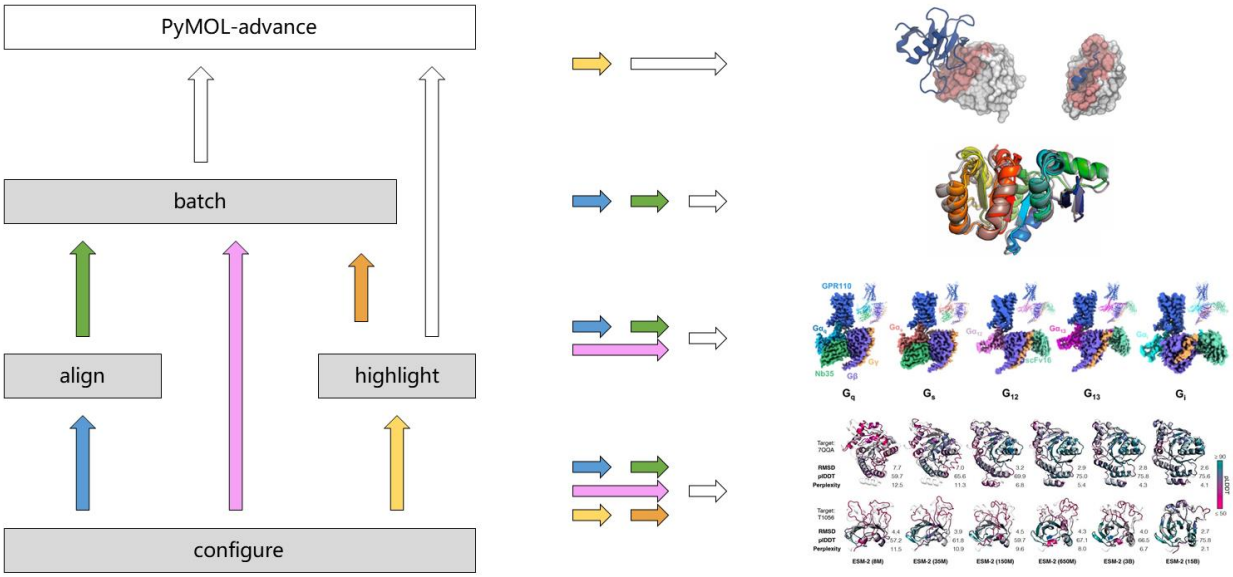


Figure 1. core requirement architecture and target cases[2-5].

As we mentioned in the proposal, we need to program scripts of structures visualization module and publication-standard figures construct module in our second quarter. Besides, we need to reproduce valuable cases from figures in high-quality academic journals (like the right part of Figure 1) and provide use methods of the established interfaces.

Quarterly Goal

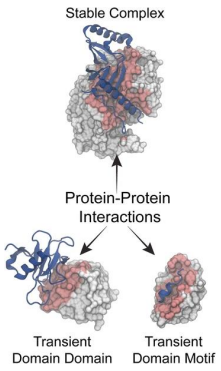
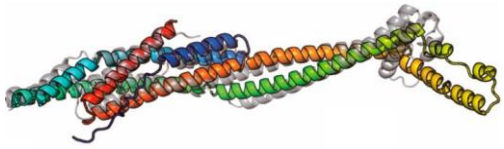
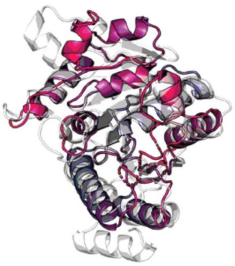
1. **structures visualization module:** This module is closely related to selective hiding, custom presentation, specific coloring and structure alignment. To improve the practicability and reliability of the module, StructureImage class provides the functions according to the conventional structure visualization process. In addition, we are prepared to provide a parameter setting interface for users to customize the rendered structure in a more advanced and elaborate manner.
2. **publication-standard figures construct module:** This module can construct a complete publication-standard figure according to the needs of users. The basic process involves initializing the figure configuration, layout structure pictures, and custom text and widgets. This series of procedures can help users to create one-step high-quality publication-standard figures to fulfill the requirements of the target academic journal like *Nature*, *Science*, and *Cell*.
3. **valuable cases reproduction:** To demonstrate how to use the interface, we reproduce valuable figures based on *PyMOL-advance* and compare them to those published in academic articles.

4. **vector graph painting feasibility:** We need to investigate the feasibility of drawing vector maps and any limitations or issues that exist, in addition to the inability of *PyMOL* to output vector graphs.

Solutions

1. **feasibility evaluation of vector graphics painting:** Early in this quarter, we provided brief feedback on vector graphics rendering. We assume that the design of the vector output interface for PyMOL has been improved. For example, when zooming and dragging the ball-and-stick model of green fluorescent protein with 3523 atoms, stagnation is obvious when using the Adobe Acrobat. The vector painting of structures needs to be compressed or optimized further to achieve our main goal of creating publication-standard figures that fulfill both journal requirements and user experience. Our final conclusion is that this is not the right time for vector graphics research and development. With more advancements in vector compression[6,7] in the future, further research and development of vector graphics may prove to be useful.
2. **cases implementation:** Upon comparing our sample figures to those published in academic articles, we find that our designs are strikingly similar, and they only require at most dozens of lines of code. The collected figures from the publications are shown in Table 1.

Table 1. Collected publication cases.

case	figure	feature	reference
1		● Multiple representations showed in one structure; ● Partial area coloring; ● Multiple sub-graph merging.	2
2		● Coloring based on amino acid position; ● Structure alignment information; ● Border reinforcement.	3
3		● Coloring based on additional properties; ● Structure alignment information.	8
4		● Batch visualization; ● Coloring by chain.	4

Moreover, the quality of certain example figures (see Figure 2) exceeds the clarity of images published in top journals.

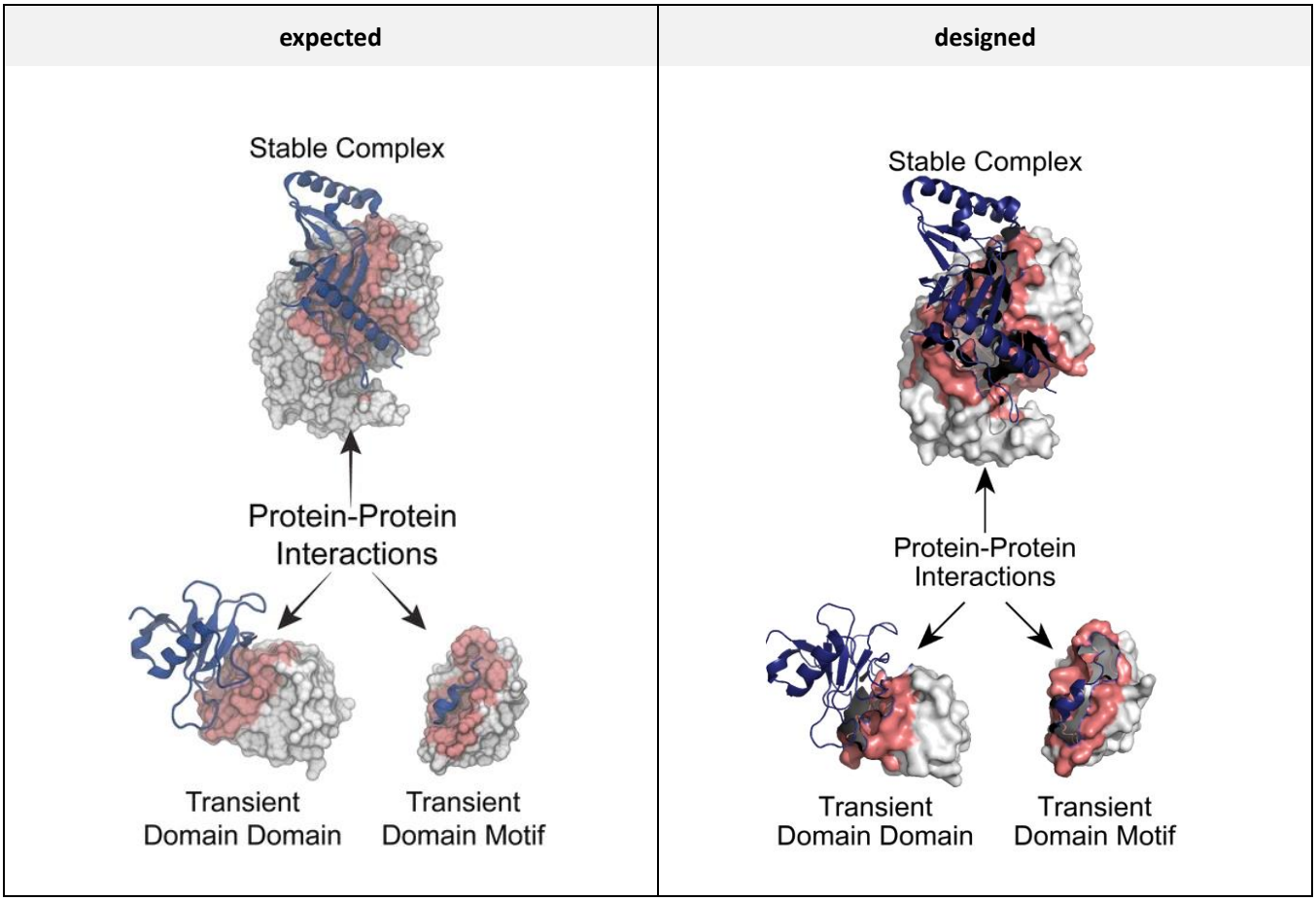


Figure 2. A comparable case.

3. **Special rendering mechanism based on additional properties:** In recent years, an increasing number of researchers have come to understand the significance of a molecule structure in relation to its physical and/or chemical properties[9]. In addition, obtaining information to compare local differences allows researchers to assess and improve prediction algorithms[3,10]. After ensuring legal use, we developed a method based on adjust-b-factor to alter the residue size and color of the structure. The adjusted values are either properties or differences, as shown "designed 2" in Figure 3.

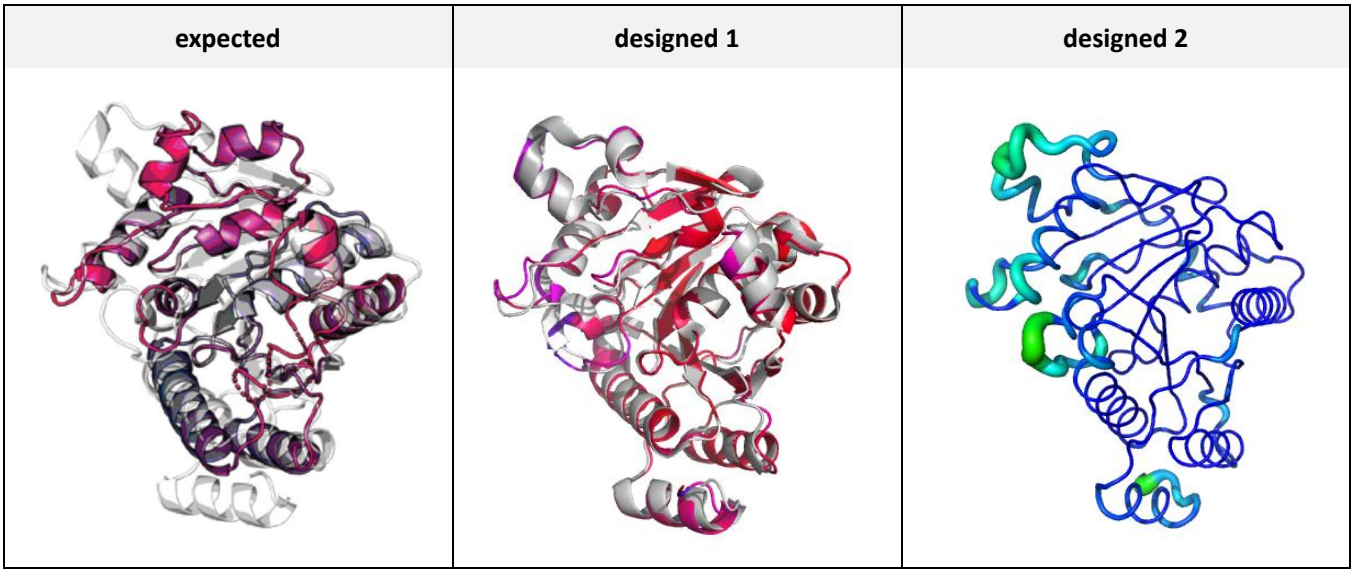


Figure 3. A comparable case. For "designed 2", the greener the thicker, the greater the RMSD.

4. **reconstruct codes for better usage:** Based on the process of case implementation, we have reflected on our interface design. To enable users to use the interface more efficiently and accurately, we have refactored our interface three times. For example, StructureImage class provides functions that follow the conventional visualization process, i.e., hiding

unimportant elements -> adjusting structures states by translating, rotating, and zooming -> setting structures representations -> coloring structures for highlight or representing the properties. We offer a scalable and readable string expression, "*shading type:target,...,target*", for configuring different targets (position, range, residue, segment, chain, and model) that allows users to customize the displayed structure in a more advanced and elaborate manner.

Remaining Issues

1. **inner and outer alignment usage:** In contrast to the internal implementation of "align" in *PyMOL*, *PyMOL-advance* offers a range of more diverse similarity evaluation mechanisms such as RMSD and TM-score, and subsequent operations like structure clustering. However, the limitation of our method is that it is unable to handle protein alignment related to sequence similarity. Therefore, our method cannot directly replace "mol.cmd.align". Besides, in order to decrease the computational complexities involved in the painting process, we attempt to input the rotation angle (calculated by our "align") into the structural drawing image. However, upon examination of the image output, it was observed that there was a substantial difference between our rotated image and the expected rotated image. Further time may be required to inspect whether any other codes impact the angle rotation results.
2. **surface rendering optimization:** As shown in Figure 2, differences exist between our designed figure and the expected figure regarding surface glossiness and features. This indicates we need to further investigate how to use PyMOL settings to better fulfill user needs.

Next Quarter

In the next quarter, we will focus on collecting configuration schemes. Currently, proficiency in Python is still a prerequisite for using *PyMOL-advance*. To ensure *PyMOL-advance* is accessible to users with less programming experience and enhance its effect, we plan to provide diversified configurations. Additionally, we will adjust some painting details and complete construction of the website, comparable cases, interface examples, and tutorials to improve user interaction.

Reference

1 DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr* **40**, 82-92 (2002).

2 Corbi-Verge, C. & Kim, P. M. Motif mediated protein-protein interactions as drug targets. *Cell Communication and Signaling* **14**, 1-12 (2016).

3 Yang, J. *et al.* Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences* **117**, 1496-1503 (2020).

4 Zhu, X. *et al.* Structural basis of adhesion GPCR GPR110 activation by stalk peptide and G-proteins coupling. *Nature Communications* **13**, 5513 (2022).

5 Lin, Z. *et al.* Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv* (2022).

6 Reina, G. Can Image Data Facilitate Reproducibility of Graphics and Visualizations? Toward a Trusted Scientific Practice. *IEEE Computer Graphics and Applications* **43**, 89-100 (2023).

7 Mousavi, S. F., Safayani, M., Mirzaei, A. & Bahonar, H. Hierarchical graph embedding in vector space by graph pyramid. *Pattern Recognition* **61**, 245-254 (2017).

8 Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023).

9 Skolnick, J. & Gao, M. The role of local versus nonlocal physicochemical restraints in determining protein native structure. *Current opinion in structural biology* **68**, 1-8 (2021).

10 Frontzek, K. *et al.* A conformational switch controlling the toxicity of the prion protein. *Nature Structural & Molecular Biology* **29**, 831-840 (2022).