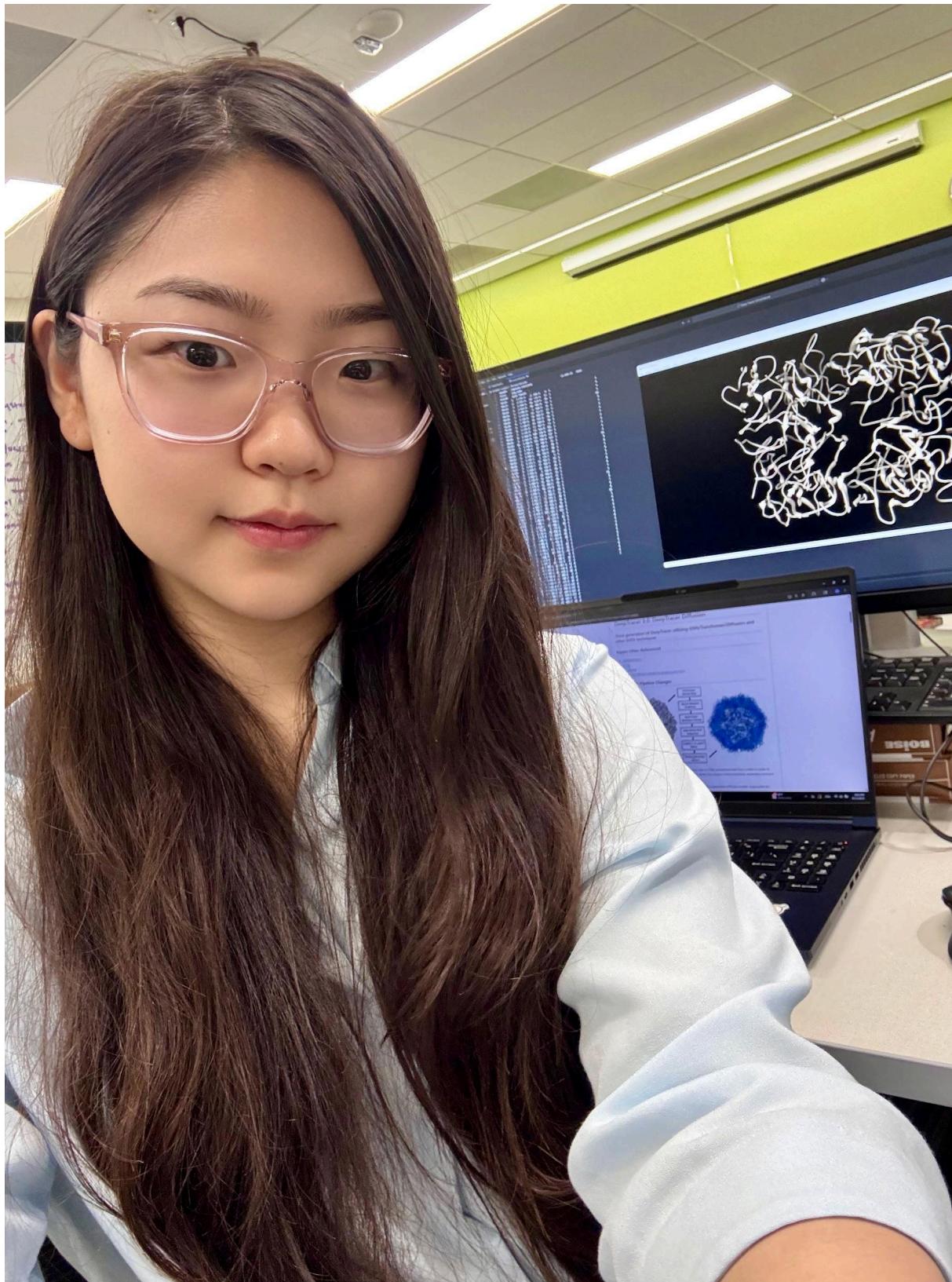


Syeon Park - Final Reflection Letter

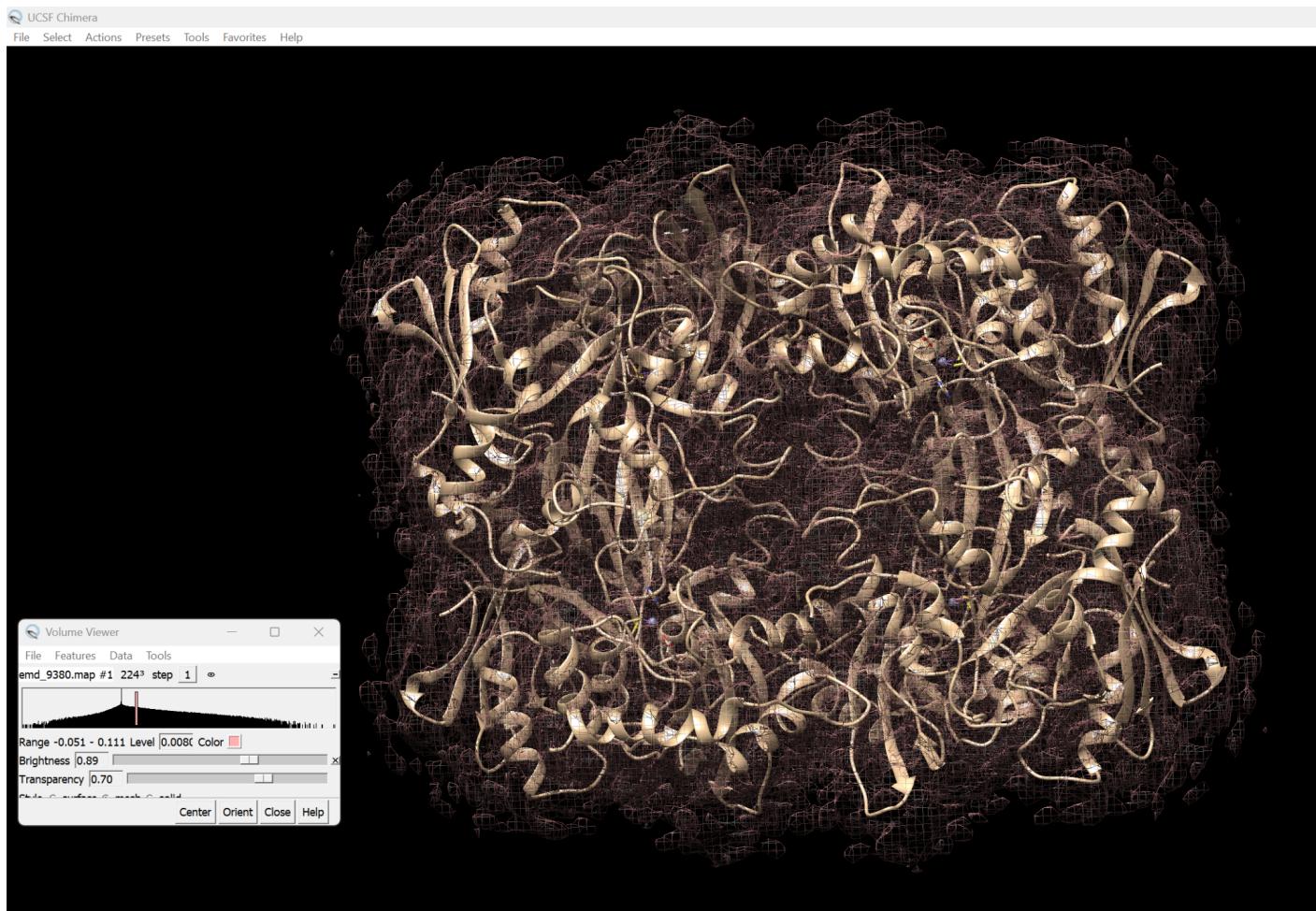


This summer, I worked on improving protein structure prediction by combining Cryo-EM experimental data with generative AI models, specifically using the DeepTracer diffusion pipeline. Our research question was about how we can generate accurate atomic models directly from 3D Cryo-EM density maps, potentially accelerating structure determination and lowering the cost of vaccine or drug design. The project matters to me because it gave me hands on experience at the intersection of AI and biology, which is a combination I want to explore further in my career. It matters to others because faster and more accessible protein modeling can speed up biomedical research.

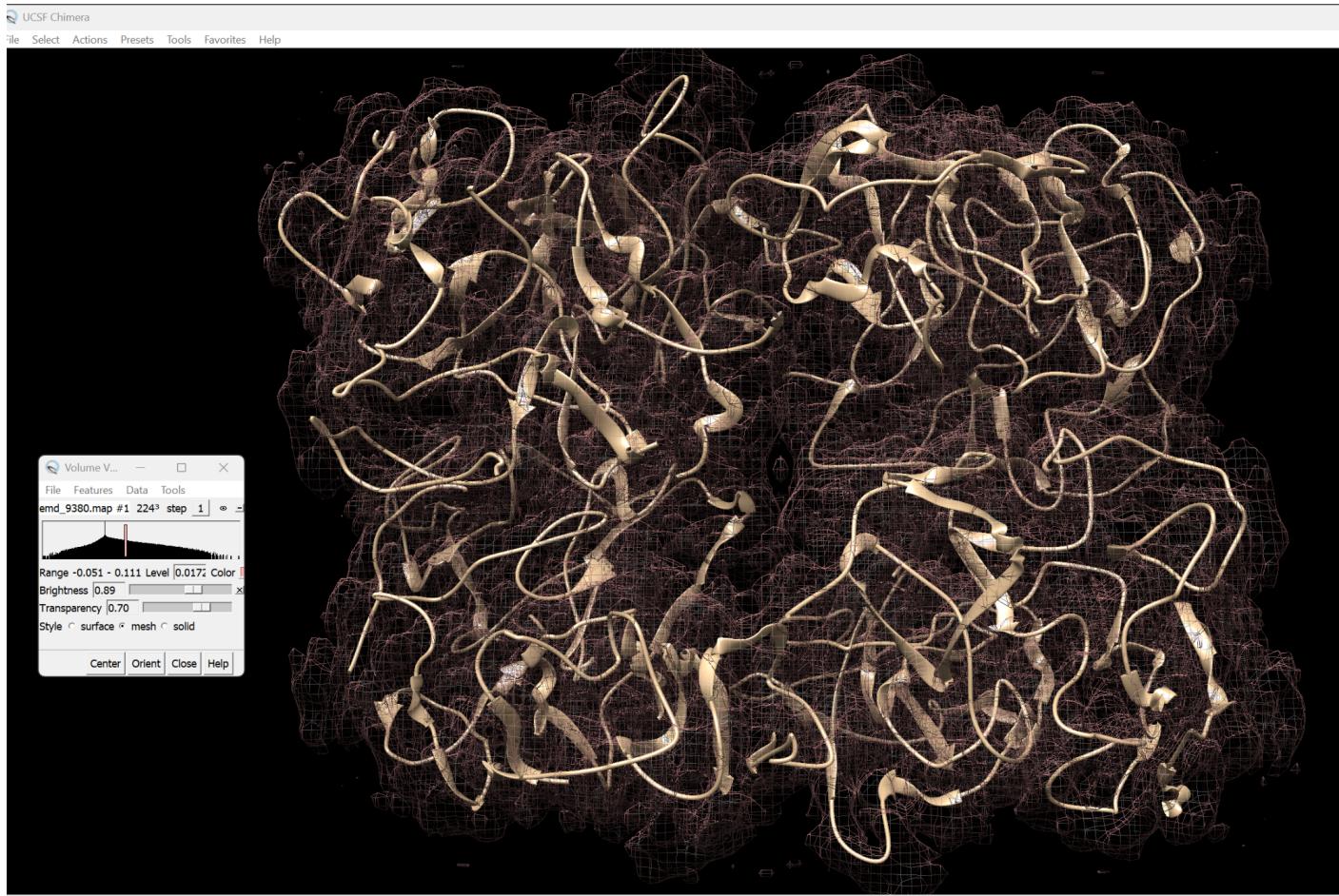
In addressing this challenge, I tested the AI pipeline on publicly available Cryo-EM maps, created my own datasets, documented the training process, and evaluated predictions by visualizing them in UCSF Chimera.

To evaluate the AI pipeline, I used the Cryo-EM density map emd_9380.map (shown as the pink mesh) and compared it against two atomic models:

- **Reference:** the deposited structure pdb6ni (“answer key”).



- **Prediction:** the AI-generated structure emd_9380.pdb.



In each screenshot, the pink mesh represents the experimental electron density, while the ribbon models show either the deposited reference structure or the AI-predicted structure. By overlaying them, we can assess how closely the AI matches the known structure. Once validated on known cases, the same approach can then be applied to new maps where no solved structure exists yet.

Beyond evaluation, I also focused on improving the project's README documentation. At the time, we had two active development branches (dt-integration and backbone) with slightly different commands and workflows, so I created branch-specific guides to help new members set up and run the pipeline without confusion. This was especially valuable during the transition period, ensuring contributors could stay productive until the branches were merged. To make the guides accurate and user-friendly, I tested the commands myself, clarified instructions, and incorporated feedback from teammates. Along the way, I experimented with map compression to handle large datasets more efficiently, explored parameter tuning to stabilize model training, and compared AI-generated models with experimental data to assess accuracy.

I learned how much structural biology depends on careful data handling and visualization, understanding the relationship between the experimental Cryo-EM density maps (like emd_9380.map) and AI-predicted atomic models (emd_9380.pdb). I also learned that in AI model development, reproducibility and clear documentation are as important as the algorithms themselves. The research process required patience with troubleshooting, as I had to deal with incomplete documentation, unexpected training errors, and huge file

sizes. I adapted by seeking advice from my mentor, diving into the literature, and running controlled tests to isolate problems.

One major challenge was the steep learning curve with both biology specific terms and with unfamiliar software like Chimera and HDF5 compression tools. I overcame this by breaking the work into smaller, testable steps and by documenting everything I learned immediately so I could build on it later. File size and compute limitations were another hurdle; I explored compression methods like gzip and lzf to keep datasets manageable, since they can get very large.

This experience confirmed for me that I enjoy working in applied AI, especially in interdisciplinary areas like computational biology. I've gained technical skills in dataset preparation, structural visualization, and AI model evaluation, as well as habits like methodical troubleshooting and documenting my work for others. Moving forward, my plan is to continue exploring AI applications in science, while focusing on projects that require both coding skills and domain-specific understanding.

One piece of advice I would like to provide for new UW Bothell undergraduate researchers is that you shouldn't wait until you "know enough" before diving into research. Rather, it is better to start, make mistakes, and learn as you go. Ask questions early and often, and take ownership of documenting your process so others (and future you) can follow it. Breaking down big problems into smaller, solvable pieces made my work less overwhelming, and sharing progress regularly with my team helped me stay on track and motivated.