Potential Pitfalls in Protein Structure Determination via Protein Crystallography

Farihah Chowdhury

1. Introduction:

Protein crystallography is a process that revolutionized the field of biochemistry. The first protein crystals were created in the 1800s, for purposes of purification (McPherson & Gavira, 2014). Then, in the early 1900s, X-ray crystallography was first used to determine the atomic structures of salt crystals. As the decades went on, X-ray crystallography was used to find the structures of many proteins as well as other molecules, such as DNA (Curry, 2015). These structures have been periodically added to the Protein Data Bank (PDB), and the number of released structures has increased exponentially in the last twenty years as various technologies have developed (RSCB, 2020).

Structural protein data can then be used for many practical purposes, such as drug discovery. A high resolution protein structure can be used in complex with ligands to analyze the mechanisms of drug activity (Zheng et al., 2014). Algorithms and code can be used to predict the possible complexes that can form between proteins and ligands, but X-ray crystallography is one of the best ways to validate this data (March-Villa, 2017).

The influx of protein structures added to the PDB is correlated to the decreased level of difficulty to crystallize a protein. Due to this vast volume of structures, there are bound to be issues with the process. Honest errors are made at every point, from almost publishing a structure that is completely wrong to forgetting to check the R factor in model refinement (Wlodawer, 2013). The aim of this project is to address the potential pitfalls associated with macromolecular crystallogenesis and structural resolution. Pitfalls can occur at any step along the meticulous process, including during crystallization, during data collection, and during data and model refinement.

2. Methodology:

Macromolecular crystallography uses the chemistry of supersaturation states of proteins to evaluate protein structure. The protein starts in the mother liquor as an undersaturated solution. Then, the particle concentration moves towards equilibrium as the solution becomes metastable. After, the solution becomes more saturated as the crystal begins to form in the nucleation zone. Finally, the crystals fully form in the precipitation zone when the solution becomes supersaturated (SINE2020, 2016). To make sense of the crystals, the diffraction patterns from the beamline are converted into a 3D protein structure (McPherson & Gavira, 2014). Throughout the crystallization and analysis of bovine thyroglobulin and human insulin structures, the pitfalls in each step were recorded and discussed.

2.1. Methods of Crystallization

There are multiple methods of crystallization, such as the hanging drop, sitting drop, micro-batch, and microdialysis methods (McPherson & Gavira, 2014). In this project, the hanging drop vapor diffusion method was used for different proteins along with the use of nucleation inducing reagents, such as gelatin. The protein solution was placed on a coverslip over a reservoir of the mother liquor. This method uses vapor diffusion, which allows the protein, buffer, and precipitate to equilibrate with the reservoir liquids to reach the necessary state of supersaturation (Rhodes, 2006).

Every protein has its own unique set of conditions that promote crystallogenesis. In order to ascertain the perfect environment for crystal growth, a multitude of conditions were tested in a systematic way. However, since they are commercially sold, the conditions are difficult to

replicate without buying the vial straight from the source, and it is tedious to test every unique solution individually.

2.2. Data Collection and Analysis

After crystals formed, they were fished and placed into a puck immersed in liquid nitrogen. In this research, the Highly Automated Macromolecular Crystallography (AMX) beamline was used to obtain diffraction data from harvested crystals (NSLS II BNL). Once diffraction patterns are obtained, these patterns can be analyzed with a multitude of programs found in CCP4i. The 2D distances between the patterns are converted into pieces of data that come together to form a 3D protein structure with electron density.

2.3. Model Refinement

After the raw data was analyzed through CCP4i, the outputted 3D model was fitted to the electron density in COOT. The model was completed using refinement, model building, and validation processes.

To refine the data, multiple techniques were used, such as Auto-zone, Sphere Refinement, Single-Atom Drag, and Ramachandran Refinement, known as Real Space Refinement collectively. Through these processes, the coordinates of the model were adjusted to match the electron density structure factors as much as possible. The bonds, planes, angles, torsions, and non-bonded contacts had to be minimized (Diamond, 1971). To complete the model, some portions needed to be built or removed (Lohkamp, 2011).

3. Results:

Along each step of protein structure resolution, several processes were unsuccessful due to pitfalls that crystallographers face on a daily basis. The issues that arose include: (1) errors during crystallization, such as using a scaffold and the crystallogenesis of salt, (2) errors during data collection, such as an electron density that doesn't match a known structure and (3) errors during data refinement, such as using the wrong sequence to analyze the structure.

3.1. During Crystallization

Pitfall of Crystallization of solely the Scaffold

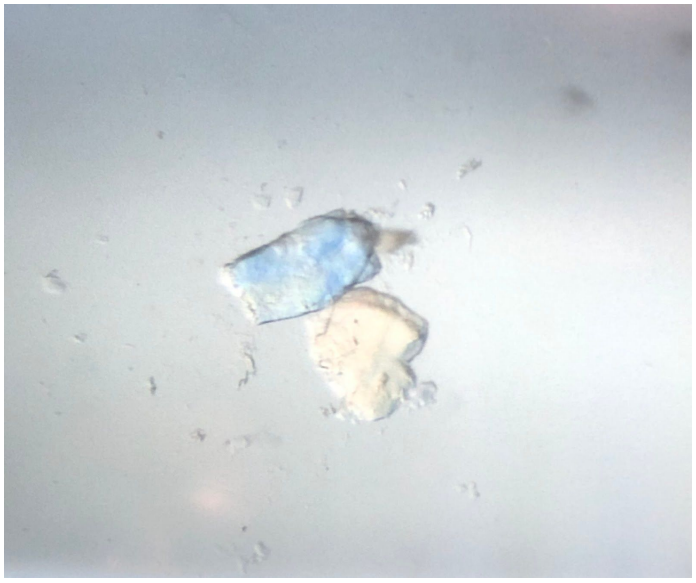

Figure 1: An image of gelatin being used as a scaffold for bovine thyroglobulin crystallization. Although diffraction was evident, no diffraction occurred yielding no protein.

Pitfall of Salt Crystals

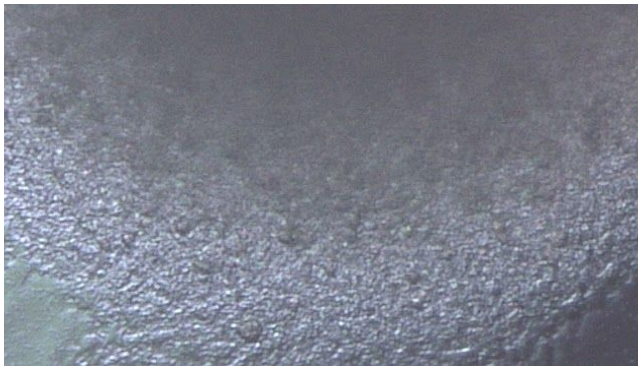

Figure 2: An image of salt crystals thought to be protein crystals until analysis.

At this point in the process, the possible pitfalls can occur with the use of a scaffold, or trying to crystallize a protein but getting a salt crystal instead. Since bovine thyroglobulin is extremely hard to crystallize, gelatin was used to try and promote the crystallogenesis of thyroglobulin. When this process was conducted, crystals were formed and analyzed on the beamline, where a diffraction pattern was depicted. However, when the data was analyzed, it was discovered that no protein was present as no diffraction occurred. In addition, when I attempted to crystallize bovine thyroglobulin by itself, what seemed like microcrystals formed (Figure 2). However, salt also has crystalline qualities, so sometimes, salt crystals could form. This could give the false pretense of protein crystals, as they also polarize, until the crystals are run on the beamline, as seen in Figure 2.

3.2. During Data Collection

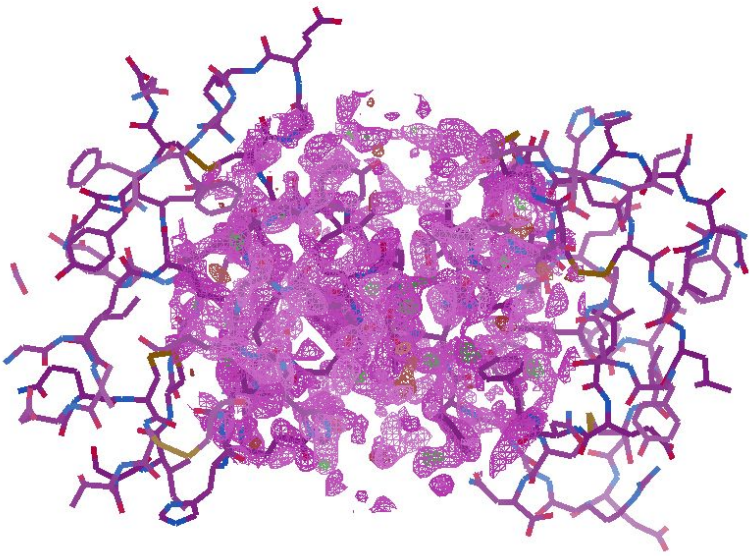

Figure 3: This figure depicts what was thought to be insulin analyzed, with an electron density that does not fit the structure.

While the crystals are being analyzed on the beamline, more issues can emerge. Figure 3 is a key depiction of how confusion can occur at the beamline. The samples being analyzed were labeled as insulin, but they were actually lysozyme, so the electron density apparently did not fit the structure.

3.3. During Data Refinement

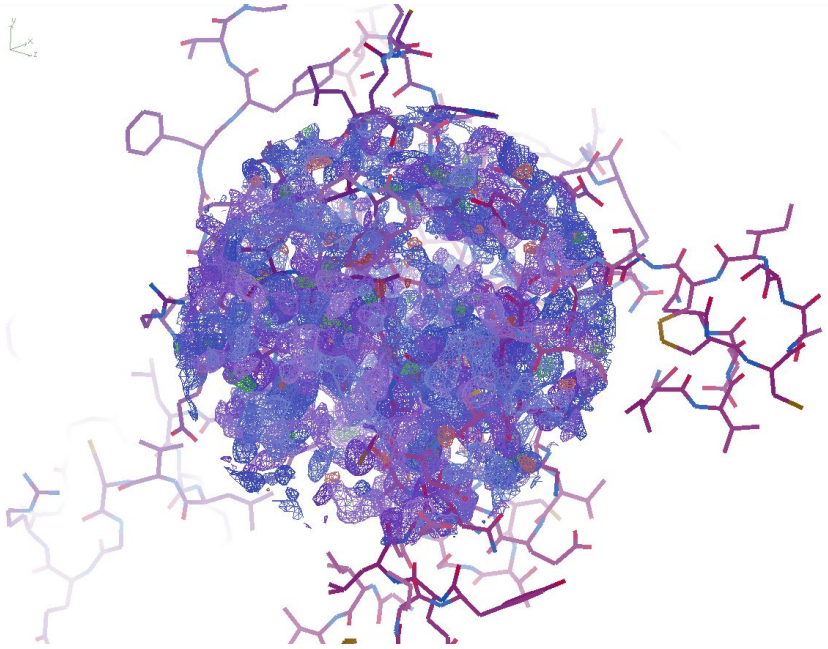

Figure 4: This figure depicts the output pdb and electron density from CCP4i of lysozyme after it was analyzed with the human insulin sequence.

Since the data was thought to be insulin as it was labeled inaccurately, the data file was analyzed through CCP4i with the insulin pdb 3i40. However, insulin and lysozyme have completely different structures, so clearly, the protein model does not fit the electron density. This was supported by the fact that the R factor and R free values were both over 0.5.

4. Discussion/Conclusion:

Clearly, there are many possible pitfalls in the protein crystallization process. Along the way, these pitfalls can be prevented. After the crystals are analyzed on the beamline, a log file is formed. The log file can be used as a reference to check if the space group and cell dimensions match up with the structure you are trying to analyze. Then, in data refinement, the Wfrac value

in Molrep should be around 0.4 and the R free and R factor should both be less than 0.25 if the model and the electron density match up.

Also, as technology has progressed, new methods of protein structure resolution have emerged. These processes either can further validate macromolecular crystal data, or replace the process entirely. One example is Small Angle X-ray Scattering, a technique that can be used to determine how the shape and size of proteins in solution averaged overall conformations. This information can be used to generate a structure of proteins, with the solution scattering data also being able to account for structural flexibility (Castellanos et al., 2016). This method was used with bovine thyroglobulin since protein crystallography has been unsuccessful.

5. Future Implications:

To circumvent the issues involved with protein crystallography and structural resolutions, other methods of structural analysis can be used, such as Cryo EM and NMR Spectroscopy. Cryo EM can be integrated into X-ray crystallography to create high resolution models of molecular assemblies. It has been successfully used to study biological compounds and macromolecules and further validates X-ray crystallography data (Milne et al., 2012). Nuclear magnetic resonance (NMR) Spectroscopy uses spectral parameters to piece together molecules to form angles and distances that convert into a 3D structure (Marion, 2013).

References:

1. Aldrich: Analytical, Biology, Chemistry & Materials Science products and services. (n.d.). Retrieved from https://www.sigmaaldrich.com/
2. Brünger, A.T.(1997). Free R Value: Cross-validation in crystallography, Methods in Enzym. 277, 366-396
3. Diamond, R. (1971). Acta Cryst. A 27, 436-452.
4. French G.S. and Wilson K.S. (1978). Acta. Cryst., A34, 517.
5. Hazes B., unpublished results
6. Lohkamp, B. (2011). Model-Building with Coot: An Introduction, low resolution tools.
7. March-Vila, E., Pinzi, L., Sturm, N., Tinivella, A., Engkvist, O., Chen, H., & Rastelli, G. (2017). On the integration of in silico drug design methods for drug repurposing. Frontiers in pharmacology, 8, 298.
8. Marion D. (2013). An introduction to biological NMR spectroscopy. *Molecular & cellular proteomics : MCP*, *12*(11), 3006–3025. doi:10.1074/mcp.O113.030239
9. McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. Acta Crystallographica Section F: Structural Biology Communications, 70(1), 2-20.
10. Milne, J. L., Borgnia, M. J., Bartesaghi, A., Tran, E. E., Earl, L. A., Schauder, D. M., … Subramaniam, S. (2013). Cryo-electron microscopy--a primer for the non-microscopist. The FEBS journal, 280(1), 28–45. doi:10.1111/febs.12078
11. RCSB Protein Data Bank. (2020). PDB Statistics: Overall Growth of Released Structures Per Year. Retrieved from https://www.rcsb.org/stats/growth/overall

12. Rhodes, G. (2006) Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models, 3rd Ed., Academic Press
13. Russo Krauss, I., Merlino, A., Vergara, A., & Sica, F. (2013). An overview of biological macromolecule crystallization. International journal of molecular sciences, 14(6), 11643-11691.
14. SINE2020. (2016). France, Germany and Sweden: Phase diagram crystal growth. Retrieved from https://sine2020.eu/about/the-road-to-the-ess/phase-diagram.html
15. Stephen Curry (2015) Structural Biology: A Century-long Journey into an Unseen World,Interdisciplinary Science Reviews, 40:3, 308-328, DOI: 10.1179/0308018815Z.000000000120
16. Vagin, AA, Steiner, RS, Lebedev, AA, Potterton, L, McNicholas, S, Long, F and Murshudov, GN. (2004). REFMAC5 dictionary: organisation of prior chemical knowledge and guidelines for its use. Acta Cryst. D60: 2284-2295
17. Vagin,A.,Teplyakov,A.(1997). MOLREP: an automated program for molecular replacement., J. Appl. Cryst. 30, 1022-1025
18. Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. The FEBS journal, 280(22), 5705-5736.
19. Zheng, H., Hou, J., Zimmerman, M. D., Wlodawer, A., & Minor, W. (2014). The future of crystallography in drug discovery. Expert opinion on drug discovery, 9(2), 125–137. doi:10.1517/17460441.2014.872623