PHYS180 Project 1
Reconstructing a 2D Structure from X-Ray Data

Owen Morehead, Karoli Clever, Rey Cervantes

April 12, 2021

1 Writing a script to reconstruct the position of a molecule

The technique of X-ray crystallography is commonly used to to study microscopic objects. When the crystal is illuminated by the plane wave that is the electromagnetic x-ray wave, resulting x-ray diffraction data provides the absolute value or amplitude of the Fourier transform of the microscopic objects electron density. However, there is a common problem involved in this technique in that there is a loss of phase information in the diffraction intensity pattern. In other words, this diffraction pattern doesn't give us the full Fourier transform. Luckily there are many ways to overcome this, allowing us to reveal the structural orientation of each atom in a given molecule or compound. One such method is explored in this assignment and is called heavy atom substitution. Substituting one atom for another gives a different diffraction pattern. With this and the original scattering data, we can solve the structure without any phase information.
The diffraction amplitudes for the native molecule is given in the file nat.txt, a 2d array of data (360x360). An extra atom is added at the origin of the coordinate system, providing point-like scattering such that the amplitude at all points is 100. Therefore this heavy atom is described by a delta function at the origin of strength A = 100. The Fourier transform of a delta function at the origin is equal to 1 as can easily be shown. The diffraction amplitude data for the native system with the addition of the heavy atom is given in hev.txt.
This heavy atom substitution method relies on the property which is that *atomic scattering is additive*. As can be visualized in Figure 1, this property allows us to use triangulation to solve for the Fourier transform of the original molecule, which we can then take the inverse Fourier transform of to obtain the molecular structure.

We start by writing:

$$\begin{aligned} FT(f_p(x,y)) &= \mathbf{F}_p \\ FT(A\delta(x-0,y-0)) &= A = 100 = \mathbf{F}_h \end{aligned} \tag{1}$$

and utilizing the fact that atomic scattering is additive:

$$FT(f_{hp}(x,y)) = \mathbf{F}_{hp} = \mathbf{F}_h + \mathbf{F}_p. \tag{2}$$

Taking the absolute value of both sides and squaring:

$$|\mathbf{F}_{hp}|^2 = |\mathbf{F}_h + \mathbf{F}_p|^2 \tag{3}$$

Where the data we are given in $nat.txt = |\mathbf{F}_p|^2$, $hev.txt = |\mathbf{F}_{hp}|^2$, and $|\mathbf{F}_h|^2 = 100^2$. Utilizing $|z|^2 = z \cdot z^*$ we expand the right hand side:

$$|\mathbf{F}_{hp}|^2 = (\mathbf{F}_h + \mathbf{F}_p) \cdot (\mathbf{F}_h^* + \mathbf{F}_p^*) \tag{4}$$

$$|\mathbf{F}_{hp}|^2 = |\mathbf{F}_h|^2 + |\mathbf{F}_p|^2 + \mathbf{F}_p\mathbf{F}_h^* + \mathbf{F}_p^*\mathbf{F}_h \tag{5}$$

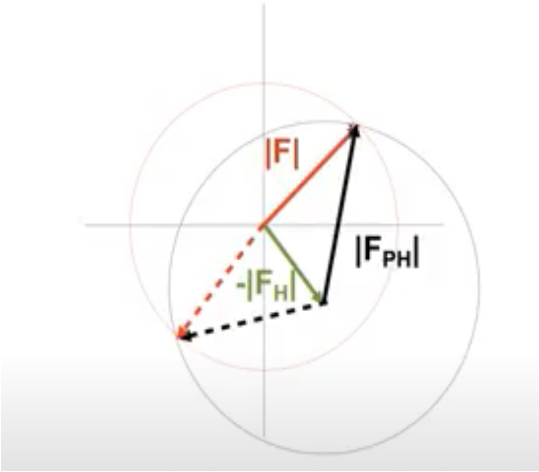


Figure 1: A visualization in the complex plane of the relationship between $\mathbf{F}_h$, $\mathbf{F}_p$, and $\mathbf{F}_{hp}$. The circles with a radius equal to the amplitude of $\mathbf{F}_p$ and $\mathbf{F}_{hp}$ indicate all the vectors that can be obtained with all the possible phase angles. There are two possible values for $\mathbf{F}_p$ that agree with the measured amplitudes and with the heavy atom models amplitudes.

We know that the diffraction data that is $\mathbf{F}_h$ is equal to 100 everywhere, $\mathbf{F}_h = 100$, and so it is a purely real value. Therefore the complex conjugate $\mathbf{F}_h^* = \mathbf{F}_h = 100$. We can also simplify the last two terms by saying $\mathbf{F}_p + \mathbf{F}_p^* = 2Re(\mathbf{F}_p)$. Now we have:

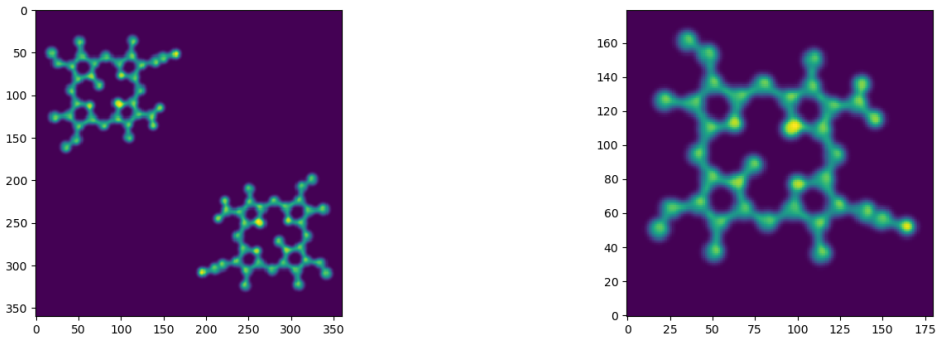
$$|\mathbf{F}_{hp}|^2 = 100^2 + |\mathbf{F}_p|^2 + 200Re(\mathbf{F}_p) \tag{6}$$

$$200Re(\mathbf{F}_p) = |\mathbf{F}_{hp}|^2 - |\mathbf{F}_p|^2 - 100^2 \tag{7}$$

Now we can take the inverse Fourier transform of both sides. Recall that from homework 1, we found that $IFT(Re(FT(f(x)))) = \frac{f(x)+f(-x)}{2}$. Therefore,

$$200\frac{f_p(x,y) + f_p(-x,-y)}{2} = IFT(|\mathbf{F}_{hp}|^2 - |\mathbf{F}_p|^2 - 100^2) \tag{8}$$

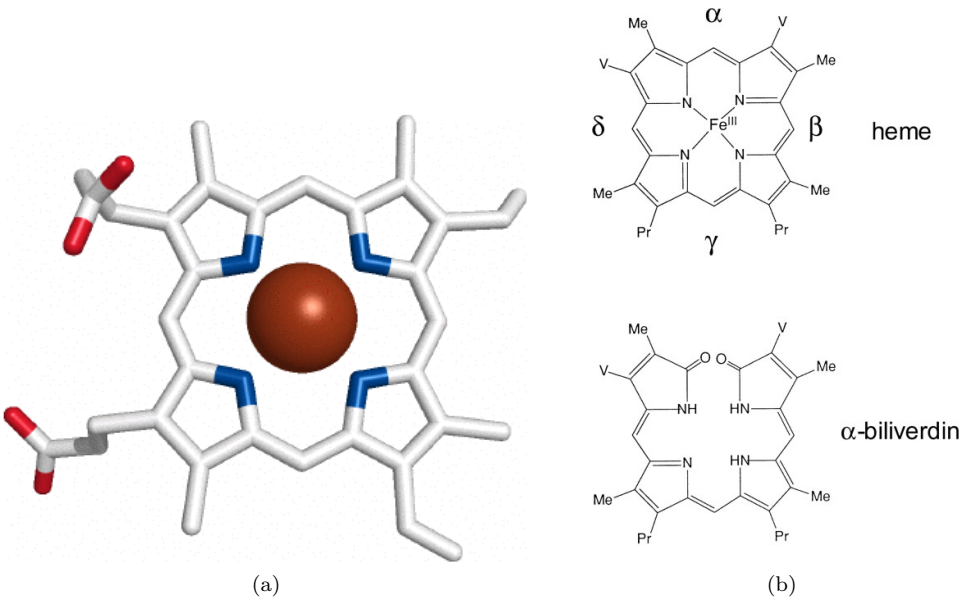
We can now use python and scipy to take the inverse Fourier transform of everything on the right hand side. Utilizing matplotlib.pyplot.imshow(), we can map the resulting array of 2d data using normalization and a colormap. As can be seen in Equation 8, what we will get is an image of the original molecule along with its mirrored image over the x and y axes (Figure 2a). This is a result of only considering the real part of the Fourier transform, $\mathbf{F}_p$. Figure 2b shows the original molecule without its mirror image. In order to clearly see both images it is important to make the system smallish compared to the region were looking at. We want to be able to see both the original and mirrored part of the fourier transform without any overlap or interference. This lets us distinguish between positive or negative phase data as can be visualised in Figure 1.



(a) (b)

Figure 2: (a) Resulting image of the original heme molecule and its mirror image after the Inverse Fourier Transform was applied as in Equation 7. (b) Original molecule with the origin now at the bottom left.

The molecule we are seeing:



(a) (b)

Figure 3: (a) Heme molecule structure with Fe(II) ion in center (uwaterloo.ca, webnotes. Web 2021). (b) β-Heme with Fe(II) ion and α-biliverdin Heme shown without the Fe(II) ion, since at a different step of the interaction with a metal ion and molecule (Pendrak et. al, 2004).

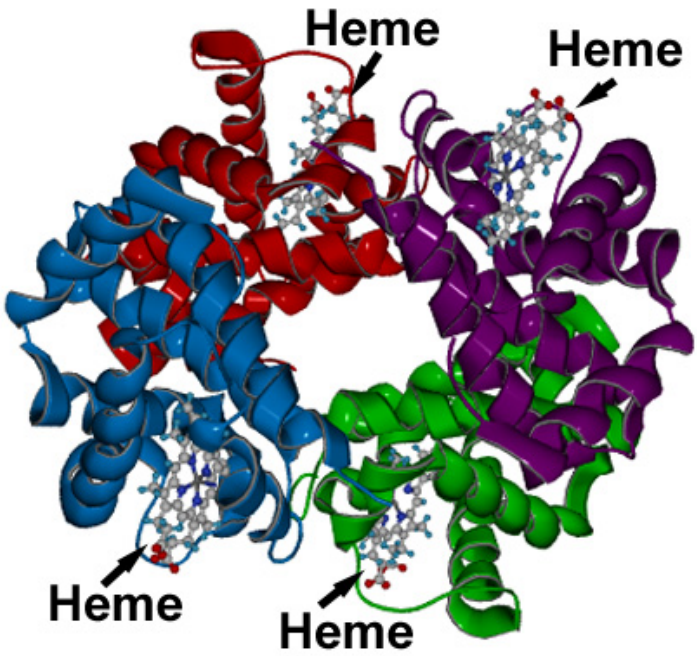


Figure 4: Hemoglobin crystalline α-helices structure with Hemes (Casiday et al., 2008).

2 The Heme Molecule in Hemoglobin

The Hemoglobin protein transports oxygen in blood from the lungs to other tissues. Polypeptide chains are created by proteins that form via linking of amino acids, where individual amino acids are known as residues. The specific arrangement and molecular interactions of these residues in every protein determine the shape and function. Hemoglobin is considered a globular protein, because it is folded into a nearly spherical shape. It has four subunits (Figure 4), where each protein subunit (blue, red, purple, green) is an individual molecule that is connected to neighboring subunits through intermolecular interactions.

Each subunit in hemoglobin contains a Heme group (Figure 3 (a)(b)). Each Heme group contains an iron atom that is able to bind to one oxygen molecule. This facilitates each hemoglobin protein to be able to bind four oxygen molecules. An important note to this process is that one of the most important classes of chelating agents in nature are the porphyrins, because they can coordinate the binding to a metal, Fe(II) in the case of a Heme, using the four nitrogen atoms as electron-pair donors (Casiday et al. 2008). Chelation is the bonding of ions and molecules to metal ions, and involves the formation or presence of two or more separate coordinate bonds between a polydentate ligand (chelant) and a single central atom (Wikipedia, 2021).

3 Further Questions and Analysis

1. Is it necessary that we be supplied with both nat.txt and hev.txt to find the structure, or can we do it from only nat.txt?

For certain specimens, we do not need any diffraction data from a heavy atom substituted specimen (hev.txt), and so the only diffraction data we need is from the original microscopic structure (nat.txt). As described in the article, *Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens*, by Miao et al, it is possible to reconstruct a microscopic sample from an X-ray diffraction pattern whether it be a crystalline specimen or a non-crystalline specimen. The article describes the latter case in which the specimen is a collection of gold dots deposited on a silicon nitride membrane arranged in the shape of letters. As we saw in this assignment, the intensity of the diffraction pattern provides data of the diffraction amplitude but not the phase (crystallography phase problem). However the situation for this non-crystalline specimen is different in that the diffraction pattern is continuous and not discrete. If this continuous pattern is oversampled on a finer scale, one can reconstruct the original data set using a specific iterative algorithm. No high resolution X-ray optical detectors are needed in this X-ray microscopy technique. It is to note that image resolutions of about 100 times lower than conventional X-ray crystallography are obtainable, but this method is also applicable to structures up to 100 times larger.

2. Experimental problem where we don't have a single crystal and have to find the structure of a system from diffraction data. Describe the difference between diffraction experiments from a single crystal and powder diffraction.

The experimental problem of trying to find the structure of a system from diffraction data without having a single crystal is a common one, especially when it comes to newly synthesized materials (which solid-state chemists constantly aim for). In these cases, single crystals may not always be readily available at the time of synthesizing, so single crystal X-ray diffraction is not a feasible method of analysis in determining the structure of the system. When looking at materials that are multi-phase (those that exist in a matrix like cement, concrete, soil, etc.), we will not have a single crystal to do the diffraction analysis with this method either. In these instances, one can look at X-ray powder diffraction as a method to determine the structure. In essence, X-ray powder diffraction is a phase analysis technique, where even if the system is a mixture and multi-phase, the individual crystalline phases and their concentration can be determined (Font-Bardia, Alcobe, 2012). Unknown material samples can also have their structure determined rather quickly with this method, due to its rapid and non-destructive analysis. This is accomplished by comparing the resulting patterns to diffraction data bases.

Single crystal X-ray diffraction differs from X-ray powder diffraction notably in the texture and composition of the sample. Regarding the process of diffraction, it occurs whenever Bragg's law is satisfied, which can occur with a variable wavelength and a fixed angle, or a fixed wavelength and a variable angle. With single crystal X-ray diffraction, the angle is fixed, and the wavelength is variable. On the other hand, with X-ray powder diffraction the wavelength is fixed, and the angle is variable (Bish et al, 2014). With the powder method, there is no need for a rotation to yield a diffraction event compared to a single crystal. This is due to the inherent randomness present in the composition of the powder, with a (theoretical) infinite number of crystalline orientations being represented equally. The assumption is that they are randomly arranged. The measurement of all the crystal orientations at the same time is extremely beneficial when it comes to short collection times, such as when the system being analyzed is inherently unstable. That being said, the usefulness of single crystal X-ray diffraction should not be ignored. Though it can be very time-consuming compared to its powder method counterpart, the accuracy of the single crystal method is much higher, making it the more powerful method for determining structures. With powder X-ray diffraction, important 3D crystallographic information is lost due to collapsing it into a 1D axis diffraction pattern (Oviedo et al, 2019).

3. How much is known about the structure of membrane proteins, and why is knowing the structure important? Is it easy to crystallize a membrane protein? How is the structure determined? Briefly describe a recent experiment or research project on membrane protein crystallography.

Around 30 % of the human genome codes for membrane proteins, meaning that approximately one third of the genetic information within our cells is solely geared toward the production of these proteins. Membrane proteins are critical for all cellular functions, in particular for cell communication and cell transport pathways. They are found to be elusive due to their hydrophobic nature, which leads to difficulty in structural studies. This is because they can't be dissolved in water, and furthermore are prevented from crystallizing, which is a necessary step in the x-ray crystallography technique. Once membrane proteins are extracted from cell membranes, they are made water-soluble through a costly and selective process by suspension in detergents mimicking the hydrophobicity of a cell membrane. Furthermore, these detergents can disrupt the structure and function of membrane proteins, since they interfere with intermolecular and intramolecular protein to protein interactions. There are approximately 8000 known membrane proteins found in human cells, but thus far only around 150 have a determined structure (Martin et al., 2019).

Many cellular processes are mediated by membrane proteins. These interactions are fundamental for the success of biological cells, and membrane-embedded transporters are used to move ions and larger molecules as well as solutes across cell membranes. Membrane proteins are the key component in successfully specified recognition of what may or may not pass through the cell membrane. Their receptors mediate communication between the cell and its environment, and specific membrane-embedded enzymes will catalyze the appropriate chemical reactions. It is therefore crucial to understand how these mechanisms are steered by different membrane proteins and their specific structure, as well as how proteins couple to their hydrated lipid membrane environment (Cournia et al., 2015).

Membrane proteins are among the most challenging targets. Cells and organelles are contained within a hydrophobic lipid bilayer where these proteins are embedded in the bilayer, and many have additional domains outside the membrane. Protein crystallization is the process of testing a large number of possible reagents, and optimization is necessary to obtain well-diffracting crystals. In the case of membrane protein crystals we find a high solvent content due to the detergent micelle, which covers the hydrophobic part of the protein. The crystals are often fragile, difficult to handle, as well as diffract to low resolution. Furthermore, they suffer from radiation damage during diffraction experiments (Carpenter et al., 2008).

The majority of structures in the Protein Data Bank (PDB) archive were determined using X-ray crystallography. When using this method, the protein is purified and crystallized, then subjected to an intense beam of X-rays. The proteins in the crystal will diffract the X-ray beam into characteristic patterns of spots. These spots will then be analyzed using tricky methods to determine the phase of the X-ray wave in each spot, and this will determine the distribution of electrons in the protein. The result is a map of the electron density which is interpreted to determine the location of each atom. The PDB archive contains two types of

data for crystal structures. The advantage is that X-ray crystallography can provide detailed atomic information, showing every atom in a protein or nucleic acid along with atomic details of ligands, inhibitors, ions, and other molecules that are incorporated into the crystal. It can have limitations on the types of proteins that can be studied by this method, since it is more optimal for rigid proteins forming ordered crystals. It has not proven to be a good method for flexible proteins since it it based on many molecules aligned in exactly the same orientation, yet these flexible parts of proteins will be invisible in crystallographic electron density maps. This is due to their electron density being dispersed over a large space (PDB, Web. 2021).

In their paper, Kwan et al. address the combination of several methods being most optimal for finding 3D structures of membrane proteins, of which X-ray crystallography is pointed out as one of the most successful methods for solving 3D protein structures at atomic level. Limitation are still found in obtaining well-ordered protein crystals that diffract at high resolution, such as with membrane proteins, yet the knowledge of membrane protein structure at the biochemical level is vital for medicine and drug discovery. The method of X-ray crystallography and determining membrane structure has improved in the last two decades, and scientists have been active and successful in solving membrane protein crystal structures. This new data helps in revealing many of their mechanisms of action. Technical maturation of biophysical methods such as X-ray crystallography and cryo-EM these past few years has allowed scientists to pursue the study of more complicated systems such as multi-protein complexes, as well as protein-DNA/RNA complexes. The arrival of serial crystallography and XFELs, has now made it possible to view structural dynamics of proteins over time effectively allowing the creation of three-dimensional movies that display crucial biological processes in the cell membrane and cell (Kwan et al., 2020).

4 References

Bish et al. "Diffraction: Powder Method." IUCrJ, vol 1, Part 6 (2014).

Carpenter, Elisabeth P et al. "Overcoming the challenges of membrane protein crystallography." Current opinion in structural biology vol. 18,5 (2008): 581-6. doi:10.1016/j.sbi.2008.07.001

Casiday, Rachel, and Regina Frey. "Hemoglobin and the heme group: Metal complexes in the blood for oxygen transport." Retrieved March 28 (1998): 2008.

Cournia, Zoe et al. "Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory." The Journal of membrane biology vol. 248,4 (2015): 611-40. doi:10.1007/s00232-015-9802-0

Font-Bardia, Merce and Xaxier Alcobe. "X-ray single crystal and powder diffraction: possibilities and applications." Handbook of instrumental techniques from CCiTUB. (2012).

https://en.wikipedia.org/wiki/Chelation

Tristan O. C. Kwan, Danny Axford, Isabel Moraes; Membrane protein crystallography in the era of modern structural biology. Biochem Soc Trans 18 December 2020; 48 (6): 2505–2524.

Martin, Joseph, and Abigail Sawyer. "Elucidating the structure of membrane proteins." (2019): 167-170.

Miao, Jianwei, et al. "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens." Nature 400.6742 (1999): 342-344.

Oviedo, et al. "Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks." (2019).

https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure

Pendrak, Michael L., et al. "Heme oxygenase in Candida albicans is regulated by hemoglobin and is necessary for metabolism of exogenous heme and hemoglobin to α-biliverdin." Journal of Biological Chemistry 279.5 (2004): 3426-3433.

http://watcut.uwaterloo.ca/webnotes/Metabolism/Iron.html