Alexander Seymour
Biology 555: Macromolecular Crystallography
Take Home Final Exam
July 31, 2012

1. a.)

$$e^{ix} = \cos(x) + isin(x)$$
$$\left(e^{ix}\right)^* = \cos(x) - isin(x)$$
$$\left(e^{ix}\right) * \left(e^{ix}\right)^* = cos^2(x) + sin^2(x)$$
$$\left|e^{ix}\right| = \sqrt{(e^{ix}) * (e^{ix})^*} = \sqrt{cos^2(x) + sin^2(x)} = 1$$

1. b.)

$$Given: z = a + ib, show\ that\ \left|e^{iz}\right| = e^{-b}$$
$$\left|e^{iz}\right| = \left|e^{i(a+ib)}\right| = \left|e^{ia-b}\right| = \left|e^{ia}\right| * \left|e^{-b}\right| = 1 * \left|e^{-b}\right| = e^{-b}$$

2. a.) Orthorhombic

2. b.) This space group has one twofold rotation axis parallel to the unit cell direction *z*, and two screw axes parallel to the unit cell directions *x* and *y*.

2. c.) There are three Harker sections:

$$(u, v, w) = (x, y, z) - \left(x + \tfrac{1}{2}, -y + \tfrac{1}{2}, -z\right) = \left(\tfrac{1}{2}, 2y + \tfrac{1}{2}, 2z\right)$$
$$(u, v, w) = (x, y, z) - \left(-x + \tfrac{1}{2}, y + \tfrac{1}{2}, -z\right) = \left(2x + \tfrac{1}{2}, \tfrac{1}{2}, 2z\right)$$
$$(u, v, w) = (x, y, z) - (-x, -y, z) = (2x, 2y, 0)$$

The conditions that will yield Harker sections are:

$$\left(\tfrac{1}{2}, v, w\right)$$
$$\left(u, \tfrac{1}{2}, w\right)$$
$$(u, v, 0)$$

3.

Rubredoxin is an iron-containing protein of relatively low molecular weight that is found in sulfur metabolizing bacteria and archaea. Rubredoxin typically participates in electron transfers in biological systems. This essay shall compare and contrast the structure of rubredoxin as determined by two different methods: neutron crystallography and X-ray crystallography. The associated PDB accessions codes are 1VCX and 3KYU, respectively.

The structure of 1VCX was determined by using the neutron single-crystal diffractometer at the JRR-3M reactor of the Japan Atomic Energy Research Institute at room temperature and a resolution of 1.5 Å with the completeness of the data being 81.9%. The model contains 413 protein atoms as well as 37 solvent molecules. The model has been refined to yield final agreement factors of R = 18.6% and R_{free} 21.7%. Formerly ambiguous assignments derived from old X-ray data are now clearly visible in the neutron structure. Additionally, Fourier mapping shows an extensive set of hydrogen bonds involving the $ND_3{}^+$ boundary that the authors suggest may contribute to the unusual thermostability of the molecule.

The PDB file 3KYU has the added advantage of joint X-ray and neutron diffraction studies. The structure of 3KYU was determined at room temperature and a resolution of 1.10 Å (X-ray) and 1.65 Å (neutron) with the completeness of the data being 95.6%. The model contains 421 protein atoms as well as 137 solvent molecules. The model has been refined to yield R and R_{free} values of 0.107 and 0.125, respectively. The authors demonstrate that the 1.65 Å resolution neutron data is eight times more likely to provide high-confidence positions for D-atoms as opposed to the 1.10 Å X-ray data. For example, at or above the 1.0-sigma level, the X-ray data defines 342 of 378 D-atom positions for D-rubredoxin, whereas the neutron data defines 291 of 365 D-atom positions for HD-rubredoxin.

4. a.) The refining process involves optimizing the model parameters to fit the target function. An error in the structure amplitude for a single amino acid side chain will not affect the overall correctness of the model, assuming the majority of the structure amplitudes have been properly refined.

4. b.) I would expect to identify this error when reviewing the electron density map for the structure.

5.

Multiwavelength anomalous diffraction (MAD) allows for an atom's electrons to absorb X-rays of a given wavelength and reemit the X-ray after a short time delay, thus inducing a phase shift in all of the reflections. Multiple isomorphous replacement (MIR) involves the insertion of heavy atoms into the structure via soaking or synthesizing the proteins with equivalents. MAD is advantageous as it requires a single crystal without model bias. However, MAD is dependent on accurately measuring rather small differences in the intensities, which then requires more data to be collected in order to measure the reflections with greater accuracy. At cost, this can lead to radiation damage. Another disadvantage of MAD is that not enough heavy atoms may be present to get a strong signal. MIR is advantageous because the data can be determined locally without model bias. Unfortunately, MIR is time consuming, as it requires the analysis of numerous crystals. After soaking, however, most crystals will not diffract well, crack or have unit cell and/or space group changes. An alternative technique known as molecular replacement relies on the existence of a previously solved protein structure that is similar to the new structure being determined. The similar molecule's phases are grafted onto the intensities that are being experimentally determined which can then lead to theoretical structure bias. This method is very quick to implement in a computer program and works well with multiple molecules in the asymmetric unit. However, running this code may take several hours to run depending on the size of the unit cells and the number of trials being performed.

6.

Twelve ways in which the parameters of two adjacent base pairs in a B-DNA double helix can vary from the average B-form structure of DNA are:

1. Shear (Sx) (Å)
2. Stretch (Sy) (Å)
3. Stagger (Sz) (Å)
4. Buckle (kappa) (degrees)
5. Propeller (pi) (degrees)

6. Opening (sigma) (degrees)
7. Shift (Dx) (Å)
8. Slide (Dy) (Å)
9. Rise (Dz) (Å)
10. Tilt (tau) (degrees)
11. Roll (rho) (degrees)
12. Twist (omega) (degrees)

7.

The repeat distance of this diffraction pattern is 232 nm. The diffraction pattern develops from the periodic array of the myosin, which contains thick filaments as well as from the actin, which contains thin filaments. If the muscle were to be contracted instead of relaxed, the diffraction pattern would show a loss of intensity of the meridian reflections.

8.

Overall, short time scales will result in the highest resolution. As the time scale increases, the resolution will decrease. Models that utilize a slower time scale may be developed from combining calibrated fast time scale models with the experimental data.

X-ray absorption fine structure (XAFS) refers to the details of how X-rays are absorbed by an atom at energies near or above the core-level binding energies of that particular atom. XAFS spectra are sensitive to the formal oxidation state, coordination chemistry as well as the distances, coordinates, and species of the atoms immediately surrounding the selected element. Crystallization is not a requirement for XAFS measurements, making it ideal for the analysis of non-crystalline and highly disordered materials. XAFS offers the possibility of fast measurements of *in situ* chemical processes, high spatial resolution as well as extreme conditions of temperature and pressure. Practical experimental limitations to XAFS include availability of synchrotron sources, energy ranges, beam sizes and intensities.

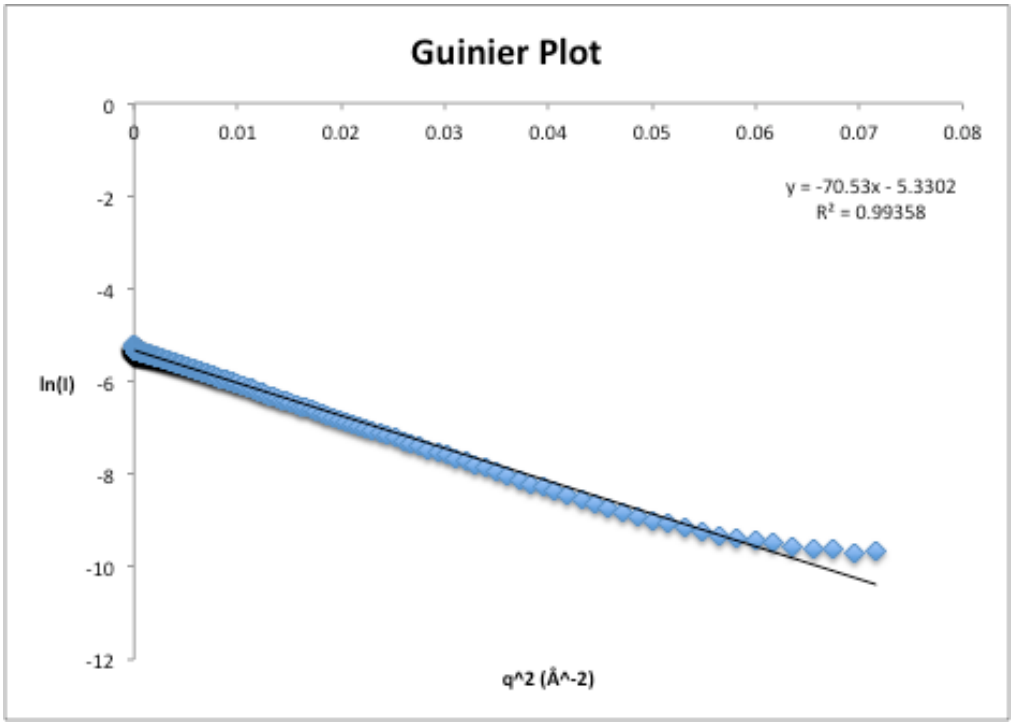
Small-angle X-ray scattering (SAXS) is a fundamental method for structure analysis of condensed matter. SAXS has a modest resolution as compared to X-ray diffraction, which makes it not sufficient to reveal the atomics structure of materials. Used in "scanning" mode, SAXS can show the large-scale organization of solid materials. It can provide structural information on inhomogeneities of the electron density with characteristic dimensions between one and a few hundred nanometers. SAXS has short response times so it can be used to follow biological processes in real time, which makes SAXS an ideal complement to time-consuming analytical techniques such as electron microscopy and X-ray diffraction.

Nuclear magnetic resonance (NMR) is the phenomenon that occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to a second oscillating magnetic field. NMR spectroscopy uses this phenomenon to study physical, chemical and biological properties of complex molecules. When irradiated with a radio frequency signal, the nuclei of a molecule can change its spin orientation. This technique can differentiate between structural isomers and can provide information about the connectivity between atoms within a molecule.

X-ray crystallography is an experimental technique that exploits the diffraction of incident X-rays on crystalline structures. Based on the diffraction pattern obtained, the electron density can be mapped. Additional phase information can be discerned from the diffraction data in order to complete the map. A model is then progressively built into the experimental electron

density map, refined against the collected data and the result is an accurate molecular structure. In contrast to NMR, X-ray crystallography offers no size limitations for the molecules or complexes being studied.

9.



From the Guinier equation:

$$I(q) = I(0)\left[1 - \frac{q^2 {R_g}^2}{3}\right]$$

As q approaches zero:

$$I(q) = I(0)\exp\left(\frac{q^2 {R_g}^2}{3}\right)$$

This can be linearized as follows:

$$\ln I(q) = \ln I(0) - q^2 \frac{{R_g}^2}{3}$$

The slope of a Guinier plot is ${R_g}^2/3$. When $q_{max} = 0.26$, R_g will become 14.5 Å.
Given "the rule of thumb" where $q_{max}\ x\ R_g < 1.3$ where $qR_g < 1.3$ for a globular protein and $qR_g < 0.8$ for an elongated protein, a new value for q_{max} can be chosen to better approximate the value for R_g. After a few iterations, the new value of q_{max} is 0.087.
If the protein is assumed to be globular, the radius of sphere can be calculated using the following equation:

$$R_g = \sqrt{\frac{3}{5}}R$$

Thus, the radius of the sphere would be 18.8 Å.

To calculate the principle axes *a* and *b* of a prolate ellipsoid, the following equations will be useful:

$$R_g{}^2 = \frac{1}{5}[a^2 + b^2 + c^2], \text{ where } c = b$$

$$R_g{}^2 = \frac{1}{3}\left[\left(\frac{a}{b}\right)^{\frac{4}{3}} + 2\left(\frac{b}{a}\right)^{\frac{2}{3}}\right]$$

Solving this system of equations with the TI-89 gives confusing results:

$$a = 32.4 \; and \; b = 0.258$$

or

$$a = 0.00409 \; and \; b = 22.9$$