Final Exam

Ed Fuchs

Macromolecular Structure Determination

1. a. $|e^{ix}| = (e^{ix}\, e^{ix^*})^{1/2}$

$e^{ix^*} = \cos x - i\sin x$

$|\cos x + i\sin x| * |\cos x - i\sin x| = (e^{ix}\, e^{ix^*})^{1/2}$

b. $|e^{iz}| = \cos z - i\sin z = \cos(a+ib) - i\sin(a+ib) = e^{-b}$

2. a. Primitive Orthorhombic

b. the two-fold symmetry exists about the Z-axis, the screw symmetry exists about the x- and y-axes.

c. $(uvw)_1 = (x,y,z)-(-x+½, y+½, -z) = 2x+ ½, ½, 2z$

$(uvw)_2 = (x,y,z)-(x+½,- y+½, -z) = ½, 2y+½, 2z$

$(uvw)_3 = (-x,-y,z)-(-x+½, y+½, -z) = ½, -2y+½, 2z$

3. Ribonuclease A has been analyzed by several different methods, including both NMR and x-ray crystallography (PDB entries 2AAS and 1 Z6S). For these experiments, the x-ray structure was determined first, then the binding of specific ligands, namely IMP (a competitive inhibitor), were examined. The NMR was used to verify the structure of the protein, and determine if the crystal structure varied from the solution structure. The structures were almost identical, aside from the following discrepancies. The NMR structure showed no motifs in the region of amino acids 87 to 96, and a large turn from 33 to 36. The crystal structure showed a short turn from amino acids 33 to 34, and a turn from 88 to 89 and 92 to 93. It also found two additional beta bridges from 90 to 91 and 93 to 94.

The structure from x-ray crystallography included more structures, however since the NMR experiments were designed to examine the protein in solution, it may give a more realistic view of how the protein exists in its natural state. Furthermore, this suggests that the aforementioned regions are more flexible in solution than as crystals, which may indicate a slight change in function. This is further evidenced by the crystallography experiments in which IMP was used as an inhibitor. The crystal structure indicates that binding occurs in a novel fashion, such that subsite B2 will not bind inosine, but the nucleobase is still attached via side chain interactions with Glu111. The NMR structure disputes some of the turns which are physically near this site, suggesting that this is not a novel binding site, but instead a situation in which crystallization has changed the binding site to make it less conducive for inosine to bind.

References

Hatzopoulos, GN, et al. “The Bidnign of IMP to Ribonuclease A.” FEBS Journal. August 2005; 272(15): 3988-4001.

Santoro, J. et al. “High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy.” Journal of Molecular Biology. Feb 5, 1993; 229(3): 722-734.

4. a. The incorrect side chain position would result in small errors on a large number of amplitudes, since the intensity of each diffracted atom contributes to all Bragg Spots, the error of one side chain would manifest as small error amongst all points.

b. This error could be identified by fitting the side chain to the most likely electron density, and by fitting within the predicted areas of the Ramashadran plot.

5. Experimental phasing, such as MIR, is used to determine phase by comparing the native protein crystal with a protein co-crystalized in heavy metal ions. This difference is mapped as a Patterson map, which allows for the metal ions to be located, identified, and properly phased. This method assumes that the metal ions will not affect the crystallization, is limited by the accessibility of the region of the protein to the metal ion, and requires an analysis of two crystals. Molecular replacement uses a model with (ideally) high sequence homology to the unknown protein structure, and orients and translates the model to fit. This can be helpful for determining structure, and is generally less labor intensive and requires overall fewer experiments and chain tracing to be done. This technique only works, however, if the model has high sequence homology, and risks model bias since the model may be critically different from the real protein structure despite the sequence homology. In this circumstance, 58% sequence homology is much higher than the minimum, and certainly a viable option, however if you have the time and resources, it would be the most accurate to do MIR.

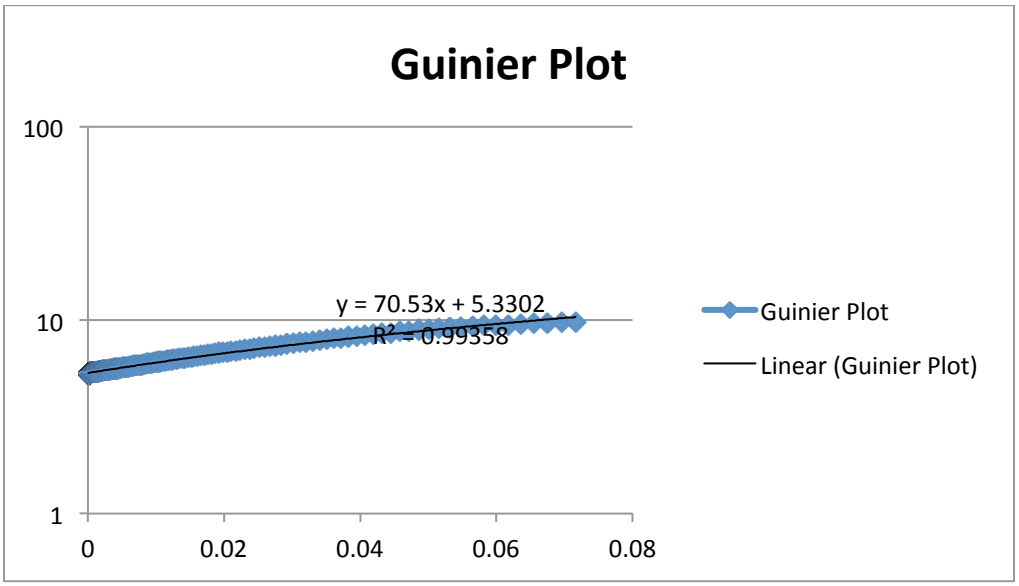
6. Two adjacent base pairs of B-DNA can vary in the following geometries: helical twist, roll, tilt, rise, slide, propeller twist, shift, stagger, stretch, shear, buckle, and opening.

7. The distance between layer lines are multiples of 0.85nm, which was found from taking the distance between layer lines, and finding the lowest multiple. In a contracting muscle, the distance between subunits would be smaller, and therefore the layer lines would be condensed along the z-axis, with little change along the R-axis. If it were a contraction due to helical twist, both the z- and R- dimensions would be condensed.

8. Many computational tools are used in structural biology in order to determine the structure of a cellular component. On the smallest scale, quantum modeling is used to examine and refine the structure with techniques such as Quantum Monte Carlo, Post-Hartree-Fock, and Kohm-Sham equations, which all focus on electrons and subatomic particles within the cellular component. The

Quantum Monte Carlo is a computer simulation of the quantum system in order to predict the interactions between many atoms working together in one system. The Post-Hartree-Fock is used to measure repulsions between electrons in a system, and the Kohm-Sham method is used to predict how the particles would react if they were completely independent. All 3 of these methods is used to predict the electron density, which aids in phasing and structure refinement of the amino acids. The next level are the atomistic models, which use a salvation models to predict the behavior of a biological structure in solvents. Since biological systems include a wide variety of solvents, from aqueous to lipid, modeling the structure in various solvents can assist in the understanding of protein structure within the cell. This aids in predicting the positions of side chains, and in some circumstances can predict the folding of the biomolecules, which is intrinsically tied to its function. The next level can take anywhere from minutes to hours, and includes coarse-grained, mean-field, kinetic, and mechanical models. These are all contributed to by the faster modeling systems, and are used in chain tracing, overall protein folding, and side chain phasing. All three levels work together to give maximal refinement and accuracy of the biomolecule, allowing for confident predictions of structure and therefore confident analyses of function.

9. The Guinier plot was made using lnI and q2 values:



The slope of this line gives R_g^2, therefore R_g= 8.398, and the maximum Q value is 0.267701. The product of these two values is 2.248, which is above the 1.3 threshold. This means that a value of Q_{max}= 0.1540 would give a product of 1.3. All Q-values above this number were removed, thus only the first 55 points were used. Assuming a prolate ellipsoid, $R_g^2 = (a^2+b^2+c^2)/5$, thus $8.398 = (a^2+b^2+c^2)/5$.

$41.99 = a^2+b^2+c^2$

As a prolate ellipsoid, b and c should be equal

6.479= a + 2b

a=3.24

b=1.62

c=1.62