***1. Structure math*:**

a. The magnitude or absolute value of a complex number is defined as the square root of the product of the complex number and its complex conjugate: That is, for a complex number $z = a + ib$, with a and b real, we define its complex conjugate z^* as $a - ib$ and $|z| \equiv \sqrt{(zz^*)}$.Use the Euler formula ($e^{ix} = \cos x + i\sin x$) to show that the absolute value of e^{ix} is one,provided that x is real.

Answer: let $e^{ix} = \cos x + i\sin x = z$,

$z^* = \cos x - i\sin x$

Therefore $|z| \equiv \sqrt{(zz^*)} = [(\cos x + i\sin x)(\cos x - i\sin x)]^{1/2}$

$= [\cos^2 x - i^2\sin^2 x]^{1/2} = [\cos^2 x + \sin^2 x]^{1/2}$

$e^{ix} = \sqrt{e^{ix}e^{ix}}^* = 1$

so $|e^{ix}| = 1$

b. Let the complex number $z = a + ib$, with a and b real. Show that $|e^{iz}| = e^{-b}$.

Answer: let $y = e^{iz} = \cos z + i\sin z$

$y^* = \cos z - i\sin z = \cos(-z) + \sin(-z) = e^{iz}$

Because $\cos(z) = \cos(-z)$, $-\sin(z) = +\sin(-z)$

$|e^{iz}| = |Y| = \sqrt{(yy^*)} = (e^{iz}\, e^{-iz})^{1/2} = [e^{i(a+ib)}\, e^{-i(a-ib)}]^{1/2}$

$= [e^{ia+i_2b}\, e^{-e+i_2b}]^{1/2}$

$= [e^{-b} \times e^{-b}]^{1/2} = [e^{-2b}]^{1/2} = e^{-b}$

*2. **Crystallography*: Space group $P2_12_12$, International Tables number 18, has the following equivalent positions:**

(x, y, z) ; $(-x, -y, z)$; $(-x + ½, y + ½, -z)$; $(x + ½, -y + ½, -z)$.

a. Is this space group triclinic, monoclinic, orthorhombic, tetragonal, trigonal /hexagonal, or cubic?

Answer: The space group is orthorhombic.

b. In which direction do you find a 2-fold axis, and in which directions do you find two-fold screw axes in this space group?

Answer: Along the Z direction is 2- fold axis and along X and Y direction is 2-fold screw axes.

c. What conditions on u, v, and / or w will define the Harker sections in this space group? Show your work. Hint: There are three Harker sections.

Answer: The Patterson map can determines the relative distance not real distance. Therefore, Harker section can explain which determine the atom in the symmetry

plane.

(x, y, z) - (-x, -y, z) = (2x, 2y, 0),

(x, y, z)- (-x + ½ , y + ½ , -z) = (2x-1/2, -1/2, 2z)

(x, y, z)- (x + ½ , -y + ½ , -z).= (-1/2, 2y-1/2, 2z)

(-x,-y,z) - (-x+1/2, y+ 1/2, -z) =(-1/2, -2y-1/2, 2z)

(-x,-y,z) - (x+1/2, -y+1/2, -z) = (-2x-1/2, -1/2, 2z)

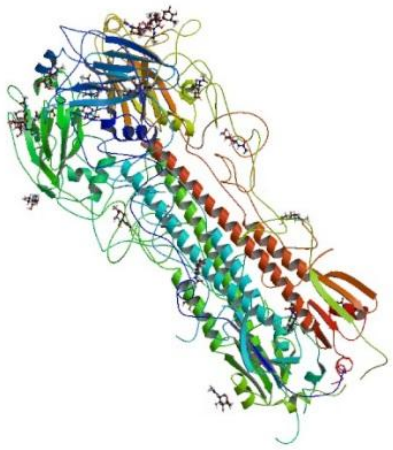
(-x+1/2, y+ 1/2, -z) - (x+1/2, -y+1/2, -z)= (-2x, 2y, 0)

Above of all, when -1/2=-1/2+1=1/2 ; 0=0+1=1.

It will got u=1/2; v=1/2; w=1.

***3. NMR and X-ray*: Search the Protein Data Bank (www.rcsb.org) for a protein whose structure has been determined both by NMR and by X-ray crystallography. Summarize in a three paragraph (roughly 15-sentence) essay the similarites and differences between the two structures. You should reference published papers about these structures as well as the PDB submissions themselves. Cite your sources**

Answer: I chose "Hemagglutitin(PDB:1HGJ)" to answer this question. The information which comes from PDB as shown as follow:



The figure shows left, which is molecule structure of 1HGJ that used X-ray diffraction to predict the structure. Results of NMR binding and X-ray crystallization show similar overall structures, and therefore similarities in structural functions may also be found. For example, the ligand affinity may be increased by the a-O-methyl glycoside with a (4'-ben zylamidocarboxybutyl) group on the protein. An X-ray study clarifies the binding affinities with the crystallographic structure. For instance, in a research, a map at 3-A resolution of the X-31(α-2,3)sialyllactose complex verifies previous research of the location of three ligand binding sites on the Hemagglutinin trimer that related to each other by threefold noncrystallographic symmetry. Furthermore, the NMR experiment allows us to substantiate the binding data from three-dimensional structural information from five BHA-sialoside complexes. NMR results in studies show the substitution of the a-O-methyl glycoside with a (4'-ben zylamidocarboxybutyl) group enlarges ligand affinity for the first site by a factor of

24. While the X-ray studies clearly suggested the location of an α-glycosidic chain (4'-benzylamidocarboxybutyl) that increases the binding affinity of sialic acid by a factor of about 3.

Slight structures were found to be diverse with the two methods, X-ray and NMR, though the ligand affinity both changed insignificantly in the experiments. The study of the BHA-sialyllactose complex suggested that without directly contacting the protein, the R4 substituent sticks out of the binding site toward solution. And NMR spectrum supports the fact that the acetyl group does not contact the protein. The X-ray structure, however, confirms that the acetyl group is highly solvent accessible. An NMR experiment indicates that the acetyl group changes the ligand affinity slight, by introducing longer chains at the 4-hydroxyl position, replacing the acetamido methyl group with an ethyl group, and removing the 7-hydroxyl group, these modifications still had insignificant effect on binding. Crystallographic refinement placed one of the atoms in the acetyl group, within 3.3 Å of the carbonyl oxygen of Gly-135 and within 3.0Å of the hydroxyl group of Ser-145. No evident changes or contacts were expected since the B values were high for the acetyl atoms.

The Existence of a Second Binding Site on the protein was evidently found by both methods; however, the particular position of the second site wasn't completely identical. All together the below information data settles the position and orientation of the Neu5Ac moiety by x-ray diffraction structure. A difference electron density map indicated a second set of three symmetry-related peaks which represent (α2-3)sialyllactose molecules. Protrusions were corresponding to the glyceryl, N-acetamido, 4-hydroxyl, and 1-carbox-ylate groups of the α-sialoside on the electron density map. In this case shows the second site locates at the interface between the HA1 and HA2 polypeptide chains. Moreover, NMR studies show that a saccharide, DSL contains both a 2,6- and a 2,3-linked Neu5Ac moiety. Interesting enough, both moieties occupy the hemagglutinin binding pocket independently, with the 2,6-linked moiety preferring to bind X-31 BHA and the 2,3-linked moiety preferring to bind X-31/HS BHA.

Reference:(1-5)

1. Sauter, N. K., Glick, G. D., Crowther, R. L., Park, S. J., Eisen, M. B., Skehel, J. J., Knowles, J. R., and Wiley, D. C. (1992) Crystallographic detection of a second ligand binding site in influenza virus hemagglutinin, Proc Natl Acad Sci U S A 89, 324-328.
2. Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S. J., Skehel, J. J., and Wiley, D. C. (1992) Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic

resonance spectroscopy and X-ray crystallography, Biochemistry 31, 9609-9621.
3. McCullough, C., Wang, M., Rong, L., and Caffrey, M. (2012) Characterization of influenza hemagglutinin interactions with receptor by NMR, PLoS One 7, e33958.
4. Eisen, M. B., Sabesan, S., Skehel, J. J., and Wiley, D. C. (1997) Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography, Virology 232, 19-31.
5. Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J., and Wiley, D. C. (1989) Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study, Biochemistry 28, 8388-8396.

***4. Structure refinement*: Suppose an X-ray crystal structure has been correctly determined, except that one amino acid's side chain has been completely mispositioned.**

a. Would you expect this error to affect the calculated structure amplitudes derived from this almost-correct structure: would a small number of structure amplitudes be completely wrong whereas the others are correct—or would a large number of structure amplitudes be slightly wrong? Explain briefly.

Answer: In this question after structure refinement, the final structure amplitudes won't be affected by this one amino acid's side chain error, the refinement process is optimizing the model parameters to fit the target function. However, the original structure amplitudes may be affected by the error. Small errors may always be refined by refinement methods. In my opinion, if the rest of the structure amplitudes are correct, small numbers of structure amplitudes errors will not affect the overall structure amplitudes.

b. How would you expect to identify this error during the structure refinement process?

Answer: There are several different methods; it might observe that the electron density map is not fitted in the building structure. Just like if Fo-Fc >0, it perhaps loss the atom in structure when the structure was building. On the other hand, Fo-Fc< 0 might build an extra atom in the structure. So, if one amino acid side chain has been mispositioned, we can rotate the side change or move the water molecular to fit the electron density map and refinement structure. Furthermore, take small amount of your refining structure amplitudes out of concern. Modify the model using most left structure amplitudes to fit the observed data. To sum, small percent of errors will not

exist in the model.

5. *Crystallographic structure determination*:

Suppose the structure of a particular protein (which we will call protein M) is already known,and you are interested in determining the structure of a protein (which we will call protein N) whose sequence is 58% identical to that of protein M. You successfully crystallize protein N. You can now proceed to determine the structure of protein N either by experimental phasing (using MIR or MAD), or by molecular replacement using protein M as a search model. In ten sentences or so, describe the advantages and disadvantages of those two approaches.

Answer: To use experiment phasing like MIR, at least two derivatives which each derivative gives two values, is needed. If peaks of the phase probability distribution which is drawn by the values, don't line up or overlap too broadly than more derivatives will be needed. However, running the computer program for molecule replacement may be time consuming depending on the unit cell size.

Moreover, advantage is that MAD needs only one sample while molecular replacement needs more. However, preparing and collecting data for better accuracy costs time and may cause radiation damage.

Data from molecular replacement is based on Patterson function to deal with the rotation and translation to fit the unknown protein structure, which protein sequences should be at least 20% of identity or the r.m.s.d. should be less than 2°A. For molecular replacement to work, the overall structure of the search model and the unknown model must be very close. Thus, disadvantage of MIR is the need to well acknowledge relative protein structure for reference phase otherwise the model bias, the inability to refine errors in a model, may appear.

***6. Nucleic acid structure*: What are 12 different 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?**

Answer: First of all, the six inter base pairs parameters shows local confirmation at every base pair step, which are rise, twist, shift, roll, tilt and slide. The other parameters which are describing the translational and rotational displacements that between the two bases of a base pair. It is describe the planar deviation within a base pair. The translational components are stagger, stretch and shear. The rotational components are propeller twist, buckle and opening.

The following table is more clearly to answer the question:

The inter base pairs parameters	rise, twist, shift, roll, tilt, slide
The translational components	stagger, stretch, shear
The rotational components	propeller twist, buckle, opening

7. Muscle diffraction: **Examine this diffraction image: The layer lines have been labeled in this diffraction pattern from relaxed insect flight muscle. They can be indexed as orders of a single long-spacing repeat. What is this repeat distance? Where does this come from? How might this diffraction pattern look different if it was from contracting muscle? Explain your answer. It may help you in answering this question to examine the paper describing this structure, which is available on the course Blackboard site.**

Answer: In this figure which the diffraction pattern arises from the periodic array of the myosin containing thick filaments and the actin containing thin filaments in the A-bands of the sarcomeres along with the accessory proteins. The fundamental long-spacing repeat distance is 232 nm, which means the layer lines will appear in this diffraction pattern at n/232 nm. Since the sarcomeres are connected end to end with the Z-band into the myofibrils, each myofibril represents a single crystallite. Thus from each myofibril the observed diffraction pattern consists the cylindrical average of the pattern. When contracting muscle, analysis of X-ray patterns from stretch-activated and skinned IFM showed that stretch activation was associated with a loss of intensity on the first row line spot, which associated between 1 and 0 that equatorial reflection, on the 38.7 nm layer lines. A gain of intensity on the first row line spot where upon the 19.3 nm layer line.

Reference*:* Thomas C. Irving. (2009) *X-Ray Diffraction of Indirect Flight Muscle from Drosophila in Vivo.ch19:1-15*

***8. Computational structural biology*:**

Summarize the types of computational tools that can be usefully employed at various time scales in characterizing biological structures. Explain how the results at the faster time scales help to inform the results derivable at the longer time scales.

Answer: There are five different level of time scale that used to analysis.

Minutes - Hours	Kinetic models
Seconds	Mechanical models
Milliseconds	Coarse-grained main Field models
Microseconds	Explicit solvent, Implicit solvent
Picoseconds	Quantum models (Monte Carlo, Post Hartree Fock, Hohenberg–Kohn: Density functional theory), XAFS

Following with the longer the time scale, the more the resolution decreases. Shorter time scale, such as picoseconds, result in highest resolution from all electrons and atoms. If only observing the α-helix movement, the atomistic model may be used to obtain the data and information of microseconds. Moreover, coarse-grained main field model may be applied to obtain milliseconds of lipid fusion. Through calibrating faster time scale models and combining the models with the experiment data, slower time models may be developed.

Reference:

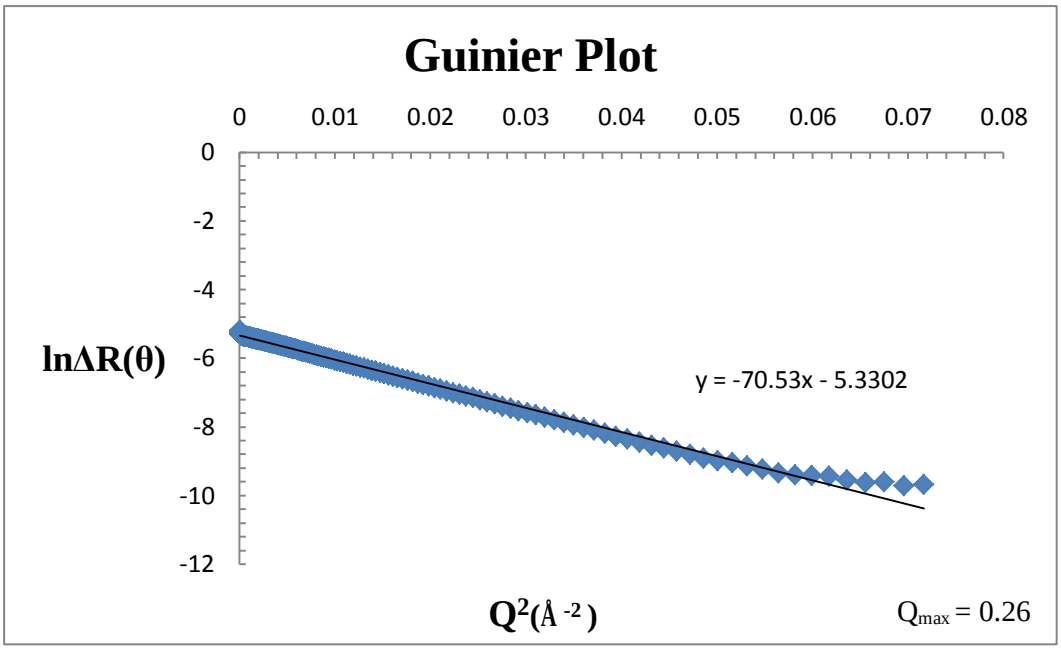
1. Johann Rohwer (2002)Concepts in Computational Systems Biology: Structural Analysis and Kinetic Modelling of Cellular Systems
2. Valentina Tozzini (2005) Coarse-grained models for proteins. 15:144–150
3. Peter L. Freddolino, Feng Liu, Martin Gruebele (2008)Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. 15; 94(10): L75–L77.
4. Eduardo M Sproviero, Jose´A Gasco´n, James P McEvoy, Gary W Brudvig and Victor S Batista(2007) Quantum mechanics/molecular mechanics structural models of the oxygen-evolving complex of photosystem II. 17:173–180

9. *X-ray Scattering*: We have uploaded a spreadsheet called "SAXS data file" to Blackboard. It has 3 columns of numbers. The first is *Q* ($4\pi\sin\theta/\lambda$) in Å-1. The second column is the scattering (intensity) from the buffer alone and the third is the scattering from the protein dissolved in buffer (buffer+protein). Taking these data and, using a Guinier plot, calculate Rg for this molecule.

Assuming a globular particle, calculate the radius of the particle. Assuming a prolate ellipsoid, calculate the principal axes *a* and *b*. In answering this question bear in mind that the Guinier approximation is only valid at low angles and will diverge at high angles. This should be apparent from the data. In answering this question bear in mind that the Guinier approximation is only valid at low angles

and will diverge at high angles. This should be apparent from the data. There is a "rule of thumb" for deciding the appropriate range of data for calculating Rg:Rg x *Q*max < 1.3 This implies that it might take a couple of iterations to get the range right. In your answer explain how you decided what range of *Q* to use to calculate Rg.

Answer: After utalize the Excel to calulate the table which form blackboard, the figure is showen as follow:

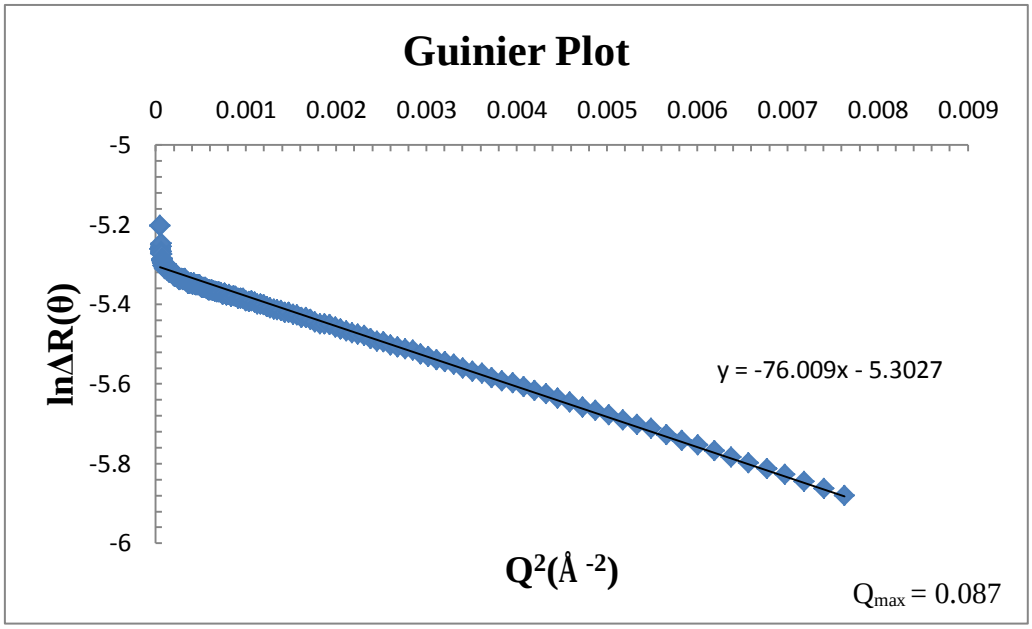


In Guinier equation, $I(Q) = I(Q)\left[1 - \frac{Rg^2Q^2}{3} + KQ^4 + \cdots\right]$,

$$\text{When } Q \to 0,\ I(Q) \cong I(0)\exp\left(\frac{-Rg^2Q^2}{3}\right)$$

After calculated the equation, the slope is -70.53 and the slope of Guinier plot is $R_g^2/3$, which means R_g is 14.54 Å when the Q_{max} is 0.26. However, the appropriate range of data which should follow "R_g x Q_{max} < 1.3", which means $QR_g < 1.3$ for globular, or $QR_g < 0.8$ for elongated.

Therefore, it should decide other range of Q_{max} which can fit the rule, and the new Q_{max} is 0.087, and New R_g^2 is approximate 15.10



Second, assuming a globular particle and calculate the radius of the particle. For a sphere of radius R it can be shown that $R_g = \sqrt{\frac{3}{5}}R$. Base on above, it could know the R is 19.48 Å after calculated.

Third, assuming a prolate ellipsoid, the equation is $Rg^2 = \frac{1}{5}(a^2 + b^2 + c^2)$, where Rg for an ellipsoid of it, c = b, and normalizing by Rg for the sphere of equal volume is $Rg^2 = \frac{1}{3}\left[\left(\frac{a}{b}\right)^{3/4} + 2\left(\frac{a}{b}\right)^{2/3}\right]$. After calculated, the a is 0.003 and b is 22.71.

Reference:

E.T. Whittaker. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944), Chapter 5.