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 (eix = cosx + isinx) to show that the absolute value of eix is one,provided that x is real.

Since $|z| \equiv \sqrt{(zz^*)}$, eix = cosx + isinx

$$
\begin{aligned}
| eix | &= \sqrt{(eix\ eix\ ^*)} \\
&= \sqrt{[(\cos x + i\sin x)(\cos x - i\sin x)]} \\
&= \sqrt{[\cos^2 x + i\sin x\cos x - i\sin x\cos x - (i\sin x)^2]} \\
&= \sqrt{(\cos^2 x + \sin^2 x)} \\
&= 1
\end{aligned}
$$

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

$e^{iz} = e^{i\,(a + ib)} = e^{(ia - b)} = e^{ia}\, e^{-b}$

$|e^{iz}| = |e^{ia}\, e^{-b}| = |e^{ia}|\, e^{-b}$ ($e^{-b} > 0$)
According to results from last question: $|e^{ia}| = 1$
So: $|e^{iz}| = e^{-b}$

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?

It is an 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?

In Z direction there is a 2-fold axis, there are two two-fold screw axes on X and Y direction relatively.

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.

Since there are four equivalent positions (x, y, z); (-x, -y, z); (-x + ½, y + ½, -z); (x + ½, -y + ½, -z).

(x, y, z)- (-x, -y, z) = (2x, 2y, 0)
(x, y, z)- (-x + ½, y + ½, -z) = (2x-1/2, -1/2, 2z)= (2x-1/2, 1/2, 2z)
(x, y, z)-(x + ½, -y + ½, -z) = (-1/2, 2y-1/2, 2z) = (1/2, 2y-1/2, 2z)
(-x,-y,z) - (-x+½, y+ ½, -z) = (-1/2, -2y-1/2, 2z) = (1/2, -2y-1/2, 2z)
(-x,-y,z) - (x+½, -y+½, -z) = (-2x-1/2, -1/2, 2z) = (-2x-1/2, 1/2, 2z)
(-x+1/2, y+ ½, -z) - (x+½, -y+½, -z) = (-2x, 2y, 0)

So: when u=1/2, Harker section: (1/2, 2y-1/2, 2z) or (1/2, -2y-1/2, 2z)
when v=1/2, Harker section: (2x-1/2, 1/2, 2z) or (-2x-1/2, 1/2, 2z)
when w=0, , Harker section: (2x, 2y, 0) or (-2x, 2y, 0)

3. Multiple methods: Search the Protein Data Bank (www.rcsb.org) for a protein whose structure hasbeen determined by two different methods. The two methods probably would be NMR and X-ray crystallography, but they could instead be NMR and cryoelectron microscopy, or X-ray and neutron crystallography, or some other permutation. 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.

X ray crystallography and small angle X ray scatterring structure of myomesin filament dimer

The X ray crystallography result reveals a tail-to-tail filament structure with My9, My10, My11, My12 forming two filaments symmetrically form a superhelical coil. In the center, two myomesin filaments' my13 dimerized in a zigzag-type arrangement forming module $(My13)_2$. The particle is a right-handed superhelix with twist angles of neighboring My domains of 26-27^o except for My12-My13 which has a average twist angle of 68^o. Between every two Ig domains there is a-helical linker which has the same orientation with the proceding Ig domain.

The SAXS result of myomesin filament shows us the general overall shape of superhelical coil and it's dimeric. At the same time it suggests the particle can be discribed as a 370 A^o long cylinder. Furthermore, the distance distribution p(r) displays a series of maxima at distances of 60 A °, 118 A °, and 165 A ° which is characteristic of extended particles with periodic domain arrangements. At last, a reduced peak sharpness appears in p(r) suggests a limited flexibility in some Ig domains which implicates the existence of Ig domain tandems.

In summary, the X ray crystallography results give us a high resolution structure of the particle with numerous details such as main components, twist angles and orientations. The structure is without doubt helpful to understand myomesin filament dimer's function and mechanism. The SAXS results, on the other hand, can't present us a full picture of the particle, however we can still get information such as volume and a low resolution shape to verify X ray crystallography results. Most importantly, it generates distance distribution function which quantitatively confirms the structure results from X ray crystallography. There are some other values, like molecule weight and radius of gyration,

can be generated from SAXS, though not used in this experiment.

Reference:
1. Superhelical Architecture of the Myosin Filament-Linking Protein Myomesin with Unusual Elastic Properties. Nikos Pinotsis, Spyros D. Chatziefthimiou, et al. PLoS Biol. 2012 February; 10(2): e1001261. Published online 2012 February 14. doi: 10.1371/journal.pbio.1001261

2. http://www.rcsb.org/pdb/explore/explore.do?structureId=2Y25

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.

Misposition of an amino acid side chain could affect the calculated structure amplitudes but I would just expect a small number of structure amplitudes are wrong. Since structure amplitudes are the summation of numerous individual parts, the error caused by a side chain misposition might be negligible**.**

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

First we can use physical principals to identify errors. For example, if the electron density at any points is negative, something must be wrong there. We can also use solvent flattening to identify errors: presence of big peaks or valleys in the solvent region must be due to errors. Non-crystallographic symmetry averaging: in asymmetric unit, differences in electron density at the corresponding places must be due to errors.
Second, check chemical realities of the model such as bond lengths and angles, torsion angles of both main-chain and side-chain (including Ramachandran angles and rotamers), hydrogen bond lengths and angles and Van der Waals interactions. Sometime disulfide bond can easily be mistaken as skeleton.

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.

Molecular replacement is effective for solving macromolecular crystal structures based upon the knowledge of homologous structures. It is time and effort saving with simplified model building process. However, this method requires the 3-dimensional structure of the search model must be very close to that of the unknown structure. Since only the sequence resemblance between N and M is provided, it is uncertain that Molecular replacement is feasible. Also the risk of model bias of Molecular replacement is greater than experimental phasing. Experimental phasing such as MIR might be the most common approach of solving the phase problem in X-ray crystallography. We can solve a structure with phases derived from experiments and the result we get from experimental phasing is statistically satisfying. But it requires a lot of experiments, some specialized facilities and certain acknowledge in software. The lack of isomorphism might also be a problem.

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?

Rise, twist, shift, roll, tilt, slide, stagger, stretch, shear, 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.

The repeat distance is 232 nm. Because these layer lines are found at an axial position of n/232 nm where n is an integer. Such as 1/38.7 nm, (n= 6), 1/23.2 nm (n=10), 1/19.3 nm (n=12), 1/14.5 nm (n=16) and 1/7.25 nm (n= 32).

If it is from contracting muscle, the first row layer line position will change, suggesting restricted binding of active myosin heads to each actin target zone forming troponin bridge. More than 30% of all myosin heads will move away from their axially ordered positions on the myosin helix. There would be an increasing intensity of tropomyosin reflection due to the conformational shift of tropomyosin.

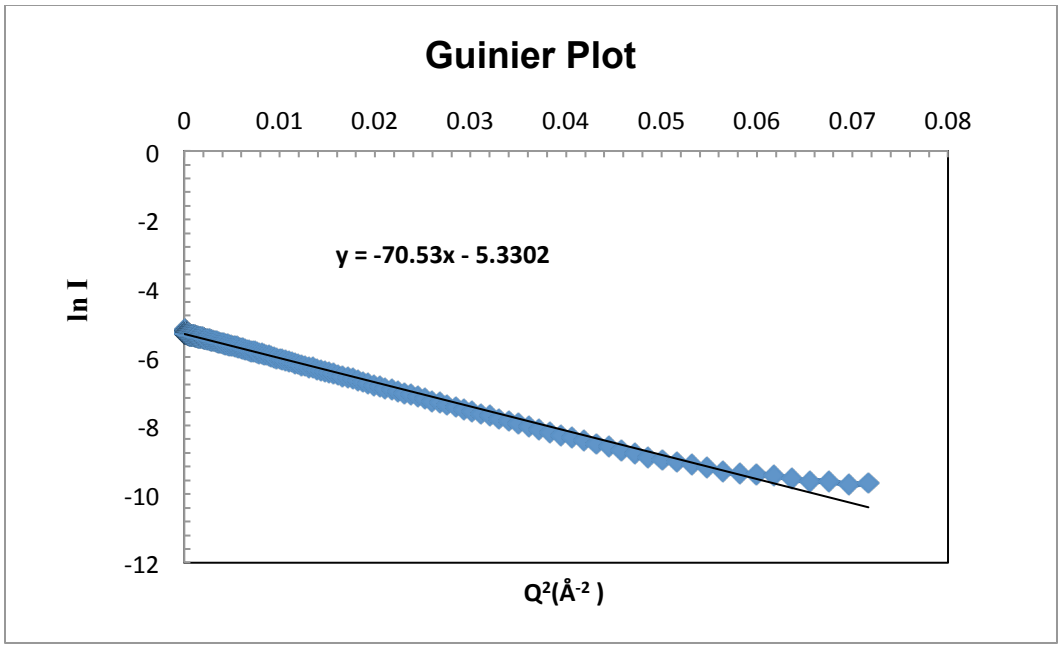
8. Computational structural biology:
Summarize the types of computational tools that can be usefully employed at various time scales in characterizing biological structures. Explain how the result at the faster time scales help to inform the results derivable at the longer time scales.

Dynamic molecule stimulation is one of the most important tools to study biological structures at various time scales.
Quantum models (picoseconds): Quantum Monte Carlo, Post Hartree-Fock, Kohn-Sham DFT
Atomistic models (microseconds): Explicit solvent, Implicit solvent
Coarse-grained/mean-field model (milliseconds)
Mechanic models (seconds)
Kinetic models (hours)

To relate microscopic information to macroscopic information requires Statistical Mechanics. Statistical mechanics provides a framework for relating the microscopic properties of individual atoms and molecules in very small time scale to the macroscopic bulk properties of materials in long time scales.
We can measure the property of micro individuals as samples over a large amount then use the weighted average to predict the property of this ensemble. Or we can run the small time scale method long enough and predict bulk properties by averaging the simulation at every time point.

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

Guinier approximation:
$\ln I = \frac{-Rg^2}{3} Q^2 + \ln I_0$ (Q small enough)
Plot ln I vs. Q^2 :



Equation generated by Excel: y = -70.53x - 5.3302

According to Guinier approximation, slope = $-\frac{Rg2}{3}$ = - 70.53.

So: Rg= 14.55 Å, since the approximation is ture only when $Q_{max}R_g < 1$, Qmax<0.069

For sphere: Rg= = $\sqrt{\frac{3}{5}}$R, so R=18.77 Å

The limit for sphere is $Q_{max}R_g < 1.3$, Qmax<0.089

For prolate ellipsoid: $Rg^2 = (a^2+b^2+c^2)/5$, a=b<c.

$3a^2/5<(a^2+b^2+c^2)/5= Rg^2=211.7$

a <18.77