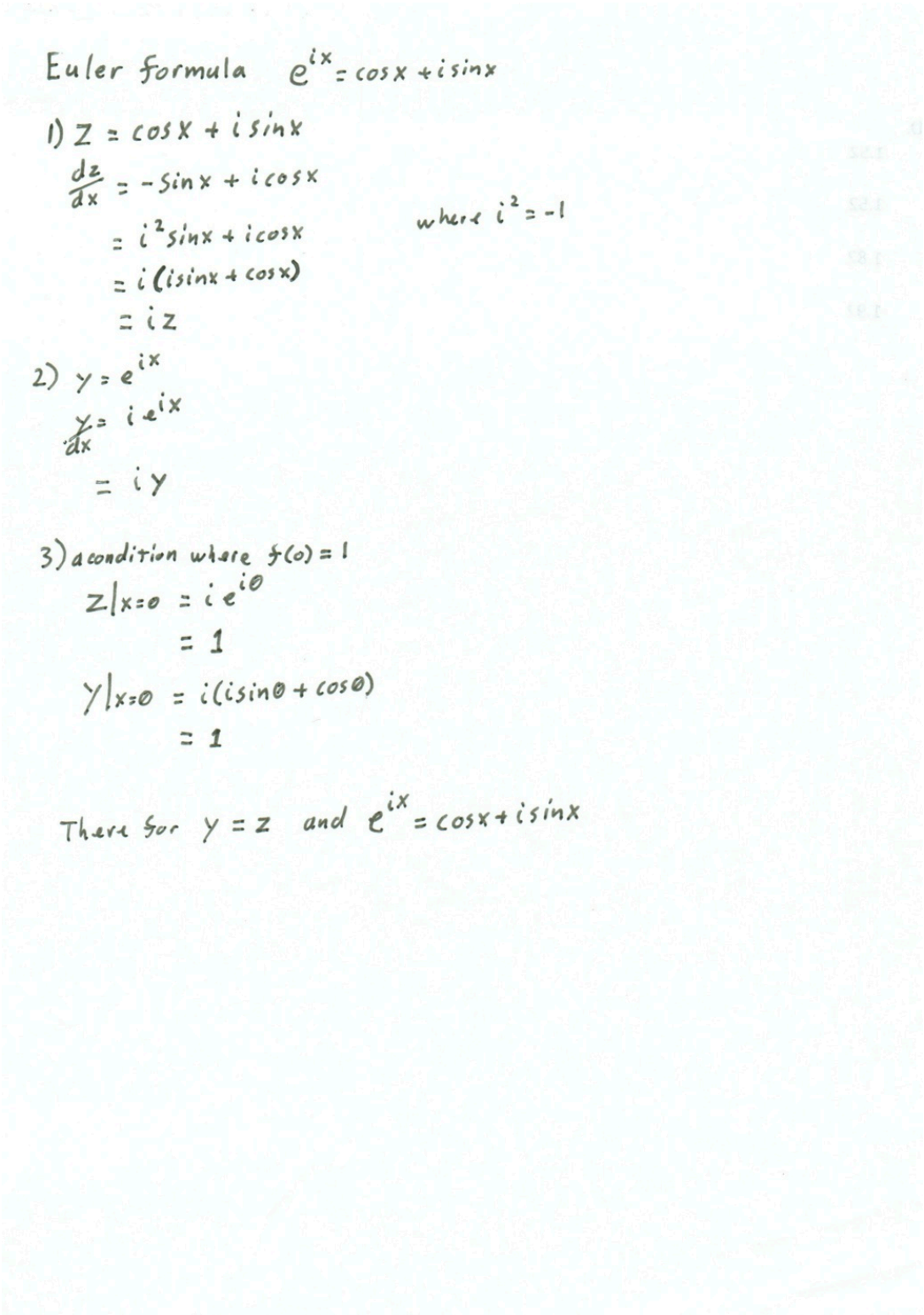
1. ***Structure math*:**

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.



Euler formula $e^{ix} = \cos x + i\sin x$

1) $Z = \cos x + i\sin x$

$\frac{dz}{dx} = -\sin x + i\cos x$

$= i^2 \sin x + i\cos x$ where $i^2 = -1$

$= i(i\sin x + \cos x)$

$= iz$

2) $y = e^{ix}$

$\frac{y}{dx} = ie^{ix}$

$= iy$

3) a condition where $f(0) = 1$

$z|_{x=0} = ie^{i0}$

$= 1$

$y|_{x=0} = i(i\sin 0 + \cos 0)$

$= 1$

There for $y = z$ and $e^{ix} = \cos x + i\sin x$

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

$|e^{iz}| = |e^{i(a+ib)}| = |e^{(ia+i^2b)}| = |e^{ia} * e^{-b}|$
a and b are real numbers
$|e^{ia}| = 1 * |e^{-b}| = e^{-b}$
Therefore $|e^{iz}| = e^{-b}$

2. *Crystallography*: Space group P21212, International Tables number 18, has the following equivalent positions: (x, y, z) ; $(-x, -y, z)$; $(-x + 1/2, y + 1/2, -z)$; $(x + 1/2, -y + 1/2, -z)$.

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

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?

The 2-fold axes are all perpendicular to each other in z.
The 2-fold screw axis is parallel to the x and y

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. If your knowledge of Patterson syntheses is incomplete, you may find it useful to consult a document I've made available: http://csrri.iit.edu/~howard/harker.html.

Strong peaks are generated from heavy atoms and are located where y=1/2
Part one: (x',y',z') = (x,y,-z)
(u,v,w) = (x-x,y-y,z-(-z)) = (0,0,2z)
Part two: (x',y',z') = (x-½,-y-½,z)
(u,v,w) = (x-x', y-y' , z-z') = (x-(x-½),y-(-y-½),z-z) = (½,2y+½,0)
Part three: (x',y',z') = (-x-1/2,-y-1/2,z)
(u,v,w) = (x-x',y-y',z-z') = (x-(x-½),y-(y-½),z-z) = (2x+½, ½,0)
The interesting peaks are where y=½ and that is (2x+1/2,1/2, 0)

3. ***Multiple methods*:** Search the Protein Data Bank (www.rcsb.org) for a protein whose structure has been 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 similarities and differences between the two structures. You

should reference published papers about these structures as well as the PDB submissions themselves. Cite your sources.

Protein: 2FA4 Crystal Structure of Oxidized Form from Saccharomyces cerevisiae
Protein: 2HSY Solution structure of Thioredoxin 2 from Saccharomyces cerevisiae

The X-ray and NMR determination of thioredoxin 2 in *Saccharomyces cervisiae* both confirmed the presence of 5 beta sheets surrounded by four alpha helices. Both methods also revealed the presence of active site cystines and C31 and C34.

The x-ray determination showed that the oxidized form of the active side chain of the Cys31 and Cys34 form a disulfide bond. It also revealed a highly conserved amino acid sequence of Asp25, Trp30, Lys35, Ala28 and Lys56 that contribute to low pKa in Cys31 and a High pKa in Cys35.

The NMR determination indicated that there was an active site at W30 to C34. The NMR determination also indicated active site loops between residues 28 to 35, 57 to 62, and 68 to 74 and the flexibility were found to be at K35, S62 and S71. Conformational diversity was also determined at W30 which contained indolic resonance and indicated the presence of conformational diversity for the loop as well as residues V61 and S62 indicated conformational diversity.

Over all both methods confirmed alpha helices and beta sheets, as well as the active sites.In all the NMR determination revealed more detailed information about the structure and functional conformations of the protein.

Reference:
Bao R, Chen Y, Tang Y, Janin J, and Zhou C. **Crystal Structure of the Yeast Cytoplasmic Thioredoxin Trx2.** Wiley interscience: Proteins: Structure, Function, and Bioinformatics 66:246-249 (2007).

Amorim G, Pinheiro A, Netto L, Valente A, Almeida F. **NMR solution structure of the reduced form of thioredoxin 2 from *Saccharomyces cervisiae***. 6 March 2007, J Boimol NMR (2007) 38:99-104 DOI 10.1007/s10858-007-9144-z

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: **Yes**
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.
The mispositioned side chain would cause a small number of the structures amplitudes to be wrong. This would be due to the mispositioned side chain would change hydrogen bonding, bond angles, bond length and Van der Waals interactions, causing a shift in its peaks.

b. How would you expect to identify this error during the structure refinement process?
The location of the mispositioned amino acid side chain would have a positive or negative density peak in its vicinity (Shift Peak). These values can be checked if the structure has similar geometries and that all bond lengths, angles… match the literature values.

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.

- **Mad Advantages:**
  - **Mad only needs a single crystal (Se-Met MAD) and has no lack of isomorphism.**
  - **No model bias**
  - **Natural state of protein**
- **MAD Disadvantages:**
  - **Depends on accurately measuring small differences in intensities, requiring the need to collect more data to measure more accurate reflections, leading to radiation damage.**
  - **Requires synchrotron to tune wavelengths of x-rays.**
  - **Requires heavy atoms to get a signal**
- **MIR Advantages:**
  - **Data collection can be done at home**
  - **No model bias**
- **MIR Disadvantages:**
  - **Requires lots of crystals and is time consuming**
  - **After soaking crystals most crystals do not defract well, and are cracked or lack isomorphism in unit cell or spacing.**
  - **Metal salts used are hazardous**
- **Molecular replacement Advantages:**
  - **Six dimensional problem**
  - **Uses already determined protein structures from the PDB**
- **Molecular Replacement Disadvantages:**
  - **Time consuming**
  - **Places solutions that are far from the origin**
  - **Trial and error**
  - **A lot of computation**

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?

- **Inter base pair**
  - **Rise (translational z axis)**
  - **Twist (Rotation about z axis)**
  - **Shift (orthogonal in helical axis)**
  - **Roll (± cleft between stacked base pairs open toward the Major or minor groove)**
  - **Tilt**
  - **Slide**
  - **Stagger**
  - **Stretch**
  - **Shear**
  - **Buckle**
  - **Propeller twist (15˚difference)**
  - **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? **The repeating distance is 38.7nm and caused by cross bridges moving axially.**
Where does this come from? **The repeat comes cross bridges moving axially and from small changes in filament length and stiffness creating elastic strain energy.**
How might this diffraction pattern look different if it was from contracting muscle? **The reflections would be brighter. This is induced from the stretch induced force between myosin and troponin coupling by cross bridges. This causes thin filament elongation, myosin tilting and troponin structural changes, resulting in strong x-ray defraction.**

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.

Molecular dynamics, Brownian dynamics, Monte Carlo, Langevin dynamics, Energy minimization.
The faster time scales can be implemented for gathering potential values for the longer time scales. If the results of the longer time scale are far from the results of the faster time scales then the results should be disposed.

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 R_g 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 R_g: $R_g \times Q_{max} < 1.3$
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 R_g.
From the data collected a linear trend line was created in excel.

Qmin: 0.00689

Qmax: 0.2677

Θ min: 0.00689

Θ max 0.0089

Trend line from excel: y=-70.53x-53302

$R_g^2/3 = 70.53$

Calculated Rg: Rg = 14.55

Rg x Qmax = < 1.3

Rg= Sqrt (⅖) *r

r = 18.78

Ellipsoid

$R_g^2 = (a^2+b^2+c^2)/5$

b=c

$a = \text{Sqrt}(1058-2b^2)$

$b = \text{Sqrt}[(1058-a^2)/2]$