PX Question 1. (8 points)
Protein crystallization and crystal freezing.

a. Describe the hanging drop method used for crystallization. Include in your answer how the droplet containing the protein changes over time to make it crystallize. (4 points)

b. For data collection a crystal is frozen to 100 K. Which effect is minimized at 100 K such that more data can be collected from 1 crystal? (2 points)

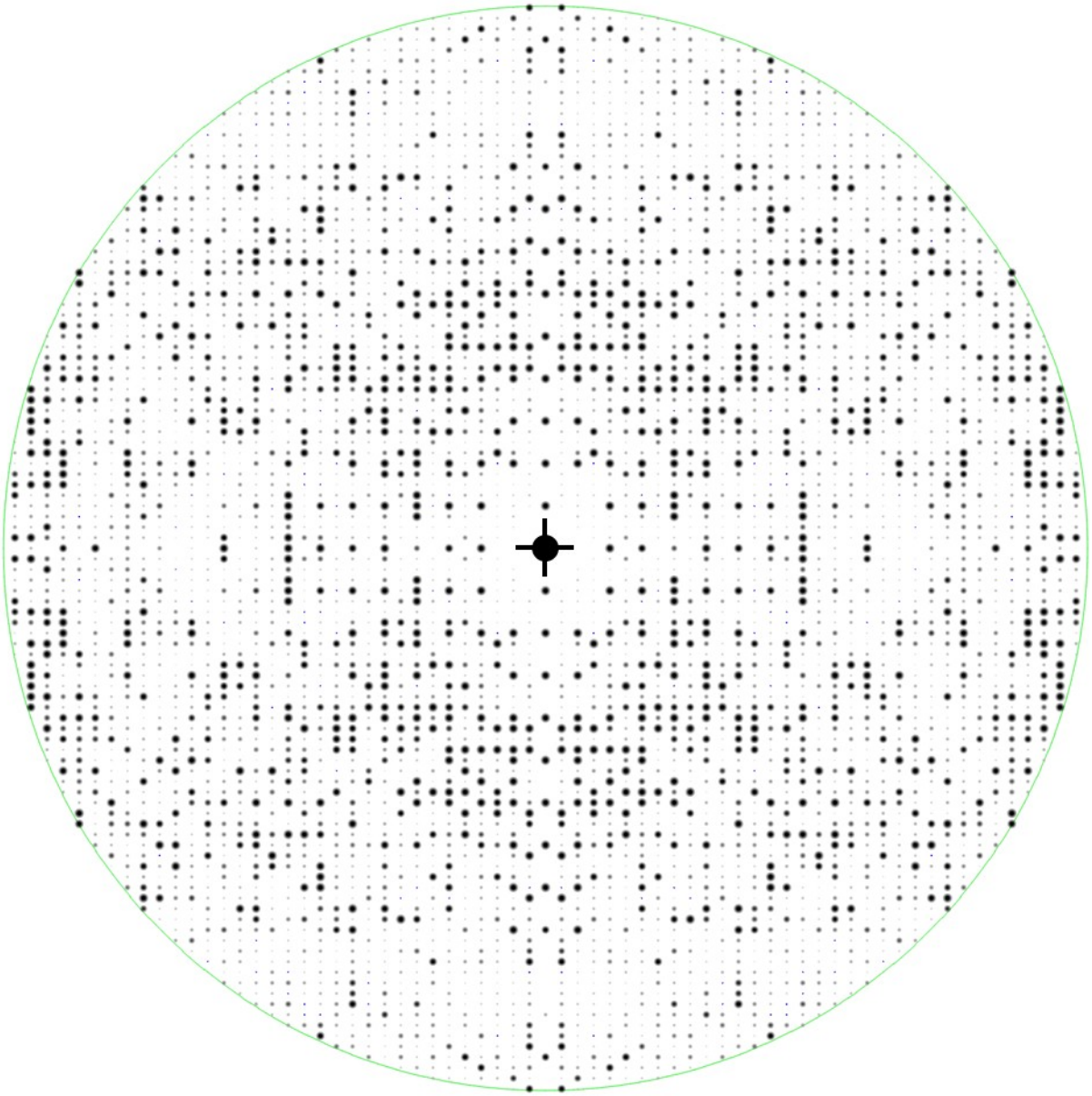
c. Often 20 to 25% v/v glycerol is added to a protein crystal before it is frozen to 100 K. Why is glycerol added? (2 points)

PX Question 2. (13 points)
Protein-crystal diffraction data collection and processing.

a. Describe the principles of how a diffraction data set is collected? Make a drawing, in which you indicate the X-ray source, position of the crystal, beam-stop and detector; indicate the path the X-rays travel; and, indicate what you do with the crystal to obtain a nearly complete data set. (4 points)

b. Indicate one property of a crystal that will determine the total number of unique reflections in a 100% complete dataset (2 points).

In the diffraction data processing of a protein crystal, you observe the following pattern for the *(h0l)* data. The central cross indicates the 0 0 0:



c. Which of the following spacegroups is the most likely for this pattern: *P*1, *P*2, *P*222, $P2_12_12$, *P*422, $P4_12_12$, $P6_1$ or $P6_12_12$? Explain why. (3 points)

In a native dataset with P1 symmetry the following intensities of five reflections have been measured.

Intensity (*h*,*k*,*l*)

33 (2, 1 , -3)
41 (-2, 1, -3)
29 (2, -1, 3)
17 (2, 1, -3)
37 (-2, -1, 3)

The following equation is given.

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_{i}^{n_{hkl}} | I_{hkl}^{i} - \langle I_{hkl} \rangle |}{\sum_{hkl} \sum_{i}^{n_{hkl}} I_{hkl}^{i}}$$

d. Calculate the R_{merge} for this dataset. (4 points)

PX Question 3. (14 points)
Experimental phasing.

a. Which information is required to calculate an electron density map. (2 points)

b. In the case of space group $P2_12_12_1$, explain why the three planes, also called Harker sections, at $u = ½$, $v = ½$ and $w = ½$ are of particular interest, when interpreting the isomorphous difference Patterson map. (2 points)

c. Assume a derivative dataset was collected from a crystal with a spacegroup that has a two-fold screw axis along y. The Harker section at $v = ½$ of this dataset shows a peak at u=0.3 and w=0.2. Determine the fractional x and z coordinates for the heavy atom. (2 points)

d. Assume you have a native protein data set (FP) and two different derivative data sets (FPH1 and FPH2). How can you derive the positions of the heavy atoms? How can you deduce phase information for the structure factors of the native protein data set FP? Also use a Harker construction to make your point clear. (5 points)

e. Describe the molecular replacement method. How is this method split in two steps? Which information is used in each of these steps? (3 points)

Question 4. (10 points)
Refinement of protein-crystal structures.

Two monoclinic datasets have been collected on two crystals of the same protein called C2a. One of the differences between the datasets is the presence of a Li^+ ion versus a Mn^{2+} ion. Two protein structures have been refined using these datasets. The statistics of the datasets and the structures are given in the table below.

	Data collection statistics	
	C2a-Li^+	C2a-Mn^{2+}
Wavelength Å	0.9330	0.9330
Space group	$P2_1$	$P2_1$
Cell dimensions (Å,°)	a=51.16	a=51.48
	b=75.08	b=77.19
	c=71.56	c=70.86
	β=110.32	β=109.55
Resolution (Å)	37.5-2.7	50.5-2.3
Completeness (%)	98.13	89.62
Multiplicity	3.6 (3.1)	3.6 (2.9)
R_{merge}	0.085 (0.42)	0.061 (0.65)
I/σI	13.7 (1.8)	15.11 (1.5)
	Refinement statistics	
# unique reflections	13177	26081
R_{work}/R_{free} (%)	21.2/28.7	18.9/24.0
Overall B-factor ($Å^2$)	30.7	22.1
RMSD bond length (Å)	0.004	0.003
RMSD bond angle (°)	0.674	0.862
#protein atoms	3993	4223
#water molecules	34	122

You can use this table to explain the following questions.

a. What is approximately *(d.w.z. bij benadering oftewel ongeveer)* the diffraction data-to-parameter ratio for the C2a-Li^+ structure? Assume one isotropic B-factor parameter per atom? (2 points)

b. Explain why additional information is needed to refine (or optimize) this protein structure? Give an example of such a type of additional information used in protein crystal-structure refinement. (2 points)

c. In both cases, one protein molecule occupies the asymmetric unit. Still, the number of protein atoms in the models is not identical in the two cases. Give a plausible explanation why the number of atoms may not be identical. (2 points)

d. What is the purpose of R_{free}? (2 points)

e. Which of the two data sets has provided a more accurate structure? Explain your choice. (2 points)