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)

See figure 2.3 and 2.4. During equilibration the drop containing the protein will shrink. The protein concentration will double in time. (If crystals appear the concentration of protein in solution will go down.)

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)

Op cryogene Temperatuur worden radicalen afgeremd
Vermindert stralingsschade

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)

Voorkomen van ijsvorming

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

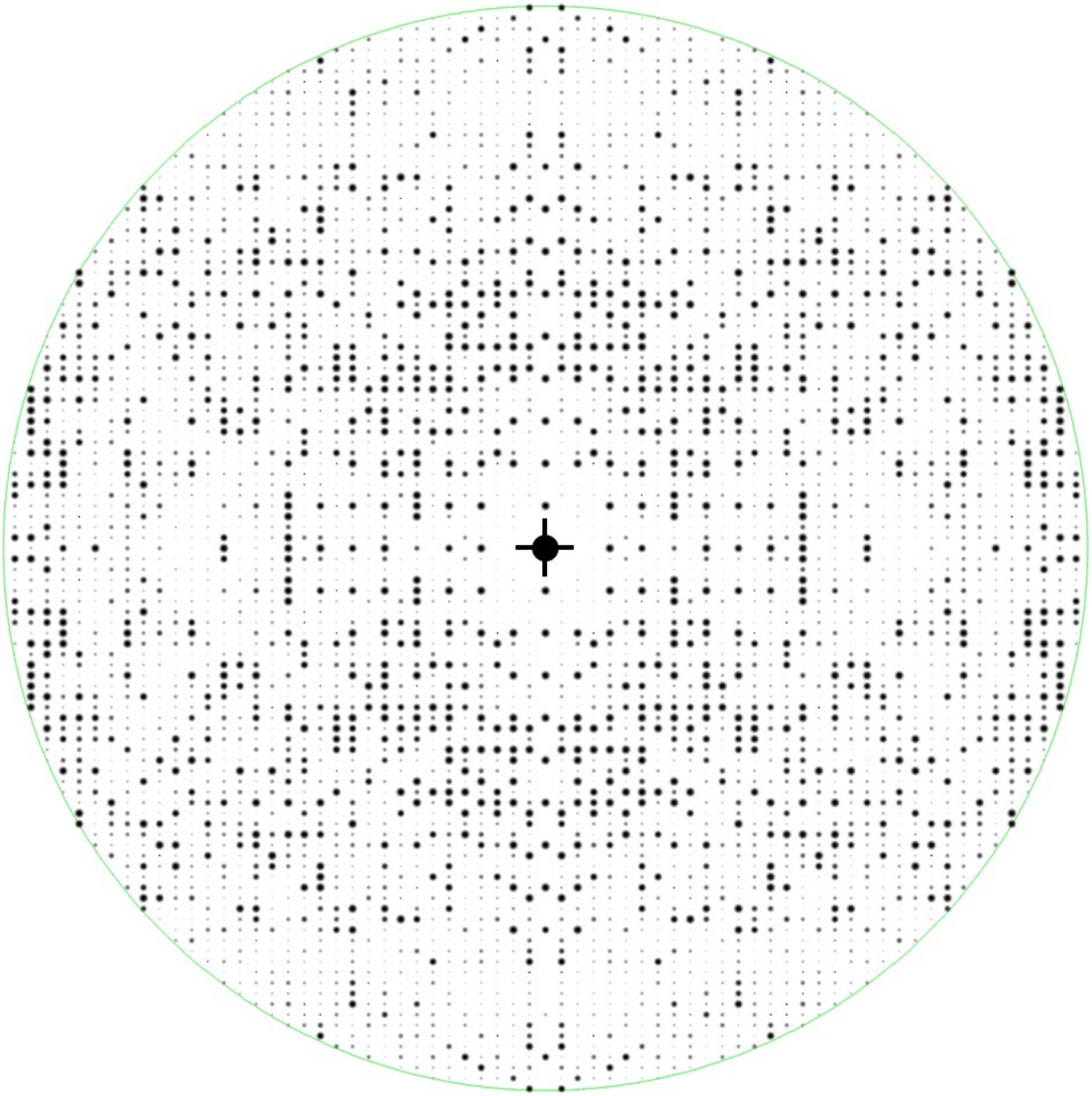
a. Describe the principles of how a diffraction data set is collected? Make a drawing, in which you indicate the X-ray source, position of the crystal, beam-stop and detector; indicate the path the X-rays travel; and, indicate what you do with the crystal to obtain a nearly complete data set. (4 points)

Source – xtal- beamstop-detector; vertrooiing/afbuiging
180 graden draaien

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

1. The resolution or the resolution range, in particular higher resolution gives more reflections. (The quality of a crystal determines the resolution.)
2. The unit cell dimensions: larger unit cell gives more reflections.
3. Symmetry of the crystal: centering gives less reflections, as do screw axes (although only a few less)

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)

$P4_12_12$ (one 4-fold screw), one two fold screw.
*P*222 (incorrect) but some points (1.5) if reasoning is the two fold symmetry in the pattern.

d. 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.

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

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

Twee sets equivalente reflecties a en b met intensiteiten a 33, 17 37 en b 41, 29.
Gemiddelde a 29 en gemiddelde b 35.

Sommatie van gemeten waarde minus gemiddelde over a (|33-29|)+(|17-29|)+(|37-29|) = 4+12+8=24
Sommatie van gemeten waarde minus gemiddelde over b (|41-35|)+(|29-35|)=6+6=12

Sommatie van afwijkingen van de twee sets 24+12=36
Sommatie van de intensiteiten van alle reflecties 157
Rmerge is 36/157 = 0.229

(Niet realiseren dat er twee sets reflecties zijn: gemiddelde 157/5 = 31.4
Sommatie van gemeten waarde minus gemiddelde 1.6+14.4+5.6+9.6+2.4 = 33.6
Rmerge is 33.6/157 = 0.214) [3 punten]

PX Question 3. (14 points)
Experimental phasing.

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

For each *hkl* the complex structure factor is required (amplitude information and phase information).

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)

Due to the symmetry present in the space group, we expect to see peaks between symmetry-related heavy-atom sites at the Harker sections.

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)

x = 0.15, z = 0.1 (or x = -0.15, z = -0.1)

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)

Positions of heavy atoms determined from patterson map. The peaks in the Patterson map indicate interatomic vectors coming from the heavy atoms. When we have obtained the positions of the heavy atoms we can calculate the complex structure factors (for each hkl) for the heavy atom contribution with the formula for the atomic structure factor. Make drawing of circles in the Harker construction to indicate how you get phase information for the native protein dataset.

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)

Make use of structure from a homologous (or similar) protein. Calculate a Patterson map from the diffraction data and from the structure. First step; use self-vectors in Patterson search to find rotation of the molecule. Second step; use cross vectors to find location of the molecule using the correct rotation.

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)	30.72	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)

13177 reflections, 4027 atoms in total (4*4027=16108 parameters), diffraction data-to-parameter ratio 0.8180

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)

Geometrische restraints zijn nodig, omdat we niet genoeg gemeten datapunten hebben
Bond restraints; angle-restraints

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)

This is most likely due to differences in disorder of parts of the protein between the two dataset. If there is too much disorder, for example in a loop, that part is left out of the model as it cannot reliably be refined.

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

It is the cross-validate the refinement. It indicates how well the model predicts the data.

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

The model refined using the highest resolution reflections, this dataset provides more information at higher resolution, thus C2a-Mn^{2+}.