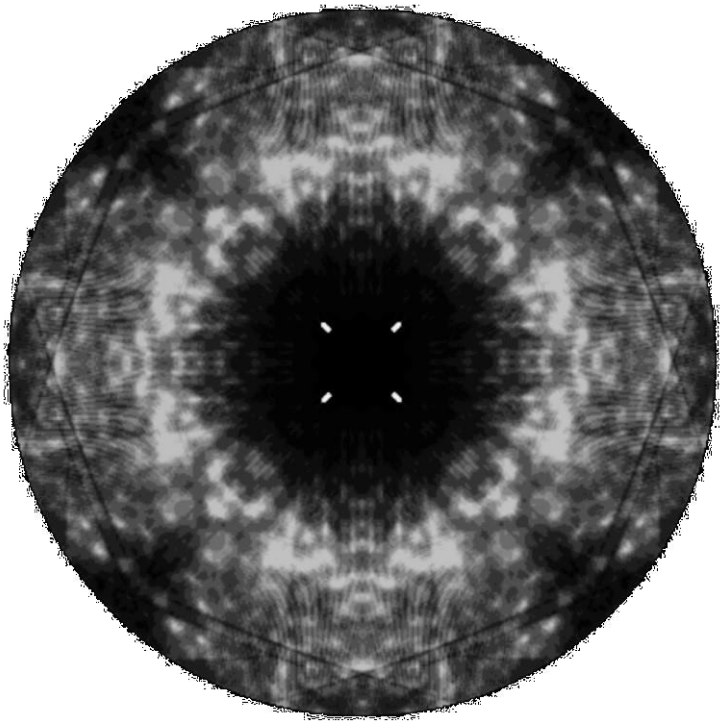
Obtaining diffuse scattering signal from X-ray diffraction experiments on lysozyme

A. Savenije

December 20, 2021



Abstract

For decades, X-ray diffraction has been used to determine protein structures in crystals to atomic level. Nevertheless, while performing X-ray diffraction experiments, the information regarding the dynamics of the protein molecule has been largely ignored. This information is contained in the so called diffuse scattering signal. However, obtaining this diffuse scattering signal remains challenging because of its low intensity and strong overlap with the signal originating from the crystal solvent. In this research we describe how to make suitable protein crystals and how to perform X-ray diffraction experiments to maximize the diffuse scattering signal. Furthermore, we describe methods to mask intense features such as Bragg reflections and ice scattering. Then we demonstrate the mapping and scaling of experimental images to the 3-dimensional reciprocal space. These experimental 3-dimensional reciprocal space maps can be compared to reciprocal space maps of models that represent the diffuse scattering signal. We observe that using method III, sharp features such as Bragg reflections can be removed as good as performing a modefilter. Next, if Any is not able to calculate scalefactors without fluctuations, the mean can be used after strong intensities that do not occur at least 1% as the most frequent occurring intensity have been removed. Using this approach the experimental reciprocal space maps show similar features compared to the model maps concluding that this newly described method can be used to isolate the diffuse scattering signal. Nevertheless, many features in the model map cannot be observed in the experimental map. We suggest that this is rather due to the low photoncount obtained during X-ray diffraction experiments than experimental data processing and asks for a different approach in performing X-ray diffraction experiments.

Layman's summary

X-ray diffraction crystallography is a technique used to determine molecular structures in crystals. The diffraction signal contains information on the atomic positions of the molecule of interest. Additionally, it contains valuable information on molecular dynamics in the form of diffuse scattering. However, this signal is usually discarded with the background. In this research we show methods to increase the contribution of this diffuse scattering and to isolate it from the other signals. Next, to explain the diffuse scattering as dynamics of the molecule of interest we compare this with theoretical models. In this comparison, similar features can be observed between the experimental and model reciprocal space maps. Therefore we concluded that our newly described method can be used to extract the diffuse scattering information. Although, various features in the model cannot be observed in the experimental map. We suggest that is because of a too low intensity obtained during the X-ray diffraction experiments.

1 introduction

Theoretical background

X-ray diffraction on crystal structures is a technique that has been around for decades and can be used to determine protein structures to atomic detail. The strength of X-ray diffraction experiments lies in the ordered arrangement of protein molecules on a crystalline lattice, resulting in intense reflected beams with well-defined directions: the Bragg reflections (Figure 1 A and C). Normally, solely these Bragg reflections are used to determine the protein structure. Although this structural arrangement is essential, there are two effects that induce disorder in the crystal. First, every single atom in the crystal has thermal motion and therefore vibrates around an average position. The higher this thermal motion of the atoms the more disorder is induced in the crystal. Secondly, not every individual protein molecule is at exactly the same position on the crystal lattice or has similar orientation (Figure 1 B).

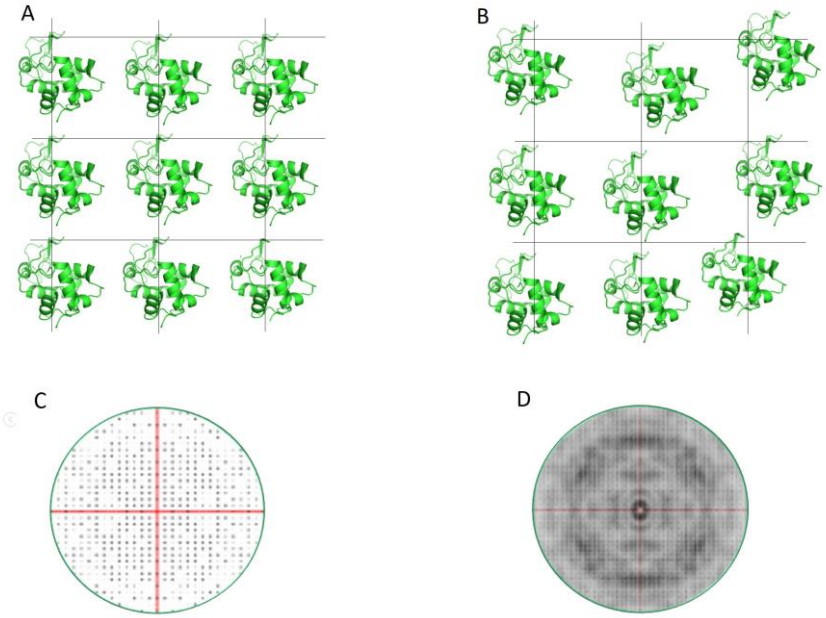


Figure 1: Lysozyme molecules on a crystal lattice and signal observed in reciprocal space. A) Lysozyme molecules perfectly arranged on a crystal lattice. B) Lysozyme molecules disorderly structured on a crystal lattice. C) Bragg reflections observed in reciprocal space as sharp intense features when the lysozyme crystal is orderly structured as in A. D) Bragg reflections and diffuse scattering observed in reciprocal space when the lysozyme crystal is disordered as in B.

Both these two phenomena contribute to disorder in the crystal. The larger the amount of disorder, the weaker the Bragg reflection intensity. This intensity loss is scattered into directions in between the Bragg peaks, resulting in increased background scattering. This is called diffuse scattering (Figure 1 D). An example of diffuse scattering in the form of a cloudy pattern is shown in Figure 2 B. As

mentioned before, usually this diffuse scattering signal remains unused, even though research has shown that beyond 3.8 Å resolution the diffuse scattering signal exceeds total amount of Bragg intensity for crystals with an average B-factor of 20 Å^2 [1].

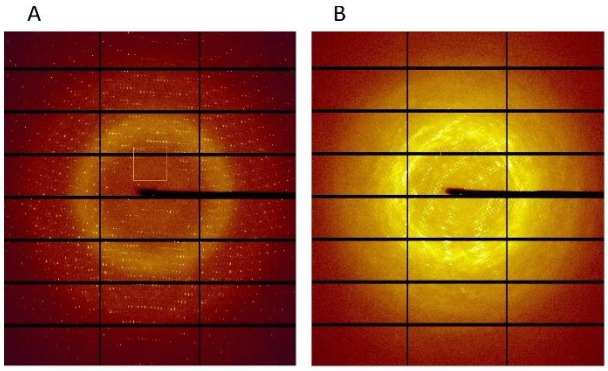


Figure 2: Two examples of X-ray diffraction patterns. A) only scattering from the solvent surrounding the crystal and that within the crystal. B) A fair amount of diffuse scattering is observed as cloudy pattern.

Additionally to the last described phenomena introducing disorder in the crystal, is that the protein in the crystal does not act as a rigid body. The protein has internal dynamics too. Smaller structural entities such as domains or secondary structural elements can be displaced in addition. This disorder contributes to the diffuse scattering too. Therefore, obtaining and analysing this diffuse scattering could give insight into the dynamics of the protein and increase our understanding of the mode of action (Figure 3).

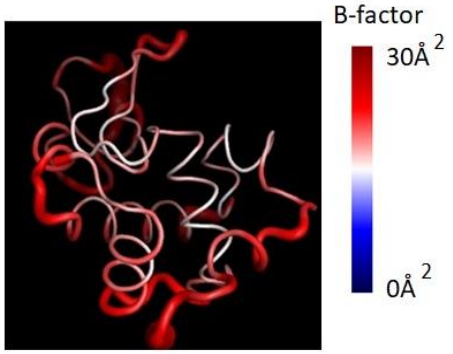


Figure 3: How the observed diffuse signal could explain the B-factors.

Objective

The goal of this research is to obtain insight into the disorder of the lysozyme molecule within the crystal. This could be achieved by performing X-ray diffraction experiments on lysozyme crystals and extracting the diffuse scattering signal. Unfortunately, difficulties exist with extracting this signal due to the fact that it remains a weak broad signal compared to the Bragg reflections. Moreover, the diffuse scattering signal is also mostly overlapping with the signal originating from the protein solvent making them even more difficult to distinguish.

Increasing this diffuse scattering signal from X-ray diffraction experiments requires a different approach in crystallizing and performing X-ray diffraction experiments. In this research we have studied if disorder can be increased by crystallizing lysozyme at higher temperature. Furthermore, we developed an approach to extract the diffuse scattering signal by masking the Bragg reflections and correcting for the intensity of the crystal solvent. From such processed images containing solely the diffuse scattering signal, we created experimental 3-dimensional reciprocal space maps.

To explain the diffuse scattering as a result of disorder of the lysozyme molecule within the crystal, a lysozyme model structure is created wherein disorder is introduced. From this disordered lysozyme structure, 3-dimensional reciprocal space maps are calculated. The correlation between the two maps gives us insight in the quality of the experimental data and the accuracy of the diffuse model (Figure 4).

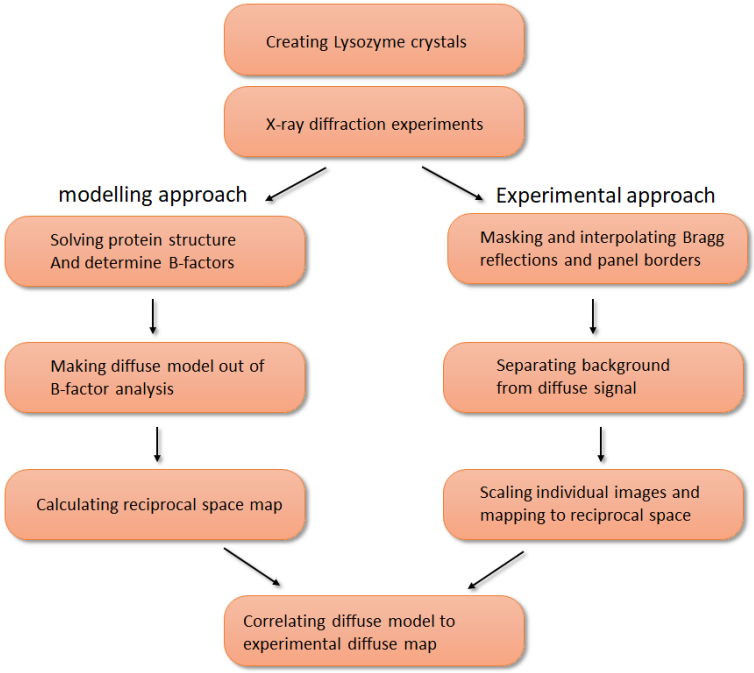


Figure 4: The experimental approach, how to go from crystallizing a protein, obtaining as much diffuse scattering signal as possible, to creating models of diffuse scattering and making experimental 3-dimensional reciprocal space maps.

2 Materials and methods

Lysozyme crystallisation and X-ray diffraction

Lysozyme crystals were grown using the hanging drop method. In here, 200 µl solution was pipetted in the reservoir, containing 10 %(m/v) NaCL and 0.2 M of NaAc/HAc buffer with pH of 4.6. The hanging drops on the cover glasses were made by adding 2 µl of 150 mg/ml lysozyme solution dissolved in 0.2 M of NaAc/HAc buffer with pH of 4.6 together with 2 µl of the reservoir solution. The hanging drops were made at room temperature and after sealing, placed at 37 ˚C for 2 days. After lysozyme crystals had formed, the cover glasses with the hanging drop were taken off and 4 µl of 20 % glycerol in water was added to the hanging drop. Hereafter, crystals were fished using different types of loops and immediately frozen in liquid nitrogen before shipment to the ESRF synchrotron facility. At ESRF a beamline of 5 µm with 14.200 keV (0.8731 Å) was used for X-ray diffraction experiments and various transmissions and exposure times were used for different crystals (Table 1). For all crystals a rotation range around omega axis of 0.1 ˚ per image was used.

Next to own X-ray diffraction experiments, more datasets were downloaded; from SBGrid 656 [2], Zenodo; ID7B2 [3] and from Proteindiffraction.org; NSLS2 [4], unfortunately some experimental crystal conditions remain unknown. Table 1 shows an overview of analysed crystals and experimental features.

Table 1: overview datasets and used experimental features

dataset	temp. crystallisation (˚C)	initial protein conc. (mg/ml)	NaCl conc. in reservoir (%m/v)	crystal size	transmission	exposure time (sec.)	omega rotation per image (˚)	total omega rotation (˚)
UU-lysozyme-85	37	150	10	100	0.2	0.006	0.1	380
UU-lysozyme-89	37	150	10	200	0.25	0.006	0.1	380
UU-lysozyme-68	37	150	10	*	0.2	0.006	0.1	380
proteindiffraction-nsls2	*	*	*	*	*	0.05	0.1	360
Zendodo-ID7B2	*	*	*	*	*	0.099	0.1	360
sbgrid-656	20	20	10	*	0.1	1	1	120

* unknown experimental feature

Peak search, integration, molecular replacement and refinement

The images of UU-lysozyme-85 were merged by 10 and for UU-lysozyme-68 and 89 images were merged by 5 using the function imagesum within Eval [5], Table 2 shows an overview of how many images are merged for each dataset. A Peak search was performed using View [5], were after dirax [5] was used to index Bragg reflections and peakref [5] was used for refinement of this matrix. Integration was done using EVAL15 [5]. Molecular replacement was done using Phaser-MR within Phenix [6] with the 6rt3 protein structure [7] with 95% sequence identity. Phenix.refine [6] was used for refinement of the determined protein structure and determining B-factors.

Experimental approach

For processing of the diffuse scattering signal all image from datasets were merged using imagesum within Eval such that the rotation increasement is 1 ° per image.

Table 2: Overview for merging datasets

dataset	Rotation increment used for peaksearch and integration (°)	Rotation increasement used for processing diffuse scattering (°)
UU-lysozyme-85	1	1
UU-lysozyme-89	0.5	1
UU-lysozyme-68	0.5	1
proteindiffraction-nsls2	0.1	1
zendodo-ID7B2	0.1	1
sbgrid-656	1	1

Masking Bragg reflections

The first method used to mask Bragg reflections was using the function maskhkl within Eval, masking all pixels within 0.25 of every HKL index.
The second method that was used is a modefilter, using the function timfilt within Eval. This method replaces the central pixel intensity with the most frequent occurring intensity in a box of 21x21 surrounding that centre pixel. If more than half of the pixels inside the box are masked as invalid (e.g. bad pixels, panel edges), then centre pixel is masked too.
method I In the third method we created a new so called median image. The median image is constructed by applying a median filter (scipy.ndimage.median_filter) to the original image, with a box size of 9x9. Next, a radial profile is generated while correcting for polarisation. Both this median image and radial profile were subtracted from the original image to create a new image (Figure 5). Now only strong intensities remain and a threshold value was used to determine whether the remaining intensity was determined to be Bragg intensity. The threshold value is chosen in such a way that all Bragg peaks are removed, but the diffuse signal is left mostly intact. To accomplish this, we tried threshold values of -17, -2 and +17. All pixels exceeding the threshold value are masked, as well as all pixels surrounding those in a box of size 9x9.
method II An addition to method I was introduced. After subtraction of the median and radial profile each pixel is averaged with its 4 orthogonally adjacent neighbours. Hereafter using a similar method, a threshold value was used to determine how much intensity was allowed to remain before masking accordingly.
method III An final addition to method II was introduced where the strong features in the image are enhanced by applying a sharpening filter to the unmodified image. The filter consists of a 3x3 pixel box, with weights 8.2 for the centre pixel and -0.9 for the eight surrounding pixels. Afterwards, the median, radial and thresholding procedures are applied as described in method II. (figure 5).

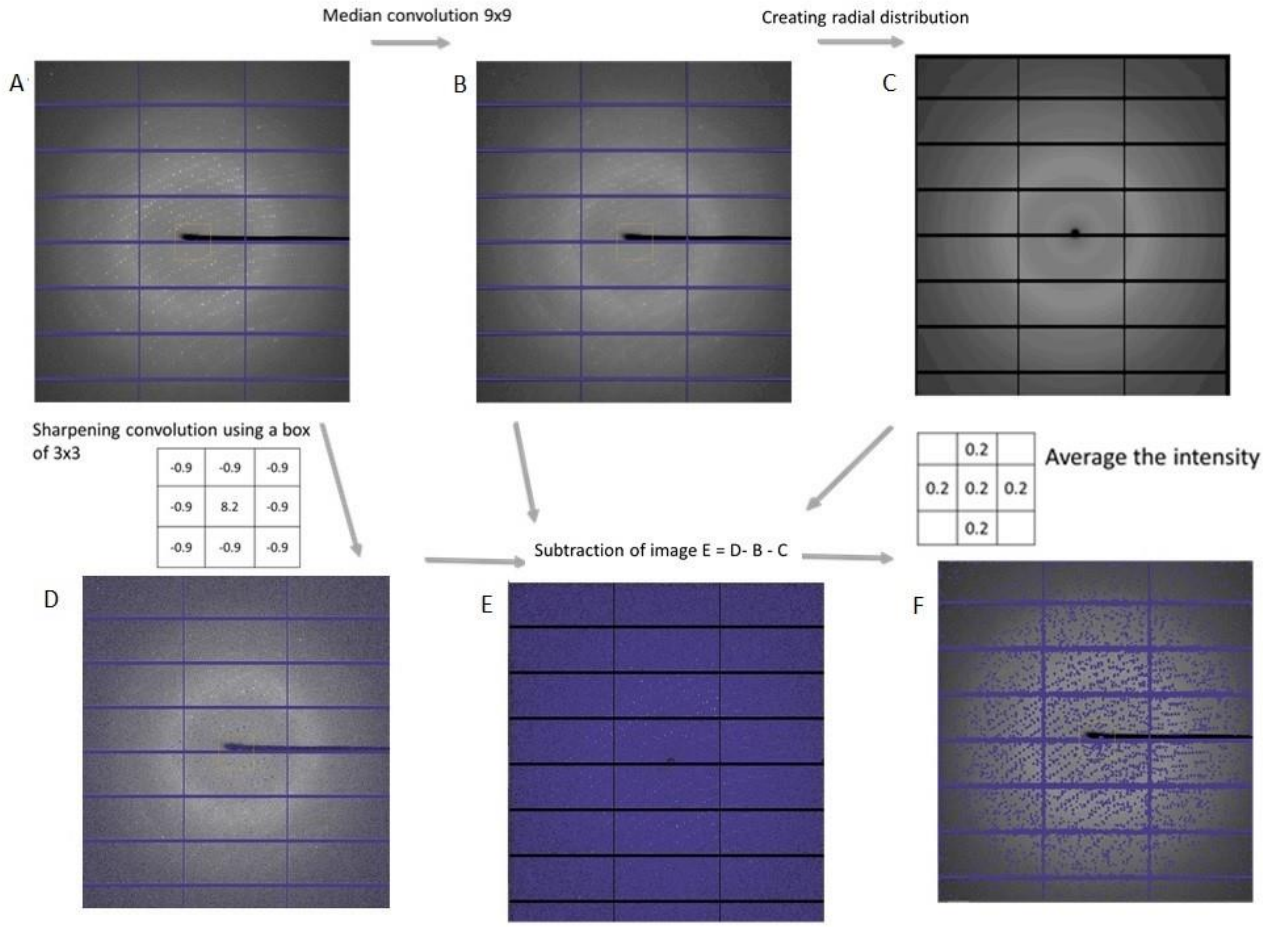


Figure 5: Procedure for identifying and masking Bragg reflections (Method III). A) UU-lysozyme-85 original detector image. B) a new so called median image is made. Herein, for every pixel the intensity is replaced with the median intensity in a box of 9x9 surrounding the centre pixel. C) From this median image a Radial profile is made using the function Radial within Eval. D) An sharpening convolution was performed multiplying the centre pixel by 8.2 combined with the 8 surrounding pixels multiplied by -0.9. E) Both the median image and radial profile were subtracted from this sharpened image to create a new image. F) all pixels were averaged using the four adjacent pixel. Whenever the intensity of a pixel was higher than the threshold value, all pixels in a box of 9x9 were masked.

Interpolating masked pixels

To interpolate the masked pixels, panel borders and beamstop, the function scipy.interpolate.griddata was used. This function uses a 2 dimensional linear function to interpolate, calculated according to pixel intensities in the near surrounding.

Scaling individual detector images

We tested various methods for image scaling. One way was using Any within the Eval suite using half window size of 1. Any calculates for every detector image a scalefactor by comparing the Bragg reflections in that image to their symmetry equivalents. For the second scaling method, we calculated the mean intensity of each image. Then, the images were scaled to make all means equal. The third method is similar to the second, but here we scaled based on the mode of the intensities instead of the mean.

For the last method, we first mask all pixels with intensity values occurring fewer times than either 1% or 10% than the most frequently intensity value occurs. Next, the images are scaled based on the mean intensity of the unmasked pixels.

Mapping images to the 3-dimensional reciprocal space

The experimental 3-dimensional reciprocal space maps were constructed using a super cell approach with the function img2hkl within the Eval suite. Parameters were the specified scale factor for every image, the super cell used was chosen to be 5x5x10 times the unit cell axis a, b, c respectively and using the "accumulate on" command.

Modelling approach

Building model of diffuse scattering signal

After solving the protein structure and determining B-factors, pandemic.adp[8] was used to determine how much of the B-factors can be explained by a correlated displacement of multiple atoms. Determined was how much of the B-factor could be explained by the 4 bigger ranges of motions: the entire protein, secondary structure elements, amino acids residues and side chains. We model the diffuse scattering using the supercell method. The supercell was constructed with a size of 5x5x10 unit cells in the a, b, and c directions, respectively. The entire supercell is filled with a transitionally and rotationally displaced lysozyme. The way the lysozyme structures are displaced is according to the chain motion contribution of the B-factor, determined by pandemic.adp. The diffuse signal is calculated as $I_{diffuse} = <F_{calc}^{2}> - <F_{calc}>^{2}$, in which we average over 128 randomly generated supercells.

3 Results

For all datasets a protein structure is solved, Table 3 showing data processing statistics regarding the protein structure (see Appendix I for extended information regarding data collection and refinement).

Table 3: overview of used datasets and experimental results

dataset	mean B-factor(Å^2)	I/σ	resolution (Å)
UU-lysozyme-85	29.7	11.6	2.0
UU-lysozyme-89	27.3	16.1	1.7
UU-lysozyme-68	18.9	15.7	1.5
Proteindiffraction-nsls2	16.4	18.0	1.3
Zenodo-ID7B2	14.0	18.7	1.8
Sbgrid-656	13.2	20.8	1.2

The three UU-lysozyme, proteindiffraction and zenodo datasets are collected using 0.1 ° rotation range around omega axis per image and have a fairly low average intensity. Therefore, to obtain the diffuse scattering signal, the images of these dataset are merged to 1 degree. Hereafter, all discussed images will be the merged images.

Experimental approach

Masking Bragg reflections

On the experimental X-ray diffraction images various types of signals are present. To obtain the diffuse scattering signal, every other type of signal needs to be removed, including the intensity observed as Bragg reflections. Various methods have been performed to remove the Bragg reflections and are compared, whereas their the advantages and disadvantages are discussed. The first method used to remove Bragg intensities is by using the expected position of Bragg reflection and masking that area. Masking all pixels within 0.25 of the location of every hkl index, results in that most of the Bragg intensities are masked (Figure 6 B). Despite being very effective in removing Bragg intensities, this approach has two disadvantages. First, this approach requires a highly accurate cell. The less accurate the cell is determined the wider the span of pixel is required to be masked, therefore masking other pixel unnecessary likewise. Second, some intense features, possible from ice scattering, remain present.

A second method to remove Bragg intensities is a method called a modefilter. A modefilter is commonly used to remove any sharp features from images. Observed is using this modefilter with a box size of 21x21 that any intense features have been removed (Figure 6 C). Interestingly, all noise seems to have been removed as well from the images.

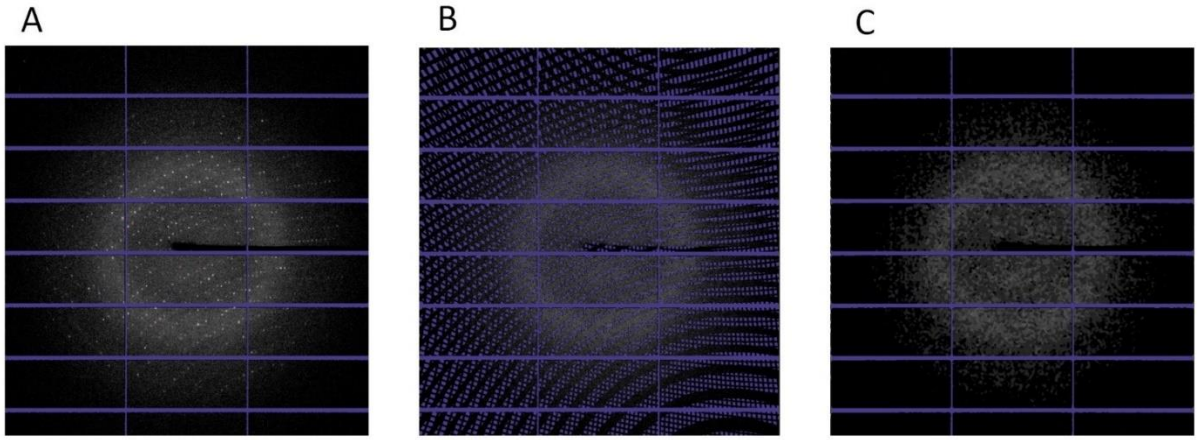


Figure 6: two different approaches of masking Bragg reflections. A) the original image. B) All pixels within 0.25 of every expected hkl index is masked. C) A mode filter is applied with a box size of 21x21.

Another approach how to mask Bragg reflections is to point out which pixels contain Bragg intensity and is based on a method described by Chapman et al. [9]. These have been described above as methods I, II and III.

Using method I, observed is that, using threshold values of -17, -2 and +17 many pixels are masked unnecessarily while expected Bragg reflections remain present (Figure 7). An close up here in is demonstrated in Figure 8.

Since method I did not yield the desired result, we continued to method II. While using the similar various threshold values of -17, -2 and 17, observed is that fewer pixels were masked. Nevertheless, several Bragg reflections remained visible whereas other pixels were masked unnecessarily. An close up of this image has been demonstrates likewise in Figure 8.

Because even method II still did not yield the desired response, a last addition had been added to method II, Method III. This time a threshold value of -17, -2, 17 and 34 is used. We show that nearly all Bragg reflection have been masked while some pixel are masked unnecessarily but the amount of pixel masked unnecessarily is acceptable. A close up of this is also demonstrated in Figure 8.

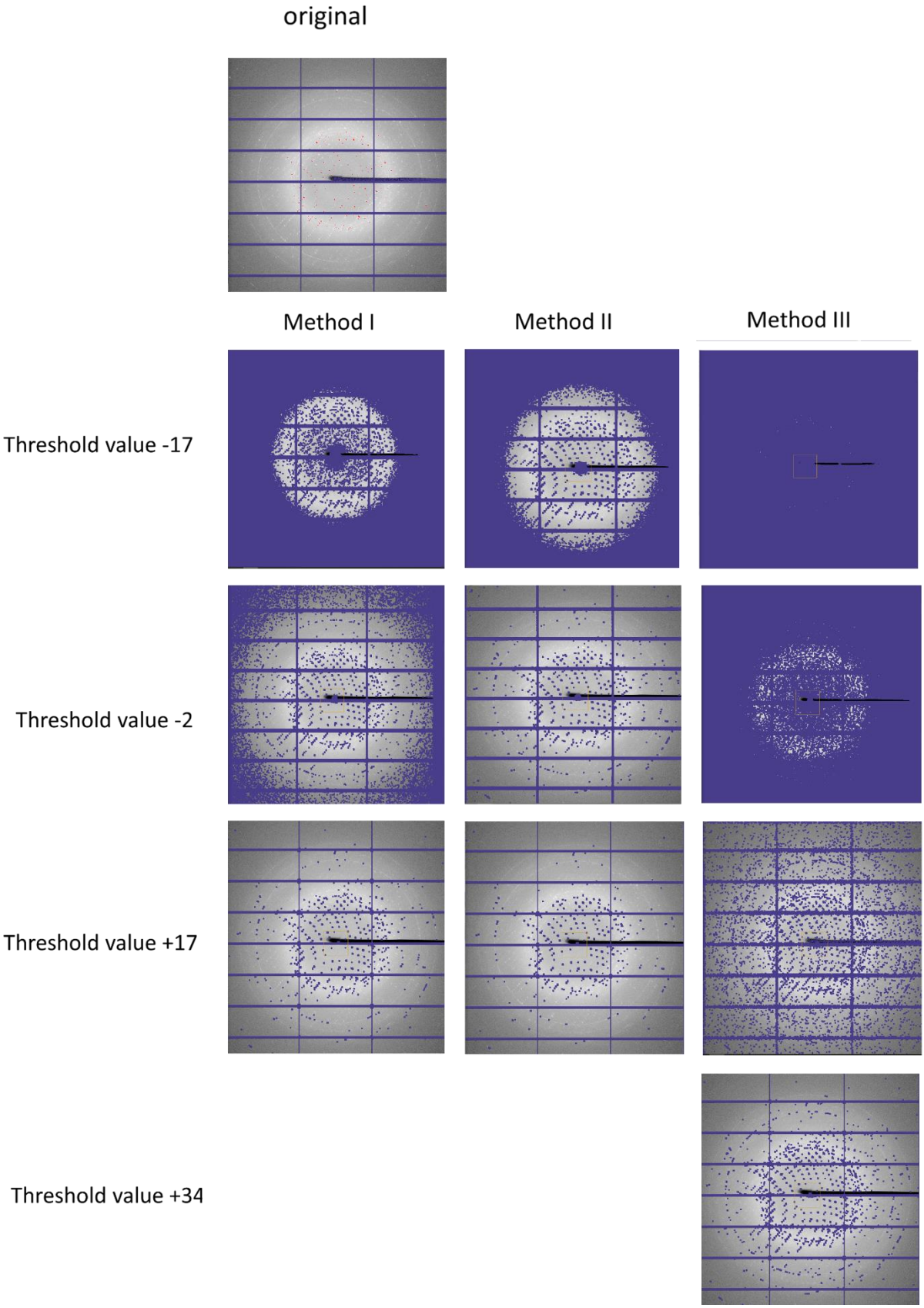


Figure 7: UU-lysozyme-85 were method I, II and III is performed using a threshold value of -17, -2, 17 and 34.

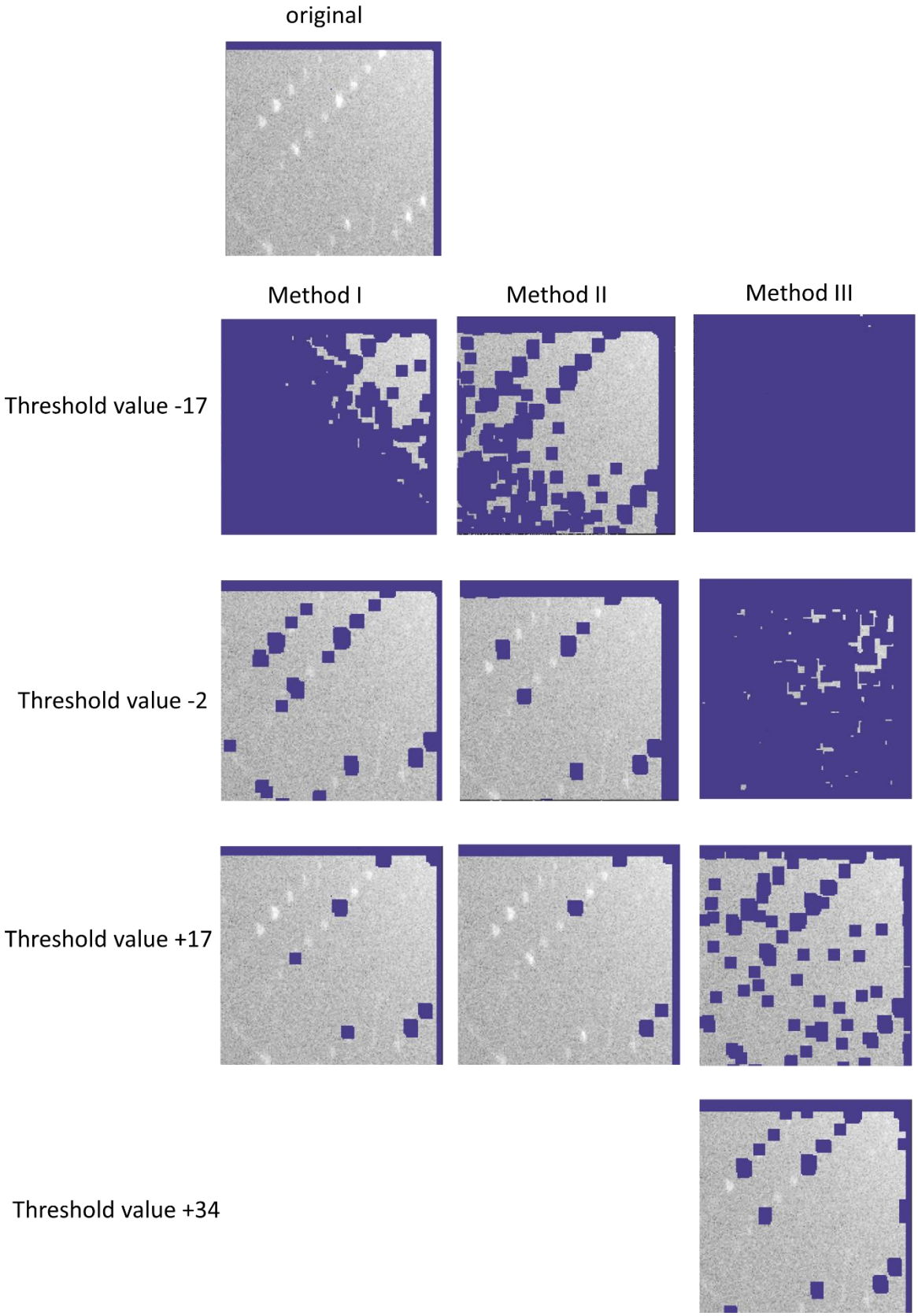


Figure 8: An close up from a part of UU-lysozyme-85 image where method I, II and III are performed using a threshold value of -17, -2, 17 and 34.

Interpolating masked Bragg reflections and panel borders

Subsequently, all the masked pixels, panel borders and the beamstop are interpolated using a 2 dimensional linear interpolation function (Figure 9). In this figure observed is that all masked pixels are interpolated according to the near surrounding.

Removing background scattering signal

The radial profile created by using the function Radial within Eval is subtracted for each image (Figure 10). After subtracting this radial profile, 43% of all intensities are below 0. Since intensities below 0 are unexpected, probably too much intensity is subtracted.

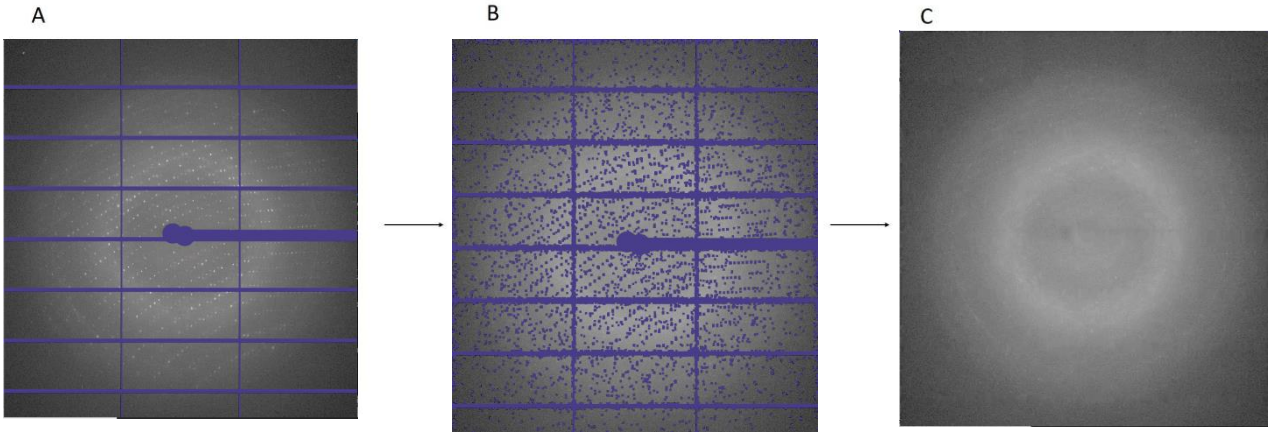


Figure 9: Interpolation of the masked Bragg reflections, panel borders and beamstop. A) original detector image UU-lysozyme-85. B) UU-lysozyme-85 after Bragg reflections have been masked. C) UU-lysozyme-85 after a 2-dimensional linear interpolation has been performed interpolating all masked Bragg reflections, panel borders and beamstop .

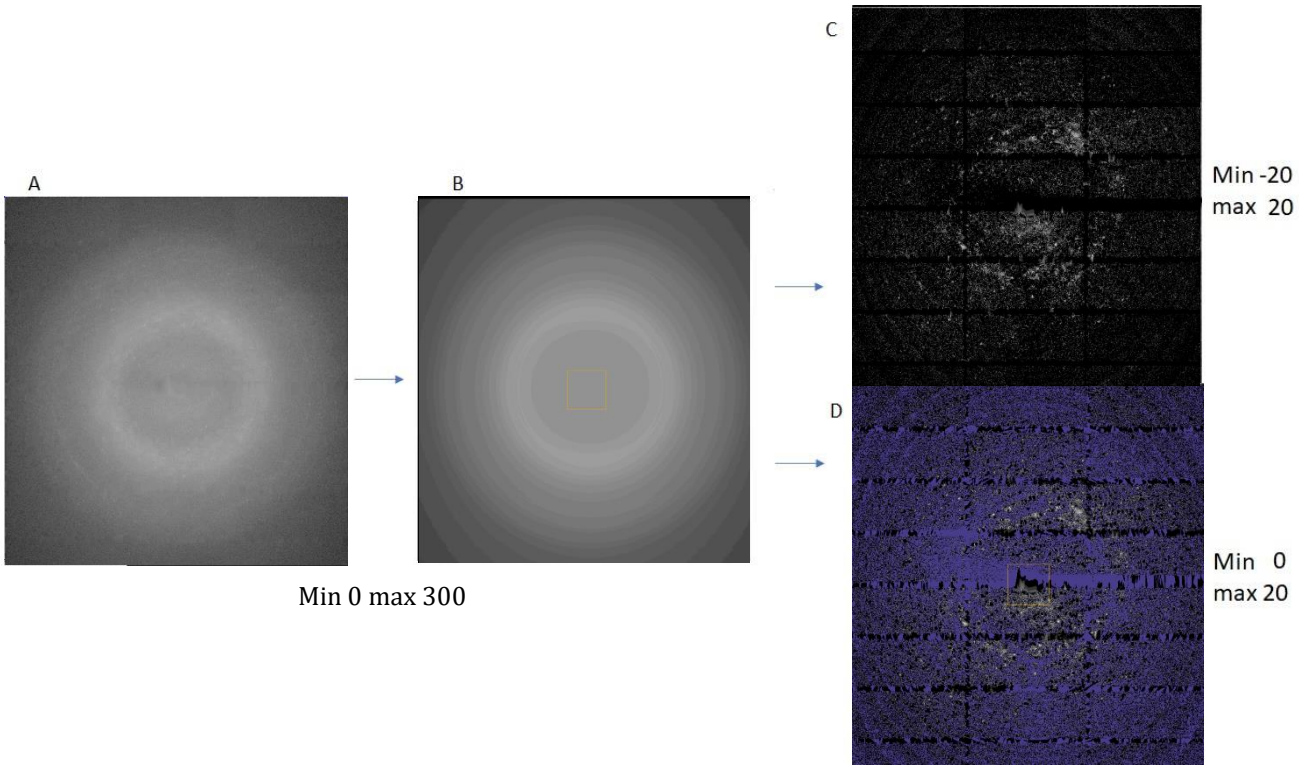


Figure 10: Removal of the solvent scattering. A) UU-lysozyme-85 interpolated. B) radial profile determined using the function radial within Eval. C and D) remaining intensity after subtraction of the radial profile.

Scaling detector images

The total observed intensity varies as a function of the rotation angle. This is caused by the varying total scattering volume, which stems from the fact that the crystal is not homogeneously shaped. Furthermore, the diffuse scattering, as well as Bragg signal will weaken due to radiation damage while exposing the crystal to X-rays. In order to negate this effect of increasing or decreasing intensity of the diffuse scattering, each individual image needs to be either scaled up or down.

Seen in Figure 11 are the scale factors calculated for the sbgrid-656 dataset and UU-lysozyme-85.

For the sbgrid-656 dataset using Any to calculated the scalefactors results in scalefactors varying within 1.4 and 0.8. If Any is used to determine scalefactors for the UU-lysozyme-85 dataset it varies between 2.25 and 0.45 and more importantly, these scalefactors are fluctuating between frames. This strong fluctuation of the scalefactors between frames is unexpected and is possible due to ice formation, a to small window size given to Any or other unintended factors.

In order to still be able to calculate a scalefactor the mean or mode is used to determine a scalefactor. For both datasets the scalefactors calculate by the

mean seems to weak compared to the results given by Any. Using the mode seems not useful since this does not results in a smooth line for both datasets. This is cause by the mode always being a integer wherefore the scalefactros are always scaled up or down 2,3,4 or 5, but never with a decimal. Although using the mode seems to be a more comparable approach to calculate scalefactors comparing to the results given by Any for UU-lysozyme-85 .

To correct for the mode always being a full integer, is by calculating how frequent the most frequently occurring intensity occurs. Next we calculate the mean intensity when all intensities are removed that do not occur at least 1% or 10% as frequent as the most frequently occurring intensity occurs. Using this alternative average of the mode creates a line quite similar to the scalefactors calculated by using the mode, although this approach does demonstrate a smooth line over the frames (Figure 11).

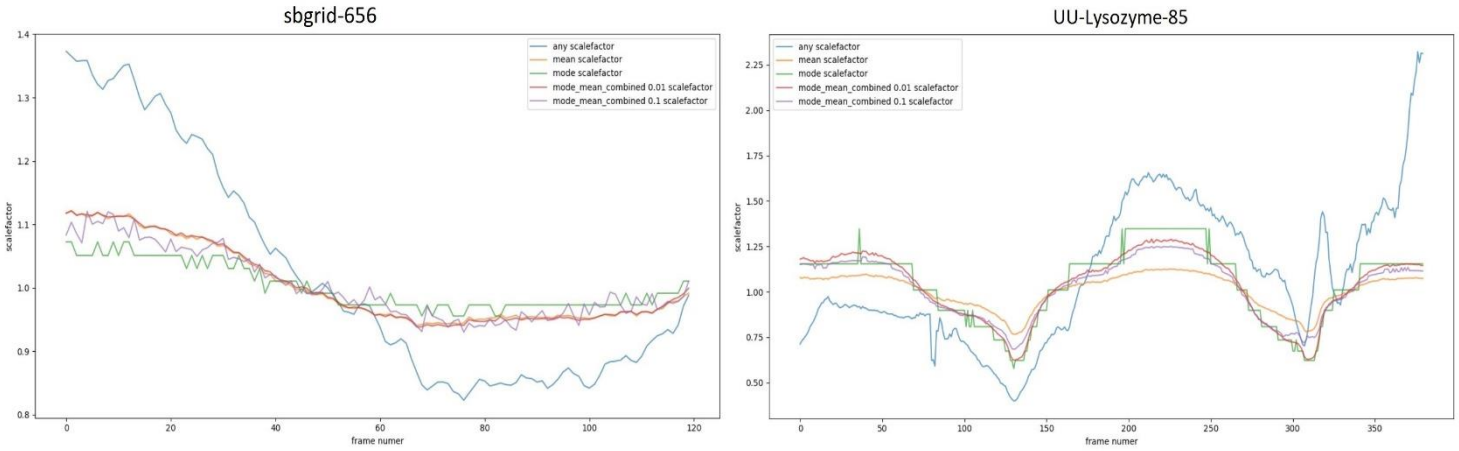


Figure 11: Scalefactors for every individual image of sbgrid-656 and UU-lysozyme-85 calculated in 5 various ways: Using Any determined along the Bragg intensities and their symmetry reflections. Multiplying all intensities in the image so the mean of every image is similar. Multiplying all intensities in the image so the mode of every image is similar. Calculating the scalefactors along the mean intensity after all pixels with intensity values occurring fewer than either 1% or 10% than the mode intensity have been masked.

Modelling diffuse scattering signal

Pandemic.adp model shows increased displacement factors

To create a diffuse model pandemic.adp [4] is used to determine how much of the B-factors of the experimental protein structure can be explained by a coherent movement of multiple atoms. To elaborate, pandemic.adp works by fitting a hierarchical series of TLS-descriptors to the atomic B-factors, such that separate levels of motions can be distinguished and reproduced. These levels are the whole chain, secondary structure elements, residues and sidechains. For the consecutive diffuse scattering calculation, we only consider the chain contribution (Figure 12). Using this displacement analysis on experimental protein structures, observed is that lysozyme structures UU-lysozyme85, UU-lysozyme-89 and UU-lysozyme-68 crystallized at 37 °C show an increased displacement over most of the levels and should theoretically contain more diffuse scattering compared to crystals grown on a lower temperature (Table 4). Having this insight into the displacement over various levels of the protein structure enables us to create a diffuse model.

Table 4: average displacement factors at different levels of movement. Using pandemic.adp it is determined how much of the atomic displacement can be assigned to a larger movement were multiple atoms are coherently translated or rotated level of displacement

dataset	total	chain	sec. struct.	residue	side chain	atomic
UU-lysozyme-85 Å(%)	29.7(100)	10.6(36)	10.4(35)	4.2(14)	2.1(7)	2.4(8)
UU-lysozyme-89 Å (%)	27.3(100)	9.5(35)	10.1(37)	3.2(12)	2.2(8)	2.3(8)
UU-lysozyme-68 Å (%)	18.9(100)	10.2(54)	5.5(29)	1.6(8)	0.9(5)	0.7(4)
proteindiffractiondata-nsls2 Å (%)	16.4(100)	7.7(47)	5.7(35)	1.5(9)	0.9(6)	0.6(3)
zendodo-ID7B2 Å (%)	14(100)	7.1(50)	3.4(24)	0.9(6)	1.4(10)	1.2(9)
sbgrid-656 Å (%)	13.2(100)	5.3(40)	2.8(21)	2.1(16)	1.4(11)	1.6(12)

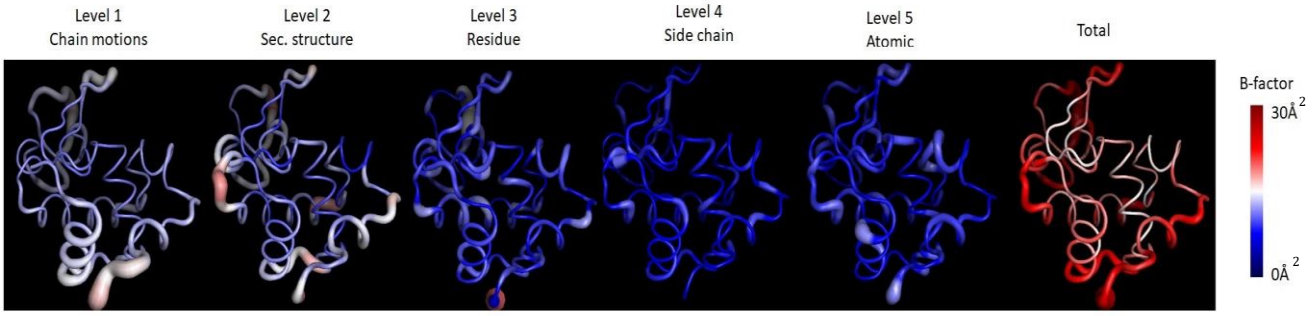


Figure 12: The portion of B-factors assigned to various levels of motions

Diffuse scattering calculation

For all six experimentally determined protein structures a disordered protein structure is created to model the diffuse scattering. In this research only displacement of the entire protein chain is used to model disorder of the lysozyme

molecule. In the supercell every protein chain is translated and rotated on the basis of the B-factor contribution of the chain level motions determined by pandemic.adp. Using this displaced super cell, a 3 dimensional reciprocal space map is calculated (Figure 13). In this figure is planes HK0, H0L, and 0KL respectively seen, demonstrating the cloudy pattern representing the diffuse scattering.

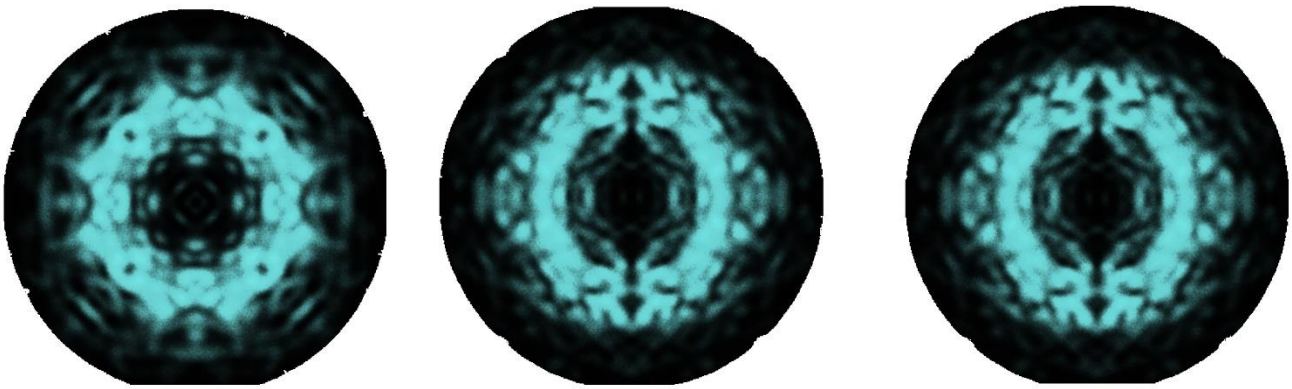


Figure 13: Showing UU-lysozyme-85 3-demensional reconstruction of reciprocal space model of the diffuse scattering signal, along HK0, H0L and 0KL respectively.

Comparing experimental 3-dimensional reciprocal space maps to diffuse scattering model maps

To explain the diffuse scattering as disorder of lysozyme within the crystal, the experimental 3-dimensional reciprocal space maps are compared to 3-dimensional reciprocal space model maps.

Comparing the experimental map of UU-lysozyme-85 (Figure 14 C) to the model map (Figure 14 E) some similar features are recognisable, nevertheless many other features cannot be observed in the experimental maps. Suggested was that the weaker diffuse scattering signal might be removed during subtracting of the solvent scattering with the radial profile. Therefore an experimental 3-demensional reciprocal space map is create using the same method except subtracting the redial profile (Figure 14 A). Despite more signal is observed, the experimental map does not show more similar diffuse scattering features compared to the model map than when the radial profile is subtracted.

Next, experimental 3-demensional reciprocal space maps are created using the a similar experimental approach except a modefilter is performed to remove intense features instead of method III (Figure 14 A and D). although the method III and modefilter maps are not identical they both show similar diffuse scattering features, providing evidence that method III yields a strong comparable result to the modefilter approach. Calculating correlation of the Method III to the modefilter using the function scud.map.correlation results in a correlation coefficient of 0.82.

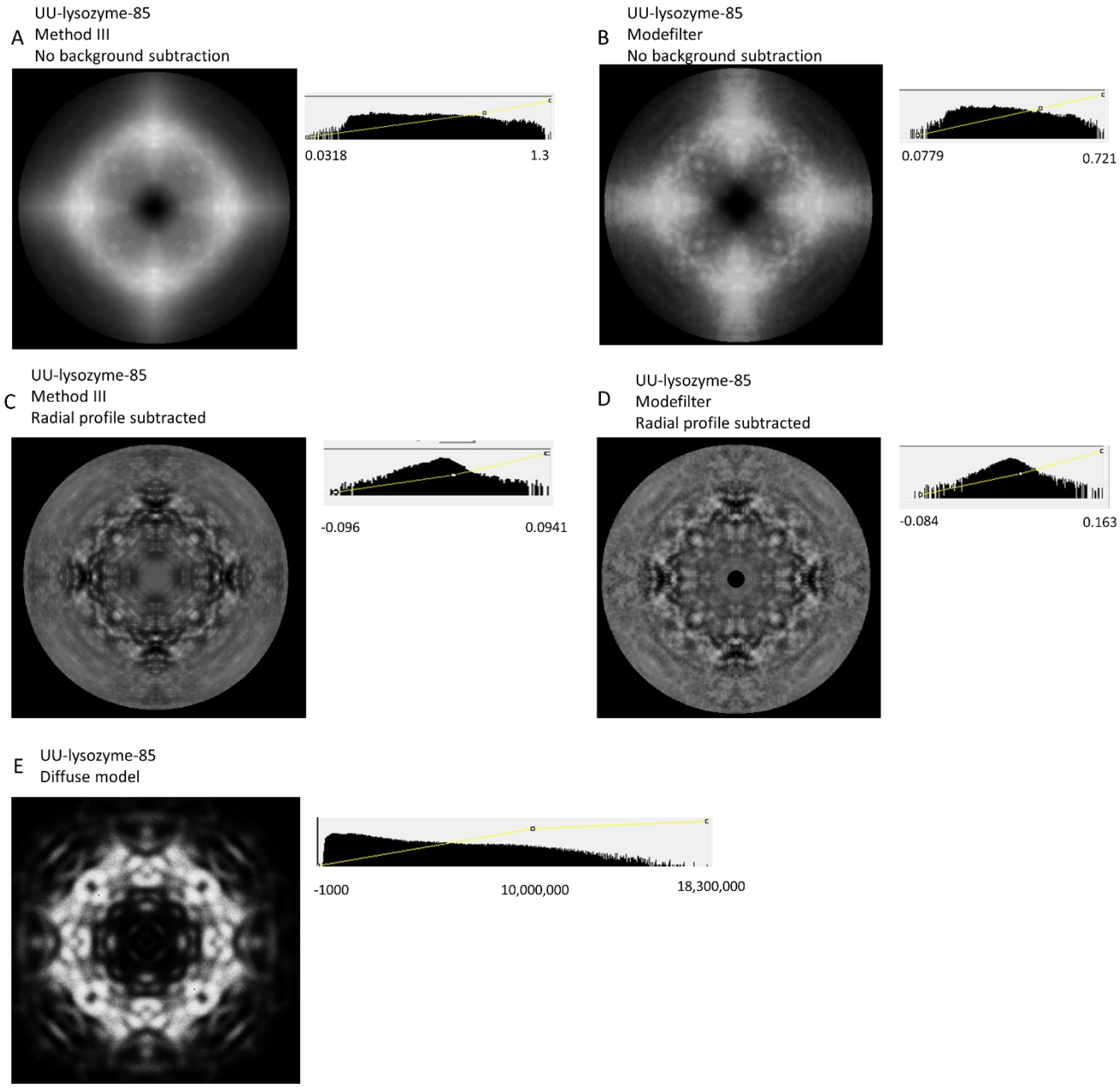


Figure 14: HK0 plane of 3-demensional reciprocal space maps of the experimental data and model. A) UU-lysozyme-85 where only method III to mask Bragg reflections has been performed. B) UU-lysozyme-85 where a modefilter has been performed. C) UU-lysozyme-85 where both method III to mask Bragg reflections and the radial profile is subtracted. D) UU-lysozyme-85 where a modefilter has been performed and a radial profile is subtracted. E) UU-lysozyme-85 diffuse model map.

One other experimental 3-demensional reciprocal space map has been made using the sbgrid-656 dataset. For this dataset a similar approach has been used, using the modefilter to remove sharper features (Figure 15). This map demonstrates many more diffuse scattering features compared to its model than UU-lysozyme-85 does. Interestingly here in is that the sbgrid-656 has an average B-factor of only 13.2 Å^2 while UU-lysozyme-85 has a B-factor of 29.7 Å^2. The increase in diffuse scattering signal while having a lower B-factor could be explained by the difference that the sbgrid-656 dataset has a higher average photon count.

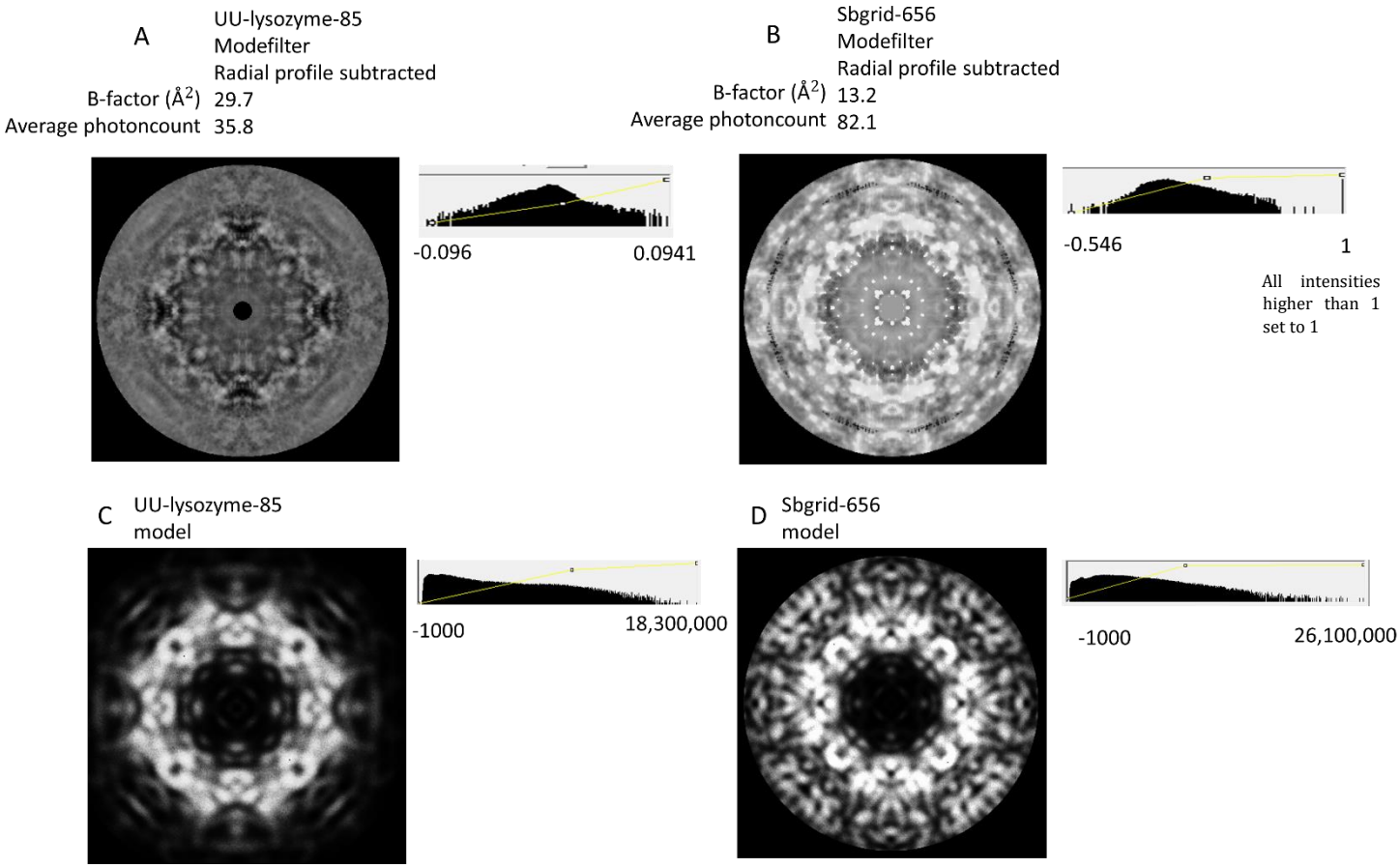


Figure 15: HK0 plane of experimental 3-dimensional reciprocal space maps of UU-lysozyme-85 and sbgrid-656. A) UU-lysozyme-85 while a modefilter has been performed and the radial profile is subtracted. B) sbgrid-656 while a modefilter has been performed and the radial profile is subtracted. C) UU-lysozyme-85 model map. D) sbgrid-656 model map.

4 Discussion

Crystallizing lysozyme

Making lysoyme crystals in order to obtain as much diffuse scattering signal as possible requires a different approach in crystallizing. In doing this, there are a couple of experimental procedures important to realize, First: an important factor is a fast crystallization process. This implies that lysozyme molecules will less likely be positioned correctly on the crystal lattice and thus increasing the B factor and diffuse scattering signal. Using a protein concentration in the hanging drop, close to the point of saturation concentration combined with a larger difference in NaCl concentration between the hanging drop and the reservoir liquid promotes a fast crystallization process. Second, temperature is an important factor. Crystallizing at a higher temperature would cause the proteins to be more dynamic. This would likely create a more disordered protein in the crystal lattice likewise. In addition, performing X-ray experiments at room temperature would be beneficial to increase the diffuse scattering signal as well. Instead of flash freezing the crystals in liquid nitrogen after fishing, X-ray experiments can be immediately performed after fishing.

Performing X-ray diffraction experiments

X-ray diffraction experiments could also be performed differently in order to obtain the diffuse scattering. To extract the diffuse scattering signal, obtaining a high photon count as diffuse scattering is essential. Increasing the transmission or exposure time to be able to obtain a higher photon count is indispensable, even at the cost of a more strongly decaying maximum resolution caused by the radiation damage. To still assure being able to fill the entire reciprocal space, rotating the crystal 90 ° is recommended, using a high order symmetry like tetragonal lysozyme as used in this research.

Method III

Although method III masked all Bragg reflections, many pixels not containing Bragg intensity were masked unintentionally as well. A possible solution to this would be to adjust the sharpening so that pixels are only enhanced if multiple stronger localised pixel can be found adjected to each other as is found in Bragg reflections and not in the noise. This can be achieved by using a different convolution for example shown in the box in Figure 16. A second improvement to masked less pixel unnecessary could be by using the predicted position of Bragg reflections. For example, the masking of pixels could only be allowed if they fall with in a radius of 10 pixels with in the expected position of the Bragg reflex.

				-3				
			-3	-1.5	-3			
		-3	-1.5	4	-1.5	-3		
	-3	-1.5	4	6	4	-1.5	-3	
-3	-1.5	4	6	10	6	4	-1.5	-3
	-3	-1.5	4	6	4	-1.5	-3	
		-3	-1.5	4	-1.5	-3		
			-3	-1.5	-3			
				-3				

Figure 16: proposed convolution

Lastly, Seen is that to much intensity is interpolated after the Bragg reflections are masked (Figure 17 A). This is likely due to the Bragg reflections not entirely being masked. One way to solve this is by masking a bigger area for example, instead of using 9x9 as is used in this research

pixels can be mased 13x13. A second way to solve this is by using not only the near surrounding to interpolate the masked area but using some more pixels for interpolation.

Subtracting signal originating from solvent scattering

Subtracting the amount of signal originating from the solvent scattering remains a challenge. In this research we demonstrated subtracting a radial profile. Although this subtraction is most likely subtracting to much intensity since the intensity of 43 % of all the pixels becomes lower than 0, which is unlikely because an intensity below 0 is theoretically impossible. Preferably we would have everywhere zero intensity observed unless diffuse scattering signal is present. This asks for a different subtraction of the solvent scattering. There are more various ways to create a background image. One way to do this is by performing an X-ray experiment on a loop, solely containing the crystal solvent. Nevertheless, due to the missing crystal there is more solvent present, thus more intensity is observed as solvent scattering signal. A downscaling according to the amount solvent scattering observed in the experimental image is therefore necessary. Nevertheless, the question remains how much the downscaling is required.

Comparing Method III with modefilter

Although experimental 3-demensional reciprocal maps of method III and the modefilter are quite comparable they are not identical. Scud has calculated the corelation between these two maps to be 82 %. There seem to be two differences between the maps. First, after Method III has been performed the Bragg reflections are stronger observed compared when a modefilter is performed (Figure 17 A and B). This could be caused by the Bragg reflection not being entirely masked and the thus the interpolating returning too much intensity. Secondly, using method III seems create a smoother map than performing a modefilter. This can originate from the fact that the modefilter is performed on image level removing all the noise and maybe therefore unsmoothing the images (Figure 17 C and D).

Moreover, while performing a modefilter on the images, 80% of all intensities are replaced with a different intensity. This should be kept in mind while performing a modefilter. The intensity observed in the reciprocal space map are most likely not the exact intensities measured. Performing method III masked 20% of all pixels which will be interpolated, meaning that four times as much pixels remain unaltered compared to the modefilter, yielding a more unadjusted result.

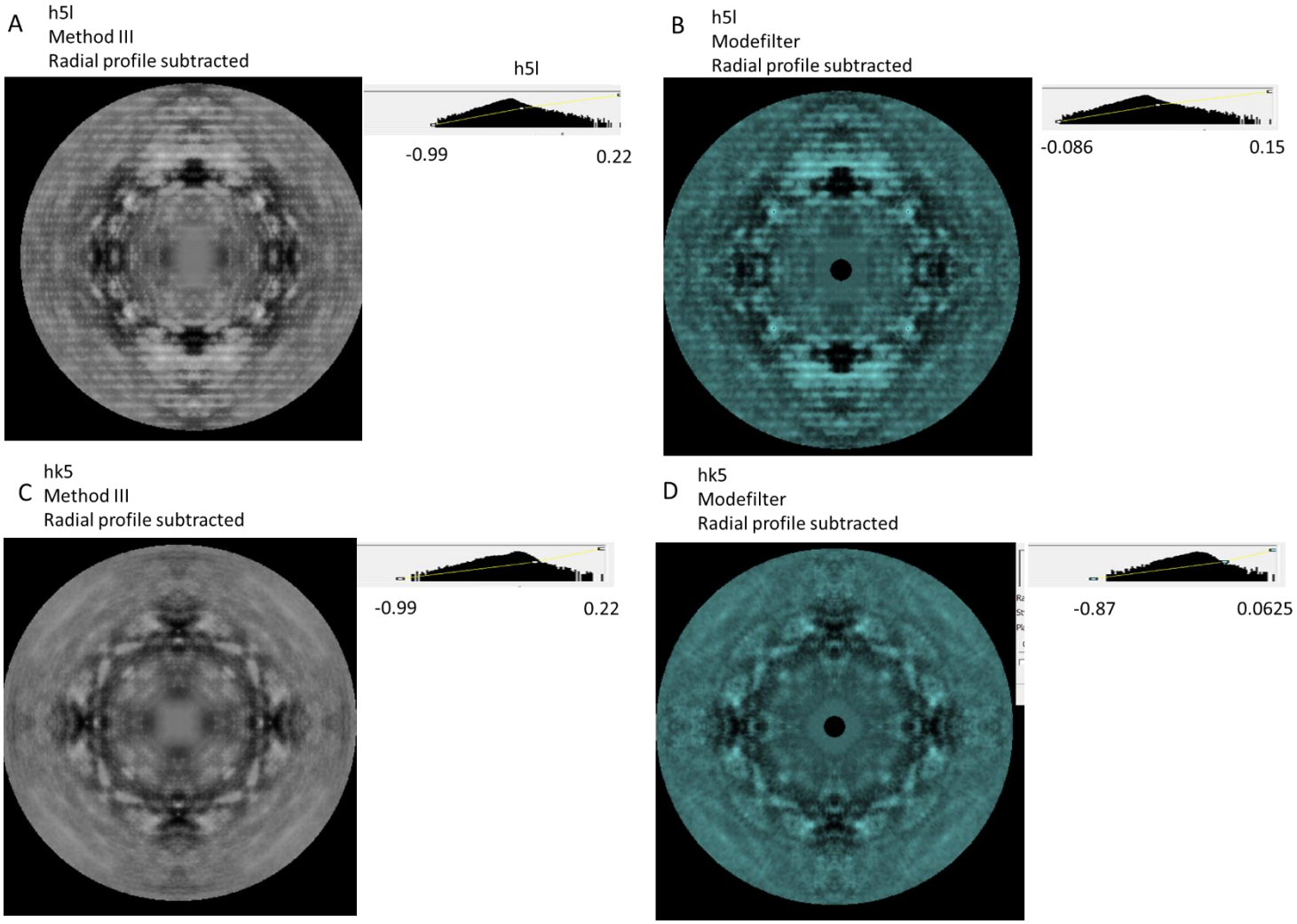


Figure 17: h5l and hk5 plane of UU-lysozyme-85 3-demensional reciprocal space maps. A) h5l plane after performing method III. B) h5l plane after performing a modefilter. C) hk5 plane after method III has been performed. D) hk5 plane after a modefilter has been performed.

5 Conclusion

To obtain the diffuse scattering signal, a different approach in performing X-ray diffraction experiments is required. For optimal extraction of the diffuse scattering signal, obtaining a photoncount as high as possible is demanded. There are multiple factors crucial to increase the photoncount of the diffuse scattering signal: First, producing a highly disordered protein crystal, which could be achieved by e.g. crystallizing as fast as possible while using a temperature as high as possible. Second, use a high transmission of the X-ray source to obtain more signal, preferably with a crystal with high order of symmetry allowing shorter rotation of the crystal while radiation damage is limited. When X-ray diffraction experiments have been performed, intense features such as Bragg reflections and from ice scattering can be removed using method III. At last, scaling individual detector images should be done by Any using the Bragg intensities. Whenever Any is not able to calculate scalefactors without strong fluctuation, scale the images to the average intensity after intensities that do not occur at least 1% of the amount of times as the most frequently occurring pixels have been removed. Reciprocal space maps created using this approach to extract the diffuse scattering signal are quite identical to the maps created using a modefilter instead of method III. Both these two experimental maps show some similar features compared to reciprocal space model map. Concluding that diffuse scattering signal can be extracted using the described method.

References

[1] T. de Klijn, A. M. M. Schreurs and L. M. J. Kroon-Batenburg *Rigid-body motion is the main source of diffuse scattering in protein crystallography*. *International Union of Crystallography*. 6:277-289, 2019

[2] Pereira, PJB. 2019. "X-Ray Diffraction data for: Chicken egg white lysozyme. PDB Code 6RT3", SBGrid Data Bank,

V1, https://doi.org/10.15785/SBGRID/656.

[3] Finke, Aaron D. Lysozyme Diffraction Data from Beamline ID7B2 (FlexX), CHESS (broadband monochromatic X-rays at 1.5% bandpass) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3462313. 2019

[4] Fuchs M., Shi W. *Lysozyme calibration data for NSLS-II 17-ID-2 beamline. proteindiffractiondata.* 2016

[5] A. M. M. Schreurs, X. Xian and L. M. J. Kroon-Batenburg *EVAL15: a diffraction data integration method based on ab initio predicted profiles. Journal of applied crystallography*. 2010\

[6] recent developments in Phenix: D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunk´oczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, and P. D. Adams *Macromolecular structure determination using*
X-rays, neutrons and electrons. Acta Cryst.. D75, 861-877, 2019

[7] Ferreira, Pinto, M.F., Macedo-Ribeiro, S., Pereira, P.J.B., Rocha, F.A., Martins, P.M. Phys Chem Chem Phys 22: 16143-16149 2020

[8] N. Pearce, P. Gros *A method for intuitively extracting macromolecular dynamics from structural disorder. bioRxiv*. 2020

[9] Chapman, Henry N. Yefanov, Oleksandr M. Ayyer, Kartik White, Thomas A. Barty, Anton Morgan, Andrew Mariani, Valerio, Oberthuer Dominik, Pande Kanupriya *Continuous diffraction of molecules and disordered molecular crystals. Journal of Applied Crystallography.*. 50, 1084-1103, 2017

Appendix

Table 5: Experimental features of the protein structures regarding data collection and refinement

	UU-lysozyme-85	UU-lysozyme-89	UU-lysozyme-68	proteindiffractiondata-nsls2	zendodo-BL2B2	sbgrid-6rt3
Data collection						
Spacegroup	P43212	P43212	P43212	P43212	P43212	P43212
Cell dimensions:						
a, b, c (Å)	77.36, 77.36, 37.24	77.24, 77.24, 37.30	77.47, 77.47, 37.86	78.66, 78.66, 37.36	79.11, 79.11, 37.05	78.87 78.87 36.80
A, B, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	2.0	1.7	1.5	1.3	1.8	1.2
Rsym and Rmerge	1.072, 1.095	0.969, 0.983	0.891, 0.938	0.589, 0.605	0.108, 0.113	0.914, 0.972
I/σ	11.6	16.1	15.7	18.0	18.7	20.8
Completeness (%)	100	100	96.9	86.6	100	100
Redundancy	24.46	36.44	10.75	17.74	11.90	8.96
Refinement						
Resolution (Å)	2.0	1.7	1.5	1.25	1.8	1.2
No. reflection	114086	188470	281616	496692	166632	577674
R-work / R-free	0.2112, 0.2843	0.1962, 0.2130	0.1819, 0.2147	0.2036, 0.2257	0.1540, 0.1891	0.1068, 0.1318
No atoms:						
Protein	1950	1950	1950	1950	1950	1950
Water			52	112		
B-factors:						
Protein	29.7	27.3	18.9	16.4	14	13.2
R.m.s. deviations:						
Bond lengths (Å)	0.007	0.011	0.010	0.008	0.009	0.006
Angles (°)	0.864	1.134	1.047	0.997	1.106	0.914