



**HAL**  
open science

# Development of serial protein crystallography with synchrotron radiation

Anastasiia Shilova

► **To cite this version:**

Anastasiia Shilova. Development of serial protein crystallography with synchrotron radiation. Biophysics. Université Grenoble Alpes, 2016. English. NNT : 2016GREAY034 . tel-01499403

**HAL Id: tel-01499403**

**<https://theses.hal.science/tel-01499403>**

Submitted on 31 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **PHYSIQUE POUR LES SCIENCES DU VIVANT**

Arrêté ministériel : 25 mai 2016

Présentée par

**Anastasiia SHILOVA**

Thèse dirigée par **Eva PEBAY-PEYROULA** et  
codirigée par **Manfred BURGHAMMER**

préparée au sein du **European Synchrotron Radiation Facility**  
dans l'**École Doctorale de Physique**

## Development of serial protein crystallography with synchrotron radiation

Thèse soutenue publiquement le **21/12/2016**,  
devant le jury composé de :

**Dr. Yves Bourne**

Directeur de recherche CNRS, AFMB Marseille, Rapporteur, président  
du jury

**Dr. Arwen Pearson**

Professeur, University of Hamburg, CUI Hamburg, Rapporteur

**Dr. Dimitri Ivanov**

Directeur de recherche CNRS, IS2M, Mulhouse. Membre

**Dr. Eva Pebay-Peyroula**

Professeur, Université Grenoble Alpes, IBS, Grenoble, Directeur de  
thèse, Membre

**Dr. Manfred Burghammer**

Chargé de recherche, ESRF, Grenoble, Co-Directeur de thèse, Membre



## THÈSE

Pour obtenir le grade de

### DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **PHYSIQUE POUR LES SCIENCES DU VIVANT**

Arrêté ministériel : 25 mai 2016

Présentée par

**Anastasiia SHILOVA**

Thèse dirigée par **Eva PEBAY-PEYROULA** et  
codirigée par **Manfred BURGHAMMER**

préparée au sein du **European Synchrotron Radiation Facility**  
dans l'**École Doctorale de Physique**

## Développement de la cristallographie sérielle de protéine utilisant le rayonnement synchrotron

Thèse soutenue publiquement le **21/12/2016**,  
devant le jury composé de :

**Dr. Yves Bourne**

Directeur de recherche CNRS, AFMB Marseille, Rapporteur, président du jury

**Dr. Arwen Pearson**

Professeur, University of Hamburg, CUI Hamburg, Rapporteur

**Dr. Dimitri Ivanov**

Directeur de recherche CNRS, IS2M, Mulhouse. Membre

**Dr. Eva Pebay-Peyroula**

Professeur, Université Grenoble Alpes, IBS, Grenoble, Directeur de thèse, Membre

**Dr. Manfred Burghammer**

Chargé de recherche, ESRF, Grenoble, Co-Directeur de thèse, Membre



## Acknowledgments

Thanks to **Eva Pebay-Peyroula** for being my university supervisor, for her helpful advice and valuable discussions, and for the opportunity to have a deep dive in the field of structural biology in the summer school in Les Houches.

I would like to thank **Manfred Burghammer** for his attentive supervision during these years. I really appreciate that he gave me an opportunity to discover the amazing field of serial crystallography and synchrotron radiation. Thanks for his guidance and support.

Thanks to the “NanoMem” Marie-Curie initial network for the financial support of my thesis. Thanks to *Gebhard Schertler* for being my mentor. Thanks to *Richard Neutze* for organizing all “NanoMem” meetings.

A special thanks to **Jacques-Philippe Colletier** who introduced me to the fields of serial crystallography and structural biology. His explanations rendered the seemingly impossible easy to understand. Thanks to **Nicolas Coquelle** for all the knowledge that I gained about data analysis and software. I really appreciate both his help and great advice. Thanks to **Joyce Woodhouse** for his help with crystallization of ubiquitin.

Thanks to **Aleksandra Woznicka** and **Hugues Nury** for their close collaboration and for providing the membrane proteins. Special thanks to Aleksandra for not only being my colleague but also my close friend.

I would like to thank **Josan Marquez**, **Florent Cipriani** and **Zuzanna Kaczmarek** for the efficient collaboration about CrystalDirect project at ID13.

Thanks to **Uwe Weierstall**, **Joerg Standfuss** and their teams for providing LCP-injectors and samples.

Thanks to **ID13 team**: **Martin Rosenthal** for his help with mechanical development of the CrystalDirect set up and with his help at the beamline; **Lionel Lardiere** for all installations and technical supports at the beamline; ID13’s former postdocs **Britta Weinhausen** and **Thomas Dane** for their help and advice

Thanks to **my friends** and to **my family** for their support and advice. Special thanks to **Pierre-Henry F.** for being patient and ready to realize my plans and ideas outside the synchrotron. Special thanks to my **mother** for

being understanding, attentive and supportive all the time. Thanks for all opportunities that you gave me. You are the best mother.

## Abstract

With the advent of X-ray free electron laser sources (XFELs), serial crystallography - a new method of data collection where thousands of crystals are delivered to the beam and exposed in random orientation at room temperature - was developed. Recently, serial crystallography approaches have started to appear on the more flexible and widely available microfocus beamlines at synchrotron sources. This thesis work is focused on the development of room temperature serial crystallography of proteins at the microfocus beamline ID13 at the European Synchrotron (ESRF), Grenoble, France.

This work describes the implementation of three different sample delivery approaches for room temperature serial crystallography. One of them is lipidic cubic phase (LCP) microjet-based serial millisecond crystallography, similar to the experiments with injectors originally developed at XFELs. A continuous LCP stream containing micro-crystals is injected into the microbeam. The LCP-jet approach will be demonstrated in this work on the example of tubulin. Another method is based on an automated system for crystal harvesting - CrystalDirect, developed at EMBL, Grenoble, France. Due to their design, CrystalDirect plates allow the collection of diffraction data in-situ with very low background. The main advantage of CrystalDirect plates is an absence of crystal transfer as they can be measured directly from the crystallization drop. This approach has been successfully applied to test samples such as lysozyme, thaumatin, ubiquitin, and also to some membrane proteins. An alternative solid support approach is based on scanning micro-crystals deposited on silicon nitride membranes, that has the lowest background scattering among all other presented methods. Successful application of this method to study room temperature conformational behaviour of ubiquitin will be discussed.

## Résumé

Les sources de rayons X produites par lasers à électrons libres (XFELs) ont permis l'émergence de la cristallographie sérielle. Cette méthode permet de collecter les données de milliers de cristaux exposés selon une orientation aléatoire à température ambiante. Récemment, la cristallographie série s'est développée au sein des synchrotrons sur les lignes de faisceaux microfocus, plus accessibles et flexibles que les XFELs. Cette thèse porte sur la cristallographie sérielle de protéines à température ambiante sur la ligne ID13 de l'ESRF (European Synchrotron Radiation Facility), Grenoble, France.

Ce travail décrit la mise en œuvre de trois méthodes de distribution d'échantillons précédant une cristallographie série à température ambiante. La première méthode utilisant les phases cubiques de lipides (LPC) en microjet pour la cristallographie milliseconde, consiste à injecter un flux continu de LCP contenant des microcristaux dans le microfaisceau. Cette méthode est similaire aux expérimentations à base d'injecteurs développées initialement dans les XFELs. Son application à la tubuline sera présentée. La deuxième méthode est fondée sur un système automatisé de récolte de cristaux, CrystalDirect, développé à l'EMBL, Grenoble, France. En raison de leur conception, les plaques CrystalDirect permettent de recueillir des données de diffraction in-situ avec un très faible bruit de fond. Le principal avantage des plaques CrystalDirect est une absence de transfert de cristaux car ils peuvent être mesurés directement à partir de la goutte de cristallisation. Cette approche a été appliquée avec succès à des échantillons d'essai tels que le lysozyme, la thaumatine, l'ubiquitine et également certaines protéines membranaires. Une approche alternative de support solide est basée sur le balayage des microcristaux déposés sur des membranes de nitrure de silicium. Cette approche a la plus faible diffusion de fond parmi toutes les autres méthodes présentées. L'application réussie de cette méthode à l'étude du comportement dynamique de l'ubiquitine sera discutée.

# Contents

List of Abbreviations .....	1
List of Figures .....	3
List of Tables .....	7
PREFACE.....	8
CHAPTER 1. INTRODUCTION .....	10
1.1 X-rays and their interaction with matter .....	10
1.2 Synchrotron Radiation .....	12
1.3 X-ray free electron lasers (XFELs) .....	14
1.4 Serial Crystallography.....	16
1.5 Proteins and membrane proteins .....	17
1.6 Structure of the dissertation and motivation.....	21
CHAPTER 2. MATERIALS AND METHODS.....	22
2.1 General methods of X-ray crystallography of the proteins .....	22
2.1.1 Diffraction of X-ray by a crystal .....	22
2.1.2 Geometry of the crystals .....	24
2.1.3 The quality of the protein crystals.....	25
2.1.4 A severe limitation for protein crystallography: radiation damage .....	26
2.1.5 Phasing .....	27
2.1.6 Refinement.....	28
2.1.7 Proteins used for this study .....	30
2.2 Methods adapted to serial crystallography .....	33
2.2.1 ID13, beamline parameters and scanning modes .....	33
2.2.2 Sample delivery .....	40
2.2.2.1 LCP-injector experiments .....	40

2.2.2.2 CrystalDirect plates .....	42
2.2.2.3 Silicon nitride membranes.....	45
2.3 Data processing in serial crystallography .....	47
2.3.1 Sorting the serial crystallography data (NanoPeakCell).....	47
2.3.2 Indexing of serial crystallography data (CrystFEL) .....	48
2.3.3. Ensemble refinement.....	50
2.3.4 Data processing in this work.....	52
CHAPTER 3. RESULTS AND DISCUSSIONS.....	54
3.1 LCP-injectors.....	55
3.1.1 Explorative LCP-jet test-experiment using the EIGER 4M detector on the example of tubulin protein.....	56
3.2 Implementation of CrystalDirect plate scanning at ID13 .....	61
3.2.1 Application to thaumatin .....	63
3.2.3 Application to ligand-gated ion channels: 5HT3 and GLIC.....	65
3.3 Improvement of the data quality using liquid removal.....	70
3.3.1 Application to lysozyme.....	71
3.3.2. Application to ubiquitin .....	74
3.3.3 Application to lysozyme with benzamidine ligand .....	78
3.4 Experimental crystal growth on silicon nitride membranes. ....	86
3.5 Dynamics of ubiquitin using silicon nitride membranes.....	89
3.5.1 Protein and solvent channels can be affected by cryocooling.....	93
3.5.2 Indication of increased disorder of flash cooled crystals using ensemble refinement.....	103
SUMMARY AND FUTURE PERSPECTIVES .....	113
BIBLIOGRAPHY .....	117

## List of Abbreviations

3D - three dimensional  
CCD - charge coupled devices  
CT - cryo-temperature  
CC - Pearson correlation coefficient  
DDM - difference distance matrix  
EMBL - The European Molecular Biology Laboratory  
ER - ensemble refinement  
ESRF - The European Synchrotron Radiation Facility  
GLIC - ligand gated ion channel  
GPCR - G-protein coupled receptor  
GTP - guanosine triphosphate  
GVNDN - gas virtual dynamic nozzle  
HPAD - hybrid pixel array detectors  
HTX - high throughput crystallization laboratory  
IBS - Institut de Biologie Structurale  
LCLS - Linac Coherent Light Source  
LCP - lipidic cubic phase  
MD - molecular dynamics simulations  
MPD - 2-methyl-2,4-pentanediol  
NMR - nuclear magnetic resonance  
PEG - Polyethylene glycol  
RT - room-temperature  
SAXS - small angle X-ray scattering  
SNR - signal-to-noise ratio

SMX - serial millisecond synchrotron crystallography

TLS - translation, libration, screw

WAXS - wide angle X-ray scattering

XFEL - X-ray free electron laser

XRD - X-ray diffraction

## List of Figures

<b>Figure 1.</b> Electromagnetic spectrum; the typical X-ray wavelength for macromolecular structure determination 1 Å marked in red .....	10
<b>Figure 2.</b> Energy loss processes. Region marked in red corresponds to the region typically used for X-ray crystallography .....	11
<b>Figure 3.</b> Left-scheme of a segment of the storage ring, right-scheme of the ESRF .....	13
<b>Figure 4.</b> Comparison of peak brilliance between synchrotrons and x-ray free electron lasers. ....	15
<b>Figure 5.</b> Schematic representation of the structure of amino acid .....	17
<b>Figure 6.</b> Four levels of protein structure. ....	18
<b>Figure 7.</b> Scheme of biological membrane. ....	19
<b>Figure 8.</b> Membrane proteins of known 3D structure at the end of 2016.....	20
<b>Figure 9.</b> Ewald sphere.....	23
<b>Figure 10.</b> The timescale of the motions inside the protein. ....	29
<b>Figure 11.</b> Scheme of the ID13 beamline. ....	33
<b>Figure 12.</b> Example of the raster-scan on lysozyme microcrystals. Red points represent the raster (mesh) scan with chosen step size and stops for the measurement at each step. XYZ axes on the right shows the direction of the motor stages.....	36
<b>Figure 13.</b> Beam signal and flux in raster-scanning mode. Left graph shows the qualitative dependence between beam intensity profile and position of the Y scanning motor. Right-dependence between flux and time.....	36
<b>Figure 14.</b> Left-working scheme of the continuous scan; parameters for quazi-continious Eiger acquisition is 110 frames of 10 ms with 10 µs readout time ; right-simulation of the fast continuous scan and composite image (hitmap) of the droplet with lysozyme crystals reconstructed after the experiment.....	38
<b>Figure 15.</b> Qualitative beam signal and exposure in continuous mode. Left graph shows a dependence between beam intensity profile and position of the Y scanning motor. Right-dependence between flux and time.....	39
<b>Figure 16.</b> An example of the composite image computed after performing a fast continuous scan of the droplets with crystals of lysozyme protein; left – micrograph from the CrystalDirect plates using CRIMS application at EMBL,Grenoble; right – hitmap calculated as number of pixels above threshold (in this case 15 detector counts), white colour means 300 or more pixels. ....	39
<b>Figure 17.</b> Sample delivery using LCP-jet. a) Scheme of the set up. The water (in blue) used to drive the injector. LCP is presented in red, gas in green. b) Crystals of bacteriorhodopsin. c) Set up at ID13 beamline, ESRF. d) Diffraction pattern of bacteriorhodopsin to 2.4 Å resolution.....	41

<b>Figure 18.</b> CrystalDirect plates set-up at ID13. To reach all compartments, the serial stage was added to the existing translational stages. Directions of the movements of the motors are shown on the right. ....	43
<b>Figure 19.</b> An example of the automatically generated input file for semi-automated data collection using CrystalDirect plates (hsize, vsize: defining horizontal and vertical size of the scan area in steps of hstep, vstep; exptime: exposure time per frame; last two columns offset corrections in y and z for the deviation of the droplet or region of interest relative to the nominal position of the compartment; all dimensions are in mm). ....	44
<b>Figure 20.</b> Liquid removal. Left- mobile device for liquid removal with 2 modes of soaking; right- manual process of liquid removal from CrystalDirect plates.....	44
<b>Figure 21.</b> Sample delivery to the X-ray beam using silicon nitride membranes. Left – mounting with still washer displayed step by step; right – mounting with several Si <sub>3</sub> N <sub>4</sub> membranes.....	45
<b>Figure 22.</b> A vapour diffusion crystallization plate adapted for the crystallization on silicon nitride membranes. ....	46
<b>Figure 23.</b> GUI interface of NanoPeakCell.....	47
<b>Figure 24.</b> Diffraction pattern of ubiquitin indexed by CrystFEL .....	52
<b>Figure 25.</b> Scheme of indexing process.....	53
<b>Figure 26.</b> Histogram of indexed parameters on the example of ubiquitin. ....	53
<b>Figure 27.</b> Left-set up of the LCP-jet experiment at ID13, right-5 cm flight tube to reduce air scattering. ....	57
<b>Figure 28.</b> A) $\alpha,\beta$ -tubulin structure. An additional domain on the top is a darphin, which was used to enhance crystallization of tubulin B) refined electron density map of tubulin (chain B) C) refined electron density map of tubulin (alpha helix region).....	59
<b>Figure 29.</b> Left- presentation of the plate; right-visualisation of the drops through the CRIMS online application.....	62
<b>Figure 30.</b> Left-drop with thaumatin crystals 15-20 $\mu\text{m}$ ; right-diffraction of thaumatin to 1.9 $\text{\AA}$ . ....	63
<b>Figure 31.</b> Refined electron density map of thaumatin at 2 $\text{\AA}$ resolution.....	65
<b>Figure 32.</b> Structure of 5-HT <sub>3</sub> receptor solved at cryo-temperature to 3.5 $\text{\AA}$ resolution. 66	
<b>Figure 33.</b> Left-crystals of 5-HT <sub>3</sub> receptor; right-maximum projection of one of the drop. ....	66
<b>Figure 34.</b> Pentameric ligand with BR-lidocaine solved to 2.4 $\text{\AA}$ resolution at 100K. ....	67
<b>Figure 35.</b> Left-crystals of membrane protein GLIC in the range from 5 to 30 $\mu\text{m}$ ; right-diffraction of GLIC to 3.8 $\text{\AA}$ .....	67
<b>Figure 36.</b> Refined structure of GLIC. Left-side view; right-top view .....	68
<b>Figure 37.</b> Left-300 $\mu\text{m}$ crystal of GLIC; right-diffraction of GLIC to 2.3 $\text{\AA}$ . ....	70

<b>Figure 38.</b> Difference in background signal for lysozyme crystals. Left-with liquid; right-liquid removed. ....	72
<b>Figure 39.</b> Refined density maps of lysozyme.....	74
<b>Figure 40.</b> Left-drop with microcrystals of ubiquitin. Right-diffraction pattern of ubiquitin to 2.9 Å.....	75
<b>Figure 41.</b> An example of the composite image obtained after performing a fast continuous scan of the droplets with crystals of ubiquitin; left – micrograph from HTX software from the CrystalDirect infra-structure, right – hitmaps calculated as number of pixels above the threshold .....	76
<b>Figure 42.</b> Electron density of ubiquitin with and without liquid.....	78
<b>Figure 43.</b> Left-drop with microcrystals of lysozyme in the 10-20 μm range; right-diffraction of the lysozyme microcrystals up to 1.6 Å.....	79
<b>Figure 44.</b> Refined density maps of lysozyme with benzamidine ligand .....	81
<b>Figure 45.</b> Left-composite image (hitmap) of a drop after data collection with solution, right- composite image of a drop without solution .....	82
<b>Figure 46.</b> Diffraction pattern of lysozyme with multiple hits from different crystals. ..	83
<b>Figure 47.</b> Nano-beam diffraction pattern of 2 μm crystal of fluorescent protein (rsEGFP2) .....	86
<b>Figure 48.</b> Crystals of AcrB grown directly on the Si <sub>3</sub> N <sub>4</sub> membranes; left-medium size membrane (2.5*2.5 mm <sup>2</sup> ), right-small size membrane (1.5*1.5 mm <sup>2</sup> ). .....	87
<b>Figure 49.</b> Crystals of 5-HT <sub>3</sub> receptor grown in-situ on the silicon nitride membranes and diffraction pattern of 5-HT <sub>3</sub> receptor to 6.4 Å.....	87
<b>Figure 50.</b> Different crystal forms of ubiquitin. A) MPD B) Cubic C) Rods.....	90
<b>Figure 51.</b> Electron density near residues 45-59 for different crystal forms of ubiquitin. ....	91
<b>Figure 52.</b> Distance differences matrix of ubiquitin MPD 100K vs. ubiquitin MPD 297K. ....	94
<b>Figure 53.</b> Distance differences matrix plot for ubiquitin cubic 100K vs. 297K, left-chainA, right-chain B.....	94
<b>Figure 54.</b> Distance differences matrix for rod-shaped ubiquitin 100k vs. 297K, chain A. ....	95
<b>Figure 55.</b> Distance differences matrix for rod shaped crystals of ubiquitin 100k vs. 297K, chain B(left) and chain C(right).....	96
<b>Figure 56.</b> Distance differences matrixes in comparison with ubiquitin rod 297K, chain C. ....	98

<b>Figure 57.</b> Rfree heat map obtained after ensemble refinement for MPD ubiquitin RT. The colour bar shows the values for Rfree with different pTLS, bath and tx parameters. Blue is the lowest values, red is the highest values. ....	104
<b>Figure 58.</b> Porcupine plots after ensemble refinement of MPD crystals of ubiquitin. Arrows are $\beta$ -strands, where the dark blue arrow is $\beta$ 1, light blue- $\beta$ 2, light green- $\beta$ 3, yellow- $\beta$ 4, red- $\beta$ 5. The $\alpha$ helices are green, and the turns are orange. ....	105
<b>Figure 59.</b> Comparison of CT (100K) and RT (297K) structures of MPD ubiquitin crystals in the area between $\beta$ 1 (dark blue) and $\beta$ 2 (light blue) strands. A-original pdb for CT structure; B-original pdb for RT structure; C- ensemble refinement for CT structure; D- ensemble refinement for RT structure; .....	106
<b>Figure 60.</b> Porcupine plots obtained after ensemble refinement of cubic crystals. ....	107
<b>Figure 61.</b> Comparison of CT (100K) and RT (297K) structures of cubic ubiquitin crystals (chain A) in the area between $\beta$ 1 (dark blue) and $\beta$ 2 (light blue) strands. Left- ensemble refinement for CT structure; right- ensemble refinement for RT structure. ....	108
<b>Figure 62.</b> Cubic crystals, area near residue 33, chain B. Left-ensemble refinement at CT structure; right- ensemble refinement at RT structure. ....	108
<b>Figure 63.</b> Porcupine plots obtained after the ensemble refinement of rod-shaped crystals. ....	109
<b>Figure 64.</b> Comparison of CT (100K) and RT (297K) structures of rod ubiquitin crystals in the area near the residues 61-64. Aoriginal pdb for CT structure; B-original pdb for RT structure; C- ensemble refinement for CT structure; D- ensemble refinement for RT structure. ....	110
<b>Figure 65.</b> Domain motions inside rod-shaped crystals of ubiquitin at room-temperature. ....	111

## List of Tables

<b>Table 1.</b> Typical primary beam parameters for serial crystallography experiments at ID13. ....	34
<b>Table 2.</b> Performance parameters of EIGER 4M detector. ....	34
<b>Table 3.</b> ID13 beamline motor parameters. ....	35
<b>Table 4.</b> Data collection parameters and crystallographic refinement statistics for LCP-jet experiment on tubulin. ....	58
<b>Table 5.</b> Data collection parameters and crystallographic refinement statistics of CrystalDirect approach on thaumatin. ....	64
<b>Table 6.</b> Data collection parameters and crystallographic refinement statistics of the membrane protein GLIC grown on CrystalDirect plates. ....	69
<b>Table 7.</b> Data collection, indexing and refinement parameters for the lysozyme with and without liquid removal. ....	73
<b>Table 8.</b> Data collection, indexing and refinement parameters for ubiquitin with and without liquid removal. ....	77
<b>Table 9.</b> Data collection, indexing and refinement parameters of the dataset “Lysozyme with benzamidine” with and without liquid removal. ....	80
<b>Table 10.</b> Parameters of the data processing for ubiquitin crystals. ....	92
<b>Table 11.</b> Alternative conformations in ubiquitin. ....	93
<b>Table 12.</b> RMSD of atomic positions of ubiquitin crystals in comparison with ubiquitin in solution. ....	97
<b>Table 13.</b> Parameters of solvent content, unit cell and protein volume for ubiquitin crystals. ....	99
<b>Table 14.</b> Potential salt bridges and hydrogen bonds inside Ubiquitin. ....	102

## PREFACE

Synchrotron radiation is one of the key factors for the tremendous success of macromolecular crystallography during the past decades. More than 90% of all protein structures in the PDB database were solved by crystallography using synchrotron radiation sources and around 95 % of them were determined from cryocooled crystals<sup>1,2</sup>. In this way, a whole data set can be collected from one flash-cooled crystal. Data collection at cryogenic temperatures drastically reduces radiation damage effects. However, structures determined using cryo-freezing techniques are limited by static nature of frozen crystals.

With the recent development of X-ray free-electron laser facilities (XFELs), which are able to produce extremely intense femtosecond X-ray pulses, the era of “Diffraction before destruction” and serial femtosecond crystallography (SFX) for micro-/nano-sized crystals has begun<sup>3,4</sup>. Employing the SFX technique a protein crystal is exposed to the X-ray beam for tens of femtoseconds before it is completely destroyed. The data are collected from diffraction patterns of many randomly oriented crystals that are exposed to the X-ray beam. In this case only one shot per crystal is possible. In order to obtain a complete data set, new data analysis techniques capable of processing large volumes of diffraction data have been developed. First experiments using the serial femtosecond crystallography (SFX) approach were performed at the Linac Coherent Light Source (LCLS, Stanford, USA) in December 2009 on photosystem I and lysozyme crystals<sup>5,6,7</sup>. At XFEL-facilities crystallographers perform room temperature structure determination of proteins almost free of radiation damage, due to the short femtosecond flashes of XFEL which pass through the sample before significant radiation damage occurs.

After these first successful experiments with the SFX technique, efforts to perform serial crystallography at room temperature on micron-sized crystals started to take place at synchrotron sources. A first attempt to perform synchrotron room-temperature serial crystallography has been done at PETRA III at DESY in Hamburg using glass capillary based microfluidics<sup>8</sup>. This method where thousands of patterns are collected from individual crystals passing through the X-ray beam<sup>9</sup>, was named synchrotron serial millisecond

crystallography (SMX). With the development of room-temperature serial crystallography techniques at the synchrotrons, the X-ray exposure is distributed over a large number of crystals in the sample, which helps compensating for the effect of increased radiation damage at ambient temperature.

Although synchrotron sources most certainly will never reach the same brightness of X-ray pulses like XFELs, they have some advantages. The main advantage of SMX is that several shots per crystal are possible. In particular, there are much less constraints for sample environments and sample delivery because synchrotron radiation is by far less destructive than XFEL. It should be also mentioned that synchrotrons are more widespread worldwide and therefore easier accessible. Currently only two XFEL facilities are open for protein crystallography users (LCLS, USA and SACLA, Japan).

The aim of this thesis is to explore and to implement methods that will allow to collect data using the serial crystallography approach at ID13, the “Microfocus Beamline” at the European synchrotron radiation facility (ESRF, Grenoble, France). In this work several serial techniques based on LCP-injections and solid state support will be described. The goal was to develop synchrotron serial crystallography based on scanning micro-diffraction to demonstrate that synchrotron sources can be used as a routine instrument to perform serial crystallography of proteins. Aspects of sample delivery, data collection and data processing will be discussed.

# CHAPTER 1. INTRODUCTION

## 1.1 X-rays and their interaction with matter

In 1895, while studying the effects of passing an electrical current through gases at low pressure, German physicist Wilhelm Conrad Röntgen accidentally discovered X-rays - highly energetic electromagnetic radiation capable of penetrating most solid objects. Electromagnetic radiation is transmitted in waves or particles at different wavelengths and frequencies. This broad range of wavelengths is known as the electromagnetic spectrum (**Fig.1**). The electromagnetic spectrum is usually divided into seven regions in order of decreasing wavelength and increasing photon energy and frequency. Here, X-ray wavelengths cover the range from 0.01 to 10 nm.

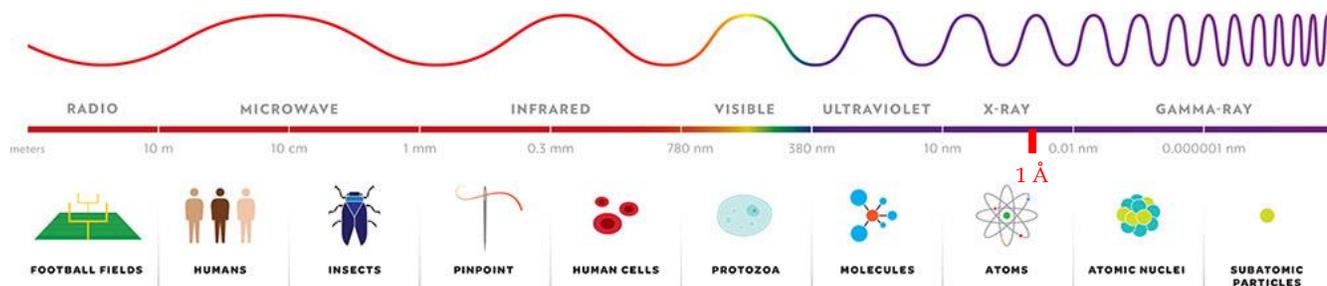


Figure 1. Electromagnetic spectrum; the typical X-ray wavelength for macromolecular structure determination  $1 \text{ \AA}$  marked in red (source <http://teachnuclear.ca/all-things-nuclear/radiation/electromagnetic-spectrum/>).

The typical wavelength used for crystallographic structure determination is around  $10^{-10} \text{ m}$  or  $1 \text{ \AA}$  which is very similar to typical interatomic distances in macromolecules and condensed matter in general. The unit  $\text{\AA}$  (Ångström) is widely used in X-ray crystallography, which corresponds to a photon energy of about 12.4 keV.

When X-rays interact with matter two types of energy transfer can occur:

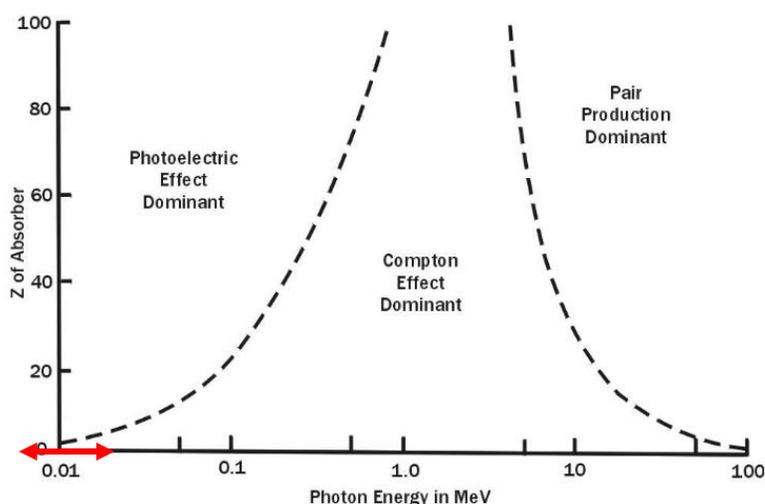
- 1) Ionization
- 2) Excitation

Ionization occurs when the incoming radiation causes the removal of an electron from an atom or molecule. Excitation happens when some of the X-ray energy is transferred to the sample leaving it in a state with higher energy.

The most common energy loss processes of X-rays interacting with matter are:

- 1) Photoelectric absorption
- 2) Compton effect (inelastic scattering)
- 3) Pair production

Which process will take place depends on the mass absorption characteristics of the sample (which is directly related to the atomic weight  $Z$ ) and the energy ( $E$ ) of X-rays (**Fig.2**).



**Figure 2. Energy loss processes. Region marked in red corresponds to the region typically used for X-ray crystallography.**

During photoelectric absorption of X-rays a photon interacts with an atom of the absorbing material and after that the photon completely disappears. This leaves an atom in an ionized state. The ionized atom then returns to the neutral state with the emission of an X-ray photon of an energy characteristic of the atom. Photoelectric absorption is the dominant process for X-ray absorption at relatively low energies with a higher probability for atoms with higher atomic numbers.

The Compton effect is also known as incoherent scattering. Compton scattering occurs between the incident X-ray photon and an electron from an atom in the absorbing material. During this process an X-ray photon of lower energy is scattered from the atom. Compton scattering does not contribute to the diffraction pattern apart from an increase of diffuse background scattering. The contribution of Compton scattering is more important for higher energies but also for samples composed of atoms with very low atomic weight such as biological macromolecules.

Pair production is not relevant for typical diffraction experiments as it is a rare process that occurs when an electron and a positron are created with the annihilation of the X-ray photon, with energies greater than 1.022 MeV.

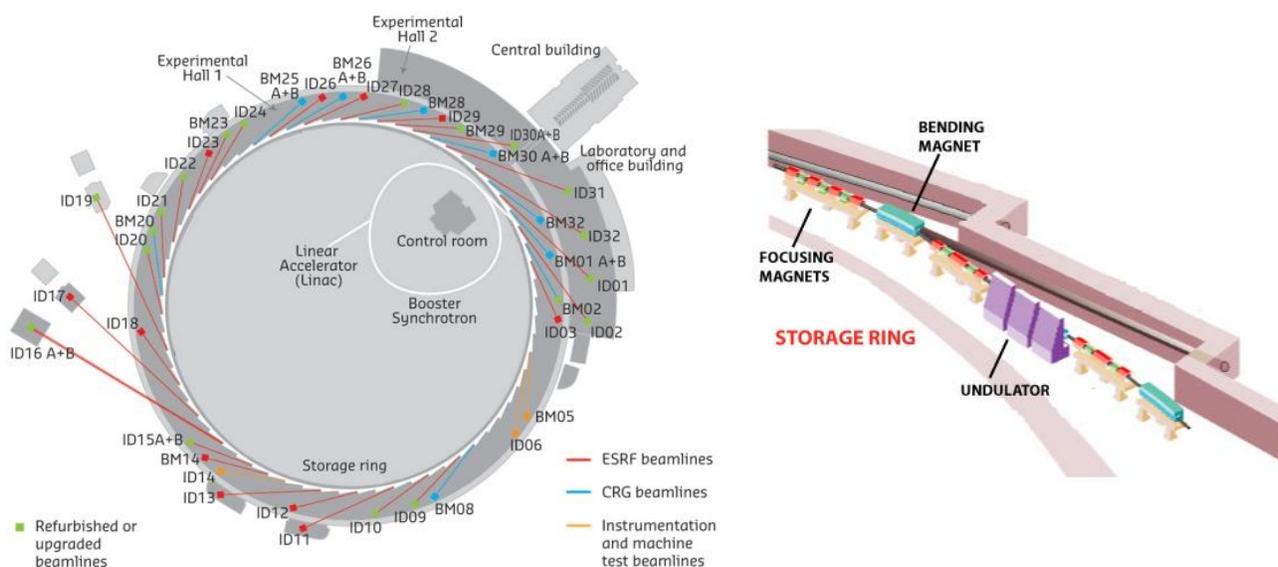
Another process called Thomson scattering or Rayleigh scattering is a coherent elastic scattering process that occurs when the X-ray photon interacts with the whole atom so that the photon is scattered with no change in energy by the scattering atom. Thomson scattering has just a minor contributor to the absorption coefficient due to the fact that scattering occurs without loss of energy. However, this is the primary effect X-ray diffraction is based on (shown in red on Fig.2).

## 1.2 Synchrotron Radiation

Due to the advent of synchrotron radiation sources a breakthrough occurred in physics, biology, medicine and materials science. The idea of the acceleration of relativistic particles inside the magnetic ring appeared from V.Veksler<sup>10</sup> in 1944 and McMillan<sup>11</sup> constructed the first electron synchrotron in 1945, shortly after leading to the discovery of synchrotron radiation at General Electric laboratory by Elder, Gurewitsch, Pollock and Langmuir<sup>12</sup>.

Synchrotron radiation is produced when electrically charged particles (electrons or positrons) are forced to travel at nearly the speed of light on a curved trajectory created by a magnetic field. Injection of the particles, usually electrons, is performed into the storage ring with a linear accelerator (linac) through the booster synchrotron (Fig.3). While electrons travel inside the ring, they pass through different arrangements of magnets, which magnetic field forces causing

particles to change their trajectory. One type of these magnetic structures are so-called undulators, which are made of small magnets that force the particles to follow certain trajectory. Another type of magnets bends particles into their orbit (bending magnets). A broad spectrum of wavelength, high flux, brilliance and stability are the main properties of the synchrotron radiation at state-of-the-art synchrotron sources.



**Figure 3.** Left-scheme of a segment of the storage ring, right-scheme of the ESRF (source <http://www.esrf.eu/about/synchrotron-science/synchrotron>).

The high importance of the synchrotron light sources research has led to the construction of a large number of dedicated synchrotron radiation facilities. Nowadays the most advanced synchrotron facilities are so-called third generation X-ray light sources. One of them is the European Synchrotron radiation facility (ESRF) that was inaugurated in 1994. Electrons for the storage ring are produced by the electron gun in the linac and then these electrons, packed in “bunches”, are accelerated to 2 Mega-electron-volts in the booster synchrotron. The 300 meters booster ring accelerates electrons to the energy of 6 GeV before they are injected into the storage ring. The length of the storage ring is 844 meters. It consists of 32 straight sections with large bending magnets and 32 curved sections (**Fig.3**). The ESRF produces hard X-rays in the energy range from a few keV to hundreds keV.

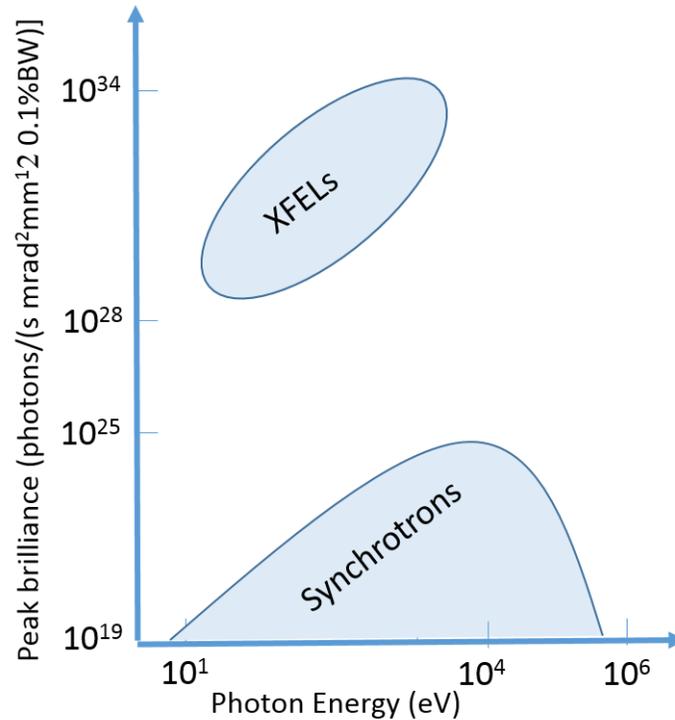
Regarding diffraction experiments, area detectors are key devices at synchrotron sources. For many years detectors have been a limiting factor to perform efficient data collection at synchrotrons. This was due to the fact that progress in improving the performance of X-ray sources was always ahead of detector development. However, detectors of the latest generation operated in single-photon-counting mode have radically transformed many research areas at synchrotron sources since 2006<sup>13</sup>. Hybrid pixel array detectors (HPADs) contain a pixelated readout chip custom-designed for a well-defined technique. HPAD devices are providing a great flexibility because the absorption and signal processing can be optimized independently. These two processes are performed by two separate pieces of material that are connected together by high-density interconnects. Every single pixel contains all necessary electronics including counters for X-ray detection, which increases the overall performance of the HPAD detectors compared to the previous generation of charge-coupled (CCD) devices<sup>14,15,16</sup>.

### **1.3 X-ray free electron lasers (XFELs)**

In 1970 John Madey developed the laser whose lasing medium consisted of very-high speed electrons moving through a magnetic field produced by a periodic arrangement of magnets with alternating poles across the beam path (wigglers and undulators)<sup>17</sup>. Later in 1976 this tunable laser was named - free-electron laser<sup>18</sup>. The principle of work is quite similar to synchrotron sources, but the electron bunches cannot be reused after each turn due to the linear nature of the accelerator. The beam of electrons is accelerated almost to the speed of light but at X-ray energies; the photons interact with the electrons in such a way that they produce extremely bright X-ray laser pulses lasting a few femtoseconds ( $10^{-15}$ ). The motion of molecules appears to be frozen due to the short time of the pulse and can be used to capture snapshots of fast events.

The intensity of each flash at XFEL sources is extremely high, containing about  $10^{12}$  photons in tens of femtosecond duration<sup>3</sup>. Typical beam parameters for a high energy XFEL are  $E=10$  keV, pulse duration = 100 fs, flux per pulse =  $5 \cdot 10^{12}$  cts, beamsize is around  $50 \cdot 50 \mu\text{m}^{19}$ . The peak brilliance of free-electron lasers

exceeds the brilliance of modern synchrotron radiation sources by several orders of magnitude, which makes these sources so interesting for the user community (Fig.4).



**Figure 4. Comparison of peak brilliance between synchrotrons and x-ray free electron lasers.**

The ability to perform experiments with very short laser pulses opened new horizons in the field of structural biology. The very first experiment with proteins at XFEL was performed at LCLS by Chapman et al.<sup>3</sup>. They showed that due to the high intensities of the X-ray free electron lasers, vaporization of the sample occurs. Nevertheless the full destruction of the sample occurs only after the pulse has passed through the sample which makes it possible to perform such measurements at room-temperature<sup>3</sup>.

## 1.4 Serial Crystallography.

X-ray crystallography is an experimental technique based on the fact that X-rays with a wavelength similar to the interatomic distances inside the crystal (around 1 Å) are diffracted by crystals. First X-ray diffraction patterns were obtained in 1912 by Friedrich and Knipping from crystals of diamond, rock salt and zinc sulphide. These experiments confirmed in a fundamental way the atomic constitution of matter and the bonding of the atoms. The first protein structure was solved by X-ray crystallography in the late 1950s, the sperm whale myoglobin by Kendrew<sup>20</sup>. Nowadays, highly detailed molecular structure models of individual proteins or large macromolecular assemblies can be routinely determined by X-ray crystallography.

From the origins of crystallography it is known, that bigger crystals diffract better than small crystals, because the diffracting power relates to the crystal volume. With proteins it is even more challenging, because the unit cells are large, while the electron density is low. Sometimes it can be hard to obtain large crystals, so a decision can be to collect data as a series of diffraction patterns from many small crystals and then to scale and merge the data into one set of structure factors. This method is called serial crystallography. Serial crystallography is usually performed at room-temperature using protein microcrystals with typical size from several microns to 50 µm. Some special requirements should be fulfilled before starting any serial crystallography experiment. The first requirement is a beam size, which should be smaller than the size of the microcrystals to optimize the signal to noise ratio. The second requirement is sample delivery, because there are only a limited number of beamlines that can propose non-standard data collection at room temperature with microcrystals. The third requirement is the need of a fast detector, which can allow the collection of millions of frames in 1 or 2 beamtime shifts. The fourth requirement is efficient data processing software, which can process huge volumes of data with intrinsically unknown orientations of collected crystals. Also sufficient amounts of high quality crystals need to be produced to perform a successful experiment. Many of these problems have been solved due to the tremendous progress in serial crystallography at XFEL sources

and recent development at synchrotron sources. These aspects are under development and constitute major challenges for the near future.

### 1.5 Proteins and membrane proteins

Proteins are the most important chemical compounds of life. They consist of linear chains with amino acids as the building units. The typical size of proteins is in the range of 100-1000 units for a single chain. The general formula of an amino acid is:

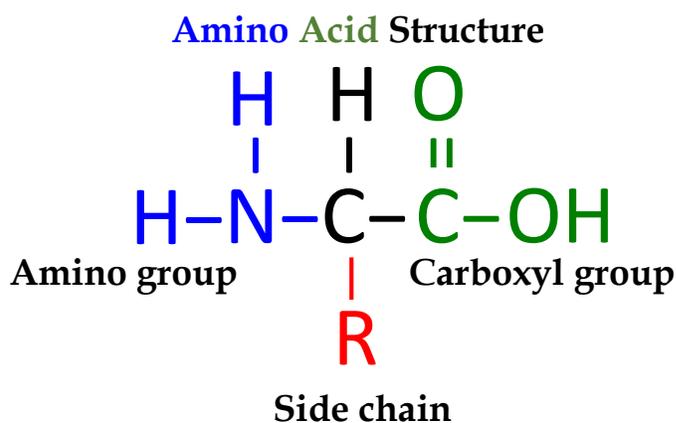
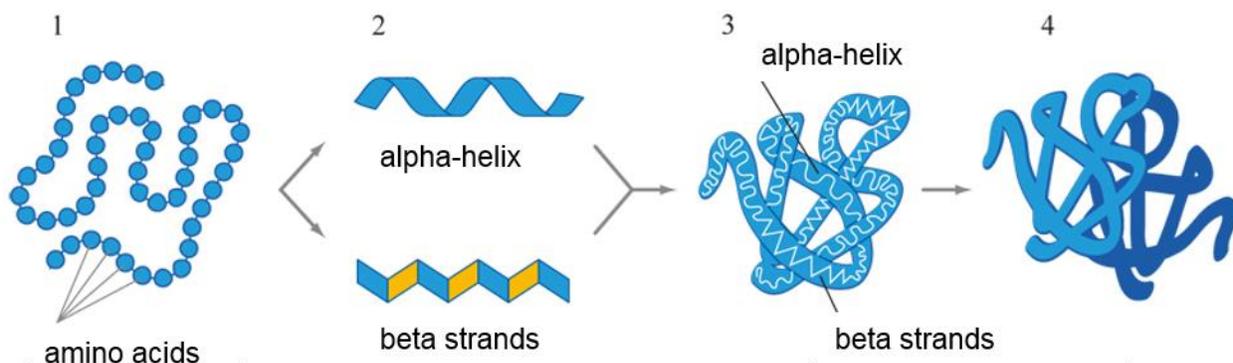


Figure 5. Schematic representation of the structure of amino acid.

There are more than 500 amino acids that can be found in nature, but proteins are only composed of 20 of them. The R group refers to the side chain of one of these 20 amino acids. Side chains can be acidic or basic, hydrophobic or hydrophilic. They also have different size and different functional groups. Hydrophobic amino-acid residues are involved only in van der Waals interactions. The polypeptide chain adopts a 3-dimensional fold that results from various non-covalent interactions between the backbone and the side-chains of all amino acids.

There are four levels of protein structure. The particular sequence of different amino acids along the polypeptide backbone, which is directly determined by the sequence of the nucleotides in the gene, is known as its primary structure. The secondary structure of the polypeptide chain is formed through hydrogen-bonding interactions between N-H and C=O groups and can be folded either as  $\alpha$  helices or  $\beta$  strands. At physiological temperatures in solution many polypeptide

chains of protein fold into a globular form<sup>21</sup>. In this form  $\alpha$  helices,  $\beta$  strands as well as loops that don't have a secondary structure are folded into a tertiary structure. The quaternary structure is a formed by association of already folded chains of more than one polypeptide.



**Figure 6. Four levels of protein structure**

(source <http://eng.thesaurus.rusnano.com/wiki/article561>).

Knowing the 3D structure of a protein is very important because it opens up the possibility of understanding a function of the protein from an analysis of this structure. Proteins can only perform their various functions when they are correctly folded, known as native structure. Protein folding occurs in a complex molecular environment within the cell, which often requires the assistance of molecular chaperones to avoid aggregation or misfolding.

Around 30 % of naturally occurring proteins are predicted to be embedded in biological membranes<sup>22</sup>. Membrane proteins are very important because they carry out the essential functions of the membrane and they represent 70 % of the known and future therapeutic drug targets.

Membrane proteins are embedded in hydrophobic membranes that separate cells, or cellular compartments as organelles, or any organism from their environment. They are responsible for all the communications and provide the biological membrane with biochemical functions including transporters, channels, receptors and signal transducers<sup>23</sup>. Transporters can perform passive or active transport. For example, they are important to generate and maintain osmotic and electrochemical gradients across the membrane. Channels facilitate

passive transport from higher to lower concentration and can be coupled to advanced mechanisms of gating that control the opening and closure of the channel pathways across the membrane<sup>24</sup>. Biological membranes are bilayers of lipid molecules, which consist of hydrophilic head-groups and hydrophobic tails (Fig.7).

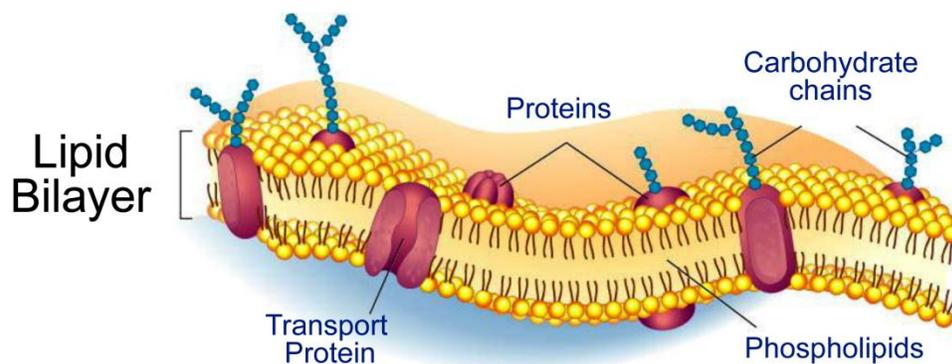
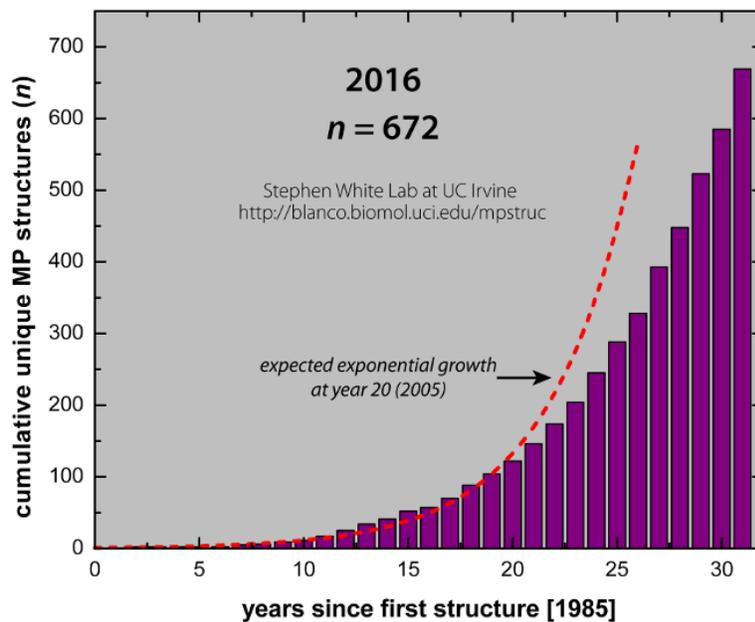


Figure 7. Scheme of biological membrane (source <http://www.acpfg.com.au>).

Very often the structure determination of a membrane protein can be a critical step to be mastered before achieving breakthroughs in life science and medicine, but it is also a notoriously difficult and time consuming procedure. This is mainly due to the fact that the protein surface is relatively hydrophobic and they can only be extracted from the cell membrane with detergents. Also in detergent, they are often flexible and unstable. This leads to challenges at all levels, including expression, solubilisation, purification, crystallisation, data collection and structure determination using standard X-ray crystallography techniques. Typically membrane proteins are expressed at low levels compared to soluble proteins and it is quite difficult to stabilize them outside of a membrane. The crystallization of membrane proteins is very difficult due to the unusual distribution of hydrophobic amino acids on a large part of the protein surface, which slows down the association of the protein in the 3D crystal.

The first membrane protein structure was solved in 1985 by Deisenhofer et al<sup>25</sup>. By the end of 2016, 672 unique structures of membrane proteins were solved (Fig.8) and can be found in the special database<sup>26</sup>.



**Figure 8. Membrane proteins of known 3D structure at the end of 2016**  
(source <http://blanco.biomol.uci.edu/mpstruc/>).

Nevertheless, recent developments of serial crystallography at synchrotrons and XFEL sources including: new sample mounting systems, ultra-fast detectors, and computational infrastructure - enabling processing of large volumes of data; combined with increasing knowledge of purification and crystallization are making it possible to study and to solve new membrane proteins structures.

## **1.6 Structure of the dissertation and motivation.**

Chapter 2 of the dissertation focuses on general methods of X-ray crystallography as well as novel methods some of which were developed during this work. Specific description of the beamline, new scanning modes, experimental set-ups and sample delivery will be presented. Proteins that can be found in this work were not chosen to investigate a specific biochemical problem but rather for their suitability as test systems and their availability (e.g. in the case of membrane proteins). The main focus of the sample delivery was on the solid support due to the technical infrastructure of the beamline, however injector-based experiments were also performed and described. Data analysis was performed for all proteins investigated in this work using a special software for serial crystallography but standard phasing and refinement programs.

In Chapter 3 results of the implementation of three different sample delivery supports will be presented with the example of soluble and membrane proteins. Discussion for each implemented system will be provided. An attempt of crystal growth on the solid support will be presented. Structural studies of protein dynamics at room temperature on the example of ubiquitin protein will be analysed.

In the final Summary & Future perspectives section, further ideas and perspectives on sample delivery supports will be presented. Future perspectives of serial synchrotron crystallography at the synchrotron sources and future upgrade of the ESRF will be discussed.

The motivation for this work was development of sample delivery methods suitable for serial crystallography at synchrotron sources. Another goal was to show to the user community that data collection with microcrystals can be performed in a simple way and that it will not be time consuming. In addition, experiments performed at room temperature can be interesting for investigating the dynamic of the proteins.

## CHAPTER 2. MATERIALS AND METHODS

This chapter comprises three parts: general methods of X-ray crystallography; standard and novel methods adapted for serial crystallography; data processing in serial crystallography. In the first part basics of crystallography and methods that are commonly used for crystallographic experiments will be described. The short description of the protein preparation used in this work will be provided. In the second part, the development of new scanning techniques and implementation of three methods based on LCP-injectors and solid supports will be described. In the third part some aspects of data processing in serial crystallography will be presented.

### 2.1 General methods of X-ray crystallography of the proteins

#### 2.1.1 Diffraction of X-ray by a crystal

The atoms in crystals interact with X-ray waves in a specific way, which leads to interference. The resulting diffraction patterns can be understood as interference of atoms in so-called lattice planes.

The law that predicts constructive interference i.e. the angle of reflection of any diffracted ray from specific atomic planes is known as Bragg's law:

$$n\lambda = 2d\sin\theta$$

n - an integer

$\lambda$  - wavelength of the ray

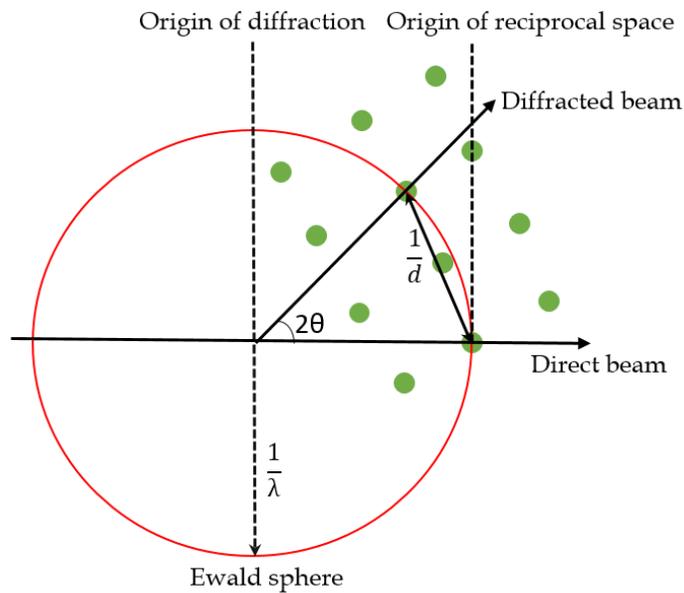
d - interplanar spacing

$\theta$  - the angle between incident wave and the diffracted wave

With crystal reorientation to a different atomic plane, it is possible to measure the d-spacing in different planes. By recording multiple X-ray diffraction patterns

at different crystal orientations, we can determine the crystal structure and the size of the unit cell of the crystal.

A very useful geometric construction to describe the diffraction conditions can be represented by the Ewald sphere. The sphere with radius  $\frac{1}{\lambda}$  demonstrates the relationship between the diffraction angle for a given reflection and the reciprocal lattice of the crystal and for the vectors of the incident and diffracted waves. The Ewald sphere can be used to determine the maximum possible resolution for a given wavelength and unit cell dimension (**Fig.9**).



**Figure 9. Ewald sphere.**

## 2.1.2 Geometry of the crystals

There are 14 different lattice types (*Bravais lattices*), which are determined from 7 crystal systems. They indicate the different symmetry configurations of the simplest repeating unit of the lattice, the unit cell. If the unit cell has only lattice points on the cell corners it is called primitive (P). If it has more lattice points than it is called non-primitive. Non-primitive unit cells can be base-centered (C), Face-centered (F) or Body-centered (I).

Symmetry is a property of a crystal, which is used to describe repetitions of a pattern within that crystal. The combination of these symmetry elements: mirror planes, rotations and inversions is known as the point group. There are 32 crystallographic point groups. The group of symmetry elements observed in crystal is known as its space group. Sometimes crystals can have an additional symmetry based on translation and rotation elements. The combination of translation and rotation axes is called screw axes. Another option is a glide plane, which is a combination of translation and a mirror plane. By combining point groups with Bravais lattices 230 possible space groups can be derived<sup>27</sup>. However, for protein crystals not all 230 groups are allowed. Only crystals without any point symmetry (triclinic) or only with rotation or screws axes are allowed, due to the fact that an application of mirror planes and inversion centers would be incompatible with the chirality of amino acids, which comprise proteins<sup>28</sup>.

### 2.1.3 The quality of the protein crystals

There are several factors that are commonly used to specify the quality of a crystals suitable for X-ray data collection:

- size
- mosaicity
- resolution limit
- radiation damage persistence
- crystal shape

To a very good approximation the intensity of a given diffraction spot is directly proportional to the crystal volume, which means that the size of the crystal has high impact on the X-ray data quality. Larger crystals allow to use shorter exposure times which can reduce the radiation damage effect. In principal the size of the crystal suitable for X-ray data collection has a limit due to the increase of the X-ray absorption. However this is not critical at a wavelength around 1 Å for organic molecules where sample absorption is quite low.

The feature that indicates the perfection of the crystals is mosaicity. The concept of mosaicity assumes that a crystal is build-up of a large number of blocks, with each block angularly slightly misaligned where a perfect crystal does not have such panels and consists only of one block. The resolution limit of the crystal is not necessarily correlated with the mosaic spread. Irradiation of the crystal leads to radiation damage. The mosaicity increases due to the destruction of the intermolecular contacts inside the crystal. Also the resolution limit decreases due to increased disorder i.e. degradation of the spatial correlation between the molecules in the crystal, for example, fusion of side chains.

Resolution is one of the main features that determine quality of the crystals and theoretically it is limited by the wavelength of X-rays (volume of the Ewald sphere), but practically the quality of the obtained crystals determines resolution. More than 70 % of the structures determined by X-ray crystallography crystals don't reach "atomic resolution", due to the fact that disorder inside the crystal is

too high. Also the intensity of diffraction spots decreases with increasing disorder while the intensity of diffuse scattering might rise. The higher the resolution, the more reflections are recorded in the data set. In protein crystallography resolution higher than 1.2 Å counts as excellent, due to the clear visualization of backbones and side chains. A resolution lower than 4.0 Å counts as poor, because only the shape of the protein can be seen from this data.

During X-ray data collection radiation damage of the crystals can occur. Some of the intermolecular bonds inside the crystal can break which increases the mosaicity and crystal disorder. These effects can be reduced via control of X-ray beam parameters. Using of the short wavelength will decrease the absorption of the beam and can reduce radiation damage<sup>29</sup>. Use of faster exposure allows the collection of diffraction data before some effects of radiation damage will occur.

#### **2.1.4 A severe limitation for protein crystallography: radiation damage**

In 1962 Blake and Phillips demonstrated that radiation damage was proportional to the dose received by the crystal<sup>30</sup>. They showed that each 8 keV photon was able to disrupt 70 molecules and to disorder 90 more. Each X-ray photon can emit only one photoelectron, but each photoelectron can result in more than 500 secondary electrons of a lower energy.

Radiation damage can be classified as specific or global. Specific radiation damage appears from inelastic scattering. The following chemical reactions that are caused by the emitted free electrons are known as secondary radiation damage. The main evidence of specific damage is decarboxylation and breakage of disulfide bonds, which can be observed via electron-density maps<sup>31,32</sup>. Global damage appears due to the destabilization of the crystal lattice and can be observed through the decrease in the Bragg reflection intensities, an increase of temperature values (B-factors) and expansion of the unit cell. However changes in the unit cell dimensions are not always caused by the radiation damage. It can be linked to the physical thermal expansion of the protein due to the higher temperatures. In addition, the dependence of unit cell volume and temperature was demonstrated recently by Weik et al<sup>33</sup>.

The mechanism of radiation damage that occurs at cryo-temperatures has already been studied quite well<sup>34,35</sup> where the mechanisms of radiation damage at room-temperatures still remain unclear<sup>36</sup>. The sensitivity of protein crystals are 20-50 times higher, than at 100 K. Warkentin et al. showed that the nature of damage changes dramatically after 180 K<sup>37</sup>. Above this temperature, solvent diffusion is becoming very important and solvent-exposed loops and turns are especially sensitive. However, recently several articles appeared, where researchers didn't find any signs of specific radiation damage at room-temperature measurements<sup>38,39</sup>. Roedig et al. claim that specific radiation damage is less temperature dependent than global radiation damage and generally occurs only at doses higher than 565 kGy<sup>38</sup>. However, global radiation damage to the lattice is a dominating effect that has to be taken into account.

### 2.1.5 Phasing

A loss of information concerning the phase that happens during the measurement causes the intrinsic difficulty of structure determination known as "phase problem". This is due to the fact that the X-ray detector can record intensities but not phases of electromagnetic waves. From these intensities and from the fact that each reflection of the diffraction pattern corresponds to a wave consisting of amplitude and phase, we can derive the amplitudes, but we lose the phase information. Nevertheless, the structure of small molecules can be solved even in the absence of phase information, because phases determine the positions of the peaks of electron density through the unit cell and the presence of a strong diffraction spot alone can give an indication that features must be present with corresponding d-spacing. If the molecule is well ordered, structure factors can be measured to high diffraction angles. So-called direct methods<sup>40</sup> can be applied. In protein crystallography, for the majority of cases, phases are derived by using the atomic coordinates of a similar protein structure (molecular replacement) or by finding the positions of heavy atoms that have been added (MIR, MAD, SAD, etc)<sup>41</sup>.

In this work all structures were solved with the method of molecular replacement<sup>42</sup>. The idea of the method is an attempt to fit the homologous structure into the unit cell of the unknown structure. The molecular replacement calculation involves finding the solution regarding the rotation and translation functions. The known molecule is first rotated in three dimensions, then for each orientation structure factors are calculated from the model. Afterwards, we use the agreement between the calculated structure factors and the observed values to identify the orientation of the known molecule that matches the unknown molecule structure. The oriented model is placed at each possible position in the unit cell and the agreement of the structure factors applied to identify the correct translation. If the correct orientation and translation can be identified, then the model may finally be used to calculate phases for all the structure factors. Electron density maps may be calculated using phases from the model structure and weighted magnitudes from the unknown structure. The resulting map helps to determine the unknown structure.

### **2.1.6 Refinement**

Refinement is the last step of the data analysis. The quality of the electron-density map is very important at this stage. In general, the clarity of electron-density maps depends on the resolution of diffraction data.

Many proteins contain regions of flexibility, which can appear in crystals in the form of static or dynamic disorder. Static disorder results from different conformations adopted by a given structural fragments in different unit cells, where dynamic disorder is the consequence of increased atomic mobility or vibrations of molecular fragments that occur in individual unit cell. The time-scale of these vibrations (Fig.10) is much shorter than the duration of the experiment, so the electron density corresponds to the average distribution of electrons in whole unit cell. In both cases (static and dynamic disorder), the electron density is smeared over different conformational states of the disordered fragments of the structure. At high resolution these fragments can appear as distinct, alternative positions<sup>43</sup>.

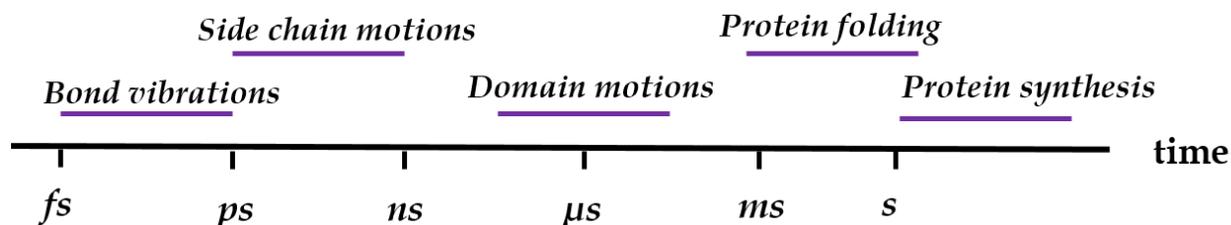


Figure 10. The timescale of the motions inside the protein.

It can be also necessary to use stereochemically restrained refinement of the proteins. The positions of an atom ( $x,y,z$ ) and their B-factors, which indicate thermal motion parameters which are unknown to be solved from the large system of equations. But using the information that proteins are made of amino acids, for which structures that are known at high resolution, we can add information on bond lengths, bond and torsion angles, planarity of the rings and etc. to our equations. The quality of the model can be assessed using **root-mean-square deviation (RMSD)**. The RMSD indicates how much the model departs from standard stereochemistry parameters. If RMSD is higher than  $0.03 \text{ \AA}$ , it indicates that something is wrong in the model.

When a good map has been obtained, a model of the protein structure can be built into it. Nowadays the mostly used software for this purpose is CCP4, Phenix, and Coot. From a structural model we can easily calculate the structure factors. If the model is perfect, the calculated amplitudes of the structure factors  $F_{\text{calc}}$  should be identical with the observed ones  $F_{\text{obs}}$ . To compare calculated and observed amplitudes, we can use the  $R_{\text{factor}}$  ( $R_{\text{work}}$ ):

$$R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

A high  $R_{\text{factor}}$  means that the model explains the data poorly, whereas a low  $R_{\text{factor}}$  means that the model explains the data well.

However, the  $R_{\text{factor}}$  might be biased by the refinement, which aims at bringing  $F_{\text{calc}}$  close to  $F_{\text{obs}}$ . Therefore, another factor that is used for the structure refinement is  $R_{\text{free}}$ .  $R_{\text{free}}$  is similar to the R-factor, but only around 5-10% randomly

selected reflections which have never been involved into model refinement are used<sup>44</sup>. If the difference between R-factor and  $R_{\text{free}}$  exceeds 6-7 %, it may indicate over-fitting of the experimental data or a serious model defect.

### 2.1.7 Proteins used for this study

In this dissertation several proteins were used for the data collection. **Ubiquitin**, **thaumatin** and **lysozyme** were purchased from *Sigma-Aldrich*. **GLIC**, **AcrB** and **5-HT<sub>3</sub> receptor** were purified and crystallized in collaboration with *Aleksandra Woznicka, Hugues Nury, Eva Pebay-Peyroula (IBS, Grenoble) and Marc Delarue (Paris)*. **Tubulin** protein was provided by *Michel Steinmetz, Natacha Olieric (PSI, Switzerland)*. For all vapour diffusion plates growth and in-situ membrane growth the volume of the protein was 1  $\mu\text{l}$  plus 1  $\mu\text{l}$  of precipitant. For CrystalDirect plates all drops consisted of 100 nl of protein plus 100 nl of precipitant.

1. **Lysozyme** was solubilized in distilled water at the following concentrations: 20, 40, 60 and 80 mg/ml.

Crystallization of lysozyme was performed using:

a) Standard vapour diffusion set up (24-well plate) with following precipitant conditions: 0.6 – 1.1 M NaCl (*Sigma Aldrich*), 0.1 M NaAcetate pH 4.2-4.8 (*Sigma-Aldrich*)

b) Crystal Direct (CD) plates with following precipitant conditions: 3.1-3.4 NaCl, 0.1 M NaAcetate pH 4.6, 10-15.5% w/v PEG 6000 (*Sigma-Aldrich*)

c)  $\text{Si}_3\text{N}_4$  membranes with the same crystallization condition as in a)

2. **Bovine ubiquitin** was solubilized in distilled water at the following concentration: 20 mg/ml. Microcrystals of ubiquitin were obtained using the batch crystallization method, where first we determined which concentration of

precipitant is required to obtain spontaneous crystal growth by direct mixing and then we found the conditions which yielded optimal crystal morphology and size.

a) MPD crystals

Ubiquitin was solubilized at 40 mg/ml in 20mM ammonium acetate at pH 4.3 to reach a final protein concentration of 20 mg/ml. The reservoir was composed of 20 mM Citrate Acid with pH 4.1, 60% v/v MPD (2-Methyl-2,4-pentanediol), hexylen glycol was mixed with an equal volume of ubiquitin solution.

b) Cubic crystals

The protein was solubilized at 100 mg/ml in 150mM HEPES pH 7.4, 150 mM NaCl to reach a final protein concentration of 20 mg/ml. The batch was composed of 100 mM MES with pH=6.3, 15% v/v PEG (polyethylen glycol) 3350, 100 mM zinc acetate. The correct volume of each compound was adjusted from the stock solutions to reach the final condition.

c) Rod crystals

20 mg/ml of ubiquitine was solubilized in 20mM ammonium acetate with pH=4.3. Final concentration of the protein was 20 mg/ml. The reservoir composed of 100 mM MES, pH 6.3, 23% w/v PEG 3350, 1 mM ZnAcetate was mixed with an equal volume of ubiquitin solution.

These crystals are named further as MPD, cubic and rods, due to the morphology of the crystals and to the precipitant agents following the nomenclature found in P. Ma et al.<sup>45</sup>.

**3. Thaumatin** was solubilized in distilled water at the following concentrations: 20, 40, 60 and 80 mg/ml. The crystallization of thaumatin was performed by using CrystalDirect plates at EMBL crystallization platform with the following conditions: 1 M Na/K tartrate (*Sigma-Aldrich*) and 0.1 M HEPES pH

7.5 (*Sigma-Aldrich*). Each drop consisted of 100 nl of protein and 100 nl of precipitant.

**4. AcrB** protein was expressed and purified by the IBS team according to the protocol described by Murakami et al.<sup>46</sup> The protein solubilized in DDM was concentrated up to 12 mg/ml, frozen and stored at -80 °C. The crystallization of AcrB was performed using standard vapour diffusion set up (24-well plate), CrystalDirect plates and Si<sub>3</sub>N<sub>4</sub> membranes. Crystals were grown at 20°C in two different conditions:

- 1) 8% w/v PEG 3350 (*Sigma-Aldrich*), 0.1M MgCl<sub>2</sub> (*Sigma-Aldrich*), 0.1M HEPES pH 8.0 (*Sigma-Aldrich*)
- 2) 20% v/v PEG 2000 (*Hampton Research*), 0.2 M KSCN (*Sigma-Aldrich*), 0.1M HEPES pH 7.5 (*Sigma-Aldrich*).

**5. GLIC** protein was kindly provided by *Marc Delarue* (Institut Pasteur, Paris).<sup>47</sup> The protein solubilized in DDM was concentrated to 10 mg/mL and stored at -80 °C. Crystallization was performed using standard vapour diffusion set up (24-well plate), CrystalDirect plates and Si<sub>3</sub>N<sub>4</sub> membranes. Crystallization conditions were: 10-16% w/v PEG 4000 (*Sigma-Aldrich*), 0.1 NaAcetate pH 4.0, 0.4 M NaSCN (*Sigma-Aldrich*).

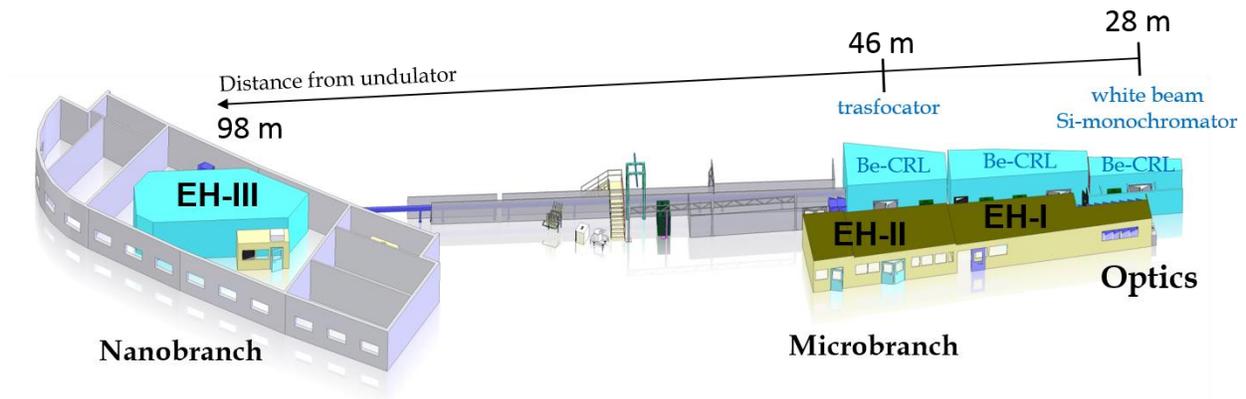
**6. 5-HT<sub>3</sub> receptor** solubilized in Cymal 6 was concentrated up to 7-8 mg/mL and mixed with the VHH15 nanobody for crystallization trials. Crystallization of 5-HT<sub>3</sub> protein was performed using standard vapour diffusion set up (24-well plate), CrystalDirect plates and Si<sub>3</sub>N<sub>4</sub> membranes. The crystallization starting conditions were: 21% PEG 10 000 w/v (*Sigma-Aldrich*), 0.1 M Na<sub>2</sub>SO<sub>4</sub> (*Sigma-Aldrich*), 0.1 M Tris pH 8.0 (*Sigma-Aldrich*). The plates were stored at 12 °C .

## 2.2 Methods adapted to serial crystallography

### 2.2.1 ID13, beamline parameters and scanning modes

Nowadays only a few beamlines in the world have succeeded in providing high-flux micro- and nanobeams. ID13 is a diversified beamline that provides ESRF users with intense micro-beams (microbranch, EH2) and nano-beams down to the 100 nm range (nanobranch, EH3). The scheme of the ID13 beamline is presented on **Fig.11**.

The primary source of ID13 is an 18 mm period in-vacuum undulator optimized for 12-13 keV energies. However, it is also possible to use a fully-tuneable undulator with 35 mm period for the energy range from 5 to 17 keV. Berillium-based compound refractive lenses (CRL-lenses) consist of 60 single lenses and generate the micro-beam. Silicon-based linear compound refractive lenses in crossed geometry (NFL-lenses) generate the nano-beam. The monochromatisation is performed with liquid nitrogen cooled Si-111 double monochromator.



**Figure 11.** Scheme of the ID13 beamline.

All experiments described in this thesis were performed at the nano- and microfocus ID13 beamline. The typical parameters of the beamline are presented in **Table 1**. Beam parameters for each experiment can be found in the results chapter in the result table for each dataset.

Parameter	Micro-beam	Nano-beam
Beam-size, $\mu\text{m}$	2*2	0.15*0.15
Beam divergence, mrad	0.3-1	0.3
Wavelength, $\text{\AA}$	0.954	0.832
Beam flux, photons $\text{s}^{-1}$	$2.0 \cdot 10^{12}$	$1.7 \cdot 10^{10}$
Flux density, photons $\text{s}^{-1} \text{mm}^{-2} \text{mrad}^{-2}$	$2.7 \cdot 10^{22}$	$6 \cdot 10^{24}$

**Table 1.** Typical primary beam parameters for serial crystallography experiments at ID13.

The Microfocus Beamline ID13, ESRF is equipped with a hybrid photon counting detector Eiger 4M (DECTRIS). Each incoming photon is individually counted and processed by the readout electronics in the pixel. The main advantages of the Eiger 4M detector are:

- small pixels and point-spread function for highest spatial resolution
- absence of readout noise and dark current for best signal-to-noise ratio
- continuous readout and frame rates close to the kilohertz range for fast data collection
- high count rates per unit area match the beamline brightness
- comparatively compact size of the device for easy manipulation at the beamline

The main parameters of the Eiger 4M detector are presented in **Table 2**:

Sensitive area, width*height ( $\text{mm}^2$ )	155.2*162.5
Pixel size ( $\mu\text{m}^2$ )	75*75
Inactive area (%)	5.6
Maximum frame rate (Hz)	750
Readout time	Continuous readout, 10 $\mu\text{s}$ dead time
Sensor thickness ( $\mu\text{m}^2$ )	450
Maximum count rate (cts/sec/pixel )	$2 \cdot 10^{12}$

**Table 2.** Performance parameters of EIGER 4M detector.

For presentation to the X-ray beam, the sample mounts were placed onto a xyz translation table (microbranch) or onto a miniature hexapod combined with a xyz piezo stage (nanobranch). For the piezo stages currently only step scanning mode is available, however the overhead is comparably short, 50 ms. The parameters for the relevant stages installed at nano- and microbranch of ID13 are presented in **Table 3**. Beam parameters and data processing parameters for each experiment can be found in the table for each dataset in the chapter with results.

Parameters	Microbranch			Nanobranch		
Motor name	X	Y	Z	X	Y	Z
Overall travel range, mm	160	100	25	0.25	0.25	0.25
Min. incremental steps, nm	200	500	100	10	10	10
Max speed, mm/s	10	5	10	N/A	N/A	N/A

**Table 3. ID13 beamline motor parameters. At the microbranch sample centring and positioning are performed with translation stages; at the nanobranch piezo stages are used for scanning.**

In this work data collections were performed using 2 different scanning modes: raster scanning and fast continuous scanning. **Raster-scanning** serial protein crystallography is an alternative approach for the data collection of high-resolution structural data from micro-sized crystals. At ID13 beamline the principal scanning stages (Micos, Germany) at the microbranch can move the sample in three directions ( $x,y,z$ ). The user chooses the step size in  $y$  direction and the distance between lines in  $z$  direction (Fig.12). This is a non-continuous scan where the translation stage stops at each point. This type of scanning is well suitable for densely packed microcrystals, in particular when high spatial resolution is required.

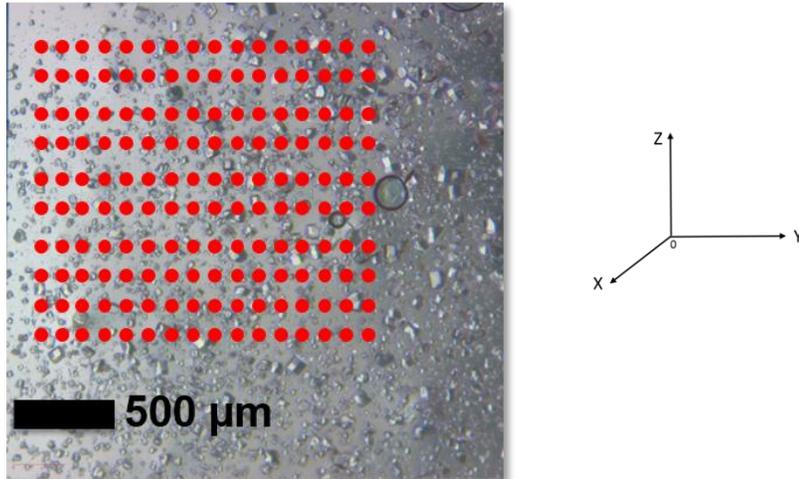


Figure 12. Example of the raster-scan on lysozyme microcrystals. Red points represent the raster (mesh) scan with chosen step size and stops for the measurement at each step. XYZ axes on the right shows the direction of the motor stages.

Beam and sample position are synchronized by default with the exposure as the motor stops after each point. The dose rate at a given position is constant during exposure of one frame (Fig.13).

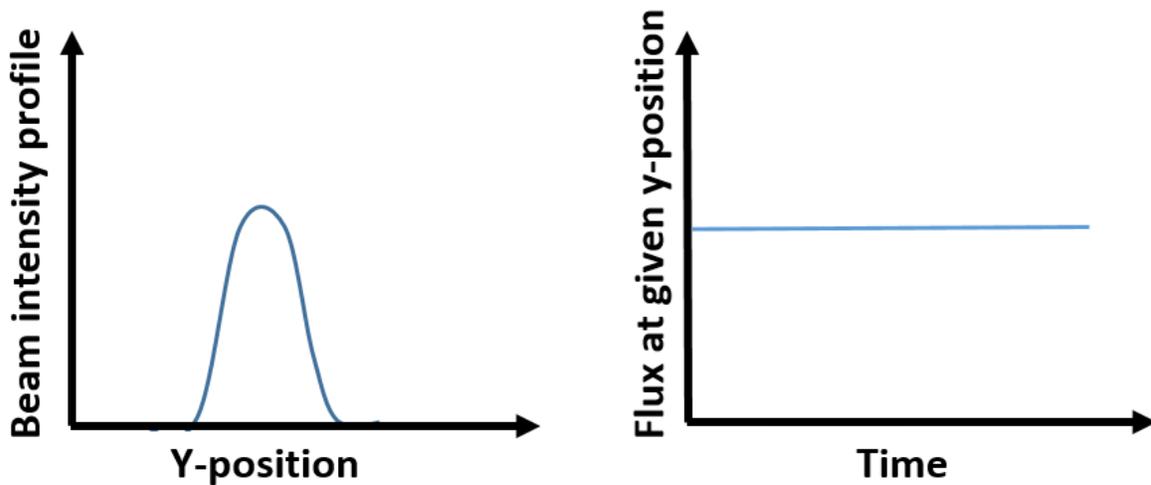
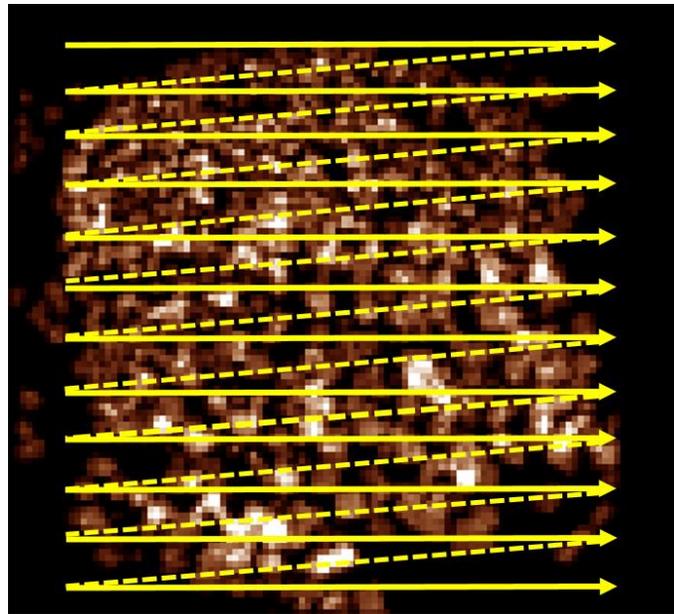
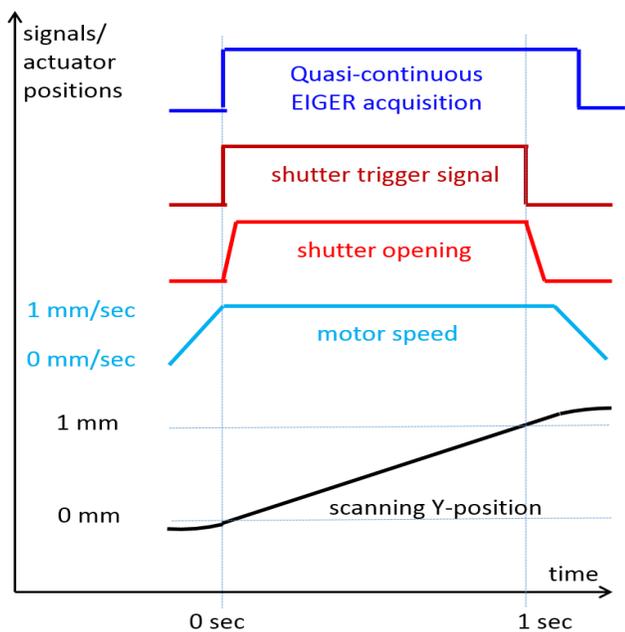


Figure 13. Beam signal and flux in raster-scanning mode. Left graph shows the qualitative dependence between beam intensity profile and position of the Y scanning motor. Right-dependence between flux and time.

**Fast continuous scanning.** During the development of the CrystalDirect approach at ID13, it became obvious that the data collection speed is limited mainly due to the non-continuous movement of the motors. To overcome this bottleneck and improve the performance of the data collection, a fast continuous scan was developed. In this collection mode the measurement starts at the beginning of a scan line. The data collection process continues until the end of the line and then the stage moves to the beginning of the next line, where this process is repeated (**Fig.14**). The detector operates asynchronously with the movement of the scanning stage. A P201 timer card (ESRF design) sends the trigger signal (TTL-signal) to the millisecond fast shutter and the EIGER detector. The shutter is automatically closed at the end of the trigger signal. However, the detector continues collecting the data for an excess time (usually 0.1 sec) in order to bracket the exposure for diagnostic purpose. Therefore a number of dark frames can be observed at the end of each line. The X-ray shutter is open during the data collection and closed while moving between two lines. On the hardware level the scanning motors work independently from the detector, trigger and shutter. The synchronization is purely software based. The scan parameters such as number of frames per line and excess movements for acceleration and deceleration of the scanning stages are calculated automatically by software adapted to the beamline control program. Inputs required by the user are the size of the scan, the spatial step width, and the exposure time.

During the scan of the plate, it is possible to start the computation of a composite image of the droplet to have an understanding about number of crystals and number of hits (**Fig.14**). The color scheme applies depending on the background signal. White area shows crystals that have hits; black-where signal is similar to the background cutoff.



**Figure 14.** Left- working scheme of the continuous scan; parameters for quasi-continuous Eiger acquisition are 110 frames of 10 ms with 10  $\mu$ s readout time ; right-simulation of the fast continuous scan and composite image (hitmap) of the droplet with lysozyme crystals reconstructed after the experiment.

For continuous scanning the beam position on the sample is moving during the exposure of one frame. Therefore the beam foot print can be obtained by convolution of a box-function modelling the movement with a Gaussian modelling the beam profile (Fig.14) i.e. the box width is the motor displacement during one frame collected by the EIGER detector. With a minimum EIGER frame time of 1.33 ms the readout time of 10  $\mu$ s can be neglected. Currently the maximum velocity is 5 mm/s. However, this is not a fundamental limitation. If required, a velocity could be achieved with moderate efforts in the order of cm/sec. The time dependence of the flux at a given y-position corresponds to the shape of the beam profile (Fig.15).

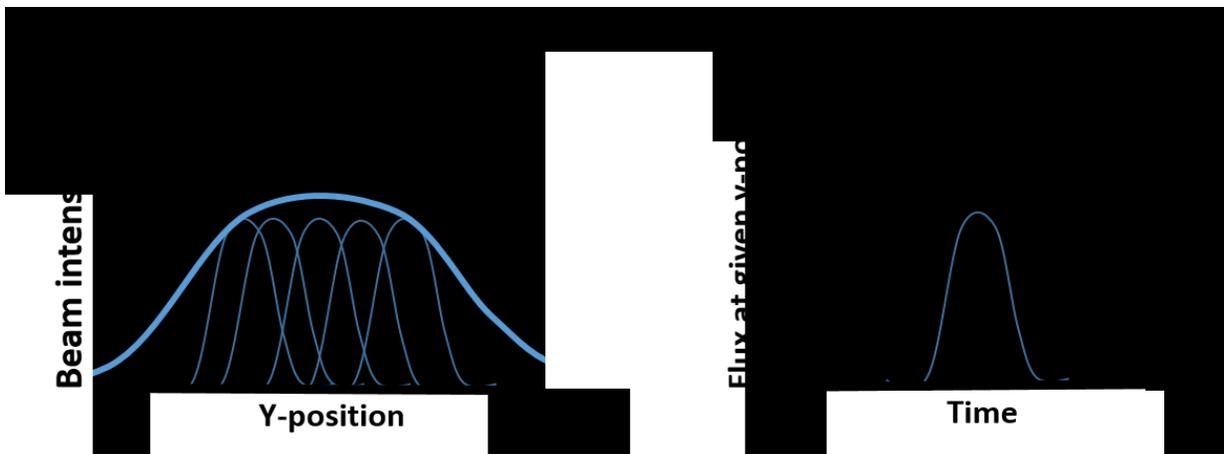


Figure 15. Qualitative beam signal and exposure in continuous mode. Left graph shows the dependence between beam intensity profile and position of the Y scanning motor. Right-dependence between flux and time.

An example of the composite image and corresponding drop are shown in Fig.16. The colour scheme for the hitmaps is based on the number of pixels that are above the background threshold, which are individual for each dataset.

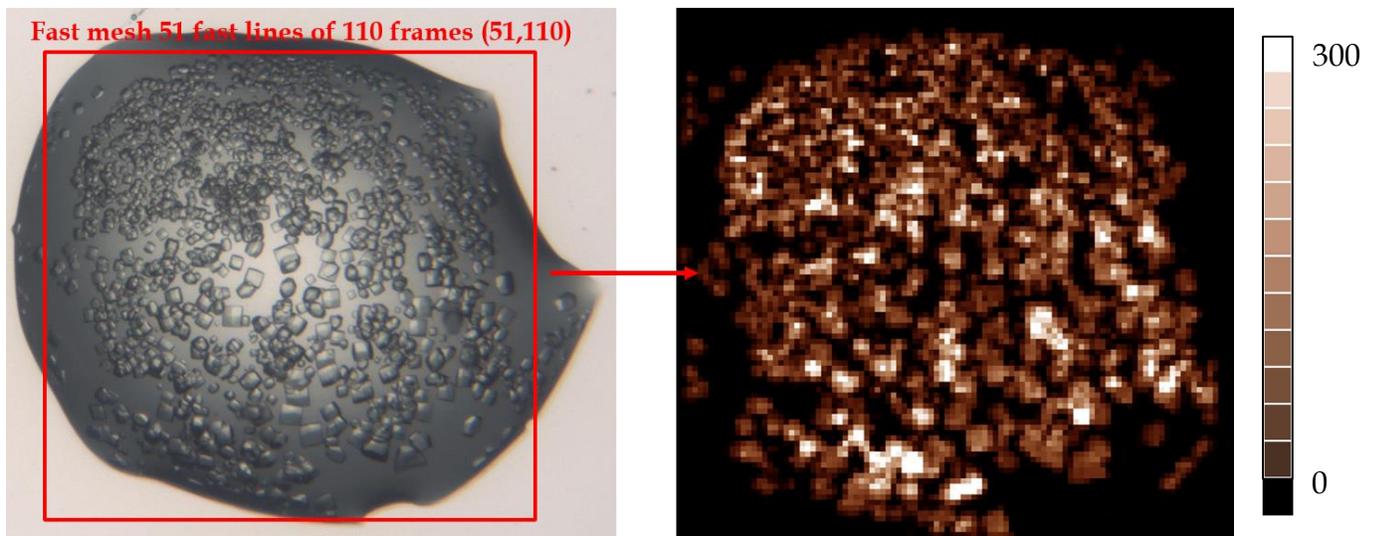


Figure 16. An example of a composite image computed after performing a fast continuous scan of the droplets with crystals of lysozyme protein; left - micrograph from the CrystalDirect plates using CRIMS application at EMBL,Grenoble; right - hitmap calculated as number of pixels above threshold (in this case 15 detector counts), white colour means 300 or more pixels.

## 2.2.2 Sample delivery

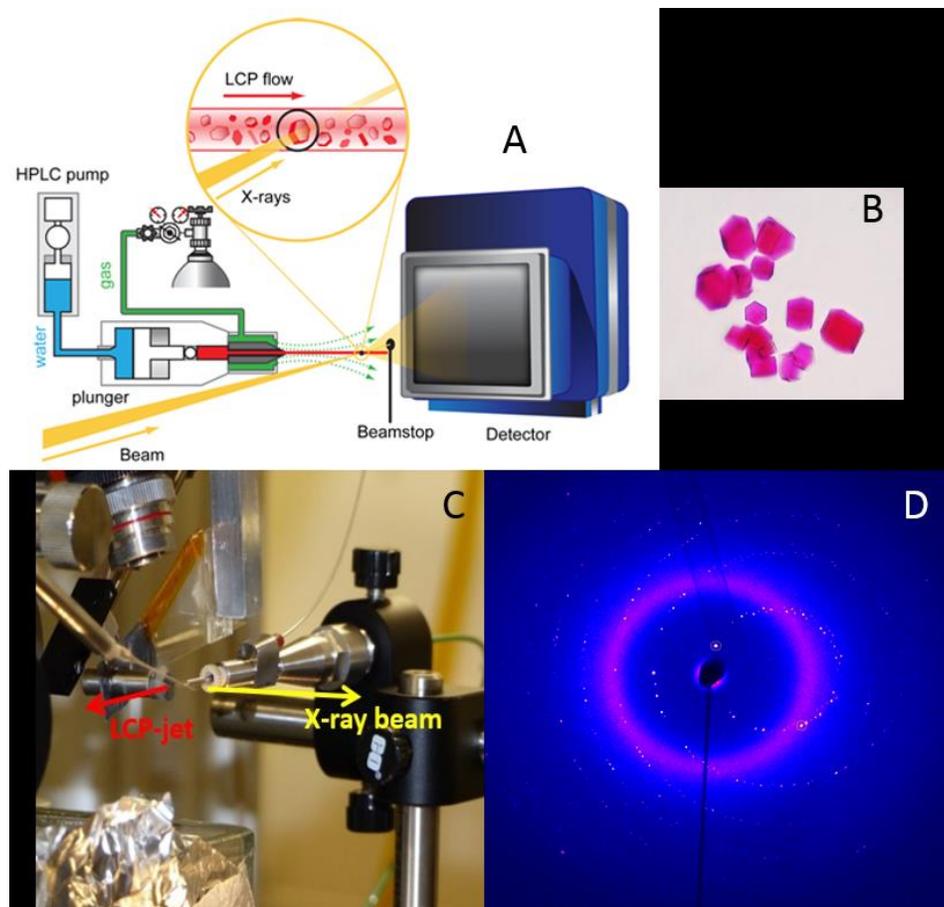
Three different sample delivery approach using two different methods have been used for serial crystallography at ID13:

- LCP-injector
- solid support on the base of silicon nitride membranes
- solid support on the base of CrystalDirect plates

The use of silicon nitride membranes for room temperature solid support serial scanning diffraction was originally developed and tested at ID13. Development and implementation of fast serial crystallography using CrystalDirect plates combined with micro-beam and liquid removal was developed in collaboration with EMBL Grenoble, France. The LCP-injector was kindly provided by a group of *J. Standfuss* from PSI, Switzerland.

### 2.2.2.1 LCP-injector experiments

The first experiment with the LCP-jet injector and a membrane protein at the synchrotron was performed in 2014. The LCP-injector was installed horizontally to the X-ray beam (**Fig.17**). Through the 50  $\mu\text{m}$  nozzle a constant LCP flow of 20-60 nl/min and a flow speed of 100  $\mu\text{m}/\text{sec}$  was stabilized by a flow of helium. A Rayonix MX-170 CCD-detector was used to collect diffraction patterns with 10-50 ms exposure times at a 10-17 Hz frame rate (**Fig.17**). The protein used for this experiment was bacteriorhodopsin, due to the already known structure, and diffraction to high resolution (**Fig.17**). Only 0.8mg of protein was needed to collect full data set. 1343092 images were collected, from which 12 982 had hits and 5691 were indexed with CrystFel. The hit rate was limited by the density of crystals in the LCP for the preparation available at that time. Series of consecutive diffraction patterns were recorded. The resolution limit of the diffraction signal was judged as 2.4  $\text{\AA}$ .



**Figure 17. Sample delivery using the LCP-jet. a) Scheme of the set up. The water (in blue) used to drive the injector. LCP is presented in red, gas in green. b) Crystals of bacteriorhodopsin. c) Set up at ID13 beamline, ESRF. d) Diffraction pattern of bacteriorhodopsin to 2.4 Å resolution.**

A limitation of this experiment was the low acquisition rate (10-17 Hz) as compared to standard 120 Hz at LCLS<sup>48</sup>. The main advantages of this experiment were low sample consumption, ability to collect 2 or 3 consecutive hits per crystal and the speed of the flow compatible with the data acquisition performance.

### 2.2.2.2 CrystalDirect plates

A specially designed set up was developed at ID13 for this project to enable in-situ room temperature plate scanning experiments.

**Set-up at ID13, ESRF.** The CrystalDirect plate was installed in a metallic support frame specifically designed for this purpose. A CD-plate can be conveniently plugged (**Fig.18**). To enable low- $q$  data collection a Helium-filled flight tube can be installed, it reduces air scattering while permitting a larger beamstop distance. With a xyz-motorized microscope installed on the scanning diffraction setup it is possible to select sample positions (crystals, droplets) and design automatic data collection schemes. The sample observation microscope with 5x and 50x magnification is used on axis with the beam direction i.e. oriented upstream. It has to be retracted during data acquisition, leaving space for the flight-tube and the detector.

Three high performance translation stages (strx, stry, strz) move the sample in  $x, y$ , and  $z$  ( $x$ : beam direction,  $y$  horizontal translation perpendicular to the beam,  $z$ : vertical translation). The minimum incremental steps of these stages are 0.2, 0.5 and 0.1  $\mu\text{m}$  respectively (**Table 3**). The fast scanning motor is stry with currently a maximum speed of 5 mm/s. To be able to conveniently reach each compartment of an entire crystallization plate two additional stages (sery, serz) were mounted on top of strz. They are used as range-extendors rather than scanning stages to be able to reach all compartments of the plate.

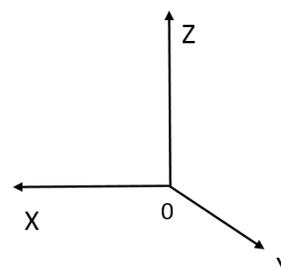
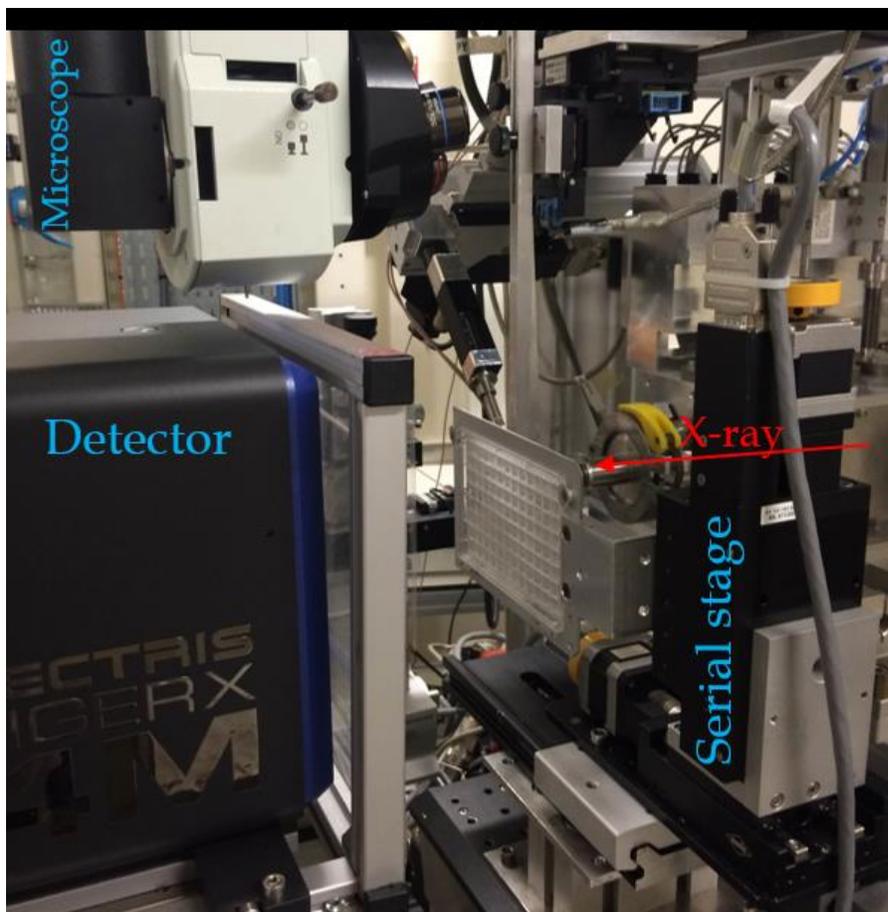


Figure 18. CrystalDirect plates set-up at ID13. To reach all compartments, the serial stage was added to the existing translational stages. Directions of the movements of the motors are shown on the right.

Data collection can be performed in an automated (scan of the full plate) or semi-automated (user selection) way. In the semi-automated mode, the user can specify a list of compartments to be examined. Then the compartments are sequentially moved into the optical field of view by the control software. At each droplet position the user is prompted for input regarding the desired scanning area, step size, and exposure time. Default parameters can be used by on a single key-press which helps making this an efficient procedure. Individual optical examination of 96 droplets typically takes 15 minutes. All parameters are written in a text file and can be always corrected afterwards. An example of such an input file is shown on **Fig.19**, where **hsize** and **vsize** are horizontal and vertical dimensions of the drop; **hstep** and **vstep** is a step size for the y and z motors respectively and **exptime** is a chosen exposure time. The two last columns show

the realistic positions of the centre of the desired scan area. Since the whole plate is delivered to the X-ray beam, this method eliminates the step of “crystal fishing” i.e. avoiding the crystal mounting process.

#		hsize	vsize	hstep	vstep	exptime		
2	A 6 :	1.000	1.000	0.0100	0.0200	0.0100	-0.00350	-0.21690
3	A 7 :	1.000	1.000	0.0100	0.0200	0.0100	-0.03600	-0.18270
4	A 8 :	1.000	1.000	0.0100	0.0200	0.0100	-0.06300	-0.18370
5	A 9 :	1.000	0.800	0.0100	0.0200	0.0100	0.01600	-0.51160
6	A 10 :	1.000	1.000	0.0100	0.0200	0.0100	0.05650	-0.32450
7	A 11 :	1.000	1.000	0.0100	0.0200	0.0100	-0.07650	-0.37830
8	B 6 :	1.000	1.000	0.0100	0.0200	0.0100	0.18850	-0.03670
9	B 7 :	1.000	1.000	0.0100	0.0200	0.0100	0.17900	-0.09960
10	B 8 :	1.000	1.000	0.0100	0.0200	0.0100	0.27800	-0.20860

Figure 19. An example of the automatically generated input file for semi-automated data collection using CrystalDirect plates (hsize, vsize: defining horizontal and vertical size of the scan area in steps of hstep, vstep; exptime: exposure time per frame; last two columns offset corrections in y and z for the deviation of the droplet or region of interest relative to the nominal position of the compartment; all dimensions are in mm).

Liquid removal was performed using a special device developed at the HTX-lab, EMBL, Grenoble. It is based on a user-controlled pump which sucks the precipitant from the drop. At the beginning of the procedure a hole is pierced with a small needle in the covering foil of the chosen well to reduce the inside pressure and to ease liquid removal. Then a second hole is created just above the drop where liquid removal occurs. This device has 2 different pressure regimes, depending on the viscosity of the drop (Fig.20).

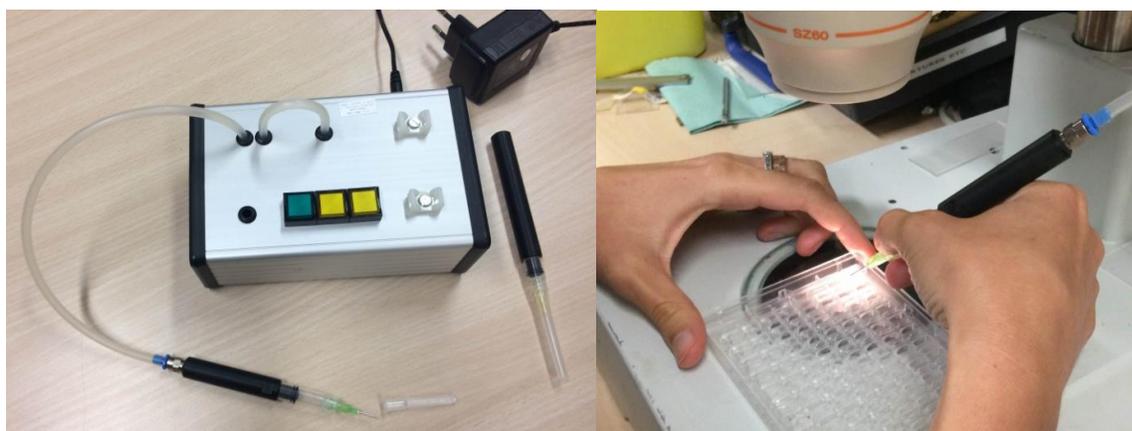


Figure 20. Liquid removal. Left- Mobile device for liquid removal with 2 modes of soaking. Right- Manual process of liquid removal from CrystalDirect plates.

### 2.2.2.3 Silicon nitride membranes

Silicon nitride membranes were purchased from Silson<sup>49</sup>. The size of the membrane windows was 1.5\*1.5 mm<sup>2</sup> (small), 2.5\*2.5 mm<sup>2</sup> (medium), and 5\*5 mm<sup>2</sup> (big). The thickness of membranes used was 500 or 1000 nm. The amount of sample that is required for one medium silicon nitride membrane is typically 500 nl but this can vary. A first successful attempt to collect data from silicon nitride membranes using lysozyme microcrystals was performed in collaboration with Nicolas Coquelle & Jacques-Philippe Colletier<sup>39</sup>. For subsequent experiments we used a variation of this protocol, attaching the membranes directly to the glass support using the following protocol:

- 1) The first set of silicon nitride membranes (base membranes) were glued on the surface of a steel washer (**Fig.21, left**) or a glass slide (**Fig.21, right**)
- 2) Sample droplets were deposited onto these base membranes
- 3) A second set of silicon nitride membranes (cover membranes) was gently deposited on the base membranes (to create so called “membrane-sandwiches”)
- 4) The sandwiches were sealed with Araldite resin to avoid drying of the sample

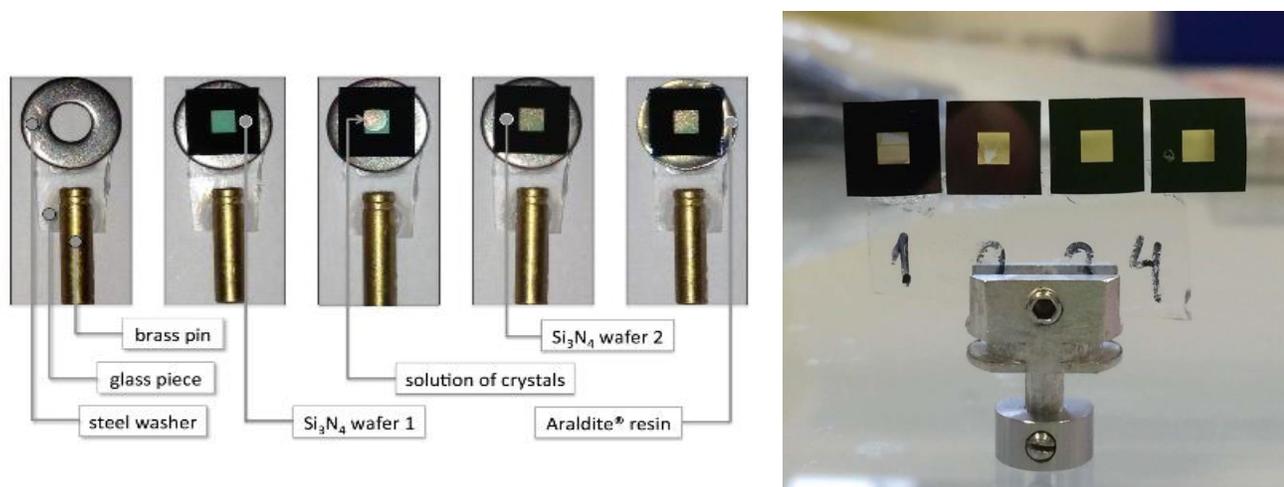
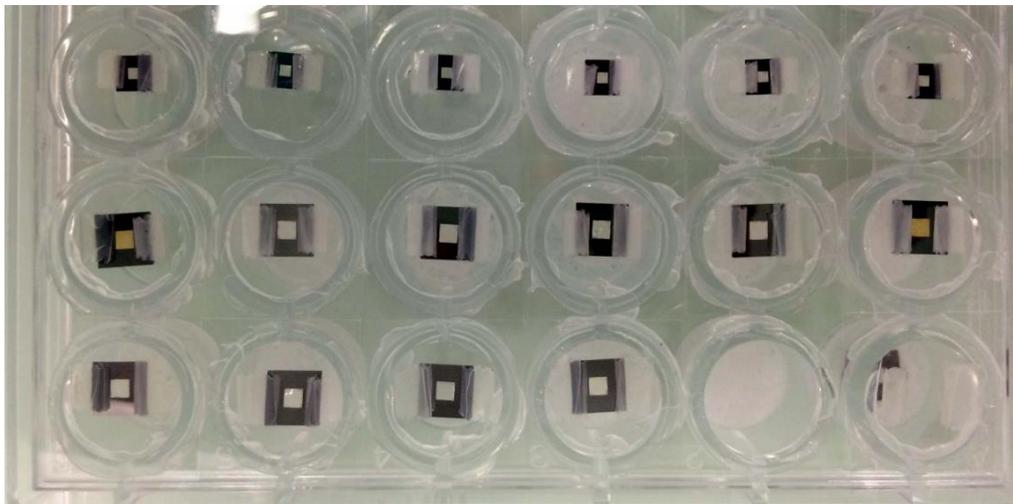


Figure 21. Sample delivery to the X-ray beam using silicon nitride membranes. Left - mounting with steel washer displayed step by step (courtesy N.Coquelle<sup>39</sup>); right - mounting with several Si<sub>3</sub>N<sub>4</sub> membranes.

If protein microcrystals are fragile or scarce, it is very important to avoid unnecessary movements or transfer of the sample. We made an attempt to grow membrane proteins (5HT<sub>3</sub>-receptor and GLIC) on the silicon nitride membrane that can be used straight after their growth for the X-ray data collection (in-place crystal growth). We took a hanging drop vapor diffusion plate, filled a 1ml reservoir with precipitant, mounted membranes with tape to the coverslip and pipetted 1 µl of protein and 1 µl of reservoir to the silicon nitride membranes. Then we carefully inverted the cover slide and put it onto the bead of grease (Fig.22). Crystals were grown in 3 days at 12 °C.



**Figure 22.** A vapour diffusion crystallization plate adapted for the crystallization on silicon nitride membranes.

## 2.3 Data processing in serial crystallography

### 2.3.1 Sorting the serial crystallography data (NanoPeakCell)

During any crystallography experiment three main steps of the data analysis should be performed: indexing, phasing and refinement. An additional step that is very important for serial crystallography is initial hit finding. Hit finding appears to be one of the major challenges in serial crystallography data analysis. In addition, background scattering can be a bottleneck in this first stage of data processing. However, to perform data correction the amount of background that we want to subtract sometimes needs to be scaled on a single pattern basis. NanoPeakCell (NPC)<sup>39</sup> is a software package recently developed for pre-processing the data meets these requirements. It is Python based and has a user-friendly interface. A screenshot is shown on **Fig.23**.

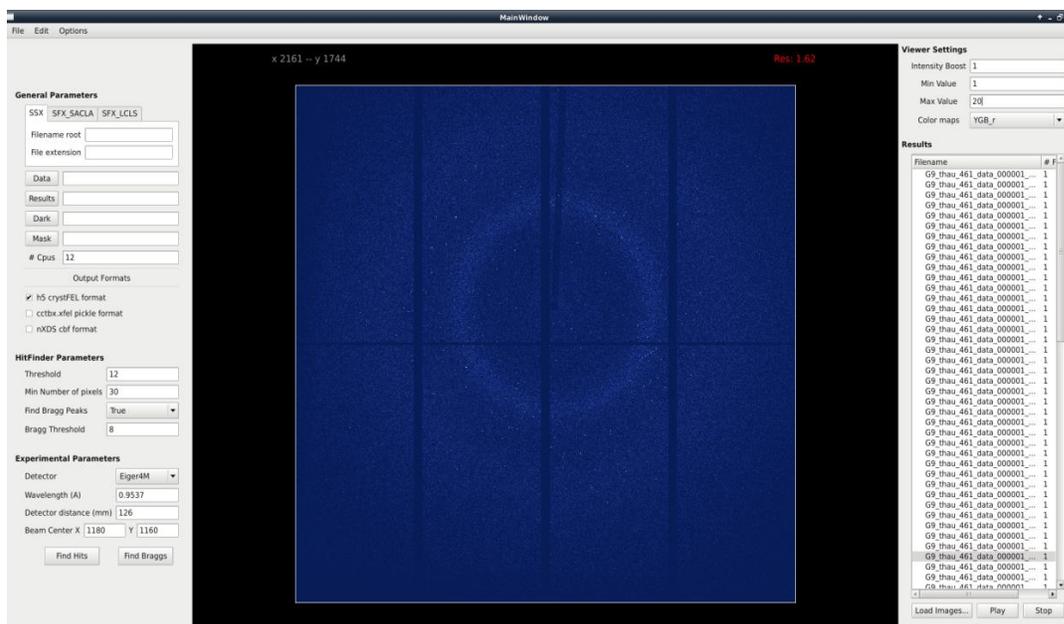


Figure 23. GUI interface of NanoPeakCell.

This software supports almost all currently existing input data formats from different detectors. It also includes a conversion of the images into a suitable format that can be used at the next steps of the data processing (for example HDF5 for CrystFEL). NPC sorts all frames of the dataset and leaves only those that have

actual diffraction spots (hits). This is done on the basis of a pixel values being higher than a threshold that user provides to the software. This step helps to reduce the amount of data for future steps of the data processing. Optionally NPC can search for the Bragg peaks using a local maximum algorithm and it can perform a maximum projection of the selected hits on fly. A maximum projection is a single computed frame containing the maximum pixel values of all diffraction patterns collected in one dataset at a given pixel position. Also if the background scattering is too high it can be optionally subtracted. Users need to provide a path to the directory containing the data, the beam centre, and the detector-sample distance.

### **2.3.2 Indexing of serial crystallography data (CrystFEL)**

According to the idea of serial crystallography, many diffraction patterns should be indexed by merging many individual measurements because the orientation of the crystals cannot be controlled and is unknown. However, the CrystFEL software allows users to perform orientation determination and to extract the intensity from Bragg peaks recorded on diffraction patterns obtained by serial crystallography<sup>50</sup>. Originally it was written to perform data analysis at XFEL sources, but now with the advent of SSX, indexing can also be performed with CrystFEL.

If the Bragg peak search was not performed in a previous step e.g. with NPC, CrystFEL identifies Bragg peaks by searching for clusters of connected pixels that are above an intensity threshold. To capture weak peaks that need a higher threshold in the water-ring region (where background level is higher), an additional criterion can be applied. This criterion is based on local noise levels, where the peak intensity must exceed a given SNR to qualify as a peak. Only diffraction patterns with a certain number of peaks are recognized as indexed and retained for further processing. This determines the hit rate that is written to the terminal.

The indexing of diffraction patterns is performed using conventional algorithms, such as data processing suite implemented in MOSFLM<sup>51</sup>. The

*indexamajig* program performs indexing of the diffraction patterns with MOSFLM<sup>51</sup>, DirAX<sup>52</sup>, XDS<sup>53</sup> and ASDf<sup>52,53</sup>. Indexing is considered to be successful if the lattice parameters found by the auto indexing tool match the known unit cell and the orientation of the matching unit-cell parameters can be used to predict a reasonable number of peaks. The final output file of *indexamajig* contains information about positions of the peaks that were found, orientation matrix, unit cell parameters, centring of the cell, etc. After indexing, the peak intensities are determined to deliver the final hkl intensities using a Monte-Carlo approach<sup>54,55</sup>.

With the recent developments of CrystFEL<sup>56</sup>, the scaling mechanism became more advanced. The merging program *partialator* - is based on multiple passes over the data. *Partialator* performs both the scaling and post-refinement steps including cross-validation and allows to split datasets after this stages.

There are a variety of parameters to judge the data obtained after indexing the diffraction experiment. The most important of these parameters are described here:

**Signal-to-noise-ratio (SNR).** The meaningfulness of measured intensities can be estimated using this parameter. The measure is not always absolutely accurate, but the general rule is that the SNR should exceed 2.

**Completeness.** Each voxel of the electron-density map is calculated correctly only with the contribution of all reflections, this is why the lack of the completeness of the dataset will negatively influence the quality and interpretability of the maps. Even with missing 2-3 % of the reflections, all crystallographic steps (from phasing to refinement) will be hard to perform well.

**$R_{\text{split}}$ .** If we split the data into two separate datasets and merge them independently, then we can examine the agreement between the two resulting intensity lists. Since the data have been split into two sets, it is expected that the degree of convergence in each subset would be lower and so this method could underestimate the quality of the combined data by a factor of  $2^{1/2}$ . A suitable figure of merit could therefore be defined as:

$$R_{split} = 2^{-1/2} \frac{\sum |I_{even} - I_{odd}|}{\frac{1}{2} \sum (I_{even} + I_{odd})}$$

Where  $I_{even}$  is the intensity of the reflection produced by merging even-numbered patterns,  $I_{odd}$  is the intensity of the equivalent reflection from the odd-numbered patterns and the sum is over all reflections.

**CC\***. Recently Karplus, Diederichs<sup>57</sup> and Evans<sup>58</sup> have suggested that the Pearson correlation coefficient (CC) of 2 split datasets ( $CC_{1/2}$ ) might be better suited than merging Rfactors for estimating data quality. **CC\*** is an experimental estimate of what could be called  $CC_{true}$ , the correlation of the averaged intensities with their underlying true values:

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}}$$

**CC\*** is an upper limit for  $CC_{work}$  and  $CC_{free}$  from a properly refined model, where the latter are correlation coefficients between intensities calculated from the model and those obtained from the experiment<sup>57</sup>.

### 2.3.3. Ensemble refinement

The dynamics of the proteins and distortion of the crystal lattice should be taken in account during X-ray diffraction experiment. The obtained diffraction data of proteins are affected by inter-molecular lattice distortions, domain motions and atomic vibrations. However, the structures, which can be found in the PDB database are usually a “snapshot” of one state. A simulation that represents a multi-conformer structure can be used to model the disorder that occurs in the protein.

A very informative way to model protein dynamics is ensemble refinement<sup>59,60</sup> (e.g. using Phenix<sup>61</sup>). It represents the structure as a set of conformers that are not interacting with each other, but each conformer accounts for a fraction of the total electron density. Basically ensemble refinement (ER) is

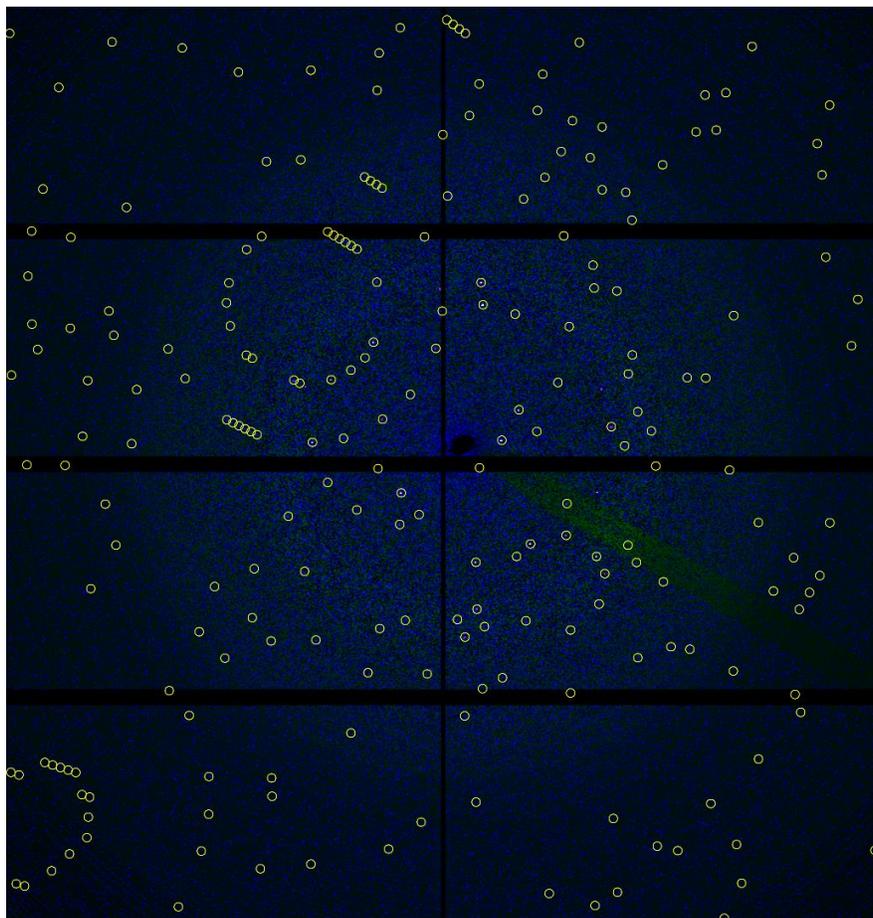
a small set of molecular dynamic (MD) simulations restrained with X-ray data. Ensembles are reduced by “ditching” structures and looking how the R-free factors evolves. At the end of the procedure, the smallest ensemble of structures that allows minimizing R-free is kept. Better electron density maps result in higher restraints on atomic positions, and therefore the diversity of structural conformations. But when the electron density is weakly defined or absent then basically no restraints apply, resulting in high conformational diversity for those atoms. These apparent dynamics are sometimes strongly influenced by crystal packing which is the second restraint imposed on atoms during ensemble refinement of X-ray diffraction data. Performing ensemble refinement on various crystal forms could allow to differentiate between real dynamics and dynamics resulting from differences in crystal packing.

ER restricts the number of structures modelled and prevents over-fitting of the data. In ensemble refinement each atom is given an individual temperature factor, all conformers - fixed occupancies and the initial separation of the conformers is achieved by torsion dynamics simulated annealing<sup>62</sup>.

To optimize an ensemble refinement different parameters can be controlled (a detailed manual can be found on the web page of Phenix ensemble refinement<sup>60</sup>). However, the most important optimization should be performed with the parameters: pTLS, tbath and tx. The PTLs parameter describes the percentage of atoms included in TLS-fitting (Translation/Libration/Screw is a mathematical model that describes the local positional displacement of the atoms in a molecule using an approximation that each atom is part of an ideal rigid body). TbatH controls an X-ray weight, which is modulated in situ such that the simulation runs at the target temperature. The Tx parameter controls the relaxation time in picoseconds of the time-average restraints used in the simulation. The optimum value for this parameter cannot be determined a priori. To determine this value, we set up an array of simultaneous simulations with different values for each dataset depending on the resolution. TbatH and pTLs values in this work were chosen as suggested in the Phenix manual.

### 2.3.4 Data processing in this work.

All data were collected at ID13 beamline, ESRF using the Eiger 4M detector. Hit-finding and Bragg-peak localization were performed using GUI-driven python-based software NanoPeakCell<sup>39</sup>. Indexing was performed by CrystFEL<sup>50</sup> using the **indexamajig** program, which relies on indexing relied on MOSFLM<sup>51</sup>, XDS<sup>53</sup> and Dirax<sup>52</sup>. An example of an indexed pattern is shown on **Fig.24**.



**Figure 24.** Diffraction pattern of ubiquitin indexed by CrystFEL

Merging of integrated intensities was performed using Monte-Carlo integration<sup>63</sup> in CrystFEL<sup>50</sup>. The peak positions are predicted by geometrical calculations, integration is performed at each predicted location and the mean of all intensities for each reflection is taken. The schematic representation of the indexing process is shown on **Fig.25**.



Figure 25. Scheme of indexing process.

CrystFEL has a graphical tool *cell\_explorer* that allows the examination of the results of the unit-cell determination. After the run of *indexamajig* without the unit-cell matching procedure, it produces a list of unit-cell parameters, which *cell\_explorer* displays like a histogram for the a, b, c and angles  $\lambda$ ,  $\beta$ ,  $\gamma$  (Fig.26).

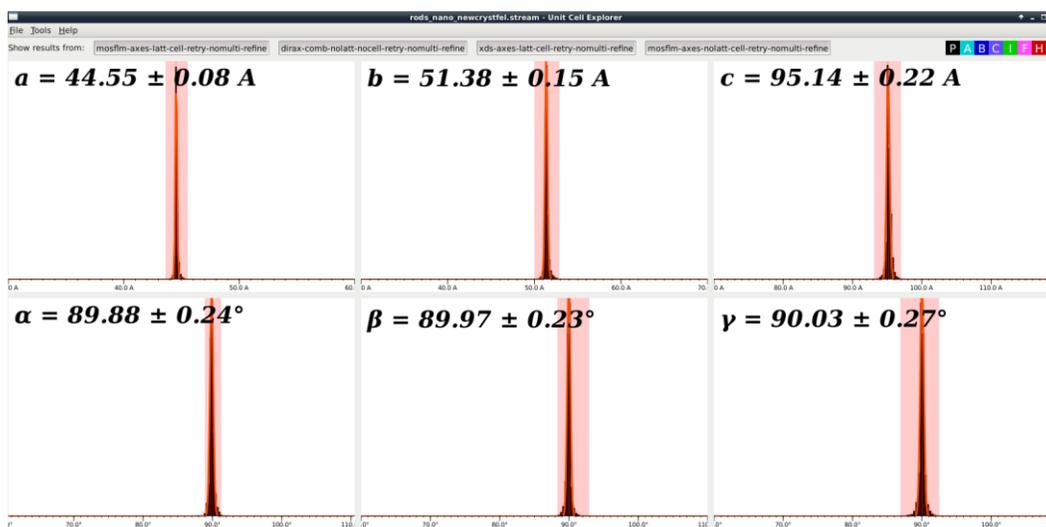


Figure 26. Histogram of indexed parameters on the example of ubiquitin.

For all datasets conservative criteria (SNR>2, completeness=100% and  $R_{split}$ <50%) were used to determine the highest resolution cut off. The quality of the datasets was judged in respect to resolution, SNR,  $R_{split}$ ,  $CC_{1/2}$  and Wilson B factor.

Data were converted to mtz format and phased by molecular replacement using Phaser<sup>64</sup>. The structures were refined by iterative cycles of reciprocal and real-space refinement using the PHENIX<sup>61</sup> software, which also includes refinement of coordinates and B factors. Manual modifications in real space were performed using the COOT<sup>65</sup> software. The figures were produced using PyMOL<sup>66</sup> or custom written Python scripts created by N.Coquelle.

## CHAPTER 3. RESULTS AND DISCUSSIONS

Sample delivery is an essential component of any biological experiment at X-ray sources. The development of sample delivery for serial crystallography at room-temperature started based on injector technology used at XFELs and will be briefly described in 3.1. Today, standard procedures for macromolecular crystallography at synchrotrons involve crystal mounts such as polymer based mesh or grid structures derived from standard nylon fiber based loop mounts which were widely used when the automation of structural biology beamlines was much less developed. Gati et al.<sup>67</sup> performed an innovative serial crystallography experiment based on a combination of simultaneous linear scanning and rotation of such a cryo-loop containing micro-crystals. Another novel way of sample delivery, different from injector technology, has been introduced by Stellato et al.<sup>8</sup>. They performed a microfluidics based experiment where a suspension of crystals was pushed at a constant rate by a syringe pump through a silica fiber into a glass capillary. Nowadays, there are already many fixed target sample delivery systems which are widely used at synchrotron and XFEL sources<sup>39,68,69,70</sup>. Another emerging method is acoustic droplet ejection technology (ADE)<sup>71</sup>. ADE uses short tone bursts of acoustic energy to eject droplets to the beam. Due to the high-efficiency of this method (over 80 % of the drops are probed by the X-ray pulse during the experiment), this method might be a competitive alternative to injectors.

In this chapter we will present the results from serial crystallography experiments based on recent ID13 Developments on injector and solid support setups.

### 3.1 LCP-injectors

In 1996 Landau & Rosenbusch<sup>72</sup> introduced crystallization of membrane proteins using lipidic cubic phase (LCP). LCP is a liquid crystalline phase that forms spontaneously upon mixing lipids with water at proper conditions. LCP consists of a single lipid bilayer that follows an infinite periodic minimal surface dividing the space into two non-intersecting networks of water channels<sup>73</sup>. This crystallization technique has led to the elucidation of the first structures of bacteriorhodopsins<sup>74</sup>, sensory rhodopsins<sup>75</sup>, as well as providing the first high-resolution details of human G protein-coupled receptors (GPCR)<sup>76</sup>.

The history of the injectors, that can be used for delivering microcrystalline samples to the X-ray beam started quite recently with a Rayleigh jet, in which a glass nozzle was used to produce the jet <sup>77</sup>. But due to many problems with clogging of the crystals while using Rayleigh jets, a liquid jet system was created. A liquid jet system uses a gas accelerated to high velocity to push the liquid samples into a jet through the aperture, so called “gas dynamic virtual nozzle” (GDVN)<sup>78</sup>. Protein crystals can be grown in lipidic cubic phase (LCP), but due to the high viscosity of this phase, it is not compatible with GDVN devices. On the basis of this, a new injector, specially designed for crystals grown in LCP was developed<sup>79</sup>. An LCP-injector extrudes a 20-50  $\mu\text{m}$  (diameter) stream into air or vacuum and also, taking advantage of the viscosity, it reduces the consumption of the sample. The first structure with LCP-injector was solved at LCLS<sup>80</sup>; nowadays scientists have already started to perform serial crystallography experiments at synchrotrons using LCP-injectors<sup>81,82</sup>.

### 3.1.1 Explorative LCP-jet test-experiment using the EIGER 4M detector on the example of tubulin protein

For membrane proteins, the main advantage of LCP is that it provides a native-like membrane environment to proteins as opposed to the rather harsh environment associated with detergent micelles. However, for soluble proteins the viscosity of LCP can be an advantage to control the delivery of crystals in the LCP-jet to collect diffraction data. The adaption of LCP-jet technology to a microfocus beamline such as ID13 can be seen as an important step towards a routine application of serial crystallography at synchrotrons. After performing a first SMX experiment on membrane proteins using a LCP-jet at ID13 in 2014<sup>9</sup> (described briefly in Materials & Methods) a further test experiment exploring the potential of the EIGER 4M detector was conducted in collaboration with the Paul Scherrer Institut (PSI), Switzerland and Arizona State University (ASU): *Joerg Standfuss* et al. (PSI) who provided the LCP-jet, *Michel Steinmetz & Natacha Olieric* (PSI) who provided the tubulin samples, and *Uwe Weierstall & Garrett Nelson* et al (ASU), who gave an advice and support for the injector operation.

An initial test of the improved injector setup using bacteriorhodopsin grown in LCP was canceled at a very early stage because the hit rate was extremely low probably due to a very low sample concentration. Therefore, we decided to focus on a second project.

Motivated by ongoing research at PSI, we decided to perform further LCP-jet test runs with the globular protein tubulin. Microcrystals of tubulin have been co-crystallized with a darphin to enhance crystallization. These tubulin crystals were not grown in LCP but the solution containing tubulin crystals was mixed with LCP shortly before the experiment in a 50% LCP:50% tubulin ratio. The size of the crystals ranged from 10 to 50  $\mu\text{m}$ . With this methodological experiment, we wanted to show that LCP-injectors can be used routinely at ID13 with globular or membrane proteins.

Tubulin is a globular protein and one of the most important building blocks of microtubules, major structural components that play an important role in intracellular transport and cell division in all eukaryotic cells. Microtubules

consist of dimers of  $\alpha$  and  $\beta$  tubulin, that bind to guanosine triphosphate (GTP) and assemble onto the ends of microtubules. Microtubule-stabilizing agents are effective chemotherapeutic drugs widely used for the treatment of cancer<sup>83</sup>.

The set up for this experiment was very similar to the previous experiment (briefly described in the Materials & Methods chapter) that was performed at ID13 in 2014<sup>84</sup>, but at this time the LCP-jet was oriented in a vertical instead of horizontal direction. A mini flight tube around 5 cm in diameter (Fig.27) was developed specially for this project to reduce air scattering. The free air-path of the beam (distance between guard aperture and the window of the He-flight tube) was around 3 mm. The velocity of the LCP-flow was 300  $\mu\text{m}/\text{sec}$ , the frame rate of the detector was about 140 Hz (7 ms frame time).

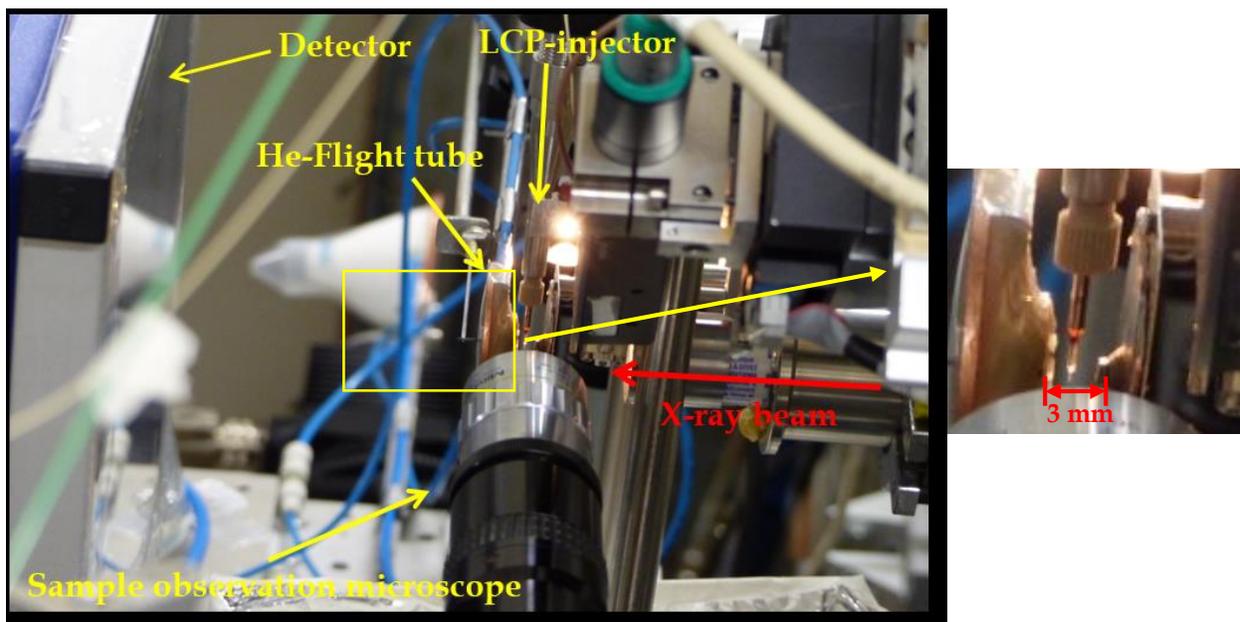


Figure 27. Left-set up of the LCP-jet experiment at ID13, right-5 cm flight tube to reduce air scattering.

The crystals that were passing through the beam ranged from 10 to 45  $\mu\text{m}$  in size, so it was possible to obtain up to 5 shots per crystals with slightly varying crystal orientation due to the small size of the beam.

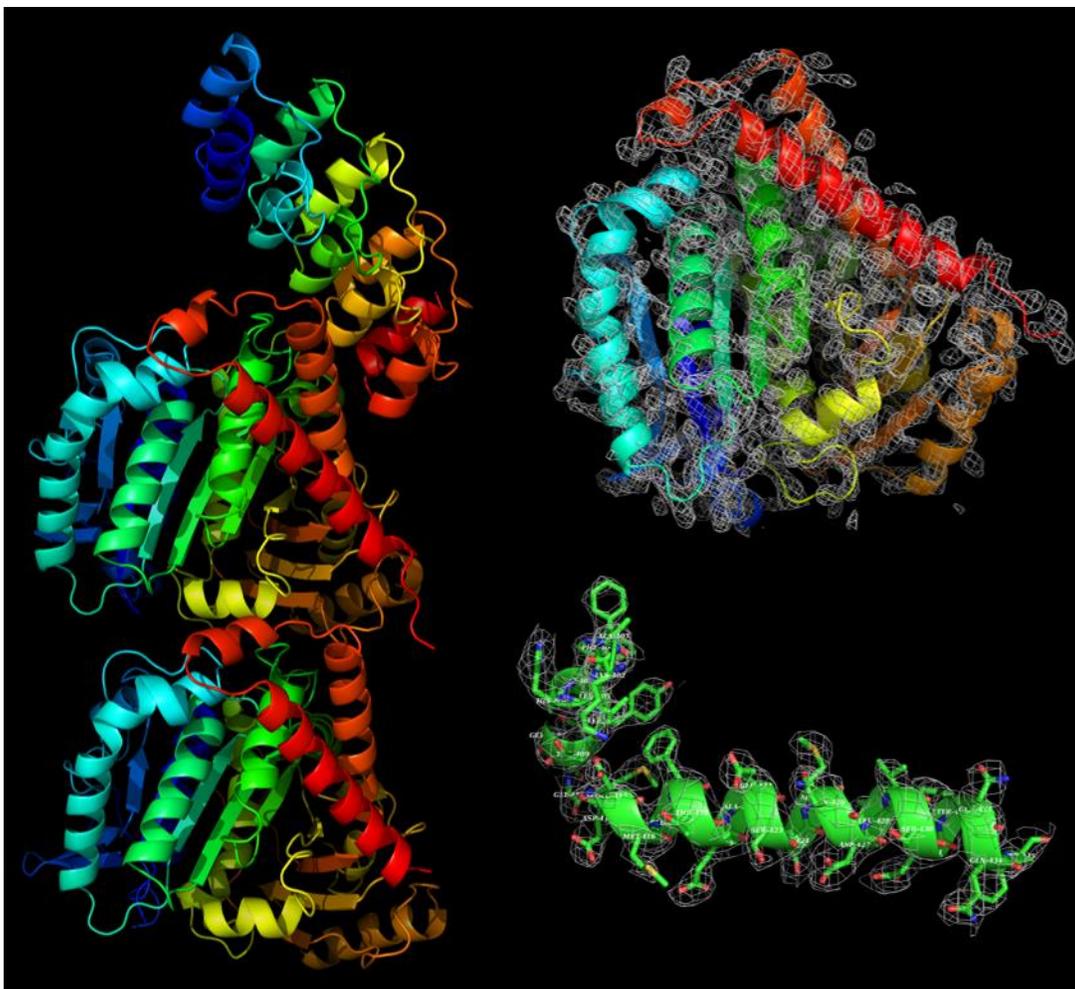
During the experiment around 2000000 frames were collected of which around 20000 had hits and 8 716 of them were indexed with CrystFEL. Based on the initial conservative data criteria the highest resolution was chosen as 3.4  $\text{\AA}$

with 100 % completeness. Data were converted to the mtz format and phased by molecular replacement using Phaser<sup>64</sup> with the PDB file from a previous experiment of the Steinmetz & Olieric group (not published yet) as a starting model. The structure of tubulin was refined by iterative cycles of reciprocal space and real space refinement. Data collection and processing parameters are presented in the **Table 4**.

Parameters	Data collection "Tubulin"
Wavelength, Å	0.954
Beam size, μm <sup>2</sup>	5*3
Beam flux, photons/sec	2*10 <sup>12</sup>
Average crystal size, μm	10-45
Space group	P 12 <sub>1</sub> 1
Unit cell, Å	a=72.84 b=89.33 c=81.73, α=90° β= 97.01° γ=90°
Exposure time, s	0.007
No. of collected frames	1980909
Average hit rate (NanoPeakCell), %	1
No. of indexed patterns (CrystFEL)	8716
Indexing rate (CrystFEL), %	44
No. of total reflections	3970854 (184894)
No. of unique reflections	19458 (1283)
Resolution range, Å	46.0 - 3.4 (3.66-3.40)
Completeness, %	100 (100)
SNR (I/σ(I))	3.95 (2)
R <sub>split</sub> intensity agreement of semi-data sets, %	23.97 (48)
Overall CC* correlation of semi-data sets	0.98 (0.82)
Rwork	0.28 (0.34)
Rfree	0.34 (0.41)
Ramachandran favored, %	96.3
Ramachandran outliers, %	0.3

**Table 4.** Data collection parameters and crystallographic refinement statistics for LCP-jet experiment on tubulin. Values in parentheses are for the highest resolution shell.

The density maps clearly show the alpha helices and the side chains of the residues are also recognizable (**Fig.28**).



**Figure 28. A)  $\alpha,\beta$ -tubulin structure. An additional domain on the top is a darphin, which was used to enhance crystallization of tubulin B) refined electron density map of tubulin (chain B) C) refined electron density map of tubulin (alpha helix region).**

Although the data collection was fast ( $\sim 140$  Hz) the following data transfer was very slow due to the incomplete integration of Eiger 4M detector at the beamline at this early stage, which led to long pauses (20% acquisition,  $\sim 80\%$  pause!, long term average acquisition rate  $\sim 30$  Hz) between the measurements. The LCP flow had to be stopped during the data transfer periods. The data collection of the full dataset took around 3 shifts (24 hours), partially due to the low hit rate but mostly due to the temporary problem with data transfer. The consumption of the pure protein used for this dataset was 1.5 mg in 200  $\mu$ l of LCP.

## Discussion

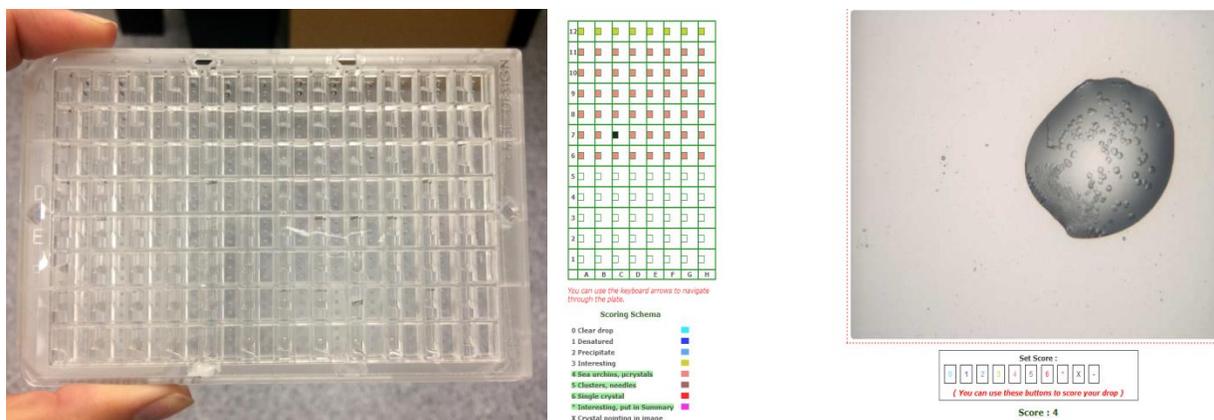
In this experiment we adapted the technology of LCP-injectors to a synchrotron beamline and showed that with this method it is possible to collect a full dataset from a globular protein with low diffraction less than in 1 day. Accounting for the data transfer problem only 5 hours would be necessary. Due to the fact that LCP-injector provides a constant stream of crystals, serial crystallography data collection is possible and the dose of received radiation is equally distributed, which reduces the effects of radiation damage. The amount of the sample that is required is much less in comparison with XFEL data collection. The possibility to collect the data from the same crystal in 2-3 different orientations is of advantage. The fast single-photon counting detector EIGER 4M allowed to collect the data with much higher frame rate, compared to the previous experiment (described in Materials & Methods). The parameters of the detector (140 Hz frame rate, 7 ms exposure time per frame, continuous readout) were well adapted for synchronization with the flow-rate of the injector. Currently, at ID13 LCP-jets experiments can be performed much faster and more efficient (simultaneous data transfer, ~500 Hz continuous frame rate) due to the full integration of the Eiger 4M detector. This would allow for time-resolved investigations.

### 3.2 Implementation of CrystalDirect plate scanning at ID13

During the last years, the field of macromolecular crystallography has been developed towards high throughput techniques for protein crystallization and data collection. An automated system for protein crystallization can improve the success of the crystal growth by reducing the amount of the required protein. Nowadays, these automated systems are widespread, and usually they include robots for the crystallization, and an automated imaging system to follow crystal growth. However, subsequent X-ray data collection requires crystal transfer from the plate to the X-ray beam. Most of the time, this is performed by manual mounting in a cryoloop which can be complicated, leading to the loss of precious crystals.

Recently, the CrystalDirect<sup>85</sup> approach was developed. It provides fully automated crystal harvesting that can be followed by X-ray data collection at cryo-temperature without the need of manual crystal mounting. In this approach, protein crystals are grown on films that are compatible with X-ray data collection. Crystals can be excised automatically by photoablation and directly attached to a pin for X-ray data collection.

The CrystalDirect plate is a vapour-diffusion 96 well microplate (MiTeGen, Ithaca, USA) with a 25  $\mu\text{m}$ -thick cyclic-olefin copolymer (COC) film as a crystallization support (**Fig.29**). COC is an amorphous polymer which minimizes background scattering. The COC film is suitable for the experiment due to the commercial availability and chemical inertness.



**Figure 29. Left- presentation of the plate; right-visualisation of the drops through the CRIMS online application.**

The user can choose how many drops per well to set up (maximum 3). Usually crystallization experiments are set up with 100 nl sample and 100 nl crystallization solution on the inner surface of the films within a CrystalDirect plate using a Cartesian PixSys robot (Cartesian Technologies)<sup>86</sup>. The plates are sealed on their upper side with CrystalClear film (Hampton Research) and the experiments can be set-up at either 20 or 5 °C and regularly imaged under either visible or UV light (**Fig.29**). The CrystalDirect system is currently in operation at the High Throughput Crystallization Laboratory (HTX lab) at EMBL Grenoble (<https://embl.fr/htxlab>)<sup>86</sup>.

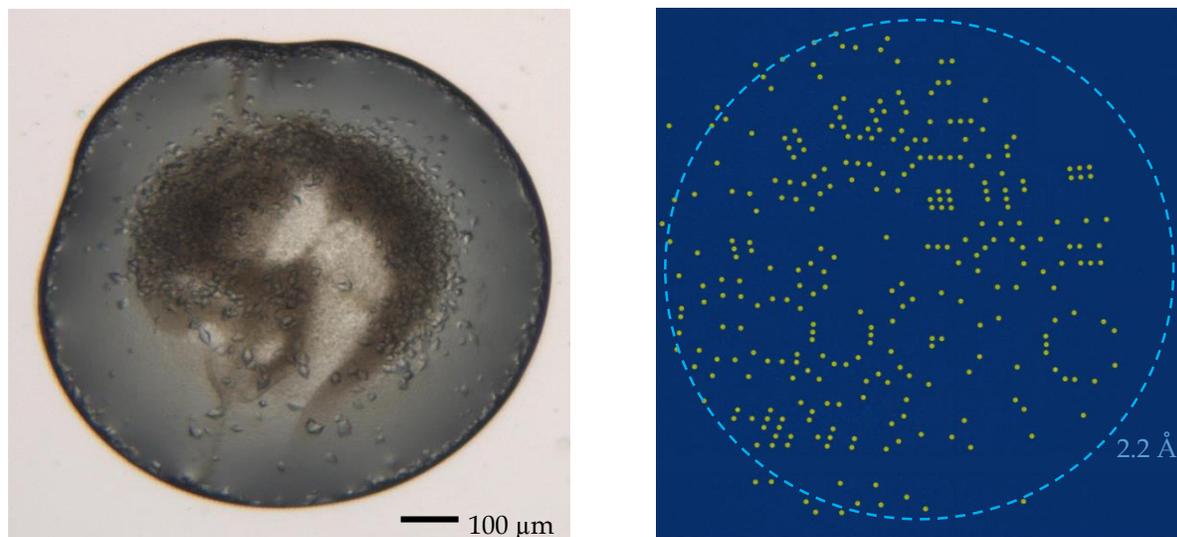
The main goal of the implementation of CrystalDirect plates at ID13 was to show that it is possible to collect high quality datasets with challenging proteins at room temperature grown and without the need of sample transfer. The in-situ approach limits any potential mechanical stress for the crystals and offers the best chances to preserve crystal quality. At the same time the extremely thin films used for CrystalDirect plates make it ideal for these types of experiments, as this considerably reduces background scattering. The specially designed CrystalDirect compatible set-up is described in the Materials & Methods chapter.

This project was performed in collaboration with *Jose Marquez, Florent Cipriani, Zuzanna Kaczmarzka* and *Guillaume Hoffmann* from EMBL, Grenoble.

### 3.2.1 Application to thaumatin

The first test of room temperature serial crystallography at ID13 with CrystalDirect plates has been performed with thaumatin. Thaumatin is a sweet-tasting protein, originally found in the katemfe fruit in West Africa <sup>87</sup>. It is often used as a model protein in crystallography due to the ease of crystallization and high diffraction power.

Crystals of thaumatin were grown on CrystalDirect plates with the help of the HTX facility at EMBL and a Cartesian robot. The average size of the crystals was around 15-20  $\mu\text{m}$  (**Fig.30**).



**Figure 30.** Left-drop with thaumatin crystals 15-20  $\mu\text{m}$ ; right-diffraction of thaumatin to 1.9  $\text{\AA}$

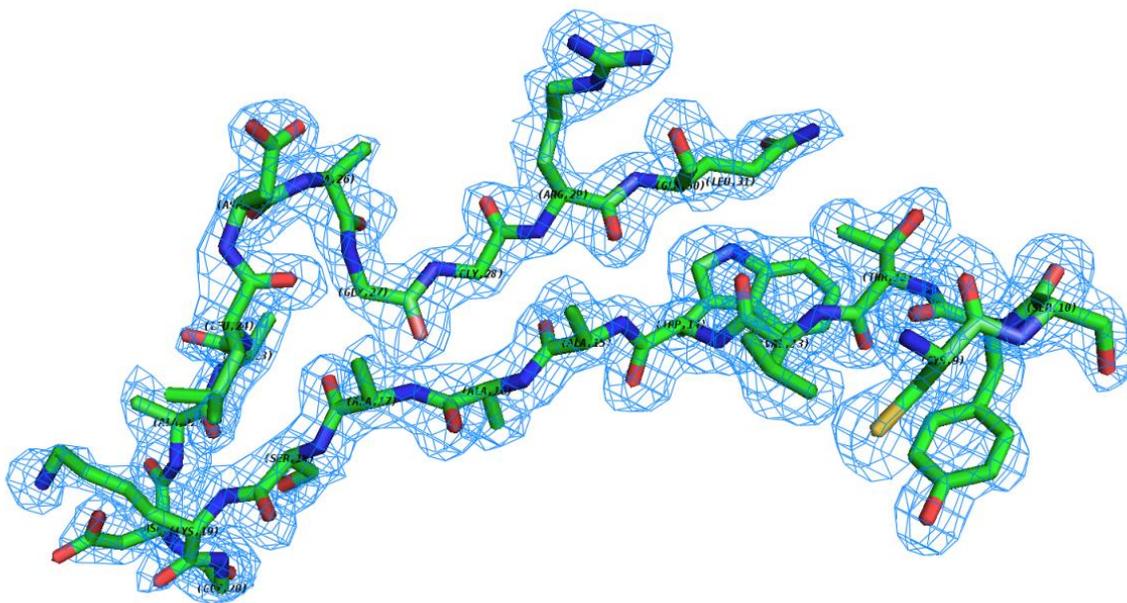
The data collection was performed at the microbranch of ID13, ESRF. The step size was 10  $\mu\text{m}$  both horizontally and vertically. The exposure time varied from 0.007 to 0.1 s and finally was chosen as 0.1 s aimed to the highest diffraction. We collected 111671 frames from 11 drops, from which 9123 images were indexed with CrystFEL with a very high (75%) indexing rate. Based on the initial conservative data processing criteria (SNR>2, Rsplit < 50%, 100% completeness) the highest resolution was chosen as 2.0  $\text{\AA}$ . Data were converted to the mtz format and phased by molecular replacement using Phaser<sup>64</sup> with PDB entry 2WBZ as the starting model. The structure of thaumatin was refined by iterative cycles of

reciprocal space and real space refinement. Data collection parameters are presented in the **Table 5**.

Parameters	Data collection "Thaumatin"
Wavelength, Å	0.954
Beam size, μm <sup>2</sup>	2*2
Beam flux, photons/sec	2.2*10 <sup>12</sup>
Average crystal size, μm	15-18
Space group	P 4 <sub>1</sub> 2 <sub>1</sub> 2
Unit cell, Å	a=b=58.53, c=151.78, α=β=γ=90°
Exposure time, s	0.007
No. collected frames	111671
Average hit rate (NanoPeakCell), %	11
No. of indexed patterns (CrystFEL)	9123
Indexing rate (CrystFEL), %	75.2
No. of total reflections	7236429 (370084)
No. of unique reflections	34115 (2272)
Resolution range, Å	46.35-2.0 (2.05-2.00)
Completeness, %	100 (100)
SNR (I/σ(I))	4 (2)
R <sub>split</sub> intensity agreement of semi-data sets, %	15.86 (48)
Overall CC* correlation of semi-data sets	0.99 (0.81)
R <sub>work</sub>	0.30 (0.30)
R <sub>free</sub>	0.34 (0.34)
Ramachandran favored, %	97.7
Ramachandran outliers, %	0

**Table 5.** Data collection parameters and crystallographic refinement statistics of CrystalDirect approach on thaumatin. Values in parentheses are for the highest resolution shell.

Refined 2mFo-DFc electron-density maps at contour level 1.2σ are shown in **Fig. 31**. As can be observed, the resulting electron density map has high quality, and the shape of all residues is explicit (for example tyrosine).



**Figure 31. Refined electron density map of thaumatin at 2 Å resolution (contour level 1.2 $\sigma$ ).**

Room temperature data collection from thaumatin crystals showed that CrystalDirect plates are suitable for scanning serial crystallography and provide a convenient way to deliver crystals to the microbeam with minimal manipulation. This experiment paved the way for further trials with more challenging proteins.

### 3.2.3 Application to ligand-gated ion channels: 5HT<sub>3</sub> and GLIC

After successfully testing the model protein thaumatin, we decided to perform data collection looking at a more “realistic” case. We focused on membrane proteins as they tend to produce microcrystals, which might be suitable for serial crystallography studies. In collaboration with *H.Nury* (IBS), we decided to test 5-HT<sub>3</sub> receptor, whose structure was just recently solved to 3.5 Å resolution<sup>88</sup> (**Fig.33**). The 5-HT<sub>3</sub> receptor is a member of the Cys-loop superfamily of ligand-gated ion channels that mediate fast neurotransmission through the nervous system. They are also first-class therapeutic targets for a number of conditions (such as depression, neurodegenerative diseases, nausea and irritable bowel syndrome). As the structure of the 5-HT<sub>3</sub> receptor was already known

(Fig.32), our goal was to investigate the dynamics of conformational changes triggered by the binding of neurotransmitters<sup>88</sup>.

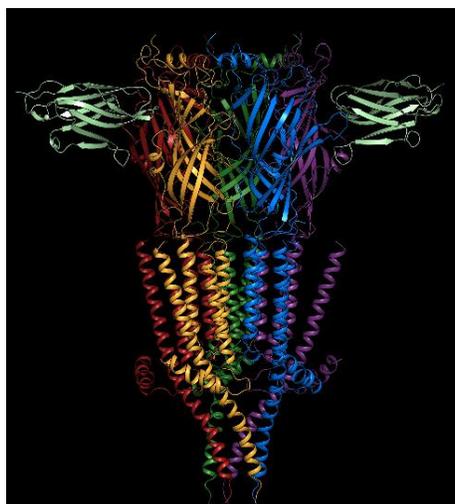


Figure 32. Structure of 5-HT<sub>3</sub> receptor solved at cryo-temperature to 3.5 Å resolution (pdb code 4PIR)<sup>88</sup>.

The crystals were grown on CrystalDirect plates, they were needle-shaped with the size around 20-60 μm (Fig.33). We performed scans on 5 drops (around 70 000 images) with 10 μm step size and with exposure time 0.05 s. Unfortunately, the hit rate was too low and the resolution did not exceed 6 Å (Fig. 33). This is probably due to the difficulty in reproducing the crystallization conditions, which is a known problem with membrane proteins.

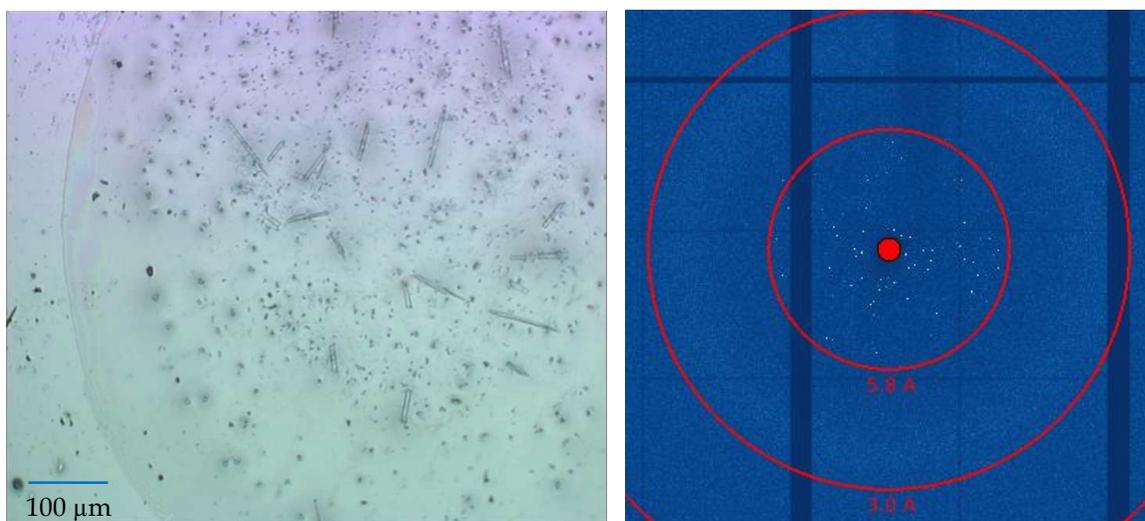


Figure 33. Left-crystals of 5-HT<sub>3</sub> receptor; right-maximum projection of one of the drop.

Due to the low diffraction quality of microcrystals of the 5-HT<sub>3</sub> receptor we decided to change the protein and we started working with GLIC, another membrane protein available at that moment, whose structure is also known at even higher resolution (2.4 Å)<sup>47</sup> than for the 5-HT<sub>3</sub>-receptor. GLIC is a ligand-gated ion channel protein which forms proton-gated channels that are activated by a pH decrease on the extracellular side and that do not desensitize after activation<sup>89</sup>. The structure builds a funnel-shaped transmembrane pore widely open on the outer side and lined by hydrophobic residues (Fig.34). On the inner side, a constriction matches with rings of hydrophilic residues that contribute to the ionic selectivity<sup>90</sup>. The crystals were grown on CrystalDirect plates and the size ranged from 5 to 30 μm (Fig. 35).

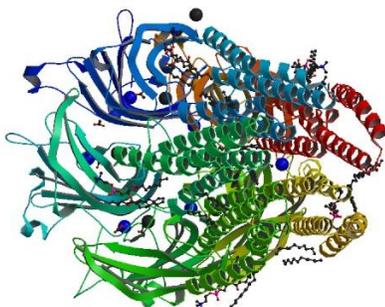


Figure 34. Pentameric ligand with BR-lidocaine solved to 2.4 Å resolution at 100K (PDB entry 4HFI).

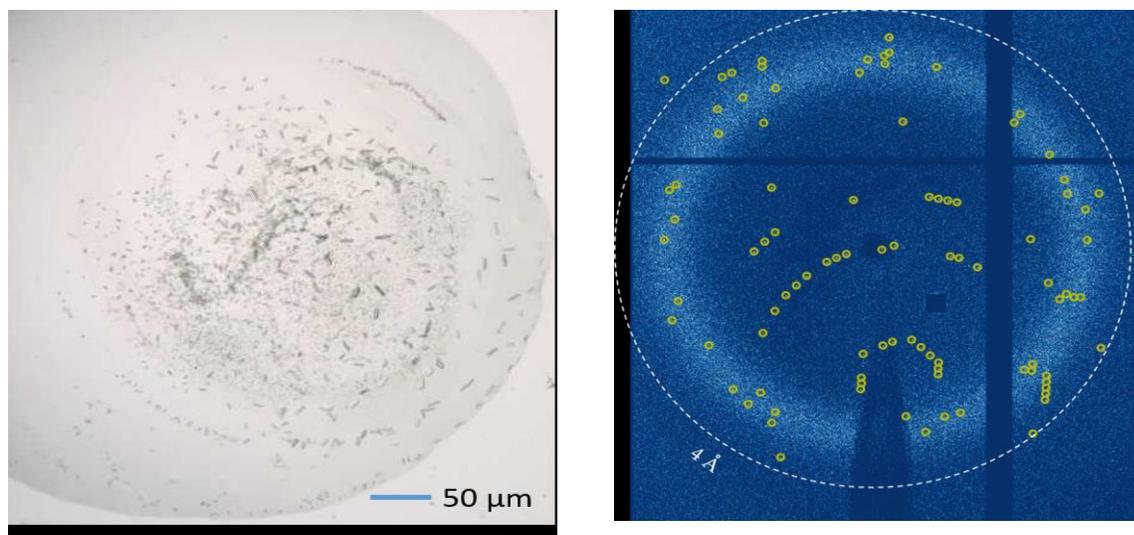
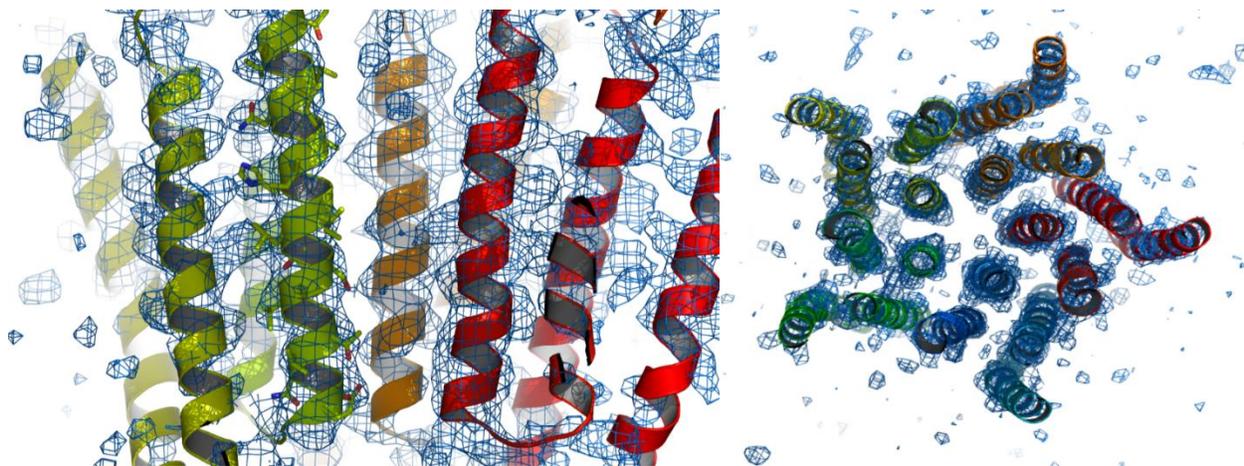


Figure 35. Left-crystals of membrane protein GLIC in the range from 5 to 30 μm; right-diffraction of GLIC to 3.8 Å.

More than one million frames were collected but the hit rate did not exceed 1% due to the low quality of the crystals. However, 10159 images were indexed with CrystFEL. Based on the initial data processing criteria the highest resolution was chosen as 4.0 Å. Alas, the completeness at this resolution was only 45 %. 100% completeness with this amount of data was observed only at 5.2 Å resolution. Due to limited amounts of the protein, we were not able to collect more data to reach the full completeness at 4 Å resolution, but from inspection of individual frames we saw that these crystals diffract even higher than 4 Å (**Fig.35**). Data were converted to mtz format and phased by molecular replacement using Phaser<sup>64</sup> with PDB entry 4HFI<sup>91</sup> as the starting model. We performed several cycles of refinement to see if with obtained phases the presence of the major structural elements can be seen. When  $R_{\text{work}}$  and  $R_{\text{free}}$  reached the values of 0.44 and 0.47, respectively, the refinement was not pursued further due to the limited resolution. Final 2mFo-DFc electron-density maps are shown in **Fig. 36**.



**Figure 36.** Refined structure of GLIC. Left-side view, right-top view.

The maps show clearly the positions of the helices (unchanged compared to the structure at CT) and even some side chain positions. The data collection and processing parameters are presented in the **Table 6**.

Parameters	Data collection "GLIC"
Wavelength, Å	0.94
Beam size, μm <sup>2</sup>	2*2
Beam flux, photons/sec	2.2*10 <sup>12</sup>
Average crystal size, μm	5-30
Space group	C 121
Unit cell, Å	a=181.84, b=132.96, c=165.00, α=90° β=104.21°, γ=90°
Exposure time, s	0.05
No. collected frames	1360000
Average hit rate (NanoPeakCell), %	1
No. of indexed patterns (CrystFEL)	10159
Indexing rate (CrystFEL), %	74.7
No. of total reflections	891735 (46131)
No. of unique reflections	41615 (2773)
Resolution range, Å	24.99-4.0 (4.2-4.0)
Completeness, %	100 (45)
SNR (I/σ(I))	2.85 (2)
R <sub>split</sub> intensity agreement of semi-data sets, %	27.81 (49.9)
Overall CC* correlation of semi-data sets	0.98 (0.80)
B factor from Wilson Plot. Å <sup>2</sup>	92.7
R <sub>work</sub>	0.44 (0.45)
R <sub>free</sub>	0.47 (0.47)
Ramachandran favored, %	96.3
Ramachandran outliers, %	0

**Table 6.** Data collection parameters and crystallographic refinement statistics of the membrane protein GLIC grown on CrystalDirect plates. Values in parentheses are for the highest resolution shell.

Even with the comparably high quality of the maps given the obtained resolution, this is a limiting factor for room temperature studies, especially if the aim is to investigate conformational changes inside the protein. With 4-5 Å resolution we can just see the shape of the molecule. Naturally, the quality of the crystals is a major challenge for membrane protein serial crystallography as for conventional crystallography. However, we decided to verify whether our low-resolution limitation was linked to the overall quality of the crystals or

depending on the size of the crystals. Upon testing the crystallization conditions, we obtained a plate with few crystals of more than 100  $\mu\text{m}$  in size, which are not very suitable for serial crystallography. We tested their quality of diffraction, also via room temperature scanning, and found that they diffract well to 2.3  $\text{\AA}$  (Fig.37).

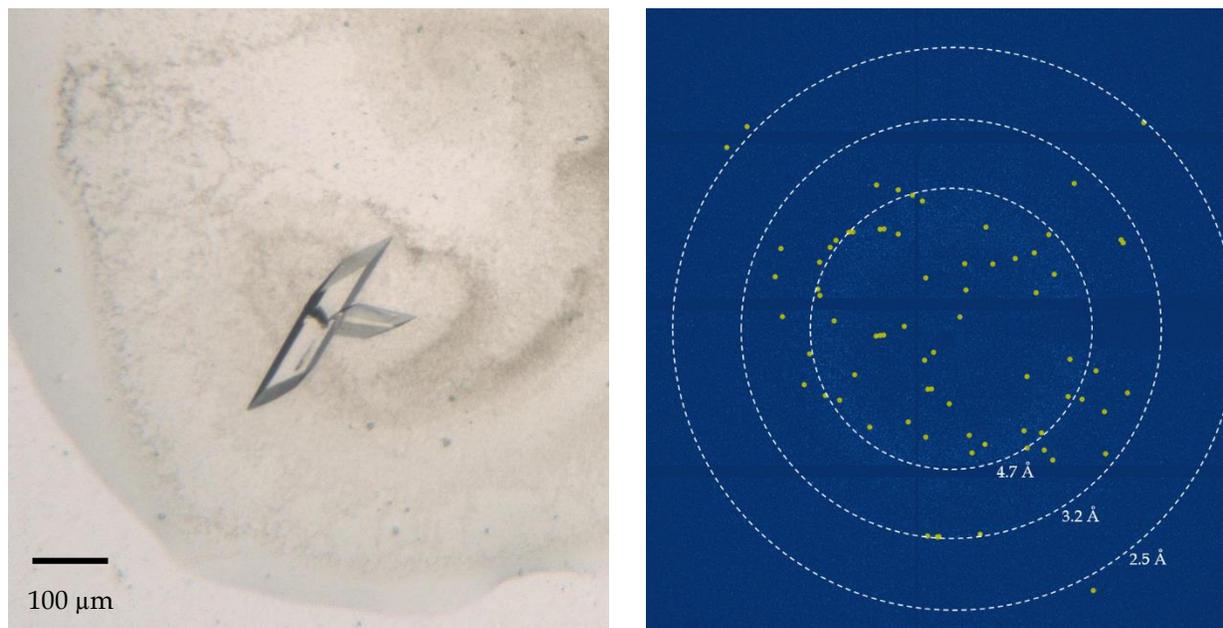


Figure 37. Left-300  $\mu\text{m}$  crystal of GLIC, right-diffraction of GLIC to 2.3  $\text{\AA}$ .

The best data for the GLIC protein in the PDB database had 2.4  $\text{\AA}$  resolution (PDB code 4HFI). From here we concluded that for this membrane protein the optimum size for the highest diffraction quality might be close to 100 microns or more.

### 3.3 Improvement of the data quality using liquid removal.

Proteins are known to remain in a native structure within the crystals and they are capable of performing short and long-range structural reorganisation within the lattice<sup>92</sup>. This reorganisation can influence the diffraction quality of the crystal. There are many post-crystallisation treatments that can be used to improve the diffraction quality and which are taken into consideration the dynamic nature of macromolecular crystals.<sup>93</sup> Sometimes reorganisation can lead

to higher internal order, lower solvent content or a change in space group that may result in a higher quality dataset.

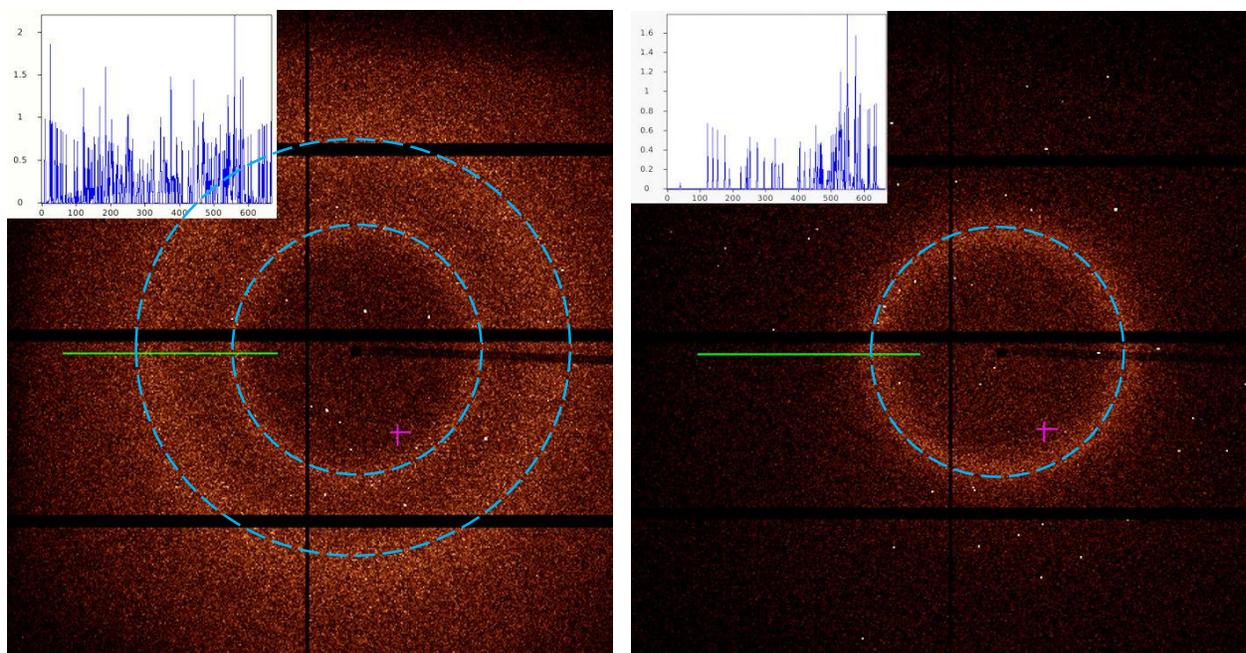
Treatments for improving data quality and crystal diffraction using dehydration of the sample are known to have successful results<sup>94,95</sup>. However, dehydration can change the crystal order and the main trademark of dehydration is a decrease in unit-cell parameters.

We wanted to address the question whether post-crystallization removal of mother liquor can increase the quality of diffraction data from a crystal without leading to evaporation and subsequent dehydration in the wells. We decided to collect the data on the same batch of protein with and without liquid excess to see if it would improve the quality of the datasets. Liquid removal was performed using a special device, developed by the EMBL, which was described earlier in the Materials & Methods chapter.

We performed data collection using CrystalDirect plates with lysozyme and ubiquitin as test proteins. The first attempt was performed with lysozyme due to the high crystallization rate and availability of this soluble protein.

### **3.3.1 Application to lysozyme**

We collected two datasets under comparable conditions on 12 drops of lysozyme microcrystals each (4 wells with 3 drops) with and without liquid. The set-up was configured with the flight tube to reduce air scattering. From the beginning of the experiment we saw a difference in the background signal (**Fig.38**). The inner (low- $q$ ) ring, which is observed in both frames, is due to scattering from the polymer film of the CD plate, on which crystals were grown. The outer ring that can be seen only on the left picture can be attributed to background scattering from the buffer.



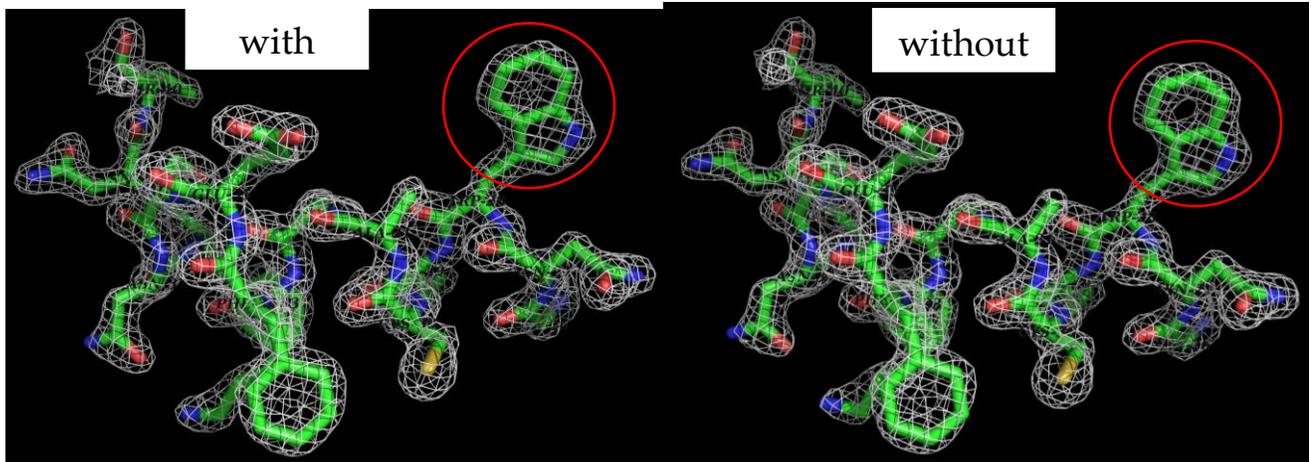
**Figure 38.** Difference in background signal for lysozyme crystals. Left-dataset with liquid; right-dataset with removed liquid.

257000 frames for each dataset were collected. The hit rate was 1 % higher for the liquid removed dataset. Based on the conservative data processing criteria the highest resolution was chosen as 1.9 Å (with liquid) and 1.8 Å (liquid removed). Data collection and processing parameters are presented in **Table 7**.

We can see that all parameters are slightly improved for the dataset without liquid. The most important difference that we observed was an increase in resolution for 0.1 Å for the lysozyme, that always diffract very well. The final 2mFo-DFc electron-density maps are shown in **Fig. 39**. These maps show a drastic improvement in the quality of the electron density near tryptophan in the dataset without solution.

Parameters	Lysozyme with liquid	Lysozyme without liquid
Wavelength, Å	0.954	
Beam size, μm <sup>2</sup>	2*2	
Beam flux, photons/sec	2.2*10 <sup>12</sup>	
Average crystal size, μm	15-18	
Space group	P 4 <sub>3</sub> 2 <sub>1</sub> 2	
Unit cell, Å	a=b=78.65, c=38.48,α=β=γ=90°	a=b=78.56, c=38.56,α=β=γ=90°
Exposure time, s	0.01	
No. collected frames	257040	
Average hit rate (NanoPeakCell), %	22.6	23.5
No. of indexed patterns (CrystFEL)	41672	47643
Indexing rate (CrystFEL), %	77.3	78.9
No. of total reflections	2204002 (79168)	5484076 (151227)
No. of unique reflections	21428 (1444)	21448 (1434)
Resolution range, Å	55.54-1.9 (1.98-1.90)	55.54-1.8 (1.82-1.80)
Completeness, %	100 (100)	100 (100)
SNR (I/σ(I)) overall	9.78 (1.97)	9.83 (2)
B factor from Wilson Plot. Å <sup>2</sup>	44.9	44.8
R <sub>split</sub> intensity agreement of semi-data sets, %	8.52 (53.7)	7.92 (44.8)
Overall CC* correlation of semi-data sets	0.996 (0.93)	0.997 (0.95)
R <sub>work</sub>	0.21 (0.44)	0.21 (0.44)
R <sub>free</sub>	0.24 (0.48)	0.25 (0.49)
Ramachandran favored, %	98.50	99.24
Ramachandran outliers, %	0	0

Table 7. Data collection, indexing and refinement parameters for the lysozyme with and without liquid removal.

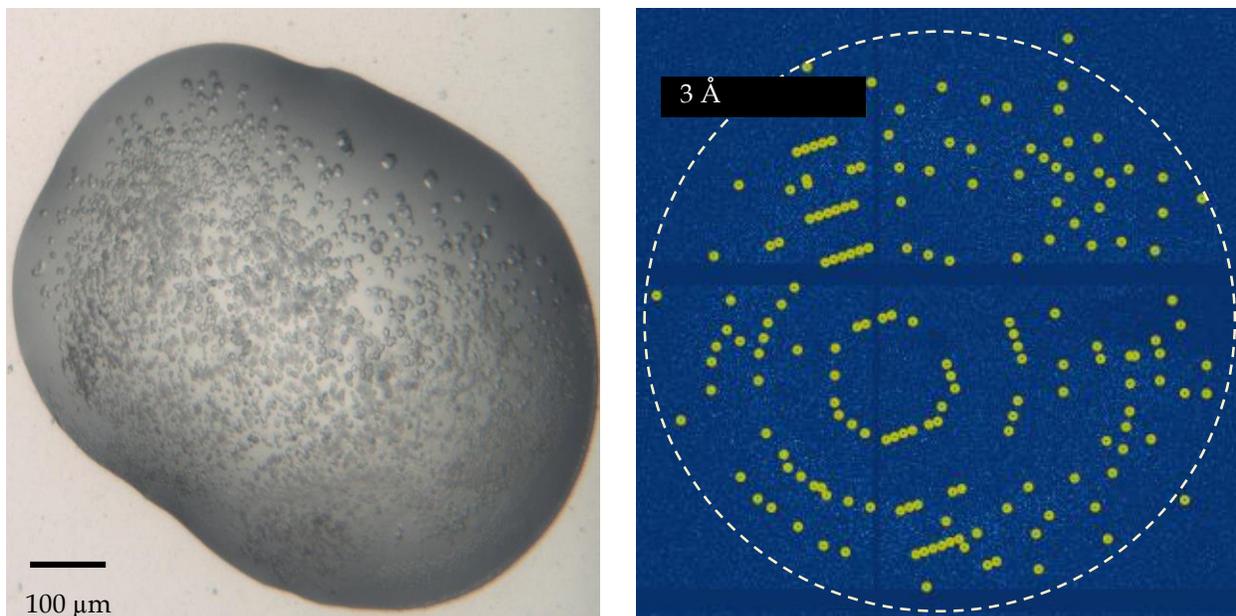


**Figure 39. Refined density maps of lysozyme, contour level  $1.8\sigma$ .**

To explore this method beyond the typical lysozyme test case, we decided to try another protein, which would diffract in the range of 2.5-3.5 Å to confirm the advantages of dehydration of the drop.

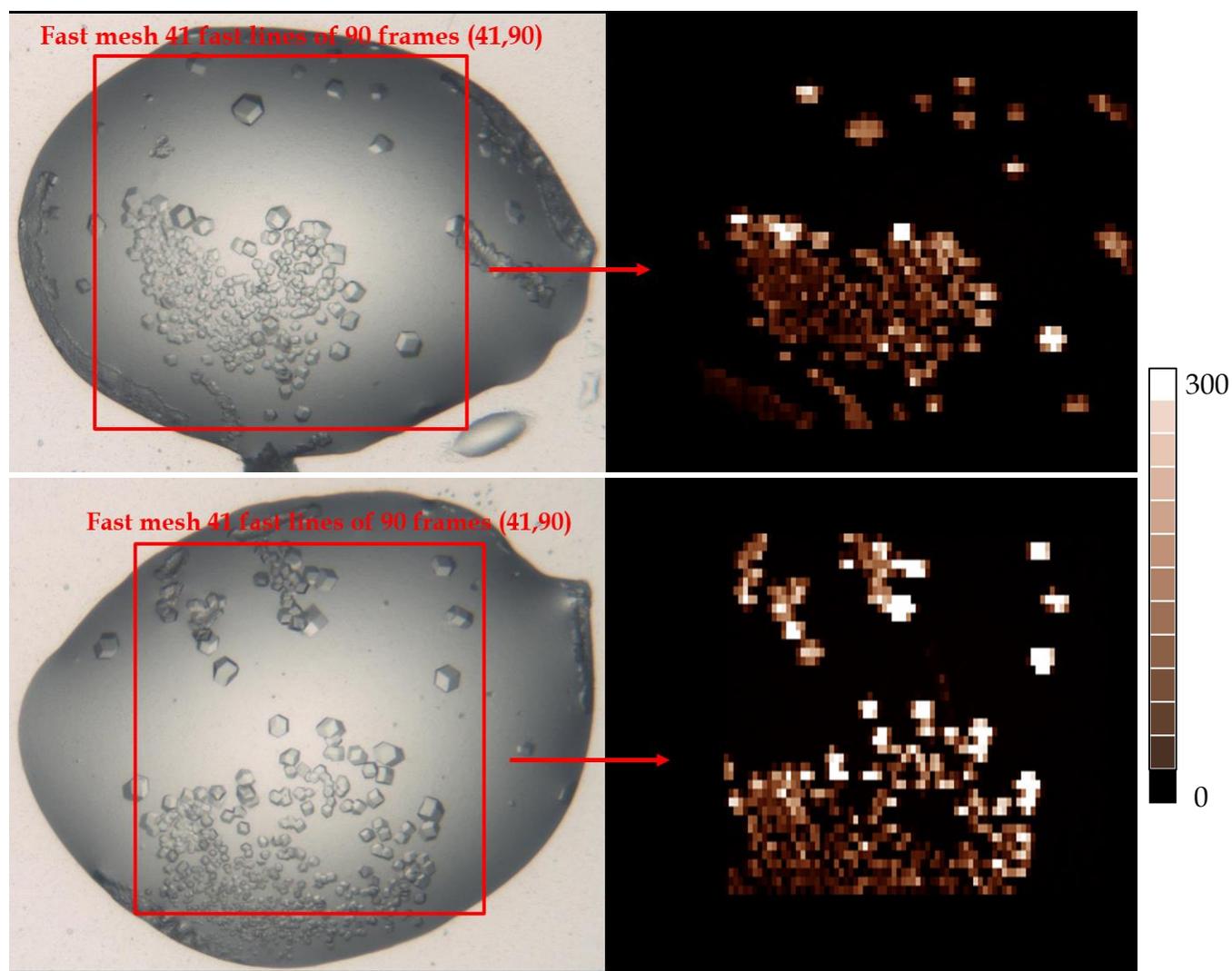
### **3.3.2. Application to ubiquitin**

Regarding liquid removal, we performed a further comparative study using cubic-shaped crystals of ubiquitin. This crystal form of ubiquitin was chosen because it typically yields a highest resolution around 3 Å (**Fig.40**), which could be helpful for understanding whether liquid removal would sufficiently reduce background scattering to improve data quality. The average size of the crystals that were grown using the CrystalDirect approach was 5-10 microns (**Fig.40**).



**Figure 40.** Left- drop with microcrystals of ubiquitin. The average crystal size was 5-10  $\mu\text{m}$ , however, several crystals exceeded 20  $\mu\text{m}$ ; right- diffraction pattern of ubiquitin to 2.9  $\text{\AA}$ .

We collected 2 datasets, 24 drops of protein microcrystals each with and without liquid. The typical size of the drops was of 1 mm in diameter. The step size for both datasets was 10  $\mu\text{m}$  both vertically and horizontally. The fast continuous scan was used for this data collection as described in Materials & Methods. The scan of one drop was approximately 5-6 minutes, depending on the scanned area, so the full dataset was collected within two hours. The composite images of the scanned drops can be seen on **Fig.41**.



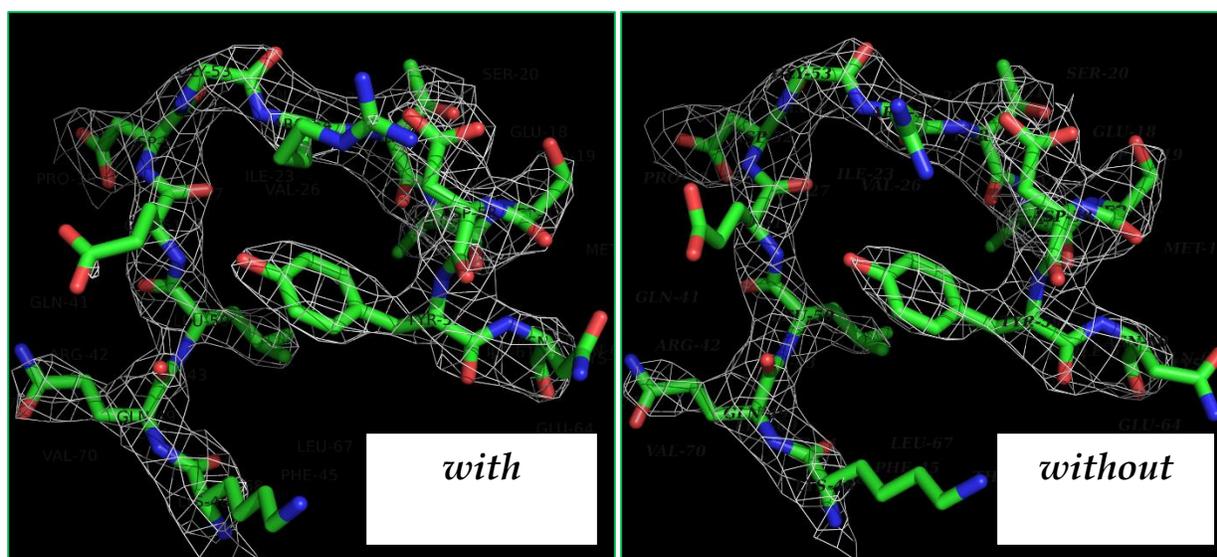
**Figure 41.** An example of the composite image obtained after performing a fast continuous scan of the droplets with crystals of ubiquitin; left - micrograph from HTX software from the CrystalDirect infra-structure; right- hitmaps calculated as number of pixels above the threshold (in this case 15 detector counts), white colour means 300 or more pixels.

Based on the initial data processing criteria the highest resolution was chosen as 3 Å (with liquid) and 2.85 Å (liquid removed). 2.85 Å was the best resolution that we found among all ubiquitin structures with this space group ( $P 4_332$ ) in the PDB database. The parameters of the data collection can be found in **Table 8**.

Parameters	Ubiquitin with liquid	Ubiquitin without liquid
Wavelength, Å	0.922	
Beam size, $\mu\text{m}^2$	2*2	
Beam flux, photons/sec	2.2*10 <sup>12</sup>	
Average crystal size, $\mu\text{m}$	3-30	
Space group	P 4 <sub>3</sub> 32	
Unit cell, Å	a=b=c=106.37, $\alpha=\beta=\gamma=90^\circ$	a=b=c=106.49, $\alpha=\beta=\gamma=90^\circ$
Exposure time, s	0.01	
No. collected frames	134640	
Average hit rate (NanoPeakCell), %	20.7	29.9 ( <b>↑ 9.2</b> )
Indexing rate (CrystFEL), %	34.8	36.5
No. of indexed patterns (CrystFEL)	9694	14672
No. of total reflections	7330119 (358498)	6022603 (268114)
No. of unique reflections	9209 (632)	9209 (607)
Resolution range, Å	47.72-3 (3.03 -3.00)	47-72-2.85 (2.88-2.85) ( <b>↓ 0.15</b> )
Completeness, %	100 (100)	100 (100)
SNR (I/ $\sigma$ (I)) overall	5.4 (2.2)	6.91 (2.1) ( <b>↑ 1.51</b> )
B factor from Wilson Plot. Å <sup>2</sup>	64	86
R <sub>split</sub> , %	14.42 (46.88)	10.81 (49.5) ( <b>↓ 3.61</b> )
CC*	0.995 (0.929)	0.996 (0.929)
R <sub>work</sub>	0.22 (0.24)	0.25 (0.27)
R <sub>free</sub>	0.27 (0.31)	0.28 (0.31)
Ramachandran favored, %	99.3	99.3
Ramachandran outliers, %	0	0

Table 8. Data collection, indexing and refinement parameters for ubiquitin with and without liquid removal. The significantly improved parameters are marked in red.

The difference in hit rate of almost 10 %, an increased SNR from 5.4 to 6.9, a much lower R<sub>split</sub> and differences on other parameters obviously indicate that liquid removal is very useful and can improve the resolution in this case by almost 0.2 Å.

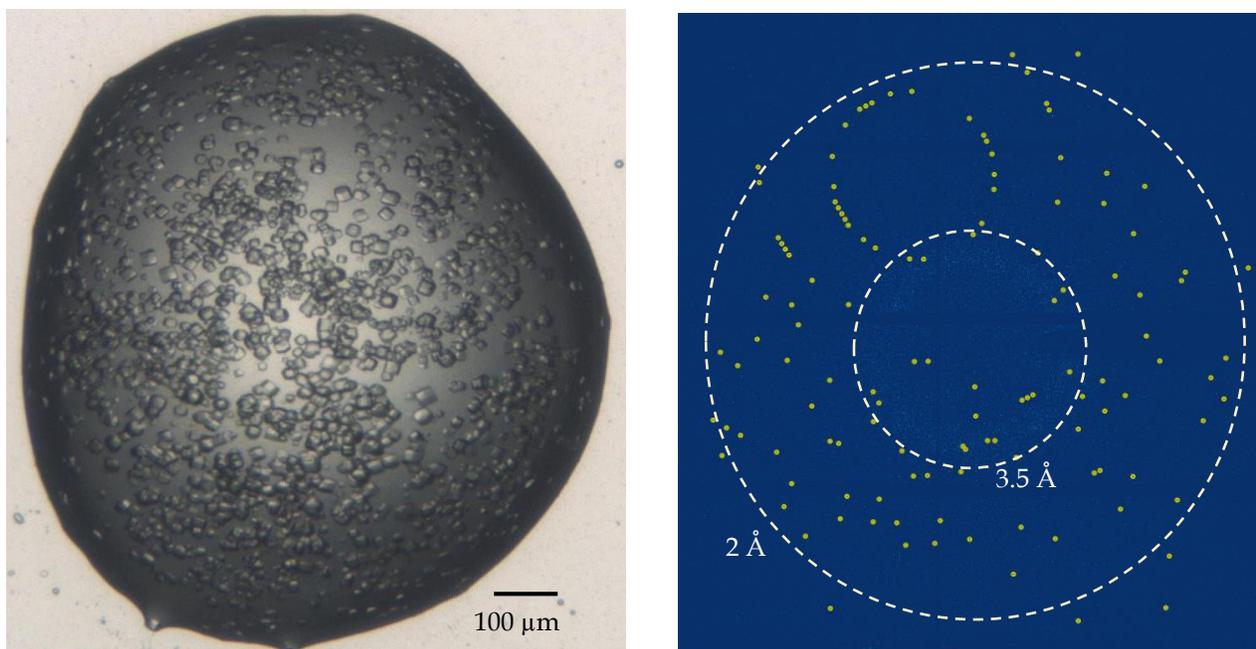


**Figure 42.** Electron density of ubiquitin with and without liquid (contour level  $1.6\sigma$ ).

The electron densities are very similar in both cases (**Fig.42**). However the density near tyrosine appears to be better defined for the dataset without liquid.

### 3.3.3 Application to lysozyme with benzamidine ligand

Protein–ligand binding is of fundamental interest in general, and for biomedical research in particular. With this experiment we wanted to show that liquid removal does not influence ligand binding. The dataset was collected after adding 50 nL of benzamidine to each 200 nL droplet containing lysozyme crystals 5 minutes before preparing the data collection. The typical size of the drops was 1 mm. The crystals ranged from 5 to 20  $\mu\text{m}$  in size (**Fig.43**) and the collected area of a drop was from 0.4 to 1  $\text{mm}^2$ . We collected two similar datasets, 9 drops of lysozyme microcrystals each with and without liquid. The step size for both datasets was 10  $\mu\text{m}$  both vertically and horizontally. We collected around 75000 images for each dataset and the hit rate was almost 10 % higher for the dataset without solution. The indexing rate was almost 3 times higher for the dataset with removed liquid. Based on the initial data processing criteria the highest resolution was chosen as 1.7  $\text{\AA}$  (with liquid) and 1.57  $\text{\AA}$  (liquid removed) (**Fig.43**).



**Figure 43.** Left- drop with microcrystals of lysozyme in the 10-20  $\mu\text{m}$  range; right-diffraction of the lysozyme microcrystals up to 1.6  $\text{\AA}$ .

Data were converted to mtz format and phased by molecular replacement using Phaser<sup>64</sup> with PDB entry 193L<sup>96</sup> as the starting model. The parameters of the data collection can be found in **Table 9**.

The structure was refined via several cycles of reciprocal space and real space refinement. Reciprocal space refinement was performed using Phenix software<sup>61</sup> including the refinement of coordinates and B-factors. Manual modifications in real space were performed by Coot<sup>65</sup>. Refined electron-density maps are shown in **Fig.44**. The figures were produced using PyMOL<sup>66</sup>.

Parameters	Lysozyme with benzamidine with liquid	Lysozyme with benzamidine without liquid
Wavelength, Å	0.922	
Beam size, μm <sup>2</sup>	2*2	
Beam flux, photons/sec	2.2*10 <sup>12</sup>	
Average crystal size, μm	5-20	
Space group	P 4 <sub>3</sub> 2 <sub>1</sub> 2	
Unit cell, Å	a=b=79.45, c=38.20, α=β=γ=90°	a=b=79.47, c=38.34, α=β=γ=90°
Exposure time, s	0.01	
No. collected frames	82940	70030
Average hit rate (NanoPeakCell), %	16.4	25.4 (↑ 9)
Indexing rate (CrystFEL), %	24.8	71 (↑ 46.2)
No. of indexed patterns (CrystFEL)	3228	13624
No. of total reflections	898202 (33119)	3473148 (108445)
No. of unique reflections	25857 (1724)	34121 (2296)
Resolution range, Å	47.72-1.7 (1.74-1.70)	47-72-1.58 (1.68-1.58) (↓ 0.12)
Completeness, %	100 (100)	100 (100)
SNR (I/σ(I)) overall	3.50 (2.16)	5.04 (2) (↑ 1.54)
B factor from Wilson Plot. Å <sup>2</sup>	62	93
Rsplit	28.88 (46.40)	18.44 (50.3) (↓ 10.44)
CC*	0.963 (0.918)	0.985 (0.913) (↑ 0.022)
Rwork	0.18 (0.21)	0.18 (0.22)
Rfree	0.21 (0.25)	0.22 (0.27)
Ramachandran favored, %	98.50	99.24
Ramachandran outliers, %	0	0

Table 9. Data collection, indexing and refinement parameters of the dataset “Lysozyme with benzamidine” with and without liquid removal. The most important improvements in parameters are marked in red. Values in parentheses are for the highest resolution shell.

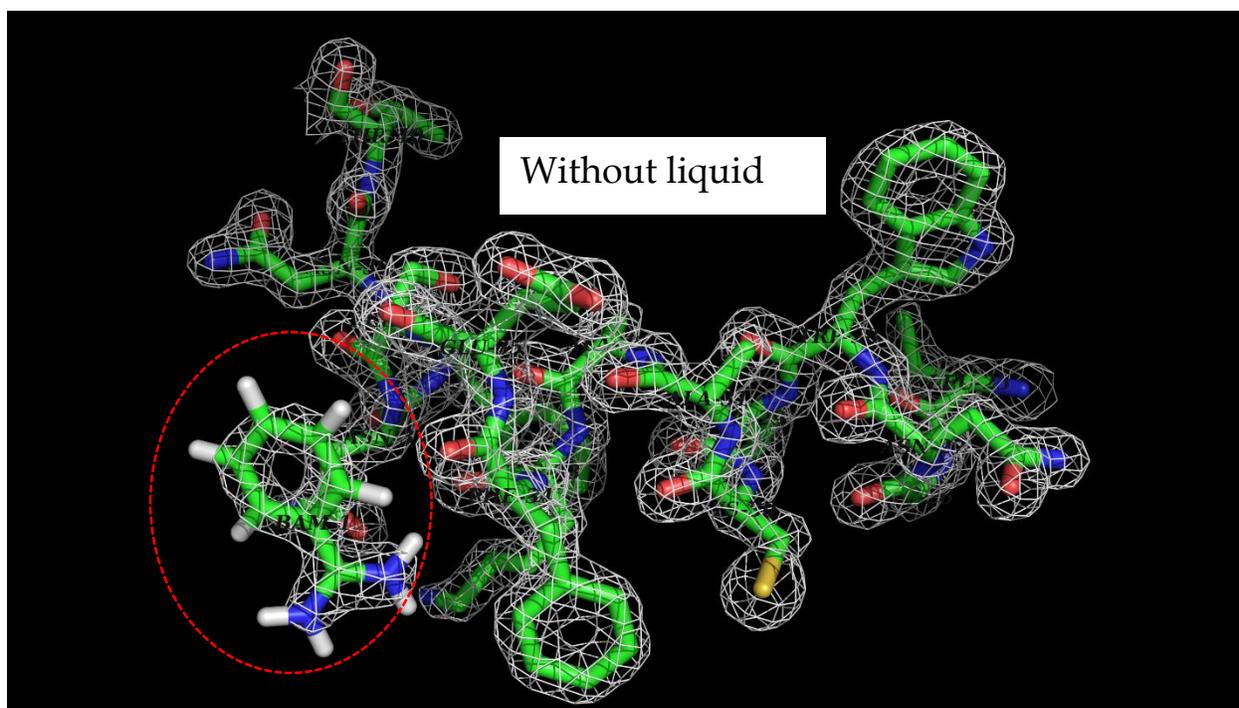
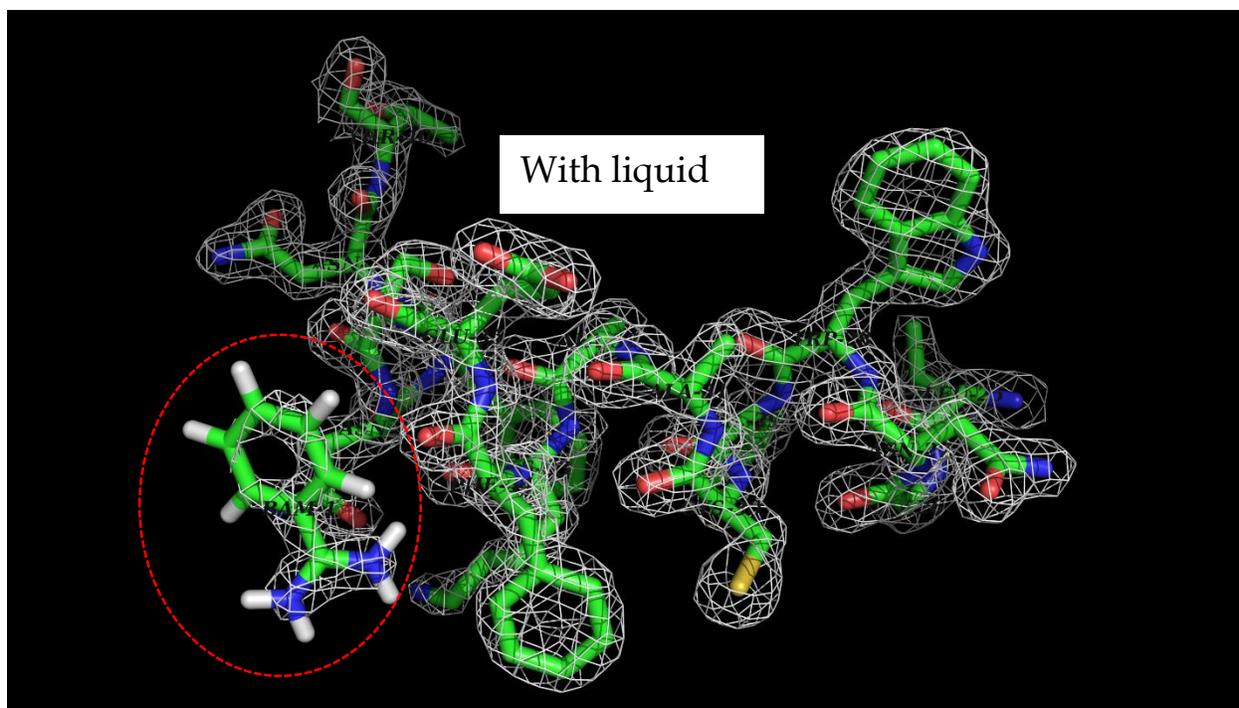
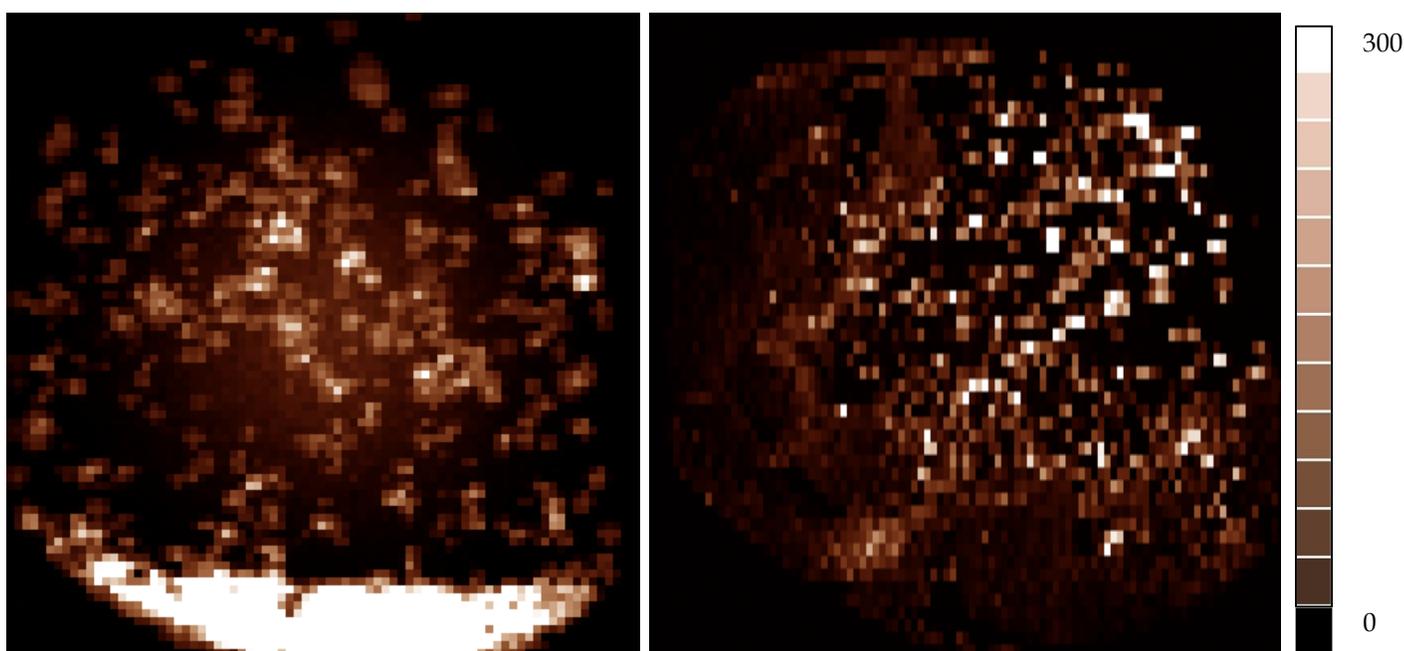


Figure 44. Refined density maps of lysozyme with benzamidine ligand (in red circle), contour level  $1.6\sigma$ .

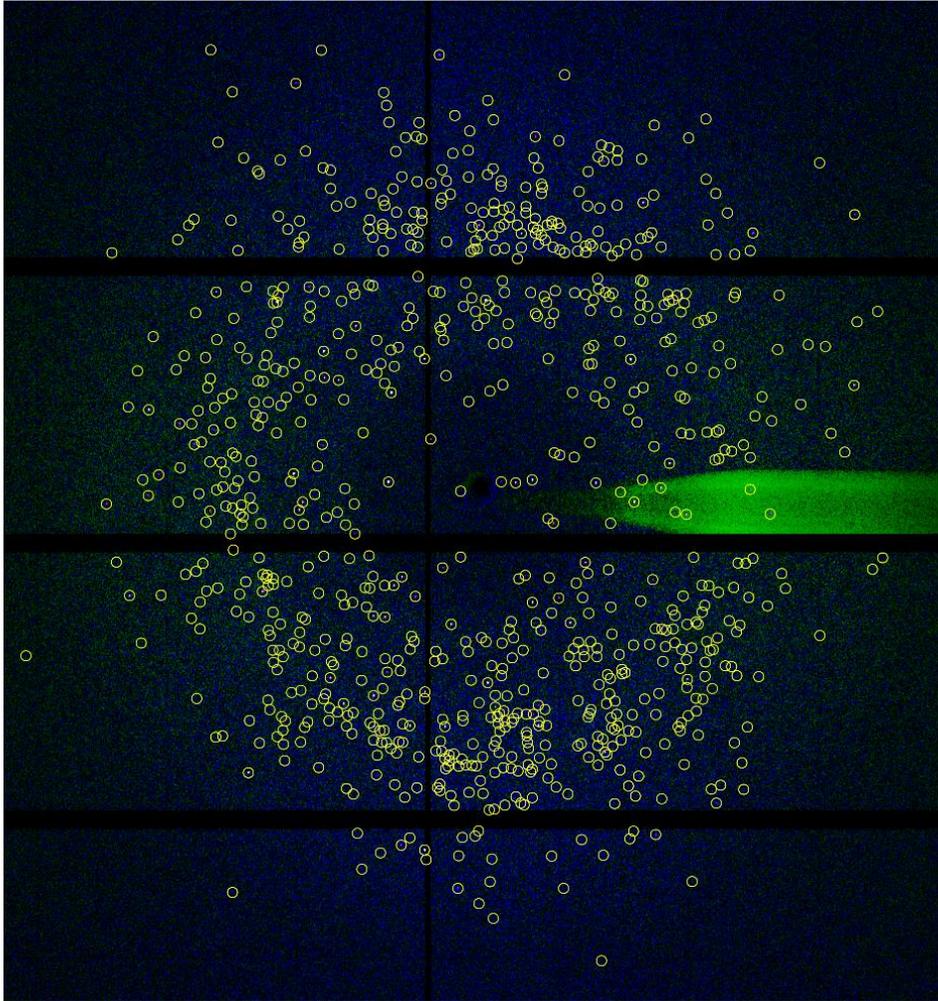
The maps are of higher quality for the dataset without liquid. It can be seen clearly when looking at the electron density near benzamidine, tryptophan and phenylalanine.

Most of the parameters improved with liquid removal. However, the origin of the huge difference of 50 % in the indexing rate between the two datasets was not immediately clear. Indeed the hit rate was not as different as seen from the composite images (hitmaps) generated for some drops for both datasets. In fact, we observed that in the drops with solution many crystals were at the bottom of the droplet while for the drops with the liquid removed the crystals were distributed quite homogeneously (**Fig.45**).



**Figure 45.** Left- composite image (hitmap) of a drop after data collection with solution; right- composite image of a drop without solution. Hitmap calculated as number of pixels above threshold (in this case 15 detector counts), white colour means 300 or more pixels.

After indexing using CrystFEL, we checked some of the diffraction patterns that the software was not able to index. We observed multiple hits from several crystals of frames where indexing failed (**Fig.46**). Sedimentation of the crystals in the drop and multiple crystals per frame limited indexing of the dataset with liquid.



**Figure 46. Diffraction pattern of lysozyme with multiple hits from different crystals.**

Adding the solution with the ligand (50 nl added to the droplet with a volume of 200 nl) obviously led to an excess of liquid and stirred up the crystals in the drop. Most likely the loss of adhesion due to the mechanical perturbation and a change in viscosity led to gravitational sedimentation for a number of crystals. The high crystal concentration in the sediment caused multiple orientations in the same diffraction pattern. These patterns count as hits for the software but cannot be indexed and these crystals are lost for the data collection. However, if after adding the ligand solution, liquid removal can be performed, it increases the chance to get a high quality dataset with higher hit rate and indexing rate.

## Discussion

The use of automated systems for crystallization and data collection are widespread now. However, those steps are separated due to the need of crystal mounting and cryocooling. In this work an implementation of the CrystalDirect plates and the data collection at room temperature without any crystal manipulation has been shown. At the current stage of development of the set-up, we can reach all compartments of the plate and if needed, a data collection of the whole plate can be performed in a straight forward way. The possibility to scan large amounts of crystals in a short time opens an opportunity to considerably speed up the process of crystal identification and characterization. Obviously, complete data collections will be dependent on the sample diffraction quality. For fragile crystals, the absence of mechanical stress during sample preparation can be expected to be a considerable advantage.

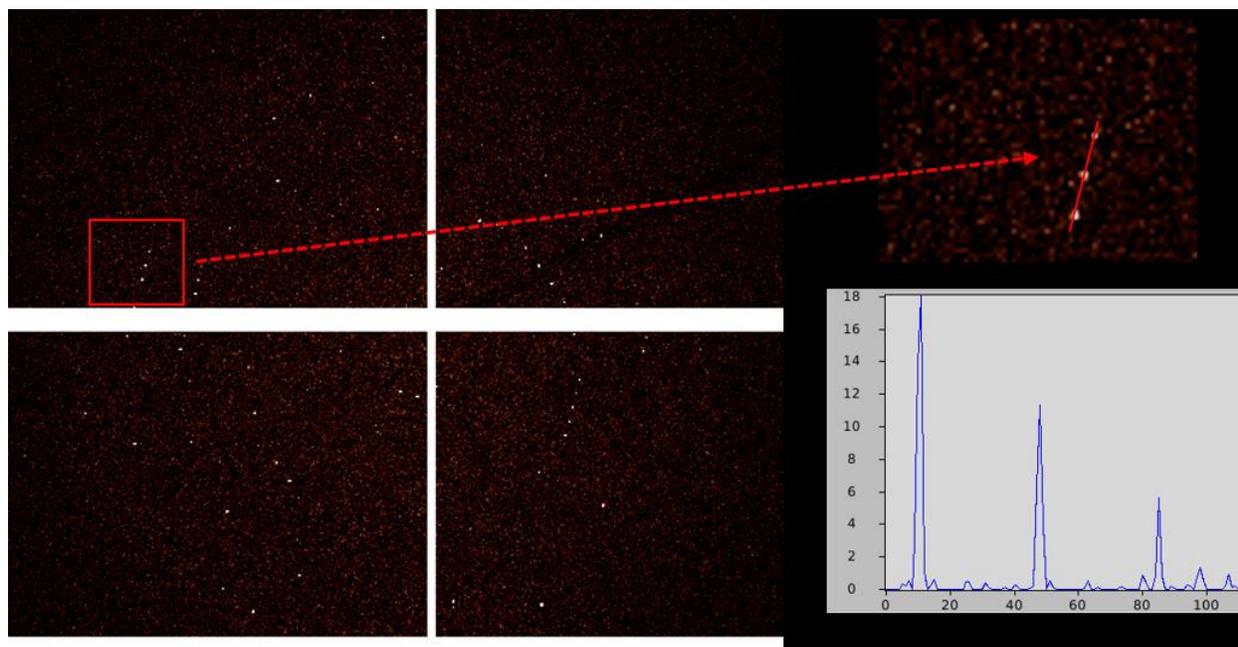
After several attempts to collect high resolution datasets with membrane proteins, we realized that data collection was suffering from the fact that the diffraction quality of the microcrystals was very low. This means that currently the data collection is limited by the inherent difficulty in obtaining high diffraction quality microcrystals. However, our results indicate that the combination of a microfocus beamline and the CrystalDirect plate scanning technique represents a very efficient system to support further studies on serial data collection on this type of samples.

We also explored data quality improvement via liquid removal for several proteins. In general, we observed an improvement of all parameters for the datasets with liquid removal, which can help to lower the sensitivity limit towards smaller crystal size. Importantly, the refined structures suggest that the liquid removal does not affect the hydration state within the crystal as seen from the conservation of the cell parameters. In this way the risk of induced disorder or change of the crystal lattice is very much reduced. As expected, there is also indication that it improved. Liquid removal clearly reduces considerably the background scattering. In addition, in some cases it allowed a more homogeneous distribution of crystals, which drastically improved indexing rates.

With high performance detectors and adapted computer software, it is now possible to collect data from a full plate with 96 drops in less than one beamtime shift (8 hours). The main advantages of this type of data collection are “crystal-handling-free” nature of setting up the plate, simple installation at the beamline, almost fully automatized data collection (The user needs to be physically at the beamline only 15 minutes during installation), and the possibility to work with fragile crystals. The integration of the CrystalDirect system approach considerably reduces the overhead for the identification of the crystals and data collection using X-rays which opens new possibilities for the related user community of synchrotron sources.

### 3.4 Experimental crystal growth on silicon nitride membranes.

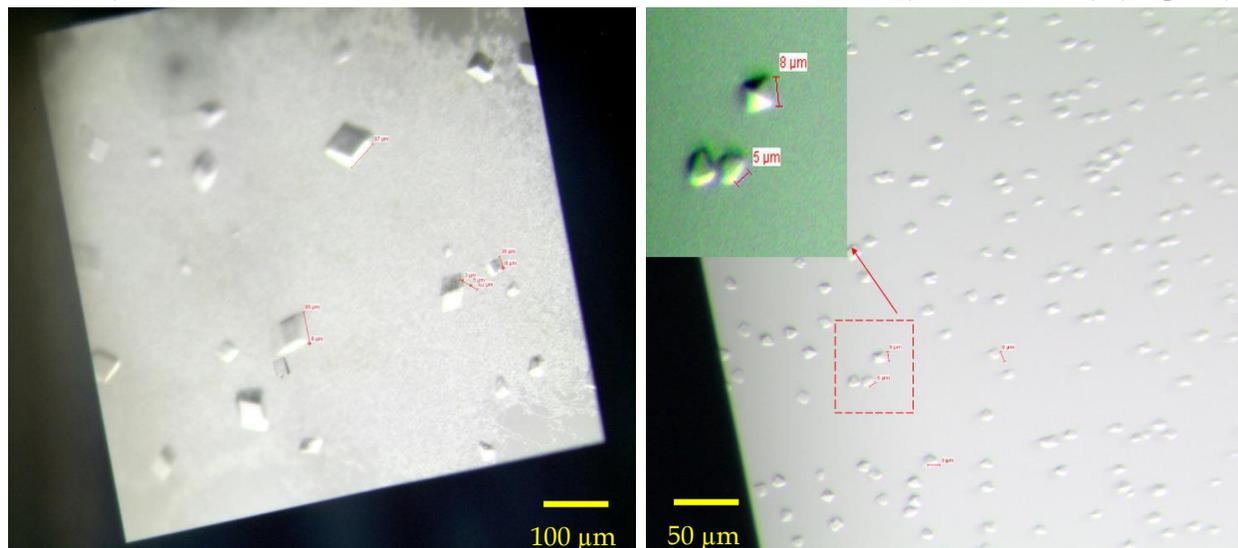
Silicon nitride membranes are well suited for raster-scanning protein crystallography due to their thriftiness and very low background scattering (**Fig.47**). For the first time implementation of silicon nitride membranes for the X-ray data collection using raster-scanning has been performed at ID13, ESRF<sup>39</sup>.



**Figure 47.** Nano-beam diffraction pattern (courtesy *Joyce Woodhouse, IBS*) of 2 micron sized crystal of fluorescent protein (rsEGFP2) (mount: silicon nitride membrane sandwich, 200 nm spot size, 14.85 keV photon energy, EIGER 4M detector). The example of a system external to this thesis work has been chosen because none of the other sample systems available had a sufficiently uniform crystal size distribution in the 1 to 2 micron range and good diffraction quality.

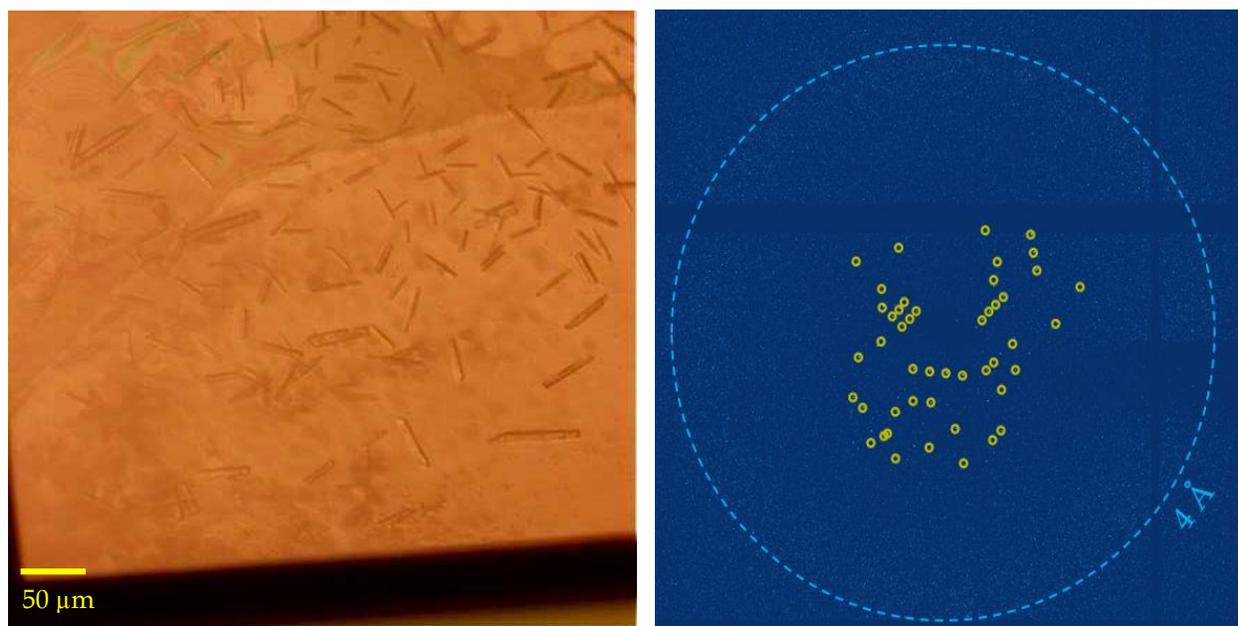
If protein microcrystals are fragile or scarce, it is very important to avoid unnecessary movements or transfer of the sample. In paragraph 3.2.3 the description of the method of crystal growth can be found. We tried to grow lysozyme and 2 membrane proteins on the silicon nitride membranes: AcrB and 5HT<sub>3</sub>-receptor. We were not able to obtain lysozyme crystals on the membrane, which might be due to incompatibility with the membrane surface. However, we refrained from further investigation as our focus was on membrane proteins. In contrast to lysozyme, membrane proteins were successfully grown after 3-4 days.

For AcrB we obtained crystals with sizes of 5-9  $\mu\text{m}$  on small membranes (size 1.5\*1.5 mm), and of 10-90  $\mu\text{m}$  on medium sized membranes (2.5\*2.5 mm) (**Fig.48**).



**Figure 48.** Crystals of AcrB grown directly on the  $\text{Si}_3\text{N}_4$  membranes; left-medium size membrane ( $2.5*2.5 \text{ mm}^2$ ), right-small size membrane ( $1.5*1.5 \text{ mm}^2$ ).

For 5HT<sub>3</sub>-receptor, crystals were from 10-60  $\mu\text{m}$  in size for medium sized membrane (**Fig.49**).



**Figure 49.** Crystals of 5-HT<sub>3</sub> receptor grown in-situ on the silicon nitride membranes and diffraction pattern of 5-HT<sub>3</sub> receptor to 6.4  $\text{\AA}$ .

Data were collected using the raster-scanning mode, however the resolution obtained from both proteins did not exceed 6 Å. We refrained from a collection of a full dataset due to the limited resolution.

## Discussion

We developed the method of raster-scanning using silicon nitride membranes and tested it at ID13. Data were collected with very low background scattering due to the very thin and amorphous silicon nitride membranes.

To our knowledge, protein crystals were grown for the first time directly on the silicon nitride membranes. Although we were not able to collect a full dataset, we demonstrated that it is possible to grow crystals straight on this solid support and without transfer of the crystals to perform an X-ray data collection at room temperature. However, silicon nitride membranes, maybe due to their comparably hydrophobic nature (e.g. as compared to glass), are obviously leading to unexpected results in crystallization (failure of lysozyme, very moderate success with membrane proteins). Further investigation in growing crystals directly on these membranes could take considerable efforts. Nevertheless, it might be an interesting approach to be considered for fragile nano- or micro-crystals when standard methods fail.

### 3.5 Dynamics of ubiquitin using silicon nitride membranes

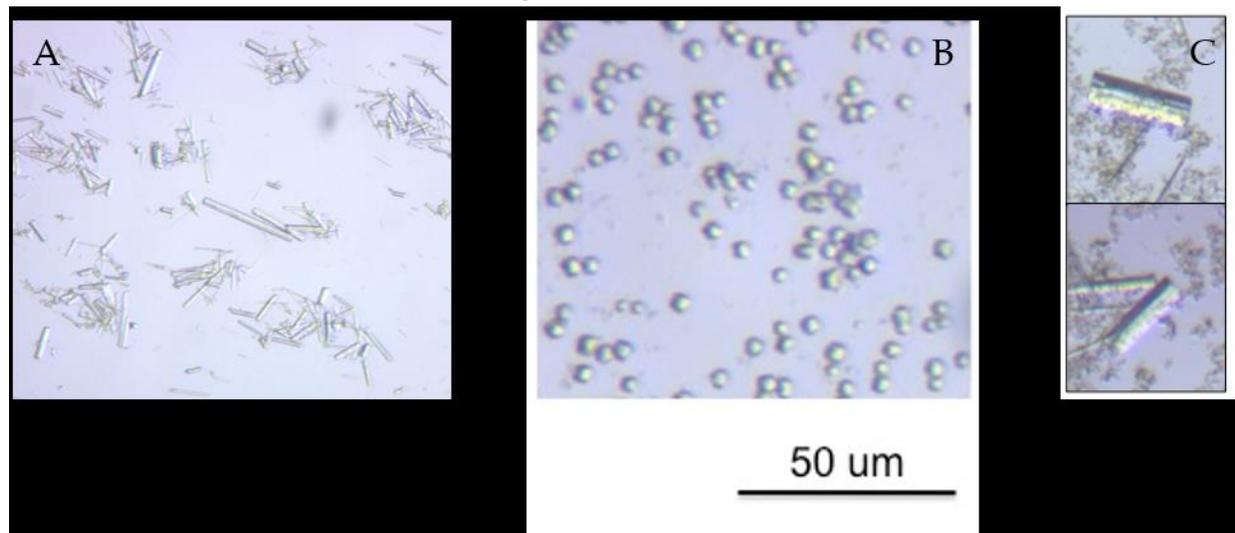
Room temperature X-ray crystallography opens the opportunity to examine the spatial distribution of individual atomic configurations at temperatures closer to the natural state (bond lengths and angles) and collective motions of the protein (torsion angles and non-bonded interactions). However, to complete structural characterization of the dynamics of a protein, X-ray diffraction methods (XRD) should be combined with other experimental techniques, for example with NMR, which can provide mechanistic detail at very high structural resolution into protein folding, catalytic turnover of enzymes, lowly populated “invisible” excited state structures and important thermodynamic quantities including entropy and heat capacity<sup>97</sup>. Nevertheless, the information derived from NMR does not always include structural coordinates, so to bypass this limitation, NMR studies are often complemented with computational simulations of protein dynamics using molecular dynamics (MD) simulations, where the three-dimensional positions of each atom in the protein and solvent are computed over time using analytically determined interaction forces. Recently P. Ma et al.<sup>45</sup> showed that a combination of NMR, MD and XRD techniques provides evidence for an overall rocking motion in protein crystals. They were able to determine the range of possible correlation times of this rocking motion to 0.1–100  $\mu$ s and they suggested that protein dynamics could be affected by the crystal packing.

Ubiquitin is a small protein that is often used as a model system for studying protein structural dynamics<sup>98,99,100</sup>. Mainly for this reason and also due to the ease of handling and crystallization, ubiquitin was chosen for this work.

Here we performed a comparison between three different crystal forms of ubiquitin collected at room temperature, further RT (this study) with structures from PDB database<sup>101</sup> (code 3ONS, 4XOL, 4XOK) collected at cryo-temperatures (CT) using a variety of analytical tools, among which is ensemble refinement.

We prepared and collected X-ray data from three different crystal forms of ubiquitin using the raster-scanning approach on silicon nitride membranes<sup>39</sup>.

These crystals are further named as ubiquitin MPD, ubiquitin cubic and ubiquitin rod, due to the morphology of the crystals and to the precipitant agents (**Fig.50**). MPD crystals of ubiquitin were needle-shaped with the size from 10 to 50  $\mu\text{m}$ . Cubic crystals of ubiquitin were typically 5-10  $\mu\text{m}$  in size. And rod shaped crystals of ubiquitin were in the range from 10-45  $\mu\text{m}$ .



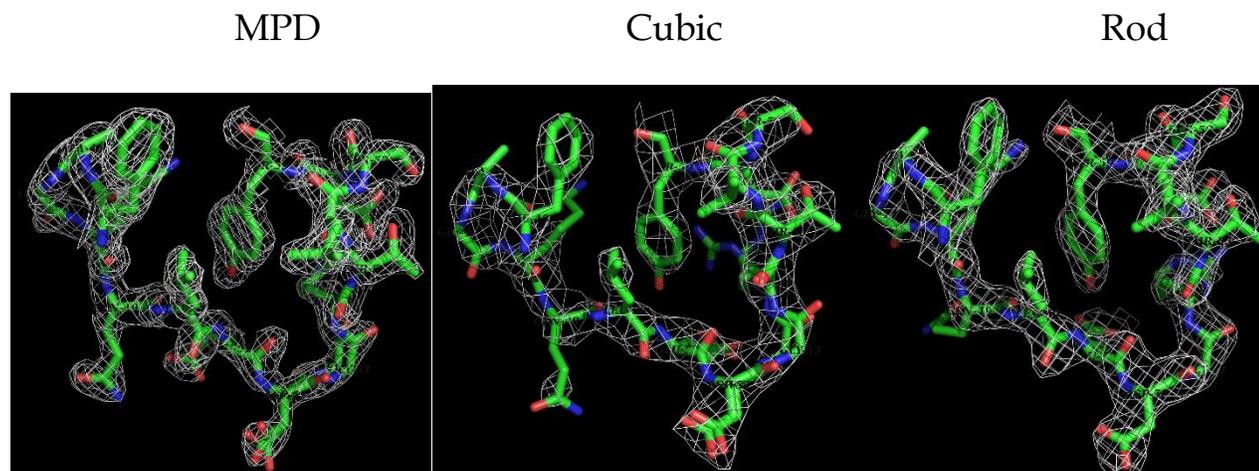
**Figure 50. Different crystal forms of ubiquitin. A) MPD B) Cubic C) Rods**

**From MPD crystals of ubiquitin** (space group  $P 3_221$ ) we collected 42164 frames from three medium sized ( $2.5 \times 2.5 \text{ mm}^2$ ) silicon nitride membranes, of which 6907 images were indexed with CrystFEL. Based on the initial data processing criteria the highest resolution was chosen as 2.0  $\text{\AA}$ . This crystal form has only one ubiquitin molecule in the asymmetric unit, in comparison to cubic and rod crystals, which have two and three molecules per asymmetric unit respectively.

**From cubic shaped crystals of ubiquitin** (space group  $P 4_332$ ) we collected 54670 frames from 2 medium sized silicon nitride membranes, from which 3970 images were indexed with CrystFEL. The highest resolution was chosen as 3.0  $\text{\AA}$ .

**From rod-shaped crystals of ubiquitin** (space group  $P 2_12_12_1$ ) we collected 173334 frames from 1 medium sized silicon nitride membrane at the nanobranh of ID13, of which 27723 images were indexed with CrystFEL. For this dataset the highest resolution was chosen as 1.8  $\text{\AA}$ . All data were converted to the mtz format and phased by molecular replacement using Phaser<sup>64</sup> with PDB entry 3ONS,

4XOK, 4XOL, respectively as the starting model. The parameters of data processing are presented in the **Table 10** in comparison with data from the PDB database. Manual modifications in real space were performed by Coot<sup>65</sup>. Refined electron-density maps are shown in **Fig.51**. The figures were produced using PyMOL<sup>66</sup>.



**Figure 51. Electron density near residues 45-59 for different crystal forms of ubiquitin (contour level 1.6 $\sigma$ ).**

Parameters	Ubiquitin MPD		Ubiquitin Cubic		Ubiquitin Rod	
Data collection temperature, K	297	100	297	100	297	100
N of molecules per unit cell	1		2		3	
Space group	P 3 <sub>2</sub> 21		P 4 <sub>3</sub> 32		P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
Unit cell, Å	a=b=49.17, c=63.16, $\alpha=\beta=90^\circ$ , $\gamma=120^\circ$	a=b=48.41, c=61.97, $\alpha=\beta=90^\circ$ , $\gamma=120^\circ$	a=b=c=106.70, $\alpha=\beta=\gamma=90^\circ$	a=b=c=104.95, $\alpha=\beta=\gamma=90^\circ$	a=44.55, b=51.37, c=95.19, $\alpha=\beta=\gamma=90^\circ$	a=43.72, b=50.36, c=93.46, $\alpha=\beta=\gamma=90^\circ$
No. collected images	42 164	N/A	54 670	N/A	173 334	N/A
Average hit rate, %	33	N/A	13	N/A	17.8	N/A
No. of indexed images	6907	N/A	3967	N/A	27 723	N/A
Indexing rate, %	50	N/A	55.8	N/A	88.9	N/A
No. of total reflections	888847 (38532)	N/A	3460856 (157377)	N/A	6097770 (264215)	N/A

No. of unique reflections	11561 (773)	N/A	8774 (591)	N/A	39175 (2639)	N/A
Resolution range, Å	42.6-2.0 (1.85-1.8)	30.0-1.8 (1.86-1.8)	46.0-3.0 (3.05-3.0)	35.0-2.9 (3.0-2.91)	47.6-2.0 (2.06-2.0)	46.7-2.2 (2.28-2.2)
Completeness, %	100	98.8	100	98.8	100	93.0
I/ $\sigma$ (I)	3.19 (1.92)	5.01 (-)	5.29 (2.37)	2.11 (1.38)	5.36 (2.46)	16.04 (7.59)
B factor from Wilson Plot, Å <sup>2</sup>	25.74	25.1	66.47	N/A	53.24	N/A
Overall Rsplit	28.7 (49.9)	N/A	15.1 (49.8)	N/A	14.4 (49.8)	N/A
Rmerge	N/A	7.1	N/A	6.7	N/A	8.3
CC*	0.97 (0.91)	N/A	0.99 (0.92)	N/A	0.99 (0.93)	N/A
Rwork/Rfree, %	22/26	18/21	25/30	24/27	26/31	30/32
Rwork/Rfree (ensemble refinement), %	21/23	16/19	22/24	21/24	25/28	28/35
Ramachandran favored, %	98.8	N/A	94.7	N/A	98.6	N/A
Ramachandran allowed, %	1.2	N/A	3.0	N/A	1.4	N/A
Rotamer outliers	0	N/A	1.40	N/A	0.5	N/A
PDB entry	-	3ONS	-	4XOL	-	4XOK

**Table 10. Parameters of the data processing for ubiquitin crystals. Initial data processing criteria was  $I/\sigma(I) > 2$ ,  $R_{split} < 50\%$  and 100% completeness. PDB entries for RT data will be deposited to the PDB as soon as these data are submitted for publication. Values in parentheses are for the highest resolution shell.**

According to the electron density maps, obtained from RT data, we added alternative conformations to some residues and compared them with the data collected at CT (**Table11**).

Protein	MPD 297K	MPD 100K (3ONS)	Cubic 297K	Cubic 100K (4XOK)	Rod 297K	Rod 100K (4XOL)
Space group	P <sub>3</sub> 2 <sub>1</sub>		P <sub>4</sub> 3 <sub>2</sub>		P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
Total N of alternative conformations	10	0	8	0	3	0
Residue numbers with alternative confirmations	6, 13, 24, 29, 34, 40, 52, 53, 63, 69	-	51, 52, 53, 54, 72-chain A 42, 51, 54-chain B	-	72 -chain B 21, 40 -chain C	-

**Table 11. Alternative conformations in ubiquitin.**

A number of alternative conformations that can be seen from RT data can be explained with more detailed electron density maps. It was already suggested before that RT allows the observation of the dynamics of protein residues that can't be seen from CT data<sup>102</sup>. Also some of these alternative conformations have already been confirmed to exist by NMR method and MD simulations<sup>45</sup>.

### 3.5.1 Protein and solvent channels can be affected by cryocooling.

To understand what is happening inside the molecules of ubiquitin we generated a difference distance matrix (DDM) for each chain of the protein and compared them in pairs of 297 K-100 K for all PDB files. A DDM is a calculation of the distances between all pairs of C-alpha atoms in the molecule of every possible pair of the amino acids common to both proteins to generate an initial distance matrix. For generation of the DDM plots in this work, we removed all alternative conformations with lower occupancy from the PDB files for RT data. The two matrices for each protein are subtracted from one another, yielding the DDM. This allows us to assess quantitatively the structural similarity or

dissimilarity between two structures where red colour means an expansion of the RT structure and blue indicates expansion of the CT structure.

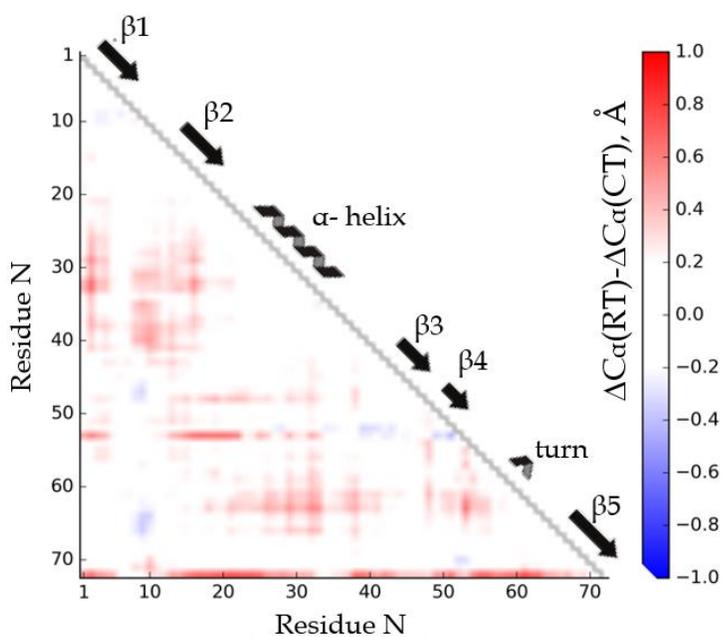


Figure 52. DDM for ubiquitin MPD 100K versus ubiquitin MPD 297K.

For the RT structure of MPD ubiquitin (Fig.52) we clearly observe an expansion of the chain near the  $\alpha$ -helix, in the region of turn, at the end of the chain, between  $\beta$ 3 and  $\beta$ 4 strands and near the residue 53.

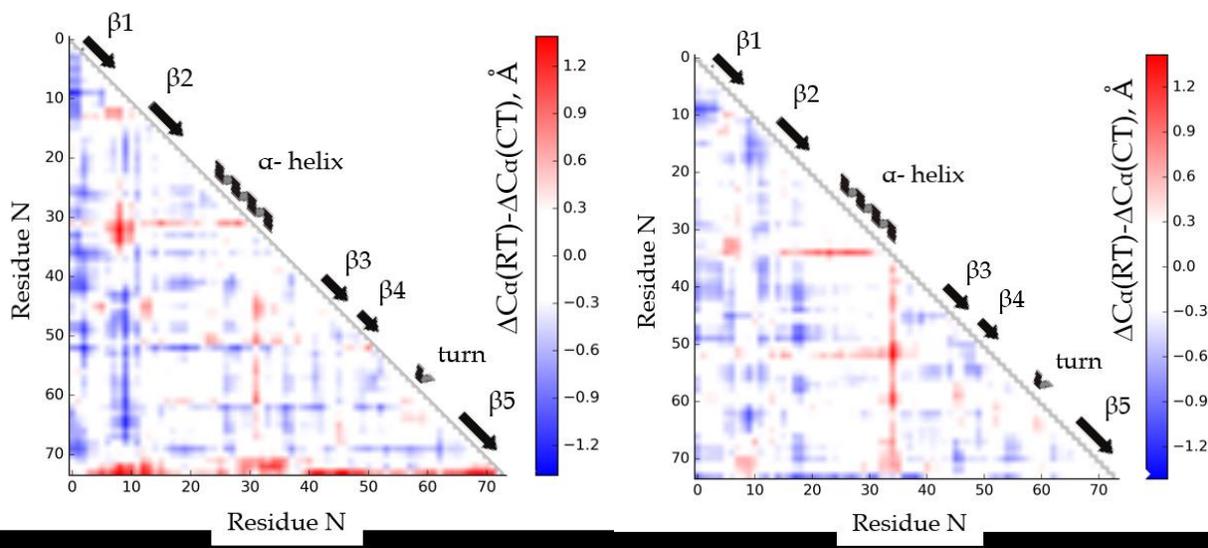
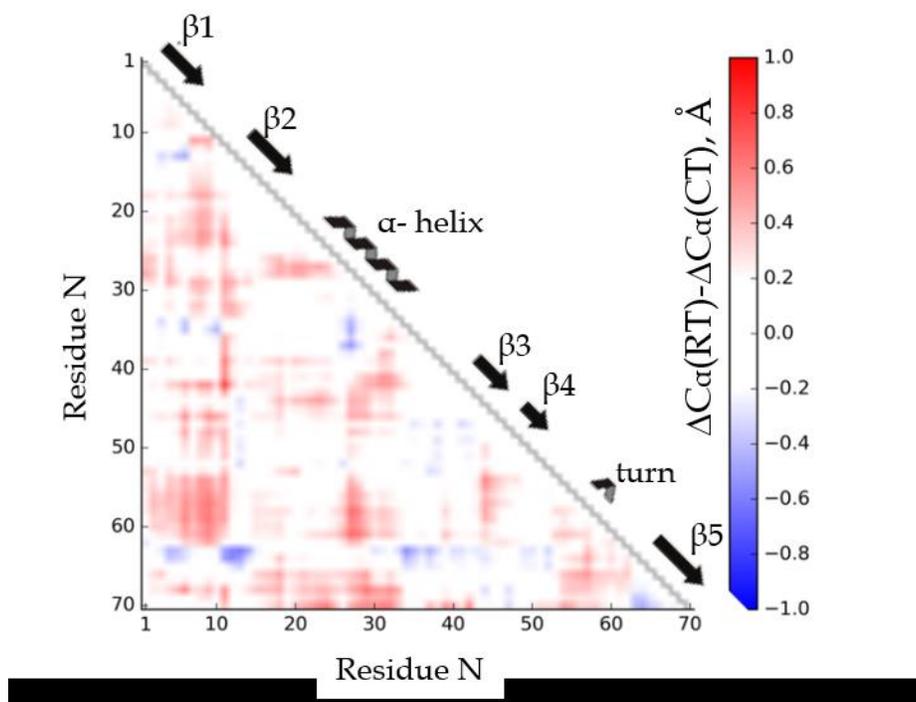


Figure 53. DDM plot for ubiquitin cubic 100K versus 297K; left-chain A, right-chain B.

For chain A of cubic ubiquitin (**Fig.53, left**) we observe an expansion which occurs in the area of the  $\alpha$ -helix and at the end of the chain for the RT structure. For the CT structure the largest expansion is observed in the beginning of the chain, near residues 8 and 52-53.

For chain B of cubic ubiquitin (**Fig.53, right**) we observe the largest difference in  $C\alpha$  distances near residue 34 for the RT structure.

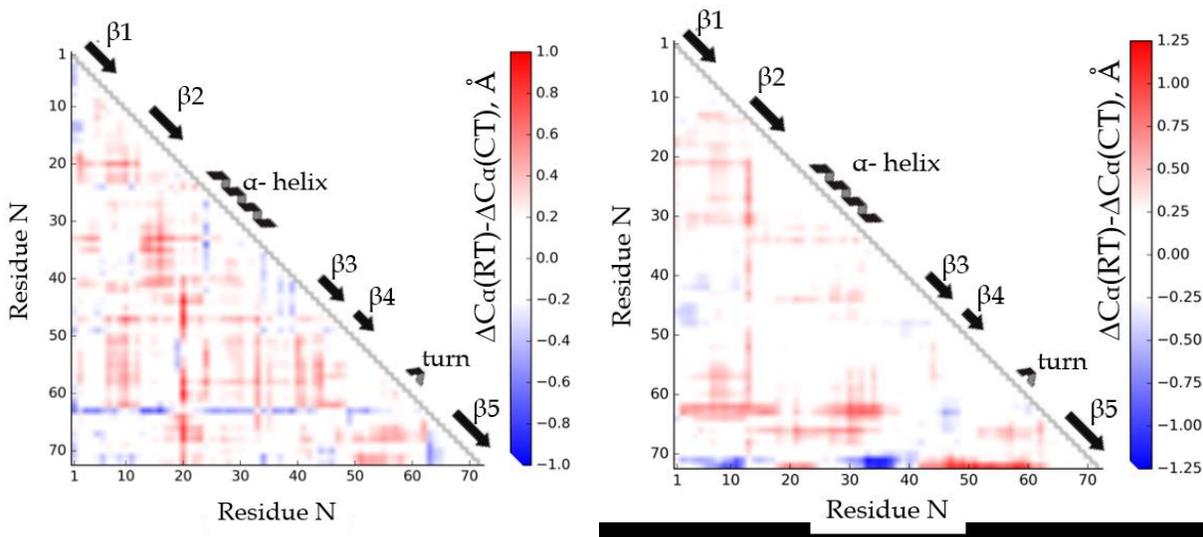


**Figure 54.** DDM for rod-shaped ubiquitin 100k versus 297K, chain A.

For the RT structure of rod ubiquitin, chain A (**Fig.54**) an expansion of the chain influences almost all regions, except the region near the residues 62-63, which shows a small expansion for the CT structure.

For chain B of rod shaped crystals of ubiquitin (**Fig.55, left**) the situation is very similar to chain A, however an expansion of the region near residues 62-63 is more pronounced. Also for the RT structure we observe an expansion near residue 20.

For chain C of rod ubiquitin (**Fig.55, right**) we observe the largest difference near residues 12, 62-63 and at the end of the chain.



**Figure 55. DDM for rod shaped crystals of ubiquitin 100k versus 297K, chain B (left) and chain C (right).**

The common feature of all DDM plots is that the largest difference in distance for all crystals occurs in the  $\alpha$ -helix regions, near the turn and at the end of the chain. Which means that residues around the  $\alpha$ -helix are the most prone to contract upon flash cooling. Also, we observed that there is an expansion in the loop regions for cubic crystals and for chain A and B of rod crystals. The residues that have the highest difference in distance and probably are involved in interactions are:

- for MPD: 2, 32, 52
- for Cubic: 1-4, 8, 34, 47, 52-53, 63-64
- for Rod: 2, 11, 20, 32-34, 41, 52-53, 63

From DDM plots we observed a large difference that occurs inside the protein near the residues 52(ASP)-53(GLY). A flip of the peptide plane 52(ASP)-53(GLY) together with the creation of a hydrogen bond between residues 24 and 53 of ubiquitin in solution was shown by NMR and MD simulations<sup>103,45</sup> techniques. In our datasets, we observed several alternative conformations near this residues in

cubic and MPD ubiquitin, however we were not able to confirm this flip from the positions of the residues in the PDB files.

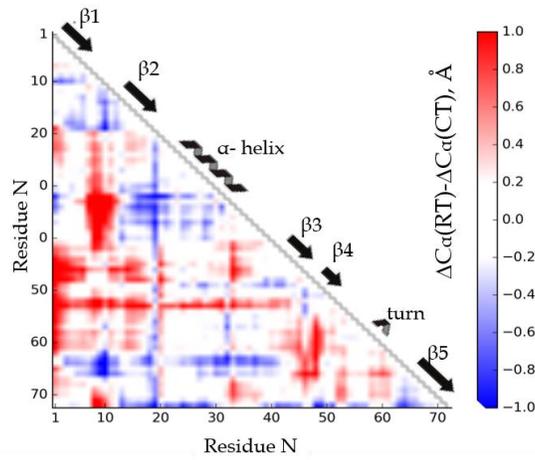
We decided to compare our coordinates with the NMR data for ubiquitin in solution (PDB code 1D3Z), where no crystal packing constraints exist. We calculated the root-mean-square deviation (RMSD) between all PDB files and compared them (**Table 12**).

Ubiquitin	MPD 100K	MPD 297K	Cubic A 100K	Cubic B 100K	Cubic A 297K	Cubic B 297K	Rod A 100K	Rod B 100K	Rod C 100K	Rod A 297K	Rod B 297K	Rod C 297K	Solution
MPD 100K	0	0.31	0.552	0.658	0.61	0.696	0.694	0.686	0.659	0.665	0.623	0.651	0.512
MPD 297K	0.31	0	0.472	0.49	0.569	0.556	0.586	0.574	0.518	0.556	0.466	0.439	0.415
Cubic A 100K	0.552	0.472	0	0.487	0.407	0.521	0.513	0.539	0.477	0.549	0.489	0.481	0.459
Cubic B 100K	0.658	0.49	0.487	0	0.614	0.473	0.573	0.613	0.564	0.614	0.525	0.483	0.525
Cubic A 297K	0.61	0.569	0.407	0.614	0	0.436	0.558	0.62	0.558	0.692	0.604	0.596	0.583
Cubic B 297K	0.696	0.556	0.521	0.473	0.436	0	0.567	0.674	0.576	0.62	0.551	0.556	0.523
Rod A 100K	0.694	0.586	0.513	0.573	0.558	0.567	0	0.36	0.434	0.311	0.318	0.38	0.476
Rod B 100K	0.686	0.574	0.539	0.613	0.62	0.674	0.36	0	0.469	0.341	0.278	0.411	0.496
Rod C 100K	0.659	0.518	0.477	0.564	0.558	0.576	0.434	0.469	0	0.44	0.403	0.345	0.413
Rod A 297K	0.665	0.556	0.549	0.614	0.692	0.62	0.311	0.341	0.44	0	0.247	0.296	0.448
Rods 297K B	0.623	0.466	0.489	0.525	0.604	0.551	0.318	0.278	0.403	0.247	0	0.292	0.458
Rod C 297K	0.651	0.439	0.481	0.483	0.596	0.556	0.38	0.411	0.345	0.296	0.292	0	0.372
Solution	0.512	0.415	0.459	0.525	0.583	0.523	0.476	0.496	0.413	0.448	0.458	0.372	0

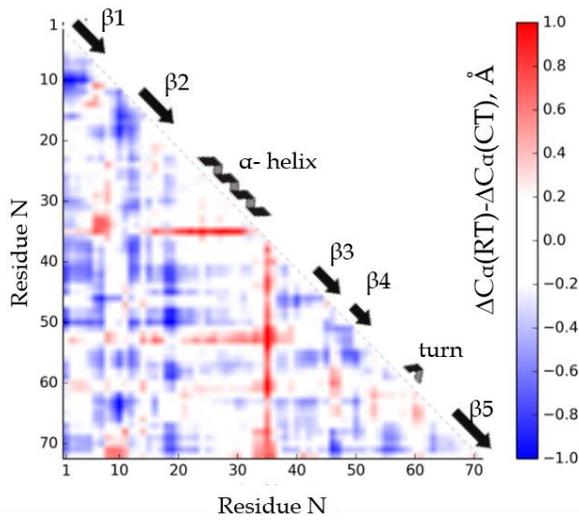
**Table 12.** RMSD of atomic positions of ubiquitin crystals in comparison with ubiquitin in solution.

Among all crystal structures of ubiquitin, the highest similarity to the solution structure of ubiquitin has been found for chain C of the rod ubiquitin 297K structure. We used this structure as a reference and compared the DDM plots with other RT structures (**Fig.57**).

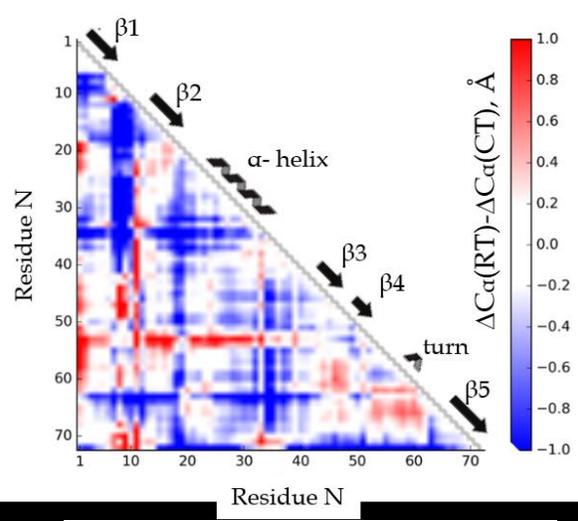
MPD 297K vs Rod chain C 297K



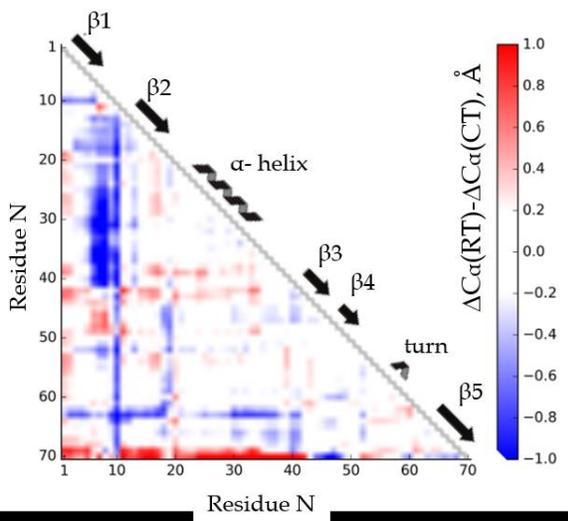
Cubic Chain A 297K vs Rod chain C 297K



Cubic Chain B 297K vs Rod chain C 297K



Rods chain A 297K vs Rod chain C 297K



Rods chain B 297K vs Rod chain C 297K

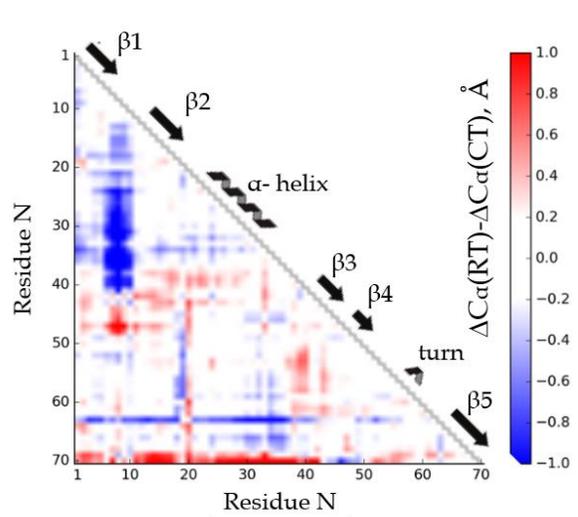


Figure 56. DDM plots in comparison with ubiquitin Rod 297K, chain C.

We observed the highest difference in the regions of the alpha helix and at the end of the chain, near the residues of interest 52-53 and near residues 10, 20, 34 and 63.

The contraction of some regions that we saw from DDM plots influences not only the protein structure, but also the solvent channels, which can probably be seen from the crystal's solvent content. To analyze the crystal's solvent content, we measured changes of unit-cell volume, using CCP4<sup>104</sup> and protein volume using VOIDOO program (<http://xray.bmc.uu.se/usf/voidoo.html> (**Table13**)).

Protein	MPD 100K	MPD 297K	Cubic 100K	Cubic 297K	Rods 100K	Rods 297K
Solvent content, %	52.2 ↑4.4%	54.5	57 ↑3.5%	59	42.2 ↑7.6%	45.4
Unit cell volume*1000, Å <sup>3</sup>	126 ↑4.8%	132	1160 ↑4.7%	1215	206 ↑6.3%	218
Protein volume*100, Å <sup>3</sup>	145 ↑13.1%	164	300 ↑6%	318	430 ↑1.2%	435

**Table 13. Parameters of solvent content, unit cell and protein volume for ubiquitin crystals.**

From the **Table 13** we can see that unit-cell volumes ratio (RT/CT) for all forms of ubiquitin are increased in average for 5.3 % at RT, when previous similar studies showed that the mean increase is 3%<sup>105,102,106</sup>. However, the protein volume increases with a different ratio from 1.2% (rod shaped ubiquitin) to 13% (MPD ubiquitin).

The solvent content increases from 3.5 % (cubic ubiquitin) to 7.6 % (rod ubiquitin) at room temperature. The highest solvent content was found for cubic crystals of ubiquitin, reaching 59 % at RT. Rod shaped crystals of ubiquitin have the lowest solvent content - 45 % at RT. This means that rod ubiquitin has higher packing density than MPD or cubic ubiquitin due to the reduced number of solvent channels. The higher packing density decreases the possibility to perform dynamical movements inside the molecule. This can explain the fact that the protein

volume for rod-shaped crystals at RT increases only by 1.2 % in comparison with MPD crystals and cubic ubiquitin. Another difference, which can be linked to such an important change in protein volume is that cubic and rods crystals were grown in PEG where MPD-ubiquitin was grown in MPD. From here two hypotheses can be established:

1) The low solvent content and reduced number of solvent channels for rod-shaped crystals leads to the creation of a denser packing, which reduces the possible movements in the protein, and causes a change in protein volume between CT and RT data of only 1.2%. At the same time the high solvent content and the increased solvent volume, that we observed for MPD and cubic data, leads to a significant change (13% and 6% respectively) in protein volume between CT and RT.

2) As crystals were grown in 2 different precipitants (MPD and PEG), this difference can have an influence on their protein volume. MPD is known to act as a cryoprotectant due to the fact that it binds to the different locations on a protein secondary structure, water is removed, which prevents ice formation for cryo-crystallography techniques<sup>107</sup>. However, at room temperature, the environment of MPD precipitant probably somehow influences on the creation of new chemical bonds with the protein. MPD is known to modify drastically the dielectric constant of the solvent<sup>108</sup>. This means that all interactions that are involving charges (or partial charges) will be affected, which can lead to the modification of weak interactions.

An increase in protein volume at room temperature can lead to the formation of new lattice contacts. To check this, we investigated the intermolecular interactions in ubiquitin crystals, using the PDBe PDBePISA (PISA) interactive web-based tool<sup>109</sup>. This tool performs a detailed analysis of surfaces, interfaces and assemblies starting from PDB files. The interface area of the residues involved in the interaction together with potential salt bridges and hydrogen bonds are presented in the **Table 14**.

Samples	Solvent accessible area, A <sup>2</sup>	Interface area, A <sup>2</sup>	Potential hydrogen bonds and salt bridges, length of the bond in Å
MPD 100K	4250.9	932.7	<ol style="list-style-type: none"> <li>Ala46-Asp32 2.83 Å°</li> <li>Glu64-Thr22 2.49 Å°</li> <li>Asn60-Gln31 3.05 Å°</li> <li>Thr9-Glu34 3.02 Å°</li> <li>Glu18-Gln49 2.80 Å°</li> <li>His68-Asp32 2.69 Å°</li> <li>Lys11-Lys33 3.44 Å°</li> <li>Lys11-Glu34 3.68 Å°</li> </ol>
MPD 297K	4275.5	858.4	<ol style="list-style-type: none"> <li>Ala46-Asp32 2.79 Å°</li> <li>Gln62-Asp39 2.86 Å°</li> <li>His68-Asp32 2.59 Å°</li> <li>Glu64-Thr22 2.84 Å°</li> <li>Asn60-Gln31 3.14 Å°</li> <li>Glu18-Gln49 3.09 Å°</li> </ol>
Cubic 100K	9176.5	2265.9 chainA=1085.8 chainB=1180.1	<ol style="list-style-type: none"> <li>Leu8-Gln49 2.88 Å° (B-A)</li> <li>Glu34-Arg72 2.40 Å° (B-A)</li> <li>Lys6-Glu64 2.48 Å° (A-A)</li> <li>Met1-Asp32 3.52 Å° (B-B)</li> <li>Ser20-Asn25 2.92-3.01 Å° (B-B)</li> <li>Ser20-Thr22 2.53-2.64 Å° (B-B)</li> <li>Lys48-Asp39 3.58 Å° (B-B)</li> <li>Arg74-Lys63 3.57 Å° (A-B)</li> <li>His68 -Glu64 3.08 Å° (A-A,B-B)</li> <li>Glu18-Lys29 3.64 Å° (B-B)</li> </ol>
Cubic 297K	9134.7	2462.8 chainA=1166.6 chainB=1296.2	<ol style="list-style-type: none"> <li>Lys63-Ala46 2.61 Å° (B-B)</li> <li>Lys11- Glu51 3.23 Å° (A-B)</li> <li>Glu18-Lys29 2.32-2.70 Å° (A-A,B-B)</li> <li>Leu8-Gln49 3.34 Å° (B-A)</li> <li>Glu34-Arg72 2.77-3.09 Å° (A-B)</li> <li>Lys6-Glu64 2.35-2.84 Å° (A-A)</li> <li>Met1-Asp32 2.40-3.49 Å° (A-A)</li> <li>Ser20-Asn25 2.83 Å° (A-A,B-B)</li> <li>Ser20-Thr22 2.35-2.38 Å° (A-A,B-B)</li> <li>Lys48-Asp39 2.41 Å° (B-A)</li> <li>Glu18- Asn25 2.79 Å° (A-A,B-B)</li> <li>Ser20-Glu24 3.12 Å° (B-B)</li> <li>Thr66-Glu64 2.46-3.54 Å° (B-B,A-A))</li> <li>Lys63-Glu51 2.79 Å° (B-C)</li> </ol>
Rod 100K	12926.5	2873.3	<ol style="list-style-type: none"> <li>Asp39-His68 3.48 Å° (A-C)</li> </ol>

		chainA=598 chainB=1826.7 chainC=448.6	<ol style="list-style-type: none"> <li>2. Glu16-Arg54 2.79-3.48 A° (B-C,C-A,B-A)</li> <li>3. Lys29-Glu18 3.50 A° (B-C)</li> <li>4. Lys63-Ala46 2.79 A° (A-C)</li> <li>5. Leu71-Leu8 3.03 A° (B-A)</li> <li>6. Lys48-Asp32 2.47-2.75 A° (B-C)</li> <li>7. Lys48-Lys33 3.19 A° (B-C,A-C)</li> <li>8. Ser57- Lys33 2.57-2.83 A° (B-C,A-C)</li> <li>9. Glu34-Asn60 2.82 A° (A-C)</li> <li>10. Ser20-Asn25 2.52-3.09 A° (B-C,C-A)</li> <li>11. Ser20-Thr22 2.35 A° (B-C,C-A)</li> <li>12. Ser20-Glu24 3.56 A° (B-C,C-A)</li> <li>13. Gln31-Lys63 2.51 A° (B-C)</li> <li>14. Glu24-Ser57 2.62 A° (C-A)</li> <li>15. Gln2-Asp52 3.18 A° (B-C)</li> </ol>
Rod 297K	13064.5	2573.4  chainA=652.8 chainB=1263 chainC=657.6	<ol style="list-style-type: none"> <li>1. Leu71-Leu8 3.04 A° (B-A)</li> <li>2. Lys48-Asp32 2.54-2.77 A° (B-C)</li> <li>3. Lys48-Lys33 3.39 A° (B-C)</li> <li>4. Ser57- Lys33 2.84 A° (B-C)</li> <li>5. Asn60-Glu34 3.59 A° (B-C)</li> <li>6. Asp58-Lys33 3.60 A° (B-C)</li> <li>7. Glu18-Asn 25 2.81-3.48 A° (A-B,C-B)</li> <li>8. Ser20-Asn25 2.61-3.14 A° (A-B,C-B)</li> <li>9. Glu24-Ser57 2.58-2.70 A° (A-B)</li> <li>10. Ser20-Thr22 2.64 A° (C-B)</li> <li>11. Asp39-His68 3.53 A° (A-C)</li> <li>12. Glu16-Arg54 3.16-3.17 A° (B-C,A-C,B-A)</li> <li>13. Lys29-Glu18 2.62 A° (A-B,C-B)</li> <li>14. Glu51-Met1 3.51 A° (B-C,A-C,B-A)</li> <li>15. Glu51-Lys63 2.68-2.97 A° (A-C)</li> </ol>

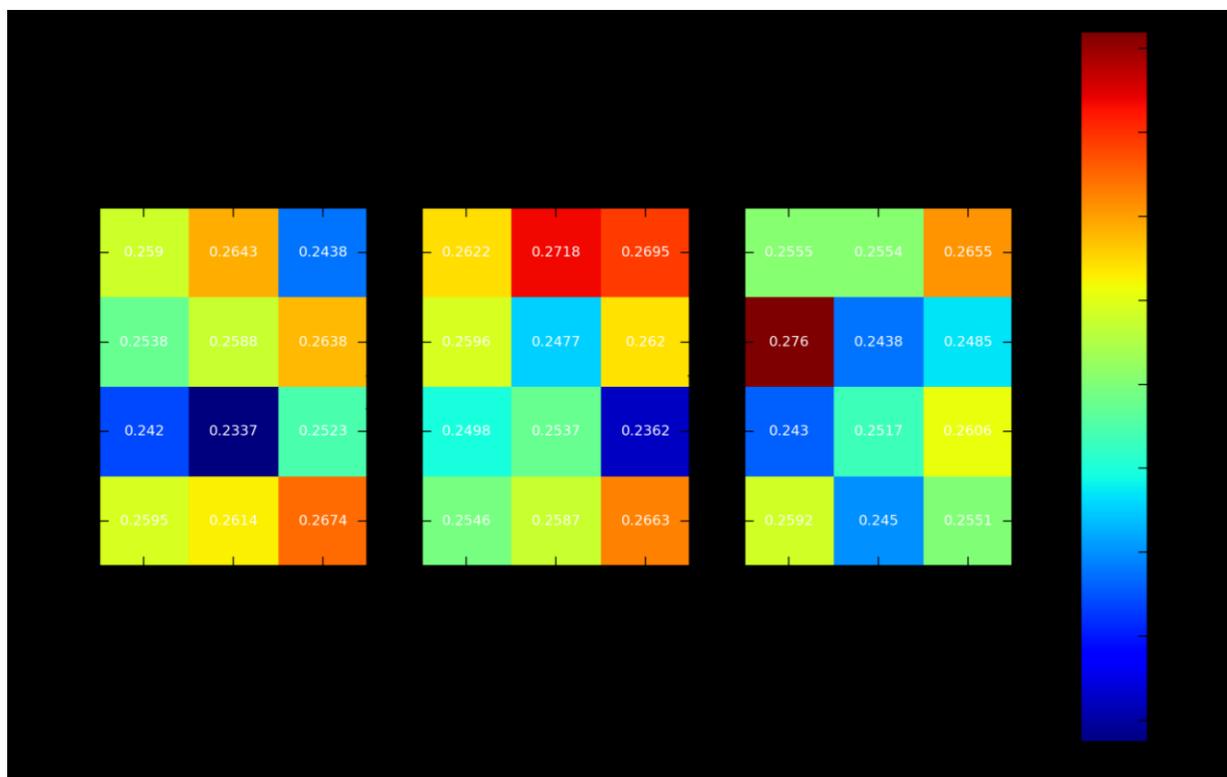
**Table 14. Potential salt bridges and hydrogen bonds inside Ubiquitin. In this table the distance for the strong hydrogen bonds counts from 2.2-2.5 Å, for salt bridges from 2.5-3.7 Å. Bonds coloured in blue can be found in both datasets (RT and CT), bonds coloured in red mean that new interaction occurred.**

The number of intermolecular bonds at RT increases only for cubic crystals. However, the formation of new interactions occurs in all cases. The solvent accessible area increases for MPD crystals and rod ubiquitin when the interface area decreases. For cubic crystals the results are reversed, the solvent accessible area

decreases with an increase of the interface area. In 2011 Fraser et al.<sup>102</sup> suggested that cryocooling creates a larger contact area by reducing the solvent channels. However, this is not the case for cubic RT ubiquitin. This can be explained with formation of four new bonds, which decrease the solvent accessible area. Also in the case of cubic crystals of ubiquitin, this can be linked to the resolution (3 Å), because the information about solvent content and interface area is becoming less defined at lower resolution. Low resolution also can explain why the DDM graphs for the cubic crystal form look different from MPD ubiquitin and rod shaped ubiquitin.

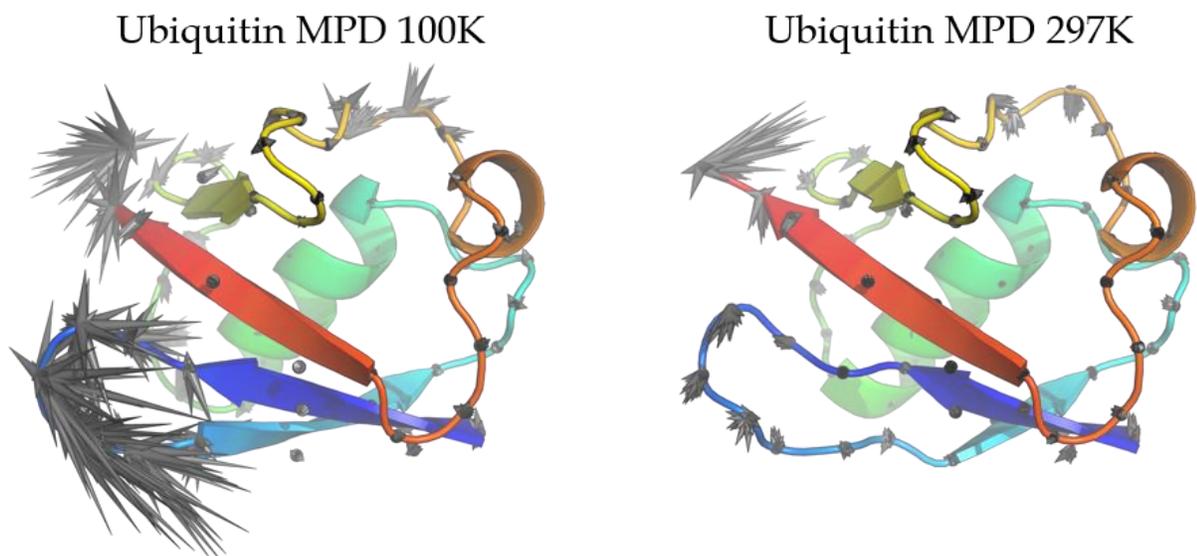
### **3.5.2 Indication of increased disorder of flash cooled crystals using ensemble refinement.**

Ensemble refinement is a useful tool to estimate atomic fluctuations in the proteins. In this work, we performed ensemble refinement using PHENIX software for all 6 pdb files: Ubiquitin-MPD, Ubiquitin-PEG-cubic, Ubiquitin-PEG-rod, 3ONS, 4XOK, 4XOL. For optimization of the parameters we used different ptls, tx and bath values which were described in the Materials & Methods chapter. We set up an array of simultaneous simulations for tx values in the range from 0.1 to 1.2 depending on the resolution and R-factors. We tested bath values at 2.5, 5, 10 K. PTLS values were in the range from 0.6 to 1.0 suggested by Phenix<sup>61</sup>, depending on resolution. An example of the obtained heat map with different tx, bath and ptls values is presented in **Fig.57**.



**Figure 57.** Rfree heat map obtained after ensemble refinement for MPD ubiquitin RT. The colour bar shows the values for Rfree with different pTLS, bath and tx parameters. Blue is the lowest value, red is the highest value.

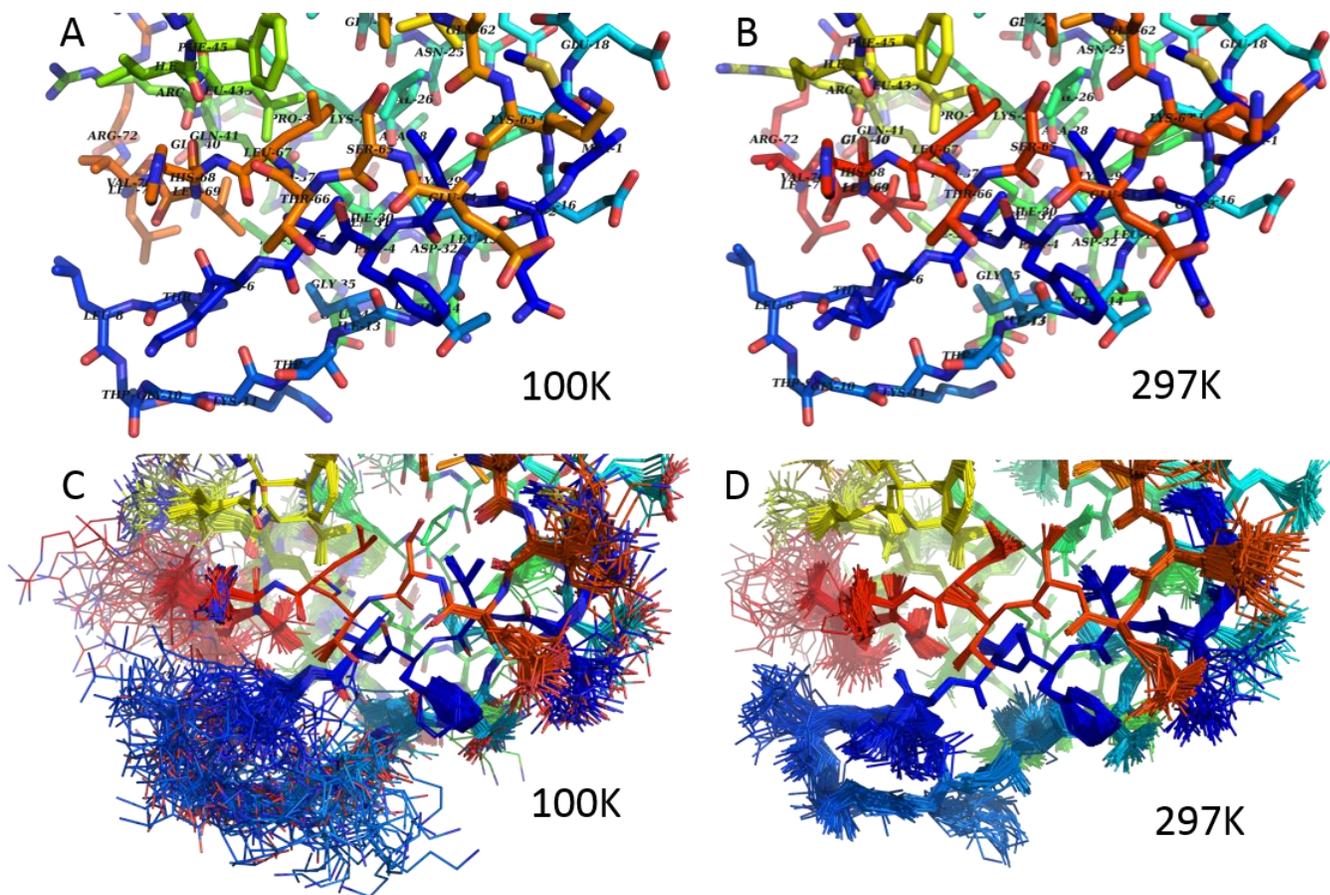
Using a Python/Pymol script written by *N.Coquelle* we generated porcupine plots for these ensembles which represent a graphical summary of the motions along the trajectory. In a porcupine plot, each C-alpha atom has a cone pointing in the direction of its motion. The length of the cone reflects the amplitude of the motion and the width of the cone indicates the number of such C-alpha atoms (**Fig.60**)<sup>110</sup>.



**Figure 58. Porcupine plots after ensemble refinement of MPD crystals of ubiquitin. Arrows are  $\beta$ -strands, where the dark blue arrow is  $\beta$ 1, light blue-  $\beta$ 2, light green-  $\beta$ 3, yellow-  $\beta$ 4, red-  $\beta$ 5. The  $\alpha$  helices are green, and the turns are orange.**

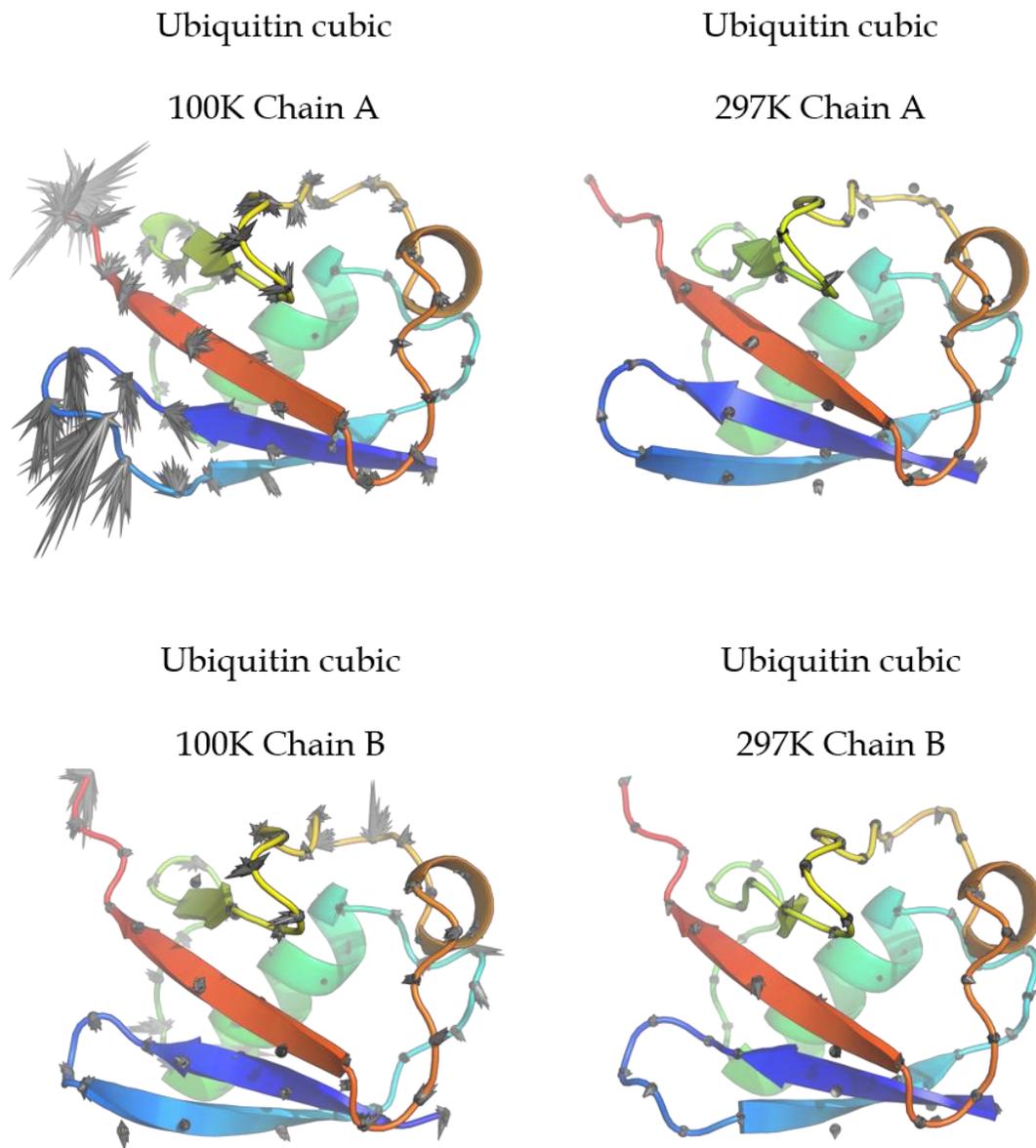
From porcupine plots for MPD crystals we observed the largest motion between  $\beta$ 1 and  $\beta$ 2 strands for MPD crystals of ubiquitin at CT structure and for both structures at the end of the chain. Probably the large motion that we observed for the structure at CT for some residues may cause a following disorder of the main chain. Evidence for an overall rocking motion in protein crystals of ubiquitin collected at 100 K was shown to exist by NMR, MD and XRD data<sup>45</sup> which can be the case here for the MPD crystals.

After comparison of the porcupine plots, we examined how the regions with the largest motion and the main chains look after ensemble refinement (**Fig.59**).



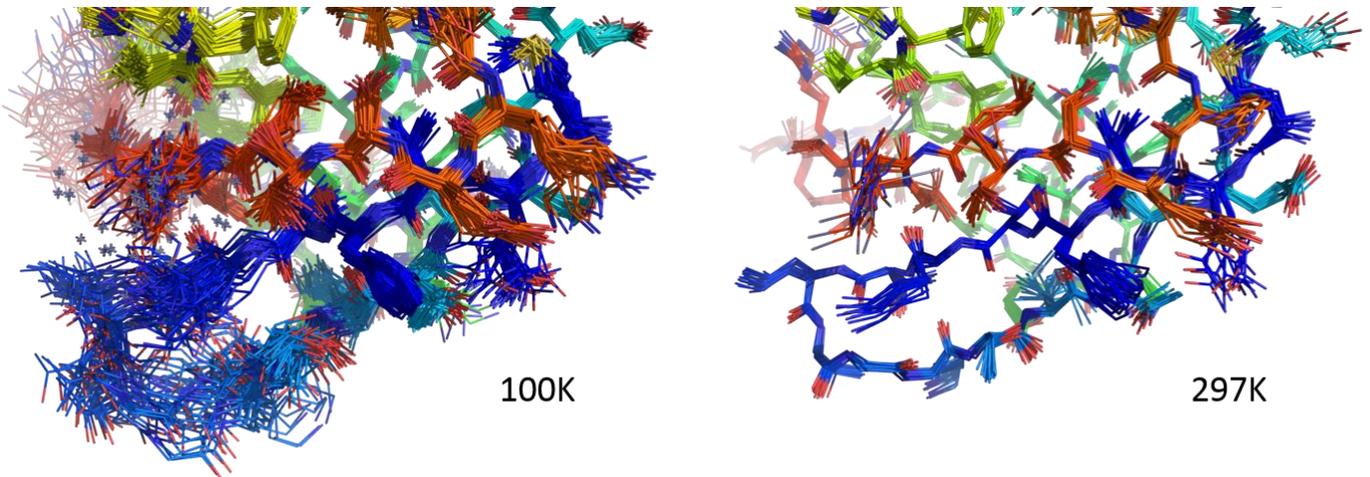
**Figure 59.** Comparison of CT (100K) and RT (297K) structures of MPD ubiquitin crystals in the area between  $\beta 1$  (dark blue) and  $\beta 2$  (light blue) strands. A-original pdb for CT structure; B-original pdb for RT structure; C- ensemble refinement for CT structure; D- ensemble refinement for RT structure.

A comparison of the ensemble refinement data with the original PDB files confirmed our hypothesis about degree of the disorder. An increase of the disorder of the residues leads to the increase of the disorder of the main chain.



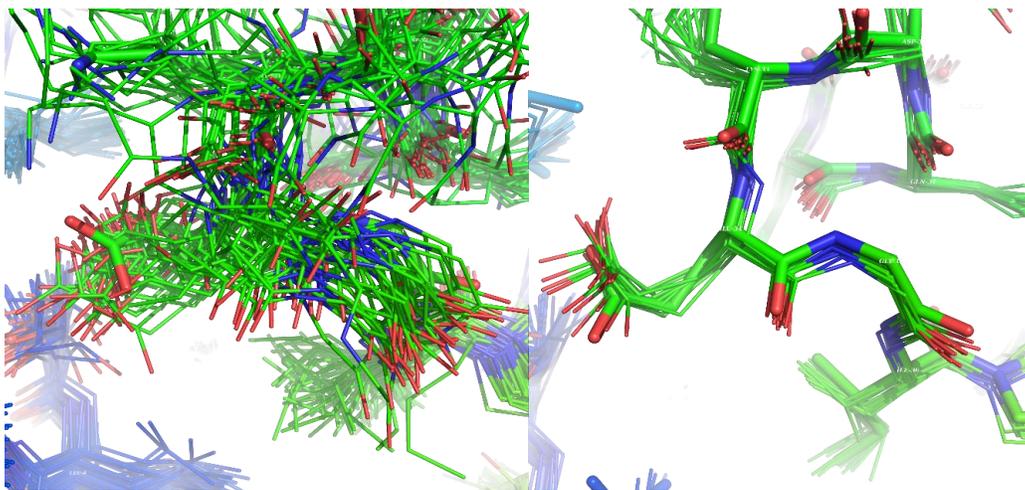
**Figure 60. Porcupine plots obtained after ensemble refinement of cubic crystals.**

From the porcupine plots for cubic crystals of ubiquitin (**Fig.60**) we can see that at CT chain A is more disordered than chain B. The highest motion for CT cubic shaped crystals of ubiquitin is observed between  $\beta 1$  and  $\beta 2$  strands and at the end of the chain. As for the MPD ubiquitin dataset, we compared the most interesting region and residues for cubic ubiquitin in PyMOL<sup>66</sup>.



**Figure 61.** Comparison of CT (100K) and RT (297K) structures of cubic ubiquitin crystals (chain A) in the area between  $\beta 1$  (dark blue) and  $\beta 2$  (light blue) strands. Left- ensemble refinement for CT structure; Right- ensemble refinement for RT structure.

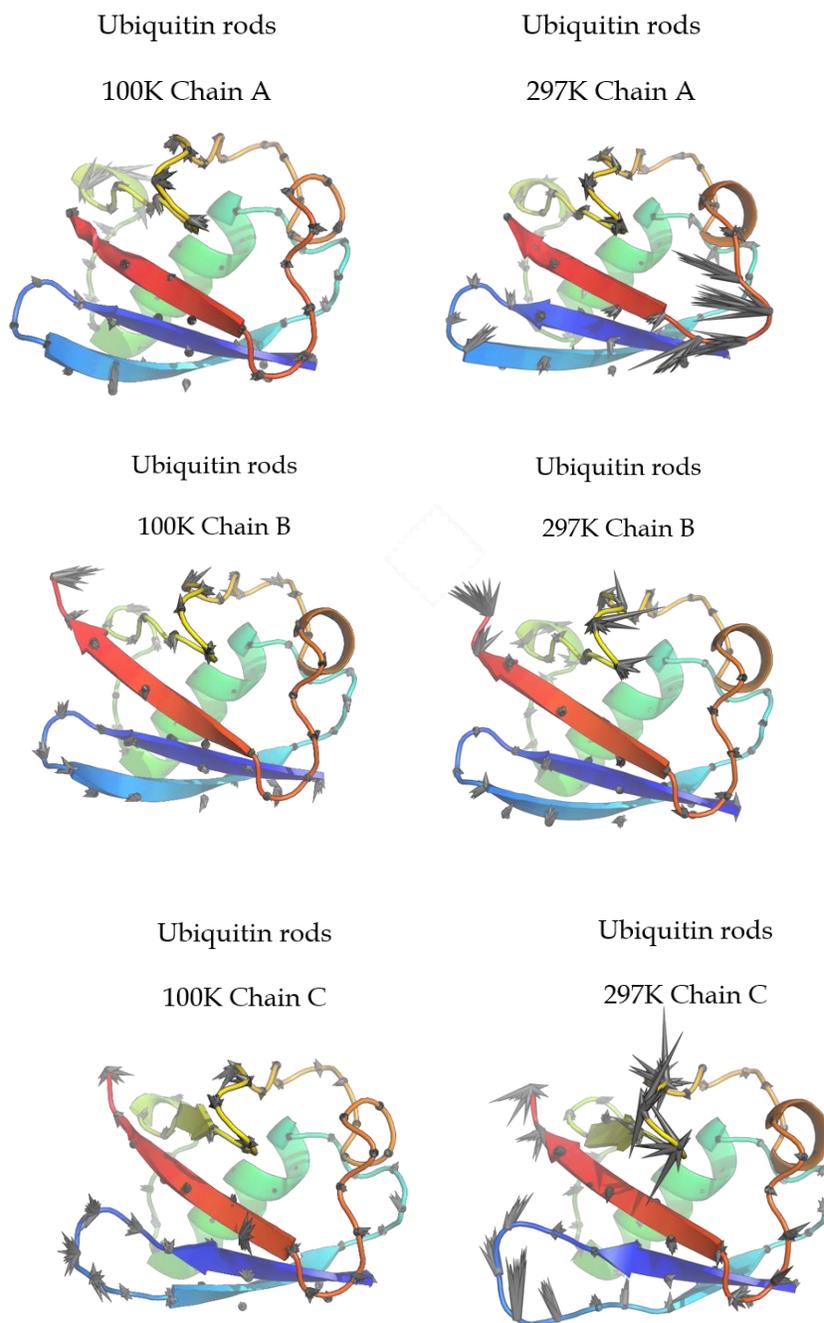
The main chain of cubic CT structure of ubiquitin looks more disordered (**Fig.61**) than the room temperature structure. The residues of chain A in the area between  $\beta 1$  and  $\beta 2$  strands cause the disorder in the main chain, the same as for MPD ubiquitin, described above. Also, the comparison of some residues (**Fig.62**) presents higher disorder of the side chain in the CT structure of cubic ubiquitin.



**Figure 62.** Cubic crystals, area near residue 33, chain B; Left-ensemble refinement of CT structure; Right- ensemble refinement of RT structure.

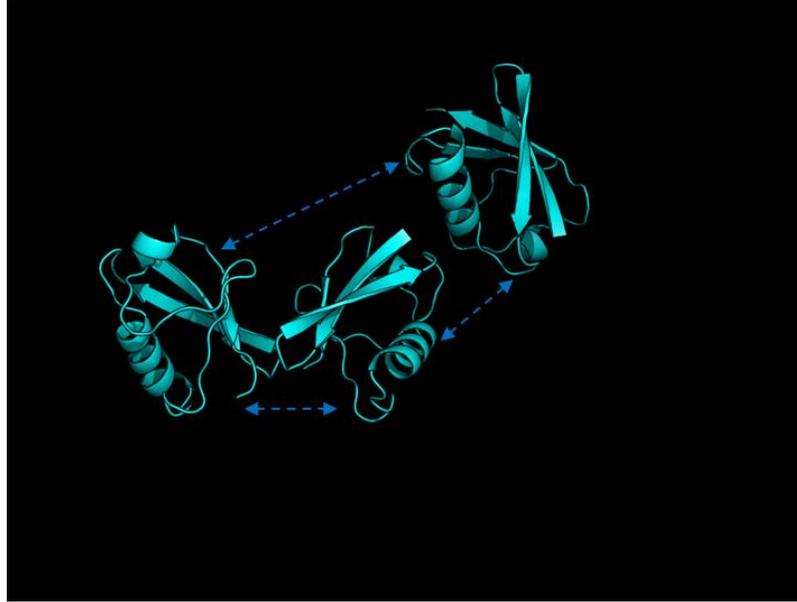
Results of the ensemble refinement performed with rod crystals (**Fig.63**) differ from MPD crystals and Cubic ubiquitin. Chain A for CT rods has the highest motion in the region near residues 37-40; for RT structure the highest motion observed near

residues 61-64 (the largest porcupines). Chain B looks similar in both cases. Chain C is more disordered than A and B. This can be explained due to their crystal packing because crystal packing was shown to have an influence on the mobility of the protein<sup>111</sup>.



**Figure 63. Porcupine plots obtained after the ensemble refinement of rod-shaped crystals.**





**Figure 65. Domain motions inside rod-shaped crystals of ubiquitin at room-temperature.**

The angle of movement between Chain A and B is almost three times smaller than between A and C or B and C, which gives the Chain C more space for dynamic movements. This fact can explain why we observe more dynamics for chain C than for chain A or B.

## Discussion

We performed an analysis on 3 different crystal forms of ubiquitin protein with different space groups, number of molecules per asymmetric unit, and packing.

From the electron density maps of the data collected at room temperature we were able to observe and fit alternative conformations for some residues. We analyzed the PDB files from the database collected at cryo-temperatures and did not find any alternative conformations.

From the crystal content analysis we observed that ubiquitin protein collected at RT on average has 5% more solvent content than at CT. The increase in the protein volume at RT was observed for all crystal forms, up to 13 % for MPD crystals of ubiquitin. We observed a negative correlation between the number of chains per asymmetric unit and the protein volume. Also, a difference in precipitant MPD vs PEG may have an influence on the protein volume, due to the specific properties of MPD. We found that there are less internal interactions inside the molecule at RT than at CT (for MPD ubiquitin and rods ubiquitin), which can be linked to the fact that cryocooling reduces the solvent volume and creates a bigger interface area. However, the situation is reversed with cubic ubiquitin. New bonds are created and interface area decreases, which could be linked to the lower resolution of ubiquitin crystals (3 Å).

DDMs plots showed that protein contracts upon cryocooling, which leads to an increase of the disorder inside the crystal. From ensemble refinement simulations we observed that at CT disorder of the residues could cause a disorder of the main chain. This can lead to the conclusion that despite the apparently good data quality of CT datasets, their dynamics might be sometimes affected by intrinsic problems of the sample quality induced by the process of cryo-freezing. Where possible, room temperature serial crystallography might provide a more accurate description of the dynamics or at least help identifying such problems.

## SUMMARY AND FUTURE PERSPECTIVES

The development of ultra-bright X-ray sources such as XFELs has enabled room temperature protein crystallography data collection based on the delivery of a series of micron- or submicron-sized crystals into the beam. Serial crystallography has also become of interest for synchrotron radiation (SR) sources, in particular with the introduction of high frame rate, single photon sensitive X-ray pixel detectors.

The availability of advanced sample delivery systems minimizing sample consumption is currently a bottleneck in the development of synchrotron based serial crystallography. Systems that are currently becoming available to the user community can be divided into 2 groups: liquid injectors and solid state supports. All systems have specific advantages and disadvantages. The ideal sample-delivery system would be a pipe-line combining crystal growth and delivery functions. Background scattering should be kept as low as possible in view of weakly scattering protein crystals.

In this thesis work three different methods of performing serial crystallography at the microfocus beamline (ID13) of the European Synchrotron (ESRF) were experimentally installed and evaluated:

1. LCP-injector experiments were performed, demonstrating that the setup developed for XFEL sources can be implemented and used at the more widely available synchrotron microfocus beamlines.

2. Raster-scanning technique developed at ESRF-ID13 shows that high-quality data collection with very low background for an aqueous protein microcrystal slush on a silicon nitride membrane support is possible. Solid support systems on the base of silicon nitride membranes in combination with raster-scanning were used in this work to study in-situ structural dynamics.

3. The CrystalDirect plates approach, originally developed at EMBL (HTX-lab) allows automated crystal growth and subsequent data collection by default without any crystal manipulation apart from the optional liquid removal, which has been successfully applied. Room temperature scanning diffraction

using this approach enables rapid screening as well as full data collection. It has been demonstrated that it could be used in the future for investigating the conformational diversity of protein molecules.

The results obtained from these three approaches might help current and future users of synchrotron microfocus beamlines to perform fast data collection of microcrystals at room temperature using routinely injector-based systems and solid supports.

However, every year new systems for sample delivery are developed in respect of the needs of users and with exceptional ideas of the beamline scientists and realization of engineers. Recently new methods, like ultrasonic acoustic levitation<sup>113</sup> of crystals and systems using movable conveyor belts<sup>114</sup> with crystals were shown to have promising results.

This work might be also interesting in the context of planned upgrades of 3<sup>rd</sup> generation synchrotron sources such as the ESRF. In this case an increase of brilliance of a factor of up to 50 can be expected after the ongoing upgrade phase: “ESRF Extremely Brilliant Source (ESRF-EBS)”. Looking at the example of ID13 additional gains in flux for micro-/nano-beams of the order of a factor 10000 are in reach for a typical beamsize of one micron or 100 nm respectively. This estimate is based on opportunities for accompanying improvements in undulator developments (~factor 10 for higher energies) and possibly a pink beam source based on directly using a single undulator harmonic (up to a factor of 100). When compromising on the gains in flux the increased brilliance can also be used to obtain smaller nano-beams. A reduction in beam size and exposure times will possibly allow extending serial protein crystallography to the scale of 100 nm crystals or less. With the EBS upgrade and a further development of the plate scanning approaches, a specialized highly automated dedicated beamline conceived for serial crystallography similar to the ESRF MASSIF beamlines could be implemented. Such an end station would allow to perform very fast in-situ screening where the screening procedure would already turn into the final data collection if enough crystals of sufficient quality were in the plate.

With the future upgrade of synchrotrons the development of detectors also has to improve further. For example, at ID13 the performance of the current EIGER 4M detector at an acquisition rate of up to 750 Hz is well matched to the current beam parameters. However, adequate detector systems could easily be a limiting factor. They have to be able to handle the severely increased count rates and match the potential of the source in terms of speed and dynamic range while maintaining the single photon counting sensitivity of today's detector models. Current ID13 data collections can be performed with exposures down to about 1.4 ms. In the future exposures of a few bunches or even a single bunch ( $< 1$  ns, 3 ns time interval) per frame in serial mode could be required.

Such an enhanced acquisition rate could lead to a further reduction of radiation damage effects on the quality of the data e.g. by better outrunning the damage induced by the diffusion of radicals generated by X-ray absorption. Already in this work we did not observe any specific radiation damage in the electron-density maps for all structures within the limits of the respective resolution obtained. However, the radiation damage issue should be always taken into account before starting an X-ray experiment to optimize the exposure time depending on the parameters of the beam. Also fast efficient data collection could be very promising for a phasing approach based on diffuse scattering<sup>115</sup>.

In the future, it should be also possible to perform serial in-situ structural dynamics experiments and time-resolved studies on the sub-millisecond to even sub microsecond time scale at synchrotrons, still much slower but more closely resembling XFEL experiments than today. It can be expected that there will be an advantage over XFELs in terms of sample consumption. However, the radiation damage effects to protein and sample delivery environment (e.g. radius of damage per shot) have to be further investigated at these improved beam conditions. With the increased time-resolution room temperature studies at the synchrotron sources can help to understand better the dynamics of the proteins. Regarding cryocrystallography, an accompanying room temperature serial plate-scanning based data collection could have a high potential for a better understanding the

scientific problem in questions and for reducing the risk of misinterpretation of data.

Evidently synchrotron sources will never reach the brightness of X-ray free electron lasers. Nevertheless, the more widespread availability of synchrotron radiation sources and development of advanced sample delivery systems make application of serial crystallography at synchrotron microfocus beamlines highly interesting.

## BIBLIOGRAPHY

1. Henderson, R. Cryo-Protection of Protein Crystals against Radiation Damage in Electron and X-Ray Diffraction. *Proc. R. Soc. London B Biol. Sci.* **241**, (1990).
2. Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 339–51 (2010).
3. Chapman, H. N., Caleman, C. & Timneanu, N. Diffraction before destruction. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **369**, 20130313 (2014).
4. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
5. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
6. Lomb, L. *et al.* Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys. Rev. B* **84**, 214111 (2011).
7. Barty, A. *et al.* Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photonics* **6**, 35–40 (2011).
8. Stellato, F. *et al.* Room-temperature macromolecular serial crystallography using synchrotron radiation. *IUCrJ* **1**, 204–212 (2014).
9. Nogly, P. *et al.* Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ* **2**, 168–176 (2015).
10. Veksler V.I. *A New Method of the Acceleration of Relativistic Particles.* (1944).
11. McMillan, E. M. The Synchrotron – A Proposed High Energy Particle Accelerator. *Phys. Rev.* **68**, 143–144 (1945).
12. Elder, F. R., Langmuir, R. V. & Pollock, H. C. Radiation from electrons accelerated in a synchrotron. *Phys. Rev.* **74**, 52–56 (1948).
13. Brönnimann, C. & Trüb, P. in *Synchrotron Light Sources and Free-Electron*

*Lasers* 1–29 (Springer International Publishing, 2015). doi:10.1007/978-3-319-04507-8\_36-1

14. Castelli, C. M., Allinson, N. M., Moon, K. J. & Watson, D. L. High spatial resolution scintillator screens coupled to CCD detectors for X-ray imaging applications. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **348**, 649–653 (1994).
15. Tate, M. W. *et al.* A Large-Format High-Resolution Area X-ray Detector Based on a Fiber-Optically Bonded Charge-Coupled Device (CCD). *J. Appl. Crystallogr.* **28**, 196–205 (1995).
16. Walter, R. L. *et al.* High-resolution macromolecular structure determination using CCD detectors and synchrotron radiation. *Structure* **3**, 835–844 (1995).
17. Madey, J. M. J. Stimulated Emission of Bremsstrahlung in a Periodic Magnetic Field. *J. Appl. Phys.* **42**, 1906 (1971).
18. Elias, L. R., Fairbank, W. M., Madey, J. M. J., Schwettman, H. A. & Smith, T. I. Observation of Stimulated Emission of Radiation by Relativistic Electrons in a Spatially Periodic Transverse Magnetic Field. *Phys. Rev. Lett.* **36**, 717–720 (1976).
19. Seeck, O. H. & Murphy, B. M. *X-ray diffraction: modern experimental techniques.*
20. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).
21. Petsko, G. a & Ringe, D. From Sequence to Structure. *Protein Struct. Funct.* 2–48 (2004). doi:10.1088/1674-1056/23/7/078705
22. Tan, S., Tan, H. T. & Chung, M. C. M. Membrane proteins and membrane proteomics. *Proteomics* **8**, 3924–32 (2008).
23. John R.Helliwell. *Macromolecular Crystallography with Synchrotron Radiation.* (1992).
24. Liljas, A. *et al.* *Textbook of Structural Biology.* (WORLD SCIENTIFIC, 2009). doi:10.1142/6620
25. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of

- the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3Å resolution. *Nature* **318**, 618–624 (1985).
26. White, S. H. Membrane Proteins of Known 3D Structure. [Http://Blanco.Biomol.Uci.Edu/Mpstruc/](http://Blanco.Biomol.Uci.Edu/Mpstruc/) **1085**, 3–5 (2016).
  27. Otwinowski, Z. & Minor, W. *International Tables for Crystallography. International Tables for Crystallography Volume F: Crystallography of biological macromolecules F*, (2006).
  28. Drenth, J. & Mesters, J. *Principles of protein X-ray crystallography: Third edition. Principles of Protein X-Ray Crystallography: Third Edition* (2007). doi:10.1007/0-387-33746-6
  29. Sliz, P., Harrison, S. C. & Rosenbaum, G. How does Radiation Damage in Protein Crystals Depend on X-Ray Dose? *Structure* **11**, 13–19 (2003).
  30. Blake, C. C. F. & Phillips, D. C. In Biological Effects of Ionizing Radiation at the Molecular Level. in 183–191 (1962).
  31. Burmeister, W. P. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 328–341 (2000).
  32. Weik, M. *et al.* Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci.* **97**, 623–628 (2000).
  33. Weik, M. *et al.* Solvent behaviour in flash-cooled protein crystals at cryogenic temperatures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 566–573 (2001).
  34. Nave, C. & Garman, E. F. Towards an understanding of radiation damage in cryocooled macromolecular crystals. *J. Synchrotron Radiat.* **12**, 257–260 (2005).
  35. Carugo, O. & Carugo, K. D. When X-rays modify the protein structure: Radiation damage at work. *Trends in Biochemical Sciences* **30**, 213–219 (2005).
  36. Garman, E. F. *et al.* Macromolecular crystallography radiation damage research: what's new? *J. Synchrotron Radiat.* **18**, 313–317 (2011).

37. Warkentin, M., Badeau, R., Hopkins, J. B. & Thorne, R. E. Spatial distribution of radiation damage to crystalline proteins at 25-300 K. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 1108–17 (2012).
38. Roedig, P. *et al.* Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *J. Appl. Crystallogr.* **49**, 968–975 (2016).
39. Coquelle, N. *et al.* Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 1184–1196 (2015).
40. Usón, I. & Sheldrick, G. M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648 (1999).
41. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 325–38 (2010).
42. Rossmann, M. G. *et al.* Molecular replacement – historical background. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 1360–1366 (2001).
43. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **275**, 1–21 (2008).
44. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
45. Ma, P. *et al.* Observing the overall rocking motion of a protein in a crystal. *Nat Commun* **6**, 1–24 (2015).
46. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002).
47. Bocquet, N. *et al.* X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* **457**, 111–114 (2009).
48. Weierstall, U. Liquid sample delivery techniques for serial femtosecond crystallography. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **369**, 20130337 (2014).

49. Silson. Available at: <http://www.silson.com/>.
50. White, T. A. *et al.* CrystFEL: A software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012).
51. Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 1690–1695 (1999).
52. Duisenberg, A. J. M. Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Crystallogr.* **25**, 92–96 (1992).
53. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
54. METROPOLIS, N. & ULAM, S. The Monte Carlo method. *J. Am. Stat. Assoc.* **44**, 335–341 (1949).
55. Kirian, R. A. *et al.* Femtosecond protein nanocrystallography – data analysis methods. *Opt. Express* **18**, 5713 (2010).
56. White, T. A. *et al.* Recent developments in CrystFEL. *J. Appl. Crystallogr.* **49**, 680–689 (2016).
57. Diederichs, K. & Karplus, P. A. Better models by discarding data? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1215–22 (2013).
58. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 282–92 (2011).
59. Levin, E. J., Kondrashov, D. A., Wesenberg, G. E. & Phillips, G. N. Ensemble refinement of protein crystal structures: validation and application. *Structure* **15**, 1040–52 (2007).
60. Burnley, B. T., Afonine, P. V, Adams, P. D. & Gros, P. Modelling dynamics in protein crystal structures by ensemble refinement. *Elife* **1**, e00311 (2012).
61. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
62. Levin, E. J., Kondrashov, D. A., Wesenberg, G. E. & Phillips, G. N.

- Ensemble refinement of protein crystal structures: validation and application. *Structure* **15**, 1040–52 (2007).
63. Kirian, R. A. *et al.* Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta Crystallogr. A.* **67**, 131–40 (2011).
  64. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
  65. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
  66. DeLano, W. L. The PyMOL Molecular Graphics System, Version 1.1. Schrödinger LLC <http://www.pymol.org> (2002). doi:10.1038/hr.2014.17
  67. Gati, C. *et al.* Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* **1**, 87–94 (2014).
  68. Hunter, M. S. *et al.* Fixed-target protein serial microcrystallography with an x-ray free electron laser. *Sci. Rep.* **4**, 6026 (2014).
  69. Roedig, P. *et al.* A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering. *Sci. Rep.* **5**, 10451 (2015).
  70. Oghbaey, S. *et al.* Fixed target combined with spectral mapping: approaching 100% hit rates for serial crystallography. *Acta Crystallogr. Sect. D, Struct. Biol.* **72**, 944–55 (2016).
  71. Roessler, C. G. *et al.* Acoustic Injectors for Drop-On-Demand Serial Femtosecond Crystallography. *Structure* **24**, 631–40 (2016).
  72. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14532–14535 (1996).
  73. The Cherezov Lab. LCP. Available at: <http://cherezov.usc.edu/resources.htm>.
  74. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals

- grown in lipidic cubic phases. *Science* **277**, 1676–81 (1997).
75. Gordeliy, V. I. *et al.* Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature* **419**, 484–487 (2002).
  76. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–65 (2007).
  77. Weierstall, U. *et al.* Droplet streams for serial crystallography of proteins. *Exp. Fluids* **44**, 675–689 (2008).
  78. DePonte, D. P. *et al.* Gas Dynamic Virtual Nozzle for Generation of Microscopic Droplet Streams. *J. Phys. D. Appl. Phys.* **41**, 7 (2008).
  79. Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014).
  80. Liu, W. *et al.* Serial Femtosecond Crystallography of G Protein-Coupled Receptors. *Science* (80-. ). **342**, 175–180 (2013).
  81. Nogly, P. *et al.* Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ* **2**, 168–76 (2015).
  82. Botha, S. *et al.* Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 387–397 (2015).
  83. Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer* **4**, 253–265 (2004).
  84. Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014).
  85. Cipriani, F. *et al.* CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 1393–9 (2012).
  86. Zander, U. *et al.* Automated harvesting and processing of protein crystals through laser photoablation. *Acta Crystallogr. Sect. D Struct.*

- Biol.* **72**, 454–466 (2016).
87. Wel, H. & Loeve, K. Isolation and Characterization of Thaumatin I and II, the Sweet-Tasting Proteins from *Thaumatococcus daniellii* Benth. *Eur. J. Biochem.* **31**, 221–225 (1972).
  88. Hassaine, G. *et al.* X-ray structure of the mouse serotonin 5-HT<sub>3</sub> receptor. *Nature* **512**, 276–281 (2014).
  89. Bocquet, N. *et al.* A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. *Nature* **445**, 116–119 (2007).
  90. Imoto, K. *et al.* A ring of uncharged polar amino acids as a component of channel constriction in the nicotinic acetylcholine receptor. *FEBS Lett.* **289**, 193–200 (1991).
  91. Sauguet, L. *et al.* Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO J.* **32**, 728–41 (2013).
  92. Morozov, V. N. & Morozova, T. Y. Viscoelastic properties of protein crystals: Triclinic crystals of hen egg white lysozyme in different conditions. *Biopolymers* **20**, 451–467 (1981).
  93. Newman, J. *et al.* A review of techniques for maximizing diffraction from a protein crystal *in stilla*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 27–31 (2006).
  94. Russo Krauss, I., Sica, F., Mattia, C. A. & Merlino, A. Increasing the X-ray diffraction power of protein crystals by dehydration: the case of bovine serum albumin and a survey of literature data. *Int. J. Mol. Sci.* **13**, 3782–800 (2012).
  95. Heras, B. & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 1173–1180 (2005).
  96. Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducriux, a. High-resolution structure (1.33 Å) of a HEW lysozyme tetragonal crystal grown in the APCF apparatus. Data and structural comparison with a crystal grown under microgravity from SpaceHab-01 mission. *Acta Crystallogr. D. Biol. Crystallogr.* **52**, 505–17 (1996).
  97. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based

- approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814**, 942–68 (2011).
98. Lindorff-Larsen, K., Maragakis, P., Piana, S. & Shaw, D. E. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B* **120**, 8313–8320 (2016).
  99. Johnson, E. C., Lazar, G. A., Desjarlais, J. R. & Handel, T. M. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure* **7**, 967–976 (1999).
  100. Ermolenko, D. N., Dangi, B., Gvritishvili, A., Gronenborn, A. M. & Makhatadze, G. I. Elimination of the C-cap in ubiquitin—structure, dynamics and thermodynamic consequences. *Biophys. Chem.* **126**, 25–35 (2007).
  101. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
  102. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16247–52 (2011).
  103. Huang, K. Y., Amodeo, G. A., Tong, L. & McDermott, A. The structure of human ubiquitin in 2-methyl-2,4-pentanediol: a new conformational switch. *Protein Sci.* **20**, 630–9 (2011).
  104. Cowtan, K., Emsley, P. & Wilson, K. S. From crystal to structure with CCP4. *Acta Crystallographica Section D: Biological Crystallography* **67**, 233–234 (2011).
  105. Frauenfelder, H. *et al.* Thermal expansion of a protein. *Biochemistry* **26**, 254–261 (1987).
  106. Keedy, D. A. *et al.* Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure* **22**, 899–910 (2014).
  107. Berejnov, V. *et al.* Effects of cryoprotectant concentration and cooling rate on vitrification of aqueous solutions. *J. Appl. Crystallogr.* **39**, 244–251 (2006).
  108. Dumetz, A. C., Chockla, A. M., Kaler, E. W. & Lenhoff, A. M.

- Comparative Effects of Salt, Organic, and Polymer Precipitants on Protein Phase Behavior and Implications for Vapor Diffusion. *Cryst. Growth Des.* **9**, 682–691 (2009).
109. Krissinel, E. & Henrick, K. in 163–174 (Springer Berlin Heidelberg, 2005). doi:10.1007/11560500\_15
  110. Wang, S. *et al.* Molecular Dynamics Analysis Reveals Structural Insights into Mechanism of Nicotine N-Demethylation Catalyzed by Tobacco Cytochrome P450 Mono-Oxygenase. *PLoS One* **6**, e23342 (2011).
  111. Phillips, G. N. *Comparison of the dynamics of myoglobin in different crystal forms.* *Biophysical Journal* **57**, (Cell Press, 1990).
  112. Hingefind. Available at: <http://biomachina.org/disseminate/hingefind/hingefind.html>.
  113. Tsujino, S. *et al.* Ultrasonic acoustic levitation for fast frame rate X-ray protein crystallography at room temperature. *Sci. Rep.* **6**, 25558 (2016).
  114. Soares, A. S. *et al.* Solvent minimization induces preferential orientation and crystal clustering in serial micro-crystallography on micro-meshes, in situ plates and on a movable crystal conveyor belt. *J. Synchrotron Radiat.* **21**, 1231–9 (2014).
  115. Ayyer, K. *et al.* Macromolecular diffractive imaging using imperfect crystals. *Nature* **530**, 202–206 (2016).