

# Data processing and analysis in serial crystallography at advanced X-ray sources

Dissertation with the aim of achieving a doctoral degree  
at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Chemistry  
of Universität Hamburg

submitted by  
**Cornelius Gati**

2015 in Hamburg

Day of oral defense:

05. November 2015

The following evaluators recommend the admission of  
the dissertation:

Prof. Dr. Christian Betzel

Prof. Dr. Henry N. Chapman

Author email: [cornelius.gati@gmx.de](mailto:cornelius.gati@gmx.de)

*Apunak*

# Abstract

In the recent past a tremendous effort has been devoted to designing X-ray sources producing radiation pulses with unique properties. With the advent of X-ray Free-Electron Lasers (XFELs), inherent limits of bioimaging can be overcome by delivering X-ray pulses short enough to outrun conventional radiation damage processes and to allow probing of timescales which were previously unthinkable to investigate. An important milestone on the way to the visionary goal of single-particle imaging using XFEL pulses was the development of serial femtosecond crystallography (SFX). Under the conditions provided at an XFEL, a protein crystal is only exposed to the X-ray beam for tens of femtoseconds before it is completely destroyed. In the novel technique of SFX, data is collected from randomly oriented crystals that are exposed to the X-ray beam one at a time. In order to assemble a complete data set, it was essential to develop data analysis techniques that are capable of dealing with large quantities of snapshot diffraction data, which significantly differ from data obtained by conventional rotation series.

The primary focus of this dissertation is the investigation of SFX datasets, with a strong emphasis on aspects of data collection and analysis, as well as an outlook on possible improvements to the currently most broadly applied SFX data analysis pipeline. General differences between conventional macromolecular crystallography and SFX with respect to data processing and resulting biomolecular structures are discussed. The case studies used to illustrate the SFX technique in this dissertation are manifold, ranging from the demonstration of a high resolution structure from X-ray diffraction on the smallest crystals (*in vivo* polyhedrin nanocrystals) known so far, to the first human membrane protein (5-Hydroxytryptamine-2B receptor) structure at ambient temperatures, and the first completely unknown structure solved by SFX (angiotensin II type I receptor).

Inspired by the success of serial femtosecond crystallography, the first proof-of-concept demonstration of serial crystallography data collection schemes at modern microfocus synchrotron beamlines, both at cryogenic and room temperature, is presented. Since synchrotron light sources are readily available, these synchrotron based serial crystallography approaches will benefit a broad spectrum of the structural biology community and will be applicable to a broad range of samples.



# Zusammenfassung

In den letzten Jahren war ein enormer Aufwand für die Konzeption von Röntgenquellen mit einzigartigen Eigenschaften, wie zum Beispiel außerordentlich hoher Strahlintensität und räumlicher Kohärenz, zu verzeichnen. Mit dem Aufkommen von Freie-Elektronen-Röntgenlasern (XFELs) wurden die Limitierungen von bildgebenden Verfahren für die Strukturanalyse von Biomolekülen überwunden, bei denen Röntgenpulse generiert werden die kurz genug sind um konventionelle strukturelle Strahlenschäden zu umgehen. Gleichzeitig wurden Experimente mit zeitlicher Auflösung ermöglicht, die bisher undenkbar waren. Ein Meilenstein auf dem Weg zum visionären Ziel der Strukturanalyse von einzelnen Biomolekülen durch XFEL Pulse war die Entwicklung der Seriellen Femtosekunden Kristallographie (SFX). Wegen der einzigartigen Bedingungen, die an einem XFEL gegeben sind, übersteht ein Proteinkristall den Röntgenstrahl nur für Bruchteile einer Sekunde, bevor der Kristall komplett verdampft ist. Um das volle Potential dieser bahnbrechenden Errungenschaften auf dem Gebiet der Beschleunigerphysik optimal auszunutzen, war es notwendig diese neuartige Methode der Datenerfassung auszubauen, indem der Kristall für jeden Röntgenstrahlpuls ausgewechselt wird und jeweils nur ein Diffraktionsmuster pro Kristall aufgenommen werden kann. Zusätzlich war es essenziell Datenanalysemethoden zu entwickeln, die in der Lage sind mit einer großen Menge an Diffraktions Schnapschüssen von willkürlich orientierten Kristallen umzugehen, welche sich signifikant von Daten unterscheiden die mit konventionellen Rotationsserien generiert wurden. Der Hauptaugenmerk dieser Dissertation liegt auf der Untersuchung von mehreren solcher Datensätze, die mit der SFX Methode generiert wurden, wobei ein starker Fokus auf der Datenerfassung und -analyse, sowie auf Verbesserungsmöglichkeiten, liegt. Generall werden Unterschiede der Datenprozessierung, sowie der resultierenden Molekülstrukturen mit konventionellen Methoden verglichen.

Die Fallbeispiele, welche in dieser Dissertation beschrieben werden, sind vielfältig. Beginnend mit den momentan kleinsten Proteinkristallen, die jemals einen Datensatz zu atomarer Auflösung ermöglicht haben, wobei ein Kristall aus lediglich 10000 Einheitszellen besteht, über die erste Struktur eines humanen Membranproteins bei Raumtemperatur, bis hin zur ersten komplett unbekanntem Proteinstruktur, welche mit SFX gelöst wurde, wobei wichtige Einsichten über einen pharmakologisch äußerst wichtigen Rezeptor erlangt wurden, der eine große Rolle bei der Behandlung von Bluthochdruck spielt, werden im Detail analysiert und beschrieben.

Inspiriert durch den Erfolg der Seriellen Femtosekunden Kristallographie wird als letzter Teil dieser Dissertation die Adaptation dieser Methodik an eine moderne Mikrofokus Synchrotronmessstation, jeweils bei kryogenen, als auch bei Raumtemperatur, präsentiert und im Detail beschrieben. Der Hauptvorteil dieser synchrotronbasierten Methode ist die Option, diese Ansätze auf andere

Synchrotronquellen zu übertragen, was diese für ein breites Spektrum der Strukturbiologengemeinde zugänglich macht, bei dem ermöglicht wird, kleine Proteinkristalle, in der Größenordnung von Mikrometern, sowohl bei kryogenen Temperaturen, als auch bei Raumtemperatur zu untersuchen.

# Declaration on oath

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

I hereby declare that the dissertation has neither been accepted nor graded 'failed' in a previous doctoral procedure.

Hamburg, 07. November 2015 \_\_\_\_\_

(Cornelius Gati)

## Acknowledgements

Given all the possible paths I could have chosen for the first stages of my scientific career, I would like to begin by thanking Prof. Gebhard F.X. Schertler for directing me towards this exciting research field. Without his exploratory nature I wouldn't be where I am today.

I would like to specially thank my supervisor Prof. Henry N. Chapman, not only for being adventurous enough to accept an unexperienced biology student as a PhD candidate, but also for his scientific enthusiasm and optimism, out-of-the-box thinking and for generally being one of the most pleasant persons I ever had the pleasure to work with. I am very grateful to him for giving me the freedom and support I needed to finish this work - I could not have made a better choice for my graduate studies.

As head of the supervising committee of my PhD thesis, I am very grateful for the support and advice of Prof. Christian Betzel. His organizational skills and persistence were absolutely essential for the final stages of this thesis.

It has been a true pleasure to work with all the past and current members in the group at CFEL and the large SFX collaboration. Regardless of who I approached, my coworkers would drop everything and support me, for which I would like to thank everyone individually. In addition, I would like to specially thank two people from the group. Dr. Thomas A. White has been a great mentor during my work - with his invaluable knowledge, in particular in crystallography and academic ethics, as well as his patience of a saint - working with him has been a true inspiration and clearly has influenced me a lot for my future career. I also thank Dr. Anton Barty for always having an open door and giving me important feedback on all levels of my work: scientifically, personally and politically.

An important part of my work was carried out in close collaboration with Prof. Vadim Cherezov, with whom I spent a lot of time both collecting and evaluating data for very exciting projects. I am deeply thankful for his trust in my work and judgement. His incredibly strong work ethic and critical thinking made the work with him one of the best experiences I had during my scientific career. I certainly hope our fruitful collaboration will continue one way or another.

I would like to thank Dr. Gleb Bourenkov for his support and invaluable knowledge from the greatest details of crystallography to the very last screw in his beamline. There was not a single question he wasn't able to answer, while being one of the most humble scientists I have ever met.

A very special thanks to Irmtraud Kleine. Without her organizational skills, this group wouldn't be nearly as successful. On the side a very special thanks for helping me organizing the summer schools over the last two years.

The person I have to thank the most is my father Dr. Georg Gáti, to whom this thesis is dedicated. I hope I inherited at least some of his unbreakable optimism, patience, sense of justice and devotion to people.

## List of abbreviations

5-HT <sub>2B</sub>	5-hydroxytryptamine receptor 2B
AA	Amino acid
ADU	Arbitrary detector unit
AGIPD	Adaptive gain integrating pixel detector
APS	Advanced Photon Source
ARB	Angiotensin II receptor 1 blockers
ASIC	Application specific integrated circuits
AT <sub>1</sub> R	Angiotensin II receptor 1
AU	Asymmetric unit
CBF	Crystallographic Binary File
CC	Correlation coefficient
CCD	Charged-coupled device
CFEL	Center for Free-Electron Laser Science
CpGV	<i>Cydia pomonella</i> granulovirus
cryoEM	cryo electron microscopy
CSPAD	Cornell-SLAC pixel array detector
CXI	Coherent X-ray Imaging
DESY	Deutsches Elektronensynchrotron
DNA	Deoxyribonucleic acid
DM	Direct methods
EMBL	European Molecular Biology Laboratory
ESRF	European Synchrotron Radiation Facility
FEH	Far experimental hall
FLASH	Free-electron laser in Hamburg
FWHM	Full width half maximum
GDP	Guanosine diphosphate
GDVN	Gas dynamic virtual nozzle
GPCR	G-protein coupled receptor
GTP	Guanosine triphosphate
HDF5	Hierarchical Data Format, version 5
HEWL	Hen egg-white lysozyme
HPLC	High performance liquid chromatography
ID	Inner diameter
KB	Kirkpatrick-Baez mirrors
kDa	Kilo Dalton (1 kDa = 1000 g/mol)
LASER	Light amplification by stimulated emission of radiation
LINAC	Linear accelerator
LCLS	Linac coherent light source
LCP	Lipidic cubic phase
MAG	Monoacylglycerol
MR	Molecular replacement
MW	Molecular weight (1 g/mol or 1 Da)
MX	Macromolecular crystallography
OB	Occlusion bodies

PEG	Polyethylene glycol
PETRAIII	Positron-Electron Tandem Ring Accelerator
PM	Patterson map
PDB	Protein data bank
RF	Radio frequency
SASE	Self-amplified spontaneous emission
SEM	Scanning electron microscopy
SFX	Serial femtosecond crystallography
SLAC	Stanford Linear Accelerator Center
SLS	Swiss Light Source
SNR	Signal-to-noise ratio
SONICC	Second-order nonlinear optical imaging of chiral crystals
SSX	Serial synchrotron crystallography
SYN	Synchrotron
TbCatB	<i>Trypanosoma brucei</i> proCathepsinB
TEM	Transmission electron microscopy
UC	Unit cell
UV	Ultra violet
XFEL	X-ray Free-Electron Laser

## List of publications

**I** - Zhang H, Unal H, **Gati C**, Han GW, Liu W, Zatsepin NA, James D, Wang D, Nelson G, Weierstall U, Sawaya MR, Xu Q, Messerschmidt M, Williams GJ, Boutet S, Yefanov OM, White TA, Wang C, Ishchenko A, Tirupula KC, Desnoyer R, Coe J, Conrad CE, Fromme P, Stevens RC, Katritch V, Karnik SS, Cherezov V. (2015) **Structure of the Angiotensin Receptor Revealed by Serial Femtosecond Crystallography**. *Cell*. 2015 Apr 21. pii: S0092-8674(15)00428-6

**II** - Stellato F, Oberthuer D, Liang M, Bean R, **Gati C**, Yefanov O, Barty A, Burkhardt A, Fisher P, Galli L, Kirian L, Meyer J, Panneerselvam S, Yoon CH, Chervinskii F, Speller E, White TA, Betzel C, Meents A, Chapman HN (2014) **Room temperature macromolecular serial crystallography using synchrotron radiation**. *IUCrJ*. 2014 1, 204-212

**III** - Weierstall U, James D, Wang C, White TA, Wang D, Liu W, Spence JCH, Doak RB, Nelson G, Fromme P, Fromme R, Grotjohann I, Kupitz C, Zatsepin N, Liu H, Basu S, Wacker D, Han GW, Katritch V, Boutet S, Messerschmidt M, Williams GJ, Koglin JE, Seibert MM, Klinker M, **Gati C**, Shoeman RL, Barty A, Chapman HN, Kirian RA, Beyerlein KR, Stevens RC, Li D, Shah STA, Howe N, Caffrey M, Cherezov V (2014) **Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography**. *Nature Communications*. 2014, 5:3309

**IV** - **Gati C**, Bourenkov G, Klinge M, Rehders D, Stellato F, Oberthuer D, Yefanov O, Sommer BP, Mogk S, Duszhenko M, Betzel C, Schneider TR, Chapman HN, Redecke L (2014) **Serial crystallography on in vivo grown microcrystals using synchrotron radiation**. *IUCrJ*. 2014 1, 87-94

**V** - Liu W, Wacker D, **Gati C**, Han GW, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A, Zatsepin NA, Li D, Messerschmidt M, Boutet S, Williams GJ, Koglin JE, Seibert M, Wang C, Shah STA, Basu S, Fromme R, Kupitz C, Rendek KN, Grotjohann I, Fromme P, Kirian RA, Beyerlein KR, White TA, Chapman HN, Caffrey M, Spence JHC, Stevens RC, Cherezov V (2013) **Serial Femtosecond Crystallography of G Protein-Coupled Receptors**. *Science*. 2013; 342 (6165):1521-4



## List of additional publications

**VI** Yefanov O, Mariani V, Gati C, White TA, Chapman HN, Barty A (2015) **Accurate determination of segmented X-ray detector geometry.** Optics Express 23(22) 28459-28470

**VII** - Kang Y., Zhou EX, Gao X, He Y, Liu W, Ishchenko A, Barty A, White TA, Yefanov O, Han GW, Xu Q, de Waal PW, Ke J, Tan MHE, Zhang C, Moeller A, West GM, Pascal B, van Eps N, Caro LN, Vishnivetskiy SA, Lee RJ, Suino-Powell KM, Gu X, Pal K, Ma J, Zhi X, Boutet S, Williams GJ, Messerschmidt M, **Gati C**, Zatsepin NA, Wang D, James D, Basu S, Roy-Chowdhury S, Conrad C, Coe J, Liu H, Lisova S, Kupitz C, Grotjohann I, Fromme R, Jiang Y, Tan M, Yang H, Li J, Wang M, Zheng Z, Li D, Howe N, Zhao Y, Standfuss J, Diederichs K, Dong Y, Potter CS, Carragher B, Caffrey M, Jiang H, Chapman HN, Spence JCH, Fromme P, Weierstall U, Ernst OP, Katritch V, Gurevich VV, Griffin PR, Hubbell WR, Stevens RC, Cherezov V, Melcher K, Xu HE (2015) **Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser.** Nature 523, 561–567 (30 July 2015)

**VIII** - Conrad CE, Basu S, James D, Wang D, Schaffer A, Roy-Chowdhury S, Zatsepin NA, Aquila A, Coe J, **Gati C**, Hunter MS, Koglin JE, Kupitz C, Nelson G, Subramanian G, White TA, Zhao Y, Zook J, Boutet Sebastien, Cherezov V, Spence JCH, Fromme R, Weierstall U, Fromme P (2015) **A novel inert crystal delivery medium for serial femtosecond crystallography.** IUCrJ. 2015 Jun 30;2(Pt 4):421-30

**IX** - Wu W, Nogly P, Rheinberger J, Kick LM, **Gati C**, Nelson G, Depi X, Standfuss J, Schertler G, Panneels V (2015) **Batch crystallization of rhodopsin for structural dynamics using an X-ray free electron laser (X-FEL)** Acta Cryst. F (2015) Jun 27; F71

**X** - Panneels V, Wu W, Tsai CJ, Nogly Przemek, Rheinberger J, Jaeger K, Cicchetti G, **Gati C**, Kick LM, Sala L, Capitani G, Milne C, Padeste C, Pedrini B, Li XD, Standfuss J, Abela R, Schertler G (2015) **Time-resolved structural studies with serial crystallography: A new light on retinal proteins.** Structural Dynamics 2, 041718 (2015)

**XI** - Galli L, Son SK, Klinge M, Bajt S, Barty A, Bean R, Betzel C, Beyerlein KR, Caleman C, Doak RB, Duszynski M, Fleckenstein H, **Gati C**, Hunt B, Kirian RA, Liang M, Nanao MH, Nass K, Oberthuer D, Redecke L, Shoeman R, Stellato F, Yoon CH, White TA, Yefanov O, Spence J, Chapman HN (2015) **Electronic damage in S atoms in a native protein crystal induced by an intense X-ray free-electron laser pulse.** Structural Dynamics 2.4

(2015): 041703.

**XII** - Barends T, White TA, Barty A, Foucar L, Messerschmidt M, Alonso-Mori R, Botha S, Chapman H, Doak RB, Galli L, **Gati C**, Gutmann M, Koglin J, Markvardsen A, Nass K, Oberthur D, Shoeman RL, Schlichting I, Boutet S (2015) **Effects of self-seeding and crystal post-selection on the quality of Monte Carlo-integrated SFX data.** *J Synchrotron Radiat.* 2015 May 1;22(Pt 3):644-652. Epub 2015 Apr 22.

**XIII** - Fenalti G, Zatsepin N, Betti C, Giguere P, Ishchenko A, Liu W, Guillemyn K, Han GW, Zhang H, James D, Wang D, Weierstall U, Spence JCH, Boutet S, Messerschmidt M, Williams GJ, **Gati C**, White TA, Oberthuer D, Metz M, Yoon CH, Barty A, Chapman H.N., Basu S, Coe J, Conrad C, Fromme R, Fromme P, Tourwé D, Schiller PW, Roth BL, Ballet S, Katritch V, Stevens RC, Cherezov V (2015) **Structural basis for bifunctional peptide recognition at human delta-opioid receptor.** *Nature Structural and Molecular Biology.* 2015 Mar;22(3):265-8

**XIV** - Tenboer J, Basu S, Zatsepin N, Pande K, Milathianaki D, Frank M, Hunter M, Boutet S, Williams G, Koglin JE, Oberthuer D, Heymann M, Kupitz C, Conrad C, Coe J, Roy-Chowdhury S, Weierstall U, James D, Wang D, Grant T, Barty A, Yefanov O, Scales J, **Gati C**, Seuring C, Srajer V, Henning R, Schwander P, Fromme R, Ourmazd A, Moffat K, Van Thor J, Spence JHC, Fromme P, Chapman HN, Schmidt M (2014) **Time Resolved Serial Femtosecond Crystallography Captures High Resolution Intermediates of Photoactive Yellow Protein.** *Science.* 2014, 346(6214), 1242-1246

**XV** - Yefanov O, **Gati C**, White TA, Spence JCH, Chapman HN, Barty A (2014) **Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography.** *Phil. Trans. R. Soc. B.* 2014 369

**XVI** - Demirci H, Sierra RG, Laksmono H, Shoeman RL, Botha S, Barends TRM, Nass K, Schlichting I, Doak RB, **Gati C**, Williams G, Boutet S, Messerschmidt M, Jogle G, Dahlberg AE, Gregory ST, Bogan MJ (2013) **Serial femtosecond X-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an X-ray free-electron laser.** *Acta Cryst. F.* 2013; 69 (Pt 9):1066-9



# Contents

<b>1</b>	<b>Motivation</b>	<b>1</b>
<b>2</b>	<b>Introduction to X-ray crystallography</b>	<b>3</b>
2.1	Definition of a crystal and reciprocal lattice . . . . .	6
2.2	Scattering of X-rays by atoms and molecules . . . . .	9
2.3	Scattering by an infinite crystal . . . . .	11
2.4	The phase problem . . . . .	13
2.5	Temperature factor . . . . .	14
2.5.1	Wilson plot . . . . .	15
2.6	Ewald construction . . . . .	16
2.7	Partiality . . . . .	18
<b>3</b>	<b>Protein crystallography</b>	<b>21</b>
3.1	‘Conventional’ macromolecular crystallography . . . . .	21
3.2	Molecular replacement . . . . .	23
3.3	Crystallographic metrics and model refinement . . . . .	24
3.4	‘Conventional’ data collection in macromolecular crystallography . .	26
3.5	Radiation damage in X-ray crystallography experiments . . . . .	27
3.5.1	The Henderson limit . . . . .	29
3.6	Protein crystallization . . . . .	31
3.6.1	Principals of protein crystallization . . . . .	31
<b>4</b>	<b>Advanced X-ray Sources</b>	<b>39</b>
4.1	Synchrotron radiation . . . . .	39

4.2	X-ray detectors in ‘conventional’ macromolecular crystallography . .	42
4.3	Free-Electron laser radiation . . . . .	43
4.3.1	Self-Amplified Spontaneous Emission (SASE) effect . . . . .	44
4.4	Linac Coherent Light Source (LCLS) . . . . .	45
4.4.1	Coherent X-ray Imaging (CXI) beamline . . . . .	46
<b>5</b>	<b>Serial femtosecond crystallography (SFX)</b>	<b>47</b>
5.1	Introduction to SFX . . . . .	47
5.2	Instrumentation in SFX . . . . .	50
5.2.1	Gas Dynamic Virtual Nozzle . . . . .	50
5.2.2	Lipidic Cubic Phase Injector . . . . .	51
5.2.3	Detectors in SFX . . . . .	53
5.3	Data analysis in SFX . . . . .	55
5.3.1	Pre-processing of raw SFX data: Cheetah . . . . .	55
5.3.2	From diffraction patterns to hkl intensities: CrystFEL . . . . .	57
5.4	High resolution structure from the smallest protein crystals . . . . .	63
5.4.1	The structure of granulovirus (GV) polyhedrin . . . . .	63
5.4.2	Sample production and characterization . . . . .	65
5.4.3	Data collection and processing . . . . .	66
5.4.4	Structure determination and comparison to homologous structures . . . . .	68
5.4.5	GV as a reference for analyzing the limits of SFX . . . . .	71
5.5	Effects of experimental parameters and improvements of SFX data quality . . . . .	74
5.5.1	Detector Saturation . . . . .	74
5.5.2	Detector calibration . . . . .	77
5.5.3	Improved integration approaches . . . . .	78
5.5.4	Effect of averaging of diffraction patterns on data quality . .	78
5.5.5	Effects of per pattern resolution cutoff on data quality . . . .	80
<b>6</b>	<b>G protein-coupled receptors in LCP-SFX</b>	<b>85</b>
6.1	Physiology of G protein-coupled receptors . . . . .	85

<i>CONTENTS</i>	iii
6.2 Structural biology of G protein-coupled receptors . . . . .	89
6.3 First structure of human membrane protein at room temperature . . . . .	90
6.4 First completely unknown structure solved by SFX . . . . .	100
6.4.1 Structure of the human Angiotensin II - receptor 1 (AT <sub>1</sub> R) . . . . .	100
<b>7 Serial synchrotron crystallography</b>	<b>107</b>
7.1 First successful serial synchrotron crystallography experiment . . . . .	109
7.1.1 Recombinant <i>in vivo</i> crystallization of <i>T.brucei</i> procathepsinB . . . . .	111
7.1.2 Experimental setup and data collection . . . . .	112
7.1.3 Data processing, analysis and quality assessment . . . . .	114
7.1.4 Discussion . . . . .	119
7.2 Sample delivery by flowing approaches . . . . .	123
7.2.1 Capillary based data collection . . . . .	124
7.2.2 Experimental setup and data collection scheme . . . . .	125
7.2.3 Data processing, analysis and quality assessment . . . . .	126
7.2.4 Discussion . . . . .	126
<b>8 Summary and outlook</b>	<b>131</b>
8.1 Improvements in sample consumption . . . . .	131
8.2 Illuminating medically important receptors - GPCRs in LCP-SFX . . . . .	132
8.3 Opening the method to a broader community . . . . .	133
8.4 On the road to single-particle X-ray imaging . . . . .	134
<b>9 Appendix</b>	<b>137</b>
9.1 Reprint of published articles . . . . .	137
9.2 Chemicals and Formulations . . . . .	164
9.2.1 Chemicals - Hazard and Precautionary Statements . . . . .	164
9.3 Curriculum Vitae . . . . .	164
<b>Bibliography</b>	<b>172</b>



# Chapter 1

## Motivation

A fundamental question in biological science is to understand life on an atomic level. Structural biology combines the fields of molecular biology, biochemistry and biophysics, and is concerned with providing information about the atomic architecture of biomolecules. The structure of a protein does not only define its size and shape, but ultimately dictates its function and interactions with other biomolecules. Hence, the knowledge of the biomolecular structure is vital to an understanding of the kinetics and dynamics of the macromolecule.

Protein X-ray crystallography is the most broadly used technique to determine macromolecular structure to near-atomic resolution. This lens-less imaging technique overcomes the problem of weak scattering signal from an individual macromolecule by using an ordered three-dimensional lattice of biomolecules. This results in a drastic increase in the signal to noise ratio of the scattered waves, due to constructive interference between scattered X-rays from millions of copies of the molecule. X-ray crystallography has contributed about 90% of all high-resolution protein structures (more than 100,000) in the Protein Data Bank (PDB), making it the most important technique for biomolecular structure determination. Despite this tremendous success, one major limitation of this technique is the necessity to grow well-ordered crystals of suitable size, depending on the intensity of the radiation source. In addition, the effect of radiation damage induced in the sample during exposure to X-rays limits the recordable information content from each crystal. While the use of synchrotron light sources has decreased the crystal size limitation, the problem of radiation damage due to a high radiation dose has significantly increased. Even by counteracting this problem by cryocooling of the sample, the crystal typically survives in the X-ray beam only for milliseconds at a modern synchrotron source.



The advent of X-ray Free-Electron Lasers (XFELs) has opened a new era in X-ray imaging. The high brilliance of X-ray pulses from XFELs makes it possible to collect data from crystals of nanometer size, and their femtosecond duration enables overcoming structural radiation damage. These characteristics of radiation from XFELs have led to the development of the method of serial femtosecond crystallography (SFX), in which diffraction signal from protein nanocrystals in random orientation is merged to produce a three-dimensional diffraction volume. One of the primary motivations of this dissertation is to explore the current limitations of SFX with respect to the lower limit on the size of protein crystals that can still provide high-resolution diffraction information. The main focus of the work is on the SFX data analysis pipeline, which has been developed and established over the past years. Various bottlenecks and problems associated with the technique are highlighted and discussed. In parallel, the application of SFX to answer important biological questions using delicate membrane proteins as target systems is discussed. I investigate the SFX approach on the medically important class of G protein-coupled receptors to underline the importance and broad applicability of this technique.

Another aspect of this work is the application of the serial crystallography approach to modern synchrotron beamlines to enable the investigation of sample size limitations at this readily available light source. This adaptation of the serial crystallography approach will broaden the applicability of serial data collection for the structural biology community.

## Chapter 2

# Introduction to X-ray crystallography

The interest of humankind to understand the composition of matter in its greatest details can be traced back to ancient cultures. However, experimental data for three-dimensional atomic models of matter were not available until the discovery of X-ray radiation on the 8th of November 1895 [1] by Wilhelm C. Roentgen. His discovery of novel, invisible rays led to the groundbreaking experiments by Max von Laue and coworkers [2], followed by the experiments of William Henry and William Lawrence Bragg [3], which fundamentally changed our understanding of life on earth – the technique of X-ray crystallography was born.

Shortly after the discovery of X-rays by Roentgen [1], a variety of experiments were carried out to investigate their true nature. These included experiments trying to prove a particle-like behavior by applying magnetic fields, and experiments to test a wave-like behavior by investigating the refractive index of X-rays by slit-experiments, of which none was successful. In 1896 Stokes and Wiechert hypothesized that the energy-loss which the electron experiences upon impact on the cathode is transformed into an ‘impulse’, short waves consisting of electromagnetic pulses that spread spherically around the electron, at the speed of light [4]. Based on this hypothesis, in 1897 Dorn tried to quantify the amount of energy released from the retarded electron upon impact. Subsequently, Wien determined the approximate wavelength of X-rays on the order of  $10^{-10}$  cm [4], which is a factor of thousand shorter than the wavelength of visible light. This estimated very short wavelength was one possibility to explain all failed experiments regarding the refractive index

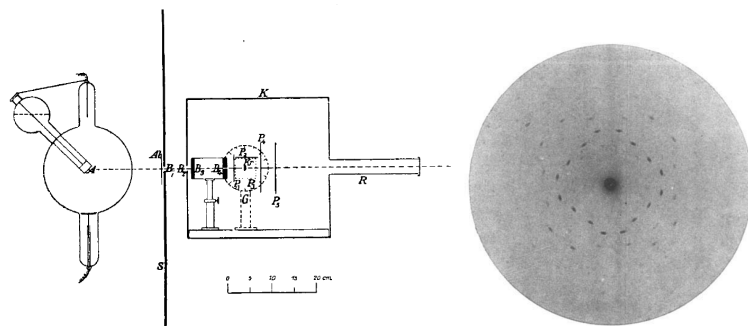


Figure 2.1: Scheme of the first X-ray crystallography experiments by Friedrich, Knipping and von Laue. (left) experimental setup with X-ray tube and photosensitive film (right) first diffraction pattern obtained from  $\text{CuSO}_4$  (from Friedrich *et al.* 1913)

of X-rays. Following Wien's results, Max von Laue postulated an experiment to test the wave-like nature of X-rays by using crystals as a grating to observe interference effects, since their inter-atomic distance was estimated to be on the same order of magnitude as the wavelength of X-rays. The experiment was carried out by his coworkers Friedrich and Knipping in 1912 [2] in Munich, who placed copper sulfate ( $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ ) and other crystalline minerals into a white X-ray beam and recorded the X-rays exiting the crystal on photographic film (Figure 2.1), now called Laue photography. Even though von Laue's initial interpretation of the obtained diffraction pattern was proven to be wrong (he explained it by an interference of the induced fluorescence by the atoms in the crystal), this discovery did not only demonstrate the wave-like nature of X-rays, but made scientists aware of one of the most important features of X-ray crystallography. As a direct consequence of the collected diffraction pattern, it was possible to derive inter-atomic distances within the crystal. Hence it was now, for the first time, possible to derive an atomic model of matter. Due to its impact for the scientific community, this experiment was awarded with the Nobel Prize in Physics in 1914. In the following two years, W. L. Bragg and his father W. H. Bragg successfully interpreted von Laue's results, designed a similar experimental setup (replacing the film by an ionization chamber, observing individual peaks at a time) and were able to solve the first crystal structures of NaCl, KCl, KBr and KI [3] which was awarded with the Nobel Prize in Physics 1915.

In the last century our knowledge of matter and X-rays has improved dramatically, even though the basic experimental geometry of von Laue and the Braggs

has not changed significantly in modern crystallography experiments. X-rays are electromagnetic waves located between UV and  $\gamma$ -rays on the spectrum of light, with photon energies in the range of 100 eV to 100 keV. The border between soft and hard X-rays is roughly defined between 2 - 5 keV. As is the case with any kind of radiation, an extensive exposure of matter to X-rays can cause damage ranging from the atomic to medically relevant levels, causing burns, hair loss or even cancer. The interaction probability between X-rays with a photon energy  $E$  and an atom, more precisely the core-shell electrons, with atomic number  $Z$  scales as  $Z^3/E^3$ . As a result hard X-rays with  $E > 5$  keV have a reduced photoelectric absorption cross section, resulting in a longer penetration depth. This makes them ideally suited for structural imaging of materials, including medical radiography, tomography and X-ray crystallography, the latter being the subject of this dissertation. Interaction between X-rays and the electrons of atoms within the crystal causes X-rays to be diffracted spherically around the scattering source. Constructive and destructive interference between these diffracted rays gives rise to a specific diffraction pattern, which can be recorded on the detector. By combining the information obtained from the diffraction pattern, including angles, relative positions and intensities of those diffracted rays (omitting the phase problem for now), it is possible to calculate an electron density map (Section 2.3). The resulting electron density provides crucial information about atomic positions, variability and disorder of the crystalline material.

Crystallographic structural investigation can be performed by X-rays, neutrons and electrons. X-ray diffraction is the most widely used technique in the scientific community. Electron crystallography has been successfully applied to structure determination of 2D protein crystals [5], but due to limitations on the availability of samples, it is rarely used nowadays. Recently, scientists from Janelia Farm Research Campus [6] and Leiden University [7][8] are reviving the field by applying electron crystallography to 3D nanocrystals. The major advantage of using electrons for diffraction experiments is the much larger cross section due to strong electron to charge density interactions, making it possible to analyze protein nanocrystals or 2D crystals with conventional electron sources and detectors. Nevertheless, there are two main limitations associated with using electron beams for diffraction experiments. The first fundamental limitation is the problem of multiple scattering events, limiting possible sample thickness (approx. 100 nm for biomolecules) and resulting in data analysis being difficult or even impossible for thicker samples. A second limitation is severe radiation damage induced by the strong interaction between electric fields.

The method of neutron scattering has been a growing field in the past decade,

due the possibility to observe the position of hydrogen atoms, which are not ‘visible’ in any of the other techniques. Because of a very small cross section between neutrons and atomic nuclei it is typically necessary to use crystals in the millimeter size range, in order to overcome signal-to-noise limitations, while experiments can take up to several weeks, making both sample preparation and experimental work very challenging. Novel sources, such as the European Spallation Source (ESS) are estimated to produce neutron beams strong enough to allow crystal sizes on the order of a few hundred micrometers, which would make this technique applicable to a much wider range of samples.

The technique of X-ray crystallography is extensively used in the field of materials science, since many properties of materials, such as conductivity, are directly correlated to their atomic structure and symmetry and often crystalline materials consist of atoms with a high atomic number ( $Z$ ). The method of X-ray crystallography also represents the most widely used bio-imaging technique, with more than 100.000 solved protein structures in the protein data bank (PDB), corresponding to more than 90% of all deposited structures. This is of great interest in medical and pharmaceutical research, because of the direct relationship between the structure and function of proteins.

## 2.1 Definition of a crystal and reciprocal lattice

A conventional crystal is a solid, homogenous material, which consists of an ordered array of atoms or molecules, which extends in either two or three dimensions. It can also be defined as the convolution of the Fourier transform of the lattice function with the contents of the unit cell (electron density), to which we will come back later on (Section 2.3). First attempts in the field of crystallography began more than 2000 years ago, during the ancient Greek and Arabic era. Before the discovery of X-rays, crystals were characterized by their macroscopic appearance, overall symmetry and angles of their facets. An important reference on X-ray crystallography is the book ‘*Fundamentals of Crystallography*’ by Carmelo Giacovazzo *et al.*, from which most of the equations in this chapter have been derived [9][10].

The smallest repeating unit of a crystal is the unit cell (UC), whose axes are defined by nomenclature as  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  spanning the angles  $\alpha, \beta, \gamma$ . Several combinations of axes and angles are defined as the 14 Bravais lattices (in 3D), which are grouped into 7 lattice systems (Triclinic, Monoclinic, Orthorhombic, Tetragonal, Cubic, Rhombohedral, Hexagonal - Figure 2.2). In addition to the lattice system,

Bravais lattices include a centering component, which can be either Primitive (P), Face (F), Body (I), Base (A, B or C) or Rhombohedral (R) centered. The symmetry operators of a crystalline lattice include rotations, screw axes, mirrors, inversions, roto-reflections and glide planes. Leonhard Euler identified all possible combinations, due to geometrical limitations, of symmetry operators within a finite object, which are called point groups and only include symmetry operators that do not contain a translational component.

By combining 32 available point groups with 14 Bravais lattices (Figure 2.2) together with additional screw axis and glide planes, a total of 230 space groups can be derived for any given ordered material. Proteins are made up of amino acids, which have a chiral center around the  $C_\alpha$  atom. Because of this chirality, it is impossible to observe glide planes, mirrors or inversion centers in protein crystals, since these would result in a mirror image of the original molecule, and hence only 65 space groups are observable in protein crystallography. The smallest building block that can fill the entire volume of a unit cell by symmetry operators is called an asymmetric unit (AU).

The reciprocal lattice represents the Fourier transform of the real-space lattice. It is a mathematical concept, introduced by Paul P. Ewald in 1921, which describes the relationship between the real-space lattice and the symmetry of the diffraction pattern in reciprocal space. We defined  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  as the directions of translation inside the crystal in direct space, while the corresponding translations in reciprocal space are  $\mathbf{a}^*$ ,  $\mathbf{b}^*$ ,  $\mathbf{c}^*$  which satisfy the following conditions:

$$\mathbf{a}^* \cdot \mathbf{b} = \mathbf{a}^* \cdot \mathbf{c} = \mathbf{b}^* \cdot \mathbf{a} = \mathbf{b}^* \cdot \mathbf{c} = \mathbf{c}^* \cdot \mathbf{a} = \mathbf{c}^* \cdot \mathbf{b} = 0 \quad (2.1)$$

$$\mathbf{a}^* \cdot \mathbf{a} = \mathbf{b}^* \cdot \mathbf{b} = \mathbf{c}^* \cdot \mathbf{c} = 1 \quad (2.2)$$

the first equation suggests that  $\mathbf{a}^*$  is normal to the plane formed by  $(\mathbf{b}, \mathbf{c})$  etc., while the second equation fixes the modulus of each vector. Due to this reciprocity, long translational repeats within a crystal will result in small spacing of Bragg peaks observable on the detector. Finally, the reciprocal lattice of a primitive cubic crystal lattice with an axis  $\mathbf{a}$  shows the same symmetry: a cubic reciprocal lattice with axis length  $\frac{1}{a}$ . The point symmetry of the reciprocal lattice is always identical to the Bravais lattice of the crystal, while the centering of the two lattices can differ - e.g. an Face-centered cubic (FCC) Bravais lattice results in an I-centered cubic (BCC) reciprocal lattice.

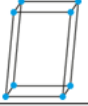
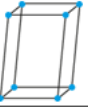
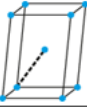
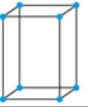
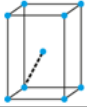
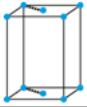
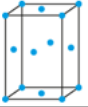

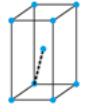
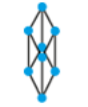
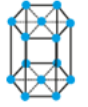

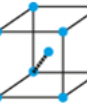

Bravais lattice	Parameters	Simple (P)	Volume centered (I)	Base centered (C)	Face centered (F)	Point groups	Laue groups
<b>Triclinic</b>	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$					$1, \bar{1}$	$\bar{1}$
<b>Monoclinic</b>	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ$ $\beta \neq 90^\circ$					$2, m, 2/m$	$2/m$
<b>Orthorhombic</b>	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$					$222, mm2, mmm$	$mmm$
<b>Tetragonal</b>	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$					$4, \bar{4}, 4/m$	$4/m$
						$422, \bar{4}2m, 42m, 4/mmm$	$4/mmm$
<b>Rhombohedral</b>	$a = b = c$ $\alpha = \beta = \gamma < 120^\circ$					$3, \bar{3}$	$\bar{3}$
						$32, 3m, \bar{3}m$	$\bar{3}m$
<b>Hexagonal</b>	$a = b \neq c$ $\alpha = 120^\circ$ $\beta = \gamma = 90^\circ$					$3, \bar{3}, 321, 312, 31m, 3m1, \bar{3}1m, 6, \bar{6}, 6/m$	$6/m$
						$622, 6mm, 6m2, 6/mmm$	$6/mmm$
<b>Cubic</b>	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$					$23, m\bar{3}$	$m\bar{3}$
						$432, \bar{4}3m, m\bar{3}m$	$m\bar{3}m$

Figure 2.2: Representation of 14 Bravais lattices and their corresponding point groups and Laue classes.

The key information obtained from a diffraction experiment is the observation of inter-atomic distances in a crystal. Because inter-atomic distances are always centrosymmetric (distance AB is equal to BA), the point group symmetry of crystallographic diffraction data is hidden and always contains a mirror symmetry. Hence, the resulting symmetry of a diffraction pattern of any crystal system will be represented by the 11 Laue groups, instead of 32 point groups (Figure 2.2). The final symmetry of a crystal, including its space group symmetry, can only be determined by the analysis of Bragg peak intensities, most importantly by the identification of systematic absences, which is described in detail in the International Tables for Crystallography, Volume A: Space-group Symmetry [11][12].

## 2.2 Scattering of X-rays by atoms and molecules

X-ray scattering experiments are a group of analytical methods to analyze chemical compositions (X-ray spectroscopy) and structural information of matter from low (Small Angle X-ray Scattering – SAXS) to atomic resolution (X-ray crystallography). Scattering events occur due to the interaction of electromagnetic waves, in this particular case X-rays, with electrons in the inner atomic shells, namely core shell electrons. After impact with these photons, the affected electrons start to oscillate, resulting in electromagnetic dipole radiation (secondary waves). This is the typical form of scattering, referred to as elastic scattering (Thomson scattering), where the scattered wave has the same wavelength and energy as the incoming photon:

$$I_{Th} = I_0 \frac{e^4}{m^2 r^2 c^4} \sin^2 \theta \quad (2.3)$$

where  $I_{Th}$  is the intensity of the resulting wave,  $I_0$  is the intensity of the incoming wave,  $e$ ,  $m$  and  $r$  are the charge, mass and classical radius of an electron, respectively, and  $\theta$  is the angle between the direction of acceleration of the electron and the direction of observation. The dependence on  $\theta$  is due to a varying path length for different angles, which results (together with other effects explained later) in a dampening of the scattered intensities of high scattering angles.

Unlike Thomson scattering, during inelastic scattering events, the energy of the incident photon is not conserved. This is called Compton scattering. The resulting emitted wave shows a decrease in energy (increase in wavelength) or vice versa (inverse Compton scattering). Compton scattering does not contribute to the diffraction pattern apart from diffuse background scattering, but is generally important because it demonstrates that light cannot purely be described in terms of a wave-like behavior. Another effect of photons interacting with electrons is absorption, where



the energy of the photon is transferred to the electron, which can get excited into a higher state, leading to the photon completely losing its energy and finally causing secondary effects, such as fluorescence, Auger decay, etc. (Section 3.5).

The scattering of X-rays by atoms depends on the interaction between the electromagnetic field of the incident X-ray beam and the charged particle (electron). Due to the probabilistic distribution of electrons surrounding an atom, the scattering of X-rays can be estimated and is commonly referred to as the atomic scattering factor or atomic form factor:

$$f(\mathbf{q}) = \int_S \rho(\mathbf{r}) e^{2\pi i \mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \quad (2.4)$$

where  $S$  is the region of the atom where the probability to observe an electron is higher than 0,  $\rho(\mathbf{r})$  is the electron density which is centrosymmetric about the origin ( $\rho(\mathbf{r}) = \rho(-\mathbf{r})$ ), and  $\mathbf{q}$  is the momentum transfer.

In terms of the unit vectors  $\mathbf{k}_0$  and  $\mathbf{k}$  representing the directions of the incident X-ray beam and the elastically scattered wave, respectively, and the wavelength  $\lambda$ , the momentum transfer or scattering vector  $\mathbf{q}$  is given as:

$$\mathbf{q} = \frac{1}{\lambda} (\mathbf{k} - \mathbf{k}_0). \quad (2.5)$$

The relationship between the magnitude of the scattering vector and the scattering angle  $\theta$  is given as:

$$q = \frac{2 \sin \theta}{\lambda} \quad (2.6)$$

The scattered wave from a continuous scattering center is represented as:

$$F(\mathbf{q}) = \int_V \rho(\mathbf{r}) e^{2\pi i \mathbf{q} \cdot \mathbf{r}} d\mathbf{r} = FT[\rho(\mathbf{r})] \quad (2.7)$$

where  $FT$  is the Fourier Transform operator of the electron density  $\rho(\mathbf{r})$ . Due to the nature of Fourier transforms, its inverse also holds true:

$$\rho(\mathbf{r}) = \int_V F(\mathbf{q}) e^{-2\pi i \mathbf{q} \cdot \mathbf{r}} d\mathbf{q} = FT^{-1}[F(\mathbf{q})] \quad (2.8)$$

which results in the fact, that by applying an inverse FT of the continuous structure amplitude it is possible to calculate the electron density  $\rho(\mathbf{r})$ . It is crucial to know both the amplitudes of the scattered waves, as well as their phases to completely

define the electron density and vice versa.

The wave scattered by a molecule or unit cell can be described as the summation over the waves scattered by all the electrons inside the molecule or unit cell. It is dependent both on the number of electrons as well as their positions in the electron cloud of an atom or molecule, which is described as the electron density. Defining  $\rho(\mathbf{r}_j)$  as the electron density of the  $j^{\text{th}}$  atom of a molecule  $M$ , at position  $\mathbf{r}_j$  with  $N$  atoms in the molecule (or unit cell):

$$\rho_M(\mathbf{r}) = \sum_{j=1}^N \rho_r(\mathbf{r} - \mathbf{r}_j) \quad (2.9)$$

with the total amplitude of the scattered wave:

$$F_M(\mathbf{q}) = \sum_{j=1}^N f_j(\mathbf{q}) e^{2\pi i \mathbf{q} \cdot \mathbf{r}_j} \quad (2.10)$$

where  $f_j(\mathbf{q})$  is the atomic scattering factor of the  $j^{\text{th}}$  atom of the molecule and  $F_M$  is the structure factor of the molecule (unit cell), which depends on the arrangement of atoms in the molecule (unit cell).

## 2.3 Scattering by an infinite crystal

One of the simplest but most important equations in crystallography has been derived by W. L. Bragg. It describes the interaction between the incoming wave and the crystal, which consists of a set of parallel planes of atoms with a lattice spacing  $d$ . If waves are scattered by those planes, the maximal positive interference is described by a path difference  $d$  of (Figure 2.3):

$$2d \sin \theta = n\lambda \quad (2.11)$$

where  $2\theta$  is the angle between incident wave  $\mathbf{k}_0$  and the diffracted wave  $\mathbf{k}$  (need to change theta in the figure),  $d$  is the distance between the lattice planes, and  $\lambda$  the wavelength. Only in the case of  $n$  being an integer order of the reflection, waves will be scattered in phase (constructive interference). If Bragg's law is not obeyed, the reflected waves will scatter out of phase (destructive interference), canceling out measurable diffraction intensities.

A crystal consists of a molecular unit which repeats upon translation:

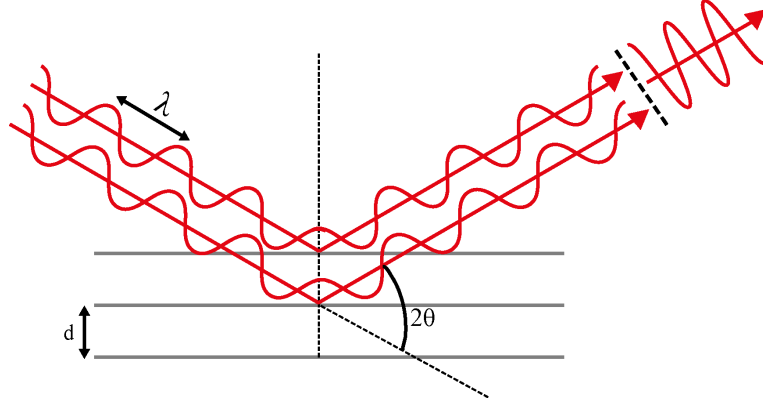


Figure 2.3: Schematic of Bragg's law, with the incident and diffracted beam (red), with a wavelength  $\lambda$  and a scattering angle  $\theta$ . Illustration of positive interference due to the diffracted rays scattering in phase.

$$\mathbf{R}_n = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c}$$

where  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  are lattice vectors and  $n_i$  are integers. The total scattering by the crystal is a summation of the scattering by all unit cells with respect to a single origin, which is arbitrary.

$$F(\mathbf{q}) = F_M(\mathbf{q}) \sum_{n_1, n_2, n_3=0}^{+\infty} e^{2\pi i \mathbf{q} \cdot \mathbf{R}_n} \quad (2.12)$$

Since  $n_1, n_2, n_3$  tend to infinity, the summation  $\sum_{n_1=0}^{\infty} e^{2\pi i \mathbf{q} \cdot \mathbf{a}}$ , and that over  $\mathbf{b}$  and  $\mathbf{c}$  is equal to 0, except when  $\mathbf{q} \cdot \mathbf{a}$ ,  $\mathbf{q} \cdot \mathbf{b}$ ,  $\mathbf{q} \cdot \mathbf{c}$  are integers  $h, k$  and  $l$ , respectively. These are known as the Laue conditions.

As a consequence of the Laue conditions, the amplitude of all summed scattered waves by a crystal is:

$$F_{\infty}(\mathbf{q}) = \frac{1}{V} F_M(H) \sum_{h, k, l=-\infty}^{+\infty} \delta(\mathbf{g} - \mathbf{g}_{hkl}) \quad (2.13)$$

with  $\delta(\mathbf{x})$  as the Dirac delta function and  $\mathbf{g}_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$  as the generic lattice vector of the reciprocal lattice. A crystal with axis  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  with an infinite number of unit cells along those  $n_1, n_2, n_3$ . Diffraction of a crystal is the summation of the scattered waves by all unit cells inside the crystal with respect to the origin.

The main aim of a diffraction experiment is to reproduce the electron density

of the molecule in order to obtain the average positions of its atoms. The electron density of the crystal is the inverse Fourier transform of its complex structure factor. Applying  $FT^{-1}$  on equation 2.12, the electron density of a crystal can be written as:

$$\rho_{\infty}(\mathbf{r}) = \rho_M(\mathbf{r}) * L(\mathbf{r}) \quad (2.14)$$

which is the convolution of its lattice function  $L(\mathbf{r})$  with the electron density of the unit cell  $\rho_M(\mathbf{r})$ .

The structure factor equation 2.13 is a Fourier transform. Hence, it is possible to regenerate the electron density of a molecule by calculating the inverse Fourier transform of the experimentally obtained structure factors  $F_H$  and phases  $\alpha_h$  (Section 3.1), which are typically obtained from different sources. At a point  $(x, y, z)$ , by substituting the Euler relation  $F_{hkl} = F_{hkl}e^{i\alpha_h}$  into the electron density equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} F_{hkl}e^{2\pi i\alpha_{hkl}} * e^{-2\pi i(hx+ky+lz)} \quad (2.15)$$

dividing by the volume  $V$  of the unit cell to recover the electron density on the correct scale. The electron density equation can be directly used for practical calculation of the electron density at any given point in the unit cell. By calculating the Fourier transform of each reciprocal lattice point for each grid point of the unit cell, the electron density can be recovered to a resolution limited by the grid size and the observed resolution information in the diffraction experiments.

## 2.4 The phase problem

Equation 2.15 can be used to calculate the electron density of the unit cell of an investigated crystal, which involves a complex conjugate, including a real and imaginary part. Since X-ray crystallography is a lens-less imaging technique, a diffraction pattern only contains spatial frequency information in form of structure factor amplitudes, that are the square root of the observed Bragg peak intensities, and the unit cell parameters. The imaginary part, called the phases of the diffraction data, are lost due to the lack of a suitable lens for X-ray radiation. This mathematical problem is known as the ‘Phase problem’, where the electron density can only be calculated by determining both the structure factor modulus ( $F_{hkl}$ ) and the phases ( $\alpha_{hkl}$ ) of the scattered waves (equation 2.15).

For ‘small molecule’ crystallography, typically less than 1000 atoms per UC, this problem is usually addressed by applying **direct methods** ‘DM’, which solely use information from structure factor amplitudes and exploit chemical constraints to derive the phases of different Fourier components. Additionally, **Patterson maps** provide information about inter-atomic distances. The Patterson function ‘PF’ is the auto-correlation (Fourier transform) of the intensities (not the structure factors). It represents inter-atomic distances in the unit cell and peaks in the PF are inter-atomic distance vectors weighted by the number of electrons in the respective atom.

Due to the complexity of biomolecules, more challenging approaches are needed to derive phases for macromolecular crystallography. One of the techniques is **isomorphous replacement** (SIR, MIR), where heavy atom derivative crystals are generated that produce diffraction data with different intensities than those obtained from the native, untreated crystals. This technique requires that the lattice, as well as the molecules assembling the crystal, and hence the position of the spots, are identical within a limited tolerance. By calculating the isomorphous difference  $|F_H| \cong |F_{PH}| - |F_P|$  between the two datasets, it is possible to locate heavy atom sites in the unit cell by direct methods or Patterson maps. This procedure is called the substructure determination, providing an initial reference point to calculate the missing phase information needed to calculate the electron density map of the entire unit cell. Alternatively, differences in intensities are introduced by exploiting the fact that heavy atoms give rise to an **anomalous contribution** (SAD, MAD), given that the incoming intensity is adjusted to the ionization energies of the core shell electrons (e.g. ‘K-edge’) of the respective heavy atom species. This process leads to the result, that Friedel’s law does not hold true, which again enables the substructure determination by direct methods or Patterson maps. An alternative approach, which is the most frequently used phase retrieval algorithm in macromolecular crystallography, named molecular replacement, is discussed in section 3.2.

## 2.5 Temperature factor

Matter consists of atoms which interact through various chemical bond types (VdW, ionic, salt bridges, etc.) with an arrangement typically in an energetic minimum. As soon as atoms are disturbed in this lattice, they will immediately go back into the conformation of lowest possible energy and finally start to oscillate around this local minimum. These oscillations will modify the electron density, typically broadening the volume which would be occupied by electrons at 0 K, under perfect conditions. Due to very short timescales of atomic vibrations (hundreds of femtoseconds), it is sufficient to observe the averaged position of electrons relative to equilibrium. This

average position of electrons inside crystals is observable both over time (with exposure times, in conventional crystallography experiments, of ms to s) and the position of particular electrons sampled over all unit cells of the crystal. Vibrational components perpendicular to the reflecting plane directly reduce the scattered intensity, which is particularly important for high resolution information. In the simplest approximation these vibrational effects are treated as isotropic, meaning equal in all directions. The correction factor  $T$ , also referred to as Debye-Waller factor, for the atomic scattering factor is:

$$T_{iso} = \exp \left[ -B_{iso} \left( \frac{\sin \theta}{\lambda} \right)^2 \right] \quad (2.16)$$

with  $B$  ( $\text{\AA}^2$ ) as the atomic temperature factor, related to the mean square displacement of electrons due to atomic vibration:

$$B_{iso} = 8\pi^2 U \quad (2.17)$$

with  $U$  as the square mean shift of the atom with respect to equilibrium.

As a consequence, the atomic form factor will gain a Gaussian, wavelength and angular dependent term. Additionally, atomic bonds restrain these isotropic motions, such that a more accurate approximation is an anisotropic thermal factor, represented by an ellipsoid around each atom. However, typical protein crystallographic data are not of sufficient quality (or resolution) to apply anisotropic thermal correction.

### 2.5.1 Wilson plot

It is possible to estimate the isotropic effect of thermal motion on the intensity of the scattered wave, namely a drop in intensity with increasing  $q$ , taking the Debye-Waller factor from equation 2.17 into account:

$$f_j^B = f_j^0 e^{-B_{iso} \left( \frac{\sin \theta}{\lambda} \right)^2} \quad (2.18)$$

with the sum of all scattering atoms is leading to the observed intensity, knowing that structure factors on an absolute scale are the square root of the observed intensity ( $I_o$ ), assuming a perfect crystal at 0 K:

$$I_0^B = I_{obs} k e^{-2B_{iso} \left( \frac{\sin \theta}{\lambda} \right)^2} \quad (2.19)$$

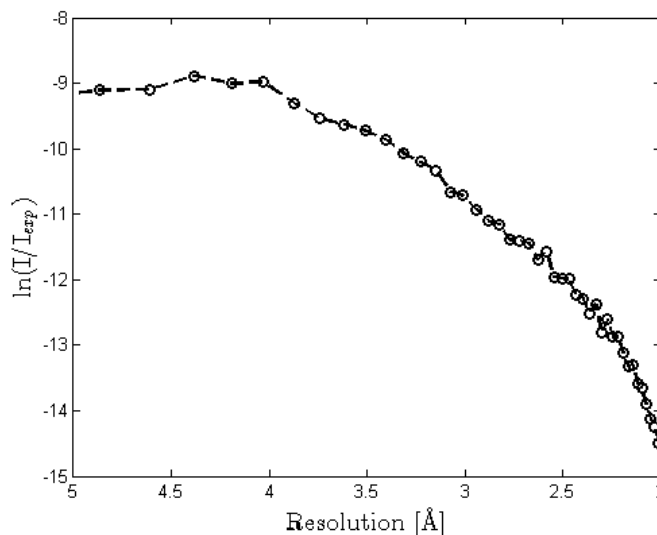


Figure 2.4: Typical Wilson plot of a well-diffracting crystal, with  $\ln\left(\frac{I_{obs}}{I_{exp}}\right)$  as a function of resolution  $\text{\AA}$ .

taking the logarithm, yielding:

$$\ln \frac{I_{obs}^B}{I_0} = \ln(k) - 2B_{iso} \left( \frac{\sin \theta}{\lambda} \right)^2 \quad (2.20)$$

Equation 2.20 represents a straight line  $y(x) = \ln(k) - 2B_{iso}x$ , hence it is possible to obtain the scale factor  $k$  from the intercept. The overall B-factor (in contrast to B-factors of individual atoms or residues) can be directly obtained from the slope of a plot  $\ln\left(\frac{I_{obs}^B}{I_{obs}}\right)$  versus  $\left(\frac{\sin \theta}{\lambda}\right)^2$  which is routinely used in crystallographic data analysis and is referred to as Wilson plot. Wilson plots obtained from data of protein crystals generally don't follow a straight line, because atoms are not completely randomly distributed in the unit cell. Particularly in protein structures at low resolution, distances such as  $C\alpha - C\alpha$  of  $3.8 \text{ \AA}$  in the polypeptide chain are prominent and therefore Wilson plots of protein crystals usually are represented similar to the Wilson plot in Figure 2.4.

## 2.6 Ewald construction

A direct consequence of Bragg's equation is the Ewald sphere with a radius of  $1/\lambda$ . The condition for a possible Bragg reflection to occur is given whenever a reciprocal lattice point meets the Ewald sphere. The Ewald sphere demonstrates the

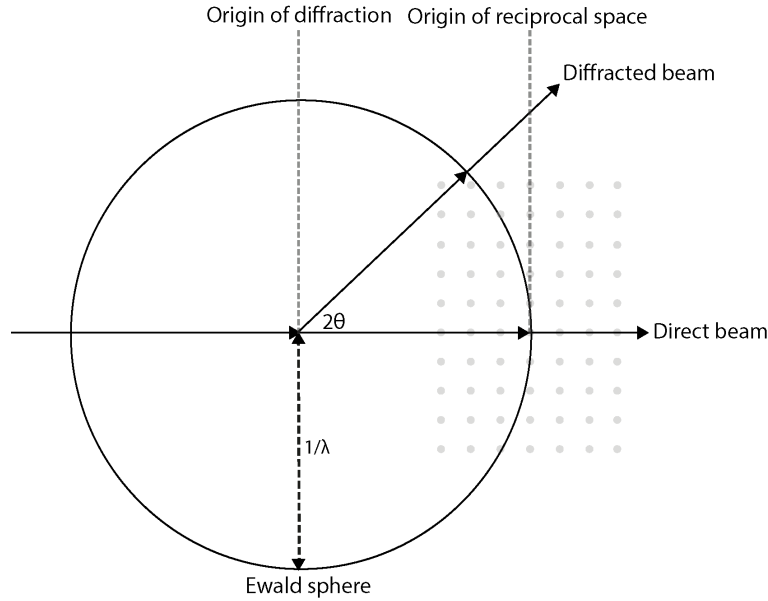


Figure 2.5: Schematic of Ewald sphere construction. The sphere of radius  $1/\lambda$  is termed ‘Ewald sphere’.

relationship between the vectors of the incident and diffracted waves, the diffraction angle for a given reflection and the reciprocal lattice of the crystal. It can be used to derive the maximum possible resolution for a given wavelength, unit cell dimension and detector distance/area. Due to the direct relationship between the symmetry of a crystal in real and reciprocal space, e.g. a primitive cubic lattice being represented as a primitive cubic lattice in reciprocal space, it is possible to predict the conditions (scattering angle) of constructive interference of the diffracted waves from each unit cell. The difference between the incident wave-vector  $\mathbf{k}_0$  and the diffracted wave  $\mathbf{k}$  is defined as the scattering vector  $\Delta\mathbf{s}$ , in case of elastic scattering, which means no energy loss during the scattering event, both  $\mathbf{k}_0$  and  $\mathbf{k}$  are of identical length. In order to probe entire reciprocal space it is necessary to rotate the Ewald sphere around the origin  $O$  (or more practically rotating the crystal rather than the X-ray source). In such a way more reciprocal space lattice points can reach the Ewald sphere, giving rise to a Bragg reflection, and hence to more interpretable data points. Finally, only reflections which have a scattering vector smaller than  $2/\lambda$  (diameter of the Ewald sphere) can potentially give rise to Bragg reflections - this construct is called the limiting sphere and can be used to predict which Bragg reflection can be observed in a given crystal orientation during the exposure.



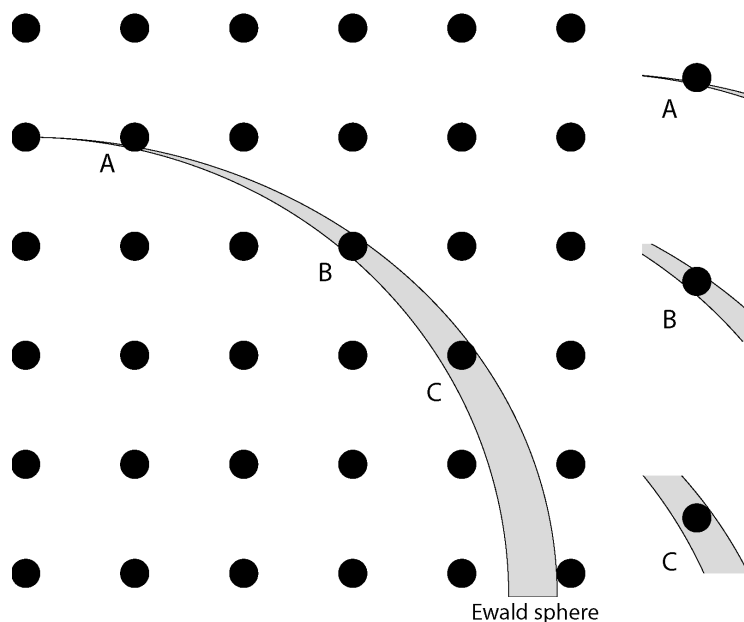


Figure 2.6: Illustration of the problem of partiality estimation, given the finite size of the Ewald sphere intersecting with reciprocal lattice points.

## 2.7 Partiality

With the introduction of rotation/oscillation ‘photography’ by Arndt & Wonnacott at the end of the 1970s [13], where crystals were rotated during the exposure and adjacent frames were probing a consistent slice through reciprocal space, partiality became a problem that needed to be addressed. The first and last frame of these series contained partial reflections (in contrast to solely ‘full’ reflections in Weissenberg photography [14]), which initially were simply discarded, but finally compensated for by estimating the partiality ( $p$ ) of each reflection, allowing an estimate of the full intensity ( $I_{\text{obs}}/p$ ). The calculation of partiality, however, requires accurate knowledge about the experimental geometry, including crystal orientation, unit cell dimensions and the effective mosaic spread, while making use of successively recorded patterns during rotation series [15]. Another prerequisite is that scaled data are available, which is performed only on full reflections (low resolution), while the final data-set includes the partiality corrected reflections [16][17].

Identical to the case of the first and last frame of a rotation series, all reflections of a ‘still’ diffraction image of a crystal in fixed orientation illuminated with perfectly monochromatic X-rays will be partially recorded (Figure 2.6). The partiality is determined by the thickness of the Ewald sphere and mosaicity of the crystal (leading to a broadening of the Bragg peak, rather than a delta-function). Since X-ray beams are neither perfectly monochromatic, nor perfectly parallel (divergence), the Ewald sphere is represented by an imperfect annulus. Figure 2.6 shows a possible experimental geometry of a still snapshot, where the relative thickness of the Ewald sphere increases with the diffracted angle  $\vartheta$ , due to the divergence of the beam, while the thickness of the Ewald sphere at the origin is determined by the bandwidth. Laue diffraction makes use of this phenomenon by simply not using monochromatic, but polychromatic ‘white’ X-ray beams in order to thicken the Ewald sphere to probe as many full reflections as possible, with the limitation of difficult orientation determination (‘indexing’) and running into the problem of overlapping Bragg reflections on the detector.



## Chapter 3

# Protein crystallography

### 3.1 ‘Conventional’ macromolecular crystallography

Biological macromolecules, which were orders of magnitude larger objects than previously studied molecules, were first analyzed crystallographically in the 1940s. The first atomic resolution structures of biological macromolecules were sperm whale myoglobin and horse heart hemoglobin, experimentally derived and solved by John C. Kendrew [18] and Max Perutz [19], who shared the Nobel prize in chemistry in 1962. After finishing his studies with John D. Bernal (‘JD’) and William L. Bragg on first attempts of protein crystallography of hemoglobin, Perutz continued his work on obtaining the atomic structure of this biomolecule. In total, this project lasted almost 15 years - the limiting factor not being the difficulty to obtain diffraction patterns from hemoglobin crystals. In fact, it was relatively easy to grow well-ordered, large crystals that gave rise to diffraction patterns with very sharp Bragg spots showing regular arrays, while to that point it was impossible to translate these measured intensities into a three-dimensional arrangement of an atomic structure of the molecule, due to the unsolved phase problem for complex molecules (Section 2.4). Diffraction data were sufficient to solve the atomic arrangement of crystals consisting small molecules by the Bragg’s, partly by trial and error approaches, while trying to obtain the atomic arrangement of more complex molecules had failed up to this point.

In isomorphous replacement (Section 2.4), a technique first described by John M. Robertson [20] initially for small organic molecules, Perutz saw a promising method

to solve the phase problem in protein crystallography. By introducing heavy atoms with a larger cross-section than the light atoms of the organic molecule, the crystals scattered X-rays more strongly. At the time, it was discussed whether few heavy atoms would produce a large enough signal within thousands of atoms from the protein. Additionally, it was discussed whether or not the attachment of heavy atoms to the protein would leave the protein intact and not introduce anisomorphism - a problem which is still apparent to date. In 1953 Perutz was finally able to show, that the phase problem could be overcome by comparing the diffracted intensities from crystals of the protein with and without heavy atom derivatives. In his case he used mercury, due to a publication that showed the binding mercury atoms to hemoglobin did not affect its capability of oxygen transfer and therefore the structure was likely to stay intact upon treatment. In parallel, Kendrew, who was working for Perutz for many years, used the same technique of isomorphous replacement in combination with anomalous scattering 'SIRAS' to solve the phase problem in the determination of the structure of myoglobin, which is roughly one quarter in size of hemoglobin, making the analysis overall less tedious and complicated.

The experiments by Kendrew and Perutz marked the birth of structural biology. Protein structures are nowadays still regularly obtained in a similar fashion, by collecting diffraction patterns of macromolecular crystals in several orientations and calculating their electron density. The experimental equipment has become much more sophisticated, including light sources orders of magnitude brighter than achievable back in the 1950s, as well as highly sensitive X-ray detectors, routinely allowing data collection from crystals smaller than 10  $\mu\text{m}$  in diameter. In combination with fast detector technology, it is possible to collect complete datasets in times on the order of milliseconds [21], allowing the structure determination of several hundred crystals on a regular day at one instrument. Another major advance was the introduction of cryocooling crystals for a longer possible exposure time, especially important for radiation sensitive samples in experiments using high doses [22], which allowed the collection of full datasets from individual crystals, rather than merging data from multiple crystals. One of the most significant improvements in this field was the gradual implementation of high-performance scientific computing, allowing automatized data collection, data reduction, phasing, model building and refinement, within a matter of seconds.

## 3.2 Molecular replacement

Molecular replacement (MR), as an alternative phasing approach to experimental phasing described above (Section 2.4), is the most widely used method to obtain initial phases for protein crystallographic data, representing about 70% of all structures deposited in the PDB [23]. Experimental phasing and direct methods are a generally applicable solution of the phase problem in protein crystallography, however its practical implementation is often difficult or time- and labor consuming. The MR method was first proposed by Michael G. Rossmann [24], which makes use of the structural similarity of different (parts of) proteins. Generally this similarity is evaluated on sequence identity between the search model and the target of interest, which should typically be more than 30%. However, a more sophisticated approach is the comparison of predicted secondary structures, since sequence homology does not necessarily lead to a high similarity of structural features. In MR, the search model of the known structure is rotated and translated in the unit cell, until the solution with the best fit between calculated diffraction data from the replaced model and the observed data from the unknown structure is obtained. The rotation step leads to the spatial orientation of the known and unknown molecule, while the translation step superimposes the (theoretically) correctly oriented molecule onto the target structure. There are two approaches to calculate the rotation and translation steps. The first approach for MR solutions makes use of the Patterson function (Section 2.4), which can quantify and evaluate the overlap between two structures. Another method uses maximum-likelihood approaches to find the best fit for the orientation and translation of the molecule. By rotating the PF of the search model and calculating its correlation coefficient to the target structure, it is possible to score every possible solution, depending on the sampling, and finally to pick the best. To accomplish the translation step the molecule is probed in the AU and structure factors  $F_{calc}$  are calculated and compared to the observed set of structure factors by calculating their R-factor where  $k$  is a scale factor to put the intensities on the same level. The lowest R-factor will most likely represent the correct translation vector. As a result one places the atomic model of the known structure in the unit cell of the unknown structure in such a way as to best reproduce the observed structure factors. The initial phases are simply calculated ('borrowed') from the known model in the overlaid orientation and are applied to the measured structure factors. By this it is possible to obtain an initial electron density map of the model which is of reasonable quality, depending on the structural similarity of the two structures, the achieved resolution and quality of the diffraction data and the accuracy of the MR solution. Initially MR has been used as a phasing method for identical proteins

crystallized in different space groups, mutant screenings or multiple ligand-target complexes. Because a large number of protein structures is readily available in the PDB ( $\sim 100,000$ ), the probability of finding a reasonably good starting model for MR is quite high. Even partial search models can be successfully used for phasing with MR, making it a generally very powerful technique to obtain phases for crystallographic data.

### 3.3 Crystallographic metrics and model refinement

The quality of crystallographic data can be extremely diverse. Comparing diffraction data obtained from crystals of a small organic compound to data obtained from a large biomacromolecule can have drastic differences. Even comparing crystals grown within the same crystallization drop, significant difference can be observed, making it important to introduce crystallographic figures of merit in order to be able to compare datasets quantitatively.

One important figure-of-merit for crystallographic data processing is the **completeness** of the dataset, which is the percentage of the number of observed Bragg reflections, compared to the theoretically achievable number of reflections for the crystal at a given resolution and symmetry. Typically, crystallographic data are only accepted with very high completeness of  $> 95\%$ , since lower completeness can introduce artifacts in the final electron density. Another metric along the same line is the **multiplicity or redundancy** of a given resolution shell which indicates the average number of observations of a given reflection  $hkl$ .

An important metric is  $\langle I \rangle / \langle \sigma(I) \rangle$  (or **signal-to-noise ratio** - SNR) which is the ratio of the average intensity of the Bragg peaks to the average standard error of the same intensities. This has been frequently used to estimate the usable resolution limit of a given dataset, since this metric illustrates the accuracy to which data are measured.

Crystallographic data for macromolecular crystals are *underdetermined*. To determine the structure at the atomic level three positional parameters and a single isotropic displacement parameter 'B-factor' are required, which corresponds to 10,000 parameters for a typical protein with 2500 non-hydrogen atoms in the AU in a unit cell of  $100,000 \text{ \AA}^3$ . Given a common space group  $P2_1$  and diffraction to  $2.5 \text{ \AA}$  resolution, in an ideal case, 5,000 reflection intensities might be observed, leading to

a data-to-parameter ratio of 0.5. Given the under-determination, chemical restraints are applied, using prior knowledge, to match a model to the given electron density map, which is commonly referred to as **refinement**.

After the initial phasing step, the model of a protein is usually far from perfect. To improve phases and the interpretation of the electron density map, refinement methods are a very important step in the interpretation of the diffraction data. The basic approach of refinement is the statistical adjustment of atomic coordinates in the model to fit the diffraction data as well as possible. The most frequently used metric in crystallographic data are R factors (Equation 3.1) to quantify the differences between structures or models. The main purpose of crystallographic refinement is the reduction of  $R_{work}$ , which is a figure of merit that was first proposed by Pendry *et al.* [25]. It compares intensities between the observed structure factors and the calculated structure factors from the model. The main purpose of crystallographic refinement is the reduction of the R factor:

$$R_{work} = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|} \quad (3.1)$$

and depending on data quality, most importantly resolution, values for  $R_{work}$  generally range between 15% and 30%.

Prior knowledge of a biomolecule begins with the linear amino acid sequence, also called the primary structure, which is essential in order to build an accurate atomic model. The secondary structure represents the three-dimensional form of local segments of a protein, most commonly defined by alpha-helices and beta-strands, which are characterized by well known parameters, such as 0.15 nm translation, 0.23 nm radius and 0.54 nm pitch of a typical alpha-helix. The secondary structure of an amino acid sequence has a direct influence on the protein backbone, limiting the possible angle between carbon atoms, which is illustrated by the Ramachandran plot [26]. The tertiary structure refers to a protein's geometric shape, which is the overall three-dimensional shape of a single amino acid chain, while the quaternary structure depicts the static interaction of multiple subunits. Both tertiary and quaternary structures are the final goal of protein X-ray crystallography structure determination. In addition to these more global aspects, side-chains of each amino acid can only be present in a limited number of conformations, given geometrical restraints, which are called rotamers, that are also tested and improved during refinement.

Given the underdetermined nature of crystallographic data, the danger of solely using R factors for model validation is over-fitting, which decreases  $R_{work}$  while the actual model is of low quality and presents limited biological information. In 1992,



Axel Bruenger introduced the figure of merit  $R_{free}$ , which is calculated from 5-10% of the reflections which are omitted from the model calculation, allowing to monitor over-fitting. As a rule of thumb,  $R_{work}$  and  $R_{free}$  should not deviate from each other by more than 5%.

Equally valid is the use of the Pearson correlation coefficient (CC) to determine both the quality of a model, as well as an appropriate measure for a realistic resolution estimate, as described by Karplus & Diederichs [27] :

$$CC = \frac{\sum_{hkl} (|F_{obs}|^2 - \overline{|F_{obs}|^2}) \times (|F_{calc}|^2 - \overline{|F_{calc}|^2})}{\sqrt{\sum_{hkl} (|F_{obs}|^2 - \overline{|F_{obs}|^2}) \times (|F_{calc}|^2 - \overline{|F_{calc}|^2})}} \quad (3.2)$$

The resolution of an atomic model represents the resolvability of details in the electron density map. It depends most importantly on the amount of disorder and mosaicity of the crystal, but may also be limited by the wavelength, detector distance, pixel size and detector area used in the experiment. Generally speaking, proteins are flexible molecules and therefore tend to form disordered crystals only allowing low resolution reconstructions, compared to small molecule crystallography (< 200 atoms per UC). In protein crystallography certain ranges of resolutions can be grouped loosely according to features of the map. A range of 0.5 - 1.5 Å is considered ‘atomic’ resolution, since individual atoms are resolvable and all rotamer conformations should be modeled correctly, even though this resolution is rarely encountered. The majority of PDB entries ranges from 1.5 - 2.5 Å (67%) - ‘high’ resolution, where a few errors in rotamer conformations might be present, while water molecules and organic ligands are still clearly visible. Medium resolution ranges from 2.5 - 4 Å (23%), where surface loops might be modeled incorrectly and long side-chains, such as Arg or Lys are likely to have wrong rotamer conformations. Below 4 Å, a map is considered ‘low’ resolution (0.5%), which are mostly used for secondary structure determination, while any sort of refinement is difficult to impossible. <sup>1</sup>

### 3.4 ‘Conventional’ data collection in macromolecular crystallography

Conventional data collection in macromolecular crystallography are based on rotation series, implemented by Arndt & Wonacott in 1977 [13]. The principle behind is to probe subsequent slices of reciprocal space to assemble the full 3D volume with very accurate structure factor estimates since almost all Bragg peaks are ‘fully’

<sup>1</sup>Percentages are calculated from the number of protein structures deposited in the PDB (Date of query: 19.02.2015) which were determined by X-ray crystallography.

recorded (Section 2.7), except peaks on the first and last detector frame. A prerequisite information is the sample position, in order to align the crystal relative to the X-ray beam. Usually the center of the beam is aligned on the center of the crystal to illuminate the largest volume possible. Other scenarios are crystals with distinct shapes, such as needles, where helical scans are necessary to illuminate the entire crystal subvolume.

One of the most stringent criteria for data collection is that one wants to cover all unique reflections for a given space group to the maximum achievable resolution. Hence, to design an ideal experiment, one would want to obtain prior knowledge about the crystal symmetry, unit cell parameters and the resolution limit. Using this prior knowledge it is possible to design a data collection strategy that can maximize the resolution and completeness of the dataset, but obviously requires several isomorphous crystals.

During the data collection, the crystal is rotated with respect to the beam and for each small increment, typically 0.1 to 1 degree, a diffraction pattern is collected. With weak light sources or less advanced detectors, this would mean to rotate the crystal by a certain angle, open the shutter to shine X-rays through the crystal for seconds up to minutes, readout the frame from the detector, erase the detected intensities and start with the next iteration. Nowadays with strong light sources and sophisticated detectors, the crystal is continuously rotated in the X-ray beam in shutter-less mode, while the rotation angle per diffraction pattern is determined by the detector readout.

If the unit cell is large, the number of Bragg peaks per diffraction pattern is large and will increase with the rotational degree of each diffraction pattern. In order to avoid overlapping Bragg peaks, this rotational degree has to be limited so that each spot can be resolved on the image. Depending on the symmetry, the total rotation has to be at least 30 degrees for high symmetry crystal systems (rhombohedral, hexagonal) and sometimes has to be as much as 180 degree for low symmetries (triclinic, monoclinic). Hence, the lower the symmetry, the more data are required.

### **3.5 Radiation damage in X-ray crystallography experiments**

In the recent past, improvements in accelerator and undulator engineering of modern synchrotron sources has allowed the collection of data from very small protein crystals on the order of  $\sim \mu\text{m}$  in diameter. However, the inherent problem of these experiments is deterioration of the crystalline lattice and the chemical composition

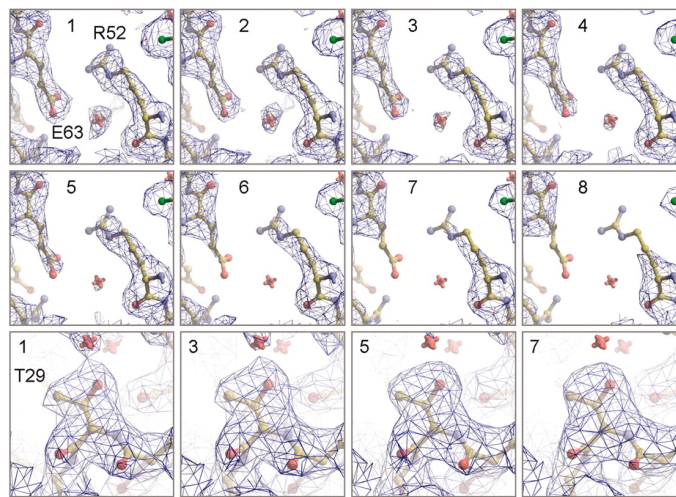


Figure 3.1: Examples of specific radiation damage in X-ray Macromolecular Crystallography. By collecting 8 subsequent datasets on a holo3 crystal, Owen *et al.* were able to experimentally quantify a dose dependent specific radiation damage. The figure shows residues Arg52 and Glu63, where the electron density of the charged side-chains are decaying with increasing applied radiation dose, whereas the electron density of the uncharged side-chain of Thr29 is stable. (Owen *et al.* 2006 [28] - Copyright (2006) National Academy of Sciences, U.S.A.)

of the sample. This effect is also known as radiation induced damage. Radiation damage is generally divided into primary and secondary effects. Primary effects are dose-dependent and are caused by direct interaction of X-ray photons with core shell electrons of atoms. Secondary effects occur due to electrons leaving a vacancy through primary effects, causing electrons from a higher energy level to fill the hole, resulting in a release of energy. Typically this energy is released in form of an emitted photon, but the energy can also be transferred to another electron, which would be ejected from the atom. This effect is called Auger decay and the secondarily emitted electron is called Auger electron. Both primary and secondary effects can be either direct (as in directly altering the protein) or indirect, altering the surrounding solvent. For example, the radiolysis of water upon X-ray photon absorption creates peroxides which are particularly destructive for the protein. In protein crystallography there are two defined classes of radiation damage, specific and general radiation damage. Specific effects are explained by the induction of free electrons propagating along the peptide backbone, providing a mechanism for local, structural damage. These include prominent examples, such as disulphide bond breakage, decarboxylation of aspartates and glutamates (Figure 3.1) and loss of OH groups from tyrosines - usually effecting areas of the protein with higher elec-

tron concentration and large solvent exposed area. In parallel, proteins with metal centers are particularly prone to specific radiation damage, due to the high cross-section of heavy atoms with X-rays. It is known that many PDB entries contain structures with reduced metal centers, which lead to the calculation of unexpected inter-atomic distances and hence interpretation of the nature of these metals, which is particularly problematic in the active center of enzymes. As the second class, global radiation damage leads to more general and therefore more severe problems in data collection. It is usually associated with large changes inside the crystal, leading to a loss of order, up to macroscopically visible damage. These include changes in unit cell dimensions, increase of the Wilson B factor, decreased diffraction power of crystals and loss of high resolution data.

### 3.5.1 The Henderson limit

The SI unit of radiation dose is typically quantified in Grays, which is defined as the deposited/absorbed energy per unit mass:

$$1 \text{ Gy} = 1 \text{ J/kg}; \text{ dose} = \frac{\text{energy}}{\text{mass}}$$

Since the dose quantifies the absorbed energy in the sample, there is a correlation between the dose and radiation damage that the sample has suffered from. In the single atom case, the absorbed dose  $D_{\text{atom}}$  is directly dependent on the probability of photons being absorbed by an electron, given by its cross-section ( $Z$ )  $\sigma_A$ , depending on the incoming intensity of the X-ray beam:

$$D_{\text{atom}} = \frac{N_{\text{photons}} h\nu}{A} \sigma_A \quad (3.3)$$

Given multiple atoms, calculating the deposited energy per unit mass, assuming a sample which is thinner than the penetration depth of a particular photon energy:

$$D = \frac{I_O N_a}{m_a} \sigma_A \quad (3.4)$$

with the incoming intensity  $I_O = \frac{N_a h\nu}{A}$ ,  $N_a$  being the number of atoms and  $m_a$  corresponding to the mass.

The equation for calculating radiation dose can be simplified as:

$$D = \frac{I_O \mu}{\rho},$$

with  $I_O$  as the incoming X-ray intensity per  $\mu\text{m}^2$ ,  $\mu$  as the attenuation length at a

given photon energy and atom cross-section [29] (on the order of 50-250  $\mu\text{m}$  for 5-10 keV) and  $\rho$  the density of protein crystals (roughly 1.35  $\text{g}/\text{cm}^3$ ).

Electron crystallography experimentalists have been the first to observe that the diffracted intensity can drastically decrease during the experiment, which happened to correlate with the applied dose. In fact, half of the intensity of the observed Bragg peaks of 2D protein crystals was fading after 1  $\text{electrons}/\text{\AA}^2$  was applied to the sample at room temperature [30] and 5  $\text{electrons}/\text{\AA}^2$  at  $-173^\circ\text{C}$  [31]. The dose corresponding to 5  $\text{electrons}/\text{\AA}^2$  is 20 MGy, which is the so-called ‘Henderson limit’ [32] that is often referred to in X-ray crystallography of biomolecules. Radiation damage arose as a major difficulty, when synchrotron radiation became routinely available. This problem led to the dogma change of cryo-crystallography [22][32][33][34], where protein crystals are cryo-cooled in liquid nitrogen and are kept at  $-173^\circ\text{C}$  during data collection, which increases the diffraction lifetime of crystals obtained by X-ray radiation tenfold [22]. An experimental determination of the Henderson limit, which is the radiation dose limit of cryo-cooled protein crystals, has been presented by Owen *et al.* 2006 [35] (Figure 3.1). The authors concluded a limit of 43 MGy as a maximum dose for an X-ray crystallography experiment, where the crystals suffered from radiation damage to the point where the scattered intensities had halved. The discrepancy between these two values was explained by Owen *et al.* by the fact that after careful analysis 10  $\text{electrons}/\text{\AA}^2$  are a generally accepted value for electron microscopy experiments, corresponding to about 40 MGy, which was a very good match to the observations by Owen *et al.* A commonly used program for calculating absorbed dose by a crystal is RADDPOSE [36]. Instead of calculating the absorbed dose, RADDPOSE uses the measure of ‘diffraction-dose efficiency’  $I_{DE}$ , which is the diffracted intensity per absorbed dose:

$$I_{DE} \propto \frac{\exp(-\mu_{att}t)}{[1 - \exp(-\mu_{abs}t)]}$$

where  $\mu_{att}$  is the attenuation coefficient of the sample,  $\mu_{abs}$  the X-ray absorption coefficient,  $t$  the path length in the beam and  $\lambda$  the wavelength [36]. This value  $I_{DE}$  can be utilized to optimize the experiment to balance the signal-to-noise ratio during an experiment, while minimizing radiation damage effects.

## 3.6 Protein crystallization

### 3.6.1 Principals of protein crystallization

The first experiments in macromolecular X-ray crystallography were motivated by the observation of macroscopically visible myoglobin crystals, which spontaneously grew in dried pools of a mixture of sea-water and whale blood on ship decks. By being able to grow crystals of myoglobin and hemoglobin in combination with the early experiments on X-ray crystallography by von Laue and the Braggs, it was now for the first time possible to even consider to solve an atomic resolution structure of a biomacromolecule. Besides whale myoglobin crystals on ship decks, there are many described cases of naturally occurring protein crystals (Section 3.6.1.2), while most of protein crystallization is typically carried out *in vitro*, artificially grown, using isolated and purified proteins. The principle idea of protein crystallization is that by adding a precipitating agent to a supersaturated protein solution, the hydrating shell of the protein is being gradually removed, favoring hydrophobic interactions between proteins. Depending on the kinetics of this process the outcome of a protein crystallization experiment can be:

1. No reaction - too low concentration of protein and/or precipitant.
2. Amorphous precipitation - non-specific hydrophobic reactions leading to denaturation of the protein. Typically due to fast dehydration or 'wrong', unfavorable precipitating agent.
3. Crystallization - specific interactions of proteins leading to an ordered array of macromolecules in either 2D or 3D.

These options are based on the assumption that the purified protein itself is conformationally homogeneous and all other involved solutions are highly pure. Conformational heterogeneity is one of the limiting factors of successful crystallization experiments. Biophysical studies of the right buffer conditions can lead to an enhanced stability and homogeneity of a protein in solution. However, there is only a limited effect on the protein by the buffer environment. Particularly membrane proteins tend to be conformationally flexible, which is typically circumvented by introducing mutations to the sequence to either thermally stabilize the protein or favor one particular conformation. The most common approach is to cut a multidomain protein into several combinations of subdomains, trying to find the most

stable fragment, which usually correlates with the highest achievable yields in protein expression. In parallel to that, point mutations can lock the protein in a certain conformation, e.g. a ligand bound state which would otherwise represent a rather loose binding. Both ways of altering the protein are controversially discussed [37], as they artificially induce a rigidity of the protein which is not present under physiologic conditions. By applying other biochemical and biophysical assays it is possible to counteract this argument, showing that e.g. ligand binding kinetics are not altered, however cannot be fully excluded.

Given a homogenous sample, in order to produce a pure protein solution a multitude of options of biochemical purification steps are available, such as affinity tag purification, Ion Exchange Chromatography, etc., but will not be discussed here in detail. Typically, prior to protein crystallization experiments, the final purification step is Size Exclusion Chromatography using an HPLC system, where proteins are diffusing through a polymeric material (dextran, agarose, polyacrylamide etc.) which separates individual protein (complexes) by their molecular weight, or to be more precise, by their Stokes radius. This is particularly useful to obtain homogenous mono/multimeric protein fractions, in case several multimeric states are present in solution. This final step does not have a direct influence on structural homogeneity e.g. due to flexible loop regions in the protein. The actual crystallization process is based on the phase diagram of each crystallization trial (Figure 3.2).

The phase diagram shown in Figure 3.2 is divided into undersaturation and supersaturation phase, as well as metastable, nucleation and precipitation zone. Crystallization can only occur when the crystallization solution has reached the nucleation zone, which has the right balance of concentrations between protein and precipitant to allow the formation of crystalline nuclei. This can only be achieved when the solution is supersaturated, but is still balanced to the extent that the protein does not precipitate, for example by the nucleation process being initiated too rapidly. Once nuclei are grown, the protein concentration in solution will drop gradually because fewer protein molecules are available in solution. By decreasing the protein concentration of the solution, it reaches the metastable zone which is the concentration range that allows crystals to grow. After initial crystals are grown, the summed crystal volume per crystallization solution (few, large crystals vs. many, small crystals) is limited by the solubility of the protein and the concentration of protein in the initial solution. As mentioned above the kinetics of this process strongly influence the outcome of the process. Too high protein or precipitant concentrations lead the solution in the precipitation zone, resulting in amorphous aggregates of proteins.

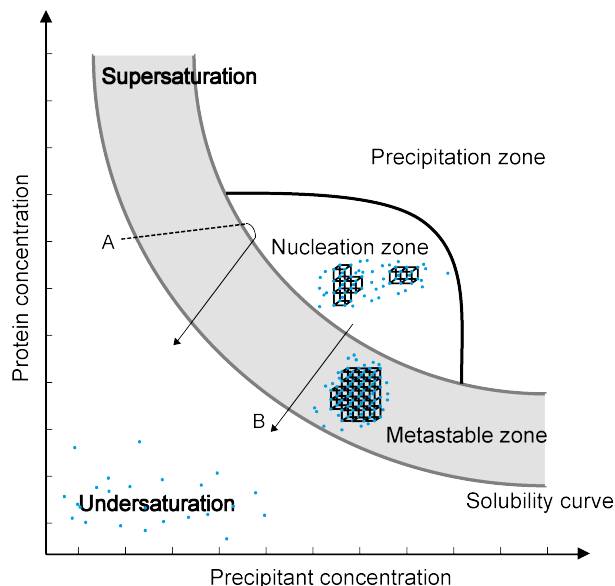


Figure 3.2: Schematic phase diagram, illustrating the crystallization process and its dependence on the equilibrium between protein and precipitant concentration.

In general it is currently impossible to predict the outcome of crystallization trials. Typically a very large number of conditions are screened with high-throughput techniques, usually applying robotic support. Modern pipetting robots can dispense crystallization drops of hundreds of nanoliters in volume and it is easily possible to screen thousands of conditions using only a small quantity of proteins, typically a few  $\mu\text{g}$ . Commercially available screening kits, mostly on 96-well plate basis, can be purchased and especially in the initial phase of a crystallization project thousands of conditions are tested until initial crystallization ‘hits’ are observed. These initial ‘hits’ are then further refined by varying protein concentration, ratio between protein and precipitant solution, precipitant concentrations and additive screens. In the following sections common crystallization principles are described in further detail.

### 3.6.1.1 Vapor diffusion and batch crystallization

Vapor diffusion experiments represent the most commonly used protein crystallization technique, especially for conventional MX experiments, where large crystals are desirable. The principal mechanism is that the protein and precipitant solution are mixed, placed in a sealed well which additionally contains a spatially separated



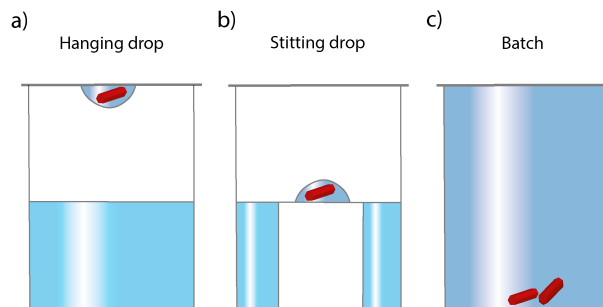


Figure 3.3: Schematic of standard crystallization techniques.

larger drop of the precipitant solution. By the precipitant concentration being diluted in the crystallization drop, the sealed environment results in a vapor pressure, causing the crystallization drop to decrease in volume. This slowly increases both protein and precipitant concentration, until the nucleation zone is reached (Figure 3.2- A) in a successful experiment. The technique can be carried out in ‘hanging’ (Figure 3.3 a) and ‘sitting’ drop (Figure 3.3 b) experiments, where the crystallization drop is either ‘sitting’ on a specially designed part of the well or ‘hanging’ typically from a siliconized cover slip, which seals the well. The choice of method influences the experimental kinetics, so both are usually tested in parallel. However, high-throughput experiments solely use sitting drop approaches due to easier implementation in robotics.

It is crucial to begin with the right initial protein and precipitant concentrations. Otherwise, the start may be in the metastable zone, where no nucleation occurs. This might result in the precipitation zone being reached, which is undesirable. On the other hand, too low concentrations of precipitant and protein might mean the nucleation zone is never reached. As a rule of thumb, the right initial protein concentration is reached when about 30% of all screening drops contain amorphous aggregates. Vapor diffusion experiments are the most prominent and successfully used techniques, because an entire range of conditions within the phase diagram is screened even within one crystallization drop (dotted line; Figure 3.2- A).

Batch crystallization is a kinetically much simpler approach (Figure 3.3 c), but typically requires more prescreening. After simply mixing both protein and precipitant solution, the well is sealed in order to avoid any evaporation. In this context it is necessary to start the crystallization solution in the nucleation zone of the phase diagram to perform a successful crystallization experiment (Figure 3.2- B). This

requirement makes this type of crystallization more difficult but it is often used in serial crystallography experiments, where a larger volume of small sized crystals is needed, which is described in detail later on.

### 3.6.1.2 *In vivo* protein crystallization

In parallel to artificial *in vitro* crystallization, several systems of naturally occurring protein crystals have been described in the literature. The function of these '*in vivo*' grown crystals ranges from storage of a protein and its rapid availability (e.g. insulin), defense against toxic proteins (e.g.  $\delta$ -Endotoxin) to crystalline shells surrounding nucleocapsids to protect themselves from environmental damage (e.g. polyhedrin). Baculoviruses are a class of insect infecting viruses, which have the common feature of one or many nucleocapsid(s) being embedded in a crystalline protein shell, in order to protect themselves from environmental damage. These viruses endure on leaves of plants, which are the corresponding nutrition source of their specific insect host. Hence, it is necessary for the virus to be protected against UV radiation, temperature fluctuations and most importantly dehydration. Polyhedrin crystals have been shown to be extremely stable, withstanding dehydration, freezing and enzymatic degradation [38], mostly due to their extremely low solvent content (~23%) and overall tight packing of the crystal (Section 5.4). As with many viral promoters, the promoter, which is the genetic sequence that initiates transcription, of polyhedrin is extremely efficient and overrules the host cell transcription system upon infection. By this, the entire host cell metabolism is 'abused' to produce more viruses. This very efficient promoter has been successfully exploited for the recombinant production of proteins using insect cell lines e.g. called *Spodoptera frugiperda* (Sf9). In this method, a virus-infection approach is being used to deliver the genetic code for the protein of interest to the host cell, which is then overexpressed in large quantities. Insect cell line expression systems have the advantage of enabling physiologically important glycosylation sites and typically yield much higher protein quantities than achievable with eukaryotic cell line expression systems (Section 7.1.1).

### 3.6.1.3 Lipidic cubic phase

Structural knowledge about membrane proteins is valuable, due to a better understanding of the detailed workings and interaction of these macromolecules, while less than 1% of the more than 100.000 protein structures deposited in the PDB are

derived from membrane proteins, despite their pharmacological relevance (Section 6.1). This mismatch can be explained by the fact that membrane proteins are often structurally flexible and present a large hydrophobic belt, usually embedded in the lipidic bilayer of the cell. These hydrophobic parts have to be shielded from the solvent by a layer of e.g. detergents, which decreases the area for potential crystal contacts and are hence difficult to crystallize. Protein expression yields are generally lower, compared to soluble proteins. The most important reason is that due to hydrophobic residues of the trans-membrane part, it is challenging to preserve these proteins in solution. The handling of membrane proteins needs to be very delicate, but purification steps, such as removing them from their native lipid bilayers by detergent solubilization, are harsh and often cause aggregation or denaturing of the protein. Indeed, most membrane proteins are vulnerable after solubilization, since the aqueous solution represents a different environment compared to their native lipid bilayer and even slight perturbations of their environment or structure can cause denaturation or aggregation.

The most commonly used approach for any biochemical experiment on membrane proteins are surfactant micelle based aqueous solutions, where a detergent solution is added to the cell culture to isolate the protein of interest from the membrane, followed by detergent exchange to stabilize the membrane protein. Detergent exchange requires a screening of multiple detergents, while often not leading to satisfying results. One explanation for their low stability in solution and their high conformational flexibility can be the release of lateral pressure of the membrane on the protein [39], which is usually applied by the cell membrane. Once a purification protocol is established, tedious crystallization trials are necessary and often do not result in crystal formation, either because of a lack of material, conformationally inhomogeneity, or the protein is simply unstable in solution.

An alternative technique to detergent-based crystallization of membrane proteins is the use of lipidic cubic phase as a carrier. Lipidic cubic phase (LCP) is a spontaneously assembling three-dimensional matrix of lipids, dividing the volume into two non-intersecting channel systems (Figure 3.4). Because it mimics the native lipid bilayer environment of a cell membrane, it keeps the lateral pressure on the protein, allowing the membrane protein to retain its physiologic structure and activity.

**Lipidic cubic phase crystallization** A resulting feature of LCP embedded membrane proteins is the possibility to nucleate and facilitate crystal growth [39]. LCP crystallization [39] is an increasingly important and successful technique for mem-

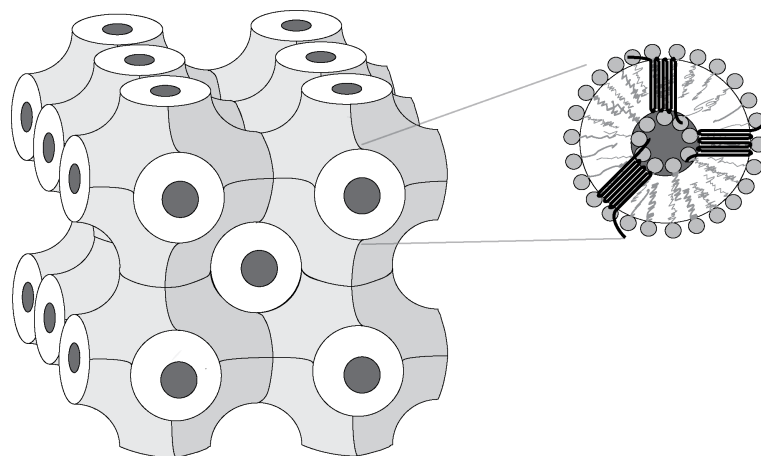


Figure 3.4: Schematic illustration of the bicontinuous lipidic cubic phase composed of water, monoolein and a GPCR. The two compartments of the matrix consist of a curved lipid layer, arranged in a three-dimensional lattice and water channels depicted in dark. The inset (right) illustrates a close-up of the lipid bilayer with embedded membrane proteins.

brane protein crystallization, in particular for the structure determination of G protein-coupled receptors (Section 6.1). Additionally, LCP based crystallization favors type I crystal packing, allowing not only hydrophilic, but hydrophobic interactions between protein molecules, leading to higher symmetry and hence lower solvent content and higher stability of the crystal, in turn resulting in higher diffraction quality. The first successful LCP crystallization trials were demonstrated by Landau and Rosenbusch, which led to macroscopic bacteriorhodopsin crystals - a model system for membrane protein biochemistry. After this initial success, the average number of structures solved from LCP grown crystals was low for several years, but is currently in a phase of exponential growth. The total number of PDB entries from unique membrane protein structures deposited from LCP grown crystals was 257 in 2015, from 81 unique structures (<http://cherezov.usc.edu/structures.htm/>).

There are 18 different commercially available lipids forming LCP, differing in molecular weight, temperature range and curvature. Monoolein (9.9 MAG - 1-(9Z-octadecenoyl)-rac-glycerol) is the most commonly used lipid, with a molecular weight of 356.6 Da. Setting up a crystallization trial is simply done by mixing two parts of protein solution with three parts lipid at 20° C. Assuming no interference effects between the protein/crystallization buffer and the lipid, a spontaneous self-assembly into LCP and simultaneous embedding of the protein should occur. High concen-

trations of detergent, salt, or residual lipids from native membrane, can hinder the LCP formation. Due to geometric restrictions within the LCP matrix, an additional limitation of the LCP crystallization process are membrane proteins of generally large overall size, or large loop regions, which do not fit into the channels formed by the LCP. These size restrictions can be relieved by transforming the LCP into a sponge phase, in order to allow diffusion of larger proteins and enable crystallization [40].

The crystallization process is typically done in batch crystallization (Section 3.6.1.1), where the protein embedded in reconstituted LCP is suspended onto a sandwich-plate - a spacer matrix between two glass plates, typically providing 96 wells for high-throughput screening. One crucial observation was the direct correlation between mobility of the protein embedded in the LCP matrix and its crystallization success rate. A technique developed to monitor and characterize this mobility in LCP is a FRAP assay, established by the Cherezov lab [41]. In this assay, a fluorescent dye, specifically binding proteins, is added to the protein/LCP mixture. Initially, the sample shows homogenous fluorescence throughout the drop and after bleaching of a certain spot in the sandwich plate with an optical laser, the time until the fluorescence is recovered is measured.

After successful crystallization, the first challenges of the LCP technique arise: due to the very high viscosity, handling of the LCP matrix is inconvenient, where embedded crystals have to be scraped off the glass plate before flash-freezing, potentially damaging the crystals. An additional difficulty of this method is that LCP turns opaque upon cryo-cooling, hampering data collection. Automated rastering and centering approaches are necessary for successful crystal detection and data collection [42][43].

## Chapter 4

# Advanced X-ray Sources

The first X-ray experiments were carried out by Roentgen using a so-called ‘Crookes tube’ [44][45] - an electrical discharge tube consisting of a cathode and anode with an applied voltage of 1-100 kV, placed inside a glass container under partial vacuum. He discovered that a fluorescent screen, wrapped in black cardboard to protect it from visible light, showed a faint glow, when placed close to the tube. As hypothesized later, the accelerated electrons are hitting an anode inside the tube, resulting in an energy loss of the electrons, which is partially converted into X-rays. The X-ray flux density of the X-ray tubes used in the very first crystallographic experiments of von Laue and coworkers are estimated around  $10^3$  photons/s/mm<sup>2</sup>. Modern X-ray tubes with a similar geometry can achieve flux densities of  $10^{10}$  photons/s/mm<sup>2</sup>, while the currently strongest, commissioned synchrotrons (PETRAIII/SPring-8) reach  $10^{25}$  photons/s/mm<sup>2</sup> peak brilliance [46]. This tremendous improvement of 22 orders of magnitude in achievable X-ray photon flux has allowed innumerable experiments during the last century, driving our knowledge of matter at atomic resolution by steadily moving the borders of feasibility of experiments. The latest developments in accelerator, undulator and beamline engineering led to the realization of hard X-ray Free-Electron lasers (Section 4.3), which produce femtosecond pulses with peak brilliances of  $10^{33}$  photons/s/mm<sup>2</sup> and unique properties that were previously not achievable. The following chapter will focus on the concept of radiation sources in context of their influence in the field of macromolecular crystallography.

### 4.1 Synchrotron radiation

Macromolecular crystallography experiments are generally performed to overcome the bottleneck of the scattering strength from individual biomolecules. In case of

well-ordered crystals on the order of hundreds of  $\mu\text{m}$  in diameter, it is well possible to use a laboratory source, such as a rotating anode or an X-ray tube, to collect high resolution data. However, particularly in the process of investigating difficult protein targets, resulting crystals are often much smaller. For this reason, sources with more intense X-ray beams are necessary to collect data from these crystals.

The generation of X-rays is realized by the fact that accelerated charged particles emit electromagnetic radiation with an energy dependent on the velocity of the particles. Synchrotron radiation is generated by charged particles, typically electrons, that are accelerated to relativistic velocities along closed orbits of large radii on the orders of dozens to hundreds of meters and are finally deflected into a curved trajectory by applying a magnetic field. The radiated energy is proportional to the particle speed and inversely proportional to the radius of its path. By accelerating to very high velocities, the frequency can reach the X-ray regime by using so-called synchrotron accelerators.

Practically this is achieved by free electrons that are generated by thermal emission from a metal cathode, injected into a linear accelerator for initial acceleration and finally into a circular or ellipsoid synchrotron ring with a circumference of up to several kilometers. The electron bunches are kept on a circular path by bending magnets to their operating energy ranging between 2-6 GeV. Radio-frequency cavities compensate the energy-loss of the electrons due to the interaction with the electromagnetic field of the bending magnet. Finally, X-rays for usage at individual beamlines are generated by tangentially installed magnetic structures, also referred to as ‘insertion devices’. Three types of magnet geometries are used for this purpose: bending magnets, wigglers and undulators (Figure 4.1). By using a bending magnet, the trajectory of an electron bunch is curved, creating a strong radiation fan around the bend. The resulting radiation cone shows a large bandwidth, which is the difference between the highest and lowest frequencies of the X-ray beam, typically larger than 1%. The high intensities and spectral bandwidth of bending magnets are frequently used for experiments that make use of polychromatic light, such as ‘Laue crystallography’, but will not be discussed here in further detail.

Wigglers and undulators are based on the same principle and are periodic arrangements of alternating magnets in order to achieve a harmonic oscillation of the electrons which results in a narrow radiation cone with a narrow bandwidth ( $\sim 0.1\%$ ). The first experimental results on synchrotron radiation have been shown by Elder and coworkers [47], who performed experiments at the General Electric synchrotron accelerator in 1946. The undulator period  $\lambda_u$  directly defines the wavelength of the emitted radiation. Hence, designing an undulator with a spacing between magnets on the order of centimeters, can emit hard X-ray photon energies. The radiation

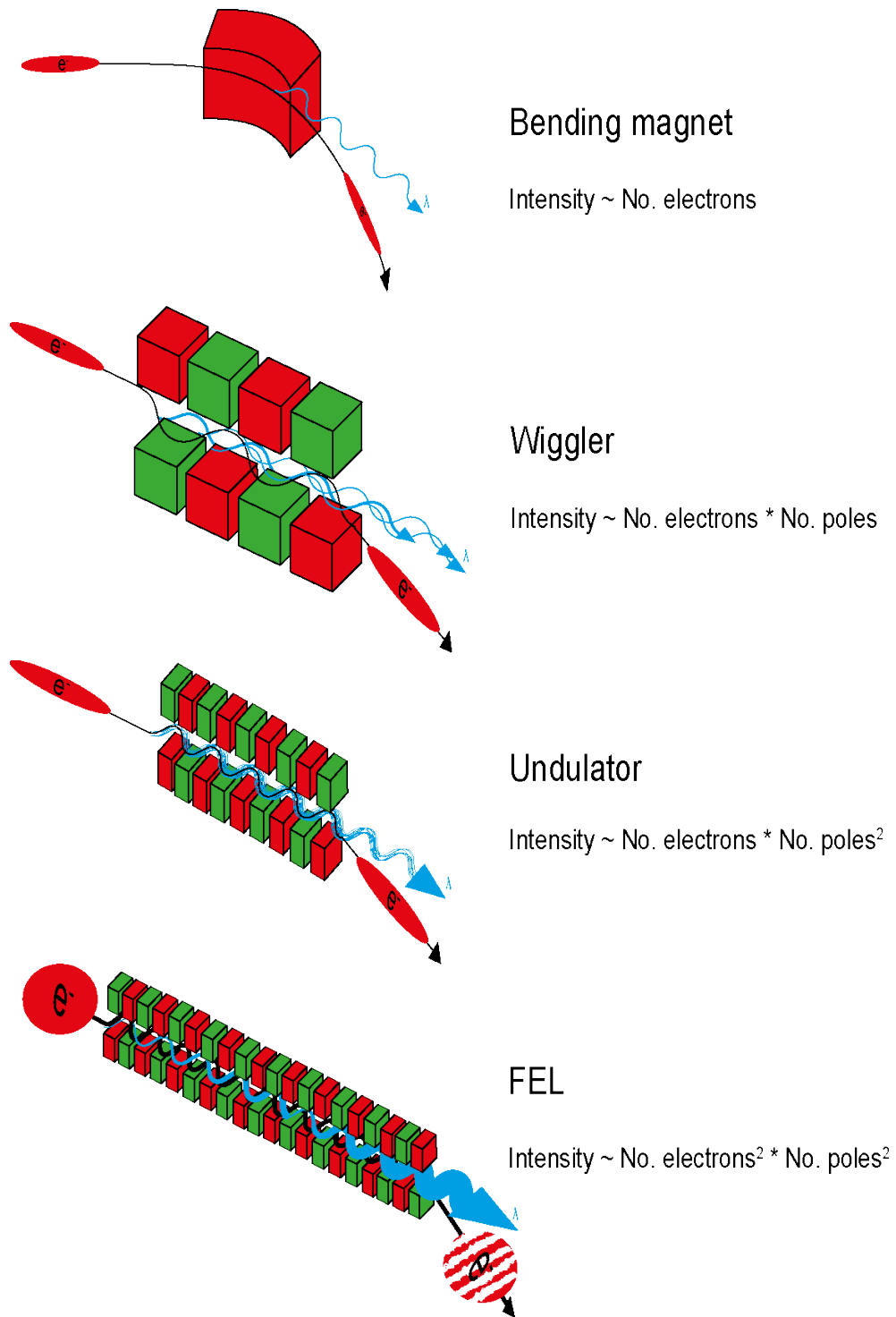


Figure 4.1: Radiation sources and the influence of number of poles and electrons on the radiation power. (red and green) magnetic poles (red ellipse) incoming electron bunch (blue) emitted X-rays with thickness illustrating the intensity gains.



power of the emitted waves is directly proportional to the number of electrons and the number of undulator segments squared. After the generation of radiation from the undulator, a double crystal monochromator, typically silicon based, is installed as a filter to obtain an even narrower bandwidth. The resulting, monochromatized X-ray beam typically has a large focal area of hundreds of  $\mu\text{m}$ , which is further focused to improve the scattering from the sample while lowering background scattering from the surrounding mount and buffer. For MX experiments, typically two perpendicular Kirkpatrick-Baez mirrors [48] are used to focus the beam horizontally and vertically, which can achieve X-ray beams with FWHM of sub-micrometer in diameter.

Synchrotron radiation is currently the most widely used radiation source for protein crystallography experiments. The most important advantages of MX synchrotron beamlines are achievable intensities compared to laboratory sources, overall stability of the X-ray beam, the availability of beamlines worldwide, the possibility to tune the wavelength of the X-ray beam, which is particularly important for experimental phasing, and a high level of automation of data collection allowing high throughput structure determination.

## 4.2 X-ray detectors in ‘conventional’ macromolecular crystallography

For the first decades of X-ray crystallography, diffraction patterns were collected on photosensitive film. They have the advantage of being very sensitive, while throughput is strongly limited since individual exposures have to be developed (semi-)manually. With the development of CCD detectors in the 1990s, the first electronic detectors became broadly available. Due to their high dynamic range, the ratio between minimum detectable to maximum measurable intensity, very small pixel size and rapid read-out, data quality and collection time was drastically improved. In most systems, a thin phosphor screen converts incident X-ray intensities into optically detectable photons, referred to as a ‘scintillator’, which are finally detected by the CCD chip. In case of macromolecular crystallography detectors, the phosphor screen is coupled to the CCD with a tapered optical fiber [49]. The major limitations of this technique are the high electronic background, which leads to the ‘leaking’ of intensities into surrounding pixels that artificially increases the detected Bragg peak sizes and a relatively long readout time limits throughput.

A further development of CCD detectors are so-called silicon hybrid pixel detector systems that are capable of single-photon counting mode. The most popular

implementations is this approach is realized in the PILATUS detector by Dectris [50]. The basic concept is a single pixelated silicon sensor which is bump-bonded to an array of CMOS readout chips. Incident photons are directly transformed into an electric charge in the silicon sensor, which is then detected by the CMOS pixel. Each CMOS pixel contains a pre-amplifier that increases sensitivity. After impact of the X-ray photon, the created charge is compared to a global threshold voltage in a comparator. If this charge exceeds the threshold, a digital signal is produced and collected. In order to improve performance these detectors are made up of smaller modules which can be assembled into large area detectors. The principle of X-ray detection of photon counting detectors is free of dark current and readout noise effects, which, in comparison to CCDs is superior in sensitivity to single photons, reliability and speed of data acquisition due to very short dead-times.

### 4.3 Free-Electron laser radiation

The invention of hard X-ray Free-Electron Lasers (XFELs), which emit X-ray pulses of femtosecond duration with unique coherence properties, had a large influence on the X-ray imaging community. XFELs are capable of producing highly brilliant X-ray pulses of femtosecond duration and full spatial coherence, which has a dramatic effect on the experimental procedure and the characteristics of collected data, which will be discussed in further detail in the next chapter 5. XFELs consist of a linear accelerator, followed by a long undulator structure. FELs have the broadest achievable spectrum of any laser type, ranging from microwaves, through terahertz and infrared, to visible light, UV and X-rays. The principle [51] and first experimental proof of FELs [52] by stimulated emission of Bremsstrahlung have been driven by John Madey.

In order to produce X-ray FEL radiation, electrons are emitted from a copper photo-cathode by an UV laser, which are subsequently accelerated to relativistic velocities in a linear accelerator by radio-frequency (RF) cavities. The first soft X-ray Free-Electron laser in Hamburg (FLASH) reached around 1 GeV, while the first hard X-ray Free-Electron laser, the Linac Coherent Light Source (LCLS) at SLAC, Stanford, currently reaches up to 14 GeV. After the acceleration, the electrons then pass through a very long undulator architecture. In case of synchrotron beamlines, typical undulator lengths are between 2-20 m, while the LCLS undulator is 132 m long [53] and the European X-FEL, which is currently under construction in Hamburg, will have a total length of 652 m. The long time between the principal idea and final realization was partly due to a lack of precision of the undulator design and calibration, where distances of hundreds of meters have to be aligned to

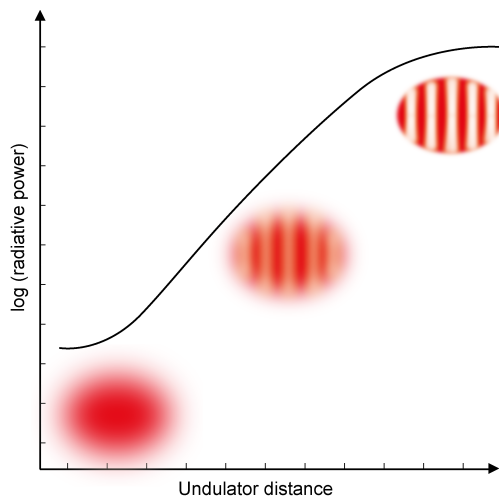


Figure 4.2: Scheme of the SASE microbunching effect achieved at FELs, illustrating the dependence of the undulator distance and the saturation length

micrometer accuracy, which was not achievable until recently [53].

### 4.3.1 Self-Amplified Spontaneous Emission (SASE) effect

The transverse (perpendicular to the propagation direction) acceleration of electrons in an undulator results in a release of photons. However, these X-ray waves are incoherent (not in phase), due to the random distribution of electrons within the electron cloud, which only allows uncorrelated emission of photons. These waves have a higher velocity than the electrons, overtaking the electron beam within one undulator period  $\lambda_u$ . As the radiation becomes sufficiently strong and given the length of the undulator, an interaction between the generated X-rays and the electromagnetic fields of the electrons occurs, called ponderomotive force, resulting in slower electrons being accelerated and faster electrons being decelerated. Because of the relativistic properties, space-charge effects, which would normally lead to electron-electron repulsion, are less dominant. This process is referred to as ‘microbunching’ and results in a periodic modulation of density within the initial electron cloud (Figure 4.2). The microbunching process leads to a positive feedback loop, since individual microbunches with  $\lambda$  spacing, coherently emit X-rays with the same  $\lambda$  wavelength, enabling nearly full transversal coherence [54][55]. Given a sufficiently long undulator and bright electron beam, the radiated power can exponentially grow with the number of poles within the undulator. Eventually, the microbunching process reaches saturation, since only a certain number of electrons can be forced into

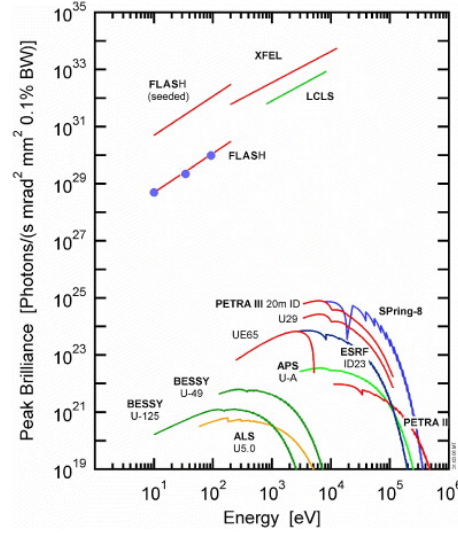


Figure 4.3: Comparison of peak brilliance of advanced X-ray sources worldwide. (from Robinson *et al.* 2010 - Copyright New Journal of Physics) [46]

individual microbunches. The full process of microbunching and reaching saturation is referred to as Self-Amplified Spontaneous Emission (SASE) effect (Figure 4.2).

## 4.4 Linac Coherent Light Source (LCLS)

The Linac Coherent Light Source (LCLS) at the Stanford Linear Accelerator Laboratory (SLAC) is the first hard X-ray Free-Electron laser in the world and was first proposed by Claudio Pellegrini in 1992 [56]. It is based on the last third of the two mile linear accelerator built by SLAC in the early 1960s, which was designed for high energy physics experiments. X-rays produced by the LCLS in SASE mode range from 270 eV to 10 keV, with a pulse duration of 30 - 300 fs and an X-ray beam energy of 1 to 4 mJ with an adjustable repetition rate ‘pulse picker’ of up to 120 Hz and a peak X-ray power of 40 GW. The pulses are fully spatially coherent and have a spectral bandwidth of about 0.2-0.5%, with a wavelength jitter (comparing the peak of the spectrum) of approximately 0.3% on a shot-to-shot bases [53]. The LCLS is operated in two charge modes, where the low charge mode operates at 20-40 pC, resulting in 0.1-0.2 mJ pulse energy and high charge mode operating at 0.25 nC, producing between 1-5 mJ X-ray pulses.

#### 4.4.1 Coherent X-ray Imaging (CXI) beamline

The SFX experiments described in this thesis were carried out at the Coherent X-ray Imaging (CXI) beamline [57], which is a dedicated beamline for Coherent (Bio)-Imaging of the LCLS. Its hardware is designed to provide the maximum flexibility of experiments for hard X-ray diffraction in a vacuum or in-air environment, perfectly suited for serial crystallography or other imaging techniques. Typical experiments, which are described in detail in the following chapter, are carried out at around 8-9 keV, 30 fs pulse duration and an average of 2 mJ, corresponding to about  $10^{12}$  photons per pulse. It is necessary to focus the X-ray beam to a size matching the sample, in order to obtain the best scattering strength of the sample. The CXI endstation is equipped with two Kirkpatrick-Baez (KB) mirror systems, enabling focal sizes of nominal 100 nm and 1  $\mu\text{m}$  FWHM, which simultaneously act as an energy filter.

Due to the specifications and geometry of the KB system, the possible photon energy range is between 2 and 25 keV, while only 10 keV can routinely be achieved by the first harmonic of the undulator [57]. A focal spot of 100 nm is achieved by a very steep convergence angle, leading to a focal length of only few  $\mu\text{m}$ , compared to 1.6 mm focal length of the 1  $\mu\text{m}$  X-ray optics [57]. The two separate KB systems are each connected to a vacuum chamber which is equipped with separate sample injecting systems (Section 5.2.1 and 5.2.2) and a slot for the CSPAD (Section 5.2.3). Recently an in-air/helium setup was designed downstream of the two vacuum chambers, which allows the usage of a marCCD Rayonix detector.

## Chapter 5

# Serial femtosecond crystallography (SFX)

### 5.1 Introduction to SFX

One of the major limitations of bio-imaging experiments, in particular using X-rays and electrons, is radiation induced damage (Section 3.5). Besides sample quality, experiments are often hampered by inelastic scattering and absorption effects, which trigger chemical reactions that will eventually damage the sample. The loss of crystalline order would result in the fading of spot intensities and therefore lose high resolution information. In case of X-ray crystallography, this problem drastically increased with the development of synchrotron radiation sources, achieving higher X-ray intensities than the sample can possibly withstand. By the development of cryo-crystallography, where crystals are flash-frozen and cryo-cooled in liquid nitrogen during the experiment, the sample life-time was extended significantly, but still is inherently limited. Especially in case of very small crystals ( $< 10 \mu\text{m}$ ) or particularly radiation sensitive proteins, e.g. including metal-clusters, the method of conventional X-ray crystallography using one or few crystals is often not practical. An alternative approach, used since the earliest days of macromolecular crystallography, is to distribute the required dose either over many crystals [58], or several parts of one crystal [59]. By this strategy, the volume of each crystal is exposed to fewer photons, finally allowing a more accurate structure determination.

The method of serial femtosecond crystallography (SFX) [60] using X-ray Free-Electron Lasers is taking the multi-crystal data collection scheme to an extreme.

XFELs have unique X-ray beam properties (Section 4.3), having both the shortest pulse duration, typically on the order of tens of femtoseconds, as well as the highest peak brilliance, 2-4 mJ per pulse ( $\sim 10^{12} \frac{\text{photons}}{\text{pulse}}$ ), of any X-ray light source available to date. In the year 2000, Neutze and coworkers [61] simulated the interaction between ultra-short, intense X-ray pulses with biological molecules. The main topic of their study was ionization effects on the sample after impact of the X-ray pulse. The most important result from these simulations was, that a pulse duration of 10 fs should potentially be short enough to obtain a diffraction pattern, while the sample is still intact. Due to the high intensity of the X-ray pulse, the sample is completely vaporized upon impact of the pulse. However, due to the short pulse duration and inertia of the atoms contained in the sample, the X-ray pulse traverses the crystal before movement would be possible that would affect the diffraction pattern. This is because the Coulomb explosion of the sample is rate limited by the acceleration of ionized particles and its atoms. This process is referred to as ‘diffract-before-destroy’, where an interpretable diffraction pattern of nanocrystals, without structural or global radiation damage (Section 3.5), can be obtained.

As a consequence of these high intensities and the following decay of the sample, only one diffraction pattern can be obtained from each individual crystal and it is necessary to integrate Bragg intensities over thousands of diffraction patterns in order to obtain an accurate estimate of structure factors. The experimental geometry of a serial femtosecond crystallography experiment is identical to the first experiments of Max von Laue and coworkers, more than 100 years ago. The X-ray beam is diffracted off a crystalline sample and scattered waves are observed by a detector. Because of the necessity to collect data from many crystals, the sample delivery technique was modified compared to conventional experiments, where single or few crystals are mounted on a fixed substrate. In order to replenish the sample continuously, in order to avoid the collection of more than one diffraction pattern per crystal, a sample delivery technique of using microjets has been adopted (Section 5.2.1 and 5.2.2) for SFX experiments. A thin liquid stream of crystallization solution, containing a sufficient quantity and concentration of protein micro- or nanocrystals, is injected into the interaction region of the XFEL. Beyond the interaction point the remaining sample and debris are collected in a ‘catcher’, which is a container where sample residues can safely be deposited and easily cleaned. The injection device, sample column, catcher and detector are all contained either in a vacuum chamber, or in a helium/air environment.

As a stochastic process following a poissonian distribution, crystals are hit by chance and in random orientations, which has an effect on data quality and data analysis procedures (Section 5.3). All details of the experimental setup, as well as

the data processing step will be discussed in detail later in this chapter.

Due to the special properties of the X-ray pulses and experimental setup, the method of SFX has the following advantages compared to conventional crystallographic experiments:

- Radiation damage ‘free’ data collection (crystal order limited, not damage limited)
- Room temperature data collection (no cryo-protection needed)
- Nanocrystals containing only a few thousand unit cells are sufficient for high resolution structural information
- Time-resolved studies of irreversible reactions with sub-ps time resolution, higher efficiency of reaction initiation due to small crystals
- Crystal handling unnecessary

Methodological limitations are:

- Beamtime applications highly competitive, due to linear geometry and limited availability of XFELs
- Sample consumption currently very high, especially using GDVN
- Limited dynamic range of the detectors, due to usage of integrating detectors
- Possible damage of the crystals during injection

The first successful implementation of a serial crystallography experiment is described in the publication of Chapman *et al.* in Nature 2011 [60]. The corresponding experiment was performed during commissioning of the AMO beamline at LCLS. The major purpose was a proof-of-principle of the possibility to collect interpretable diffraction pattern before the sample is vaporized by the X-ray pulses. PhotosystemI (PSI) nanocrystals were injected into the interaction region and hundreds of thousands of diffraction patterns were collected using pnCCD detectors [62]. The maximum resolution of the PSI structure was estimated to 8 Å resolution, but was limited due to the long wavelength available (2 keV) and the experimental geometry (detector size and distance). Nevertheless, these groundbreaking results illustrated that diffraction-before-destruction and room-temperature measurements of



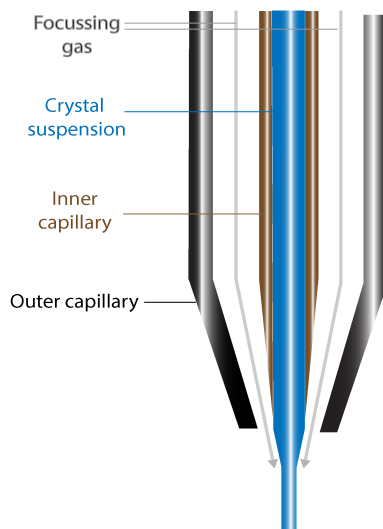


Figure 5.1: Scheme of a Gas Dynamic Virtual Nozzle setup, used in a typical SFX experiment.

highly complex macromolecules are possible using this novel X-ray source. During commissioning of the far experimental hall and the first availability of hard XFEL pulses at the CXI beamline, the limitations of the technique were further investigated, which lead to the first high resolution SFX protein structure of lysozyme [63] at 1.8 Å. Shortly after, followed by the first novel structural information obtained by an XFEL, where a 2.1 Å structure of *in vivo* grown proCathepsinB microcrystals was solved from almost 180 000 single crystal diffraction pattern [64]. In the following sections the experimental setup, as well as details of data handling from these unique experiments will be discussed.

## 5.2 Instrumentation in SFX

### 5.2.1 Gas Dynamic Virtual Nozzle

Due to damage that the sample suffers from upon impact of the X-ray pulse, the crystal containing solution has to be constantly replaced. In case of the LCLS, with a repetition rate of 120 Hz, it is essential to have a high velocity jet in order to avoid hitting each crystal multiple times. Single crystal diffraction patterns can be obtained with a ‘hit’ rate (number of patterns containing a minimum number of Bragg peaks per time unit) proportional to the dimension of the interaction volume (depending on X-ray beam size and jet diameter) and the crystal concentration. The

quality and resolution of each diffraction pattern strongly depends on the sample quality (crystalline order, mosaicity, solvent content etc.), as well as on the level of background scattering, mostly arising from scattered photons by the crystallization buffer. For aqueous solutions, the solvent scattering gives rise to a diffuse ring on the diffraction pattern at approximately  $3 - 3.5 \text{ \AA}$ , representing the average distance between oxygen atoms in liquid water. By reducing the diameter of the jet the background scattering will decrease accordingly. The first implementation of a liquid microjet is the so-called Gas Dynamic Virtual Nozzle (GDVN) [65], developed by DePonte and coworkers. It is a very small device consisting of a fused silica capillary (20 - 100  $\mu\text{m}$  inner diameter) delivering the crystal solution. The inner capillary is inserted into a borosilicate capillary, which is connected to a gas line (typically N or He), streaming coaxially flowing sheath gas which is focusing the liquid stream to 1-4  $\mu\text{m}$  in diameter. The tip of the inner fiber is sharpened by a fine grinder, ideally matching the shape to the aperture of the outer capillary, where a constriction is achieved by melting of the capillary. The tip of the outer capillary is typically conically reduced by grinding, in order to avoid background scattering from the borosilicate interacting with the X-ray beam. GDVN's are pressurized by two independent gas or HPLC lines, typically with applied pressures between 100 - 1000 psi (7 - 70 bar) on each of the lines, resulting in a flow rate between 5-50  $\mu\text{l}/\text{min}$ , on the order of 10-20 m/s. Due to the high density of the crystals compared to their liquid environment, settling of the nano- and microcrystals is an experimental limitation and several devices have been designed to overcome this problem, by constantly rotating the sample containing reservoir [66]. Samples with high concentrations of salts or viscous agents, which are often present in crystallization buffers, cause difficulties during sample injection using the GDVN. These can cause built-ups at the end of the nozzle or inside the capillary, eventually obstructing the flow. Cooling and heating elements are connected to the sample reservoir in order to maintain the temperature used in the delicate crystallization process (typically  $4^\circ$  or  $18^\circ \text{ C}$ ) [66].

### 5.2.2 Lipidic Cubic Phase Injector

The GDVN allows the delivery of aqueous crystallization solutions of low to medium viscosity. Previous experiments, which targeted membrane protein structures involved detergent-based solutions [60] and liquid-like lipidic sponge phase media [67]. It was necessary to adapt the design of the GDVN, in order to stream highly viscous media into the XFEL interaction region, such as using a sample derived from LCP crystallization (Section 3.6.1.3). This was achieved by including hydraulic amplifica-

tion of pressure on the sample, in order to enable pressures high enough to extrude such viscous material. Weierstall and coworkers [68] have successfully designed a Lipidic Cubic Phase Injector, which is compatible with such highly viscous media and can operate at pressures between 2000 and 10000 psi (140 - 700 bar). It consists of a hydraulic stage (amplifying the applied pressure), a sample reservoir of 20-50  $\mu\text{l}$  and a nozzle run by the same coaxial sheath gas as the GDVN. The main differences in design compared to the GDVN are much higher achievable pressures, the sample reservoir being embedded in the injector, instead of being placed upstream of the sample chamber, and the sheath gas only being needed for keeping the extruded LCP column straight, rather than focusing it. The typical inner capillary diameter used in the LCP injector is 50  $\mu\text{m}$ , leading to a similar diameter of the extruded LCP column, finally resulting in higher background scattering compared to the GDVN. In fact, the scattering of the LCP itself is strong enough to damage the detector and in combination with generally higher background scattering, the crystal size for LCP based experiments is typically on the order of 5-10  $\mu\text{m}$  in diameter.

What is intuitively thought to be a disadvantage, the high viscosity leads to the key benefit of the LCP injector, resulting in low jetting velocities and hence very low sample consumption. The velocity of the GDVN is far beyond what would be needed for the repetition rate of 120 Hz and as a consequence only a fraction of the sample is being probed. An average flow rate of the GDVN is roughly 10  $\mu\text{l}/\text{min}$ , with a data collection time of approximately 5-6 hours, it consumes about 10-100 mg of protein, while only 1 out of 10 000 crystals is being probed. Compared to 10 - 20 m/s velocities in the GDVN, the LCP injector can be adjusted to produce a jet velocity on the order of 20 - 30  $\mu\text{m}/\text{s}$ , corresponding to a flow rate of 1-300  $\text{nl}/\text{min}$ , up to 1  $\mu\text{l}/\text{min}$ . The lower limit in flow-rate during an experiment is given by the repetition rate of the LCLS, where each part of the sample has to be replaced before it is hit by another XFEL pulse. Full datasets using the LCP injector were collected from  $\sim 0.3$  mg [69][70], which is usually achievable for any target of interest.

The most commonly used lipid for LCP crystallization is monoolein (9.9 MAG - Section 3.6.1.3). However, this lipid is not suited for the delivery into a vacuum chamber, due to its phase transition at 18°C from cubic phase to lamellar crystalline phase ( $L_c$ ). After entering the vacuum chamber, the jet is immediately cooled and the  $L_c$  phase gives rise to sharp, strong Debye-Scherrer rings, covering a large area, ultimately posing a danger to the detector. Monoolein can be doped with shorter chain MAGs (7.9 or 9.7 MAG), which prevent this phase transition, allowing unperturbed data collection. Additionally, the medium of microcrystals grown in 9.9 MAG can be doped with shorter chain MAGs, which has the same effect, allowing the growth in different lipids and still being able to overcome the phase transition

problem. The possibility to expand the LCP injector technique to use other viscous media [71], such as agarose, in order to decrease sample consumption of soluble protein microcrystals, has successfully been tested at the LCLS [72].

### 5.2.3 Detectors in SFX

A detector represents the heart piece of a beamline. Photon-counting detectors, such as the PILATUS (Dectris), have been implemented with great success in the recent past for synchrotron based experiments. Their capabilities to count single photons with basically no electronic background, a small pixel size, while covering a large area and fast readout (in the ms range), enabling high repetition rates, make them an ideal detector for MX experiments using synchrotron radiation. However, the comparably long counting time does not satisfy the necessities of SFX experiments, since a large number of photons arrive at the detector within a short period of time, while the integration could not keep up and would cause instantaneous saturation of individual pixels. Due to the special requirements for the detector in SFX experiments, Gruner and coworkers have developed the Cornell-SLAC Pixel Array Detector [73], which is an integrating detector based on CMOS technology and will be discussed in the following section.

**The CSPAD detector** A detector for SFX experiments has to withstand a high intensity of diffracted X-rays, integrated in a very short time, while the instantaneous count rate for pixels in an XFEL diffraction pattern can reach up to  $10^{17}$  photons/sec/px, which is the main reason why photon counting detectors are not suitable (Figure 5.2). At the same time, the necessity of a fast readout time allowing a high repetition rate of 120 Hz in the case of the LCLS arises, in order to use the full potential of the source. Finally, the electronic noise of the readout electronics has to be limited, in order to avoid detector artifacts and increased background levels, which is particularly important for single particle experiments. The CSPAD is a 2D detector, consisting of 64 individual ASICs (Application Specific Integrated Circuits) bump-bonded on 32 silicon wafers, with an overall geometry separated into 4 quadrants (Figure 5.4). The relative position of each panel on one quadrant is fixed, each being mounted on a motorized stage, which allows an adjustable center hole size (1 - 10 mm), since conventionally used beam-stops would not withstand the direct beam intensities. Each ASIC is comprised of 185 x 194 bump-bonded pixels, with a pixel size of 110 x 110  $\mu\text{m}$ , giving a total of 1516 x 1516 pixel detector area (including the gaps approximately 18 cm in diameter). The CSPAD is mounted on a motorized stage inside the vacuum chamber, where the detector distance can be

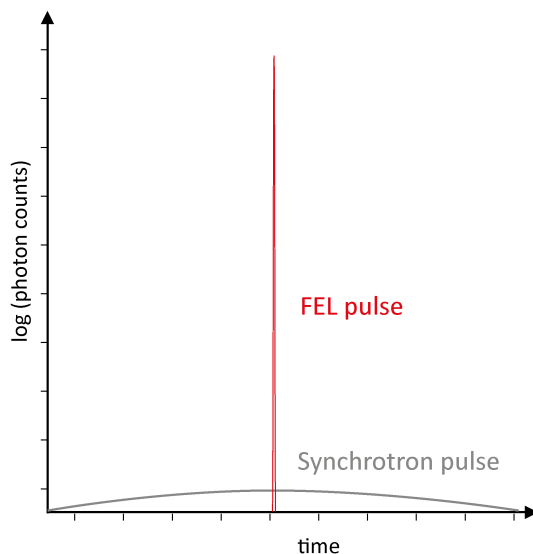


Figure 5.2: Comparison between the photon counts per pixel of a diffraction pattern obtained from an XFEL pulse and a synchrotron source, where pulses of ps duration are typically integrated over milliseconds to seconds. Femtosecond XFEL pulses can produce counts greater than  $10^{17}$  photons/s, which is the reason why photon counting detectors are not suitable for XFEL imaging experiments.

varied between 50 and 500 mm. Given a photon energy range of 6 keV to 10 keV at CXI, it is theoretically possible to collect diffraction data in a resolution range between 200 and 1.2 Å using the main, front detector. In case of the microfocus chamber, a second CSPAD detector can be installed downstream, enabling data collection of very low angle scattering data. The CSPAD has been specifically designed for X-ray scattering experiments with single particles, with a good signal-to-noise ratio and a dynamic range of approximately  $10^3$  ADU. Given the complexity of this detector, in particular the alignment of the 64 ASICs and the adjustable central hole, it is essential to know the exact arrangement of the detector, ideally to sub-pixel accuracy. As described in a following chapter (Section 5.3) there are options to align the detector tiles either semi-manually, or by newly developed algorithms to sub-pixel accuracy in a lab-coordinate system [74].

There is a strong motivation to improve detectors in SFX. The low dynamic range of  $10^3$  is one of the major bottlenecks why most experiments cannot use the full intensity of the XFEL pulse, limiting achievable resolution. Crystal sizes in the micrometer range diffract more photons than the detector could withstand, so current developments, such as the AGIPD (Adaptive Gain Integrating Pixel Detector) [75] at CFEL, as well as the Jungfrau detector at DeCTRIS and Paul-Scherrer Institute

[76] are aiming towards a dynamic range of  $10^4$ , while trying to deal with the special properties of the high repetitions rates of the European XFEL.

### 5.3 Data analysis in SFX

Resulting from the overall setup of the experiments, SFX data analysis schemes significantly differ from the strategy of software packages available for conventional MX experiments. Protein microcrystals are delivered into the interaction region with a microjet (GDVN or LCP) presenting a random orientation of crystals, where the probability  $P_\lambda$  of hitting a crystal with the X-ray beam is following a Poisson distribution:

$$P_\lambda = \frac{\lambda^k}{k!} e^{-\lambda} \quad (5.1)$$

with  $\lambda$  as a real, positive value describing the variance of the distribution and  $k$  as the mean number of observed crystals per diffraction pattern. The ‘hit-rate’ is defined as the number of diffraction patterns per time, since only a fraction of the 120 Hz readouts contain useful information. The average hit-rate depends both on the sample concentration, as well as the X-ray beam profile and size of the focal spot. Assuming a random distribution of crystals in solution, under perfect experimental conditions, an ideal hit-rate would be 63%, since at that point the highest fraction (37%) of single crystal diffraction pattern per shot can be recorded [77]. Practically, this value is rarely achieved, mostly due to relatively large crystal sizes (1-5  $\mu\text{m}$  in diameter), which are limiting the crystal concentration. The average hit-rate during experiments varies between 5 - 10%, while as little as 0.5% have been observed for LCP injector based experiments. Test datasets of lysozyme microcrystals have been collected within minutes with hit-rates of up to 70-80% (data not shown).

#### 5.3.1 Pre-processing of raw SFX data: Cheetah

One CSPAD frame occupies about 6 MB of disk space. Given a repetition rate of 120 Hz (432 000 detector readouts per hour) and a regular beamtime of 5 x 12 hour shifts at the LCLS, the total amount of disk space occupied would be approximately 150 Tb, ignoring possible downtimes due to experimental limitations. Assuming a processing time of only one second for each frame, the total data processing step of those 26 mio. detector frames would take up to 300 days in a serial manner. As mentioned above, on average, only 10% of all detector frames contain diffraction patterns. Thus, by filtering the raw data for potential diffraction patterns, it is

possible to both improve the amount of disk space needed, as well as downstream data processing time by an order of magnitude. For this initial pre-processing step, the software package cheetah has been developed by Barty *et al.* [78].

Cheetah is an open-source, C++ based software suite for processing and filtering raw data obtained from SFX and similar experiments. Raw data at LCLS are stored in XTC format, including detector frames and information about the experiment. All recordable experimental parameters are stored, such as motor positions, X-ray beam information (energy spectra or intensities), multiple detectors e.g. fluorescence screens, timing tool for pump-probe delays, etc. Cheetah extracts needed information from these XTC files for individual ‘hits’ and stores the detector image together with additional experimental parameters in HDF5 file format (Hierarchical Data Format, version 5 - <http://www.hdfgroup.org/HDF5/>).

Cheetah’s general work-flow is to correct for detector artifacts, such as electronic noise, subtracting local background on a pixel to pixel basis and finally to detect Bragg peaks of each detector frame. In a first step, cheetah corrects for a static offset (‘pedestal’) which is estimated by the acquisition of a series of dark (absence of X-rays) frames, which can be explained by electronic noise. Additionally, common mode corrections (uncorrelated shot-to-shot noise) and coupled drifts (offset of the total signal) are accounted for. ‘Bad’ pixels, which constantly give a high or low value due to physical damage of the pixel, are identified and masked for subsequent analysis. The detector correction is followed by the removal of background scattering of the liquid surrounding the crystal. The crystal is surrounded by a volume of crystallization solution which is usually larger than the crystal volume, hence, the resulting diffraction pattern includes a significant portion of background. Local background is estimated by calculating the median value of a box (typically 2-5 pixel radius) around each pixel and subtracting this median from the measured intensity. This background corrected frame is then used for final Bragg peak detection.

Bragg peak detection follows an approach where all pixel values are stored in memory, compared to several thresholds, such as minimum ADU, signal-to-noise ratio etc. in order to identify clusters of connected pixels which belong to the potential Bragg peak. By applying these thresholds it is possible to avoid false positive peaks, such as single ‘hot’ pixels which were not identified in the previous step, or large, diffuse Bragg peaks arising from ice nucleation.

The number and size of these identified Bragg peaks is used to decide, whether or not a detector frame contains a diffraction pattern, which could be useful for the following indexing step. Finally the peak information is passed on to the subsequent analysis steps. The main reason for the approach of using a single threshold, as well as simply looking for connected pixels above a certain background, in a simple algo-

rithm is computational speed. Even though the cheetah software package is highly parallelized, it is difficult to reach the speed of the incoming data. Cheetah is not only used for data processing after an experiment, but also during the experiment, in order to provide online feedback. It is a crucial part of the experiment of being able to optimize different experimental parameters, such as hit-rates by adjusting crystal concentration, analyzing detector saturation to adjust the X-ray beam transmission, or simply deciding which batch of crystals provides the best data.

### 5.3.2 From diffraction patterns to hkl intensities: CrystFEL

The analysis steps of SFX data following the initial pre-processing of the raw images are indexing and integration. H5 files with detector frames identified by cheetah as 'hits', contain the information about the position of each Bragg peak. The extraction of intensity of Bragg peaks, orientation determination of the crystals giving rise to each diffraction pattern and the final merging of intensities is performed by the software suite CrystFEL [79], which is specifically designed for data analysis in serial crystallography. An alternative software package is cctbx.xfel, mostly developed by the group around Sauter *et al.* in Berkeley [80], which will not be further discussed here. Due to certain specificities of SFX experiments, such as individual partial ('still') snapshots of single crystals in random orientations with low signal-to-noise ratios and incoming X-ray intensities fluctuations, conventional software cannot be applied to obtain structure factor estimates.

Indexing, the assignment of Miller indices to Bragg spots in a diffraction pattern, is performed by the tool *indexamajig*, a part of the CrystFEL software suite. This program passes on the location of each Bragg peak to several conventional (FFT-based) indexing algorithms, such as MOSFLM [15][81][82], DirAx [83] and XDS [84]. The next step is to evaluate whether the indexing of a diffraction pattern was successful, is achieved by comparing the obtained parameters from the algorithm with a known unit cell. If the unit cell is not known, it is also possible to obtain initial estimates of the unit cell constants, as in any conventional software package available. After these initial lattice parameters are determined, the second iteration including the cell constant comparison can be performed. In either case the orientation matrix is used to predict Bragg spot locations matching the lattice. The orientation is considered correct, only if a certain percentage of peak locations matches the obtained orientation matrix, which is the second criterion used to evaluate the indexing success. One metric to monitor the success rate of the indexing procedure is the indexing rate, which is defined as the number of indexed patterns over the number of initially identified diffraction hits. The final yield is strongly



dependent on the quality and thresholds of the hitfinding process. Extremely strict hitfinding can allow indexing rates up to 100%, while discarding useful diffraction patterns. On the other hand, extremely loose parameters may yield false positive 'hits', which can not be indexed. A better approach might be to compare the number of indexed patterns over a certain time (e.g. # indexed patterns per minute) in order to monitor both sample and data analysis quality. The number of indexed patterns is typically lower in case of very large unit cells (small spacing of Bragg peaks on the detector), low resolution diffraction patterns, or a large number of false positive 'hits'.

### 5.3.2.1 Indexing ambiguities

A current limitation of indexing algorithms are indexing ambiguities, where the space group is of lower symmetry than its Bravais lattice. For example within the point group 4 (Laue group 4/m) the mirror plane  $m$  acts as a twinning operator. Rotation of a crystal structure (often by  $180^\circ$ ) might result in an overlay of the initial lattice with the rotated lattice, while the rotated molecule would be in a different orientation. In this case the transformation belongs to the lattice symmetry, but not of the underlying structure itself. In other words, crystals with a point group showing an indexing ambiguity might produce two diffraction patterns with identical Bragg peak positions, while the intensities differ because of a difference in the underlying structure. Currently available indexing algorithms solely rely on the geometry of Bragg peak positions and calculate distances thereof. Hence, the algorithm can only identify the lattice orientation, but it is not possible to distinguish the underlying structure, which would result in changes in Bragg peak intensities. Previously this has been a limitation in SFX data analysis, since 38 out of 65 possible space groups in protein crystallography belong to merohedral point groups - one out of six crystal structures in the PDB [85]. By not resolving this ambiguity the final dataset will artificially be of higher symmetry, because of an additional symmetry axis despite the real point group (Figure 5.5). Brehm & Diederichs [85] have proposed and implemented an algorithm, which can break the indexing ambiguity by applying prior knowledge. The algorithm can overcome the ambiguity by calculating average Pearson correlation coefficients between a single diffraction pattern and all diffraction patterns in the dataset, as well as the transpose of every diffraction pattern applying the apparent twin law. This twin law is related to the additional symmetry axis gained. By this approach it is possible to assign all diffraction patterns to subgroups of certain orientation, transpose all of them according to their twin law and finally merging them in their real Laue group. This

approach is implemented in the CrystFEL package *ambigator* and uses a similar algorithm.

### 5.3.2.2 Integration

The final output of the program *indexamajig*, is a ‘stream’ file, which is a text file containing information about the filename, position of found peaks, saturated peaks, orientation matrix, unit cell parameters, reflection list of integrated intensities from predicted peak positions, etc. These initially integrated intensities are obtained from single Bragg reflections on individual snapshots. Due to variations in crystal size and shape, as well as fluctuations of the X-ray beam profile and inherent limitations, such as partiality, CrystFEL is using a Monte Carlo approach [86][87] for merging data into the final hkl intensities, as well as to estimate the errors of the measurements. This is achieved by taking the average of a large quantity of individually integrated hkl intensities of symmetrically unique reflections from different crystals. In contrast to conventional crystallography, where goniometer based data collection schemes are applied, Bragg peaks are only recorded from ‘still’ images with a certain orientation. The Monte Carlo integration method is applied in order to allow 3D sampling of hkl intensities of these ‘still’ images. To counteract the partiality problem (Section 2.7) of Bragg peaks on ‘still’ diffraction pattern, it may be possible to increase the bandwidth by changing the properties of the electron beam in the accelerator (‘chirping’), or to increase the divergence angle of the focusing mirrors. For  $m$  crystallites, the Monte Carlo integrated intensity of a Bragg reflection  $hkl$  of wave vector transfer  $\Delta k_j$  can be calculated as follows:

$$I_{hkl}^{exp}(m) = \sum_{n=1}^m \sum_{\{j\}_{m,hkl}} I'_n(\Delta k_j) \quad (5.2)$$

with  $\{j\}_{m,hkl}$  as a set of pixels in a diffraction pattern  $m$ , which contributes to the integrated Bragg peak intensity.  $I'_n(\Delta k_j)$  is the diffracted intensity after background subtraction and polarization correction on the  $n^{th}$  diffraction pattern:

$$I'_n(\Delta k_j) = \frac{I_n(\Delta k_j) - I_{bg}(\Delta k_j)}{P(k_{o,j})\Delta\Omega_j} \quad (5.3)$$

, where  $P(k_{o,j})$  describes the polarization factor of an incoming wave  $k_o$  and  $\Delta\Omega$  is the solid angle between the incoming wave vector and the pixel position. The final Monte Carlo integrated intensity represents the average of all integrated intensities from a particular Bragg peak hkl and its symmetry equivalent reflections. Kirian *et al.* [88] have studied the convergence behaviour of the Monte Carlo integration

method with varying integration volume, showing an optimum between integration domain and the crystallite size. The initial Monte Carlo integration method [86] was using the pixel-wise averaged intensity, corrected for the solid angle  $\Delta\Omega$ . The current implementation of this Monte Carlo algorithm in CrystFEL is taking the average total peak intensity as the integrated intensity without solid angle correction, which is less dependent on the size of the integration volume [79]. Additionally, the Monte Carlo algorithm implemented in CrystFEL uses a peak prediction based on geometrical calculations.

The integration of each Bragg peak is performed by concentric circles around the centroid of the predicted Bragg peak position, with specified radii for: integration and outer and inner radius of pixels surrounding the peak for a background estimate. The radii of the latter have to be optimized for every sample in every experiment, since it strongly depends on crystal size, quality (such as mosaicity), detector distance, etc. Usually pixel radii of 3 (integration), 4 (inner), 5 (outer) have been used as a default. It is crucial to set the size of the integration area radius large enough to not miss detectable intensity, as well as to allow for possible errors in the centroid determination, while it is undesirable to integrate background together with signal. Positions of predicted peaks are not included in any background estimate.

Estimating the errors  $\sigma$  of a Bragg peak integrated on an individual snapshot diffraction pattern, is crucial to estimate general data quality. It is calculated by the variance of the background ( $\sigma_{bg}$ ) within the integration region, normalized by the number of photons in the peak ( $N$ ) and :

$$\sigma_{total}^2 = \sigma_{Poisson}^2 + N\sigma_{bg}^2 \quad (5.4)$$

with  $\sigma_{Poisson} = k \sum I$  with  $k \sum I$  representing the integrated intensity of a Bragg peak measured in arbitrary units (ADU), where  $k$  is the conversion factor of an incident photon into detector units at a given photon energy ('photon per ADU').

### 5.3.2.3 Merging and data quality metrics in SFX

Due to the above mentioned differences between SFX and conventional MX data collection and assembly, additional figures of merit were introduced in order to evaluate SFX data quality and consistency. The Monte Carlo integration approach relies on probing a broad variety of intensities from each symmetry unique reflection, since 'full' reflections are not available. Individual measurements of the same reflection might not be of similar intensity. In fact, it is not desirable to achieve a high similarity within individual measurements, since the main aim is to fully probe the underlying distribution.  $R_{merge}$  is a conventional data metric that is used to

indicate the similarity of each dataset merged into the final dataset and is given by:

$$R_{merge} = \frac{\sum_{hkl} \sum_{j=1}^N |I_{hkl}(j) - I_{hkl}|}{\sum_{hkl} \sum_{j=1}^N I_{hkl}(j)} \quad (5.5)$$

or simplified as:

$$R_{merge} \approx \frac{\sum(|I - \langle I \rangle|)}{\sum |I|} \quad (5.6)$$

where  $R_{merge}$  represents the weighted difference between the average intensity and individual Bragg reflections. It typically used for MX datasets merged from multiple crystals, where each dataset contributing to the final intensities is varying completeness and redundancy. Hence,  $R_{merge}$  strongly depends on the distribution of intensities from each subset and the higher the variance of measured intensities of each hkl reflection, the higher the final  $R_{merge}$  value. Therefore, it is not a meaningful figure or merit for SFX data analysis, because of the inherently high variance and multiplicity. The SFX community has generally followed an approach where the consistency of a dataset is evaluated by splitting the initial dataset into two subsets, either randomly or even and odd-numbered patterns. White *et al.* [79] have introduced the metric of  $R_{split}$  where the two half datasets are compared:

$$R_{split} = \sum_{hkl} \frac{1}{\sqrt{2}} \frac{\sum |I_{even} - I_{odd}|}{\frac{1}{2} \sum (I_{even} + I_{odd})} \quad (5.7)$$

, where  $I_{even}$  is the intensity of an hkl reflection merged from the even-numbered patterns, while  $I_{odd}$  is the merged intensity from the equivalent reflection of the odd-numbered patterns. A very similar approach of comparing two half datasets has been introduced by Karplus & Diederichs [27]:

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}} \quad (5.8)$$

where  $CC_{1/2}$  is the Pearson correlation coefficient between two random halves of a dataset:

$$CC_{1/2} = \frac{\sum_{hkl} (|F_{even}|^2 - \langle F_{even} \rangle^2) \times (|F_{odd}|^2 - \langle F_{odd} \rangle^2)}{\sqrt{\sum_{hkl} (|F_{even}|^2 - \langle F_{even} \rangle^2) \times (|F_{odd}|^2 - \langle F_{odd} \rangle^2)}} \quad (5.9)$$

which is comparing the correlation between the two subsets and is now frequently used in crystallographic data analysis to determine the final resolution cutoff of a dataset [27]. The usage of correlation coefficients to estimate the resolution cutoff

of a dataset has been extensively and successfully used in the field of cryo-EM [89].

Conservatively, the resolution cutoff of a dataset has been determined by the  $\frac{I}{\sigma(I)}$  in the highest resolution shell. As a general rule of thumb the  $\frac{I}{\sigma(I)}$  cutoff was chosen between 3-4 to obtain “reliable” data. This rule is still a matter of discussion, since choosing a lower resolution cutoff, than the data would theoretically allow, artificially drops R-factors to low values. Due to the nature of data collection in SFX, all measured reflections are partials and therefore it is not possible to strictly calculate  $\frac{I}{\sigma(I)}$  as it is done in traditional crystallography, where all measured reflections are full or could be easily extrapolated. After the initial integration of individual 2D diffraction pattern, the Monte Carlo approach is used to obtain accurate structure factor estimates. The ‘signal-to-noise ratio’ (SNR) is calculated to estimate the error in the final structure factor intensity estimate of each reflection and is given by:

$$\sigma_{hkl} = \left[ \sum (I_{spot} - \langle I_{hkl} \rangle)^2 \right]^{1/2} / N_{hkl}$$

where  $I_{spot}$  is an individual intensity measurement of a reflection hkl,  $\langle I_{hkl} \rangle$  the mean of all measurements of that reflection hkl and  $N_{hkl}$  the number of observations of that reflection hkl. This ‘post-merging’ SNR value cannot be compared to conventional  $\frac{I}{\sigma(I)}$  value.

This SNR value estimates the accuracy of each integrated hkl intensity and essentially reflects the spread of individually observed intensities of one hkl reflection, accounting for varying multiplicity of measurement. Due to the nature of SFX experiments, namely shot-to-shot intensity differences of the XFEL beam, crystal size variability and hitting crystals at random positions of the Gaussian beam profile, there is a broad spread of observed intensities. While SNR values calculated this way appear low in high resolution shells when compared with  $\frac{I}{\sigma(I)}$  values reported in traditional crystallography, all other figures of merit usually indicate a significant signal in the highest resolution shells.

#### 5.3.2.4 Differences of SFX data and conventional crystallographic data

As mentioned earlier, the data obtained from SFX experiments are in principle comparable to conventional crystallographic data. After Monte Carlo integration, a list of hkl intensities is fed into conventional software for phasing or model refinement. However significant differences in the experimental setup and X-ray source are perceptible in the final data.

To summarize, there are three key differences between SFX and MX.

1. **Crystals can not be manipulated** during the experiment. Because of the sample delivery using liquid jets, current technology does not allow individual crystals to be mounted or oriented during the experiment. Hence, probing the crystals is a stochastic process following a Poisson distribution, resulting in not every detector frame containing a diffraction pattern. Additionally, detector frames might include diffraction pattern from multiple crystals, which can complicate the indexing and integration step, e.g. because of possible overlaps in Bragg spot positions.
2. Because of the very high peak brilliance, only **one diffraction pattern** can be collected **per crystal**. Most experiments are currently limited by the dynamic range of the detector and potential damage of the detector by strong Bragg peak intensities. But in case of using the unattenuated beam, the sample will be turned into a plasma during the exposure. A consequence of this limitation is the necessity to average over thousands to hundreds of thousands of diffraction patterns in order to accurately determine structure factor amplitudes.
3. Each diffraction pattern only contains **partially integrated Bragg reflections** resulting from single X-ray pulses with no rotation of the sample during the exposure. This affects data quality and data processing approaches, since many analysis steps have to be performed on incomplete, individual 'still' snapshots, rather than more complete datasets obtained from standard MX experiments.
4. Due to the high repetition rate of XFELs, a **drastically increased data volume** leads to the possibility to average diffraction patterns with the necessity to develop sophisticated, parallelized algorithms to cope with terabytes of data in a reasonable time (Section 5.5.4).

## 5.4 High resolution structure from the smallest protein crystals

### 5.4.1 The structure of granulovirus (GV) polyhedrin

Cypoviruses and baculoviruses are families of viruses that parasitize a variety of insect species. All members of these virus classes have the common feature of the virion being embedded in a robust crystalline shell (polyhedra), which is typically a few  $\mu\text{m}$  in diameter ( $10^{-3}$  to  $10^3 \mu\text{m}^3$ ) and which can either host one or several ( $\sim 5 - 15$ ) nucleocapsids. Given their nature of infecting insects these viruses have to

persist on leaves or fruits and hence have to be able to survive harsh environmental conditions, such as exposure to UV-radiation, dehydration etc. Polyhedra insect viruses are of economical importance, because of their wide usage as biological pesticides [90] since the early 1990s, but also remain a significant problem for the silk industry [91]. Structures of polyhedrin from larger microcrystals, some of which were recombinantly expressed and grown *in vivo* (Section 3.6.1.2) and hence lacking the virus, were solved using modern synchrotron microfocus beamlines [38][92][93][94]. Of those, one structure was solved *in cellulo*, without the need for crystal purification from the cells [93]. The most recent structure of CPV17 was solved with the technique of SFX [94]. These insect viruses are estimated to be genetically more than 100 mio. years old [95], while the large variety of the origins of these viruses, and hence their amino-acid sequence, converging to a similar lifecycle, is representing classical analogous evolution. Another remarkable feature of these viruses is that all polyhedrin structures solved to date show cubic I 2 3 symmetry, have similar unit cell constants ( $\sim 105 \text{ \AA}$ ), very similar overall shape and extremely low solvent content ( $\sim 20\%$ ), despite the low homology in amino acid sequence ( $\sim 50\%$ ) [96]. The reason for the low solvent content is the extensive intermolecular network, of which more than 70% of the protein surface are shielded. Generally, the outstanding stability of these protein crystals, surviving dehydration, freezing and enzymatic degradation [97], concentrated urea, acid and detergent treatment [38], indicates that this crystalline lattice, as well as covalent and non-covalent interactions between the polyhedrin (28-29 kDa) trimers, allow the virus to survive under such harsh conditions. This stability allows a lifecycle of the virus by oral-faecal infections and resisting solubilization until exposed to the alkaline pH of the insect's midgut [97]. Given their extremely stable phenotype, small size, easy and cheap production and the fact that polyhedra are naturally binding small molecules, such as nucleotides or PEG, polyhedra insect viruses were proposed to be used as a framework for nano-technology applications [38][92].

*Cydia pomonella granulovirus* (CpGV), species of the genus *betabaculoviridae* (family: *Baculoviridae*), has a circular, double-stranded DNA genome, specifically infects *Cydia pomonella* ('codling moth') and is mostly used as a pesticide for apple and pear trees (Carpovirusine®) [98]. Each particle contains only one cylindrical nucleocapsid (30–50 nm diameter, 200–300 nm long) is surrounded by a crystalline shell where only 60% of the particle consists of crystalline polyhedra.

### 5.4.2 Sample production and characterization

*Cydia pomonella* granulosis virus (CpGV) (family: *Baculoviridae*), is a double-stranded DNA virus with 123,500 base pairs with 143 ORFs (Open Reading Frames, individual fragments of genetic code that can potentially encode a protein). Single nucleocapsids are embedded in the crystalline shell of polyhedrin, while the crystalline fraction of the entire particle is about 60%, corresponding to an average number of 10,000 unit cells per virus particle. The host organisms of CpGV are invertebrates, of which the main host is *Cydia pomonella* (Codling moth), which is an economically relevant parasite, affecting apple and pear trees. The caterpillars of *C. pomonella* burrow into fruits, which diminishes the economic value of the harvest. At the same time, there are no adverse effects of CpGV on humans and therefore, CpGV is being used as a biological ('organic') pesticide, although resistances against CpGV have been observed amongst those insects.

Virus production has been described by Eberle *et al.* [99] and performed by the Jehle group at the Julius-Kuehn-Institute in Darmstadt. *C. pomonella* neonate codling moth larvae were being fed for 10 days on virus-free artificial diet, reaching the instar L4-L5 stage. A virus suspension was added to their regular diet, while keeping individual larvae in separate wells and incubated at 26°C until the infection was visible (5-6 days, discoloration of the larvae, slow-moving etc.). Dead insects were homogenized using 0.5% SDS detergent solution to allow tissue and cell rupture and release of the occlusion bodies (OBs), followed by a filtering step to discard macroscopic contaminations. Several washing steps of centrifuging the sample at 18,000 g for 30 minutes were applied to remove microscopic contaminations, such as fatty acids, bacteria etc. The final purification step was a density gradient using glycerol, where OBs are concentrated in the 60% glycerol layer, followed by final washing steps with deionized water and stored in a freezer (-18°C or colder).

In order to monitor the diffraction quality, the pure CpGV sample was analyzed prior to the SFX experiment by dynamic light scattering, scanning electron microscopy (SEM) and powder diffraction. To allow an estimate of purity of the sample, particle concentration, average size distribution and more detailed features, high-resolution cryo-EM images of the CpGV particles were obtained using a Titan Krios microscope. All cryo-EM experiments were performed by Ken Goldie, in the Stahlberg lab at the Center for Cellular Imaging and NanoAnalytics in Basel. In order to estimate the average size distribution, low magnification cryoEM images were obtained from multiple CpGV batches (Figure 5.3 a). By using an automated image processing algorithm, boundaries were estimated and particle sizes were cal-



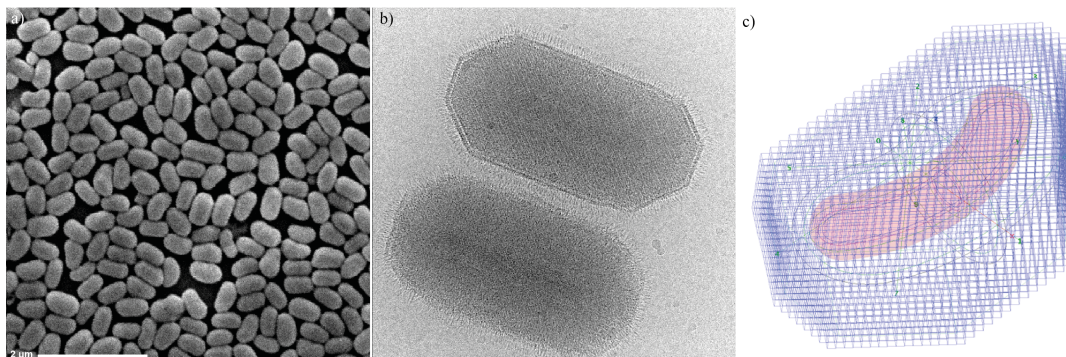


Figure 5.3: Electron microscopy images of granulovirus particles (a) SEM image showing a very homogenous size distribution of CpGV particles with only few outliers (b) high magnification cryo-EM image, showing facets and internal structure of the crystalline shell, its surrounding protein layer ('spikes'), as well as the structure of the embedded nucleocapsid. (c) Model combining all obtained information about the virus particle, including the cubic crystalline lattice and nucleocapsid. Courtesy of K. Goldie (Univ. Basel, CH) and P. Metcalf (Univ. Auckland, NZ) (from Gati *et al. in preparation*)

culated. From high magnification images of individual CpGV particles, the average fraction of crystalline material of the total volume was estimated to be 45% (Figure 5.3 b). Low resolution images of several hundred particles were analyzed, which led to an average particle size of  $210 \pm 20 \times 210 \pm 20 \times 400 \pm 30 \text{ nm}^3$ , corresponding to approximately 10 000 unit cells, or 240 000 protein molecules per virus particle. A final concentration of  $10^{13} \text{ particles/ml}$  was adjusted by centrifugation, which was the highest concentration achievable in order to allow optimal hit rates during data collection.

### 5.4.3 Data collection and processing

Serial Femtosecond Crystallography experiments with CpGV OBs were performed at the LCLS (beamtime ID: L767) at the Stanford National Accelerator Facility (SLAC - Menlo Park, USA). The LCLS X-ray beam with a pulse duration of 50 fs was focused to a beam size of  $1.3 \times 1.3 \text{ }\mu\text{m}$  FWHM at an average pulse energy of 2.7 mJ, a photon energy of 7.9 keV and a repetition rate of 120 Hz. CpGV particles with an average size of  $210 \times 210 \times 400 \text{ nm}$  were injected into the X-ray interaction region by a Gas Dynamic Virtual Nozzle system (GDVN - Section 5.2.1) [65], where the crystal suspension was focused to 3-4  $\mu\text{m}$  diameter with an average flow rate of 10 to 15  $\mu\text{l/min}$ . Individual 'hits' (diffraction pattern containing more than 20 Bragg peaks with a signal-to-noise larger than 6) were identified using

#### 5.4. HIGH RESOLUTION STRUCTURE FROM THE SMALLEST PROTEIN CRYSTALS67

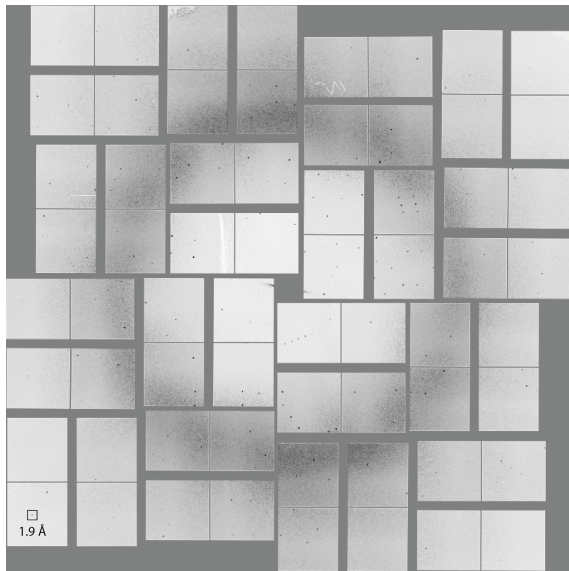


Figure 5.4: The strongest diffraction pattern in the dataset of polyhedrin from CpGV OBs collected by SFX. The corner of the detector corresponds to  $< 1.9 \text{ \AA}$  resolution, indicating that theoretically a higher resolution limit could have been achieved. (from Gati *et al. in preparation*)

cheetah [78]. Hit rates varied between 10 and 60%, with a total of 487085 detector frames recorded. Diffraction data were collected on a CSPAD [73] (Section 5.2.3), set to a total detector distance of 130 mm, which corresponds to a maximum of  $1.8 \text{ \AA}$  in the corners of the detector. The detector geometry was chosen since powder diffraction experiments as well as previous SFX experiments, suggested a resolution estimate of  $3 \text{ \AA}$ . However, in hindsight, the applied detector distance limited the resolution of the final dataset (Figure 5.4).

After the identification of diffraction hits and the following conversion of individual diffraction pattern to the HDF5 format by cheetah (Section 5.3.1), the software suite CrystFEL (version 0.5.2+93d0472) was used for crystallographic data analysis. The peak location information was used for the indexing of individual, randomly oriented crystal diffraction pattern using several indexing algorithms. The integration step was performed using the in-built Monte-Carlo algorithm (Section 5.3.2) [86][87] to determine accurate structure factors from ten-thousands of measured Bragg peak intensities. The final set of indexed patterns, containing 82603 frames, was merged into a dataset (Overall  $CC^* = 0.9994$ ;  $2 \text{ \AA}$  cutoff) for further analysis (unit cell:  $a = 103.3 \text{ \AA}$ ;  $\alpha = 90$ ).

The space group of the polyhedrin crystal of CpGV is  $I 2 3$ . This space group

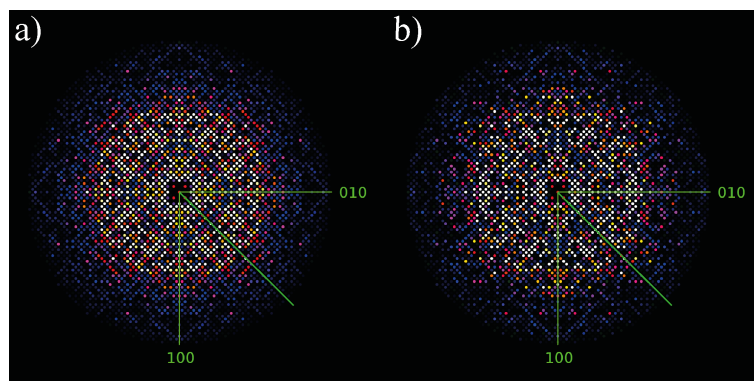


Figure 5.5: Results of the ‘detwinning’ process of the CpGV dataset by CrystFEL, showing  $hk0$ -type reflections plotted with a color scale. Green lines indicate the axis  $h00$  and  $0k0$ . a) before running ‘ambigator’ with an additional, artificial symmetry axis b) ‘ambigator’ output, where the artificial symmetry axis is clearly broken. (from Gati *et al.* *in preparation*)

presents an indexing ambiguity (Section 5.3.2.1), due to existing choice of two crystal orientations ( $180^\circ$  rotation) that give rise to identical spot locations of Bragg peaks on the diffraction pattern. The ambiguity can be overcome by examining the relative intensities of peaks, which is complicated by the degree of partiality of peaks and experimental background, visible in a pattern. As described in section 5.3.2.1, Brehm & Diederichs [85] devised an algorithm that overcomes this problem by comparing the average Pearson correlation between a single diffraction pattern and every other in the dataset in order to decide whether the chosen assignment is correct, which allowed to ‘detwin’ this dataset. Of the initial 82 603 frames, 42 143 assigned indexing results were corrected according to the suggestion of the average Pearson correlation coefficients (Figure 5.5).

#### 5.4.4 Structure determination and comparison to homologous structures

Molecular replacement was performed using PHASER MR [100] using a search model of a homologous polyhedrin (WNPV) model with 52% sequence identity (PDB: 3JVB). After an initial MR solution of the SFX data, positive densities in three previously undefined regions were identified, which enabled us to model loop regions in the central part (146-149, 176-190, 200-207) and additional residues (6-13 and 41-43) at the N-terminus, compared to the search structure. It was possible to resolve most residues which were missing in the search model, although 5 amino acids at

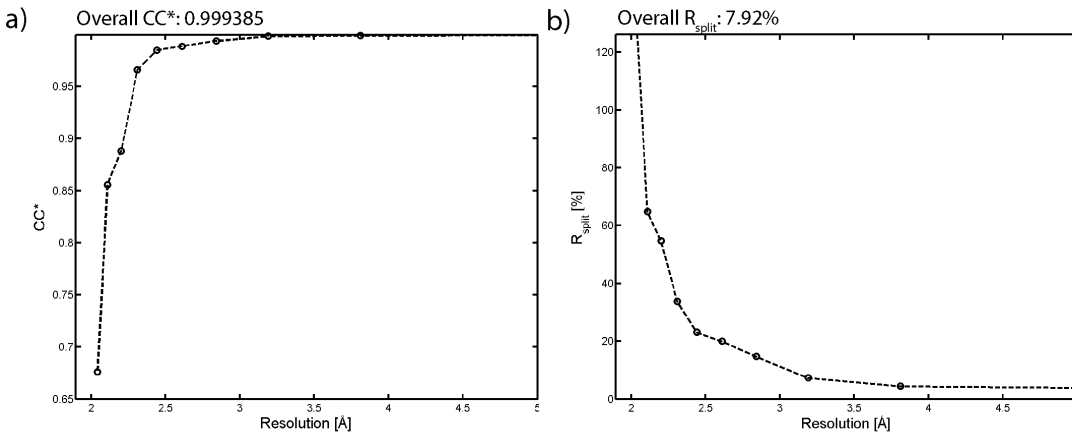


Figure 5.6:  $CC^*$  and  $R_{split}$  statistic plots of the CpGV polyhedrin structure

the N-terminus remained unresolved.

**Structural analysis, comparison and discussion** The central part of the CpGV polyhedrin structure consists of a compact  $\beta$ -sandwich, with an additional two  $\alpha$ -helices H2/3. Perpendicular to the main body, CpGV polyhedrin shows an extruding N-terminus, perpendicular to the central part, beginning with a short  $\beta$ -hairpin structure, followed by a long, bent  $\alpha$ -helix H1. CpGV polyhedrin shows a remarkable structural homology to *Wiseana nucleopolyhedrovirus* (WNPV) polyhedrin (PDB IDs: 3JVB), which was used as a search model for the MR solution, despite the sequence homology of only 52%. However, three unstructured loop regions in the central part and several residues at the N-terminus are well-defined in the SFX CpGV polyhedrin structure, which are absent in the WNPV polyhedrin search model. These missing regions are facing the origin, as well as the core of the unit cell, which are the only cavities containing solvent. Cryogenic cooling of the crystals in the WPNV structure might have induced disorder in these areas of the crystal, even though further analysis would be necessary to support this hypothesis. Comparing the relative B-factor distribution of these two structures, both polyhedrin structures show increased B factors in all five sites, indicating a dynamic environment of these regions (Figure 5.7). A limited structural similarity of CpGV polyhedrin to *Autographa californica* nucleopolyhedrosis virus (AcNPV – PDB ID: 2WUY) can be observed, while mostly the  $\beta$ -sandwich and one  $\alpha$ -helical structure resemble a similarity. The overall crystal packing is very similar to the polyhedrin structures described previously, such as very low solvent content and intertwined trimeric polyhedrin molecules, allowing a variety of crystal contacts. Homologous

Table 5.1: Crystallographic statistics of the CpGV polyhedrin structure (from Gati *et al. in preparation*)

<b>Data collection</b>	<b>CpGV polyhedrin</b>
Temperature, K	294
Wavelength, eV	7900
Beam size, $\mu\text{m}^2$	1.3 x 1.3
Average particle size, $\mu\text{m}^3$	0.24 x 0.24 x 0.4
Crystalline fraction	60%
Flux	$1 \times 10^{12}$ photons/pulse
Max dose per crystal, MGy	1300
Space group	I23
Unit cell, Å	a = 103.3, $\alpha$ = 90
Pulse duration, fs	50
No. collected frames	1.535.619
No. hits	487.085
No. indexed patterns	82.603
No. total / unique reflections	77177221 / 12576
Resolution range, Å	40 - 2
Completeness, %	100 (100)
Multiplicity	6008 (1258)
I/ $\sigma$ (I)	9.59 (0.92)
CC*	0.999 (0.60)
R <sub>split</sub> , %	7.89 (116.2)
<b>Refinement</b>	
No. reflections / test set	16,025 / 814
R <sub>work</sub> / R <sub>free</sub> , %	16.5 / 20.1
No. atoms	2121
Polypeptide	2031
B-factors, Å <sup>2</sup>	36.86
Wilson B/ Overall B	36.31
R.m.s bonds, Å / angles, °	0.003° / 0.58
Ramachandran plot stats, %	
Favored	97
Allowed	2.10
Disallowed	0.9

Table 5.2: Crystallographic statistics of the CpGV polyhedrin structure

Completeness [%]	No. measurements	Multiplicity	SNR	Resolution shell [ave. Å]
100	13616878	10315.8	34.3	7.82
100	11064878	8610.8	21.62	3.81
100	10653215	8441.5	13.49	3.19
100	9488845	7646.1	7.00	2.84
100	8964071	7137.0	5.21	2.61
100	8271654	6601.5	4.50	2.44
100	5996385	4778.0	3.21	2.31
100	3856055	3127.4	2.01	2.20
100	2587320	2055.1	1.67	2.11
100	1544938	1249.9	0.90	2.04

to the CPV17 and WPNV structure, an intermolecular disulphide bond is found at Cys135, which adds to the mechanical stability of these nanocrystals on an atomic level (Figure 5.7 d). These details form the basis of the chemically and mechanically one of the most stable protein crystals known to date.

#### 5.4.5 GV as a reference for analyzing the limits of SFX

Because of the small size of its crystalline shell ( $\sim 10,000$  unit cells) and its narrow particle size distribution, CpGV is an optimal sample for the investigation of current limitations of the method of serial femtosecond crystallography. This unique experiment represents the smallest crystal which ever produced an atomic resolution structure to date, while using the highest flux density available. As mentioned earlier, current experiments are limited by the dynamic range of the CSPAD and due to relatively large crystals used in SFX experiments (1-5  $\mu\text{m}$ ), the beam has to be attenuated in order to not damage the detector with low resolution reflections. Hence, there is currently no published dataset with unattenuated X-ray beam at the CXI instrument. By being able to accurately model the sample and simulate diffraction patterns from those particles, it is possible to monitor and guide simulation approaches from even smaller crystals than CpGV.

Figure 5.8 shows a simulated diffraction pattern of granulovirus polyhedrin with 123 unit cells of polyhedrin - 123 was chosen to achieve a sphere of similar shape than the original CpGV OB. The simulation was performed assuming identical experimental conditions (photon energy, pulse intensity, etc.), but omitting experimental noise, such as background scattering arising from the buffer or electronic noise from the detector. The inset in figure 5.8 b) shows the shape of the crystalline structure and the atomic coordinates that were used for the simulations. The nucleocapsid,

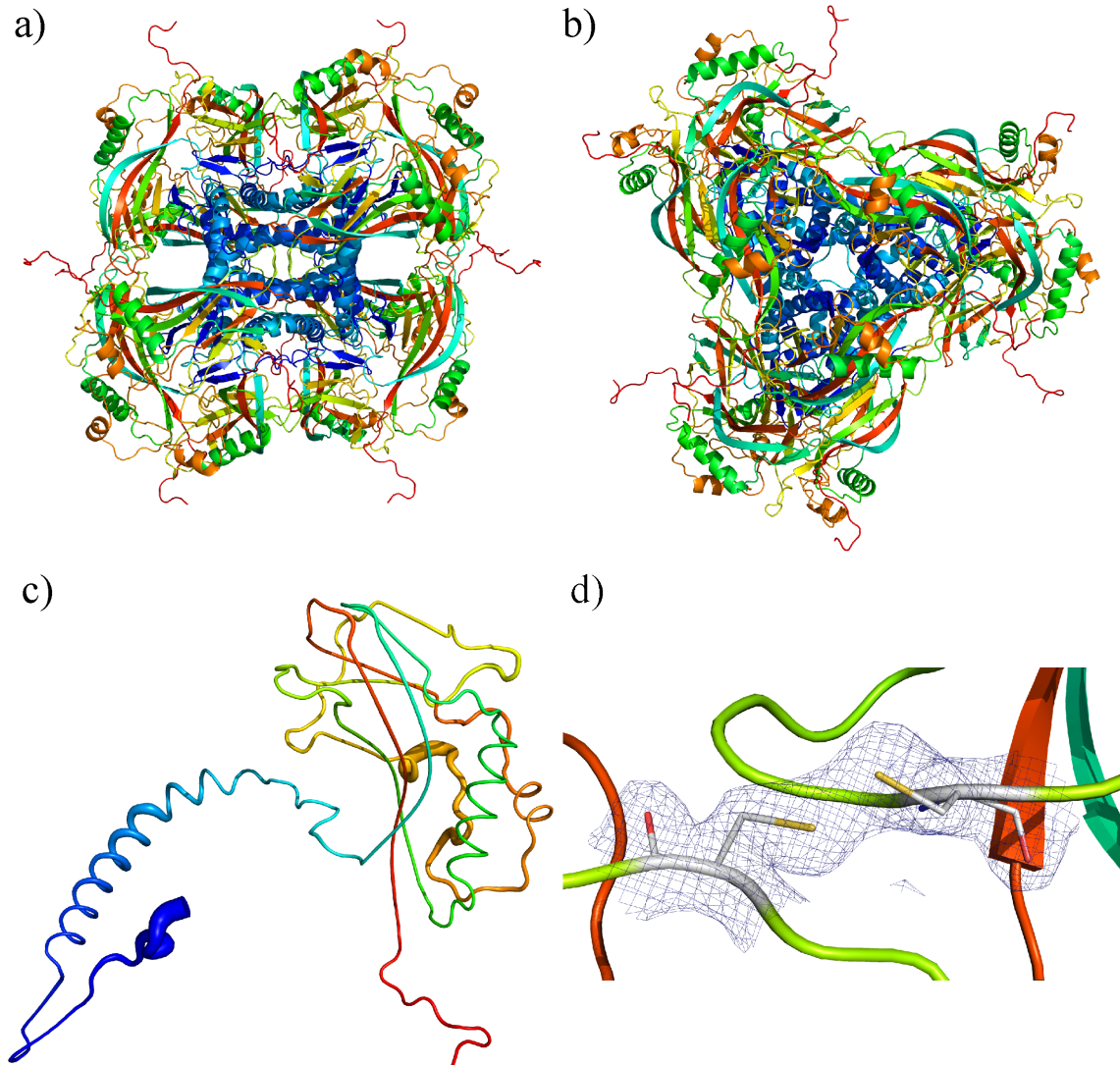


Figure 5.7: Overview of the CpGV polyhedrin structure. a-b) Top view of the biological unit consisting of 12 individual polypeptides resembling the building blocks of the polyhedra shell. c) Cartoon illustration of the CpGV polyhedrin structure representing relative B-factors with thickness. d) Intermolecular disulphide bond at Cys<sub>135</sub> adds to the robustness of the crystal (from Gati *et al.* *in preparation* - Courtesy of D. Oberthuer CFEL)



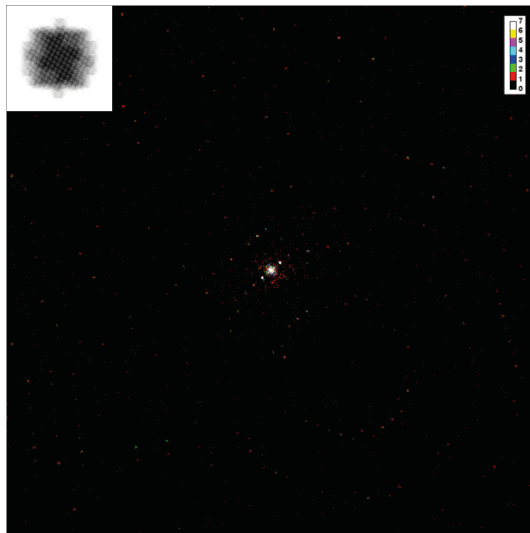


Figure 5.8: Simulated diffraction pattern of crystalline CpGV polyhedrin, omitting experimental noise (right) Simulated diffraction pattern from a crystal with a total of 123 unit cells (inlet) shape of the crystalline particle for the simulation with 123 unit cells. Courtesy of O. Yefanov (CFEL) (from Gati *et al. in preparation*)

which would be present in CpGV, is absent here for simplicity, which should not significantly alter the diffracted peak intensities.

Despite the lack of experimental background in these simulations, it could potentially be possible to obtain a high resolution structure of crystalline samples as little as 123 unit cells with currently available light sources. In fact, the resolution of this simulated dataset was solely limited by the detector size. These results lead to the conclusion that bioimaging using XFELs is already close to the visionary goal of single particle diffraction at ambient temperatures. Two of the most important experimental limitations of single particle imaging are background scattering and hitrate of current injector systems [101]. By further reducing background to increase the signal from single biomolecules, improving sample delivery techniques e.g. by reducing background by using an aerodynamic lense and to increase the overall hitrate and novel algorithms to extract structural information from diffraction patterns with very sparse data [102], it might be possible to reach the exciting regime of high-resolution single particle imaging with XFELs. This technique would revolutionize the field of structural biology, as it would overcome the need of tedious crystallization. In contrast to single-particle cryoEM it would enable room temperature measurements of individual molecules, allowing structural investigation without visible structural radiation damage and a unique time resolution to



investigate time-resolved dynamics.

## 5.5 Effects of experimental parameters and improvements of SFX data quality

### 5.5.1 Detector Saturation

For SFX experiments, the currently most frequently used detector is the CSPAD (Section 5.2.3) with a dynamic range of  $10^3$ , meaning the ratio between minimum detectable to maximum measurable scattered intensities between e.g. simultaneously detecting 10 and 10,000 ADU (arbitrary detector units), which translates to a wavelength-dependent limit of photons per pixel (e.g. 24 ADU per each 8 keV photon). Saturation of a detector pixel is reached when too many photons arrive at a pixel in a given time and the downstream electronics are not able to integrate the full intensity of a certain Bragg peak. In a diffraction pattern, the scattered intensity drops as a function of the scattering angle  $q$ , following the Wilson B-factor (Section 2.20). SFX experiments represent diffraction limited experiments, where the resolution achievable in the diffraction pattern is limited by the order and quality of the crystal, as opposed to limitations occurring from radiation damage or a lack of sufficiently high X-ray intensities.

As mentioned earlier, most SFX experiments are carried out with crystals in the range of  $1 - 10\mu\text{m}$  in diameter, since detection and monitoring of protein nanocrystal growth is very challenging, with currently no technique reliable enough to provide sufficient accuracy. With crystals of such size, almost regardless of the number of unit cells per crystal, using an unattenuated beam at the LCLS, all low resolution reflections might be saturated or even damaging the detector. In fact, even the smallest crystals used so far (granulovirus occlusion bodies - section 5.4), show saturated Bragg peaks.

The most important information extracted from diffraction experiments are relative variations between scattered hkl intensities, rather than absolute intensities. In case of saturated low resolution peaks, those intensities are at the maximum reachable value of each pixel and therefore all have identical values. Hence, if many saturated low resolution reflections are recorded, the low resolution information content is practically lost. This represents a crucial part of a dataset, since these reflections are essential for e.g. experimental phasing.

One result of saturated diffraction patterns merged into a dataset is that conventional twinning tests (L-tests etc.) falsely indicate, that the obtained data are

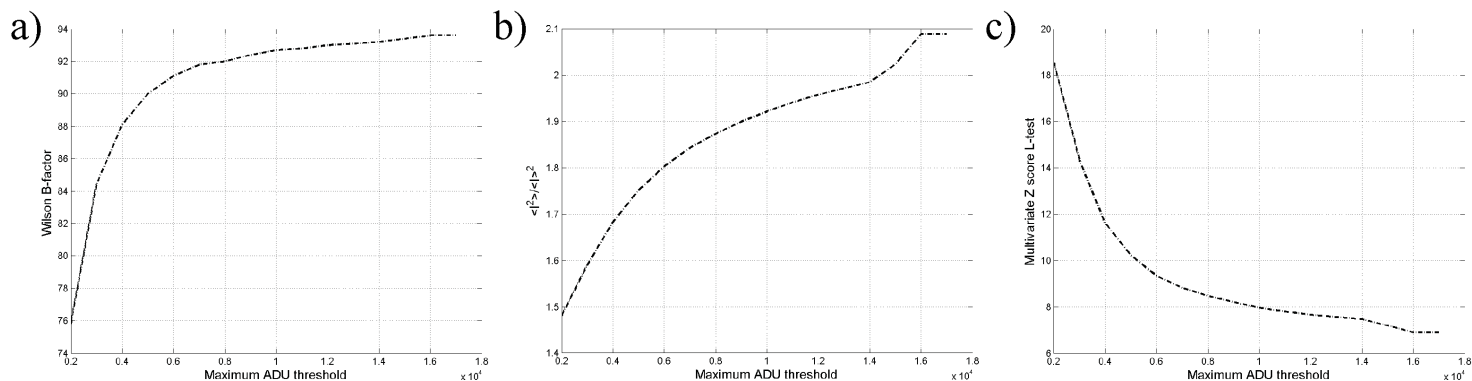


Figure 5.9: Effects of thresholding of data derived from saturated diffraction pattern on data quality. Maximum ADU threshold discards Bragg peaks which have one or more pixel intensities above the value. a) Wilson B-factor against maximum ADU b) and c) are results from conventional twinning tests.

‘twinned’ (Figure 5.9). Twinning tests rely on a certain intensity distribution comparing ‘close’ (similar indices) hkl intensities. Wilson showed that for a single crystal the mean and higher moments of centric and acentric intensities and amplitudes follow a predictable pattern [103]. Twinned data would show an unnaturally low variance in intensities, due to superimposed lattices and the resulting Bragg peak intensities are scattered from different crystal lattices.

Figure 5.9 b) and c) show results from twinning tests from a saturated dataset derived from AT<sub>1</sub>R. This dataset was collected with an average transmission of 16%, which is most likely beyond an intensity that the detector can cope with, in particular using crystal sizes of 5-10  $\mu\text{m}$ . A way to deal with saturation in crystallography data is thresholding. By applying a maximum threshold in the merging step, where Bragg peaks including one or more pixel value above this threshold are discarded. This feature is one way to computationally avoid including saturated peaks, so only Bragg peak intensities within the dynamic range of the detector are included.

The intensity distribution  $\frac{\langle I^2 \rangle}{\langle I \rangle^2}$  (Figure 5.9 b) should vary between theoretical values of 2.0 for ‘untwinned’ data and 1.5 for perfectly twinned data. Beginning with a low maximum ADU threshold of 2000, the  $\frac{\langle I^2 \rangle}{\langle I \rangle^2}$  value is beyond what could theoretically be true, even for a perfect twin, while with increasing threshold this value reaches a maximum at about 2.0. A similar trend can be seen in Figure 5.9 c), where a multivariate Z score was calculated for each maximum ADU threshold, which should ideally reach a value of lower than 3.5, even though the value never drops below 6. The intensity statistics observed in L-tests [104] for two different

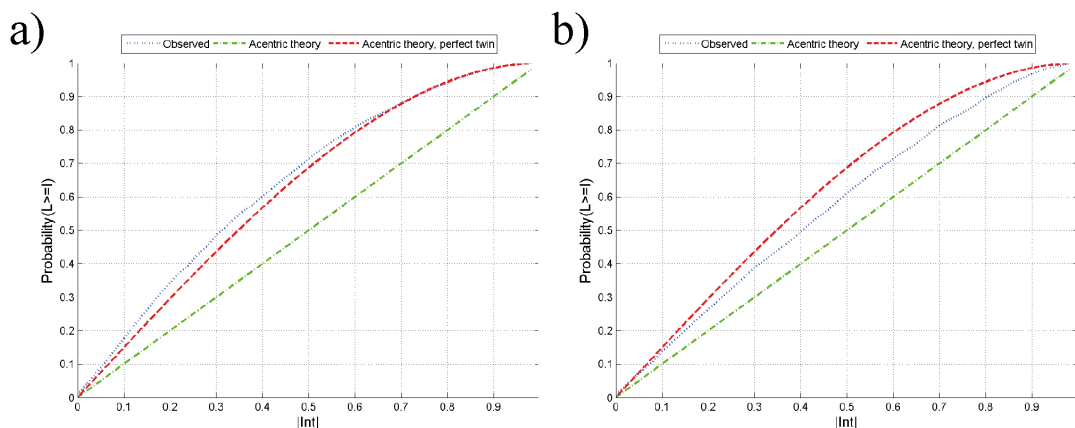


Figure 5.10: Effects of detector saturation on L-test results. a) Maximum ADU threshold of 2000 b) Maximum ADU threshold of 16000.

datasets with thresholds of 2000 (Figure 5.10 b) and 16000 ADU (Figure 5.10 c) show an identical trend, where including a higher threshold improves the shape of the plot of the acentric reflections, but still never reaches the theoretically achievable value. In case of data derived from saturated detector frames, twinning tests appear to be well suited to monitor the severity of saturation (Figure 5.10), even though this could only be used with great care and only on non-twinned datasets. One additional metric to examine saturation is the Wilson B-factor, which will artificially be lowered by saturated detector frames (Figure 5.9 a).

To counteract the problem of saturation, all currently published SFX structures at the LCLS are recorded with an attenuated X-ray beam, typically on the order of 0.1 - 3% transmission. The flux for each recorded diffraction pattern is comparable to the synchrotron case, applying doses well below the Henderson limit (3.5.1), even though it would theoretically not be necessary. By limiting the incoming intensity, the theoretically achievable resolution is limited, which is an obviously unwanted condition.

One option to overcome this problem is the usage of the two available gain-modes of the CSPAD, which can be applied to individual pixels. This ‘low gain’ mode increases the limit of number of arriving photons per pixel by a factor of about 7, which was used for the first atomic resolution time-resolved experiment using SFX on PYP [105]. However, this gain factor needs to be corrected for, since it does vary depending on the incoming beam intensity, which might introduce artifacts to the final data and requires more thorough characterization.

Currently all novel detector designs take this limitation into account, aiming

for dynamic ranges of  $10^4$  (or higher), while trying to improve repetition rates for future XFEL and synchrotron sources [75][76]. Alternatively to in vacuum SFX experiments, a setup for in helium/air experiments is both available at LCLS and SACLA, where CCD detectors with higher dynamic range can be used. However, these suffer from background scattering from long path lengths in helium/air and additionally from low repetition rates due to long readout times.

### 5.5.2 Detector calibration

One difficulty in data analysis of SFX data is the accurate calibration of the detector distance. An exact motor position of the detector on its stages can be determined to nm accuracy, while the absolute distance between the sample and the detector stage can vary. The current design of the injector-rod which is placed inside the vacuum chamber through a load-lock results in non-reproducibility of the distance. The most broadly used strategy to counteract this problem is the calibration of the detector distance by using a sample with known unit cell dimensions and Bragg's law  $L = \frac{D}{\tan(2 \sin^{-1} \frac{\lambda}{2d})}$  (L: detector distance, D: diameter of the detector, d: recordable resolution). Even though this strategy can achieve a certain accuracy, the unit cell dimensions can significantly vary for different crystallization batches and conditions, e.g. depending on the hydration level of the crystal. In fact, the significant discrepancy between the unit cell dimensions of *in vivo* grown TbCatB microcrystals solved at the LCLS [64] and PETRAIII [106] was thought to be induced due to the temperature difference during the experiment (294 vs. 100 K). But in fact after careful analysis and detector calibration using the software described by Yefanov *et al.* [74], the difference in unit cell dimensions was found to be due to a wrong detector distance assumed in the SFX experiment. Alternative ways of experimentally deriving the detector distance would be desirable, in order to overcome this problem and accurately determine this value. Laser interferometry could be one possibility to achieve a highly accurate determination of the real detector distance.

Due to the flexibility of the CSPAD, each quadrant being movable separately in all directions, its constant development and reassembly, as well as the physical exchange between the two sample chambers, a way to determine the exact detector metrology is crucial. Experimentally this can only be achieved to a certain accuracy. Before the detector is assembled, a light-microscopy based metrology of each quadrant is recorded, to an accuracy which is not sufficient for accurate prediction of Bragg peak positions. In order to obtain a first estimate of the relative positions of each quadrant, a virtual powder pattern of a calibration sample is recorded, which needs to be available in sufficient quantity and has to have a limited unit cell size in

order to have a sufficient spacing of Bragg peaks, which would otherwise result in overlapping Debye Scherrer rings. With this first rough refinement of the detector metrology it is usually possible to obtain the first indexing results that are needed for the final sub-pixel accuracy refinement of the detector geometry. The basic principle of this algorithm is the comparison of the found Bragg peak positions and the predicted peak position. With a sufficiently large dataset, it is possible to resolve systematic errors of the detector geometry. Typically, the highest accuracy geometry is achieved by iteratively applying this algorithm and rerun the indexing and integration step using CrystFEL. The geometry refinement step has been one of the most crucial improvements in SFX data analysis, since it drastically improved data quality, in case of CpGV leading to a 5-fold increase of the number of indexed patterns and a resolution improvement from 3 to 2 Å.

### 5.5.3 Improved integration approaches

The currently used 2D integration step implemented in CrystFEL is a simple three ring approach, where three non-dynamic radii are specified prior to the analysis step. The first radius specifies the number of pixel values that are summed for the integration, while the second and third specify a circle that is used to estimate the local background of each Bragg peak. Even though this has been the most successful approach so far using the CrystFEL software package, it is counter-intuitive, since fluctuations in crystal sizes and shapes, as well as X-ray beam fluctuations will have an influence on the shape of each Bragg peak. 2D and 3D profile fitting approaches have been successfully implemented for synchrotron rotation data, which have the advantage of observing one particular Bragg reflection fully over several frames from different angles with high accuracy. These algorithms are only of limited use for SFX data, since most Bragg peaks are only partially recorded and only to low accuracy, due to background scattering and experimental noise. Nevertheless, several algorithms and software packages have been implemented in order to resolve the peak shape for each reflection to allow the best possible integration mask [80][94][84].

### 5.5.4 Effect of averaging of diffraction patterns on data quality

As mentioned earlier, CpGV polyhedrin nanocrystals are an exceptionally good sample to model experimental parameters, due to their small size and the homogeneous size distribution. Averaging of many, weak datapoints to obtain an accurate dataset is a well known phenomenon and is particularly important in the field of (bio)imaging. Conventional crystallography uses the strength of averaging over a

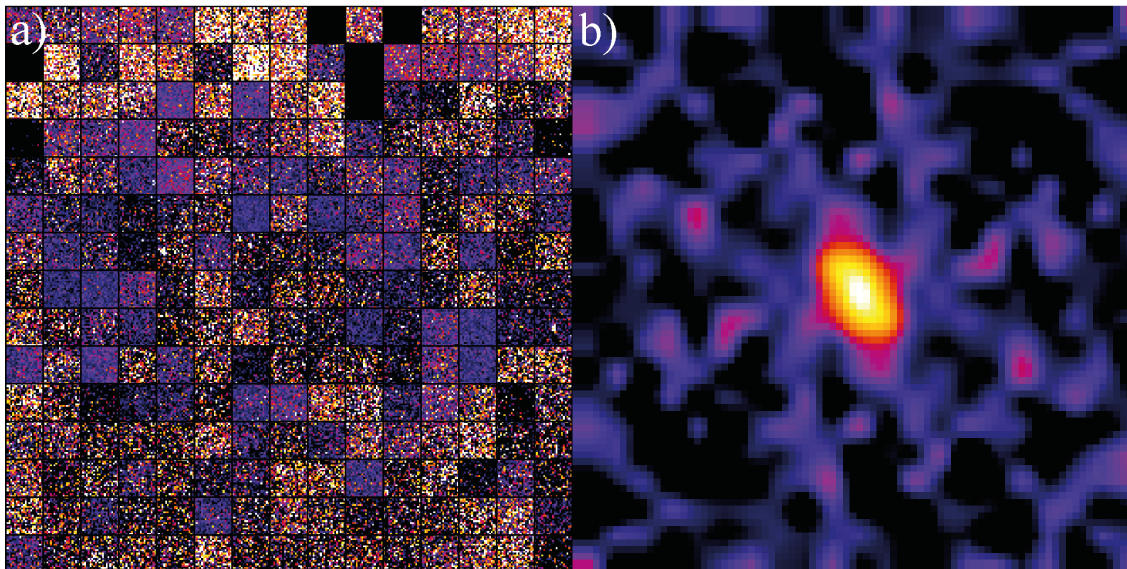


Figure 5.11: Illustration of the averaging over thousands of diffraction patterns (a) Random selection of Bragg peaks with index  $hkl$  (0 22 24) (b) final averaged, merged Bragg peak intensity. Courtesy of O. Yefanov (CFEL)

large number of unit cells, in order to overcome signal-to-noise ratio limitations in X-ray diffraction experiments. As a different approach, single particle cryo-electron microscopy relies on the averaging of thousands of very noisy, weak, individual particle projections in order to finally obtain high resolution electron density maps [89][107]. A comparable trend has been followed in the field of macromolecular crystallography, where merging multiple incomplete datasets can improve the overall resolution of the merged dataset significantly [38]. The technique of serial crystallography makes use of this averaging process and brings it to an extreme. By Monte Carlo integration (5.3.2) of thousands of weak, partial diffraction patterns, it is possible to fully probe reciprocal space and accurately determine structure factor amplitudes to an accuracy, where experimental phasing is possible [108], however still challenging and limited to model cases.

Partial Bragg reflection intensities extracted from thousands of diffraction patterns are merged in order to estimate the ‘true’ intensity of each Bragg peak. This results in an improvement of signal-to-noise ratio and finally in a more accurate structure factor determination with increased confidence. In fact, it is possible to bring this to a point where the Bragg peak intensity is not visible in an individual diffraction pattern, but can still be recovered by averaging over several thousand observations, since the signal accumulates at a faster rate as the background. Figure

5.11 illustrates this averaging approach of one particular hkl reflection (0 22 24).

Figure 5.12 shows several plots of figures of merit, as functions of an increasing number of diffraction patterns in the case of CpGV. Three different lines represent different dataset, which were sorted according to their scattering strength. Generally all statistics (SNR,  $CC^*$  and  $R_{\text{split}}$ ) steadily improve with an increasing number of diffraction patterns for all three datasets. Obviously they converge to the same statistic, when including all data points. As one would expect, the SNR value for the random dataset in (a - black) increases roughly as the  $\sqrt{n}$ , while the ascending dataset (red - weak to strong) starts off with a less steep slope and the descending dataset (blue - strong to weak) begins with a steep slope which eventually decreases.

The most interesting point from these plots is the fact that in case of the descending sorting, from strong to weak diffraction patterns, in particular comparing only the highest resolution shell (b), (d) and (f)), the best value is reached not by merging all patterns, but by including the strongest 65,000 diffraction patterns. All three figures of merit, SNR,  $CC^*$  and  $R_{\text{split}}$  reach the best value, followed by a slight decrease in data quality by including the weaker diffraction patterns. This can be explained by the fact that these weak diffraction pattern have a lower signal-to-noise ratio for individual Bragg reflections, making the structure factor estimate less reliable. Additionally, it is more likely that weaker patterns are wrongly indexed, which present an additional source of noise in the final data. Hence, one would like to reject these outliers, which is a common strategy in conventional multi-crystal crystallography. Finally, the decrease in data quality by adding the weak diffraction pattern after the strong patterns only has a marginal influence on the final data quality and mostly affects the highest resolution shell.

### 5.5.5 Effects of per pattern resolution cutoff on data quality

Experimental fluctuations are the major difficulties of data processing in serial crystallography. In particular with crystals of varying diffraction quality, either due to order, or illuminated volume of the crystal, as well as variations of the incoming intensity of the X-ray pulse, the final diffraction limit of each diffraction pattern can vary significantly. With the goal of reducing noise in the final dataset, we applied a per pattern resolution cutoff to obtain the structure of AT<sub>1</sub>R (Section 6.4), which became a broadly used procedure with later datasets. The principal idea is to estimate the resolution limit of an individual diffraction pattern to avoid the integration of non-existing Bragg peaks. The conservative resolution limit of each diffraction pattern is estimated by matching the initially found Bragg spot positions with the predicted spot locations. The resolution limit is then calculated by taking the 98th

5.5. EFFECTS OF EXPERIMENTAL PARAMETERS AND IMPROVEMENTS OF SFX DATA QUALITY 81

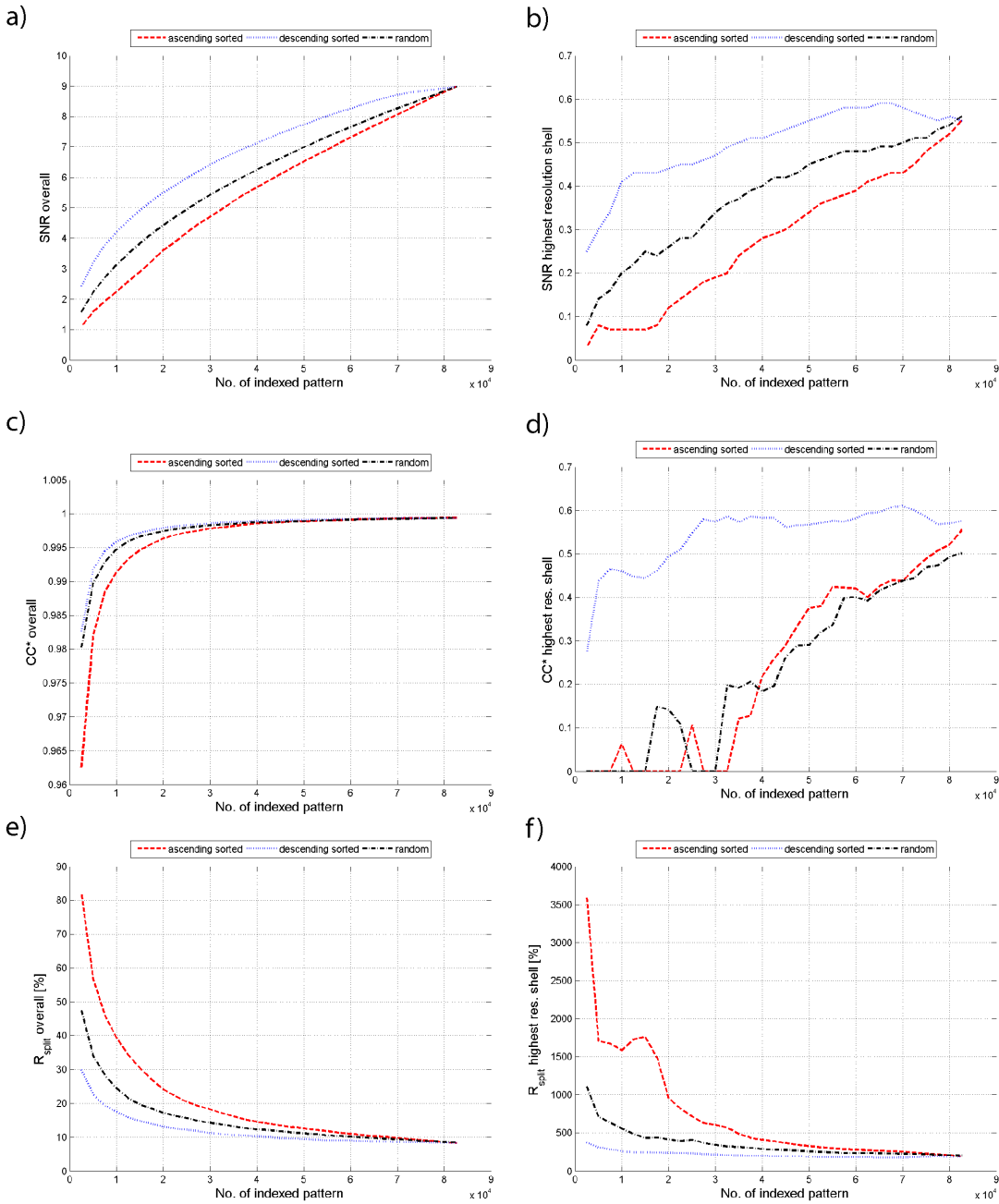


Figure 5.12: Effects of averaging diffraction patterns. Figures of merit are plotted as a function of increasing number of included diffraction patterns. For each curve diagram three different datasets were analyzed: (black) the final, untreated dataset of CpGV and the dataset sorted by its scattering strength (judged by the number of peaks per diffraction pattern) from weak to strong (ascending - red) and sorted from strong to weak (descending - blue).



percentile of the scattering angles within the matching peaks.

Figure 5.13 compares conventional figures of merit of the AT<sub>1</sub>R GPCR structure obtained by LCP-SFX. In addition to the conservative per pattern resolution cutoff, the CrystFEL option *'pushres'* adds a certain resolution bin to each resolution limit with the idea to avoid losing meaningful data by applying a too strict cutoff. It increases the estimated resolution cutoff by a value in  $\text{nm}^{-1}$ , this value often has its optimum around 1, e.g. increasing a conservative resolution estimate from 3 Å to 2.4 Å. The plots begin with 0.8  $\text{nm}^{-1}$ , as lower values only include a small number or no observations in the highest resolution shells, leading to unreliable statistics. One interesting observation from the plots in Figure 5.13 is that the SNR in the highest resolution shell,  $R_{\text{split}}$  in the highest resolution shell and the LLG value for a MR solution show a parabolic trend. By including additional data beyond the conservative resolution cutoff, it reaches a maximum at 1.2  $\text{nm}^{-1}$ , after which the data quality drops again. As the Monte Carlo integration calculates a weighted average of individual Bragg peak intensities, it is strongly dependent on data quality. By adding intensities integrated from non-existent Bragg peaks ('zeros'), the average Bragg peak estimate is reduced, which finally leads to a lower signal to noise ratio and generally worse data quality. A possible explanation for the initial increase in data quality is the inclusion of previously missing Bragg reflections with meaningful intensities, as the initial resolution estimate is too conservative. Beyond this optimum, too weak or non-existent Bragg reflections are being merged into the data. This would ultimately lower the final merged intensity, as well as the data quality, of the hkl intensities, compared to the dataset with applied per pattern resolution cutoff. An interesting aspect is that the  $\text{CC}^*$  in the highest resolution shell does not follow the same trend. It appears to only follow the trend 'the more the better', as the statistics improve by simply including more observations. Another interesting point is the Wilson B-factor which seems to oscillate around a maximum after reaching the value of 1.2  $\text{nm}^{-1}$ . It is intuitive that by including more low-SNR Bragg peaks in the highest resolution shell, the B-factor increases.

From these statistics one can conclude that there is an optimal *'pushres'* value for this dataset and it has been successfully shown for several other datasets. Regardless, it has the limitation that this estimate of a per pattern resolution cutoff strongly depends on the initial peak detection during hitfinding. For example, it is quite common to use a mask for peak detection during hitfinding in order to avoid false positive Bragg peaks e.g. in the noisy water ring. Applying this mask can strongly influence the statistics, since the conservative resolution cutoff will only extend to the point where peaks are found. This can ultimately lead to artifacts in the data, such as artificially low B-factors, if one constantly discards all high resolution

5.5. EFFECTS OF EXPERIMENTAL PARAMETERS AND IMPROVEMENTS OF SFX DATA QUALITY83

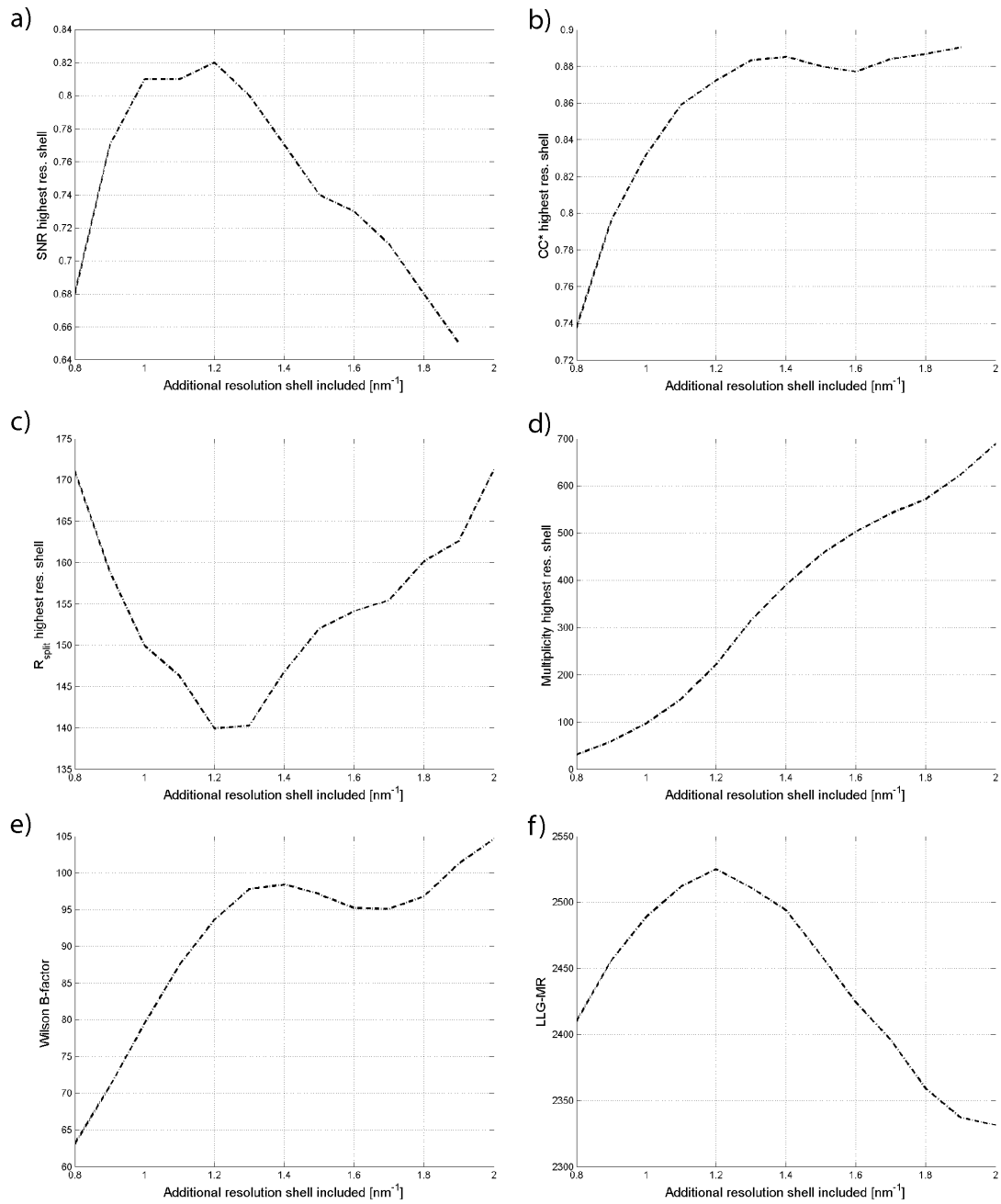


Figure 5.13: Effects of the per pattern resolution cutoff. Figures of merit are plotted as a function of increasing resolution included.

reflections, regardless of their strength. It would be interesting to test alternative ways to estimate a per pattern resolution cutoff by e.g. estimating a B-factor for each diffraction pattern using only low resolution Bragg peaks, in order to estimate the point where the signal is below background.

In general the topic of outlier rejection is a delicate task in serial crystallography, as sometimes the number of diffraction patterns is limited, or processing can be heavily biased e.g. by using an external reference ('Einstein from noise' [109]), in particular when working with noisy data. On the other hand including diffraction patterns or reflections of limited quality can drastically drop the final data quality. Nevertheless, the last two sections of this chapter indicate that careful outlier rejection either on individual pattern or reflection basis can improve data quality drastically. A possibility would be to introduce more sophisticated algorithms, such as hierarchical cluster analysis, which was introduced both for conventional [110], as well as serial crystallography datasets [111]. It is always best to manipulate the data as little as possible, while trying to improve the overall data quality as much as possible, while balancing the two is usually not easy. For this it is absolutely crucial to understand the information content one is trying to optimize, while it is necessary to have a solid knowledge about the possible sources of errors and sufficient statistics, in order to be able to justify 'data massaging'.

## Chapter 6

# G protein-coupled receptors in LCP-SFX

### 6.1 Physiology of G protein-coupled receptors

Membrane proteins represent one of the biomedically most important classes of proteins, as they represent one third of the human proteome and perform critical cellular and physiological functions. Developing drugs targeting intracellular, soluble proteins is often challenging, due to a limited permeability of most compounds through the cell membrane, general pharmacokinetic limitations and alternative signaling pathways which bypass the inhibition of individual proteins. Drugs targeting membrane proteins are more favorable and in fact represent 60% of all current drug targets in humans [112]. This can be explained by an easier accessibility of the drug to the cell membrane and the possibility to block a signaling pathway before any extracellular signal can be sensed, or an intracellular response can be triggered. Despite their importance, structural information of membrane proteins is clearly lagging behind, as less than 1% of the structures currently deposited in the PDB are derived from membrane proteins.

G protein-coupled receptors (GPCRs), integral membrane proteins consisting of seven transmembrane helices, represent a large membrane protein family and play a crucial role in survival and accurate responses to diverse stimuli of a cell. Hence, GPCRs are involved in a variety of diseases and are the most targeted group of proteins for prescriptive drugs on the market (30-50%) [113], while two of the top 5 best selling drugs in the US target GPCRs. Given the importance of this group of receptors, the Nobel prize in Chemistry was awarded to Brian Kobilka and Robert

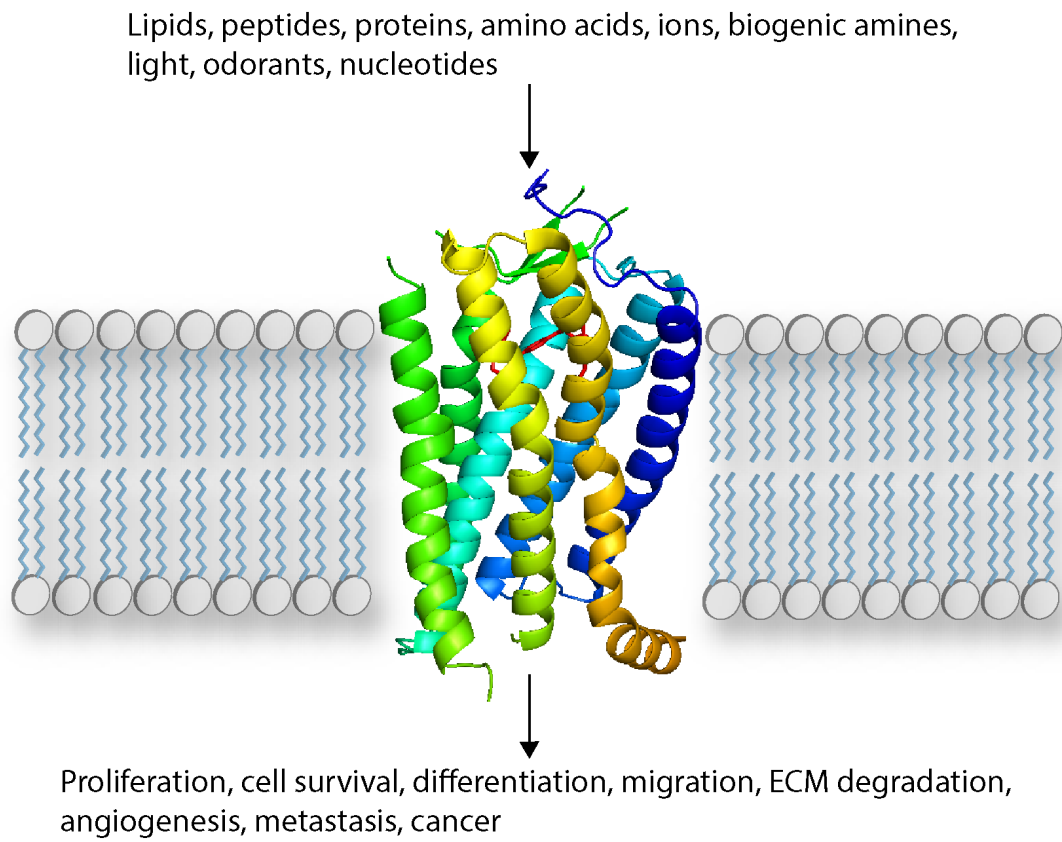


Figure 6.1: Physiological processes of G protein-coupled receptors, embedded in lipid bilayer

Lefkowitz "for studies of G protein-coupled receptors". Lefkowitz was the first one to discover this important group, identifying  $\beta$ -adrenergic receptors by radioisotope labeling. Kobilka was awarded for his discovery of gene expression regulation of the synthesis of these receptors [114], his structural work, after solving the first GPCR structure [115] and a complex structure between a G-protein and a GPCR [116].

Members of this group of receptors consist of seven transmembrane domains and generally induce a signaling cascade in the intracellular lumen upon arrival of extracellular signals. These signals are ligand-driven and include odors, pheromones, hormones and neurotransmitters with a varying size from small molecules to peptides or small proteins. Receptors are generally in an equilibrium between inactive and active state, while extracellular signals are capable of altering that equilibrium. In general, ligands interacting with GPCRs can be grouped into three classes:

- **Agonist:** shifting equilibrium towards the **activated state** of the receptor
- **Inverse Agonist:** shifting equilibrium towards the **inactive state** of the receptor
- **Antagonist:** do not affect the equilibrium, but **block the binding** of another ligand

GPCRs are only found in eukaryotes, however structurally similar receptors exist in evolutionary older organisms, such as bacteriorhodopsins which are mostly found in Halobacteria, and show a similar behavior (sensing) and overall shape (seven transmembrane helices). Physiological sensory functions of GPCRs are diverse, beginning with the visual (rhodopsin), olfactory and gustatory senses, neurotransmitter detection and behavioral effects, immune system regulation, embryogenesis, regulation of the endocrine/autonomic nervous system, cell density sensing, homeostasis (e.g. water and electrolyte balance) and metastasis growth [117]. There are a total of 800 annotated genes of GPCRs in the human genome, classified into five subgroups (A, B, C, F, adhesion), of which class A accounts for 85% of all GPCRs. Over half of all class A receptors are predicted to be olfactory receptors, of which little is known physiologically, as well as structurally. Interestingly, these groups do not show any detectable sequence homology, while structural features are common within the group, such as large extracellular domains and dimer formation in class C GPCRs.

All effects driven by GPCRs are achieved upon interaction with a ligand, including the sensing of physical signals, such as light in the case of rhodopsin, which induces a *cis-trans* isomerisation of its bound ligand retinal, followed by a conformational change of the receptor (Figure 6.1) [118]. During the initial conformational

change of a GPCR upon activation by a ligand, the relative positions of the transmembrane helices are being altered, leading to an increased exposed intracellular surface of the receptor. Heterotrimeric G-proteins are hypothesized to be bound to the inactive GPCR prior to activation [119], however this is controversially discussed. There are two different types of transducers known in the cell, which communicate between the activated receptor and the cell by direct interaction to the GPCR upon activation - either G protein ('G protein biased') or arrestin ('arrestin biased') mediated.

After the new intracellular domain of the receptor is being exposed, the GPCR acts as a guanine nucleotide exchange factor (GEF), by exchanging the bound GDP in the G-protein to GTP. The heterotrimeric G-protein dissociates upon GDP to GTP exchange, the  $\alpha$  subunit (with bound GTP) dissociates from the  $\beta$  and  $\gamma$  subunit to induce signaling pathways inside the cell. There are three main signaling pathways activated by the  $G_\alpha$ -subunit, mediated by either of the four types of G-proteins:

- **cAMP** production by adenylate cyclase ( $G_s$ -protein mediated) - effect on ion channels and protein kinase A (PKA)
- **phosphatidylinositol** production by phospholipase C ( $G_q$ -protein mediated) - protein kinase C (PKC) and calmodulins
- **Rho** (small GTPase) activation ( $G_{12/13}$ -protein mediated) - cytoskeleton regulation

The  $G_{\beta\gamma}$ -subunit primarily activates ion channels, such as G-protein-regulated inwardly rectifying K<sup>+</sup> channels (GIRKs) or P/Q- and N-type voltage-gated Ca<sup>2+</sup> channels. After dissociation of the G-protein, depending on the type of receptor and its conformational change, a new heterotrimeric G-protein can bind to the receptor which allows a new activation cycle.

Alternatively to the G-protein mediated pathway, again depending on the type of receptor and its conformational change upon activation, the receptor is very strictly regulated by an alternative binding protein called arrestin, which binds the active form of the GPCR. Several conserved sites (Ser or Thr) on the C-terminal tail of the GPCR can be phosphorylated, drastically increasing the affinity of the receptor to  $\beta$ -arrestin and enables an interaction. Arrestin sterically blocks G-protein binding and finally induces alternative pathways, such as extracellular-signal regulated kinase (ERK), inactivation or receptor internalization. Drugs can either work through the recruitment of G-proteins or arrestin, or both of them simultaneously. Most drugs induce both pathways, while drugs preferentially inducing one out of the two possible

pathways ('functional signalling' or 'biased agonism') are associated with fewer and less severe side-effects [120][121].

## 6.2 Structural biology of G protein-coupled receptors

An enormous variety of biochemical assays and techniques is available to study membrane proteins in great detail, but in order to fully understand the physiology of a receptor and its disease related function, the accurate knowledge of the structure of the receptor is a key advantage. By being able to investigate and analyze the interaction of a ligand with its target protein in atomic detail, it is possible to understand its affinity, resulting effects upon binding and conformational changes and more global impacts, such as side effects of a drug by binding affinities towards unwanted targets. Additionally, *in silico* approaches are recently being used in the field of drug design, where the affinity and the pharmacodynamics of a ligand and its binding pocket are calculated using molecular dynamics simulations. The accuracy of the outcome of the predictions of protein-ligand interactions are heavily dependent on the quality of the initial structure of the protein, which is another reason why structural information of proteins are so important.

The structural investigation of membrane proteins is particularly challenging. Most GPCRs (and other membrane proteins) are structurally heterogeneous, unstable in detergent-based solution, fragile to physical or chemical stress and typical expression yields are very low, making them a difficult target for crystallographic structure determination. Commonly used (thermo-)stabilization of the receptor is achieved by introducing specific point mutations in hinge-regions, which lock the protein in one conformation, usually supported by co-purification of the receptor with a ligand of interest. In addition to point mutations, antibody fragments are used to both lock the receptor in a certain conformation, as well as to increase the overall solubility of the complex [115] [116]. Flexible loop regions of the receptor are entirely deleted and replaced by either T4 lysozyme [122] or more recently a thermostabilized apocytochrome, b562RIL (BRIL) domain from *E. coli* [123] in order to increase the solubility of the receptor, enforce a certain conformation and to facilitate crystal contacts. In addition to mutagenesis studies, it is crucial to identify the best detergent for solubilization and different detergents for the crystallization of each membrane protein, which can be very time and resource consuming. Sf9 insect cell lines with a baculovirus system are most frequently used to obtain the largest yield of GPCRs possible, in order to provide the protein quantity needed for



the crystallization process. The currently most commonly used crystallization technique for the crystallization of GPCRs are LCP based approaches [122], as discussed previously in section 3.6.1.3.

One of the most prominent targets in the field of GPCR pharmacology are  $\beta$ -adrenergic receptors, which play a crucial role in cardiovascular diseases, the leading cause of death globally [124]. The first structure of a human GPCR was  $\beta_2$ -adrenergic receptor, which was solved by Kobilka and coworkers by designing an antibody fragment to increase the solubility of the receptor in the year 2007 [115], followed by a higher resolution structure (2.4 Å) with an engineered  $\beta_2$ -adrenergic receptor, genetically fused to a T4 lysozyme and crystallized in an LCP environment [122].

Structurally, GPCRs are characterized by a topology with an extracellular N-terminus, seven transmembrane helices connected with each three extra- and intracellular loops (IL1-3, EL1-3) and an intracellular C-terminus (Figure 6.1). Extracellular loops are often glycosylated and show a conserved disulphide bond to improve receptor stability, while the C-terminus is mostly involved in downstream signaling. The overall shape of a GPCR is comparable to a barrel, where the transmembrane domains form a cavity inside the membrane presenting a ligand binding pocket, while voluminous, large peptides and proteins bind to extracellular loops or the large, N-terminal tail in case of class C GPCRs.

In the inactive form GPCRs bind a heterotrimeric G-protein. Upon agonist binding of the GPCR, which results in an activation of the receptor, most prominently the cytosol facing parts of transmembrane helix V and VI, are moving outwards, resembling a twisting motion. Only after that GDP is exchanged to GTP in the  $\alpha$  subunit, followed by a dissociation of the three subunits. On the other hand, in case of binding of inverse agonists or antagonists, it is necessary to prevent the conformational change of transmembrane helices to disable G-protein binding. A further common feature is one or several palmitoylation sites of the C-terminal tail. This is achieved by covalently binding alanyl groups to a cysteine of the receptor, triggering the targeting of the receptor to cholesterol- and sphingolipid-rich microdomains in the cell membrane, called lipid rafts.

### 6.3 First structure of human membrane protein at room temperature

Membrane protein structures are of great importance, however their structure determination is usually hindered by the difficulty to obtain large, well-diffracting crystals.

One promising technique to overcome this bottleneck is LCP based crystallization (Section 3.6.1.3). Initial crystallization trials usually lead to small crystals, with a high nucleation rate, but optimization to obtain crystals of sufficient size for conventional crystallography is often tedious or not successful. Microfocus beamlines have been developed and improved steadily with great success [59][125][126], however the achievable resolution of small crystals at synchrotrons is ultimately limited by radiation damage of the specimen. Even at a modern micro-focus beamline, typically long exposure times of several seconds have to be applied per diffraction pattern, only allowing slices of reciprocal space from each crystal due to global radiation damage. This usually leads to the necessity of merging data from multiple crystals.

Because of the high nucleation rate, small crystal size and the viscous medium, LCP grown microcrystals are an ideal sample for serial femtosecond crystallography experiments. As described earlier (Section 5.2.2), the adaptation of the injector to deliver this viscous material, by using the 'LCP-injector' [68], opened up the possibility to perform experiments on these very important target proteins using XFELs.

#### 6.3.0.1 Structure of the 5-hydroxytryptamine receptor 2B (5-HT<sub>2B</sub>)

The following experiment is described in the publication by Liu *et al.* 'Serial femtosecond crystallography of G protein-coupled receptors', published in Science 2013 [69]. My main contribution to this work was help during data collection, mostly by online data analysis for fast feedback to guide decisions during the experiment and final data processing from hitfinding to initial molecular replacement solution and refinement.

5-Hydroxytryptamine receptor 2B (5-HT<sub>2B</sub>), also known as serotonin receptor 2B, is a class A GPCR with the main function of mediating the effects of the neurotransmitter serotonin and is of high pharmacological interest, because of a variety of target areas. 5-HT<sub>2B</sub> is abundant in the central-, peripheral nervous and cardiovascular system. One of its most important roles is the influence on behavior and sensitization to tactile stimuli by presynaptic inhibition and is hence target of recreational drugs, such as LSD. Its cardiovascular function includes pulmonary vasoconstriction (causing pulmonary hypertension) and regulation of the cardiac structure.

The very first successful example of the novel technique of LCP-SFX was the determination of the structure of the 5-HT<sub>2B</sub>receptor. The first structure was solved just shortly before the experiments at the LCLS by multi-crystal MX, allowing the

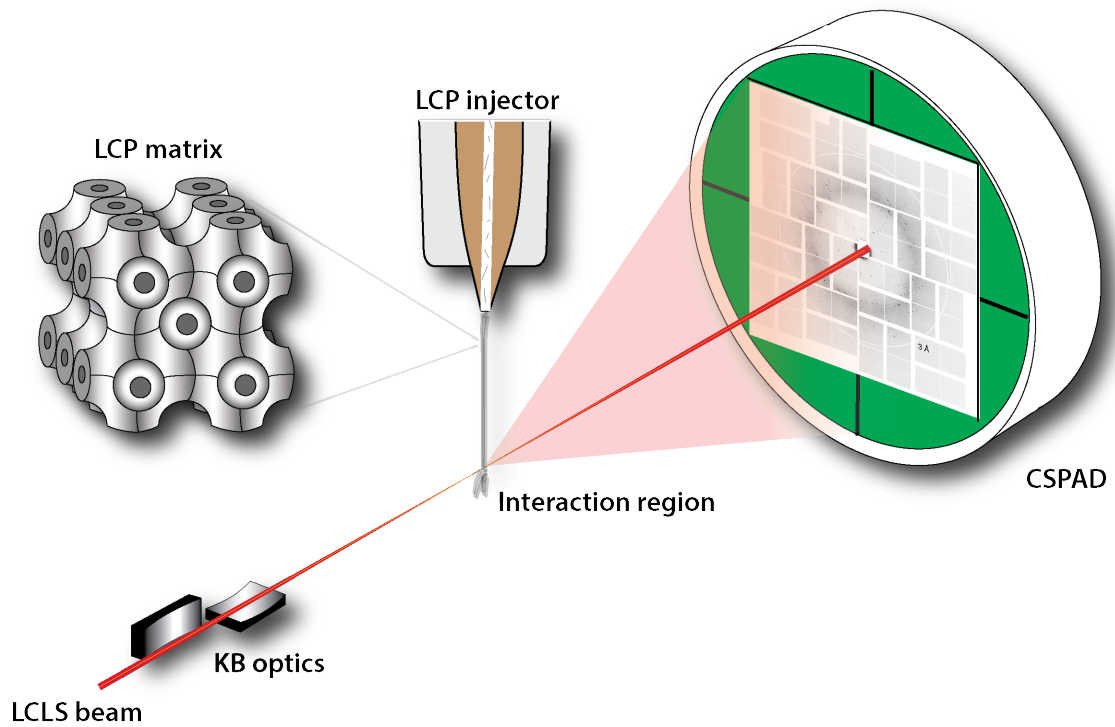


Figure 6.2: Illustration of the experimental setup of a LCP-SFX experiment. LCLS beam propagating from left to right, focused by KB optics and interact with the LCP jet within the interaction region. Diffraction patterns are recorded on the CSPAD detector.

validation of the obtained results.

**Experimental details** Protein expression, purification and crystallization was performed by the Stevens and Cherezov lab at The Scripps Research Institute, La Jolla, CA, USA ([69] Supplementary Material). Briefly, Sf9 cells were infected with a recombinant baculovirus encoding a construct of 5-HT<sub>2B</sub>, containing a thermostabilized BRIL domain in ICL3. 5-HT<sub>2B</sub> and its ligand ergotamine (ERM) were incubated for complex formation, extracted using DDM and cholesterol hemisuccinate, purified (IMAC) and prepared for crystallization trials by buffer exchange and concentrated to 20 mg/ml. The purified protein was reconstituted in 9.9 MAG with 10% cholesterol, mixed with precipitant solution and the final crystallization was set up in Hamilton syringes at room temperature. Sample quality and crystal size was monitored by UV-fluorescence imaging, SONICC and optical microscopy. Overall, a total of 100  $\mu$ l of sample was used for the experiment, corresponding to approximately 0.3 mg of the purified protein.

Data collection of the 5-HT<sub>2B</sub> receptor was performed at the Coherent X-ray Imaging beamline [57] at the LCLS, SLAC, USA. X-ray pulses were tuned to a photon energy of 9.5 keV, with a pulse duration of  $\sim 50$  fs and a total of  $8 \cdot 10^{11} \frac{\text{photons}}{\text{pulse}}$ , focused to a beam with  $1.5 \mu\text{m}^2$  FWHM, using X-ray optics in a Kirkpatrick-Baez geometry [48]. LCP containing 5-HT<sub>2B</sub> microcrystals, ranging in size from 1-10  $\mu\text{m}$  in the longest dimension, with an average diameter of 5  $\mu\text{m}$ , were mixed with 7.9 MAG to prevent a phase transition of 9.9 MAG at 18°C and loaded into the LCP injector as batches of 20  $\mu\text{l}$ . The injector was set to a flow-rate between  $50 - 200 \frac{\text{n}\ell}{\text{min}}$ , with a co-axial flow of nitrogen between 300-500 psi to stabilize the jet. Single shot diffraction patterns of microcrystals in random orientation were collected with a repetition rate of 120 Hz on a CSPAD [73]. The sample to detector distance was set to 100 mm, allowing a maximum of 1.7 Å resolution data in the corners of the detector. Due to the use of comparably large crystals, saturation effects were limiting the beam transmission, which was attenuated to 3-6%, corresponding to about  $3 \cdot 10^{10} \frac{\text{photons}}{\text{pulse}}$  with an applied radiation dose to the sample of maximum 25 MGy. A total of 4.217.508 detector frames were collected in approximately 10 hours of beamtime. Initial pre-processing and Bragg peak identification was performed by Cheetah [78], where a minimum of 15 peaks per frame was used as a threshold value to define a ‘hit’. A total of 152.651 hits were obtained, which corresponded to an average hit-rate of 3.6%, while the latter was fluctuation between 0.7% and 24%, depending on both sample quality and the jetting conditions of the LCP injector (data not shown). Subsequent autoindexing, integration and merging was performed

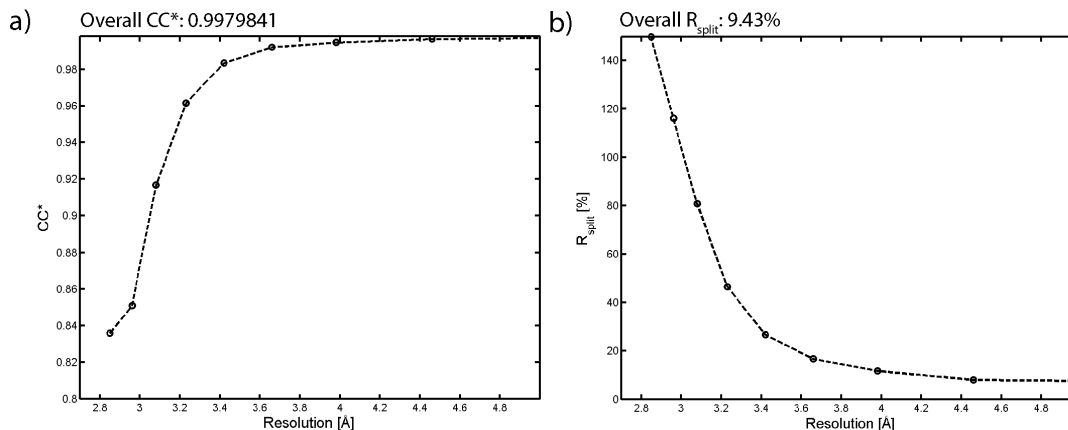


Figure 6.3: CC\* and R<sub>split</sub> plots for 5-HT<sub>2B</sub>-SFX (from Liu et al. 2013)

with CrystFEL [79] (version 0.5.0+5143ea), using the autoindexing algorithms of mosflm [81] and DirAx [83]. A total of 32 819 diffraction patterns were successfully indexed, while the majority of the data collected originated from one single batch (22.109 indexed patterns) within a total of 2.5 hours of data collection time.

By the Monte Carlo integration step implemented in CrystFEL, a three-dimensional dataset of structure factors was merged into a set of hkl intensities. The resolution of the 5-HT<sub>2B</sub>-SFX dataset ranged from 35 to 2.8 Å (compared to 2.7 Å of the 5-HT<sub>2B</sub>-MX structure) covering a total of 16.052 unique reflections. An appropriate resolution cutoff was chosen by analyzing the Pearson correlation coefficient [27], as well as monitoring  $R_{work}$  and  $R_{free}$ . The consistency of the data was underlined by an overall  $R_{split}$  of 9.5%, a  $CC^*$  of 0.998 with 100% completeness over the entire data range. The space group of the LCP grown microcrystals was determined as  $C222_1$  given the present unit cell constants and systematic absences along the c-axis, which is identical to the synchrotron structure. Structure factors were used to calculate a molecular replacement solution using the 5-HT<sub>2B</sub>-MX structure as a search model (PDB: 4IB4) using PHASER MR [100], followed by several rounds of refinement using phenix.refine [127] to a final  $R_{work} / R_{free} = 22.7 \% / 27.0\%$ . A summary of the data collection and crystallographic figures of merit can be found in table 6.1.

**Structural analysis, comparison and discussion** The electron density of the 5-HT<sub>2B</sub>-SFX receptor, and most importantly the ligand ergotamine, is overall very well defined and the obtained model is similar to the 5-HT<sub>2B</sub>-MX structure (overall

6.3. FIRST STRUCTURE OF HUMAN MEMBRANE PROTEIN AT ROOM TEMPERATURE 95

Table 6.1: Crystallographic statistics of the 5-HT<sub>2B</sub>-SFX-ERG structure (from Liu *et al.* 2013)

<b>Data collection</b>	<b>5-HT<sub>2B</sub>-SFX</b>
PDB ID	4NC3
Temperature, K	294
Wavelength, eV	9500
Beam size, $\mu\text{m}^2$	1.5
Average crystal size, $\mu\text{m}^3$	5 x 5 x 5
Flux	$3 \times 10^{10}$ photons/pulse
Max dose per crystal, MGy	25
Space group	C222 <sub>1</sub>
Unit cell, Å	61.5, 122.2, 168.5
Pulse duration, fs	50
No. collected frames	4,217,508
No. hits	152,651
No. indexed patterns	32,819
No. total / unique reflections	18,515,376 / 16,052
Resolution range, Å	35 - 2.8
Completeness, %	100 (100)
Multiplicity	1.150 (1035,6)
I/ $\sigma$ (I)	5.9 (0.64)
CC*	0.998 (0.74)
R <sub>split</sub> , %	9.5 (161.9)
<b>Refinement</b>	
No. reflections / test set	16,025 / 814
R <sub>work</sub> / R <sub>free</sub> , %	22.7 / 27.0
No. atoms	
Receptor / BRIL	2,856
Ligand	43
Lipids and others	224
B-factors, Å <sup>2</sup>	
Wilson B/ Overall B	115.7 / 98.7
Receptor / BRIL	88.4 / 133.7
Ergotamine	68.1
Lipids and other	110.4
R.m.s bonds, Å / angles, °	0.002 / 0.60
Ramachandran plot stats, %	
Favored	96.4
Allowed	3.6
Disallowed	0

Table 6.2: Crystallographic statistics of the 5-HT<sub>2B-SFX</sub>-ERG structure over 10 resolution shells with equal number of reflections.

Completeness [%]	No. measurements	Multiplicity	SNR	Resolution shell [ave. Å]
100	2336945	1407.0	15.71	10.44
100	2159750	1368.7	11.68	5.33
100	1983813	1273.3	11.75	4.46
100	1623737	1053.7	7.15	3.98
100	1799995	1169.6	4.92	3.66
100	1784933	1156.8	3.31	3.42
100	1573247	1032.3	1.98	3.23
100	1602345	1046.6	1.15	3.08
100	1574594	1035.2	0.87	2.96
100	1495880	979.6	0.66	2.85

0.68 Å RMSD). However, after thorough comparison of both structures, there are some significant differences. The volume of unit cell of the 5-HT<sub>2B-SFX</sub> structure (a = 61.5, b = 122.2, c = 168.5 Å) is 2.1% larger than the cryo-cooled synchrotron structure (a = 60.57, b = 119.75, c = 170.61 Å), with an expansion along the a- and b- axis, while a slightly shorter c-axis was observed. This might be explained by the cryo-cooled state of the crystals in conventional experiments, where the a-b plane perpendicular to the lipidic membrane allows translational motion of the receptor at room temperature. At the same time the BRIL fusion domain is rotated towards the receptor by ~2.5°, leading to a slightly altered packing of the unit cell. It is a common phenomenon that protein crystals show different unit cell dimensions upon flash-freezing, which is a frequently faced problem in merging data from several crystals, due to induced anisomorphism [128][129]. Hence, room-temperature measurements often allow a more consistent merging of data [130], but are typically limited in usage, due to an increased susceptibility to radiation damage in conventional crystallography experiments.

Significant structural differences, as described by Liu et al., between the two obtained structures are the following:

- Several side-chains are fully resolved in the 5-HT<sub>2B-SFX</sub>, which are absent in 5-HT<sub>2B-MX</sub>, potentially due to a lack of radiation damage (Figure 6.3.0.1 a and 6.3.0.1)
- The overall difference in distribution of relative B-factors in 5-HT<sub>2B-SFX</sub>, thought to represent the dynamical nature of the receptor is closer to physiological conditions (Figure 6.3.0.1 a and b)
- Salt bridge between Glu319 and Lys247 links intracellular parts of helices

### 6.3. FIRST STRUCTURE OF HUMAN MEMBRANE PROTEIN AT ROOM TEMPERATURE 97

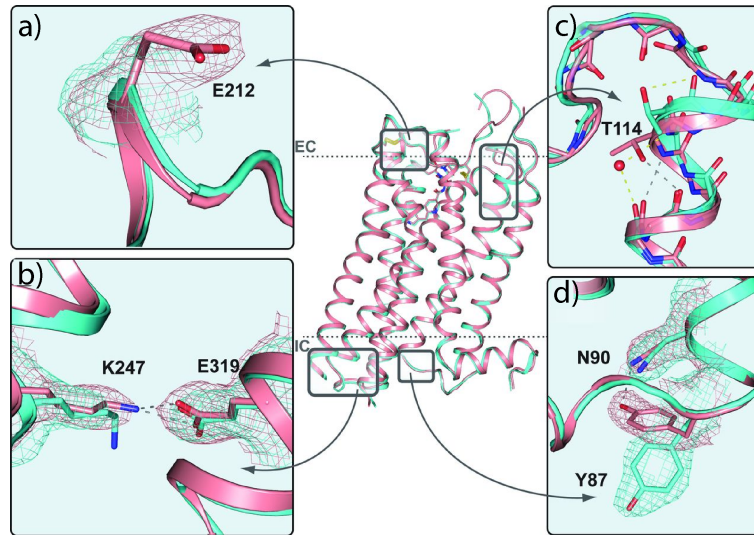


Figure 6.4: Comparison of the model of 5-HT<sub>2B</sub>-SFX (PDB: 4NC3) and 5-HT<sub>2B</sub>-MX (PDB: 4IB4). a) Glutamate212 side-chain visible in SFX structure b) Salt bridge between Lys247 and Glu319 visible in SFX structure c) Conformational difference of ECL2 d) Rotamer difference of Tyr87 and formed salt bridge in SFX structure (from Liu *et al.* Science 2013).

V and VI in 5-HT<sub>2B</sub>-SFX, which are involved in receptor activation (Figure 6.3.0.1 b)

- Extracellular tip of helix II forms a regular alpha-helix in 5-HT<sub>2B</sub>-SFX (Figure 6.3.0.1 c)
- Tyr87 forms hydrogen bond with Q90 and has generally different rotamer (Figure 6.3.0.1 d) in 5-HT<sub>2B</sub>-SFX

Analyzing the relative B factor distributions of the 5-HT<sub>2B</sub>-SFX model, structural dynamics of certain areas seem more likely representing the dynamical flexibility of the receptor. As one would expect, the transmembrane domain of the receptor is more rigid than the loop regions in both structures (Figure 6.3.0.1 a and b). In the SFX model, the transmembrane helices in direct contact with the ligand have the lowest B-factors, while all other domains are slightly more flexible. Higher B-factors are most pronounced in the N-terminus, ICL2, ECL1 and a part of ECL2. None of these highly flexible domains are directly interacting with the ligand, but might be influencing the binding kinetics of the ligand to the receptor, as e.g. ECLs act as a lid to prevent ligands from dissociation.

One significant difference of the B-factor distribution between the two structures



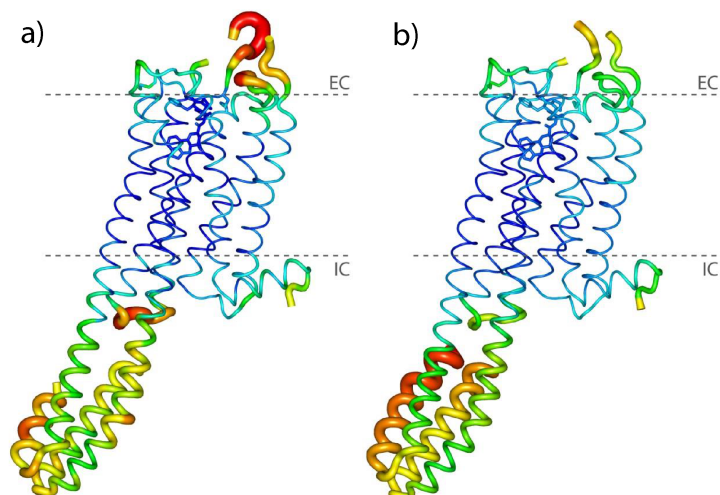


Figure 6.5: Cartoon illustration of B-factor distribution of the 5-HT<sub>2B</sub> structures with a relative scale from green ( $B = 100 \text{ \AA}^2$ ) to red ( $B = 200 \text{ \AA}^2$ ) b) B-factor distribution of the 5-HT<sub>2B</sub>-MX structure with a relative scale from green ( $B = 100 \text{ \AA}^2$ ) to red ( $B = 200 \text{ \AA}^2$ ) (from Liu *et al.* Science 2013)

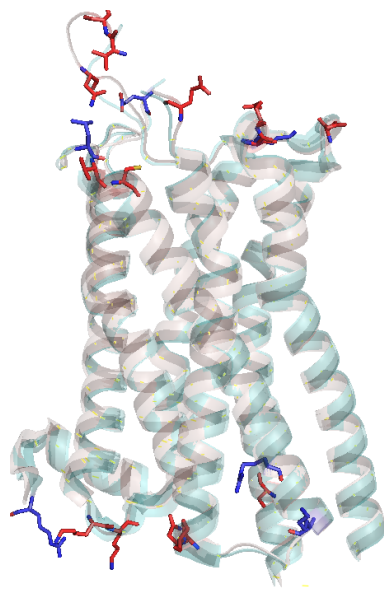


Figure 6.6: Stick representation of the side-chains resolved in one of the two structures (red - SFX, blue - MX)

regarding the ligand is the phenyl-group of ergotamine. In the SFX structure, the average B-factor of the phenyl group is 78,86 Å<sup>2</sup>(SYN 61,06), compared to 65,94 Å<sup>2</sup>(SYN 57,06) of the remaining ligand. These increased B-factors might be due to the (hydrophobic) interactions with flexible residues in the N-terminal ends of loop V and VI, as well as ECL2 [131] (Figure 6.3.0.1 e). By replacing this phenyl group with a chemical group, which interacts more strongly with the receptor, e.g. by allowing additional hydrogen bonds, an increased affinity and specificity of the ligand might be achieved. A comparison of the inter-atomic distances between ergotamine and 5-HT<sub>2B</sub> in the binding pocket with the online tool SimiCon [132] with a threshold of distances < 4Å shows an overall very similar binding mode. The SYN structure shows 96 interactions with an average inter-atomic distance of 2.88 Å, while the SFX structure presents only 75 interaction points with an average of 3.05 Å distance between the ligand and the receptor, with only 60 interaction points overlapping.

In addition to the observations of dynamical flexibility of the structure, a total of thirteen residues are more completely resolved in the 5-HT<sub>2B-SFX</sub> and six in the 5-HT<sub>2B-MX</sub> structure (Figure 6.3.0.1). The residues, which are better resolved in the 5-HT<sub>2B-SFX</sub> are located at the surface of the receptor (both intracellular and extracellular), while the residues better resolved in the 5-HT<sub>2B-MX</sub> structure are buried.

The additional salt bridge observed between Lys247 and Glu319 represents an additional link between helix V and VI. In the SYN structure, Lys247 forms a salt bridge with the BRIL domain, which is clearly artificial. Wacker and coworkers [133] suggested a mechanism for arrestin biased signaling, where ergotamine locks helix V/VI in an inactive-like state, while III/VII induce arresting signaling [133]. It has been observed in the previous synchrotron structure, that Phe333 of the PIF motif and the conformation of helix V/VI, are adopting an inactive state-like conformation. This newly resolved salt bridge in the SFX structure appears to lock the receptor in the inactivated state, which has to be disrupted upon receptor activation. Both the conformation of Phe333 and helix V/VI, might be induced or preserved by the salt bridge between Lys247 and Glu319.

The extracellular domain, including ECL2 and nine better defined residues are not directly interacting with the ligand, but might be involved in binding kinetics, as they might act like a lid closing the barrel. On the other hand, the intracellular domain with a more dynamic ICL2 and four better resolved side-chains are giving more detailed insights into the binding pocket of the transducer molecules (G-protein or arrestin). The residues better defined in the 5-HT<sub>2B-MX</sub> might indicate that cryo-cooling induces side-chains to be locked in certain rotamers, while these side-chains are more flexible at ambient temperatures and can adopt multiple conformations.

All differences between 5-HT<sub>2B-SFX</sub> and 5-HT<sub>2B-MX</sub> are likely caused by differences in thermal motion of the receptor, due to room temperature measurements and the lack of structural radiation damage in SFX experiments. All the above mentioned points are allowing a clearer understanding of the nature of the receptor. Overall it appears that the 5-HT<sub>2B-SFX</sub> structure gives a structure, which is closer to physiologically relevant conditions and might be more useful for follow-up studies, such as structure based drug design, because of multiple differences in the binding pocket and domains involved in receptor activation and downstream signaling.

## 6.4 First completely unknown structure solved by SFX

Due to a combination of the necessity of requiring relatively large volumes of samples and the beamtime application process being extremely competitive, most SFX experiments until now focused on either methodological development, reproduction of radiation-damage ‘free’ structures of known proteins, or pump-probe based experiments for the elucidation of structural dynamics in biomolecules. Structurally significantly novel insights were so far only seen in the structure of ProcathepsinB from *T. brucei*, the parasite causing the African sleeping sickness, where it was possible to model a density of a pro-peptide and glycosylation sites of the enzyme, which were not detectable in the *in vitro* crystallized synchrotron structure [64]. In the following chapter the structure of the human Angiotensin II type 1 receptor in complex with the selective antagonist ZD7155 will be discussed, representing not only a pharmacologically very important target molecule, but the first completely unknown structure solved by SFX so far.

### 6.4.1 Structure of the human Angiotensin II - receptor 1 (AT<sub>1</sub>R)

The following experimental results are described in the publication of Zhang *et al.*, which was published in Cell in 2015 [70]. My main contribution to this research was the help during data collection, mainly focusing on online data analysis, as well as guidance of the experiment, followed by pre-processing (‘hitfinding’), indexing and merging, to initial molecular replacement solutions.

Cardiovascular diseases are the leading cause of death worldwide [124], even though diagnostics and treatments improve steadily. High blood-pressure (hyper-

tension) is the most common modifiable cardiovascular disease, though usually with treatable symptoms. As intuitively expected, it has been shown that the treatment of hypertension drastically decreases the risk of myocardial infarction, stroke, heart failure and nephropathy [134].

There are several pathways involved in hypertension, while the renin-angiotensin system (RAS) is representing the best studied one, since over-stimulation of the latter induces several cardiovascular disease related symptoms. The initiation of RAS stimulation is achieved by proteolytic degradation of precursor proteins: angiotensinogen is cleaved by **renin** to angiotensin I, which is cleaved by **angiotensin converting enzyme (ACE)** to angiotensin II, III and 1-7. Angiotensin II and III both act as vasoconstrictors, while most pathological and physiological effects of Ang II are mediated by **Angiotensin II receptor 1 (AT<sub>1</sub>R)**. Given this proteolytic cascade, it is not surprising that the three main groups of drugs, that are used in the treatment of hypertension are Renin inhibitors, ACE inhibitors and AT<sub>1</sub>R blockers (ARBs). As described previously, ligands binding to GPCRs are able to induce either both or either of the two transducer proteins (G-protein and/or arrestin). The G-protein dependent pathway is crucial for physiological processes, while it has been shown that drugs inducing an arrestin biased pathway show fewer side effects [120][121]. ARBs, or sartanes, are one of the most widely used drugs on the market, with about 600 mio. EUR sales in Germany in 2010. They show very few side-effects, due to the high specificity of these ligands to the receptor. The main indicators for sartanes are hypertension, cardiac hypertrophy, arrhythmia and heart failure. Despite its medical relevance, the structure of AT<sub>1</sub>R has not been solved until now. Many attempts of crystallization to sufficient size for data collection at synchrotrons failed mostly due to radiation damage, twinning of large crystals and a lacking electron density for the ligand in any structure. Hence, the binding mode of ligands and AT<sub>1</sub>R were not known, which was limiting our insights of AT<sub>1</sub>R function and control, as well as inhibited rational design and optimization of lead compounds for further improvements in the treatment of hypertension.

**Experimental details** The experimental details for the structure solution of AT<sub>1</sub>R by LCP-SFX experiment were almost identical to the ones described in section 6.3.0.1. Briefly, the X-ray beam was focused to 1.5 x 1.5  $\mu\text{m}^2$  FWHM, using the micro-focus X-ray optics at the CXI endstation. A pulse duration of nominal 36 fs, repetition rate of 120 Hz, photon energy of 7.9 keV and an average pulse energy of 2.7 mJ, at an average transmission of 16% resulted in 75 MGy maximum dose at the sample.

With a sample volume of 65  $\mu\text{l}$ , a total of 2.764.739 detector frames were col-

lected within 6.5 hours of beamtime. Preprocessing by Cheetah [78] with a threshold of minimum 15 Bragg peaks (min. SNR of 4) defining a 'hit', resulted in 457.275 positive 'hits', corresponding to an average hitrate of 17%. The detector geometry was refined using a recently developed algorithm [74], which compares the location between found and predicted peaks, in order to extrapolate possible deviations in the detector geometry. After further refinement of peak detection, CrystFEL [79] (version 0.5.3) was used for the indexing and integration step. A total of 73.130 diffraction images were successfully indexed and merged into a final set of hkl intensities. A newly implemented algorithm for discarding Bragg reflections beyond a certain resolution cutoff, as well as a threshold for minimum Pearson correlation coefficient between individual diffraction pattern and an initially merged set of intensities has been applied to remove noise and will be discussed later (Section 5.5.3).

Finally, a dataset with a resolution range of 32 to 2.9 Å, with a total of 11.190 unique reflections was assembled and used for structure determination. The resolution was estimated using the Pearson correlation coefficient approach, where the  $CC^*$  for the highest resolution shell was 0.872. An overall  $R_{\text{split}}$  of 9.5% and  $CC^*$  of 0.999 underlines the consistency of the dataset. All initial attempts for finding a molecular replacement solution failed. An alternative technique was used by firstly identifying six sequence homologues, taking the superposition of all and trimming poorly conserved regions. The original receptor structure, as well as the trimmed model of each, were used each together with the BRIL domain (PDB: 1M6T) as an input for PHASER MR [100]. Automated scripts were testing different combinations of search models with varying resolution cutoff, followed by rigid body refinement. The best solution was chosen according to the lowest  $R_{\text{free}}$  of 0.345, achieved with a search model using CCR5 (PDB: 4MBS), which was used for further refinement using Refmac5 and autoBUSTER. During the refinement, data processing parameters were constantly improved (peak detection, integration radii, filtering thresholds etc.) and the quality of the resulting datasets was evaluated by visual examination of the electron densities, e.g. of the ligand and crystallographic figures of merit. The model refinement was performed to a final  $R_{\text{work}} / R_{\text{free}}$  of 22.8 / 27.4. Several docking approaches, combined with biochemical assays, were performed in order to be able to expand the biochemical knowledge about the newly discovered binding pocket and the interaction with several other important ligands. Table 6.3 summarizes the data collection, as well as analysis statistics for the AT<sub>1</sub>R-ZD7155 structure.

**Structural insights of AT<sub>1</sub>R** The overall structure of AT<sub>1</sub>R, containing the canonical 7-transmembrane domain, has a certain similarity to other peptide receptors, namely chemokine (36% sequence homology to CXCR4) and opioid receptors

Table 6.3: Crystallographic statistics of the AT<sub>1</sub>R - ZD7155 SFX structure (from Zhang *et al.* 2015)

<b>Data collection</b>	<b>AT<sub>1</sub>R-ZD7155</b>
PDB ID	4YAY
Temperature, K	294
Wavelength, eV	7900
Beam size, $\mu\text{m}^2$	1.5 x 1.5
Average crystal size, $\mu\text{m}^3$	10 x 2 x 2
Flux	$1 \times 10^{11}$ photons/pulse
Max dose per crystal, MGy	75
Space group	C2
Unit cell, Å; $\beta$ -angle ( $^\circ$ )	72.8, 41.0, 167.7; 99,4
Pulse duration, fs	36
No. collected frames	2,764,739
No. hits	457,275
No. indexed patterns	73,130
No. total / unique reflections	14,415,424 / 11,190
Resolution range, Å	32.64 - 2.9
Completeness, %	100 (100)
Multiplicity	1.288 (215)
I/ $\sigma$ (I)	8.2 (0.84)
CC*	0.999 (0.872)
R <sub>split</sub> , %	9.85 (140)
<b>Refinement</b>	
No. reflections / test set	11.167 / 576
R <sub>work</sub> / R <sub>free</sub> , %	22.8 / 27.4
No. atoms	
Receptor / BRIL	3,077
Ligand	33
B-factors, Å <sup>2</sup>	
Wilson B	76.1
Receptor / BRIL	90.0 / 111.7
Ligand	61.0
R.m.s bonds, Å / angles, $^\circ$	0.010 / 0.95
Ramachandran plot stats, %	
Favored	96.1
Allowed	3.9
Disallowed	0

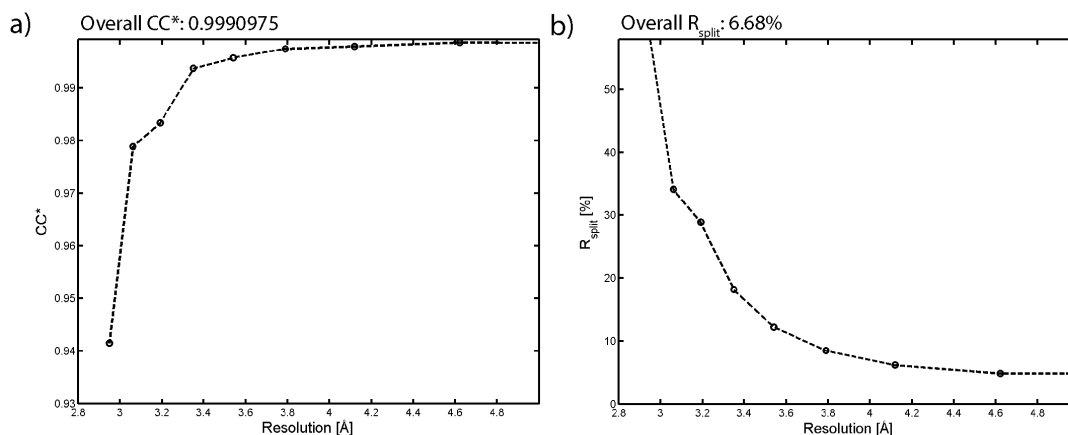


Figure 6.7: CC\* and R<sub>split</sub> plots of the AT<sub>1</sub>R - ZD7155 SFX structure (from Zhang *et al.* 2015)

Table 6.4: Crystallographic statistics of the AT<sub>1</sub>R - ZD7155 SFX structure

Completeness [%]	No. measurements	Multiplicity	SNR	Resolution shell [ave. Å]
100	4468423	7586.5	35.67	10.50
100	2552078	4761.3	20.70	5.52
100	2342577	4453.6	20.01	4.62
100	1641568	3114.9	15.59	4.12
100	1074567	2123.7	10.81	3.79
100	757820	1485.9	7.54	3.54
100	671832	1272.4	4.94	3.35
100	545047	1092.3	3.60	3.19
100	448409	879.2	2.81	3.06
100	324858	638.2	2.00	2.95

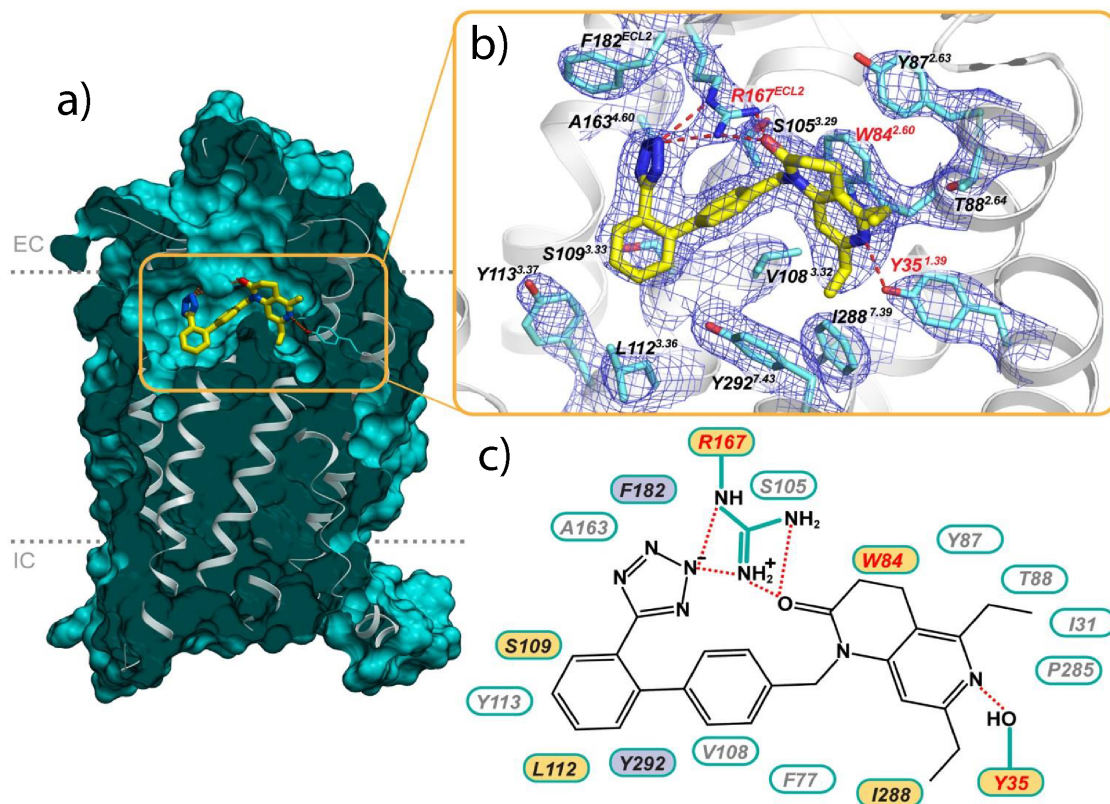


Figure 6.8: Illustration of the AT<sub>1</sub>R - ZD7155 SFX structure, highlighting the binding pocket and the protein-ligand interaction (from Zhang *et al.* 2015)

(33% sequence homology to  $\alpha$ -OR). However, structural differences in helices I, V, VI and VII are observable, comparing peptide receptors of different classes, whereas helices II and III have an almost identical conformation. Interestingly ECL2 of the AT<sub>1</sub>R has been shown to contain the target epitope of harmful autoantibodies in preclampsia and malignant hypertension, which could be exploited as a drug target in future research [135][136]. As in many other class A GPCRs, the a D(E)RY motif in helix III and the NPxxY motif in helix VII in AT<sub>1</sub>R are present, which are both known to be involved in receptor activation [137].

The binding pocket shows a complex network of residues interacting with the ligand ZD7155. Three residues being absolutely critical in the binding of the ligand: R167 in ECL2, W84 in helix II and Y35 in helix I. R167 interacts with the ligand through a hydrogen bond and is a unique residue compared to other structurally similar receptors, which might be a key residue for defining the selectivity and affi-



ity of certain ligands to this receptor. Residue Y35 interacts with the ligand through a hydrogen bond, too, while being well conserved amongst other peptide GPCRs. A mutation of this residue to alanine completely abolishes both the peptide, as well as the non-peptide binding capacity of this receptor. Finally, residue W84 interacts with the ligand through hydrophobic  $\pi$ - $\pi$  interaction, while being a conserved residue, too, and being essential for the ligand binding. A total list of fourteen different residues might be involved in ligand binding, of which four are well conserved amongst peptide receptors. Two residues are chemically similar to other conserved regions (hydrophobic residues). The remaining eight residues are completely unique compared to other peptide receptors, resulting in the special binding selectivity of AT<sub>1</sub>R ligands and a very low cross-reactivity of ligands.

To summarize, the AT<sub>1</sub>R structure obtained from this LCP-SFX experiment allowed the following insights:

- Detailed interaction of residues within the binding pocket and the ligand ZD7155
- General selectivity, correlating with low side effects of ARBs, can be explained
- The activation mechanism previously proposed was structurally confirmed (hydrogen bond between Asn111 - Asn295)
- Two targets for drug design: binding pocket for novel ARBs or ECL2, since target of auto-antibodies

Finally, the structure of a very important drug target has been solved with the help of the technique of LCP-SFX. Previous attempts using conventional MX only allowed crystal structures of less than 4 Å resolution, even after extensive screening, trying to improve the crystal size and quality. Additionally, the resulting electron density maps were all lacking a density for the ligand ZD7155, while being able to analyze the detailed interaction of the ligand with the receptor is of highest priority. For the first time in this relatively new field of serial femtosecond crystallography, it was possible to obtain structural information of a protein, which is of high biomedical relevance and which was not possible to achieve at a synchrotron light source.

## Chapter 7

# Serial synchrotron crystallography

Macromolecular crystallography (MX) is the most frequently used technique for the structure determination of biomolecules to atomic resolution, representing 90% of all deposited structures in the PDB. Beginning with the advance of 3<sup>rd</sup> generation synchrotron sources in the 1990s, tremendous efforts have been made to achieve the highest possible flux densities. X-ray beams can be focused to micrometer size, in order to match the X-ray beam to small crystal sizes, reducing background scattering necessary for small crystal data collection. All modern synchrotrons are equipped with a microfocus beamline for MX experiments, where microfocus means X-ray beams of smaller than 10  $\mu\text{m}$  FWHM readily available for the structural biology community. These achievements have pushed the limits of MX experiments, which allowed the structure determination of proteins from micrometer-sized crystals in several cases [38][115][122]. Nevertheless, structure determination of biomolecules by macromolecular crystallography is generally limited by the growth of crystals large enough for conventional MX using rotation techniques [13].

One difficulty of working with very small crystals is the detection of individual crystals in the mount, which is typically a nylon-loop or a kapton-based support. Currently, the most frequently used approach for very small crystals is rastering of the sample, which begins with a low-dose exposure following a predefined grid on the sample loop. By scoring individual detector frames by the maximum resolution or average strength of Bragg peaks over background, it is possible to localize crystals and derive a data collection strategy by automated centering on each of them. This difficulty is most prominent for LCP grown crystals, since the LCP matrix turns

opaque upon flash-freezing, making it difficult to impossible to detect even large crystals [42]. An alternative approach to rastering techniques is to obtain a low-dose X-ray radiograph of the sample loop, a technique that is particularly suitable for crystals grown in LCP [138].

In principle it is possible to obtain high resolution diffraction patterns from well-ordered microcrystals with a suitable setup, but a major limitation of diffraction experiments from very small crystals is radiation damage. Due to unwanted effects during experiments, namely inelastic and absorption events, energy is deposited into the crystal. This deposited energy initiates chemical reactions, e.g. radiolysis of water, which damage the crystal and ultimately lead to a loss of crystallinity (Section 3.5). This loss in crystallinity leads to a loss of obtainable information during the exposure, most importantly high resolution diffraction. Hence, a direct correlation between the applied dose and a loss of information content and quality of the diffraction data can be derived (Henderson limit 3.5.1). To limit radiation damage by suppressing thermal motion, the method of cryo-crystallography was introduced in the 1990s, where the sample is flash-frozen and cooled by liquid nitrogen during data collection. However, flux densities currently achievable at microfocus beamlines are already far beyond the point, where radiation damage would not be an issue, preventing the collection of a full dataset from individual crystals. Usually the X-ray beam is either strongly attenuated or defocused (or a combination of both) in order to match the X-ray spot size to the crystal, as well as to reduce the applied X-ray intensity to the crystal.

To overcome the bottleneck of a limited scattered intensity by radiation damage, beginning in the earliest days of macromolecular crystallography, partial datasets were obtained and merged from multiple crystals. By this, the dose is distributed, in order to keep the quality and therefore the information content, of the diffraction pattern as high as possible. Several approaches for the optimal data collection by collecting partial datasets from multiple crystals [139] and multiple, separated parts of one large crystal [59] have been designed in the recent past. These approaches are becoming more frequently used and data collection schemes, detectors and data analysis are steadily improving. As mentioned earlier, the major problem of multi-crystal approaches is the merging of anisomorphous data, which is still not resolved and often limits experiments. The probability of obtaining a sufficient number of isomorphous partial datasets decreases with the degree of anisomorphism.

Most of the above mentioned problems of using small crystals, such as radiation damage, crystal detection and handling, are overcome by the method of serial

femtosecond crystallography (SFX), which makes use of the unique properties of hard X-ray Free-Electron Lasers (Chapter 5). However, due to the design of currently available XFELs, the X-ray beam can only be delivered to one (or a few) user group(s) at a time. Beamtime applications are hence extremely competitive, since currently only two hard XFELs are available (SACLA and LCLS), of which three beamlines are available for bio-imaging experiments. Additionally, the running cost of XFEL beamtime is very high, compared to other light sources, with an annual budget of 128 mio. USD in 2012 for the LCLS.

This following chapter describes two experiments, which made use of two of the most powerful microfocus beamlines currently available for macromolecular crystallography - P14 and P11 at PETRAIII. By combining high flux densities, micrometer sized beams, high-precision goniometer and state-of-the-art, shutter-less data collection using pixel-array detectors, it was possible to obtain high resolution structures of one of the smallest protein crystals published to date. Here, I will describe these first two successful experiments of serial crystallography using synchrotron radiation [106][140].

## 7.1 First successful serial synchrotron crystallography experiment

This chapter describes the structure solution of *in vivo* grown microcrystals of *T. brucei* ProcathepsinB (Section 7.1.1), solved during the first successful experiment of serial crystallography using synchrotron radiation, which was published in Gati *et al.* 2014 in *IUCrJ*. My contribution to this experiment was beamtime proposal, experimental layout, sample characterization, data collection, data processing and analysis, up to initial molecular replacement solutions and refinement.

African trypanosomiasis (African sleeping sickness) is a parasitic, infectious disease, caused by the protozoa *Trypanosoma brucei*, which infects humans or other animals. The disease mostly occurs in sub-Saharan Africa (80% in Congo) with about 9000 deaths in 2010 (as opposed to 34000 in 1990). The most common way of infection is a bite of infected tse-tse flies (*Glossina morsitans*). The final stage of the untreated disease is constant lethargy and fatigue, which is the eponym of the disease. Currently available drugs are available for the disease, while their molecular mechanism is mostly unknown. Additionally, these treatments are limited in efficacy, some of which have severe side effects and as with any other antibiotic, resistances appeared to all currently available drugs. Given that only few drugs are available,

it is becoming an increasingly severe problem for the treatment [141]. Hence, novel compounds specifically inhibiting essential pathways of the parasite are urgently required. One promising target for the treatment of trypanosomiasis are lysosomal, papainlike cysteine proteases, which are crucial for host-protein degradation during the uptake of the parasite into the cell [142]. The knockout of the gene encoding for *T. brucei* procathepsinB (TbCatB) in a mouse model by RNA interference led to the clearance of infected mice from the parasite, highlighting it as a potential target for the treatment of the disease [143]. Cysteine proteases are generally synthesized in an inactive form, where the propeptide (initial amino acids at the N-terminus) are binding at the active site and act as a native inhibitor of the enzyme. After uptake of the latter into the lysosomal compartment, the drop in pH releases the propeptide and the enzyme is present in its active, mature form, degrading host cell proteins [144]. Hence, the structural knowledge of this enzyme is of great importance for biomedical purposes.

One very interesting observation during the overexpression of TbCatB in Sf9 insect cells was that the expressing cells formed intracellular microscopic needles, which were accidentally detected using a microscope. Subsequent investigation of the cells with scanning- and transmission electron microscopy supported the theory that these microstructures were indeed microcrystals grown inside the Sf9 cells, given their very regular facets. Even though *in vivo* crystallization was observed before, this was one of the first pieces of evidence that recombinantly expressed proteins can form crystals heterologously [145], since earlier studies involved e.g. polyhedrin microcrystals which grow natively in insect cells (Section 3.6.1.2). The first structure of mature TbCatB was initially solved by re-solubilizing the *in vivo* grown microcrystals to obtain a saturated protein solution, followed by *in vitro* re-crystallization of TbCatB for conventional MX experiments. In parallel, powder diffraction experiments confirmed the crystalline order of these microcrystals, even though the resolution of those powder diffraction patterns was limited.

The first high resolution structure solved from *in vivo* grown TbCatB microcrystals was published by Redecke *et al.* in Science 2013. By using the method of serial femtosecond crystallography, it was possible to obtain a structure from one of the smallest protein crystals ( $\sim 9 \mu\text{m}^3$  average volume) to date. The most important difference between the *in vitro* and *in vivo* crystallized structures was an additional density in the substrate binding cleft, which was identified as the native propeptide of 62 amino acids. This shows the inactive premature form of the enzyme. Additionally, extra densities were found for two glycosylation sites (Asn58 and Asn 216). Glycosylation sites are very rarely visible in protein crystallography, due to usually very high heterogeneity of these chemical modifications. The glycosylation

sites were also present in the *in vitro* recrystallized protein.

Inspired by the technique of SFX and the successful solution of TbCatB, with the limitation of competitive beamtime proposal reviews, we attempted to solve the structure of TbCatB using a cryo-loop containing a solution of *in vivo* grown microcrystals at a microfocus synchrotron beamline. Given the very small size of the crystals, homogenous size distribution and the available structure, TbCatB microcrystals posed an ideal model system for our investigation.

### 7.1.1 Recombinant *in vivo* crystallization of *T.brucei* procathepsinB

*In vivo* crystallization in *Spodoptera frugiperda* (Sf9) insect cells is achieved by replacing the polyhedrin gene on a baculovirus plasmid, with a gene of interest. The crystallization process is facilitated by the constitutively active promoter of polyhedrin, yielding very high local concentrations of the recombinantly expressed protein, most likely in specific cellular compartments (e.g. lysosome or peroxisome), which is a prerequisite for crystallization. Naturally occurring *in vivo* crystallization is monitored through changes in the ionic environment, specific proteolysis or binding partners. *In vivo* crystallization of recombinantly expressed proteins has been used to crystallize polyhedrin lacking a nucleocapsid [38], fragment of polyhedrin attached to a protein of interest [146], as well as two polyhedrin-free subunits of calcineurin [147], of which some were considered to be too small for X-ray crystallographic experiments. Koopmann *et al.* [145] have investigated the *in vivo* crystallization of recombinantly expressed (non-polyhedrin) proteins can lead to crystal sizes large enough for structure determination using X-ray Free-Electron Lasers. They have presented two pharmacologically relevant proteins derived from the parasite *T. brucei*, the parasite causing African sleeping sickness in humans, namely Inosine-5'-monophosphate dehydrogenase (IMPDH) and procathepsinB (TbCatB), of which only the latter one will be discussed in the following.

A detailed protocol for *in vivo* crystallization can be found in the Supplementary Information of Koopmann *et al.* [145]. The gene encoding TbCatB was inserted into a pFastBac1 expression plasmid and transformed into DH10Bac-competent *E. coli* strains, producing infectious viruses for Sf9 cell transformation. The gene encoding polyhedrin, as well as the polyhedrin nuclear import signal peptide were replaced by the TbCatB gene, including a new signal peptide sequence. Sf9 cells were transformed with Bacmid DNA by lipofection, producing a recombinant virus stock for further infection of Sf9 cells and final crystal production. The virus stock was used

to infect a monolayer of Sf9 insect cells (0.1 p.f.u. per cell), incubated for 72-96 hours. After harvesting the cells, the cells were lysed by vortexing, protein crystals were isolated similar to CpGV, where a mild detergent (SDS) was applied, followed by gradient centrifugation steps for high purity samples. In vivo grown microcrystals had an average size of  $1 \times 1 \times 9 \mu\text{m}^3$  and were characterized by SEM and powder diffraction prior to the experiment.

### 7.1.2 Experimental setup and data collection

The final crystal concentration was estimated to  $5 \times 10^8 \text{crystals/ml}$  and prior to data collection was supplemented with 50% (v/v) glycerol as cryo-protectant. The crystal solution was mounted in a standard nylon-loop (Hampton Research, USA) with 0.7 mm diameter and  $20 \mu\text{m}$  thickness, which ultimately influences the final sample thickness and meniscus. An estimated 13 nl were mounted in the loop, corresponding to an estimated total of about 5000 crystals available for data collection.

The P14 beamline at PETRAIII is a fully dedicated microfocus beamline for Macromolecular Crystallography and is equipped with one of the most sophisticated microfocus setups currently available. The U29 undulator was tuned to a photon energy of 10 keV, with the smallest achievable beam size of  $5(h) \times 4(v) \mu\text{m}^2$  FWHM ( $1.2 \times 10^{12} \text{photons/s}$ ) achieved by Rh-coated adaptive bimorph mirrors in KB geometry [48]. This photon energy was chosen to allow a high photon flux, high sensitivity of the detector, while limiting the power load on the double crystal monochromator (DCM, Si 111), ultimately minimizing the variation in beam position and intensity over the long experimental time of 8 hours. The crystal containing nylon-loop was mounted with a magnetic base onto a MK3 mini-kappa goniometer head attached to an MD3 microdiffractometer (ARINAX, Moirans, France), with a downward pointing vertically mounted omega axis allowing very high accuracy with a sphere of confusion on the order of 100 nm. The sample is cooled to 110 K with a liquid nitrogen stream. (Cryojet, XL, Oxford Instruments, England). Diffraction data were collected on a Pilatus 6M-F detector in shutter-less mode at 25 Hz (DECTRIS Ltd., Baden, Switzerland).

During an initial phase we investigated the diffraction properties, as well as an adequate cryo-protectant concentration for the experiment. It appeared to us that microcrystals in general tend to nucleate ice formation and first attempts using regular cryoprotectant concentration between 20-30% were not successful. In this particular case, remaining cell debris from the *in vivo* crystallization process might have favored the ice formation by heterogeneous nucleation. Finally, it was necessary to increase the final glycerol concentration in the loop to 50 % (v/v), in order to avoid

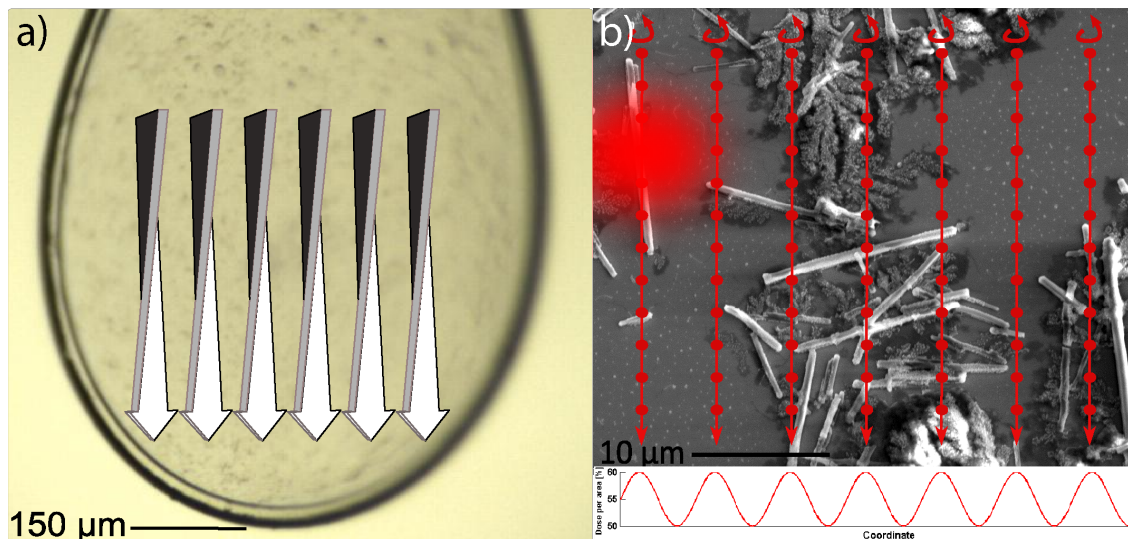


Figure 7.1: Macroscopic (left) and microscopic (right) illustration of the data collection schemes used for the structure solution of TbCatB. Red flare illustrates the beam as a 2D gaussian with a FWHM of  $5 \times 4 \mu\text{m}^2$ . Red dots represent the spacing between exposures. The plot in the lower right corner illustrates the distributed maximum dose per area, ranging between 50 and 60 MGy. (from Gati *et al.* IUCrJ 2014)

crystalline ice. It is currently unknown to us how these high glycerol concentrations affect the quality of the crystals, but it is possible that it has a negative impact on the obtained diffraction patterns.

After 2 s exposure times with  $1^\circ$  oscillation after manually centering the X-ray beam on a crystal, it was possible to obtain diffraction pattern with visible Bragg peaks beyond  $3\text{\AA}$  resolution. By short test rotation series, visual inspection of the collected data and being able to index the short oscillation series appropriately using XDS [84], we estimated the lifetime of a crystal in the beam of about 0.5 - 1 s. Processing of several datasets was successful, while fine-sliced series collected with less than 0.5 s exposure (or more than 1 s total exposure), always failed. This experimentally estimated lifetime matched the predicted lifetime using the software RADDOSE [148], assuming an estimated dose rate of  $34 \text{ MGy s}^{-1}$  with homogenous crystal illumination with a photon flux density of  $5.3 \times 10^{10} \text{ photons/s}/\mu\text{m}^2$ .

Cell debris in solution were preventing the detection of a sufficient number of crystals for successful structure solution by conventional multi-crystal approaches. Therefore we were motivated to use an alternative data acquisition approach, by



applying a series of helical line scans (Figure 7.1). By simultaneously rotating and translating the sample loop through the beam, we attempted to probe the largest volume containing crystals, while distributing the dose as homogeneously as possible. The major advantage of this approach is the possibility to collect short rotation series from randomly distributed and oriented crystals. By this, in contrast to ‘still’ snapshots from other serial crystallography experiments, it is possible to fully observe Bragg reflections (Section 2.7, Figure 2.6). This enables the usage of software packages developed for conventional MX data. The combination of using these well established programs with less partial data gives much more accurate structure factor estimates allowing complete datasets from much fewer crystals.

Our final data collection scheme covered a region of  $600 \times 600 \mu\text{m}^2$ , where a series of 120 parallel helical line scans with a spacing of  $5 \mu\text{m}$  was applied. The spacing of  $5 \mu\text{m}$  matches the FWHM of the X-ray beam, which was chosen to keep the applied dose as homogenous as possible over the entire exposed area, while keeping the applied dose to the crystals in the following line scan to a minimum. Each helical scan covered a range of  $90^\circ$  ( $\pm 45^\circ$  relative to the aligned flat loop), 240 exposures of 1 s each, with a total translation of  $600 \mu\text{m}$ . Therefore, each diffraction pattern corresponded to a rotation of  $0.375^\circ$  with a translation of  $2.5 \mu\text{m}$  of the sample. Our final estimate of the maximum radiation dose per frame was 50-60 MGy, by trying to optimize the highest achievable resolution. However, further detailed analysis and experimental evidence will be necessary to support and justify this high dose. In fact in subsequent experiments we confirmed that an exposure time beyond 50 ms for  $4 \mu\text{m}$  lysozyme microcrystals is significantly damaging crystals in the following line scan. Finally, 28 800 detector frames were acquired over a period of 8 hours of total beamtime.

### 7.1.3 Data processing, analysis and quality assessment

The nature of this data collection scheme enforced a significantly different data processing approach compared to conventional MX. Protein crystals, embedded in vitreous ice, are in random orientation. However, due to the needle-shape of the crystals, their angle relative to the plane of the aligned flat loop is limited. Because of the random distribution, not every detector frame contained a diffraction pattern. This would not be accepted as an input by conventional MX data reduction software. Additionally, the relative position of X-ray beam and each crystal will be different, which will result in a broadened spectrum of radiation dose applied to each crystal, leading to higher variances in SNR and ultimately a higher  $R_{\text{merge}}$  value, comparable to the description in section 5.3.2.3. Another result from the random orientation

Table 7.1: Data collection and crystallographic statistics of the TbCatB structure (from Gati *et al.* IUCrJ 2014)

<b>Data collection</b>	<b>TbCatB</b>
PDB ID	4N4Z
Light source, beamline	PETRA III, P14
Wavelength, eV	10000
Beam size, $\mu\text{m}^2$	4 x 5
Average crystal size, $\mu\text{m}^3$	1 x 1 x 9
Flux	$1.2 \times 10^{12}$ photons/s
Max dose per crystal, MGy	50-60
Space group	P4 <sub>2</sub> 2 <sub>1</sub> 2
Unit cell, Å	a = b = 123.5, c = 54.3
Exposure time, s	1
No. collected frames	28.800
No. indexed patterns, CrystFEL	2233
Final no. indexed patterns, XDS	426
No. unique reflections	8881
Resolution range, Å	88.1-3.0 (3.16-3.00)
Completeness, %	99.8 (99.9)
Multiplicity	12.3 (12.6)
I/ $\sigma$ (I)	3.7 (1.0)
CC*	0.97 (0.79)
R <sub>merge</sub>	0.71 (2.69)
<b>Refinement</b>	
No. reflections / test set	8484 / 399
R <sub>work</sub> / R <sub>free</sub> , %	22.3 / 26.4
No. atoms	
Protein	2392
Carbohydrate	67
B-factors, Å <sup>2</sup>	
Protein (main chain/side chain)	38/43
Carbohydrate	54
R.m.s bonds, Å / angles, °	0.01 / 1.32
Ramachandran plot stats, %	
Favored	91.2
Allowed	8.2
Disallowed	0.66

and size distribution of the crystals is the fact that each ‘wedge’ (short series of subsequent diffraction patterns from one crystal) contains a different number of patterns.

The initial complete dataset, consisting of 28 800 detector frames, saved in CBF file format [149], were converted into hdf5 format, using a script provided by O. Yefanov (CFEL), for further analysis using CrystFEL ([79], Section 5.3.2). In this particular case, in contrast to most other serial crystallography experiments, the input files for CrystFEL were not pre-filtered. A ‘hit’-finding step was performed using CrystFEL, where the definition of a ‘hit’ was not the detection of a certain number of Bragg peaks above background, but rather the ability to index individual frames. This step resulted in a total of 2 233 indexed diffraction patterns (Figure 7.1 a). As expected from the previous experiment, the tetragonal symmetry was confirmed with unit cell constants of  $a = b = 123.5$ ,  $c = 54.3$ . The approach of applying helical line scans allow processing using software developed for rotation series, such as XDS [84], which should be a better approach for this type of data compared to integrating diffraction patterns separately. The following step was the definition of a wedge, which was necessary to proceed as an individual input for XDS, A ‘wedge’ was defined as a short rotation series of horizontally adjacent diffraction patterns (more than 1), which were obtained from the same line scan. This limited the number of diffraction patterns initially used for processing with XDS to 1 734, grouped into 595 wedges containing between two and ten diffraction patterns each (Figure 7.1 b), corresponding to a collection range of  $0.75^\circ$  and  $3.75^\circ$ . Of these 595 wedges, only 130 (557 frames) were successfully indexed and integrated by XDS, following the standard 3D profile fitting procedure to integrate fully and partially recorded reflections. The default settings of XDS were used for this step (allowing  $\geq 2$  pixels per Bragg spot and  $\geq 2$  sigma per pixel) with a criterion of successful processing on the basis of the fraction of indexed spots larger than 50%. The decreased indexing yield compared to CrystFEL was most likely due to low signal-to-noise ratios of the diffraction data contained in most of the wedges, since XDS uses relatively strict thresholds.

The initial 130 partial data sets were scaled and merged in two iterations using XSCALE [84], where during the second iteration 10 wedges showed an overall Pearson correlation coefficient of less than 0.7 with respect to the initially merged dataset. Hence, the final dataset contained 120 wedges with a total of 426 diffraction patterns consisting of 109,661 reflections. By manual observation of the distribution of the final 120 wedges (Figure 7.1 c), the final data were obtained from a total of 80 crystals. The  $CC^*$  calculated for each resolution shell indicated that statistically significant data are still measurable up to 3 Å resolution, which was used as the final

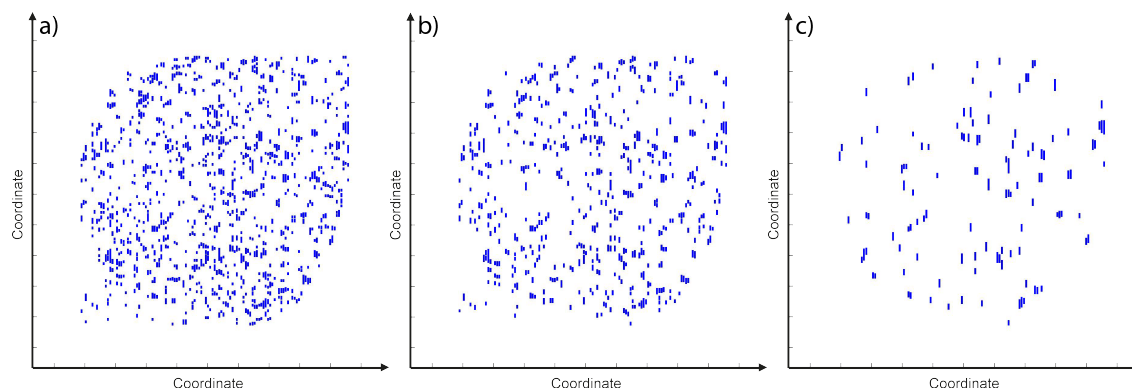


Figure 7.2: Illustration of the data processing scheme with (a) initially indexed patterns by CrystFEL (b) adjacent frames clustered as an input for XDS (c) finally indexed, integrated and merged frames by XDS

resolution cutoff [27]. After merging and scaling the final dataset consisted of 8881 unique reflection intensities in a resolution range from 88 to 3 Å, with an overall completeness of 99.8%. The distribution of the number of patterns per wedge not only reflects the size distribution, but also the orientation of the crystals relative to the scanning direction. The majority of the wedges which were merged into the final dataset included between three and five diffraction patterns, corresponding to 1.125 to 1.875° of reciprocal space.

An initial molecular replacement solution was calculated using Phaser [100], with the *in vitro* recrystallized mature TbCatB structure as a search model (PDB ID: 3MOR [145]) with 100% sequence identity, but lacking the propeptide of 62 amino acids, which was used for the SFX structure solution of the *in vivo* grown TbCatB microcrystals [64]. During a stepwise manual model building, the 62 residues of the propeptide, as well as five carbohydrate residues were modeled into positive difference electron densities. The model was refined to a final  $R_{\text{work}} / R_{\text{free}}$  of 22.3% / 26.4% using REFMAC5.5 [150], with the same set of free reflections as the SFX structure. Ordered water and glycerol molecules were present in the electron density, but could not be modeled with confidence due to the limited resolution and were hence omitted completely in this structure.

Overall, the electron density is well defined by the model. The TbCatB structure shares a papain-like fold, which is characteristic for cathepsinB enzymes. It includes the propeptide residues 27-72 and 79-85, with no defined density between the two parts. Additionally, carbohydrate chains can be observed on both Asn58 in the propeptide, including two N-acetylglucosamine (NAG) monomers, as well as Asn216

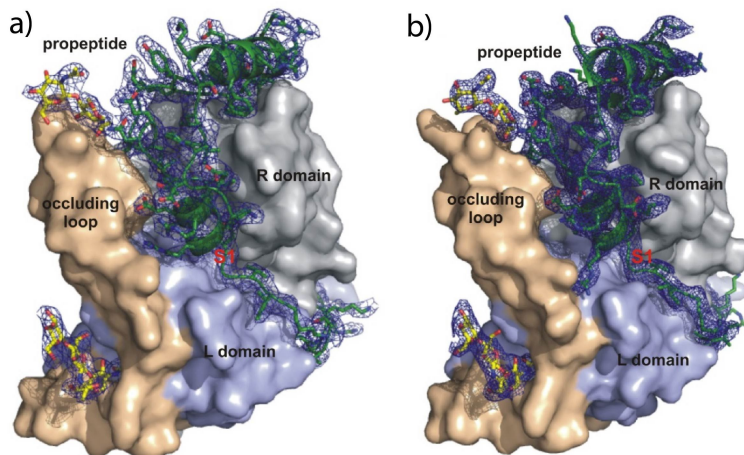


Figure 7.3: Overall structural comparison between TbCatB solved by a) SMX and b) SFX, the SFX structure was truncated to 3 Å resolution for comparison (from Gati *et al.* IUCrJ 2014)

of the enzyme domain, containing two NAG and one  $\beta$ -mannose (BMA) monomer.

To perform the structural comparison between the SFX obtained TbCatB structure (PDB ID: 4HWY [64]) was artificially truncated to 3 Å resolution (upon request of a referee), followed by the same refinement protocol compared to the synchrotron structure, resulting in an R-factor of 17% ( $R_{\text{free}} = 19.6\%$ ). A shrinking of the unit cell of the synchrotron structure by 3.7%, compared to the room temperature derived SFX structure was observed and was first attributed to the cryogenic conditions during data collection. However it was finally confirmed as a detector distance error in the SFX structure (data not shown). A superposition of the backbone trace of both models shows a high degree of similarity with an r.m.s.d. 0.35 Å. This is comparable to the estimated final error of 0.32 Å of the model, based on the limited resolution of the data. Overall no significant structural differences are present, including a lack of features which are usually attributed to specific radiation damage (Section 3.5). Finally, the largest main-chain deviations of more than 0.8 Å are present in flexible regions, such as the N- and C-terminus, loop regions, as well as residues flanking the proteolytic cleavage site of the propeptide between Ser78 and Ile79. Another result underlining the similarity of the two structures is the fact that the electron density of 9 side-chains are not defined in this structure, while 11 are missing in the SFX structure. A detailed comparison between rotamers of the two structures, as well as specific radiation damage, was not performed due to the limited resolution of the synchrotron structure.

#### 7.1.4 Discussion

The structure determination of crystals of small dimensions (more precisely a low number of unit cells), as well as crystals containing heavy-metals, is inherently limited by radiation damage. As a compromise, when acquiring data from a single crystal, the balance between the obtained resolution and the completeness of the data is adjusted accordingly. In such way one can distribute the dose over many crystals, while keeping the dose of individual crystal volumes low. In the recent past, multi-crystal approaches have led to the successful determination of a number of structures [38][59][115][116]. However, several limitations are associated to this approach, for example centering is required for conventional data processing steps. This has the disadvantage that some initial dose is applied for crystal detection, limiting the possible applied dose during final data collection. Crystal detection for centering is limited to crystals of a certain size, due to sensitivity and radiation damage limitations, which probably would have been out of reach for *in vivo* grown TbCatB microcrystals.

Here, we demonstrated a data collection and processing scheme to obtain diffraction data of micron sized crystals close to the diffraction limit of an individual crystal, without the need for an initial detection and centering step. The basic principal of this method is the systematic illumination of sub-volumes of a sample, containing a large quantity of individual microcrystals embedded in vitrified ice. The applied dose is adjusted to obtain the highest possible resolution, while fully exploiting the lifetime of the crystal in the X-ray beam. This method produces short rotation series of each crystal (called ‘wedges’) [13], such that by collecting data from an adequate number of crystals, all of reciprocal space is probed. The required number of crystals is defined by the crystal size and therefore the resulting number of diffraction patterns per wedge, and the symmetry of the crystal. Finally, these individual wedges are scaled and merged into a complete dataset, suitable for further analysis and refinement.

Microcrystals used in the above described study were smaller than almost all previously described successful structure solutions at synchrotron light sources. Tb-Catb crystals contained an average of  $\sim 10^7$  unit cells in  $9 \mu\text{m}^3$ . Compared to this, cypovirus polyhedra (CPV), which were until then the smallest studied protein crystals solved at a synchrotron [38], had an average of  $\sim 10^8$  unit cells (in  $125 \mu\text{m}^3$ ). A recent study published by Stuart and coworkers [94] published a new structure of polyhedrin - CPV17, which shows an average crystal diameter of  $1.5 \mu\text{m}$ , corre-

sponding to about  $\sim 3 \times 10^6$  unit cells. The structure factors obtained from CPV17 solved at a synchrotron (ID24 at DLS) were not sufficient to calculate a molecular replacement solution. In parallel, data from CPV17 crystals were collected by SFX data, which was followed by MR solution with a homologue polyhedrin structure and model refinement. This XFEL structure was placed into the synchrotron data using rigid body refinement and which allowed a structure solution. Overall CPV17 crystals with  $3.375 \mu\text{m}^3$  were slightly smaller than *in vivo* grown TbCatB microcrystals, but the synchrotron data quality was not sufficient to solve the structure.

One of the largest differences between the dataset obtained from SFX [64] and SMX [106], is the difference in maximum achievable resolution from similar crystals - 2.1 Å vs. 3 Å. There are several possible reasons for this dramatic difference in achievable resolution, either background scattering from the cryogenically cooled crystallization solution, radiation damage or sample quality. Background scattering in this experiment was comparably high, due to the thickness of the loop (20  $\mu\text{m}$ ), which ultimately limited the minimal thickness of the sample. Because the ice is at least 20 times thicker than the sample, background scattering is one major limitation of this experiment. By either replacing the nylon loop with thinner material, e.g. micro-meshes and grids with micron sized apertures, or more sophisticated ways, such as using graphene [151], or blotting and vitrification using approaches for grid preparation in cryo-EM [7], it might be possible to reduce this limitation of the experiment. Regarding the limitation by radiation damage, it will be necessary to systematically investigate the maximum resolution achieved with varying dose, since crystal acquired in this dataset were exposed to a maximum of 50-60 MGy, which is well beyond the Henderson limit (Section 3.5.1). Nevertheless, only minor features of the map revealed specific radiation damage. The issue of sample quality is extremely difficult to address. Batch to batch variations from *in vivo* grown crystals are frequently observed, and the high concentration of cryo-protectant might have limited the resolution of the crystals in this experiment.

An interesting observation was that only 130 wedges of the initial 595 were successfully indexed and integrated by XDS. This is most likely caused by the relatively strict criteria XDS applies, such as a minimal signal-to-noise ratio threshold, at which a feature of a diffraction pattern is considered a Bragg peak. XDS uses much more conservative cutoffs compared to CrystFEL to accept an indexing solution. As a result, only strong data acquired from relatively large crystals have been used for the final dataset. Including some of the remaining weaker data might improve the overall data quality. However, any attempt to relax these criteria did

not significantly improve the overall data quality.

A possible explanation for the successful data collection despite the high dose mentioned earlier, most of the diffraction patterns might have been collected by the tails of the X-ray beam (outside of the FWHM focus) from subsequent line scans, while crystals directly exposed to the focus were immediately destroyed. If this is the case, data were actually collected from the crystals, which were meant to be collected in the following line scan. This point has to be addressed by separate experiments by varying exposure times on each spot of the sample. It might explain, that it was possible to collect data with such high maximal radiation dose, while the actual data collection was performed with a lower dose. In fact, we believe that data collected with a dose larger than 30 MGy are generally negligible, due to global radiation damage effects. Typically, in MX experiments, the diffracted signal at 3 Å resolution is reduced by a factor of 5 after the exposure to 30 MGy (Section 3.5.1). Diffraction pattern affected by such a factor of lost diffraction power were most likely sorted out during the data reduction step (595 to 130 wedges, see above), or were down-weighted during data scaling and merging. Any attempt to reduce the high-dose contribution by truncating the last few frames from each group did not lead to any improvement in the final data statistics. Nevertheless it is impossible to exclude, that the optimal exposure time for this sample would have been shorter than 1 second. One possibility to answer this question would be to vary the spacing between individual line scans, which was chosen to be one FWHM for this experiment, or the exposure time for each frame.

*In vivo* grown TbCatB crystals are extremely stable, due to a low solvent content. Additionally, this low solvent content lowers the susceptibility to radiation damage, as well as decreases the possible heterogeneity between crystals. With several model systems (namely HEWL and insulin) we observed that some crystals tend to aggregate, which hampers data collection and processing, due to multiple crystal lattices per frame. Another advantage of using TbCatB microcrystals is their needle shape, which, in the case of the crystal being oriented along a scanning axis, increases the number of diffraction patterns one can obtain from an individual crystal, e.g. compared to a cubic crystal shape. One can therefore conclude that *in vivo* grown TbCatB microcrystals represent an ideal case for methodological development, but extrapolation to more delicate target proteins is currently difficult or impossible.



#### 7.1.4.1 Advantages of the SMX technique

As mentioned above, the handling of microcrystals is particularly difficult. Because of the random orientation of the crystals on 2D supports, preferential orientation would limit the accessible coverage of reciprocal space. This problem has been faced by Miller and coworkers [152]. By applying helical line scans, it is possible to fully overcome this problem, since any rotation width can be covered during data collection. Furthermore, no dose is ‘wasted’ by initial low-dose exposure for the crystal detection. Most microfocus beamlines have implemented line scans, so this technique could be accessible at most synchrotrons and is easy to implement.

An obvious advantage, which was the main motivation for this experiment, is the possibility to use a synchrotron beamline. Our approach not only circumvents the necessity of FELs for the structure determination of sub 10  $\mu\text{m}$  sized crystals, but also makes use of the full potential of modern microfocus synchrotron beamlines. Most current experiments by users either need to defocus or attenuate the beam, in order to not completely destroy the sample before a full dataset can be collected. Another, very important advantage compared to SFX experiments is the extremely low sample consumption, with a difference in volume of two orders of magnitude (nl vs. ml), making it more accessible for the structural biology community, compared to SFX experiments.

#### 7.1.4.2 Outlook

A team around Colletier and coworkers at IBS (Grenoble) have recently published results from an experiment [153], where tetragonal lysozyme crystals of  $> 20 \mu\text{m}$  in diameter were sandwiched between two  $\text{Si}_3\text{N}_4$  membranes as a sample mount allowing room temperature measurements. Raster scanning (without oscillation) with an exposure time of 0.2 s was applied to collect 139985 frames with a microfocus setup, followed by indexing and integration using CrystFEL [79]. Very large crystals were used (tens of  $\mu\text{m}$ ), which easily could have been measured at RT on an individual crystal, potentially at an in-house X-ray source. Given the diffraction strength of crystals of this size, background scattering limitations are completely overcome. Additionally the collection of ‘still’ images, instead of oscillation series, increases the number of diffraction patterns needed, which increases the sample consumption, as well as the data collection time, finally showing the limited impact of this publication.

First experiments by colleagues from the EMBL using an *in situ* data collection approach at P14, using a series of helical line scans of 5  $\mu\text{m}$  insulin crystals enabled the successful collection of a room temperature dataset with the same technique as described above. In general, this approach could be implemented for high-

throughput screening of crystallization trials by e.g. initial crystal detection using non-invasive methods, such as SONICC, followed by helical line scans for *in situ* data collection.

Possible further improvements of the above described technique are manifold. For example, it might be advantageous to merge data from multiple loops, in order to improve signal-to-noise ratios and the maximum resolution of the dataset. Additionally, it will be important to investigate the effect of different exposure times, attenuation levels, as well as spacing between the line scans in order to better understand the effect of applied dose, and dose rate, on final data quality.

## 7.2 Sample delivery by flowing approaches

This chapter describes the structure solution of very small lysozyme microcrystals solved during the first successful room temperature experiment of serial crystallography using synchrotron radiation, which was published in Stellato *et al.* 2014 in *IUCrJ* [140]. My contribution to this experiment was experimental design, data collection, data processing and analysis.

The development of serial femtosecond crystallography (Section 5) enabled structural information of biomolecules from small protein crystals at room-temperature. The point of room temperature measurements is an important limitation of MX experiments, in parallel to the crystallization limiting the conformational freedom of the biomolecule. Typically data are collected at cryogenic conditions, while biology mostly happens at ambient temperatures and hence cryo-cooling of the sample artificially lowers the conformational freedom of the molecule.

In order to enable serial crystallography diffraction pattern from small microcrystals at ambient temperatures at a storage ring, several technical hurdles have to be overcome. A very bright X-ray source, fast readout detectors with low electronic noise, a suitable sample delivery system and software packages are necessary in order to process this type of data.

As described earlier, the size of a crystal and its sensitivity to radiation damage are coupled. The diffracted intensity is directly proportional to the incident X-ray fluence and scales with the number of illuminated unit cells squared. At the same time, radiation damage from ionizing radiation directly correlates with dose, which is the energy absorbed per unit mass. The Henderson limit of a protein crystal at cryogenic conditions is 10 MGy per Å of resolution [32][35] (Henderson limit: 30 MGy for 3 Å resolution, Section 3.5.1), at which the disorder dampens the highest resolution reflections by a factor of five. However, radiation damage processes at

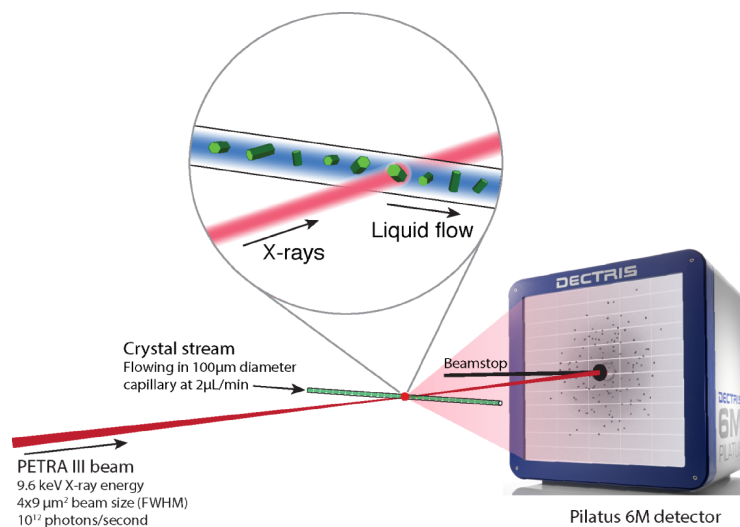


Figure 7.4: Setup of the liquid-flow based serial synchrotron crystallography experiment of lysozyme microcrystals (from Stellato *et al.* *IUCrJ* 2014)

room temperature are more complex, e.g. due to increased thermal motion and diffusion rates. As for any radiation damage limited experiment, since the earliest MX experiments by Perutz and Kendrew, multi-crystal approaches for room temperature MX experiments have been implemented to overcome these limitations, with the above described limitations of anisomorphism.

### 7.2.1 Capillary based data collection

Here we present the first adaptation of the SFX approach to a microfocus synchrotron beamline at room temperature, where a stream of small lysozyme microcrystals flows through a quartz capillary into the interaction region of the X-ray beam (Figure 7.4). With the high photon flux of the PETRA III storage ring, it was possible to implement the serial crystallography approach at room temperature from crystals smaller than 5 μm in diameter and exposure times shorter than 10 ms per detector frame. In this setup, millions of detector frames were collected in a continuous, shutter-less acquisition mode from a large number of lysozyme microcrystals. Due to the stochastic nature of hitting a crystal with the X-ray beam, the exposure time and the dose applied to each crystal is dependent on the flow-rate. In order to avoid multiple crystals per diffraction pattern, the detector frame rate, the sample flow-rate and crystal concentration were adjusted accordingly. The pro-

cessing pipeline for this experiment was very similar to typical SFX experiments, where first detector frames containing a diffraction pattern were identified, followed by indexing, integration and merging using the CrystFEL software package [79].

### 7.2.2 Experimental setup and data collection scheme

Microcrystals of HEWL were obtained by batch crystallization following an adapted protocol from Falkner *et al.* [154]. In order to systematically explore the limits of this technique, several microcrystal sizes were produced, ranging from 1 to 10  $\mu\text{m}$  diameter. The final dataset was collected from nominally  $3 \times 3 \times 6 \mu\text{m}$  diameter lysozyme crystals, obtained by mixing three parts of precipitant (14.7% (w/v) NaCl, 22% (w/v) PEG8000, 500 mM acetate buffer at pH 3) with one part of highly concentrated lysozyme solution (100 mg/ml). The crystal density was monitored and adjusted to  $1 \times 10^8 \frac{\text{crystals}}{\text{ml}}$  by light microscopy. From our previous experience with lysozyme microcrystals during several other experiments, we knew that the crystals settle in the reservoir within a matter of seconds. To overcome this problem, we increased the viscosity of the solution by adjusting the final PEG concentration to 28% (w/v).

Data collection was performed at the P11 Bio-Imaging beamline at PETRAIII. Focusing the X-ray beam to a  $6 \times 9 \mu\text{m}^2$  FWHM was achieved by a Fresnel zone plate. An X-ray energy of 9.8 keV was adjusted (0.01% bandwidth, 1 mrad divergence), resulting in a flux at the focal spot of  $2 \times 10^{12} \text{photons/s}$ . A slurry of lysozyme microcrystals was pushed into the interaction region through a 100  $\mu\text{m}$  ID quartz capillary (10  $\mu\text{m}$  wall thickness) by a syringe pump. The capillary was mounted on the in-house MX goniometer head, allowing precise adjustment of the capillary position relative to the beam. During initial tests we realized that when keeping the capillary at the same position for several shots ( $\sim 10$ ), the quartz of the capillary was damaged by the high intensity of the beam. We also observed protein crystal debris accumulating on the inner wall of the capillary, which lead to rapid clogging of the flow. Hence, it was necessary to scan the beam along the capillary during the experiment, in order to avoid exposing one spot for a longer time. This allowed a constant flow rate and illumination of the sample. The scans were adjusted to a velocity of roughly 100  $\mu\text{m}$  per second and covered the entire area reachable with the motors of the capillary, omitting the top and bottom part to avoid high background scattering from the quartz. By taking the average flow of the sample (3 - 10  $\text{mm/s}$ ) as well as the FWHM of the beam spot into consideration, the crystals were estimated to be exposed to X-rays for about 1-3 ms. In contrast to experiments at XFELs with femtosecond exposures, the synchrotron experiments in the millisecond range allow

a certain rotation of the crystal during the exposure, enabling the collection of full reflections rather than partial Bragg peaks in an individual frame. Due to crystal motion, long exposure times would lead to a spreading of Bragg peaks into radial arcs, hampering the integration step. Hence, the smallest crystal size was limited by the exposure time, which was limited to up to 10 ms. The diffraction patterns were collected on a Pilatus 6M detector with a distance of 300 mm from the interaction region, allowing 2.1 Å resolution data at the central edge of the detector. The detector exposure was set to 10 ms, while the frame rate was limited to 20 Hz, leading to a dead-time of the detector of 75%. A final dose for each crystal was estimated to approximately 0.1 - 0.3 MGy by the program RADDPOSE, corresponding to an exposure time of 3 - 10 ms [155]. This applied dose is well below the Henderson limit, allowing structural radiation damage-‘free’ data. During a total of 17 hours of beamtime 1,500,000 detector frames were collected. The total sample consumption was estimated to 2.5 ml of crystal suspension, which corresponds to about 250 mg of pure protein.

### 7.2.3 Data processing, analysis and quality assessment

The standard file format for the PILATUS 6M detector is CBF, which was first converted into HDF5 with using an in-house script (O. Yefanov - CFEL). The experiment resulted in 40 233 diffraction patterns successfully indexed (2.7% of all detector frames) and integrated using the CrystFEL software package (version 0.5.3). The data were merged in the Laue group 4/mmm. In contrast to the fluctuations observed in SFX regarding beam fluctuations (intensity and photon energy jitter) are much less dominant in synchrotron experiments. However, in addition to crystal variability, variance in the data are due to different exposure times of each crystals, which makes it necessary to use the Monte Carlo integration approach. The data were merged with an estimated maximum resolution of 2.1 Å, derived using standard figures of merit for data consistency checks, such as  $R_{\text{split}}$  and  $CC^*$ . A molecular replacement solution was calculated by PHASER [100] using the human lysozyme model (PDB ID: 2ZIL) followed by automatic model building [156], in order to proof the sufficient quality of the data and to overcome the problem of phase bias. A summary for data collection parameters, as well as refinement statistics can be found in Table 7.2.

### 7.2.4 Discussion

The results presented here are a proof-of-principle of the adaptation of the SFX approach to a synchrotron beamline. A high intensity microfocus beamline enabled us

Table 7.2: Data collection and crystallographic statistics of the lysozyme microcrystal room-temperature structure, solved at a synchrotron light source (from Stellato *et al.* IUCrJ 2014)

<b>Data collection</b>	<b>TbCatB</b>
PDB ID	4O34
Light source, beamline	PETRA III, P11
Wavelength, eV	9800
Beam size, $\mu\text{m}^2$	6 x 9
Average crystal size, $\mu\text{m}^3$	3 x 3 x 6
Flux	$2 \times 10^{12}$ photons/s
Max dose per crystal, MGy	0.3
Space group	P4 <sub>3</sub> 2 <sub>1</sub> 2
Unit cell, Å	a = b = 79.5, c = 38.4
Max. exposure time, ms	10
No. collected frames	1.500.000
No. indexed patterns, CrystFEL	40.233
Resolution range, Å	39.65-2.09
Completeness, %	93.4 (82.0)
Multiplicity	1755 (1281)
I/ $\sigma$ (I)	8.1 (1.9)
CC*	0.9986 (0.9007)
R <sub>split</sub>	7.65 (53.98)
<b>Refinement</b>	
No. reflections / test set	6411 / 709
R <sub>work</sub> / R <sub>free</sub> , %	18.0 / 23.0
No. atoms	
Protein	1000
Ions	3
Water	12
B-factors, Å <sup>2</sup>	
Protein	51.7
Ions	59.1
Water	45.1
R.m.s bonds, Å / angles, °	0.007 / 1.08
Ramachandran plot stats, %	
Favored	97.6
Allowed	2.4
Disallowed	0.0

to collect a full dataset of lysozyme microcrystals at room temperature, without visible structural damage. This data collection scheme shows several advantages over conventional data collection. Flowing crystals completely overcomes sample manipulation and handling, making it possible to use fully automated capillary based approaches. An important aspect of room temperature measurements is the possibility to perform time-resolved studies, such as pump-probe experiments [105][157] or diffusion based studies of chemical and biological reactions using mixing jets [158]. In fact, in contrast to time-resolved studies using polychromatic ‘Laue’ diffraction at synchrotron bending magnet beamlines with polychromatic light, with this serial crystallography approach it is possible to study irreversible reactions, since only one diffraction pattern is collected from each crystal. This initial result has already influenced similar projects, such as the adaptation of the LCP injector at the ID13 microfocus beamline at ESRF [159], where bacteriorhodopsin crystals grown in LCP environment were probed in a similar fashion. The combination of Laue diffraction with a serial delivery approach using a microfluidic crystallization device has also been successfully implemented at the APS [160].

There are several possible improvements of this capillary based delivery technique for serial crystallography at a storage ring. One of the major problems during data collection was the ablation of sample and damage of the quartz in the capillary. A possibility would be to coat the capillary and to use a more radiation hard material, overcoming the problem of the damage and debris, clogging the flow of the sample. Background scattering could also be improved by using for example kapton based capillaries in a helium environment. Using free-flowing liquid jets, or drop-on-demand techniques, would drastically decrease the background scattering from the quartz capillary, which will allow even smaller crystals and shorter exposure times, which will also be of interest for time-resolved experiments. The only technical limitation would be the necessity of applying a vacuum that would require installing a chamber at the beamline. Faster detectors, such as the EIGER detector [161] with a kHz readout rate will diminish the dead-time, which was 75% in the experiment described in section 7.2.1, greatly reducing the time needed to collect a full dataset. A combination of Laue diffraction and serial crystallography, using a broad bandwidth beam ( $\sim 1.5\%$ ) would increase the available flux [160]. One small technical problem was the settling of the sample which required the increase of viscosity by adding PEG to the crystals. This limits the general applicability and has to be overcome by anti-settling devices, such as developed for SFX [66].

Recently, Wolfgang Kabsch introduced a novel adaptation of his well-known software package XDS, named nanoXDS (‘nXDS’) [162]. In contrast to the traditional

software, which assumes rotation series, nXDS is specifically designed to handle individual snapshots from crystals in random orientation. Currently, no study has been published investigating its potential use for SFX, which Kabsch attributes to shot-to-shot variation of the X-ray beam and varying background noise. However, a recent publication about serial crystallography using the LCP injector at the SLS synchrotron [163], has thoroughly compared the advantages of nXDS versus CrystFEL for data processing. It shows a clear advantage of processing data with nXDS over CrystFEL, in particular for the experimental determination of phases by SAD techniques. Again, like in almost all other experiments of serial crystallography using synchrotron radiation, large crystals producing diffraction pattern with very low background were studied in this publication. Nevertheless, it would be interesting to process other datasets of our synchrotron experiments with nXDS and compare the outcome. Unfortunately the program is not yet publicly available.





## Chapter 8

# Summary and outlook

This dissertation has described research in the field of serial crystallography at the Linac Coherent Light Source (LCLS) and the synchrotron light source PETRAIII. An investigation of the strengths and limitations of the technique of serial crystallography at advanced light sources has been presented, with the aim to propose possible improvements. A major focus of this dissertation has been the analysis and comparison of data from conventional rotation series approaches and serial crystallography. This chapter re-examines the results of the experiments and analysis presented in Chapters 5 and 6 and highlights the unique advantages of the SFX technique.

### 8.1 Improvements in sample consumption

Currently, one of the most severe technical limitations of SFX experiments using the GDVN is sample consumption. Typical experiments require crystallized protein in the range of milligrams to grams, which is out of reach for most target molecules, such as membrane proteins and large macromolecular complexes. Initial expression yields are often low, and in combination with losses during purification and crystallization the final outcome usually ranges on the order of  $\mu\text{g}$ . The flow velocity of the GDVN is well beyond what would be needed for currently available XFEL light sources, where most of the sample is sprayed into the vacuum chamber without ever being hit by any X-rays. The most commonly followed approach is the usage of low flowrate injectors, such as the LCP injector [68] or the electrospin injector [164]. A common feature of these injectors is the usage of high viscosity material as a carrier medium, in order to lower the flow rate, thereby reducing sample consumption. The major difference between the two is the force applied to inject the material; while the LCP injector is pressure driven, the electrospin injector uses an electric field to pull the sample

out of the capillary. Both injectors rely on viscous media, which are not compatible with all crystallization conditions, ultimately limiting their general applicability. Additionally, the flow velocity of the electrospin injector is so low that ice formation in the jet is a severe problem, necessitating cryoprotection of the sample. Because of the high viscosity, very thick columns of sample ( $> 50 \mu\text{m}$ ) are squirted into the interaction region, leading to high background scattering. This puts a lower limit on the size of the crystal that can be used with these injectors. Current development is mainly focused on trying to apply these injectors more generally, e.g. by using more inert viscous materials, such as agarose [72].

In parallel to high viscosity injectors, there is a variety of experimental alternatives to reduce sample consumption. One option is a jet-in-jet approach, where a second stream of fluid surrounds the sample column, which reduces the diameter of the inner jet and therefore the flowrate. This has been achieved both with the GDVN and the nanoflow electrospin injector, but the analysis of data quality, e.g. background limitations, is still ongoing. Another approach is a pulsed jet, where a discontinuous jet is formed by piezo elements, which can be synchronized with the repetition rate of the XFEL in order to only inject sample when X-rays are available. Ultimately this would be the most desirable way to reduce sample consumption of liquid injectors, but it is by far the most complicated to achieve.

A second route for reaching this goal is fixed target based approaches, where a low-Z, and hence low background, support is used to deliver crystals in the beam. Several successful implementations of the fixed target approach in SFX have been described [77][165][166] in the literature. The sample consumption of fixed target approaches can barely be beaten by flowing approaches. The bottleneck of achievable repetition rates has recently been overcome and experiments can now be run at 120 Hz [167]. A major difficulty with this approach is dehydration of the sample, since thin, low-background material is usually permeable to water, in particular in vacuum. Hence, the easiest workarounds are experiments in air or helium, which increases background scattering and has to be dealt with by reduced air paths.

## 8.2 Illuminating medically important receptors - GPCRs in LCP-SFX

G protein-coupled receptors represent one of the medically most important classes of biomolecules. Given their nature of sensing extracellular signals and passing them on to the intracellular lumen, they are an ideal target for specific blockage of signalling pathways. This is illustrated by the fact that every second newly approved

drug on the market is targeting a GPCR. With the hope to understand this important class of receptors on an atomic level and to improve currently available drugs, great efforts have been made to obtain structural knowledge of this receptor class. Conventional protein crystallography approaches are mostly limited due to radiation damage, since the resulting crystals are of a limited size. The success rate of high resolution GPCR structures determined by serial femtosecond crystallography experiments at X-ray Free Electron Lasers using a lipidic cubic phase injector (Section 6.2) shows the ability of XFELs to outrun structural radiation damage on small well-ordered crystals. It also shows the effectiveness of the LCP injector in reducing sample consumption and matching the LCLS repetition rate. GPCRs in LCP-SFX represent one of the largest classes amongst all SFX structures that have so far been solved, with five high resolution structures deposited in the PDB. These include the first human membrane protein structure of a serotonin receptor 5HT<sub>2B</sub> at ambient temperatures (Section 6.3), which shows a more realistic picture of the structural dynamics of this receptor. Knowledge obtained from this newly high resolution structure has direct implications on our understanding of the receptor physiology and its interaction with ligands. A second GPCR structure described in this dissertation (Section 6.4) represents the first completely unknown structure of a biomolecule solved by the technique of SFX. The angiotensin II receptor I is of great pharmacological importance, since drugs targeting this receptor (ARBs) are the most broadly prescribed drugs for hypertension, due to their extremely low side-effects. For the first time it is now possible to understand this molecular mechanism and the specificity of the drug on an atomic level. The most recently published GPCR structure solved by LCP-SFX is the structure of bovine rhodopsin in complex with  $\beta$ -arrestin, which gives detailed insights into the mechanism of how GPCRs are silenced upon activation. Several ongoing projects are focused on the reconstruction of data collected from more GPCRs. Further research in this field is focused on methodological developments, such as smaller crystals, lower protein consumption and faster data collection, experimental phasing, structure solution of more complexes and ultimately time-resolved crystallography experiments to elucidate protein dynamics.

### 8.3 Opening the method to a broader community

The power of distributing X-ray dose on a large number of crystals has been frequently used to overcome limitations of crystallography due to radiation damage. This is the key advantage of the serial crystallography approach, which allows data collection from very small crystals that are out-of-reach for conventional data col-

lection schemes. Beamtime proposals at XFELs are extremely competitive and even with a successful proposal review beamtime is very limited and usually does not allow room for improvements. The last two experiments described in this dissertation summarize our efforts on adapting the serial crystallography approach to modern synchrotron beamlines. This alternative approach is of great value since beamtime is more readily available at synchrotron light sources which allow data collection from micrometer sized crystals at both cryogenic and room temperature. An additional advantage of the cryogenic approach described in section 7.1 is the extremely low sample consumption of a few nanoliters and the relatively fast data collection.

Further improvements in the X-ray sources, beamlines, sample delivery, detectors and data analysis will definitely reduce both the limitation on crystal size and experimental time. Planned synchrotron upgrades and novel storage ring designs will allow flux densities which are orders of magnitude larger than what is currently available. Even though these improvements will be of limited value to the MX community since conventional MX experiments typically already use only a fraction of the flux possible, serial crystallography experiments would greatly benefit from these novel light sources. Given the very competitive nature of SFX proposal reviews, alternative approaches for data collection at synchrotrons are extremely valuable and will be of great interest for the structural biology community.

## 8.4 On the road to single-particle X-ray imaging

SFX experiments described in Chapter 5.4 of this dissertation demonstrate high resolution structure determination with the smallest crystals and highest dose so far used in X-ray crystallography. The sample used during the experiment was CpGV occlusion bodies with a crystalline shell of 10,000 unit cells (Section 5.4). The main conclusion from the experiment is the fact that high resolution data can be collected from protein crystals of much smaller size than currently possible at any other light source, while collecting data with an applied dose much higher than applied for any other high resolution structure determination. The achieved resolution of 2 Å was in fact limited by the experimental geometry and could have been improved by a shorter detector distance.

With the help of simulated data, it was possible to extrapolate the current size limitations of the technique of SFX. We estimated the smallest crystal size usable for high resolution structures using SFX with currently available beam parameters to about 100 unit cells, omitting experimental background. In order to achieve the visionary goal of single particle X-ray imaging to near atomic resolution, we predict that it will be necessary to increase the number of photons per pulse to at least 100

GGy, while it will be necessary to shorten the pulse duration to sub-fs in order to overcome the problem of electronic radiation damage of the sample [168][169][170]. Currently proposed XFEL designs do aim for higher repetition rates with only minor improvements on pulse energy, while shorter pulse durations in the attosecond regime always come with a drastic reduction in flux, due to technical limitations. Hence, in the short term, background scattering and sample delivery have to be optimized to provide higher hit rates and improved resolution reconstructions. Additionally, algorithm development is needed for accurate orientation determination from noisy diffraction pattern and being able to detect and deal with structural heterogeneity of the sample. Given improvements in all these three research fields (X-ray source, sample delivery, algorithms) single-particle X-ray imaging at ambient temperatures might become a reality, opening unpredictable advances for the structural biology community.



## Chapter 9

# Appendix

### 9.1 Reprint of published articles



developmental potential, even before their physical disappearance. The role of CRL4<sup>VPRBP</sup> in mammalian oocytes is summarized in Fig. 4E.

Although CRL4<sup>VPRBP</sup> is crucial for TET activities, our results suggest that TET1, 2, and 3 are not the only CRL4<sup>VPRBP</sup> substrates in oocytes. More than one-third of the embryos derived from TET3-deleted oocytes could develop to term (11). However, all embryos that were derived from DDB1-deleted oocytes died before the eight-cell stage, which indicated that they had defects other than TET3-mediated genome reprogramming. CRL4 might also recruit other substrate adaptors, poly-ubiquitinate a number of protein substrates, and direct them toward degradation. Identifying other CRL4<sup>VPRBP</sup> substrates will shed new light on the molecular regulatory mechanisms of oocyte functions.

#### References and Notes

1. D. Adhikari, K. Liu, *Endocr. Rev.* **30**, 438–464 (2009).
2. S. Jackson, Y. Xiong, *Trends Biochem. Sci.* **34**, 562–570 (2009).
3. S. Angers *et al.*, *Nature* **443**, 590–593 (2006).
4. S. A. Pangas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8090–8095 (2006).
5. Y. Choi, D. Yuan, A. Rajkovic, *Biol. Reprod.* **79**, 1176–1182 (2008).
6. A. Rajkovic, S. A. Pangas, D. Ballow, N. Suzumori, M. M. Matzuk, *Science* **305**, 1157–1159 (2004).
7. S. M. Soyak, A. Amleh, J. Dean, *Development* **127**, 4645–4654 (2000).
8. J. J. Eppig, *Reproduction* **122**, 829–838 (2001).
9. Y. F. He *et al.*, *Science* **333**, 1303–1307 (2011).
10. S. Ito *et al.*, *Science* **333**, 1300–1303 (2011).
11. T. P. Gu *et al.*, *Nature* **477**, 606–610 (2011).
12. D. H. Castrillon, L. Miao, R. Kollipara, J. W. Horner, R. A. DePinho, *Science* **301**, 215–218 (2003).
13. P. Reddy *et al.*, *Hum. Mol. Genet.* **18**, 2813–2824 (2009).
14. P. Reddy *et al.*, *Science* **319**, 611–613 (2008).

**Acknowledgments:** We thank K. Guan and X. Feng for discussions and critical reviews of the manuscript. J.J. Chen and Z. Xia provided VPRBP and TET1, 2, and 3 plasmids, respectively. G. Xu and L. Jia provided TET3 antibody and MLN4924, respectively. Y. Xiong provided *Vprbp* floxed mice and VPRBP-2RA plasmid. This study was supported by National Basic Research Program of China [2011CB944504 and 2012CB944403 (to H.-Y.F. and Q.-Y.S.) and 2012CB966600 (to C.T.)], National Natural Science Foundation of China (81172473 and 31371449 to H.-Y.F.), and Basic Scientific Research Funding of Zhejiang University (2011QN81001 to H.-Y.F.).

#### Supplementary Materials

www.sciencemag.org/content/342/6165/1518/suppl/DC1  
Materials and Methods  
Figs. S1 to S11  
Tables S1 to S3  
References (15–24)

12 August 2013; accepted 15 November 2013  
10.1126/science.1244587

## Serial Femtosecond Crystallography of G Protein–Coupled Receptors

Wei Liu,<sup>1</sup> Daniel Wacker,<sup>1</sup> Cornelius Gati,<sup>2</sup> Gye Won Han,<sup>1</sup> Daniel James,<sup>3</sup> Dingjie Wang,<sup>3</sup> Garrett Nelson,<sup>3</sup> Uwe Weierstall,<sup>3</sup> Vsevolod Katritch,<sup>1</sup> Anton Barty,<sup>2</sup> Nadia A. Zatsepin,<sup>3</sup> Dianfan Li,<sup>4</sup> Marc Messerschmidt,<sup>5</sup> Sébastien Boutet,<sup>5</sup> Garth J. Williams,<sup>5</sup> Jason E. Koglin,<sup>5</sup> M. Marvin Seibert,<sup>5,6</sup> Chong Wang,<sup>1</sup> Syed T. A. Shah,<sup>4</sup> Shibom Basu,<sup>7</sup> Raimund Fromme,<sup>7</sup> Christopher Kupitz,<sup>7</sup> Kimberley N. Rendek,<sup>7</sup> Ingo Grotjohann,<sup>7</sup> Petra Fromme,<sup>7</sup> Richard A. Kirian,<sup>2,3</sup> Kenneth R. Beyerlein,<sup>2</sup> Thomas A. White,<sup>2</sup> Henry N. Chapman,<sup>2,8,9</sup> Martin Caffrey,<sup>4</sup> John C. H. Spence,<sup>3</sup> Raymond C. Stevens,<sup>1</sup> Vadim Cherezov<sup>1\*</sup>

X-ray crystallography of G protein–coupled receptors and other membrane proteins is hampered by difficulties associated with growing sufficiently large crystals that withstand radiation damage and yield high-resolution data at synchrotron sources. We used an x-ray free-electron laser (XFEL) with individual 50-femtosecond-duration x-ray pulses to minimize radiation damage and obtained a high-resolution room-temperature structure of a human serotonin receptor using sub-10-micrometer microcrystals grown in a membrane mimetic matrix known as lipidic cubic phase. Compared with the structure solved by using traditional microcrystallography from cryo-cooled crystals of about two orders of magnitude larger volume, the room-temperature XFEL structure displays a distinct distribution of thermal motions and conformations of residues that likely more accurately represent the receptor structure and dynamics in a cellular environment.

**G**protein–coupled receptors (GPCRs) represent a highly diverse superfamily of eukaryotic membrane proteins that mediate cellular communication. In humans, ~800 GPCRs respond to a variety of extracellular signaling mol-

ecules and transmit signals inside the cell by coupling to heterotrimeric G proteins and other effectors. Their involvement in key physiological and sensory processes in humans makes GPCRs prominent drug targets. Despite the high biomedical relevance and decades of dedicated research, knowledge of the structural mechanisms of ligand recognition, receptor activation, and signaling in this broad family remains limited. Challenges for GPCR structural studies include low-expression yields, low receptor stability after detergent extraction from native membranes, and high conformational heterogeneity. Many years of developments aimed at receptor stabilization, crystallization, and microcrystallography culminated in a series of breakthroughs in GPCR structural biology leading to the structure determination of 22 receptors, some of which were solved in several conformational states and one in complex with its G protein partner (1–5).

Nonetheless, crystallographic studies of GPCRs remain difficult because many of them produce only microcrystals. Most GPCR structures to date have been obtained by using crystallization from the membrane-mimetic environment of a lipidic cubic phase (LCP) (6, 7). LCP crystallization has proven successful for obtaining high-resolution structures of a variety of membrane proteins, including ion channels, transporters, and enzymes, in addition to GPCRs (8, 9). This method leads to highly ordered crystals that are, however, often limited in size. Microfocus x-ray beams of high intensity (~10<sup>9</sup> photons/s/μm<sup>2</sup>) and long exposures (~5 s) are typically required in order to obtain sufficient intensity for high-resolution data from weakly diffracting microcrystals. The high-radiation doses induce severe radiation damage and require merging data from multiple crystals in order to obtain complete data sets of sufficient quality. Accordingly, sub-10-μm GPCR crystals are currently not suitable for high-resolution data collection, even at the most powerful synchrotron microfocus beamlines (7, 10).

Serial femtosecond crystallography (SFX) (11), which takes advantage of x-ray free-electron lasers (XFEL), has recently demonstrated great promise for obtaining room-temperature high-resolution data from micrometer- and sub-micrometer-size crystals of soluble proteins, with minimal radiation damage (12, 13). The highly intense (~2 mJ, 10<sup>12</sup> photons per pulse) and ultrashort (<50 fs) x-ray pulses produced by XFELs enable the recording of high-resolution diffraction snapshots from individual crystals at single orientations before their destruction. SFX data collection, therefore, relies on a continuous supply of small crystals intersecting the XFEL beam in random orientations—typically provided by a fast-running liquid microjet (12)—which is incompatible with streaming highly viscous gel-like materials such as LCP and requires tens to hundreds of milligrams of crystallized protein for data collection (11). For many membrane proteins, including most human membrane proteins, obtaining such quantities is not practical.

<sup>1</sup>Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>2</sup>Center for Free Electron Laser Science, Deutsches Elektronen-Synchrotron, 22607 Hamburg, Germany. <sup>3</sup>Department of Physics, Arizona State University, Tempe, AZ 85287, USA. <sup>4</sup>School of Medicine and School of Biochemistry and Immunology, Trinity College, Dublin, Dublin 2, Ireland. <sup>5</sup>SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. <sup>6</sup>Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-751 24 Uppsala, Sweden. <sup>7</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287, USA. <sup>8</sup>Department of Physics, University of Hamburg, 22761 Hamburg, Germany. <sup>9</sup>Center for Ultrafast Imaging, 22607 Hamburg, Germany.

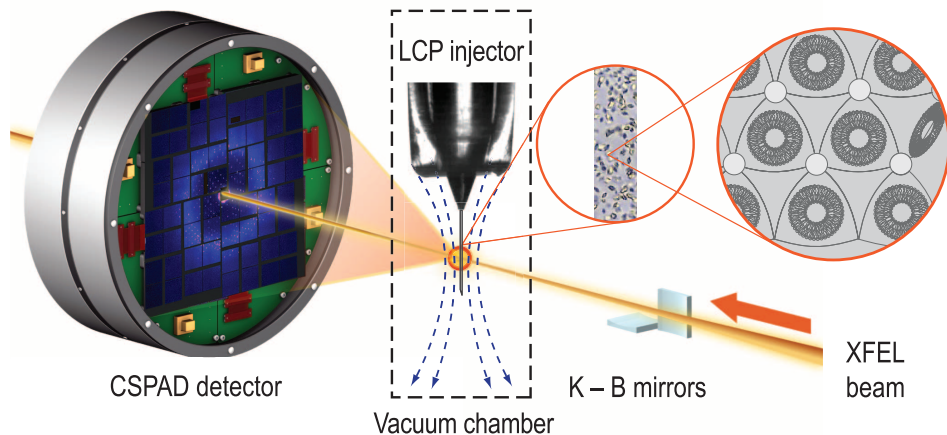
\*Corresponding author. E-mail: vcherezov@scripps.edu

We have modified the SFX data collection approach (Fig. 1) and obtained a room-temperature GPCR structure at 2.8 Å resolution using only 300 µg of protein crystallized in LCP. SFX experiments were performed at the Coherent X-ray Imaging (CXI) instrument of the Linac Coherent Light Source (LCLS) (14). LCP-grown microcrystals (average size of 5 by 5 by 5 µm) (fig. S1) (15) of the human serotonin 5-HT<sub>2B</sub> receptor (16) bound to the agonist ergotamine were continuously delivered across a ~1.5-µm-diameter XFEL beam by using a specially designed LCP injector. LCP with randomly distributed crystals was ex-

truded through a 20- to 50-µm capillary into a vacuum chamber (10<sup>-4</sup> torr) at room temperature (21°C) (17) and a constant flow rate of 50 to 200 nL/min and was stabilized by a co-axial flow of helium or nitrogen gas supplied at 20 to 30 bar. We recorded single-pulse diffraction patterns (fig. S2) using 9.5-keV (1.3 Å) x-ray pulses of 50 fs duration at a 120 Hz repetition rate by means of a Cornell-SLAC pixel array detector (CSPAD) (18) positioned at a distance of 100 mm from the sample. The XFEL beam was attenuated to 3 to 6% so as to avoid detector saturation. The average x-ray pulse energy at the sample was 50 µJ

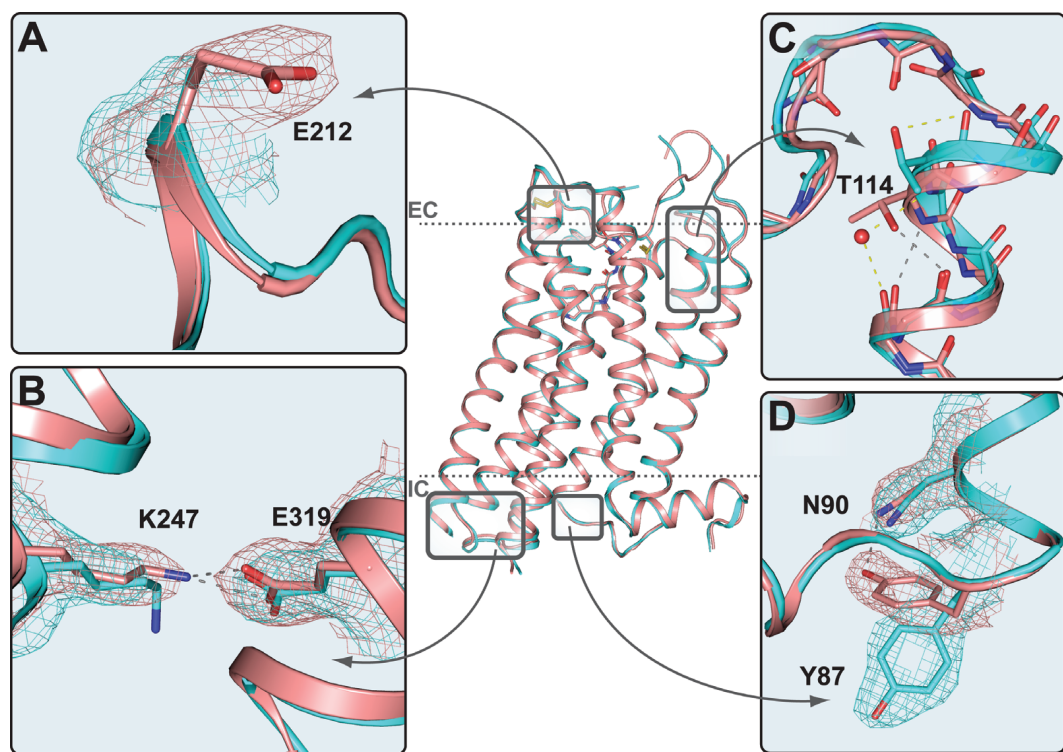
(3 × 10<sup>10</sup> photons/pulse), corresponding to a radiation dose of up to 25 megagrays per crystal. A total of 4,217,508 diffraction patterns were collected within 10 hours by using ~100 µL of crystal-loaded LCP, corresponding to ~0.3 mg of protein. Of these patterns, 152,651 were identified as crystal hits (15 or more Bragg peaks) by the processing software Cheetah (<http://www.desy.de/~barty/cheetah/>), corresponding to a hit rate of 3.6%. Of these crystal hits, 32,819 patterns (21.5%) were successfully indexed and integrated by CrystFEL (19) at 2.8 Å resolution (table S1). The structure was determined through molecular replacement and refined to  $R_{\text{work}}/R_{\text{free}} = 22.7/27.0\%$ . Overall, the final structure (fig. S3) has a well-defined density for most residues, including the ligand ergotamine (fig. S4).

We compared the XFEL structure of the 5-HT<sub>2B</sub> receptor/ergotamine complex (5-HT<sub>2B</sub>-XFEL) with the recently published structure of the same receptor/ligand complex obtained by means of traditional microcrystallography at a synchrotron source [Protein Data Bank (PDB) ID 4IB4; 5-HT<sub>2B</sub>-SYN] (21). Synchrotron data were collected at 100 K on cryo-cooled crystals of a much larger size (average volume, ~10<sup>4</sup> µm<sup>3</sup>) than those used for the XFEL structure (average volume, ~10<sup>2</sup> µm<sup>3</sup>) (fig. S1). Other differences between data collection protocols are listed in table S1. Both data sets were processed in the same spacegroup C222<sub>1</sub>, which is expected given the very similar crystallization conditions. However, the lattice parameters for the room-temperature XFEL crystals are slightly longer in the *a* and *b* directions and slightly shorter in the *c* direction, resulting in a 2.1% larger unit cell volume. Concomitant with these lattice changes, we observed a ~2.5° rotation of



**Fig. 1. Experimental setup for SFX data collection using an LCP injector.** 5-HT<sub>2B</sub> receptor microcrystals (first zoom level) dispersed in LCP (second zoom level) are injected as a continuous column of 20 to 50 µm in diameter—stabilized by a co-axial gas flow (blue dash curved lines)—inside a vacuum chamber and intersected with 1.5-µm-diameter pulsed XFEL beam focused with Kirkpatrick-Baez (K-B) mirrors. Single-pulse diffraction patterns were collected at 120 Hz by using a CSPAD detector. The entire XFEL beam path and CSPAD are under vacuum.

**Fig. 2. Comparison between 5-HT<sub>2B</sub>-XFEL (light red) and 5-HT<sub>2B</sub>-SYN (teal) structures.** Central image represents a backbone overlay of the two structures. Dashed lines correspond to membrane boundaries defined by the Orientation of Proteins in Membrane database (<http://opm.phar.umich.edu>) (28). (A) Electron density for the Glu212 side chain is missing in 5-HT<sub>2B</sub>-SYN and fully resolved in 5-HT<sub>2B</sub>-XFEL. (B) A salt bridge between Glu319 and Lys247 links intracellular parts of helices V and VI in the 5-HT<sub>2B</sub>-XFEL structure. In the 5-HT<sub>2B</sub>-SYN structure, Lys247 makes a hydrogen bond with Tyr1105 from the BRIL fusion protein. (C) Extracellular tip of helix II forms a regular helix in 5-HT<sub>2B</sub>-XFEL with Thr114, making a stabilizing hydrogen bond with the backbone carbonyl, whereas in 5-HT<sub>2B</sub>-SYN, a water-stabilized kink is introduced at this position. (D) Tyr87 forms a hydrogen bond with Asn90 in 5-HT<sub>2B</sub>-XFEL; this hydrogen bond is broken, and Tyr87 adopts a different rotamer conformation in the 5-HT<sub>2B</sub>-SYN structure. 2mF<sub>obs</sub>-DF<sub>calc</sub> maps (contoured at 1σ level) are shown only around described residues.





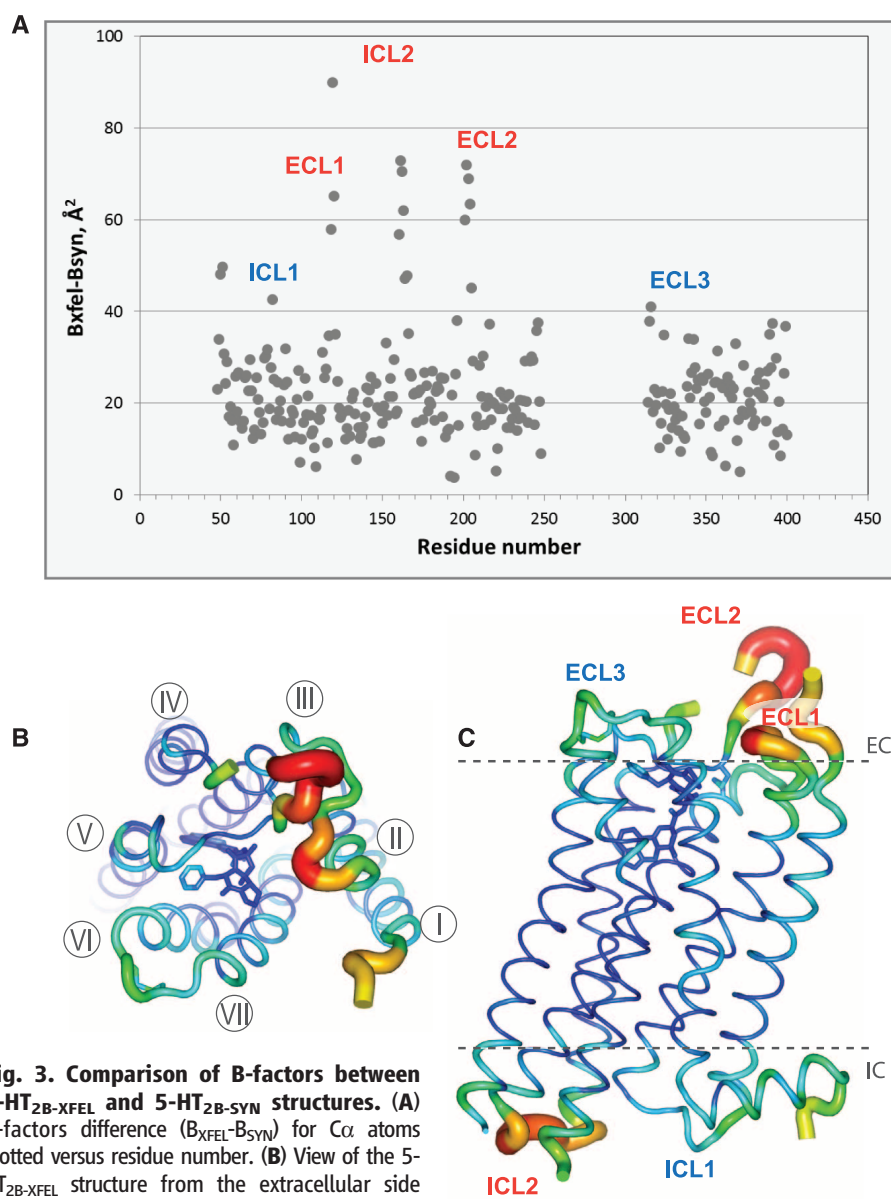
the  $b_{562}$  RIL (BRIL) fusion domain with respect to the receptor (fig. S5). Otherwise, the receptor domains of the 5-HT<sub>2B-XFEL</sub> and 5-HT<sub>2B-SYN</sub> structures are very similar [receptor C $\alpha$  root mean square deviation (RMSD) = 0.46 Å, excluding flexible residues at the N terminus, 48 to 51, and in the extracellular loop 2 (ECL2), 195 to 205] (Fig. 2). The ligand ergotamine has indistinguishable electron density and placement (total ligand RMSD = 0.32 Å) in both structures (Fig. 2 and fig. S4B). The largest backbone deviations were observed in the loop regions, especially in the stretch of ECL2 between helix IV and the Cys128–Cys207 disulfide bond, which is apparently very flexible. We observed an unexpected backbone deviation at the extracellular tip of helix II (Fig. 2C), which adopts a regular  $\alpha$ -helix in the 5-HT<sub>2B-XFEL</sub> structure, with Thr<sup>114</sup> forming a stabilizing hy-

drogen bond with the main chain carbonyl of Ile110. In the 5-HT<sub>2B-SYN</sub> structure, however, a water-stabilized kink was found at this location, which results in the two structures deviating by 2.0 Å (at C $\alpha$  atom of Thr114) at the tip of helix II and up to 3.4 Å (at O atom of Phe117) in ECL1.

Although absolute B- (or temperature) factor values can be affected by errors associated with experimental conditions, their distribution generally represents the relative static and dynamic flexibility of the protein in the crystal (22). Because both structures were obtained from similar samples and at similar resolutions, we analyzed their B-factor distributions so as to study the effect of the different temperatures on the thermal motions of the receptor. The average B-factor for the receptor part in the room-temperature 5-HT<sub>2B-XFEL</sub> structure (88.4 Å<sup>2</sup>) is 21 Å<sup>2</sup> larger than that in the

cryo 5-HT<sub>2B-SYN</sub> structure (67.2 Å<sup>2</sup>), which is consistent with larger thermal motions at higher temperature and possible effects of Bragg termination during the XFEL pulse (20). The distribution of B-factors highlights a more rigid core of the seven transmembrane helices in comparison with loops, with more pronounced B-factor deviations observed in the room-temperature 5-HT<sub>2B-XFEL</sub> structure (Fig. 3 and fig. S6). N terminus, intracellular loop 2 (ICL2), ECL1, and part of ECL2 between helix IV and the Cys128–Cys207 disulfide bond show much larger deviations in B-factors (50 to 100 Å<sup>2</sup>) between the two structures as compared with the average difference of 21 Å<sup>2</sup>. These parts of the structure are not involved in direct interactions with the ligand ergotamine, but their mobility may affect the kinetics of ligand binding and interactions with intracellular binding partners (23). In contrast, ICL1, part of ECL2 between the Cys128–Cys207 disulfide bond and helix V, and ECL3 display just an average increase in the B-factors, suggesting that the relative range of their thermal fluctuations was adequately captured in the cryo structure. As previously established with cryocrystallography, one of the most pronounced differences between the two subtypes of serotonin receptors, 5-HT<sub>2B</sub> and 5-HT<sub>1B</sub>, occurs at the extracellular tip of helix V and ECL2, which forms an additional helical turn stabilized by a water molecule in 5-HT<sub>2B</sub> (21). This additional turn pulls the extracellular tip of helix V toward the center of the helical bundle and was suggested to be responsible for the biased agonism of ergotamine at the 5-HT<sub>2B</sub> receptor. The 5-HT<sub>2B-XFEL</sub> structure confirms the rigid structured conformation of ECL2, stabilized by a comprehensive network of hydrogen bonds, involving residues Lys193, Glu196, Arg213, Asp216, and a lipid OLC (monoolein) (fig. S7); however, no ordered water molecule was observed, emphasizing that water is more disordered and probably does not play a substantial structural role at this location.

Several side chains have partly missing electron density in both room-temperature and cryo structures (table S2). Such lack of density is most likely related to disorder of the corresponding side chains (such as residues at the N terminus, ECL2, and ICL2) (Fig. 2A). Two disulfide bonds, Cys128–Cys207 and Cys350–Cys353, are intact and well resolved in both structures; however, the B-factor increase in the 5-HT<sub>2B-XFEL</sub> structure compared with 5-HT<sub>2B-SYN</sub> for each of these disulfide bonds (11.1 and 5.7 Å<sup>2</sup>, respectively) is lower than that of the average B-factor increase (21 Å<sup>2</sup>). Several side chains have different rotamer conformations between the two structures (Fig. 2D and table S3), which is consistent with a partial remodeling of the side chain conformational distribution upon cryo-cooling observed in soluble proteins (24). Several interactions involving charged residues appear stronger and better defined in 5-HT<sub>2B-XFEL</sub> compared with the 5-HT<sub>2B-SYN</sub> structure (table S4). This strengthening of the charged interactions at higher temperatures potentially can be explained



**Fig. 3. Comparison of B-factors between 5-HT<sub>2B-XFEL</sub> and 5-HT<sub>2B-SYN</sub> structures.** (A) B-factors difference ( $B_{XFEL} - B_{SYN}$ ) for C $\alpha$  atoms plotted versus residue number. (B) View of the 5-HT<sub>2B-XFEL</sub> structure from the extracellular side and (C) in the lateral-to-membrane orientation.

Structure in (B) and (C) is shown in putty representation and colored in rainbow colors by the C $\alpha$  B-factors (range 60 to 170 Å<sup>2</sup>). Loops for which the B-factor difference is above 50 Å<sup>2</sup> are labeled in red, and those with a difference below 50 Å<sup>2</sup> are in blue in (A) and (C). Helices are labeled in (B).

by a decrease in the dielectric constant of water with temperature, reducing the desolvation penalty (25, 26). In particular, the salt bridge between Glu319 and Lys247 is well defined in the 5-HT<sub>2B</sub>-XFEL structure but appears broken in the cryo 5-HT<sub>2B</sub>-SYN structure (Fig. 2B). Because GPCR activation has been associated with large-scale structural changes in the intracellular parts of helices V and VI, this salt bridge may play a role in the receptor function and is likely to be more accurately resolved and represented in the 5-HT<sub>2B</sub>-XFEL structure recorded at room temperature.

Overall, the observed differences likely originate from effects related to thermal motions, cryo-cooling (24), and radiation damage (27). Thus, the XFEL source enables access to a room-temperature GPCR structure, which more accurately represents the conformational ensemble for this receptor under native conditions. Because dynamics are an integral part of GPCR biology, the use of SFX to accurately determine GPCR structural details at room temperature can make an important contribution to understanding the structure-function relationships in this superfamily.

#### References and Notes

- V. Cherezov *et al.*, *Science* **318**, 1258–1265 (2007).
- S. G. Rasmussen *et al.*, *Nature* **469**, 175–180 (2011).
- V. Katritch, V. Cherezov, R. C. Stevens, *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
- A. J. Venkatakrishnan *et al.*, *Nature* **494**, 185–194 (2013).
- M. Audet, M. Bouvier, *Cell* **151**, 14–23 (2012).
- E. M. Landau, J. P. Rosenbusch, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14532–14535 (1996).
- M. Caffrey, V. Cherezov, *Nat. Protoc.* **4**, 706–731 (2009).
- V. Cherezov, *Curr. Opin. Struct. Biol.* **21**, 559–566 (2011).
- M. Caffrey, D. Li, A. Dukkipati, *Biochemistry* **51**, 6266–6288 (2012).
- J. L. Smith, R. F. Fischetti, M. Yamamoto, *Curr. Opin. Struct. Biol.* **22**, 602–612 (2012).
- H. N. Chapman *et al.*, *Nature* **470**, 73–77 (2011).
- S. Boutet *et al.*, *Science* **337**, 362–364 (2012).
- L. Redecke *et al.*, *Science* **339**, 227–230 (2013).
- S. Boutet, G. J. Williams, *New J. Phys.* **12**, 035024 (2010).
- Materials and methods are available as supplementary materials on Science Online.
- The engineered-for-crystallization construct is based on the sequence of the human 5-HT<sub>2B</sub> receptor with the following modifications: (i) Residues Tyr249–Val313 within ICL3 were replaced with Ala1–Leu106 of the most stabilized apo cytochrome BRIL; (ii) N-terminal residues 1 to 35 and C-terminal residues 406 to 481 were truncated; and (iii) a thermostabilizing Met144 Trp mutation was introduced.
- The temperature measured in the CXI hutch during the experiments was 294 K (21°C). The actual crystal temperature was likely a few degrees lower because of the evaporative cooling upon injection of crystal-loaded LCP in vacuum.
- P. Hart *et al.*, *Proc. SPIE* **8504**, 85040C (2012).
- T. A. White *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1231–1240 (2013).
- A. Barty *et al.*, *Nat. Photonics* **6**, 35–40 (2012).
- D. Wacker *et al.*, *Science* **340**, 615–619 (2013).
- B. T. M. Willis, A. W. Pryor, *Thermal Vibrations in Crystallography* (Cambridge Univ. Press, London, 1975).
- A. C. Pan, D. W. Borhani, R. O. Dror, D. E. Shaw, *Drug Discov. Today* **18**, 667–673 (2013).
- J. S. Fraser *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16247–16252 (2011).
- A. H. Elcock, *J. Mol. Biol.* **284**, 489–502 (1998).
- S. Kumar, R. Nussinov, *ChemBioChem* **3**, 604–617 (2002).
- E. F. Garman, *Acta Crystallogr. D Biol. Crystallogr.* **66**, 339–351 (2010).
- M. A. Lomize, A. L. Lomize, I. D. Pogozheva, H. I. Mosberg, *Bioinformatics* **22**, 623–625 (2006).

**Acknowledgments:** Parts of this research were carried out at the LCLS, a National User Facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences and at the General Medicine and Cancer Institute Collaborative Access Team of the Argonne Photon Source, Argonne National Laboratory. This work was supported by the National Institutes of Health Common Fund in Structural Biology grants P50 GM073197 (V.C. and R.C.S.), P50 GM073210 (M.C.), and R01 GM095583 (P.F.); National Institute of General Medical Sciences PSI:Biologics grants U54 GM094618 (V.C., V.K., and R.C.S.) and U54 GM094599 (P.F.); and NSF Science and Technology Center award 1231306 (J.C.H.S.). We further acknowledge support from the Helmholtz Association, the German Research Foundation, the German Federal Ministry of Education and Research (H.N.C.), and Science Foundation Ireland (07/IN.1/B1836, 12/IA/1255) (M.C.). We give special thanks to G. M. Stewart, T. Anderson, SLAC Infomedia, and K. Kadyshchik from The Scripps Research Institute for preparing Fig. 1; T. Trinh and M. Chu for help with baculovirus expression; H. Liu and M. Klinker for help with data processing; A. Walker for assistance with manuscript preparation; and I. Wilson for reviewing the manuscript. Coordinates and the structure factors have been deposited in PDB under the accession code 4NC3. The diffraction patterns have been deposited in the Coherent X-ray Imaging Data Bank <http://cxi.db.org> under the accession code ID-21. U.W. and J.C.H.S. are inventors on a patent application filed by Arizona State University titled “Apparatus and Methods for Lipidic Cubic Phase (LCP) Injection for Membrane Protein Investigations.” W.L. developed protocols of producing high-density microcrystals in LCP; prepared samples; and helped with testing LCP injector, data collection, and writing the paper. Da.W. prepared 5-HT<sub>2B</sub>

microcrystals in LCP and helped with data collection, structure refinement, analysis, and writing the paper. C.G. participated in data collection and processed and analyzed data. G.W.H. performed structure refinement. D.J., Di.W., and G.N. helped develop and operate the LCP injector. U.W. conceived, designed, and developed the LCP injector. V.K. analyzed the results and helped with writing the paper. A.B. participated in data collection, wrote data processing software, and helped with data processing and writing the paper. N.A.Z. and Sh.B. participated in data collection and helped with data processing. D.L. helped with data collection. Se.B., M.M., G.J.W., J.E.K., and M.M.S. set up the XFEL experiment, beamline, controls, and data acquisition; operated the CXI beamline; and performed the data collection. C.W. helped with sample preparation. S.T.A.S. synthesized and purified 7.9 MAG. R.F., C.K., K.N.R., and I.G. participated in data collection and contributed to sample characterization. P.F. was involved in the initiation and planning of the experiments, assisted with sample characterization and data collection, and contributed to writing the paper. R.A.K. developed the Monte Carlo integration method and contributed to data processing. K.R.B. contributed to software development and data processing. T.A.W. developed the Monte Carlo integration method, wrote data processing software, and contributed to data processing. H.N.C. supervised software development and data processing and helped with writing the paper. M.C. provided the 7.9 MAG and helped with data collection and with writing the paper. J.C.H.S. helped develop the LCP injector and developed the Monte Carlo integration method with R.A.K. R.C.S. supervised GPCR production and contributed to writing the paper. V.C. conceived the project, designed the experiments, supervised data collection, performed structure refinement, analyzed the results, and wrote the paper.

#### Supplementary Materials

[www.sciencemag.org/content/342/6165/1521/suppl/DC1](http://www.sciencemag.org/content/342/6165/1521/suppl/DC1)  
Materials and Methods  
Figs. S1 to S8  
Tables S1 to S4  
References (29–41)

2 August 2013; accepted 13 November 2013  
10.1126/science.1244142

## mTOR Inhibition Alleviates Mitochondrial Disease in a Mouse Model of Leigh Syndrome

Simon C. Johnson,<sup>1</sup> Melana E. Yanos,<sup>1,2</sup> Ernst-Bernhard Kayser,<sup>3</sup> Albert Quintana,<sup>4</sup> Maya Sangesland,<sup>1</sup> Anthony Castanza,<sup>1</sup> Lauren Uhde,<sup>1</sup> Jessica Hui,<sup>1</sup> Valerie Z. Wall,<sup>1</sup> Arni Gagnidze,<sup>1</sup> Kelly Oh,<sup>1</sup> Brian M. Wasko,<sup>1</sup> Fresnida J. Ramos,<sup>1</sup> Richard D. Palmiter,<sup>4</sup> Peter S. Rabinovitch,<sup>1</sup> Philip G. Morgan,<sup>3</sup> Margaret M. Sedensky,<sup>3</sup> Matt Kaeberlein<sup>1\*</sup>

Mitochondrial dysfunction contributes to numerous health problems, including neurological and muscular degeneration, cardiomyopathies, cancer, diabetes, and pathologies of aging. Severe mitochondrial defects can result in childhood disorders such as Leigh syndrome, for which there are no effective therapies. We found that rapamycin, a specific inhibitor of the mechanistic target of rapamycin (mTOR) signaling pathway, robustly enhances survival and attenuates disease progression in a mouse model of Leigh syndrome. Administration of rapamycin to these mice, which are deficient in the mitochondrial respiratory chain subunit Ndufs4 [NADH dehydrogenase (ubiquinone) Fe-S protein 4], delays onset of neurological symptoms, reduces neuroinflammation, and prevents brain lesions. Although the precise mechanism of rescue remains to be determined, rapamycin induces a metabolic shift toward amino acid catabolism and away from glycolysis, alleviating the buildup of glycolytic intermediates. This therapeutic strategy may prove relevant for a broad range of mitochondrial diseases.

**L**eigh syndrome is a clinically defined disease resulting from genetic defects that disrupt mitochondrial function. It is the

most common childhood mitochondrial disorder, affecting 1 in 40,000 newborns in the United States (1). Leigh syndrome is characterized by



## Serial Femtosecond Crystallography of G Protein–Coupled Receptors

Wei Liu *et al.*

*Science* **342**, 1521 (2013);

DOI: 10.1126/science.1244142

*This copy is for your personal, non-commercial use only.*

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of July 14, 2015):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/342/6165/1521.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2013/12/19/342.6165.1521.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/342/6165/1521.full.html#related>

This article **cites 36 articles**, 8 of which can be accessed free:

<http://www.sciencemag.org/content/342/6165/1521.full.html#ref-list-1>

This article has been **cited by** 9 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/342/6165/1521.full.html#related-urls>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>



Received 7 November 2013

Accepted 16 December 2013

Edited by J. L. Smith, University of Michigan, USA

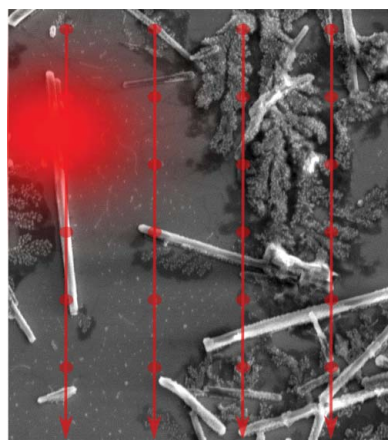
‡ These authors contributed equally to this study.

**Keywords:** protein microcrystallography; serial crystallography; *in vivo* grown microcrystals**PDB reference:** cathepsin B, 4n4z**Supporting information:** this article has supporting information at [www.iucrj.org](http://www.iucrj.org)Serial crystallography on *in vivo* grown microcrystals using synchrotron radiationCornelius Gati,<sup>a,‡</sup> Gleb Bourenkov,<sup>b,‡</sup> Marco Klinge,<sup>c</sup> Dirk Rehders,<sup>c</sup> Francesco Stellato,<sup>a</sup> Dominik Oberthür,<sup>a,d</sup> Oleksandr Yefanov,<sup>a</sup> Benjamin P. Sommer,<sup>d,e</sup> Stefan Mogk,<sup>e</sup> Michael Duszenko,<sup>e</sup> Christian Betzel,<sup>d</sup> Thomas R. Schneider,<sup>b,\*</sup> Henry N. Chapman<sup>a,f,\*</sup> and Lars Redecke<sup>c,\*</sup><sup>a</sup>Center for Free-Electron Laser Science (CFEL), Deutsches Elektronensynchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany, <sup>b</sup>European Molecular Biology Laboratory (EMBL), Hamburg Outstation, Notkestrasse 85, 22607 Hamburg, Germany, <sup>c</sup>Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, Notkestrasse 85, 22607 Hamburg, Germany, <sup>d</sup>Institute of Biochemistry and Molecular Biology, University of Hamburg, Notkestrasse 85, 22607 Hamburg, Germany, <sup>e</sup>Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse 4, 72076 Tübingen, Germany, and <sup>f</sup>Institute of Experimental Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. \*Correspondence e-mail: [thomas.schneider@embl-hamburg.de](mailto:thomas.schneider@embl-hamburg.de), [henry.chapman@desy.de](mailto:henry.chapman@desy.de), [redecke@biochem.uni-luebeck.de](mailto:redecke@biochem.uni-luebeck.de)

Crystal structure determinations of biological macromolecules are limited by the availability of sufficiently sized crystals and by the fact that crystal quality deteriorates during data collection owing to radiation damage. Exploiting a micrometre-sized X-ray beam, high-precision diffractometry and shutterless data acquisition with a pixel-array detector, a strategy for collecting data from many micrometre-sized crystals presented to an X-ray beam in a vitrified suspension is demonstrated. By combining diffraction data from 80 *Trypanosoma brucei* procathespin B crystals with an average volume of 9  $\mu\text{m}^3$ , a complete data set to 3.0 Å resolution has been assembled. The data allowed the refinement of a structural model that is consistent with that previously obtained using free-electron laser radiation, providing mutual validation. Further improvements of the serial synchrotron crystallography technique and its combination with serial femtosecond crystallography are discussed that may allow the determination of high-resolution structures of micrometre-sized crystals.

## 1. Introduction

Macromolecular crystallography (MX) is a powerful method for obtaining structural information about biological macromolecules and their assemblies. Since the 1990s, advanced third-generation synchrotrons have been used to produce micrometre-sized high-flux X-ray beams whose focus size matches the size of small crystals (Cusack *et al.*, 1998; Riek *et al.*, 2005; Evans *et al.*, 2011; Smith *et al.*, 2012). X-ray beams with dimensions of less than 10  $\mu\text{m}$  are now in routine use at many synchrotron-radiation facilities (Evans *et al.*, 2011) and enable the determination of crystal structures from crystals with volumes of less than 1000  $\mu\text{m}^3$  (Cusack *et al.*, 1998). Using these microbeams, the structures of the cypovirus polyhedra protein (Coulbaly *et al.*, 2007), amyloid-like fibres (Nelson *et al.*, 2005) and a number of complexes addressing the structure and function of G-protein-coupled receptors (GPCRs; Cherezov *et al.*, 2007; Rasmussen *et al.*, 2007, 2011) have been determined from micrometre-sized crystals.



Inelastic and absorption events deposit energy in the crystal, triggering a variety of chemical reactions that lead to a repositioning of atoms (during the experiment), ultimately reducing the crystalline order and corresponding to increasing levels of ‘radiation damage’. As a consequence, even under cryogenic conditions, the quality of the diffraction data, in particular those at high resolution, quickly deteriorates as a function of the X-ray dose applied to the crystal. Since the earliest days of macromolecular crystallography, this problem has been addressed by combining partial data sets collected from multiple crystals into a full data set to high resolution. In the 1990s cryogenic cooling became routine in macromolecular crystallography (Henderson, 1990; Garman & Schneider, 1997), extending the lifetime of individual crystals, but with the implementation of undulator beamlines on third-generation synchrotrons the tolerable X-ray dose could be deposited in a short time even for cryocooled crystals (Ravelli & Garman, 2006). Data-collection schemes employing multiple positions on a single crystal (Riekel *et al.*, 2005), to distribute the delivered dose over the entire crystal, or partial data sets from multiple crystals (Brodersen *et al.*, 2003) were developed. The main complication with multicrystal methods is that systematic differences between data measured on different crystals, or at different positions on a single crystal, can prevent the derivation of a consistent assembled data set. Nevertheless, Hendrickson and coworkers (Liu *et al.*, 2012) have recently shown that by removing sources of experimental error and by carefully selecting consistent partial data sets for merging, data sets of sufficient accuracy for phasing using the anomalous signal from S atoms in native proteins can in fact be assembled. Systematic errors in data collections from multiple positions in the same crystal can, amongst other techniques, be reduced by ‘helical scan’ procedures (Flot *et al.*, 2010), in which a single crystal, usually needle-shaped, is continuously moved during data collection to expose fresh parts of the same crystal.

When working with small crystals, it is often not possible to locate the crystals accurately using optical microscopy owing to their small size, optical distortions caused by materials surrounding the crystal, or the crystal being embedded in an opaque matrix. All of these problems are present during the data collection of GPCR crystals grown from lipidic cubic phase (LCP). Cherezov *et al.* (2009) described a method in which the mounted LCP containing crystals is first systematically rastered using a small beam to test each position for diffraction. Data are then collected in a second step using the rotation method (Arndt & Wonacott, 1977) from positions where diffraction was actually detected. The application of this method has been crucial in the structure determination of GPCRs, where the crystals are small (of the order of  $25 \times 6 \times 4 \mu\text{m}^3$ ) and invisible to optical light because the lipidic mesophase in which the crystals grow turns opaque upon cryogenic cooling. Recently, an alternative procedure was proposed in which radiographs are taken of the entire sample mount to localize crystals (Warren *et al.*, 2013).

Extremely short pulses of X-rays, each of tens of femtoseconds duration or even shorter, generated by a free-electron

laser (FEL), have recently been shown to overcome the dose limitations mentioned above. Referred to as ‘diffraction before destruction’, the inertia of the atoms prevents significant atomic displacement during the short exposing pulse (Neutze *et al.*, 2000; Boutet *et al.*, 2012; Barty *et al.*, 2012), even though the initial photoabsorption eventually leads to plasma formation and the complete vaporization of the sample. A typical serial femtosecond crystallography (SFX) experiment collects tens or hundreds of thousands of single-crystal X-ray diffraction ‘snapshots’ from a liquid suspension of protein microcrystals that flows across the focused X-ray beam at room temperature (Chapman *et al.*, 2011). Since each FEL pulse destroys the sample, only one diffraction pattern can be collected per crystal. New data-processing tools, such as *Cheetah*, *CrystFEL* (White *et al.*, 2012, 2013; Kirian *et al.*, 2011) and *cctbx.xfel* (Kern *et al.*, 2013), have been developed to process these large data volumes and have been successfully applied to several examples (Chapman *et al.*, 2011; Boutet *et al.*, 2012; Redecke *et al.*, 2013; Kern *et al.*, 2013).

The crystal structure of *Trypanosoma brucei* procathepsin B (TbCatB) in complex with its native propeptide represents the first novel bioinformation obtained by applying the SFX approach at a free-electron laser (Redecke *et al.*, 2013). This enzyme is of scientific and medical interest, since the knockout of its encoding gene has been shown to be lethal to the parasite that causes sleeping sickness in Africa (Abdullah *et al.*, 2008; Bryant *et al.*, 2009), which turns TbCatB into an urgently required potential new drug target (Fairlamb, 2003). TbCatB crystals grew spontaneously within living baculovirus-infected insect cells during protein overexpression to a size of 10–15  $\mu\text{m}$  in the longest dimension (Koopmann *et al.*, 2012). In a recent experiment (Redecke *et al.*, 2013), these crystals were used to obtain 178 875 single-crystal diffraction patterns by SFX at LCLS, which enabled structure determination of the enzyme to 2.1 Å resolution.

Inspired by the SFX methodology and the new capabilities to process large data sets, as well as the successful structure determination of TbCatB in the pro-form using the SFX approach, we embarked on an experiment to determine the crystal structure of *T. brucei* procathepsin B using a suspension of *in vivo* grown microcrystals mounted in a standard nylon loop for crystallographic data collection on a microfocus synchrotron beamline. Owing to the presence of cell remnants in the *in vivo* crystal suspension, individual TbCatB crystals are difficult to detect in the cryocooled sample. Our procedure combined elements from SFX with a helical line-scan approach used in microcrystallography. As with the SFX approach, an initially unknown subset of the recorded detector frames contain diffraction signals, which are selected for further processing into a crystallographic data set. Unlike the snapshots recorded at an FEL, the sample is rotated during exposure and multiple exposures from the same crystal can be obtained and subsequently processed in a consistent manner. In the following, we describe the diffraction experiment, data processing and structure determination. The independent determination of the crystal structure of procathepsin B from *T. brucei* via two different methods using

synchrotron and FEL radiation sources provided a unique opportunity to validate the results obtained against each other, and we present a comparison of the crystallographic models.

## 2. Results

### 2.1. Sample preparation and data collection

Spontaneous crystallization of TbCatB was obtained in baculovirus-infected Sf9 insect cells following our previously established protocol (Fig. 1; Koopmann *et al.*, 2012). The *in vivo* crystals were isolated and purified by cell lysis and stepwise centrifugation based on their high mechanical and chemical stability. The purity of the crystal preparation was verified by scanning electron microscopy (SEM). Prior to diffraction data collection, the suspension of needle-shaped crystals containing approximately  $5 \times 10^8$  crystals per millilitre with average dimensions of  $0.9 \times 0.9 \times 11 \mu\text{m}$  (approximately  $9 \mu\text{m}^3$  in volume) was supplemented with 40% (v/v) glycerol as a cryoprotectant. A small volume of approximately 13 nl of the crystalline suspension was mounted after settling in a standard  $20 \mu\text{m}$  thick nylon loop (Hampton Research, USA; 0.7 mm diameter) containing approximately 5000 crystals.

Diffraction experiments were conducted on the P14 microfocus beamline at the PETRA III storage ring (DESY, Hamburg) with a  $4 \times 5 \mu\text{m}$  (FWHM) microfocus beam, a total photon flux of  $1.2 \times 10^{12}$  photons  $\text{s}^{-1}$  at the sample position and a photon energy of 10.00 keV. The nylon loop was mounted on an MK3 mini-kappa goniometer head attached

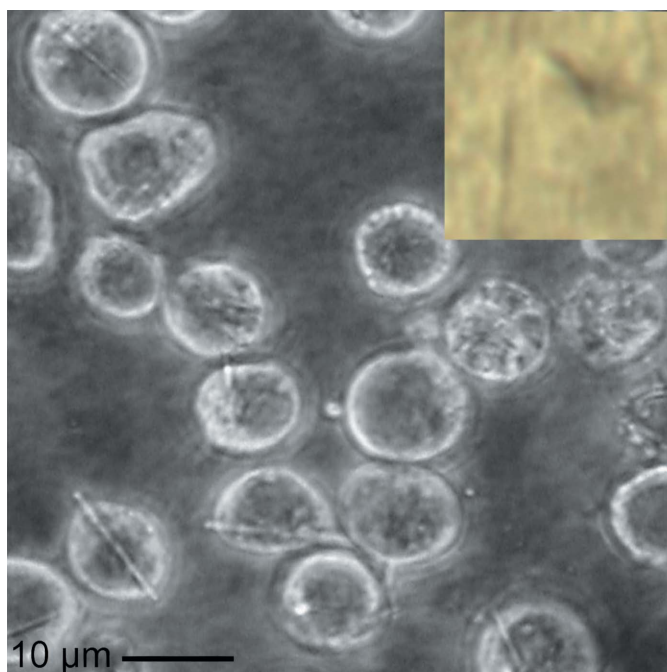
to an MD3 microdiffractometer (ARINAX, Moirans, France) and kept at 110 K using a gaseous nitrogen stream (Cryojet XL, Oxford Instruments, England). Diffraction data were recorded on a PILATUS 6M-F detector (DECTRIS Ltd, Baden, Switzerland).

In a set of initial experiments, the diffraction properties of individual TbCatB crystals were characterized. After optically centring selected crystals with respect to the X-ray beam, single diffraction images collected with an oscillation range of  $1^\circ$  at an exposure time of 2 s showed diffraction spots extending to a resolution higher than  $3 \text{ \AA}$  (Supporting Fig. S2). Using a series of short rotation exposures, by visual inspection of diffraction images and using the ability to index them as a criterion (see Supporting Information), we empirically determined a crystal lifetime of 0.5–1 s under the conditions at hand. This observed crystal lifetime corresponded well to the lifetime of 0.88 s calculated *via RADDOS* (Paithankar & Garman, 2010) at an estimated maximum dose rate of  $34 \text{ MGy s}^{-1}$ .

To collect a complete data set, we employed a data-collection strategy in which a region of interest of  $600 \times 600 \mu\text{m}$  was raster scanned with rotation exposures (Fig. 2). 120 parallel helical scans were performed, spaced  $5 \mu\text{m}$  apart. During each helical scan, the goniostat was rotated from  $\Omega = -45^\circ$  to  $\Omega = +45^\circ$  (where  $\Omega = 0^\circ$  corresponds to the orientation of the loop surface perpendicular to the incoming beam) and translated by  $600 \mu\text{m}$ . Taking 240 exposures of 1 s duration over the course of a each helical scan, each individual frame recorded on the detector corresponded to a rotation of  $0.375^\circ$  and a translation of  $2.5 \mu\text{m}$  of the sample. Under these conditions, every crystal within the region of interest received a dose of between 50 and 60 MGy. This high total (integrated over multiple exposures) dose was chosen to collect the highest possible resolution data from each crystal. Using the above strategy, 28 800 detector frames were acquired in a period of 8 h.

### 2.2. Data processing and structure determination

Adopting the recently established methods for the processing of diffraction patterns from SFX experiments, the *CrystFEL* software suite (White *et al.*, 2012) was used as a first step to identify and index single-crystal diffraction patterns within the large set of detector frames. Frames containing diffraction patterns that were recorded consecutively during the same helical scan were considered as originating from the same crystals and were assembled into 595 groups containing between two and ten consecutive frames (Fig. 2d and Supporting Fig. S2). These groups were further treated as regular rotation data for re-indexing and integration applying *XDS* (Kabsch, 2010). Horizontally adjacent groups of diffraction images, potentially containing diffraction patterns from the same crystal, were treated independently. In a standard three-dimensional profile-fitting procedure, both fully and partially recorded reflections were integrated. Processing was successful for 130 groups containing a total of 557 frames.

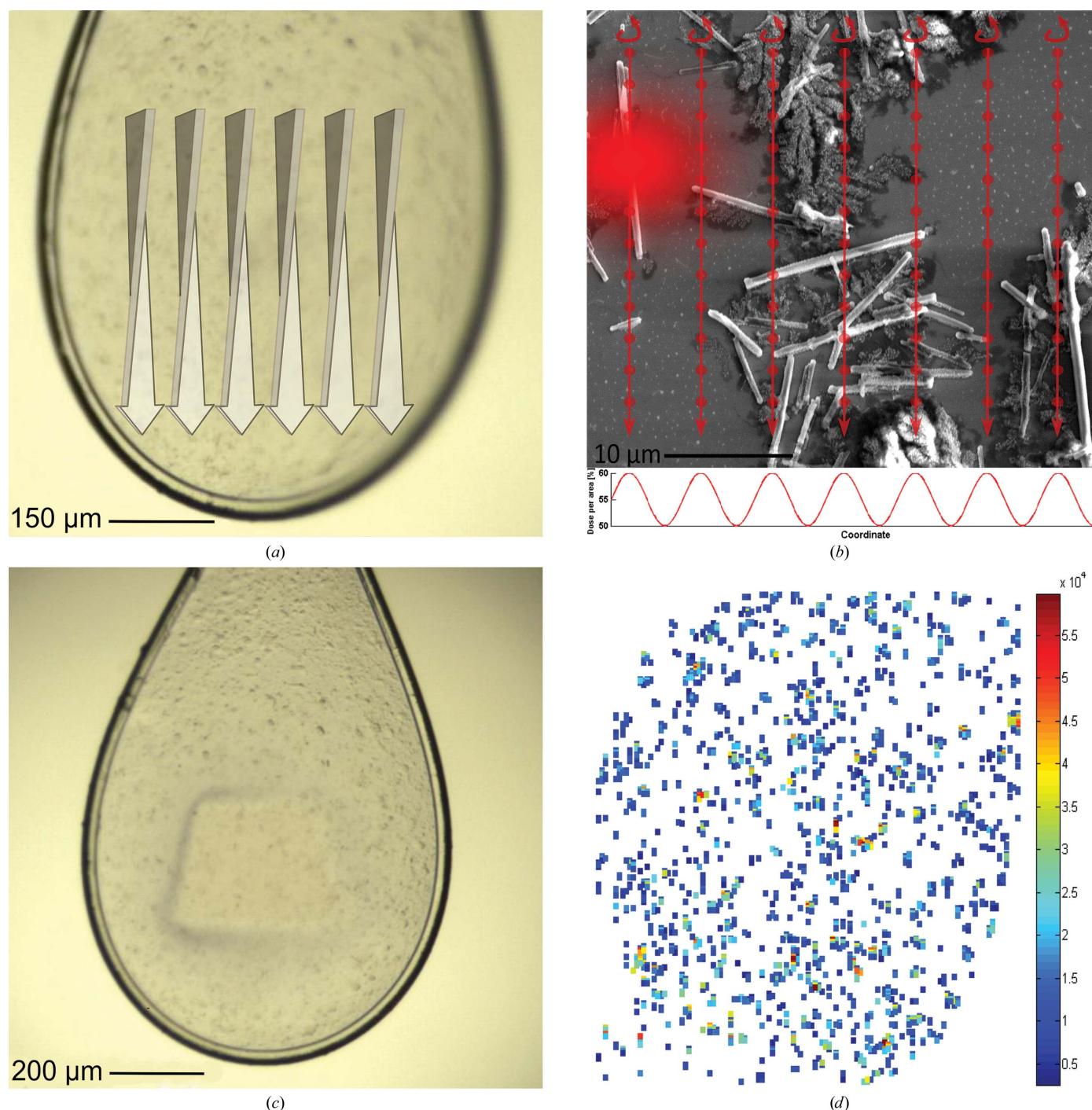


**Figure 1**  
Light micrograph of Sf9 cells spontaneously crystallizing trypanosomal cathepsin B. The isolated and purified crystals (inset) were mounted on a standard cryoloop for the serial synchrotron diffraction experiments.



After iterative merging and scaling, 109 661 reflection intensities in the resolution range from 88 to 3.0 Å were merged into a final data set consisting of 8881 merged

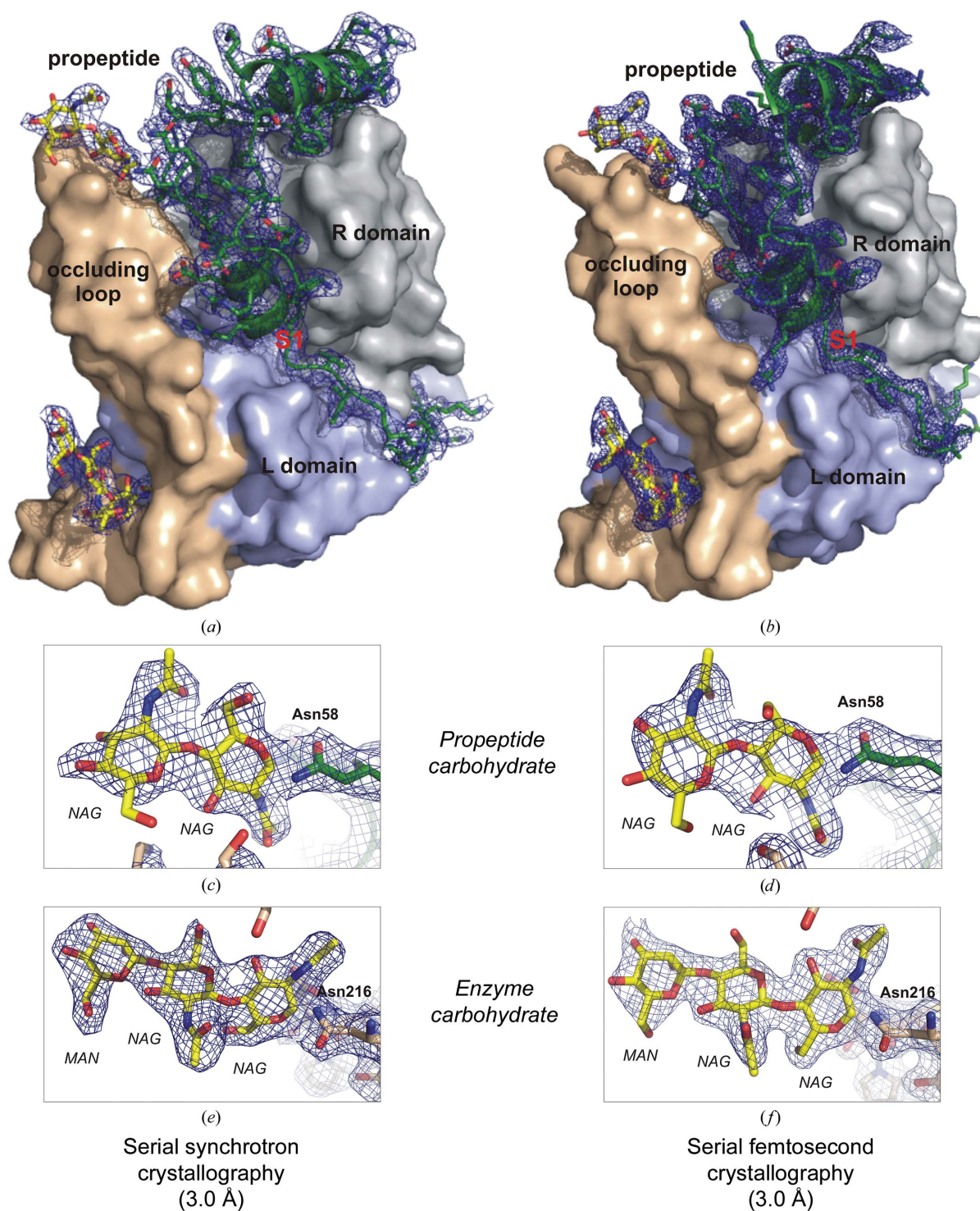
reflection intensities with an overall completeness of 99.8%. This final set of reflection intensities included data from 426 diffraction patterns collected from 80 individual TbCatB



**Figure 2** Experimental setup of the serial synchrotron crystallography experiment. (a) Schematic macroscopic illustration of the serial helical line-scan approach using a standard cryogenic loop, imaged with the inline microscope. (b) SEM image of isolated *in vivo* grown cathepsin B microcrystals on a silicon support. Red arrows illustrate the serial helical line scan. The incident beam is represented by the red 'flare'. The colour density in the flare is proportional to a calculated two-dimensional Gaussian function with FWHM  $4 \times 5 \mu\text{m}$ , with relative size to the  $10 \mu\text{m}$  scale bar, showing a significant fraction of photon flux away from the centre of the beam. Red dots illustrate the positions of collected frames during the line scan with an oscillation width of  $0.5^\circ$  each. The graph (lower part) visualizes the delivered dose per area against arbitrary coordinates, indicating a total dose per area fluctuating between 50 and 60% owing to the ratio of FWHM of the beam and the gap between each line-scan position. (c) After the serial helical line scan, the photoinduced ionization at the exposed part of the sample is macroscopically visible. (d) Heatmap of diffraction images in the crystal loop after pre-selection using *CrystFEL*. The colour bar codes the average intensity of Bragg peaks in each diffraction pattern as an indication of the diffraction strength in each pattern.

crystals in 120 groups. The distribution of the size of the groups (Supporting Fig. S2) apparently reflects the variation in the crystal size in TbCatB preparations (see Fig. S1 of

Redecke *et al.*, 2013). Most of the data were derived from groups of three to five consecutive frames corresponding to a total rotation range of 1.125–1.875° of one crystal. A small



**Figure 3**

Quality of the calculated electron density from diffraction data sets of *in vivo* grown TbCatB crystals collected using serial synchrotron crystallography (3.0 Å resolution; left) and SFX (refined at 3.0 Å resolution; PDB entry 4hwy; right) techniques. (a, b) Surface representation of the TbCatB–propeptide complexes independently solved by molecular replacement using the mature TbCatB structure (Koopmann *et al.*, 2012) as a search model. The solutions consistently revealed additional electron density ( $2F_{\text{obs}} - F_{\text{calc}}$ ,  $1\sigma$ , blue) of the propeptide (green) that is bound to the V-shaped substrate-binding cleft and of two carbohydrate structures (yellow) N-linked to the propeptide (c, d) and to the mature enzyme (e, f). Considering the difference in maximum resolution, the propeptide, as well as both carbohydrates, are well defined within the electron-density maps, confirming that the phases are not biased by the search model.



fraction of data originated from groups containing eight to ten frames, while the majority of groups contained three to five consecutive frames.

The quality and internal consistency of the data were judged on the basis of standard  $\langle I/\sigma(I) \rangle$  statistics and on the basis of the CC\* criteria recently advocated as a single statistically valid guide for deciding the resolution cutoff of the obtained data (McCoy *et al.*, 2007; Karplus & Diederichs, 2012; Evans, 2012). The CC\* calculated in resolution shells for the TbCatB data set (Supporting Fig. S1) indicated the presence of statistically significant data to a resolution of 3.0 Å and below.

Following the same strategy as for the previous determination of the *T. brucei* pro-cathepsin B crystal structure *via* SFX (Redecke *et al.*, 2013), initial phases were obtained by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the structure of the nonglycosylated and *in vitro* crystallized TbCatB (PDB entry 3mor; Koopmann *et al.*, 2012) that lacks the propeptide and the carbohydrate chains as a search model. During stepwise model building and refinement, 62 propeptide residues and five carbohydrate residues were manually placed in difference electron-density maps.

The refined TbCatB structure ( $R$  factor = 22.3%,  $R_{\text{free}}$  = 26.4%) shares the papain-like fold which is characteristic of cathepsin B enzymes, including the propeptide residues 27–72 and 79–85 without defined electron density in between, as well as a carbohydrate chain consisting of two *N*-acetylglucosamine (NAG) monomers N-linked to Asn58 (in the propeptide) and another carbohydrate chain consisting of two NAG monomers and one  $\beta$ -mannose (BMA) molecule N-linked to Asn216 of the enzyme domain. Overall, the 3.0 Å resolution electron-density map is well defined by the TbCatB model. No electron density is observed for nine flexible amino-acid side chains mainly located within a loop region spanning residues His195–Asn209 or for ten atoms of the carbohydrate structures. In particular, as for the SFX structure determination, the expected features of the electron-density map that were not part of the search model are well defined by the propeptide and two carbohydrate chains after manual model building and refinement (Fig. 3).

### 2.3. Comparison of the structural TbCatB models

For detailed comparison of the *T. brucei* procathepsin B structure solved in this study at 110 K using synchrotron radiation with that previously obtained at room temperature using the FEL-based SFX technique (PDB entry 4hwy; Redecke *et al.*, 2013), electron-density maps were generated using the SFX data truncated at 3.0 Å resolution. Applying an identical refinement protocol that omits solvent atoms resulted in an  $R$  factor of 17.0% ( $R_{\text{free}}$  = 19.6%). A slight shrinking of the unit-cell parameters of the TbCatB *in vivo* crystals observed for the synchrotron data set (Table 1) can be attributed to the cryogenic data-collection conditions. At room temperature, unit-cell parameters of  $a = b = 125.5$ ,  $c = 54.6$  Å were previously obtained by SFX. The superposition of the peptide backbone atoms of both structures revealed a high degree of consistency, resulting in an average r.m.s.d. value of  $0.35 \pm 0.19$  Å, which is comparable to the

**Table 1**

X-ray data-collection and refinement statistics for *in vivo* crystallized TbCatB analyzed at the P14 beamline of the PETRA III synchrotron source (DESY, Hamburg, Germany).

Values in parentheses are for the highest resolution shell.

Data collection	
Light source, beamline	PETRA III, P14
Maximum dose (MGy)	50–60
Space group	$P4_22_12$
Unit-cell parameters (Å)	$a = b = 123.5$ , $c = 54.3$
$V_M$ (Å <sup>3</sup> Da <sup>-1</sup> )	2.99
Solvent content (%)	58.6
Resolution range (Å)	88.1–3.0 (3.16–3.00)
No. of unique reflections	8881
Completeness (%)	99.8 (99.9)
$R_{\text{merge}}$	0.71 (2.69)
$\langle I/\sigma(I) \rangle$	3.7 (1.0)
CC*	0.97 (0.79)
Multiplicity	12.3 (12.6)
Refinement	
Resolution range (Å)	88.1–3.0
No. of reflections used in refinement	8482
No. of reflections used for $R_{\text{free}}$	399
$R_{\text{work}}/R_{\text{free}}$	0.223/0.264
No. of atoms	
Protein	2392
Carbohydrate	67
$B$ factors (Å <sup>2</sup> )	
Protein (main chain/side chain)	38/43
Carbohydrate	54
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.32
Average r.m.s. $B$ factor (main/side chain)	1.6/1.8
Ramachandran plot (%)	
Most favoured	91.2
Allowed	8.2
Disallowed	0.66

overall coordinate error of 0.32 Å estimated based on maximum likelihood by *REFMAC5.5* (Murshudov *et al.*, 2011). No significant structural differences are present, including no major features related to radiation damage (Supporting Fig. S3). Main-chain deviations of more than 0.8 Å are limited to nine residues located at the N-terminus and C-terminus, in flexible loop regions and at positions flanking the disordered part of the propeptide region that results from an increased flexibility of the residues after proteolytic cleavage between Ser78 and Ile79 (Redecke *et al.*, 2013). Even the two carbohydrate chains are clearly defined and largely superimposable between the two models (Figs. 3c–3f). Slight differences were only observed for the second *N*-acetylglucosamine residue of the propeptide carbohydrate, which represents the most flexible carbohydrate within the model. This is further reflected by the almost identical number of amino-acid side chains/carbohydrate atoms not defined by electron density in both TbCatB structures (nine side chains and ten carbohydrate atoms in this structure *versus* 11 side chains and eight carbohydrate atoms in the SFX structure).

Despite the overall similarity in atomic coordinates, systematic differences were observed in the relative heights of the electron-density peaks at the 12 Cys SG atoms involved in disulfide bridges. Considering refined Debye–Waller factors as an (anticorrelated) measure of the height of electron-density maxima, we note that in the synchrotron structure the average

Debye–Waller factor ( $\langle B_{\text{SG}} \rangle = 57 \text{ \AA}^2$ ) is higher than that averaged over all atoms ( $\langle B_{\text{All}} \rangle = 41 \text{ \AA}^2$ ). In the SFX TbCatB structure refined using an identical protocol at 3.0 Å resolution,  $\langle B_{\text{SG}} \rangle$  (38 Å<sup>2</sup>) is lower than  $\langle B_{\text{All}} \rangle$  (45 Å<sup>2</sup>). This observation is consistent with a significant diffraction contribution from reduced disulfide bonds in the synchrotron data but not in the SFX data.

### 3. Discussions and conclusions

Particularly for crystals with dimensions in the low micrometre range, the determination of macromolecular crystal structures is inherently limited by radiation damage. In most cases, when an X-ray flux sufficient to measure Bragg reflections to the highest resolution as defined by the degree of crystalline order in a given crystal is used, the crystal will be severely damaged before complete diffraction data can be collected. As a consequence, in practice a compromise is sought balancing the resolution and the completeness of the data to be measured on a single crystal. In recent years, combining data from multiple crystals has enabled the determination of a number of important structures (Rasmussen *et al.*, 2011; Siu *et al.*, 2013; Li *et al.*, 2013), despite the difficulties arising from systematic errors when data from multiple crystals are merged. The recent introduction of SFX exploiting X-rays from an FEL to collect single diffraction images to the maximum resolution from large numbers of crystals in the ‘diffraction-before-destruction’ regime has realised an extreme approach to overcoming the radiation-damage problem.

Here, we have demonstrated a strategy for the collection of complete diffraction data close to the diffraction limit from micrometre-sized crystals using synchrotron radiation. This method is based on the serial illumination of subvolumes of a sample consisting of a cryogenically vitrified suspension of microcrystals mounted in a standard nylon loop. During exposure, the sample is rotated, resulting in ‘classical’ rotation frames (Arndt & Wonacott, 1977) for subvolumes presenting crystalline material to the X-ray beam. The X-ray dose received by an exposed subvolume is chosen to fully exploit the crystal lifetime during the exposure time used, a requirement to achieve maximum resolution under these conditions. The combination of microcrystals assuming quasi-random orientations in the suspension and continuous rotation of the loop during each helical scan will present a variety of different crystal orientations to the beam, effectively covering the rotation space and thus providing complete diffraction data in the end.

The crystals used in this study are smaller in volume ( $\sim 10^7$  unit cells in a crystal of volume  $9 \mu\text{m}^3$ ) than those of cypovirus polyhedra (CPV) studied by Metcalf and coworkers (Coulibaly *et al.*, 2007;  $\sim 10^8$  unit cells in  $125 \mu\text{m}^3$ ). In addition, the TbCatB crystals are embedded in a matrix giving rise to a high scattering background, similar to the situation of GPCR crystals mounted in LCP, where the crystals are commonly of larger volume ( $\sim 10^9$  unit cells in  $600 \mu\text{m}^3$ ; Cherezov *et al.*, 2007). While systematic strategies collecting diffraction data on individually pre-centred crystals were employed for both the CPV and GPCR cases, the serial strategy suggested here

has delivered data of comparable quality in terms of signal to noise of the measured diffraction intensities without the need to identify and centre micrometre-sized crystals before data collection (see Supporting Information).

The serial synchrotron-radiation diffraction data allowed us to phase and refine the TbCatB structure using standard technologies. With respect to the data set collected from TbCatB using FEL radiation, the data collected at the synchrotron extend to lower resolution, reflecting the difference between measuring radiation-damage-free ‘infinitesimally still’ data with femtosecond laser pulses and a macroscopic rotation of the crystal over  $0.375^\circ$ , with a similar dose in both cases. Additionally, different levels of background influence the data quality. While a liquid jet of approximately 4 μm diameter delivered the sample for SFX, a 20 μm nylon loop suspending a film of 20 μm thickness was used to mount the crystals in this study. The structural model built against the synchrotron data is consistent with the model obtained with FEL radiation, providing mutual validation.

The present implementation of the serial synchrotron strategy can be improved in many ways. The embedding of the TbCatB crystals with a maximum dimension of approximately 10 μm in a vitrified matrix of approximately 20 μm thickness consisting of buffer and cell debris causes significant background scatter that could be minimized by employing different mounting technologies, such as adaptations of cryo-EM techniques (Nederlof *et al.*, 2013) or graphene supports (Wierman *et al.*, 2013). Reducing the beam size to  $1 \times 1 \mu\text{m}$  will reduce the background scattering 20-fold. Increasing the flux density of the X-ray beam will reduce the data-collection time, which is currently several hours for a single loop-mounted drop of crystalline suspension. This can be accomplished using improved X-ray optics and the implementation of wide-bandpass monochromators. Merging data collected on multiple loops would increase the signal-to-noise ratio, while decreasing the rotation increment per X-ray dose unit could increase the resolution to which diffraction data can be recorded closer to the limit posed by the degree of order in the crystal.

The method presented here is conceptually simple and could be implemented at many microfocus synchrotron-radiation beamlines employing existing helical scan schemes. It lends itself to data collection on small crystals in suspension, such as those obtained from *in vivo* preparations, as it avoids the centring of hardly visible (or invisible) crystals. Parameters can be tuned to maximize accuracy (*e.g.* applying a larger rotation range per exposed subvolume may allow more accurate integration and scaling) or to maximize resolution (by using a smaller rotation range during the application of the tolerable X-ray dose).

In addition to the promising application as a standalone approach, the combination of serial synchrotron and SFX data collected for a given crystallized protein further offers a new strategy for scaling and phasing of SFX data. In comparison to SFX data collection, serial synchrotron crystallography allows the extraction of accurate diffraction data, albeit to lower resolution owing to the finite rotation range and the onset of

radiation damage during the exposure, from a small number of microcrystals by the systematic acquisition of structure-factor amplitudes followed by the application of well defined scaling models modelling a finely controlled experimental process. In contrast, at present, a three orders of magnitude larger number of microcrystals is required for the convergence of Monte Carlo intensity integration when arithmetic means of partially recorded intensities are used without scaling (Kirian *et al.*, 2011). The use of complete and accurate low-resolution data sets obtained using synchrotron radiation for bootstrapping scaling procedures for SFX data could improve the convergence behaviour of these procedures. If diffraction data can be collected on the same system using X-rays from both synchrotron and free-electron laser sources, the combined use of these data therefore has the potential to provide more accurate crystallographic data than those originating from only one of the two methods, ultimately resulting in higher quality macromolecular structures from micrometre-sized crystals.

### Acknowledgements

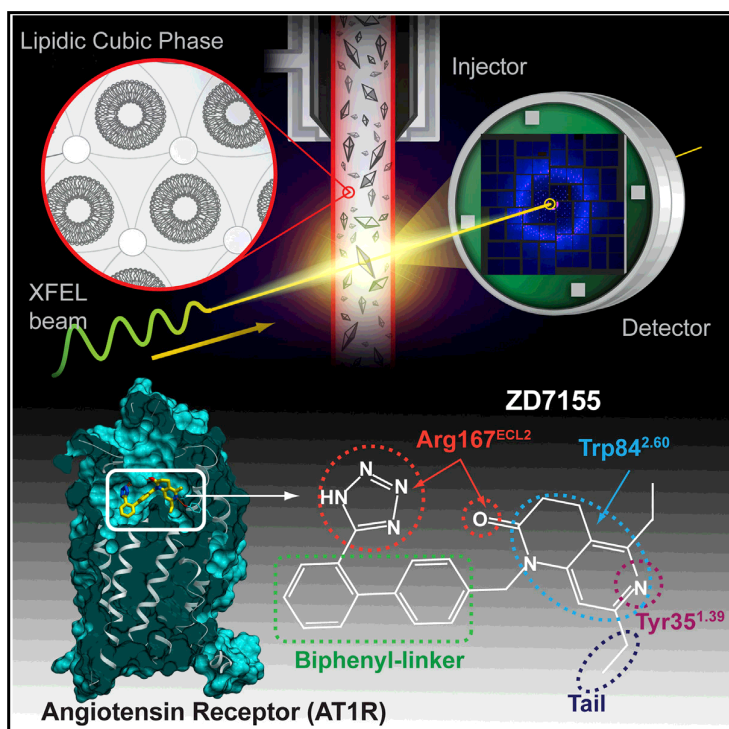
The X-ray diffraction experiments were carried out at beamline P14 of the PETRA III synchrotron source operated by the European Molecular Biology Laboratory (EMBL) at the German Electron Synchrotron (DESY) in Hamburg, Germany in May and June 2013. We gratefully acknowledge discussions with Ilme Schlichting, and we thank Thomas White for help with using *CrystFEL*. CG acknowledges support from the PIER Graduate School, Helmholtz Association Scholarship. LR, MK, DR and CB thank the German Federal Ministry for Education and Research (BMBF) for funding (grants 01KX0806 and 01KX0807). BPS, DO, LR, MD, CB and HNC acknowledge support from the BMBF in the context of the Röntgen-Angström-Cluster (grant 05K12GU3). Support from the Hamburg Ministry of Science and Research and Joachim Herz Stiftung as part of the Hamburg Initiative for Excellence in Research (LEXI) and the Hamburg School for Structure and Dynamics (SDI) as well as the DFG Cluster of Excellence 'Inflammation at Interfaces' (EXC 306) is gratefully acknowledged.

### References

- Abdullah, M. H., O'Brien, T., Mackey, Z. B., Sajid, M., Grab, D. J. & McKerrow, J. H. (2008). *PLoS Negl. Trop. Dis.* **2**, 298–304.
- Arndt, U. M. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North Holland.
- Barty, A. *et al.* (2012). *Nature Photonics*, **6**, 35–40.
- Boutet, S. *et al.* (2012). *Science*, **20**, 362–364.
- Brodersen, D. E., Clemons, W. M., Carter, A. P., Wimberly, B. T. & Ramakrishnan, V. (2003). *Acta Cryst.* **D59**, 2044–2050.
- Bryant, C., Kerr, I. D., Debnath, M., Ang, K. K. H., Ratnam, J., Ferreira, R. S., Jaishankar, P., Zhao, D., Arkin, M. R., McKerrow, J. H., Brinen, L. S. & Renslo, A. R. (2009). *Bioorg. Med. Chem. Lett.* **19**, 6218–6221.
- Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
- Cherezov, V., Hanson, M. A., Griffith, M. T., Hilgart, M. C., Sanishvili, R., Nagarajan, V., Stepanov, S., Fischetti, R. F., Kuhn, P. & Stevens, R. C. (2009). *J. R. Soc. Interface*, **6**, S587–S597.
- Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). *Science*, **318**, 1258–1265.
- Coulibaly, F., Chiu, E., Ikeda, K., Gutmann, S., Haebel, P. W., Schulze-Briese, C., Mori, H. & Metcalf, P. (2007). *Nature (London)*, **446**, 97–101.
- Cusack, S., Belrhali, H., Bram, A., Burghammer, M., Perrakis, A. & Riek, C. (1998). *Nature Struct. Biol.* **5**, 634–637.
- Evans, G., Axford, D., Waterman, D. & Owen, R. L. (2011). *Crystallogr. Rev.* **17**, 105–142.
- Evans, P. (2012). *Science*, **336**, 986–987.
- Fairlamb, A. H. (2003). *Trends Parasitol.* **19**, 488–494.
- Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
- Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.
- Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **25**, 1030–1033.
- Kern, J. *et al.* (2013). *Science*, **26**, 491–495.
- Kirian, R. A., White, T. A., Holton, J. M., Chapman, H. N., Fromme, P., Barty, A., Lomb, L., Aquila, A., Maia, F. R. N. C., Martin, A. V., Fromme, R., Wang, X., Hunter, M. S., Schmidt, K. E. & Spence, J. C. H. (2011). *Acta Cryst.* **A67**, 131–140.
- Koopmann, R. *et al.* (2012). *Nature Methods*, **9**, 259–262.
- Li, D., Lyons, J. A., Pye, V. E., Vogeley, L., Aragão, D., Kenyon, C. P., Shah, S. T., Doherty, C., Aherne, M. & Caffrey, M. (2013). *Nature (London)*, **497**, 521–524.
- Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancía, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Nederlof, I., Li, Y. W., van Heel, M. & Abrahams, J. P. (2013). *Acta Cryst.* **D69**, 852–859.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. Ø., Riek, C., Grothe, R. & Eisenberg, D. (2005). *Nature (London)*, **435**, 773–778.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature (London)*, **406**, 752–757.
- Paithankar, K. S. & Garman, E. F. (2010). *Acta Cryst.* **D66**, 381–388.
- Rasmussen, S. G. F., Choi, H.-J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I. & Kobilka, B. K. (2007). *Nature (London)*, **450**, 383–387.
- Rasmussen, S. G. F. *et al.* (2011). *Nature (London)*, **477**, 549–555.
- Ravelli, R. B. G. & Garman, E. F. (2006). *Curr. Opin. Struct. Biol.* **16**, 624–629.
- Redecke, L. *et al.* (2013). *Science*, **339**, 227–230.
- Riek, C., Burghammer, M. & Schertler, G. (2005). *Curr. Opin. Struct. Biol.* **15**, 556–562.
- Siu, F. Y. *et al.* (2013). *Nature (London)*, **499**, 444–449.
- Smith, J. L., Fischetti, R. F. & Yamamoto, M. (2012). *Curr. Opin. Struct. Biol.* **22**, 602–612.
- Warren, A. J., Armour, W., Axford, D., Basham, M., Connolley, T., Hall, D. R., Horrell, S., McAuley, K. E., Mykhaylyk, V., Wagner, A. & Evans, G. (2013). *Acta Cryst.* **D69**, 1252–1259.
- White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. A., Zatsepin, N. A. & Chapman, H. N. (2013). *Acta Cryst.* **D69**, 1231–1240.
- White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *J. Appl. Cryst.* **45**, 335–341.
- Wierman, J. L., Alden, J. S., Kim, C. U., McEuen, P. L. & Gruner, S. M. (2013). *J. Appl. Cryst.* **46**, 1501–1507.

# Structure of the Angiotensin Receptor Revealed by Serial Femtosecond Crystallography

## Graphical Abstract



## Authors

Haitao Zhang, Hamiyet Unal, ..., Sadashiva S. Karnik, Vadim Cherezov

## Correspondence

cherezov@usc.edu

## In Brief

Structure determination of human Angiotensin II type 1 receptor bound to an antagonist using serial femtosecond crystallography with X-ray free-electron laser and docking studies of other common anti-hypertensive drugs into the structure offer insights into design of blood pressure modulators.

## Highlights

- Crystal structure of the human Angiotensin II type 1 receptor at 2.9-Å resolution
- Structure is solved by X-ray laser serial femtosecond crystallography
- Antagonist ZD7155 forms critical interactions with Tyr35, Trp84 and Arg167
- Docking reveals binding modes of common angiotensin receptor blockers



# Structure of the Angiotensin Receptor Revealed by Serial Femtosecond Crystallography

Haitao Zhang,<sup>1</sup> Hamiyet Unal,<sup>2</sup> Cornelius Gati,<sup>3</sup> Gye Won Han,<sup>4</sup> Wei Liu,<sup>5</sup> Nadia A. Zatsepin,<sup>6</sup> Daniel James,<sup>6</sup> Dingjie Wang,<sup>6</sup> Garrett Nelson,<sup>6</sup> Uwe Weierstall,<sup>6</sup> Michael R. Sawaya,<sup>7</sup> Qingping Xu,<sup>8</sup> Marc Messerschmidt,<sup>9</sup> Garth J. Williams,<sup>10</sup> Sébastien Boutet,<sup>10</sup> Oleksandr M. Yefanov,<sup>3</sup> Thomas A. White,<sup>3</sup> Chong Wang,<sup>11</sup> Andrii Ishchenko,<sup>4</sup> Kalyan C. Tirupula,<sup>2</sup> Russell Desnoyer,<sup>2</sup> Jesse Coe,<sup>5</sup> Chelsie E. Conrad,<sup>5</sup> Petra Fromme,<sup>5</sup> Raymond C. Stevens,<sup>1,4,12</sup> Vsevolod Katritch,<sup>1</sup> Sadashiva S. Karnik,<sup>2</sup> and Vadim Cherezov<sup>4,\*</sup>

<sup>1</sup>Department of Biological Sciences, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA

<sup>2</sup>Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

<sup>3</sup>Center for Free Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany

<sup>4</sup>Department of Chemistry, Bridge Institute, University of Southern California, Los Angeles, CA 90089 USA

<sup>5</sup>Department of Chemistry and Biochemistry, Center for Applied Structural Discovery at the Biodesign Institute, Arizona State University, Tempe, AZ 85287, USA

<sup>6</sup>Department of Physics, Arizona State University, Tempe, AZ 85287, USA

<sup>7</sup>Department of Chemistry and Biochemistry, the UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA

<sup>8</sup>Joint Center for Structural Genomics, Stanford Synchrotron Radiation Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>9</sup>BioXFEL Science and Technology Center, Buffalo, NY 14203, USA

<sup>10</sup>Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

<sup>11</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

<sup>12</sup>iHuman Institute, ShanghaiTech University, Shanghai, 201210 China

\*Correspondence: [cherezov@usc.edu](mailto:cherezov@usc.edu)

<http://dx.doi.org/10.1016/j.cell.2015.04.011>

## SUMMARY

Angiotensin II type 1 receptor (AT<sub>1</sub>R) is a G protein-coupled receptor that serves as a primary regulator for blood pressure maintenance. Although several anti-hypertensive drugs have been developed as AT<sub>1</sub>R blockers (ARBs), the structural basis for AT<sub>1</sub>R ligand-binding and regulation has remained elusive, mostly due to the difficulties of growing high-quality crystals for structure determination using synchrotron radiation. By applying the recently developed method of serial femtosecond crystallography at an X-ray free-electron laser, we successfully determined the room-temperature crystal structure of the human AT<sub>1</sub>R in complex with its selective antagonist ZD7155 at 2.9-Å resolution. The AT<sub>1</sub>R-ZD7155 complex structure revealed key structural features of AT<sub>1</sub>R and critical interactions for ZD7155 binding. Docking simulations of the clinically used ARBs into the AT<sub>1</sub>R structure further elucidated both the common and distinct binding modes for these anti-hypertensive drugs. Our results thereby provide fundamental insights into AT<sub>1</sub>R structure-function relationship and structure-based drug design.

## INTRODUCTION

Cardiovascular disease remains one of the main causes of death throughout the world despite impressive advances in diagnosis

and therapeutics during the past few decades. Hypertension is the most common modifiable risk factor in cardiovascular disease, as myocardial infarction, stroke, heart failure, and renal disease can be greatly reduced by lowering blood pressure (Zaman et al., 2002). The best known regulator of blood pressure is the renin-angiotensin system (RAS). Over-stimulation of the RAS is implicated in hypertension, cardiac hypertrophy, heart failure, ischemic heart disease, and nephropathy (Balakumar and Jagadeesh, 2014). A cascade of proteolytic reactions in the RAS can generate various angiotensin peptides. Renin cleaves the precursor protein, angiotensinogen, releasing the inactive angiotensin I. Subsequently, angiotensin I is cleaved by angiotensin converting enzyme (ACE) to generate angiotensin II (AngII), angiotensin III, and angiotensin 1-7. These peptides exert diverse functions; angiotensins II and III act as vasoconstrictors, while angiotensin 1-7 acts as a vasodilator (Zaman et al., 2002). AngII is also responsible for cell migration, protein synthesis, endothelial dysfunction, inflammation, and fibrosis (Ramchandran et al., 2006).

In humans, AngII binds to two subtypes of angiotensin G protein-coupled receptors (GPCRs), angiotensin II type 1 receptor (AT<sub>1</sub>R) and angiotensin II type 2 receptor (AT<sub>2</sub>R) (Oliveira et al., 2007). Almost all physiological and pathophysiological effects of AngII are mediated by AT<sub>1</sub>R (de Gasparo et al., 2000), while the function of AT<sub>2</sub>R remains largely unknown (Akazawa et al., 2013). AT<sub>1</sub>R exhibits multiple active conformations, thereby activating different signaling pathways with differential functional outcomes (Shenoy and Lefkowitz, 2005). The G protein-dependent signaling by AT<sub>1</sub>R is vital for normal cardiovascular homeostasis yet detrimental in chronic dysfunction, which associates with cell death and tissue fibrosis and leads to cardiac

hypertrophy and heart failure (Ma et al., 2010). Accumulating evidence suggests that G protein independent  $\beta$ -arrestin mediated signaling by AT<sub>1</sub>R confers cardio-protective benefits (Whalen et al., 2011; Wisler et al., 2014).

Targeting the RAS cascade has proven to be effective in the treatment of hypertension, as well as specific cardiovascular and renal disorders. The most commonly used drugs include renin inhibitors, ACE inhibitors, and AT<sub>1</sub>R blockers (ARBs). ARBs, or sartans, are non-peptide antagonists and include the well-known anti-hypertensive drugs losartan, candesartan, valsartan, irbesartan, telmisartan, eprosartan, olmesartan, and azilsartan, most of which share a common biphenyl-tetrazole scaffold (Burnier and Brunner, 2000; Imaizumi et al., 2013; Miura et al., 2013a; Miura et al., 2013b). These ARBs are now extensively used for the treatment of cardiovascular diseases, including hypertension, cardiac hypertrophy, arrhythmia, and heart failure. There is additional interest in ARBs regarding their efficacy in the treatment of blood-vessel diseases such as Marfan-like syndrome, aortic dissection, and aortic aneurysms (Keane and Pyeritz, 2008; Ramanath et al., 2009).

Previous functional studies on AT<sub>1</sub>R have provided numerous clues into AT<sub>1</sub>R activation and inhibition mechanisms (Oliveira et al., 2007). Despite its high medical relevance and decades of research, the structure of AT<sub>1</sub>R and the binding mode of ARBs, however, are still unknown, which limits our understanding of the structural basis for AT<sub>1</sub>R function and modulation and precludes the rational optimization of AT<sub>1</sub>R lead compounds. One such experimental anti-hypertensive compound is ZD7155, a high-affinity antagonist and precursor to the anti-hypertensive drug candesartan. ZD7155 has a biphenyl-tetrazole scaffold similar to other ARBs and is more potent and longer-lasting than the first clinically used ARB losartan (Junggren et al., 1996). While structures of several different GPCRs have been reported, the determination of a new GPCR structure remains a significant challenge. X-ray crystallography using synchrotron radiation requires sufficiently large crystals in order to collect high-resolution data. Our extensive efforts to solve the AT<sub>1</sub>R structure were hampered by the limited size of micro-crystals grown in the membrane mimetic matrix known as lipidic cubic phase (LCP) (Caffrey and Cherezov, 2009). Nevertheless, by applying the recently developed method of serial femtosecond crystallography with LCP as a growth and carrier matrix for delivering microcrystals (LCP-SFX) into an X-ray free-electron laser (XFEL) beam (Liu et al., 2013; Weierstall et al., 2014; Liu et al., 2014a), we successfully determined the room-temperature crystal structure of the human AT<sub>1</sub>R in complex with ZD7155 (AT<sub>1</sub>R-ZD7155). Based on the AT<sub>1</sub>R-ZD7155 structure, we further performed mutagenesis and docking simulations to reveal binding modes for clinically used anti-hypertensive drugs targeting AT<sub>1</sub>R.

## RESULTS

### Structure Determination of AT<sub>1</sub>R-ZD7155 Complex Using LCP-SFX Method

To facilitate crystallization, a thermostabilized apocytochrome, b<sub>562</sub>RIL (BRIL) (Chun et al., 2012), was fused to the amino terminus (N terminus) of the human AT<sub>1</sub>R. Eleven residues were trun-

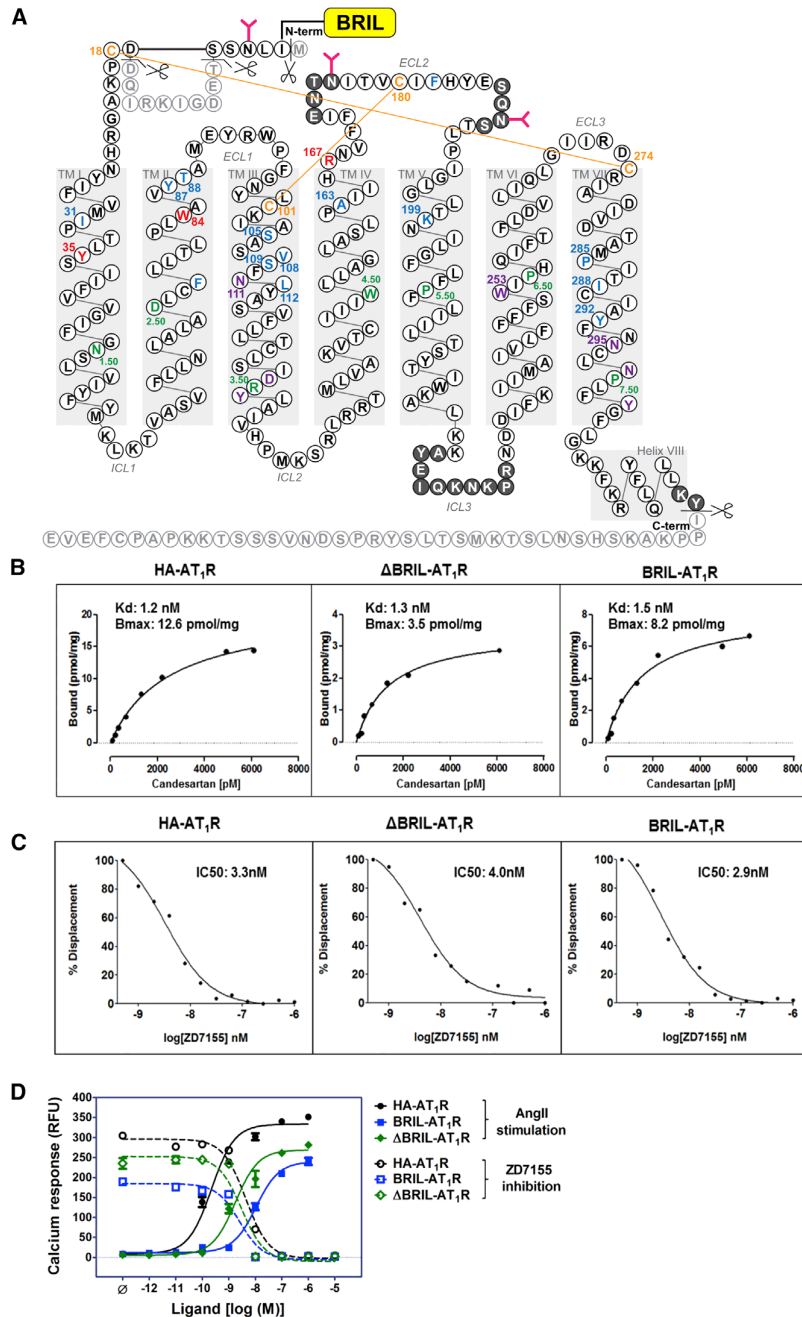
cated from the N-terminal region of AT<sub>1</sub>R (Met1, Thr7-Asp16), in order to shorten the flexible N terminus while keeping both the putative glycosylation site at Asn4 and the disulfide bond site at Cys18 intact. Forty residues were truncated from the carboxyl terminus (C terminus) after the cytoplasmic helix VIII (Figure 1A). The effect of protein engineering on AT<sub>1</sub>R function was evaluated using radio-ligand binding and calcium mobilization assays, in which neither the truncations nor BRIL insertion significantly altered the functional and pharmacological properties of the receptor for ligand binding and signaling (Figure 1B–1D). With this engineered AT<sub>1</sub>R, we obtained micro-crystals (maximum size 40 × 4 × 4  $\mu\text{m}^3$ ) in monoolein-based LCP, supplemented with cholesterol (Figure S1A). These microcrystals diffracted to only about 4-Å resolution at a synchrotron source under cryogenic conditions. To improve the resolution and avoid radiation damage and freezing, we took advantage of a recently developed LCP-SFX method and collected diffraction data at room temperature at the Linac Coherent Light Source (LCLS) using AT<sub>1</sub>R micro-crystals (average size 10 × 2 × 2  $\mu\text{m}^3$ ) grown in syringes (Figures S1B and S1C). A total of 2,764,739 patterns were collected by using ~65  $\mu\text{l}$  of crystal-loaded LCP, corresponding to ~0.35 mg of protein. Of these frames, 457,275 were identified as crystal hits, corresponding to a hit rate of 17%. Of these crystal hits, 73,130 frames (16%) were successfully indexed and integrated by CrystFEL (White et al., 2012) to 2.9-Å resolution (Table S1 and Figures S1D–S1F). The structure of the AT<sub>1</sub>R-ZD7155 complex was refined to  $R_{\text{work}}/R_{\text{free}}$  of 22.8%/27.4%. The final structure includes 289 out of 359 residues in the full-length human AT<sub>1</sub>R (Figure 1A), and it has well-defined densities for most AT<sub>1</sub>R residues and for the ligand ZD7155.

### Overall Architecture of AT<sub>1</sub>R

AT<sub>1</sub>R, being the angiotensin II octapeptide receptor, shares some sequence similarity with other peptide receptors of class A GPCRs, structures of which are known (sequence alignment is shown in Figure S2), with the closest homology to the chemokine receptors (e.g., 36% sequence identity with CXCR4) and opioid receptors (e.g., 33% sequence identity with  $\kappa$ -OR) (Wu et al., 2010; Wu et al., 2012). AT<sub>1</sub>R exhibits the canonical seven transmembrane  $\alpha$ -helical (7TM) architecture, with an extracellular N terminus, three intracellular loops (ICL1–3), three extracellular loops (ECL1–3), an amphipathic helix VIII and an intracellular C terminus (Figure 2A). The overall fold of the angiotensin receptor AT<sub>1</sub>R is most similar to the chemokine and opioid receptors (Figure 2B), with the lowest root mean square deviation for 80% of AT<sub>1</sub>R  $\alpha$ -carbon atoms ( $\text{RSMD}_{C\alpha}$ ) of about 1.8 Å to the nociceptin/orphanin FQ peptide receptor (NOP) (Thompson et al., 2012). Despite the overall similarity, a number of structural differences in the transmembrane bundle were observed between AT<sub>1</sub>R and other peptide GPCRs (Figures 2C and 2D). For example, the tilts and extensions of the extracellular ends of helices I, V, VI, and VII are substantially different among these peptide receptors, while at the intracellular side, helices IV and V adopt the most diverse conformations. The conformations of helices II and III, however, are nearly identical for all these peptide receptors.

The extracellular part of AT<sub>1</sub>R consists of the N-terminal segment, ECL1 (Glu91-Phe96) linking helices II and III, ECL2





**Figure 1. AT<sub>1</sub>R Construct Design and Functional Characterization**

(A) Snake plot of the BRIL-AT<sub>1</sub>R construct used for crystallization. Residues that occupy the most conserved positions on each helix in class A GPCRs (X.50; B&W scheme) are colored in green. The four cysteine residues that form two disulfide bonds in the extracellular region are colored in orange. Three critical residues for ZD7155 binding are colored in red. All other residues that interact with ZD7155 are colored in blue. Critical residues/motifs for AT<sub>1</sub>R activation are colored in purple. Truncated residues are shown as light gray, and residues that do not have sufficient density in the structure and therefore were not modeled are shown in dark gray circles.

(B) Saturation binding of the non-peptide antagonist <sup>3</sup>H-candesartan to the wild-type HA-AT<sub>1</sub>R, ΔBRIL-AT<sub>1</sub>R, and BRIL-AT<sub>1</sub>R.

(C) Competition binding of ZD7155 to the wild-type HA-AT<sub>1</sub>R, ΔBRIL-AT<sub>1</sub>R, and BRIL-AT<sub>1</sub>R, performed by displacement of <sup>3</sup>H-candesartan.

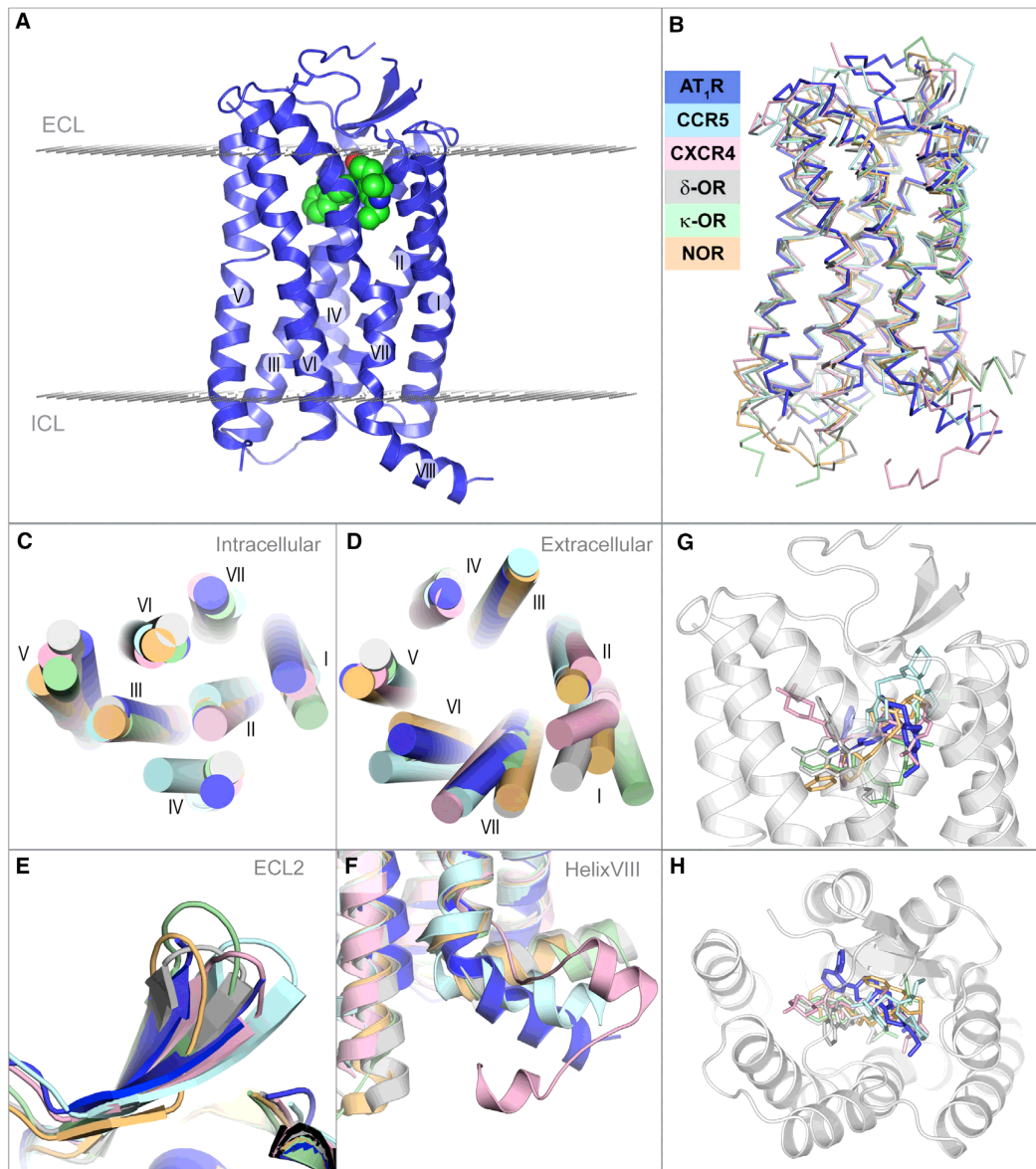
(D) Intracellular calcium responses for the wild-type HA-AT<sub>1</sub>R, BRIL-AT<sub>1</sub>R, and ΔBRIL-AT<sub>1</sub>R. The agonist AngII and the antagonist ZD7155 dose-response curves for HA-AT<sub>1</sub>R (circles), BRIL-AT<sub>1</sub>R (squares), and ΔBRIL-AT<sub>1</sub>R (diamonds) are shown in closed and open symbols, respectively. Error bars represent SEM.

mon motif among peptide GPCRs (Figure 2E). Intriguingly, ECL2 of AT<sub>1</sub>R was found to serve as an epitope for the harmful agonistic autoantibodies in pre-eclampsia and malignant hypertension (Unal et al., 2012; Xia and Kellems, 2013).

The intracellular portion of AT<sub>1</sub>R contains ICL1 (Lys58 to Val62) linking helices I and II, ICL2 (Val131 to Arg137) linking helices III and IV, ICL3 (Leu222 to Asn235) linking helices V and VI, and the C-terminal helix VIII. As in many other class A GPCRs, the conserved D(E)RY motif in helix III and the NPxxY motif in helix VII of AT<sub>1</sub>R, both at the intracellular ends of transmembrane domain, were proposed to participate in receptor activation (Oliveira et al., 2007). However, the “ionic lock” salt bridge interaction between Arg<sup>3.50</sup> (superscript indicates residue number as per the Ballesteros and Weinstein, 1995 [B&W] nomenclature) of the D(E)RY motif and Asp/Glu<sup>6.30</sup> at the cytoplasmic end of helix VI is not possible in AT<sub>1</sub>R, because the human AT<sub>1</sub>R lacks an acidic residue at the position 6.30.

The C-terminal helix VIII of AT<sub>1</sub>R was shown to bind the calcium-regulated effector protein, calmodulin (Thomas et al., 1999). Integrity of this region is also important for receptor internalization and coupling to G protein activation and signaling (Thomas et al., 1995; Sano et al., 1997). In most previously

(His166 to Ile191) linking helices IV and V, and ECL3 (Ile270 to Cys274) linking helices VI and VII (Figure 1A). Two disulfide bonds help to shape the extracellular side of AT<sub>1</sub>R, with Cys18-Cys274 connecting the N terminus and ECL3, and Cys101-Cys180 connecting helix III and ECL2, similar to the chemokine receptors CXCR4 and CCR5 (Wu et al., 2010; Tan et al., 2013). Besides engaging in the conserved disulfide bonding, ECL2 of AT<sub>1</sub>R exhibits a β-hairpin secondary structure, a com-



**Figure 2. Overview of AT<sub>1</sub>R-ZD7155 Architecture and Structural Comparison with Other Peptide GPCRs**

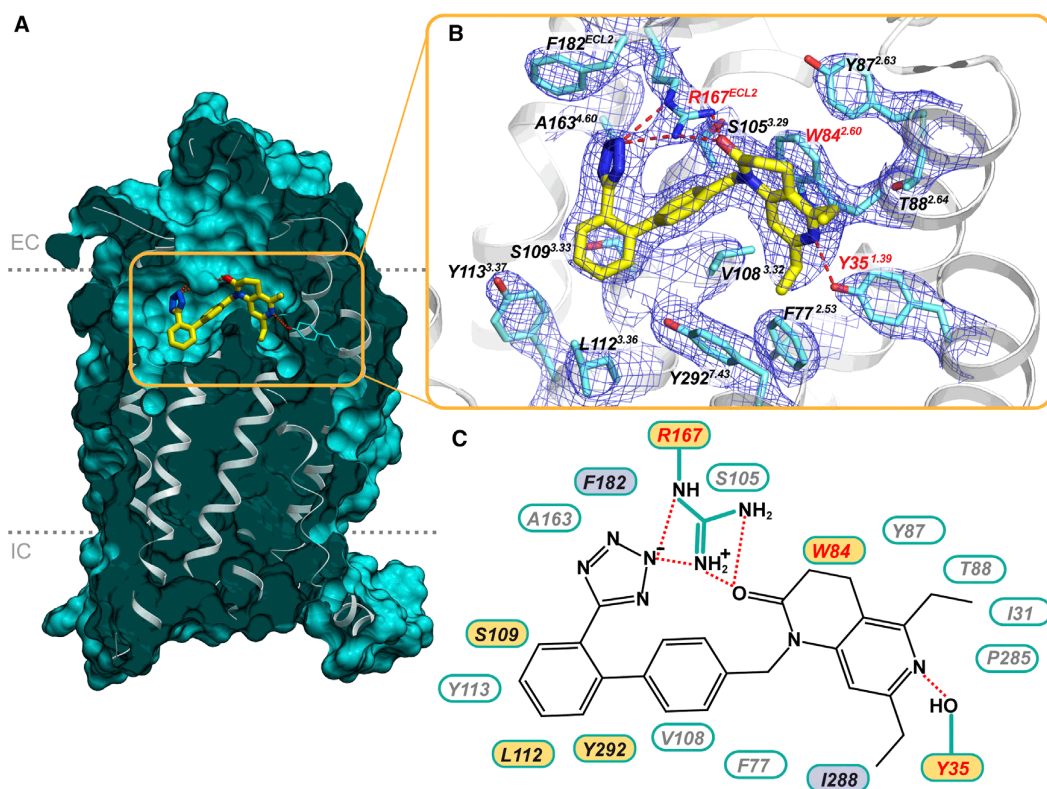
(A) Overall AT<sub>1</sub>R structure is shown as blue cartoon. ZD7155 is shown as spheres with carbon atoms colored green. Membrane boundaries, as defined by the PPM web server (Lomize et al., 2012), are shown as planes made of gray spheres.

(B–H) Superposition of AT<sub>1</sub>R with chemokine and opioid receptors, chemokine CCR5 receptor, light cyan (PDB ID 4MBS); chemokine CXCR4 receptor, light pink (PDB ID 3ODU);  $\delta$ -opioid receptor, gray (PDB ID 4N6H);  $\kappa$ -opioid receptor, light green (PDB ID 4DJH); NOP receptor, light orange (PDB ID 4EA3), comparing the whole structure (B), intracellular view (C), extracellular view (D), ECL2 (E), helix VIII (F), and the ligand binding pocket side (G) and top (H) views.

See also Figures S1 and S2 and Table S1.

solved GPCR structures, helix VIII runs parallel to the membrane bilayer, however, in AT<sub>1</sub>R it angles away from the membrane, resembling the orientation of this helix in CCR5 (Figure 2F). Experimentally, the secondary structure of AT<sub>1</sub>R helix VIII was observed to be sensitive to hydrophobic environment, thereby associating with the cytoplasmic side of the cell membrane via

a high-affinity, anionic phospholipid-specific tethering that serves to increase the amphipathic helicity of this region (Mozsolits et al., 2002). As a separate peptide, helix VIII of AT<sub>1</sub>R showed a higher affinity for lipid membranes that contained negatively charged phospholipids rather than zwitterionic phospholipids (Kamimori et al., 2005). A high concentration of positively



**Figure 3. Interactions of ZD7155 with AT<sub>1</sub>R**

(A) Cross-section view of AT<sub>1</sub>R highlighting the shape of the ligand binding pocket.

(B) Zoomed-in view of the ligand binding pocket showing all residues within 4 Å from the ligand ZD7155, along with the 2mF<sub>o</sub>-DF<sub>c</sub> electron density (blue mesh) contoured at 1 σ level. In (A) and (B) ZD7155 is shown as sticks with yellow carbons.

(C) Schematic representation of interactions between AT<sub>1</sub>R and ZD7155. Hydrogen bonds/salt bridges are shown as red dashed lines. The residues shown by mutagenesis to be critical for ligand binding are labeled red, those that are important for either peptide or non-peptide ligands binding are labeled in yellow, and the residues that discriminate between peptide and non-peptide ligands are labeled in purple.

See also [Figure S2](#) and [Table S2](#).

charged residues (306-KKFKR-312) in helix VIII of AT<sub>1</sub>R possibly defines its orientation and explains its sensitivity to the negatively charged lipids. Moreover, in AT<sub>1</sub>R there is no putative palmitoylation site that is present in many GPCRs in this region, anchoring helix VIII to the lipid membrane.

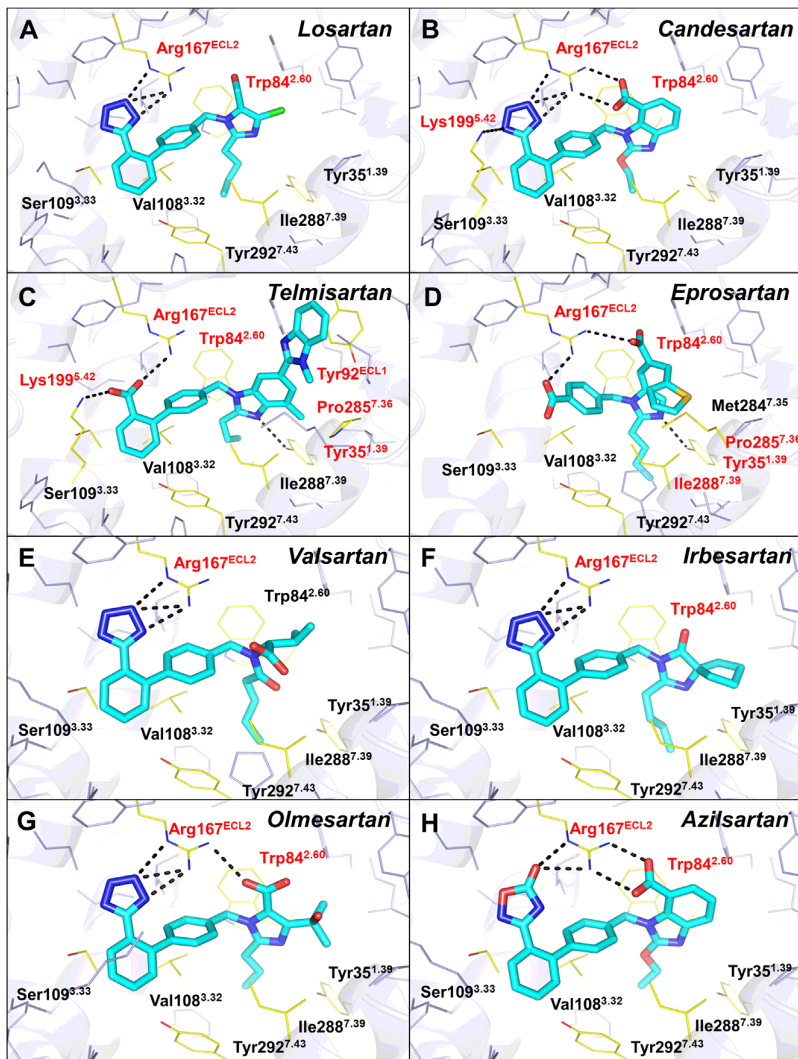
#### ZD7155 Interactions in AT<sub>1</sub>R Ligand-Binding Pocket

Small molecule antagonist ZD7155 was modeled into the prominent and well-defined electron density inside the ligand-binding pocket of AT<sub>1</sub>R (Figure 3A and 3B), interacting with residues mainly from helices I, II, III, and VII, as well as ECL2. Side chains of Arg167<sup>ECL2</sup> and Tyr35<sup>1.39</sup> were found to form ionic and hydrogen bond interactions with ZD7155. The positively charged guanidine group of Arg167<sup>ECL2</sup> forms an extensive interaction network with the acidic tetrazole and the naphthyridin-2-one moieties of ZD7155. Leveraging this information in mutagenesis studies, we found that mutation of Arg167<sup>ECL2</sup> to alanine abolished both the peptide and non-peptide ligands binding to AT<sub>1</sub>R (Table S2). However, the Arg167<sup>ECL2</sup>Lys mutant showed only 2- to 3-fold reduced binding affinities for ZD7155, which

can be explained by the ability of lysine in this position to engage in salt bridge and hydrogen bond interactions similar to Arg167<sup>ECL2</sup>, although likely with less optimal interaction geometry. The tetrazole moiety, or other acidic isostere in the ortho position of the biphenyl group comprises the most common scaffold among ARBs, and Arg167<sup>ECL2</sup> is a unique residue of AT<sub>1</sub>R compared to other structurally similar peptide GPCRs (Figure S2). This observation suggests that Arg167<sup>ECL2</sup> may play an essential role in determining AT<sub>1</sub>R ligand-binding affinity and selectivity. An additional hydrogen bond forms between Tyr35<sup>1.39</sup> and the naphthyridin-2-one moiety of ZD7155. Our data showed that the Tyr35<sup>1.39</sup>Ala mutant abolishes the binding capabilities of both peptide and non-peptide ligands with AT<sub>1</sub>R (Table S2). Tyr<sup>1.39</sup> is a well conserved residue in the angiotensin, chemokine, and opioid receptors (Figure S2). In the CCR5 structure, for example, Tyr37<sup>1.39</sup> interacts with its ligand maraviroc (Tan et al., 2013).

The ZD7155 binding site in AT<sub>1</sub>R partially overlaps with known ligand binding sites in the chemokine and opioid receptors (Figures 2G and 2H). Intriguingly, some of the residues that comprise





**Figure 4. Docking of Different Anti-Hypertensive Drugs in the AT<sub>1</sub>R Crystal Structure (A–H)** The ARBs are shown as sticks with cyan carbons. The AT<sub>1</sub>R residues interacting with ligands are labeled and shown as yellow lines, with the key residues highlighted in red. The hydrogen bonds are shown as black dashed lines. See also Table S3.

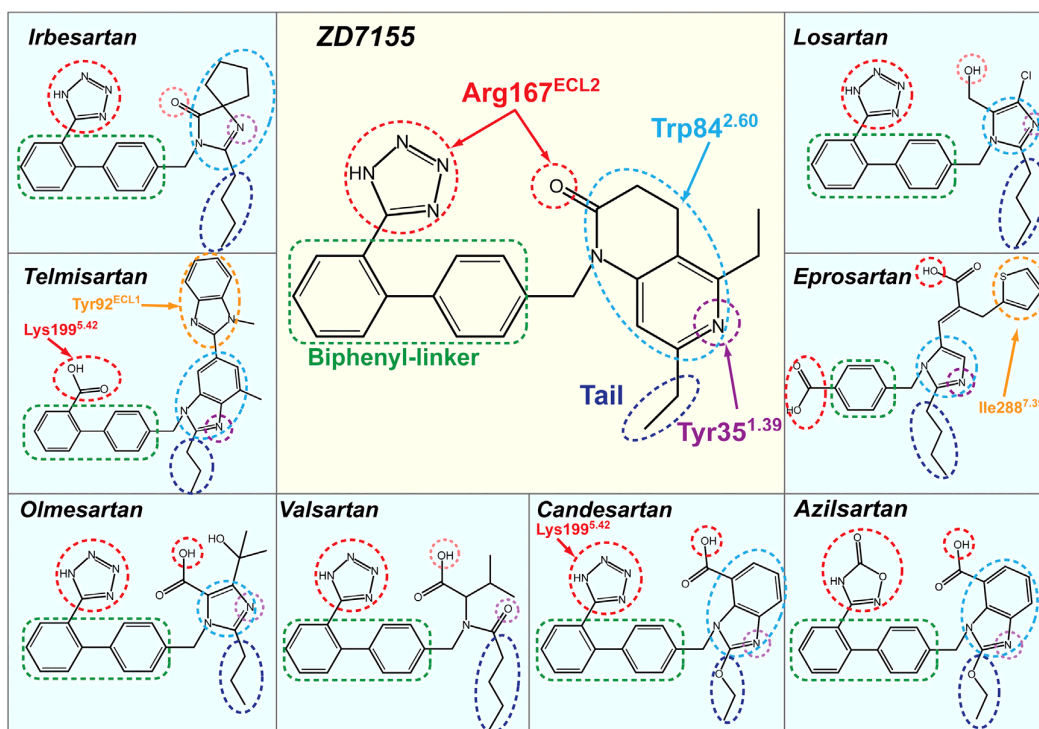
the other contacts for ZD7155 binding to AT<sub>1</sub>R, however, are mediated by non-conserved residues, including Tyr87<sup>2.63</sup>, Thr88<sup>2.64</sup>, Ser105<sup>3.29</sup>, Ser109<sup>3.33</sup>, Ala163<sup>4.60</sup>, Phe182<sup>ECL2</sup>, Pro285<sup>7.36</sup>, and Ile288<sup>7.39</sup> (Figures 3B and 3C and Figure S2). These residues along with Arg167<sup>ECL2</sup> therefore define the unique shape of the AT<sub>1</sub>R ligand-binding pocket and explain the lack of cross-reactivity between ligands binding to AT<sub>1</sub>R and other peptide receptors.

#### Binding Modes of Different ARBs toward AT<sub>1</sub>R

To analyze the common and diverse features of the binding modes for different ARBs in AT<sub>1</sub>R, we performed energy-based docking simulations of the clinically used anti-hypertensive ARBs using the AT<sub>1</sub>R structure. The docking results show robust positioning of these compounds in the AT<sub>1</sub>R ligand-binding pocket (Figure 4 and Table S3). Although the nature of the interactions with AT<sub>1</sub>R is different for each ARB given their distinct chemical structures, most of these compounds are bound in similar orientations and engage in interactions with the three residues critical for ZD7155 binding, Arg167<sup>ECL2</sup>, Trp84<sup>2.60</sup>, and Tyr351<sup>1.39</sup> (Figure 5). Residues Phe77<sup>2.53</sup>, Tyr87<sup>2.63</sup>, Ser105<sup>3.29</sup>, Val108<sup>3.32</sup>, Ser109<sup>3.33</sup>, Leu112<sup>3.36</sup>, Ala163<sup>4.60</sup>, Phe182<sup>ECL2</sup>, Ile288<sup>7.39</sup>, and Tyr292<sup>7.43</sup> also contribute to the receptor-ligand interactions and shape the ligand-binding pocket. For example, one of the common features among these ARBs is a short alkyl tail with two-four carbons extending into a narrow hydrophobic pocket formed by Tyr351<sup>1.39</sup>, Phe77<sup>2.53</sup>, Val108<sup>3.32</sup>, Ile288<sup>7.39</sup>, and Tyr292<sup>7.43</sup> (Figure 5).

Losartan is the first clinically used ARB for the treatment of hypertension. It is, however, a surmountable antagonist with lower binding affinity to AT<sub>1</sub>R compared to the later developed ARBs (Miura et al., 2011). Docking results suggest that Arg167<sup>ECL2</sup> forms a salt bridge only with the tetrazole moiety of losartan but lacks polar interactions with other groups (Figure 4 and Table S3). Although the derived imidazole moiety of losartan can also contribute to polar interactions via methanol hydrogen bond to Cys180<sup>ECL2</sup> main chain or via nitrogen interaction with

the ligand-binding pockets, including Ile<sup>1.35</sup>, Phe<sup>2.53</sup>, Trp<sup>2.60</sup>, and Tyr<sup>7.43</sup>, can be found among these structurally similar peptide GPCRs (Figure S2). Residues Phe77<sup>2.53</sup> and Trp84<sup>2.60</sup> from helix II of AT<sub>1</sub>R are conserved in the chemokine receptors CXCR4 and CCR5 (Wu et al., 2010; Tan et al., 2013). Particularly, Trp84<sup>2.60</sup> of AT<sub>1</sub>R forms  $\pi$ - $\pi$  interaction with the naphthyridin-2-one moiety of ZD7155, and mutation of Trp84<sup>2.60</sup> to alanine abolished both the peptide and non-peptide ligands binding to AT<sub>1</sub>R (Figure 3C and Table S2). Residues Ile31<sup>1.35</sup> and Tyr292<sup>7.43</sup> from helices I and VII of AT<sub>1</sub>R are conserved in the opioid receptors  $\kappa$ -OR,  $\delta$ -OR, and NOP. Additionally, residues Val108<sup>3.32</sup> and Leu112<sup>3.36</sup>, which hydrophobically interact with ZD7155 in the AT<sub>1</sub>R ligand-binding pocket, are replaced by Tyr108<sup>3.32</sup> and Phe112<sup>3.36</sup> in CCR5 and form hydrophobic interactions with its ligand maraviroc. In contrast, the position 3.32 in the aminergic and opioid receptors is occupied by a conserved aspartic acid that engages in a salt bridge interaction with ligands. Most of

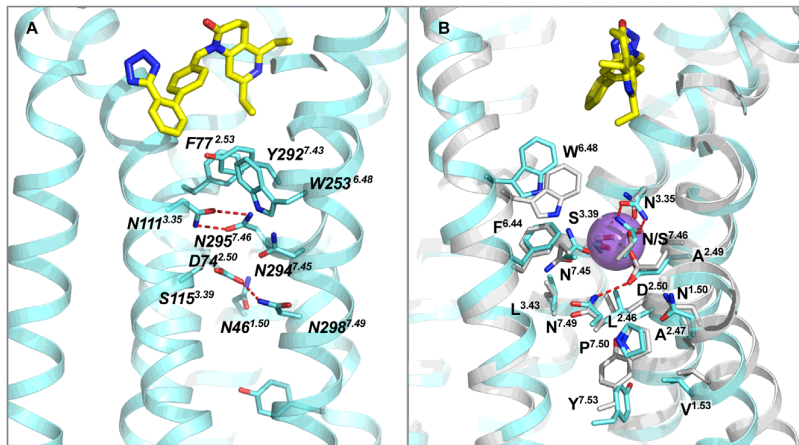


**Figure 5. Common and Distinct Binding Modes of Different ARBs with AT<sub>1</sub>R**

The ARB chemical groups that are engaged in hydrogen bonding/salt bridging with Arg167<sup>ECL2</sup> and Tyr35<sup>1.39</sup> are marked by red and purple dashed circles, respectively. Pale red and pale purple dotted circles are used for groups with sub-optimal contacts as suggested by docking. The heterocyclic groups forming  $\pi$ - $\pi$  contacts with Trp84<sup>2.60</sup> are surrounded by light-blue dashed circles. The biphenyl-linker groups for hydrophobic interactions are outlined by green dashed boxes, and the two-four carbons tails, extending into the hydrophobic pocket formed by Tyr35<sup>1.39</sup>, Phe77<sup>2.53</sup>, Val108<sup>3.32</sup>, Ile288<sup>7.39</sup>, and Tyr292<sup>7.43</sup>, are outlined by dark-blue dashed circles. Specific interactions of candesartan and telmisartan with Lys199<sup>5.42</sup> are shown by red arrows. Specific interactions between Tyr92<sup>ECL1</sup> and telmisartan, and between Ile288<sup>7.39</sup> and eprosartan are highlighted by orange dashed circles. See also Figure S3.

Tyr35<sup>1.39</sup>, distances and angles for hydrogen bonding are sub-optimal; this may explain the lower binding affinity and surmountable property of losartan at AT<sub>1</sub>R. An active metabolite of losartan, EXP3174, is predicted to bind in a similar pose as losartan, but instead of interaction with Cys180<sup>ECL2</sup>, its carboxylic group could engage in a second salt bridge interaction with Arg167<sup>ECL2</sup>, similarly to ZD7155 (Table S3). In contrast, candesartan is an insurmountable inverse agonist with a slow dissociation rate from AT<sub>1</sub>R (Takezako et al., 2004). The docking results indicate that besides interacting with the tetrazole moiety of candesartan, Arg167<sup>ECL2</sup> forms two salt bridges to the carboxylic group of the benzimidazole moiety (Figure 4 and Table S3). Moreover, Lys199<sup>5.42</sup> is predicted to form an additional salt bridge with the tetrazole moiety, which can further stabilize candesartan binding. Telmisartan lacks the conserved tetrazole moiety among ARBs. Instead, the carboxylic group of telmisartan is predicted to form salt bridges with both Arg167<sup>ECL2</sup> and Lys199<sup>5.42</sup> (Figure 4 and Table S3). Moreover, unlike other ARBs studied here, two consecutive benzimidazole moieties of telmisartan extend to Tyr92<sup>ECL1</sup>, making additional hydrophobic and  $\pi$ - $\pi$  contacts, which are likely to contribute to its high potency (Balakumar et al., 2012). This prediction was confirmed

by our mutagenesis data, which showed a dramatic decrease in affinity of telmisartan to the Tyr92<sup>ECL1</sup>Ala mutant (Figure S3A). Eprosartan is the most unique among the ARBs studied here, lacking both the tetrazole group and one of the two benzene rings of the biphenyl scaffold. As our docking results suggest, eprosartan uses its two carboxylic groups to form salt bridges with Arg167<sup>ECL2</sup> (Figure 4 and Table S3). Additionally, the specific thiophen moiety of eprosartan forms hydrophobic interactions with Pro285<sup>7.36</sup> and Ile288<sup>7.39</sup> and reaches toward Met284<sup>7.35</sup>. Mutation of Met284<sup>7.35</sup> to alanine produced minimal effect, slightly increasing the affinity for eprosartan binding, in agreement with predicted interactions of this ligand with only mainchain and C $\beta$  atoms of Met284<sup>7.35</sup> (Figure S3B). On the other hand, mutations Pro285<sup>7.36</sup>Ala and Ile288<sup>7.39</sup>Ala induced a strong decrease in the binding affinity of eprosartan (Figures S3C and S3D), highlighting essential role of these residues in eprosartan binding. Finally, both our crystal structure and docking results suggest that Lys199<sup>5.42</sup> retains some conformational heterogeneity in AT<sub>1</sub>R. Docking with the flexible side chain of Lys199<sup>5.42</sup> indicates that the amino group of this residue can reach the acidic moieties of ARBs by forming salt bridges (as interacting with candesartan and telmisartan) or water-mediated



**Figure 6. Critical Residues for AT<sub>1</sub>R Activation**

(A) A cluster of aromatic residues (F77<sup>2.53</sup>, W253<sup>6.48</sup> and Y292<sup>7.43</sup>) is located just below ZD7155, bridging the ligand binding pocket with a cluster of polar residues that includes several highly conserved in class A GPCR residues (N46<sup>1.50</sup>, D74<sup>2.50</sup>), along with N111<sup>3.35</sup> and N295<sup>7.46</sup> forming hydrogen bonds that hold helices III and VII together.

(B) Superposition of the AT<sub>1</sub>R structure with the high-resolution structure of  $\delta$ -OR (PDB ID 4N6H) reveals a high structural conservation of the putative sodium-binding site. Sodium ion is shown as purple ball.

interactions, which may explain the reduced ligand-binding capabilities of Lys199<sup>5.42</sup> mutants (Table S2).

### Mechanism of AT<sub>1</sub>R Modulation

Based on previous observations that mutations of either Asn111<sup>3.35</sup> or Asn295<sup>7.46</sup> induce constitutive activation of the receptor, it was proposed that the inactive conformation of AT<sub>1</sub>R is stabilized by interactions between Asn111<sup>3.35</sup> and Asn295<sup>7.46</sup>. Further, it was suggested that binding of AngII to the wild-type (WT) receptor disrupts the hydrogen bonds between Asn111<sup>3.35</sup> and Asn295<sup>7.46</sup>, thus allowing Asn295<sup>7.46</sup> to interact with the conserved Asp74<sup>2.50</sup> (Balakumar and Jagadeesh, 2014; Unal and Kamik, 2014). Indeed, two intramolecular hydrogen bonds are observed between Asn111<sup>3.35</sup> and Asn295<sup>7.46</sup> in the AT<sub>1</sub>R-ZD7155 structure (Figure 6A). Of particular interest, Asp74<sup>2.50</sup>, Asn111<sup>3.35</sup>, and Asn295<sup>7.46</sup>, together with two other residues, Trp253<sup>6.48</sup> from the WxP motif and Asn298<sup>7.49</sup> from the NPxxY motif, belong to the putative sodium pocket of AT<sub>1</sub>R (Katritch et al., 2014) as revealed by superposition with the sodium site in the high-resolution structure of  $\delta$ -OR (Figure 6B) (Fenalti et al., 2014). All residues lining this pocket in AT<sub>1</sub>R are conserved exactly as in  $\delta$ -OR, except for Asn295<sup>7.46</sup> (Ser in  $\delta$ -OR), which is observed at this position in a GPCR structure for the first time; therefore, its presence and the strong hydrogen bond interactions with Asn111<sup>3.35</sup> may impact the sodium binding and functional properties of AT<sub>1</sub>R. Moreover, the neighboring residue Phe77<sup>2.53</sup> from the ligand-binding pocket of AT<sub>1</sub>R was also found to be critical for the inter-helical interactions required for AT<sub>1</sub>R activation (Miura et al., 2003). Combination of Phe77<sup>2.53</sup>Ala and Asn111<sup>3.35</sup>Gly mutations resulted in an almost fully active receptor (Miura et al., 2008). Thus, multiple structural and functional data suggest that the hydrogen bond network around Asn111<sup>3.35</sup> and Asn295<sup>7.46</sup> as revealed in the current structure may play an essential role in AT<sub>1</sub>R activation, probably by relaying the conformational changes in the ligand-binding pocket to the cytoplasmic domain coupling to the downstream signaling, although further structural, functional, and biophysical studies are required to fully understand the mechanism of AT<sub>1</sub>R modulation.

### DISCUSSION

The angiotensin receptor AT<sub>1</sub>R is a therapeutic target of outstanding interest due to its important roles in cardiovascular pathophysiology. Several AT<sub>1</sub>R blockers have been developed and clinically used as anti-hypertensive drugs. Although extensive efforts were taken to delineate the pharmacophores of AT<sub>1</sub>R ligands, structure-based drug design was still hindered by the lack of structural information. By using an XFEL, we successfully determined the crystal structure of the human AT<sub>1</sub>R in complex with its antagonist ZD7155. Compared to the traditional X-ray crystallography with cryo-cooled crystals, the LCP-SFX method yields the room-temperature structure of the AT<sub>1</sub>R-ZD7155 complex, which is likely to represent more accurately the receptor conformations and dynamics in the native cellular environment. The AT<sub>1</sub>R-ZD7155 complex structure reveals a variety of key features of AT<sub>1</sub>R shared with other GPCR family members, as well as many novel and unique structural characteristics of the angiotensin receptor. Unexpectedly, three AT<sub>1</sub>R residues, which have not been previously implicated in binding small molecule ligands, were found to form critical interactions with ZD7155; Arg167<sup>ECL2</sup> and Tyr35<sup>1.39</sup> are engaged in ionic and hydrogen bonds, while Trp84<sup>2.60</sup> forms extensive  $\pi$ - $\pi$  interactions with the ligand. The antagonist-bound AT<sub>1</sub>R structure was used further for docking of several anti-hypertensive ARBs into the AT<sub>1</sub>R ligand-binding pocket, elucidating the structural basis for AT<sub>1</sub>R modulation by drugs. Our extensive mutagenesis experiments revealed that residues Tyr35<sup>1.39</sup>, Trp84<sup>2.60</sup>, Arg167<sup>ECL2</sup>, and Lys199<sup>5.42</sup> are critical for both peptide ([Sar<sup>1</sup>, Ile<sup>8</sup>]-AngII) and non-peptide (candesartan) binding. Residues Phe182<sup>ECL2</sup> and Ile288<sup>7.39</sup> discriminate between the peptide and non-peptide ligand (these mutants do not bind [Sar<sup>1</sup>, Ile<sup>8</sup>]-AngII but bind candesartan). Mutations of Ser109<sup>3.33</sup> and Tyr292<sup>7.43</sup> slightly affected non-peptide (candesartan) binding but not peptide binding (Table S2).

Among the naturally occurring amino acid variations in AT<sub>1</sub>R, reported in Uniprot (<http://www.uniprot.org/uniprot/P30556>), Ala163<sup>4.60</sup>Thr, Thr282<sup>7.33</sup>Met, and Cys289<sup>7.40</sup>Trp are located near the binding pocket for ARBs. These variants may directly



alter binding of ARBs and therefore modify the anti-hypertensive response to treatment with different ARBs in individuals carrying these variations. In contrast, Leu48<sup>1,52</sup>Val, Leu222<sup>CL3</sup>Val, and Ala244<sup>6,39</sup>Ser, which are located closer to intracellular ends of helices, may indirectly influence binding of ARBs or signaling by AT<sub>1</sub>R. Finally, Thr336Pro and Pro341His are located in the C-terminal tail that was not included in the crystallized construct. These residues, however, are known to affect GPCR kinase-dependent phosphorylation, an event that is necessary for  $\beta$ -arrestin recruitment to AT<sub>1</sub>R.

Of particular interest, the atomic details of ECL2 and the extracellular ligand-binding region, revealed in the current structure, are expected to guide design of two different types of therapeutic agents targeting AT<sub>1</sub>R, the anti-hypertensive ARBs extensively interacting with Arg167<sup>ECL2</sup> on the ligand-binding pocket side of ECL2, and the peptide-mimicking antigens against autoantibodies, which bind to the extracellular side of ECL2 in patients with autoimmune disorders, such as preeclampsia and malignant hypertension (Zhou et al., 2008; Fu et al., 2000). Therefore, our results provide long anticipated insights into the AT<sub>1</sub>R structure-function relationship and pharmacological properties and demonstrate the potential for using the LCP-SFX method at XFEL sources to accelerate structural studies of challenging targets.

## EXPERIMENTAL PROCEDURES

### Protein Engineering for Structural Studies

The sequence of the human AT<sub>1</sub>R gene was optimized for insect cell expression and synthesized by GenScript. A thermostabilized apocytochrome *b<sub>562</sub>*RIL (BRIL) from *E. coli* (M7W, H102I, R106L) was fused to the N terminus of the human AT<sub>1</sub>R, using overlapping PCR. The construct has truncations of the AT<sub>1</sub>R residues 1, 7–16, and 320–359. The resulting BRIL-AT<sub>1</sub>R chimera sequence was subcloned into a modified pFastBac1 vector (Invitrogen), which contained a haemagglutinin (HA) signal sequence, a FLAG tag and 10  $\times$  His tag, followed by a tobacco etch virus (TEV) protease cleavage site, before the N terminus of the chimera sequence.

### Protein Expression and Purification

BRIL-AT<sub>1</sub>R construct was expressed in *Spodoptera frugiperda* (Sf9) insect cells using the Bac-to-Bac baculovirus expression system (Invitrogen). Cells with a density of 2–3  $\times$  10<sup>6</sup> cells per ml were infected with baculovirus at 27°C, and harvested at 48 hr after infection.

BRIL-AT<sub>1</sub>R in complex with ZD7155 (Tooris Bioscience) was solubilized from isolated membranes using 1% (w/v) n-dodecyl-beta-D-maltopyranoside (DDM, Anatrace) and 0.2% (w/v) cholesterol hemisuccinate (CHS, Sigma-Aldrich). After purification by metal affinity chromatography BRIL-AT<sub>1</sub>R/ZD7155 complex was desalted to remove imidazole using PD MiniTrap G-25 column (GE Healthcare) and then treated overnight with His-tagged TEV protease to cleave the N-terminal FLAG/His tags from the protein. The cleaved FLAG/His tags and TEV protease were removed by TALON IMAC resin. The protein was not treated with PNGase F and therefore remained fully glycosylated. Finally, the purified protein was concentrated to 30 mg/ml with a 100 kDa cutoff concentrator (Vivaspin) and used in crystallization trials. The protein yield and monodispersity were tested by analytical size exclusion chromatography (aSEC).

### Lipidic Cubic Phase Crystallization

BRIL-AT<sub>1</sub>R in complex with ZD7155 was crystallized in LCP composed of monoolein supplemented with 10% cholesterol (Caffrey and Cherezov, 2009). LCP crystallization trials were performed using an NT8-LCP crystallization robot (Formulatrix). 96-well glass sandwich plates (Marienfeld) were incubated and imaged at 20°C using an automatic incubator/imager (RockImager

1000, Formulatrix). The crystals grew in the condition of 100 mM sodium citrate (pH 5.0–6.0), 300–600 mM NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub>, 20%–30% (v/v) PEG400 and 2%–8% (v/v) DMSO. The crystals were harvested using micromounts (MiTeGen) and flash-frozen in liquid nitrogen for data collection at a synchrotron source. These crystals diffracted only to about 4-Å resolution, even after extensive optimization of crystallization conditions.

Microcrystals for SFX data collection were prepared in gas-tight syringes (Hamilton) as described (Liu et al., 2014b), using 100 mM sodium citrate (pH 5.0), 450 mM NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub>, 28% (v/v) PEG400 and 4% (v/v) DMSO as a precipitant. Before loading microcrystals in the LCP injector the excess precipitant was removed, and 7.9 MAG was added and mixed with LCP, to absorb the residual precipitant solution and prevent formation of a crystalline phase due to a rapid evaporative cooling when injecting LCP into vacuum (Weierstall et al., 2014).

### X-Ray Free Electron Laser Data Collection

Data collection was performed at the Coherent X-ray Imaging (CXI) end station of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, using XFEL pulses of 36 fs duration focused to a size of 1.5  $\times$  1.5  $\mu$ m<sup>2</sup> by Kirkpatrick-Baez mirrors. A photon energy of 7.9 keV, an average pulse energy of 2.7 mJ and a transmission level of 16% resulted in a maximum dose of 75 MGy at the sample.

Microcrystals dispersed in LCP were delivered into the interaction region using an LCP injector (Weierstall et al., 2014) with a 50  $\mu$ m diameter nozzle at a flow rate of 170 nl per minute. Diffraction patterns were collected on a Cornell-SLAC Pixel array detector (CSPAD - version 1.5) (Hart et al., 2012) at a rate of 120 Hz.

With a total sample volume of 65  $\mu$ l, a total of 2,764,739 diffraction frames were collected within 6.4 hr. Initial frames were corrected and filtered using the software package Cheetah (Barty et al., 2014). A crystal “hit” was defined as an image containing a minimum of 15 diffraction peaks with a signal to noise ratio above 4. A total of 457,275 positive “hits” were further processed using the CrystFEL software suite (version 0.5.3) (White et al., 2012). The detector geometry was refined using an automated algorithm designed to match found and predicted peaks to sub-pixel accuracy. By further refinement of parameters (peak detection, prediction, and integration), a total of 73,130 images were indexed, integrated, and merged into a final dataset. To reduce noise and outliers and thus improve data quality we have applied two data rejection criteria: (1) per pattern resolution cutoff, and (2) rejection of patterns based on a Pearson correlation coefficient threshold, as described in the [Extended Experimental Procedures](#). A resolution cutoff was estimated to be 2.9 Å using a combination of CC\* (Karplus and Diederichs, 2012) and other parameters (Figures S1D–S1F). The final dataset had overall  $R_{split} = 9.8\%$ , and CC\* = 0.872 in the highest resolution shell.

### Structure Determination

The structure was solved by molecular replacement with Phaser (McCoy et al., 2007) using an automated script described in the [Extended Experimental Procedures](#).

Refinement and model completion were performed by repetitive cycling between Refmac5 (Murshudov et al., 1997) and autoBUSTER (Bricogne et al., 2009), followed by manual examination and rebuilding of the refined coordinates in Coot (Emsley et al., 2010). Data collection and refinement statistics are shown in [Table S1](#).

### Docking of ARBs into AT<sub>1</sub>R Ligand-Binding Pocket

Representative ARBs were docked into the AT<sub>1</sub>R crystal structure using an energy-based docking protocol implemented in ICM molecular modeling software suite (Molsoft). Molecular models of compounds were generated from two-dimensional representations and their 3D geometry was optimized using MMFF-94 force field (Halgren, 1995). Molecular docking employed biased probability Monte Carlo (BPMC) optimization of the ligand internal coordinates in the grid potentials of the receptor (Totrov and Abagyan, 1997). To ensure convergence of the docking procedure, at least five independent docking runs were performed for each ligand starting from a random conformation. The results of individual docking runs for each ligand were considered consistent if at least three of the five docking runs produced similar ligand

conformations (RMSD < 2.0 Å) and Binding Score < -20.0 kJ/mol. The unbiased docking procedure did not use distance restraints or any other a priori derived information for the ligand-receptor interactions.

#### Ligand Binding Assays

Ligand binding was analyzed using total membranes prepared from COS-1 cells transiently expressing HA-AT<sub>1</sub>R (wild-type), ΔBRIL-AT<sub>1</sub>R (crystallized construct without BRIL), and BRIL-AT<sub>1</sub>R (crystallized construct) constructs. Single mutants were constructed by a PCR-based site-directed mutagenesis strategy as previously described (Unal et al., 2010). Protein concentration was determined by Bio-Rad Protein Assay (Bio-Rad). For both saturation and competition binding assays, 10 μg of homogenous cell membrane was used per well.

Saturation binding assays with <sup>3</sup>H-candesartan were performed under equilibrium conditions, with <sup>3</sup>H-candesartan (Amersham Pharmacia Biotech) concentrations ranging between 0.125 and 12 nM (specific activity, 16 Ci/mmol) as duplicates in 96-well plates for 1 hr at room temperature. Nonspecific binding was measured in the presence of 10 μM candesartan (gift from AstraZeneca). The binding kinetics were analyzed by nonlinear curve-fitting program GraphPad Prism 5, which yielded the mean ± SD for the K<sub>d</sub> and B<sub>max</sub> values.

Competition binding assays were performed under equilibrium conditions, with 2 nM <sup>3</sup>H-candesartan and various concentrations of the ZD7155 ranging between 0.04 and 1,000 nM. The binding kinetics were analyzed by nonlinear curve-fitting program GraphPad Prism 5, which yielded the mean ± SD for the IC<sub>50</sub> values.

#### Signaling Assays in Whole Cells

Calcium levels inside COS-1 cells transiently expressing different AT<sub>1</sub>R constructs were measured using a Fluorescent Imaging Plate Reader (FLIPR) Calcium 5 assay kit (Molecular Devices). For the antagonist dose-response, the cells were first treated with different concentrations of ZD7155 for 1 hr followed by stimulation with 100 nM AngII. The EC<sub>50</sub> values for AngII dose response were 0.2, 2, and 12 nM for HA-AT<sub>1</sub>R, ΔBRIL-AT<sub>1</sub>R, and BRIL-AT<sub>1</sub>R, respectively. The IC<sub>50</sub> values for ZD7155 to inhibit AngII response were between 3 to 4 nM for all constructs. The curves from a representative experiment wherein measurements are made in triplicate are shown as mean ± SEM. Additional information is available in the [Extended Experimental Procedures](#).

#### ACCESSION NUMBERS

The coordinates and structure factors have been deposited into the Protein Data Bank under the accession code 4YAY.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.011>.

#### AUTHOR CONTRIBUTIONS

H.Z. designed, optimized, purified, and characterized receptor constructs for structural studies, crystallized the receptor in LCP; collected and processed diffraction data; determined the structure; analyzed the data; and wrote the paper. H.U. performed mutagenesis, signaling, and ligand binding studies and contributed to writing the paper. C.G. participated in the XFEL data collection and processed XFEL data. G.W.H. solved and refined the AT<sub>1</sub>R structure. N.A.Z. participated in XFEL data collection and contributed in the XFEL data processing. D.J., D.W., G.N., and U.W. designed, prepared, and operated the LCP injector during the XFEL data collection. M.M., G.J.W., and S.B. operated the CXI beamline and performed the XFEL data collection. O.M.Y. refined the geometry of the CSPAD detector. T.A.W. implemented new data processing algorithms in CrystFEL, used in this study. W.L., C.W. and A.I. helped with XFEL sample preparation and participated in the XFEL data collection. K.C.T. and R.D. participated in mutagenesis, membrane production, signaling, and ELISA data collection and analysis. M.R.S. and Q.X. helped with synchrotron

data processing, XFEL structure solution and structure validation. J.C., C.E.C. and P.F. helped with biophysical characterization of microcrystals at LCLS and participated in the XFEL data collection. R.C.S. conceived the project, supervised receptor expression and characterization, and contributed to writing the paper. V.K. designed initial AT<sub>1</sub>R constructs, analyzed the structure, performed docking studies, and wrote the paper. S.S.K. conceived the project, supervised mutagenesis and functional studies, and contributed to writing the paper. V.C. conceived the project, supervised crystallization and crystallographic data collection, analyzed the data and wrote the paper. Edits were provided by C.G., G.W.H., T.A.W., U.W. and P.F.

#### ACKNOWLEDGMENTS

This work was supported in parts by the NIH grants R01 GM108635 (V.C.), U54 GM094618 (target GPCR-11) (V.K., V.C., R.C.S.), P41 GM103393 (S.B.), R01 HL57470 (S.S.K.), R01 HL115964 (S.S.K.), U54 GM094599 (P.F.), R01 GM095583 (P.F.), U54 GM094586 (Q.X.), and National Research Service award HL007914 (H.U.). Further support was provided by the NSF BioXFEL Science and Technology center grant 1231306 (P.F., U.W., G.N.); the Helmholtz Gemeinschaft, the DFG Cluster of Excellence “Center for Ultrafast Imaging”; the BMBF project FKZ 05K12CH1 (C.G., O.M.Y., T.A.W.); the PIER Helmholtz Graduate School and the Helmholtz Association (C.G.); and the Chinese 1000 Talent Program (R.C.S.). Parts of this research were carried out at the LCLS, a National User Facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences, and at the GM/CA CAT of the Argonne Photon Source, Argonne National Laboratory. We thank J. Velasquez for help with molecular biology; T. Trinh and M. Chu for help with baculovirus expression; M. Metz, D. Oberthuer, A. Barty, S. Basu, and R. Fromme for help with the XFEL data collection and analysis, and R. Miller, K. Kadyshvskaya, and A. Walker for assistance with manuscript preparation. We are grateful to J. Spence and H. Chapman for their encouragement and support.

Received: November 6, 2014

Revised: December 30, 2014

Accepted: March 2, 2015

Published: April 23, 2015

#### REFERENCES

- Akazawa, H., Yano, M., Yabumoto, C., Kudo-Sakamoto, Y., and Komuro, I. (2013). Angiotensin II type 1 and type 2 receptor-induced cell signaling. *Curr. Pharm. Des.* 19, 2988–2995.
- Balakumar, P., and Jagadeesh, G. (2014). Structural determinants for binding, activation, and functional selectivity of the angiotensin AT1 receptor. *J. Mol. Endocrinol.* 53, R71–R92.
- Balakumar, P., Bishnoi, H.K., and Mahadevan, N. (2012). Telmisartan in the management of diabetic nephropathy: a contemporary view. *Curr. Diabetes Rev.* 8, 183–190.
- Ballesteros, J.A., and Weinstein, H. (1995). Integrated methods for the construction of three dimensional models and computational probing of structure–function relations in G-protein coupled receptors. *Methods Neurosci* 25, 366–428.
- Barty, A., Kirian, R.A., Maia, F.R., Hantke, M., Yoon, C.H., White, T.A., and Chapman, H. (2014). Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Cryst.* 47, 1118–1131.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonnrhein, C., et al. (2009). Buster version 2.8.0 (Cambridge, United Kingdom: Global Phasing Ltd).
- Burnier, M., and Brunner, H.R. (2000). Angiotensin II receptor antagonists. *Lancet* 355, 637–645.
- Caffrey, M., and Cherezov, V. (2009). Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* 4, 706–731.



- Chun, E., Thompson, A.A., Liu, W., Roth, C.B., Griffith, M.T., Katritch, V., Kunken, J., Xu, F., Cherezov, V., Hanson, M.A., and Stevens, R.C. (2012). Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* 20, 967–976.
- de Gasparo, M., Catt, K.J., Inagami, T., Wright, J.W., and Unger, T. (2000). International union of pharmacology. XXIII. The angiotensin II receptors. *Pharmacol. Rev.* 52, 415–472.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
- Fenalti, G., Giguere, P.M., Katritch, V., Huang, X.P., Thompson, A.A., Cherezov, V., Roth, B.L., and Stevens, R.C. (2014). Molecular control of  $\delta$ -opioid receptor signalling. *Nature* 506, 191–196.
- Fu, M.L., Herlitz, H., Schulze, W., Wallukat, G., Micke, P., Eftekhari, P., Sjögren, K.G., Hjalmarsen, A., Müller-Esterl, W., and Hoebeke, J. (2000). Autoantibodies against the angiotensin receptor (AT1) in patients with hypertension. *J. Hypertens.* 18, 945–953.
- Halgren, T. (1995). Merck molecular force field I-V. *J Comp Chem* 17, 490–641.
- Hart, P., Boutet, S., Carini, G., Dragone, A., Duda, B., Freytag, D., Haller, G., Herbst, R., Herrmann, S., Kenney, C., et al. (2012). The Cornell-SLAC Pixel Array Detector at LCLS. In Nuclear Science Symposium, Medical Imaging Conference.
- Imaizumi, S., Miura, S., Yahiro, E., Uehara, Y., Komuro, I., and Saku, K. (2013). Class- and molecule-specific differential effects of angiotensin II type 1 receptor blockers. *Curr. Pharm. Des.* 19, 3002–3008.
- Junggren, I.L., Zhao, X., Sun, X., and Hedner, T. (1996). Comparative cardiovascular effects of the angiotensin II type 1 receptor antagonists ZD 7155 and losartan in the rat. *J. Pharm. Pharmacol.* 48, 829–833.
- Kamimori, H., Unabia, S., Thomas, W.G., and Aguilar, M.I. (2005). Evaluation of the membrane-binding properties of the proximal region of the angiotensin II receptor (AT1A) carboxyl terminus by surface plasmon resonance. *Anal. Sci.* 27, 171–174.
- Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030–1033.
- Katritch, V., Fenalti, G., Abola, E.E., Roth, B.L., Cherezov, V., and Stevens, R.C. (2014). Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* 39, 233–244.
- Keane, M.G., and Pyeritz, R.E. (2008). Medical management of Marfan syndrome. *Circulation* 117, 2802–2813.
- Liu, W., Wacker, D., Gati, C., Han, G.W., James, D., Wang, D., Nelson, G., Weierstall, U., Katritch, V., Barty, A., et al. (2013). Serial femtosecond crystallography of G protein-coupled receptors. *Science* 342, 1521–1524.
- Liu, W., Wacker, D., Wang, C., Abola, E., and Cherezov, V. (2014a). Femtosecond crystallography of membrane proteins in the lipidic cubic phase. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130314.
- Liu, W., Ishchenko, A., and Cherezov, V. (2014b). Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nat. Protoc.* 9, 2123–2134.
- Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40, D370–D376.
- Ma, T.K., Kam, K.K., Yan, B.P., and Lam, Y.Y. (2010). Renin-angiotensin-aldosterone system blockade for cardiovascular diseases: current status. *Br. J. Pharmacol.* 160, 1273–1292.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.
- Miura, S., Zhang, J., Boros, J., and Karnik, S.S. (2003). TM2-TM7 interaction in coupling movement of transmembrane helices to activation of the angiotensin II type-1 receptor. *J. Biol. Chem.* 278, 3720–3725.
- Miura, S., Kiyu, Y., Kanazawa, T., Imaizumi, S., Fujino, M., Matsuo, Y., Karnik, S.S., and Saku, K. (2008). Differential bonding interactions of inverse agonists of angiotensin II type 1 receptor in stabilizing the inactive state. *Mol. Endocrinol.* 22, 139–146.
- Miura, S., Karnik, S.S., and Saku, K. (2011). Review: angiotensin II type 1 receptor blockers: class effects versus molecular effects. *J. Renin Angiotensin Aldosterone Syst.* 12, 1–7.
- Miura, S., Nakao, N., Hanzawa, H., Matsuo, Y., Saku, K., and Karnik, S.S. (2013a). Reassessment of the unique mode of binding between angiotensin II type 1 receptor and their blockers. *PLoS ONE* 8, e79914.
- Miura, S., Okabe, A., Matsuo, Y., Karnik, S.S., and Saku, K. (2013b). Unique binding behavior of the recently approved angiotensin II receptor blocker azilsartan compared with that of candesartan. *Hypertens. Res.* 36, 134–139.
- Mozsolits, H., Unabia, S., Ahmad, A., Morton, C.J., Thomas, W.G., and Aguilar, M.I. (2002). Electrostatic and hydrophobic forces tether the proximal region of the angiotensin II receptor (AT1A) carboxyl terminus to anionic lipids. *Biochemistry* 41, 7830–7840.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.
- Oliveira, L., Costa-Neto, C.M., Nakaie, C.R., Schreier, S., Shimuta, S.I., and Paiva, A.C. (2007). The angiotensin II AT1 receptor structure-activity correlations in the light of rhodopsin structure. *Physiol. Rev.* 87, 565–592.
- Ramanath, V.S., Oh, J.K., Sundt, T.M., 3rd, and Eagle, K.A. (2009). Acute aortic syndromes and thoracic aortic aneurysm. *Mayo Clin. Proc.* 84, 465–481.
- Ramchandran, R., Takezako, T., Saad, Y., Stull, L., Fink, B., Yamada, H., Dikalov, S., Harrison, D.G., Moravec, C., and Karnik, S.S. (2006). Angiotensinergic stimulation of vascular endothelium in mice causes hypotension, bradycardia, and attenuated angiotensin response. *Proc. Natl. Acad. Sci. USA* 103, 19087–19092.
- Sano, T., Ohyama, K., Yamano, Y., Nakagomi, Y., Nakazawa, S., Kikyo, M., Shirai, H., Blank, J.S., Exton, J.H., and Inagami, T. (1997). A domain for G protein coupling in carboxyl-terminal tail of rat angiotensin II receptor type 1A. *J. Biol. Chem.* 272, 23631–23636.
- Shenoy, S.K., and Lefkowitz, R.J. (2005). Angiotensin II-stimulated signaling through G proteins and beta-arrestin. *Sci. STKE* 2005, cm14.
- Takezako, T., Gogonea, C., Saad, Y., Noda, K., and Karnik, S.S. (2004). “Network leaning” as a mechanism of insurmountable antagonism of the angiotensin II type 1 receptor by non-peptide antagonists. *J. Biol. Chem.* 279, 15248–15257.
- Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G.W., Kufareva, I., Li, T., Ma, L., Fenalti, G., Li, J., et al. (2013). Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* 341, 1387–1390.
- Thomas, W.G., Baker, K.M., Motel, T.J., and Thekkumkara, T.J. (1995). Angiotensin II receptor endocytosis involves two distinct regions of the cytoplasmic tail. A role for residues on the hydrophobic face of a putative amphipathic helix. *J. Biol. Chem.* 270, 22153–22159.
- Thomas, W.G., Pipolo, L., and Qian, H. (1999). Identification of a Ca<sup>2+</sup>/calmodulin-binding domain within the carboxyl-terminus of the angiotensin II (AT1A) receptor. *FEBS Lett.* 455, 367–371.
- Thompson, A.A., Liu, W., Chun, E., Katritch, V., Wu, H., Vardy, E., Huang, X.P., Trapella, C., Guerrini, R., Calo, G., et al. (2012). Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395–399.
- Totrov, M., and Abagyan, R. (1997). Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*, 215–220.
- Unal, H., and Karnik, S.S. (2014). Constitutive activity in the angiotensin II type 1 receptor: discovery and applications. *Adv. Pharmacol.* 70, 155–174.
- Unal, H., Jagannathan, R., Bhat, M.B., and Karnik, S.S. (2010). Ligand-specific conformation of extracellular loop-2 in the angiotensin II type 1 receptor. *J. Biol. Chem.* 285, 16341–16350.
- Unal, H., Jagannathan, R., and Karnik, S.S. (2012). Mechanism of GPCR-directed autoantibodies in diseases. *Adv. Exp. Med. Biol.* 749, 187–199.
- Weierstall, U., James, D., Wang, C., White, T.A., Wang, D., Liu, W., Spence, J.C., Bruce Doak, R., Nelson, G., Fromme, P., et al. (2014). Lipidic cubic phase

- injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309.
- Whalen, E.J., Rajagopal, S., and Lefkowitz, R.J. (2011). Therapeutic potential of  $\beta$ -arrestin- and G protein-biased agonists. *Trends Mol. Med.* **17**, 126–139.
- White, T.A., Kirian, R.A., Martin, A.V., Aquila, A., Nass, K., Barty, A., and Chapman, H.N. (2012). CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Cryst.* **45**, 335–341.
- Wisler, J.W., Xiao, K., Thomsen, A.R., and Lefkowitz, R.J. (2014). Recent developments in biased agonism. *Curr. Opin. Cell Biol.* **27**, 18–24.
- Wu, B., Chien, E.Y., Mol, C.D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F.C., et al. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330**, 1066–1071.
- Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G.W., Vardy, E., Liu, W., Thompson, A.A., Huang, X.P., Carroll, F.I., et al. (2012). Structure of the human  $\kappa$ -opioid receptor in complex with JDTic. *Nature* **485**, 327–332.
- Xia, Y., and Kellems, R.E. (2013). Angiotensin receptor agonistic autoantibodies and hypertension: preeclampsia and beyond. *Circ. Res.* **113**, 78–87.
- Zaman, M.A., Oparil, S., and Calhoun, D.A. (2002). Drugs targeting the renin-angiotensin-aldosterone system. *Nat. Rev. Drug Discov.* **1**, 621–636.
- Zhou, C.C., Zhang, Y., Irani, R.A., Zhang, H., Mi, T., Popek, E.J., Hicks, M.J., Ramin, S.M., Kellems, R.E., and Xia, Y. (2008). Angiotensin receptor agonistic autoantibodies induce pre-eclampsia in pregnant mice. *Nat. Med.* **14**, 855–862.

## 9.2 Chemicals and Formulations

### 9.2.1 Chemicals - Hazard and Precautionary Statements

Compound	CAS-No.	Supplier	GHS Hazard	Hazard Statements	Precautionary Statements
----------	---------	----------	------------	-------------------	--------------------------

This work did not involve any handling of chemicals.

## 9.3 Curriculum Vitae

# Curriculum Vitae

## Personal Information

### Cornelius Gati

Juergen-Toepfer-Str. 2

22763 Hamburg

☎ ++49176/82007181

✉ cornelius.gati@desy.de

Title	Dr. rer. nat.
Name	Cornelius Gati
Date of Birth	30 <sup>th</sup> of June 1987
Nationality	German, Hungarian



## Education & Training

**01.09.2012 – 30.11.2015 Ph.D. thesis - Prof. Dr. Henry N. Chapman**

*Institution* Deutsches Elektronensynchrotron (DESY)

*Department* Center for Free-Electron-Laser Science (CFEL)

*Degree* PhD – Department of Chemistry, University of Hamburg (UHH)

*Year* 2012 - 2015

***Final grade* Summa cum laude**

*Title* Data Processing and Analysis in Serial Crystallography at Advanced Light Sources

*For reference please contact* Prof. Dr. Chapman, Henry N.

☎ ++494089984155

✉ henry.chapman@desy.de

**13.09.2013 – 23.10.2013 Internship at Coherent X-ray Imaging Beamline (CXI) – Dr. Sebastien Boutet**

*Institution* Linear Coherent Light Source (LCLS), SLAC, CA, USA

*Year* 2013

*Title* Characterization of the beam profile of a nanofocused x-ray free-electron-laser beam

*Field of research/study* SEM, AFM, Wire scans, Imprint series, X-ray beam alignment, Instrumentation

**October 2010 – September 2012 Master degree program at the Technical University Munich**

*Institution* Research centre Weihenstephan, TU Munich, Germany

*Degree* Master of Science - Biology

*Year* 2010-2012

*Final average degree* 1.4

*Field of study* Protein biochemistry, Molecular Oncology, Biotechnology

<b>01.04.2012 – 31.08.2012</b>	<b>M.Sc. thesis at Paul Scherrer Institute – Prof. Dr. Schertler</b>
<i>Institution</i>	Schertler Group, Department of Biomolecular Research, PSI, CH
<i>Year</i>	2012
<i>Title</i>	Preparation of Rhodopsin Nanocrystals for Kinetic Studies using X-ray Free-Electron-Lasers
<i>Field of research/study</i>	Membrane protein chemistry – Protein purification (Solubilization of Membrane proteins, Affinity Chromatography, Ion Exchange Chromatography), Crystallization (Vapor diffusion, Batch), Characterization of crystals by diffraction measurements at the Swiss Light Source
	<p style="text-align: right;"><i>For reference please contact</i>   Prof. Schertler, Gebhard    ☎ ++41563104265    ✉ gebhard.schertler@psi.ch</p>
<b>01.10.2011 – 27.02.2012</b>	<b>Internship at Roche Diagnostics GmbH, Pharma Research and Early Development – Dr. Hoffmann</b>
<i>Institution</i>	Hoffmann Group, Department of Protein Sciences
<i>Year</i>	2011/2012
<i>Field of research/study</i>	Protein chemistry – Protein purification, MabSelect Affinity Chromatography, SEC, Disulfide coupling reactions
<b>01.08.2011 – 25.09.2011</b>	<b>Internship at the Technical University of Munich, Department Biological Chemistry – Prof. Dr. Skerra</b>
<i>Institution</i>	Skerra Group, Technical University of Munich, Germany
<i>Year</i>	2011
<i>Field of research/study</i>	Purification and crystallization of two Fibronectin specific Anticalines, in complex with their targets
	E. coli Fermentation, periplasmic and cytoplasmatic Expression, Streptavidin-Affinity Chromatography, IMAC, IEC, SEC, SDS-PAGE, Cloning, PCR, Transformation, DNA-Sequencing, Crystallization approaches
<b>14.02.2011 – 01.04.2011</b>	<b>Internship at the Technical University of Munich, Department Chemistry – Chair in Biotechnology - Prof. Dr. Buchner</b>
<i>Institution</i>	Buchner Group, Technical University of Munich, Germany
<i>Year</i>	2011
<i>Field of research/study</i>	Purification and characterization of the human surrogate light chain (SLC)
	E. coli Fermentation, Inclusion bodies purification, IMAC, IEC, SEC (Superdex™), SDS-PAGE, Western Blot, Cloning, PCR, Transformation, Circular Dichroism, Thermal Shift Assay
<b>October 2007 – August 2010</b>	<b>Bachelor degree program at the Eberhard Karls University of Tübingen</b>
<i>Institution</i>	Faculty of Biology, University of Tübingen, Germany
<i>Degree</i>	Bachelor of Science - Biology
<i>Year</i>	October 2007 – August 2010
<i>Title of Bachelor thesis</i>	“HLA ligands and gene expression of renal oncocytoma and chromophobe renal cell carcinoma” – Department of Immunology – Prof. Dr. Rammensee
<i>Methods</i>	Immunoprecipitation, Stable isotope labeling of peptides, Liquid-Chromatography Mass spectrometry (LC-MS/MS), Microarray data analysis, quantitative ELISA
<i>Final average degree</i>	1,6
<b>08.03.2010 – 16.04.2010</b>	<b>Internship at the Universitätsklinikum Tübingen, Department Medical Virology – Prof. Dr. Iftner</b>
<i>Institution</i>	Stubenrauch Group, Universitätsklinikum Tübingen, Germany
<i>Year</i>	2010
<i>Field of research/study</i>	Interaction partners of the viral (HPV) protein E8 <sup>E2C</sup>
	Cloning of DNA, Cell culture, Co-Immunoprecipitations, Western Blot, PCR,

	Transformation of bacteria, Transfection of eukaryotic cells
<b>10.08.2009 – 18.09.2009</b>	<b>Internship at the DKFZ Heidelberg, Department Molecular Immunology – Prof. Dr. Hämmerling</b>
<i>Institution</i>	Momburg Group, DKFZ Heidelberg, Germany
<i>Year</i>	2009
<i>Field of research/study</i>	Lentiviral Transduction of NK92 Cells with siRNA targetting NCRs
	Cloning of cDNA-fragments, DNA preparation, Cell culture, MACS-Technique, IFN- $\gamma$ capture assay, ELISA, FACS, Cr-release assays (NK-Cytotoxicity assay), Lentiviral transduction
<b>09.03.2009 – 20.04.2009</b>	<b>Internship at the Universitätsklinik Heidelberg, Department Tumor cellbiology – Prof. Dr. Zöller</b>
<i>Institution</i>	Zöller Group, Uniklinik Heidelberg, Germany
<i>Year</i>	2009
<i>Field of research/study</i>	Cloning of the RCC associated antigen MAGE-A9 into in-vitro transcription-plasmid pST1-p(120)A
	Cloning, Transformation, in vitro Transcription, mRNA Transfection, Electroporation, basic cell culture
<b>September 1997 – June 2006</b>	<b>Abitur at Gymnasium Plochingen</b>
<i>Degree</i>	Secondary school Diploma (Abitur)
<i>Year</i>	September 1997 – August 2006
<i>Field of study</i>	Secondary school education

## *Positions*

<b>November 2010 – January 2011</b>	<b>Student assistantship at Antes Group (Protein Modeling), Technical University of Munich</b>
<b>May 2011 – July 2011</b>	
<i>Main activities</i>	Evaluation and development of MHC prediction algorithms
<i>Holding Position</i>	Student assistant
<i>Institution</i>	TU Munich, Theoretical Chemical Biology and Protein Modelling Group – Prof. Dr. Antes
<b>November 2008 – August 2010</b>	<b>Student assistantship at Stevanovic Group, University of Tübingen</b>
<i>Main activities</i>	Supervision of SYFPEITHI database ( <a href="http://www.syfpeithi.de">www.syfpeithi.de</a> ), Literature research
<i>Holding Position</i>	Student assistant
<i>Institution</i>	University of Tübingen, Department of Immunology – Prof. Dr. Rammensee
<b>October 2006 – June 2007</b>	<b>Civilian Service at the otorhinolaryngological operating theatre Katharinenhopsital Stuttgart</b>
<i>Main activities</i>	Patients care
<i>Position held</i>	Conscientious objector
<i>Institution</i>	Katharinenhospital Stuttgart

## Publications

- Nature methods* 2015 (accepted) Sierra RG, **Gati C**, Laksmono H, Han Dao E, Gul S, Fuller F, Kern J, Chatterjee R, Ibrahim M, Brewster A, Young ID, Michels-Clark T, Aquila A, Liang M, Hunster MS, Koglin JE, Boutet S, Junco EA, Hayes B, Bogan MJ, Hampton CY, Puglisi EV, Sauter NK, Stan CA, Zouni A, Yano J, Yachandra VK, Soltis SM, Puglisi JD, DeMirici H
- Serial Femtosecond Crystallography of Ribosome-Antibiotic and Photosystem II Complexes at Ambient Temperature using a Concentric Electrospinning Injector.**
- Optics Express* 23(22) 28459-28470 (2015) Yefanov O, Mariani V, **Gati C**, White TA, Chapman HN, Barty A
- Accurate determination of segmented X-ray detector geometry**
- Nature* 523,561–567 (2015) Kang Y., Zhou EX, Gao X, He Y, Liu W, Ishchenko A, Barty A, White TA, Yefanov O, Han GW, Xu Q, de Waal PW, Ke J, Tan MHE, Zhang C, Moeller A, West GM, Pascal B, van Eps N, Caro LN, Vishnivetskiy SA, Lee RJ, Suino-Powell KM, Gu X, Pal K, Ma J, Zhi X, Boutet S, Williams GJ, Messerschmidt M, **Gati C**, Zatsepin NA, Wang D, James D, Basu S, Roy-Chowdhury S, Conrad C, Coe J, Liu H, Lisova S, Kupitz C, Grotjohann I, Fromme R, Jiang Y, Tan M, Yang H, Li J, Wang M, Zheng Z, Li D, Howe N, Zhao Y, Standfuss J, Diederichs K, Dong Y, Potter CS, Carragher B, Caffrey M, Jiang H, Chapman HN, Spence JCH, Fromme P, Weierstall U, Ernst OP, Katritch V, Gurevich VV, Griffin PR, Hubbell WR, Stevens RC, Cherezov V, Melcher K, Xu HE
- Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser**
- IUCrJ* Jun 30;2(Pt 4):421-30 (2015) Conrad CE, Basu S, James D, Wang D, Schaffer A, Roy-Chodhury S, Zatsepin NA, Aquila A, Coe J, **Gati C**, Hunter MS, Koglin JE, Kupitz C, Nelson G, Subramanian G, White TA, Zhao Y, Zook J, Boutet Sebastien, Cherezov V, Spence JCH, Fromme R, Weierstall U, Fromme P
- A novel inert crystal delivery medium for serial femtosecond crystallography**
- Acta Cryst. F* (2015) Jun 27; F71 Wu W, Nogly P, Rheinberger J, Kick LM, **Gati C**, Nelson G, Deupi X, Standfuss J, Schertler G, Panneels V
- Batch crystallization of rhodopsin for structural dynamics using an X-ray free electron laser (X-FEL)**
- Structural Dynamics* 2, 041718 (2015) Panneels V, Wu W, Tsai CJ, Nogly Przemek, Rheinberger J, Jaeger K, Cicchetti G, **Gati C**, Kick LM, Sala L, Capitani G, Milne C, Padeste C, Pedrini B, Li XD, Standfuss J, Abela R, Schertler G
- Time-resolved structural studies with serial crystallography: A new light on retinal proteins**
- Structural Dynamics* 2.4 (2015): 041703. Galli L, Son SK, Klinge M, Bajt S, Barty A, Bean R, Betzel C, Beyerlein KR, Caleman C, Doak RB, Duszenko M, Fleckenstein H, **Gati C**, Hunt B, Kirian RA, Liang M, Nanao MH, Nass K, Oberthuer D, Redecke L, Shoeman R, Stellato F, Yoon CH, White TA, Yefanov O, Spence J, Chapman HN
- Electronic damage in S atoms in a native protein crystal induced by an intense X-ray free-electron laser pulse**

- J Synchrotron Radiat.* 2015 May 1;**22**(Pt 3):644-652. Epub 2015 Apr 22. Barends T, White TA, Barty A, Foucar L, Messerschmidt M, Alonso-Mori R, Botha S, Chapman H, Doak RB, Galli L, **Gati C**, Gutmann M, Koglin J, Markvardsen A, Nass K, Oberthur D, Shoeman RL, Schlichting I, Boutet S.
- Effects of self-seeding and crystal post-selection on the quality of Monte Carlo-integrated SFX data.**
- Cell.* 2015 Apr 21. S0092-8674(15)00428-6 Zhang H, Unal H, **Gati C**, Han GW, Liu W, Zatsepin NA, James D, Wang D, Nelson G, Weierstall U, Sawaya MR, Xu Q, Messerschmidt M, Williams GJ, Boutet S, Yefanov OM, White TA, Wang C, Ishchenko A, Tirupula KC, Desnoyer R, Coe J, Conrad CE, Fromme P, Stevens RC, Katritch V, Karnik SS, Cherezov V.
- Structure of the Angiotensin Receptor Revealed by Serial Femtosecond Crystallography.**
- Nature Structural and Molecular Biology.* 2015 Mar;**22**(3):265-8 Fenalti G, Zatsepin N, Betti C, Giguere P, Ishchenko A, Liu W, Guillemy N, Han GW, Zhang H, James D, Wang D, Weierstall U, Spence JCH, Boutet S, Messerschmidt M, Williams GJ, **Gati C**, White TA, Oberthuer D, Metz M, Yoon CH, Barty A, Chapman H.N., Basu S, Coe J, Conrad C, Fromme R, Fromme P, Tourwé D, Schiller PW, Roth BL, Ballet S, Katritch V, Stevens RC, Cherezov V
- Structural basis for bifunctional peptide recognition at human delta-opioid receptor**
- Science.* 2014, **346**(6214), 1242-1246 Tenboer J, Basu S, Zatsepin N, Pande K, Milathianaki D, Frank M, Hunter M, Boutet S, Williams G, Koglin JE, Oberthuer D, Heymann M, Kupitz C, Conrad C, Coe J, Roy-Chowdhury S, Weierstall U, James D, Wang D, Grant T, Barty A, Yefanov O, Scales J, **Gati C**, Seuring C, Srajer V, Henning R, Schwander P, Fromme R, Ourmazd A, Moffat K, Van Thor J, Spence JHC, Fromme P, Chapman HN, Schmidt M
- Time Resolved Serial Femtosecond Crystallography Captures High Resolution Intermediates of Photoactive Yellow Protein**
- Phil. Trans. R. Soc. B.* 2014 369 Yefanov O, **Gati C**, White TA, Spence JCH, Chapman HN, Barty A
- Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography**
- IUCrJ.* 2014 **1**, 204-212 Stellato F, Oberthuer D, Liang M, Bean R, **Gati C**, Yefanov O, Barty A, Burkhardt A, Fisher P, Galli L, Kirian L, Meyer J, Panneerselvam S, Yoon CH, Chervinskii F, Speller E, White TA, Betzel C, Meents A, Chapman HN
- Room temperature macromolecular serial crystallography using synchrotron radiation**
- Nature Communications.* 2014, 5:3309 Weierstall U, James D, Wang C, White TA, Wang D, Liu W, Spence JCH, Doak RB, Nelson G, Fromme P, Fromme R, Grotjohann I, Kupitz C, Zatsepin N, Liu H, Basu S, Wacker D, Han GW, Katritch V, Boutet S, Messerschmidt M, Williams GJ, Koglin JE, Seibert MM, Klinker M, **Gati C**, Shoeman RL, Barty A, Chapman HN, Kirian RA, Beyerlein KR, Stevens RC, Li D, Shah STA, Howe N, Caffrey M, Cherezov V
- Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography**



<i>IUCrJ.</i> 2014 <b>1</b> , 87-94	<b>Gati C</b> , Bourenkov G, Klinge M, Rehders D, Stellato F, Oberthuer D, Yefanov O, Sommer BP, Mogk S, Duszenko M, Betzel C, Schneider TR, Chapman HN, Redecke L  <b>Serial crystallography on in vivo grown microcrystals using synchrotron radiation</b>
<i>Science.</i> 2013; <b>342</b> (6165):1521-4	Liu W, Wacker D, <b>Gati C</b> , ... , Chapman HN, Caffrey M, Spence JHC, Stevens RC, Cherezov V  <b>Serial Femtosecond Crystallography of G Protein-Coupled Receptors</b>
<i>Acta Cryst. F.</i> 2013; <b>69</b> (Pt 9):1066-9	Demirci H, Sierra RG, Laksmono H, Shoeman RL, Botha S, Barends TRM, Nass K, Schlichting I, Doak RB, <b>Gati C</b> , Williams G, Boutet S, Messerschmidt M, Jogle G, Dahlberg AE, Gregory ST, Bogan MJ  <b>Serial femtosecond X-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an X-ray free-electron laser</b>
<i>PLoS Pathog.</i> 2011; 7(8):e1002195	Jarahian M, Fiedler M, Cohnen A, Djandji D, Hämmerling GJ, <b>Gati C</b> , Cerwenka A, Turner PC, Moyer RW, Watzl C, Hengel H, Momburg F  <b>Modulation of NKp30- and NKp46-Mediated NK Cell Responses by Poxviral Hemagglutinin</b>

### Grants and Fellowships

2015	<b>Human Frontiers Science Program</b> – Long-Term Fellowship for Postdoctoral stay at the Laboratory of Molecular Biology – MRC, Cambridge University, UK  <b>Novel methods for the investigation of DNA replisome structural dynamics by single-particle cryo-EM</b>
2012	<b>Helmholtz Association Fellowship</b> – PhD thesis at CFEL / DESY, D  <b>Data processing and analysis in serial crystallography at advanced light sources</b>
2012	<b>ERASMUS SMP Fellowship</b> - M.Sc. thesis at Paul-Scherrer-Institute, Villigen, CH  <b>Preparation of Rhodopsin Nanocrystals for Kinetic Studies using X-ray Free-Electron-Lasers</b>

### Talks and Conferences

25.07.2015 – 29.07.2015	American Crystallographic Association (ACA) meeting, Philadelphia, PA, USA ( <b>Session chair:</b> Advances in Multi-crystal approaches and Serial Crystallography)
07.06.2015 – 13.06.2015	Second International Summer School of Crystallography at DESY Hamburg, D (Prof. C. Giacovazzo), <b>Main Organizer: C. Gati</b>
01.06.2015 – 02.06.2015	Synchrotron and Neutron Workshop (SyNew) 2015, Utrecht, NL ( <b>Invited talk:</b> Serial femtosecond crystallography of G protein-coupled receptors)
09.04.2015 – 11.04.2015	CSSB Symposium – Systems in Infection Biology From Molecules to Organisms, Hamburg, D ( <b>Presentation title:</b> Serial crystallography using advanced light sources: pushing the limits)
25.02.2015 – 28.02.2015	2nd Ringberg Workshop on Structural Biology with FELs, Ringberg, D
09.11.2014 – 14.11.2014	NRAMM Workshop - Advanced Topics in EM Structure Determination, The Scripps Research Institute, La Jolla, CA, USA
06.10.2014 – 09.10.2014	PIER Graduate week, DESY, Hamburg, D

14.08.2014 – 26.08.2014 14.07.2014	The 6 <sup>th</sup> Brazil School for Single Particle Cryo-EM, Socorro, BR Laboratory of Macromolecular Biology, Medical Research Council, Cambridge University, GB ( <b>Invited talk</b> : Pushing the limits: Serial crystallography using advanced light sources)
30.05.2014 – 08.06.2014	International School of Crystallography, Erice, IT ( <b>Poster</b> : Serial crystallography using synchrotron radiation)
11.05.2014 – 17.05.2014	First International Summer School of Crystallography at DESY Hamburg, D (Prof. C. Giacovazzo), <b>Main Organizer: C. Gati</b>
07.05.2014	DESY Photon Science Meeting ( <b>Invited talk</b> : Serial crystallography using synchrotron radiation)
23.04.2014 – 24.04.2014	FEI and NeCEN Cryo-workshop, Leiden, NE
10.04.2014 – 12.04.2014	8th International Workshop on X-ray Radiation Damage to Biological Crystalline Samples, Hamburg, D ( <b>Poster</b> : Serial crystallography using synchrotron radiation)
08.04.2014 – 10.04.2014	British Crystallographic Association Spring Meeting 2014 Loughborough, Leicestershire, UK ( <b>Invited talk</b> : Serial crystallography using synchrotron radiation)
17.03.2014 – 20.03.2014	22 <sup>nd</sup> Annual Conference of the German Crystallographic Society (DGK) 2014 Berlin, D ( <b>Presentation title</b> : Serial crystallography using synchrotron radiation)
01.10.2013 – 04.10.2013	SSRL / LCLS Users' Meeting & Workshops 2013, SLAC, USA
08.09.2013 – 11.09.2013	International Conference on Biology and Synchrotron Radiation – EMBL Hamburg, D ( <b>Poster</b> : Serial crystallography of in vivo grown microcrystals using synchrotron radiation)
10.06.2013 – 14.06.2013	EMBO Practical Course on Exploiting Anomalous Scattering in Macromolecular Structure Determination, ESRF, FR ( <b>Poster</b> : Serial femtosecond crystallography of G-protein coupled receptors)
22.04.2013	Nanocrystal data analysis workshop (Barty, White) DESY, D
01.04.2013 – 07.04.2013	International Workshop on Fundamentals of Crystallography (Prof. C. Giacovazzo), University of Patras, GR
03.03.2013 – 05.03.2013	Fragments 2013 - Fourth RSC-BMCS Fragment-based Drug Discovery meeting, Diamond Light Source, GB
23.01.2013 – 25.01.2013	XFEL & Photon Science Users Meeting and Satellite Meeting DESY, D
20.11.2012	SwissFEL Photonics: Sample Mounting and Injection Workshop PSI, CH
18.06.2012 – 21.06.2012	Ultrafast X-ray Summer School SLAC, USA
13.04.2012	SwissFEL X-ray Free Electron Laser: Follow-up Workshop on Hard X-Ray Instrumentation at SwissFEL PSI, CH
05.05.2010	Proteomics: From technologies to targets and biomarkers NMI, D

### *Personal skills and competences*

<i>Languages</i>	German   Native Hungarian   Native English   Fluent Latin   Intermediate Latin certificate
<i>Computer skills and competences</i>	Good command of bioinformatical tools (Cheetah, CrystFEL, phenix, CCP4, XDS, PYMOL, COOT, VMD, DynaDock, AutoDock, IRECS, FlexX, Protein Discoverer, MassLynx, Mascot, Mayday, ApE, Unicorn, Uniprot etc.) Good command of Linux, C, Matlab, Python
<i>Memberships</i>	Gesellschaft für Biochemie und Molekularbiologie (GBM) e.V., Viva con Agua St. Pauli (Charity), DKMS, TASSO e.V., Hundehilfe Marika e.V. (Animal welfare in Hungary)



# Bibliography

- [1] Röntgen, W. C. *Annalen der Physik* **1898**, *300*, 1–11.
- [2] Friedrich, W.; Knipping, P.; Laue, M. *Annalen der Physik* **1913**, *346*, 971–988.
- [3] Bragg, W.; Bragg, W. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1913**, 428–438.
- [4] Norman, H. F.; Lonsdale, K.; et al. *International tables for X-ray crystallography*; Kynoch Press, 1952.
- [5] Amos, L.; Henderson, R.; Unwin, P. *Progress in biophysics and molecular biology* **1982**, *39*, 183–231.
- [6] Nannenga, B. L.; Shi, D.; Leslie, A. G.; Gonen, T. *Nature Methods* **2014**, *11*, 927–930.
- [7] Nederlof, I.; Li, Y. W.; van Heel, M.; Abrahams, J. P. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 852–859.
- [8] Nederlof, I.; van Genderen, E.; Li, Y.-W.; Abrahams, J. P. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 1223–1230.
- [9] Giacovazzo, C.; Monaco, H.; Viterbo, D.; Scordari, F.; Gilli, G.; Zanotti, G.; Catti, M. *Fundamentals of Crystallography*; IUC-Oxford University Press, 1992.
- [10] Giacovazzo, C.; Monaco, H.; Viterbo, D.; Scordari, F.; Gilli, G.; Zanotti, G.; Catti, M. *Acta Cryst* **1993**, *49*, 373–374.
- [11] Hahn, T.; Paufler, P. *International tables for crystallography, Vol. A. space-group symmetry, Dordrecht, Holland/Boston, USA, 1983, ISBN 90-277-1445-2*; Wiley Online Library, 1984; Vol. 19.
- [12] Aroyo, M. I. **1987**.

- [13] Arndt, U. W.; Wonacott, A. *Rotation method in crystallography*; North-Holland Pub. Co., 1977.
- [14] Weissenberg, K. *Zeitschrift für Physik* **1924**, *23*, 229–238.
- [15] Rossmann, M. G.; van Beek, C. G. *Acta Crystallographica Section D: Biological Crystallography* **1999**, *55*, 1631–1640.
- [16] Winkler, F.; Schutt, C. t.; Harrison, S. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **1979**, *35*, 901–911.
- [17] Rossmann, M. G. *Journal of Applied Crystallography* **1979**, *12*, 225–238.
- [18] Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666.
- [19] Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G. *Nature* **1960**, *185*, 416–422.
- [20] Robertson, J. M. *Nature* **1939**, *143*, 75–76.
- [21] Owen, R. L.; Axford, D.; Nettleship, J. E.; Owens, R. J.; Robinson, J. I.; Morgan, A. W.; Doré, A. S.; Lebon, G.; Tate, C. G.; Fry, E. E.; et al. *Acta Crystallographica Section D: Biological Crystallography* **2012**, *68*, 810–818.
- [22] Haas, D.; Rossmann, M. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **1970**, *26*, 998–1004.
- [23] Abergel, C. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 2167–2173.
- [24] Rossmann, M. G. *Acta Crystallographica Section D: Biological Crystallography* **2001**, *57*, 1360–1366.
- [25] Pendry, J. *Journal of Physics C: Solid State Physics* **1980**, *13*, 937.
- [26] Ramachandran, G.; Ramakrishnan, C.; Sasisekharan, V. *Journal of molecular biology* **1963**, *7*, 95–99.
- [27] Karplus, P. A.; Diederichs, K. *Science* **2012**, *336*, 1030–1033.
- [28] Garman, E. F.; Owen, R. L. *Acta Crystallographica Section D: Biological Crystallography* **2006**, *62*, 32–47.

- [29] Henke, B. L.; Gullikson, E. M.; Davis, J. C. *Atomic data and nuclear data tables* **1993**, *54*, 181–342.
- [30] Stenn, K.; Bahr, G. *Journal of ultrastructure research* **1970**, *31*, 526–550.
- [31] Chiu, W.; Downing, K.; Dubochet, J. *Journal of Microscopy* **1986**.
- [32] Henderson, R. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **1990**, *241*, 6–8.
- [33] Petsko, G. A. *Journal of molecular biology* **1975**, *96*, 381–392.
- [34] Hope, H. *Acta Crystallographica Section B: Structural Science* **1988**, *44*, 22–26.
- [35] Owen, R. L.; Rudiño-Piñera, E.; Garman, E. F. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 4912–4917.
- [36] Zeldin, O. B.; Gerstel, M.; Garman, E. F. *Applied Crystallography* **2013**, *46*, 1225–1230.
- [37] Heydenreich, F. M.; Vuckovic, Z.; Matkovic, M.; Veprintsev, D. B. *Frontiers in pharmacology* **2015**, *6*.
- [38] Coulibaly, F.; Chiu, E.; Ikeda, K.; Gutmann, S.; Haebel, P. W.; Schulze-Briese, C.; Mori, H.; Metcalf, P. *Nature* **2007**, *446*, 97–101.
- [39] Landau, E. M.; Rosenbusch, J. P. *Proceedings of the National Academy of Sciences* **1996**, *93*, 14532–14535.
- [40] Cherezov, V.; Clogston, J.; Papiz, M. Z.; Caffrey, M. *Journal of molecular biology* **2006**, *357*, 1605–1618.
- [41] Cherezov, V.; Liu, J.; Griffith, M.; Hanson, M. A.; Stevens, R. C. *Crystal Growth and Design* **2008**, *8*, 4307–4315.
- [42] Cherezov, V.; Hanson, M. A.; Griffith, M. T.; Hilgart, M. C.; Sanishvili, R.; Nagarajan, V.; Stepanov, S.; Fischetti, R. F.; Kuhn, P.; Stevens, R. C. *Journal of The Royal Society Interface* **2009**, *6*, S587–S597.
- [43] Huang, C.-Y.; Olieric, V.; Ma, P.; Panepucci, E.; Diederichs, K.; Wang, M.; Caffrey, M. *Acta Crystallographica Section D: Biological Crystallography* **2015**, *71*, 0–0.
- [44] Crookes, W. *Proceedings of the Royal Society of London* **1878**, *28*, 102–111.

- [45] Crookes, W. *Proceedings of the Royal Society of London* **1878**, *28*, 347–352.
- [46] Robinson, I.; Gruebel, G.; Mochrie, S. *New Journal of Physics* **2010**, *12*, 035002.
- [47] Elder, F.; Gurewitsch, A.; Langmuir, R.; Pollock, H. *Physical Review* **1947**, *71*, 829.
- [48] Kirkpatrick, P.; Baez, A. V. *JOSA* **1948**, *38*, 766–773.
- [49] Phillips, W. C.; Stanton, M.; Stewart, A.; Qian, H.; Ingersoll, C.; Sweet, R. M. *Journal of applied crystallography* **2000**, *33*, 243–251.
- [50] Kraft, P.; Bergamaschi, A.; Broennimann, C.; Dinapoli, R.; Eikenberry, E.; Henrich, B.; Johnson, I.; Mozzanica, A.; Schlepütz, C.; Willmott, P.; et al. *Journal of synchrotron radiation* **2009**, *16*, 368–375.
- [51] Madey, J. M. *Journal of Applied Physics* **1971**, *42*, 1906–1913.
- [52] Deacon, D. A.; Elias, L.; Madey, J. M.; Ramian, G.; Schwettman, H.; Smith, T. I. *Physical Review Letters* **1977**, *38*, 892–894.
- [53] Emma, P.; et al. *Proceedings of PAC09, Vancouver, to be published in <http://accelconf.web.cern.ch/AccelConf>* **2009**.
- [54] Saldin, E.; Schneidmiller, E.; Yurkov, M. *Proceedings FEL 2006, on JACOW* **2006**, 206–209.
- [55] Saldin, E.; Schneidmiller, E.; Yurkov, M. *Optics Communications* **2008**, *281*, 4727–4734.
- [56] Pellegrini, C. *Proc. of the Workshop on Fourth Generation Light Sources* **1992**, 364.
- [57] Boutet, S.; Williams, G. J. *New Journal of Physics* **2010**, *12*, 035024.
- [58] Pothineni, S. B.; Venugopalan, N.; Ogata, C. M.; Hilgart, M. C.; Stepanov, S.; Sanishvili, R.; Becker, M.; Winter, G.; Sauter, N. K.; Smith, J. L.; et al. *Journal of applied crystallography* **2014**, *47*, 1992–1999.
- [59] Riekkel, C.; Burghammer, M.; Schertler, G. *Current opinion in structural biology* **2005**, *15*, 556–562.
- [60] Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; DePonte, D. P.; Weierstall, U.; et al. *Nature* **2011**, *470*, 73–77.

- [61] Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. *Nature* **2000**, *406*, 752–757.
- [62] Meidinger, N.; Andritschke, R.; Hartmann, R.; Herrmann, S.; Holl, P.; Lutz, G.; Strüder, L. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **2006**, *565*, 251–257.
- [63] Boutet, S.; Lomb, L.; Williams, G. J.; Barends, T. R.; Aquila, A.; Doak, R. B.; Weierstall, U.; DePonte, D. P.; Steinbrener, J.; Shoeman, R. L.; et al. *Science* **2012**, *337*, 362–364.
- [64] Redecke, L.; Nass, K.; DePonte, D. P.; White, T. A.; Rehders, D.; Barty, A.; Stellato, F.; Liang, M.; Barends, T. R.; Boutet, S.; et al. *Science* **2013**, *339*, 227–230.
- [65] DePonte, D.; Weierstall, U.; Schmidt, K.; Warner, J.; Starodub, D.; Spence, J.; Doak, R. *Journal of Physics D: Applied Physics* **2008**, *41*, 195505.
- [66] Lomb, L.; Steinbrener, J.; Bari, S.; Beisel, D.; Berndt, D.; Kieser, C.; Lukat, M.; Neef, N.; Shoeman, R. L. *Applied Crystallography* **2012**, *45*, 674–678.
- [67] Johansson, L. C.; Arnlund, D.; White, T. A.; Katona, G.; DePonte, D. P.; Weierstall, U.; Doak, R. B.; Shoeman, R. L.; Lomb, L.; Malmerberg, E.; et al. *Nature Methods* **2012**, *9*, 263–265.
- [68] Weierstall, U.; James, D.; Wang, C.; White, T. A.; Wang, D.; Liu, W.; Spence, J. C.; Doak, R. B.; Nelson, G.; Fromme, P. *Nature Communications* **2014**, *5*.
- [69] Liu, W.; Wacker, D.; Gati, C.; Han, G. W.; James, D.; Wang, D.; Nelson, G.; Weierstall, U.; Katritch, V.; Barty, A.; et al. *Science* **2013**, *342*, 1521–1524.
- [70] Zhang, H.; Unal, H.; Gati, C.; Han, G. W.; Liu, W.; Zatsepin, N. A.; James, D.; Wang, D.; Nelson, G.; Weierstall, U.; et al. *Cell* **2015**.
- [71] Sugahara, M.; Mizohata, E.; Nango, E.; Suzuki, M.; Tanaka, T.; Masuda, T.; Tanaka, R.; Shimamura, T.; Tanaka, Y.; Suno, C.; et al. *Nature methods* **2015**, *12*, 61–63.
- [72] Conrad, C. E.; Basu, S.; James, D.; Wang, D.; Schaffer, A.; Roy-Chowdhury, S.; Zatsepin, N. A.; Aquila, A.; Coe, J.; Gati, C.; et al. *IUCrJ* **2015**, *2*, 421–430.



- [73] Philipp, H. T.; Koerner, L. J.; Hromalik, M. S.; Tate, M. W.; Gruner, S. M. *Nuclear Science, IEEE Transactions on* **2010**, *57*, 3795–3799.
- [74] Yefanov, O.; Mariani, V.; Gati, C.; White, T. A.; Chapman, H. N.; Barty, A. *Optics Express* **2015**, *23*, 28459–28470.
- [75] Becker, J.; Bianco, L.; Göttlicher, P.; Graafsma, H.; Hirsemann, H.; Jack, S.; Klyuev, A.; Marras, S. L. A.; Trunk, U.; Klanner, R.; et al. *arXiv preprint arXiv:1303.2523* **2013**.
- [76] Mozzanica, A.; Bergamaschi, A.; Cartier, S.; Dinapoli, R.; Greiffenberg, D.; Johnson, I.; Jungmann, J.; Maliakal, D.; Mezza, D.; Ruder, C.; et al. *Journal of Instrumentation* **2014**, *9*, C05010.
- [77] Hunter, M. S.; Segelke, B.; Messerschmidt, M.; Williams, G. J.; Zatsepin, N. A.; Barty, A.; Benner, W. H.; Carlson, D. B.; Coleman, M.; Graf, A.; et al. *Scientific reports* **2014**, *4*.
- [78] Barty, A.; Kirian, R. A.; Maia, F. R.; Hantke, M.; Yoon, C. H.; White, T. A.; Chapman, H. *Journal of applied crystallography* **2014**, *47*, 1118–1131.
- [79] White, T. A.; Kirian, R. A.; Martin, A. V.; Aquila, A.; Nass, K.; Barty, A.; Chapman, H. N. *Journal of Applied Crystallography* **2012**, *45*, 335–341.
- [80] Hattne, J.; Echols, N.; Tran, R.; Kern, J.; Gildea, R. J.; Brewster, A. S.; Alonso-Mori, R.; Glöckner, C.; Hellmich, J.; Laksmono, H.; et al. *Nature methods* **2014**, *11*, 545–548.
- [81] Powell, H. R. *Acta Crystallographica Section D: Biological Crystallography* **1999**, *55*, 1690–1695.
- [82] Powell, H. R.; Johnson, O.; Leslie, A. G. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 1195–1203.
- [83] Duisenberg, A. J. *Journal of Applied Crystallography* **1992**, *25*, 92–96.
- [84] Kabsch, W. *Acta Crystallographica Section D: Biological Crystallography* **2010**, *66*, 125–132.
- [85] Brehm, W.; Diederichs, K. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *70*, 101–109.
- [86] Kirian, R. A.; Wang, X.; Weierstall, U.; Schmidt, K. E.; Spence, J. C.; Hunter, M.; Fromme, P.; White, T.; Chapman, H. N.; Holton, J. *Optics express* **2010**, *18*, 5713–5723.

- [87] White, T. A.; Barty, A.; Stellato, F.; Holton, J. M.; Kirian, R. A.; Zatsepin, N. A.; Chapman, H. N. *Biological Crystallography* **2013**, *69*, 1231–1240.
- [88] Kirian, R. A.; White, T. A.; Holton, J. M.; Chapman, H. N.; Fromme, P.; Barty, A.; Lomb, L.; Aquila, A.; Maia, F. R.; Martin, A. V.; et al. *Acta Crystallographica Section A: Foundations of Crystallography* **2011**, *67*, 131–140.
- [89] Scheres, S. H. *Journal of structural biology* **2012**, *180*, 519–530.
- [90] Kunimi, Y. *Journal of invertebrate pathology* **2007**, *95*, 181–186.
- [91] Goldsmith, M. R.; Shimada, T.; Abe, H. *Annu. Rev. Entomol.* **2005**, *50*, 71–100.
- [92] Ji, X.; Sutton, G.; Evans, G.; Axford, D.; Owen, R.; Stuart, D. I. *The EMBO journal* **2010**, *29*, 505–514.
- [93] Axford, D.; Ji, X.; Stuart, D. I.; Sutton, G. *Acta Crystallographica Section D: Biological Crystallography* **2014**, *70*, 1435–1441.
- [94] Ginn, H. M.; Messerschmidt, M.; Ji, X.; Zhang, H.; Axford, D.; Gildea, R. J.; Winter, G.; Brewster, A. S.; Hattne, J.; Wagner, A.; et al. *Nature communications* **2015**, *6*.
- [95] Poinar, G.; Poinar, R. *Journal of invertebrate pathology* **2005**, *89*, 243–250.
- [96] Anduleit, K.; Sutton, G.; Diprose, J. M.; Mertens, P. P.; Grimes, J. M.; Stuart, D. I. *Protein science* **2005**, *14*, 2741–2743.
- [97] Rohrmann, G. *Journal of General Virology* **1986**, *67*, 1499–1513.
- [98] Hunter-Fujita, F. R.; Entwistle, P. F.; Evans, H. F.; Crook, N. E.; et al. *Insect viruses and pest management.*; John Wiley & Sons Ltd, 1998.
- [99] Jehle, J. A. *Manual of Techniques in Invertebrate Pathology* **2012**, 15.
- [100] McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Journal of applied crystallography* **2007**, *40*, 658–674.
- [101] Seibert, M. M.; Ekeberg, T.; Maia, F. R.; Svenda, M.; Andreasson, J.; Jönsson, O.; Odić, D.; Iwan, B.; Rocker, A.; Westphal, D.; et al. *Nature* **2011**, *470*, 78–81.
- [102] Ayyer, K.; Philipp, H. T.; Tate, M. W.; Wierman, J. L.; Elser, V.; Gruner, S. M. *IUCrJ* **2015**, *2*, 29–34.

- [103] Wilson, A. *Acta Crystallographica* **1949**, *2*, 318–321.
- [104] Chandra, N.; Acharya, K. R.; Moody, P. *Acta Crystallographica Section D: Biological Crystallography* **1999**, *55*, 1750–1758.
- [105] Tenboer, J.; Basu, S.; Zatsopin, N.; Pande, K.; Milathianaki, D.; Frank, M.; Hunter, M.; Boutet, S.; Williams, G. J.; Koglin, J. E.; et al. *Science* **2014**, *346*, 1242–1246.
- [106] Gati, C.; Bourenkov, G.; Klinge, M.; Rehders, D.; Stellato, F.; Oberthür, D.; Yefanov, O.; Sommer, B. P.; Mogk, S.; Duszynski, M.; et al. *IUCrJ* **2014**, *1*, 87–94.
- [107] Crowther, R. *Philosophical Transactions of the Royal Society B: Biological Sciences* **1971**, *261*, 221–230.
- [108] Barends, T. R.; Foucar, L.; Botha, S.; Doak, R. B.; Shoeman, R. L.; Nass, K.; Koglin, J. E.; Williams, G. J.; Boutet, S.; Messerschmidt, M.; et al. *Nature* **2014**, *505*, 244–247.
- [109] Henderson, R. *Proceedings of the National Academy of Sciences* **2013**, *110*, 18037–18041.
- [110] Giordano, R.; Leal, R. M.; Bourenkov, G. P.; McSweeney, S.; Popov, A. N. *Acta Crystallographica Section D: Biological Crystallography* **2012**, *68*, 649–658.
- [111] Zeldin, O. B.; Brewster, A. S.; Hattne, J.; Uervirojnangkoorn, M.; Lyubimov, A. Y.; Zhou, Q.; Zhao, M.; Weis, W. I.; Sauter, N. K.; Brunger, A. T. *Acta Crystallographica Section D: Biological Crystallography* **2015**, *71*, 0–0.
- [112] Yildirim, M. A.; Goh, K.-I.; Cusick, M. E.; Barabási, A.-L.; Vidal, M. *Nature biotechnology* **2007**, *25*, 1119–1126.
- [113] Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. *Nature reviews Drug discovery* **2006**, *5*, 993–996.
- [114] Rohrer, D. K.; Kobilka, B. K. *Physiological reviews* **1998**, *78*, 35–52.
- [115] Rasmussen, S. G.; Choi, H.-J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R.; Sanishvili, R.; Fischetti, R. F.; et al. *Nature* **2007**, *450*, 383–387.

- [116] Rasmussen, S. G.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; et al. *Nature* **2011**, *477*, 549–555.
- [117] Wettschureck, N.; Offermanns, S. *Physiological reviews* **2005**, *85*, 1159–1204.
- [118] Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; et al. *science* **2000**, *289*, 739–745.
- [119] Qin, K.; Dong, C.; Wu, G.; Lambert, N. A. *Nature chemical biology* **2011**, *7*, 740–747.
- [120] Whalen, E. J.; Rajagopal, S.; Lefkowitz, R. J. *Trends in molecular medicine* **2011**, *17*, 126–139.
- [121] Wisler, J. W.; Xiao, K.; Thomsen, A. R.; Lefkowitz, R. J. *Current opinion in cell biology* **2014**, *27*, 18–24.
- [122] Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; et al. *science* **2007**, *318*, 1258–1265.
- [123] Chun, E.; Thompson, A. A.; Liu, W.; Roth, C. B.; Griffith, M. T.; Katritch, V.; Kunken, J.; Xu, F.; Cherezov, V.; Hanson, M. A.; et al. *Structure* **2012**, *20*, 967–976.
- [124] Mendis, S.; Puska, P.; Norrving, B.; et al. *Global atlas on cardiovascular disease prevention and control*; World Health Organization, 2011.
- [125] Cusack, S.; Belrhali, H.; Bram, A.; Burghammer, M.; Perrakis, A.; Riekel, C. *Nature Structural & Molecular Biology* **1998**, *5*, 634–637.
- [126] Smith, J. L.; Fischetti, R. F.; Yamamoto, M. *Current opinion in structural biology* **2012**, *22*, 602–612.
- [127] Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. *Acta Crystallographica Section D: Biological Crystallography* **2010**, *66*, 213–221.
- [128] Juers, D. H.; Matthews, B. W. *Journal of molecular biology* **2001**, *311*, 851–862.

- [129] Halle, B. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 4793–4798.
- [130] Dunlop, K. V.; Irvin, R. T.; Hazes, B. *Acta Crystallographica Section D: Biological Crystallography* **2005**, *61*, 80–87.
- [131] Wang, C.; Jiang, Y.; Ma, J.; Wu, H.; Wacker, D.; Katritch, V.; Han, G. W.; Liu, W.; Huang, X.-P.; Vardy, E.; et al. *Science* **2013**, *340*, 610–614.
- [132] Rueda, M.; Katritch, V.; Raush, E.; Abagyan, R. *Bioinformatics* **2010**, *26*, 2784–2785.
- [133] Wacker, D.; Wang, C.; Katritch, V.; Han, G. W.; Huang, X.-P.; Vardy, E.; McCorvy, J. D.; Jiang, Y.; Chu, M.; Siu, F. Y.; et al. *Science* **2013**, *340*, 615–619.
- [134] Zaman, M. A.; Oparil, S.; Calhoun, D. A. *Nature Reviews Drug Discovery* **2002**, *1*, 621–636.
- [135] Unal, H.; Jagannathan, R.; Karnik, S. S. In *Biochemical Roles of Eukaryotic Cell Surface Macromolecules*; Springer, 2012; pp 187–199.
- [136] Xia, Y.; Kellems, R. E. *Circulation research* **2013**, *113*, 78–87.
- [137] Oliveira, L.; Costa-Neto, C. M.; Nakaie, C. R.; Schreier, S.; Shimuta, S. I.; Paiva, A. C. *Physiological reviews* **2007**, *87*, 565–592.
- [138] Warren, A. J.; Armour, W.; Axford, D.; Basham, M.; Connolley, T.; Hall, D. R.; Horrell, S.; McAuley, K. E.; Mykhaylyk, V.; Wagner, A.; et al. *Biological Crystallography* **2013**, *69*, 1252–1259.
- [139] Brodersen, D.; Clemons, W.; Carter, A.; Wimberly, B.; Ramakrishnan, V. *Acta Crystallographica Section D: Biological Crystallography* **2003**, *59*, 2044–2050.
- [140] Stellato, F.; Oberthur, D.; Liang, M.; Bean, R.; Gati, C.; Yefanov, O.; Barty, A.; Burkhardt, A.; Fischer, P.; Galli, L.; et al. *IUCrJ* **2014**, *1*, 0–0.
- [141] Fairlamb, A. H. *Trends in parasitology* **2003**, *19*, 488–494.
- [142] Bryant, C.; Kerr, I. D.; Debnath, M.; Ang, K. K.; Ratnam, J.; Ferreira, R. S.; Jaishankar, P.; Zhao, D.; Arkin, M. R.; McKerrow, J. H.; et al. *Bioorganic & medicinal chemistry letters* **2009**, *19*, 6218–6221.

- [143] Abdulla, M.-H.; O'Brien, T.; Mackey, Z. B.; Sajid, M.; Grab, D. J.; McKerrow, J. H. *PLoS neglected tropical diseases* **2008**, *2*, e298.
- [144] Lecaille, F.; Kaleta, J.; Brömme, D. *Chemical reviews* **2002**, *102*, 4459–4488.
- [145] Koopmann, R.; Cupelli, K.; Redecke, L.; Nass, K.; DePonte, D. P.; White, T. A.; Stellato, F.; Rehders, D.; Liang, M.; Andreasson, J.; et al. *Nature methods* **2012**, *9*, 259–262.
- [146] Ijiri, H.; Coulibaly, F.; Nishimura, G.; Nakai, D.; Chiu, E.; Takenaka, C.; Ikeda, K.; Nakazawa, H.; Hamada, N.; Kotani, E.; et al. *Biomaterials* **2009**, *30*, 4297–4308.
- [147] Fan, G.; Maldonado, F.; Zhang, Y.; Kincaid, R.; Ellisman, M. H.; Gastinel, L. N. *Microscopy research and technique* **1996**, *34*, 77–86.
- [148] Paithankar, K. S.; Garman, E. F. *Acta Crystallographica Section D: Biological Crystallography* **2010**, *66*, 381–388.
- [149] Bernstein, H.; Hammersley, A. In *International Tables for Crystallography Volume G: Definition and exchange of crystallographic data*; Springer, 2005; pp 37–43.
- [150] Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallographica Section D: Biological Crystallography* **2011**, *67*, 355–367.
- [151] Wierman, J. L.; Alden, J. S.; Kim, C. U.; McEuen, P. L.; Gruner, S. M. *Journal of applied crystallography* **2013**, *46*, 1501–1507.
- [152] Zarrine-Afsar, A.; Barends, T. R.; Müller, C.; Fuchs, M. R.; Lomb, L.; Schlichting, I.; Miller, R. D. *Acta Crystallographica Section D: Biological Crystallography* **2012**, *68*, 321–323.
- [153] Coquelle, N.; Brewster, A. S.; Kapp, U.; Shilova, A.; Weinhausen, B.; Burghammer, M.; Colletier, J.-P. *Biological Crystallography* **2015**, *71*.
- [154] Falkner, J. C.; Al-Somali, A. M.; Jamison, J. A.; Zhang, J.; Adrianse, S. L.; Simpson, R. L.; Calabretta, M. K.; Radding, W.; Phillips, G. N.; Colvin, V. L. *Chemistry of materials* **2005**, *17*, 2679–2686.
- [155] Paithankar, K. S.; Owen, R. L.; Garman, E. F. *Journal of synchrotron radiation* **2009**, *16*, 152–162.

- [156] Terwilliger, T. C. *Acta Crystallographica Section D: Biological Crystallography* **2004**, *60*, 2144–2149.
- [157] Kupitz, C.; Basu, S.; Grotjohann, I.; Fromme, R.; Zatsepin, N. A.; Rendek, K. N.; Hunter, M. S.; Shoeman, R. L.; White, T. A.; Wang, D.; et al. *Nature* **2014**, *513*, 261–265.
- [158] Wang, D.; Weierstall, U.; Pollack, L.; Spence, J. *Journal of synchrotron radiation* **2014**, *21*, 0–0.
- [159] Nogly, P.; James, D.; Wang, D.; White, T. A.; Zatsepin, N.; Shilova, A.; Nelson, G.; Liu, H.; Johansson, L.; Heymann, M.; et al. *IUCrJ* **2015**, *2*, 0–0.
- [160] Perry, S. L.; Guha, S.; Pawate, A. S.; Henning, R.; Kosheleva, I.; Srajer, V.; Kenis, P. J.; Ren, Z. *Journal of applied crystallography* **2014**, *47*, 1975–1982.
- [161] Johnson, I.; Bergamaschi, A.; Billich, H.; Cartier, S.; Dinapoli, R.; Greiffenberg, D.; Guizar-Sicairos, M.; Henrich, B.; Jungmann, J.; Mezza, D.; et al. *Journal of Instrumentation* **2014**, *9*, C05032.
- [162] Kabsch, W. *Acta Crystallographica Section D: Biological Crystallography* **2014**, *70*, 2204–2216.
- [163] Botha, S.; Nass, K.; Barends, T. R.; Kabsch, W.; Latz, B.; Dworkowski, F.; Foucar, L.; Panepucci, E.; Wang, M.; Shoeman, R.; et al. *Acta Crystallographica Section D: Biological Crystallography* **2015**, *71*, 0–0.
- [164] Sierra, R. G.; Laksmono, H.; Kern, J.; Tran, R.; Hattne, J.; Alonso-Mori, R.; Lassalle-Kaiser, B.; Glöckner, C.; Hellmich, J.; Schafer, D. W.; et al. *Acta Crystallographica Section D: Biological Crystallography* **2012**, *68*, 1584–1587.
- [165] Roedig, P.; Vartiainen, I.; Duman, R.; Panneerselvam, S.; Stübe, N.; Lorbeer, O.; Warmer, M.; Sutton, G.; Stuart, D.; Weckert, E.; et al. *Scientific Reports* **2015**, *5*.
- [166] Feld, G. K.; Heymann, M.; Benner, W. H.; Pardini, T.; Tsai, C.-J.; Boutet, S.; Coleman, M. A.; Hunter, M. S.; Li, X.; Messerschmidt, M.; et al. *Journal of Applied Crystallography* **2015**, *48*, 0–0.
- [167] Mueller, C.; Marx, A.; Epp, S.; Zhong, Y.; Kuo, A.; Balo, A.; Soman, J.; Schotte, F.; Lemke, H.; Owen, R.; et al. *Structural Dynamics* **2015**, *2*, 054302.
- [168] Caleman, C.; Timneanu, N.; Martin, A. V.; Jönsson, H. O.; Aquila, A.; Barty, A.; Scott, H. A.; White, T. A.; Chapman, H. N. *Optics express* **2015**, *23*, 1213–1231.

- [169] Chapman, H. N.; Caleman, C.; Timneanu, N. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2014**, *369*, 20130313.
- [170] Son, S.-K.; Young, L.; Santra, R.; et al. *Physical Review A* **2011**, *83*, 033402.