



HAL
open science

Powder diffraction studies of proteins

Yves Watier

► **To cite this version:**

Yves Watier. Powder diffraction studies of proteins. Structural Biology [q-bio.BM]. Université de Grenoble, 2011. English. NNT: 2011GRENV009 . tel-00605443

HAL Id: tel-00605443

<https://theses.hal.science/tel-00605443>

Submitted on 1 Jul 2011

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Yves Watier

Thèse dirigée par « **Andrew N. Fitch** »

préparée au sein de l'**European Synchrotron Radiation Facility**
dans l'**École Doctorale Chimie et Sciences du Vivant**

Etudes de diffraction de poudres de protéines.

Thèse soutenue publiquement le **19 Avril 2011**,
devant le jury composé de :

Mr Serge PEREZ

Professeur, European Synchrotron Radiation Facility , Président.

Mr John HELLIWELL

Professeur, University of Manchester, Rapporteur.

Mr Marc SCHILTZ

Professeur, École polytechnique fédérale de Lausanne, Rapporteur.

Mr Andrew N. FITCH

Professeur, European Synchrotron Radiation Facility, Membre.



à mes parents et à Céline.

Acknowledgements :

I would like first to thank my direct team, Andy Fitch, Irène Margiolaki and Jon Wright for all the inestimable help, all the advice and ideas during the last three years.

I would like to thank the three students from the University of Bath who worked hard with me, Sotonye Dagogo, Ines Collings and Lucy Saunders.

The ID31, ID11, ID14 and BM16 teams for their support and help during experiments.

John Helliwell, Serge Pérez and Marc Schiltz for accepting to be part of the jury of this thesis.

Marion Giffard, Richard Kahn, Françoise Bonneté, Mohamed El-Hajji and Sanofi-Aventis for their collaboration on the polymorphism of urate oxidase.

Mathias Norrman, Gerd Schluckebier and Novo Nordisk for the insulin samples.

Elaine Chiu and Peter Metcalf for their collaboration on the baculovirus project.

Nicolas Papageorgiou, Bruno Coutard, Violaine Lantéz and Ernest Gould for their collaboration on the macro domain of the Mayaro virus project.

Finally, I want to thank the ESRF for financial support and the provision of synchrotron beam time.

Abstract:

Knowledge of the structure of proteins helps in understanding their biological function. The main technique used, single crystal diffraction, requires the limiting step of growing a single crystal. On the other hand, powder diffraction requires only a crystalline precipitate made of many microcrystals such as those often discarded during the search for suitable single crystal growth conditions. We present in this thesis different studies of proteins by powder diffraction. Development of new methods for sample preparation and data acquisition are also presented, as they have been crucial steps to obtain high quality diffraction data.

We studied the polymorphism of Urate oxidase by observing the different crystallographic phases resulting from the changes in the crystallisation conditions. A crystallographic phase of pharmaceutical interest has been identified.

Also one phase of urate oxidase complexed with its inhibitor gave a powder pattern sufficient to re-determine and refine its structure.

A protocol for cryocooling protein powder samples has been found, extending the lifespan of the sample in the intense X-ray beam. This allowed the refinement from powder data of two forms of cryocooled human insulin.

We present also the determination of a preliminary structure of the mayaro virus macro domain, based on a powder diffraction pattern obtained on a single urchin-like bundle of needles.

A study of the protective protein matrix of two baculoviruses is presented, showing some current limits of the method.

In annexes, the steps for preparing and analysing protein powders are described.

Résumé de la thèse :

La technique de diffraction de monocristaux est la plus utilisée pour déterminer une structure tridimensionnelle de protéine, permettant ainsi la compréhension de sa fonction biologique. Cependant, cette technique nécessite l'obtention d'un monocristal. La détermination d'une structure par diffraction de poudre ne requiert que l'obtention d'un précipité cristallin, souvent obtenu lors de la recherche d'une condition de cristallisation.

Nous présentons dans cette thèse plusieurs protéines étudiées par diffraction de poudre. Le développement de nouvelles méthodes pour la préparation des échantillons et l'acquisition des données sont présentées, étant donnée leur importance cruciale dans ces études.

Nous avons étudié le polymorphisme de l'urate oxidase par la détermination des différentes phases cristallines résultant des modifications des conditions de cristallisation. Cette étude a débouché sur l'identification d'une forme cristalline nouvelle d'intérêt pharmaceutique.

Une des phases d'urate oxidase a permis l'obtention d'un cliché de diffraction de qualité suffisante à la redétermination et l'affinement de sa structure.

Un protocole pour le refroidissement cryogénique de poudre de protéine est présenté, offrant une durée de vie accrue de l'échantillon lors de l'acquisition de données. Ce protocole a permis l'affinement de deux structures d'insuline humaine.

Nous présentons également la détermination d'une structure préliminaire du domaine macro du virus Mayaro, basé uniquement sur des données acquises sur un unique échantillon polycristallin en forme d'oursin. Une étude de la matrice de protection de deux baculovirus illustre les limites auxquelles se heurte actuellement la technique de diffraction de poudre de protéines.

En annexes, les étapes nécessaires pour la préparation et l'analyse des données sont présentées.

Contents

1	Introduction	9
2	Instrumentation and theory	17
2.1	Instrumentation	18
2.1.1	Synchrotron radiation sources	18
2.1.2	ID31	19
2.1.3	ID11	21
2.1.4	ID14	22
2.1.5	The choice of the beamline for protein powder diffraction	23
2.2	Crystallography - Theory	25
2.2.1	Definitions	25
2.2.2	Diffraction	25
2.2.3	The phase problem	28
2.2.4	Rietveld refinement	33
2.3	Crystallisation	36
2.3.1	Crystallisation	36
2.3.2	Cryocooling role	38
3	Polymorphism of urate oxidase	41
3.1	Introduction	42
3.1.1	Function of the enzyme	42
3.1.2	Signification of the absence of Uox in primates	43
3.1.3	Drawbacks of the absence of Uox in humans	44
3.1.4	Pharmaceutical interest	44
3.2	Materials and methods	47
3.2.1	Crystallisation	47
3.2.2	Sample preparation	47
3.3	Results - urate oxidase polymorphism	49
3.3.1	Summary	49
3.3.2	A new way of obtaining an orthorhombic ($P2_12_12$) phase	52
3.3.3	Trigonal ($P3_121$) phase	55

3.3.4	Three monoclinic ($P2_1$) phases	59
3.3.5	Orthorhombic ($I222$) phase	62
3.3.6	Polymorphism of Urate oxidase - Conclusion	63
4	Structure of Urate oxidase	65
4.1	Integrated intensity extraction	66
4.2	Molecular replacement	67
4.3	Refinement on extracted intensities	71
4.4	Rietveld refinement	71
4.5	Conclusion	75
5	The Macro domain of Mayaro Virus	77
5.1	Introduction	78
5.1.1	The Mayaro virus	78
5.1.2	The macro domain	80
5.1.3	Aim of the study	82
5.2	Results	83
5.2.1	Sample preparation	83
5.2.2	First experiment, on ID31	86
5.2.3	Second experiment, on ID11	88
5.2.4	Third experiment, on ID14-eh1	90
5.3	Conclusion	103
6	Cryocooling of protein powders	105
6.1	Introduction	106
6.1.1	History of cryocooling of protein crystals.	106
6.1.2	Physical explanation of radiation damage.	107
6.1.3	Main challenges for protein powder diffraction	108
6.2	Results	109
6.2.1	Why protein powder diffraction needs cryocooling?	109
6.2.2	Sample holder for cryocooling	112
6.3	Structural refinement of cryocooled human insulin.	125
6.3.1	Insulin microcrystals	125
6.3.2	Rietveld refinement	126
6.4	Conclusion	129
7	AcMNPV and CpGV	131
7.1	Introduction	132
7.1.1	Baculoviruses : a short history	133
7.1.2	Baculovirus life cycle	135
7.1.3	Structure of Baculovirus	135

7.1.4	Aim of the protein powder diffraction studies	136
7.2	Results	138
7.2.1	Lattice parameter comparison	138
7.2.2	Rietveld fit	141
7.2.3	Structural variation	153
7.3	Discussion	158
8	Conclusions and further work	161
A	Sample preparation for room temperature experiments	167
A.1	Glass capillary	167
A.2	Silson membrane	169
B	Extraction of the microcrystals from the mother liquor	175
B.1	Electrophoresis	176
B.2	With heavy oil	176
C	Humidity variation experiments	181
C.1	Protocol	183
C.2	Test case : Glucose isomerase	184
C.2.1	Experiment	184
C.2.2	Results obtained on glucose isomerase	187
D	Xray data reduction	189
D.1	ID31	189
D.1.1	Refining the offsets between the detectors	190
D.1.2	Summing with id31sumall and id31sum	191
D.1.3	Clustering of multiple scans	194
D.2	On 2D detectors	195
D.2.1	Fable - Imageviewer	196
D.2.2	FIT2D	197
D.2.3	fit2dcake on Frelon camera	202
D.2.4	fit2dcake for ID14	204
E	Background subtraction	207
E.1	Background subtraction	208
E.2	Quality evaluation	211
F	Topas	215
F.1	Introduction	215
F.2	Indexing	216
F.3	Indexing without the graphical interface	217

F.4	Pawley fit	218
F.5	Pattern calculation with protein structures	221
G	Extracting intensities with PRODD	231
G.1	Description of the .ccl command file	232
G.1.1	Comments	234
G.1.2	Cycles and scales	234
G.1.3	Datasets and background	234
G.1.4	Phase information	235
G.1.5	Refining parameters	236
G.2	Executing the software	237
G.3	Plotting the results	238
H	Rietveld refinement with GSAS	241

1

Introduction

Au cours de ces dernières années, la biologie structurale a permis de comprendre plus en détail le fonctionnement de la vie. Cette évolution des connaissances a été possible notamment par la détermination à l'échelle de l'atome de la structure d'un certain nombre de protéines.

La technique la plus utilisée pour déterminer la structure de protéines est la diffraction aux rayons X de monocristaux. Cette technique est très puissante et permet d'obtenir des modèles de la structure de la protéine observée de très haute qualité, cependant, elle requiert l'obtention préalable d'un monocristal.

Obtenir un monocristal d'une protéine est très délicat et nécessite beaucoup d'efforts, tant humain que matériel. Lors du processus de recherche de conditions de cristallisation adéquates pour un monocristal, un grand nombre de précipités cristallins sont obtenus. Ces précipités cristallins sont constitués d'un grand nombre de microcristaux. Pris un par un pour une expérience de diffraction de monocristal, ces cristaux ne diffracteraient pas suffisamment pour permettre l'obtention d'un jeu de données de bonne qualité. Mais l'ensemble du précipité, composé d'un grand nombre de cristaux, est suffisant pour la mesure d'un signal de diffraction. Étant donné que ces cristaux sont tous orientés différemment et de manière aléatoire, le cliché de diffraction obtenu, constitué d'anneaux concentriques, est typique d'une expérience de diffraction de poudre.

Depuis l'avènement d'instruments à haute résolution couplés à de puissantes sources de rayonnement synchrotron, la technique de diffraction de poudre a progressé énormément ces dernières années. La taille des systèmes étudiés par diffrac-

tion de poudre n'a cessé d'augmenter, jusqu'à ce qu'il soit envisagé de l'appliquer aux protéines. Bien que certaines protéines aient été analysées par diffraction de poudre au cours du siècle dernier, ce n'est qu'en 1999 que R.B. Von Dreele a réussi à affiner la structure d'une protéine uniquement à partir de données de diffraction de poudre.

La technique de diffraction de poudre appliquée aux protéines possède de nombreuses débouchées. Depuis son utilisation comme outil qualitatif lors de la production d'une protéine sous forme cristalline par des industriels ou des chercheurs en biologie structurale, jusqu'à son utilisation en technique complémentaire pour la résolution d'une structure, lorsqu'un monocristal n'a pu être obtenu.

L'intention de cette thèse est de montrer les diverses possibilités de cette technique complémentaire aux cristallographes en biologie, en leur proposant un ensemble de projets dans lequel cette technique a été utilisée. En annexe nous expliquons étapes par étapes les méthodes qui ont été utilisées et développées au cours de cette thèse depuis la préparation des échantillons jusque l'analyse des données collectées.

Over the past decades, structural biology has opened new horizons in understanding the mechanisms of life.

Knowing the structure at the atomic scale of the constituents of life is very important to understand and predict the behaviour of each component.

The most frequently and the most powerful technique to obtain the atomic scale structure of proteins is single crystal X-ray diffraction. Proteins targeted by bioinformatic methods tend to be those that are likely to be expressed as soluble stable species in recombinant systems. Even considering this bias, only 67% of the selected targets can be expressed, less than 25% of those ever produce soluble protein, and less than 3% result in a crystal structure.

One of the main difficulties in protein crystallography is to obtain a diffracting single crystal of suitable size and quality. This means also that the inability to obtain a single crystal is the major cause of failure for 22% of the interesting proteins selected by bioinformatics¹.

The structural coverage of the proteomes may remain in the end incomplete if the only method used is single crystal diffraction alone. When this single crystal cannot be obtained, structural biologists tend to focus on other

1. Swiss Institute of Bioinformatics <http://www.isb-sib.ch/>

structural techniques. Among others, we can cite nuclear magnetic resonance (NMR) which by avoiding passing through a solid state gives information on the atomic scale but also very useful information about the internal motions of a protein. Unfortunately, NMR has a size limit for the protein studied (<40kDa). Small angle X-ray scattering (SAXS) has no limitation for the size of the protein and needs only a pure solution, but can obtain only the number and shape of the ordered domains of the studied protein without detail of the atoms' positions.

Electron microscopy is a good candidate also for obtaining the external envelope of the protein but reveals no details of the internal structure.

However during the process of obtaining a single crystal, which often takes a long time, numerous by-products are discarded, among them, crystalline precipitates are containing a large amount of information.

Crystalline precipitates consist of a large number of very small microcrystals. These microcrystals are of exactly the same nature as a single crystal, except in size. A single crystal needs to be big enough to obtain a recordable diffracting signal. One microcrystal alone taken out of the precipitate would not have the diffracting power suitable for a single crystal experiment. However the precipitate taken as a whole, containing thousands of microcrystals, will definitely have enough diffracting power to be recorded. In this case, as multiple crystals are recorded at the same time, it can be seen as an X-ray powder diffraction experiment. Compared to a single crystal, it is not possible to orient all the microcrystals in the same orientation, they are, most of the time, randomly orientated. The resulting pattern consists of rings, typical for powders, which are the addition of all the intensities of the reflections having the same d-spacing.

As protein powders are often quickly available during the search for a single crystal, powder diffraction can be used in parallel during this search to help rationalize the crystallisation process. In the cases where no single crystal can be found, but precipitates are observed, powder diffraction should be the nearest alternative.

The method of powder diffraction was developed in 1916 by Debye and Scherrer. For more than half a century, this technique has been only used for qualitative phase analysis, semi-quantitative phase analysis and also stress measurements in materials.

Historically this delay in using the technique for structure determination was due to the intrinsic difficulty of powder diffraction. This difficulty is mainly due to the accidental or systematic overlap of reflections. As three dimen-

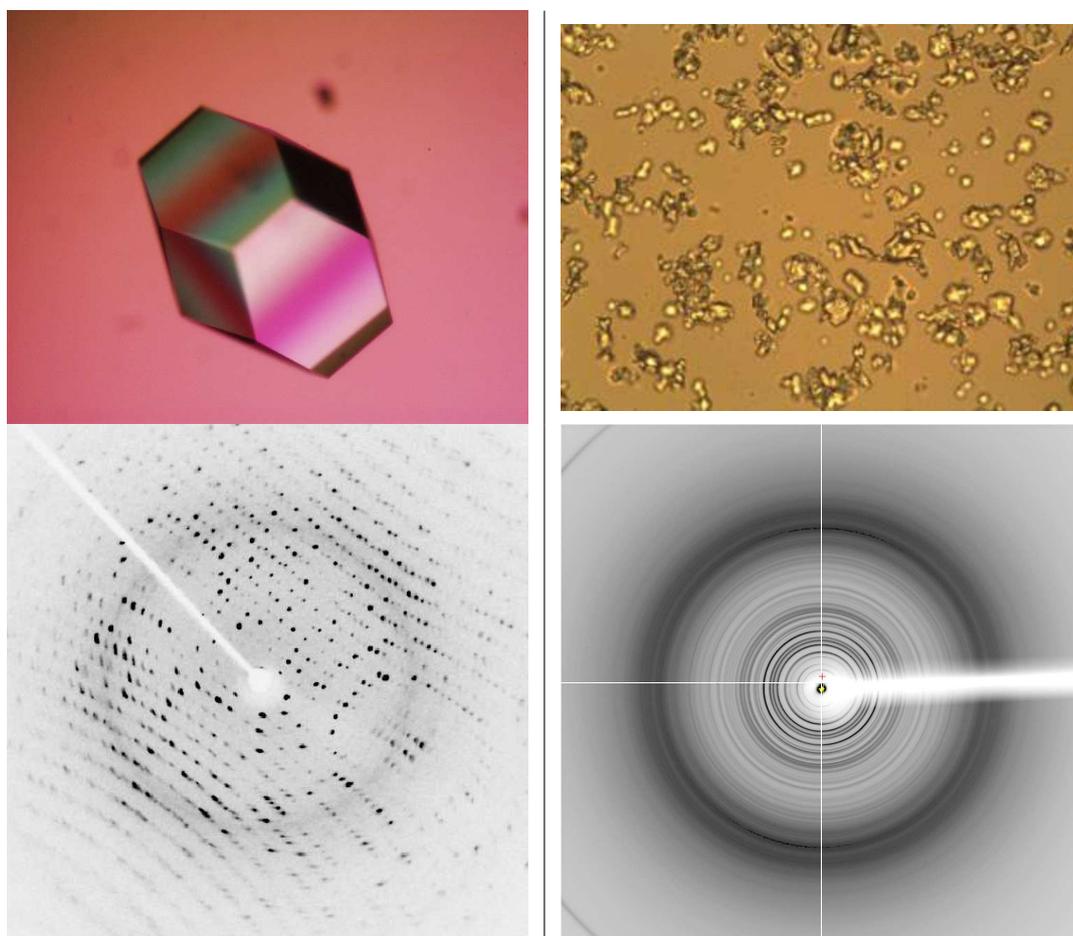


Figure 1.1: Top left is a single crystal (Photographed by Mathias Klode), top right is a crystalline protein precipitate. Bottom left are the resulting X-ray image of one single crystal at one orientation, bottom right is the rings observed when a protein powder is diffracting.

sional reciprocal space is projected onto a one dimensional space, the 2θ axis, the reconstruction of a three dimensional structure is more difficult.

The improvement in instrumentation and X-ray sources combined with the evolution of modern computers and software has allowed the use of powder diffraction for structure solution in most of modern chemistry.

The important evolutionary step in this technique was the quantitative method developed by Rietveld in 1969[1]. With this method, the whole powder pattern is used during the analysis and not only the non overlapped

reflections (which would be very limited for a large system). This necessitates the modelling of the peak shape, background parameters, instrumental specific parameters, lattice parameters together with the structure itself. All the parameters are refined using a least squares approach in order to minimise differences between the modelled and the observed powder pattern. While this method minimises the impact of having overlapped peaks, the need for a good initial guess of the structure is essential, as only small adjustments will be refined.

Simple analysis of powder diffraction on proteins started a long time ago. The first observations of protein powder diffraction came from Corey and Wyckoff in 1936 [2], where it was stated that the peaks observed between 80 and 3\AA were compatible with those from proteins.

Then Bernal and Fankuchen in 1941 [3] report the first unit cell determination from plant virus proteins.

Two large forms of tubulin were observed and a unit cell proposed by Amos in 1984 [4].

In the 1990s, powder diffraction evolved with successful analysis of bigger and bigger crystal structures.

But until the end of the 1990s, no protein structures has ever been refined using powder diffraction. This delay can be explained by a missing necessary step, which was the availability of suitable instruments for data acquisition. The appearance of high resolution powder diffractometers at synchrotron radiation sources opened the way for solving bigger structures.

In 1999, R.B. Von Dreele obtained the first protein crystal structure refined from high-resolution X-ray powder diffraction[5]. This was the first time that a Rietveld refinement was carried out on a protein structure. To realise this, a set of stereochemical restraints, as are normally used in low-resolution refinements with single-crystal data, were implemented in the software GSAS[6].

In 2000, the first determination of a novel form of Insulin produced by mechanical grinding was solved[7].

Multi-pattern refinement played a significant role in resolving the non systematic overlapping reflections. A multi-pattern refinement is performed using slight changes in lattice parameters between isomorphous samples. In 2004 the software PRODD[8] was modified to allow different lattice parameters for intensity extraction. In 2005, a Rietveld refinement of turkey egg white lysozyme was carried out based on ID31 high resolution powder diffraction data[9].

Anisotropic lattice strains were used advantageously for solving the SH3 domain of ponsin[10] in 2007.

In 2008 the anisotropic lattice strains were used also associated to the single isomorphous replacement method to generate phase information for structure factors up to 6 Å resolution sufficient to compute molecular envelopes of the protein molecule [11].

Many applications can be imagined for the study of proteins with the powder diffraction technique. Protein powder diffraction may be used as a quality evaluation tool when crystalline forms are produced, in research or in industrial processes. Protein diffraction can be imagined to be used as a routine lab technique to guide when crystallisation conditions screening is performed, in order to select rationally the most promising conditions. Also finally, solving new structures by the use of protein powder diffraction, while being long and delicate, may become the most interesting use of protein powder diffraction.

One of the limitations of the technique is the amount of protein sample required to do powder diffraction. Protein expression and protein purification required for diffraction studies take a lot of time. This man-time together with the need for major equipment have a big impact on the price of the research in structural biology. Producing more of this protein because the standard amount of sample in powder diffraction is higher than for a single crystal experiment is unacceptable for a mere screening and optimisation tool in most of the projects. Still as a structure solution tool, this amount of sample needed may become acceptable in some projects. Nonetheless, reducing the amount of sample needed will greatly help interest more crystallographers in trying powder diffraction.

These are the reasons why finding new protocols, keeping high quality information when only a small amount of sample is present, are needed. Also being able to perform simple screening with the by products of single crystal crystallisation condition screening would be a must, as the cost in man-time and money would be minimal.

The first aim of this thesis is to show the various possibilities offered by the technique for structural biologists. The second aim is to propose in the appendices a set of new protocols for sample preparation for protein powder diffraction. This I believe to be important as actually, there is no step by step descriptions of how to perform such tasks. Also in the appendix, a personal choice of the software needed today for the analysis of protein powder

patterns will be presented. Adaptation and modification of these software tools will be briefly discussed.

2

Instrumentation, theory and crystallisation

L'obtention de clichés de diffraction de haute qualité est une étape nécessaire et cruciale pour l'étude de protéines.

La première partie de chapitre est destiné à la description des instruments utilisés pour obtenir les données présentées dans cette thèse. Le rayonnement synchrotron et les diverses lignes de lumières utilisées à l'European Synchrotron Research Facility (ESRF - ID31, ID11, ID14) seront décrites. Ensuite je discuterai de leur avantages et inconvénients pour l'étude de diffraction de poudre de protéines.

La seconde partie de ce chapitre est consacrée aux bases théoriques nécessaire à la compréhension de la diffraction aux rayons-X.

La dernière partie de ce chapitre présente brièvement les techniques utilisées pour l'obtention de cristaux de protéines.

2.1 Instrumentation

2.1.1 Synchrotron radiation sources

While the theory of synchrotron radiation emitted by an electron in a curved trajectory was given by Liénard and Wiechert in 1898, the first practical observation was on the first synchrotron ever built in 1947 at the General Electric company laboratory[12].

Back in those days, particle accelerators were not aimed at producing X-rays, which were at first considered as a nuisance. Soon the interest of the emission of such a powerful light was evident and in the 1970s second generation synchrotrons dedicated to the use of the light emitted were built. In the late 1980s and early 1990s, the third generation of synchrotrons appeared.

In the last generation of synchrotrons, electrons are accelerated to relativistic speeds by a LINAC linear pre-injector accelerator. Then they are accelerated until they reach high enough energy in a booster synchrotron. Once the energy (6 GeV at the ESRF) is reached, they are injected into the storage ring where they will turn in a closed orbit inside a chamber under vacuum. A radio-frequency cavity system will periodically restore the energy lost by the emission of radiation.

In the storage ring, the electrons will be either in straight sections or turned by a bending magnet. While bending magnets emit usable light, the most intense light is obtained by the insertion devices (ID) in the straight portions. Insertion devices use wigglers or undulators to produce a sinusoidal magnetic field where the electrons will obtain a sinusoidal trajectory. The wigglers emit a broad spectrum from the infrared to X-rays, while an undulator spectrum consists of strong peaks two orders of magnitude higher in brilliance than a wiggler.

The optics of the beamline are adapted to the type of experiments done. Compared to an X-ray lab source, an X-ray beam on a synchrotron beamline will have a lower divergence and a higher intensity. For powder, the important fact is that it will result in higher angular resolution than with a conventional X-ray tube.

Each beamline has its own specifications. We used for the data presented in this thesis three beamlines, ID31, ID11 and ID14.

2.1.2 ID31

ID31 is the High Resolution Powder Diffraction beamline of the ESRF[13]. The main strengths of ID31 for protein powder diffraction came of the possibility to record very high angular resolution diffraction patterns. This results in the possibility to distinguish close but not totally overlapped reflections. This will allow high reciprocal space resolution, this is more important for powders than for single crystals. For the same maximal d-space resolution achieved by a sample, having higher angular resolution will increase the number of distinguishable reflections, allowing a more precise structure determination.

Powder diffraction requires a large beam able to illuminate a large volume of sample. This is why ID31 optics are kept simple, without any focusing optics. The X-rays are produced by three 11-mm-gap undulators, $1 \times u35$ and $2 \times u32$, which cover the energy range from 5 keV to 60 keV. The white beam from the source is incident to a cryogenically cooled monochromator composed of two Si(111) crystals.

Precision of the diffractometer is important, this is why ID31's is made of two heavy duty rotation tables. This mechanism is stable, accurate and reliable and has been used with great success continuously by users over the last 14 years.

To obtain high resolution powder patterns, the beamline uses a nine channel multianalyser detector. Each detector is preceded by a Si(111) analyser crystal, and are separated two degrees apart and are measuring in parallel the diffracted beam as a complete unit. Resulting measured diffracted peaks are very narrow, with an instrumental contribution to the FWHM of 0.003° , while positions are accurate and reproducible to a few tenths of a millidegree.

All detector scans are recorded independently and need to be summed together. This allows a fine summation required for analysing protein samples. As the nine detectors are collecting in parallel over 16 degrees, it is possible to collect more rapidly, which is better for samples with high radiation damage.

The main advantage of ID31 is the finest quality of the information collected, enough to perform structural refinement. The drawback is the amount of sample needed to get a recordable signal, and the minimum amount of time needed to collect one pattern. This is why for most of the protein projects, other beamlines use 2D detectors, requiring less sample and faster data ac-

quisition.

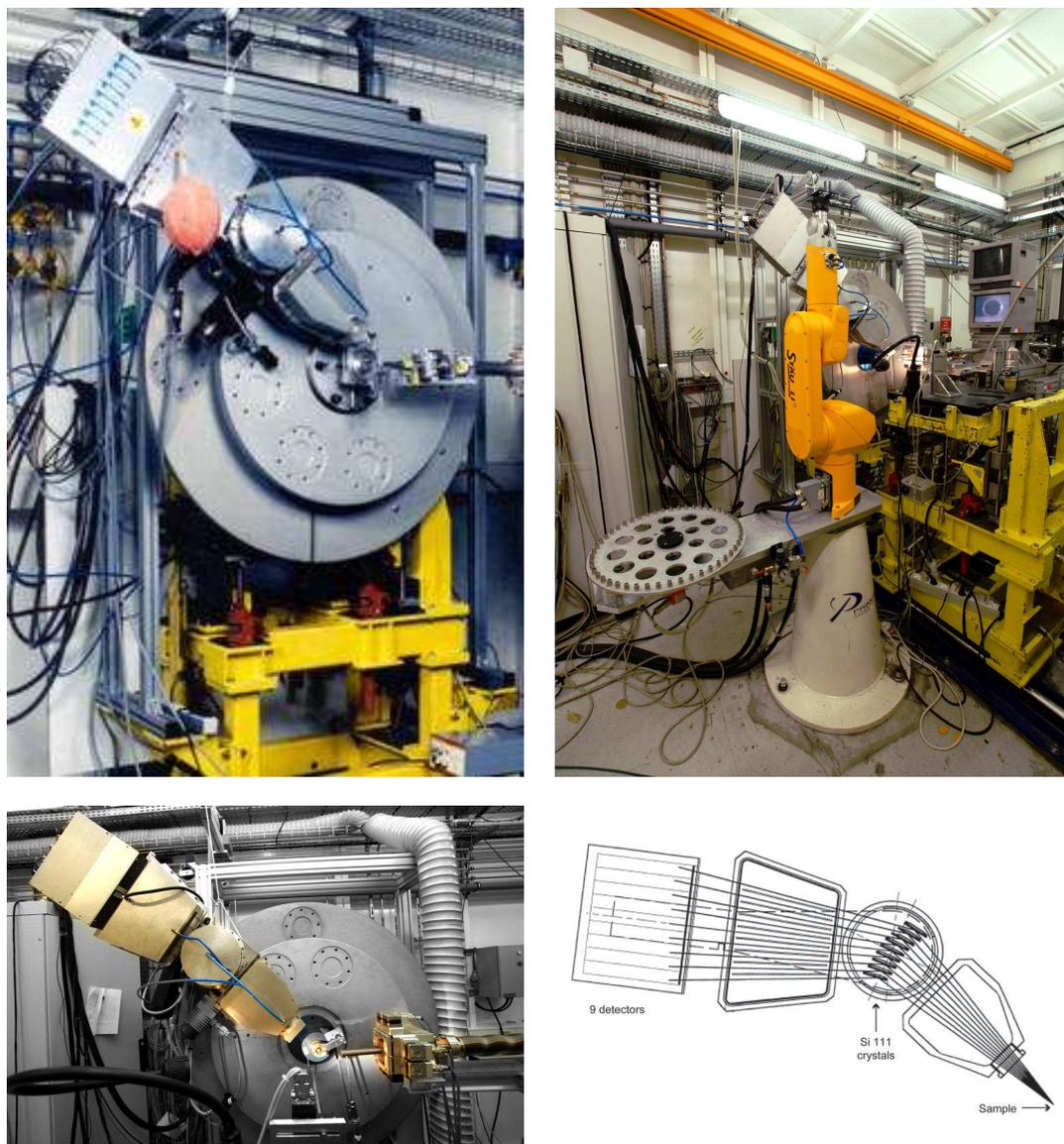


Figure 2.1: Views of ID31 experimental hutch. Top left, the two circle diffractometer when installed. Top right, the sample changer robot (orange) and the experimental table (yellow) in front of the diffractometer. Bottom left, the detector arm. Bottom right, schematic view of the multianalyser crystal principle.

2.1.3 ID11

ID11 is the Material Science beamline of the ESRF. The general name given to the beamline is representative of the broad variety of applications possible there.

ID11 is one of the newly renovated beamlines of the ESRF. It has a second separate experimental hutch outside the ring to ensure maximum focusing possibilities. Mechanical and thermal stability have been designed to allow sub-micron operation with a thermally isolated inner-hutch for the micro-focusing optics.

Two undulators (u23 and u22) under vacuum are used to produce the X-ray beam. A horizontal Laue-Laue monochromator, consisting of two Si(111) bent crystals allow an energy range between 25 and 140 keV. A new transfocator consisting of 7 groups of Al parabolic lenses allows the beam to be precisely collimated or focused.

ID11 has high flux and is capable of high energy while having the focusing optics necessary for small samples. We used the Frelon 2D CCD detector to collect powder rings. The Frelon camera is based on Kodak chips with 14-bit image depth with a resolution of 2048×2048 pixels, pixel size is $50 \mu\text{m}$. The use of a 2D detector is ideal when radiation damage is severe, as an image can be taken instantaneously so that the sample is not destroyed before the end of the acquisition. For protein powder experiments, ID11 is used when a precise beam is needed on samples in small quantities or very fragile. Drawbacks on using ID11 include the difficulty to perform an experiment. As the beamline is designed to perform many different experiments, there is no “ready to use” sample environment. Diffractometer, detector and sample need to be aligned manually before each experiment, and calibrations need to be redone for each movement of the detector.

The strengths of ID11 are the versatility of the possible experiments, perfectly compatible with conditions required for some difficult samples (weakness of the diffraction, size of sample, damage due to radiation). Unfortunately its use for screening experiments, like for example when large numbers of different crystal conditions need to be tested, or for quick experiments, is less suitable. Also, if enough sample is available, the use of a high resolution diffractometer will allow data collection with better angular resolution.

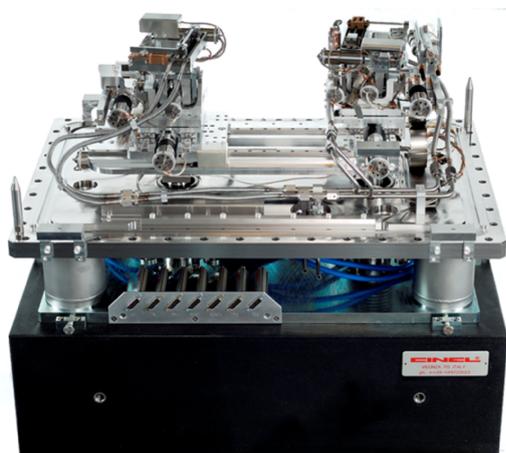


Figure 2.2: ID11 Laue/Laue monochromator

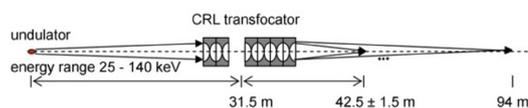


Figure 2.3: ID11 translocator allow to obtain a precise collimated or focused beam.

2.1.4 ID14

ID14 is a beamline specifically dedicated to macromolecules, ID14 is composed of four stations operating in parallel.

ID14-3 is a new dedicated Bio-SAXS (Small angle Xray Scattering) and the only non crystallographic station of ID14. ID14-1 and ID14-2 are fixed energy beamlines while ID14-4 is a tunable energy beamline between 9 and 14.5 keV dedicated to anomalous scattering experiments.

2D detectors depend on the station used. ID14-1 has a ADSC Q210 CCD detector (4096×4096 pixels), ID14-2 has a ADSC Q4 (2304×2304 pixels), ID14-4 has a ADSC Q315r detector (6144 x 6144). All these detectors have phosphor screen optimised for 1 Å wavelength.

All three crystallographic stations are equipped with automated sample environments. Screening a large number of samples is natural, as it is possible to use a robotic sample changer able to deal with already cryocooled mounted samples. Also the interface between humans and the beamline is very well conceived. The loading of a sample while the hutch is closed is automated, the centring of the sample in the beam performed in seconds and is painless due to the combination of high quality motorised goniometers and well conceived software. Most of the users coming to ID14 are able to use completely the beamline in a very short time without prior knowledge. Many small modifications have been done in order to improve the robustness of the

sample environment in order to deal with the large number of users coming there for only few hours.

Protein powder diffraction on this beamline is ideal to screen a large number of powder conditions. Unfortunately, it is to be noted that the strength of having a robust unique setup for single crystal experiment reduces versatility. For example, it is not possible to translate the detector perpendicular to the beam. This makes perfect sense for the experiments to which the beamline is dedicated, but would be interesting in a powder context where centring the beam in the corner of the detector would allow a better angular resolution with the same detector.

It is to be noted that the first time we used ID14 for a powder experiment was because of the availability of an exotic sample environment. This equipment, a humidifier device, is able to keep hydrated and change the humidity level of a sample in open air (figure C.1, p. 183).



Figure 2.4: ID14-1 experimental hutch



Figure 2.5: ID14-2 experimental hutch

2.1.5 The choice of the beamline for protein powder diffraction

A crucial step in a powder diffraction experiment is the choice of the beamline. Depending of the amount of protein powder sample available and the quality expected, we can sort the beamlines as following:

<i>Beam line</i>	<i>Detector type</i>	<i>Versatility</i>	<i>Amount of sample needed</i>	
ID14	Large 2D detector	- -	extremely small (one crystallisation drop precipitate)	Screening, initial phasing, medium angular resolution, fast acquisition, easy to use.
ID11	multi purpose 2D detector	+ + +	small to medium (several crystallisation drop precipitates)	Possibility to test and perform easily any kind of diffraction experiment, complicated to use.
ID31	High angular resolution diffractometer	-/+	large at room temperature, medium under cryocooled conditions (20-3 μ l of crystals packed)	Highly specialised for powder, the highest quality achievable for protein powder patterns. Ideal for structure solution.

The beamlines used have their own strengths and limitations for protein powder diffraction. ID14 would be ideal for screening in order to optimise the best crystallisation conditions, ideal also on very small and fragile samples, when centrifuging the sample in a capillary must be avoided. The powder pattern in this case can be of a sufficient quality for a first phasing and preliminary model building.

ID11 is more versatile and allows also to collect quickly on a 2D detector. More possibilities are open in ID11 hutches, but at the price of a more complicated experiment.

ID31, when enough sample is available, remains the best choice for collecting high angular resolution diffraction patterns of a quality necessary for accurate structure refinement.

2.2 Crystallography - Theory

Crystallography can be seen as a microscopy method, where the lenses have been replaced by a mathematical tool, the Fourier transform.

A microscope needs a light adapted to the size of the object that is under observation. For atoms, the light which has the adapted wavelength is in the X-ray range. Unfortunately, there is no lens able to reproduce an image from the X-rays scattered by an object. Only intensity and the direction of a scattered X-ray can be measured. Phases needed in order to obtain an image of an object are lost.

2.2.1 Definitions

A **crystal** consists of atoms periodically repeated in space. To describe a crystal 3 vectors are used (\vec{a} , \vec{b} , \vec{c}) defining one **unit cell**. Multiples of these vectors define a series of points called a **lattice**. The **primitive cell** is the smallest cell including only one point of the lattice.

The **asymmetric unit** is the smallest part of the unit cell which is enough to describe the whole cell by using symmetry elements.

The phenomenon of **diffraction** is due to the ondulatory property of light. When light is passing through a lattice, the light scattered will be affected by constructive or destructive interference.

When an incident beam of light is passing into a crystal, diffracted beams, called **reflections**, are produced. Each reflection has its own intensity and direction measurable, while its phase is not measurable.

The whole set of **directions of the reflections** obtainable from a crystal is depending only from the 3 main vectors \vec{a} , \vec{b} , \vec{c} , compared to the incident beam. Meaning that the directions give information of the crystal form.

The proportion between **the intensities of the reflections** is only due to density of the matter interacting with the X-rays, the electron density, inside the unit cell. Meaning that the intensities will give information on the electron density inside one cell.

2.2.2 Diffraction

A diffraction experiment can be described by the fundamental laws of Laue and Bragg.

Laue's law

Diffraction by a lattice is observed only in directions that match Laue's equations. With \vec{s}_0 the incident vector, with the direction the incident beam and length $|\vec{s}_0| = \frac{1}{\lambda}$, diffraction is observed along the diffuse vector \vec{s} , when the diffraction vector $S = s - s_0$ is obeying Laue's equations:

$$S \cdot \vec{a} = h \quad (2.1)$$

$$S \cdot \vec{b} = k \quad (2.2)$$

$$S \cdot \vec{c} = l \quad (2.3)$$

h, k, l , called Miller indices, are integers.

Inside a theoretical reciprocal lattice made by the base vector $\vec{a}^*, \vec{b}^*, \vec{c}^*$, Laue's law is transformed into :

$$S = h \vec{a}^* + k \vec{b}^* + l \vec{c}^* \quad (2.4)$$

The relation between the reciprocal lattice and the direct lattice is obtained by:

$$\vec{a}^* = \frac{\vec{b} \wedge \vec{c}}{\vec{a} \cdot \vec{b} \wedge \vec{c}} \quad (2.5)$$

$$\vec{b}^* = \frac{\vec{c} \wedge \vec{a}}{\vec{a} \cdot \vec{b} \wedge \vec{c}} \quad (2.6)$$

$$\vec{c}^* = \frac{\vec{a} \wedge \vec{b}}{\vec{a} \cdot \vec{b} \wedge \vec{c}} \quad (2.7)$$

As $a \cdot b \wedge c$ is the volume of the unit cell, the bigger the unit cell is in real space, the smaller the corresponding reciprocal lattice will be.

Bragg's law

Bragg's law introduces the notion of parallel lattice planes perpendicular to S , having h, k, l , as indices and d_{hkl} as the distance between them. Diffraction becomes possible when the angle between s_0 and s called 2θ satisfies :

$$|S| = 2|s_0| \sin \theta = \frac{2 \sin \theta}{\lambda} = \frac{1}{d_{hkl}} \quad (2.8)$$

$$\lambda = 2d \sin \theta \quad (2.9)$$

The positions of Bragg reflections in a powder diffraction pattern are a function of 2θ . This means that between different instruments, at different wavelengths, the positions will be different. It is then more common to plot diffraction patterns as a function of d in \AA or Q in \AA^{-1} .

$$d = \frac{\lambda}{2 \sin \theta} \quad (2.10)$$

$$Q = \frac{2\pi}{d} \quad (2.11)$$

For a single crystal, different reflections having the same d-spacing will give separated diffraction spots easily recordable independently. With powders, as the orientation of the large number of microcrystals is random, the reflections having the same d-spacing will be merged on the same Debye-Scherrer ring.

Structure factors

The diffuse wave from an object is the sum of the scattering of all the object atoms. This is why this scattered wave can mathematically be represented as a Fourier transform of this object. A crystal can be presented as a convolution of two functions, one describing the lattice, one describing the contents of one cell.

The Fourier transforms of a convolution is equal to the product of the Fourier transforms. Diffracted waves are therefore the Fourier transform of the electronic density sampled by the reciprocal lattice. Reflections are represented by a complex number \overrightarrow{F}_{hkl} called the structure factor:

$$\overrightarrow{F}_{hkl} = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(x, y, z) e^{-2\pi i(hx+ky+lz)} dx dy dz \quad (2.12)$$

It is possible to compute the image of the density at a certain point x,y,z of the cell :

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l \overrightarrow{F}_{hkl} e^{2\pi i(hx+ky+lz)} \quad (2.13)$$

A structure factor \overrightarrow{F}_{hkl} can be described by an amplitude $F_{hkl} = |\overrightarrow{F}_{hkl}|$ and a phase Φ_{hkl} , which has to be known at the same time to build the

density.

It is possible to compute a structure factor from an atomic model containing N atoms, with f_j atomic factor scattering, B_j the agitation thermic factor, q_j the occupancy, x_j, y_j, z_j the coordinates of each atom:

$$\overrightarrow{F}_{hkl} = F_{hkl} e^{i\Phi_{hkl}} = \sum_{j=1}^{j=N} q_j f_j e^{-B_j \left(\frac{\sin \theta}{\lambda}\right)^2} e^{2\pi i(hx_j + ky_j + lz_j)} \quad (2.14)$$

The measurable intensity is due only from the amplitude on the structure factor:

$$I_{hkl} = F_{hkl} \overline{F_{hkl}} = |F_{hkl}|^2 \quad (2.15)$$

Assymmetric unit

The asymmetric unit is the smallest volume of the unit cell which is enough to describe the whole cell by using symmetry elements.

Due to the chirality of the protein amino-acids, only 64 spacegroups can appear in protein crystals.

Also it is possible that a non crystallographic symmetry exist within an asymmetric unit.

It is possible to estimate the number of molecules in an asymmetric unit, by using the Matthews coefficient:

$$V_m = \frac{V}{M.n.Z_a} \quad (2.16)$$

V is the unit cell volume, n the number of equivalent positions, Z_a the number of molecule in the asymmetric unit, and M the molecular weight of the protein. V_m is typically between 2 and 4 depending of the solvent content.

It is possible to estimate the solvent proportion in the crystal :

$$S = 1 - \frac{1.23}{V_m} \quad (2.17)$$

2.2.3 The phase problem

As illustrated before, the phases are lost during the acquisition process. It's not half the information that is lost, but much more than that. Indeed, if for example we compute an electron density map with the phases based on a putative atomic model while giving random amplitudes, we will obtain

an interpretable map around the model input. This is called model bias, the phases dominating what we see in a real space map.

If we consider Parseval's theorem:

$$rms(\rho) = \frac{1}{V} \left(\sum_h |F(h)|^2 \right)^{1/2} \quad (2.18)$$

$$rms(\rho) \propto rms(|F|) \quad (2.19)$$

$$rms(\rho_r - \rho_c) \propto rms(|F_r - F_c|) \quad (2.20)$$

To put this in words, the error of the electron density calculated is dependent of the structure factor (real and imaginary part).

If we look at the difference vector between the real structure factor and the one calculated for an actual model, we can see that a random choice of amplitude will have less importance than a random choice of phase.

But also, if the initial phases are good, the error on the amplitude will have not a big impact on the map quality. This is interesting where powder pattern intensities extraction does not give reliable errors at high angles.

It is therefore very important to have initial phases of good quality, which can be obtained by molecular replacement or experimental phasing.

Molecular replacement

Molecular replacement is a method which works well when a good proportion of the structure is already known or correctly guessed.

When two proteins have a high level of sequence identity, they will often share similar structures. Based on this assumption a first putative model can be built. This model needs to be placed correctly in the unit cell. Three parameters for the rotation and three parameters for the translation need to be found. This is a 6 dimensional problem often split into two 3 dimensional problems. A first rotation function is computed to find the 3 rotation angles, and then the orientated model is translated within the cell. This search will be done in Patterson space.

As we don't know the phases yet, we can ignore them and compute a Fourier transform of the squared amplitudes with all the phases set at 0.

This is called the Patterson map, defined by the Patterson function:

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)|^2 e^{-2i\pi(hu+kv+lw)} \quad (2.21)$$

u, v, w values between two atoms 1 and 2 are defined by :

$$u = x_1 - x_2 \quad v = y_1 - y_2 \quad w = z_1 - z_2 \quad (2.22)$$

The Patterson space is periodic and defined by a unit cell identical to the crystal unit cell. The Patterson map gives a series of position vectors (relative positions) between each pair of atoms in the structure. The value of the function at these maxima is proportional to the product of the atomic numbers of the two atoms considered.

The Patterson map vectors can be divided in intra- or inter-molecular vectors.

Intra-molecular vectors are between atoms from the same protein, and depend only on the orientation of the protein in the cell. Intra-molecular vectors can so be used for the rotation search.

Inter-molecular vectors are between two atoms of different protein molecules. Inter-molecular vectors depend on the orientation and the position of the molecule in the cell. Once the orientation has been fixed based on the intra-molecular vector, the inter-molecular vectors can be used for the position search.

The cut-off between intra and inter-molecular vectors is a distance that can be freely chosen in molecular replacement software.

Experimental phasing

Multiple isomorphous replacement Isomorphous replacement is based on the fact that a protein structure will remain the same if only few atoms are replaced by other atoms. It is therefore possible to swap some atoms for atoms having a much larger atomic number. To give a diffraction signal, the heavy atoms need to be bonded to the protein, and not present only in the bulk solvent for example. Also, to remain interpretable, the structure must not change too much (isomorphism).

The scattered intensity of a native crystal and a derivative (with heavy atoms) crystal will be different. As there are only a few atoms changed, it is possible to calculate a Patterson map that is simple and easy to deconvolute. Once the position of heavy atoms is known, it is possible to calculate a first set of structure factors.

Considering that the structure factor for the derivative crystal $\overrightarrow{F_{Prot+Heavy}}$, is the sum of the protein structure factor $\overrightarrow{F_{Prot}}$ and the structure factor from the heavy atom $\overrightarrow{F_{Heavy}}$, we have one remaining problem. The amplitude and phase of $\overrightarrow{F_{Heavy}}$ are known, but only the amplitude of $\overrightarrow{F_{Prot}}$ and $\overrightarrow{F_{Prot+Heavy}}$, meaning that there are two possible phases for $\overrightarrow{F_{Prot}}$. This is the ambiguity of the Single Isomorphous Method (SIR).

To remove this ambiguity, another crystal must be prepared, with heavy atoms binding at different sites. This is the multiple isomorphous replacement method (MIR).

The SIR method has been successfully applied to a protein powder sample of hen egg white lysozyme with gadolinium heavy atom complex (Gd-Hp-Do3A)[14].

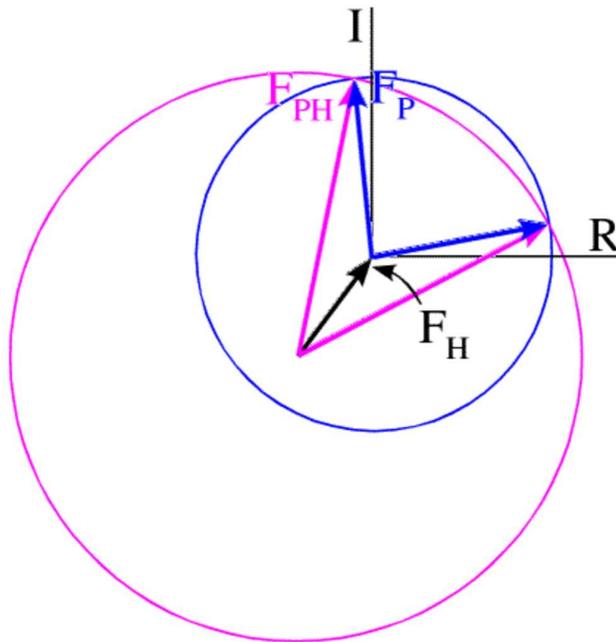


Figure 2.6: The Harker construction for the MIR method. The black vector \vec{F}_{Heavy} , (while having an error bar not represented here) is fully defined in both amplitude (length) and phase (orientation). The blue vector \vec{F}_{Prot} has only its amplitude known (represented as a blue circle centred on the \vec{F}_{Heavy} vector). The purple vector $\vec{F}_{Prot+Heavy}$ representing the derivative dataset has only its amplitude known, therefore represented as the purple circle. The intersections of the two circles represent the two possible phases for \vec{F}_{Prot} and demonstrate the ambiguity of the method.

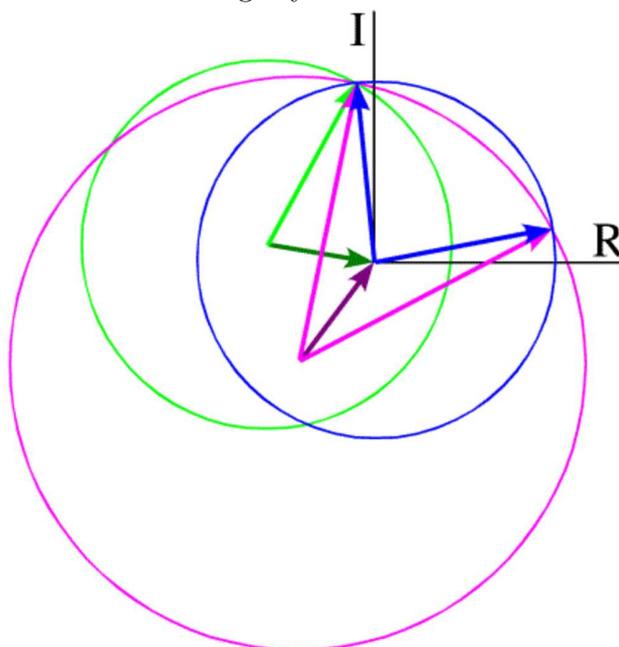


Figure 2.7: The addition of a third crystal (in green) allows resolution of the ambiguity, only one phase remains possible for \vec{F}_{Prot} .

Multiple wavelength anomalous dispersion In order to have initial phases it is possible to do a multiple wavelength anomalous dispersion (MAD) experiment. Anomalous scattering refers to a change in a diffracting X-ray's phase that is unique from the rest of the atoms in a crystal due to X-ray absorption at a given wavelength. The protein needs to contain atoms able to give such an anomalous signal. Metal ions in metalloproteins, or selenium in the classical exchange of the methionine amino acid by selenomethionine, are good candidates in single crystal measurements.

The scattering factor f for a given atom is :

$$f = f_0 + f'_\lambda + i f''_\lambda \quad (2.23)$$

Where f_0 is the usual scattering factor of the element, f' and f'' are the real and imaginary parts of the anomalous scattering.

In this experiment, the wavelength is first changed to the atom's absorption edge where $\Delta f'$ is at its maximum amplitude (as $\Delta f'$ is a negative quantity, it is in fact a minimum). It is then changed to the white line, where f'' is largest, and finally to a remote location where both $\Delta f'$ and f'' are small.

As the protein atoms have no contribution to the anomalous signal observed, only the anomalously scattering atoms will break Friedel's law and make amplitudes different between h,k,l and $-h,-k,-l$.

A problem occurs when this is applied to powders. With powder samples, Friedel pairs (h,k,l and $-h,-k,-l$) are exactly overlapped, the anomalous difference (f'') cannot be measured.

Fortunately, with powders, the variation in dispersive difference $\Delta f'$ can still be measured. This method has been demonstrated recently on $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$ at the K edge of Ni to do ab initio structure determination using dispersive differences with a powder sample[15].

2.2.4 Rietveld refinement

Refinement of a structure from an X-ray powder pattern can be done either with extracted intensities, or using the Rietveld method[1].

With extracted intensities, the overlapped reflections will lead to a loss of information. The sum of the reflections observed will be split between the

reflections without knowledge of their internal repartition.

On the other hand, using the Rietveld method will help in taking care of this problem.

A model of the structure of the sample studied is created and modelled as a powder pattern. For this other parameters depending of the instrument and the sample characteristics need to be modelled, like peak shape, scale, solvent scattering and background.

The modelled pattern is then compared to the observed, and a least squares minimisation is carried out in order to minimise their differences by changing the parameters of the original model. The least squares minimisation ends when the residual S_y is minimised.

With S_y the residual, w_i a weighting factor, y_{meas} the measured intensities and y_{calc} the modelled intensities.

$$S_y = \sum_i w_i \times (y_{meas} - y_{calc})^2 \quad (2.24)$$

Statistical parameters exists to follow the evolution of the refinement of a powder pattern.

$$R_p = \left[\frac{\sum_i |y_{meas_i} - y_{calc_i}|}{\sum_i w_i y_{meas_i}} \right] \quad (2.25)$$

$$R_{wp} = \sqrt{\left[\frac{\sum_i w_i (y_{meas_i} - y_{calc_i})^2}{\sum_i w_i (y_{meas_i})^2} \right]} \quad (2.26)$$

$$R_e = \sqrt{\left[\frac{N - P}{\sum_i w_i (y_{meas_i})^2} \right]} \quad (2.27)$$

$$\chi^2 = \left[\frac{R_{wp}}{R_e} \right]^2 \quad (2.28)$$

The R_{wp} factor is the most used when refining. It is compared to the R_e to which it can theoretically go (N=number of observation, P number of parameters). χ^2 indicates how well a refinement is going. Following R_{wp} and χ^2 during the refinement gives an idea to whether the refinement is working or if it has fallen into a false minimum for example. Taken alone, these parameters are not enough to tell if the refinement is finished. Plotting the difference curve between the observed and calculated values gives more information than the statistical values. Indeed, their dependance on the background and the total number of points included can lead to misleading values. For example, a high background, with many overlapped peaks

2.2. CRYSTALLOGRAPHY - THEORY

35

at high angle, will lead to low statistical values. This is why, when a powder refinement is presented, it is always followed by a plot of the observed, calculated and difference curves.

2.3 Crystallisation

The crystallisation of a protein needs three steps.

The two first steps are the protein expression and purification needed to obtain an extremely pure amorphous protein sample. These two steps were, in the case of the results presented in this thesis, carried out by collaborators or samples were bought commercially. In all cases we received protein of the finest quality, and this is a crucial step in order to allow crystallisation.

The third step is obtaining a crystal of this protein. This step is a major problem for protein single crystal crystallography, as the growth of a perfect crystal requires much effort.

2.3.1 Crystallisation

The idea behind protein crystallography is to change a protein from a free amorphous state in a liquid to an ordered solid state, the crystal. This state transition is done while an equilibrium is established between the two states. Two major steps will be necessary, the crystal nucleation and the crystal growth.

Some of the surface amino acids of the protein, in contact with the solvent when the protein is in an amorphous state, will need to be able to make protein-protein interactions to initiate crystal nucleation. To start this phenomenon, the protein needs to be in a thermodynamically unstable state, of supersaturation. This state is achieved by diminishing the solubility of the protein leading to destabilised interactions of the protein and the solvent. The nucleation can be homogeneous or more often heterogeneous. A homogeneous nucleation results in local modification of the concentrations in the liquid. A heterogeneous nucleation appears when the liquid contains foreign particles such as dust.

Up to this moment two scenarios may occur. If the interactions between two proteins are unspecific, it will lead to an unusable amorphous precipitate. If the interactions between the proteins are specific, nucleation will occur.

After the nucleation, the protein is still in a thermodynamically unstable state and this will lead the protein nuclei to grow to a bigger solid, the crystal.

A large number of factors need to be tested in order to obtain a perfect supersaturated state, leading to a good crystal. Among them:

1. Protein purity, absence of contaminants or isomers of the same protein in a different conformation.
2. Protein solubility, is dependent on buffers, ions, substrates, inhibitor, cofactor, detergents.
3. pH which will influence the charges at the surface of the protein.
4. Temperature will change the solubility and also the kinetics of the crystallisation process.
5. Precipitant agents
 - (a) Polymers like PEGs, which will exclude solvent around the protein.
 - (b) Non chaotropic salts, changing the ionic strength by augmenting the hydrophobic interaction by excluding solvent.
 - (c) Organic solvent which diminishes the dielectric constant of the mother liquor.

Several methods exist for growing protein crystals. The most used for single crystals is the hanging drop method, while batch crystallisation is the most used method for protein powder samples.

Vapour diffusion, the hanging drop method

The principle behind this method is the fact that two solutions of different initial concentration will equilibrate if they are inside a closed environment.

In practice, the bottom of a well is filled with the mother liquor containing all the components to test at a certain concentration. A drop hanging above the well is made with a part of this mother liquor and some protein. At this moment, the precipitant concentration is lower, diluted by the addition of the protein. To return to an equilibrated state, vapour diffusion will lead to evaporation of the drop, diminishing its volume, therefore leading to a slow concentration of the protein.

The hanging drop method is very well suited when small amounts of crystalline material are needed, such as required for a single crystal diffraction experiment. To have enough material for a typical powder experiment in a capillary, one needs to mix several independent drops. This mixing of different crystals formed independently may lead to poorer results due to local variation of the samples' growth. Therefore another method, more suited to producing more material of the same condition, is needed.

Batch crystallisation

Batch crystallisation is much simpler and ideal when large volumes of protein crystals are needed. In a batch experiment there is no equilibration against another well. All the components are mixed at their final concentrations at the beginning of the experiment. Usually, concentrations required for batch crystallisation are higher in precipitants, but also sometimes in protein. Due to the higher concentrations, precipitation is usually faster, and often happens overnight. Unfortunately, some crystalline forms don't like the stress of a fast crystallisation. In some cases, a lower crystallisation speed leads to a crystal diffracting better (finer peak width and better d-space resolution).

2.3.2 Cryocooling role

When a protein crystal is inside an intense X-ray beam like at a synchrotron, a phenomenon called radiation damage occurs.

During data collection, the crystal degrades, releasing free radicals broken from the protein and the solvent.

These species diffuse inside the crystal, and therefore interact with the components. These movements are facilitated by the fact that the immediate environment of the protein is bulk water. The result is a lattice quickly damaged, and diffraction properties affected.

Cryocooling is a general term used when a crystal is cooled before and while inside an X-ray beam. By doing so, the free radicals will still be created, as it has been demonstrated in some studies, but will not move so rapidly inside the crystal. The resulting effect is a crystal able to support a longer data acquisition.

Unfortunately, as an important component of our crystal is bulk water, cooling the crystal without any protection will in most cases result in crystalline water ice outside the crystal. This ice will break the fragile protein crystal and its diffraction properties will be lost.

To avoid having crystalline ice, two main parameters need to be taken into account.

First, the speed of the cooling is important. If the sample is cooled slowly, crystalline ice will be able to form. So the sample needs to release quickly its internal heat. While this is easy for a single crystal directly in contact with

2.3. CRYSTALLISATION

an open environment, we will see that it is much more difficult for a protein powder sample.

Second, the mother liquor used to grow the crystal must be modified in order to allow the solvent to be cooled without forming crystalline ice. This is done by adding a cryoprotectant to the mother liquor. Unfortunately modifying the fragile equilibrium of the protein may dissolve the crystal. Much consideration should be given to choosing and improving cryocooling conditions. Also new conditions by mixing two different cryoprotectants appeared recently. By differentiating the effect of the cryoprotection inside and outside the crystal, a non-damaging cryoprotectant goes inside the solvent channel of the crystal, while another cryoprotectant stays outside and prevents crystalline ice surrounding the crystal.

3

Polymorphism of urate oxidase

Ce chapitre est dédié à l'étude du polymorphisme de l'urate oxidase par diffraction de poudre aux rayons-X. Urate oxidase est une enzyme catalysant l'acide urique en 5-hydroxyisourate, conduisant à l'allantoïne. L'urate oxidase est présente chez de nombreux organismes vivants, mais pas dans l'espèce humaine. En effet, au cours de l'évolution, la sélection naturelle a conservée des délétions et mutations précises du génome pour empêcher son expression. L'avantage de la perte de cette protéine est due à la présence finale d'acide urique dans le sang qui ne sera pas dégradé en allantoïne. L'acide urique est un puissant anti-oxydant, et, de par sa présence, il va protéger l'organisme contre les radicaux libres. Les maladies neurologiques ainsi que les cancers dûs à l'âge sont moindre. L'espérance de vie des organismes incapables de dégrader l'acide urique est donc accrue.

Malheureusement une trop forte concentration d'acide urique peut entraîner certaines maladies comme l'hyper-uricémie (goutte). L'acide urique a donc été utilisée comme protéine-médicamenteuse depuis 20 ans. Une forme microcristalline du médicament permettrait une conservation au domicile du patient. Ceci éviterait une hospitalisation, les patients pourraient s'injecter eux-même leur traitement. Différentes préparations possibles de microcristaux d'urate oxidase ont été analysées par diffraction aux rayons-X. L'étude du polymorphisme de ces microcristaux par diffraction de poudre est l'objet de ce chapitre.

3.1 Introduction

3.1.1 Function of the enzyme

Urate oxidase is an enzyme catalysing the oxidation of uric acid to 5-hydroxyisourate. The production of allantoin from a metastable 5-hydroxyisourate is via a non enzymatic pathway[16]:

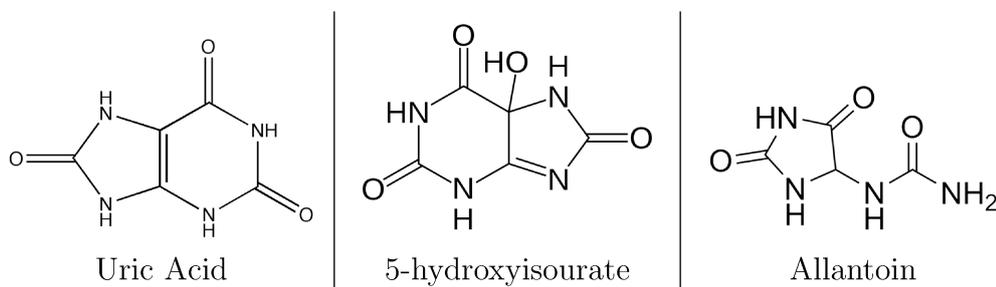
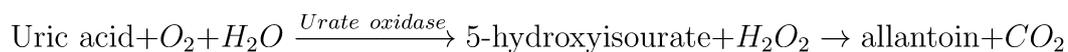


Figure 3.1: Urate oxidase catalyzes uric acid into 5-hydroxyisourate, a metastable compound which decomposes into allantoin nonenzymatically[16].

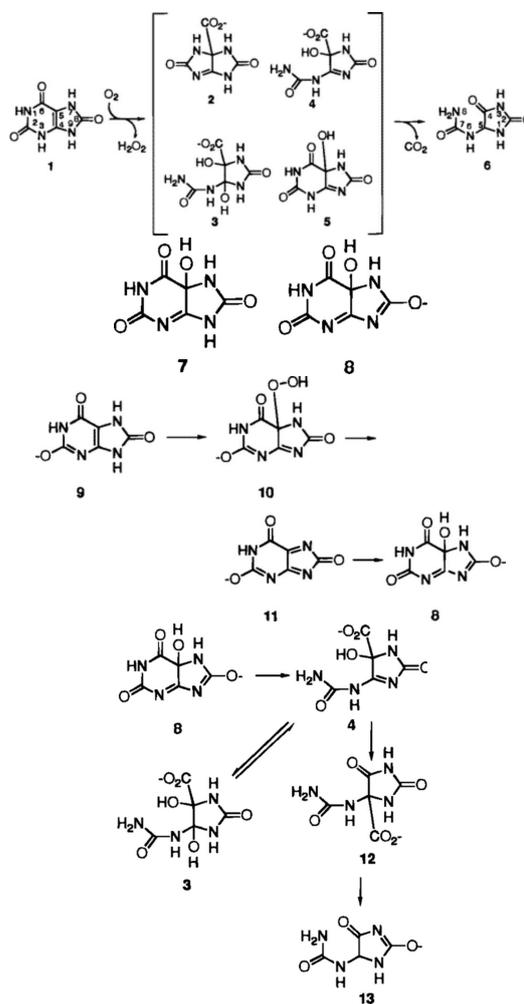


Figure 3.2: Complete urate oxidase reaction as shown by *Kahn et al.*[16].

3.1.2 Signification of the absence of Uox in primates

Urate oxidase is found in a very large variety of organisms, from bacteria to plants and mammals. However in humans and in many primates urate oxidase is absent.

The absence is due to precise mutations and deletions conserved during evolution[17]. The evolutionary advantage of the loss of this protein is due to the benefit of having uric acid as the end product of catabolism of purines in humans. Indeed, uric acid is a powerful antioxidant, and will therefore protect the organism against singlet oxygen and other radicals. This an-

tioxidant property helps in decreasing neurological diseases and age-specific cancers[18]. This has led to a lengthening of the life-span of organisms unable to degrade uric acid.

3.1.3 Drawbacks of the absence of Uox in humans

Unfortunately, an excessive concentration¹ in the blood of uric acid can lead to acute hyperuricaemia and gout. According to [19] the main factors promoting hyperuricemia are obesity, purine-rich diet, insulin resistance, diuretic use, hypertension, renal insufficiency, genetic makeup, dehydration while performing exercise, but most of all, alcohol consumption. The causes of hyperuricemia, accelerated generation of uric acid and/or impaired excretion in the kidney, lead to a classification of hyperuricaemia into three types, overproduction, underexcretion and mixed.

3.1.4 Pharmaceutical interest

As urate oxidase can degrade the uric acid responsible for the previously cited diseases, an enzyme based drug seems a logical possibility.

Actual formulation of urate oxidase

Uricozyme® was the first urate oxidase formulation for drug use produced from *Aspergillus flavus*. It has been produced at a rate of 1 kg per year over the last 20 years, allowing treatment for 15000 patients per year[20]. Fasturec®, consisting of a liquid formulation of rasburicase stable at 25°C[21] replaced Uricozyme® which was a lyophilised urate oxidase preparation requiring the dissolution inside a solvent before injection.

The actual Fasturec® formulation[22], a recombinant urate oxidase enzyme (raburicase), made by Sanofi-Synthelabo, can be used to prevent renal failure in tumor lysis syndrome (TLS). In this syndrome, patients commencing chemotherapy produce a high level of uric acid due to the tumor lysis or shrinkage[23].

1. If the concentration in the blood exceeds $360 \mu\text{mol.L}^{-1}$ for women and $400 \mu\text{mol.L}^{-1}$ for men.

Towards a crystalline form of the drug ?

For the fabrication of Fasturec®[®], the rasburicase is purified using multiple steps of chromatography. These expensive steps of purification of rasburicase could be replaced by purification through crystallisation.

Crystallisation also has several advantages over liquid formulation. The drug release profile is modified by a microcrystalline formulation, due to the time taken by the crystals to dissolve in the blood, avoiding a strong release peak at the drug's administration. The stability for storage is better than in liquid formulations. The viscosity of the formulation is lower, allowing higher concentration of the protein. The manufacturing cost is lower than with lyophilisation.

The main difficulty in engineering crystalline protein drugs is to obtain crystallisation conditions allowing the growth of microcrystalline protein, while being administrable. Small isogranular and isometric crystals are required. The crystallisation conditions must contain only non-toxic approved additives at physiological osmolarity. The most difficult parameter to adjust is the osmolarity, which should be kept to an equivalent of 130 mM NaCl concentration, which could be a limiting factor. Actually the only crystalline protein drug meeting these requirements is insulin[24][25].

Interest of polymorphism

The protein size and polymorphism will theoretically control the dissolution of the injected crystal, and therefore the drug release rate. The variation of the crystallising agent concentration will lead to a modification in the polymorphism and the crystal size.

In order to have injectable crystals, crystal size has to be kept under 50 μm [26], and the bigger the crystal is, the longer the release will be. The second parameter to govern release rate will be the polymorphism of the crystal. Polymorphism is studied for drugs in solid form, to evaluate the impact of different polymorphs [27][28]. This is due to the fact that crystal packing governs crystal solubility, and thus will govern the bioavailability. As the crystal packing is different, binding between protein molecules will be different. The more binding, the lower the release rate of the protein in the blood.

3.2 Materials and methods

3.2.1 Crystallisation

Sanofi-Aventis provided us with recombinant urate oxidase from *Aspergillus flavus*, expressed in *Saccharomyces cerevisiae*. The initial phosphate buffer was swapped for a Tris buffer (50 mM Tris buffer pH 7.0, 50 mM Tris buffer pH 9.0, or pure water) using gel-filtration chromatography. For this three conditions, pH was adjusted with HCl prior to protein transfer, and checked afterwards. No pH changes were observed after transfer (pH 7.0 and pH 9.0).

8-Azaxanthine was bought from Sigma-Aldrich and was incubated with the protein prior to crystallisation, any excess of non-bound 8-Azaxanthine is removed during the gel filtration.

Urate oxidase at a concentration of 3 mg.ml⁻¹ and in the absence of salt crystallises in a week. The crystallisation can be made faster by adding more than 2% of PEG8000. Keeping crystals in a cold room (277 K) leads to better diffraction patterns.

Different crystallization conditions containing salts were established. It was previously known [29] that the addition of salt between 10 mM and 1 M dissolves pre-existing crystals or prevents their crystallisation. The addition of pEG8000 counterbalances this effect and allows the growth of new crystal forms with salt.

Stock solutions of NaCl, KCl, NH₄Cl and (NH₄)₂SO₄ (Sigma-Aldrich) were prepared by dilution of the appropriate amounts of salt in 50 mM Tris buffer pH 8.0. A solution of 40 % w/v of PEG 8000 in 50 mM Tris buffer pH 8.0 was prepared from a 50 % w/v solution (Hampton Research). All salt and rasburicase solutions for crystallisation trials were filtered through 0.22 μm Millipore filters.

Crystallisation was done using the batch crystallisation method (2.3.1, p.38), directly inside a 2ml eppendorf tube.

3.2.2 Sample preparation

Crystals and their mother liquor were loaded into 1 mm and 1.5 mm diameter borosilicate capillaries and centrifuged (A.1, p.167).

In order to exploit cryocooled conditions we deposited inside 96-well plates urate oxidase crystals mixed with all the cryoprotectants contained in the “cryopro” set, obtained from Hampton Research.

3.3 Results - Lattice parameter study of the various phases

3.3.1 Summary

Using the software TOPAS[30] (F.1, p. 215) we precisely determined the lattice parameters of all the phases we observed (3.1, p. 50 and 3.2, p. 49). Even if all the crystal forms observed conserved the tetrameric form of urate oxidase, the crystal packing is different and the variation of crystallisation conditions induces noticeable changes in the lattice parameters.

Uox	1WS3	1XY3	1XXJ	1R51
Ligand	Uracil	Guanine	Benzene	8-azaxanthine
Space group	$P3_121$	$P2_1$	$P2_12_12$	$I222$
Unit-cell parameters				
a (Å)	140.60	82.75	126.25	81.30
b (Å)	140.60	141.94	142.27	96.30
c (Å)	151.08	135.08	81.32	105.60
α (°)	90	90	90	90
β (°)	90	92.67	90	90
γ (°)	120	90	90	90
Matthews coefficient (Å ³ Da ⁻¹)	3.14	2.66	2.60	2.95
Solvent content (%)	60.8	53.8	52.7	58.3
No. of monomers per AU	4	8	4	1

Table 3.2: Different structural models of Uox available in the PDB relevant to the phases reported in this study. All crystallisation was done in the presence of 5-8 %(*w/v*) PEG 8000 at pH 8.5 with the inhibitors added in large excess.

Uox crystallised with PEG concentration	<i>NH₄Cl</i> 15 %	<i>NaCl</i> 15 %	<i>NaCl</i> 8 %	<i>(NH₄)₂SO₄</i> 15 %	<i>H₂O</i> 15 %	<i>KCl</i> 10 %	complexed AZA 10 %
Data Collection							
Wavelength (Å)	1.30000(6)	1.30000(6)	1.53986(6)	1.30000(6)	1.30000(6)	1.30000(6)	1.30000(4)
Exposure time per scan (s)	60	60	60	60	60	60	180
LeBail refinement							
Space group	<i>P</i> 3 ₁ 21	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 3 ₁ 21	<i>I</i> 222
Unit-cell parameters							
<i>a</i> (Å)	140.4002(9)	81.8712(9)	82.593(2)	75.4153(8)	133.6473(19)	141.455(3)	80.2944(3)
<i>b</i> (Å)	140.4002(9)	124.7628(13)	141.723(4)	141.8930(15)	135.9379(16)	141.455(3)	96.2244(3)
<i>c</i> (Å)	151.1053(13)	142.9454(15)	134.919(3)	149.5638(15)	78.9043(9)	151.273(5)	105.5280(3)
α (°)	90	90	90	90	90	90	90
β (°)	90	93.7280(6)	92.714(14)	94.4041(6)	90	90	90
γ (°)	120	90	90	90	90	120	90
Volume (Å ³)	2579560(40)	1457020(30)	1577490(70)	1595740(30)	1433510(30)	2621400(100)	815340(5)
Matthews coefficient (Å ³ Da ⁻¹)	3.14	2.66	2.88	2.90	2.60	3.19	2.95
Solvent content (%)	60.8	53.8	57.3	57.6	52.7	61.4	58.3
No. of monomers per AU	4	8	8	8	4	4	1
Resolution range (Å)	121.6–8.0	142.6–10.0	134.8–12.0	149.0–10.0	95.3–8.2	121.5–7.5	106.4–3.6
Closest single crystal PDB structure	1R51	1WS3	1XXJ	1XY3	1XY3	1XY3	1XY3
Figure 3.3 label	(b)	(e)	(d)	(f)	(c)	(g)	(a)

Table 3.1: Details of sample crystallisation, data collection, and LeBail [31] analysis of the powder profiles.

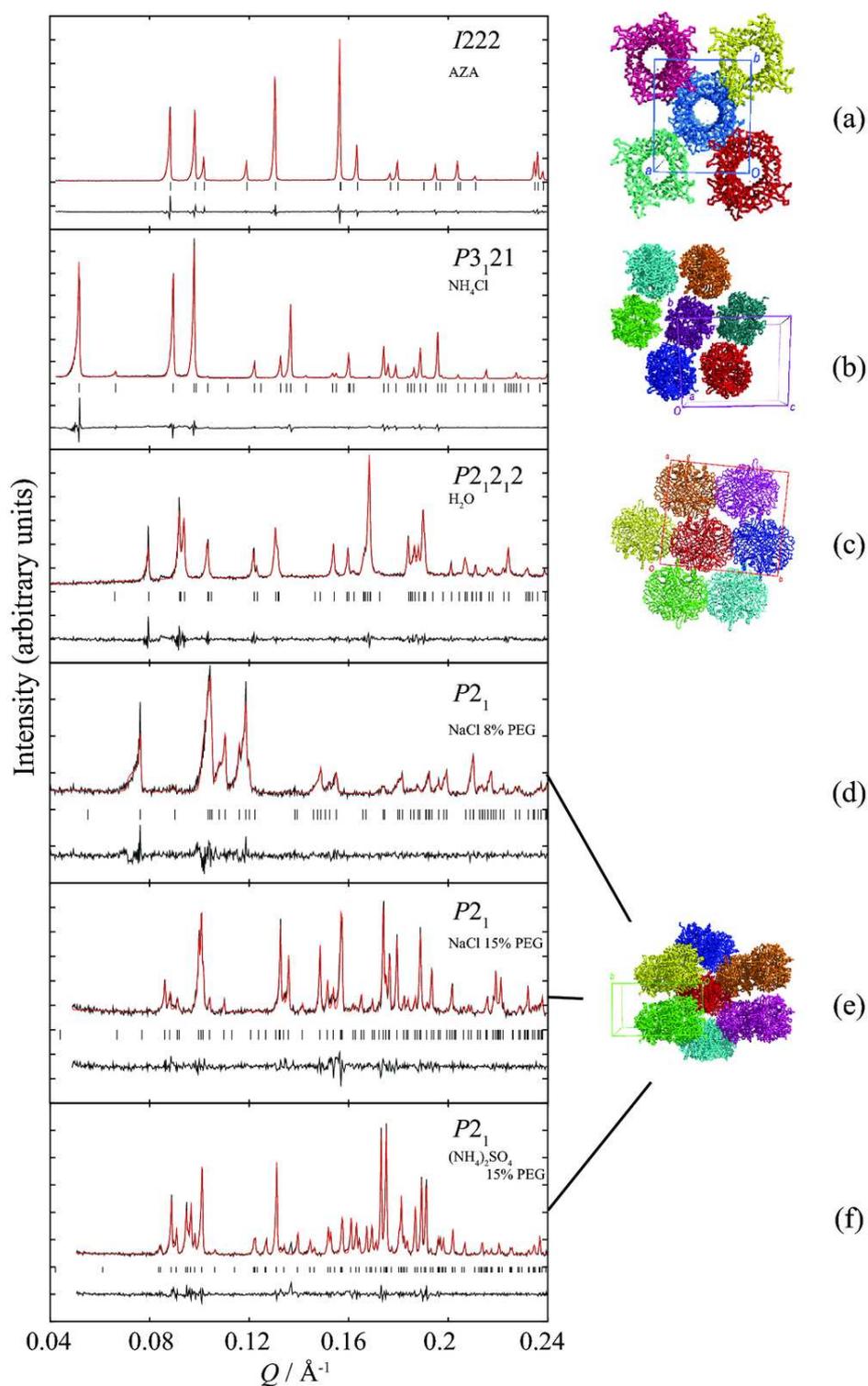


Figure 3.3: **Left panel:** Selected Q -region of the LeBail fits for six distinct phases of Uox reported in this study. **Right panel:** Projection in the ab plane of candidate structure models available in the pdb (a) 1R51 [32], (b) 1WS3 [33], (c) 1XXJ [32], (d), (e), and (f) 1XY3 [32].

3.3.2 A new way of obtaining an orthorhombic ($P2_12_12$) phase

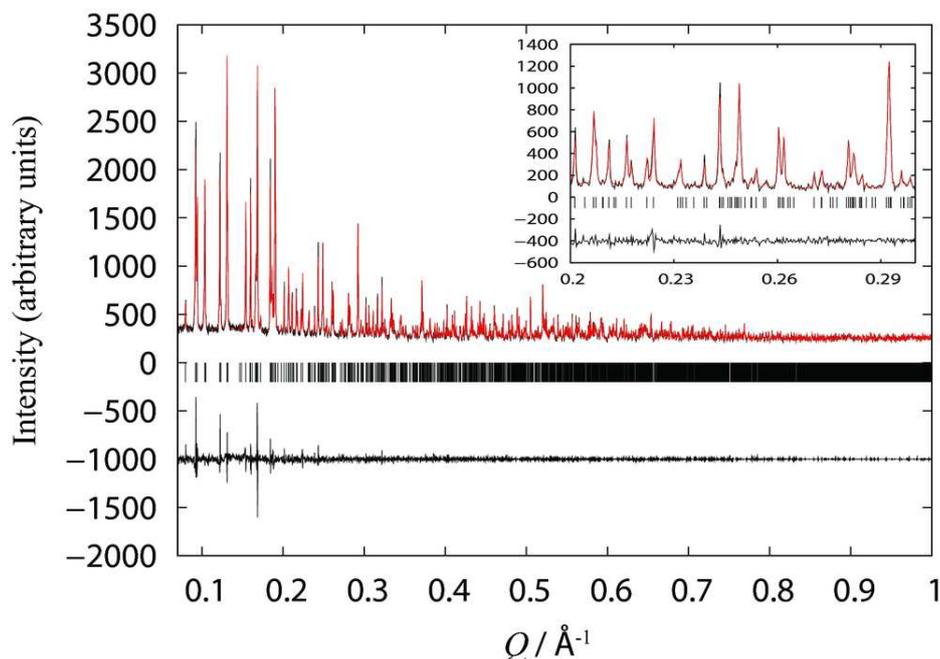


Figure 3.4: LeBail fit of Uox ligand-free Uox crystallised in water with 10 % PEG 8000 ($P2_12_12$) (ID31: $\lambda = 1.30000(6)$ Å, 295 K). The black, red, and lower black lines represent the experimental data, calculated pattern, and difference between experimental and calculated profiles respectively. The vertical bars correspond to Bragg reflections compatible with its particular space group. The insets correspond to magnification of the fits in selected Q ranges.

Two new crystallisation conditions were found leading to the same phase. The crystallisation conditions have the remarkable originality to be pure water or Tris buffer, without any salt or ligands. This is potentially of significant pharmaceutical interest.

In the absence of salt, we saw that the addition of PEG 8000 is not essential, but leads to variations with lattice parameters (figure 3.5, p. 54),

$$\frac{\Delta V}{V_i} = \frac{V_{[PEG8K]=25\%} - V_{[PEG8K]=0\%}}{V_{[PEG8K]=25\%}} = -7\%$$

In the absence of salt or polymer, when the pH is changed a small anisotropic variation of the lattice parameters is observed. When passing from pH 7.2 to 9.0 a increases while b and c decrease.

(3.7 (a), p. 54).

In the presence of PEG8K, when the pH is changed between pH 7.4 and pH 8.0 a clear variation of the lattice parameters is observed. At a PEG concentration of 8% or 16% all the lattice parameters increase (except b at 8% PEG8K) (3.7 (b and c), p. 54).

But when the pH is changed between pH 8.0 and pH 8.5 another anisotropic effect is observed, a and b decrease while c increases (3.7 (b and c), p. 54).

When looking only at the unit cell volume when varying the pH, the presence and concentration of PEG8K will induce different effects (3.6, p. 54). The biggest effect could be seen between pH8.0 and pH8.5 where the shrinkage of the unit cell $\frac{\Delta V}{V_i} = -0.280 \%$

No PDB structure matches the cell parameters found for this phase (figure 3.2, p. 49).

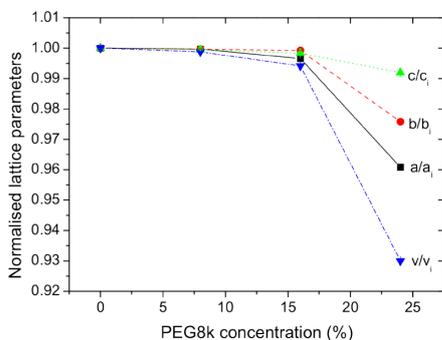


Figure 3.5: Evolution of the normalised cell parameters of orthorhombic ($P2_12_12$) ligand-free Uox crystallised in Tris buffer without salt at pH 8 with increasing PEG 8000 concentration.

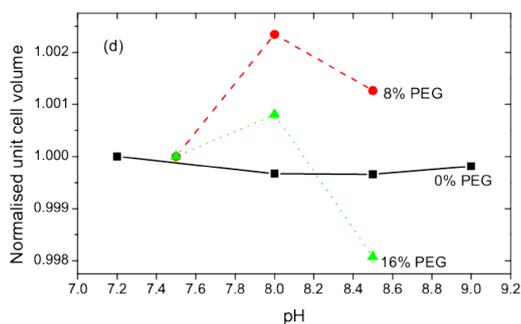


Figure 3.6: pH evolution of the normalised unit cell volume of orthorhombic ($P2_12_12$) ligand-free Uox crystallised in Tris buffer without salt using (a) 0 % PEG, (b) 8 % PEG, (c) 16 % PEG.

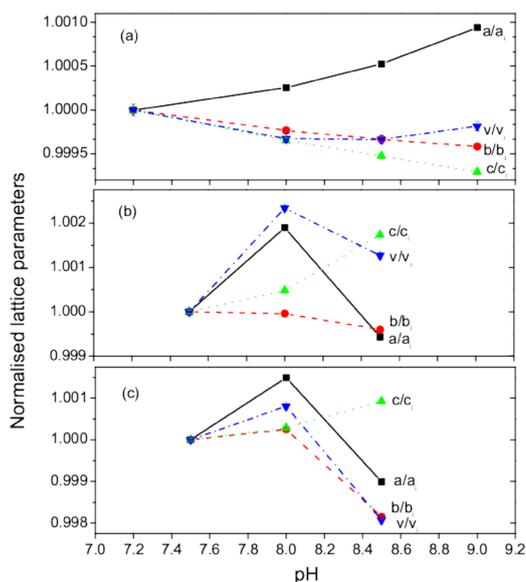


Figure 3.7: pH evolution of the normalised lattice parameters and unit cell volume of orthorhombic ($P2_12_12$) ligand-free Uox crystallised in Tris buffer without salt using (a) 0 % PEG, (b) 8 % PEG, (c) 16 % PEG.

3.3.3 Trigonal ($P3_121$) phase

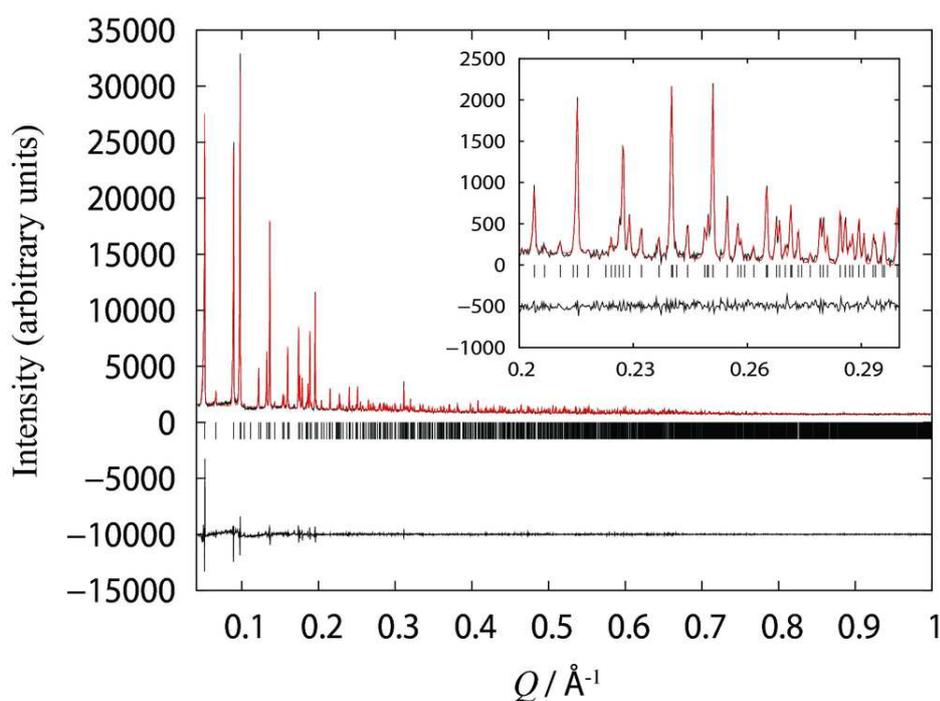


Figure 3.8: LeBail fit of Uox ligand-free Uox crystallised with NH_4Cl and 15 % PEG 8000 ($P3_121$) (ID31: $\lambda = 1.30000(6)$ \AA , 295 K). The black, red, and lower black lines represent the experimental data, calculated pattern, and difference between experimental and calculated profiles respectively. The vertical bars correspond to Bragg reflections compatible with its particular space group. The inset correspond to magnification of the fit in the selected Q range..

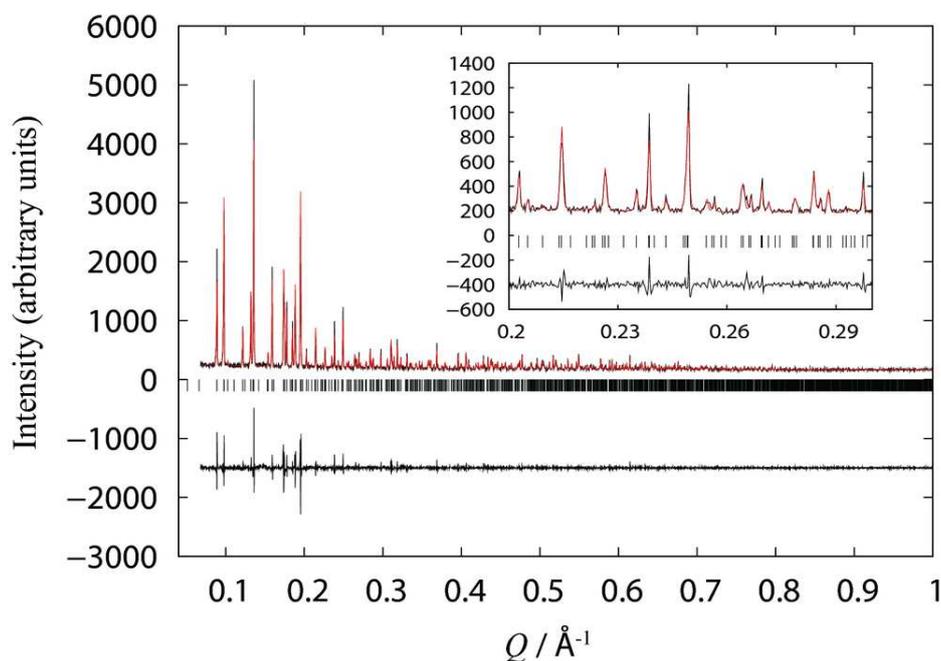


Figure 3.9: LeBail fit of ligand-free Uox crystallised with *KCl* and 10 % PEG 8000 ($P3_121$) (ID31: $\lambda = 1.30000(6)$ Å, 295 K).

This phase can also be obtained in the presence of salt but only combined with the presence of a polymer. We tried two different salts NH_4Cl and KCl combined with PEG8000.

This phase is very interesting due to the salting-in effect observed[29]. Indeed, when the salt concentration is increased, the protein becomes more and more soluble. In our case, when salts are present, only the addition of PEG8K allows crystallisation.

Two different anisotropic effects are observed during the data collection time (figure 3.3.3, p. 58).

The first is observed only for Urate oxidase crystallised with KCl imitating a second-order phase transition between 15 s and 75 s exposure time, where we saw a decrease along the c axis ($\Delta c/c_i = -1.22\%$) The second effect appeared in both conditions tested (NH_4Cl or KCl , without inhibitor, pH 8.0, PEG8K) between 45 s and 75 s in the beam. While the a crystallographic axis did not evolve much ($\Delta a/a_i = 0.39\%$), the c axis decreases abruptly ($\Delta c/c_i = -1.16\%$).

The extracted lattice parameters are closely related to the PDB models of Uox complexed with uracil (1WS3) and 5,6-diaminouracil (1WS2) listed

in Table (figure 3.2, p. 49) [32]. However, even with these starting models, the data quality (8 Å) was insufficient for redetermining the structure via MR.

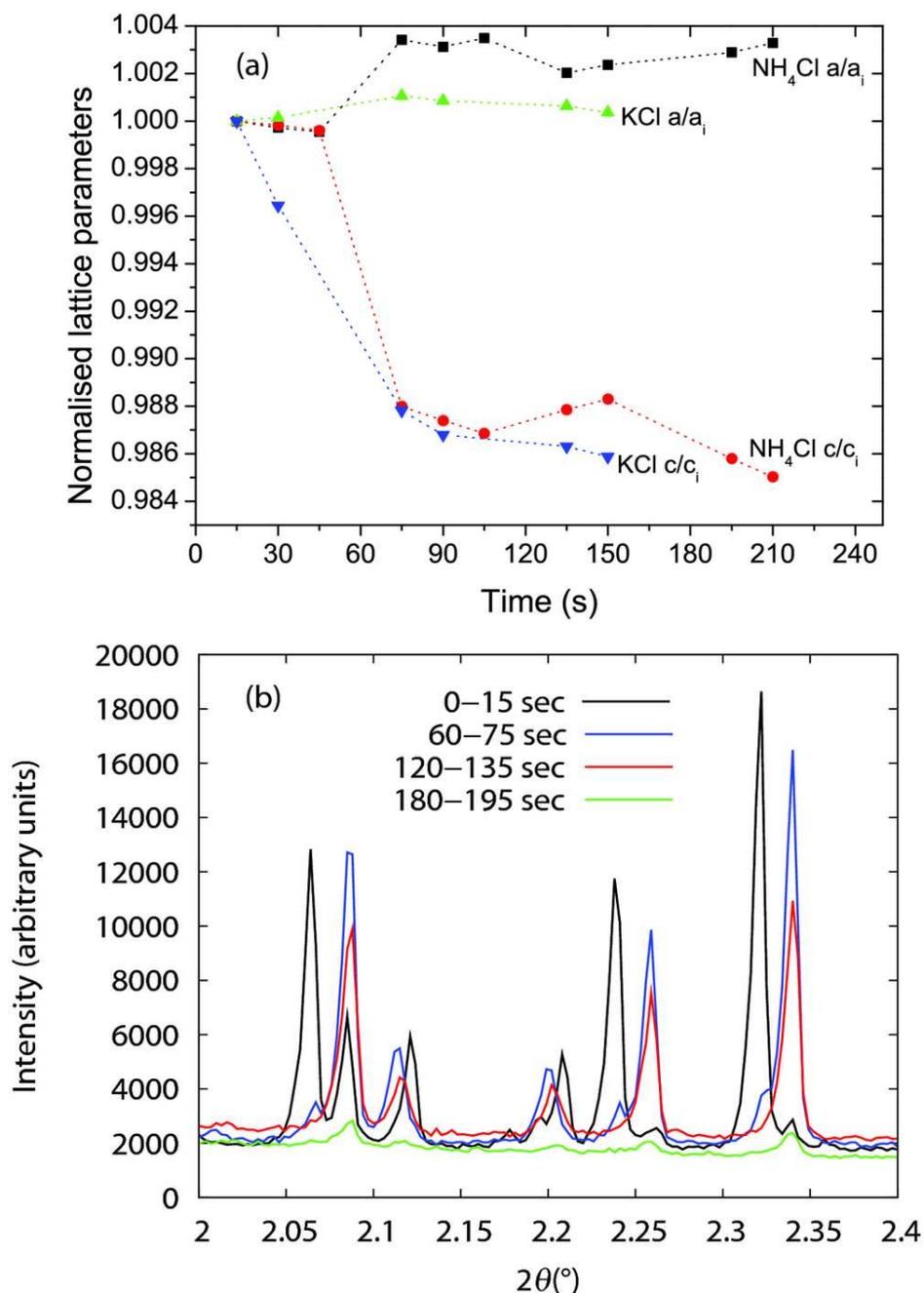


Figure 3.10: (a) Anisotropic variation of the normalised lattice dimensions of trigonal ligand free Uox crystallised with NH_4Cl or KCl at pH 8 ($P3_121$) with increasing sample exposure time to the synchrotron beam. (b) Selected 2θ region of powder diffraction profiles of Uox crystallised with NH_4Cl at pH 8 (ID31: $\lambda = 1.30000(6)$ Å, 295 K) showing a gradual evolution of the peak positions and widths with increasing irradiation time. The colours correspond to different sample exposure times varying from 15 to 195 seconds.

3.3.4 Three monoclinic ($P2_1$) phases

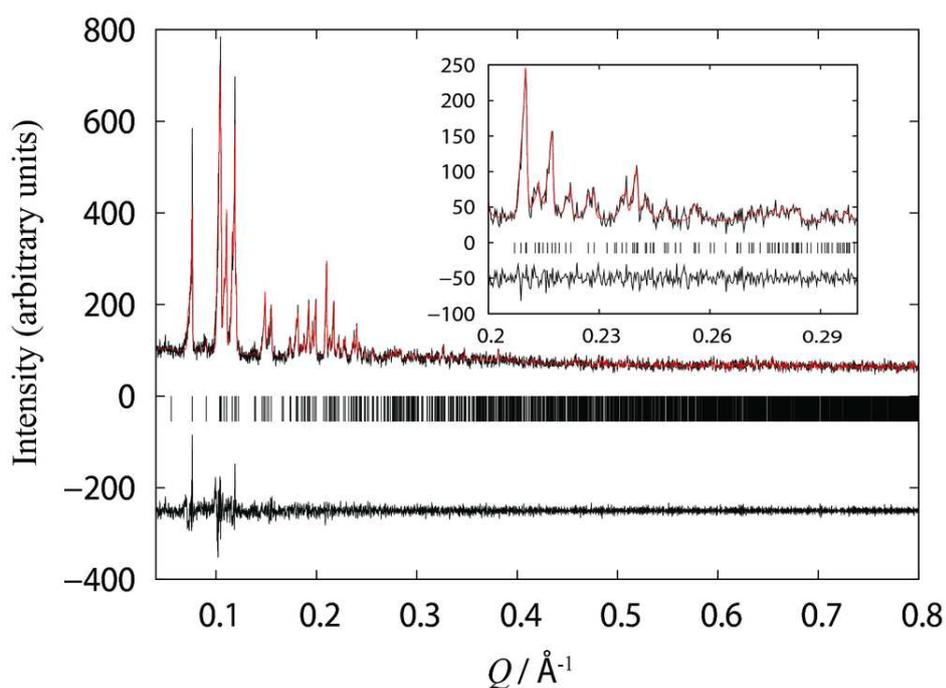


Figure 3.11: LeBail fit of ligand-free urate oxidase crystallised with NaCl and 8 % PEG 8000 ($P2_1$) (ID31: $\lambda = 1.30000(6)$ Å, 295 K). The black, red, and lower black lines represent the experimental data, calculated pattern, and difference between experimental and calculated profiles respectively. The vertical bars correspond to Bragg reflections compatible with its particular space group. The inset correspond to magnification of the fit in the selected Q range.

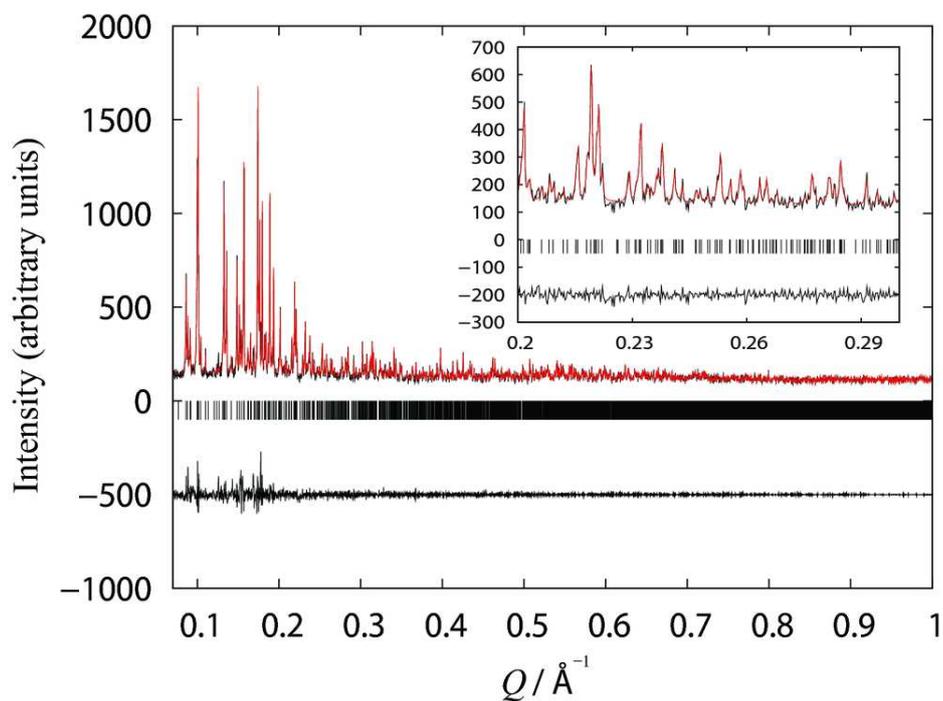


Figure 3.12: LeBail fit of ligand-free urate oxidase crystallised with NaCl and 15 % PEG 8000 ($P2_1$) (ID31: $\lambda = 1.30000(6)$ Å, 295 K). The black, red, and lower black lines represent the experimental data, calculated pattern, and difference between experimental and calculated profiles respectively. The vertical bars correspond to Bragg reflections compatible with its particular space group. The inset correspond to magnification of the fit in the selected Q range.

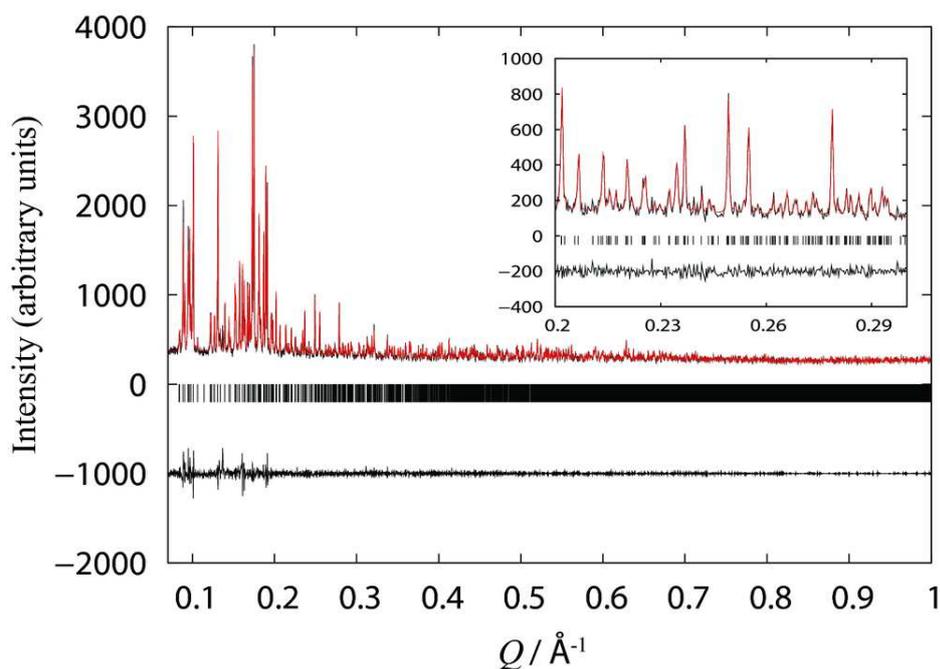


Figure 3.13: LeBail fit of ligand-free urate oxidase crystallised with $(NH_4)_2SO_4$ and 15 % PEG 8000 ($P2_1$) (ID31: $\lambda = 1.30000(6)$ Å, 295 K). The black, red, and lower black lines represent the experimental data, calculated pattern, and difference between experimental and calculated profiles respectively. The vertical bars correspond to Bragg reflections compatible with its particular space group. The inset correspond to magnification of the fit in the selected Q range..

Two different salts, $NaCl$ and $(NH_4)_2SO_4$ gave monoclinic symmetry when associated with the polymer PEG8K. Variation of lattice parameters depending of the crystallisation conditions are seen (figure 3.1, p. 50). Three $P2_1$ phases are observed:

Uox crystallised with PEG concentration	$NaCl$ 15 %	$NaCl$ 8 %	$(NH_4)_2SO_4$ 15 %
Space group	$P2_1$	$P2_1$	$P2_1$
Unit-cell parameters			
a (Å)	81.8712(9)	82.593(2)	75.4153(8)
b (Å)	124.7628(13)	141.723(4)	141.8930(15)
c (Å)	142.9454(15)	134.919(3)	149.5638(15)
α (°)	90	90	90
β (°)	93.7280(6)	92.714(14)	94.4041(6)
γ (°)	90	90	90
Volume (Å ³)	1457020(30)	1577490(70)	1595740(30)

In each of these three phases, two tetramers make up the asymmetric unit. Even if this doesn't have any drawbacks for the determination of the lattice parameters, this could be a problem for obtaining enough high d-space resolution diffraction data for structural refinement, as more atoms will result in more overlapped peaks.

When crystallised with $(NH_4)_2SO_4$ we observed an increase of 8.7% of the unit-cell volume compared to crystallisation conditions containing *NaCl*. With $(NH_4)_2SO_4$ the phase obtained has significantly different lattice parameters than the published corresponding pdb model 1XY3 (3.1, p. 50 and 3.2, p. 49).

The lattice parameters for the crystallisation condition containing *NaCl* and 8% PEG8K are very close to the published pdb structure 1XY3 [32]. This phase is not very "robust". For example variations of crystallisation time affect the lattice parameters and overall quality of the diffraction pattern collected, without reproducibility. Also radiation damage occurs quickly for this phase while collecting data on ID31.

With *NaCl* the PEG concentration variation induces a continuous phase change. At 10% PEG8K we observed the coexistence of the 8% and 15% PEG8K phases.

3.3.5 Orthorhombic (*I*222) phase

The *I*222 phase is the most robust that we observed during this study. If an excess of 8-azaxanthine, an inhibitor of Urate oxidase, is present during the crystallisation the *I*222 phase is observed.

Attempts at varying pH, salt and PEG8K concentration did not significantly alter the lattice parameters. The unit-cell parameters are also essentially the same as the published pdb structure 1R51 (3.1, p. 50 and 3.2, p. 49).

Radiation damage is also lower than for the other phases and combined with very good crystallinity, this allowed us to collect our best diffraction pattern of urate oxidase.

The high counting statistics, the very good angular resolution, and high diffraction limit (in d-space) allowed us to attempt structural determination on this phase (figure 4, p. 65).

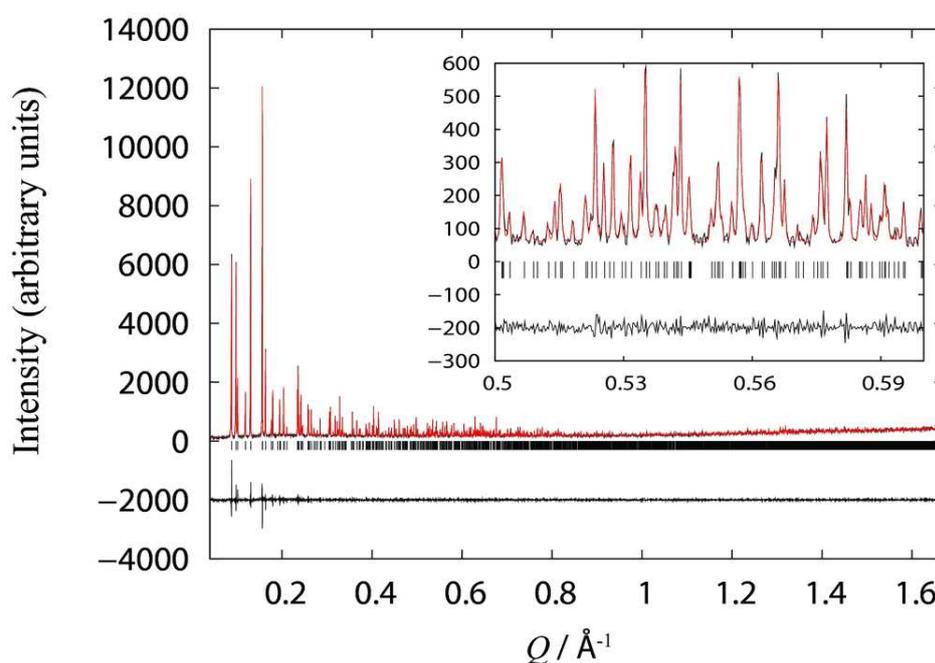


Figure 3.14: LeBail fit of Uox complexed to AZA and crystallised with *NaCl* (*I*222) (ID31: $\lambda = 1.30000(6)$ Å, 295 K).

3.3.6 Polymorphism of Urate oxidase - Conclusion

Several effects due to the presence in the crystallisation medium of salts, pH, PEG8000, presence of AZA, have been illustrated.

Several distinct phases were distinguished based on high resolution powder diffraction measurements.

When Uox was crystallised in Tris buffer or pure water in the absence of salt, a distinct polymorph of orthorhombic symmetry ($P2_12_12$) was obtained.

Cell dimensions are modified when small variations of pH and PEG 8000 concentration are made, revealing the importance of clearly defining exact crystallisation conditions for protein powder diffraction.

Without ligands but with PEG 8000 at pH 8.0, the most important factor is the salt employed for crystallisation.

It is known from the structure of Uox complexed with AZA that surface cations, not in the proximity of the active site where AZA is located, attach to specific locations on the surface of the protein [29]. The resolution of the powder diffraction data was inadequate to verify that the same phenomenon was the case for ligand-free Uox. But the salt dependent polymorphism reported here is compatible with the observed (in single crystal structures) surface cations engaged in protein packing. Changes in these salts could lead

to modification of the way the proteins pack as well as the protein contacts in the 3D network of the crystal, leading to the polymorphism observed. When no salt is present in the crystallisation medium, Uox crystallised in an orthorhombic unit cell adopting the smallest volume and solvent content observed. Small variations of the lattice parameters are observed when Uox is uncomplexed when the pH changes between 7.0 and 9.0, and the PEG 8000 concentration changes from 0 to 16 %.

When very high concentrations of PEG8000 are used for crystallisation, a collapse of the unit cell was observed (mainly a), interpreted as replacement or removal of water molecules by PEG 8000. Also here unfortunately, the resolution of patterns measured is not high enough to observe directly the phenomenon.

Also, the coexistence of two phases is reported at 10 % PEG, with lattice parameters comparable to those observed at 8 % and 15 % respectively, indicating a transition between the two forms.

As far as drug design is concerned, currently Fasturtec is a drug intended for a single injection in the case of Tumor Lysis Syndrome. In this case, a Uox lyophilisate, which is reconstituted in a hospital for an intravenous injection, is suitable. In the case where regular injections are required, it would be more efficient to develop a subcutaneous formulation. Indeed subcutaneous, contrary to intravenous, allows for self administration avoiding hospitalisation. The main limitation of subcutaneous formulation is the volume of injection (250 mL for intravenous versus 1 ml for subcutaneous), which requires a highly concentrated formulation. This study demonstrates the feasibility of production and structural characterisation of Uox microcrystalline suspensions. All the chemical components we employed for crystallisation are approved by the Food and Drug Administration (FDA) and can be used for drug formulation at isotonicity. PEG and salt may be varied to control the crystal size, shape, and solubility, hence to adjust the bioavailability.

4

Structure re-determination of Urate oxidase complexed with 8-azaxanthine

Dans ce chapitre nous présentons la re-détermination de la structure de la protéine Urate oxidase complexée avec son inhibiteur 8-azaxanthine. Cette protéine, de près de 300 acides aminés est la plus grosse structure pour laquelle un affinement structural à partir de données de poudre a été possible.

Nous avons sélectionné la phase cristallographique donnant les meilleurs clichés de diffraction. Il s'agit de la phase orthorhombique I222 d'Urate oxidase, cristallisée en présence de NaCl et de son inhibiteur. La résolution maximale de la diffraction obtenue dans l'espace réel est de 3.6Å. Nous avons utilisé une extraction simultanée des réflexions sur quatre profils de diffraction à divers niveaux de dommages dus aux radiations. Cette extraction a permis d'avoir des données suffisantes pour re-déterminer la structure par remplacement moléculaire. Le modèle utilisé pour la recherche par remplacement moléculaire est le modèle 1R51 muté en une chaîne de polyalanine. Les cartes de densités obtenues après le remplacement moléculaire ont permis de positionner certaines chaînes latérales. Les logiciels phenix.autobuild et phenix.refine ont permis l'obtention d'un modèle préliminaire basé sur les intensités extraites. Le logiciel GSAS a permis l'affinement de ce modèle en utilisant la méthode Rietveld.

Following the study of polymorphism of recombinant Urate oxidase (Uox) from *Aspergillus flavus*, we selected the phase which gave the best diffraction quality in order to perform structural refinement. The orthorhombic $I222$ phase given by recombinant Urate oxidase complexed with the inhibitor 8-azaxanthine and crystallised with NaCl was selected.

We report the solution, model building, and refinement of Uox complexed with AZA using a polyalanine model as the search model for the molecular replacement. The combined use of single crystal and powder software allowed us to refine successfully this structure using only powder diffraction profiles. The structure re-determination of Urate oxidase complexed with its inhibitor is an excellent test case to push the structure refinement limits of the protein powder diffraction technique.

The use of AZA is not currently relevant from a pharmaceutical development point of view. However this inhibitor was used in the past at low concentration as a stabilizer [34]. In this case, the use of AZA is interesting from a methodological point of view. Indeed, the use of AZA for crystallisation of Uox improves the diffraction quality of the protein crystals. To our knowledge, this is the largest protein structure re-solved and refined from powder diffraction data.

4.1 Integrated intensity extraction

Significant changes in the lattice parameters accompanied by a gradual increase of peak broadening and significant loss of intensity are common characteristics of radiation damage effects. In order to increase the counting statistics without compromising the data quality, the capillary was translated to give a fresh region of sample after 4 scans, as by the fifth scan, peak positions and widths were largely affected by radiation damage. The first scans of each position were combined together, and then the second scan of each position, etc. The maximum d-space resolution obtained for this sample was 3.6Å.

Integrated intensities were extracted by a multi-pattern Pawley [35] refinement using the program PRODD [8], which is based on the CCSL subroutine library [36], from four datasets of the same Urate oxidase sample complexed with 8-azaxanthine at different radiation damage levels. The profile coefficients, background and instrumental zero shift were optimised together with the values of the intensities of the Bragg peaks using a damped least-squares procedure [8]. In this method, each diffraction pattern was calculated as a sum of overlapping reflections, the intensities of which are variables in the least-squares procedure. Radiation-induced processes are

identified from the gradual alteration of the diffraction peak positions associated with anisotropic expansion of the three lattice parameters. Each pattern is refined with its own lattice parameters and a scale factor while identical peak intensities are kept constant between all patterns. This has been exploited in order to improve the separation of overlapping reflections and improve the accuracy of the extracted intensities [37].

4.2 Molecular replacement

The extracted intensities were imported into the CCP4i software suite [38] and converted to the MTZ format to provide input into the molecular-replacement programs MOLREP [39] and phaser [40]. We used only a reduced model, mutated to a polyalanine chain in order to perform the molecular replacement. The extracted intensities were accurate enough to obtain a reliable solution for the rotation ($R_f/\sigma = 5.17$) and translation functions, clearly distinguishing the first solution from those following (figure 4.2, p. 67). The software phenix.refine was used then to refine the position of the polyalanine model, which enabled us to locate certain side chains in the electron density maps (4.3, p. 69 and 4.4, p. 70).

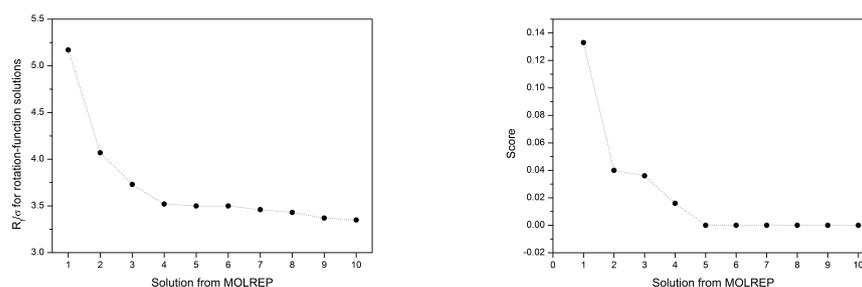


Figure 4.2: Left, R factor/ σ computed for rotation-function solutions and correlation coefficients computed for molecular replacement using model 1R51 (orthorhombic $I222$ Uox) and intensities extracted from 4 profiles using PRODD. The top ten rotation-function peaks were each used to generate ten translation-function peaks using the MOLREP program. Right, final scoring of the solutions from Molrep with a polyalanine model for the search.

CHAPTER 4. STRUCTURE OF URATE OXIDASE

68

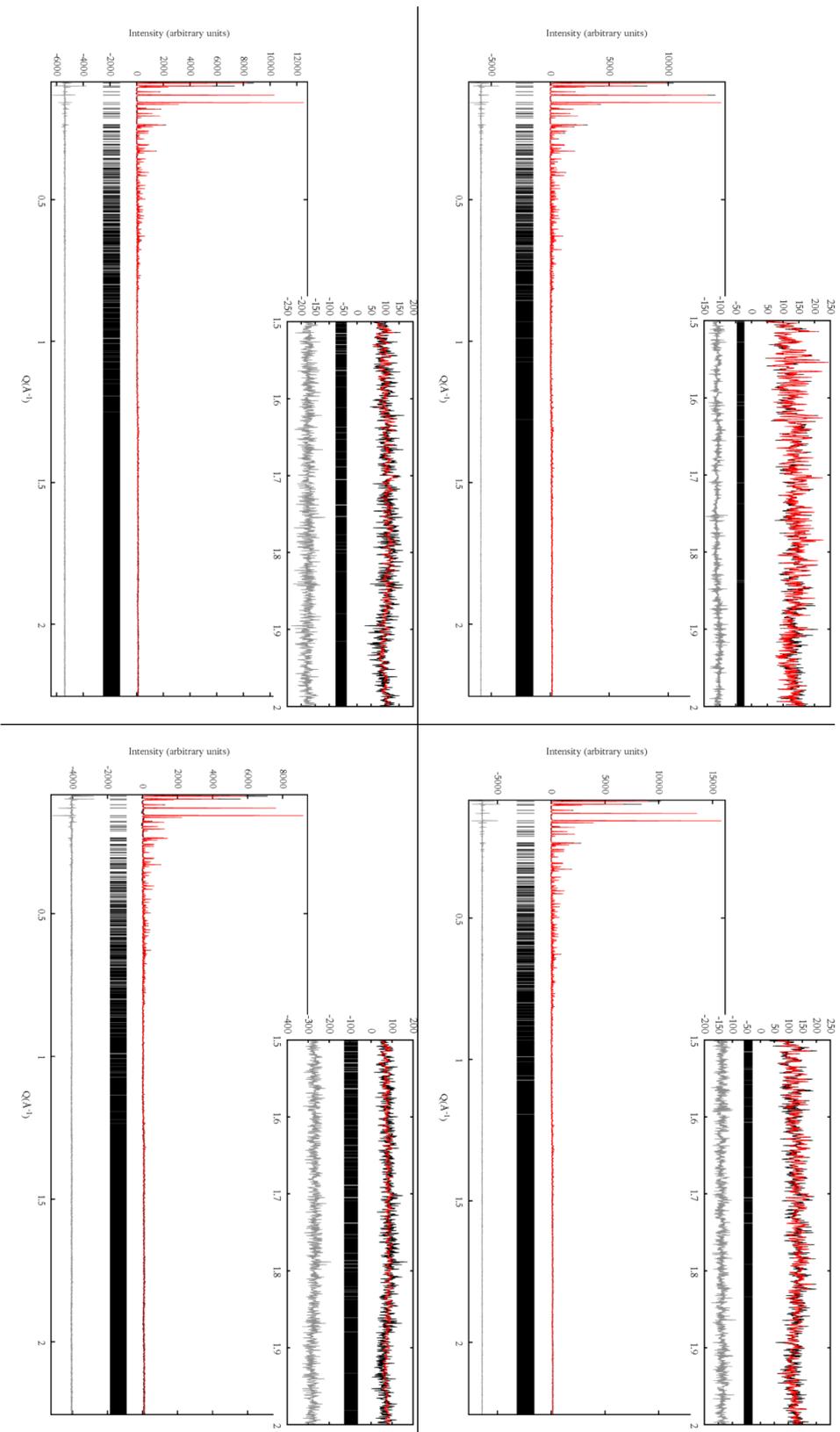


Figure 4.1: The four patterns used for the intensity extraction using the Pawley method. Different radiation damage levels of the same sample were used in the extraction.

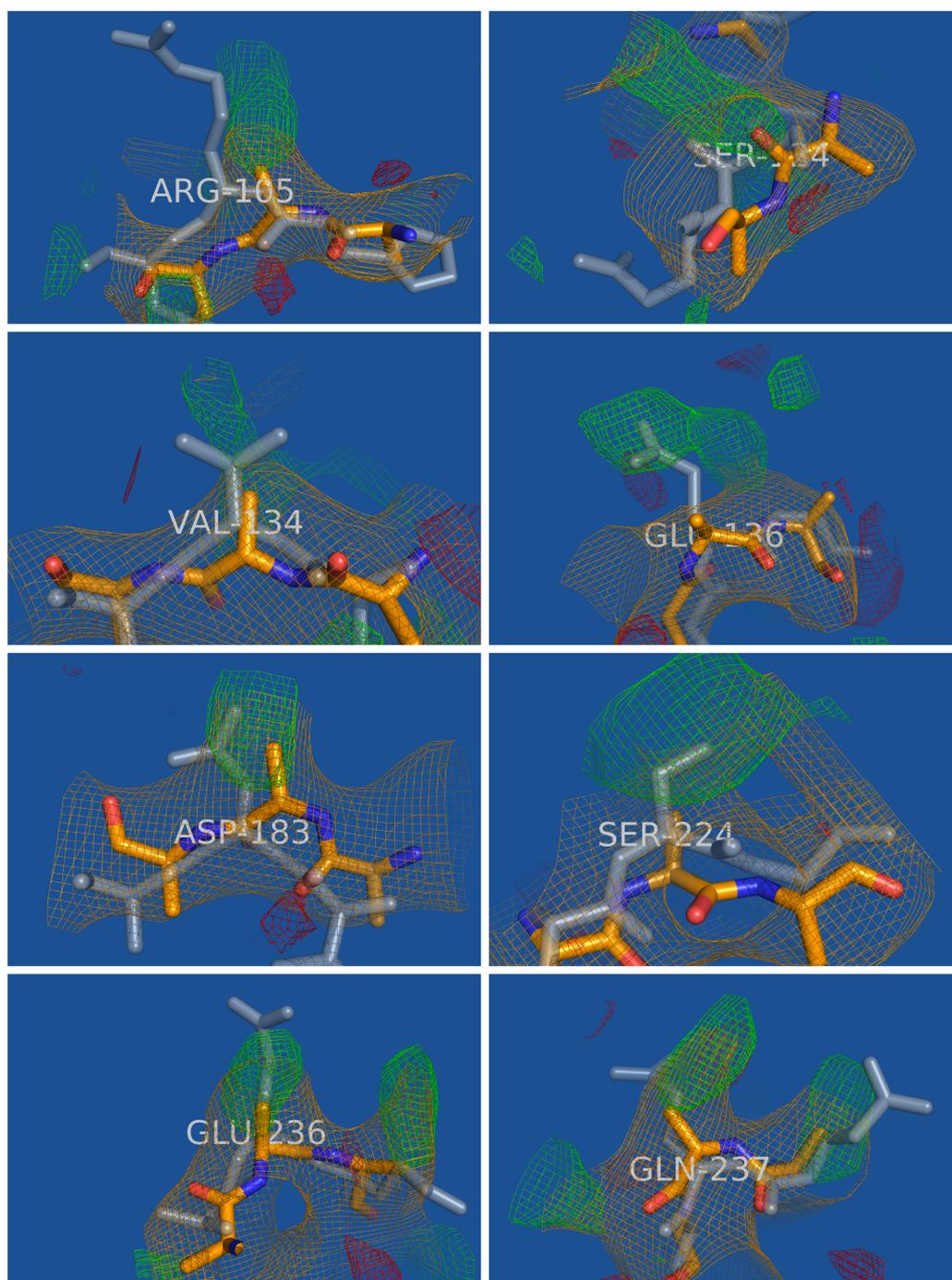


Figure 4.3: Density maps obtained immediately after a molecular replacement (with only a polyaniline model as input). It is the first time that we show some unmodelled side chains density in molecular replacement maps using powder diffraction extracted intensities. In orange : polyaniline model used for molecular replacement search, in grey : complete single crystal model (1R51). Orange map : $2 F_{obs} - F_{calc}$; green/red: $1 F_{obs} - 1 F_{calc}$ (green positive, red negative).

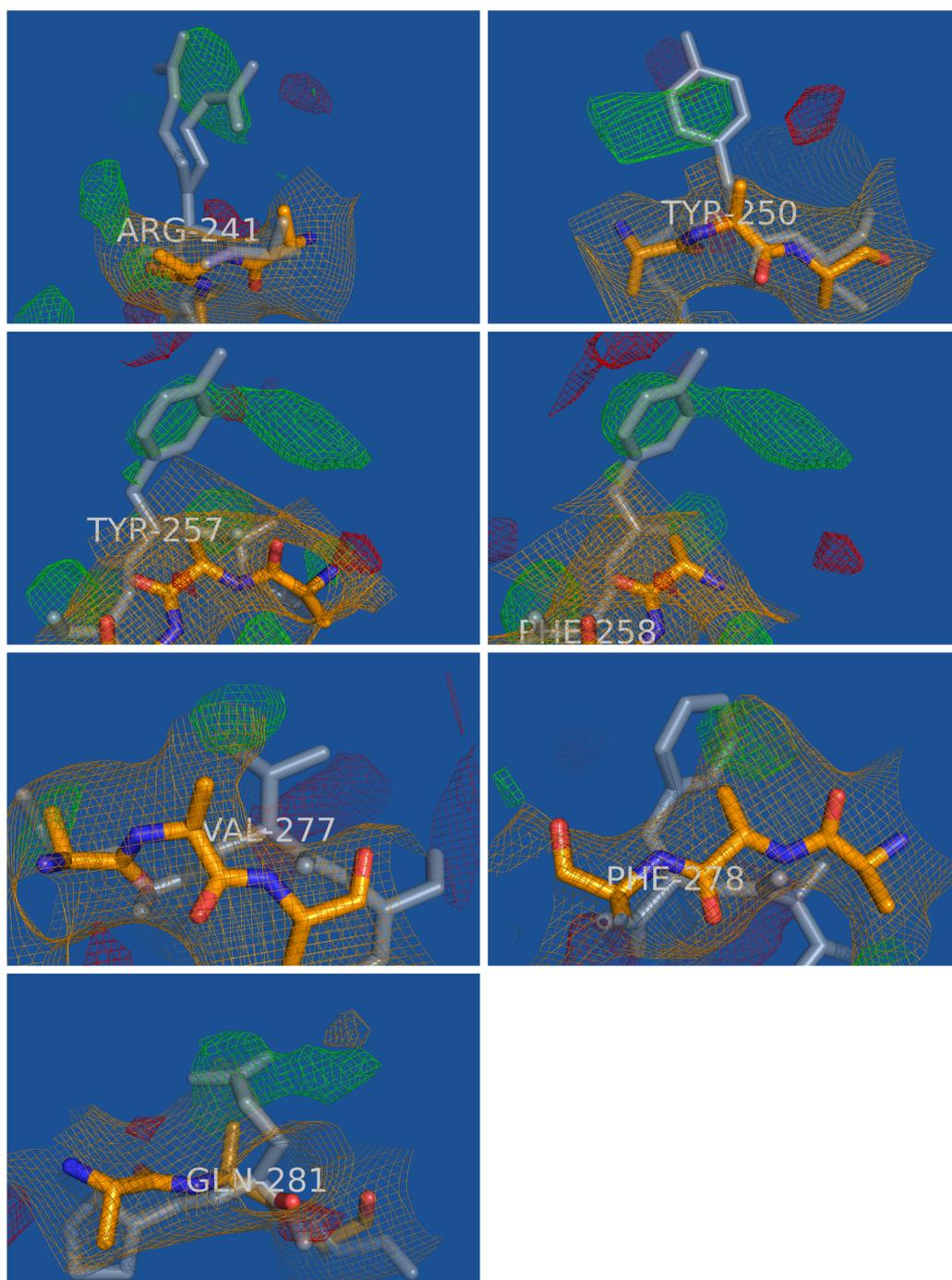


Figure 4.4: Density maps obtained immediately after a molecular replacement (with only a polyalanine model as input). It is the first time that we show some unmodelled side chains density in molecular replacement maps using powder diffraction extracted intensities. In orange : polyalanine model used for molecular replacement search, in grey : complete single crystal model (1R51). Orange map : $2 F_{obs} - F_{calc}$; green/red: $1 F_{obs} - 1 F_{calc}$ (green positive, red negative).

4.3 Refinement on extracted intensities

Rietveld refinement is more complicated and more time consuming than refinement on extracted intensities. Also a Rietveld refinement needs a very good starting model to be possible. The software `phenix.autobuild` was used on the molecular replacement solution found to get a preliminary model based on the extracted intensities. `phenix.autobuild` built the side chain residues and performed several cycles of building and refinement. Then more refinement cycles were carried out with `phenix.refine`.

The statistics for this model were $R_{work} = 0.3206$ $R_{free} = 0.4262$ and cannot be improved further using only the integrated intensities.

4.4 Rietveld refinement

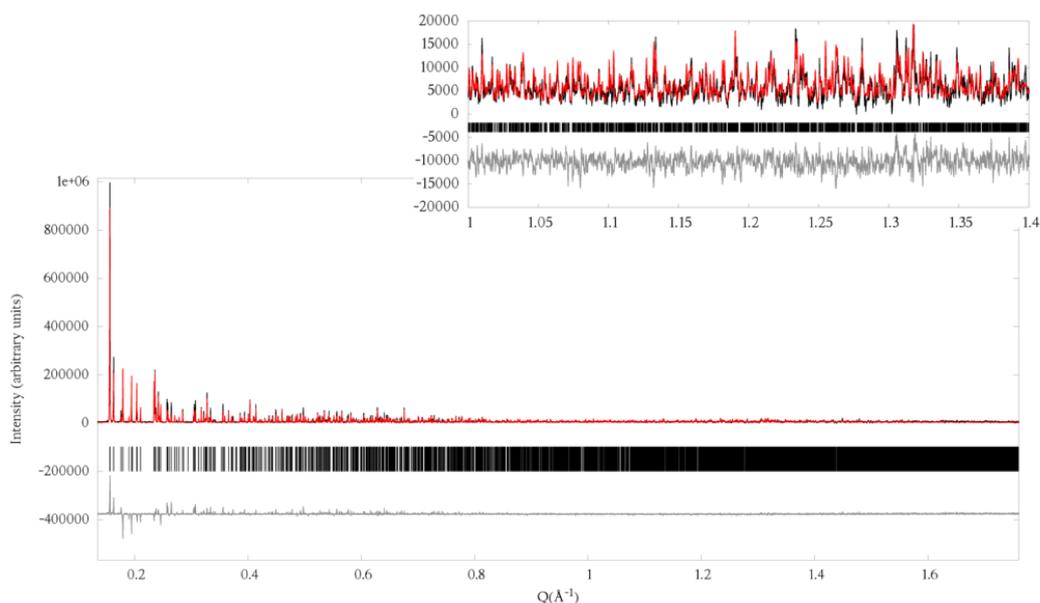


Figure 4.5: Rietveld fit of Urate oxidase refined using GSAS. $R_{wp} = 0.1967$
 $R_p = 0.1491$

Rietveld refinement was performed with GSAS. The model refined with `phenix` was imported and refined in GSAS. Care is taken to adjust the restraints weight in order to allow refinement while keeping bond distances and

angles to reasonable values. Every 20 cycles, an omit map is built and the structure modified with the visualisation software coot. Many cycles between coot and Gsas are necessary to obtain a good model.

Small differences between the single crystal and the powder model are observable (figure 4.6, p. 73). The whole protein is shifted a little in the cell and some of the side chains are slightly different. Also, it is important to note that some sheets disappeared during refinement in regions where density was less clear (figure 4.7, p. 74).

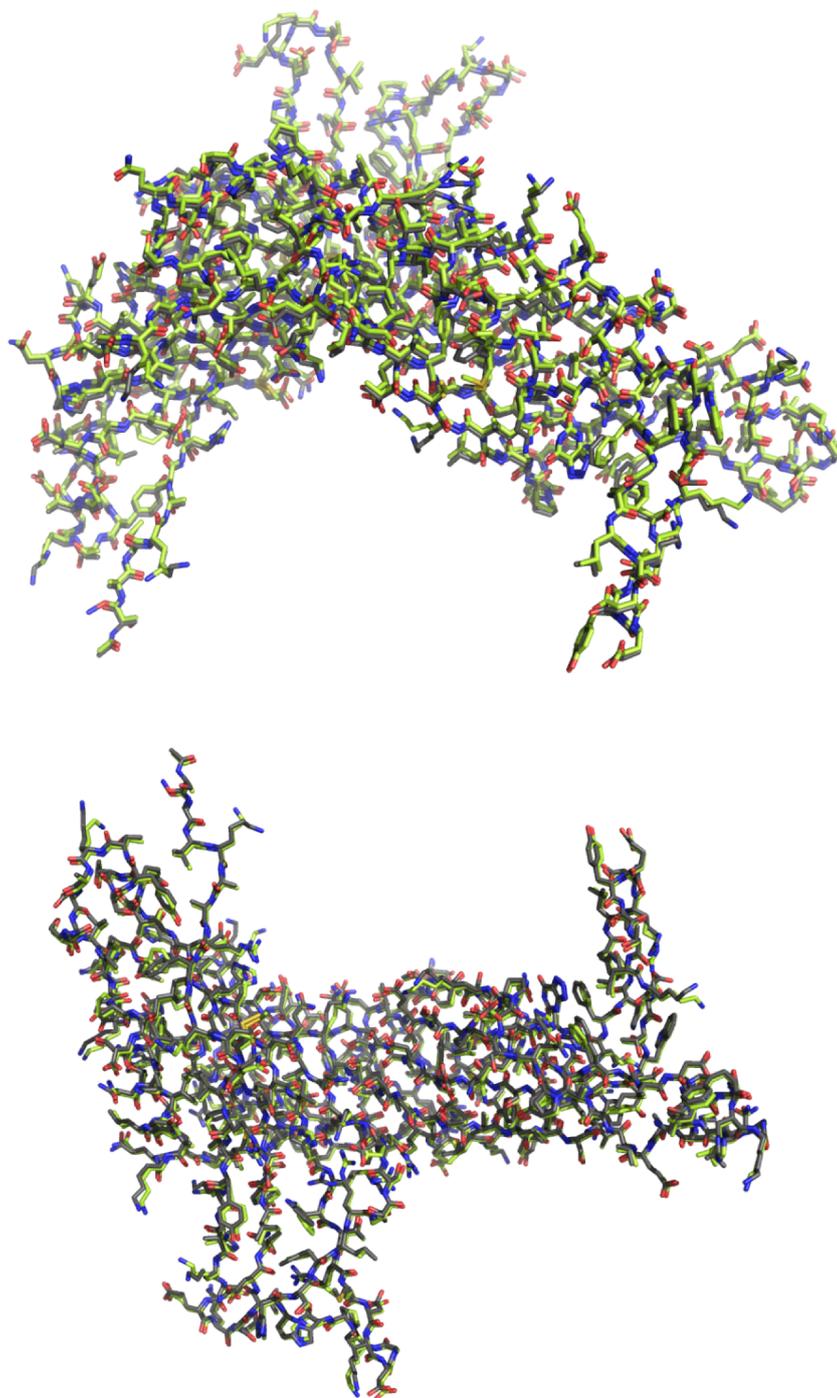


Figure 4.6: Single crystal structure of Urate oxidase 1R51 in grey, and the model refined using powder data in green. A small shift in position together with some side chain differences are visible.

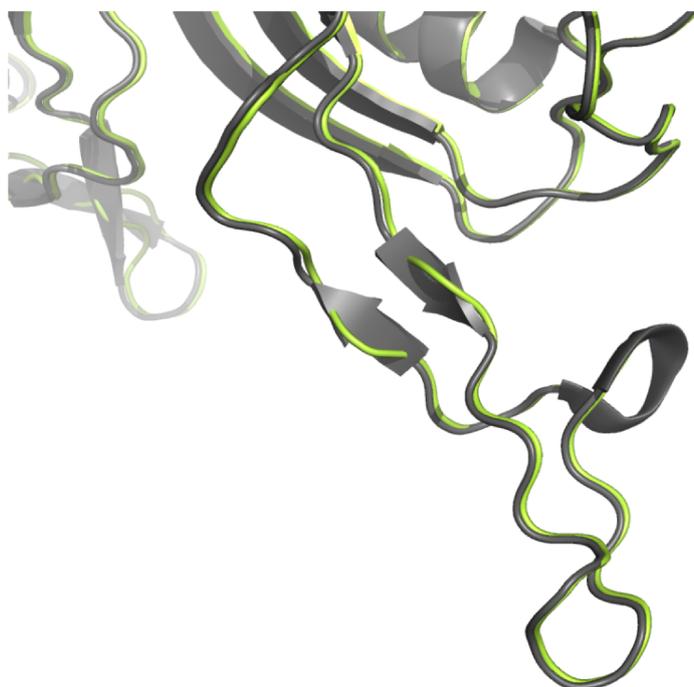
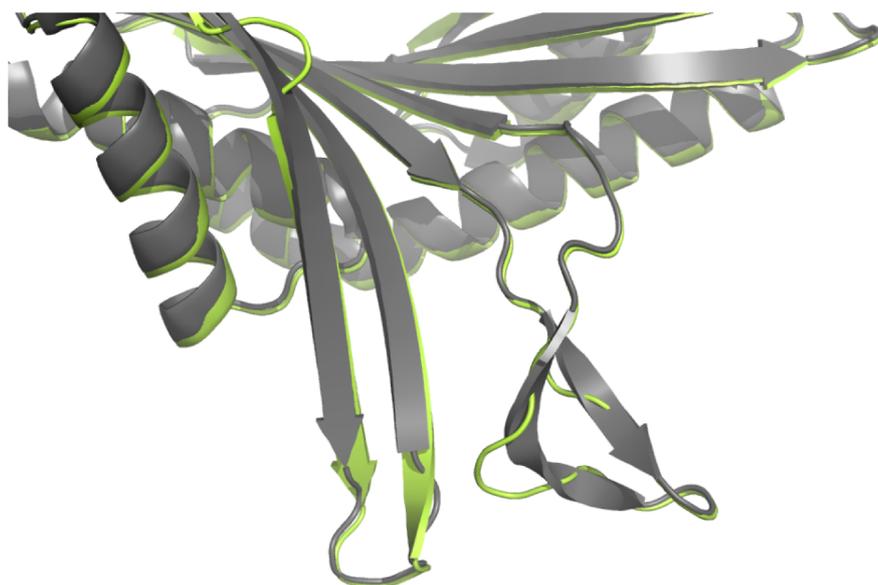


Figure 4.7: Single crystal structure of Urate oxidase 1R51 in grey, and the model refined using powder data in green. Some differences in the secondary structure appeared during refinement. The Rietveld refinement lost some β -sheets conformations and extended some others.

	before refinement	after refinement
Rwp	0.3062	0.1965
Reduced χ^2	19.40	7.991
d-space resolution	40.14 - 3.58 Å	
Number of observations	6466	
Number of refined parameters	2590	
Number of restraints	3141	

Table 4.1: Urate oxidase, current structural refinement using GSAS, using the 1R51 pdb structure as starting model.

4.5 Conclusion

We succeeded in refining a structure as large as Urate oxidase using powder diffraction. This was possible for only one polymorph, when Urate oxidase was complexed with its inhibitor 8-azaxanthine. In this case, the diffraction quality of the microcrystals was sufficient to redetermine and refine the structure at 3.6 Å. A few differences are observable between the single crystal structure and the powder structure, probably due to the difference in temperature between the two samples. The single crystal structure was based on data collected under cryocooled conditions, while we obtained our powder pattern at room temperature. Unfortunately we never succeeded in cryocooling this protein without losing d-space and angular resolution to be able to confirm that the differences observed are the consequence of the difference in temperature. The failure to obtain a cryocooled powder pattern may be caused by the large solvent channel present in the structure.

I would like to thank M. El-Hajji and F. Ragot from Sanofi-Aventis (France) for providing the rasburicase; Françoise Bonneté and Marion Giffard for the initial idea of studying Urate oxidase sample with powder diffraction; Ines Collings and Sotonye Dagogo for the sample preparation and help in the analysis of the polymorphism.

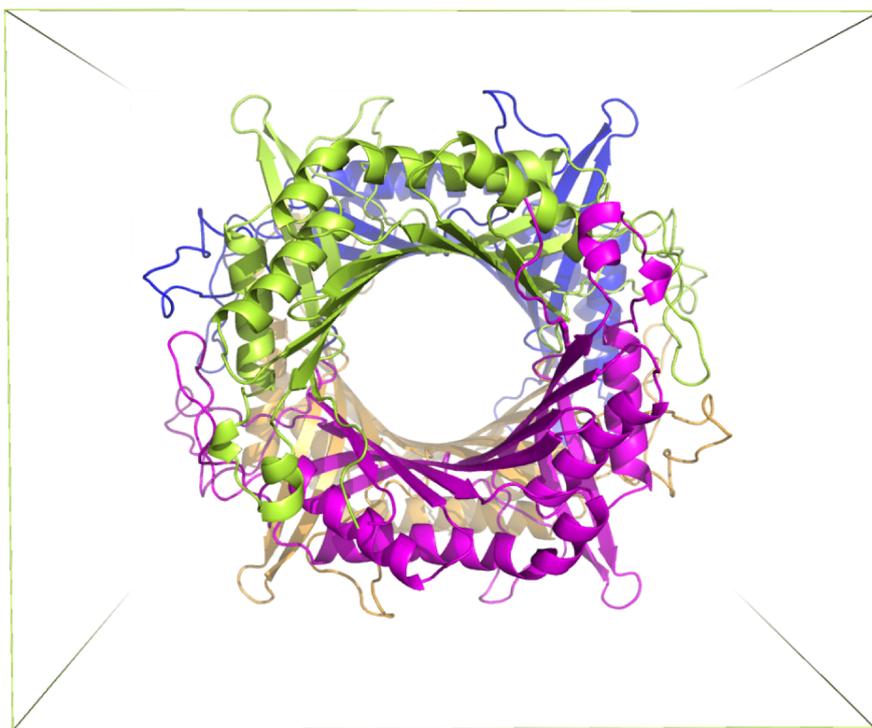


Figure 4.8: 4 molecules of Urate oxidase are forming this characteristic shape of “doughnut” around a large solvent channel. This solvent channel might be the cause of failure of the cryocooling attempts.

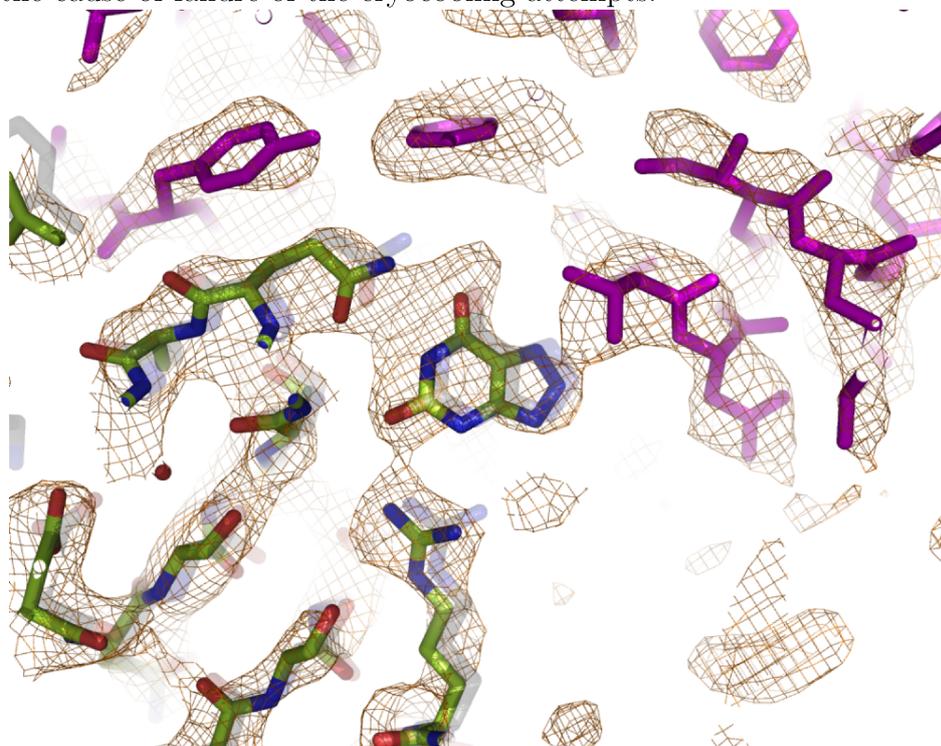


Figure 4.9: The inhibitor, 8-azaxanthine (centre) observable in the omit electron density map (orange).

5

Investigation of the nsP3 macro domain from the Mayaro virus (MAYV)

La fièvre de Mayaro est une maladie virale due à un arbovirus transmis par un moustique, Aedes aegypti, principalement en Amérique du sud. Le virus de Mayaro est très proche du virus du chikungunya et fait partie du genre Alphavirus, genre contre lequel il n'existe aucun médicament antiviral actuellement. Nous n'avons pas pu à l'heure actuelle obtenir des monocristaux du domaine macro du virus de Mayaro. Par contre, nous avons observé dans certaines conditions de cristallisation des polycristaux en forme d'oursins. Nous avons réussi à collecter, sur un unique oursin, des données de diffraction de poudre de très bonne qualité. Pour réussir à collecter sur un si petit échantillon nous avons utilisé un contrôleur d'humidité pour prévenir la déshydratation de l'oursin pendant l'acquisition de données à température ambiante. Le protocole développé pour ce projet pourra être réutilisé routinemement sur des lignes de lumière MX pour tout type de précipités cristallins en faible quantité.

5.1 Introduction

5.1.1 The Mayaro virus

The Mayaro virus (MAYV) disease is an emerging mosquito-borne zoonosis in rural areas of tropical South America. The disease is a non-fatal, dengue-like illness characterized by fever, chills, headache, eye pain, generalized myalgia, arthralgia, diarrhea, vomiting, and a rash of 3 to 5 days duration [41]. Joint pain called arthritides can stay for months and is incapacitating [42].

The Mayaro virus transmission to humans is due to tree-canopy-dwelling Haemagogus species mosquitoes, like *Aedes aegypti*, while primates like monkeys are a reservoir of the virus [43].

This infection via mosquitos, combined with the demographic changes observed in South America, make the Mayaro disease one of the major emerging diseases in the neotropics.

There are actually no effective antiviral drugs to treat animals infected by any of the alphaviruses.

Also an adaptive mutation recently allowed the Chikungunya virus to be spread by another culicidae of the same genus *Aedes albopictus* instead of *Aedes aegypti* in La Reunion island in 2005, infecting one third of the island's population and causing 327 deaths. This *Aedes albopictus*, new vector of the Chikungunya has been observed since 1999 in metropolitan France [44].

Along with this mosquito expansion, the possible risk of mutation from the Mayaro virus in a similar way to the Chikungunya and the recent emergence of Chikungunya fever in the world (Indian ocean in 2005, Singapore and Australia in 2007) has raised the alphavirus profile significantly.

Mayaro Virus	
Group :	group IV (+)ssRNA
Family :	Togaviridae
Genus :	Alphavirus

The Mayaro virus is a member of the genus alphavirus of the Togaviridae family.

The viruses of the genus alphavirus are very widespread, infecting terrestrial



Figure 5.1: *Aedes aegypti* is the vector spreading dengue fever, chikungunya, yellow fever, Mayaro virus and various other diseases.

animals, insects and fish. On earth they can be found mainly in the sylvatic area, but also in urban and peri-urban environments.

Like all group IV viruses, MAYV is a single strain RNA positive-sense virus, therefore the virus is seen by the cell as a mRNA and expressed directly. Riboviruses, RNA viruses excluding retroviruses, have a high mutation rate compared to DNA viruses, due to the absence of the proof reading ability of the RNA polymerase. This causes this group of viruses to be extremely changeable, and causes new emerging diseases like for example Chikungunya, the MAYV's closest homologue.

Like a mRNA MAYV 5'-terminus consists of a methylated nucleotide cap and its 3'-terminus has a polyadenylated tail. The MAYV genome is 11,429 nucleotides in length, excluding the 5' cap nucleotide and 3' polyadenylated tail.

A spherical 70 nm diameter particle, with a bilipidic layer containing E3 and heterodimeric protein spikes, encloses the MAYV nucleocapsid. The heterodimeric spikes consist of an E1-E2 glycoprotein heterodimer. E1 is the fusion protein for the virus entry in the acidic cytoplasmic endosomes, while E2 is covering E1 at neutral pH [47]. The enclosed nucleocapsid consists of an individual RNA and a capsid made of 240 proteins.

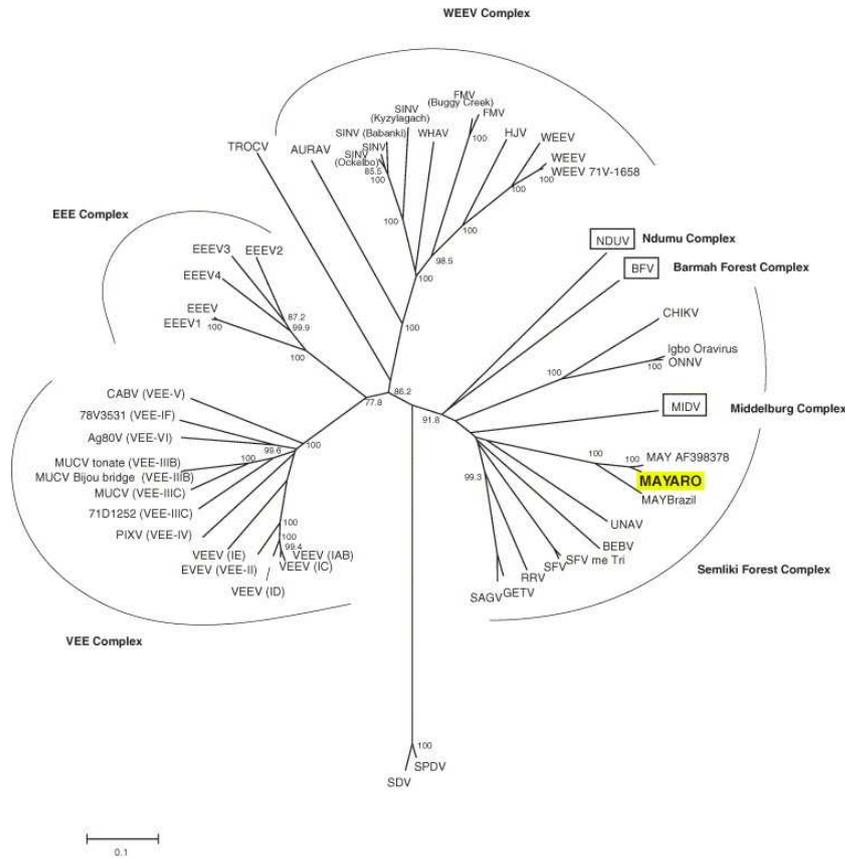


Figure 5.2: Unrooted phylogenetic tree of Alphavirus species, generated from E1 envelope glycoprotein gene sequences by using the Neighbour-Joining method [45].

Riboviruses have a very simple life cycle. The binding of the glycoprotein of the envelope allows the entrance of the virus through the cell barrier. All the replications and expressions are done in the cytoplasm of the cell.

5.1.2 The macro domain

The ADP-ribose binding process occurs in a variety of biological processes, including DNA repair, transcription, chromatin biology and long-term memory formation. Until recently, no protein domain has been successfully identified to bind ADP-ribose.

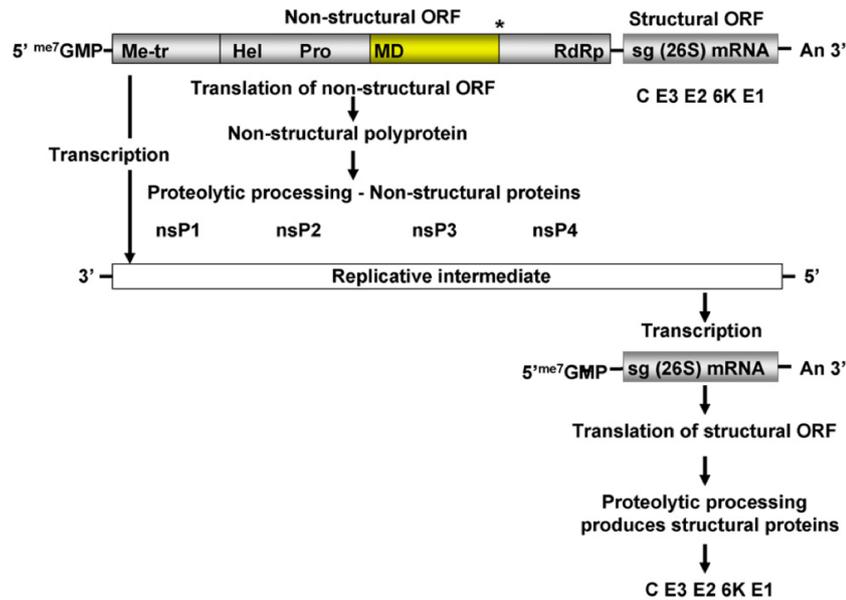


Figure 5.3: Alphavirus genome coding strategy, open-reading frame (ORF) represented as an open box, and untranslated regions as solid black lines. Me-tr (methyltransferase), Hel (helicase), Pro (protease), MD (**macro domain**), RdRp (RNA-dependent RNA polymerase), C (capsid), E (envelope) [46].

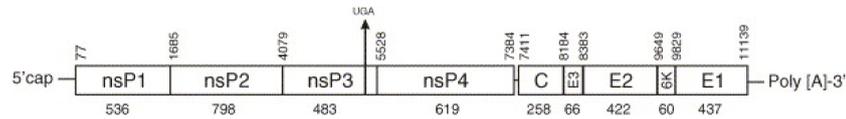


Figure 5.4: Organization of Mayaro virus genome according to [45].

The macro domains, called also X-domains, are protein domains that have been shown to have an ADP-ribose binding module [48].

These domains are conserved through evolution in various species, both eukaryotes and prokaryotes. Macro domains are also encoded by a set of positive-strand RNA viruses that replicate in the cytoplasm of animal cells, including coronaviruses and alphaviruses.

Recently this macro domain has been observed in Chikungunya Virus (CHIKV) and Venezuelan Equine Encephalitis Virus (VEEV) by single crystal crystallography [49].

5.1.3 Aim of the study

This study follows the work made by the AFMB team in Marseilles on determining CHIKV and VEEV nsp3 macro domains [49].

Another virus contains the same macrodomain, the Mayaro Virus, and the study of this domain by single crystal crystallography has not been possible as it has not been possible to grow a single crystal yet, but several crystallisation conditions result in precipitates where urchins can be observed.

This project is an excellent test case to see how far the powder diffraction technique can go on small amounts of protein sample, of proteins that are currently under research. Very small amounts of sample, typically the size of protein crystalline precipitates, are found in numerous crystallisation trial drops. Often discarded as by-products during the search for ideal single crystal crystallisation conditions, this protein crystalline precipitate can be used to obtain powder diffraction data.

5.2 Results

5.2.1 Sample preparation

Protein expression, purification, crystallisation

The protein expression and purification was done by our collaborators at the AFMB laboratory using the following protocol:

The sequence coding for the N-terminus domain of nsP3 (amino acid position 1 to 160) was amplified on cDNA of MAYV (strain TRLV4675) and cloned into the pOPIN plasmid [50] using the In-Fusion cloning method.

```

APAYTVKRADIATAIEDAVVNAANHRGQVGDGVCRAVARK
WPQAFRNAATPVGTAKTVKCDETYIIHAVGPNFNNTSEAE
GDRDLAAAYRAVAEINRLSISSVAIPLLSTGIFSAGKDR
VHQSLSHLLAAMDTTEARVTIYCRDKTWEQKIKTVLQNR

```

Table 5.1: Mayaro N-terminus domain of nsP3 sequence

The protein was expressed and purified according to the same protocol as for the CHIKV and VEEV macro domains [49].

At the end of the process, the pure protein was concentrated to 14 mg/ml in 10 mM Hepes pH 7.5, NaCl 300 mM. ADP-ribose was then added at a final concentration of 2 mM.

Initial crystallisation screenings at 20°C were performed with a nano-drop dispenser (Honeybee; Genomic Solutions) in 96-well sitting drop plates (Greiner Bio One) using three commercial crystallization kits: Structure Screen combination, Stura footprints (Molecular Dimensions Limited), and Nextal SM1 (Qiagen) using the vapour diffusion method with sitting nanodrops.

Protein powder sample preparation

From several crystallisation drops mixed together Production of crystals at a larger scale was performed in Linbro plates from two screening conditions:

- Imidazole malate 0.2M pH6, 20% (v/v) PEG 4000
- Sodium acetate 0.1M pH4.6, 40% PEG 200

Both conditions resulted in needle-like crystals of similar morphology. However, only the first crystallisation method utilising imidazole malate led to sufficient amount of crystalline material for performing powder diffraction measurements.

An adequate quantity of microcrystals, prepared with imidazole malate 0.2M pH6, 20% (v/v) PEG 4000, was achieved by mixing 40 drops made with $1\mu\text{l}$ of protein and $0.5\mu\text{l}$ of precipitation solution.

The microcrystals, loaded into a Kapton capillary tube of 2mm inner diameter, were densely packed by centrifuging. Excess liquid was removed and the tube was sealed with wax to prevent protein dehydration.

It is important to note that the mixed drops, while supposed to be identical in condition, are not necessarily producing identical crystals. A large number of parameters are to be considered at this stage. For example, small volume variations while preparing all the drops led to small differences in concentration. The time taken to seal the crystallisation well will allow different initial dehydration of the drop. Dust presence on the slide cover may help start the nucleation process. Crystal growth speed could be different in all the “identical” drops.

These differences are not a problem in single crystal diffraction, but will be in powder diffraction if we merge different drops together.

It is also interesting to note that the crystals were inbedded inside a drop’s “skin”. This skin is made of amorphous protein that did not crystallise, and makes a barrier between the drop and the air. It is difficult to fish crystals out of the drop because of the high viscosity of the skin. We took the whole drop, including the skin, by pipetting, and merged them all together in the capillary. The centrifuging moved the crystals out of the skin to pack them at the end of the capillary. It is so possible that this procedure degrades the crystals.

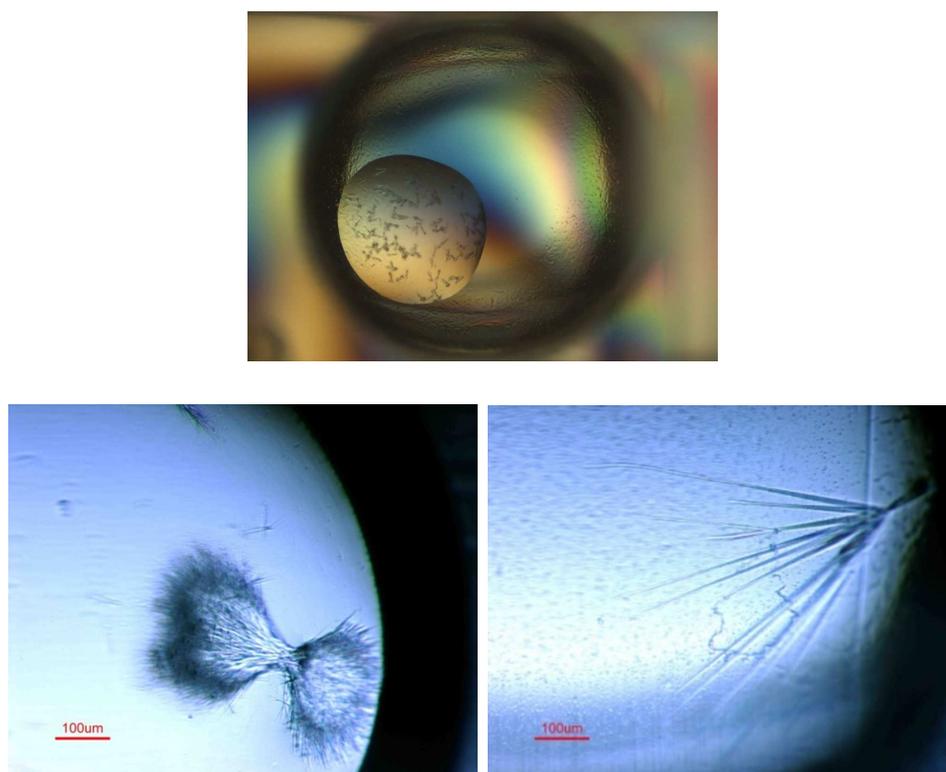


Figure 5.5: Different crystallisation drops of the mayaro virus, showing the typical urchin-like microcrystals we packed together for the ID11 and ID31 experiments.

Direct batch crystallisation To exclude variation from the different drops mixed together, and to avoid having problem from the amorphous protein from the skin, we made a sample via batch crystallisation.

5.2.2 First experiment, on ID31

The amount of sample was visually enough for ID31, but the crystallinity was low, and therefore the diffracting signal weak. Due to the weakness of the diffracted signal, 320 scans were collected. Scans were sorted into clusters using `cluster.py` (figure D.1.3, p. 194).

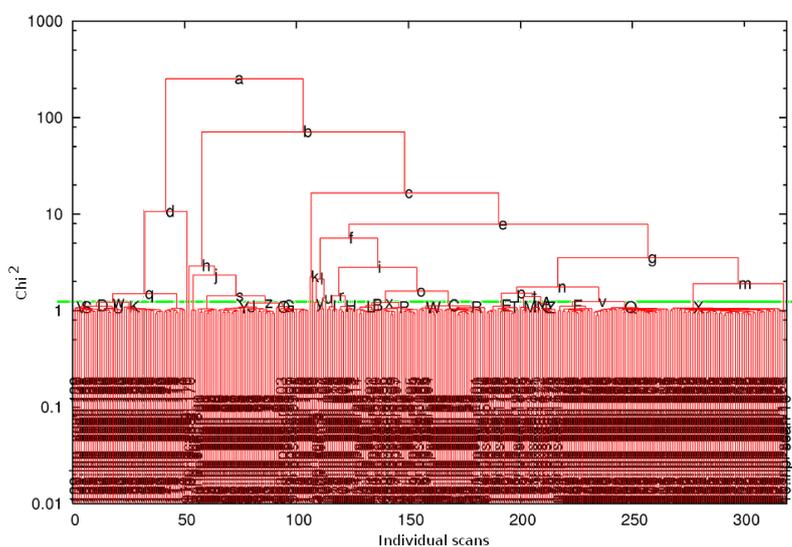


Figure 5.6: The clustering of all the patterns collected on the first attempt on ID31.

Clusters having a χ^2 score lower than the cluster label u (1.3529) (figure 5.6, p. 86) were summed together. This allows similar scans to be summed together. After examining the different summed files, the pattern having the largest number of peaks, cluster “l” summed over 6 scans, was used for indexing.

The data, characterised by restricted d-spacing resolution, low signal-to-noise ratio, increased background and a limited number of diffraction peaks, are indicative of low sample crystallinity. Using positions extracted for the

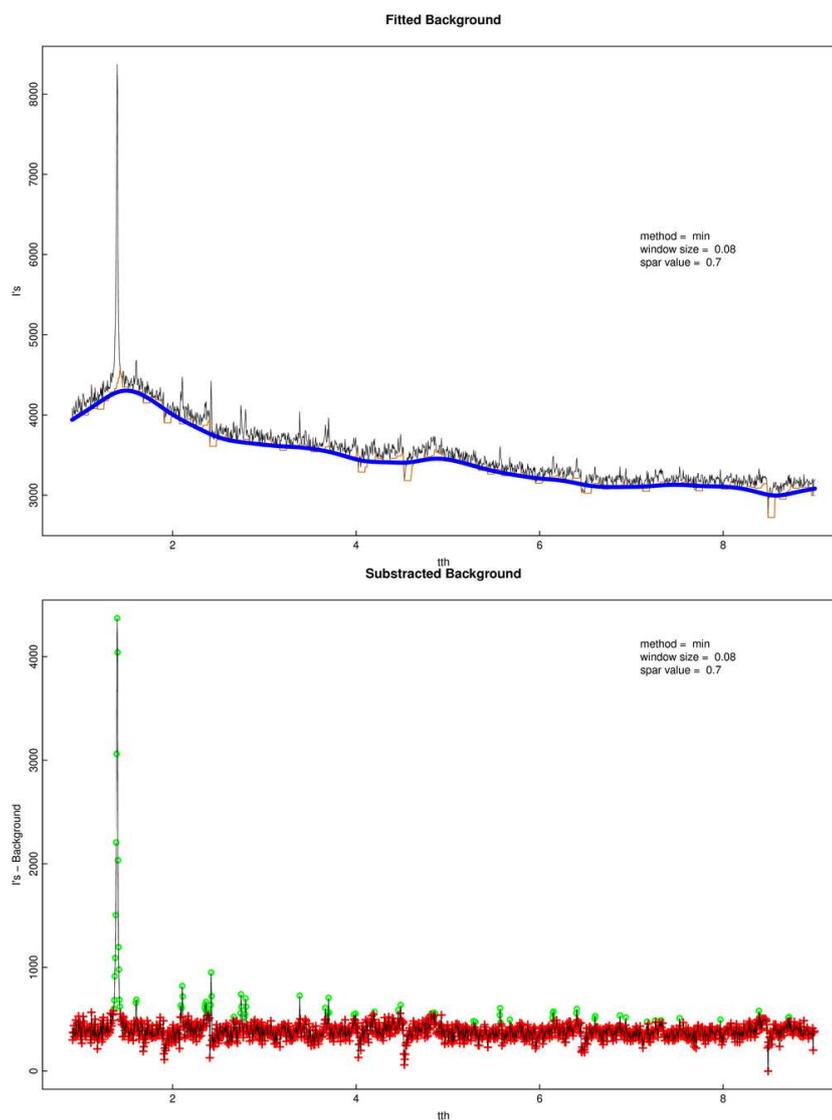


Figure 5.7: The first pattern collected on ID31 of the mayaro virus nsp3 macro domain. Beam size is $2 \text{ mm} \times 1 \text{ mm}$, $\lambda = 1.29983(2) \text{ \AA}$. The background is subtracted using the method described in appendix E, p 207 (using a window size of 0.08° and a smoothing parameter value of 0.7).

first 19 diffraction peaks a trigonal/hexagonal unit cell with dimensions of $a = 61.503 \text{ \AA}$, $c = 94.488 \text{ \AA}$ was found using DICVOL [51] with figures of merit $M(19) = 9.5$ and $F(15) = 55.3$. Owing to the limited information content of this profile, additional data were necessary in order to verify this result and continue with our analysis.

5.2.3 Second experiment, on ID11

The sample used for the ID11 experiment was produced using the batch crystallisation method directly inside a 0.5ml eppendorf tube. The sample was left for two months at 4°C.

Then, the crystals were gently packed by centrifuging inside a 0.8 mm diameter glass capillary.

The beamsize was $120\mu\text{m}^2$, silver behenate (AgBeh) was used for calibration [52], two energies were used, 42 keV ($\lambda=0.2952$) and 35.5 keV ($\lambda=0.3492$).

The first pattern collected on ID11 at 42 keV was indexed with TOPAS as $P3_1$, using the procedure which excludes non chiral spacegroups (figure F.2, p. 216). This confirmed the sample quality, and encouraged, we changed energy to 35.5 keV and moved the beamstop forwards to obtain a lower background.

Comments on the indexing

With powder, space group identification can be difficult as different space groups have identical hkl 2θ positions. In this case, as $P3_1$ was indexed, other space groups, $P3_2$, $P3_112$, $P3_121$, $P3_212$, $P3_221$, $P6_2$, $P6_4$, $P6_222$ and $P6_422$ have identical extinctions.

Meaningful values of Matthews coefficient and solvent content are obtained only if the protein crystallises in $P3_1$, $P3_112$, $P3_121$ or $P6_2$. In $P3_1$, the highest probability corresponds to two molecules in the asymmetric unit (asu) whereas one molecule per asu is expected in $P3_112$, $P3_121$ or $P6_2$ with a Matthews coefficient of 2.94 for crystals with 58% solvent content. Thus, the trigonal/hexagonal unit cell should consist of six protein molecules. We note that at this stage of analysis the exact symmetry cannot be deduced from powder diffraction as all four space groups have the same reflection condition ($00l, l=3n$).

A possible solution for the $P3_1$ phase

In order to differentiate the different possible space groups a series of molecular replacement trials on all these space groups was attempted using the softwares MOLREP[39] and PHASER[40] and all non-clashing solutions found in these space groups modelled as a powder pattern using TOPAS.

The only one giving a sensible fit was for a solution in $P3_1$. A rigid body analysis was done on it and the resulting fit shows agreement. Nevertheless the low angular resolution of the diffraction data and resolution limit (6.5\AA) didn't allow one to conclude firmly on the solution found.

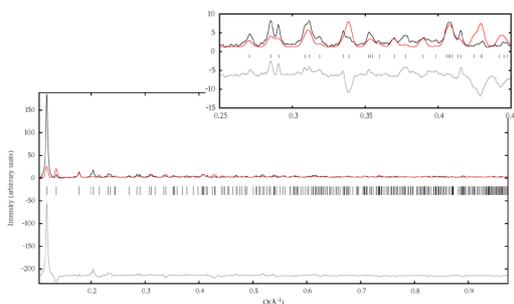


Figure 5.8: Rietveld fit of a molecular solution found for the $P3_1$ space group using phaser. We refined 3 variables : a scale factor, 2 parameters for solvent scattering.

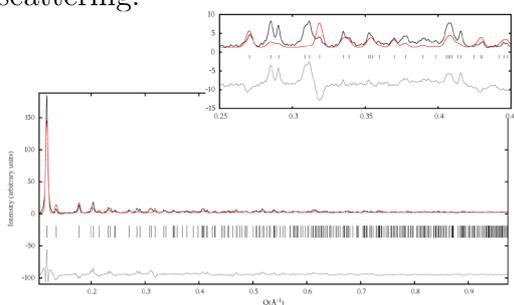


Figure 5.10: The same solution after a rigid body Rietveld refinement in TOPAS. We refined 9 variables : a scale factor, 2 parameters for solvent scattering, 3 variables for the orientation of the molecule in the cell, 3 for the rotation around the centre of mass of the molecule.

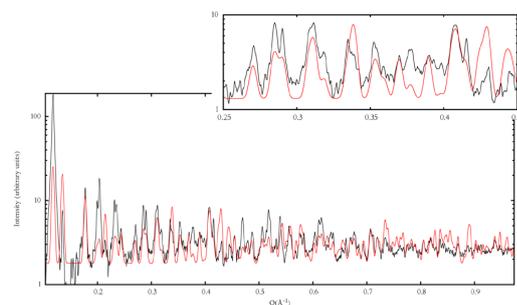


Figure 5.9: Idem fig.5.8 on log scale for the intensities.

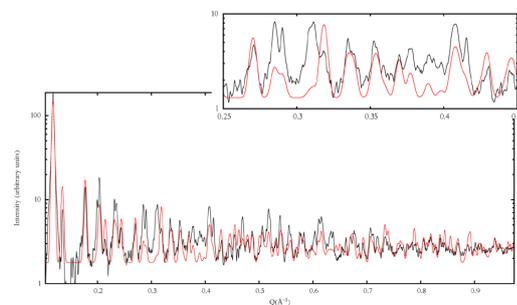


Figure 5.11: Idem fig.5.10 on log scale for the intensities.

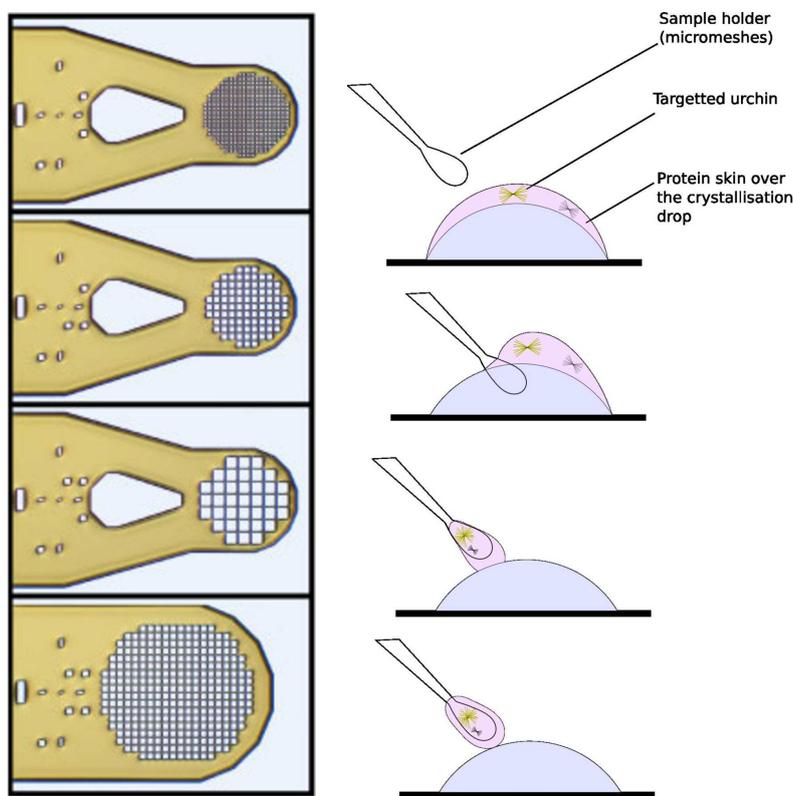


Figure 5.12: Increased difficulty in fishing the sample was due to the viscous skin of amorphous protein over the drop where the crystals were embedded. The skin was inseparable from the urchin and enveloping it on the micromesh.

5.2.4 Third experiment, on ID14-eh1

Method developed for acquisition

A third series of experiments was attempted on ID14-eh1. The main idea here was to collect on the smallest volume of sample as possible. As the studied protein crystallises in the shape of urchins, collecting on only one urchin seemed ideal. In order to realise this, we kept the urchin hydrated with a humidity controller (figure C.1, p. 183). This method was tested on glucose isomerase microcrystals, and showed that it was possible to get very good information, enough to phase the structure by molecular replacement (figure C.1, p. 184). To fish and keep the sample centred in the beam we used Mitegen MicroMeshesTM of 400 μm in diameter and 25 μm opening.

Unexpectedly good diffraction were observed compared to the previous

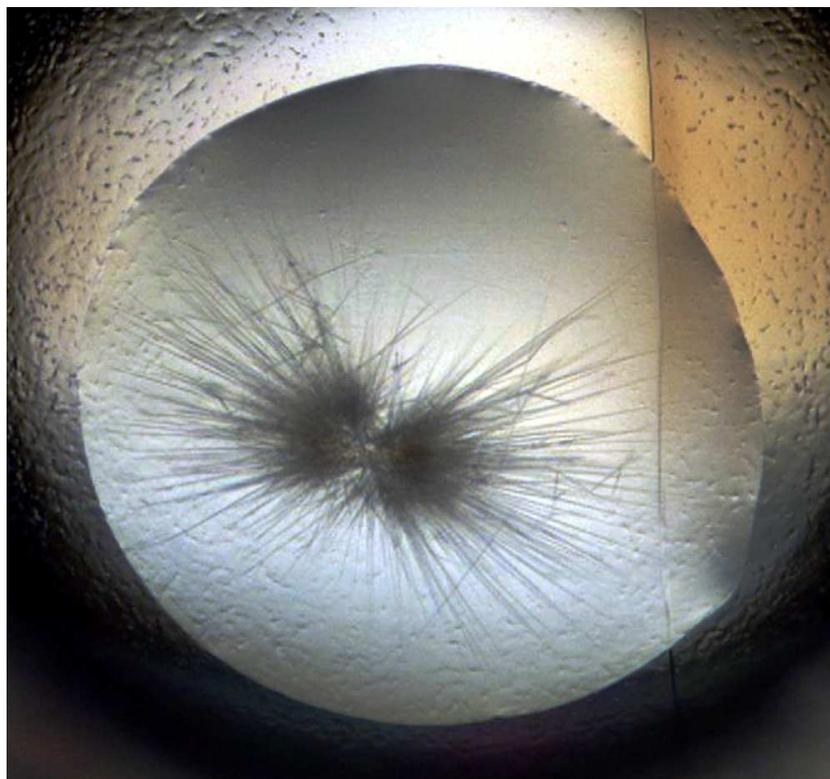


Figure 5.13: A crystallisation drop of the mayaro virus, showing the typical urchins. Crystallisation conditions are Imidazole malate 0.2M pH6, 20% (v/v) PEG 4000. The drop is made by mixing $0.5\mu\text{l}$ of mother liquor and $1\mu\text{l}$ of protein. When fished out, the drop's skin envelops the urchin which loses the outer needles, so on the micromesh only the cores remain.

experiments and considering the extremely low amount of material.

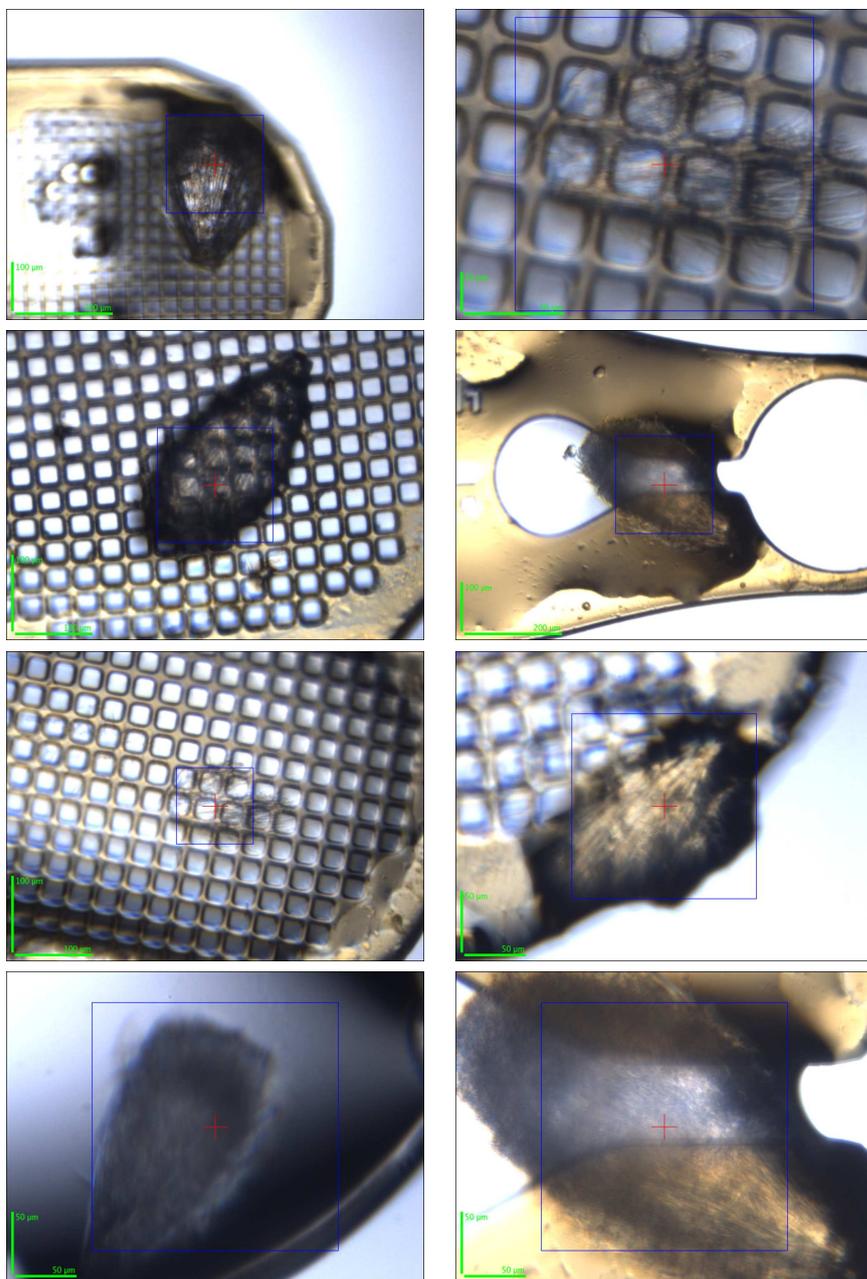


Figure 5.14: Mayaro crystals in the shape of urchins mounted on micromeshes kept to a 99% humidity in air to avoid dehydration.

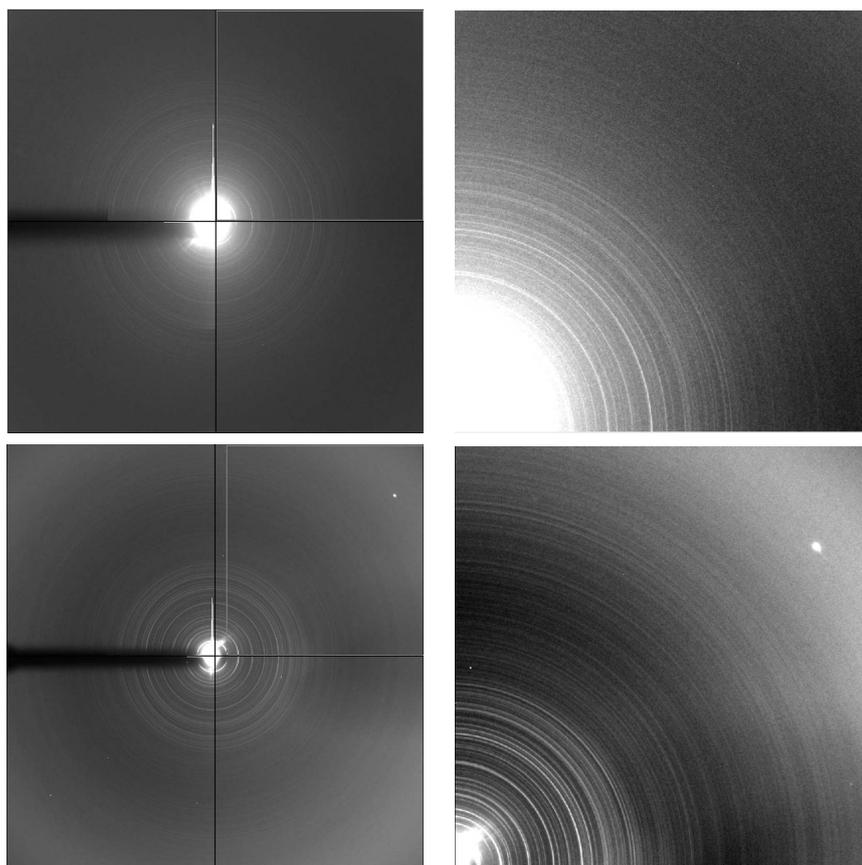
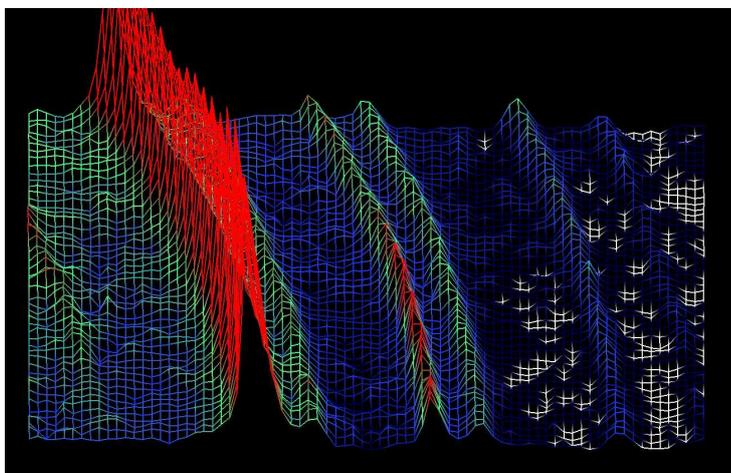
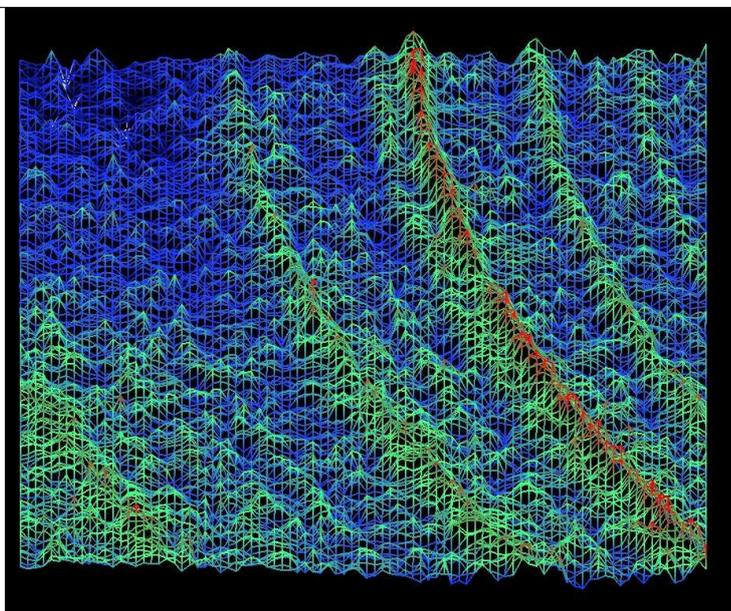


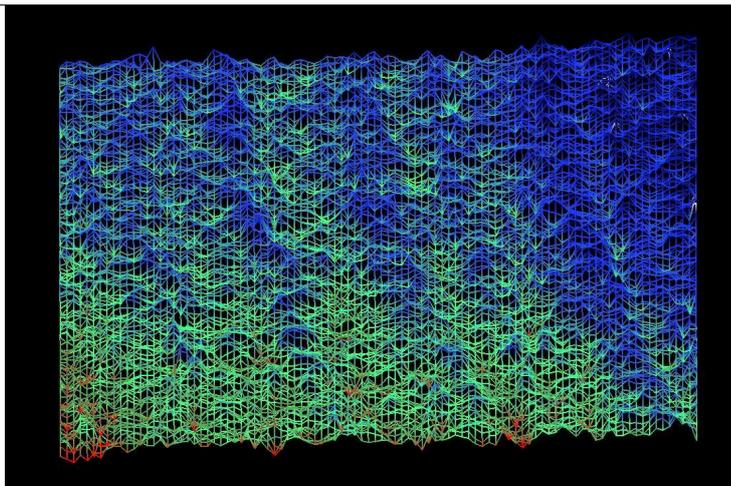
Figure 5.15: On the left, two images taken on the same urchin, at two different beam positions, with two different sample to detector distances. On the right, a zoomed region of a quarter of the detector. On some images, a clear spot of a salt single crystal is seen (and will be excluded).



At low angles, the rings are clearly separated, pixel size is adapted to the peak width. But less pixels are available for the radial integration.



At medium angles peaks are still well separated.



At high angles, peaks are merged together due to the low angular resolution of the detector.

Figure 5.16: 3D views of zoomed regions of the detector. Colour depends on the measured intensity, at low angles the rings are well separated, at high angles several peaks are measured by the same pixels, this is the drawback of using 2D detector instead of high angular resolution instruments.

Indexing

Powder patterns were obtained from the 2D images by radial integration using Fit2D. Indexing was performed using TOPAS on a restricted set consisting of chiral spacegroups (figure F.2, p. 216). The best answer gave a goodness of fit of 33.62. For this urchin, the indexed space group solution found was $P6_1$ with $a=61.642483$ and $c=95.540131$. This solution is very close to another possible $P3_1$ solution, except for one differentiating peak. While the low angle peaks for the $P6_1$ solutions are all observed, one peak corresponding to the 003 of the $P3_1$ possible solution is not observed (figure 5.17, p. 96).

As this peak was clearly observed in the ID11 powder pattern collected previously, it is possible that this urchin was made of a different crystalline phase than previously. On ID11 batch crystallisation was used for obtaining the crystals, whereas on ID31 the usual hanging drop method was used. This is another case of possible polymorphism when preparing protein powder microcrystals.

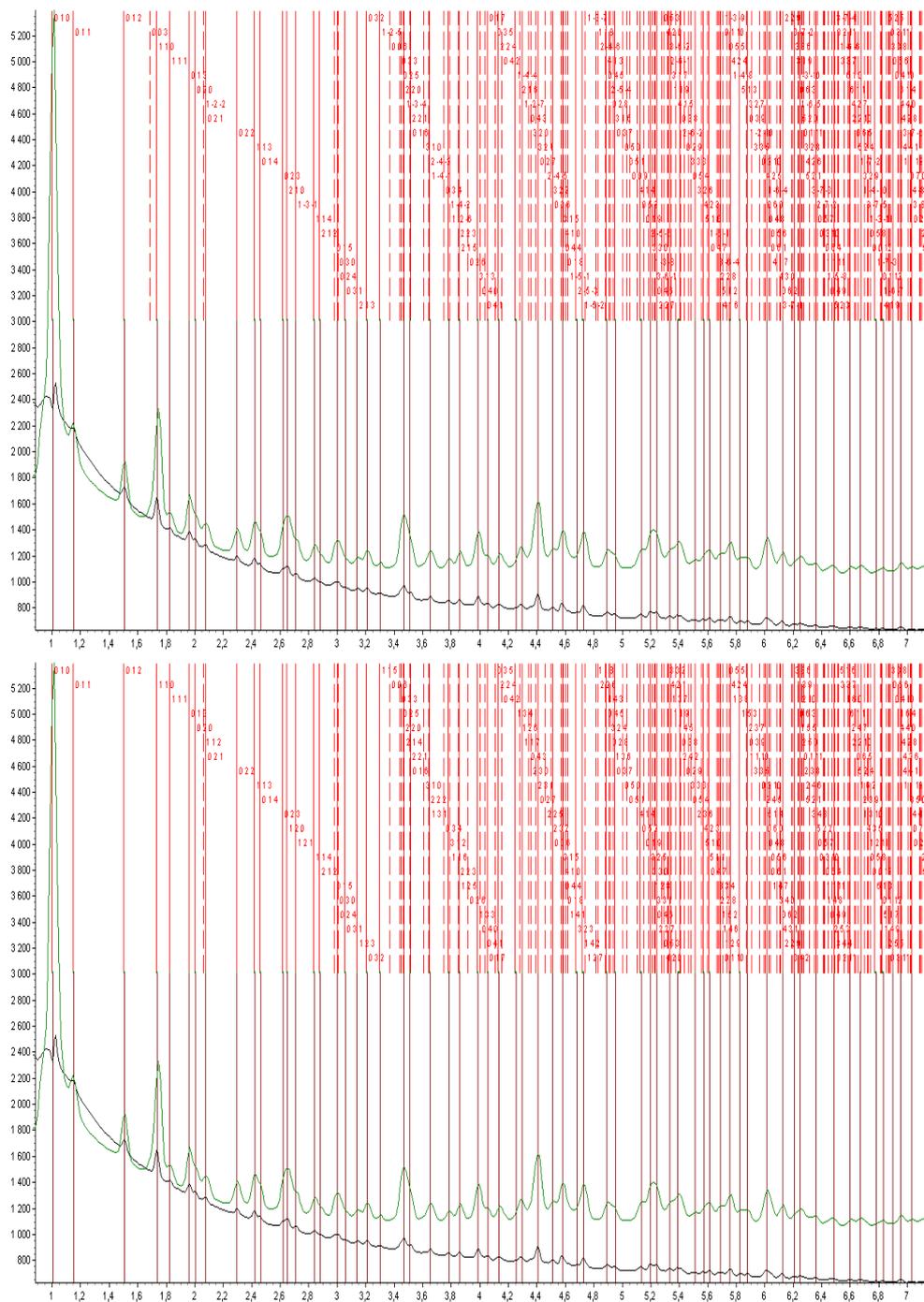


Figure 5.17: Indexing of two patterns collected on ID14-eh1 on a single urchin of mayaro virus with two different detector-sample distance (for the green curve detector distance is 659.180 mm, for the black curve detector distance is 434.485mm). On top is the $P3_1$ solution as found in previous experiments, below is the $P6_1$ solution. The upper dark red lines are the positions of the peaks used for the indexing procedure. The lower light red are the computed lines for the solution. The dashed red lines are the theoretical positions of peaks which have not been input for the indexing (unobserved peaks). This is a view of all the peaks used for the indexing.

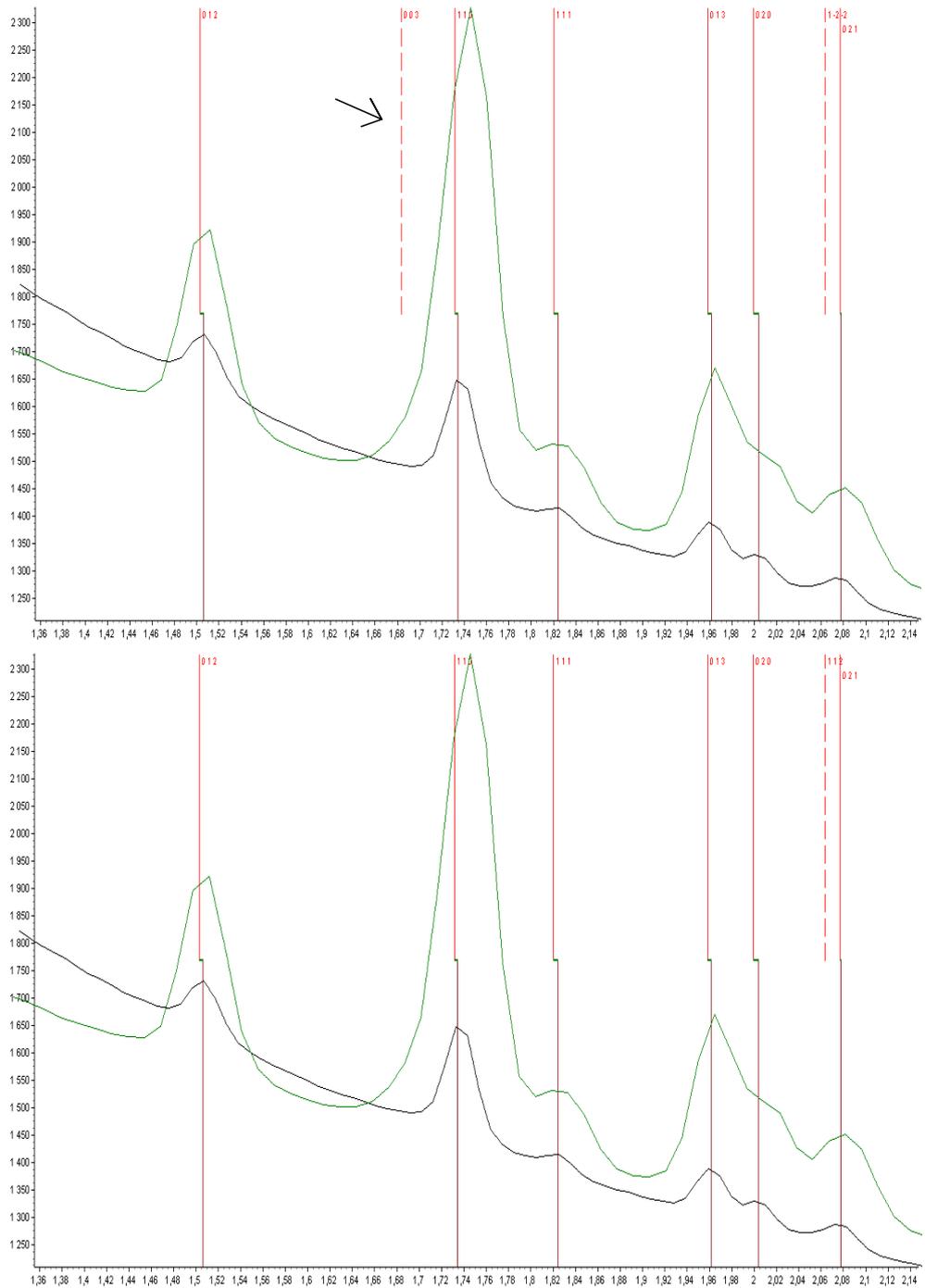


Figure 5.18: This is a zoomed view of 5.17 showing the first 6 peaks used for the indexing, to show the difference between the $P3_1$ solution and the $P6_1$ solution. In $P3_1$ the 003 peak though predicted is not observed. Its presence would exclude the $P6_1$ for which the extinction condition $00l, l=6n$ applies. Therefore this absence doesn't allow to exclude $P3_1$, but suggest $P6_1$.

Preliminary structure of Mayaro virus macro domain.

Molecular replacement was attempted using the softwares MOLREP[39] and PHASER[40] on the extracted intensities on two images collected on the same urchin at two different positions (to avoid radiation damage) and with two different detector distances. When the detector is close to the sample, information at high angle is obtainable. When the detector is further away from the sample, the integration of the low angle peaks is better (more pixels).

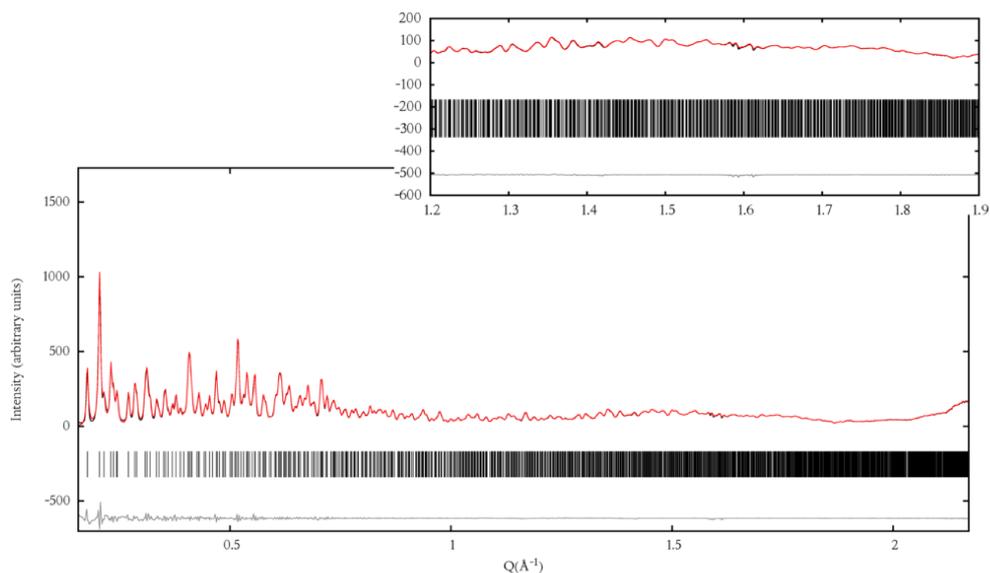


Figure 5.19: The first of the profiles used for intensity extraction using PRODD[8].

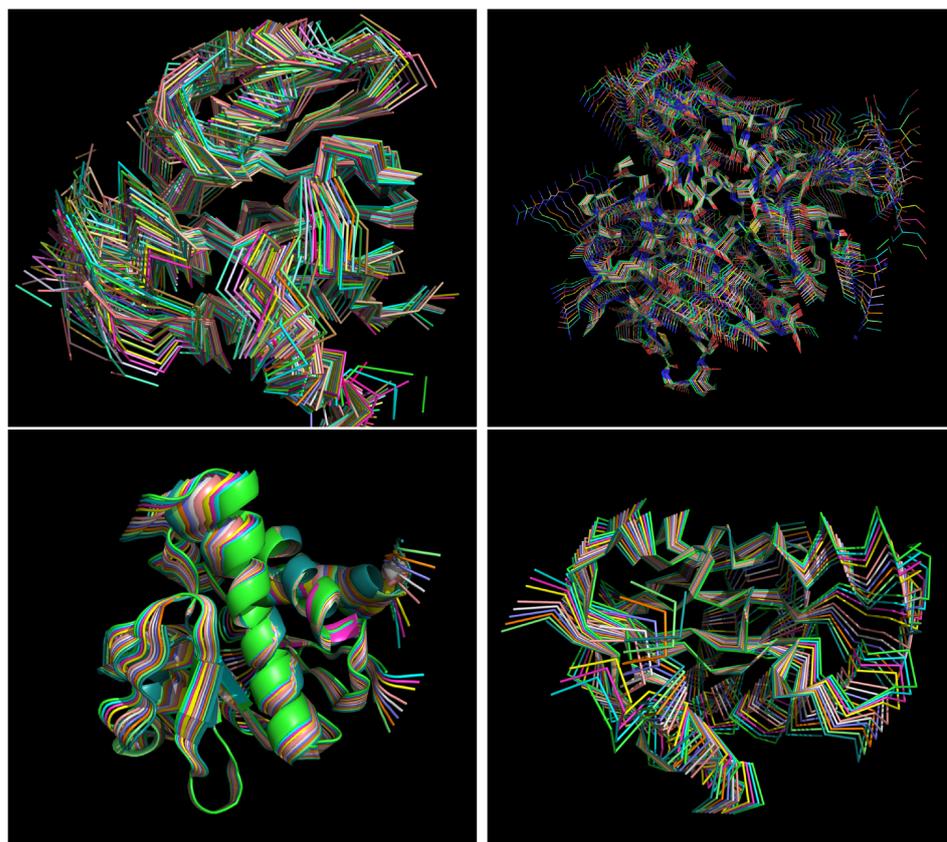


Figure 5.20: The normal mode models computed based on a mutated 3GPG model. A first model based on the chikungunya 3GPG structure is mutated to correspond to a mayaro sequence. Then normal modes 7/8/9/10/11 with perturbation amplitude between -100 and +100 (step=20) are computed using the Elnemo web server [53]. Top left images represent all the models (all modes) used for the molecular replacement search. The three others are different representations of the search ensembles giving the best score during the molecular replacement search.

We tried to solve by molecular replacement in the $P6_1$ spacegroup using a model based on the chikungunya virus macro domain (PDB code : 3GPG[49]) with residues mutated according to the mayaro sequence. The best molecular solution found with the software Phaser was by using a low frequency perturbed model (normal mode 8), the log-likelihood gain obtained was 57.830.

Based on the phaser solution, a preliminary model was built using phenix.autobuild from the extracted intensities, R-work = 0.3148 R-free = 0.4079. Then this model was refined using phenix.refine in order to save time for the Rietveld refinement. phenix.omit.maps were used to build omit maps used for the building of the structure during the Rietveld refinement performed with Gsas.

The actual model refined using the Rietveld method is not complete and suffers from the lack of angular resolution of the detector (figure 5.16, p. 94). Also, new experiments are under preparation to validate the molecular replacement solution by experimental phasing techniques.

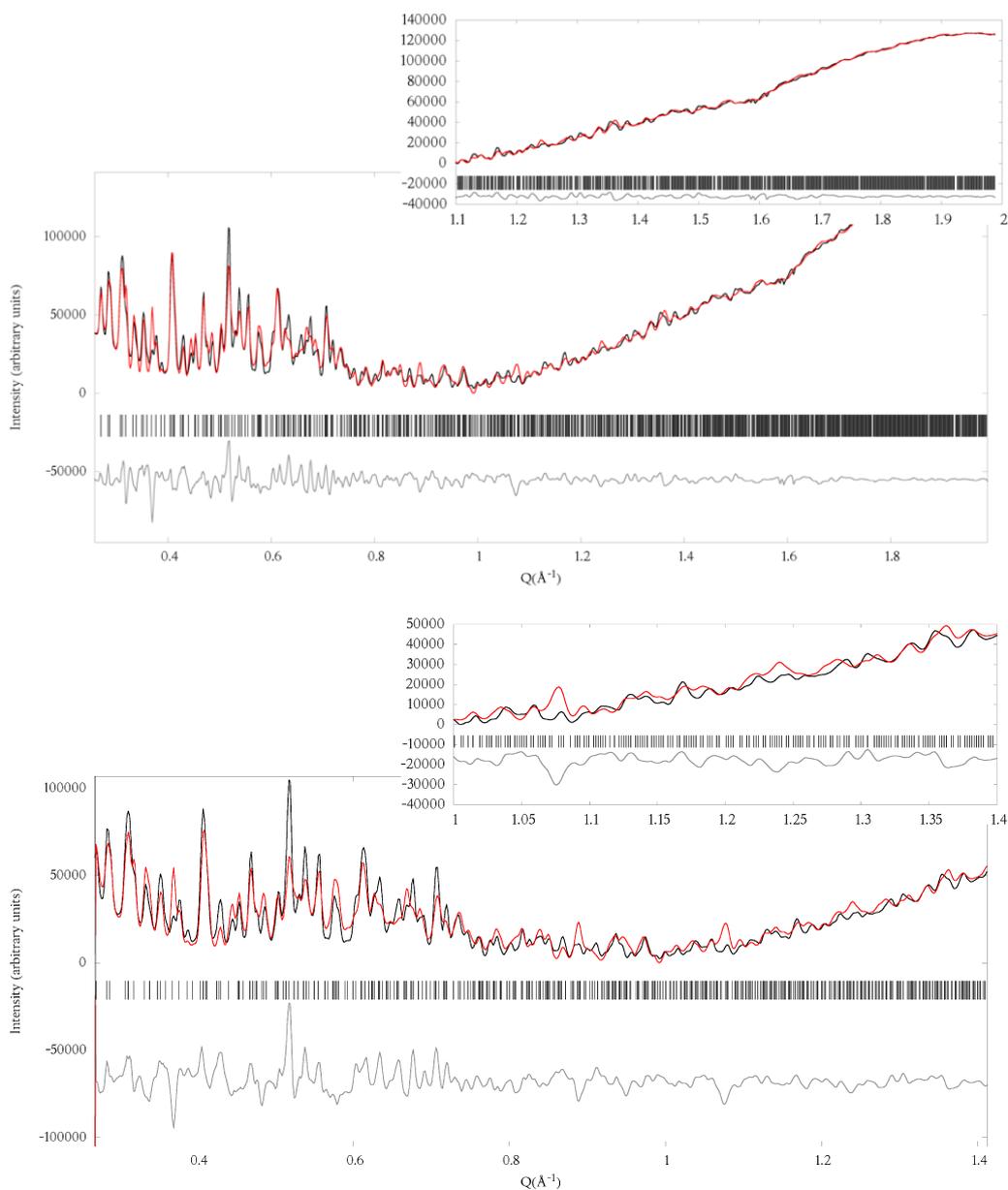


Figure 5.21: Actual fit of the Mayaro virus macro domain model. The model was built using phenix.autobuild, then phenix.refine and GSAS for the Rietveld fit. It is clear by looking at the difference curve that the model is not complete. Further work needs to be carried on to validate and finish this model.

	after molecular replacement	current refinement
wRp	0.0319	0.0244
Reduced χ^2	1119	659.7
d-space resolution	23.3 - 3.22 Å	
Number of observations	1571	
Number of refined parameters	1241	
Number of restraints	1243	

Table 5.2: Mayaro structural refinement with GSAS, comparison between the molecular replacement solution and the current model.

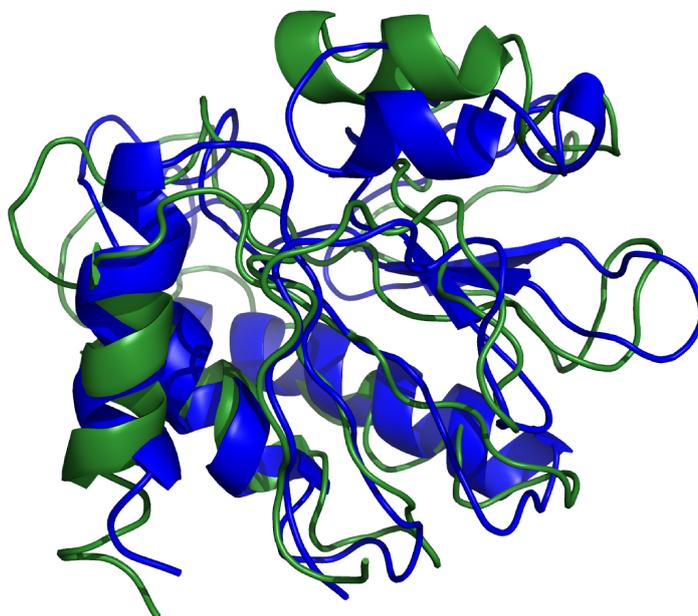


Figure 5.22: Incomplete model of the mayaro virus macro domain. In blue the actual refinement of the structure, in green the initial molecular replacement model. While it is interesting to see that a single urchin can produce information sufficient to build a preliminary model, further work needs to be carried on to validate and complete this model.

5.3 Conclusion

Measuring on a single urchin at room temperature used to be unrealistic due to the absence of a reliable available setup. In air, the samples were dehydrated, and measuring a small amount in a capillary is also difficult. For fragile samples, the centrifuging of microcrystals in a capillary may be fatal for diffraction. We show that the design of the humidity control device available at ID14 allowed us to collect high quality information. The possibility to preserve small protein powder samples at room temperature without losing diffraction properties is a new step for protein powder diffraction. This allows several possibilities, first, to screen very easily and routinely crystalline precipitates in order to determine their diffraction quality, on any PX beamline; second, to get reliable intensities at low to medium resolution, suitable to solve a structure by molecular replacement; third, to build and refine a preliminary model using these extracted intensities or via a Rietveld refinement.

Future development of this technique would require cryocooling of the sample, which has so far been unsuccessful on the Mayaro virus nsp3 macro domain, probably due to maladapted cryocooling conditions unable to pass through the drop's skin enveloping the urchins tried. We also do not expect a big improvement of the overlapping-peaks problem at high d-space resolution, because of the pixel size of the detector, but cryocooling would definitely improve the lifespan of the sample in the beam and allow us to obtain better statistics.

Also the severe overlap of the peaks at high angles will allow multiple models to fit the same set of overlapped reflections with the same statistics.

A more automated procedure for the structural refinement would be ideal to build several models in parallel in order to obtain a set of different structures fitting the same observed powder pattern.

6

Cryocooling of protein powders

Lorsqu'un échantillon de protéine sous forme cristalline est exposé au flux intense d'un rayonnement de rayons-X, le phénomène de dommages dûs aux radiations apparait. Cela résulte en un accroissement du facteur d'agitation thermique, d'une augmentation de la taille de la maille cristalline ainsi que de la rupture de liaisons chimiques de la molécule entraînant l'apparition de radicaux libres qui se déplacent et endommagent le cristal. Nous présentons dans ce chapitre une méthode permettant de congeler des températures cryogéniques des échantillons de poudre de protéine sans formation de glace cristalline. Cette technique permet de prolonger la durée de vie de l'échantillon pendant l'acquisition de données, mais aussi de ce fait permet d'obtenir des clichés de diffraction de poudre partir d'un volume d'échantillon moindre. Nous présentons également la possibilité d'affiner la structure tridimensionnelle d'une protéine à partir des données obtenues dans ces conditions.

6.1 Introduction

When a crystalline protein sample is exposed to an intense X-ray beam, damage due to the effect of ionizing radiation can be observed. Radiation damage results in making atomic B factors increase, unit-cell volumes increase, rotations and translations of molecules in the cell, rupture of disulphide bonds and decarboxylation of acid residues[54][55].

6.1.1 History of cryocooling of protein crystals.

A first experiment on myoglobin crystals was carried out in 1962 by Blake and Philips[56]. Their observation was that radiation damage was dose dependant. They measured that for each photon at 8 keV absorbed by the crystal, approx 70 molecules were disrupted, and 90 were disordered. A first model for losses in diffraction intensities was proposed, but wasn't enough to describe the "small but significant changes in the diffracted intensities which may indicate structural effects of the irradiation".

The usefulness of cryocooling samples was demonstrated in 1970 by Haas and Rossmann on lactate dehydrogenase[57]. They reported a "tenfold reduction in the rate of intensity loss of two reflections" compared to a room temperature experiment. To protect against the cooling effect, the crystal was soaked in a 3 M sucrose solution.

In 1975 a general method for "the replacement of the normal crystal mother liquor with salt-free aqueous/organic liquids of low freezing point" was presented[58]. A confirmation of the observation made in 1962 by Blake and Philips concerning the structural effect of the irradiation was made in 1988 by Helliwell on insulin [59].

The idea of using a thin-wire loop where the crystals are held by surface tension [60] and the availability of commercial cryostats allowed the technique to become routine.

With the availability of cryocooled measurements, it became possible to observe in more detail the radiation damage phenomena. In 2000 by using 100 K cryocooled crystals of myrosinase, radiation damage was observed in Fourier difference maps "in the form of breakage of disulfide bonds, decarboxylation of aspartate and glutamate residues, a loss of hydroxyl groups from tyrosine and of the methylthio group of methionine" [54].

Observed the same year on *Torpedo californica* acetylcholinesterase and hen egg white lysozyme, the effect of radiation damage is not the same for all the residues of the same type in a protein. Some sites, like in these cases the catalytic triad residue of acetylcholinesterase, appear to be more sensitive than other similar residues[61]. This led to the idea that some information about possible “weak links” is observable with radiation damage effects. These observations made clear that much care should be taken of the “fingerprint” left imprinted on structures by X-rays even at cryogenic temperatures. Especially “care must be taken when assigning structural significance to features that might easily be radiation-damage-induced changes” [55].

In 2005, the damage due to radiation has been used in a multiple wavelength anomalous experiment around the Se K-edge[62] in order to demonstrate the possibility to solve a structure; this is known as the radiation-damage-induced phasing (RIP).

The first successful attempt to collect a protein powder pattern under cryocooled conditions was made in 2007[63]. The microcrystalline tetragonal chicken egg white lysozyme used for the experiment diffracted for a much longer time (30x) improving statistics and allowing collection of useful information to a better d-spacing. Unfortunately, the presence of big crystalline ice peaks contaminating the sample was observed, leading to a loss of measured reflections.

6.1.2 Physical explanation of radiation damage.

When a crystal is exposed to a 12.4 keV synchrotron beam (1\AA), 98% of the X-rays pass through without interacting[64] and go on to the beam stop. The remaining 2% of X-rays interacting with the crystal are divided according to three effects¹ :

- **Thomson scattering (Rayleigh) 8%**, elastic and coherent, contribute to the observed diffraction pattern.
- **Compton scattering 8%**, inelastic and incoherent, contribute to adding background in the images.
- **Photoelectric effect 84%**, is where most of the energy absorbed by the crystal has effect. The photon energy received will eject a lower

1. The calculation is for a 100 micron thick lysozyme crystal which has 38% solvent and the remaining volume is full of amino acids considered to have average composition: $5\text{C} + 1.35\text{N} + 1.5\text{O} + 8\text{H}$ per amino acid (129 l.a.a. long). Also, the lysozyme having 8 cysteines and two methionines, so 10 sulphurs per molecule too (8 molecules per unit cell). Calculation made with RADDOSSE[65].

shell electron from the atom. While losing some energy in the process of ionisation, the electron will carry away the remaining energy from the photon. This photo-electron will move through the crystal losing energy while creating free radicals.

The free radical produced, OH, H, H^+ and e_{aq}^- (hydrated electrons due to radiolysis of water) move inside the crystal and destroy the crystal packing by interacting directly with the protein. The gain in diffraction resistance between a cryocooled and a room temperature sample is not obtained by a protective X-ray shielding effect. The gain is in that under cryocooled condition, the free radicals cannot move so freely, and therefore their impact is reduced.

6.1.3 Main challenges for protein powder diffraction

The actual methods developed for single crystals seemed to be an excellent starting point for obtaining cryocooled protein powder samples. Exchanging the mother liquor by a mixture which will give vitreous ice for example, or flash freezing the sample in liquid nitrogen before the experiment, seemed to be directly applicable to the protein powder diffraction technique. Unfortunately, when tried[63], if diffraction properties were conserved, it was at the price of crystalline ice peaks in the powder pattern.

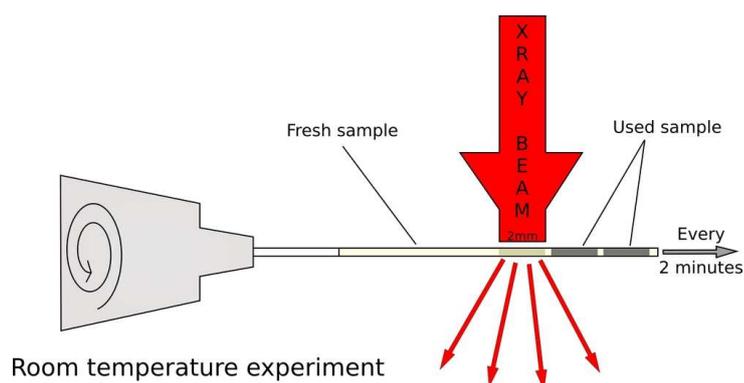
An understanding of this ice formation when the mother liquor was exchanged for a mixture supposed to be cryocooled was needed.

The addition of a cryoprotectant in the crystallisation condition is always disturbing for the protein crystals. Fortunately, it is known that the critical concentration of cryoprotectant is dependent of the cooling rate[66], if cooling is fast enough, the amount of cryoprotectant can be reduced. The design of simple and robust protocols was needed to avoid the waste of samples and time in various attempts.

6.2 Results

6.2.1 Why protein powder diffraction needs cryocooling?

A typical protein powder sample loaded in a borosilicate capillary, mounted on the ID31 diffractometer at room temperature, often can't resist more than 2 minutes due to radiation damage. The sample is moved every two minutes to a fresh position, where the sample has not yet been irradiated, in order to collect another pattern.



When a data set is collected from a sample at the same place for several scans, radiation damage effects are clearly observed resulting in :

- Reduction of the peak intensities
- Shifts of the peak positions
- Broadening of the peak width

After some time in the beam, we observe a colour change of the sample (darkening). Also it is possible to observe bubble formation inside the capillary (figure 6.5, p. 111).

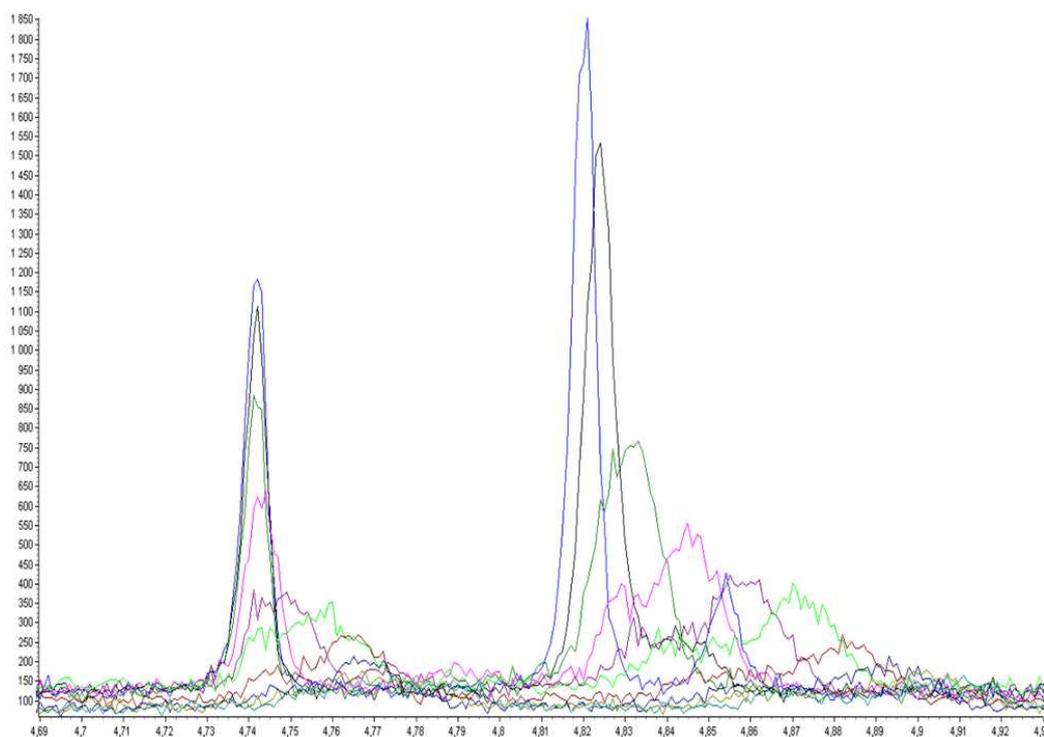


Figure 6.1: Effect of radiation damage on a human insulin T6 sample at room temperature. A reduction of the peak intensity, a shift of the peaks and a broadening of the peak widths are observed.

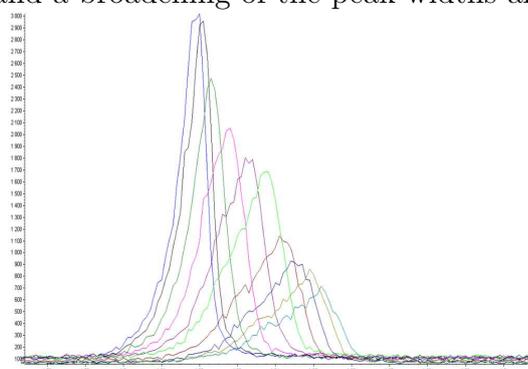


Figure 6.2: Effect of radiation damage on a human insulin T6 sample at room temperature. While the two first scans are of comparable quality (same intensity and same peak width), a slight shift of the peak is evident.

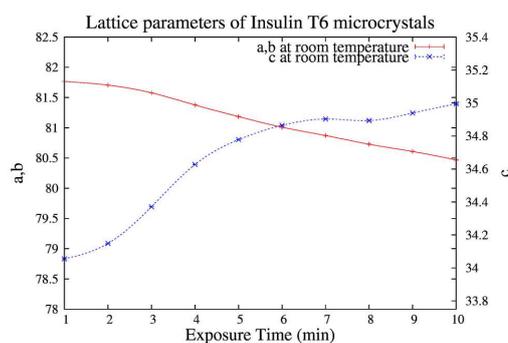


Figure 6.3: An important shift in lattice parameters is observed at room temperature, only after a few fast scans of 1 minute.

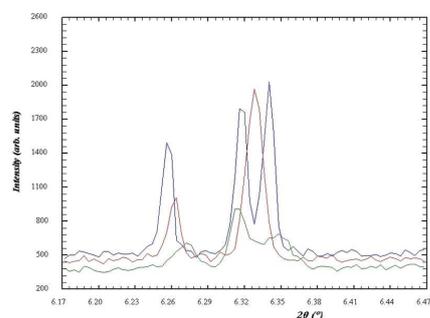


Figure 6.4: Anisotropic shifts in peak positions caused by radiation damage can be used with overlapping peaks to determine their relative proportion. After the third collection at the same sample position there is no diffraction signal exploitable.

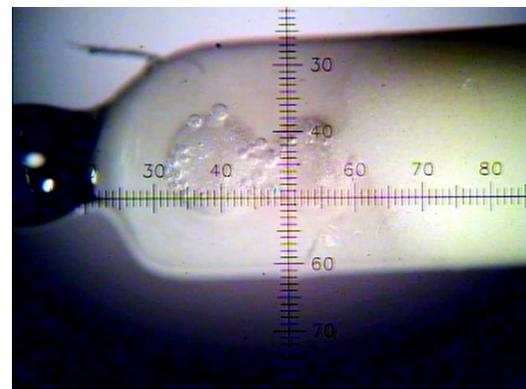


Figure 6.5: Effect of radiation damage on sample.

At room temperature, gas formation is observed while the sample is in the beam.

6.2.2 Finding a correct sample holder for cryocooling of protein powders.

Cryocooling with glass capillaries.

In order to reduce the diffusion of free radicals released by the X-ray beam (figure 6.1.2, p. 107) we tried using a Cryostream to keep cooled our typical sample holder, borosilicate (glass) capillary.

A classical test in protein single crystallography is to hold a nylon loop in a cryostream flow. The loop is containing only the mixture of cryoprotectant and mother liquor. If the mixture is cryocompatible, meaning only that it will not give any ice formation when cooled, the loop remains clear. If the mixture is not cryocompatible, the loop becomes white (due to the ice).

A similar test is applied to our glass capillary in order to find the optimal condition for a given mother liquor. It shown that even the cryoprotectant itself alone (without mother liquor) give ice in a 1mm borosilicate capillary (figure 6.8, p. 113).

It was then clear that the classical borosilicate capillary wasn't suited for this experiment. Several attempts were then made to find a way to prepare samples, including among others, the idea of spraying the crystals onto a cold surface or using a nanodrop dispenser on top of liquid nitrogen. While some were sometimes successful, these attempts were not compatible with the sample studied. Protein microcrystals are fragile, rare and therefore expensive. The need for a more reliable setup, keeping the amount of sample needed reasonable, was evident.

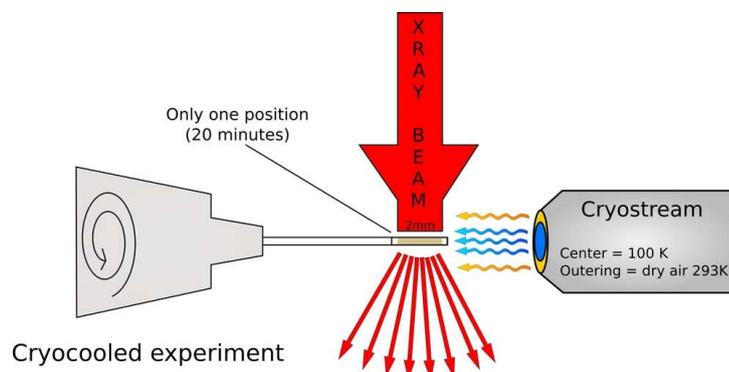


Figure 6.6: The cryostream produces a cold nitrogen flow at 100 K surrounded by dry nitrogen. The cold nitrogen flow keeps the sample cryocooled. The dry nitrogen is important in order to avoid ice formation from the humidity in the air outside of the tube.



Figure 6.7: A typical sample holder for room temperature protein powder diffraction. The brass parts are made to hold the 1 mm borosilicate capillary centred in the beam while being spun at high speed (typically 1000 rpm).



Figure 6.8: Even filled with only 100% glycerol (a classical cryoprotectant), and flash cooled in liquid nitrogen, a white crystalline form of glycerol is formed inside the borosilicate (glass) capillary.

Cryocooling with polyimide tubes.

Among the different materials used in synchrotron radiation facilities, kapton foils made of a polyimide film developed by DuPont attracted us. Kapton films remain stable under a wide range of temperatures, are flexible but more importantly are transparent to X-rays. These foils are commonly used as windows for equipment where X-rays need to pass through without interaction. Also it was known that kapton's thermal conductivity ($0.10 \text{ Wm}^{-1}\text{K}^{-1}$) properties make it ideal for use in dissipating and managing heat in electronic assemblies.

We first used a foil of kapton kept on a very cold surface (a metal block cooled with liquid nitrogen) as a receptacle for sprayed microcrystals. But soon the idea to use this material in the shape of a capillary was evident. The use of a kapton capillary was already described for room temperature experiments as a sample holder with an embedded filter [67].

Tube preparation The company GoodFellow sells polyimide tube. Our first choice was for a tube with an internal diameter of 1.0 mm (wall thickness is 0.025 mm).

The kapton capillaries are open to both ends, and need to be sealed. Several trials were made to find the best way to do this. Flame sealing was impossible, and resulted in releasing in air carbon monoxide. Sealing with all kinds of waxes results in cracks when cooling at low temperatures. Sealing with Araldite, a bi-component glue, allows the capillary to be cooled down and warmed up for a limited amount of time. Araldite can be used if the sample will not be kept and reused afterwards. The best sealing material found was a glue used in some applications under vacuum Stycast 2850FT with catalyst 8% (weight/weight) of 4,7,10-trioxyatridecane-1,3-diamine. Stycast is a two component epoxy resin, which doesn't shrink when cooled down. As this epoxy resin takes overnight to get solid, it is a necessary to seal one end of the capillary prior to the experiment. This will allow the centrifuging of the microcrystals to the sealed end of the capillary. Unfortunately the time to get solid is too slow to seal the sample once it is prepared, or for attaching the capillary to the brass sample holder. For this silicone vacuum grease is enough, and will keep the capillary on the brass sample holder (if diameters are well chosen), and protect the sample from contamination or evaporation at room temperature.



Figure 6.9: The complete setup for a cryocooling experiment (the cryostream is removed for a clearer view). The sample (in white) is centrifuged to the end of a kapton capillary. The capillary is sealed at one end with Sty-cast resin (black), and the capillary is held inside the brass holder with grease. The brass holder is held via a strong magnet on the spinner of the diffractometer.



Figure 6.10: While the use of the Sty-cast resin is recommended, it is still possible to use Araldite to seal the capillary. After cycling between cryotemperatures and room temperature, the sample will no longer be sealed. The sample is in white (centre), the araldite is in orange (right side).

Sample preparation Even when using kapton capillaries, the choice of cryoprotectant and the amount to be mixed with the mother liquor need to be adjusted (figure 6.12, p. 118).

The first thing to do is to define a possible cryoprotectant, by filling the end of a kapton capillary with the mixture of cryoprotectant and mother liquor. A good idea is to start with a high concentration of cryoprotectant. If the kapton tube inside liquid nitrogen remains clear, there is a good chance that the cryoprotectant is usable. Decreasing the amount of cryoprotectant until ice is formed will give the minimum amount needed.

Once one condition with the minimum amount of cryoprotectant to avoid ice is found, the crystals' resistance to this change of condition needs to be tested.

As the preparation of a capillary will need a centrifuging step, the removal of excess mother liquor and sealing prior to cooling, the crystalline sample needs to be able to remain stable under the cryocooled conditions.

In numerous cases, we saw dissolution of the protein crystals during preparation, which take on average 5 minutes. In this case it is very easy to see, as no crystals remain, that there will be no diffraction. In other cases, crystalline sample seemed to remain stable, but didn't diffract properly afterwards.

Our favourite starting point is the CryoPro Cryoprotectant reagent kit from Hampton Research. Composed of 36 different agents, polyols, organics, oils, polymers, sugars and salts, this kit is a useful tool for screening purposes when simple agents are unsuccessful.

In some particular case we didn't succeed in finding a good cryocooling condition, even after screening in a 96 well plate (figure 6.11, p. 116) to see under which conditions the crystals were dissolving.

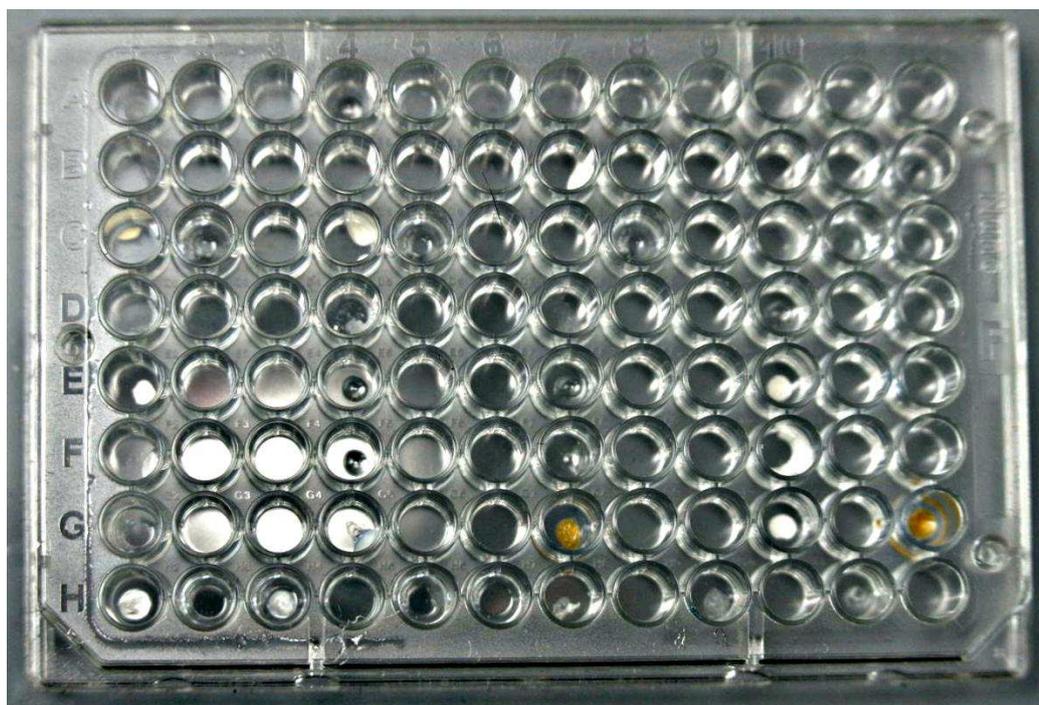


Figure 6.11: Finding a proper cryoprotectant which has a high enough concentration for avoiding ice formation inside the capillary while not dissolving the crystals can require some effort. Here in this 96 well plate (normally used for crystallisation screening purposes) several cryoconditions mixed with crystals are tested. All the suitable conditions for a capillary from a CryoPro kit are mixed with protein microcrystals at different concentrations. Transparent wells demonstrate conditions where crystals dissolve, whereas the wells where a remaining white precipitate is seen are conditions that should be tested with X-rays (C4, H1, E10, F10, G10 for example).

It is important to note that an isotropic effect is always seen when samples are cryocooled. Also, in a cryocooled experiment, parameters difficult to

monitor, such as the cooling speed, may lead to small variations in lattice parameters. This is important to keep in mind when comparable accurate lattice parameters are needed. For a polymorphism study for example, or when optimal crystallisation conditions need to be defined, an experiment with only room temperature measurements would be more suited.

Also it is important to note that the use of full capillaries implicate prior cooling in liquid nitrogen. If a cryostream is used directly to cool the capillary, crystalline ice will appear, even if the cryoconditions are good. The use of prior cooling in liquid nitrogen (often called “flash cooling”) is mandatory. Unfortunately, this step cannot be accomplished by ID31’s robotic sample changer.

When a good cryocooling condition is obtained, an extension in d-space of observable reflections is obtained (figure 6.14, p. 120). Peak broadening compared to room temperature is severe, but allows data collection for a longer time, allowing a larger number of scans collected on the same sample volume to be summed together.

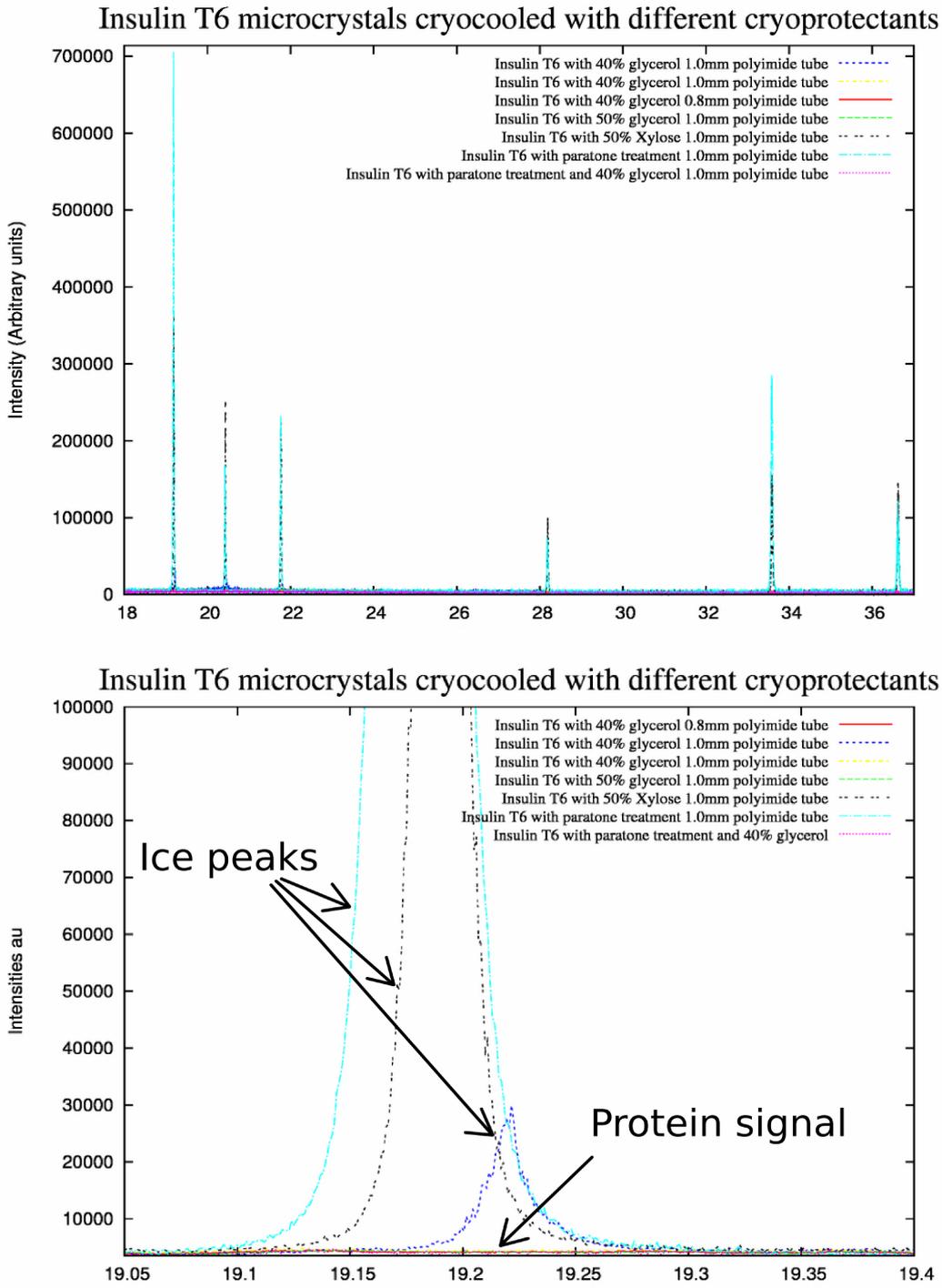


Figure 6.12: Intense and broad (compared to the almost invisible protein signal) crystalline ice peaks are visible when suitable cryocooling conditions are not found.

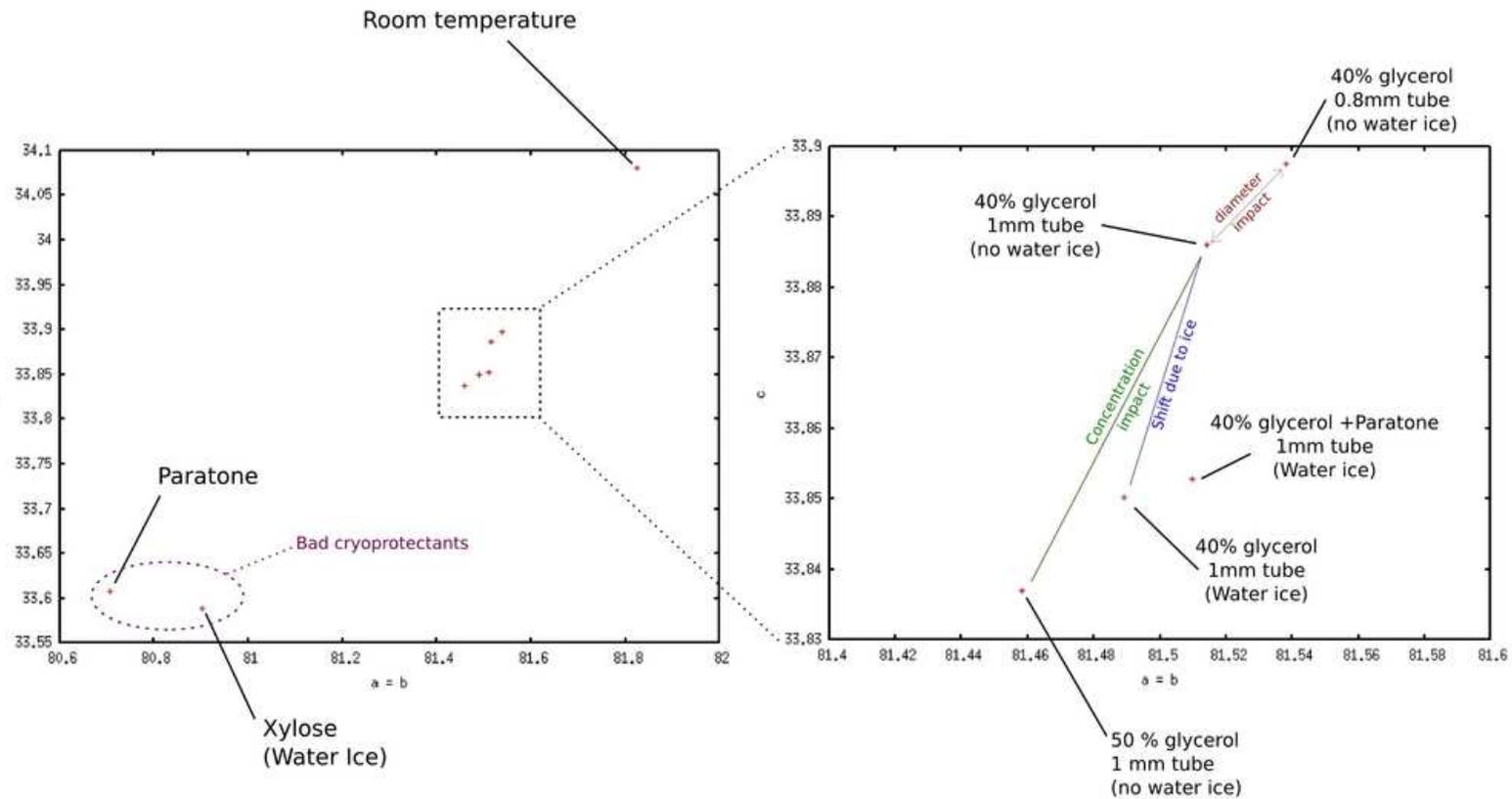


Figure 6.13: The high angular resolution of ID31 diffractometer allows us to see shifts in lattice parameters of human Insulin T6 due to cryoprotectants. The same batch of microcrystals is used in all experiment removing crystallisation variation bias. It is also interesting to see that for the same condition (40% glycerol), in one case when the sample wasn't plunged inside liquid nitrogen (but directly cooled with the cryostream) the cooling speed reached wasn't sufficient and gave crystalline ice.

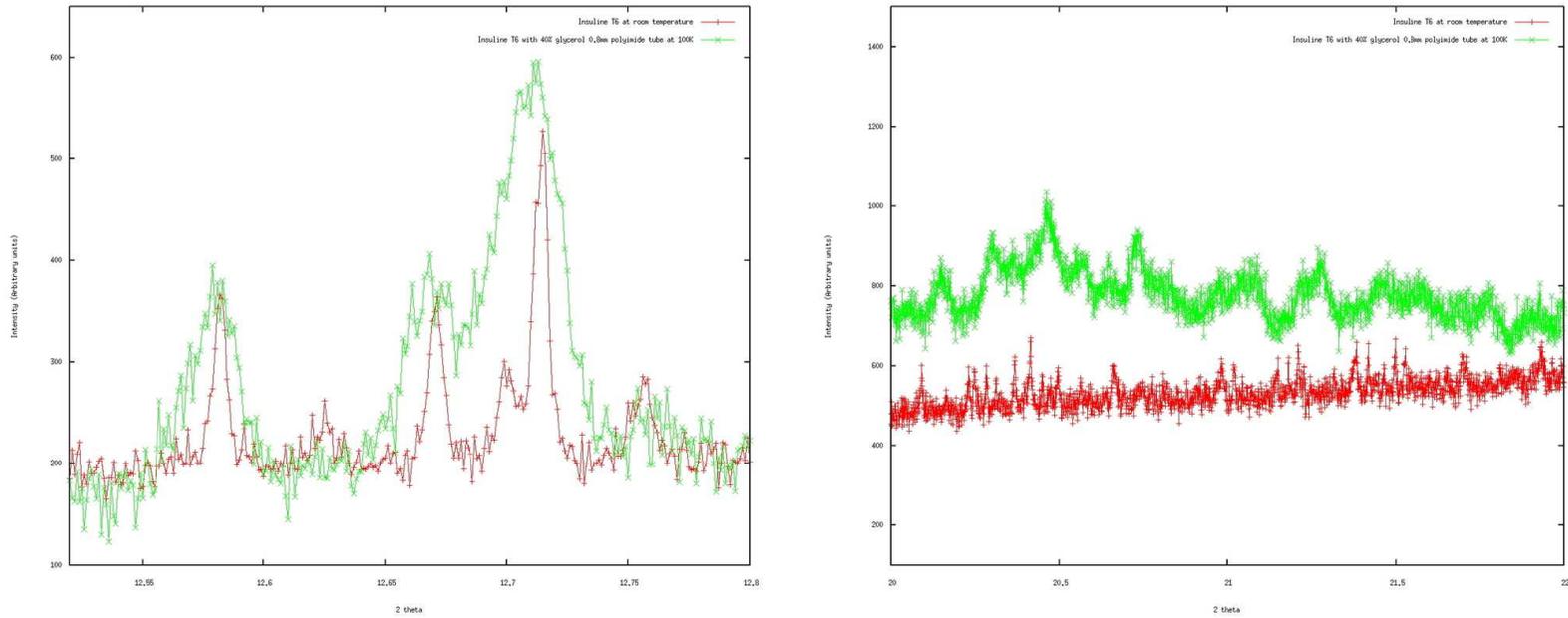


Figure 6.14: Comparison of two scans of human insulin T6, at room temperature (red) and at cryocooled conditions (green). At room temperature, the peaks are much thinner, the angular resolution is better. The cryocooled pattern has broader peaks, but new peaks are seen at high angle, the d-space resolution is better. Also, this is only one scan, when summing multiple scans together, which is easily possible with the cryocooled sample, the noise will become smaller.

Cryocooling a cylindrical annulus of protein powder.

Several motivations pushed us to improve the kapton tube set-up. Using even less sample was of course the first motivation, obtaining less peak broadening was the second, and finally having a set-up compatible with our actual sample changing robot was interesting too.

Peak broadening appears when the sample is cooled. A possible explanation is that the lattice parameter variations, observed when the sample is cooled are not identical for all the crystals. In this scenario, several peaks slightly shifted will add together and lead to a broader peak (figure 6.15, p. 121).

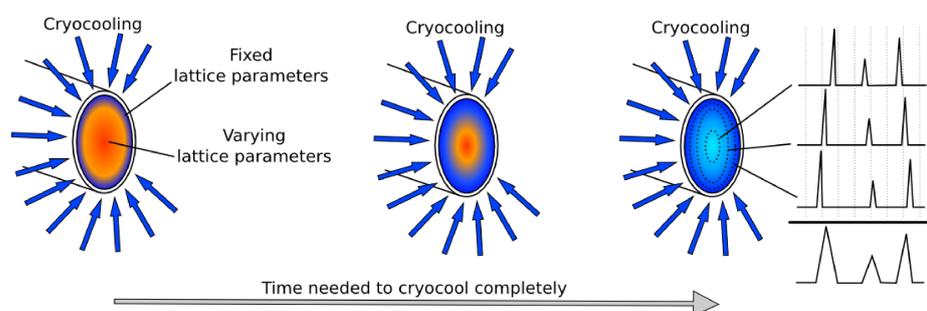


Figure 6.15: The time needed to cryocool all the grains of the sample leads to non-uniform variation of the lattice parameters. When irradiated, the different peaks are summed together, worsening the peak broadening.

These variations were probably due to the amount of time needed to cryocool the sample. Manipulating smaller capillaries was possible, the 0.8mm inner diameter is now a standard in our laboratory. But using even smaller tube size, like 0.4mm or 0.2mm inner diameters proved impossible. Indeed the viscousness of most of the samples, made of a mixture of cryoprotectant and protein, makes very difficult the sample preparation and centrifuging. Another way of reducing the amount of sample needed, and the time to cool it completely, has been found by inserting another capillary inside a kapton capillary containing the sample (figure 6.16, p. 122).

The resulting cylindrical annulus of protein powder volume is nearly 3

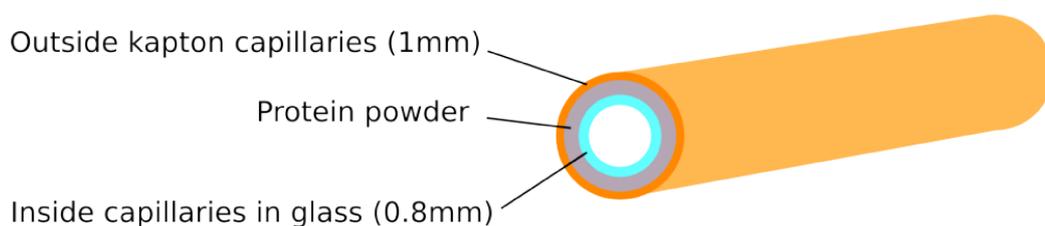


Figure 6.16: An empty glass capillary (light blue) is inserted inside a kapton tube (in orange) containing the protein powder (in grey). The resulting shape of the protein powder sample is a cylindrical annulus. The amount of sample needed is smaller.

times smaller than with a full capillary. To collect with a horizontal beam size of 2mm (optimal for the detectors) the final volume needed of protein powder is only 0.57 mm^3 .

A test to see the evolution of the cell parameters over a long period has been made with human insulin T3R3, showing a slow and regular cell expansion (figure 6.17, p. 122). The diffracting pattern collected after 13h was of the same angular quality as the first one.

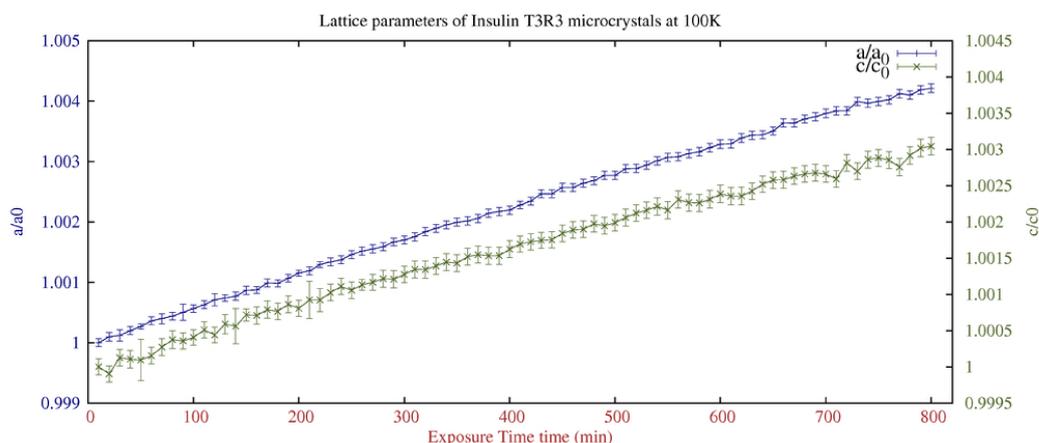


Figure 6.17: Evolution of the cell parameters during 13.3 hours of collection time on the same sample of cryocooled human insulin T3R3. After 13h the cell parameters expand by less than 0.4%. Parametric refinement[68] was performed on the 80 patterns simultaneously.

But some effects of the radiation damage can be observed during collection time over such long periods. While some reflections remain constant some vary dramatically (figure 6.18, p. 123).

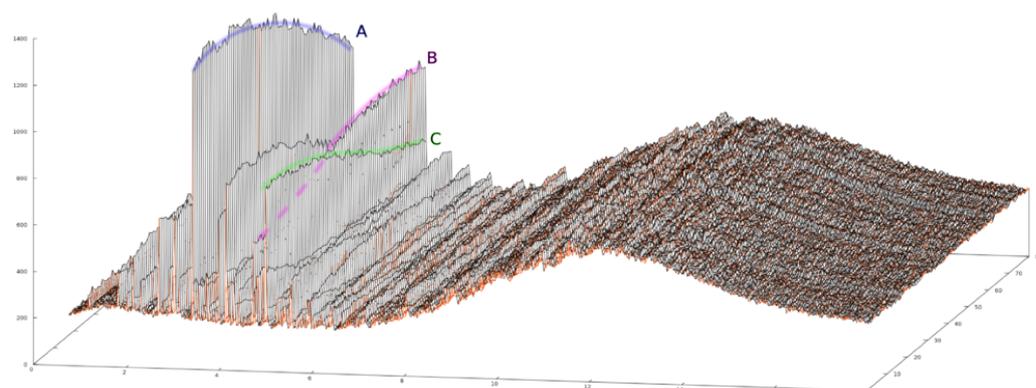


Figure 6.18: Evolution of the diffraction pattern during collection time. Sample of human insulin T3R3 collected at the same position for 13.3 hours, cryocooled to 100K, without prior flash freezing. The z-axis shows the number of scans (10 minutes/scan). While the lattice parameters are slightly moving (figure 6.17, p. 122), the structure itself is changing. Some reflection intensities stay nearly constant over time (A), some vary significantly (B & C).

A selective summation must be done in this case, as it is not possible to sum all the scans together, since the structure clearly changed.

Protocol for protein microcrystal cryocooling.

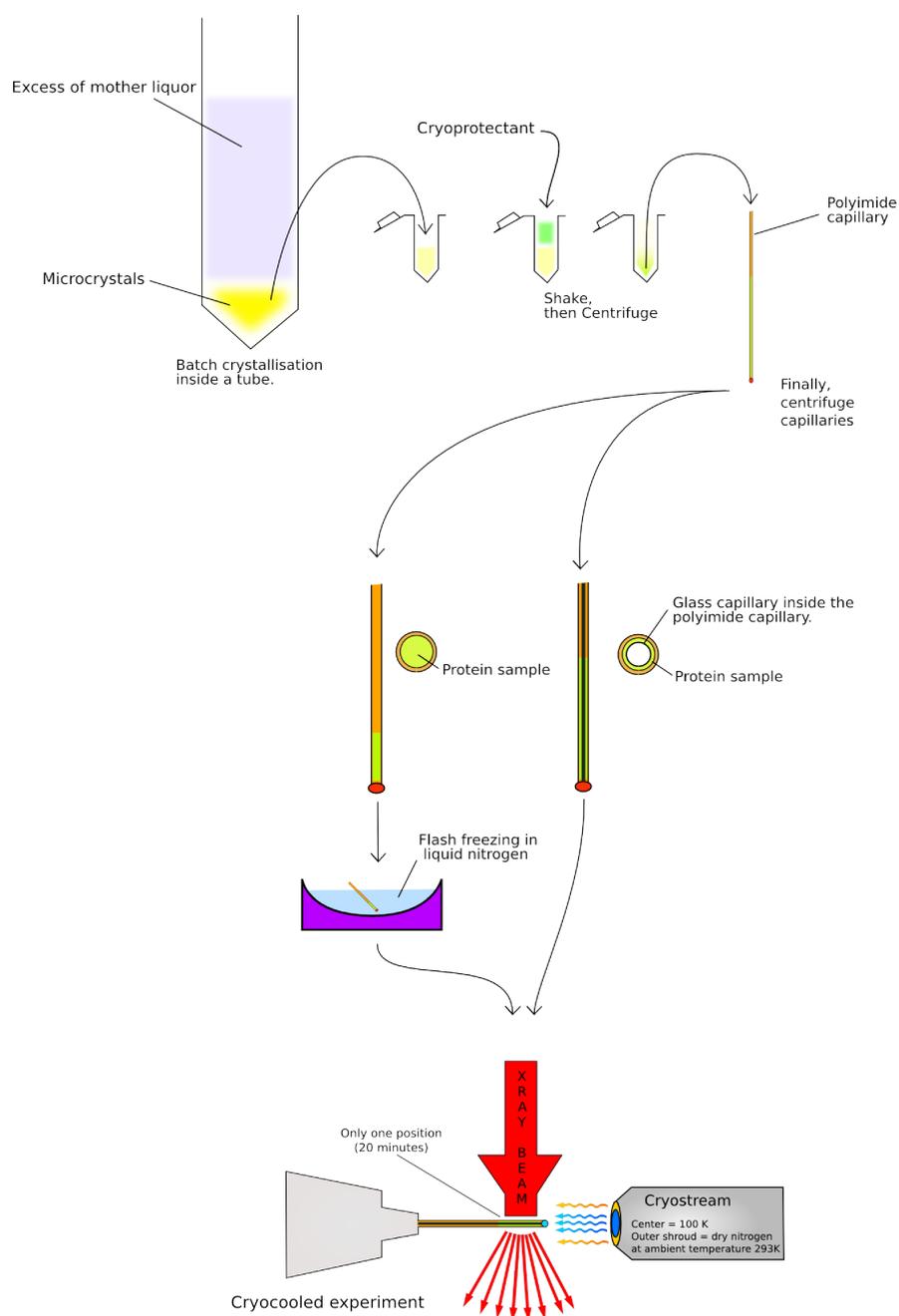


Figure 6.19: An easy protocol for cryocooling of a protein powder sample.

6.3 Structural refinement of cryocooled human insulin.

In order to validate the protocol of cryocooling we refined two human insulin structures (T6 and T3R3). These two structures were already known by single crystal diffraction studies on microcrystals under cryocooled conditions.

6.3.1 Insulin microcrystals

Insulin is a well known hormone whose key role is to regulate energy and glucose metabolism. Insulin is composed of 51 amino acids and has a molecular weight of 5808 Da. It is produced in the islets of Langerhans in the pancreas.

When insulin is present, the cells in the liver, muscle, and fat tissue will store glucose from the blood as glycogen. When insulin is absent, glucose is not stored by cells and the organism begins to use fat as an energy source.

People having low levels of insulin or low sensitivity suffer from diabetes mellitus. Type I diabetes mellitus is the consequence of a non production of insulin by the organism. It can be considered as an auto-immune disease[69], as the immune system is attacking the β -cells of the islets of Langerhans. Patients require a life-long supply of exogeneous insulin. Type II diabetes mellitus appears when cells lose the ability to respond adequately to the blood level of insulin. The mechanisms that lead to this insulin resistance are still elusive, but obesity is often associated with type II diabetes. As in this case the level of insulin is normal, the treatment doesn't always require injection, but most of the time is treated by a healthier diet and more physical activities. All diabetic patients will have less glucose entering the cells, leading to a very high level of glucose in the blood, up to the point where glucose will be excreted by the kidneys in the urine.

Injection of insulin under a microcrystalline formulation is interesting because the insulin crystals will need time to dissolve in the blood, and therefore the release time will be longer. Modern preparations of insulin are composed of an amorphous and microcrystalline mixture.

Novo Nordisk, a pharmaceutical company and leader in the area of diabetic care, provided us with large quantities of two forms of microcrystalline human insulin (T6 and T3R3). We used these high quality preparations to

optimise and find protocols for protein powder cryocooling. As insulin T6 and T3R3 structures were already determined, we checked that our powder cryocooled samples were giving the same information by doing a Rietveld refinement.

The samples were prepared following the protocol described in 6.19, p. 124, with 40% glycerol as cryoprotectant.

6.3.2 Rietveld refinement

A Rietveld refinement was carried out in GSAS using as starting model the corresponding single crystal model. The Rietveld refinement is straightforward due to the large number of chemical restraints.

The differences between single crystal and powder structures show only small differences in side chains.

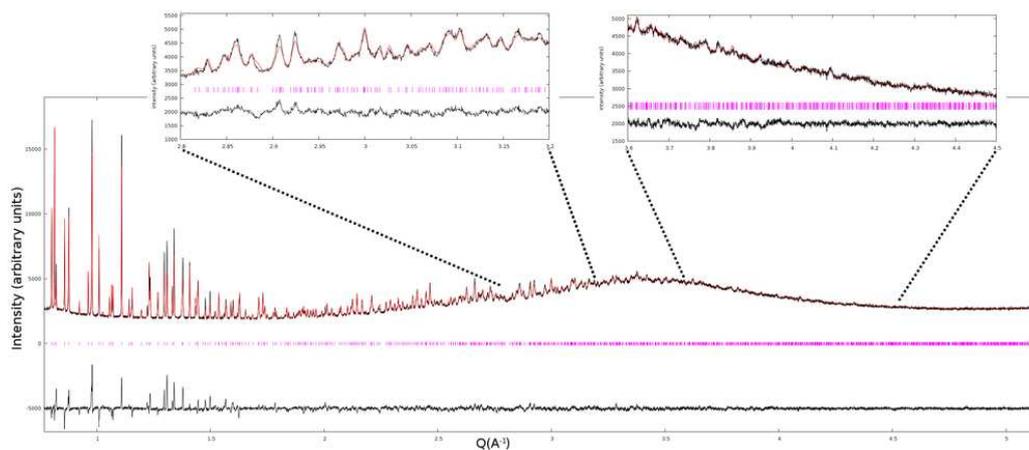


Figure 6.20: Rietveld fit of human insulin T6 cryocooled at 100K. The fit was obtained after refining the single crystal structure 1MSO with GSAS.

I would like to thank Gerd Schluckbier and Mathias Norrman for providing us with this high quality samples, and Sotonye Dagogo for her help during sample preparation.

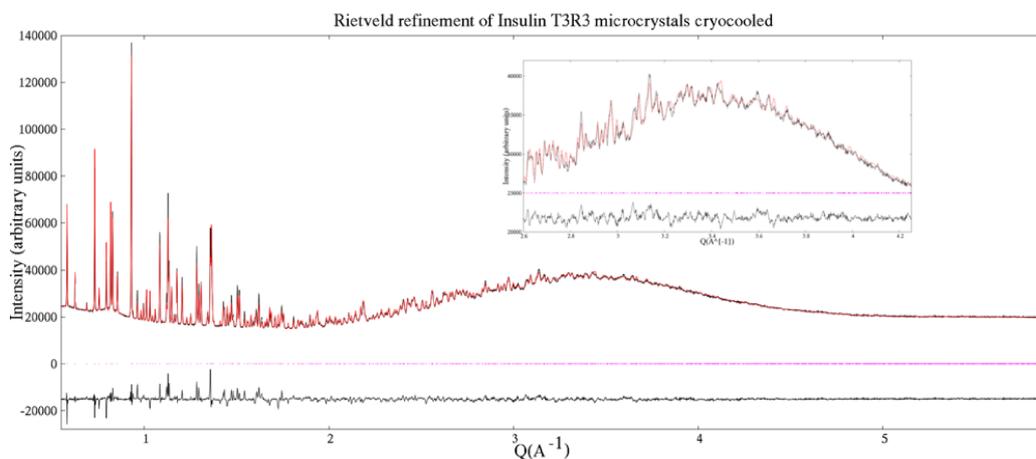


Figure 6.21: Rietveld fit of human insulin T3R3 cryocooled at 5K using a Liquid-helium-cooled cryostat. The fit was obtained after refining the room temperature powder crystal structure 1FUB with GSAS.

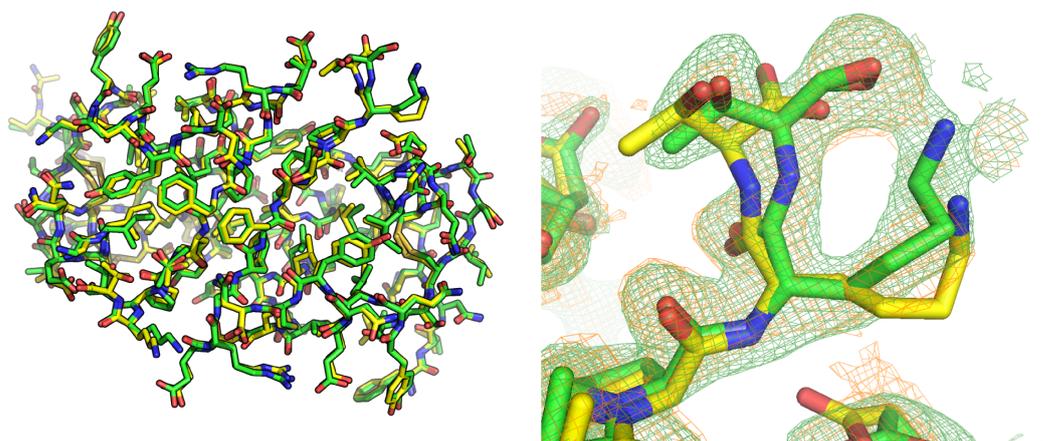


Figure 6.22: Human insulin T6 structure refined (green) compared to the 1MSO (yellow) single crystal structure. Green map is the Fobs-Fcalc map from GSAS, the orange map is a total omit map obtained with sfcheck.

	before refinement	after refinement
wRp	0.0524	0.0371
Reduced χ^2	194.8	4.794
d-space resolution	13.5 - 2.64 Å	
Number of observations	5413	
Number of refined parameters	891	
Number of restraints	1046	

Table 6.1: Cryocooled human insulin T3R3 at 5 K, structural refinement using GSAS, using the 1FUB pdb structure as starting model.

	before refinement	after refinement
wRp	0.0401	0.0357
Reduced χ^2	0.1032E+06	24.20
d-space resolution	17 - 2.36 Å	
Number of observations	7210	
Number of refined parameters	891	
Number of restraints	1046	

Table 6.2: Cryocooled human insulin T6 at 100K, structural refinement using GSAS, using the 1MSO pdb structure as starting model.

6.4 Conclusion

A new protocol for protein microcrystals cryocooling sample mounting have been developed, consisting of cooling only a cylindrical annulus of protein powder. This protocol made possible a higher cooling rate of the complete solution containing the protein crystals, allowing the possibility to use smaller amount of cryoprotectant agents.

The diffraction patterns obtained on insulin microcrystals shows the possibility to use this technique routinely. The analysis of these diffraction patterns shows that after refinement of the single crystal structure against powder data, only small variations concerning the protein side chains atoms positions were observed.

In mitigation, the added value of having more collection time and a higher d-space resolution is counterbalanced by having a lower angular resolution critical to powder patterns. The intensities can be considered more reliable because of the larger number of scans that can be collected on the same sample, but the informations obtained can be more complicated to interpret, especially when the effect of overlapped reflections became more severe due to the lower angular resolution.

Also, it is important to keep in mind that the absence of a good diffraction pattern from a cryocooled sample doesn't mean that the sample itself is not suitable for powder diffraction, but that the crystals could have been degraded by the cryoprotection procedure.

Other techniques in order to improve the quality of the current setup have been tested. For example annealing of temperature[70] have been tried, but didn't show improvement for the Insulin microcrystals. Still this can be of interest in other test cases.

To conclude, instead of, like in a room temperature experiment, summing together many diffraction patterns collected on different samples in order to have sufficient statistics, here with cryocooling, only a small amount of sample is required. The main advantage of this new protocol is then to reduce the amount of sample needed to collect an accurate diffraction pattern, which is a crucial key to the use of powder diffraction with biological samples.

7

Investigation of AcMNPV and CpGV polyhedrin with powder diffraction

Les baculovirus et les cypovirus ont la remarquable habileté à produire eux-même, une matrice cristalline, pendant l'infection de leur hôte. Ces microcristaux servent de protection contre les agressions extérieures et permettent la préservation du virion en dehors de l'hôte. De ce fait il est intéressant d'observer les informations que peuvent nous apporter un cristal "naturellement" formé in-vivo, puis de les confronter à celles d'un cristal d'une protéine recombinante équivalente. Nous avons étudié la diffraction de la matrice cristalline de deux virus, AcMNPV et CpGV. Nous avons observé que les informations des cristaux obtenus naturellement étaient suffisantes pour discriminer l'espèce du virus infectant, ce qui peut être utile comme contrôle qualité dans l'industrie agro-alimentaire où les baculovirus sont de plus en plus utilisés. Nous avons également observé les limites de la diffraction de poudre de protéine dans le cadre de symétries cubiques. Dans ce cas, il est en effet très difficile d'affiner la structure de la protéine formant la matrice cristalline, malgré la haute qualité des clichés de diffraction obtenus.

7.1 Introduction

Baculoviruses and cypoviruses cause diseases in a whole range of insects, on earth and underwater. Even though they are present in all environments, in our food, in the air, in water for example, only members of the phylum *Arthropoda* are infected.

The family *Baculoviridae* is a member of the Group I of DNA viruses (double stranded DNA - dsDNA, with no RNA stage) with a genome size between 88 kbp and 160 kbp. There are 184¹ known members of the *Baculoviridae* family.

Baculoviruses are all occluded in a protein matrix, but they are divided in two genera depending of the protein used for the occlusion matrix. Granuloviruses (GV) have a granulin crystalline matrix while nuclear polyhedrosis viruses (NPV) have a polyhedrin crystalline matrix. For the GVs, each inclusion body contains only one virion, while for the NPVs it could be either one (SNPV) or several (MNPV) envelopped nucleocapsids per inclusion body.

The genus cypovirus, even though it shares a lot of structural similarity for its occlusion body with the MNPV, is not considered to be baculovirus. This is because the genetic information is in this case a double stranded RNA (dsRNA). There are 33 members² of the cypovirus genus.

This crystalline matrix protecting the virions allows the viruses to stay viable for a long time outside the host. This feature exists also in completely different viruses, like within the Poxviridae, with the entomopox viruses (EPV). The occlusion mechanism is very similar between these groups, with a release in the host midgut due to the alkali-solubility of the occlusion body. Perhaps surprisingly, it seems that these three virus groups have evolved independently to achieve the same protection mechanism[71].

The understanding of such a convergence toward a unique mechanism of protection is a motivation for using diffraction techniques to study a natural (in vivo) protein crystal.

1. On the 2nd of February 2010, 184 baculoviruses: NCBI Taxonomy ID 10442
2. On the 2nd of February 2010, 33 Cypoviruses NCBI Taxonomy ID 10981

Table 7.1: The Baculovirus and Cypoviruses

Baculoviridae	$\alpha/\gamma/\delta$ -baculovirus	NPV Nuclear polyhedrosis virus Polyhedrin matrix	SNPV One nucleocapside per envelope MNPV Multiple nucleocapsides per envelope
	β -baculovirus	GV Granulovirus Granulin matrix	 One nucleocapside per envelope
Reoviridae	cypovirus	CPV Polyhedrin matrix	 One nucleocapside per envelope

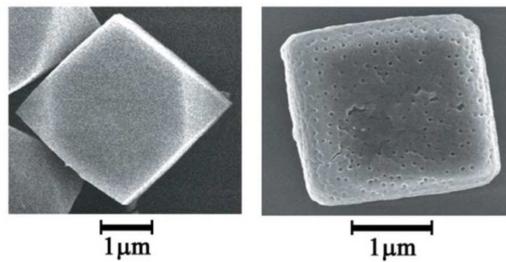


Figure 7.1: Crystals of baculovirus. Left a recombinant crystal, right a wild (natural) crystal.

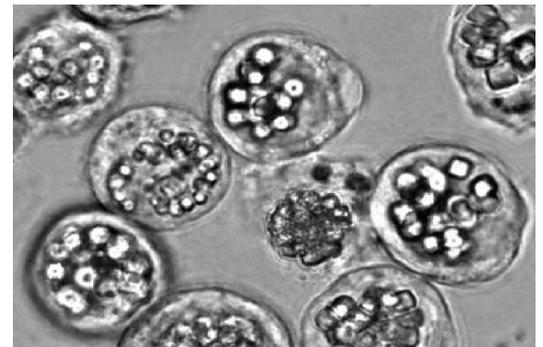


Figure 7.2: Insect cells infected by a baculovirus, crystals are formed inside the cell.

7.1.1 Baculoviruses : a short history

Man has been aware of the effects of baculoviruses for a long time. An interesting historical fact is that a description of a disease due to a baculovirus is found as far back³ as the 16th century⁴. It could be very surprising to have such an early description of a non-human disease, but the described effect, which was liquefaction of larvae, was on very useful and economically important insects, silkworms.

3. apart from some ancient Chinese literature[72]

4. poem by Marco Vida of Cremona[73]

Another interesting historical fact is that the first structural discovery concerning the baculoviruses was made two centuries ago. During the 19th century, with the improvement of microscopy techniques, it was possible to observe the details of the liquefaction observed by the silk moth farmer. A large number of small polyhedral crystals were observable. At the beginning of the 20th century emerged the idea that virus particles were present in the crystals. Rod shaped virions were observed by Bergold in the 1930s, which started the modern study of the baculoviruses.

The first motivation for these studies was to find an effective class of pesticide agents against insect pests. The first use of baculovirus as “biological pesticide” was successfully conducted against the “European spruce sawfly”, a pest that appeared in the 1940s.

Later even if biological pesticides were mainly replaced by “superior” chemical pesticides, baculovirus were still studied for their wonderful molecular biology. In the 1970s, the separation of the baculovirus into two different types, one occluded (ODV) and one budded (BV), was correlated with the infected part of the insect. ODV are limited to the midgut of the insect while BV can infect other tissues and *in vitro*.

The ability of the BV form to spread in cell culture was used with the *Autographa californica* nuclear polyhedrosis virus (AcMNPV). The large number of identical viruses obtained in cell culture allowed study of each independent gene in *Escherichia coli*.

The improvement in molecular biology in the 1990s allowed the development of powerful methods for biotechnology, for example, using baculoviruses as vectors for expression of heterologous genes. The very high level of transcription needed to produce the matrix protein is mainly due to the viral α -amanitin-resistant RNA polymerase. This has been exploited for the development of the baculovirus expression system, by replacing the polyhedrin gene by the gene that one wants to be over-expressed, while keeping the original polyhedrin promoter. This expression system allows production of eukaryotic proteins in high quantities while keeping the folding and the post transcriptional processing of the native protein.

Today, baculovirology is very active in various fields. For example, organic farmers are in crucial need of efficient biological pest control agents to replace chemical pesticides. Also, the vaccine for chickens against the H5N1 avian influenza is produced in a baculovirus expression vector. One vaccine produced by recombinant technology using a baculovirus expression system for human use has been approved by the US Food and Drug Administration

(FDA). It is Cervarix (GlaxoSmithKline Biologicals) used in the prevention of the Human Papillomavirus (HPV).

From another point of view, it is interesting to see that while a lot of progress has been made in one century, in order to produce baculoviruses, the initial problem of stopping the baculoviruses infesting beneficial arthropods is not yet resolved. This is actually very important due to the fact that some new economic activities, like the shrimp aquaculture industry[74], are confronted with baculovirus diseases.

7.1.2 Baculovirus life cycle

The first phase of infection of an insect (for example) by a baculovirus starts when the insect feeds on a leaf contaminated with the occlusion derived virus (ODV) form of the baculovirus. The crystalline protein matrix of the ODV is dissolved in the midgut of the animal by the high pH. The ODV enters the cell by endocytosis, the intestine cell membrane forming an endosome around the ODV. Then nucleocapsids escape from the endosome and are transported to the nucleus.

The occluded form is not efficient at spreading the infection within insect tissue. That's why, once in the nucleus, after transcription and replication has occurred, budded viruses (BV) are first produced, spreading the infection throughout the insect.

The second phase of the infection starts when occluded virus (ODV) starts to be formed. Nucleocapsids are transported from the nucleus to the cytoplasm, and obtain the envelope from the host plasma membrane modified with a viral glycoprotein called gp64. Then the enveloped nucleocapsids are packed inside the protein matrix and become an occluded body. The cell lysing will literally melt the insect, which will release the virus outside its host, in its ODV form, spreading the infection to another insect.

7.1.3 Structure of Baculovirus

Even though most viruses are fragile outside their host, Baculoviruses have the remarkable ability to resist over a long period in harsh conditions. For example, on land, baculoviruses are known to be able to stay outside their host for several years in soil[75], resisting harsh temperatures, pH, chemicals, light inactivation. This capability is needed because of the seasonal nature of their hosts, for example insects often have a population expansion peak each

year due to seasonal parameters. To understand how this class of viruses is able to survive in these conditions, structural studies have been made over the last 40 years.

With over 150 open reading frames (ORF) in its genome (for example 154 for AcMNPV), baculoviruses have the possibility to encode a large number of viral proteins. There are several structural proteins that can be sorted into three categories depending on their location in the virus structure, one for the polyhedra, one for the envelope and the third for the nucleocapsid.

Table 7.2: Major Structural proteins of Baculoviruses.

Major structural proteins	Envelope	gp64: Baculovirus fusion protein gp64 mediates the fusion between host and viral membranes. gp64 is concentrated within the plasma membrane in order to envelop the virion when budded outside the cell. This protein has the unusual ability to promote entry in insect and mammalian cells.
	Nucleocapsid	DNA binding protein Replaces the histone for hyper-condensation of the large genome in the nucleocapsid. Allows the formation of a biochemically inert chromatin complex.
		p39, p87, p24, p61 ...: Major component of the capsid
	Occlusion body	Polyhedrin and Granulin: Major components of polyhedra because of their hyperexpression, not needed for replication in cell culture. 245 amino acids (29KDa).
		PE protein: electron dense protein, surrounding the protein matrix. 290 amino acids.
		p10: forms fibrillar structures found in nucleus and cytoplasm of infected cell, then trapped in the occlusion matrix.

7.1.4 Aim of the protein powder diffraction studies

While microcrystals of baculovirus polyhedrin are found naturally, most of the structures of these proteins were solved from single crystals made in

vitro of recombinant protein.

A first aim of this study was to see how much information could be found from a “natural” sample. Normally protein microcrystals are studied from ultra pure conditions strictly defined. Here, as the microcrystals grow from a living insect cell, the environment is far from being humanly defined. A broad range of different chemicals, in different concentrations depending on the cell studied, is present. Salts, sugars, proteins, RNAs, among others are present, but in the end a crystal embedding only a virus is produced in this environment. Does this environment allow a diffracting crystal to be obtained, and if so, of what quality ?

Will these growth environment differences result in structural differences ? This is a very important question as protein crystals made from recombinant proteins have been believed for years to be structurally equivalent to a protein produced in its natural environment.

Also, diffraction from good protein microcrystals on a high resolution diffractometer allows better angular resolution, and therefore more accurate lattice parameters. This method can be used to compare at room temperature the lattice parameters of these crystals. This can help in understanding why several different organisms evolved the same type of crystal coating as a protection mechanism.

Also another, more concrete aim was to see if the powder diffraction method could be used as a quality control tool when baculoviruses are produced industrially as natural, “organic” pesticides.

In this study we collect information about two viruses, *Autographa californica* Multicapsid Nucleopolyhedro Virus (AcMNPV) which is the most studied baculovirus, and *Cydia pomonella* Granulosis Virus (CpGV). We determine their lattice parameters accurately, then see if it is possible to distinguish one virus structure from the other based on the observed powder patterns.

7.2 Results

7.2.1 Lattice parameter comparison

Indexing and Pawley fits were performed with Topas[30], the peak shape parameters, background and lattice parameters were then fixed and used for modelling the calculated diffraction patterns of different related viruses based on their single crystal pdb structure.

AcMNPV

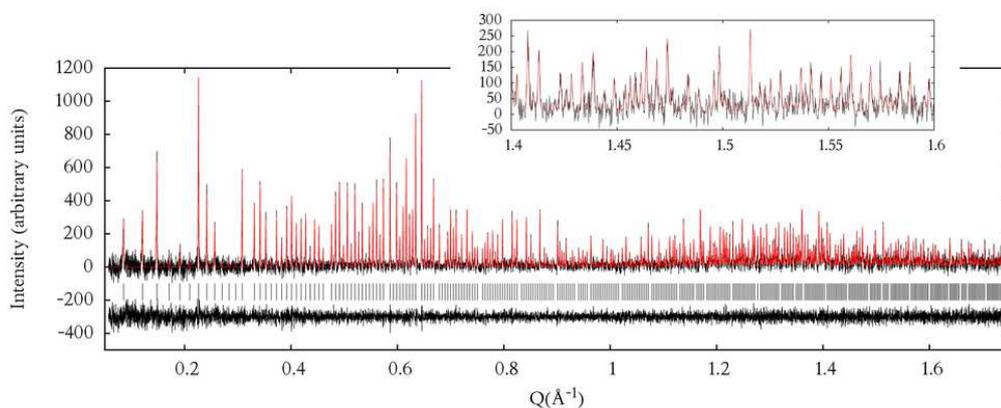


Figure 7.3: Final fit of one data set of **AcMNPV** employed for intensity extraction using the Pawley method. The data were collected at a temperature of 295 K (ID31, $\lambda = 1.53985(4)\text{\AA}$). The black, red, and lower black lines represent the experimental data, calculated pattern, and difference between experimental and calculated profiles, respectively. The vertical bars correspond to Bragg reflections of the phase. The inset corresponds to magnification of the observed and calculated profiles. The background intensity has been subtracted for clarity. Space group is $I23$, lattice parameter is $a=103.9079(5)\text{\AA}$

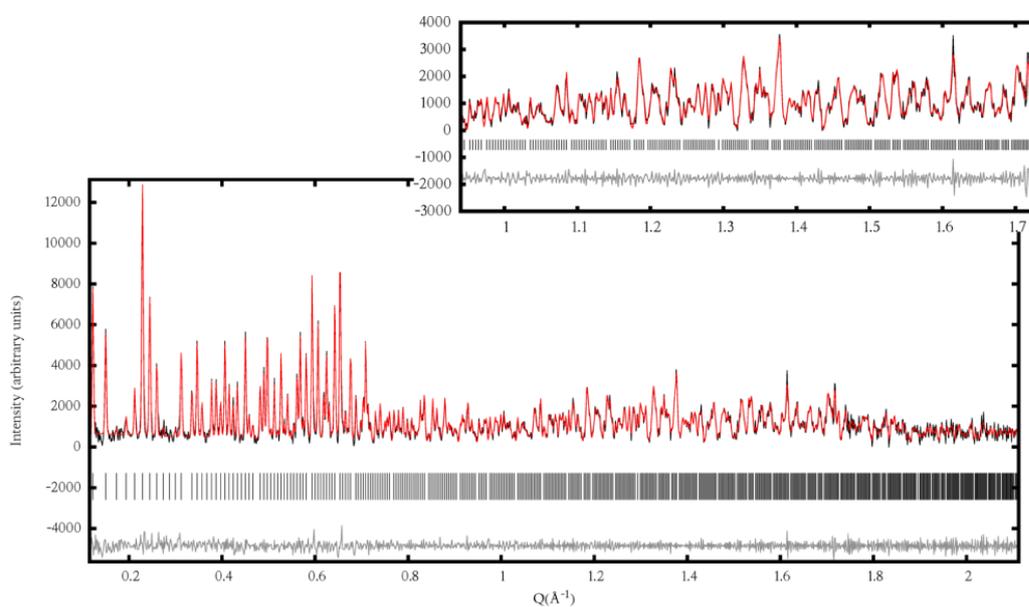
CpGV

Figure 7.4: Final fit of the **CpGV** using the Pawley method. The data were collected at a temperature of 295 K (ID31, $\lambda = 1.53985(4) \text{ \AA}$), Space group is $I23$ lattice parameter is $a = 102.6108(4) \text{ \AA}$.

AcMNPV cryocooled

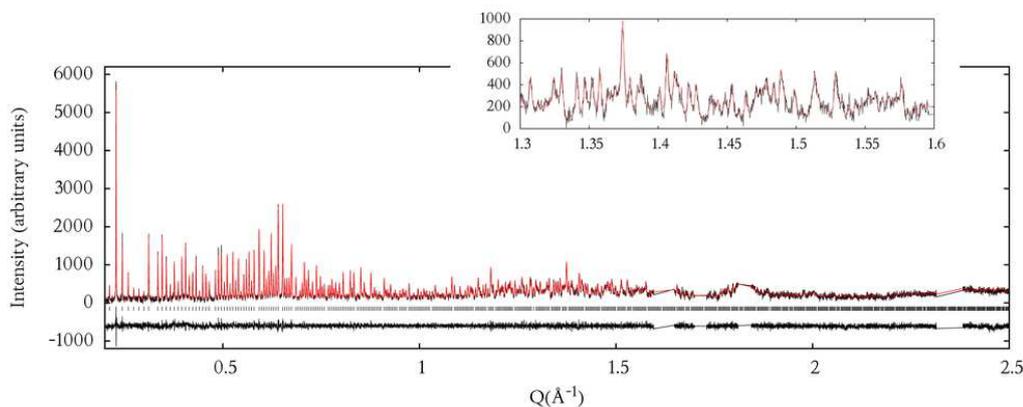


Figure 7.5: Final fit of one data set of **AcMNPV** employed for intensity extraction using the Pawley method. The data were collected at a temperature of 80 K (ID31, $\lambda = 1.30000 \text{ \AA}$), the cryoprotectant is ethylene glycol 50%. Three ice peaks were observed and excluded from the intensities extraction.

CpGV cryocooled

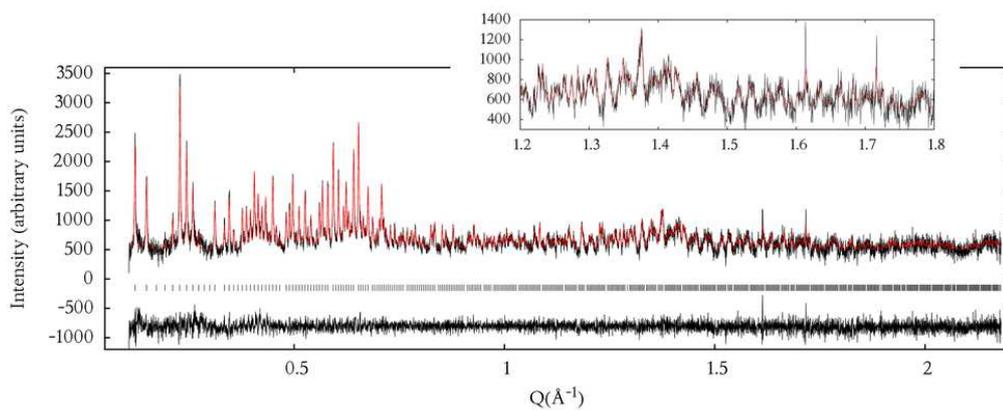


Figure 7.6: Final fit of one data set of **CpGV** employed for intensity extraction using the Pawley method. The data were collected at a temperature of 80 K (ID31, $\lambda = 1.30000 \text{ \AA}$). Space group is $I23$ lattice parameter is $a=102.64(3)\text{\AA}$.

7.2.2 Rietveld fit of different pdb structures compared to AcMNPV powder diffraction at room temperature

Several specific parameters are needed to model a powder pattern of a protein. One way to have an objective comparison of different structures would be to change only the structure and compare the R factors. The main advantage is that the only variation comes from the structure alone. The main disadvantage would be that all the others parameters have to be guessed beforehand. If the parameters like solvent scattering from the sample are refined for one structure, then fixed for the other structures, the statistics for this structure will be biased. The structure where the parameter has been adjusted will appear better. Another way would be to refine freely all sets of parameters for each structure tested. The two approaches will be tested and discussed on the AcMNPV powder pattern at room temperature. Another aspect shown here is the comparison of a background subtraction prior to the modelling, or the refinement of the background parameters as variables. The R factors of the modelling will be given for both.

With prior background subtraction

In order to reduce the number of variables of the least squares refinement and to have more comparable values between datasets, background could be subtracted from the dataset.

We used the method described in appendix E, p. 207 to subtract the background. This method determines first a window containing the intensity of the points close to the point we're studying (± 0.5). Then the value of the first quantile of the points in this window is chosen for the two theta value. Applied to all points this method will result in a crenellated function. Then, a smooth function based on this crenellated function is determined. The profile is then subtracted by the values of this smooth function and an offset is added to have the lowest point at an intensity of 0.

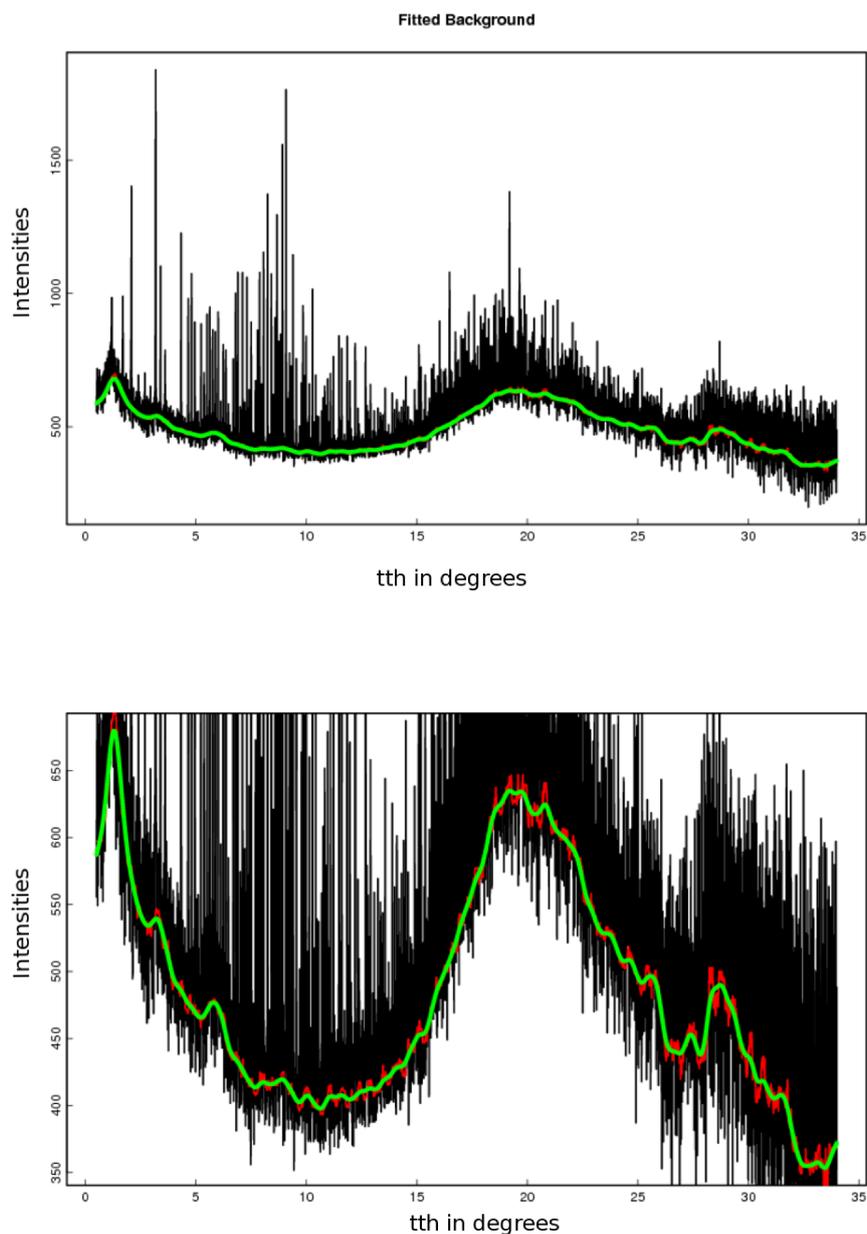


Figure 7.7: Background subtraction of AcMNPV at room temperature. The observed profile will be subtracted by the green curve, which is a smoothing of the red curve. The red curve is obtained by taking for each point the first quantile value in a window of ± 0.5 around the studied point.

Model comparison, without solvent scattering individual refinement

In order to see if the different methods of background modelling can have an impact of the interpretation of the statistics we will look at two datasets. One has a prior subtraction of the background, the other has a background modelled within the least squares software.

For both methods, the three models that fit best our datasets are 3JVB ($R_{wp}=36.034$), 2WUX ($R_{wp}=35.802$) and 3JW6 ($R_{wp}=35.397$).

For both methods, 2OH5/6/7 and 3CNF seems to form a group with worse statistics (table 7.3).

Only 3CNF gave a problem. A look at the fit reveals clearly that the bad statistics might be biased. The problem came from the first peak (figure 7.25, p. 152), which is the only one to be fitted correctly. This is typical of an incorrect scale value, correct for the first peak, but refined to too small a value for the rest of the pattern.

This is the limitation of the refinement method consisting of fixing all parameters and allowing only the structure to change.

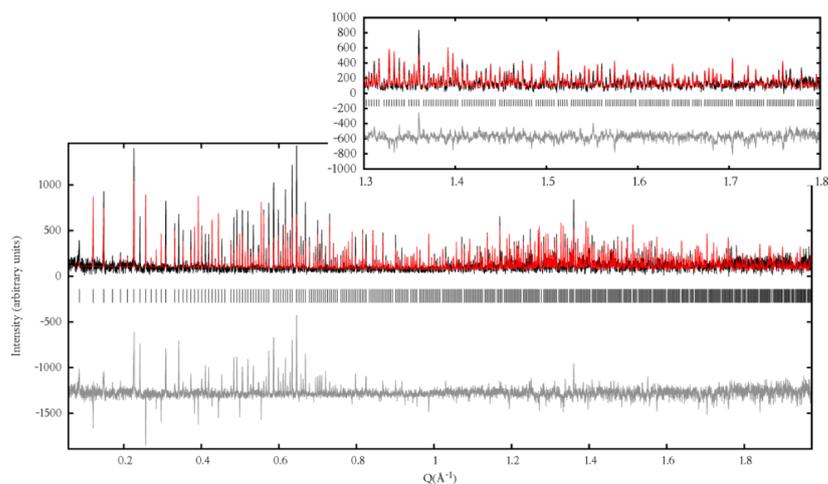


Figure 7.8: 3JVB structure, with prior background subtraction and with solvent scattering not refined.

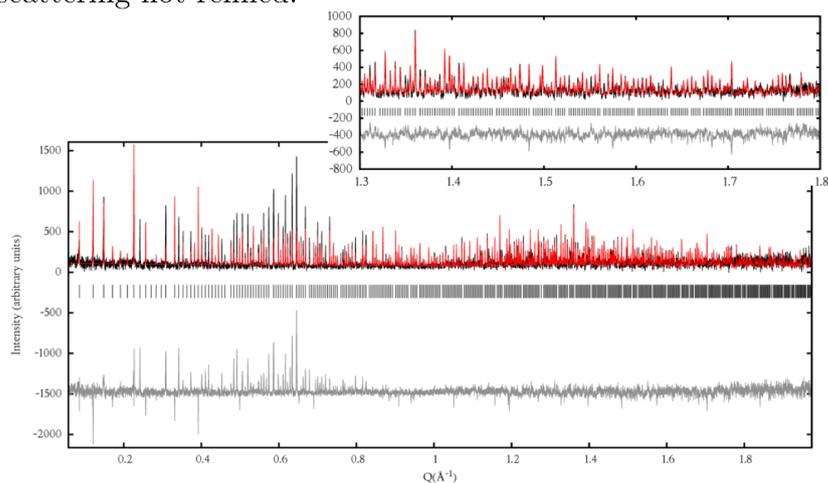


Figure 7.9: 3JW6 structure, with prior background subtraction and with solvent scattering not refined.

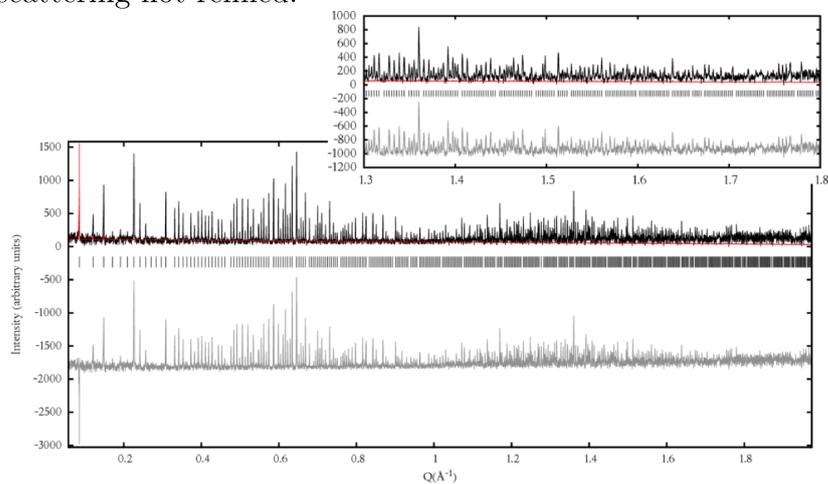


Figure 7.10: 3CNF structure, with prior background subtraction and with solvent scattering not refined.

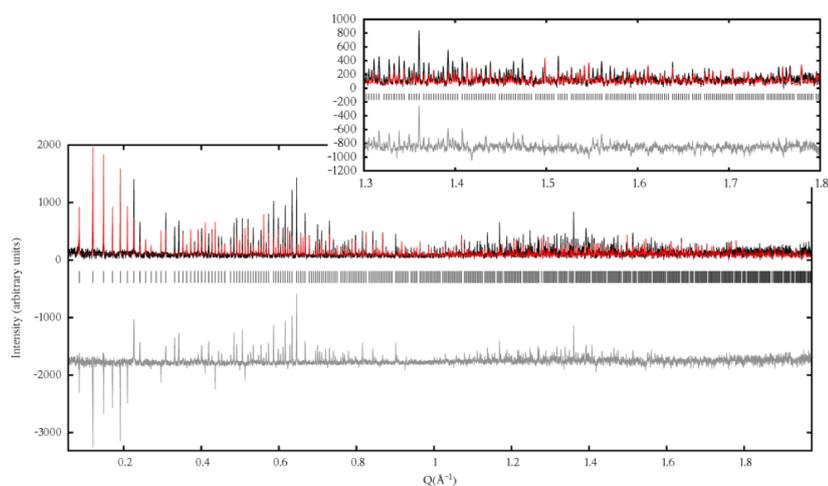


Figure 7.11: 2OH5 structure, with prior background subtraction and with solvent scattering not refined.

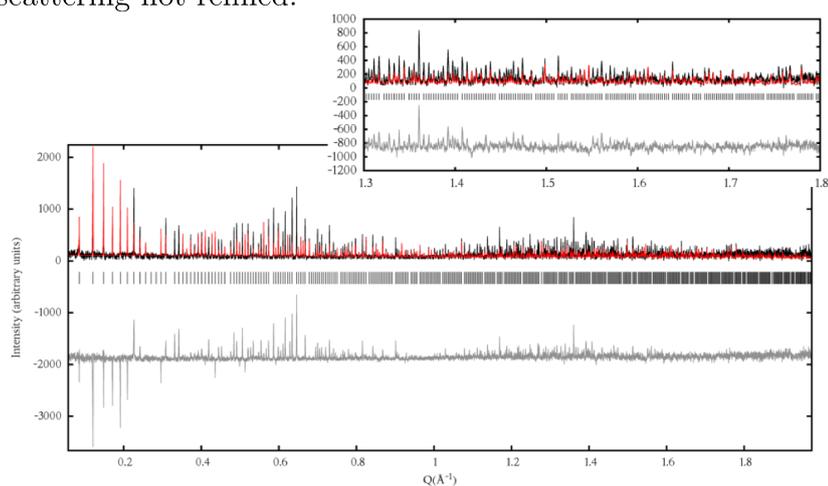


Figure 7.12: 2OH6 structure, with prior background subtraction and with solvent scattering not refined.

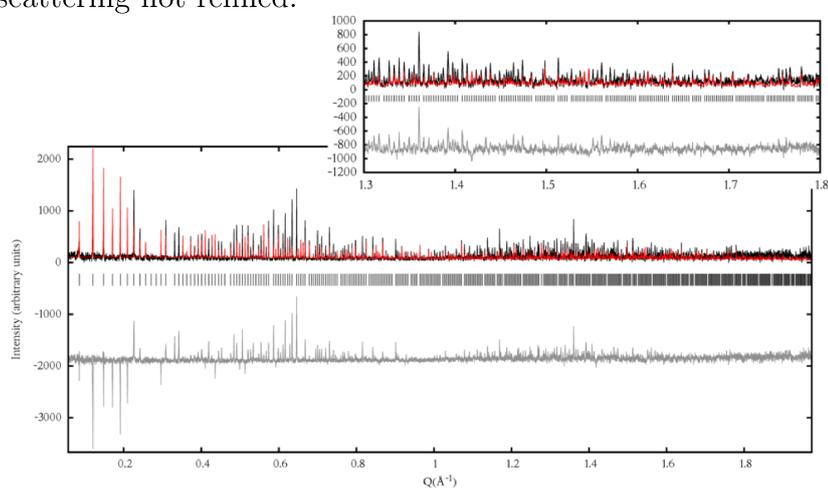


Figure 7.13: 2OH7 structure, with prior background subtraction and with solvent scattering not refined.

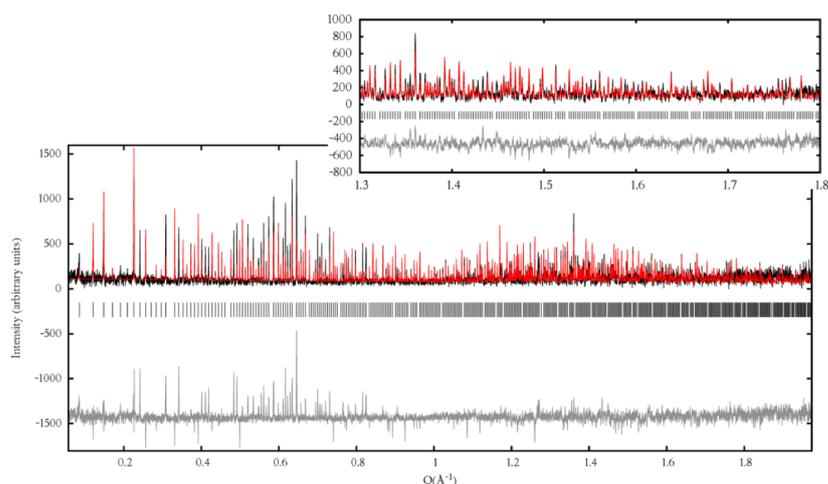


Figure 7.14: 2WUX structure, with prior background subtraction and with solvent scattering not refined.

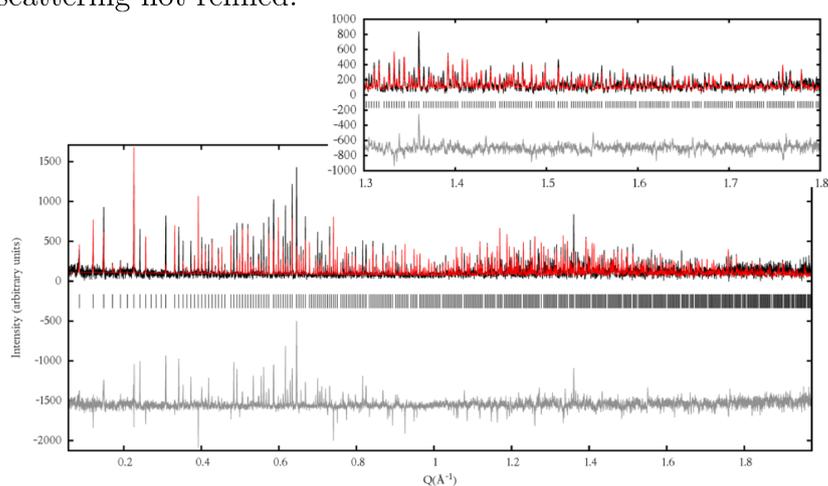


Figure 7.15: 2WUY structure, with prior background subtraction and with solvent scattering not refined.

Model comparison, with solvent scattering individual refinement

In order to avoid biasing the statistics, favouring one model which would have all the parameters ideally refined for it, we will refine solvent scattering for each individual structure and see if this would change our interpretation. Instead of having three clear groups, we can now merge the second and third groups. 2WUX, 3JVB and 3JW6 are still clearly the three structures that fit

best our pattern. 2WUY, 2OH5/6/7 and 3CNF form a group of structures where the fit is not as good.

The case of 3CNF, which is seen as a structure very different in one method of comparison, and as a structure more similar with the method will be discussed separately (figure 7.2.2, p. 153).

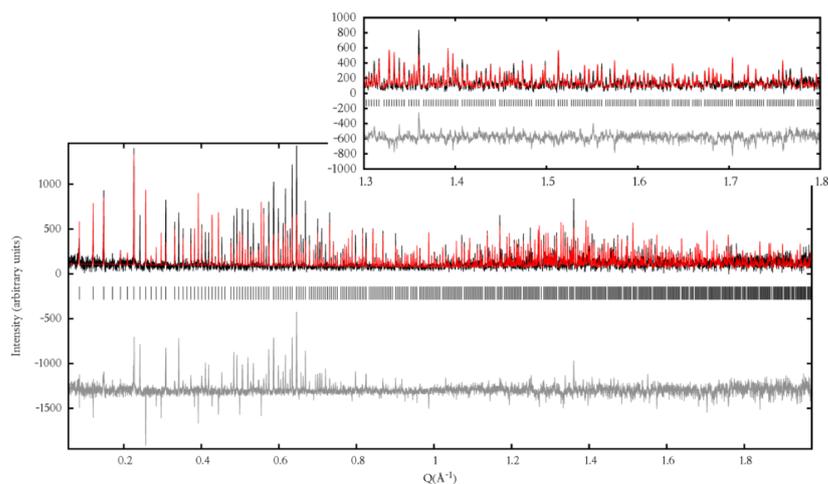


Figure 7.16: 3JVB structure, with prior background subtraction and solvent scattering refined.

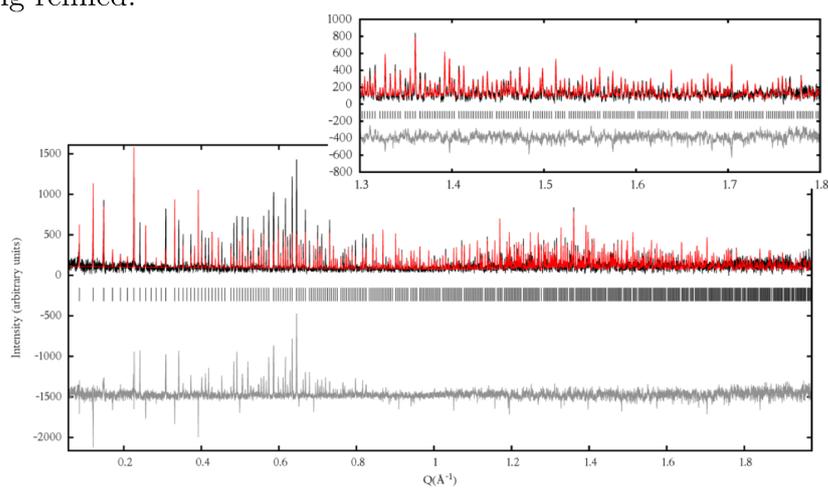


Figure 7.17: 3JW6 structure, with prior background subtraction and solvent scattering refined.

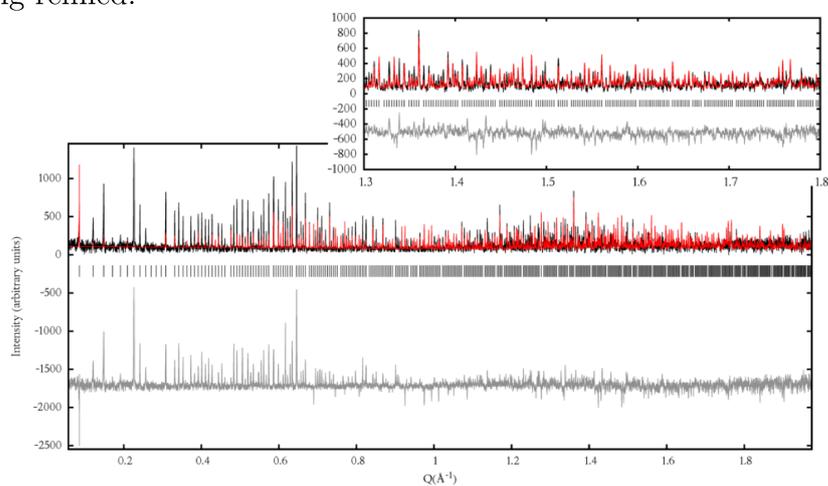


Figure 7.18: 3CNF structure, with prior background subtraction and solvent scattering refined.

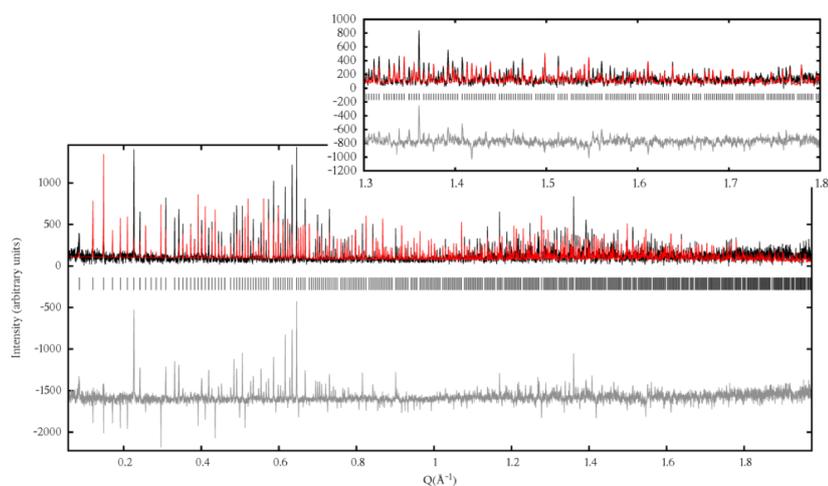


Figure 7.19: 2OH6 structure, with prior background subtraction and solvent scattering refined.

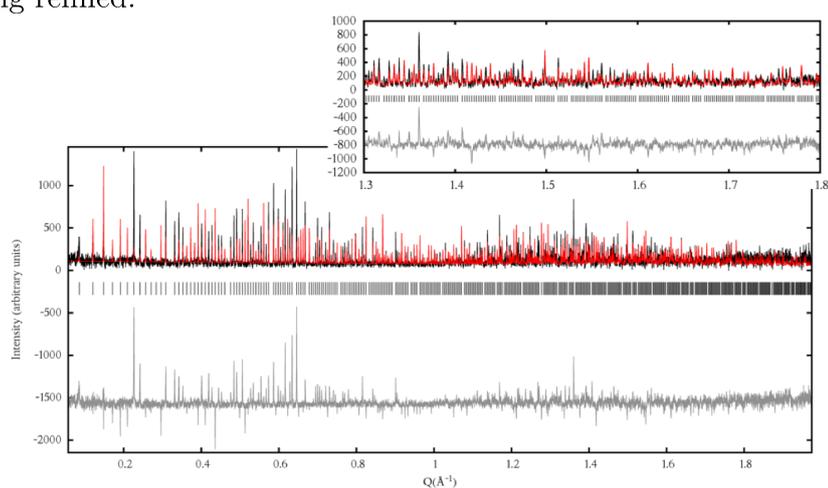


Figure 7.20: 2OH5 structure, with prior background subtraction and solvent scattering refined.

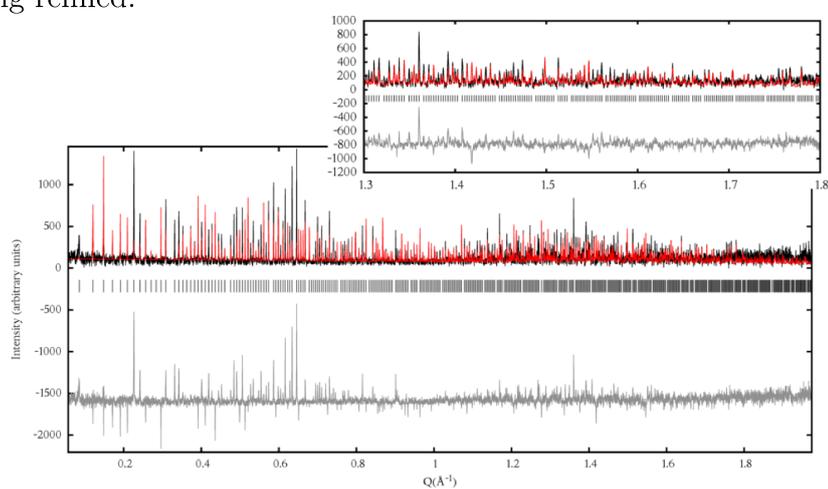


Figure 7.21: 2OH7 structure, with prior background subtraction and solvent scattering refined.

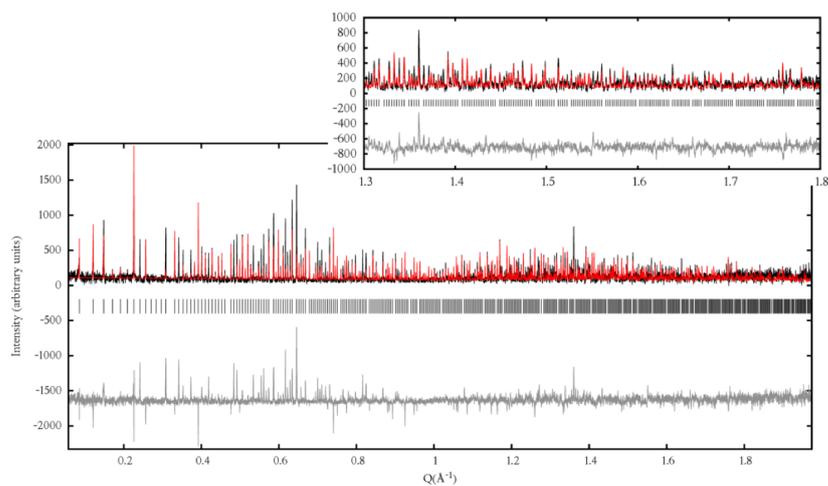


Figure 7.22: 2WUY structure, with prior background subtraction and solvent scattering refined.

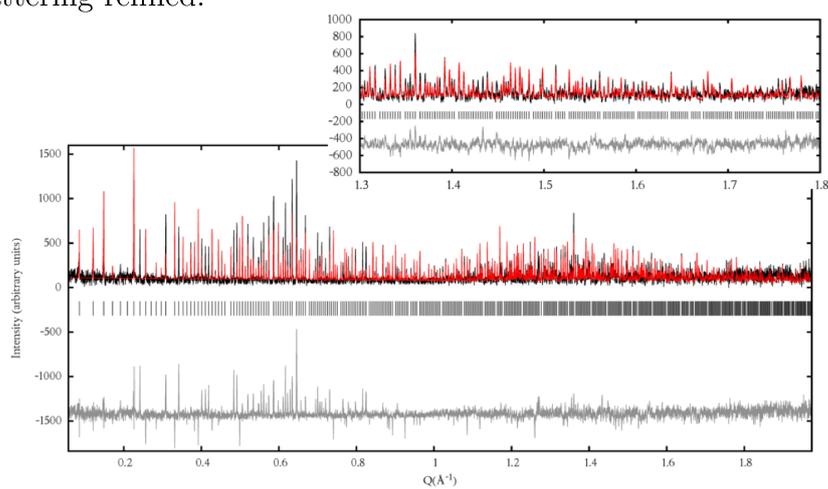


Figure 7.23: 2WUX structure, with prior background subtraction and solvent scattering refined.

PDB reference	background subtracted		background modelled	
	Rwp	Rexp	Rwp	Rexp
2OH5	45.763	19.658	12.000	5.157
2OH6	48.199	19.658	12.608	5.157
2OH7	48.919	19.658	12.794	5.157
2WUX	35.802	19.658	9.230	5.157
2WUY	39.067	19.658	10.107	5.157
3CNF	64.465	19.658	16.866	5.157
3JVB	36.034	19.658	9.353	5.157
3JW6	35.397	19.658	9.143	5.157

Table 7.3: Statistical values for different models against AcMNPV collected at room temperature. The background has been removed prior to the analysis or modelled.

PDB	Solvent scattering refined on 3JW6			Solvent scattering freely refined		
	R_{wp}	A_{Solv}	B_{Solv}	R_{wp}	A_{Solv}	B_{Solv}
2OH5	45.763	6.586(2)	62.7(7)	40.757	6.30(1)	98(2) (max)
2OH6	48.199	6.586(2)	62.7(7)	41.482	6.29(1)	99(2) (max)
2OH7	48.919	6.586(2)	62.7(7)	41.744	6.30(1)	99(2)
2WUX	35.802	6.586(2)	62.7(7)	35.642	6.559(2)	60.2(7)
2WUY	39.067	6.586(2)	62.7(7)	38.895	6.577(3)	56.3(7)
3CNF	64.465	6.586(2)	62.7(7)	43.182	5.891(5)	99(2)(max)
3JVB	36.034	6.586(2)	62.7(7)	35.903	6.622(3)	59.3(7)
3JW6	35.397	6.586(2)	62.7(7)	35.397	6.586(2)	62.7(7)

Table 7.4: Résumé of R_{wp} for different structures, with or without solvent scattering refinement.

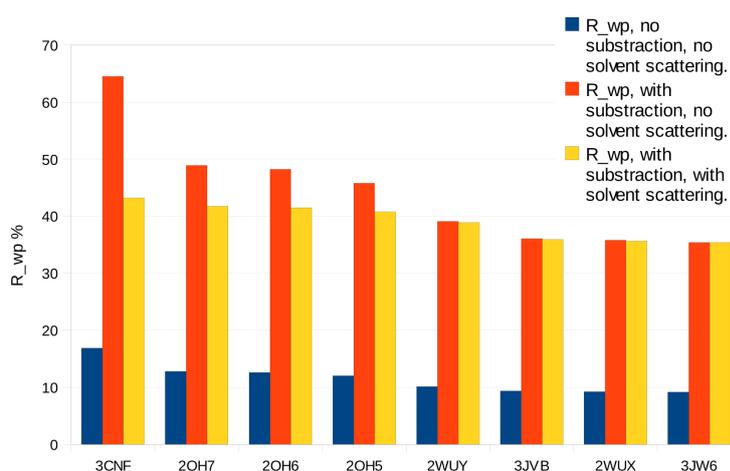


Figure 7.24: Statistics for each model simulation. The lower the better.

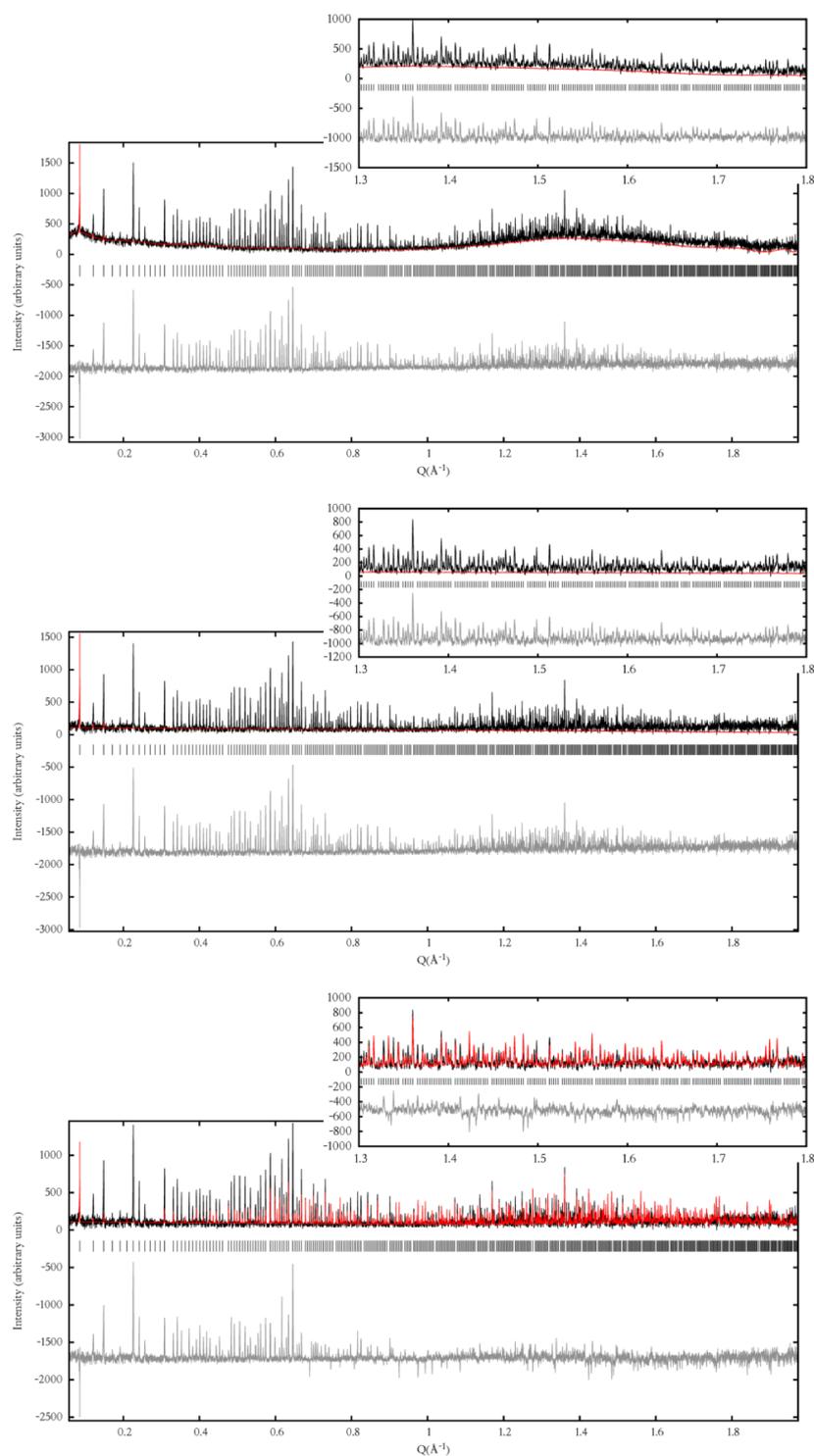


Figure 7.25: AcMNPV at room temperature with 3CNF structure. First pattern is with modelled background. Second pattern is with previously subtracted background. Last pattern is with subtracted background and solvent scattering refined. Without good solvent scattering parameter values, the scale factor is refined to a wrong value, with only the first peak fitted. It is therefore important for comparing structures to have scale factors and solvent scattering refined simultaneously.

Discussion

The single crystal structures that gave the best fit for the AcMNPV powder pattern are 2WUY, 2WUX, 3JVB and 3JW6, all AcMNPV polyhedra. Surprisingly, the crystals we received are wild type AcMNPV, but the diffracted pattern is fitted slightly better by a recombinant single crystal structure (2WUX) than a wild type (2WUY).

It is interesting to note that if the solvent scattering parameters are not refined for each structure, but fixed, this could lead, like for the 3CNF, to a misleading conclusion. In the 3CNF case without solvent scattering the scale factor refined to fit the most intense peak (figure 7.25, p. 152). Statistics and the fit wrongly implied that the structure was very different. After refining the solvent scattering parameters for each structure independently, it was shown that the structure is just a little more different than the other non-AcMNPV structures.

It is possible even in a most difficult case like cubic symmetry with a powder pattern of one polyhedron, to distinguish between very close homologous structures. Care must be taken when comparing these structures to refine scale factor and solvent scattering for each.

7.2.3 Structural variation between natural and recombinant protein

AcMNPV

The Rietveld fit of the AcMNPV single crystal structure compared to the powder pattern collected on natural microcrystals shows differences. While the high angle part of the pattern is well fitted, the low angle part is less concordant.

Several parameters can be the cause, low angle misfitting indicates a problem in the low resolution shell. The protein can be slightly misfitted in the cell, or have some small shift between domains of the protein structure, or the solvent scattering model can be improved.

A rigidbody refinement is attempted in order to improve the fit at low angle.

Rigid body refinement A rigid body refinement can be performed with Topas directly against the powder data, without changing the molecular structure. This is done using the method described in F.5, p. 221. Scale and solvent scattering parameters are freely refined in order to avoid bias over

one structure.

The 3JW6 structure is used as a test case of a rigidbody refinement of the AcMNPV data at room temperature.

A refinement of 9 parameters, (3 for rotation, 3 for translation, 2 for solvent scattering and 1 for the overall scale), allowed only a small improvement of the fit of the low angular region:

The translation in the cell were 0.11% along the a axis, 0.16% along the b axis, 0.42% along the c axis, rotations were quasi absent (0.006% maximum). These displacements represents less than 0.5\AA , the resulting improvement is considered to be insignificant and is probably due to the added variables.

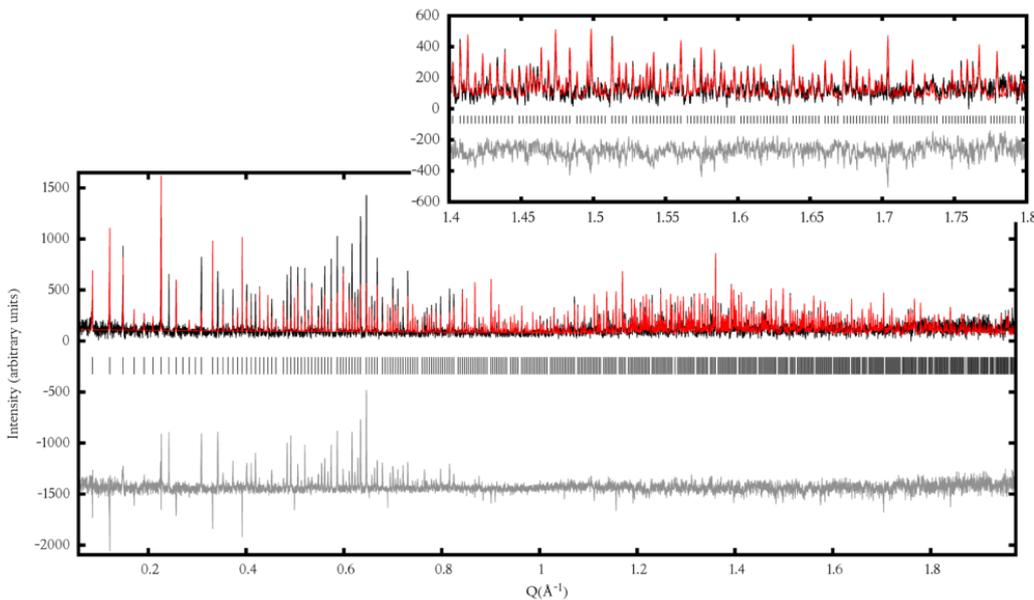


Figure 7.26: Rietveld fit of the 3JW6 structure to the AcMNPV powder pattern, after a small rigid body refinement. Small improvements at low angles are noted, but statistically not significant, due to the added parameters.

CpGV

Correlation between extracted intensities By looking at correlations between extracted intensities ($\frac{I_{single\ crystal} - I_{powder\ extracted}}{I_{single\ crystal}}$), as it is often done between intensities from different single crystals, we miss the information in the overlapped intensities. In this case the intensities distributed under one

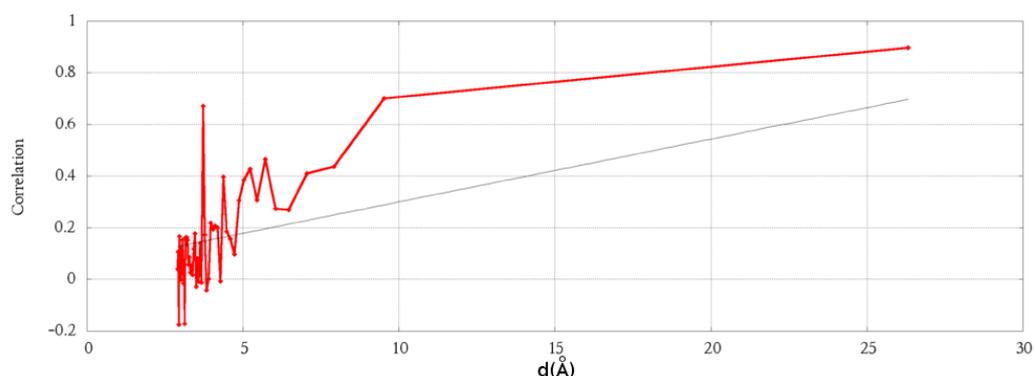


Figure 7.27: Correlation between single crystal extracted intensities from Peter Metcalf (model unpublished) and powder extracted intensities. Reflections are divided in 50 bins, and correlation is made as $\frac{I_{single\ crystal} - I_{powder\ extracted}}{I_{single\ crystal}}$.

peak are not known, only their observed sum. This worsens the correlations between powder and single crystal intensities at high angles.

Modelling the single crystal structure as a powder pattern will rescue the repartition of the overlapped peak intensities based on the structure. This will show if the absence of correlation is due to bad modelling of the high resolution shell, or to the inexact repartition induced by the extraction of intensities without structural knowledge.

We tried a rigid body refinement of the whole protein within the unit cell using the Rietveld method, which leads to a insignificant improvement of the fit of the CpGV model.

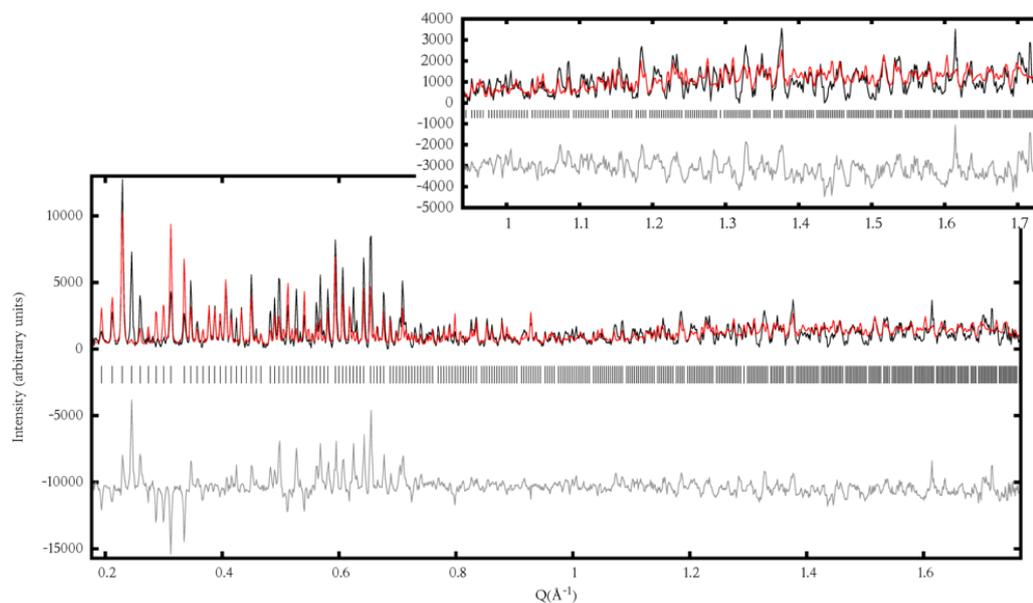


Figure 7.28: CpGV single crystal structure (current single crystal model from Peter Metcalf, unpublished) Rietveld fit compared to powder data, after a rigid body refinement. A very small improvement of the fit is seen.

Rietveld refinement We didn't succeed to perform a complete Rietveld refinement of the individual atomic positions neither for AcMNPV nor for the CpGV. While the powder pattern obtained is of high quality, the symmetry here poses a problem.

In a cubic system, the systematic reflection overlap is very high and it is not possible to decorrelate the reflections by varying the sample preparation in order to change lattice parameter. In the high angle part of the diffracted pattern, for example at a dspace of 3.53\AA , there is an exact overlap of 23 reflections (table 7.5 p.157). There is no possibility to decorrelate experimentally these reflections, as when the cell parameters will change, the reflections will be still overlapped.

Even the information collected on the single crystal was known to be difficult to interpret (low contrast due to small amount of solvent, maps more biased by the cubic symmetry (figure 7.29, p. 157)). An attempt to obtain initial phases by molecular replacement on extracted intensities didn't succeed. A powder Rietveld refinement, based on an initial single crystal model, would have been extremely complicated, and in fact not worth the trouble, as single crystal data were already obtained.

Ref. nb.	h	k	l	d-space	tth	Ref. nb.	h	k	l	d-space	tth
2392	29	0	5	3.53093	25.18930	2404	21	20	5	3.53093	25.18930
2393	0	29	5	3.53093	25.18930	2405	16	23	9	3.53093	25.18930
2394	4	29	3	3.53093	25.18930	2406	24	13	11	3.53093	25.18930
2395	28	9	1	3.53093	25.18930	2407	13	24	11	3.53093	25.18930
2396	9	28	1	3.53093	25.18930	2408	23	16	9	3.53093	25.18930
2397	24	17	1	3.53093	25.18930	2409	19	21	8	3.53093	25.18930
2398	27	11	4	3.53093	25.18930	2410	29	4	3	3.53093	25.18930
2399	11	27	4	3.53093	25.18930	2411	21	19	8	3.53093	25.18930
2400	17	24	1	3.53093	25.18930	2412	21	16	13	3.53093	25.18930
2401	15	25	4	3.53093	25.18930	2413	16	21	13	3.53093	25.18930
2402	25	15	4	3.53093	25.18930	2414	19	19	12	3.53093	25.18930
2403	20	21	5	3.53093	25.18930						

Table 7.5: Cubic symmetry at high angle show exact overlap of numerous reflections. Under one peak at $2\theta = 25.18930$ there are 23 symmetry independent reflections (ignoring friedel pairs).

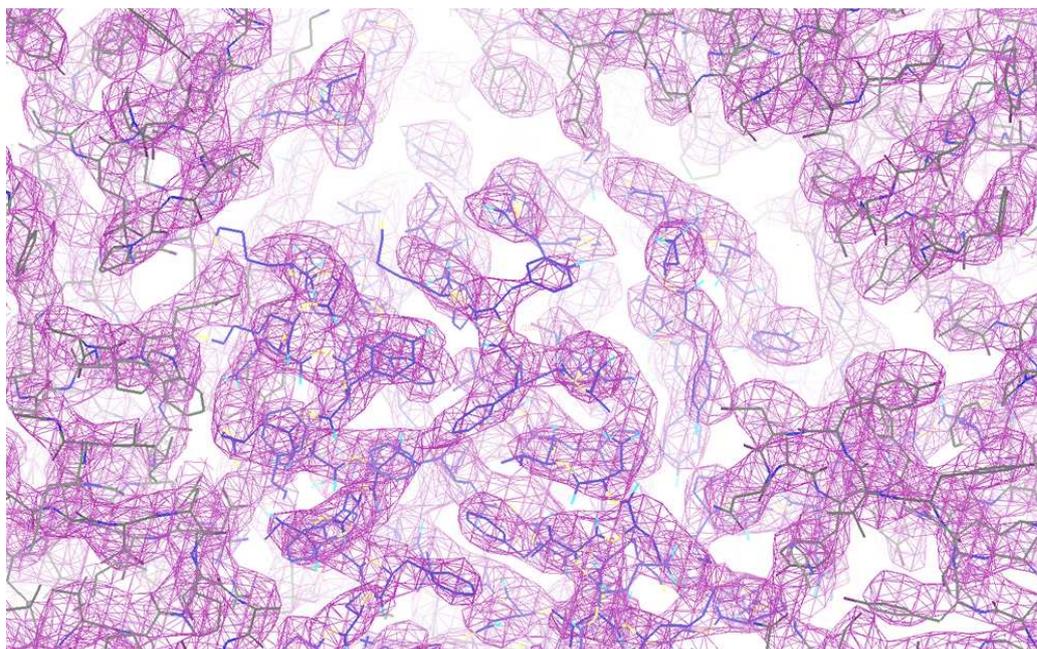


Figure 7.29: Omit maps from a Rietveld refinement are too much biased by the model. It is not exploitable in finding problematic regions in order to explain the structural differences observed in the powder pattern.

7.3 Discussion

The lattice parameters of two different viruses have been determined using powder diffraction on a high resolution powder diffractometer.

AcMNPV at room temperature	a=103.9015(5)
AcMNPV cryocooled	a=103.9078(4)
CpGV at room temperature	a=102.6108(4)
CpGV cryocooled	a=102.64(3)

It is interesting to relate that the lattice parameters of the crystalline matrix protecting the viruses are very close between a nuclear polyhedrosis virus and a granulovirus. Smaller changes than with other proteins are observable when the sample is cryocooled, minimising the impact of the method. Nonetheless, the loss in quality for the cryocooling of CpGV results in a lower precision on the lattice parameter determination. As several other viral groups (CPV, EPV) have converged towards the same protection mechanism, it would be interesting to compare more species at room temperature.

We saw also that it was possible to identify correctly, by using single crystal pdb structures, which virus was collected. This can be interesting for industrially producing these viruses in large quantities as a control tool. As manufacturers are using live insects for production, it is important to check if the virus granulin produced is correct, as it will probably have an impact on conservation. We saw that the background can be subtracted before modelling without influencing the choice of the structure. We saw also that it was preferable to refine for each individual structure fitted, the solvent scattering parameters and scale factor, as fixing them would bias the resulting choice.

When analysing differences between single crystal structures compared to recombinant protein and natural microcrystalline powder we saw differences. Intensities extracted from a Pawley refinement were difficult to interpret due to the absence of knowledge of the repartition of overlapped reflections. The correlation in this case seemed poor. Modelling the structure and refining via a rigid body of the whole protein in the cell didn't improve the fit significantly. Even considering the d-space resolution obtained with AcMNPV (3Å), it was not possible to perform a complete Rietveld refinement of the atomic positions. This is probably due to the increased difficulty raised by cubic symmetry in powder diffraction, having a high degree of systematic peak overlap.

It is therefore difficult to conclude on the difference between the recombinant

protein and the virus powder naturally found in the infected insect. Differences in the intensities seemed small, but only a complete refinement will give a clear picture of any structural differences.

8

Conclusions and further work

Obtaining a protein structure using crystallography is a long and expensive process which requires much effort and time. It is still believed by many structural biologists that crystallography can be done only using information collected on a high quality single crystal. However we saw that many protein precipitates usually discarded during the search for a single crystal crystallisation condition can be used for powder diffraction.

One major limitation to this technique was the amount of sample required to perform a powder diffraction experiment. Obtaining sufficient information using a high angular resolution powder diffraction beamline such as ID31 was requiring large volumes of microcrystals. In order to solve this critical problem we developed new protocols for sample preparation and data acquisition.

New protocols allowing a rapid extraction of the microcrystals have been developed (annex B), which allow microcrystals to be separated in large quantities from their mother liquor while keeping them protected from dehydration. Protocols for collecting at cryocooled conditions using capillaries have been developed (chapter 6). Protocols for data acquisition on PX beamlines at room temperature on a very small amount of sample (typically of the size of a precipitate obtained in a single crystallisation drop) have been defined using a humidity controller to keep the sample hydrated (annex C). Also, adaptation of existing powder software has been made to help the analysis of a protein powder pattern by non experts (annexes E and F).

All these developments have made it possible to study many different systems and demonstrate the broad applications in which protein powder diffraction can be used.

We have shown that it is possible to cryocool a protein powder sample, extending life time in an intense X-ray beam and thus the maximal d-spacing resolution observable. The information collected under cryocooled conditions allowed us to refine the structure of two different forms of human insulin by using the Rietveld method (chapter 6).

We have shown that powder diffraction data collected on high quality microcrystals were sufficient for refining large protein structures such as Urate oxidase (chapter 4) at moderate resolution (3.6Å).

We have also shown that powder diffraction on proteins can be used for example by pharmaceutical companies for polymorphism studies (chapter 3) allowing the possible development of new crystalline forms of proteins suitable for intravenous administration.

We have shown also that it is possible to obtain preliminary structural information by collecting on only a single urchin of Mayaro virus macro domain (chapter 5).

In these successful studies we saw that powder diffraction on proteins can pass from an occasional to a more routine technique in structural biology.

However limitations still exists for extending the use of the technique to even more difficult systems. Some limitations depend on the crystallographic phase of the protein studied. We saw that for the cubic crystalline matrix of the baculoviruses (chapter 7) the structures were very difficult to re-determine even with high quality diffraction profiles collected under cryocooled conditions.

Also, whereas indexing and solving a protein structure by molecular replacement are possible with high angular resolution powder diffraction patterns, it becomes even more challenging with low angular resolution powder patterns such as those obtained with 2D detectors (chapter 5). In these cases, molecular replacement and structural refinements became much more difficult and uncertain.

Also, whereas the determination of powder crystallisation conditions are much easier reducing the time spent in the laboratory preparing samples, the amount of time spent to solve and refine the structure is still substantial and of the same time scale as when dealing with low resolution data from a protein single crystal.

While it is very important to stress out that the main use of the pro-

tein powder diffraction technique stay in polymorphism studies rather than structure determination, research must be done to optimise further the experimental methods to solve protein structures, such as for example finding new bigger complexes for multiple wavelength anomalous dispersion techniques.

Also, new cryocooling protocols must be elaborate to obtain better angular resolution and lower quantities of cryoprotectant. In addition, a great potential for routine use of the technique on PX beamlines has been revealed and should be developed in the near future. This will require improvement of the software used for analysis in order to have better integration of powder data within macromolecular crystallographic suites of software such as ccp4 and phenix.

Finally it is important to note that all the research presented here requires the use of powerful X-ray sources combined with instruments of high quality available only at large research facilities.

Conclusions

Obtenir la structure d'une protéine est un processus complexe, long et coûteux, requérant beaucoup d'efforts. De nombreux biologistes structuraux pensent encore qu'une expérience de cristallographie ne peut être faite uniquement sur un monocristal. Cependant, nous avons vu que de nombreux précipités cristallins, souvent mis de côté lors de la recherche d'une condition de cristallisation, peuvent être utilisés pour une expérience de diffraction de poudre.

La principale limite de cette technique était la quantité d'échantillons requise pour réaliser une expérience de diffraction de poudre. Un très grand nombre de microcristaux étaient nécessaires pour réaliser une expérience sur des instruments à haute résolution angulaire comme ID31. Afin de résoudre ce problème critique nous avons développé de nombreux protocoles pour la préparation et l'acquisition de données.

De nouveaux protocoles autorisant une extraction rapide des microcristaux ont été développés (annexe B), séparant les microcristaux de leur liqueur mère en grande quantité, tout en les protégeant d'une déshydratation néfaste. Des protocoles permettant l'acquisition de données à partir d'échantillons refroidis à des températures cryogéniques ont été mis en œuvre (chapitre 6). Des protocoles d'acquisition de données de poudres sur des lignes de diffraction de monocristaux de protéines ont été développés. Ces protocoles permettent

de collecter à température ambiante des données sur des échantillons de très petites tailles, typique des précipités observés dans une seule goutte de cristallisation, tout en maintenant l'échantillon hydraté (annexe C). De plus, des adaptations de logiciels informatiques existants ont été réalisées, permettant l'étude de clichés de diffraction de poudre par de non-experts (annexe E et annexe F).

Toutes ces évolutions techniques ont permis d'étudier de nombreux différents systèmes et de démontrer le large champ d'application dans lequel la technique de diffraction de poudre peut être utilisée.

Nous avons montré qu'il était possible de refroidir à des températures cryogéniques un échantillon de poudre de protéine, allongeant son espérance de vie un intense faisceau de rayons-X et améliorant la résolution maximale observée. Les informations collectées ont permis d'affiner la structure de deux formes d'insuline humaine en utilisant la méthode Rietveld (chapitre 6).

Nous avons montré qu'il était possible d'affiner à l'aide de données de poudre des structures de protéines telles que l'Urate oxidase à moyenne résolution (3.6Å).

Nous avons montré également que la diffraction de poudre de protéine pouvait être utilisée par des industries pharmaceutiques pour réaliser des études de polymorphisme afin d'orienter un processus de développement de nouvelles formes injectables de protéines.

Nous avons vu qu'il était possible d'obtenir des informations structurales préliminaires en collectant des données sur seulement un cristal en forme d'oursin du domaine macro du virus Mayaro.

Nous avons montré par l'ensemble de ces études que la diffraction de poudre de protéine pouvait quitter son statut de technique occasionnelle pour devenir une technique de routine en biologie structurale.

Cependant, certaines limites existent toujours dans l'optique d'une utilisation de la technique pour des problèmes encore plus complexes. Certaines limitations sont dues à la phase cristallographique de la protéine étudiée. Nous avons vu que lors de l'étude de la matrice cristalline des baculovirus, la redétermination de la structure de la protéine pouvait être très difficile, malgré l'utilisation de données à haute résolution obtenues à des températures cryogéniques.

De plus, bien que l'indexation et la résolution d'une structure par remplacement moléculaire sont souvent possible avec des données à haute résolution angulaire, cela devient plus ambitieux avec des données à plus basse résolution telles celles obtenues avec des détecteurs 2D. Dans ces cas, le remplacement moléculaire et l'affinement de la structure deviennent plus compliqués et in-

certain.

Enfin, bien que la détermination des conditions de cristallisation de poudres sont plus faciles et que le temps passé en laboratoire en est d'autant réduit, le temps nécessaire pour résoudre et affiner une structure est toujours très long et du même ordre de grandeur que lors de l'analyse de données à basse résolution d'un monocristal.

Afin de résoudre ces problèmes, de nouvelles recherches doivent être faites afin d'optimiser encore plus les méthodes expérimentales de résolution de structures de protéine, en trouvant des plus grands complexes pour la technique de dispersion anormale à plusieurs longueurs d'onde par exemple.

L'amélioration des logiciels utilisés pour l'analyse doit également être poursuivie afin d'avoir une meilleure intégration des données de poudre dans les suites logiciels tels que ccp4 et phenix.

De plus, un grand potentiel pour une utilisation de routine de la technique de diffraction poudre sur une ligne de diffraction de monocristaux de protéine a été révélée et devrait continuer à être développée dans un futur proche.

Finalement, il est important de noter que toutes les recherches présentées dans cette thèse nécessitent l'utilisation de puissantes sources de rayons-X combinées à des instruments de très haute qualité, disponible uniquement dans de grands centres de recherche.

A

Sample preparation for room temperature experiments

A.1 Glass capillary

Glass capillaries are the most common sample holders for powder diffraction. Here is described the method we currently use on ID31 for room temperature. This preparation method is easy to perform, allows spinning on the diffractometer, provides a relatively low background for most of the systems and its cost is reasonable.

For protein powder diffraction some precautions need to be taken during sample preparation. First, microcrystals need to be packed at the end of the capillary to ensure a measurable diffracted signal. Then the excess of the mother liquor needs to be removed to ensure that the crystals will not move away from the sealed end of the capillary. Also, a thin layer of mother liquor must be left to avoid dehydration of the sample. In the end, sealing of the open end of the capillary requires also precautions.

1. Microcrystals are gently centrifuged to the bottom of the tube used for crystallisation.
2. Microcrystals are then pipetted into the funnel of the glass capillary (if the sample is sticky, cut the end off the pipette tip to enlarge the hole).

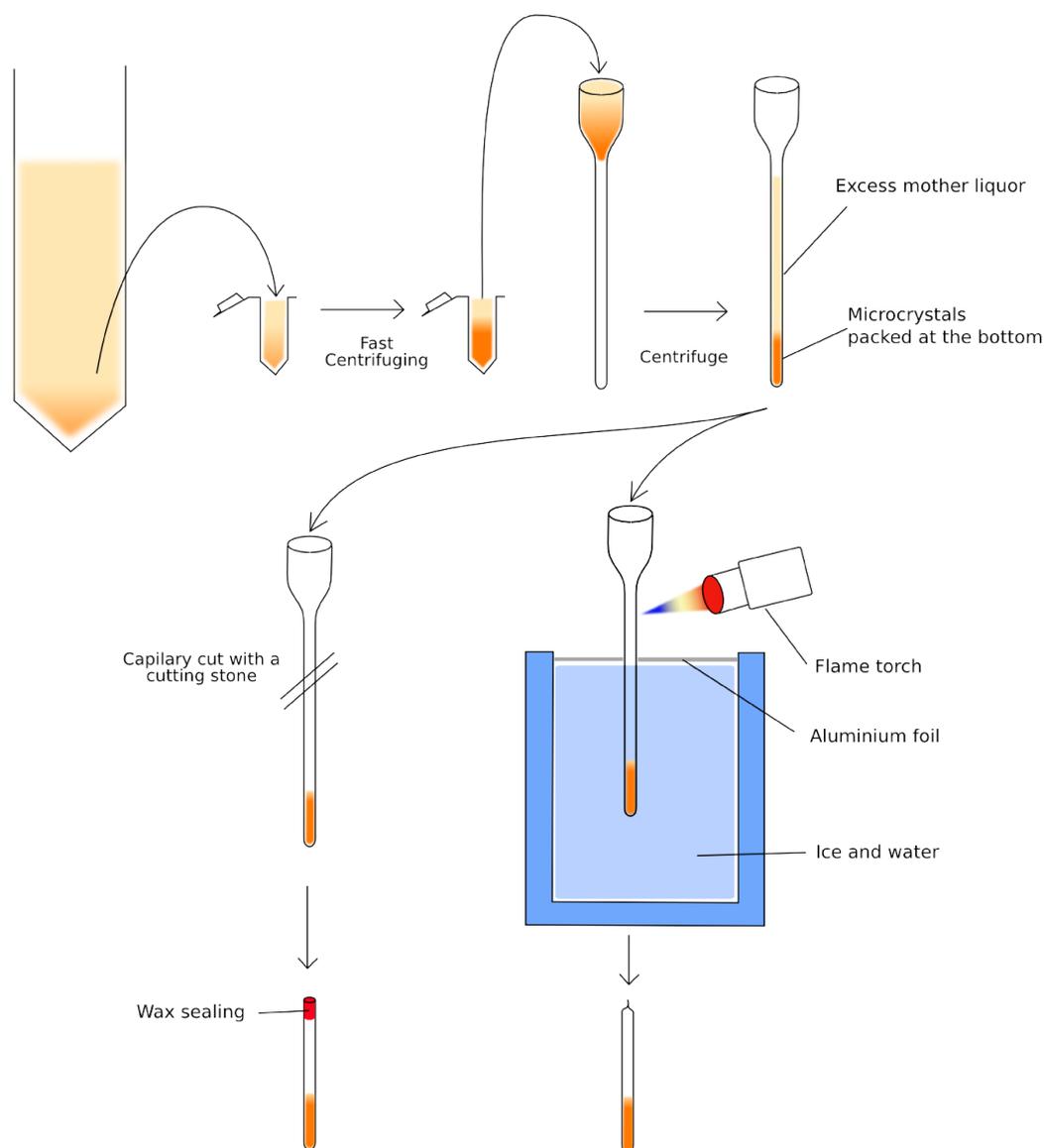


Figure A.1: Protocol for sample preparation at room temperature. Two different sealing methods for glass capillaries are presented, sealed with wax, or flame sealed.

3. Capillary is centrifuged in order to pack the crystals to the bottom of the capillary.
4. Excess mother liquor is removed with a syringe, while keeping a small layer of mother liquor to avoid any dehydration.

5. Capillary is then sealed :
 - (a) with wax
 - i. Gently rub a Hampton Research capillary stone on the outside of the capillary to cut it without breaking the whole capillary by applying too much pressure.
 - ii. Seal with beeswax.
 - (b) by flame
 - i. Put the bottom of the capillary inside a bucket of ice and water to keep cool the protein during sealing.
 - ii. Flame the capillary while pulling the funnel.

A.2 Silson membrane

A Silson silicon nitride membrane window consists of a square silicon nitride membrane in a square supporting frame.

For ID31 we chose a window size of 3 mm x 3 mm, slightly larger than our typical beamsizes of 2 mm x 1 mm. The window thickness is 500 nm (available from 50 nm to 1000 nm).

The sample is pipetted directly on top of the window. But as a protein sample needs to avoid dehydration, a second Silson membrane is added on top and sealed with grease.

This double silicon nitride window allows collection while spinning (as with a capillary).



Figure A.2: Though such large volumes are unusual, protein microcrystals can be grown in large quantities by batch crystallisation.



Figure A.3: A typical 1.5 mm capillary mounted on an ID31 brass holder. The brass holder allows use of the spinner mounted on the axis of the diffractometer, but also the sample changer robot.

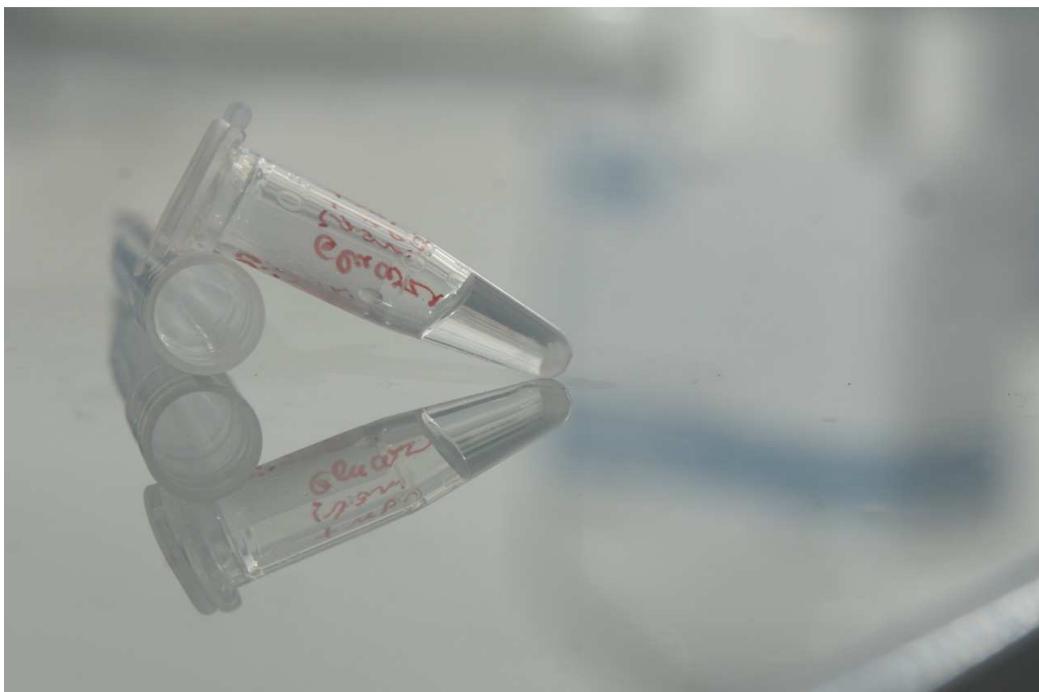


Figure A.4: Microcrystals are quickly centrifuged in their crystallisation tube prior to capillary transfer.



Figure A.5: A capillary centrifuge is used to pack microcrystals to the bottom of the capillary.

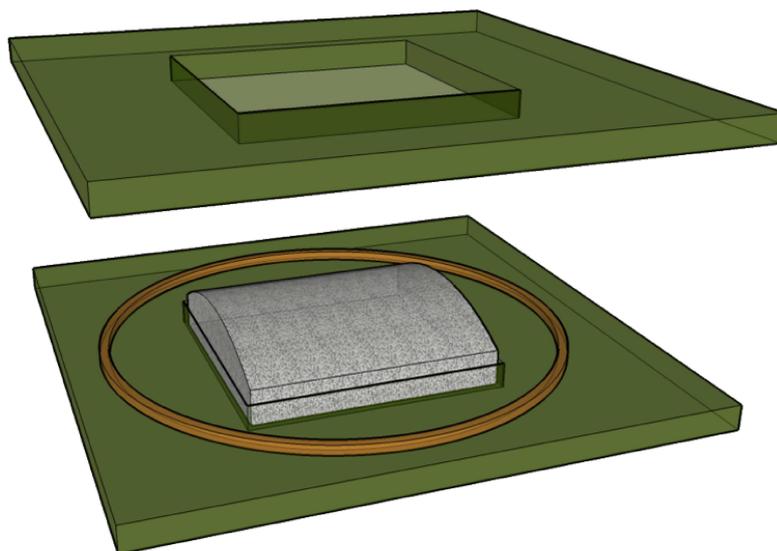


Figure A.6: A sample is prepared by adding the crystals onto the window of a first Silson membrane. Then a second window is added on top. Both windows are held together by grease, sealing the sample which avoids dehydration.

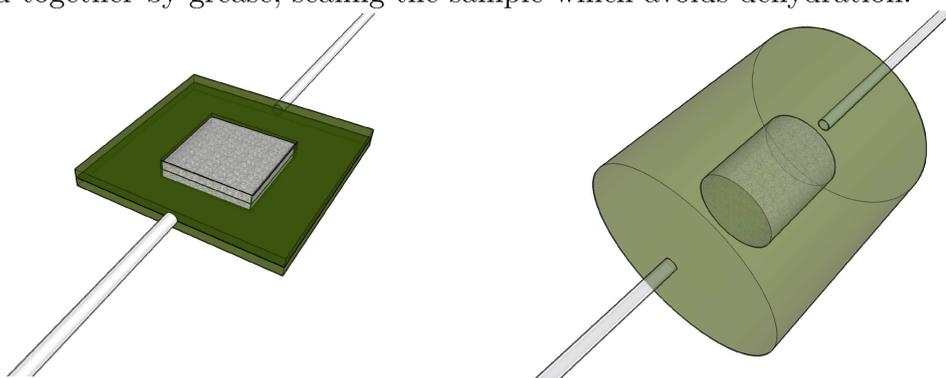


Figure A.7: The sample can be rotated if necessary.

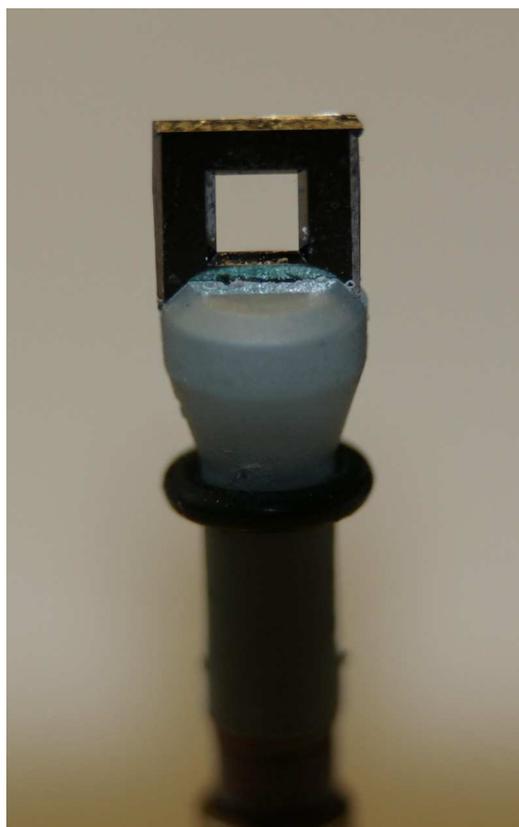


Figure A.8: A Silson membrane on a sample holder, adapted from a scapel blade holder, used on the ID31 diffractometer.

B

Extraction of the microcrystals from the mother liquor

The goal of this annex is to show two possible protocols able to reduce the background contribution due to the amorphous protein remaining in the mother liquor surrounding the sample.

One of the major difficulties with protein powder samples comes from the fact that it is not conceivable to fish manually all the crystals and put them on the sample holder. Taking account of the number of crystals to manipulate, automated procedures need to be considered, and new protocols need to be defined. For most of the projects, a simple centrifuging is enough to remove the excess of mother liquor from the sample. In some projects, improvement of the quality of the signal measured was seen by excluding from the sample the excess mother liquor, containing, among other things, amorphous proteins.

Care must be taken not to dry the crystalline part of the sample; if filtered for example, the resulting sample is always dehydrated and does not diffract any longer. Undesirable constituents of the sample must be removed without damaging the crystals. We explored two ways of doing this; either by migrating the amorphous protein out of the sample, like we did with the electrophoresis protocol (figure B.1, p. 176); by isolating the crystals from the mother liquor by centrifuging them through high density oil (figure B.2,

p. 176).

B.1 Electrophoresis

Native Gel Electrophoresis is a technique used mainly when different proteins need to be separated without being denatured and therefore separated only by their charge-to-mass ratio.

Native gel electrophoresis does not use a charged denaturing agent (no use of SDS). Therefore the protein in the crystals will not be denatured. The “free” proteins from the mother liquor solution will move through the agarose gel pores, while the crystals will remain at the interface with the gel.

The set-up presented here consists of a kapton capillary loaded at both ends with agarose and containing the sample previously packed by centrifuging in the middle.

High voltage is applied to both ends to promote migration of the amorphous protein out of the sample through the agarose.

This method is very delicate to prepare, and many hazards can occur, like for example, the sample being in contact with the liquid of the electrophoresis tank, resulting in a change of mother liquor conditions. Also an increase in temperature in the capillary is observed and could lead to dissolving crystals, etc...

This method can be of use in some projects, but has been mainly replaced now by the separation using heavy oil, wasting less sample during preparation.

B.2 With heavy oil

The idea to use oil to embed the crystals came from a single crystal preparation protocol using paratone-N to coat a single crystal. Sometimes, mainly to solve difficult cryocooling experiments, a single crystal is passed, after being mounted in a loop, through a drop of paratone-N, a very viscous oil. As we can't manually pass all microcrystals one by one through paratone-N, we tried to adapt this procedure. Centrifuging the sample on top of a layer of paratone never worked (figure B.2, p. 178), the reason being that the density of paratone is lower than water, and therefore than most mother liquors.

Several conditions are required to obtain a suitable liquid to filter the

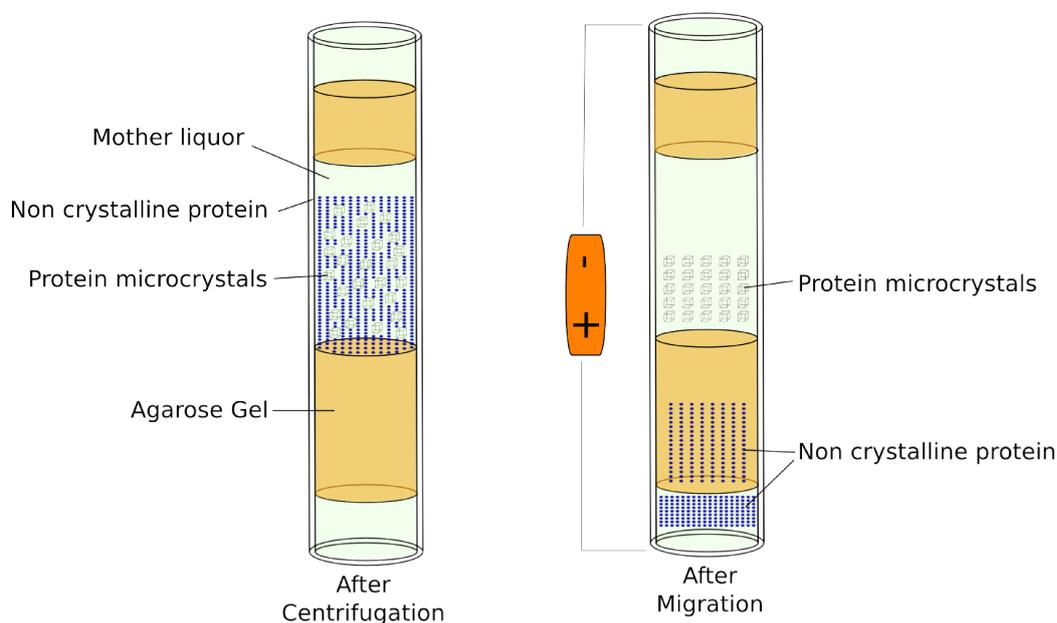


Figure B.1: Electrophoresis protocol : Crystals are packed on top of an agarose layer inside a kapton capillary. When voltage is applied, amorphous protein passes through agarose gel pores while keeping the crystals hydrated. A second layer of agarose (not represented) on top of the sample separating the sample from the electrophoresis tank is mandatory.

crystals:

- Not miscible with the mother liquor, to avoid changing crystallisation conditions.
- A density higher than the mother liquor.
- A density lower than the crystals.
- A density easily adjustable.

The solution proposed here consisted of an oil whose density is adjusted with tetra-bromomethane (figure B.3, p. 179). CBr_4 is dissolved at saturation inside a silicone or paraffin oil solution. The solution obtained is the highest density achievable and should be kept as stock solution. Addition of silicone or paraffin oil will reduce the density as required to have it just higher than the mother liquor.

Centrifuging can be done directly inside a capillary, minimising the number of preparation steps.

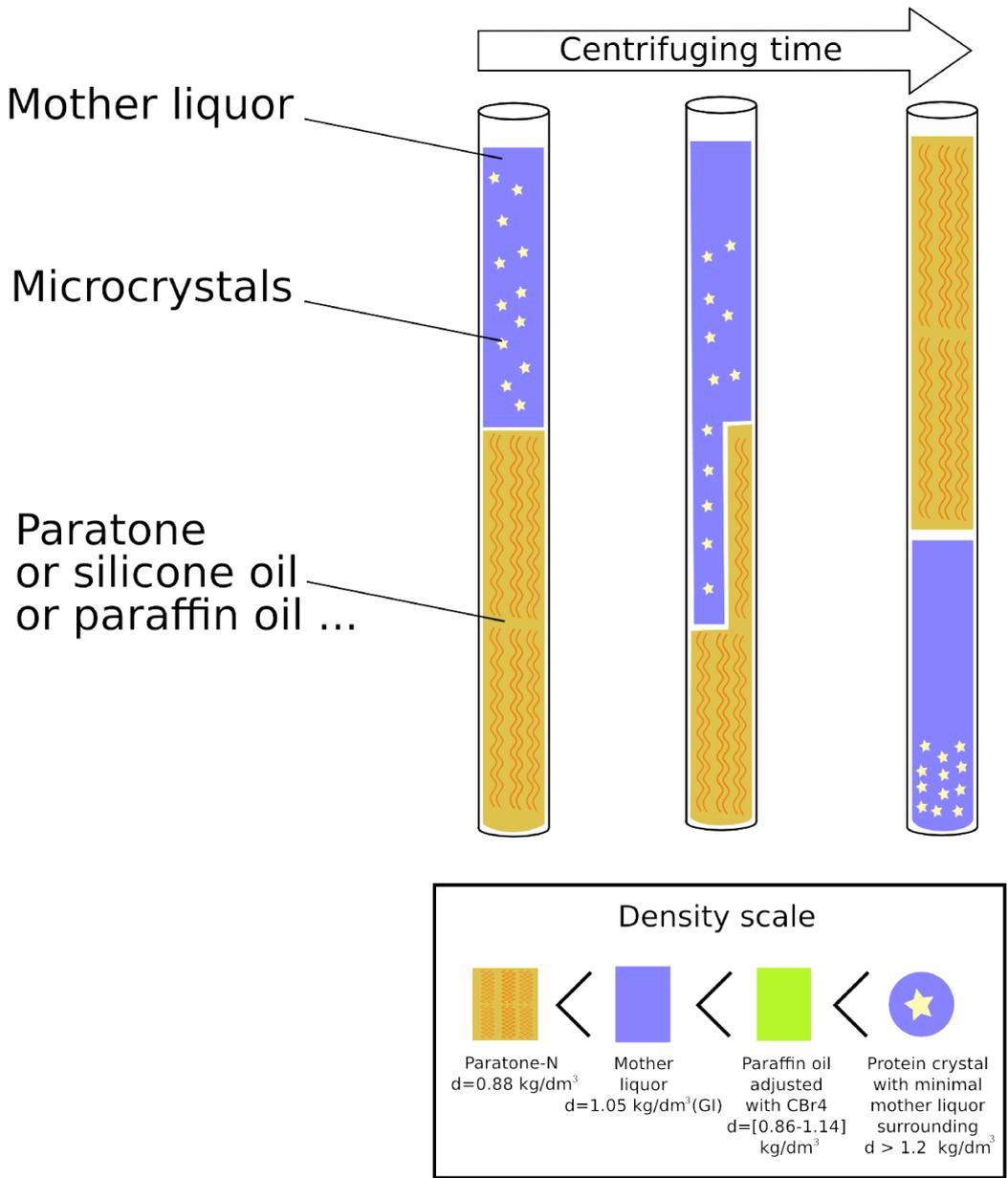


Figure B.2: Centrifuging on top of common oils, like paraffin oil or silicone oil, or even viscous ones like paratone-N does not work. The oil density is lower and when centrifuged, the oil passes over the sample without affecting it.

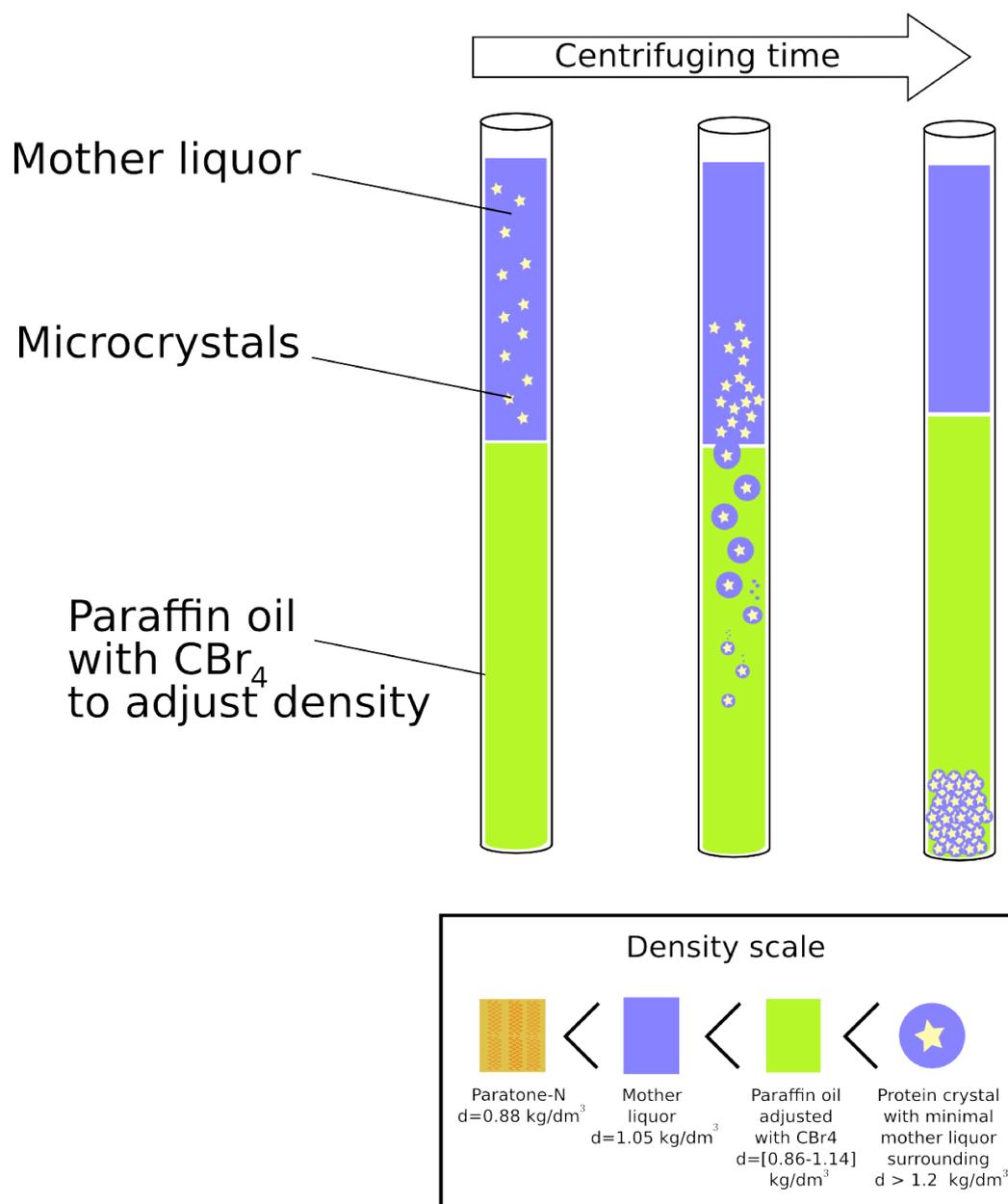


Figure B.3: Centrifuging over a layer of high density oil, adjusted to have a density slightly higher than the mother liquor. The crystals will be embedded in oil and go through it. We used CBr_4 for adjusting the oil density.



Figure B.4: Stock solution of heavy oil, consisting of paraffin oil saturated with CBr_4 .

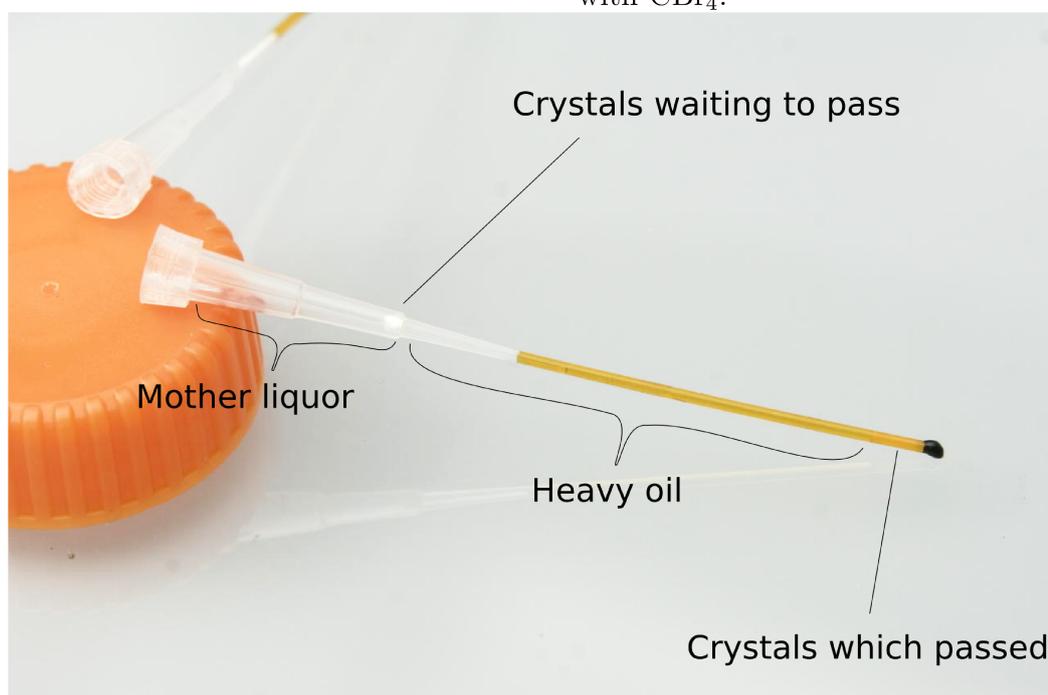
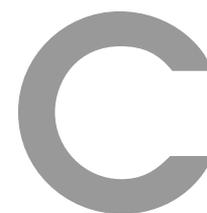


Figure B.5: State of a capillary during the protocol described (figure B.3, p. 179). Microcrystals are packed to the bottom of the capillary after passing through oil adjusted with CBr_4 .



Humidity variation experiments

At room temperature, a protein microcrystalline sample is quickly degraded if not sealed, due to dehydration. To avoid this degradation, the common method is to seal the sample inside a capillary. If the sample is too fragile, and does not survive the capillary transfer, it is better to use another sample holder. In this case dehydration must also be avoided.

On ID14 a device made by the EMBL is able to produce a continuous flow of hydrated air at a controlled humidity level [76]. This equipment's original use was to change the humidity of a single crystal in order to improve diffraction quality in situ. As with a single crystal the sample volume is much smaller, we were curious to see if the control of dehydration was possible with this equipment.

The chosen sample holder has to be large enough for our sample and can be open in order to allow exchange with the humid air. Our first attempt was to use a Silson membrane (figure C.1, p. 183) and see if changing the set-point of the humidifier resulted in a controlled sample size change (dehydration). We saw that the set-up was good enough to allow dehydration and rehydration afterwards. When dehydration is too severe, we saw near complete evaporation of the mother liquor, leading to concentration of salt. Salt crystals appear in the sample and rehydration didn't redissolve them.

C.1 Protocol

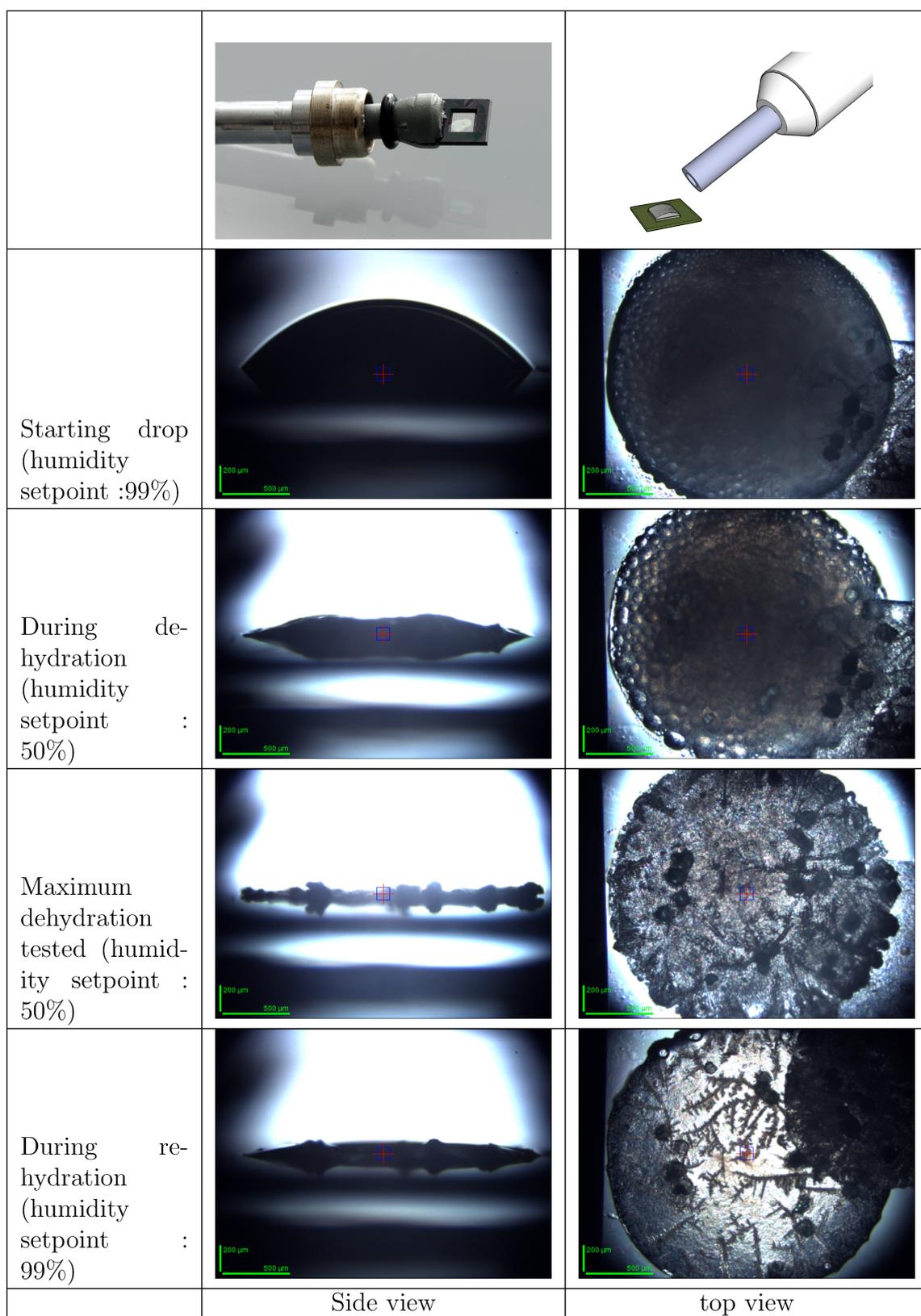


Figure C.1: Changing the humidity of a protein powder sample (glucose isomerase microcrystals) is possible. If too severe, the dehydration leads to near complete evaporation of the mother liquor, concentrating the precipitant, resulting in salt crystal growth.

The main interest is that it is possible to keep a small amount of sample at room temperature without dehydration. The Silson membrane used (3 mm x 3 mm window) was not really adapted to the ID14 beamsize (200 μm x 200 μm). This is why we changed for a micromesh from MiTeGen, with a diameter of 400 μm and pore size of 25 μm .

The actual protocol used for collecting protein powder diffraction from micromeshes is:

1. Find the relative humidity of the crystallisation conditions. This is done by putting a drop of the mother liquor on a micromesh and tracking the drop size. By adjusting the humidity set point the drop will grow (if too high) or shrink (if too low). When the drop size is stable, the humidity set point found is kept as reference.
2. Scoop the microcrystals from the crystallisation drop with a micromesh.
3. Immediately put the micromesh inside the humid air to avoid dehydration.
4. Using a paper wick or the corner of an absorbing paper, remove the mother liquor in excess until the microcrystals are left alone on the micromesh. As they are under humid air, they will not dehydrate but still give a diffraction signal.

C.2 Test case : Glucose isomerase

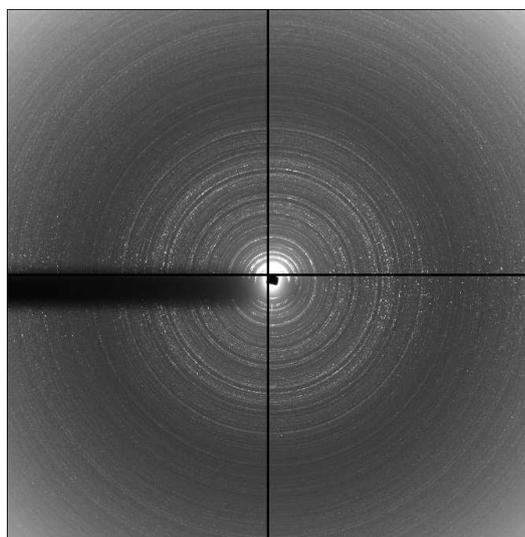
C.2.1 Experiment

In order to test the method before its use with the macro domain of the Mayaro virus (figure 5.2.4, p. 90) we performed a series of tests with glucose isomerase crystals. Glucose isomerase is an enzyme which converts glucose into fructose, which we bought already crystallised from Hampton Research. The sample was transferred using a pipette on top of a micromesh (700 μm in diameter, with pore size of 25 μm). The sample was immediately transferred into the humid air flow of the HC1b[76]. Excess mother liquor was then removed with a thin paper wick, while in the mean time the humid air was preventing dehydration.

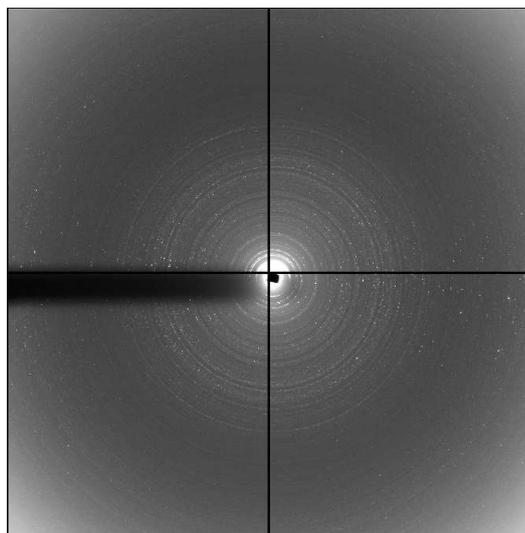
Images were collected while the humidity was decreased until no diffraction was observable (figure C.2, p. 186). In this particular case, we didn't observe any improvement in diffraction quality when the humidity was decreased. Worse, other components of the mother liquor saw their concentrations increase when the sample was dehydrated. This results in crystallisation of

salts in situ.

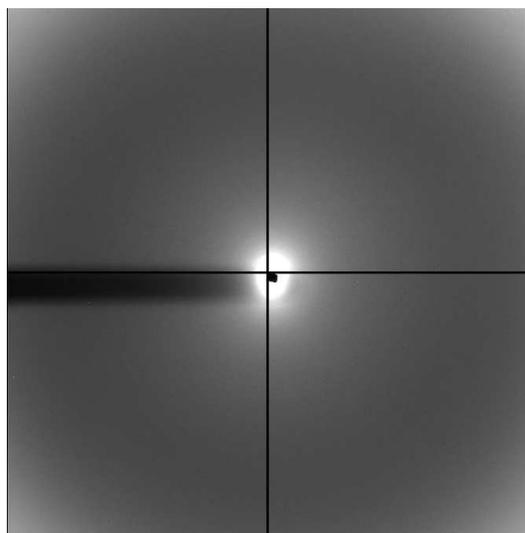
In order to reduce preferred orientation effects, care must be taken during the image acquisition to spin the sample by rotating the phi axis. The sample centring and the beam size must be adjusted in order to enclose the maximum amount of sample during a full rotation.



At 99% humidity rings are observable up to the corner of the detector.



When humidity is decreased, high resolution rings are lost.



Bellow a certain humidity (here 75%) diffraction disappeared.

Figure C.2: Controlled dehydration in situ of glucose isomerase protein powder.

C.2.2 Results obtained on glucose isomerase

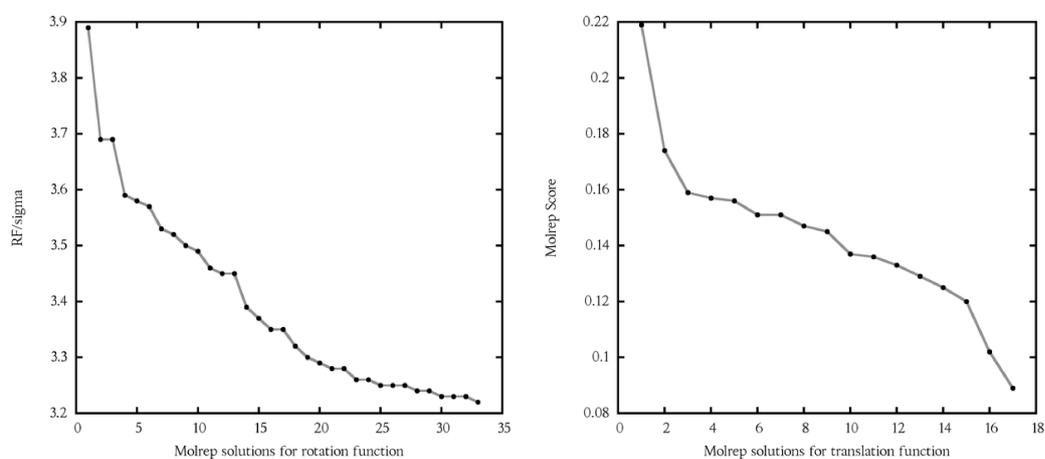


Figure C.3: Scores for the different molecular replacement solutions found by Molrep using a glucose isomerase single crystal search model (1MNZ). A clear first solution distinguishing itself from the others is found, for both rotation and translation.

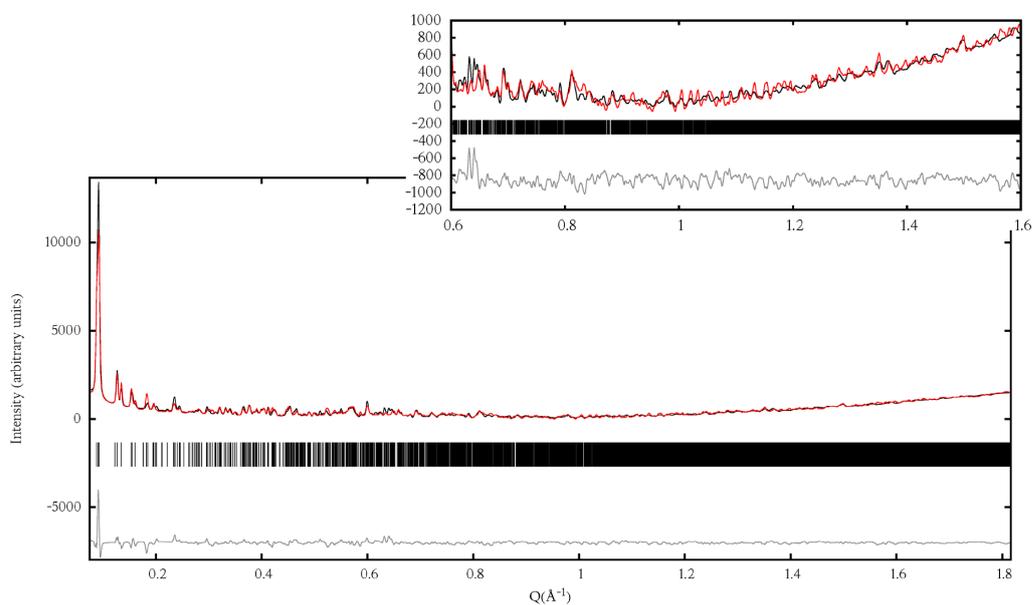


Figure C.4: Rietveld fit of the molecular replacement solution found for Glucose Isomerase.

D

Xray data reduction

Whereas the methods described for sample preparation and data analysis are specific for protein powder diffraction, the necessary data reduction step presented here is identical with other powder diffraction experiments.

D.1 ID31

On ID31, there are 9 detectors, scanning vertically to measure the diffracted intensity as a function of 2θ [77]. The scanning is continuous, the accumulated detector counts and the detector arm's angular position recorded simultaneously. This continuous scanning requires a binning step to get a typical histogram, with angle, intensity and error[78]. Before each detector there is a Si 111 analyser crystal, and between each analyser crystal there is an angle of approx $2^\circ 2\theta$.

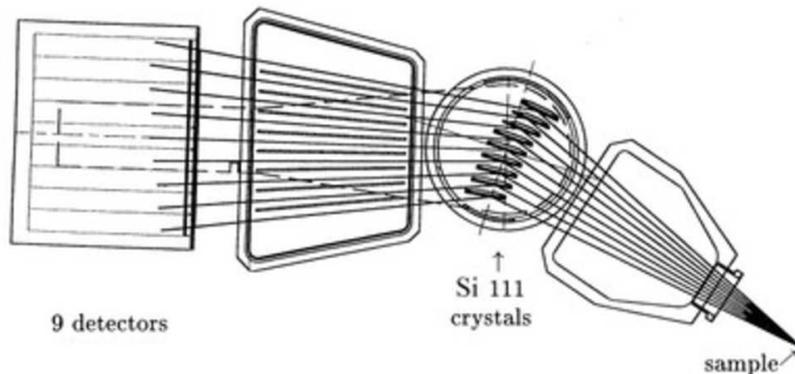


Figure D.1: The nine crystal multianalyser stage [77], was conceived by the Laboratoire de Cristallographie at the CNRS, Grenoble, for BM16

It is therefore necessary to define the exact angle between each detector. Then, the sum of all channels (all detectors counts) for the exact two-theta angle can be made for one scan, or for several scans to increase counting statistics. These operations are done by the `id31sum` program, on the machine `diffRACT31`.

D.1.1 Refining the offsets between the detectors

The offsets between detectors are contained in a file named `temp.res`:

```
8.05717844, 1.00000000, 0.00000000, 0.00000000,
5.98518955, 1.00000000, 0.00338029, 0.00000000,
4.00825584, 1.00000000, 0.00267771, 0.00000000,
2.07423734, 1.00000000, 0.00226371, 0.00000000,
0.00000000, 1.00000000, 0.00205394, 0.00000000,
-1.89610946, 1.00000000, 0.00199657, 0.00000000,
-4.00512519, 1.00000000, 0.00192428, 0.00000000,
-5.94307239, 1.00000000, 0.00189654, 0.00000000,
-7.97315111, 1.00000000, 0.00176422, 0.00000000,
```

The first column is the offset compared to the nominal angle of the diffractometer arm. That file needs to be tuned for each new beamline setup. To refine the offsets between detectors we will use `id31offsets` usually on a silicon dataset collected at the end of the setup.

```
grep -e "#S" inhouse_24_11_07_si.dat
#S 1 hookscan tth 15 100 5 6
```

For example, the file `inhouse_24_11_07_si.dat` contains a scan (number 1) between `tth=15` and `tth=100`, it is a good candidate for `id31offsets`.

```
% id31offsets inhouse_24_11_07_si.dat 0.001 1 1 lowtth=38 hightth=75
temp.res file found and read in
 38.000 < tth < 75.005 step= 0.10000E-01 npts= 3700
Binning scan 1
-0.7916738068E-04 8.057257607
 0.5938405413E-04 5.985130166
-0.6871721390E-04 4.008324557
-0.5266127826E-04 2.074290001
 0.000000000 0.000000000
 0.2585618028E-03 -1.896368022
-0.2967232904E-03 -4.004828467
-0.6261278584E-03 -5.942446262
-0.2979678969E-03 -7.972853142
temp.new file to be created (copy over temp.res)
```

The console output informs for each detector what is the offset (second column) and the change calculated (first column).

A `temp.new` file created by `id31offsets` has to be copied over `temp.res`. `id31offsets` should be rerun until the changes estimated are under 10^{-5} .

Cycling `id31offsets` and `cp` can be automated using `auto_offset_for_Si.pl`, a very simple Perl script that will call `id31offsets` several times:

To test on the first scan of `inhouse_24_11_07_si.dat` between `tth=38` and `tth=75` do :

```
auto_offset_for_Si.pl inhouse_24_11_07_si.dat 1 1 38 75
```

To check the validity of the correction, `pld` can be called to plot all the detectors together.

`pld` will over-plot all the channels with the correction. This is useful to check that a diffraction peak is seen at the same position on all detectors.

Some options common to `id31sum`/`id31sumall`/`id31offsets` can be added (table D.1.2 p. 193).

D.1.2 Summing with `id31sumall` and `id31sum`

Identifying the scans in a .dat file In order to sum all the scans independently we will use `id31sumall`. The first thing to do is to know how

many scans are contained in a `.dat` file:

```
grep -e "#S" myexperiment.dat
#S 1 hookscan tth 1 30 10 3
#S 2 hookscan tth 1 30 10 3
#S 3 hookscan tth 1 30 10 3
#S 4 hookscan tth 1 30 10 3
#S 5 hookscan tth 1 30 10 3

...

#S 47 hookscan tth 1 30 10 3
#S 48 hookscan tth 1 30 10 3
#S 49 hookscan tth 1 30 10 3
#S 50 hookscan tth 1 30 10 3
#S 51 hookscan tth 1 30 10 3
#S 52 hookscan tth 1 30 10 3
#S 53 hookscan tth 1 30 10 3
```

In this example we have 53 scans, recorded at the speed of 10 degrees per minute, between 1 and 30 degrees.

Summing all individual scans with `id31sumall` `id31sumall` needs as arguments the filename of the `.dat` file, the stepsize of the final pattern, the first scan and the last scan to include. Some optional arguments can be added after (`lowtth=1` for example) (figure D.1.2, p. 193).

Here for example to obtain 53 `.xye` files (`tth / intensity / error on intensity`) for each scan between scan 1 and scan 53, with a stepsize of 0.003 degrees excluding signal below 1 degree:

```
id31sumall T6_Novo.dat 0.003 1 53 lowtth=1

temp.res file found and read in

0.99900 < tth < 160.00 step= 0.30000E-02 npts=53000
Processing scan 1
Determining detector efficiencies
```

Channel efficiencies found from 2515 points where all detectors overlap

Efficiencies from temp.res file, the values found now are compared

Det	Offset	Effic	<Effic>	current	unused values
0	8.0580263	1.0315431	0.0042134	0.9708797	0.0055310
1	5.9858285	1.0686351	0.0042997	1.0684182	0.0058164
2	4.0092737	1.0752279	0.0043126	1.0886128	0.0058522
3	2.0751736	1.0546568	0.0042687	1.0701667	0.0057703
4	0.0000000	1.0040601	0.0041474	1.0112710	0.0055714
5	-1.8962349	0.9801896	0.0040710	0.9860485	0.0054798
6	-4.0044830	0.9622571	0.0040355	0.9657536	0.0054014
7	-5.9425488	0.9240694	0.0039532	0.9237719	0.0052622
8	-7.9726553	0.8993609	0.0038796	0.9150776	0.0052153

R_exp = 11.252 with 252 pts having I/<I> less than 3, from 12328 pts obs

Reduced chi**2 for channel merge=0.8229 from 79647 pairs of pts 0.47% differ by >3 sigma, 0.0025% by >6 sigma (ideally 0.04% and 0.0000%)

Wrote 1.inp

...

Wrote 53.inp

Plotting all the scans individually On the diffract31 machine, use plmany to plot multiple .xye or .inp files.

plmany_*.inp will plot all the scans summed by id31sumall.

plmany_1[0-5].inp will plot scans from 10 to 15.

Summing several scans together - id31sum The usage for id31sum is identical to id31sumall.

id31sum filename stepsize firstScan lastScan [options]

Where the most useful [options] are :

<code>ed=n1,n2</code>	to exclude detectors n1 and n2
<code>es=m1,m2</code>	to exclude scans m1 and m2
<code>is=m1,m2</code>	to include only scans m1 and m2
<code>lowtth=xx.xx</code>	to set min two theta to use
<code>hightth=xx.xx</code>	to set max two theta to use
<code>zap=xx</code>	for esd level in filtering operation
<code>superzap=xx</code>	useful to filter crazy points
<code>renorm</code>	to use current efficiency instead of values from temp.res
<code>gsas</code>	to output a .gsa file for gsas
<code>window=counter,low,high</code>	to select points for example window=Epoch,0,10 to have only the points collected in the 10 first seconds of each scan.

For example, to sum the scans between 1 and 20 in `myexperiment.dat`, with a 0.003 degrees stepsize, excluding points below 0.8 degrees and above 22, and excluding detector number 0, outputting the summed profile inside a `.xye` file and a `.gsa` file, one should type :

```
id31sum myexperiment.dat 0.003 1 20 lowtth=0.8 hightth=22 ed=0 gsas
```

D.1.3 Clustering of multiple scans

When multiple scans are collected on id31, at different sample positions, with different radiation damage levels, a good way to sort them all is to compute a dendrogram of similarities.

This is possible by using the C clustering library : <http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/cluster.pdf>. The `treecluster` routine implements pairwise single-, complete, average-, and centroidlinkage clustering. In `cluster_rescale.py` we are using average.

After having summed all scans individually using `id31sumall`, it is possible to call `cluster_rescale.py` on all the individual scans (`*.inp`) and the `id31sumall` logfile.

On diffract31 machine :

```
cluster_rescale.py *.inp logfile=inhouse_24_11_07_y25.log
```

A picture, `dendogram.ps` is created, and a file `clusters.scans` containing the scans (`is=1,2,3,4,5,...`) referred to via letter in the picture :

```

...
e score=0.157676 is=8,9,10,18,19,20,28,29,30
f score=0.122567 is=4,5,14,15,24,25,34,35
g score=0.121742 is=1,2,11,12,21,22,31,32
h score=0.112534 is=8,9,18,19,20,28,29,30
i score=0.097045 is=6,7,16,17,26,27,36,37
...

```

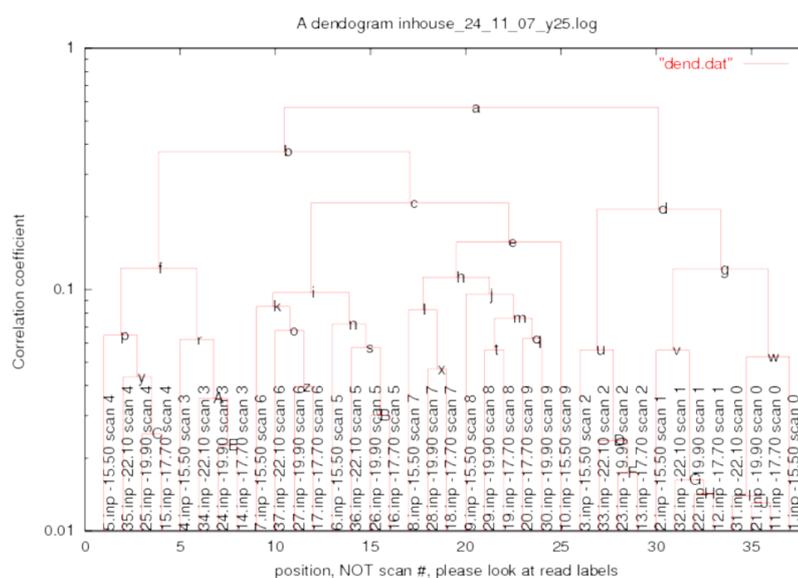


Figure D.2: A dendrogram of similarities for several scans of the same sample at different positions on id31.

It is then easy to sum the scans for the similarity branch *f* by doing :

```
id31sum inhouse_24_11_07_y25.dat 0.003 4 35 is=4,5,14,15,24,25,34,35
```

which will sum the scans 4,5,14,15,24,25,34 and 35 together.

This method allows one to sum together similar scans in order to improve the statistics, and obtain different profiles (for example due to radiation damage) which will be useful for intensity extraction using anisotropic lattice parameter shifts, with for example prodd (figure G, p. 231).

D.2 On 2D detectors

To obtain a powder pattern from the rings observed on a 2D detector image, we must integrate the image.

The easiest way to perform this operation is to use `imageviewer` from `fable` to identify useful images, then `FIT2D` to find the correct parameters for the integration and finally `fit2dcake.py` to apply the integration on series of images.

D.2.1 Fable - Imageviewer

`Imageviewer` is a detector image visualisation software part of the `Fable` framework. The `fable` framework has been created originally for the Total-Cryst European project aiming for solving and refining arbitrary polycrystalline samples¹.

Starting fable At the ESRF `fable` is already installed on the “coral” machines but the refresh rate of the graphical interface is a lot slower than if run locally on the PC. To start `Fable` on coral type `/sware/exp/fable/bin/fable`.

Opening image files During an experiment or after, when multiple images have been taken, it is easier to open all the images contained in a directory and not each individual file:



Figure D.3: Use *Open file* to get individual file, or *Open Directory* (recommended) to obtain a list of all the images in a directory inside the image navigator.

1. Information about TotalCryst and `fable` can be found at <http://www.totalcryst.dk/> and <http://sourceforge.net/apps/trac/fable/wiki>.

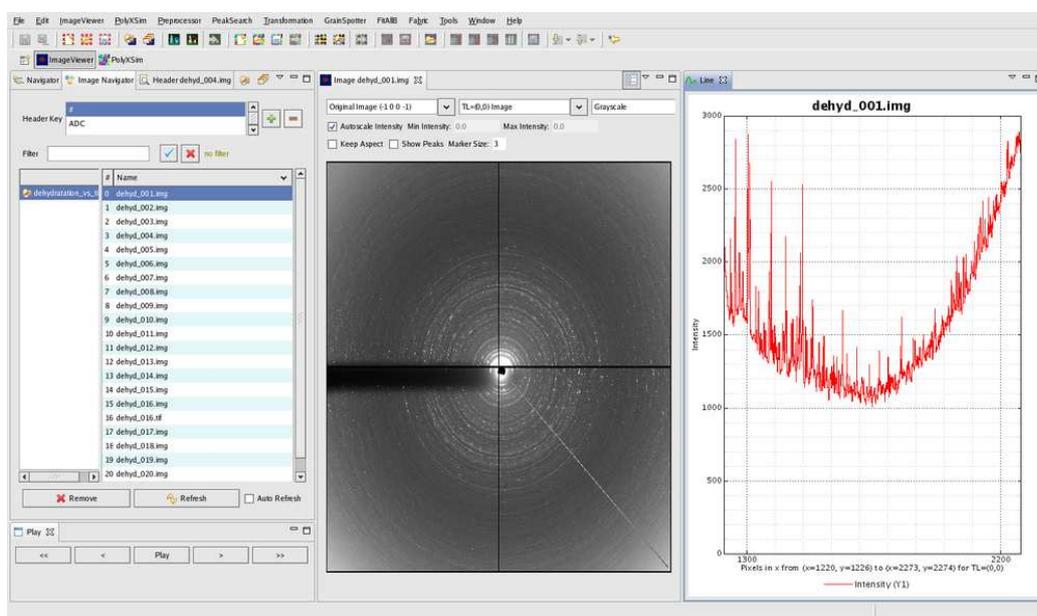


Figure D.4: A screen capture of fable. The images contained under the opened directory are visible in image navigator. A zoom line obtained by right-clicking on the image gives a first idea of the diffraction quality of the sample.

D.2.2 FIT2D

All the steps from finding the correct parameters to performing the transformation of a 2D image to an integrated pattern are done by using FIT2D[79]. FIT2D is a graphical software, allowing one to perform complicated geometrical operations on the image files coming from 2d detectors.

FIT2D can be downloaded at <http://www.esrf.eu/computing/scientific/FIT2D/>. A complete manual can be found at http://ftp.esrf.fr/pub/expg/FIT2D/fit2d_ref_12_012.ps.

Starting FIT2D At the ESRF, FIT2D is already installed on coral computers, typing `/usr/local/bin/fit2d` will start it.

Opening a file The first thing asked by FIT2D is the array size that should be created for the image memory. The array size corresponds to the number

of pixels of the detector used.

DIMENSIONS OF PROGRAM ARRAYS (need to be big enough to store and work on data)		
O, K.	CANCEL	? HELP INFO
DESCRIPTIONS	VALUES	CHANGE
FIRST DIMENSION OF ARRAYS	2304	X-DIMENSION
SECOND DIMENSION OF ARRAYS	2304	Y-DIMENSION
CREATE MEMORY ARRAYS	YES	MEMORY
CREATE VARIANCE ARRAYS	NO	VARIANCES

Click on variable to change, or 'O, K.'

Figure D.5: The array size depends of the number of pixel of the detector, here 2304 for an ADSC detector on ID14-2.

Then enter the *POWDER DIFFRACTION (2-D)* menu, and open the file selector by clicking on *INPUT*. After choosing a file we obtain :

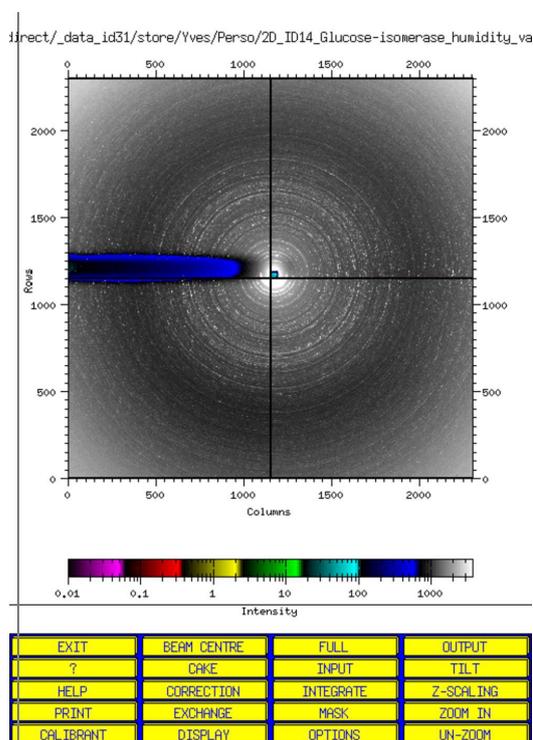


Figure D.6: A image opened in FIT2D.

The most important menus are :

- Z-SCALING To adjust the displayed colours and grayscale.
- BEAM CENTER To define and refine the beam centre.
- MASK To remove beamstop shadow, direct beam and virtual pixels at the edges of detectors.
- CAKE To draw a radial transform of the 2D image.
- INTEGRATE To do the radial integration of the 2D image.

Identifying the beam centre This is easily performed by entering the BEAM CENTER / CIRCLE COORDINATES menu. Clicking on an intense ring several times (3 minimum) is the quickest way of finding a beam centre. To obtain a precise result, the use of the two click method is recommended (one in the image, one in the magnified region called *spy-glass*).

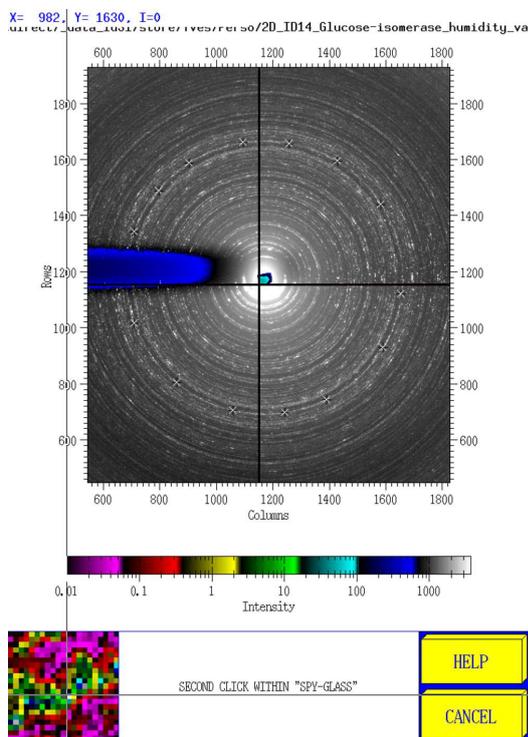


Figure D.7: Defining a beam centre with one ring. A first click in the main window followed by a precise positioning of the ring in the magnification window (green is the ring, purple the background).

In the console a line should appear with the beam centre :

INFO: Beam/symmetry centre (pixels) = (1170.962 , 1178.146)

Obtaining a radial transform A routine radial transform is a good thing to evaluate the precision of some parameters, for example, each time a new beam centre is refined.

To obtain a radial transform, go into the CAKE menu, answer the questions and click on INTEGRATE (in the cake menu, not in the powder diffraction menu). For a control quality test done on a radial transform, an integration over the whole image should give this :

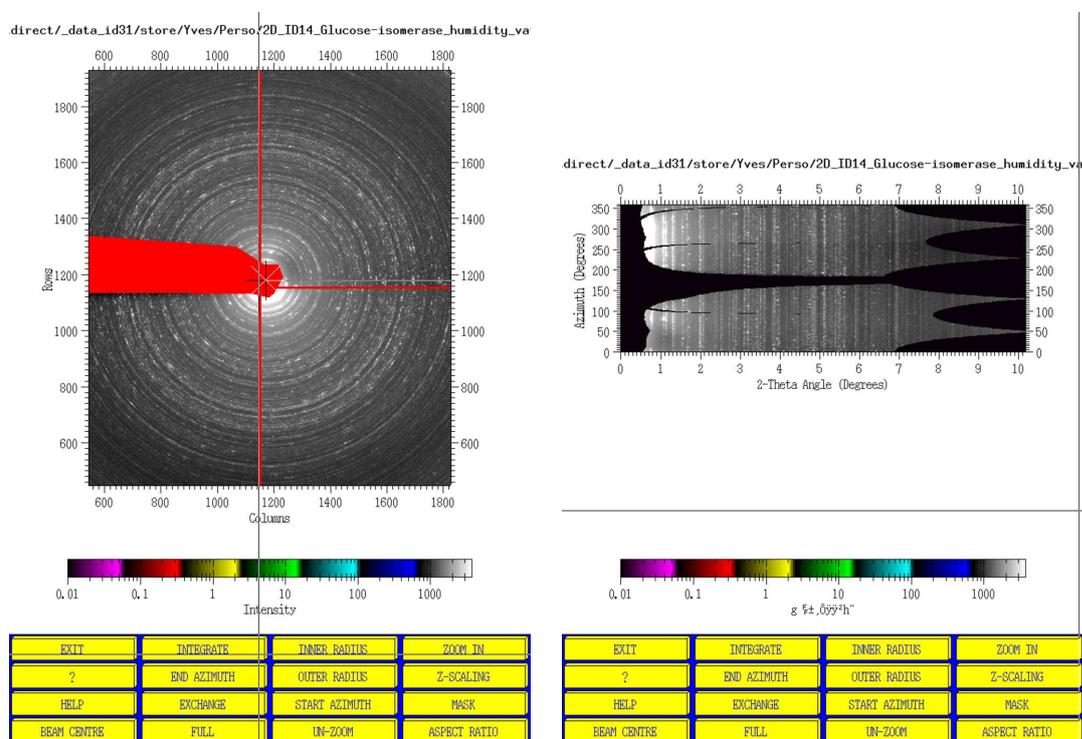


Figure D.8: A radial transformation (called *cake* in FIT2D) is useful for checking the quality of a beam centre position. The red colour is a mask over the beam stop shadow and the detector borders. The black in the radial transform corresponds to this mask.

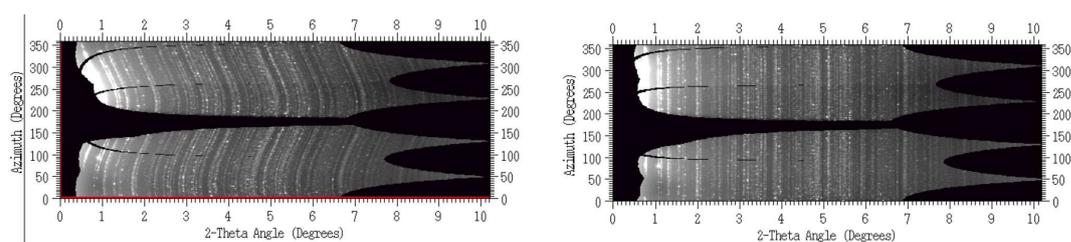


Figure D.9: In the case of a badly defined beam centre (left), the radial transformed rings will produce wavy lines. If the beam centre is ok, straight lines are observed.

Radial integration A radial integration is performed in the **INTEGRATE** menu of the powder diffraction menu. Once the pattern is obtained, it can be

saved as a .xy file by going into OUTPUT (in the integrate menu) and choosing CHIPLLOT file format.

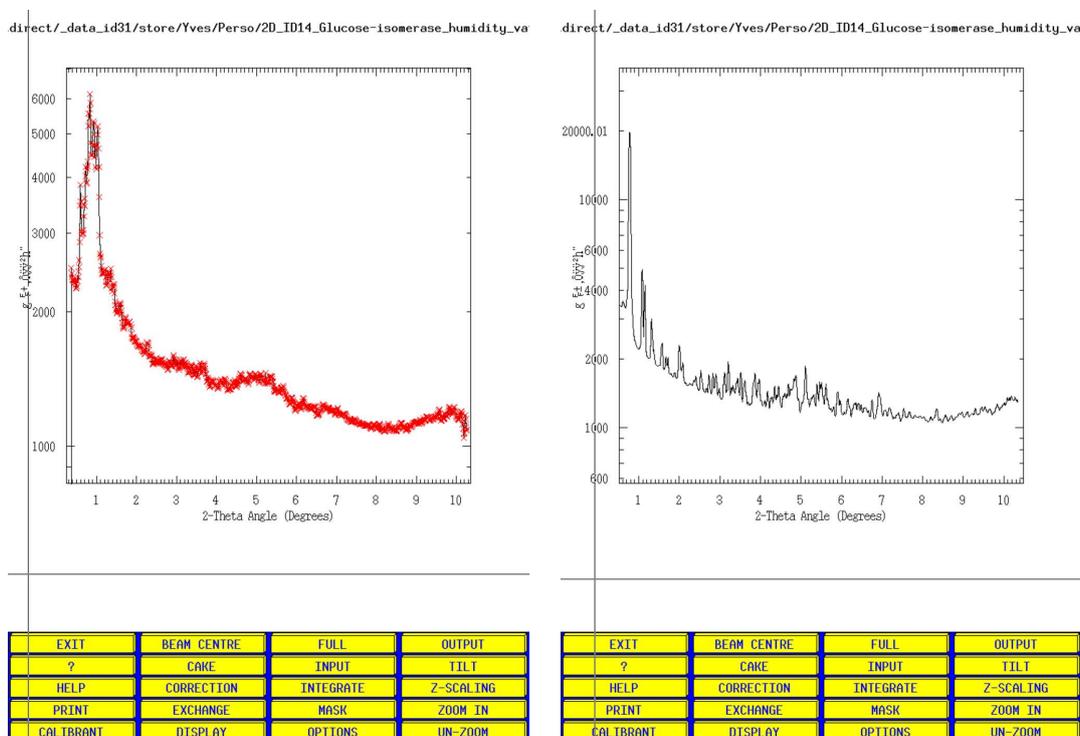


Figure D.10: The resulting radial integration pattern when a beam centre is badly defined (left), and when the beam centre is correctly defined (right).

D.2.3 fit2dcake on Frelon camera

Using fit2d on multiple files, or with different detectors, or different detector positions, is not really optimal. While FIT2D can perform a file series, it is often easier to use fit2dcake.py (written by Jon Wright) with a set of parameters corresponding to one experiment.

Parameter file On coral, to have fit2dcake.py accessible from any directory the simplest way would be to add this line to your `.bashrc_private` file:

```
export PATH=/sware/exp/fable/standalone/bin:$PATH
```

In order to have a starting input file for `fit2dcake.py` one should use the “-c” option with a new filename. This will create a parameter file corresponding to the last execution of FIT2D :

```
coral30:~ % fit2dcake.py -c test.params

coral30:~ % cat test.params
# Image correction parameters
DARK CURRENT = YES
DC FILE = /mntdirect/_data_id11_inhouse/jon/aug09/dark10_/dark10_0002.edf
FLAT-FIELD = YES
FF FILE = /mntdirect/_data_id11_inhouse/Frelon4M/F4M_Sm_July08.edf
FF SCALE = NO
FF MULTIPLIER = 1000.000
SPATIAL DIS. = YES
SD FILE = /mntdirect/_data_id11_inhouse/Frelon4M/frelon4m.spline

# Integration parameters
START AZIMUTH = -154.804773765
END AZIMUTH = -26.785605035
INNER RADIUS = 32.0
OUTER RADIUS = 1800.0
SCAN TYPE = 2-THETA
1 DEGREE AZ = NO
AZIMUTH BINS = 1
RADIAL BINS = 2048
CONSERVE INT. = NO
POLARISATION = YES
GEOMETRY COR. = YES

# Experiment geometry parameters
X-PIXEL SIZE = 50.0
Y-PIXEL SIZE = 50.0
DISTANCE = 1437.443
WAVELENGTH = 0.2952
X-BEAM CENTRE = 1020.207
Y-BEAM CENTRE = 1764.416
TILT ROTATION = 81.648090088
ANGLE OF TILT = -0.519972548998

# Mask and dimensions
USE MASK = NO
MASK FILE = None
DIM1_DATA = 2048
DIM2_DATA = 2048

# I/O parameters
input_extn = edf
saving_format = CHIPLLOT
output_extn = xy
```

Then to obtain the radial integration of all the images at once, use the command : `fit2dcake.py_stem_name_first_last_parfile`. The argument `stem_name` is the base name of the image without the serial number and the extension (for ex: `Xdomain_id11_first_position_0001.edf`). The arguments `first` and `last` are the numbers for the first image and the last image of the series (for ex: **1 135**). The last argument, `parfile`, is the name

of the parameter file created and modified previously (for ex: **test.params**).

```
fit2dcake.py Xdomain_id11_first_position_ 1 135 test.params
```

will produce 135 radial integrations, as **.xy** files.

D.2.4 fit2dcake for ID14

The images produced by ID14-2 detector, an ADSC Q4 detector, cannot be treated directly by `fit2dcake.py`. The reason is that the detector format is fixed for the Frelon camera, and the file serial made for 4 digits (`_xxxx.edf`). Some quick modifications were made in `fit2dcake.py`, to detect automatically the ADSC detector when an image has the extension `.img`, then to obtain the correct file serial number (`_xxx.img`) and also to have the possibility to be in a directory with longer names than normally allowed by `fit2d`. In order to use ADSC detectors with `fit2dcake.py` one should make the following changes on its local copy:

In `cakefileglob`, function used to read the file names we changed from:

```
filelist=glob.glob("%s????.%s"%(stem_arg,self.input_extn))
```

to

```
filelist=glob.glob("%s*.%s"%(stem_arg,self.input_extn))
```

In `cakefileseries` the corrected file series number is obtained by changing :

```
filein = "%s.%04d%s" % (stem_arg, i, self.input_extn)
for
filein = "%s.%03d%s" % (stem_arg, i, self.input_extn)
```

In run the directories that could be too long (too many characters) are opened by :

```
tmpfile = tempfile.NamedTemporaryFile(dir=".")
then, only the last part of the directory is kept with :
stringFileName = re.sub("/.*\/([\^\/]+)", "\\1", tmpfile.name)
and finally launched like this:
cmd = FIT2D + " -dim%s -mac%s.mac"%(array, stringFileName)
```

An updated working executable for ADSC Q4 detector can be found on coral under the name `fit2dcakeid14.py`. Adding to `~/.bashrc_private` the line

```
export_PATH=/data/id31/store/Yves/bin:$PATH
```

will allow access to the modified version executable from any directory.

For example,
`fit2dcakeid14.py` `dehyd_1_16_fit2dcake.pars`
will read the files `dehyd_001.img` to `dehyd_016.img` using the `fit2dcake.pars`
parameter file.

An example of parameter file for `fit2dcakeid14.py`:

```
# Image correction parameters
DARK CURRENT = NO
DC FILE = Dark_30sec_A0005.edf
FLAT-FIELD = NO
FF FILE = flat_field.bin
FF SCALE = NO
FF MULTIPLIER = 1000.000
SPATIAL DIS. = NO
SD FILE = spatial.spline

# Integration parameters
START AZIMUTH = -178
END AZIMUTH = 177
INNER RADIUS = 54
OUTER RADIUS = 1594
SCAN TYPE = 2-THETA
1 DEGREE AZ = NO
AZIMUTH BINS = 355
RADIAL BINS = 1540
CONSERVE INT. = NO
POLARISATION = NO
GEOMETRY COR. = YES

# Experiment geometry parameters
X-PIXEL SIZE = 82.0
Y-PIXEL SIZE = 82.0
DISTANCE = 446.668
WAVELENGTH = 0.933
X-BEAM CENTRE = 1164
Y-BEAM CENTRE = 1178
TILT ROTATION = 0.0
ANGLE OF TILT = 0.0

# Mask and dimensions
```

```
USE MASK = YES
MASK FILE = id14-2_fit2d.msk
DIM1_DATA = 2304
DIM2_DATA = 2304
```

```
# I/O parameters
input_extn = img
saving_format = CHIPLLOT
output_extn = xy
```

The result will be a list of `.xy` files corresponding to the `.img` files.

The `.xy` files produced by `fit2d` have some useless header lines at the beginning that can be a problem for some software (Topas for example). For example to remove the 5 first lines of all the `.xy` files in the current directory:

```
find . -name "*.xy" -exec tail -n+5 {} > {} \;
```



Background subtraction

Background subtraction, in powder diffraction, is usually modelled as a polynomial and refined via a set of least squares variables during pattern fitting. For smooth backgrounds, when few amorphous components are in the sample, this is an easy task. This task can even be performed in the early stages when all the other parameters are not well refined. Unfortunately, for most protein powder diffraction data, several problems complicate the background determination.

When many peaks overlap at high angles, the background is not visible between the peaks. This often causes divergence when refining parameters such as thermal B factors and a complicated background function. For this we propose a preliminary background subtraction in order to have a simple background function in the least squares refinement.

A unix command line software, `background_remover`, using simple statistical methods for determining the background and quality evaluation, has been created.

The software we propose here is using the GNU project called “R Project for Statistical Computing” <http://www.r-project.org/>, commonly called R. R was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues to replace “S” which was not open source.

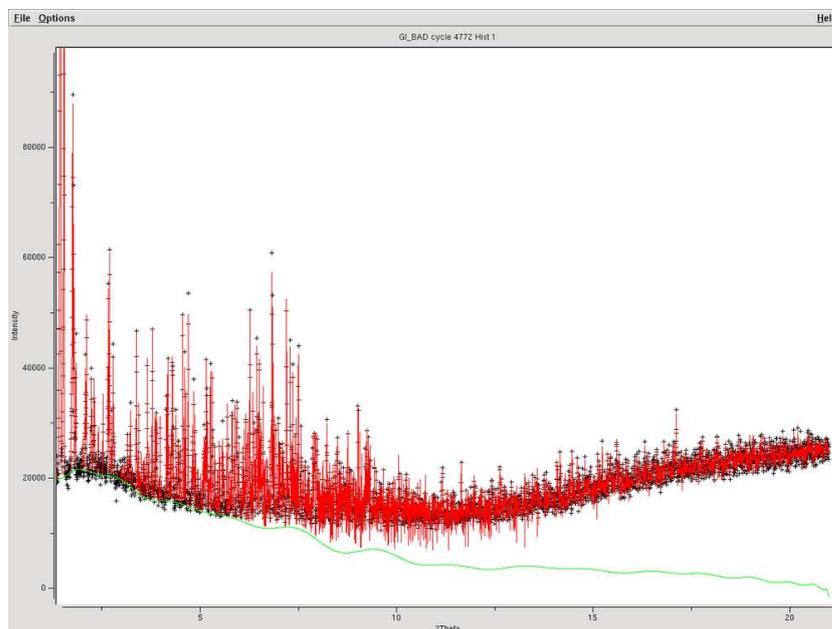


Figure E.1: An example of a poor Le Bail refinement, with a high order background function; when peaks exhibit severe overlapping, the background is not determined accurately.

E.1 Background subtraction

The idea to determine the background of a powder profile is to draw a smooth function which will underlie the powder pattern.

The originality in `background_remover` is to apply a cubic spline smoothing function, not on all the points of the powder pattern, but on a subset of the points of the pattern.

The points where the smoothing function is applied are determined by the result of a rolling function:

- First, determine a window around the point we're considering.
- Second, determine the result based on the points' intensities in this window in order to obtain the result of the rolling function. There are two different rolling functions, selecting inside the window either the minimum or alternatively any quantile value of the intensities (usually

the first quartile, but this can be adjusted).

- Third, determine a smooth function based on the result of the rolling function.
- Fourth, subtract the profile with this function and add an offset to have the lowest point at 0 (to avoid bugs in some analysis software).

We implemented the rolling function in the software while the smoothing function is already implemented in R as a cubic smoothing spline fitting the supplied data¹.

Several parameters can be adjusted, the window size, the method and the quantile value for determining the result of the rolling function; also a smoothing parameter for the smoothing applied over the rolling function.

1. Complete description of the smoothing function : <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/smooth.spline.html>

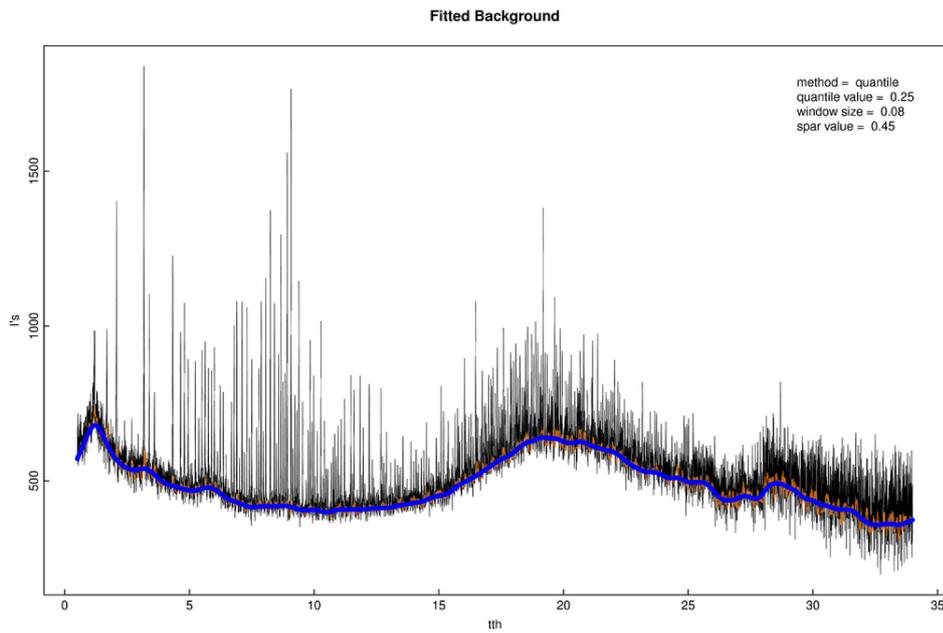


Figure E.2: In orange, the result of the rolling function determined by taking the first quartile in a window of 0.08 degrees around the point. In blue the smoothed function using a smoothing parameter value of 0.45.

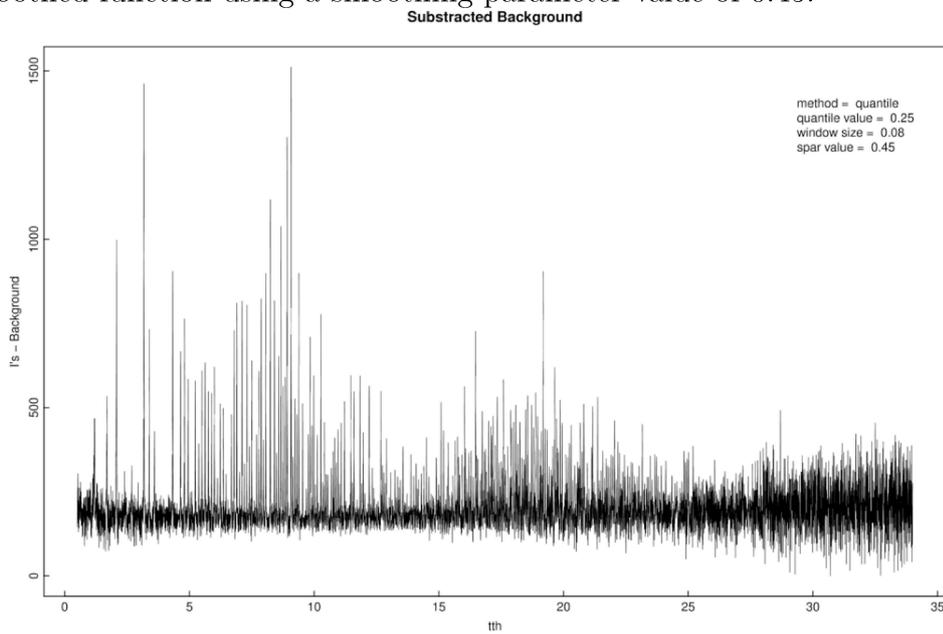


Figure E.3: The resulting subtracted pattern.

E.2 Quality evaluation

Very often, the d-space resolution limit of a protein powder pattern is problematic and hard to discriminate. The noise in the case of a single scan collection, or the overlapping high angles peaks for big cells, lead to problematic interpretation where some powder experts do believe seeing a peak while some other expert doesn't.

Together with the background subtraction, a quality evaluation tool is provided in `background_remover`, using the subtracted profile. The way to evaluate the quality here is to split the pattern into several bins, of equal numbers of observed points, and find the number of points describing peaks in each bin. The number of bins can be chosen freely in `background_remover`, a too small number of bins will not help evaluate precisely the locations of the last observed peaks, and a too big number of bins will result in having a bin with a 2θ width too small to include a peak.

The observed intensity in one bin can be separated into two groups. A first group, which is a normal distribution (for the points between two peaks), coloured in red on the next figures. A second group of intensities which are outliers of this normal distribution (the intensities belonging to a peak), coloured in green in the next figures.

A point here is considered as an outlier of the distribution of the bin, if its intensity is greater than the third-quartile value plus the interquartile value, $I > \text{thirdquartile} + (\text{thirdquartile} - \text{firstquartile})$.

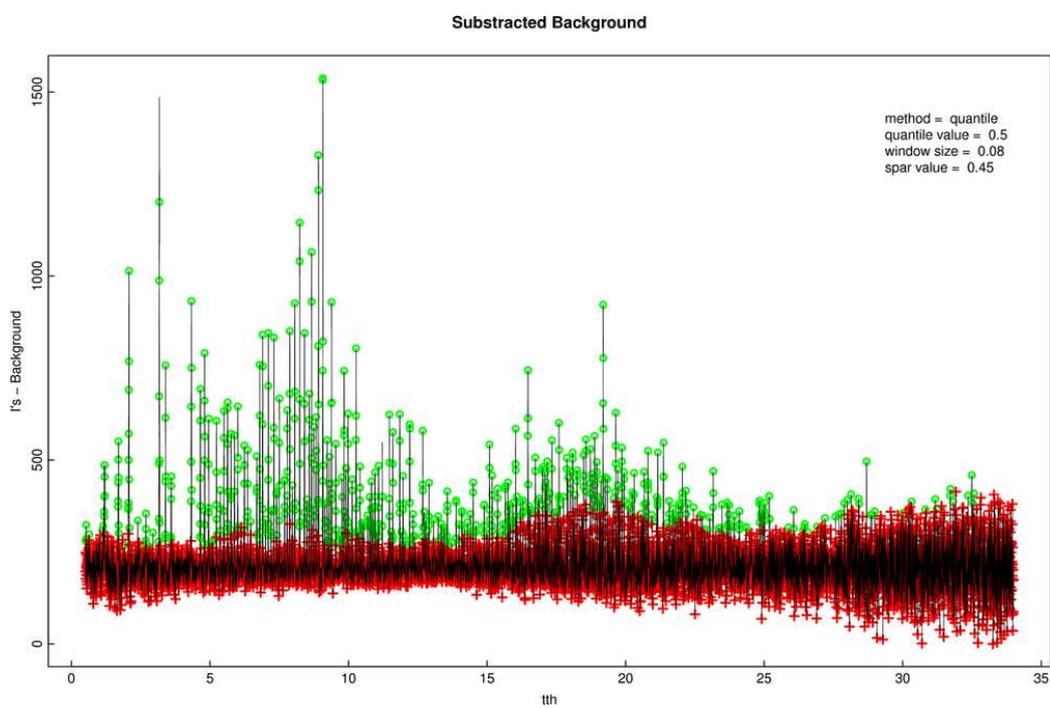


Figure E.4: Subtracted background with in red the points belonging to a normal distribution, and in green the points considered as outliers.

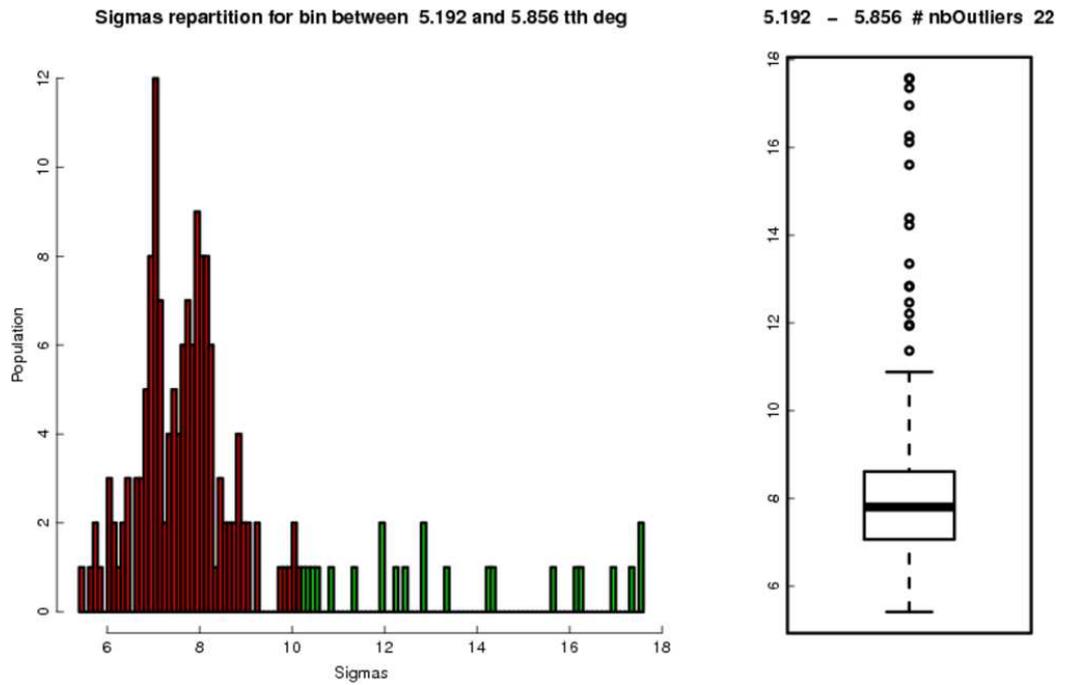
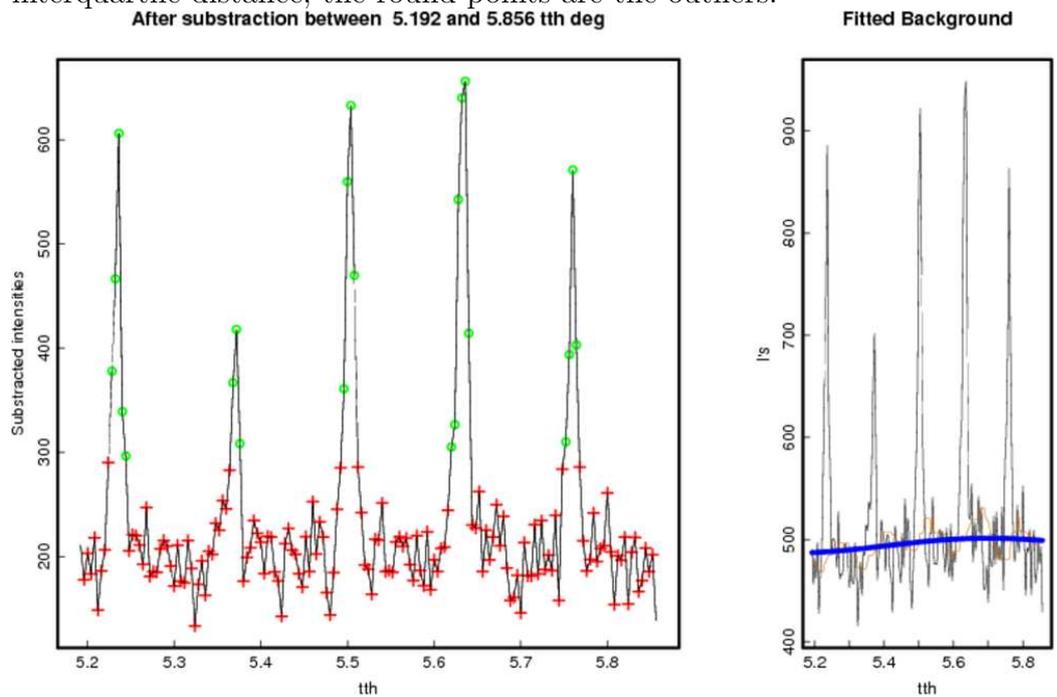


Figure E.5: An example of one bin, between 5.192 and 5.856 degrees. On the left, a histogram of the sigma repartition for the bin with (in green) the outliers and a graphical box-and-whisker diagram of the population. The lower line of the box-and-whisker diagram represents the first quartile, the upper line of the box the third quartile, the bold line the median; the whiskers represent the min and max values if they are not greater than 1.5 times the interquartile distance; the round points are the outliers.



The software is called `background_remover` and is freely available and requires R and perl to be installed.

`background_remover`

Minimum USAGE: `background_remover -f filename.xye`

USAGE: `background_remover -f filename.xye [-w windowSize]`
`[-s sparSize] [-m method] [-qv quantile_value]`
`[-b nbBins] [-B nbBreaks] [-pngs] [-onepdf]`

[windowSize] is 0.1 by default (in tth) adjust the crenallated function.

[sparsize] is 0.5 by default adjust the smoothing function.

[method] can be `min` or `quantile`

[quantile_value] if not given, will be 0.25 by default (for a first quantile).

[nbBins] Is the number of bins you want for the distribution plot (and the number of different histograms (10 by defaults)

[nbBreaks] Is the number of breaks for the histograms (100)

[-pngs] make pictures from the pdf for illustrations.

[-onepdf] supress the output of long filenames, only `last.pdf` and `last_histogram.pdf` are created (useful to start)

This method for background subtraction is working well with various kinds of pattern. However care must be taken when the profile has broad peaks at high angles, like when for example the sample is cryocooled or collected on a 2D detector. In these cases, the background may actually fit the broad peaks instead of fitting a background curve. Also, in these cases, the evaluation tool proposed below is less useful, as the broad peaks would be seen as a normal distribution with too small bin sizes.

Therefore the use of the evaluation tool based on outliers must be reserved for high angular resolution patterns, like room temperature powder patterns collected at ID31. In this case, after subtraction, the modelled background in the following analysis software can be represented as a line. This will avoid background related divergence of refinement in LeBail, Pawley or Rietveld analysis.

F

Indexing and modelling powder patterns with TOPAS

Here we describe a set of macros created or adapted to use the analysis software TOPAS with powder patterns. To help generating the macro for topas for each protein structure, we wrote a software `pdb_to_topas`, freely available [80].

F.1 Introduction

While various software packages exist to study powder patterns, few are accessible to beginners and yet remain highly customisable.

TOPAS is a powder diffraction software originally for small molecules that can perform various operations by default. These methods include Indexing (LSI and LP-Search methods), whole Powder Pattern Decomposition (Pawley and Le Bail methods), Ab-initio structure determination in direct space from powder and single crystal data, Quantitative Rietveld analysis and Rietveld structure refinement.

While TOPAS is a commercial software with a closed source code, it is possible to adapt it to specific needs. Indeed TOPAS includes a macro language and algebra system that allow the user to create his own macros, extending

the usefulness of the software.

It is therefore possible to restrain the indexing of a powder pattern to a selection of space groups meaningful in structural biology, create an input file for TOPAS with protein atoms and solvent scattering modelling, perform a rigid-body refinement of the whole protein while avoiding false minima.

However the fact that the TOPAS source code is closed and non adaptable (other than by the author, who is often willing to oblige), and the fact that the macro language is limited to very basic and linear functions (no loops, no input processing, no regular expression, no debugging procedure for example), force us to use external scripting methods to generate TOPAS input files.

F.2 Indexing

Proteins are composed of a long chain of chiral amino-acids, which are naturally found in living systems as the L-form. This homochirality results in there being a restricted set of space groups for protein structures. In optimal cases, when peak positions are clearly defined, the restriction of indexing to chiral space groups is not necessary, however, when peaks are broad and weak, peak positions become less accurate and biologically incompatible space groups (with mirrors, glide planes, etc.) can appear in the solutions. We saw that in difficult cases performing the indexing with a restricted set of space groups saved time in finding the correct answer.

In order to obtain an accurate list of peak positions, the GUI mode of TOPAS is the most logical choice. After having loaded the scan file, filled in the instrumental parameters, and zoomed in on the low-angle part of the pattern, a peak search ([View]→[Search Peaks]) will give a first set of peaks (see fig. F.1). In some cases, cleaning of unwanted peaks or adding missed peaks might be needed. For this the peak detail window ([View]→[Peak detail window]) allows peaks to be added and F9 deletes the closest peak.

Then clicking on the peaks phase submenu, “Create Indexing Range”, will keep previously defined parameters (see fig. F.2 - (1) and (2)). It is then necessary to uncheck the checkbox (fig. F.2 - (3)) of the profile, in order to see only the peak positions. In the indexing menu (fig. F.2 - (4)), the “Try space group” option (fig. F.2 - (5)), insert all the chiral space groups and press the run button. Go to the “Solutions” (fig. F.2 - (6)) button to see all the possible solutions sorted by Goodness of fit *vs* volume.

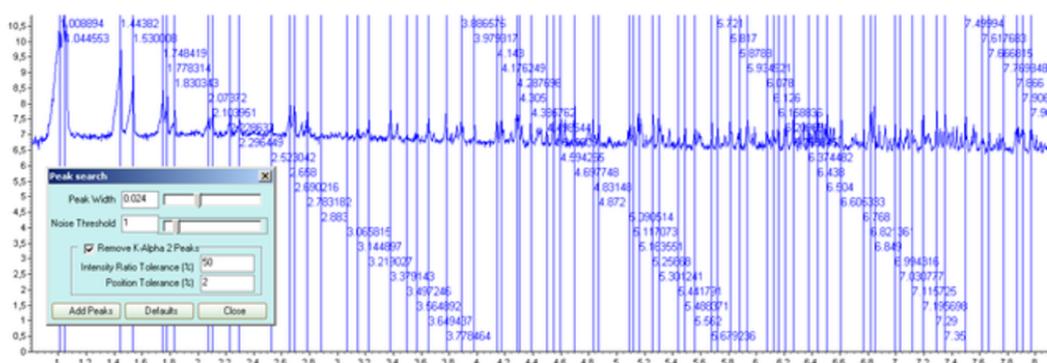


Figure F.1: The peak search view when using the graphical interface.

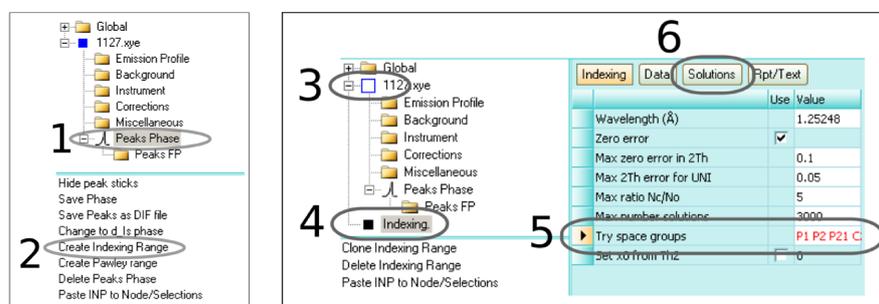


Figure F.2: The peak phase submenu (1) with create indexing range (2), the profile checkbox (3), the indexing menu (4), the try space group option (5) and the solutions menu (6).

F.3 Indexing without the graphical interface

If you don't want to use the graphical interface to define your peak positions (for example if you have already done it with another software) it is possible to use the launch mode in TOPAS[81]. For this you need an input file with the following lines:

```
''' Simplified input for the indexing of a list of
''' peak positions on a restricted group of space groups.
,
' Insert your wavelength here :
index_lam 1.25248
index_zero_error
index_max_zero_error 0.1
' insert all the space group you want to try :
try_space_groups "P212121 C2221 I222 P3221" ' [etc...]
' Insert peak positions in two theta
load index_th2 {
```

```

1.01094152
1.04069078
1.05557656
' [etc...]
}

```

F.4 Pawley fit

On one dataset The launch mode in TOPAS needs an input file (.inp) with the necessary commands, and will output a copy of this input file (.out) with the modifications made to the refined parameters. The input file is preprocessed and then the variable parameters are refined.

In a new session of TOPAS it is first necessary to set the location of the input file with ([Launch] → [Set INP file]), and then to press the run button to load, preprocess and execute the input file. At the end or after stopping the refinement, if you are satisfied with the fit, TOPAS will propose to overwrite or not the original input file with the refined parameters.

The input file is composed of different blocks. A first instrumental block defines the location of the diffraction profile (xdd), the background parameters (bkg), the limits of data to include in the processing (start_X, finish_X), the zero shift (Zero_Error), axial divergence (Simple_Axial_Model), wavelength (lo), calculation step and the extra range beyond finish_X to use for the inclusion of Bragg reflections in the calculated profile (extra_X_right). If a fundamental parameters peak shape is to be used, primary and secondary diffractometer radius (Rp Rs) can be defined (default is 217.5 mm).

A second part is composed of the Pawley parameters (hkl_Is), space_group, lattice parameters, and peak profiles (TCHZ_Peak_Type for example).

To refine a parameter, add a @ in front of the parameter. For example, a @ 93.92614 will allow a to be refined. To refine a parameter called by a function, a comma separating the @ is needed,

```
Simple_Axial_Model( @ , 6.33659).
```

By definition in the TOPAS.inc file, Simple_Axial_Model will refine between 0.00001 and 50, but it is possible to limit the refinement of any parameter to certain values easily, by adding min/max,

```
Simple_Axial_Model( @ , 6.33659 min = 1; max = 10;).
```

```
''' Simplified input for a pawley refinement
''' for glucose isomerase:
'''

```

```

,
xdd "glucose_isomerase.xye"
  bkg      864.847633  1.84438664  295.810305
  start_X  0.9
  finish_X  20.89

  Zero_Error( , 0.00839 min =-0.1; max =0.1;)
  Simple_Axial_Model( , 6.33659)
  lam
    ymin_on_ymax  0.001
    la 1 lo 1.25248 lh 0.1
  x_calculation_step 0.001
  extra_X_right 0.001

  hkl_Is
    a    93.92614
    b    99.54783
    c    102.94948
    phase_name "glucose_isomerase"
    space_group "I222"

    TCHZ_Peak_Type( , -0.00959',
                    , -0.00065', , -0.01223',
                    , 0.01227', , 0.02902',
                    , 0.00116')

```

On multiple datasets When refining parameters with multiple patterns a judicious choice is to use parametric refinement [68]. For example, for three datasets collected during the same experiment, if we want to have the same value for the zero error, the axial divergence, and the same background curve, we can write our `.inp` file like this :

```

prm myZerro  0.00839 min =-0.1; max =0.1;
prm !mySAM   6.33659

prm back1    864
prm back2    1.88
prm back3    295

xdd "first.xye"
...
Zero_Error( myZerro )
Simple_Axial_Model( mySAM )
bkg      =back1; =back2; =back3;
...
hkl_Is

```

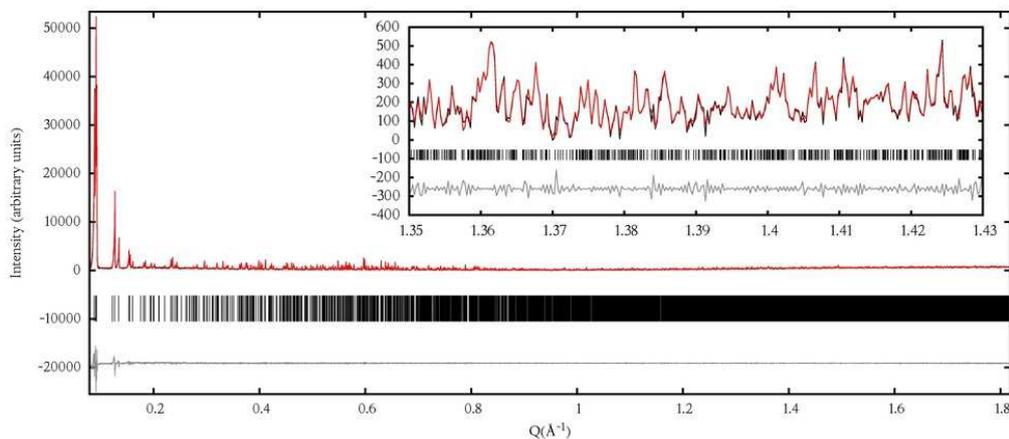


Figure F.3: Pawley fit of Glucose isomerase, I222, $a=93.92614$, $b=99.54783$, $c=102.94948$. Powder pattern collected at ID31 ESRF.

```

...
...
xdd "second.xye"
...
Zero_Error( myZerro )
Simple_Axial_Model( mySAM )
bkg      =back1; =back2; =back3;
...
hkl_Is
...
...
xdd "third.xye"
...
Zero_Error( myZerro )
Simple_Axial_Model( mySAM )
bkg      =back1; =back2; =back3;
...
hkl_Is
...
...

```

When using user defined parameters, ! in front of a parameter stops it refining. If ! is absent the parameter will be refined. In this previous example, myZerro will be refined to the same value for the 3 datasets and mySam

will not be refined and is kept to the same value in all 3 datasets. When using a parameter outside a function call, like the background parameters in this example, it is necessary to define the parameters by using `=parameter;`

If you want to constrain intensities to be equal accros patterns, use a scale factor for each pattern and the `I_parameter_names_have_hkl` keyword which gives each extracted intensity a unique parameter name composed from its hkl indices.

F.5 Pattern calculation with protein structures

Structure definition, scale factor and solvent scattering Once a Pawley fit has been done it is very easy to modify the input file to add a structure. The `hkl_Is` (Pawley) is changed for `str` (structure) and the hkl variables (`hkl_m_d_th2 ... I @ ...`) need to be removed. A scale factor is required, given by `scale`. Protein crystals are always in a hydrated state, water molecules surrounding the protein. These water molecules are either bound to the protein atoms, in this case they should be defined in the atom section as part of the crystal structure, or are in a disordered state. For the disordered water in the crystal, two parameters describing the solvent scattering are defined within TOPAS:

```
protein_As @ 4 min 3 max 6
protein_Bs @ 30 min 10 max 100
```

These parameters modify all the atoms' scattering factors following Babinet's principle[82].

$$f = f_0 - A_s e^{\frac{-8\pi^2 \times B_s \times \sin^2 \theta}{\lambda^2}}$$

Atoms and rigid body Protein structures are commonly described by the .pdb format (protein data bank) with atom positions in orthogonal Å coordinates. Because TOPAS can convert these into fractional coordinates via

the “point_for_site” definition in Cartesian coordinates, it is very easy to incorporate a pdb file into a TOPAS input file.

A TOPAS atom description starts with `site`, then a unique name for the atom, `x y z` and the fractional coordinates (that can be set to 0 if a `point_for_site` is defined), `occ` atom type and occupancy, `beq` = atom B factor.

In order to refine easily an overall B factor, two parameters `Bover` and `DeltaBover` can be defined in the input file. Like this it is possible, by changing these parameters, to have either a scaling of an already known set of B factors from the model, or an average B factor for the whole protein.

A rigid body definition is then necessary to define the atoms’ positions from the pdb. This is done by using `point_for_site` with the Cartesian coordinates of the pdb file. `Ux`, `uy` and `uz` fix the position of the atom, declared in the “site” section, within the rigid body.

For example, 3 atoms and a rigid body with a B factor of 30 are defined like this:

```
prm !Bover 0
prm !DeltaBover 30

''' PDB file format :
'' ATOM      1  N  GLY  A   1      13.749  47.133  32.172  1.00 29.44      N
'' ATOM      2  CA GLY  A   1      13.950  46.145  31.092  1.00 27.61      C
'' ATOM      3  C  GLY  A   1      15.306  45.457  31.262  1.00 38.42      C

''' Topas file format:
site 1_ATOM_1_N_GLY_1_A  x 0.00 y 0.00 z 0.00 occ N 1.00 beq = 29.44 * Bover + DeltaBover;
site 2_ATOM_2_CA_GLY_1_A x 0.00 y 0.00 z 0.00 occ C 1.00 beq = 27.61 * Bover + DeltaBover;
site 3_ATOM_3_C_GLY_1_A  x 0.00 y 0.00 z 0.00 occ C 1.00 beq = 38.42 * Bover + DeltaBover;
rigid
point_for_site 1_ATOM_1_N_GLY_1_A ux 13.749 uy 47.133 uz 32.172
point_for_site 2_ATOM_2_CA_GLY_1_A ux 13.950 uy 46.145 uz 31.092
point_for_site 3_ATOM_3_C_GLY_1_A ux 15.306 uy 45.457 uz 31.262
```

Rotation In most case a pdb structure will need to be (slightly) shifted and rotated in fitting to the powder data. The `Rotate_about_axies()` and `Translate()` functions for a rigid body used together with a good simulated annealing strategy should be sufficient in most cases. However these are correlated as the rotations act about the Cartesian axes. Thus a rotation can move a molecule a considerable distance spatially if its centroid (or centre of mass) is not situated at the origin of the Cartesian co-ordinate system. To avoid this, and reduce correlations between the rotation and the translational parameters, it is prudent to move the centre of mass of the molecule taken from the pdb to the origin.

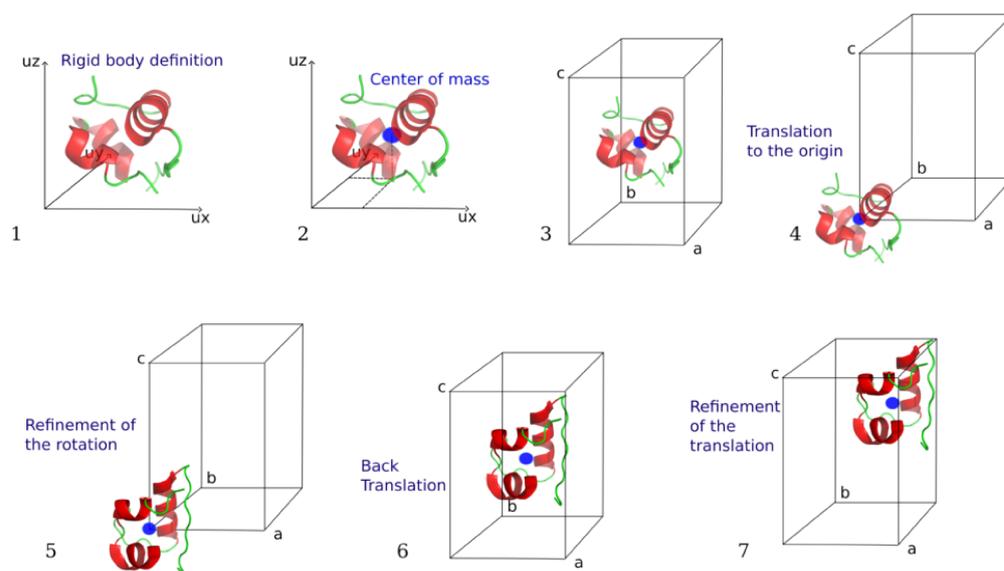


Figure F.4: Pipeline for the refinement of a rigid body in TOPAS. 1) Define point_for_site for all the atoms. 2/3) define the centre of mass of the molecule. 4) Translate the centre of the molecule to the origin. 5) Refine rotation. 6) translate the molecule back in its original place. 7) Refine the translation.

In order to have the rotation of the rigid body centred on the centre of mass of the molecule we need to do the following :

- Define a dummy atom at the centre of mass.
- A first translation of all the atoms except the dummy, to be centred about the origin¹.
- The rotation of all the atoms except the dummy.
- A second translation to move back the atoms.

```
' center of mass
site CenterOfMass x 0.0 y 0.0 z 0.0 occ C 0.00 beq 0
point_for_site CenterOfMass ux 20.1480373831776 uy 38.5541644859813 uz 28.666738317757

Translate_point_amount(CenterOfMass, -) operate_on_points "*" !CenterOfMass"
  rotate ra 0 qa 1 operate_on_points "*" !CenterOfMass"
  rotate rb 0 qb 1 operate_on_points "*" !CenterOfMass"
  rotate rc 0 qc 1 operate_on_points "*" !CenterOfMass"
Translate_point_amount(CenterOfMass, +) operate_on_points "*" !CenterOfMass"
```

1. Dummy is not moved in order to keep a record of where to move back to, after the rotation.

Translation Then it is possible to refine a translation of the molecule within the cell, including this time the centre of mass.

```

translate
  ta @ 0
  tb @ 0
  tc @ 0
' or
Translate( @ 0, @ 0, @ 0 )

```

Including separate structure files To reduce the size of the main input file to a few lines, we can store the atom site and the rigid body point_for_site definitions in a separate file. When TOPAS pre-processes the input file, we instruct it to include the file. However, by doing so, the included file is not modified at the end of the refinement. In order to obtain the final atom fractional positions, it is necessary to use `Out_CIF_STR(out.cif)` to obtain the cif format file. The fractional coordinates can be converted back to pdb format using an external program such as Mercury. We hope that in the future a macro command within TOPAS will allow output of the pdb-compatible Cartesian coordinates directly.

```

' Include while preprocessing atom.inc containing
' the atom and point_for_site definitions.
#include "atoms.inc"

```

Adding #define directives Another possibility inherited from the C language is the ability to use `#define` directives. This allows a block of instructions to be toggled on or off. This block has to be surrounded by `#ifdef` and `#endif` keywords.

For example, to perform a rigid-body refinement on the atoms defined in “atoms2.inc” the TOPAS main input file could look like :

```

' #define structure1  ':means commented line, structure is not loaded
#define structure2    ' use structure2
#define rigidBody     ' do a "rigidBody"

#ifdef structure1
  #include "atoms.inc"
#endif

#ifdef structure2

```

```

#include "atoms2.inc"
#endif

#ifdef rigidBody
  Translate_point_amount(CenterOfMass, -) operate_on_points "* !CenterOfMass"
  rotate ra 0 qa 1 operate_on_points "* !CenterOfMass"
  rotate rb 0 qb 1 operate_on_points "* !CenterOfMass"
  rotate rc 0 qc 1 operate_on_points "* !CenterOfMass"
  Translate_point_amount(CenterOfMass, +) operate_on_points "* !CenterOfMass"
#endif

```

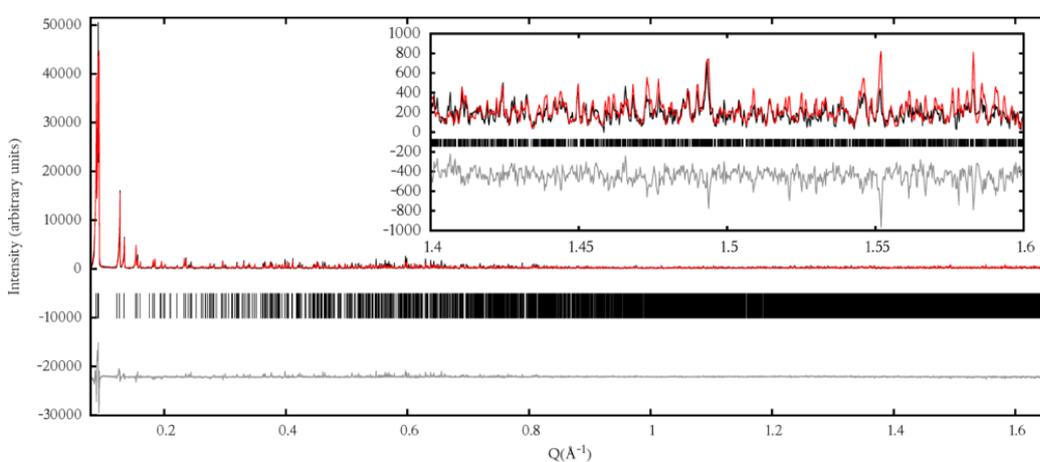


Figure F.5: Glucose isomerase fit with 2G4M pdb structure after rigid-body refinement with solvent-scattering parameters refined, an ideal starting point for a Rietveld refinement.

Grid of starting configurations. In TOPAS, the usual way to carry out simulated annealing is to use “continue_after_convergence” together with “randomize_on_errors” keywords, or the macro “Auto_T(t)”. By using this method, after several steps, when convergence is achieved, all refined parameters are randomized based on their errors, and a new refinement cycle starts. We saw that in some desperate cases, replacing the simulated annealing by a grid of starting configurations could be useful. This involves starting each refinement cycle with a different value of the translation and rotation parameters.

Unfortunately there is no loop possibility in the TOPAS macro language. The only possibility is to test, using the IF() statement, the value of the refinement Cycle. This leads to a long list of nested IF() statements, which for clarity and brevity can be written in a separate file.

In the example below, the “dummy” parameter allows the re-evaluation of the if statement at each new cycle. This is necessary to avoid TOPAS stopping after only one refinement cycle.

For example, in order to have two different starting points along each axis and 2 different orientations about each rotation axis, it is necessary to have $nbsteps_translation^3 \times nbsteps_rotation^3 = 2^3 \times 2^3 = 64$ starting points for the grid search.

```
#ifdef gridsearch
continue_after_convergence
prm dummy 0
Translate_point_amount(centerOfMass_2G4M, -) operate_on_points "* !centerOfMass_2G4M"
rotate = If(Cycle == 0 + dummy 0 , 0,
If(Cycle == 1 + dummy 0 , 0,
If(Cycle == 2 + dummy 0 , 0,
...
If(Cycle == 62 + dummy 0 , 180,
If(Cycle == 63 + dummy 0 , 180,
0))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))
)))))) ; qa 1 operate_on_points "* !centerOfMass_2G4M"
rotate = If(Cycle == 0 + dummy 0 , 0,
If(Cycle == 1 + dummy 0 , 0,
If(Cycle == 2 + dummy 0 , 0,
...
If(Cycle == 62 + dummy 0 , 180,
If(Cycle == 63 + dummy 0 , 180,
0))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))
)))))) ; qb 1 operate_on_points "* !centerOfMass_2G4M"
rotate = If(Cycle == 0 + dummy 0 , 0,
If(Cycle == 1 + dummy 0 , 0,
If(Cycle == 2 + dummy 0 , 0,
...
If(Cycle == 62 + dummy 0 , 180,
If(Cycle == 63 + dummy 0 , 180,
0))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))
)))))) ; qc 1 operate_on_points "* !centerOfMass_2G4M"
Translate_point_amount(centerOfMass_2G4M, +) operate_on_points "* !centerOfMass_2G4M"
translate
ta = If(Cycle == 0 + dummy 0 , 0,
If(Cycle == 1 + dummy 0 , 0,
If(Cycle == 2 + dummy 0 , 0,
...
If(Cycle == 62 + dummy 0 , 0.5,
If(Cycle == 63 + dummy 0 , 0.5,
0))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))
))))))));
```

```

    tb = If(Cycle == 0 + dummy 0, 0,
    If(Cycle == 1 + dummy 0, 0,
    If(Cycle == 2 + dummy 0, 0.5,
    ...
    If(Cycle == 62 + dummy 0, 0.5,
    If(Cycle == 63 + dummy 0, 0.5,
    0))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))
    )))))));

    tc = If(Cycle == 0 + dummy 0 , 0,
    If(Cycle == 1 + dummy 0 , 0.5,
    If(Cycle == 2 + dummy 0 , 0,
    ...
    If(Cycle == 62 + dummy 0 , 0,
    If(Cycle == 63 + dummy 0 , 0.5,
    0))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))
    )))))));

' Cycle 0  $\to$   ra = 0   rb = 0   rc = 0   ta = 0  tb = 0   tc = 0
' Cycle 1  $\to$   ra = 0   rb = 0   rc = 0   ta = 0  tb = 0   tc = 0.5
' Cycle 2  $\to$   ra = 0   rb = 0   rc = 0   ta = 0  tb = 0.5 tc = 0
...
' Cycle 62 $\to$   ra = 180  rb = 180  rc = 180 ta = 0.5 tb = 0.5 tc = 0
' Cycle 63 $\to$   ra = 180  rb = 180  rc = 180 ta = 0.5 tb = 0.5 tc = 0.5
#endif

```

Using the same file with several structures Often when various models exist for one protein structure, an easy way to test several structures without too many files is desirable. It is possible to define several structures in one input file, and by un-commenting the `#define` definition to switch quickly from one to another. Common value for each structure can be defined in a block for `strs{}` which will be called for each structure.

```

#define str_9INS ' uncomment to use 9INS atoms.
#define str_7INS ' uncomment to use 7INS atoms.

#ifdef str_9INS
str
'' original file : 9INS
phase_name 9INS
#include "cubic_insulin_atoms_9INS.inc"
r_wp 16.360
#endif

```

```

#ifdef str_7INS
str
'' original file : 7INS
phase_name 7INS
#include "cubic_insulin_atoms_7INS.inc"
r_wp 15.243
#endif

for xdds {
  for str {
    ''' This will be called for each structures
extra_X_right 0.01
    ''' parameters for proteins
protein_As @ 3 min 3 max 6
protein_Bs @ 10 min 10 max 100
prm !Bover 1
prm !DeltaBover 0 min 10 max 80
    ''' Add your peak shape for all structures :
CS_L(, 2687.81270)
CS_G(, 8116.70362)
    ''' And SG and co ... :
space_group "I213"
Cubic( 78.99610)
scale @ 2.45720973e-007'
} }

```

Generation of a Topas input file from a pdb structure file In order to convert automatically a pdb file into a TOPAS atom input file, create the rigid body definition and perform a grid search, we wrote a perl script called `pdb_to_topas`.

Usage The minimal use of the software is with the `-pdb` and `-topas` options to define the input file and the base name for the outputed files.

```

pdb_to_topas -pdb PDBFILE -topas OUT-TOPAS-BASENAME
-pdb "*.pdb" You can add multiple files surrounding with double quotes or several -pdb
-topas base_name for files generated
[-hoh] if you want to keep water molecules
[-rigidbody] will include the translation and rotation commands
              for the rigid-body refinement of the whole molecule
[-gridsearch] will compute input for different starting rotations and translations of the molecule.
[-stepsrotation{X} 3] for [-gridsearch] define number of steps for rotation (default 3)
[-stepstranslation{X} 3] for [-gridsearch] define number of steps for translation (default 3)
[-minrotation{X} 0] min value for rotation gridsearch (default 0)
[-maxrotation{X} 360] min value for rotation gridsearch (default 360)
[-mintranslation{X} 0] min value for translation gridsearch (default 0)
[-maxtranslation{X} 1] min value for translation gridsearch (default 1)
where {X} = a or b or c.

```

Example For example launching the software with :

```
pdb_to_topas -pdb 9INS.pdb -pdb 7INS.pdb -topas cubic_insulin -rigidbody -gridsearch
```

will create 5 files. `cubic_insulin.inp` is the main file that will need to be edited.

`cubic_insulin_atoms_7INS.inc` and `cubic_insulin_atoms_9INS.inc` contain the atom and rigid-body definitions.

`cubic_insulin_7INS_gridsearch.inc` `cubic_insulin_9INS_gridsearch.inc` contain the starting points for each cycle for the rigid body.



Extracting intensities from powder patterns with PRODD

We are using PRODD in order to extract integrated intensities [8] from multiple patterns of proteins simultaneously. The patterns can have different lattice parameters. Having slightly different lattice parameters, like with different radiation damage levels for example, will give information on the repartition of the reflections overlapped under one peak, provided the “strain” is anisotropic .

PRODD is an interactive command line tool. A set of different files must be prepared before its execution.

- **.dat** files, one per pattern, containing two or three columns, 2θ angle and counts of the powder pattern, and optionally the errors, following this format:

```
7.0846349E-01    2.0304198E+03
7.2307098E-01    1.9806311E+03
7.3767847E-01    1.9261326E+03
7.5228596E-01    1.8776517E+03
```

If no errors are supplied in a third column (like in this example) a square root of the counts is used instead.

- a **.ccl** file (following the Cambridge crystallography subroutine library format[36]) containing the instructions needed for the execution of

PRODD.

G.1 Description of the .ccl command file

```

I NCYC 3 MCOR 0 PRIN 0 PRFO 0 PRPR 2 PRFC 2 ZBAK 1 PVAL 0 FRIE 1
L SCAL 7.1828 1.0000
L SORC CN CN
Z
ZZZZZZZZZZZZZZZZZZZZ DATASET 1 ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZzz
L PKCN *S1 TYPE 2
L RTYP 1 1.333 18.568 0.000
L WVLN 0.9334000
L WGHT 3
L EXCL 0 1.33310
L EXCL 18.5676 100
L ZERO 0.0083
L BACK 5 1.33310 1676.22729
L BACK 1.88138 1270.63953
L BACK 2.19800 1207.78577
L BACK 3.35500 1020.59705
L BACK 4.22000 1048.79150
L BACK 5.00000 1047.15759
L BACK 5.93300 1031.19177
L BACK 6.55000 1008.00378
L BACK 8.47000 1028.26099
L BACK 9.27000 1039.54333
L BACK 13.57000 1346.85474
L BACK 13.93000 1436.76685
L BACK 15.96000 1725.55090
L BACK 16.63000 1750.57605
L BACK 17.42000 1710.92065
L BACK 18.12000 1591.83594
L BACK 18.56760 1408.90698
Z
ZZZZZZZZZZZZZZZZZZZZ DATASET 2 ZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZZzz
L PKCN *S2 TYPE 2
L RTYP 1 1.150 12.100 0.000
Z 1 means read hkl file
Z 2 means generate new hkls (==powpref)
L WVLN 0.9334000

```

```

L WGHT 3
L EXCL 0 1.41
L EXCL 12.1 100
L ZERO 0.0031
L BACK 5 1.41000 1738.09888
L BACK 1.61500 1524.75415
L BACK 1.90473 1325.22485
L BACK 2.12201 1203.00928
L BACK 3.07000 926.04126
L BACK 3.82000 823.43658
L BACK 5.00000 718.21051
L BACK 6.53000 635.44250
L BACK 7.60000 605.13007
L BACK 8.53000 587.39490
L BACK 9.13300 580.34589
L BACK 9.97000 580.88672
L BACK 12.10000 618.91943
**
N Xdomain on ID14-eh1, ESRF P61 61.7434 95.2478
C 61.642483 61.642483 95.540131 90.00000 90.00000 120.00000
S GRUP P 61
L REFI DILS
L PKFN *S1 TYPE 4
Z U V W
L PKFN SIGX 0.0000 0.00000 0.00000 0.00000
Z X Y
L PKFN GAMX 5.0000 0.428507E-01 0.253550
L PKFN SOVL 0.0000 0.00000
L PKFN DOVL 0.0000 0.00000
Z
Z
L PKFN *S2 TYPE 4
Z U V W
L PKFN SIGX 0.0000 0.00000 0.00000 0.00000
Z X Y
L PKFN GAMX 5.0000 0.263794E-01 0.628890E-01
L PKFN SOVL 0.0000 0.00000
L PKFN DOVL 0.0000 0.00000
Z
Z
Z

```

```

ZL VARY *S1  GAMX X
ZL VARY *S1  GAMX Y
ZL VARY *S1  ALL BACK
ZL VARY *S2  GAMX X
ZL VARY *S2  GAMX Y
ZL VARY *S2  ALL BACK
Z
ZL VARY *S1  ALL CELL
ZL VARY *S2  ALL CELL
ZL VARY ZERO 1
ZL VARY ZERO 2
ZL VARY SCALE 1
Z
X DILS 0 0.50 0.50

```

G.1.1 Comments

Any line starting by Z will be ignored by PRODD.

G.1.2 Cycles and scales

```

I NCYC 3 MCOR 0 PRIN 0 PRFO 0 PRPR 2 PRFC 2  ZBAK 1 PVAL 0 FRIE 1
L SCAL  7.1828      1.0000
L SORC CN CN

```

The header of the file contains the number of refinement cycles that PRODD will execute (NCYC₃ for 3 cycles). The scale between the datasets (L_{SCAL}_{7.1828}_{1.00}, 7 for the first dataset, 1 for the second), and when using multiple datasets, L_{SORC}_{CN}_{CN} will indicate how many patterns to expect (one CN must be added per dataset).

G.1.3 Datasets and background

```

L PKCN *S1 TYPE 2
L RTYP  1      1.333      18.568      0.000
L WVLN  0.9334000
L WGHT  3
L EXCL  0 1.33310
L EXCL  18.5676 100

```

```

L ZERO      0.0083
L BACK      5          1.33310   1676.22729
L BACK      1.88138   1270.63953
L BACK      2.19800   1207.78577
L BACK      3.35500   1020.59705
L BACK      4.22000   1048.79150
L BACK      5.00000   1047.15759
L BACK      5.93300   1031.19177
L BACK      6.55000   1008.00378
L BACK      8.47000   1028.26099
L BACK      9.27000   1039.54333
L BACK     13.57000   1346.85474
L BACK     13.93000   1436.76685
L BACK     15.96000   1725.55090
L BACK     16.63000   1750.57605
L BACK     17.42000   1710.92065
L BACK     18.12000   1591.83594
L BACK     18.56760   1408.90698

```

The dataset described in the following line will be called *S1 for the variables (L_PKCN_*S1_TYPE_2).

L_RTYP_1_1.333_18.568_0.000 the first number 1 or 2 is to select or not the generation or refinement of hkls (1=refine, 2=generate (equivalent to powpref in gsas)) The second and third numbers are the 2θ limits for the hkls.

The wavelength is defined by :L_WVLN_0.9334000

The regions that need to be excluded from the pattern L_EXCL_0_1.33310 and L_EXCL_18.5676_100.

The zero error is defined by L_ZERO_0.0083

A first background point is defined with a first number 5 meaning background function number 5 and the two-theta position and value of the background :L_BACK_5_1.33310_1676.22729, and the following background points are defined like this L_BACK_1.88138_1270.63953.

G.1.4 Phase information

**

N Xdomain on ID14-eh1, ESRF P61 61.7434 95.2478

C 61.642483 61.642483 95.540131 90.00000 90.00000 120.00000

```

S GRUP P 61
L REFI DILS
L PKFN *S1 TYPE 4
Z
L PKFN SIGX 0.0000 U 0.00000 V 0.00000 W 0.00000
Z
L PKFN GAMX 5.0000 0.428507E-01 0.253550
L PKFN SOVL 0.0000 0.00000
L PKFN DOVL 0.0000 0.00000
Z

```

The line starting with N is the phase name where is noted any information about the phase (it is not used by the software, but nonetheless required).

The line starting with C is where is defined the lattice C $a, b, c, \alpha, \beta, \gamma$.

S GRUP P 61 is the line where is defined your space group.

L PKFN *S1 TYPE 4 is for defining the peak function type 4 for the pattern *S1.

L PKFN SIGX is the Lorentzian parameters of the peak shape, the first number is the amount in % of the tail of the peak to exclude, then the parameters U, V and W.

L PKFN GAMX is the Gaussian parameters of the peak shape, the first number is the amount in % of the tail of the peak to exclude, then the parameters X and Y.

L PKFN SOVL and L PKFN DOVL are to model the peak shape asymmetry, the first number is the amount in % of the tail of the peak to exclude, then the asymmetry parameter S or D.

Very clear information about peak shape functions can be found in the GSAS manual[6].

G.1.5 Refining parameters

```

L VARY *S1 GAMX X
L VARY *S1 GAMX Y
L VARY *S1 ALL BACK
L VARY *S1 ALL BACK
L VARY ZERO 1
L VARY SCALE 1
Z
X DILS 0 0.50 0.50

```

Adding `L_VARY` + pattern name + variable will refine this variable. `X_DILS 0 0.50 0.50` are parameters for the behaviour of the least squares refinement, first number for the background (the lower the number, the closer to the profile the background will be), second and last numbers are damping factors.

G.2 Executing the software

This is done by answering the questions asked by PRODD:

```
mprodd
```

```
This is MPRODD version 06.07.06
Give name for Output file           output
Give name of Crystal data file      input.ccl
Give name of File to read h,k,l list *P1 *S1  input1.hkl
Give name of File to read h,k,l list *P1 *S2  input1.hkl
Give name of Observations file      a.dat
Give name of Next obs file          b.dat
T
>>> Starting cycle    1
Processing phase 1

          Total  Src  1  Src  2
R_p      0.251  0.210  0.335
R_pb     4.543  2.920 15.552
R_wp     0.466  0.469  0.459
R_wpb    6.096  5.147 15.950
R_e      3.0418 2.7096 3.6237
R_eb    39.7891 29.7121125.9060
Nobs      2220   1180   1040

Chi^2    0.0235  0.0300  0.0160
For      0 variables
Time in seconds for this cycle:      1.25
DILS refined 4622 ints + 0 basic vars,
using 1497848 matrix elements
About 0.85 /s for DILS, 0.61/s for
the matrix, 1 iterations
```

```

LSQ variables converged, continuing for ints

...

More cycles (Y/N)?n
>>> Starting cycle    3
Processing phase 1
Give name for File to output Profiles  output

      Total  Src  1  Src  2
R_p    0.251  0.210  0.335
R_pb   4.540  2.917 15.546
R_wp    0.466  0.469  0.459
R_wpb   6.095  5.148 15.943
R_e    3.0418 2.7096 3.6237
R_eb  39.7891 29.7121125.9060
Nobs    2220   1180   1040

Chi^2  0.0235  0.0300  0.0160
For    0 variables
Do you want to write out the DILS matrix?          n
Time in seconds for this cycle:    4.84
DILS refined  4622 ints +    0 basic vars,
      using    1497848 matrix elements
About    0.84 /s for DILS,    0.60/s for
      the matrix,    1 iterations
Give name for New Crystal Data file output
Give name for HKL list for source  1 and phase  1    hkl1
Give name for HKL list for source  2 and phase  1    hkl2

output.ccn will contain the refined parameters and can be copied over
your .ccl file if you're satisfied with it.

```

G.3 Plotting the results

You can use any plotting software you want on the .pro file output by PRODD.

For example with `gnuplot` :

```
gnuplot>  
plot "a.dat" u 1:2 w l lw 5 lc 2, "b.dat" u 1:2 w l lw 5 lc 5, \  
"output.pro" u 1:2 w l lc 3, "output_old.pro" u 1:4 w l lc 1 lw 2
```

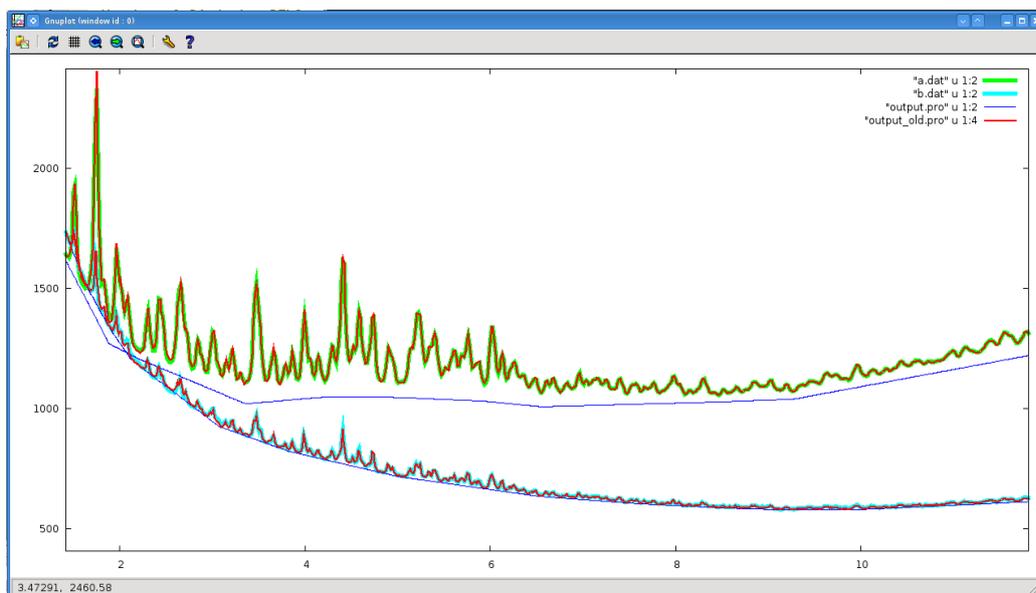


Figure G.1: Gnuplot can be used to see the fitted powder patterns refined within PRODD.



Rietveld refinement with GSAS

The software “General Structure Analysis System” (GSAS)[6] is actually the only software able to perform a Rietveld refinement of a protein with an automatic set of chemical restraints (bond, angles, dihedrals). GSAS is an interactive software, but we are also using the graphical interface called EXPGUI[83] when it is possible to perform routine tasks.

GSAS has a very well documented user manual which should be used as reference for each step. However, a beginner using GSAS for the first time might be lost in the order to proceed. Very general steps to perform a protein powder analysis can be :

1. Prepare the .EXP file :
 - (a) Insert a new histogram.
 - (b) Define limits for the histogram.
 - (c) Set the wavelength for the histogram.
 - (d) Insert a new phase, define lattice parameters and spacegroup.
2. LeBail refinement :
 - (a) Add a “dummy” atom (for example C at $x=0$ $y=0$ $z=0$).
 - (b) Set initial peak shape “profile” values.
 - (c) Define graphically an initial background.

- (d) Execute powpref.
- (e) Execute genles with 0 cycles until intensities are seen in liveplot.
- (f) Refine using genles until obtaining a satisfactory LeBail fit:
 - i. peak shape parameters.
 - ii. lattice parameters and zero shift.
 - iii. background parameters.
- (g) Fix these parameters for the Rietveld refinement.

3. Rietveld refinement :

- (a) Remove the dummy atom used for the LeBail refinement.
- (b) Insert pdb atoms.
- (c) Define restraints using the .mac files in the directory gsas/macros/.
- (d) Insert and search for Z-matrix constraints.
- (e) Set $\Delta f'$ and $\Delta f''$ for each atom type for the wavelength used.
- (f) Refine Scale factor
- (g) Set and refine initial solvent scattering parameters.
- (h) Refine atom positions.
- (i) Extract refined pdb.
- (j) Check geometrical parameters and adjust restraint weights.

List of Figures

1.1	A single crystal	12
2.1	ID31 hutch	20
2.2	ID11 Laue/Laue monochromator	22
2.3	ID11 transfocator	22
2.4	ID14-1 experimental hutch	23
2.5	ID14-2 experimental hutch	23
2.6	Harker construction for MIR	32
2.7	Harker construction for MIR	32
3.1	Urate oxidase catalyses the conversion of uric acid into 5-hydroxyisourate.	42
3.2	Complete catalysis of uric acid into allantoin.	43
3.3	Low angle part for all urate oxidase phases.	51
3.4	LeBail fit of urate oxidase $P2_12_12$	52
3.5	Cell parameter variation with PEG8K	54
3.6	pH effect on urate oxidase $P2_12_12$	54
3.7	pH effect on urate oxidase $P2_12_12$	54
3.8	LeBail fit of urate oxidase $P3_121$	55
3.9	LeBail fit of urate oxidase $P3_121$	56
3.10	Radiation effect on urate oxidase $P3_121$	58
3.11	LeBail fit of urate oxidase $P2_1$	59
3.12	LeBail fit of urate oxidase $P2_1$	60
3.13	LeBail fit of urate oxidase $P2_1$	61
3.14	LeBail fit of urate oxidase $I222$	63
4.2	Urate oxidase, Molrep results.	67
4.1	Urate oxidase Pawley extraction	68
4.3	Urate oxidase, molecular replacement maps.	69
4.4	Urate oxidase, molecular replacement maps.	70
4.5	Urate oxidase, Rietveld fit	71
4.6	Differences between single crystal and powder structures.	73

4.7	Differences between single crystal and powder structures. . . .	74
4.8	Urate oxidase, doughnut shape	76
4.9	8-azaxanthine in Urate oxidase structure	76
5.1	<i>Aedes aegypti</i>	79
5.2	80
5.3	Alphavirus genome coding strategy	81
5.4	Organization of Mayaro virus genome	81
5.5	Mayaro virus, crystallisation drop.	85
5.6	Mayaro first attempt on ID31, cluster	86
5.7	Mayaro first attempt on ID31, summed file	87
5.8	Rietveld fit of mayaro virus molecular replacement solution (ID11).	89
5.9	Rietveld fit of mayaro virus molecular replacement solution (ID11).	89
5.10	Rietveld fit of mayaro virus molecular replacement solution (ID11), after rigid body refinement.	89
5.11	Rietveld fit of mayaro virus molecular replacement solution (ID11), after rigid body refinement.	89
5.12	The skin problem for fishing urchin like crystals.	90
5.13	Mayaro virus, crystallisation drop.	91
5.14	Mayaro crystals in the shape of urchins mounted on micromeshes kept to a 99% humidity in air to avoid dehydration.	92
5.15	Mayaro virus : 2D images collected on ID14	93
5.16	3D views of zoomed regions of the detector	94
5.17	Mayaro indexing of ID14-eh1 data	96
5.18	Mayaro indexing of ID14-eh1 data	97
5.19	Mayaro prodd extraction of ID14-eh1 data	98
5.20	99
5.21	Mayaro Rietveld fit ID14	101
5.22	Preliminary model of the Mayaro macro domain.	102
6.1	Effect of radiation damage on a human insulin T6 sample. . .	110
6.2	Effect of radiation damage on a human insulin T6 sample. . .	110
6.3	An important shift in lattice parameters is observed at room temperature, only after a few fast scans of 1 minute.	110
6.4	Anisotropic shifts in peak positions caused by radiation damage	111
6.5	Effect of radiation damage on sample.	111
6.6	Cryocooling protocol	113

6.7	A typical sample holder for room temperature protein powder diffraction. The brass parts are made to hold the 1 mm borosilicate capillary centred in the beam while being spun at high speed (typically 1000 rpm).	113
6.8	100% glycerol cryocooled	113
6.9	Cryocooling setup	115
6.10	Cryocooling setup	115
6.11	96 well plate used to determine suitable cryoprotectants	116
6.12	ice peaks in diffractino patterns	118
6.13	Lattice parameters shifts due to cryoprotectants	119
6.14	Cryocooled pattern ot Insulin T6 vs room temperature	120
6.15	Cryocooling broadening effect	121
6.16	The cylindrical annulus shape of protein sample for cryocooling	122
6.17	Evolution of the cell parameters during 13.3 hours of collection time on the same sample of cryocooled human insulin T3R3. .	122
6.18	Evolution of the diffraction pattern of a insulin sample cryocooled during collection time.	123
6.19	An easy protocol for cryocooling of a protein powder sample. .	124
6.20	Rietveld fit of human insulin T6 cryocooled at 100K.	126
6.21	Rietveld fit of human insulin T3R3 cryocooled at 5K.	127
6.22	The human insulin T6 structure refined	127
7.1	Crystals of baculovirus. Left a recombinant crystal, right a wild (natural) crystal.	133
7.2	Insect cells infected by a baculovirus, crystals are formed inside the cell.	133
7.3	Pawley fit of AcMNPV at Room temperature	138
7.4	Pawley fit of CpGV at room temperature	139
7.5	Pawley fit of AcMNPV cryocooled	140
7.6	Pawley fit of CpGV at 80K	140
7.7	Background subtraction of AcMNPV room temperature. . . .	142
7.8	3JVB structure, with prior background subtraction and with solvent scattering not refined.	144
7.9	3JW6 structure, with prior background subtraction and with solvent scattering not refined.	144
7.10	3CNF structure, with prior background subtraction and with solvent scattering not refined.	144
7.11	2OH5 structure, with prior background subtraction and with solvent scattering not refined.	145
7.12	2OH6 structure, with prior background subtraction and with solvent scattering not refined.	145

7.13	2OH7 structure, with prior background subtraction and with solvent scattering not refined.	145
7.14	2WUX structure, with prior background subtraction and with solvent scattering not refined.	146
7.15	2WUY structure, with prior background subtraction and with solvent scattering not refined.	146
7.16	3JVB structure, with prior background subtraction and solvent scattering refined.	148
7.17	3JW6 structure, with prior background subtraction and solvent scattering refined.	148
7.18	3CNF structure, with prior background subtraction and solvent scattering refined.	148
7.19	2OH6 structure, with prior background subtraction and solvent scattering refined.	149
7.20	2OH5 structure, with prior background subtraction and solvent scattering refined.	149
7.21	2OH7 structure, with prior background subtraction and solvent scattering refined.	149
7.22	2WUY structure, with prior background subtraction and solvent scattering refined.	150
7.23	2WUX structure, with prior background subtraction and solvent scattering refined.	150
7.24	Statistics for each model.	151
7.25	AcMNPV at room temperature with 3CNF structure.	152
7.26	Rietveld fit of the 3JW6 structure	154
7.27	Correlation between single crystal and powder intensities	155
7.28	CpGV single crystal structure Rietveld fit	156
7.29	Omit maps biased by model	157
A.1	Protocol for sample preparation at room temperature	168
A.2	Batch crystallisation	170
A.3	A typical 1.5 mm capillary	170
A.4	Microcrystals are quickly centrifuged in their crystallisation tube prior to capillary transfer.	171
A.5	A capillary centrifuge is used to pack microcrystals to the bottom of the capillary.	171
A.6	Silson membrane preparation	172
A.7	The sample can be rotated if necessary.	172
A.8	A Silson membrane on a sample holder, adapted from a scapel blade holder, used on the ID31 diffractometer.	173

B.1	Electrophoresis protocol	177
B.2	Centrifuging on top of common oils	178
B.3	Centrifuging over a layer of high density oil	179
B.4	Stock solution of heavy oil, consisting of paraffin oil saturated with CBr_4	180
B.5	Heavy oil : during centrifuging	180
C.1	Changing the humidity of a protein powder sample	183
C.2	Controlled dehydration in situ of glucose isomerase protein powder.	186
C.3	Scores for the different molecular replacement solutions found	187
C.4	Rietveld fit of the molecular replacement solution found for Glucose Isomerase.	187
D.1	ID31 analyser crystals and detector.	190
D.2	A dendogram of similarities.	195
D.3	Fable, open images.	196
D.4	Fable, zoom line.	197
D.5	FIT2D, arrays.	198
D.6	FIT2D, image.	199
D.7	FIT2D, beam centre, circle coordinates.	200
D.8	FIT2D, radial transformation.	201
D.9	FIT2D, a bad beam centre example, radial transform.	201
D.10	FIT2D, a bad beam centre example, radial integration.	202
E.1	An example of a poor Le Bail refinement, with a high order background function; when peaks exhibit severe overlapping, the background is not determined accurately.	208
E.2	Rolling and smoothing function.	210
E.3	The resulting subtracted pattern.	210
E.4	Subtracted background with in red the points belonging to a normal distribution, and in green the points considered as outliers.	212
E.5	A quality evaluation of one bin	213
F.1	The peak search view	217
F.2	The peak phase submenu	217
F.3	Pawley fit of Glucose isomerase	220
F.4	Pipeline for the refinement of a rigid body in Topas	223
F.5	Glucose isomerase fit with 2G4M pdb structure	225

- G.1 Gnuplot can be used to see the fitted powder patterns refined within PRODD. 239

List of Tables

3.2	Urate oxidase, structural models.	49
3.1	Urate oxidase, LeBail analysis.	50
4.1	Urate oxidase, current structural refinement using GSAS, using the 1R51 pdb structure as starting model.	75
5.1	Mayaro N-terminus domain of nsP3 sequence	83
5.2	Mayaro structural refinement with GSAS, comparison between the molecular replacement solution and the current model. . .	102
6.1	Cryocooled human insulin T3R3 at 5K	128
6.2	Cryocooled human insulin T6 at 100K, structural refinement using GSAS, using the 1MSO pdb structure as starting model.	128
7.1	The Baculovirus and Cypoviruses	133
7.2	Major Structural proteins of Baculoviruses.	136
7.3	Statistical values for different models against AcMNPV at room temperature.	151
7.4	Résumé of R_{wp} for different structures, with or without solvent scattering refinement.	151
7.5	Cubic symmetry at high angle shows exact overlap of numerous reflections.	157

Bibliography

- [1] H. Rietveld, “A profile refinement method for nuclear and magnetic structures,” *Journal of Applied Crystallography*, vol. 2, no. 2, pp. 65–71, 1969.
- [2] R. Corey and R. Wyckoff, “Long spacings in macromolecular solids,” *Journal of Biological Chemistry*, vol. 114, no. 2, p. 407, 1936.
- [3] J. Bernal and I. Fankuchen, “X-ray and crystallographic studies of plant virus preparations: I. Introduction and preparation of specimens II. Modes of aggregation of the virus particles,” *Journal of General Physiology*, vol. 25, no. 1, p. 111, 1941.
- [4] L. Amos, J. Jubb, R. Henderson, and G. Vigers, “Arrangement of protofilaments in two forms of tubulin crystal induced by vinblastine,” *Journal of Molecular Biology*, vol. 178, no. 3, pp. 711–729, 1984.
- [5] R. Von Dreele, “Combined Rietveld and stereochemical restraint refinement of a protein crystal structure,” *Journal of Applied Crystallography*, vol. 32, no. 6, pp. 1084–1089, 1999.
- [6] A. C. Larson and R. B. Von Dreele, “General structure analysis system (gsas),” *Los Alamos National Laboratory Report LAUR 86-748*, Los Alamos, USA, 2004.
- [7] R. Von Dreele, P. Stephens, G. Smith, and R. Blessing, “The first protein crystal structure determined from high-resolution X-ray powder diffraction data: a variant of T3R3 human insulin-zinc complex produced by grinding,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 56, no. 12, pp. 1549–1553, 2000.
- [8] J. Wright, “Extraction and use of correlated integrated intensities with powder diffraction data,” *Zeitschrift für Kristallographie/International journal for structural, physical, and chemical aspects of crystalline materials*, vol. 219, no. 12/2004, pp. 791–802, 2004.
- [9] I. Margiolaki, J. Wright, A. Fitch, G. Fox, and R. Von Dreele, “Synchrotron X-ray powder diffraction study of hexagonal turkey egg-white

- lysozyme,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 61, no. 4, pp. 423–432, 2005.
- [10] I. Margiolaki, J. Wright, M. Wilmanns, A. Fitch, and N. Pinotsis, “Second SH3 domain of ponsin solved from powder diffraction,” *J. Am. Chem. Soc.*, vol. 129, no. 38, pp. 11865–11871, 2007.
- [11] J. P. Wright, C. Besnard, I. Margiolaki, S. Basso, F. Camus, A. N. Fitch, G. C. Fox, P. Pattison, and M. Schiltz, “Molecular envelopes derived from protein powder diffraction Molecular envelopes derived from protein powder diffraction data,” *Journal of Applied Crystallography*, vol. 41, 2008.
- [12] F. Elder, A. Gurewitsch, R. Langmuir, and H. Pollock, “Radiation from electrons in a synchrotron,” *Physical Review*, vol. 71, no. 11, pp. 829–830, 1947.
- [13] A. Fitch, “The high resolution powder diffraction beam line at ESRF,” *Journal of research-national institute of standards and technology*, vol. 109, pp. 133–142, 2004.
- [14] J. Wright, C. Besnard, I. Margiolaki, S. Basso, F. Camus, A. Fitch, G. Fox, P. Pattison, and M. Schiltz, “Molecular envelopes derived from protein powder diffraction data,” *Journal of Applied Crystallography*, vol. 41, no. 2, pp. 329–339, 2008.
- [15] J. Helliwell, M. Helliwell, and R. Jones, “Ab initio structure determination using dispersive differences from multiple-wavelength synchrotron-radiation powder diffraction data,” *Acta Crystallographica Section A: Foundations of Crystallography*, vol. 61, no. 6, pp. 568–574, 2005.
- [16] K. Kahn, P. Serfozo, and P. Tipton, “Identification of the true product of the urate oxidase reaction,” *J. Am. Chem. Soc.*, vol. 119, no. 23, pp. 5435–5442, 1997.
- [17] X. Wu, D. Muzny, C. Chi Lee, and C. Thomas Caskey, “Two independent mutational events in the loss of urate oxidase during hominoid evolution,” *Journal of molecular evolution*, vol. 34, no. 1, pp. 78–84, 1992.
- [18] B. Ames, R. Cathcart, E. Schwiers, and P. Hochstein, “Uric acid provides an antioxidant defense in humans against oxidant-and radical-caused aging and cancer: a hypothesis,” *Proceedings of the National Academy of Sciences*, vol. 78, no. 11, p. 6858, 1981.
- [19] T. Yamamoto, Y. Moriwaki, and S. Takahashi, “Effect of ethanol on metabolism of purine bases (hypoxanthine, xanthine, and uric acid),” *Clinica Chimica Acta*, vol. 356, no. 1-2, pp. 35–57, 2005.

- [20] N. Colloc'h, M. El Hajji, B. Bachet, G. L'Hermite, M. Schiltz, T. Prangé, B. Castro, and J. Mornon, "Crystal structure of the protein drug urate oxidase-inhibitor complex at 2.05 Å resolution," *Nature Structural & Molecular Biology*, vol. 4, no. 11, pp. 947–952, 1997.
- [21] C. Aleman, B. A., B. T., and P. Dupin, "Stable liquid composition containing urate oxidase and lyophilized composition for its preparation," September 1999.
- [22] B. McGrath and G. Walsh, *Directory of therapeutic enzymes*. Taylor & Francis.
- [23] L. Cammalleri and M. Malaguarnera, "Rasburicase represents a new tool for hyperuricemia in tumor lysis syndrome and in gout," *International Journal of Medical Sciences*, vol. 4, no. 2, p. 83, 2007.
- [24] S. Havelund, "Pulmonary insulin crystals," July 6 2007. US Patent App. 11/774,027.
- [25] M. Brader and M. Sukumar, "Crystalline compositions for controlling blood glucose," June 30 2004. US Patent App. 10/880,884.
- [26] B. Rabinow, "Nanosuspensions in drug delivery," *Nature Reviews Drug Discovery*, vol. 3, no. 9, pp. 785–796, 2004.
- [27] A. Aguiar, J. Krc Jr, A. Kinkel, and J. Samyn, "Effect of polymorphism on the absorption of chloramphenicol from chloramphenicol palmitate," *J. Pharm. Sci.*, vol. 56, no. 7, pp. 847–853, 1967.
- [28] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, "Ritonavir: an extraordinary example of conformational polymorphism," *Pharmaceutical research*, vol. 18, no. 6, pp. 859–866, 2001.
- [29] M. Giffard, N. Colloc'h, N. Ferté, B. Castro, and F. Bonneté, "Salting-In Effects on Urate Oxidase Crystal Design," *Cryst. Growth Des*, vol. 8, no. 12, pp. 4220–4226, 2008.
- [30] A. Coelho, "Topas-academic," pp. <http://www.bruker-axs.de/topas.html> (accessed Jan 20, 2010)., 2004.
- [31] A. L. Bail, H. Duroy, and J. Fourquet, "Ab-initio structure determination of lisbwo6 by x-ray powder diffraction," *Materials Research Bulletin*, vol. 23, no. 3, pp. 447–452, 1988.
- [32] P. Retailleau, N. Colloc'h, D. Vivarès, F. Bonneté, B. Castro, M. El Hajji, and T. Prangé, "Urate oxidase from *Aspergillus flavus*: new crystal-packing contacts in relation to the content of the active site," *Acta Crystallographica Section D: Biological Crystallography*, vol. 61, no. 3, pp. 218–229, 2005.

- [33] P. Retailleau, N. Colloc'h, D. Vivarès, F. Bonneté, B. Castro, M. El Haggi, J. Mornon, G. Monard, and T. Prangé, "Complexed and ligand-free high-resolution structures of urate oxidase (Uox) from *Aspergillus flavus*: a reassignment of the active-site binding mode," *Acta Crystallographica Section D: Biological Crystallography*, vol. 60, no. 3, pp. 453–462, 2004.
- [34] P. M. Navolanic, C. Pui, R. A. Larson, M. R. Bishop, T. E. Pearce, M. S. Cairo, S. C. Goldman, S. C. Jeha, C. B. Shanholtz, J. P. Leonard, and J. A. McCubrey, "Elitek rasburicase: an effective means to prevent and treat hyperuricemia associated with tumor lysis syndrome, a meeting report, dallas, texas, january 2002," *Leukemia*, vol. 17, no. 3, pp. 499–514, 2003.
- [35] G. Pawley, "Unit-cell refinement from powder diffraction scans," *Journal of Applied Crystallography*, vol. 14, no. 6, pp. 357–361, 1981.
- [36] J. Matthewman, P. Thompson, and P. Brown, "The cambridge crystallography subroutine library," *Journal of Applied Crystallography*, vol. 15, no. 2, pp. 167–173, 1982.
- [37] I. Margiolaki and J. P. Wright, "Powder crystallography on macromolecules," *Acta Cryst. A*, vol. 64, pp. 169–180, Jan 2008.
- [38] E. Potterton, P. Briggs, M. Turkenburg, and E. Dodson, "A graphical user interface to the CCP4 program suite," *Acta Crystallographica Section D: Biological Crystallography*, vol. 59, no. 7, pp. 1131–1137, 2003.
- [39] A. Vagin and A. Teplyakov, "MOLREP: an automated program for molecular replacement," *Journal of Applied Crystallography*, vol. 30, no. 6, pp. 1022–1025, 1997.
- [40] A. McCoy, R. Grosse-Kunstleve, P. Adams, M. Winn, L. Storoni, and R. Read, "Phaser crystallographic software," *Journal of Applied Crystallography*, vol. 40, no. 4, pp. 658–674, 2007.
- [41] R. Tesh, D. Watts, K. Russell, C. Damodaran, C. Calampa, C. Cabezas, G. Ramirez, B. Vasquez, C. Hayes, C. Rossi, *et al.*, "Mayaro virus disease: an emerging mosquito-borne zoonosis in tropical South America," *Clinical Infectious Diseases*, vol. 28, no. 1, pp. 67–73, 1999.
- [42] R. Tesh, "Arthritides caused by mosquito-borne viruses," *Annual Review of Medicine*, vol. 33, no. 1, pp. 31–40, 1982.
- [43] M. Receveur, M. Grandadam, T. Pistone, and D. Malvy, "Infection with Mayaro virus in a French traveller returning from the Amazon region, Brazil, January, 2010.," *Euro surveillance: bulletin européen sur les maladies transmissibles = European communicable disease bulletin*, vol. 15, no. 18, 2010.

- [44] F. Schaffner and S. Karch, "Première observation d'*Aedes albopictus* (Skuse, 1894) en France métropolitaine: First record of *Aedes albopictus* (Skuse, 1894) in metropolitan France.," *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie*, vol. 323, no. 4, pp. 373–375, 2000.
- [45] A. Lavergne, B. Thoisy, V. Lacoste, H. Pascalis, J. Pouliquen, V. Mercier, H. Tolou, P. Dussart, J. Morvan, A. Talarmin, *et al.*, "Mayaro virus: Complete nucleotide sequence and phylogenetic relationships with other alphaviruses," *Virus research*, vol. 117, no. 2, pp. 283–290, 2006.
- [46] E. Gould, B. Coutard, H. Malet, B. Morin, S. Jamal, S. Weaver, A. Gorbalenya, G. Moureau, C. Baronti, I. Delogu, *et al.*, "Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control," *Antiviral Research*, 2009.
- [47] S. Mukhopadhyay, W. Zhang, S. Gabler, P. Chipman, E. Strauss, J. Strauss, T. Baker, R. Kuhn, and M. Rossmann, "Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses," *Structure*, vol. 14, no. 1, pp. 63–73, 2006.
- [48] G. Karras, G. Kustatscher, H. Buhecha, M. Allen, C. Pugieux, F. Sait, M. Bycroft, and A. Ladurner, "The macro domain is an ADP-ribose binding module," *The EMBO Journal*, vol. 24, no. 11, p. 1911, 2005.
- [49] H. Malet, B. Coutard, S. Jamal, H. Dutartre, N. Papageorgiou, M. Neuvonen, T. Ahola, N. Forrester, E. Gould, D. Lafitte, *et al.*, "The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket," *The Journal of Virology*, vol. 83, no. 13, p. 6534, 2009.
- [50] N. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. Stuart, and R. Owens, "A versatile ligation-independent cloning method suitable for high-throughput expression screening applications," *Nucleic Acids Research*, 2007.
- [51] A. Boultif and D. Louër, "Indexing of powder diffraction patterns for low-symmetry lattices by the successive dichotomy method," *Journal of Applied Crystallography*, vol. 24, pp. 987–993, Dec 1991.
- [52] T. Huang, H. Toraya, T. Blanton, and Y. Wu, "X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard," *Journal of Applied Crystallography*, vol. 26, no. 2, pp. 180–184, 1993.
- [53] F. Tama, F. Gadea, O. Marques, and Y. Sanejouand, "Building-block approach for determining low-frequency normal modes of macro-

- molecules,” *Proteins: Structure, Function, and Bioinformatics*, vol. 41, no. 1, pp. 1–7, 2000.
- [54] W. Burmeister, “Structural changes in a cryo-cooled protein crystal owing to radiation damage,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 56, no. 3, pp. 328–341, 2000.
- [55] R. Ravelli and S. McSweeney, “The ‘fingerprint’ that X-rays can leave on structures,” *Structure*, vol. 8, no. 3, pp. 315–328, 2000.
- [56] C. Blake and D. Phillips, “Effects of X-irradiation on single crystals of myoglobin,” *The basic aspects of radiation effects on living systems*, p. 183, 1962.
- [57] D. Haas and M. Rossmann, “Crystallographic studies on lactate dehydrogenase at 75°C,” *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, vol. 26, no. 7, pp. 998–1004, 1970.
- [58] G. Petsko, “Protein crystallography at sub-zero temperatures: cryoprotective mother liquors for protein crystals,” *Journal of Molecular Biology*, vol. 96, no. 3, pp. 381–388, 1975.
- [59] J. Helliwell, “Protein crystal perfection and the nature of radiation damage,” *Journal of Crystal Growth*, vol. 90, no. 1-3, pp. 259–272, 1988.
- [60] T. Teng, “Mounting of crystals for macromolecular crystallography in a free-standing thin film,” *Journal of Applied Crystallography*, vol. 23, no. 5, pp. 387–391, 1990.
- [61] M. Weik, R. Ravelli, G. Kryger, S. McSweeney, M. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and J. Sussman, “Specific chemical and structural damage to proteins produced by synchrotron radiation,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 2, p. 623, 2000.
- [62] R. Ravelli, M. Nanao, A. Lovering, S. White, and S. McSweeney, “Phasing in the presence of radiation damage,” *Journal of synchrotron radiation*, vol. 12, no. 3, pp. 276–284, 2005.
- [63] M. Jenner, J. Wright, I. Margiolaki, and A. Fitch, “Successful protein cryocooling for powder diffraction,” *Journal of Applied Crystallography*, vol. 40, no. 1, pp. 121–124, 2007.
- [64] E. Garman and R. Owen, “Cryocooling and radiation damage in macromolecular crystallography,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 1, pp. 32–47, 2005.
- [65] J. Murray, E. Garman, and R. Ravelli, “X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition

- on absorbed dose,” *Journal of applied crystallography*, vol. 37, no. 4, pp. 513–522, 2004.
- [66] V. Berejnov, N. Husseini, O. Alsaied, and R. Thorne, “Effects of cryoprotectant concentration and cooling rate on vitrification of aqueous solutions,” *Journal of applied crystallography*, vol. 39, no. 2, pp. 244–251, 2006.
- [67] R. Von Dreele, “A rapidly filled capillary mount for both dry powder and polycrystalline slurry samples,” *Journal of Applied Crystallography*, vol. 39, no. 1, pp. 124–126, 2006.
- [68] G. W. Stinton and J. S. Evans, “Parametric Rietveld refinement,” *J Appl Crystallogr*, vol. 40, pp. 87–95, Feb 2007.
- [69] A. Cooke, “An overview on possible mechanisms of destruction of the insulin-producing beta cell.,” *Current topics in microbiology and immunology*, vol. 164, p. 125, 1990.
- [70] S. Kriminski, C. Caylor, M. Nonato, K. Finkelstein, and R. Thorne, “Flash-cooling and annealing of protein crystals,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 58, no. 3, pp. 459–471, 2002.
- [71] G. Rohrmann, “Baculovirus structural proteins,” *Journal of General Virology*, vol. 73, no. 4, p. 749, 1992.
- [72] L. Miller, “Introduction to the Baculoviruses,” *The Baculoviruses*, pp. 1–6, 1997.
- [73] G. Benz, “Introduction: historical perspectives,” *The biology of baculoviruses*. Boca Raton: CRC, vol. 1, pp. 1–35, 1986.
- [74] D. Lightner and R. Redman, “Shrimp diseases and current diagnostic methods,” *Aquaculture*, vol. 164, no. 1-4, pp. 201–220, 1998.
- [75] R. Jaques, “Persistence, accumulation, and denaturation of nuclear polyhedrosis and granulosis viruses,” *Baculoviruses for Insect Pest Control: Safety Considerations*, pp. 90–99, 1975.
- [76] J. Sanchez-Weatherby, M. Bowler, J. Huet, A. Gobbo, F. Felisaz, B. Lavault, R. Moya, J. Kadlec, R. Ravelli, and F. Cipriani, “Improving diffraction by humidity control: a novel device compatible with X-ray beamlines,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 65, no. 12, pp. 1237–1246, 2009.
- [77] J. Hodeau, P. Bordet, M. Anne, A. Prat, A. Fitch, E. Dooryhee, G. Vaughan, and A. Freund, “Nine crystal multi-analyser stage for high resolution powder diffraction between 6 and 40 keV,” in *SPIE Proceedings*, vol. 3448, pp. 353–361, 1998.

- [78] J. Wright, G. Vaughan, and A. Fitch, “Merging data from a multi-detector continuous scanning powder diffraction system,” *Commission on Crystallographic Computing*, p. 92.
- [79] E. Hammersley, “Internal Report (1998), ESRF98HA01T, FIT2D V9.129 Reference Manual V3. 1; Hammersley, et al,” 1996.
- [80] Y. Watier and A. Fitch, “Protein Powder Diffraction Analysis with TOPAS,” in *Materials Science Forum*, vol. 651, pp. 117–129, Trans Tech Publ, 2010.
- [81] A. Coelho, “Indexing of powder diffraction patterns by iterative use of singular value decomposition,” *Journal of Applied Crystallography*, vol. 36, no. 1, pp. 86–95, 2003.
- [82] M. Babinet, “Mémoires d’optique météorologique,” *CR Acad. Sci*, vol. 4, pp. 638–648, 1837.
- [83] B. Toby, “EXPGUI, a graphical user interface for GSAS,” *Journal of Applied Crystallography*, vol. 34, no. 2, pp. 210–213, 2001.

