**X-ray crystallography structures**

**Project-based approaches and practical tips**

Dom Bellini

Biophysical techniques symposium 10/03/20

---

**X-ray crystallography facility at the LMB**

**Room 1S205**

---

**Dom Bellini**

Crystallography Facility Manager

- Tel: +44 1223 338201
- Email: d.bellini@mrc-lmb.cam.ac.uk
- Room: 1S205

---

**David Barron**

Deputy Head of Cranfield

- Tel: +44 1223 338201
- Email: d.barron@mrc-lmb.cam.ac.uk
- Room: 1S205

---

**Fabrice Corre**

Crystallography Facility Manager

- Tel: +44 1223 338201
- Email: f.corre@mrc-lmb.cam.ac.uk
- Room: 1S205

---

**Jan Lowe**

Deputy Head of Cranfield

- Tel: +44 1223 338201
- Email: j.low@mrc-lmb.cam.ac.uk
- Room: 1S205

---

“talk to us” = Dom and Fabrice
Crystallographic project workflow

Sample prep.  Xtal  Xtal hit optimis.  Xtal Soak/cryo  Xtal harvest  In-house screen  Sychrotron data collection  Data proc.  Phasing, validation  Biology

Typical time consumption

- sample prep.
- xtal./optimis.
- ?
- soak/cryo
- harvest
- in-house screen
- data collection
- structure solution

Sample quality: what to aim for?

Gel filtration (SEC-MALLS)  SDS-PAGE gel
Two crystallization key factors of protein samples: solubility and stability

Proteins, unlike DNA, have a large variation in both structural and physicochemical properties, which in turn will affect solubility and stability (key factors in crystallization).

These two key factors are defined in the “crystallography urban dictionary” as follows:

- **Soluble sample** = a single peak in gel filtration chromatograms that can be concentrated to at least around 5-10 mg/ml
- **Stable sample** = a 5-10 mg/ml soluble sample that produces around 50% clear drops in a crystallization screening experiment

If the ratio of clear/precipitation drops differs significantly from 50:50 either way, you can discuss with us what to do next

How to obtain soluble and stable samples

◆ Skills
◆ Experience
◆ Creativity

**Skills** = lab skills/tricks in working with proteins (e.g., construct design, different overexpression cell strains, different inductions levels or growth temperatures of cultures, different tags for purification, …), which can make the difference between obtaining soluble protein rather than inclusion bodies.

**Experience** = literature knowledge of specific protein families and insights into stabilising conditions, such as right detergent/chemicals, salt concentrations, phosphorylation state, cofactors (e.g., small molecule, peptide, protein-protein interactions, …), …

Once you decide on a target to crystallize, you can come to talk to us for advice on best strategies and/or we can suggest who in the LMB may have the expertise to help you.
Creativity = your talent

Example 1: Lowe’s group

Crystallized monomer (2.1 Å) (unpublished)

Filament-disrupting mutations

Example 2: Hegde’s group

Crystallized P1^*–P2 complex (2.2 Å) (unpublished)

Limited proteolysis

Example 3: Barford’s group

Crystallized 5-protein complex (2.9 Å) (unpublished)

multiple sequence alignment and secondary structure predictions

Example 4: James’ group

Stabilization of HIV hexameric capsid by engineered disulfide bridges for crystallization

Price et al., 2009 (NSMB)

Example 5: Aricescu’s group

Tethered construct to crystallize a protein complex connecting pre- and post-synaptic axons

Elegeheert et al., 2016 (Science)

Example 6: Tate’s labs

Thermostabilization of membrane proteins by systematic mutagenesis (6-12 months!)

Magnani et al., 2016 (Nat. protocols)
Creativity = your talent

Example 7: ESRF’s labs

**ESPRIT (Expression of Soluble Proteins by Random Incremental Truncation)**

Gene bidirectional digestion

Library size selection

Fluorescent colony array screen

Up to 30,000 constructs can be screened in a few weeks

ESPRIT helps to identify soluble expression constructs in difficult cases with no useful sequence homology to other proteins (e.g., the high valuable target influenza virus polymerase subunit PB2, Guilligay et al., 2008, NSMB)

---

**In situ proteolysis**

Limited proteolysis

<table>
<thead>
<tr>
<th>P1</th>
<th>P1*</th>
<th>P2</th>
</tr>
</thead>
</table>

C-
In situ proteolysis

Limited proteolysis

**In situ Protocol**
Add protease to protein sample (e.g., 1:1000 ratio)
**In situ proteolysis**

- *In situ* proteolysis can be performed at the X-ray facility with 3 available proteases to test: trypsin, chymotrypsin and subtilisin
- The protocol (quick and easy) is available on the facility internal webpage
- If you would like help with this you can talk to us

10-20 μl of 13 M K-formate can shrink drops from all different conditions in xtal. plates:

- Drop dehydration can be performed at the X-ray facility with concentrated K-formate solutions readily available
- The protocol (quick and easy) is available on the facility internal webpage
- If you would like help with this you can talk to us
Crystal soaking and cryoprotection are underestimated obstacles that lie between crystallization and crystallography. Difficulties in soaking and cryoprotection can vary greatly depending on a number of factors:

**Crystal robustness**
- anything will work with some really robust crystals
- others (majority) will require systematic screening of many cryoprotectants and/or ligand concentrations at different time courses

**Ligand solubility**
- the ideal ligand is highly soluble in aqueous solutions (however, the ligand should always be solved in a crystal compatible solution)
- poor solubility requires experimenting with chemicals/conditions suitable for both ligand and crystal (it can range from very easy to very hard to impossible)

**Heavy atom soaks (for ab initio phasing)**
- many different strategies and too long topic for this talk
- some suggestions/protocols are available on the X-ray facility webpage

Different levels of expertise can be required to optimise the above procedures depending on the case.
Once you have optimised your crystal hits, you can talk to us for advices on how to prepare the crystals for data collection.
Avoid sudden acceleration while fishing crystals

- Quick training session on technical and safety issues required before using the instrument (email me to arrange training)
- Access to online booking system (Kerio account) is granted after training
- Instructions on how to use it available both on site and on the X-ray facility webpage
How to take advantage of an in-house X-ray generator

1 – Quickly discriminate between protein crystal hits and false positives (e.g., salt crystals)
   - LV cryo-oil (i.e., universal cryoprotectant) available at the X-ray facility

2 – Optimise cryoprotection conditions (checking for ice rings and/or loss of diffraction)

3 – Minimise the time you have to spend to collect at synchrotrons (limited time available and unsocial working hours often at week-ends)

4 – Structures of large well diffracting crystals can be solved in-house (5-20 minutes per frame = 15-60 hrs for 180 images)

5 – Select the best crystals for experiments at low throughput beamlines (e.g., longwavelength beamline, I23, at Diamond)

6 – If there is no imminent synchrotron beamtime available (e.g., maintenance shutdown or overbooking), ligand or heavy atom soaks can be optimised in-house by checking different time courses and ligand concentrations

MX beamlines: ID23-1, ID29 & ID30B – tunable
ID23-2 & ID30A – fixed wavelength microfocus
MASSIF-1 – fully automatic data collection
BM29 – Small Angle X-ray Scattering (SAXS)

MX beamlines: I03 – tunable
I04 & I24 – tunable microfocus
I04-1 – fixed wavelength Se K-edge
I23 – long wavelength (S-SAD and light atoms, K, Ca, Cl and P)
VMXi – fixed wavelength in-situ data collection
VMXm – tunable nanofocus
B21 – Small Angle X-ray Scattering (SAXS)

The schedule for synchrotron trips/sessions can be found on the LMB crystallography webpage.
To request time for experiments at synchrotrons please email to: xraymgr@mrc-lmb.cam.ac.uk
Two data collection options at synchrotons

1 – travelling to the sites

2 – remote data collection

- The X-ray facility can help with data collection both remotely and at synchrotrons (please contact me if you need help with your data collection)
- Remote collection requires shipping samples by courier to synchrotrons and instructions to do it can be found on the crystallography webpage (if you are not sure about the instructions for preparing shipping airway bills please contact me)
A few basic concepts of crystallography

- Theory of crystallography is outside the scope of this talk; moreover, it is not essential to carry out experiments at the LMB facilities

Schools:
- Cold Spring Harbor (New York, USA) – Crystallography course (2 weeks during autumn)

Books:
- Outline of crystallography for biologists (David Blow)
- Protein Crystallography (Blundell & Johnson)
- Biomolecular Crystallography (Bernhard Rupp)
- Structure determination by X-ray crystallography (Ladd and Palmer) (Chemistry level)
- Elements of Modern X-ray physics (Als-Nielsen & McMorrow) (Physics level)

- Fabrice and I are always happy to discuss any issue/question you may have about crystallography

Crystals

14 Bravais lattices
(after Auguste Bravais, 1811-1863)

A crystalline arrangement is defined by a lattice, which is an infinite array of points with identical environment

Asymmetric unit (a.u.) – smallest unit that can generate u.c. via crystal symmetry

Unit cell (u.c.) – basic building (repeat) block of a crystal

Real case example of a.u. (green)
**Spacegroups**

The **spacegroup** is a concise description of all the symmetries in a given 3-dimensional crystal.

**Crystal symmetries:**

- **Rotations**
- **Inversions**
- **Reflections**
  - (mirror planes)
- **Screw axes**
  - (rotation + translation)
- **Glide planes**
  - (reflection + translation)

**Example:** spacegroup P422 4-fold rotation along c, 2-fold rotation along a and b, and a screw axis along c (the letter in front represents the type of Bravais lattice, in this case P = Primitive)

The total number of possible spacegroups = 230

---

**Spacegroups**

The **spacegroup** is a concise description of all the symmetries in a given 3-dimensional crystal.

**Crystal symmetries:**

- **Rotations**
- **Inversions**
- **Reflections**
  - (mirror planes)
- **Screw axes**
  - (rotation + translation)
- **Glide planes**
  - (reflection + translation)

**Example:** spacegroup P422 4-fold rotation along c, 2-fold rotation along a and b, and a screw axis along c (the letter in front represents the type of Bravais lattice, in this case P = Primitive)

The total number of possible spacegroups = 230

**Only 65 spacegroups are possible in protein crystals**
The 65 chiral spacegroups

<table>
<thead>
<tr>
<th>Symmetry</th>
<th>Abbreviation</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>P</td>
<td>P 1</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>P</td>
<td>P 2</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>P 2/m</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>P 21</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>P 2/c</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>P</td>
<td>P 2 2 2</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C 2 2 2</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>I 2 2 2</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>P</td>
<td>P 4</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>I 4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>P 4 2 2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>P 4 2 2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>P 4 3 3</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>P</td>
<td>P 6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>P 6 3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>P 6 3</td>
</tr>
<tr>
<td>Cubic</td>
<td>P</td>
<td>P 2 3</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>F 2 3</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>I 2 3</td>
</tr>
</tbody>
</table>

X-ray crystal diffraction

Bragg's law: $2d\sin \theta = \lambda$
X-ray crystal diffraction

Bragg's law
\[ 2d \sin \theta = \lambda \]

Data collection strategies

**in-house**
(5-20 mins per image)

1. Bravais lattice estimation from only 1-2 images with iMOSFLM (Andrew Leslie, LMB)
2. Mosaicity (Fabrice’s talk) estimation (also depending on beam divergence and resolution)
3. Radiation damage is not an issue

**at synchrotron**
(0.08-0.1 secs per image)

1. Lattice knowledge is no longer critical due to speed (unless crystals are very sensitive to radiation damage)
2. Mosaicity knowledge not critical due to fine \( \varphi \) slicing (0.1-0.2°)
3. Radiation damage always to be checked (on-the-fly tools and/or sacrificial crystal)
Fourier transform of electron density of crystal unit cell

\[ p(x,y,z) = \frac{1}{V} \sum_{h=\text{min}}^{\text{max}} \sum_{k=\text{min}}^{\text{max}} \sum_{l=\text{min}}^{\text{max}} |F_{hkl}| \cos \left( 2\pi(hx + ky + lz) - \varphi_{hkl} \right) \]

Rupp (2003)

Data processing of diffraction images

Spot finding

0°

360°

Image header info
- Detector distance
- Beam centre
- Wavelength (\(\lambda\))
Data processing of diffraction images

Spot finding

Image header info
- Detector distance
- Beam centre
- Wavelength ($\lambda$)

Index
- Unit cell
- Bravais lattice (cubic, trigonal, monoclinic, ...)
- Crystal orientation

hkl → Spot prediction and integration

Scaling and merging

\[ \sqrt{hkl} = F_{hkl} \]

rs_mapper, Takanori Nakane (LMIB)

Finding the phases

\[ p(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hk0}| \cos \left( 2\pi(hx + ky + lz) - \varphi_{hk0} \right) \]

Molecular Replacement
- Experimentally the simplest method with a single native dataset required
- Search homolog not always available
- Ideally 40% or more sequence identity between search model and target is required

Single-wavelength anomalous diffraction (SAD)
- Anomalous scatterers are part of the protein (e.g., sulfur, zinc, iron, ...) or can be incorporated (e.g., SeMet)

- Multi-wavelength anomalous dispersion (MAD)
- Single isomorphous replacement with anomalous signal (SIRAS)
- Multiple isomorphous replacement (MIR) with anomalous signal (MIRAS)
- Produce better starting phases (i.e., better initial maps for building)
- Experiments are more difficult (either for data collection (MAD) or soaking the heavy atoms (SIRAS/MIRAS))
Model **refinement** to:
1) fit chemically sensible models into observed data (electron density maps)
2) calculate the best electron density to criticise the model

Software for refinement in crystallography:
- **Refmac5** in CCP4 (Garib Murshudov, LMB)
- Phenix
- BUSTER

**Model validation in MolProbity**

- Electron density, R value
- Resolution
- Ramachandran
- Flexibility
- Stereochemistry
- Steric clashes
- Overall quality
- Bond and angle quality
- Ramachandran plot

**Crystallography software**

- Data collected using the in-house X-ray generator are usually processed with MOSFLM through the iMOSFLM graphical user interface
- Synchrotron data at Diamond are automatically processed by Xia2 using three different software, DIALS, XDS and autoPROC/Starna
- Instructions on how to launch Xia2 and scripts/tutorials for iMOSFLM, DIALS and XDS can be found on the X-ray crystallography webpage
- For Molecular Replacement the following software is available via CCP4: Phaser and MolRep as stand alone programs and Balbes, MrBump and MoRDa as automatic pipelines
- For anomalous/heavy atom search is available ShelxD via CCP4 and HySS via Phenix (or automatic phasing pipeline such as autoSHARP and Crunk2 in CCP4 or Autosol in Phenix)
- Density modification programs include Parrots and ShelxE via CCP4 and Resolve in Phenix
- For model building is available Coot (Paul Emsley, LMB) and ChimeraX
- Refinement programs include Refmac5 (Garib Murshudov) via CCP4 and phenix.refine in Phenix
- Structure validation is generally assessed with MolProbity (Duke University)

All software necessary from data processing to phasing to structure refinement is available on the LMB linux servers (hex, hal and max)

For any problem or help with the software please do not hesitate to contact me.
Tutorials

Crystallography explanatory videos
Recorded movies of seminars on data processing given by people who wrote codes for data integration (Harry Powell and Andrew Lesley), experimental phasing (Andrea Thorn) and refinement (Garib Murshudov):

https://www2.mrc-lmb.cam.ac.uk/research/scientific-training/crystallography-course-2013/

Two-day data collection and processing training
Three times a year (usually Feb, Jul and Dec) a two-day users’ training is run at Diamond (one day to visit all the MX beamlines and the other day is a workshop on data processing)

Seven-day data collection and structure solution workshop
Diamond-CCP4 jointly run a seven-day workshop (Dec) ranging from data collection to data processing, analysis and structure solution

If you are interested in the training courses above and would like more information please do not hesitate to contact me

Summary and advices

Tutorials

Crystallography explanatory videos

Two-day data collection and processing training

Seven-day data collection and structure solution workshop

Summary and advices

Target

Sample prep. → Xtal. → Xtal hit optim. → Xtal Soak/cryo → Xtals harvest → In-house Screen → Synchrotron data collection → Data proc., Phasing, Validation

Biology

• Produce the most “crystallizable” construct/sample (skills, experience and creativity)

• Prepare two 10 mg/ml protein samples for crystallization screening:
  1. low salt (NaCl) concentration (50 mM)
  2. high salt (NaCl) concentration (500 mM)

• Set up 1 or 2 sparse-matrix screens with the above samples and check that about 50% of drops are clear before proceeding to use all 23 different available screens
  If that is not the case, especially with the low salt concentration samples, try adding 10% glycerol or other stabilizers until 50% of drops will be clear in this pre-screening

• If no crystals appear after 4-6 weeks in any of the 23 screens, dehydration of drops and/or in situ proteolysis are an option before moving on to different constructs

• Once crystal hits are obtained, talk to Fabrice for crystal optimisation

• Ligand/heavy atoms soaking and cryoprotection → talk to us

• Using in-house X-ray generator or synchrotrons → talk to me

• Data processing, phasing and other structure related issues → talk to me → LMB

Talk to us → LMB
**X-ray crystallography or single particle cryo-EM?**

<table>
<thead>
<tr>
<th>X-ray crystallography</th>
<th>Cryo-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>no size limit (though MWs higher than 300 kDa are rarely crystallized, unless they are really stable, e.g. viruses)</td>
<td>Lower limit currently of about 50 kDa</td>
</tr>
<tr>
<td>Quick output (1 hour to set up and few days to get an answer)</td>
<td>Lower throughput</td>
</tr>
<tr>
<td>Majority of useful crystals produce data at resolutions of about 1-3 Å</td>
<td>Majority of useful EM grids maps are &gt;3.5 Å</td>
</tr>
</tbody>
</table>

- Regions of interest in cryo-EM maps of large complexes that cannot be resolved properly due to flexibility can be investigated separately by X-ray crystallography for high resol. data
- X-ray crystallography as method of choice for drug discovery for high-throughput capabilities to screen a large amount of compounds in very short time

**Final remark**, if sample amount is not an issue, why not to spend 1 hour to set up crystallization screens before spending months on optimizing EM grids?

---

**Real space crystal diffraction**

\[ 2d \sin \theta = \lambda \]

- hkl (140) hkl (210)
- \( d = 10 \text{ Å}, \lambda = 1 \text{ Å} \Rightarrow \sin \theta = 1/20 \)
- \( \theta = \arcsin(0.05) = 2.9^\circ \)

- X-ray Beam
- F(140) F(210)
Reciprocal space crystal diffraction

Reciprocal lattice
\[ a^* = 1/a, \ b^* = 1/b \] and \[ c^* = 1/c \]

2D projection of the reciprocal lattice points lying on the Ewald sphere at a given orientation