Lecture 2: Growing, handling and analyzing protein crystals, post-mortem methods

Bernhard Rupp
Dept. of Forensic Crystallography
k.-k. Hofkristallamt
Vista, CA 92084, USA
Innsbruck, A 6020 Austria

• Benefits of robotics and automation
  - Handling crystals - harvest, cryo-cooling
  - Real crystals: diffraction - twinning - scanning

• Analysis of large crystallization data sets
• What can be predicted and what not
• Targeted protein modifications
• Post-mortem analysis and summary

www.ruppweb.org

From previous lecture:
Principles of crystallization

Irrespective of all the theory we have learned so far, efficiency is our main concern: How to set up a maximum of successful experiments with the least amount of material and cost?

Practical methods -> Statistics -> Analysis -> Prediction -> Modification -> Success!

Summary and questions
Parallel vs. serial approaches

Early stage failures are generally cheap. It is generally more efficient to pursue multiple target variants in parallel instead of waiting until failure and then repeating the procedure. Avoid focusing of the first (often marginal) success until ultimate failure.

Sobering statistics: crystallization success

http://targetdb.pdb.org/

<table>
<thead>
<tr>
<th>Statistic</th>
<th>PSI average (%)</th>
<th>PSI Best (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected targets to clones</td>
<td>49.0 (49.9)</td>
<td>95.6 (95.6)</td>
</tr>
<tr>
<td>Clones that express</td>
<td>67.8 (33.8)</td>
<td>88.4 (84.5)</td>
</tr>
<tr>
<td>Clones to purified protein</td>
<td>16.6 (5.6)</td>
<td>81.7 (69.0)</td>
</tr>
<tr>
<td>purified protein to crystal</td>
<td>37.3 (2.1)</td>
<td>81.9 (56.5)</td>
</tr>
<tr>
<td>purified protein to data</td>
<td>14.2 (0.8)</td>
<td>33.6 (19.0)</td>
</tr>
<tr>
<td>purified protein to structure</td>
<td>12.3 (0.7)</td>
<td>22.5 (12.7)</td>
</tr>
</tbody>
</table>

- Some caveats:
  - Reporting and recording metrics are inconsistent between facilities
  - ‘best’ in one class is not same than in other class
  - Metrics are harsh on centers pursuing high hanging fruit (eukaryotics)
What is protein crystallization and how-to?

Protein crystallization is a special case of phase separation forming a protein rich phase from thermodynamically metastable (supersaturated) solution under the control of kinetic parameters.

How to: crystallization techniques

Different crystallization methods traverse the crystallization phase space differently.
It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....

Automation in the small laboratory

It is important to establish a time line during observations – crystals appear and disappear again!! Schmutz is present immediately, real stuff later....
What do we actually see?

1 beautiful but does not diffract
2 not a protein crystal
3 anisotropic
4 too small
5 crystals and urchins
6 Ostwald ripening
7 useless dendrites?
8 grainy precipitate
9 slimy precipitate
10 nothing - clear

Polarization and UV imaging are helpful

Figure 2: Initial crystallization trials of afamin. The leftmost panels show spherulites developing from a protein-rich phase typical at the transition to crystalline material, used for preparation of microseeding stock. A second round of screening trials with matrix micro-seeding and seed beads yielded about 10 different conditions containing imperfect crystals, two of which are shown in the right side panels. The top panel shows polarized visible light images, while the bottom row shows the corresponding native UV images demonstrating that the crystals are indeed protein, consistent with the absence of any high salt concentrations in the mother liquor.
Figure 2: Initial crystallization trials of afamin. The leftmost panels show spherulites developing from a protein-rich phase typical at the transition to crystalline material, used for preparation of microseeding stock. A second round of screening trials with matrix micro-seeding and seeded beads yielded about 10 different conditions containing imperfect crystals, two of which are shown in the right side panels. The top panel shows polarized visible light images, while the bottom row shows the corresponding native UV images demonstrating that the crystals are indeed protein, consistent with the absence of any high salt concentrations in the mother liquor.
Crystal harvesting and mounting

The primary reason for cryocooling is the prevention of radiation damage during X-ray exposure.

Cryo-cooling reduces radiation damage

Crystals often need cryo-protection to prevent formation of surface ice during quenching or flash-cooling. PEGs, glycerol, salts, sucrose and others are effective cryo-protectants.
Why cryo-protection is necessary

- Density of water changes abruptly upon freezing.
- We want to reach cryogenic temperatures w/o formation of crystalline ice.

![](image)

- This fast cooling is called quenching, or flash-cooling, but it is not freezing. We do not freeze crystals, because formation of ice (=freezing) destroys them.

Fast quenching requires less cryo-protectant to reach vitreous state

- Hyper-quenching: using a stream of dry nitrogen to remove warm gas cushion.
Even harvesting can be automated

Simple changes can provide major advantages

Automated well opening:
- Hole must be only as big as work envelope requires
- Significant increase in open well time
- More crystals harvestable from same well
- Resealable if good crystals remain
- Offset hole does not obstruct crystal viewing
Precise micro-manipulations of crystals with the UMR

Micro-crystal harvesting:
- Machine vision determines proper loop size selection
- Precise movements below 1 μm step size (10 μm Xtals)
- Automated zoom adjustment
- Automated focusing
- Harvest move assist
- Automated cryo-cooling and quenching

Drip cryo-protecting with PFPE

- New drip technique: drop the protectant on crystal, no swishing.
- Improved performance with Lee-valve, full control over drop size and speed. Optimized for low viscosity perfluoropolyether to be applied to a crystal in the loop immediately after harvesting.
- Crystals stay almost always! Even small ones - probably a universal cooling method
In-situ RT-characterization

• Why harvest and mount in the first place?
• There are good reasons for and against it....

![Image of in-situ RT-characterization](image)

---

Home-lab RT-characterization

**Objective:** Inspect one and the same crystal at RT and collect LT data

• RT mounting and diffraction analysis - straight forward, for the robot - harder by hand

• Systematically evaluate cryo-protocols - new insight into cryo expected.

• Did you damage the crystal through mounting or not?
In-situ characterization

- Another way to get the drop into the beam:
- Cut the drop support (kapton) and transfer it into beam!

Data processing delivers an integrated, scaled, merged set of unique data
Inspection of the initial diffraction pattern: ice rings

Figure 8-25 Ice rings. Diffraction image (1° rotation) typical for a 3 x 3 module detector. The horizontal and vertical white lines separate the nine modules of the detector as shown in Figure 8-8, but the image is zoomed and the peripheral areas are not completely shown. When crystalline ice forms in the mother liquor or through frost deposition on the crystal, the ice crystallites are generally randomly orientated and generate concentric diffraction rings typical for a powder diffraction pattern. The lattice spacing (d) for the ice rings is provided in the corresponding labels. The regions affected by the ice rings can be excluded from data processing with not too much detriment. Ice rings can also be used to verify the location of the primary beam position. Image and ice rings simulated using MUPSIM® by James Holton, UCSF.

Ice rings are a powder diffraction pattern of polycrystalline ice, most of the time frost on the sample. Severe ice formation in mother liquor surrounding the crystal almost always destroys it.

Inspection of the initial diffraction pattern: exposure time

Figure 8-26 Effects of exposure time on diffraction image quality. The same image, exposed for successive times (30 ms, 1 s, 64 ms) giving about eight times improvement in S/N between images. For each panel, the insert shows a magnification of the lower right quadrant. The first image (A) is too noisy, and we are not fully exploiting the diffraction limit of this well diffracting crystal. The second image (B) is just right; we obtain good resolution and only very few reflections are slightly saturated (yellow spot centers in insert). The third image (C) is heavily overexposed, practically all low resolution reflections are saturated (yellow spots in insert), while we are not gaining much more in terms of ultimate resolution. Images simulated using MUPSIM®.

Overexposure leading to detector saturation and thus loss of even a small percentage of high intensity reflections almost invariably leads to problems in phase determination - both experimental phasing and molecular replacement.
Inspection of the initial diffraction pattern: reflection overlap

Moving the detector back or selecting longer wavelength can improve spot resolution

Protein crystals are not perfect inside

Phenomena of mosaicity and twinning complicate data collection - even different parts may diffract differently!
Inspection of the initial diffraction pattern: high mosaicity

Note: Only large mosaicity can be detected by visual inspection from a single frame snapshot, but becomes obvious during reflection integration.

Crystals can be scanned - quality and radiation damage

Malcolm Capel and Raj Rajashankar, NE-CAT
Continuous Raster Scanning

Real crystals: Mosaicity and non-merohedral twinning

Real crystals are imperfect
Merohedral (hemihedral) twinning

Not all space groups allow for twinning, but because perfect merohedral or hemihedral twinning cannot be visually recognized in a diffraction pattern, we need twinning tests. Tests for twinning are based on deviations from expected diffraction intensity distributions (Wilson statistics).

Tests for twinning: N(H) and N(|L|)

Both centric and acentric intensity distributions show deviations from Wilson expectations in presence of twinning. Acentric (more reflection) are generally more indicative. Raw moments can also be used (Table 8-3).
Crystallization depends on **LOCAL** properties of protein — thus **GLOBAL** descriptors (MW, pI, etc) are of **LIMITED** **SPECIFIC** predictive power (same problem as R-value and local quality estimate)

**Desired:** make **specific** prediction for a given protein using all available prior information. Fat chance.

**BUT:** overall probabilities and propensities can still be determined!

**Associated problems:** experimental design bias, neglect of negatives in data collection, non-overlapping basis sets, hidden parameters (factors)

---

**Crystallization as a sampling problem**

Nearly infinite combinations of reagents and limited amount of material

**Multivariate** (multi-dimensional), sparsely populated sampling space

**We want to**
- Reduce dimensionality (use reagents that work)
- Truncate space (eliminate a priori improbably experiments)
- Sample the remaining space as efficiently as possible
- Becomes a probabilistic sampling problem with or without prior knowledge — **Bayesian Approach**
Basic mapping of crystallization space provides simple unconditional crystallization probabilities for reagents and proteins.

**Figure 3.34** Visualization of simple 3-dimensional crystallization space. The figure illustrates the varying coverage of the crystallization parameter space by different sampling protocols using 12 trials each. The large cube represents 4-fold oversampling in repeated use of the same sparse matrix-type experiment. Grid screening is not normally used to comprehensively screen for initial conditions, but deployed with a rationale to explore systematic variations of the two dimensions considered to be major factors, while keeping other parameters constant.

**Problem: sampling of clustered data**

- **Random sample**
- **Footprint screen**
- **Grid screen**
- **4x oversampled sparse screen**
Random Combinatorial Screening

Random screening is the most efficient approach when prior information is absent and clever human process analysis and manipulation can not be implemented.

- Highest likelihood of success particularly for rare events
- Defined decision point when to terminate screening
- Intended towards the application of rigorous and unbiased statistics - unbiased population of data space - necessary for quantification of results
- Fully customizable and adaptable to inclusion of prior knowledge
- Automatic setup of optimization protocols

Efficiency = Throughput*Success_Rate/Cost
(not everything that is smart is worth pursuing)

Protein crystallization propensity (from literature)

2/3 of all proteins fall into 'poor' crystallizer category
1/3 fall into 'reasonable' Crystallizer category > 0.2%
Cumulative probabilities determine effort and endpoint

Even a poor crystallizing protein (1/500) still has a 50% chance to show a crystal after 300 (288) trials.

Results from real HTPX analysis: Easy proteins come in multiple forms

Again: ~ 60% of proteins below average propensity.
Results from SG initiatives


Crystallization propensities: reagents

PEGs at neutral pH are a good Start

Probabilities are still conditioned on basis set

Proteins that like to crystallize also give better crystals - crystallizability is an inherent protein property
Optimization

Once initial conditions are determined, we have knowledge (conditions) about crystallization probabilities - conditional probabilities prob(C|I, Chem, Meth)

- Grid expansion screening
- Additive screening
- Partners, cofactors
- Context
- Drop ratio/size variation
- T-variation

- Rational optimization (response surface, ANOVA, factorials)

Interesting question:

- If individual local distribution of contacts determines crystal formation

- How then - and to what degree - can it be predicted in absence of structural knowledge?
**Crystallization prediction conditioned on protein properties**

- **Basis for crystallization predictors such as XtalPred.**
  
  [http://ffas.burnham.org/XtalPred-cgi/xtal.pl](http://ffas.burnham.org/XtalPred-cgi/xtal.pl)

- **Problem:** Prior probability of difficulty only, need to try actual screening anyhow

---

**Local modifications of proteins - SER**

- **Surface Entropy Reduction**
  
  Idea: reduce $T\Delta S$ entropy term in free energy of crystallization

  \[
  \Delta G = \Delta H - T(\Delta S_{prot} + \Delta S_{solv})
  \]

  by replacing Lys with Ala, Thr, Val

---

Chemical modification by surface Lys methylation

Lysine methylation - LYS to MLY - is largely used as a salvage strategy and also reduces solubility and changes charge distribution - multiple effects and actual causality for success is hard to establish.

Screening for protein stability

ThermoFlour assays allow rapid assessment of the stability of the protein. Scattering methods and NMR can assess conformational state and homogeneity, and CD spectroscopy estimates of secondary structure.
Light scattering and second virial coefficient

Static light scattering only measures the protein-protein interactions – not a clear predictor for crystallization success

![Diagram of light scattering and second virial coefficient](image)

Figure 4-23 Determination of molecular weight and second osmotic virial coefficient from static light scattering. The left panel represents light scattering measurements of a given protein in four different cocktails as a function of the protein concentration and the corresponding extrapolation of Equation 4-3 to $c \to 0$. The red data points are obtained in the case of repulsive interaction (positive $B_{22}$) and the open symbols in the case of a near ideal system ($B_{22} = 0$). None of these correlations can ever lead to crystallization. The filled black symbols indicate strong attractive interactions (quite negative $B_{22}$), the protein precipitates at low concentrations beyond which no data can be measured. The green points represent a case of modeled attractive interactions conducive to (but not necessarily sufficient for) crystallization. The right panel shows experimental $B_{22}$ values for various proteins observed in the “crystallization window.” Figures based on graphs in the review by Bill Wilson.

Light scattering and conformational purity

![Diagram of light scattering and conformational purity](image)

Figure 4-24 Analysis of size distribution of samples by dynamic light scattering. A desirable outcome of a DLS experiment is shown in the left panel: the size distribution is narrow, and the sample is monodisperse. Notwithstanding other hindrances, the sample is likely to crystallize. The analysis of the scattering data shown in the right panel is less promising: The sample is polydisperse and in addition the two species have a rather large broad size distribution. The probability for successful crystallization is rather low (but crystallization is not impossible). Such size distribution histograms or similar representations are typically provided by DLS instruments.

Dynamic (quasielastic) light scattering measures the hydrodynamic radius and thus aggregate size as well as size distribution. Narrow and unimodal size distribution is a good sign, but no guarantee for success. Precipitants are also chaotropes!
Sec and MALS sizing inform about conformational purity

**Constant MW**

**Decreasing MW**
Sec and MALS sizing inform about MW and conformational purity

- 67 kDa
- 120 kDa

Small angle solution scattering

Small angle solution scattering with X-rays or neutrons determines the low resolution shape function and thus also aggregate size. Useful for modeling known high resolution parts into unknown protein complexes.
2-D NMR spectroscopy - HSQC

2-D NMR HSQC spectra can provide a rapid assessment of the folding state of a protein. Useful as a diagnostic tool.

Circular Dichroism spectroscopy

Spectropolarimetry (CD spectroscopy) provides an assessment of the secondary structure content and thus whether secondary structure is present (folded) or not (unstructured).
A few points for review (I):

• Parallel approaches and miniaturization i.e. most trials (information) with least material - are the main benefits of automation

• In addition, take advantage of collecting comprehensive data for statistical analysis

• Use automation also in difficult steps like harvesting, mounting, cryo-protection, or in-situ techniques

• Take advantage of the fact that many crystals can be automatically screened at modern facilities – crystals from same drop - or even individual crystals, at different points, can diffract strikingly different

A few points for review (I):

• Accept that your chance of obtaining diffracting crystals of your protein without any additional procedural adjustments (e.g. optimization) or protein modifications is at best 10-20%.

• Accept the probabilistic nature of the crystallization game. You can win only by increasing your odds, not by seeking certainty.

• In other words, do nothing stupid, but sample everything else efficiently - most info with least material - know your material!
A few points for review (III)

• Do not oversample – if no promising results are obtained after about initial 300 trials, it is likely that your protein construct is a hard case for crystallization. Consider other constructs, residue mutations or orthologs, the cellular context, etc.

• Try to gain a rapid assessment of your protein’s crystallization propensity by using a 2-tiered approach, starting with pH-PEG or index screens and expand the sample space in the next round.

• Use robotics for repetitive and tedious tasks and for miniaturization, but do not expect your robots to think for you - or to save you much time.

A few points for review (IV)

• The choice of screening kit or reagent set is probably the least significant - well crystallizing proteins generally crystallize under multiple conditions. Most kits share the same basic reagents. The protein (even batch!!) is the major determinant!

• Do not believe any tips or claims that lack a clear rationale. Causality rules, even for statistically infrequent events.
A few points for review (V):

- Understand that crystals are real and not ideal objects - be aware of twinning, large mosaicity, local(!) diffraction properties, etc. and corresponding remedy.

- All this and more can be found in my book (best prices on Amazon):

  www.ruppweb.org

Summary in most simple words (suitable response to impatient and nagging supervisors):

“Then, I thought, Hey, hold on a minute—maybe failure is an option.”