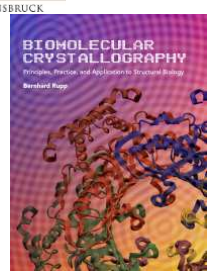
Lecture 2: Growing, handling and analyzing protein crystals, post-mortem methods



www.ruppweb.org

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

From previous lecture: Principles of crystallization

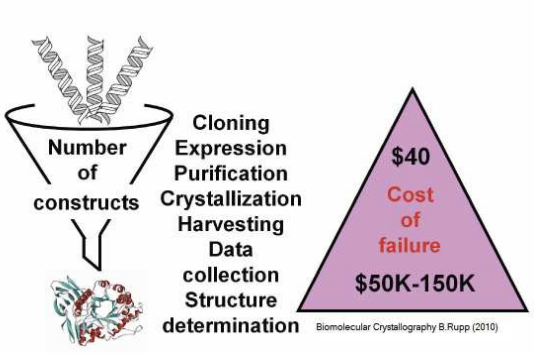
Irrespective of all the theory we have learned so far, efficiency is our main concern: How to set up a maximum of successful experiments with the least amount of material and cost?

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

Summary and questions

Parallel vs. serial approaches

Figure 1-5 Keeping the funnel full. Given the high attrition rate in early experimental steps of a protein structure determination project, a parallel approach pursuing multiple variants of the protein target significantly increases the chances of success. Parallel approaches are also less susceptible to late stage failures, because the pressure to pursue one single protein construct that already shows warning signs to the bitter end is diminished. The most expensive failure is a perfectly good diffraction data set that cannot be phased, because all the costly and time-consuming laboratory experiments have been conducted already. Developing a suitable phasing strategy early on is thus part of good study planning.



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/

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

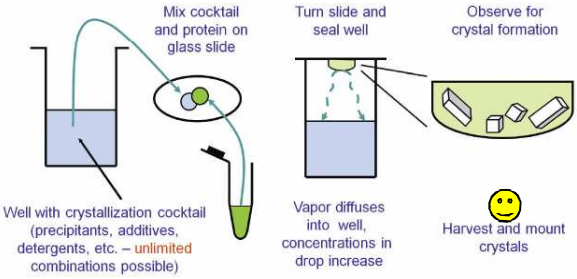
- Some caveats :
- reporting and recording metrics are inconsistent between facilities
- 'best' in one class is not same than in other class
- metrics are harsh on centers pursuing high hanging fruit (eukaryotics)

What is protein crystallization and how-to?

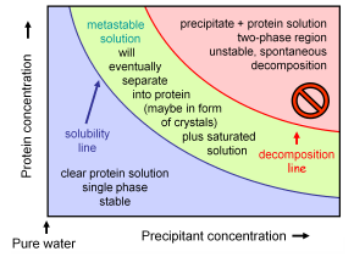
Protein crystallization is a special case of phase separation forming a protein rich phase from thermodynamically metastable (supersaturated) solution under the control of kinetic parameters

Figure 3-1 Basic hanging-drop vapor diffusion. Hanging-drop vapor diffusion has been in use for over 30 years for the manual setup of protein crystallization. The reservoir (generally one well of a multi-well assay plate) is partially filled with several hundred µl of crystallization cocktail. A small drop (a few µl or less) of this cocktail is set in the center of a siliconized cover slide, and mixed there with an equal volume of protein stock solution (green). The cover slide is then turned over and placed on the greased rim of the reservoir well. The mixing with protein has reduced the precipitant cocktail concentration to half of the original value, and the sealed system thus equilibrates by water vapor diffusion from the drop into the reservoir solution, thus effectively increasing the concentration of all constituents (protein and precipitation cocktail reagents) in the crystallization drop. During this process the drop becomes supersaturated, nucleation can occur, and protein crystals may grow from the supersaturated solution.

Figure 3-2 Crystallization plates. A 24-well Linbro plate used for a manual hanging-drop crystallization setup is shown to the left. Originally designed for cell culture work, the Linbro plates made from polystyrene are cheap and commercially available with pre-greased rims. The wells are covered with siliconized 22 mm circular cover slides. Although not suitable for automated work, the hanging-drop method using Linbro plates is still useful in small laboratories. The crystallization plate to the right is a typical 96-well sitting-drop plate in the smaller, standardized high-throughput SBS (Society for Biomolecular Sciences) format (Intelligate, Art Robbins Instruments). Crystallization robotics most frequently use sitting-drop plates with a SBS footprint.



How to: crystallization techniques



Different crystallization methods traverse the crystallization phase space differently

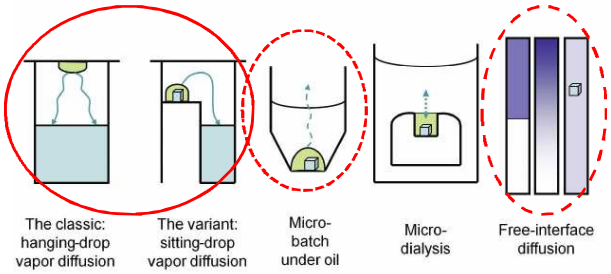


Figure 3-19 Schematic sketches of popular crystallization techniques. Hanging-drop vapor diffusion is a common method used in small-scale manual setups while sitting-drop vapor diffusion is preferred with robotic setups. The absence of additional sealing requirements and the ease of miniaturization favors automated microbatch screening under oil, although harvesting tends to be more difficult. Use of silicone oils in microbatch wells allows partial exchange of solvent vapor (indicated by the broken arrow). Microdialysis is hard to miniaturize but can be used to grow very large crystals. Miniaturized free-interface diffusion screening chips are gaining popularity, but automation and harvesting issues remain to be resolved. Each method traverses the crystallization phase space in a different path and the same chemical screening conditions can produce widely varying results.

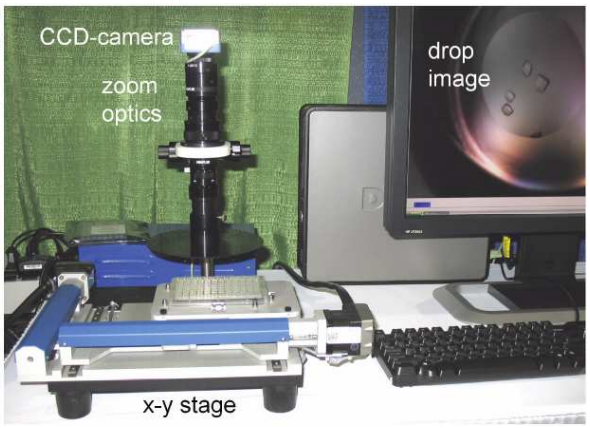
Automation in the small laboratory

Sidebar 3-13 Automated crystallization setup for the small laboratory. Based on the assumption of modest throughput requirements, and no necessity for full walk-away automation, two low-budget approaches to automation are conceivable: selection of a single system that can prepare crystallization cocktails (perhaps in a limited fashion) and also set up the crystallization plates;[83] or a dual-station layout using separate cocktail preparation with a generic liquid-handling system followed by a dedicated plate-setup robot.[84] The major reason for separating plate setup from cocktail production is differing requirements for dispensing precision, volume, and speed. Fast, small volume (µl to nl), and very accurate (also in geometric terms) dispensing is mandatory for plate crystallization setup, whereas large volume (ml) handling with modest speed and precision requirements suffices for cocktail production. Another advantage of the separation between the cocktail stage and the plate setup is that simple one-to-one dispensing into reservoir wells and drop aliquots followed by protein addition with a single needle dispenser suffices (Figure 3-33) once the cocktails are produced in a 96-well format deep-well block. Deep-well blocks prefilled with crystallization cocktails are also commercially available. In addition, compared with a single-stage setup, failure of one system component does not affect the other. For example, cocktail production can continue while the plate setup robot is inoperative. Figure 3-33 shows a popular robot for 96-well crystallization plate setup.

Figure 3-33 A robot for automated crystallization plate setup. The Phoenix robot (Art Robbins Instruments) can set up 96 crystallization trials in about one minute. On the left side, a 96-channel syringe dispenser re-arrays (100 µl each) 96 prefabricated or purchased crystallization cocktails simultaneously from a standard deep-well block into the reservoirs of an SBS-format, 96-well sitting-drop crystallization plate, and places between 1 µl and 100 nl into the drop shelves or wells. From the right side, a contact-less microvalve dispenser nozzle immediately adds the pre-aspirated protein (stock vials in the red block) rapidly and without contact onto each of the precipitant drops. To minimize evaporation, the plate is then immediately sealed with a sheet of pressure-sensitive adhesive. Taking all losses into account, about 12 to 15 µl of protein stock is required for 96 100 + 100 nl drops. The robot design has been based on a prototype developed in an academic laboratory setting.[84]

Automation in the small laboratory

Figure 3-36 A low-cost automated crystallization plate imaging station. The crystallization plate is positioned by an x-y translation stage, and a digital zoom camera takes high-resolution images of the crystallization drops. The images taken in about 2 minutes can then be manually inspected on a computer screen, or processed by automated image recognition software. The depicted instrument is the CrysCam microscope manufactured by Art Robbins Instruments.

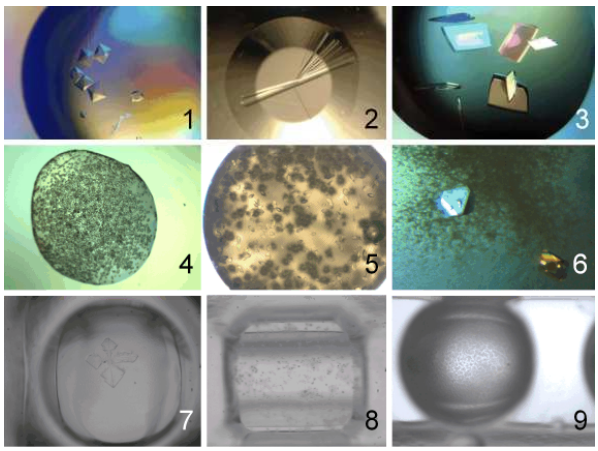


Art Robbins Instruments CrysCam

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

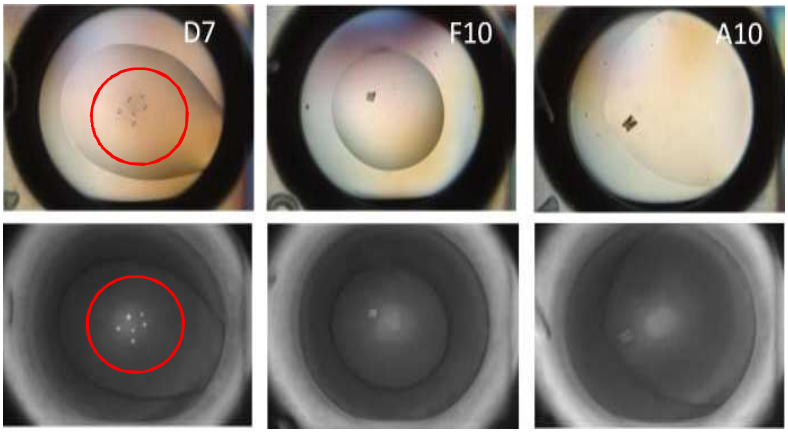


Figure 2: Initial crystallization trials of afamin. The leftmost panels show spherulites developing from a protein-rich phase typical at the transition to crystalline material, used for preparation of microseeding stock. A second round of screening trials with matrix micro-seeding and seed beads yielded 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.

Polarization and UV imaging are helpful

Figure 2: loping from a protein-ri croseeding 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.

Automation in the big laboratory

Crystal harvesting and mounting

Figure 8-12 Cryoloops. Cryomounting loops and meshes micro-fabricated from laser-etched Kapton[20] are reproducibly manufactured, sturdy but flexible, and a significant improvement over the floppy nylon or hair loops used previously.

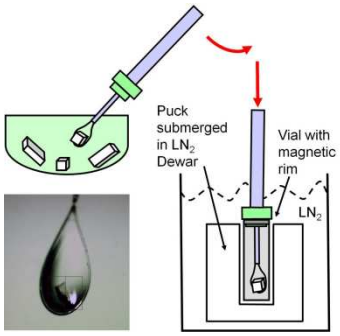


Figure 8-13 Harvesting and flash-cooling process. The crystal is scooped up from the drop in a mounting loop on a pin, and dipped quickly into liquid nitrogen. A vial with a magnetic rim placed in a submerged storage puck protects the mounting pin. The insert shows a harvested crystal in an old-style hair (cat-whisker) loop.

Figure 8-16 Robotic cryoquenching of a mounted crystal. The robot arm carries the crystal on a cryo-pin with a magnetic base and is about to dip it into the storage puck submerged in liquid nitrogen Note how the liquid nitrogen is boiling, and that the crystal will travel through a cushion of cold nitrogen gas first before it is quenched into the liquid nitrogen itself. Blowing the gas cushion away with dry nitrogen drastically increases the quenching rates.[37] This robot is also a prototype capable of semi-automatically harvesting crystals, and autonomous crystal harvesting is a remaining frontier in full automation of high-throughput crystallography.[38]

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

Cryo-cooling reduces radiation damage

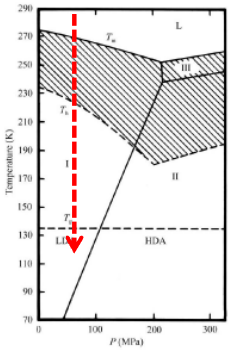
Figure 8-14 Various cryotools. The picture shows a variety of common tools used during cryomounting and cryostorage of protein crystals. The blue vessel is an unbreakable foam Dewar made by SpearLabs from closed cell foam. On top of it rests a magnetic wand used to hold and manipulate the mounting pins. A Hampton base with a copper pin is shown; stainless steel pins are more frequently used. The storage cylinder to the left of the foam vessel is an SSRL design. It fits the bore of a shipping Dewar and the pins can be auto-mounted by the SSRL beam line robotics. Below the cylinder is a simple aluminum cane that holds a series of cryo-vials with a magnetic rim securing and protecting the mounting pins. Several of these canes can be stored in a shipping or storage Dewar. At the bottom of the image is a cryo-tong that is pre-cooled under LN_2 and used to transfer pins from storage or quenching vessels onto the diffractometer. Tools and samples are kept at cryogenic temperatures at all times during mounting.

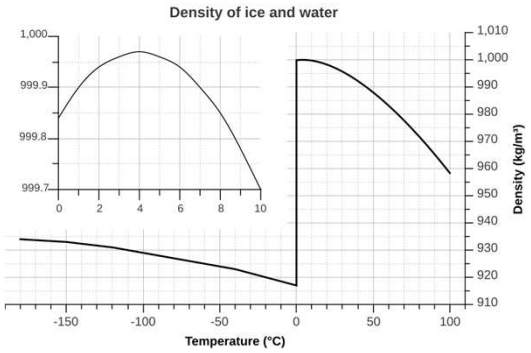
Figure 8-15 Radiation damage. The image shows massive radiation damage (black area) in a crystal of protein Atu1728 from *Agrobacterium tumefaciens* caused by intense X-ray exposure. The crystal, 350 × 350 µm, was exposed 2 min per frame for total of 180 exposures (6 h) on a Rigaku FR-E X-ray generator (one of the most powerful rotating anode laboratory X-ray generators available) equipped with VariMax HR optics and a 0.1 mm collimator. Image kindly provided by Aiping Dong and Xiaohui Xu, University of Toronto.

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

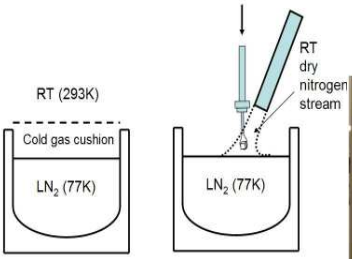




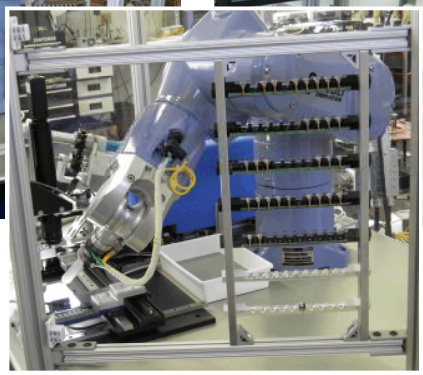
- This fast cooling is called quenching, or flash-cooling, but it is not freezing. We do not freeze crystals, because formation of ice (=freezing) destroys them.

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

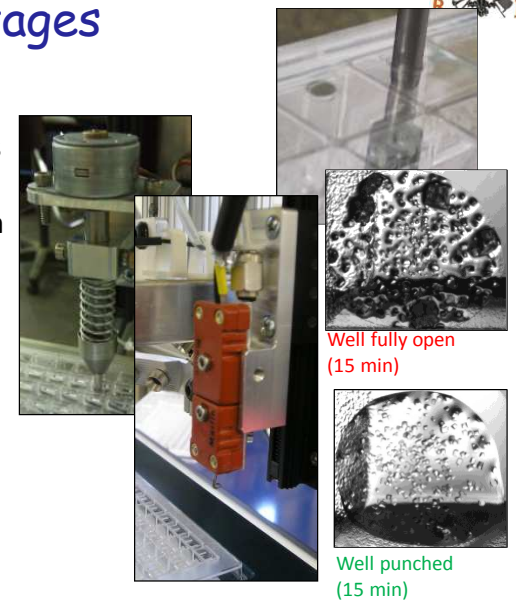


Square One
Systems Design

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



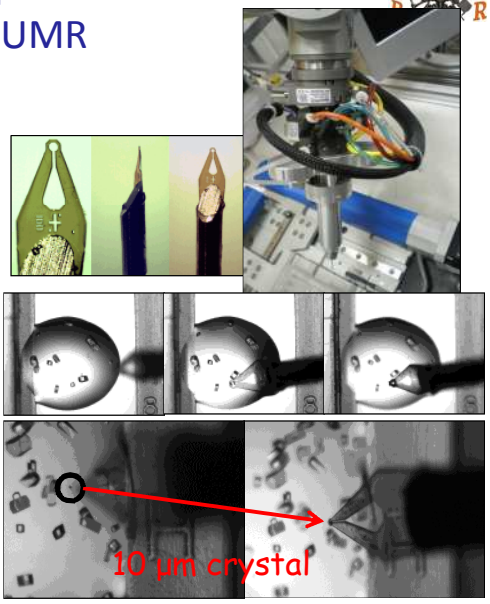
Well fully open (15 min)

Well punched (15 min)

Precise micro-manipulations of crystals with the UMR

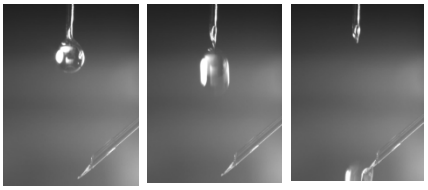
Micro-crystal harvesting:

- Machine vision determines proper loop size selection
- Precise movements below 1 um step size (10 um 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….

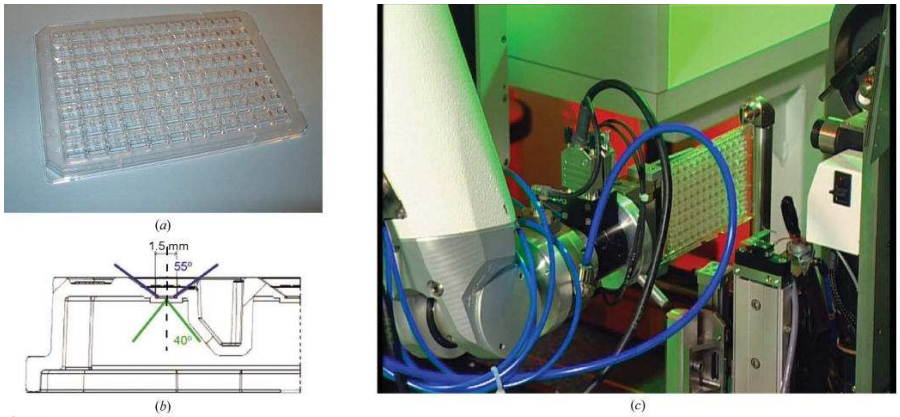
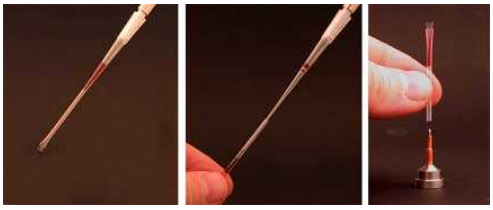


Figure 1
(*a*) View of the new 96-well CrystalQuick X plates (Greiner Bio-One) used for crystallogenesis and *in situ* diffraction. (*b*) Detailed view of the geometry of the plate well. (*c*) View of the plate hold by the CATS/G-Rob system in the X-ray beam during data collection.

Home-lab RT-characterization

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

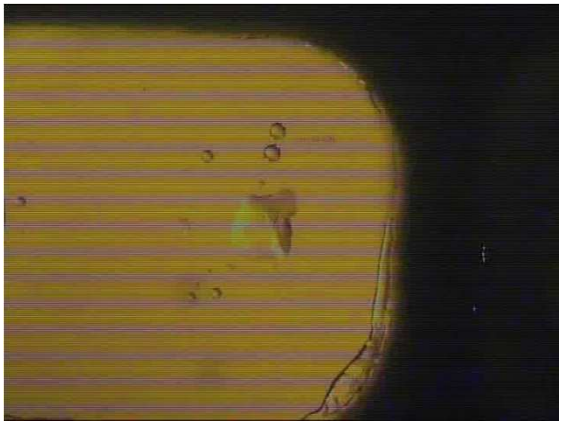
- RT mounting and diffraction analysis – straight forward, for the robot – harder by hand
- Systematically evaluate cryo-protocols - new insight into cryo expected.



- 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!



CrystalDirect: A new method for automated crystal harvesting based on laser-induced photoablation of thin films
Florent Cipriani, Martin R¨ower, Christophe Landret, Ulrich Zander, Franck Felisaz and Jose Antonio Marquez* Acta Cryst D 2010

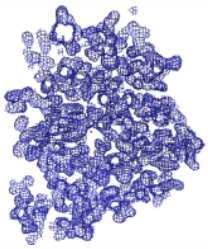
Data processing delivers an integrated, scaled, merged set of unique data

Figure 8-24 Crystal inspection and decision making during data collection. The first inspections relying primarily on visual appearance of the pattern are shaded in pink. If deemed satisfactory, the first images are then indexed and the mosaicity is estimated. Once the indexing succeeds, a strategy is developed based on the initial diffraction symmetry and data covering at least the asymmetric unit of the reciprocal space (plus the Friedel Bijvoet opposites in the case of anomalous data) are collected. A final risk factor is radiation damage of the crystal; it will manifest itself in diminishing diffraction over time.

h	k	l	F(hkl)	φ(hkl)
2	0	0	228.0	180.0
1	0	1	10.4	90.0
2	0	1	901.8	270.0
1	1	1	367.0	332.1
1	2	3	149.3	37.8
8	9	1	97.9	255.1
7	7	2	111.5	139.7

FT^{-1}

FT



Inspection of the initial diffraction pattern: ice rings

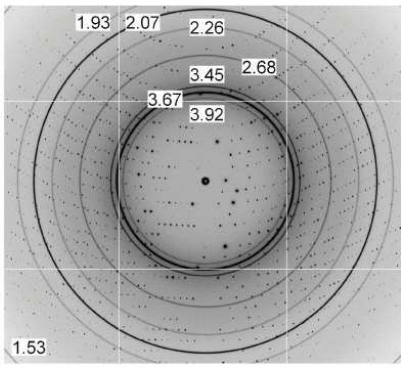


Figure 8-25 Ice rings. Diffraction image (1° rotation) typical for a 3 × 3 module detector. The horizontal and vertical white lines separate the nine modules of the detector as shown in Figure 8-8, but the image is zoomed and the peripheral areas are not completely shown. When crystalline ice forms in 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 (Å) 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 MLFSOM[53] 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

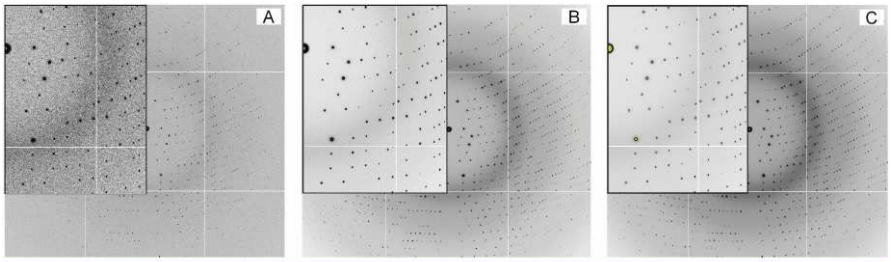
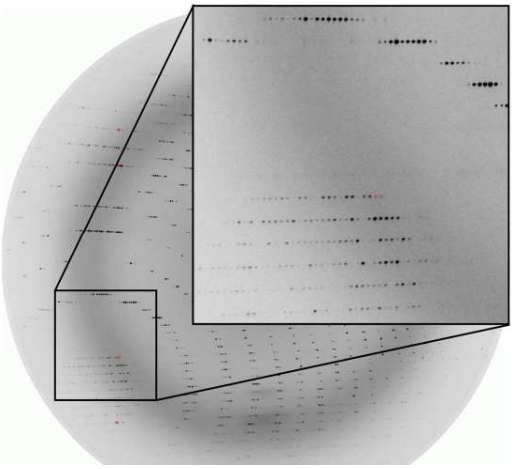


Figure 8-26 Effects of exposure time on diffraction image quality. The same image, exposed for successive times (30 ms, 1 s, 64 s) giving about eight times improvement in S/N between images. For each panel, the insert shows a magnification of the lower right quadrant. The first image (A) is too noisy, 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 MLFSOM.[53]

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

Figure 8-27 Diffraction image of crystal with 300 Å unit cell. The crystals of ^{SS}Rv1347c, a putative antibiotic resistance protein[55] from *M. tuberculosis*, belong to orthorhombic space group $P2_12_12_1$ with cell parameters a = 75.6 Å, b = 77.4 Å, and c = 297.6 Å, with eight molecules of 210 residues each per asymmetric unit. The data were recorded on a MAR345 image plate detector which was positioned as close as possible while still capturing all reflections up to the resolution limit of 2.2 Å. The insert shows that this is about the closest crystal–detector distance where the closely spaced reflections along **c*** are still resolved. Image files courtesy Clyde Smith and Graham Card, U. of Auckland, NZ, and SSRL, Stanford, CA.



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

Protein crystals are not perfect inside

Figure 3-11 Atomic force microscope images of crystal growth. (Panel A) The atomic force microscope images of the 001 surface of glucose isomerase show the two most common growth patterns observed in crystal growth: step growth starting from 2-dimensional nucleation islands (A, left image) and a double-spiral growth pattern (A, right image). Panel B shows formation of supercritical 2-dimensional nuclei on the 001 surface of cytomegalovirus (CMV), a member of the herpes virus family. As indicated by the arrows, in this case only two virions (B, left image) suffice to generate a critical nucleus from which new step growth commences (B, right image). Images courtesy of Alexander McPherson and Aaron Greenwood, University of California, Irvine.

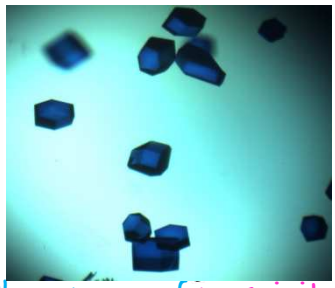
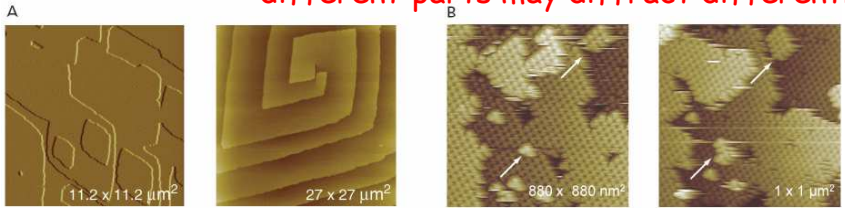


Figure 3-12 Growth of a real mosaic crystal. The schematic drawing shows a crystal growing in a solution of protein molecules (blue spheres). Small impurities (red) and some larger detritus (green squares) are also present in the solution. New molecules attach preferentially to steps and edges (red arrows) and we can recognize a growth defect in the form of a hole; impurities are enclosed at the domain boundaries; and a larger piece of detritus is incorporated at a domain boundary. Individual domains can be substantially misaligned, in this case about 6°; such a highly mosaic crystal would not be useful for diffraction experiments.

Phenomena of mosaicity and twinning complicate data collection – even different parts may diffract differently!



Inspection of the initial diffraction pattern: high mosaicity

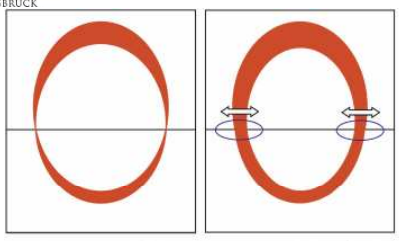
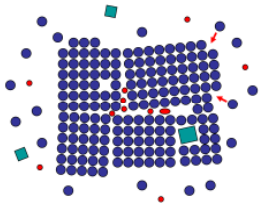
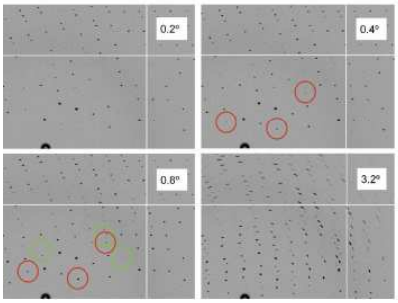


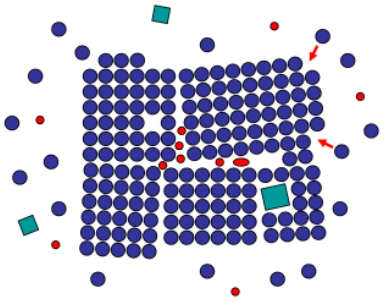
Figure 8-28 **Manifestation of mosaicity.** Estimating mosaicity from a single diffraction pattern is not trivial. In cases where the lunes are approximately centered around the rotation axis (top panel), the lunes widen along the axis direction. In a general orientation (bottom panels, same 1° rotation image with increasing mosaicity) the lunes are not as distinct and we only see an increasing number of partial reflections (extending over more than one image, red and green circles) but no distinct or tell-tale elongation of the spot shapes. Once the mosaicity becomes excessive and exceeds the rotation range, we observe the streaking as sketched in Figure 8-17. Images simulated using MLFSOM[23] courtesy of James Holton, UCSF.





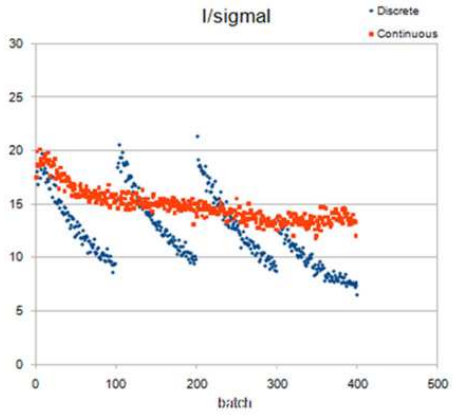
Note: Only large mosaicity can be detected by visual inspection from a single frame snapshot, but becomes obvious during reflection integration.

Crystals can be scanned - quality and radiation damage



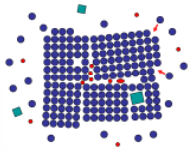
Malcolm Capel and Raj Rajashankar, NE-CAT

Continuous Raster Scanning

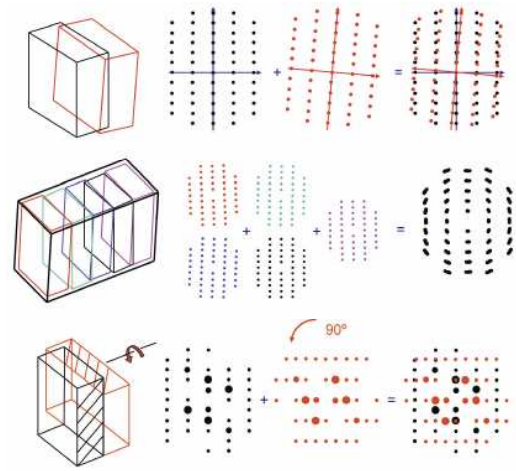


Malcolm Capel and Raj Rajashankar, NE-CAT

Real crystals: Mosaicity and non-merohedral twinning



Real crystals are imperfect



Biomolecular Crystallography B.Rupp (2010)

Figure 8-17 Macroscopic twinning, mosaicity, and non-merohedral twinning. The top panel shows macroscopic twinning of a crystal. The two growth domains are distinctly different as are their diffraction patterns. It may be possible to dissect this crystal, to expose only one domain, or (sometimes) to index and integrate one of the patterns if it is dominant. The center panel shows a mosaic crystal, with a relatively large (6°) mosaic spread. The diffraction pattern shows a corresponding mosaic spread of the spots. The bottom panel shows non-merohedral twinning, where two orientations are epitaxially related. The diffraction pattern is an interpenetration of the two patterns which can, in some cases but not always, be separately indexed.

Merohedral (hemihedral) twinning

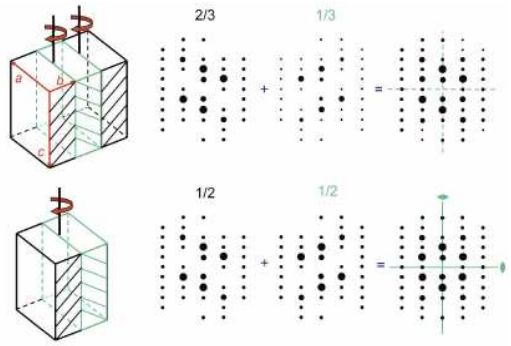


Figure 8-18 Hemihedral twinning. The superposition of two twin-related diffraction patterns in partial hemihedral twinning (merohedral twinning with a single twin operator) with a twin fraction $\alpha = 0.5$ generates a diffraction pattern that is more symmetric than the original pattern (top row). As a consequence, intensity statistics differ from expected values for Wilson statistics for untwinned structures and the distributions become sharper. For perfect hemihedry with $\alpha \neq 0.5$, the apparent symmetry is perfect and confers higher space group symmetry (lower row).

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|)

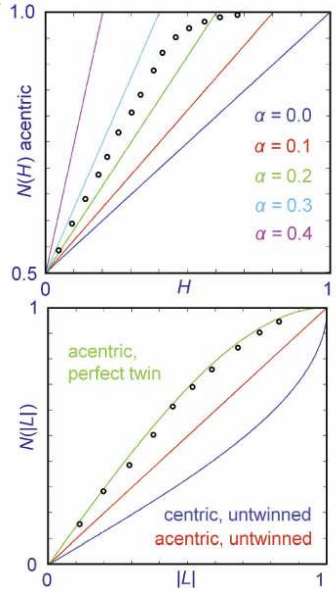


Figure 8-19 Cumulative probability distribution *N*(*H*) for acentric reflections. The plot shows the theoretical curves for acentric data and the experimental data points. A twinning ratio α of about 0.26 can be interpolated from the graph.

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).

Figure 8-20 Yeates–Padilla plot for the cumulative probability distribution *N*(|*L*|). The graph shows the expected cumulative distribution curves for acentric and centric untwinned data and acentric experimental data (open circles) for a perfect twin. For partial twins, deviations from untwinned data will result in experimental data points located between the calculated curves.

Now that we have the data: Challenges in prediction of crystallization

Crystallization depends on LOCAL properties of protein – thus GLOBAL descriptors (MW, pI, etc) are of LIMITED SPECIFIC predicitve 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!

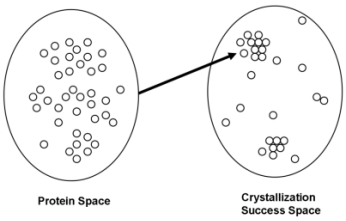
Associated problems : experimental design bias, neglect of negatives in data collection, non-overlapping basis sets, hidden parameters (factors)

Crystallization as a sampling problem

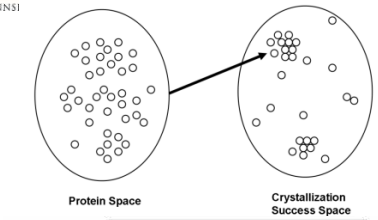
Nearly infinite combinations of reagents and limited amount of material

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



Crystallization as a sampling problem



Basic mapping of crystallization space provides simple unconditional crystallization probabilities for reagents and proteins.

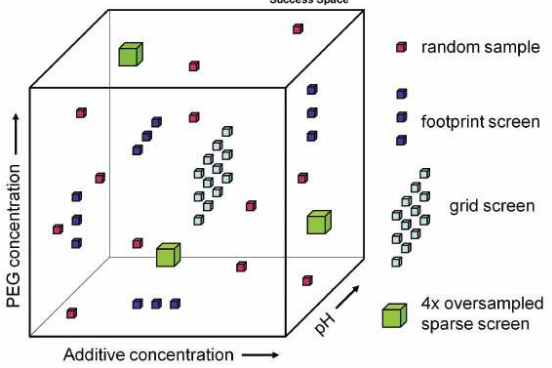
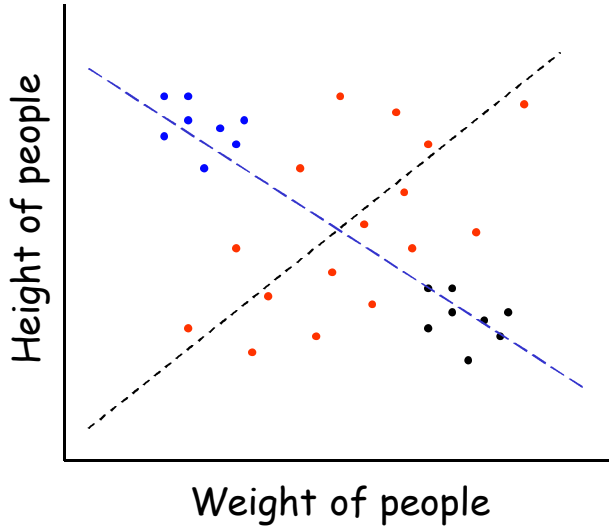


Figure 3-34 Visualization of simple 3-dimensional crystallization space. The figure illustrates the varying coverage of the crystallization parameter space by different sampling protocols using 12 trials each. The large cube represents 4-fold oversampling in repeated use of the same sparse matrix-type 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.

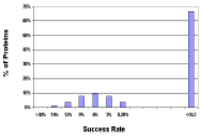
Problem: sampling of clustered data

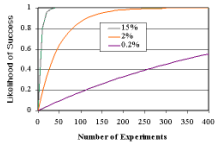


Random Combinatorial Screening

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

- Solid theoretical foundation (Segelke, 2001, J.Cryst.Growth)
- 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

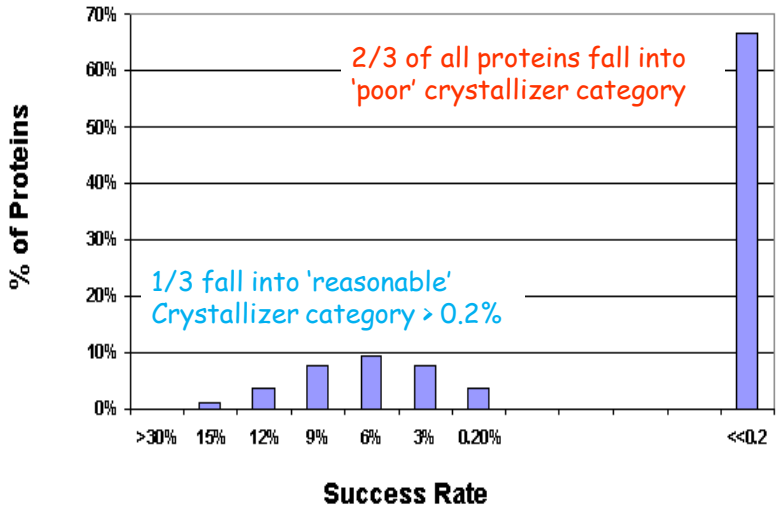




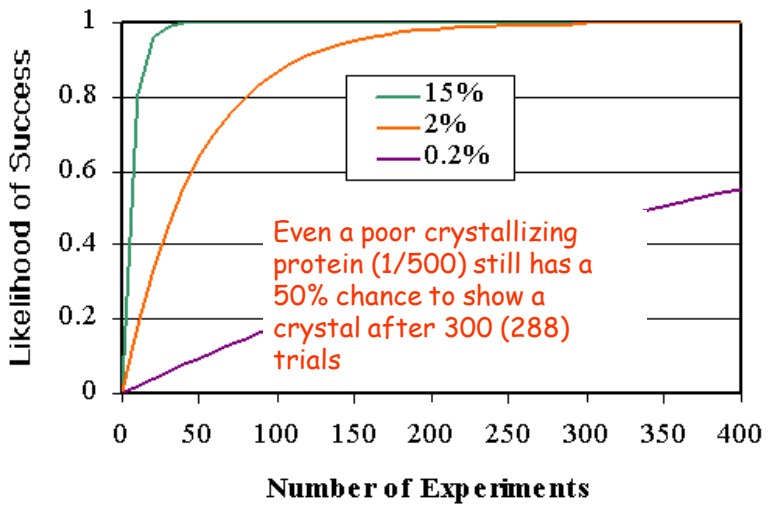
Efficiency = Throughput*Success_Rate/Cost

(not everything that is smart is worth pursuing)

Protein crystallization propensity (from literature)

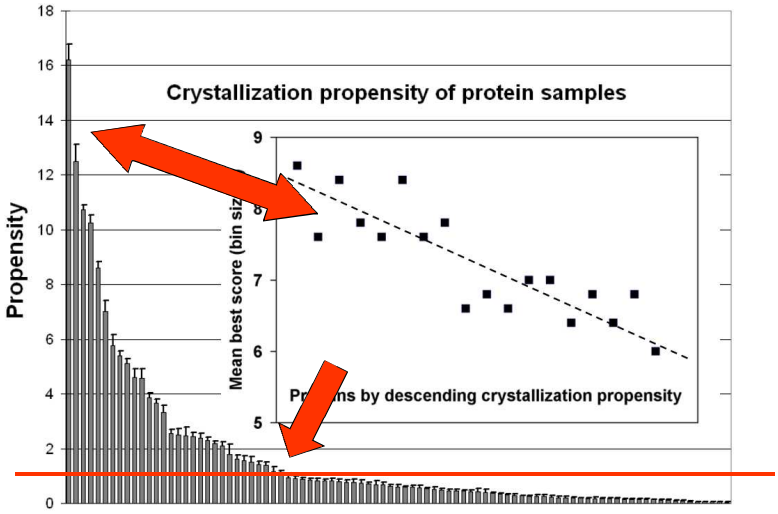


Cumulative probabilities determine effort and endpoint



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

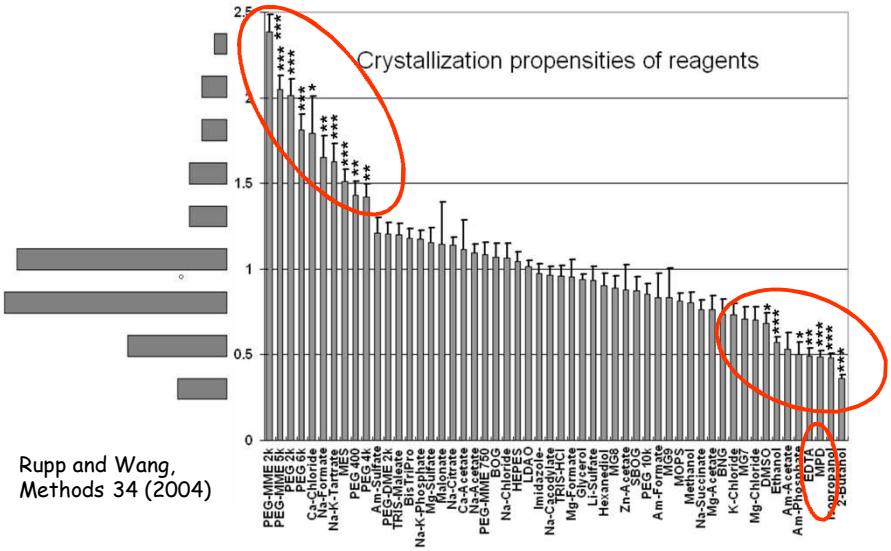
Results from real HTPX analysis: Easy proteins come in multiple forms



Again: ~ 60% of proteins below average propensity

MEDIZINISCHE UNIVERSITÄT

Results from SG initiatives

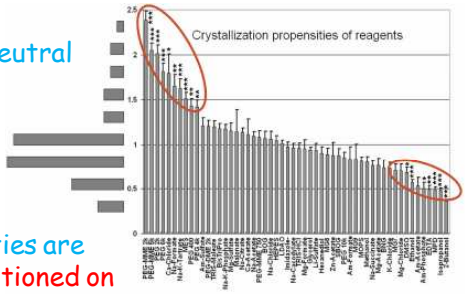


Rupp and Wang, Methods 34 (2004)

MEDIZINISCHE UNIVERSITÄT INNSBRUCK

Crystallization propensities: reagents

PEGs at neutral pH are a good Start



Probabilities are still conditioned on basis set

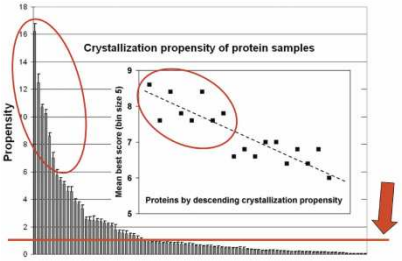


Figure 3-39 Crystallization propensity of proteins. General trends regarding unconditional crystallization probability have been derived by unbiased random sampling.[75] Proteins that crystallize well (quantified as propensity, a measure of how well a specific protein crystallizes compared with the average protein with a propensity of 1.0) also tend to crystallize with higher quality scores (circled in red) as shown in the insert. The compilation agrees well with the empirical estimate that only about one third of proteins crystallize better than average (red line, propensity of 1) in first trials, but those which crystallize do quite well and under multiple conditions. The findings emphasize the rationale that it is more efficient to improve the inherent "crystallizability" of a protein than to engage in excessive further screening that will likely produce only few more, and also in all likelihood poor, crystals. Figure adapted from Rupp and Wang.[48]

Figure 3-35 Crystallization propensities for 50 common reagents. Propensities are normalized frequencies, where an average crystallizing agent has a propensity of 1.0, and one that crystallizes better than average has a propensity above one. The data, obtained from unbiased random sampling of over 100 proteins, confirm the well-established fact that polyethylene glycols (PEGs) and their methyl ethers as a class are significantly better "crystallizers" than, for example, small molecule alcohols. The trends are statistically significant as indicated by the number of stars, but although over half a million experiments have been analyzed, shortcomings remain: only 50 reagents have been analyzed, synergistic effects are not considered, and the protein set is biased toward soluble, bacterially expressed prokaryotic proteins. Redrawn from Rupp and Wang.[48]

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

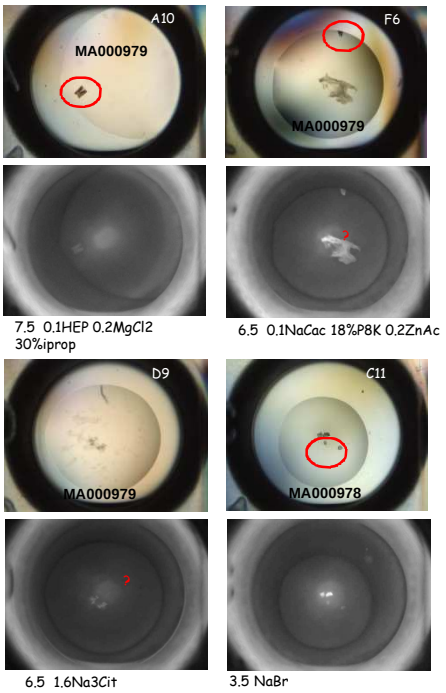
MEDIZINISCHE UNIVERSITÄT INNSBRUCK

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)



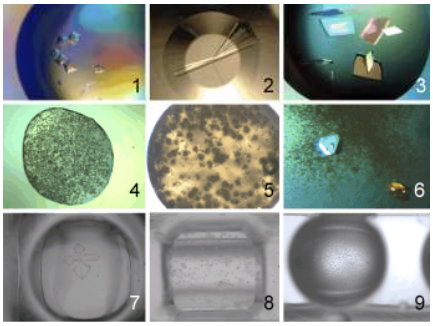
MEDIZINISCHE UNIVERSITÄT INNSBRUCK

Interesting question :

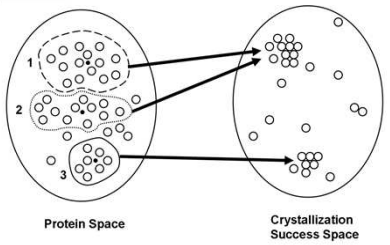
k&k Hofkristallamt B R

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

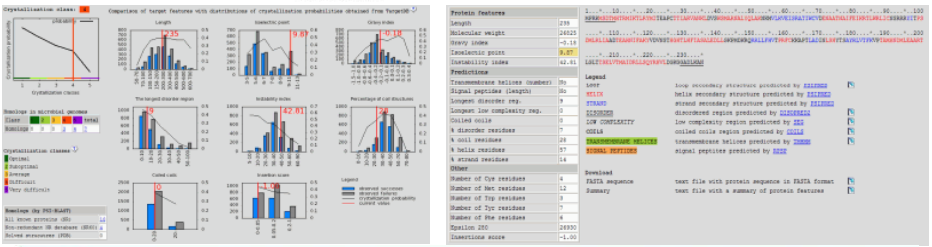


Figure 4-5 Prediction of crystallization success. The panels show the output of the XtalPred server[33] for an unknown (hypothetical) protein from *Pyrobaculum* spherical virus. The source of the poor score is not immediately obvious—the largest perceived risk factor might be the high isolectric point. Also note the contradicting predictions—the predicted N-terminal helical segment is also predicted to be disordered. One would have to conduct the actual experiment to determine with any certainty whether this protein indeed deserves a 4 (or D- grade) in Crystallization 101.

Local modifications of proteins - SER

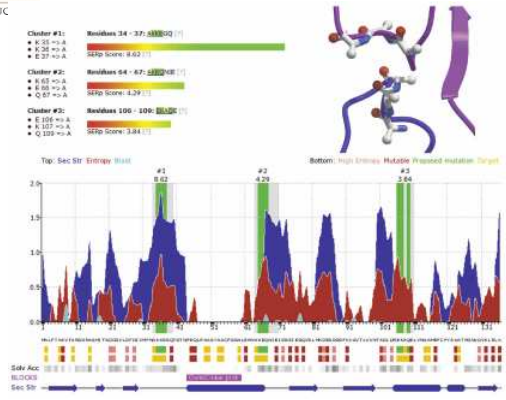


Figure 4-6 Selection of residues for surface entropy reduction. The analysis of the wild type (which could not be crystallized) of *Bacillus subtilis* organic hydroperoxide-resistance protein (OhrB) by the SER server at UCLA identifies a stretch of residues 34–37 -A(KKE)- as a cluster of high surface entropy. The inset top right shows the crystal packing in the 2.1 Å structure of the engineered triple mutant, -A(AAA)-. The arrangement of the molecules (blue) and the symmetry related mates (purple) clearly shows that this particular packing and crystal form would not be possible for the native structure given the space requirements of its extended lysine and glutamate residues. While the importance of modifying the protein is evident, it remains open which of the multiple molecular property changes that concur with the SER mutation actually are causal for successful crystallization. The details of the SER analysis and server output can be extracted from the web server instruction pages. PDB entry 2bjo.[28]

- Surface Entropy Reduction
  Idea: reduce TΔS entropy term in free energy of crystallization
  ΔG=ΔH-T(ΔSprot+ΔSsolv)
  by replacing Lys with Ala, Thr, Val

34. Goldschmidt L, Cooper DR, Derewenda ZS, et al. (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* 16(8), 1569-1576.

Chemical modification by surface Lys methylation

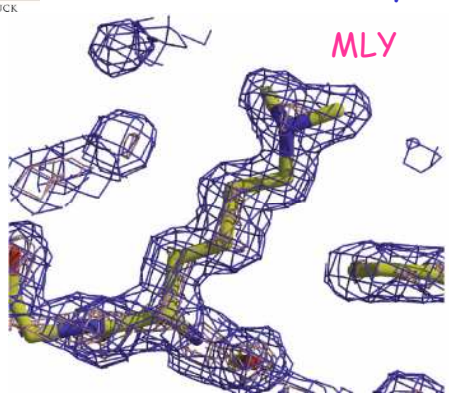
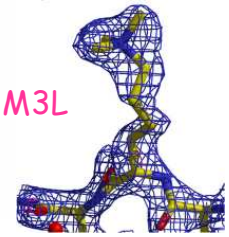


Figure 4-19 **Chemically modified lysine residue (dimethyllysine).** The electron density clearly shows that this surface residue of the alanine racemase of *Bacillus anthracis* is dimethylated (MLY). See Figure 2-27 for the density of a natively occurring, posttranslationally modified trimethyllysine (M3L) in calmodulin. PDB entry 2vd8.[107]



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

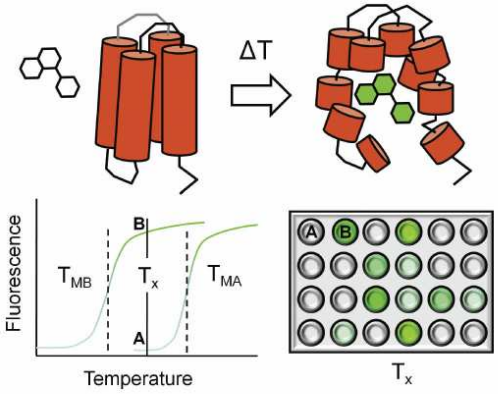


Figure 4-21 **Principle of the ThermoFluor stability assay.** Fluorescence of a hydrophobic fluoroprobe is quenched in an aqueous environment (indicated by the white molecule) but enhanced when it binds to exposed hydrophobic residues of an unfolded protein (green probe molecule). Increase in melting temperature measured by later onset of fluorescence thus indicates higher stability and additives in the protein solution that increase the protein's stability can be identified. The experiments can be elegantly carried out in 96-well format in an instrument combining a thermocycler and a fluorescence plate reader.

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

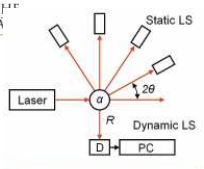


Figure 4-22 Principle of light scattering instruments. The top section of the figure illustrates a static multi-angle light scattering configuration (MALS) while the lower part shows a dynamic light scattering or photon correlation spectroscopy (DLS, PCS) setup. D indicates the detector, and PC the photon correlator (a fast electronic semiconductor circuit chip). Remaining symbols as used in Equation 4-2.

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

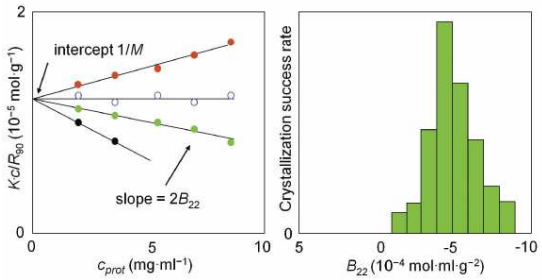
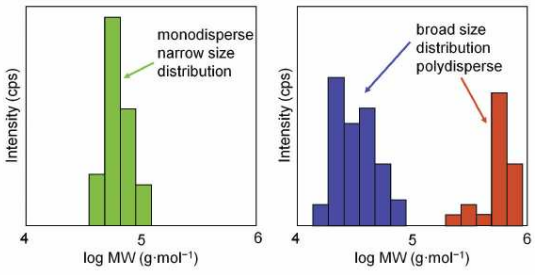


Figure 4-23 Determination of molecular weight and second osmotic virial coefficient from static light scattering. The left panel represents light scattering measurements of a given protein in four different cocktails as a function of the protein concentration and the corresponding extrapolation of Equation 4-3 to $c \rightarrow 0$. 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 ($B_{22} \sim 0$). None of these conditions 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 modest 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.[118]

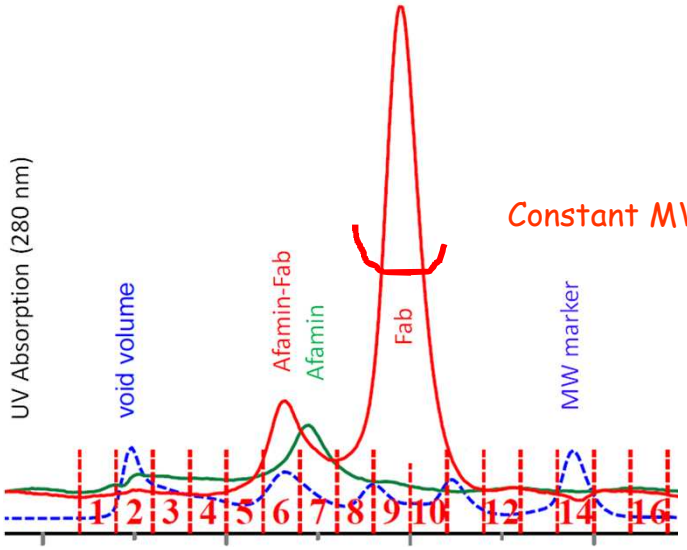
Light scattering and conformational purity

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 panels 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.

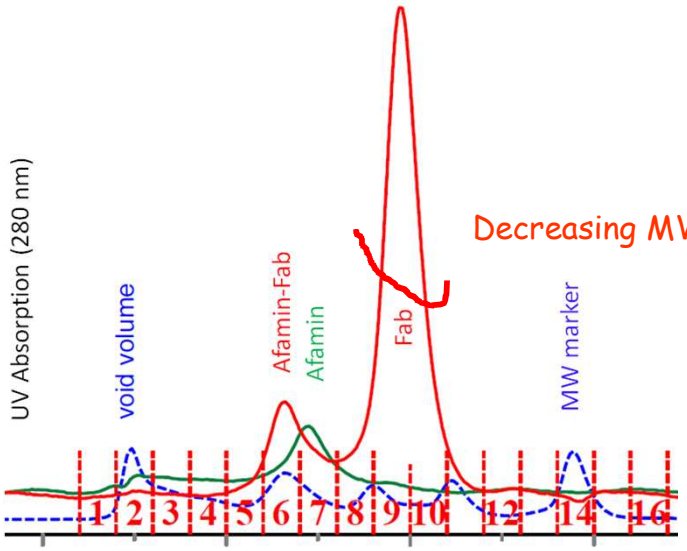


Dynamic (quasielastic) light scattering measures the hydrodynamic radius and thus aggregate size as well as size distribution. Narrow and unimodal size distribution is a good sign, but no guarantee for success. Precipitants are also chaotropes!

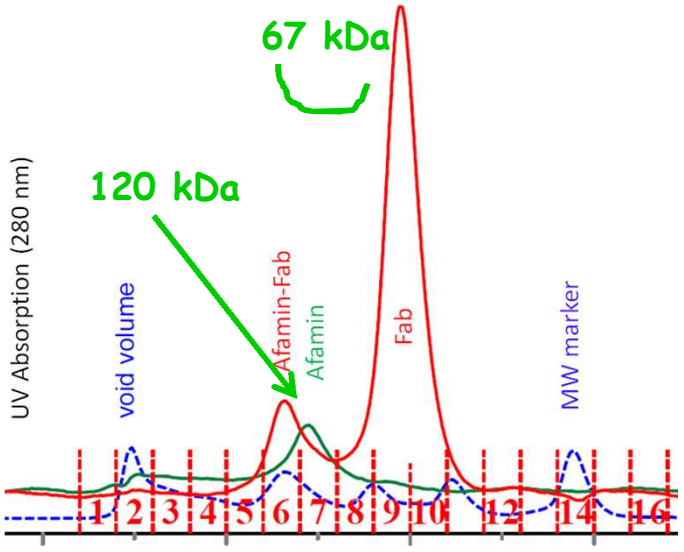
Sec and MALS sizing inform about conformational purity



Sec and MALS sizing inform about conformational purity

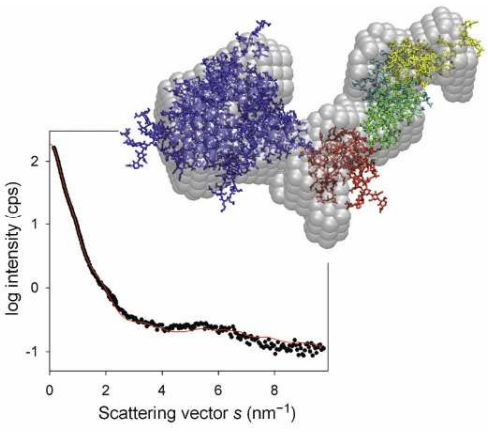


Sec and MALS sizing inform about MW and conformational purity



Small angle solution scattering

Figure 4-25 Analysis of small angle scattering. The linear part of the logarithmic plot of the relative scattering intensity (Guinier plot) provides an estimate for the overall hydrodynamic radius of gyration, while detailed analysis of the fall-off at higher scattering angles allows fitting of the curve with composite low-resolution models assembled from (partial) crystal structures. The corresponding resolution in Å can be readily calculated from $d = 2\pi/s$, that is, around 6.2 Å at the rightmost edge of the diagram (s = 10 nm^{-1}). The example shows the proposed model of the plasminogen-like overall solution structure of the multi-domain hepatocytic growth factor complex.[124] Figure courtesy of Dimitri Svergun, EMBL Outstation Hamburg.



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.

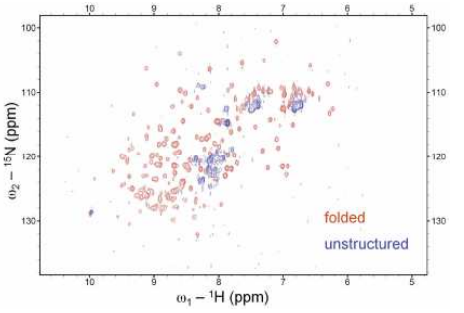


Figure 4-27 HSQC spectrum of folded and unstructured protein. The 2-dimensional ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) NMR spectrum clearly shows the distinct discrimination in the region below 8.3 ppm in ω_1 for the folded 228 residue protein (red, sharp peak contours) compared with the few wide and unresolved peaks for the unstructured protein sample (blue contours). NMR spectra courtesy of Simon Colebrook, Department of Biochemistry, Oxford University, and Joanne Nettleship, Oxford Protein Production facility.

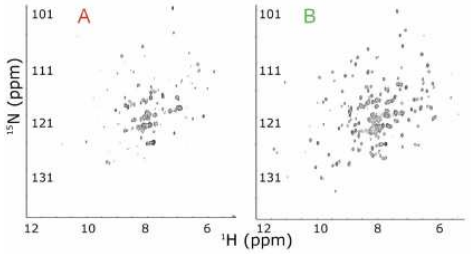
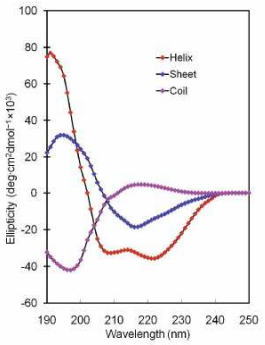


Figure 4-28 HSQC spectra of ligand-free and ligand-bound proteins. The 2-dimensional HSQC spectra of a bacterial methionine aminopeptidase (bMAP, 263 residues) without (A) and with (B) a tightly bound novel inhibitor.[131] Note the drastic increase in number of peaks and the improved discrimination of the spectrum for the bMAP–ligand complex compared with the *apo*-protein. The crystals of the bMAP–ligand complex diffracted to 0.9 Å resolution. Image courtesy of Artem Evdokimov, Procter & Gamble Pharmaceuticals, Mason, OH.

Circular Dichroism spectroscopy



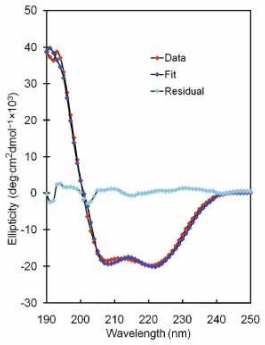


Figure 4-26 Reference CD spectra and fitted CD spectrum of myoglobin. The left panel shows the three basic reference spectra for helix, sheet, and random coil conformation as established for poly-L-lysine. The right panel shows an actual fit of a sperm whale myoglobin sample data against the standards (left) using the program CDFIT, predicting a total helix content of 80% and 20% random coil. The actual value is 77% total helix content. Note that much of the discriminating information is in the short wavelength far UV region sensitive to sample absorption, and that particularly for more complex proteins the fit results may vary depending on the extent of the fitted region.

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)

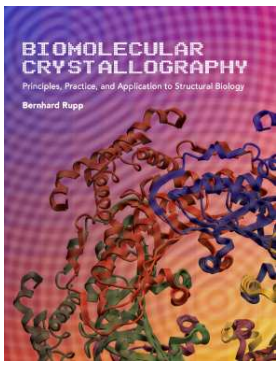
- The choice of screening kit or reagent set is probably the least significant – well crystallizing proteins generally crystallize under multiple conditions. Most kits share the same basic reagents. The protein (even batch!!) is the major determinant!
- Do not believe any tips or claims that lack a clear rationale. Causality rules, even for statistically infrequent events.

A few points for review (V):

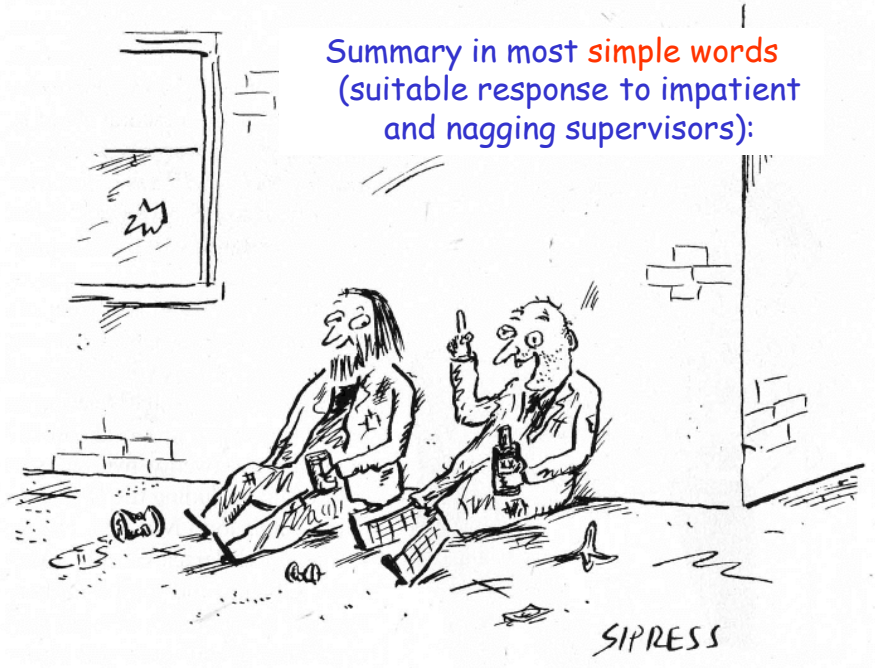
- Understand that crystals are real and not ideal objects- be aware of twinning, large mosaicity, local(!) diffraction properties, etc. and corresponding remedy.

- 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."