Crystallization

Edward H. Snell,

Hauptman Woodward Medical Research Institute, Buffalo, NY 14203, USA.

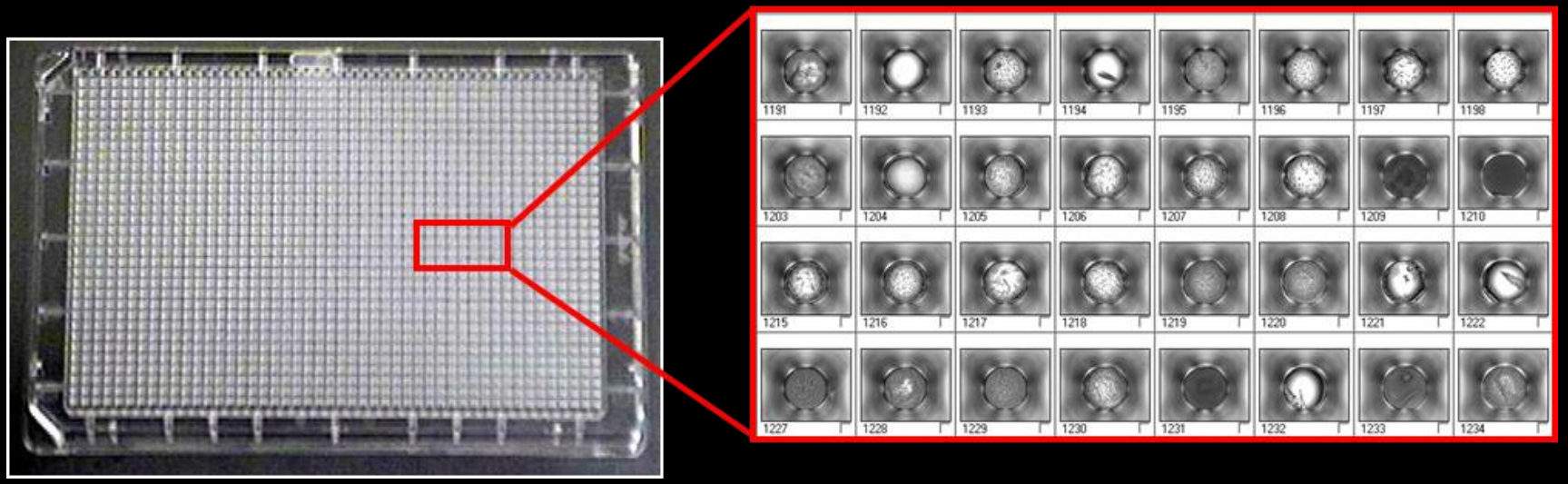
The Crystallization Screening laboratory at the Hauptman-Woodward Medical Research Institute

Since February of 2000 the High Throughput Search (HTS) laboratory has been screening potential crystallization conditions as a high-throughput service

The HTS lab screens samples against three types of cocktails:

1. Buffered salt solutions varying pH, anion and cation and salt concentrations
2. Buffered PEG and salt, varying pH, PEG molecular weight and concentration and anion and cation type
3. Almost the entire Hampton Research Screening catalog.

The HTSlab has investigated the crystallization properties of over 15,000 individual proteins archiving approximately 140 million images of crystallization experiments.

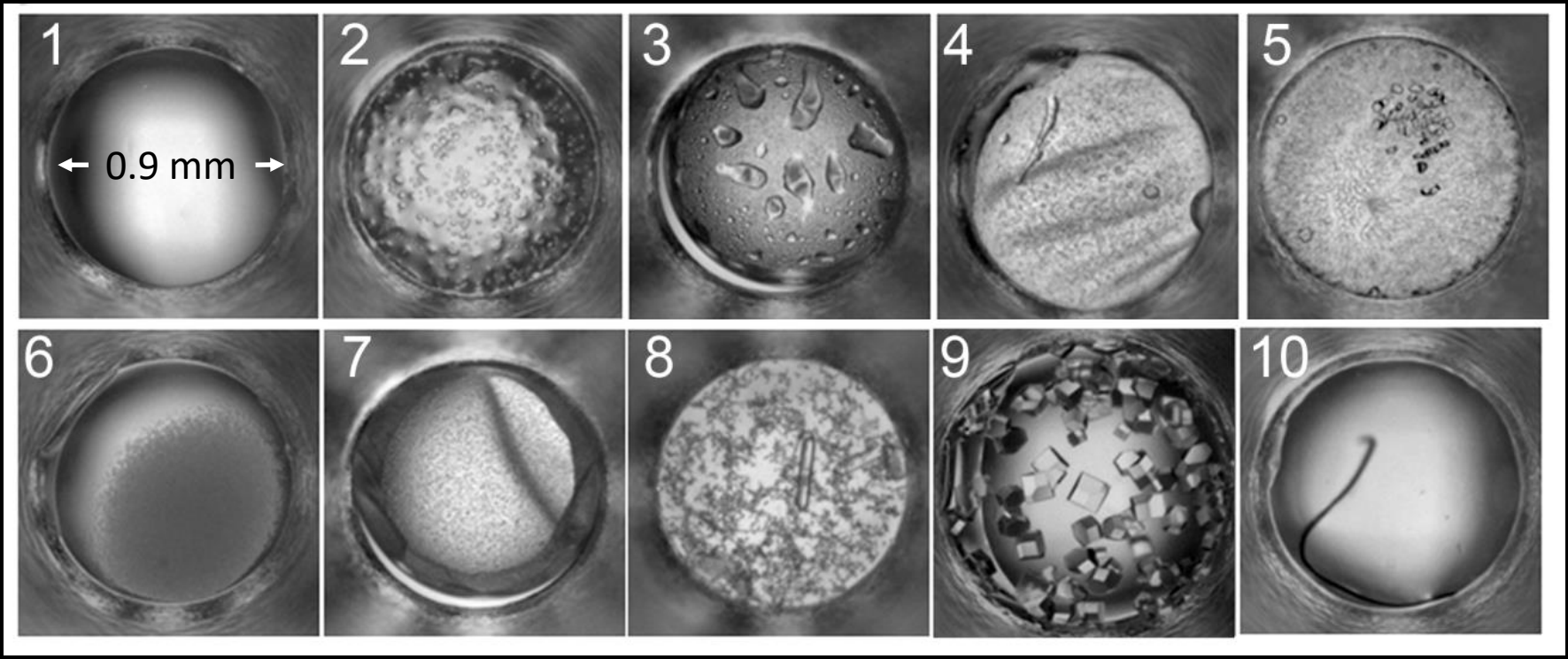


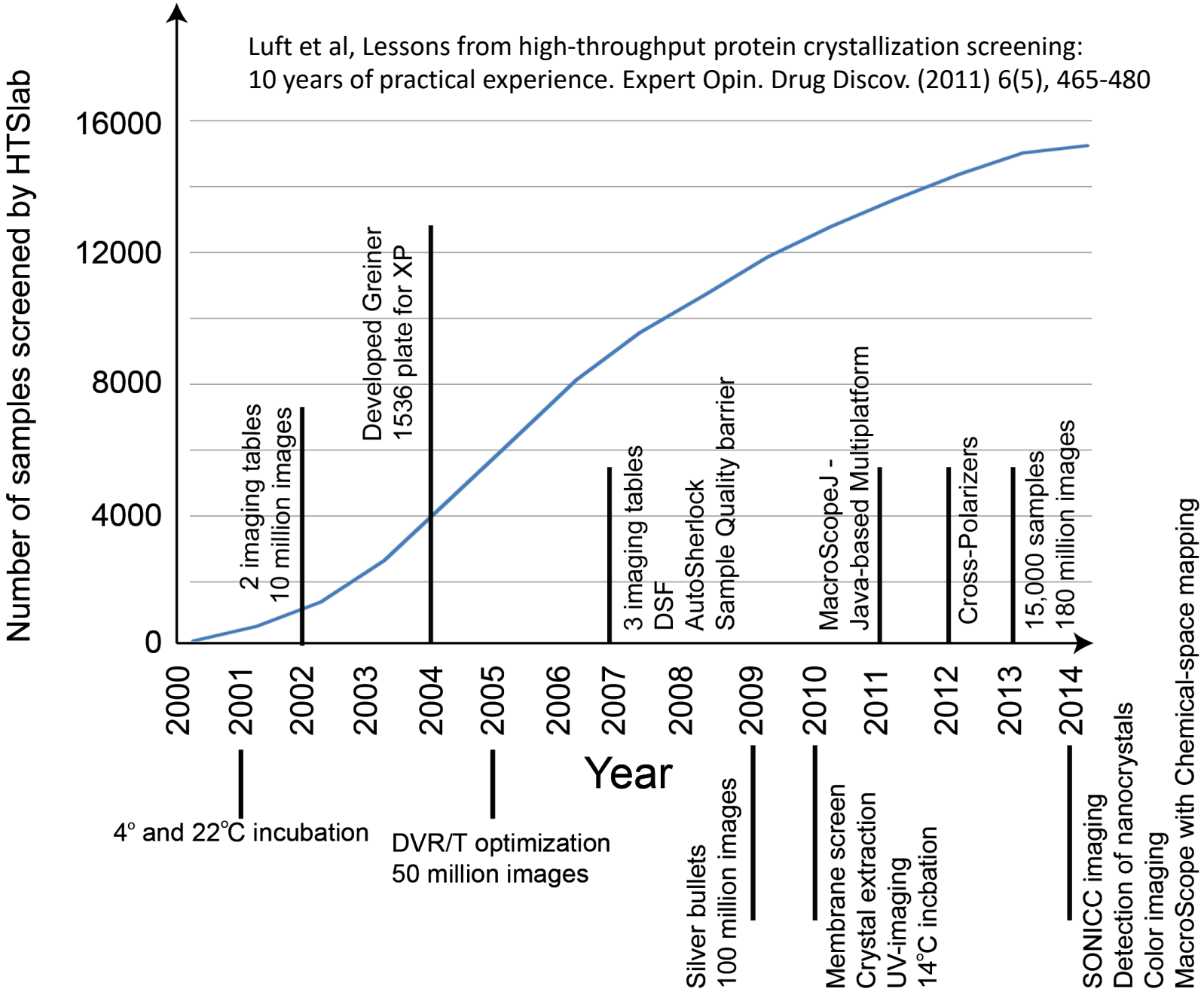
The crystallization method used is micro-batch under oil with 200 nl of protein solution being added to 200 nl of precipitant cocktail in each well of a 1536 well plate.

Wells are imaged before filling, immediately after filling then weekly for six weeks duration with images available immediately on a secure ftp server.

Several software utilities for viewing and analyzing data are available.

Outcomes




Luft et al, Lessons from high-throughput protein crystallization screening:
10 years of practical experience. Expert Opin. Drug Discov. (2011) 6(5), 465-480
Number of samples screened by HTSlab
16000
12000
8000
4000
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Year
2 imaging tables
10 million images
Developed Greiner
1536 plate for XP
3 imaging tables
DSF
AutoSherlock
Sample Quality barrier
MacroScopeJ -
Java-based Multiplatform
Cross-Polarizers
15,000 samples
180 million images
4° and 22°C incubation
DVR/T optimization
50 million images
Silver bullets
100 million images
Membrane screen
Crystal extraction
UV-imaging
14°C incbation
SONICC imaging
Detection of nanocrystals
Color imaging
MacroScope with Chemical-space mapping

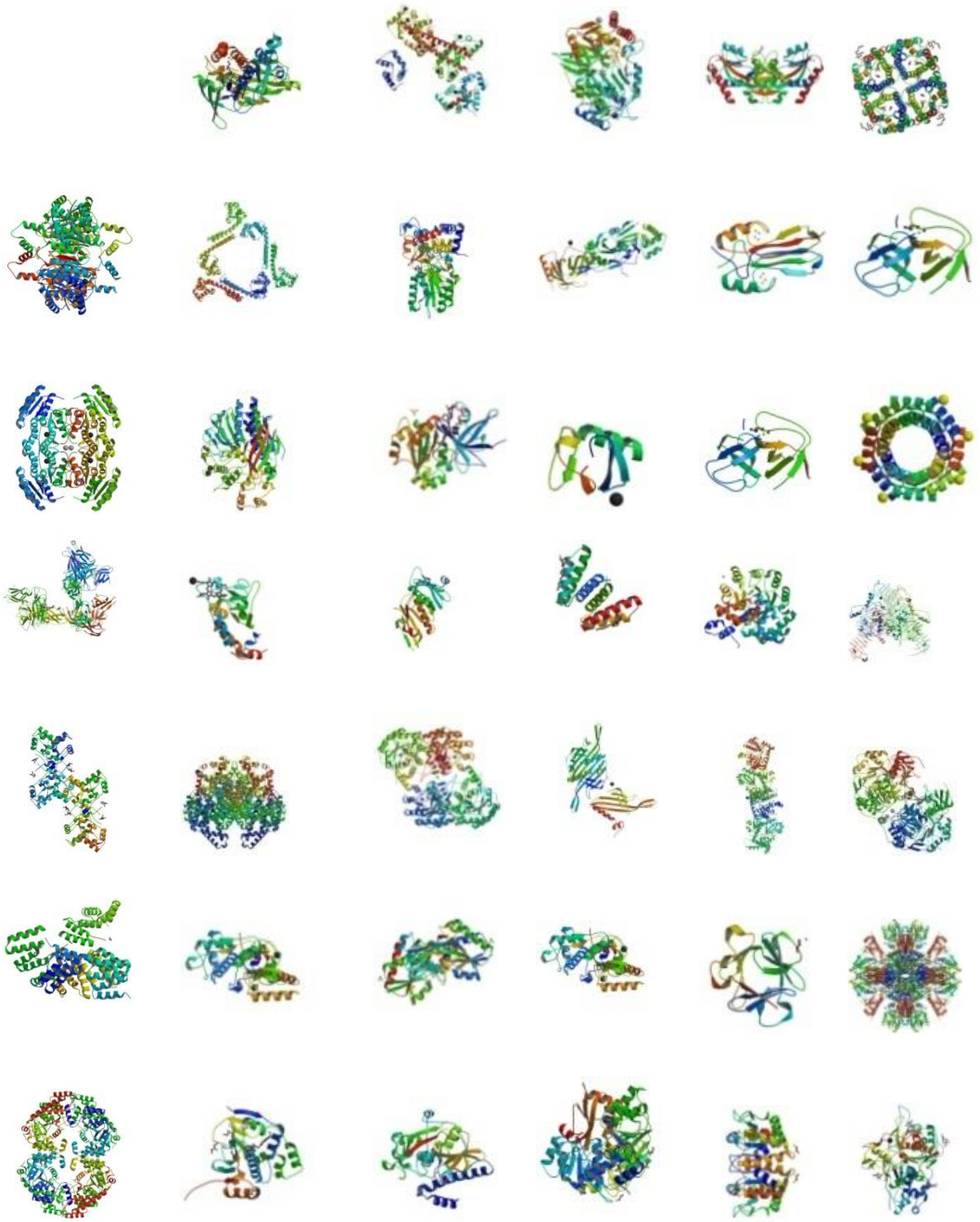
Born in Buffalo

Over 1,000 general biomedical laboratories world wide use the crystallization screening service with approximately 2,000 unique investigators.

Investigators are sent photographs of the results, analyze these images and perform their own optimization of any hits observed.

No information is released on targets. Progress is tracked by acknowledgements and citation searches. Currently no other metrics are used to measure success rates for the general biomedical community.

These images represent examples of structures from initial hits in the HTS laboratory.



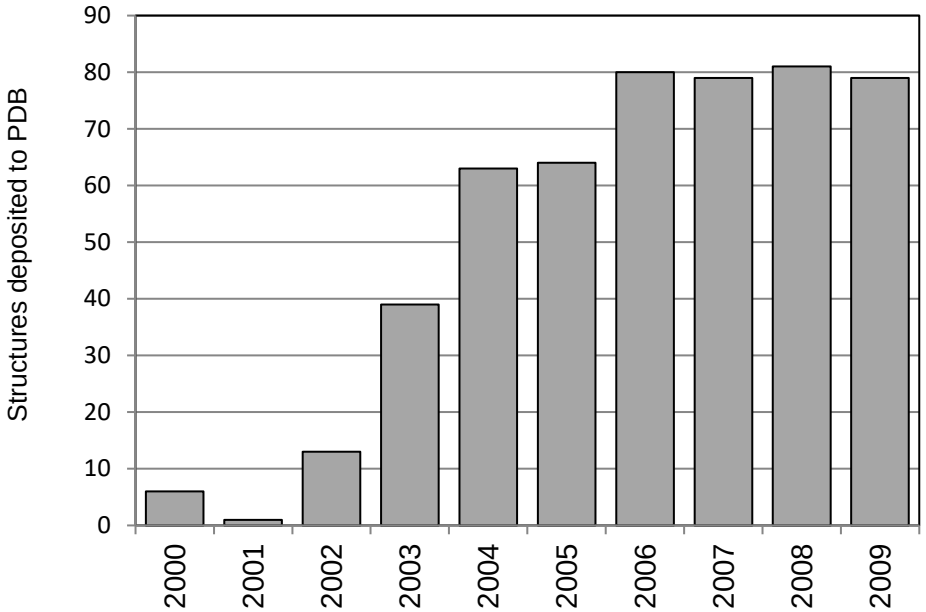
Where success is tracked.

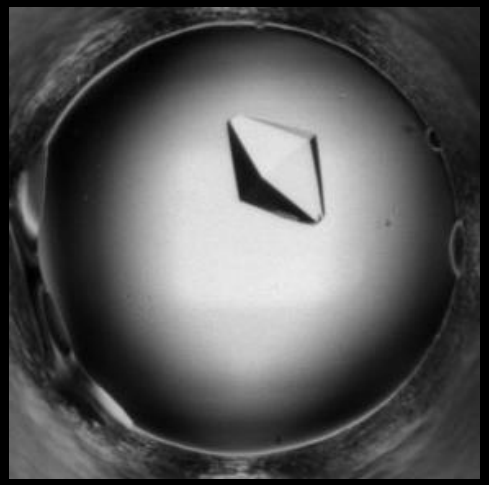
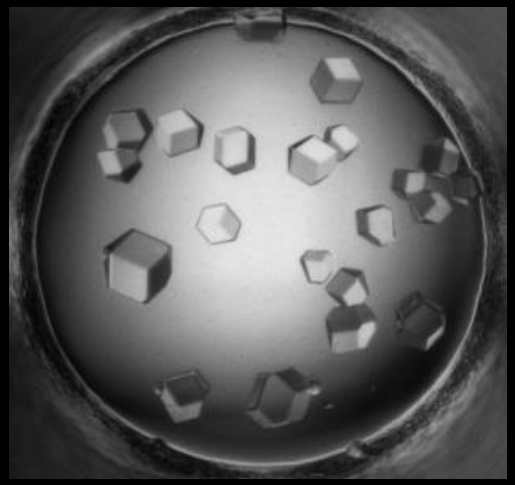
For our Protein Structure Initiative partners both success and failure is tracked. In the case of NESG our initial screening hits enable on average 80 structures per year to be deposited to the PDB.

The graph demonstrates the ramp up of operations with maximum success reached from 2006 onward.

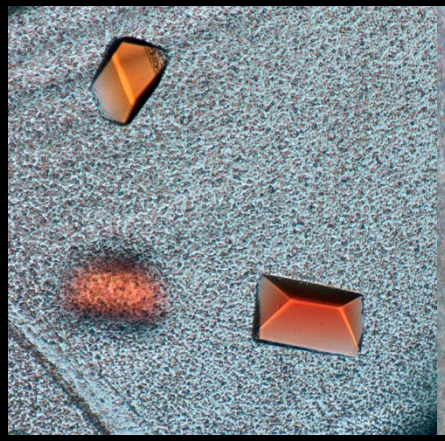
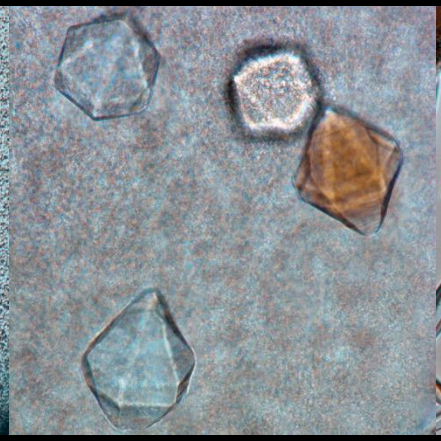
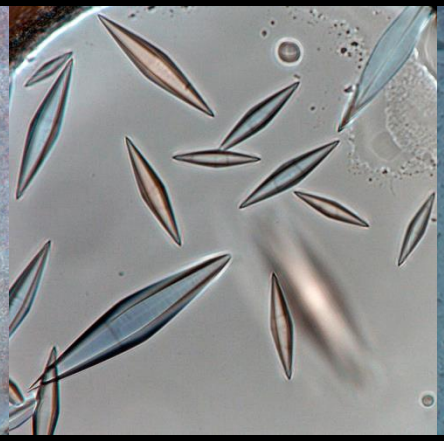
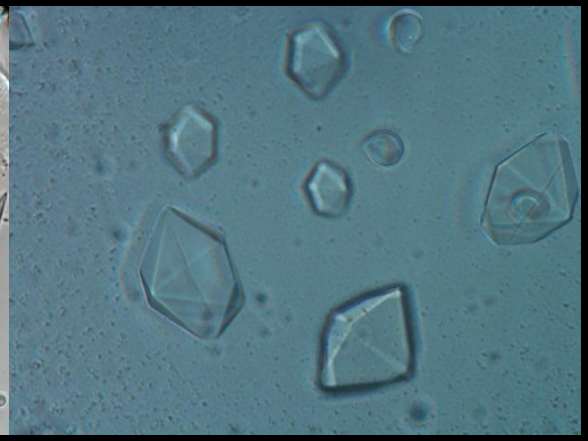
Our success rate from protein in the door to a crystallization hit leading to a PDB deposition is **22%**.

The NESG samples represent a special case in that they are well characterized beforehand – size exclusion chromatography, mass spec analysis and dynamic light scattering studies.




Crystallography Requires Crystals

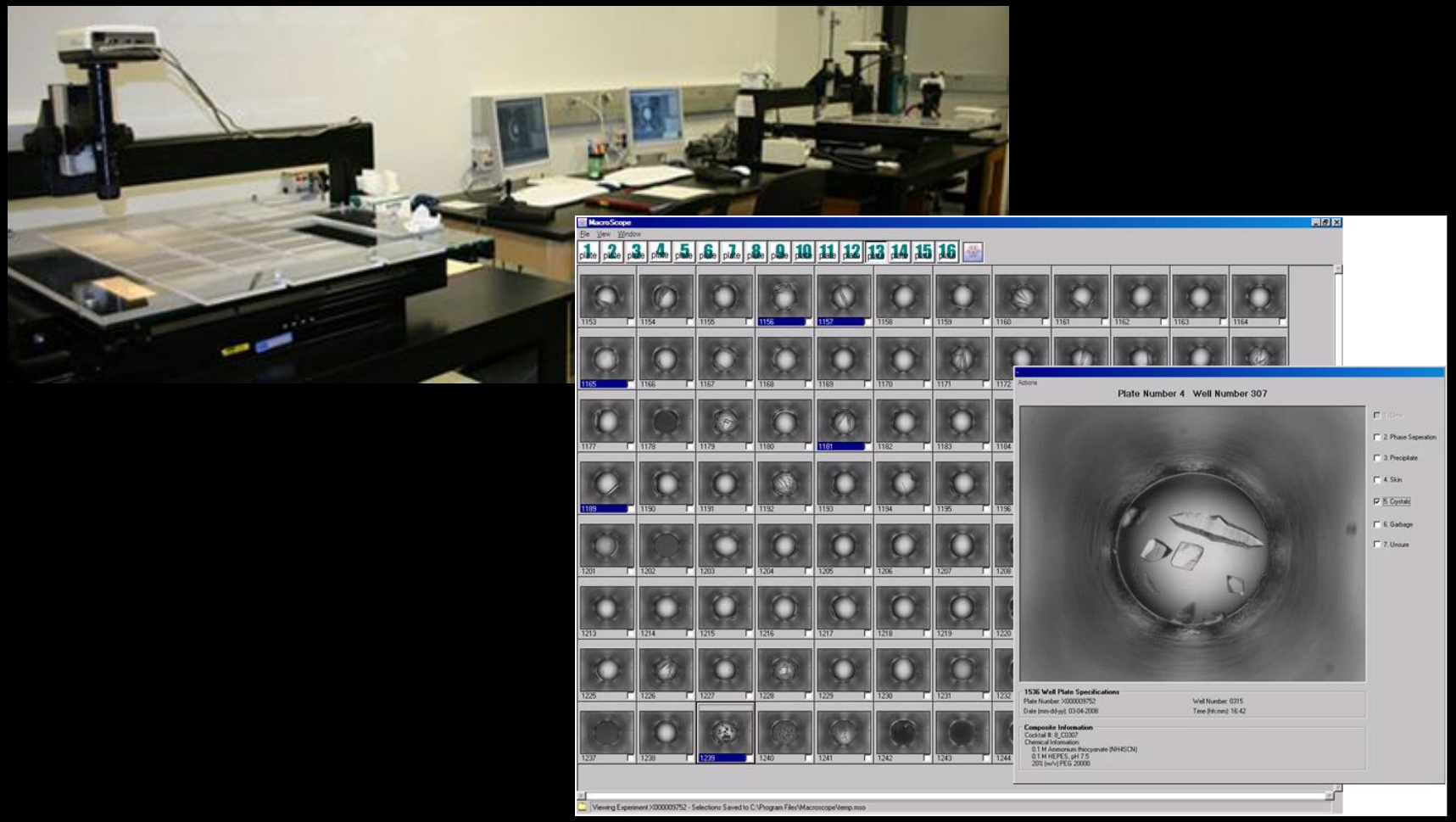





No crystal …

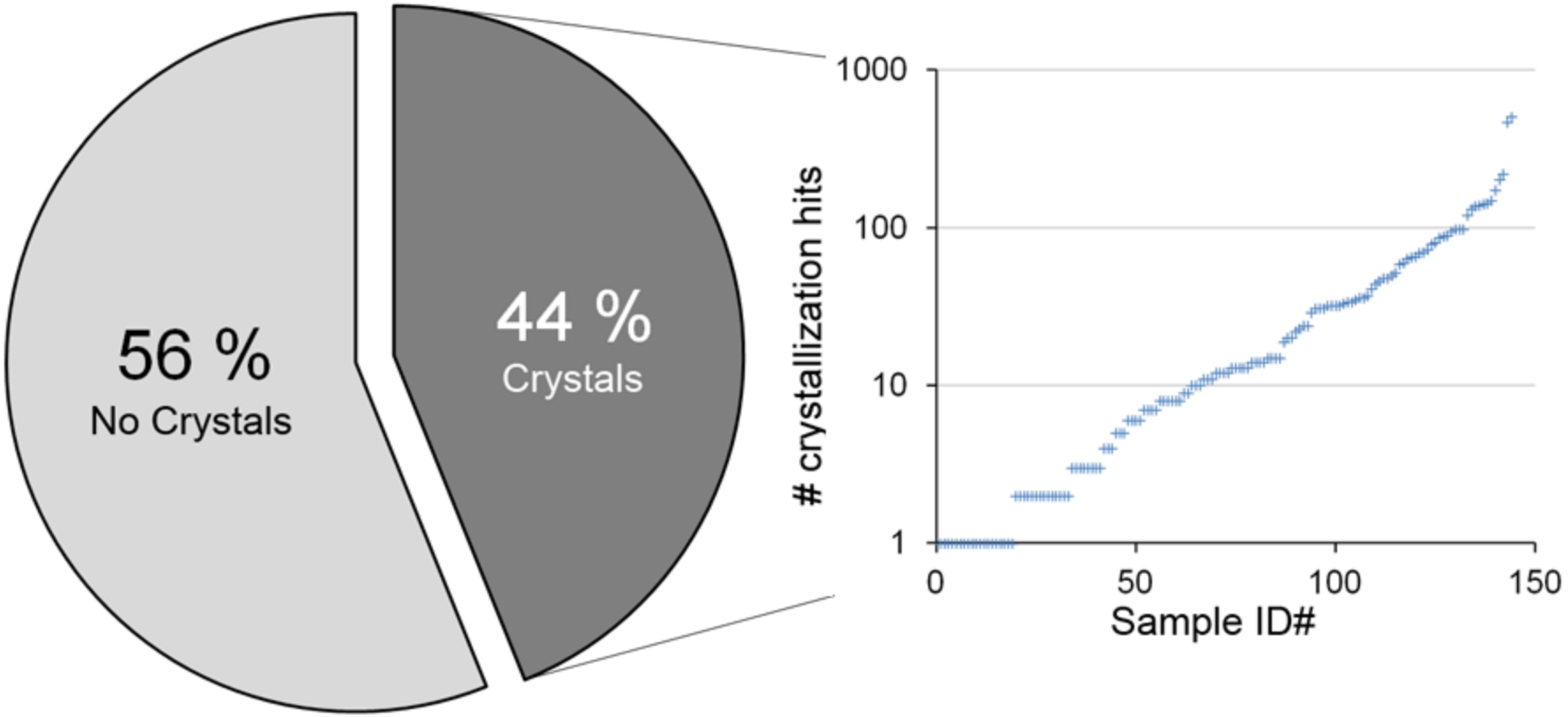
No crystallography ….

No crystallographer ….

Imaging technology based on standard microscopy – requires visible crystals

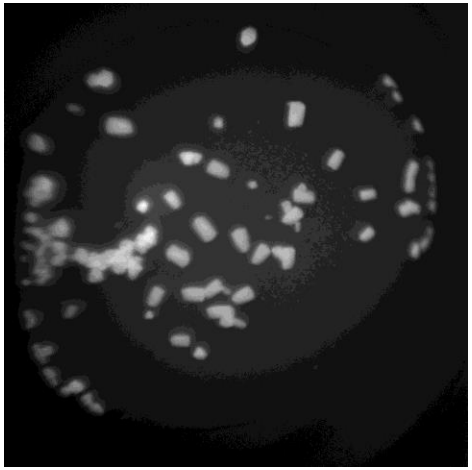


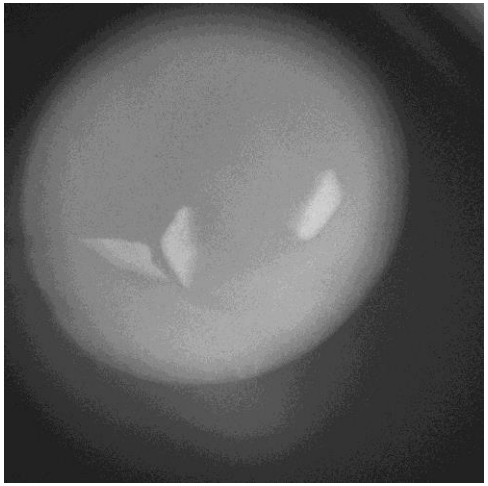
And it's still pretty good at indentifying them

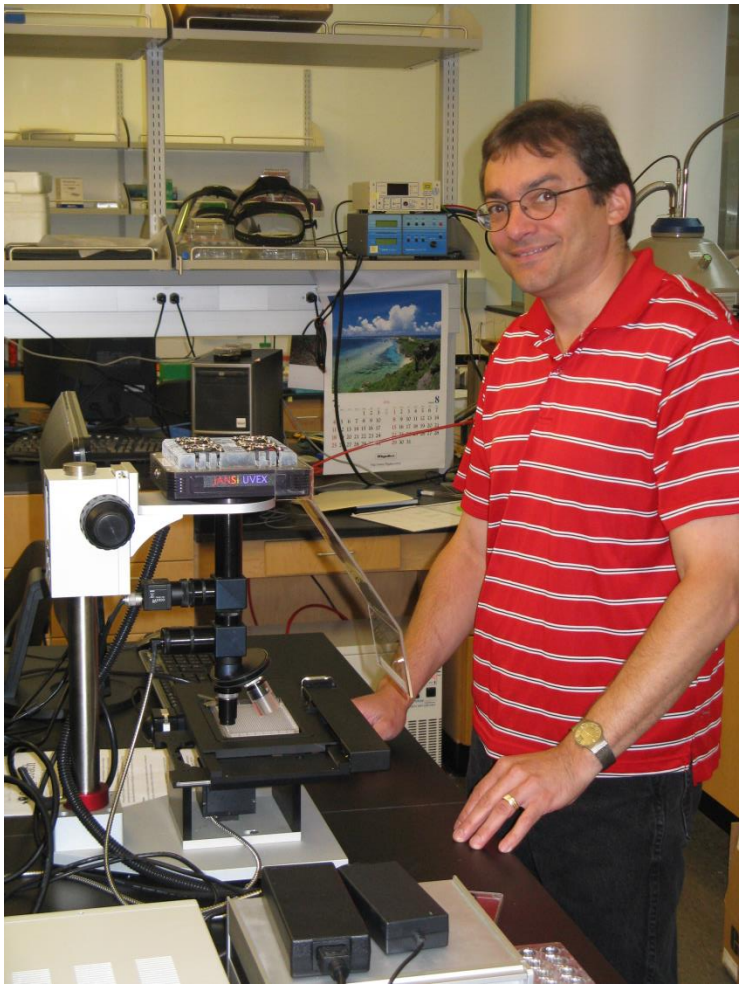


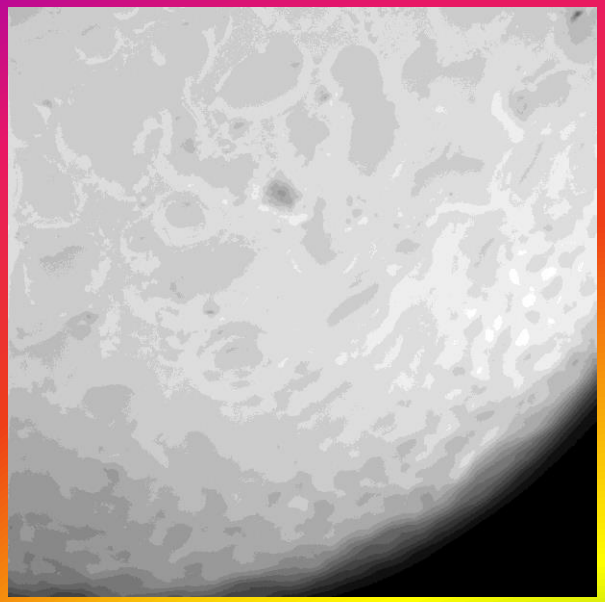
Based on data from NESG 2013, 144/328 targets with one or more verified hits

UV imaging – is it protein?

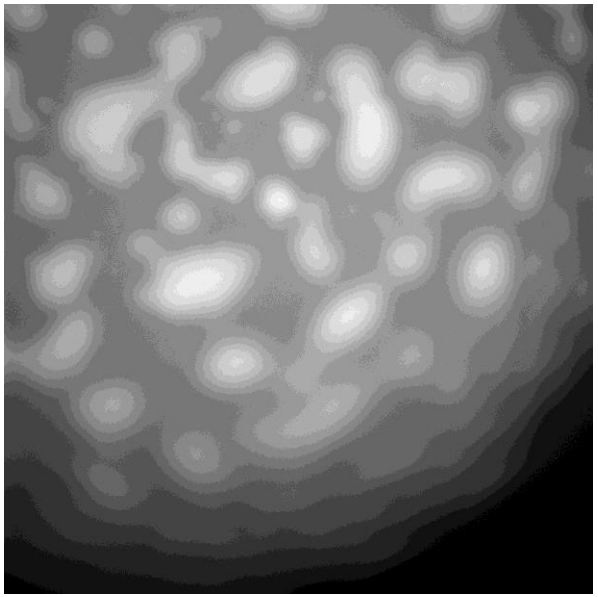


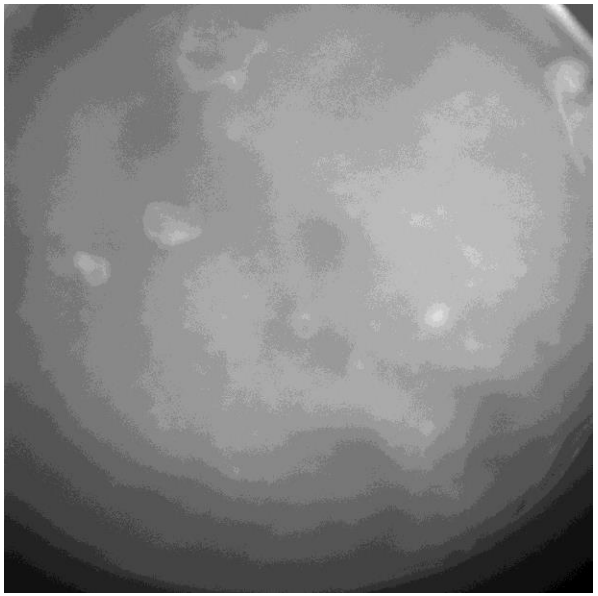






Visible

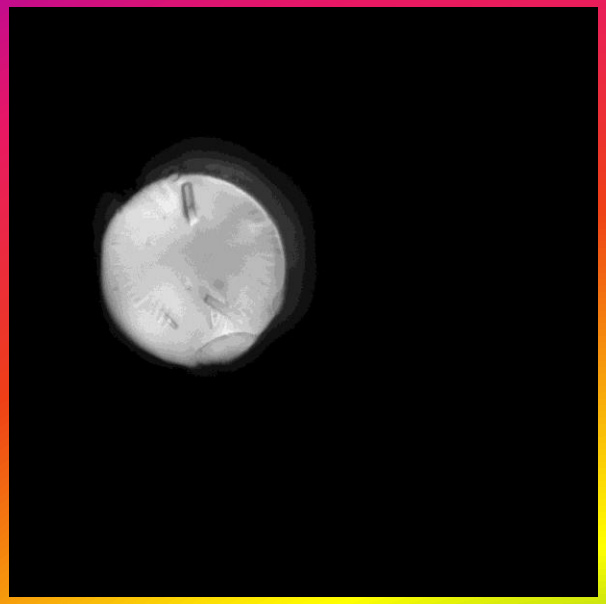


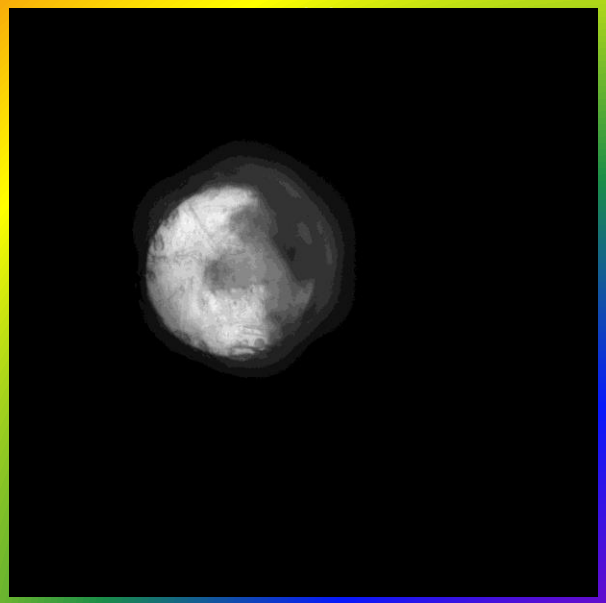


UV

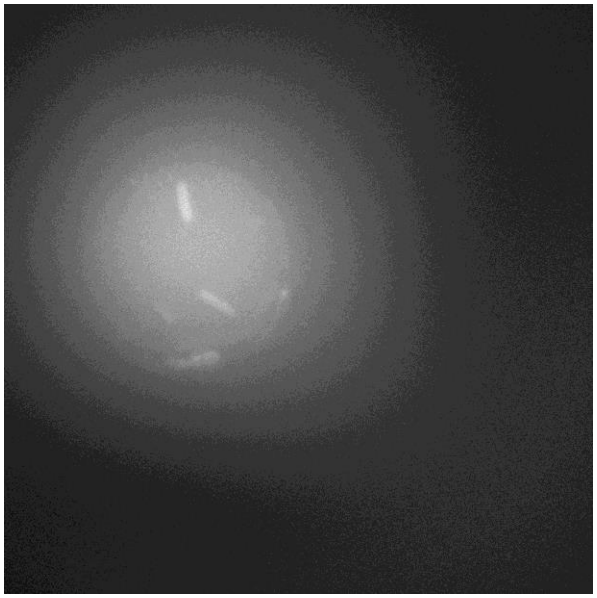
Protein phase

Protein crystal

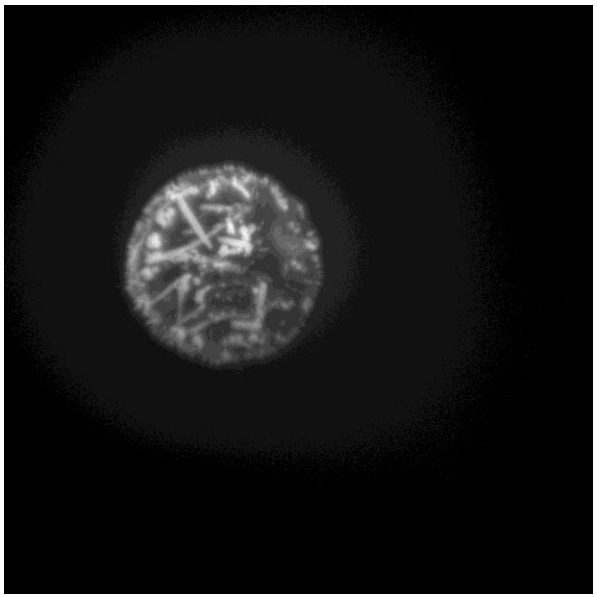




Visible

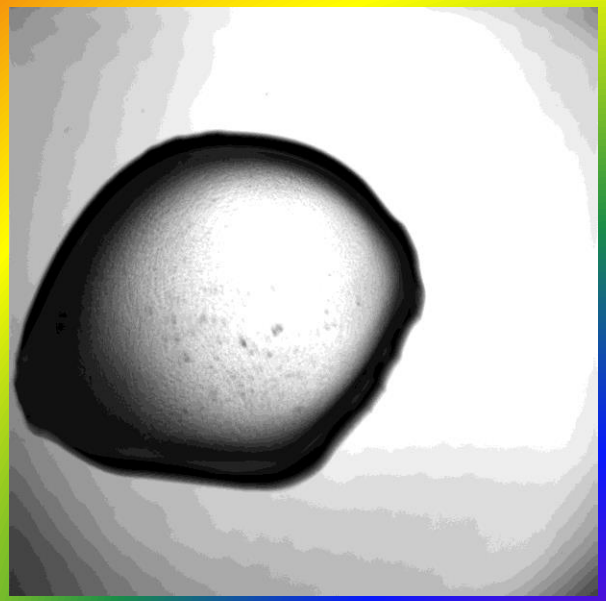


Protein crystal

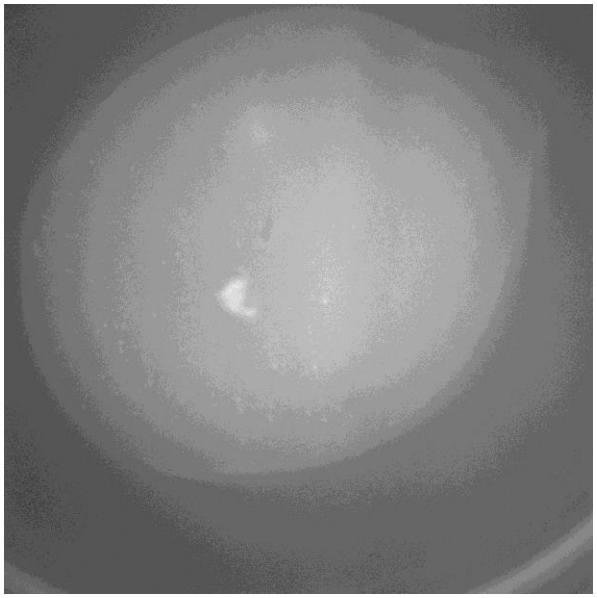


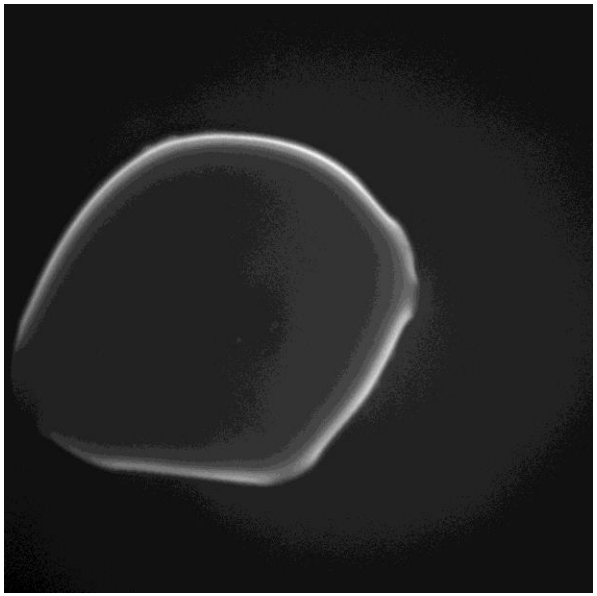
Protein crystal

UV



Visible

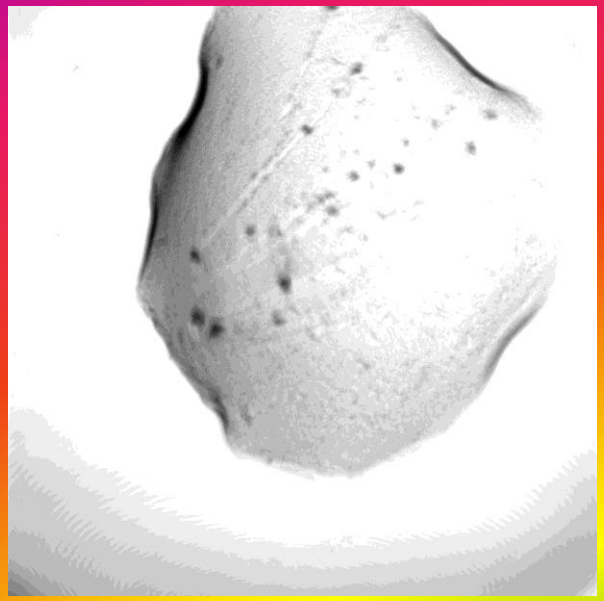


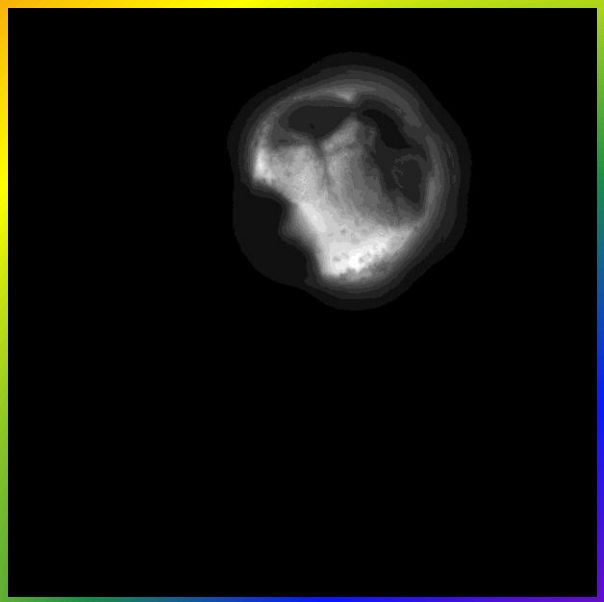


UV

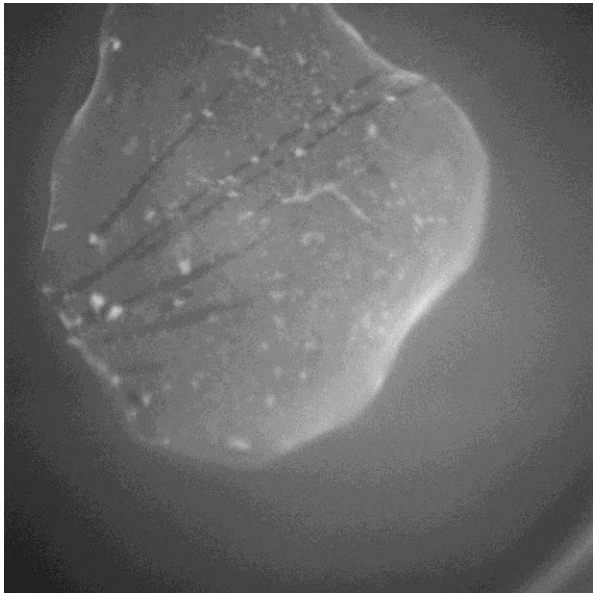
Protein crystal

Salt crystals

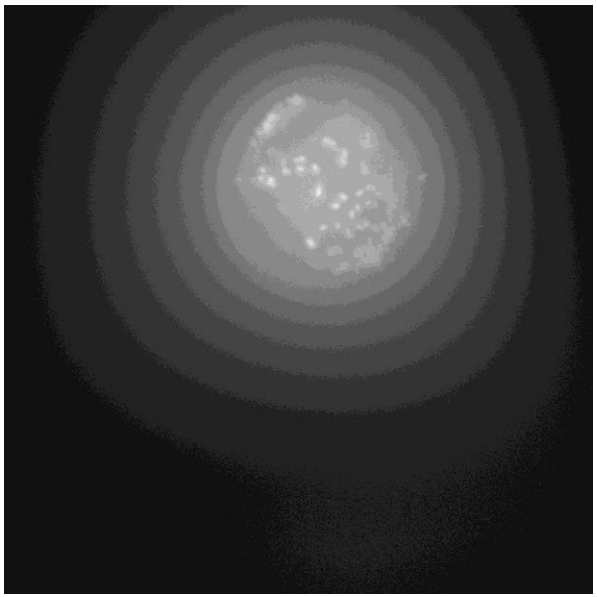




Visible



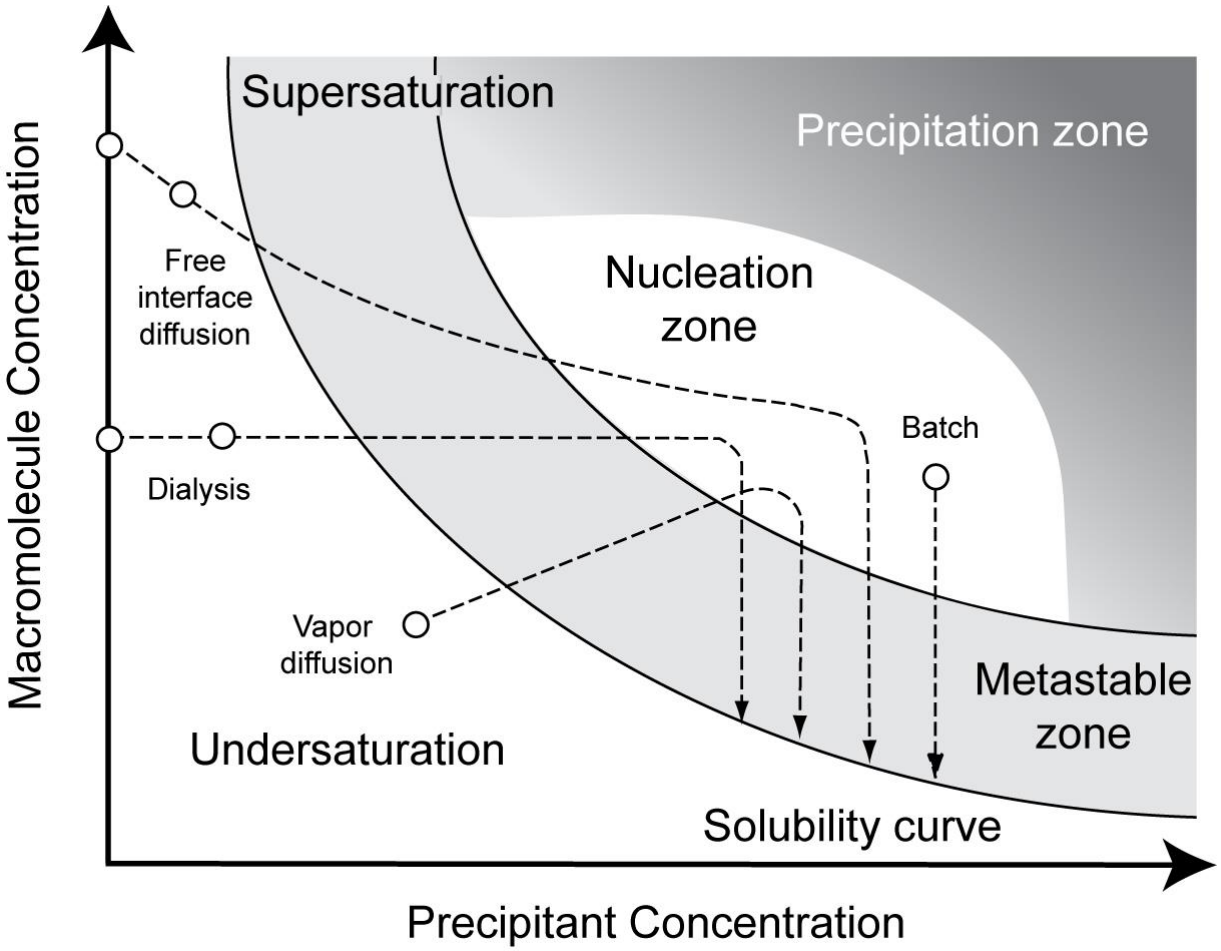
Protein crystals



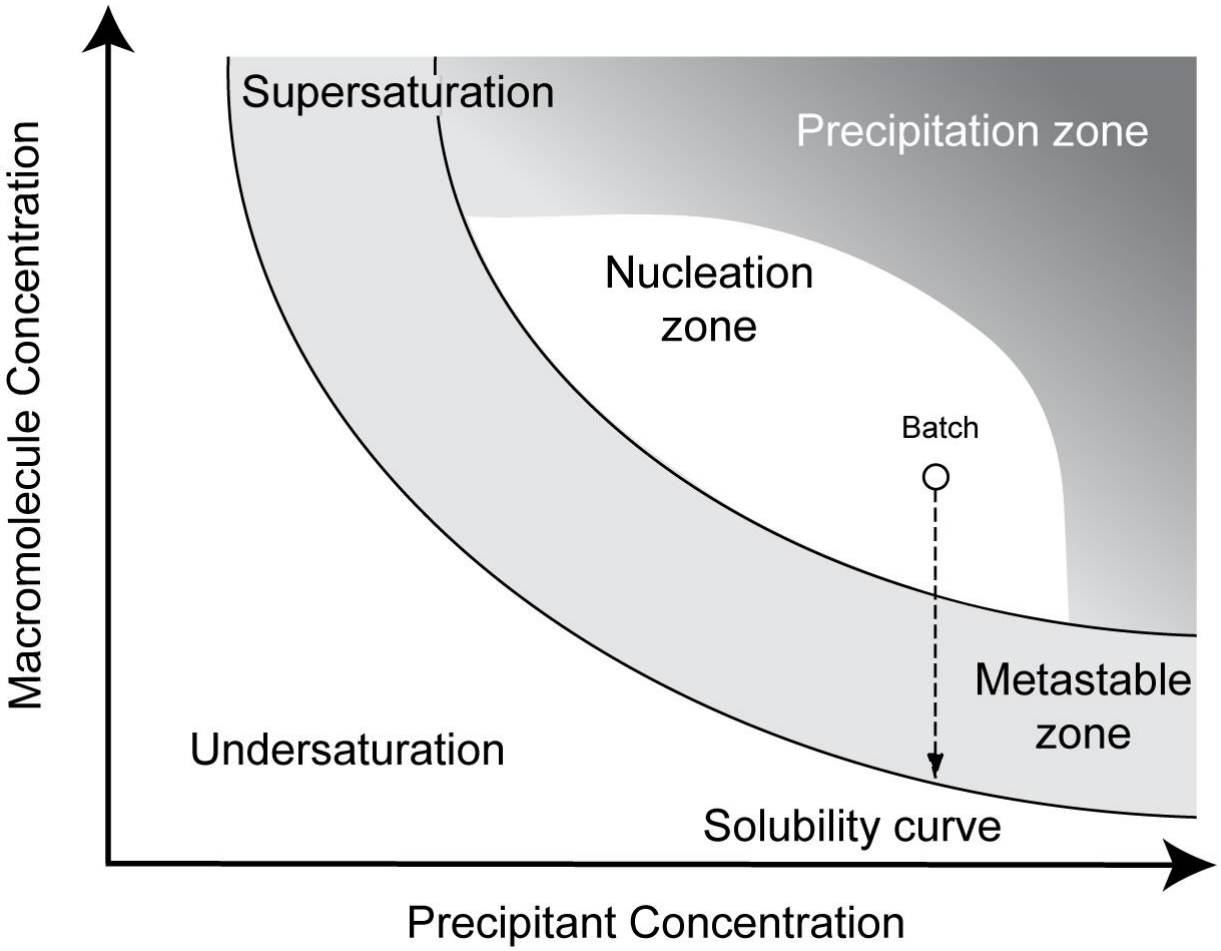
Protein crystals

UV

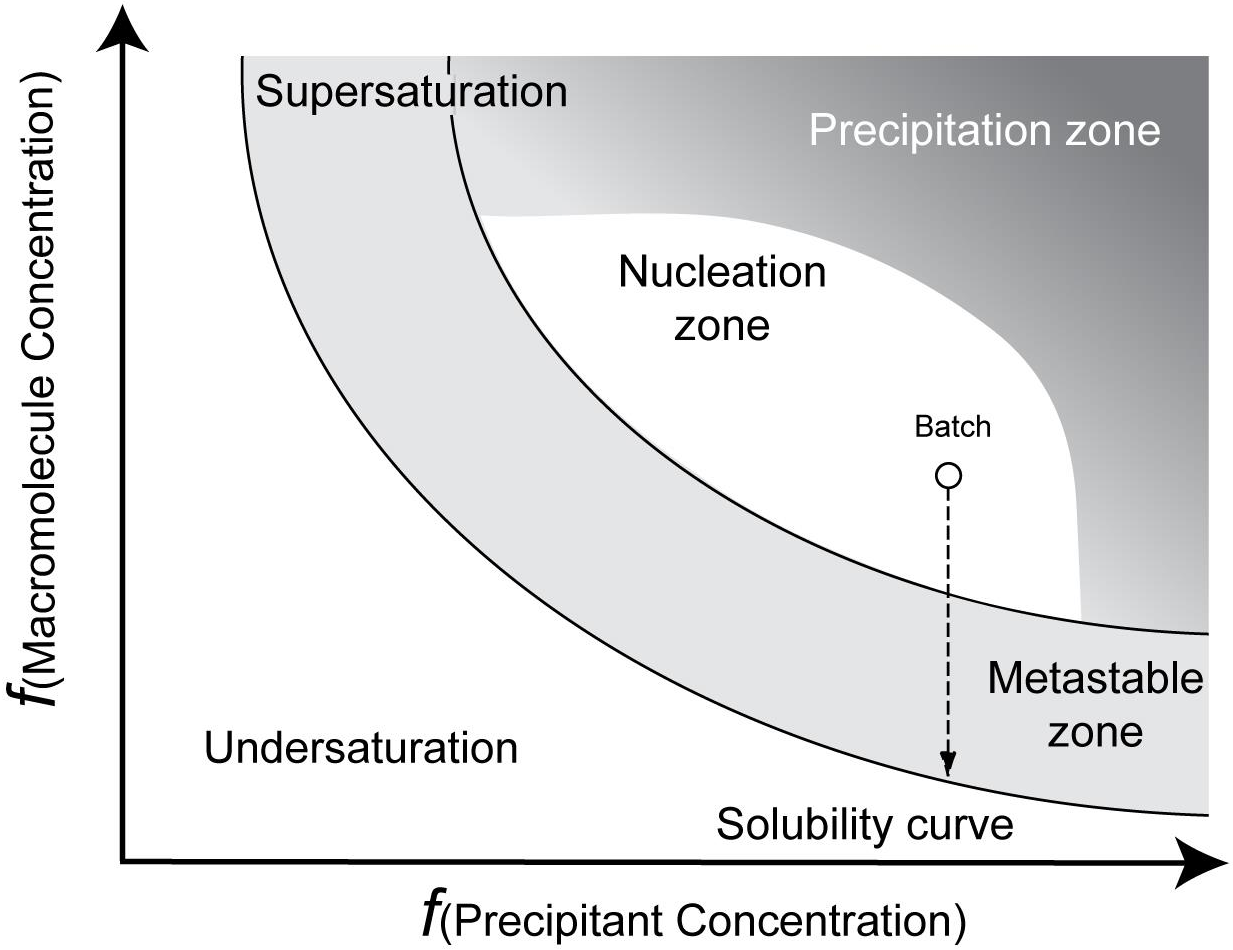
Simplified phase diagram for crystallization



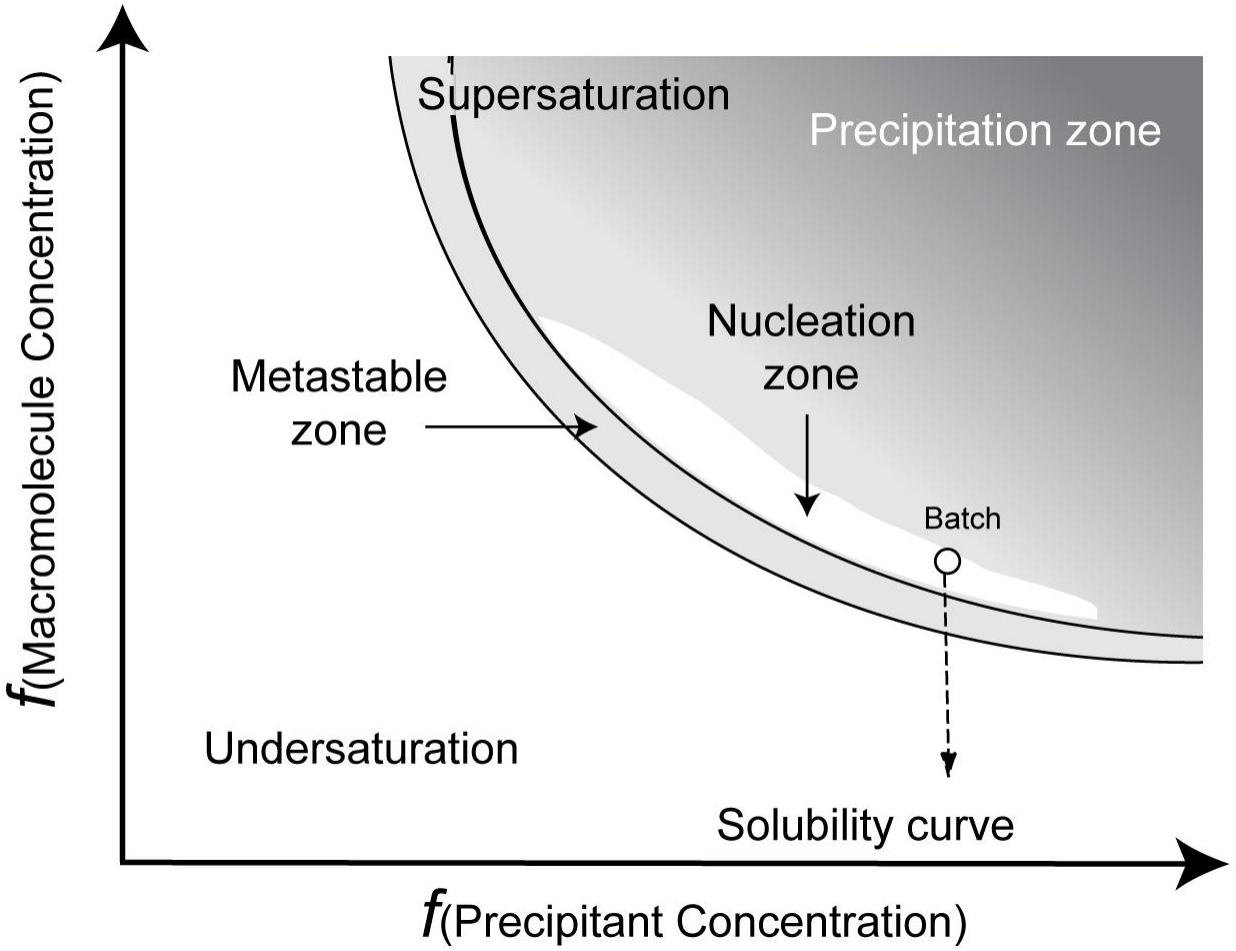
Even simpler phase diagram for crystallization



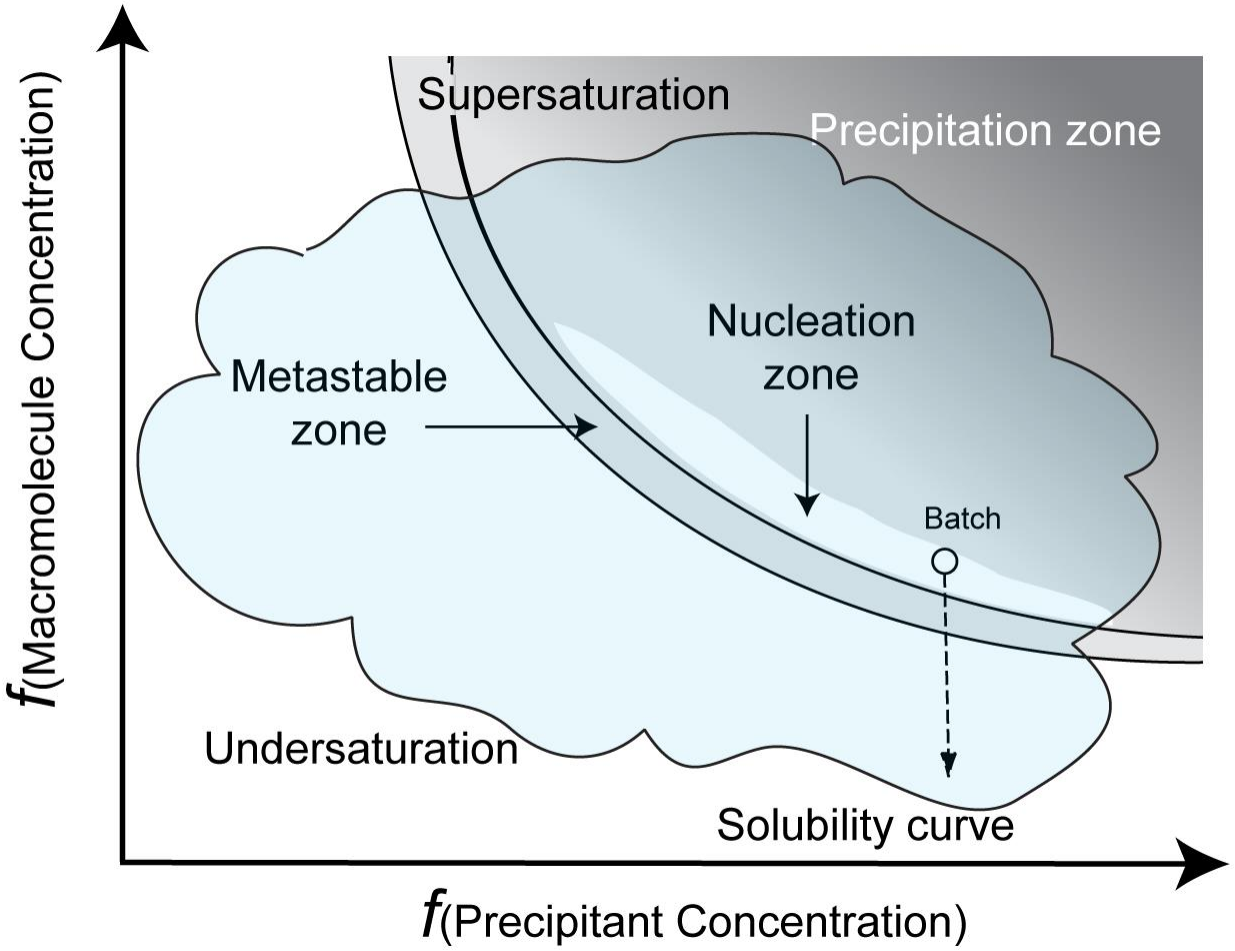
Start to throw some reality into the equation



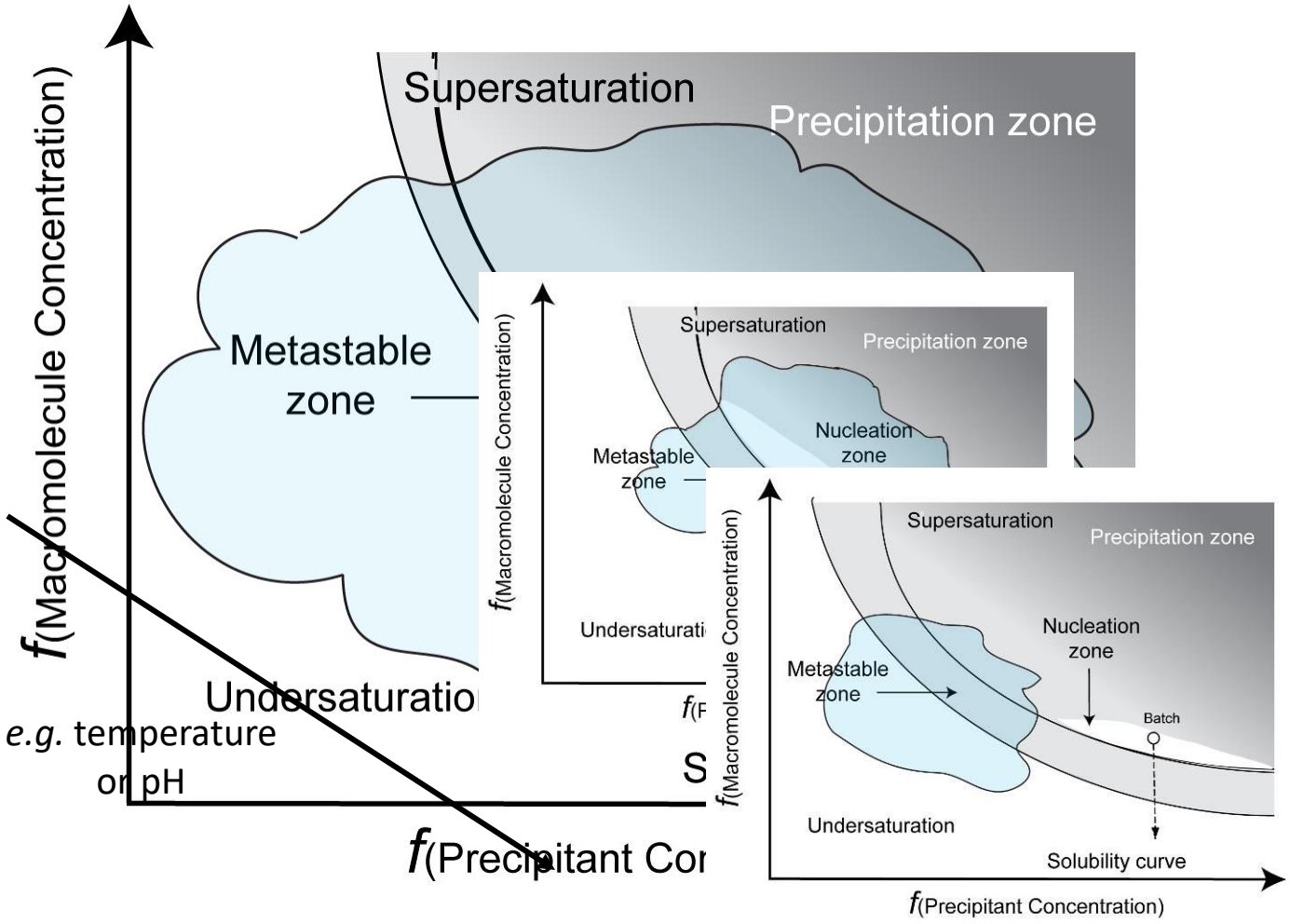
And reduce the chances of crystallization a little



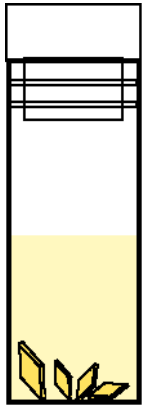
Add the experimental space we sample



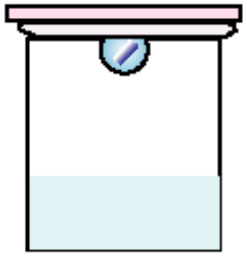
And the fact that it's not just two dimensions



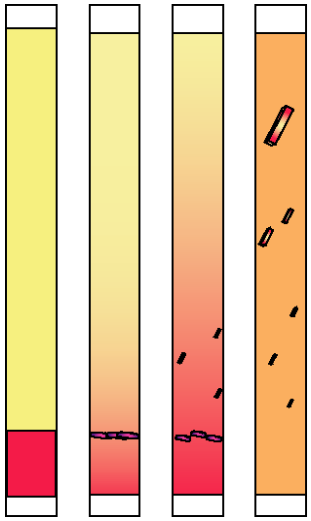
Three basic methods: batch, vapor and liquid



Batch

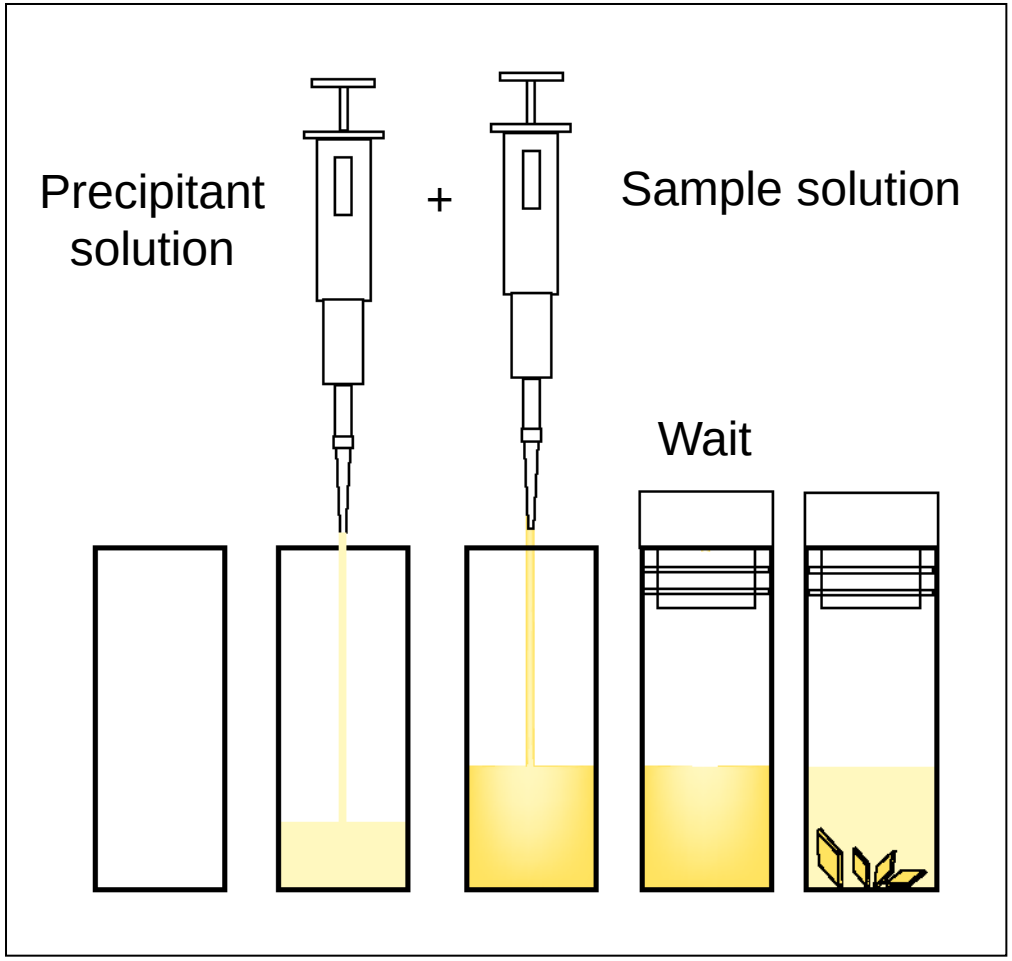


Vapor
(also known as vapour)

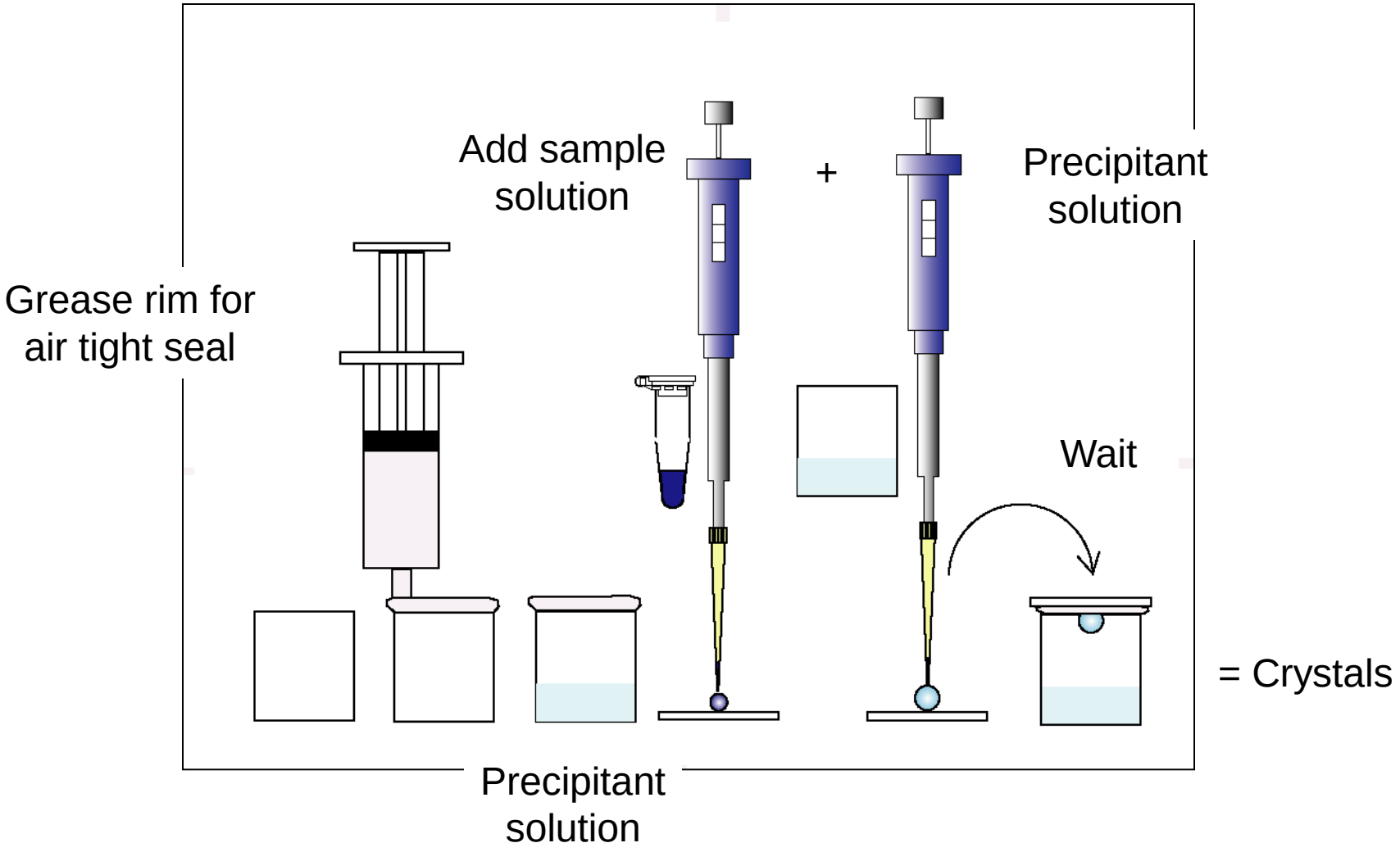


Liquid

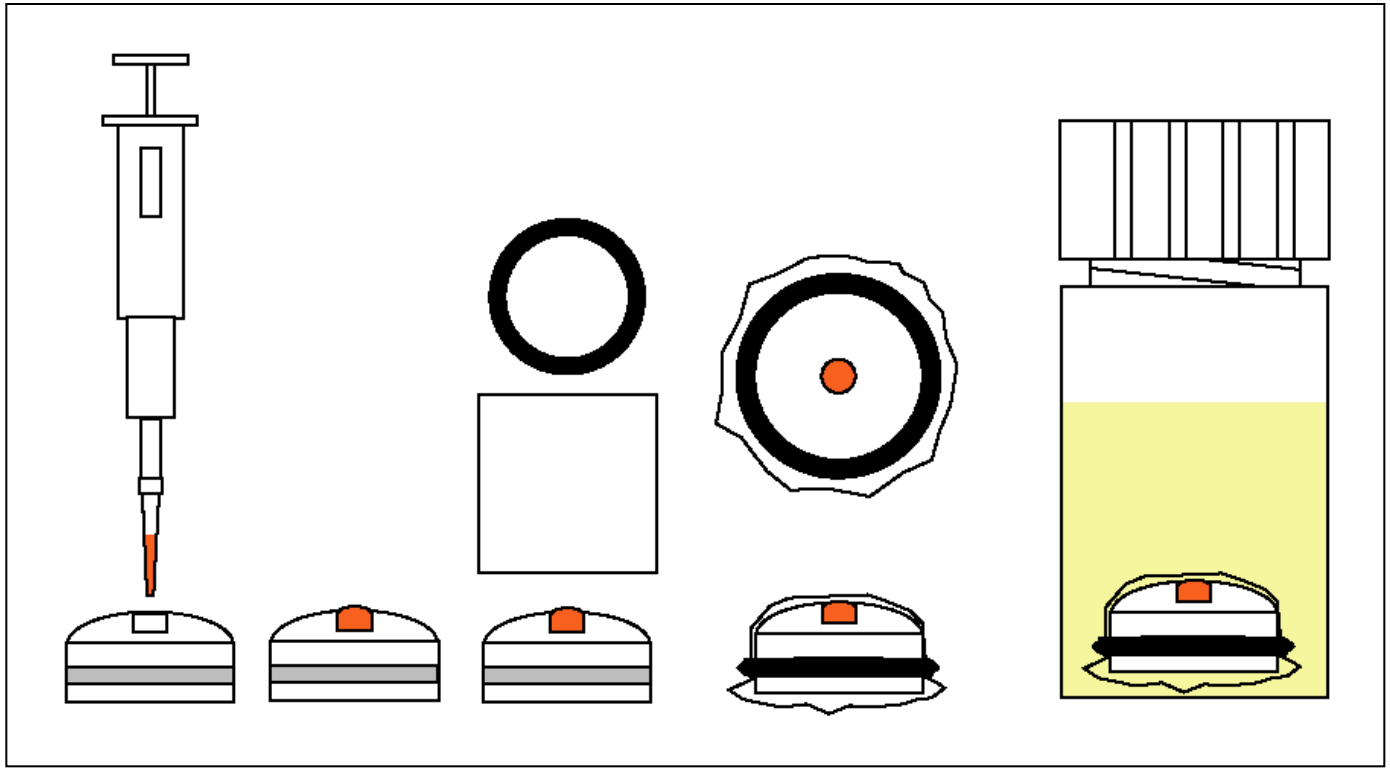
Batch in a Vial: Set up



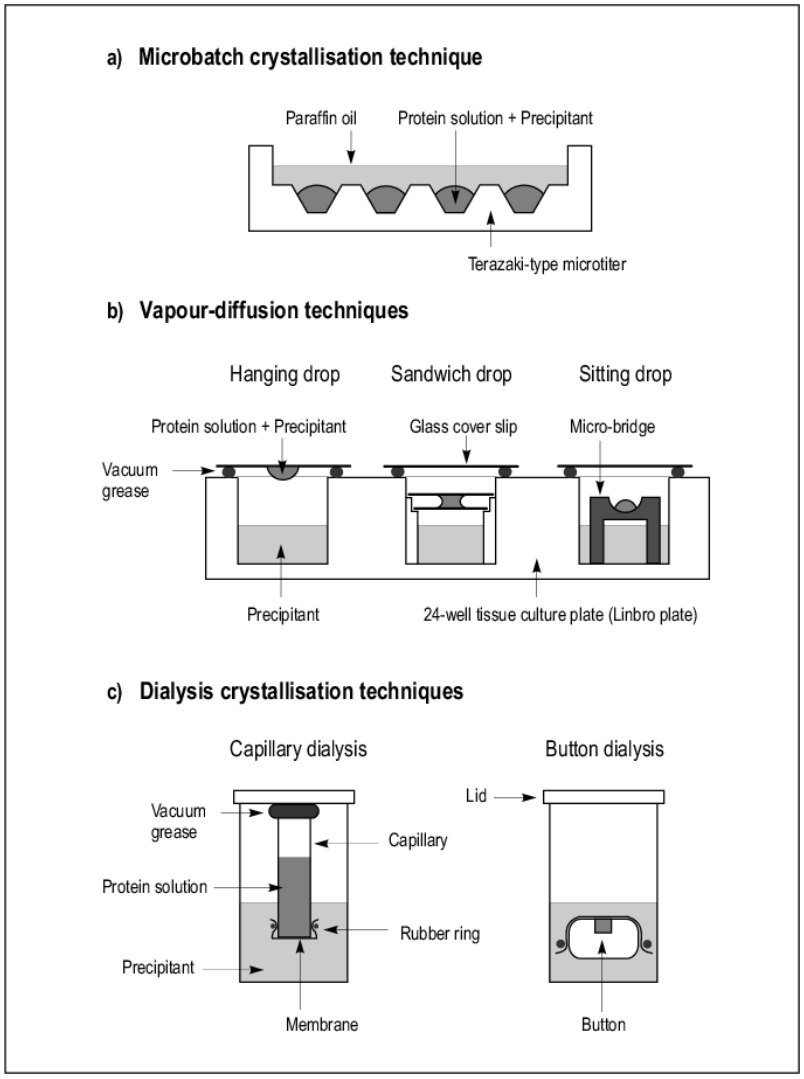
Vapor Diffusion Setup



Dialysis Experiments



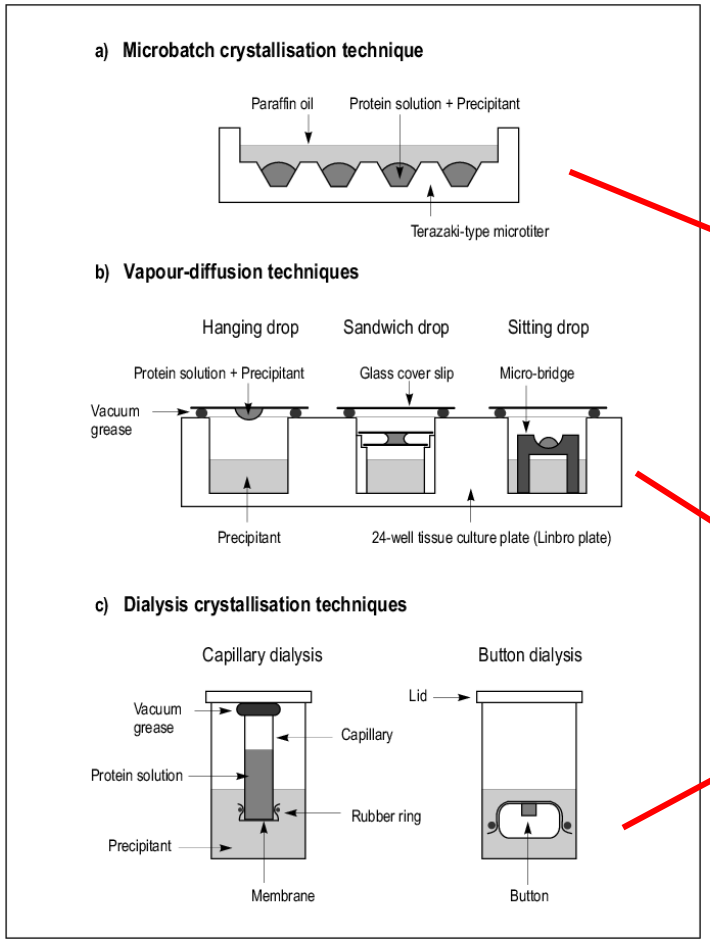
Crystallizing Macromolecules



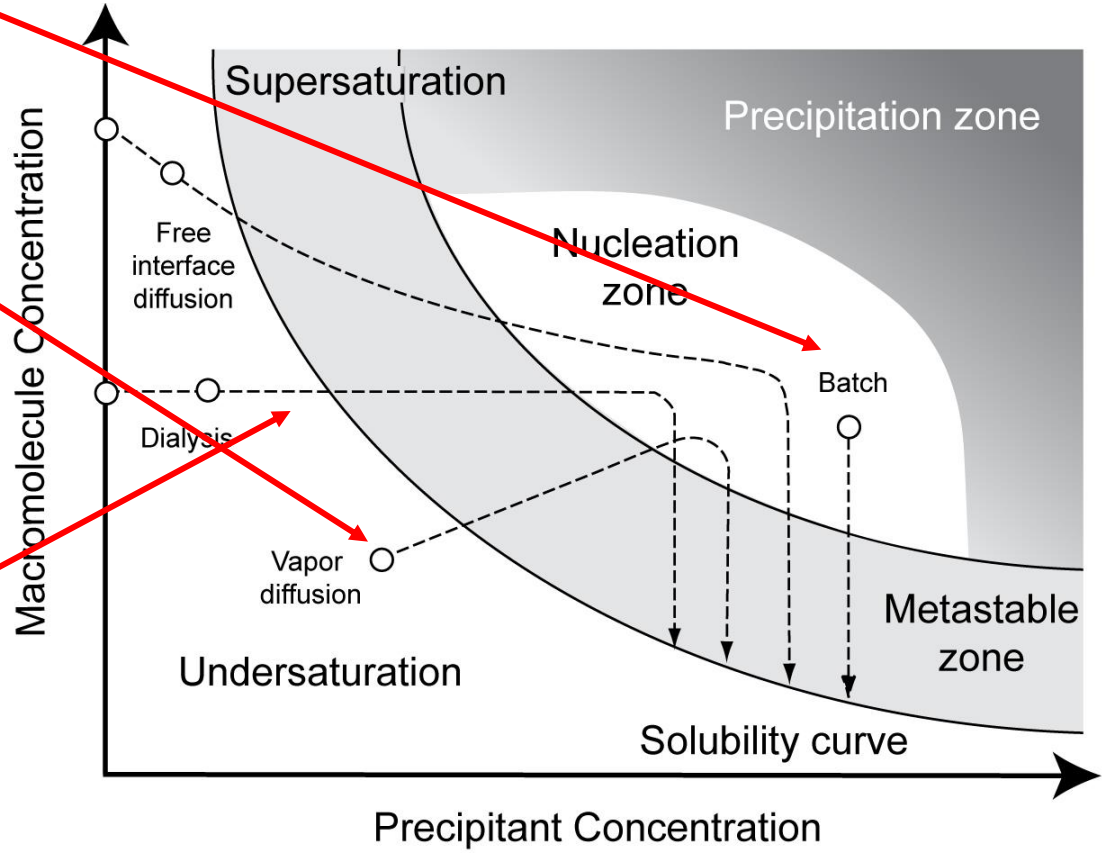
Many different methods but they all have things in common:

- They are designed to traverse the crystallization phase diagram.
- They use many different kinds of solutions to sample crystallization space at many points.

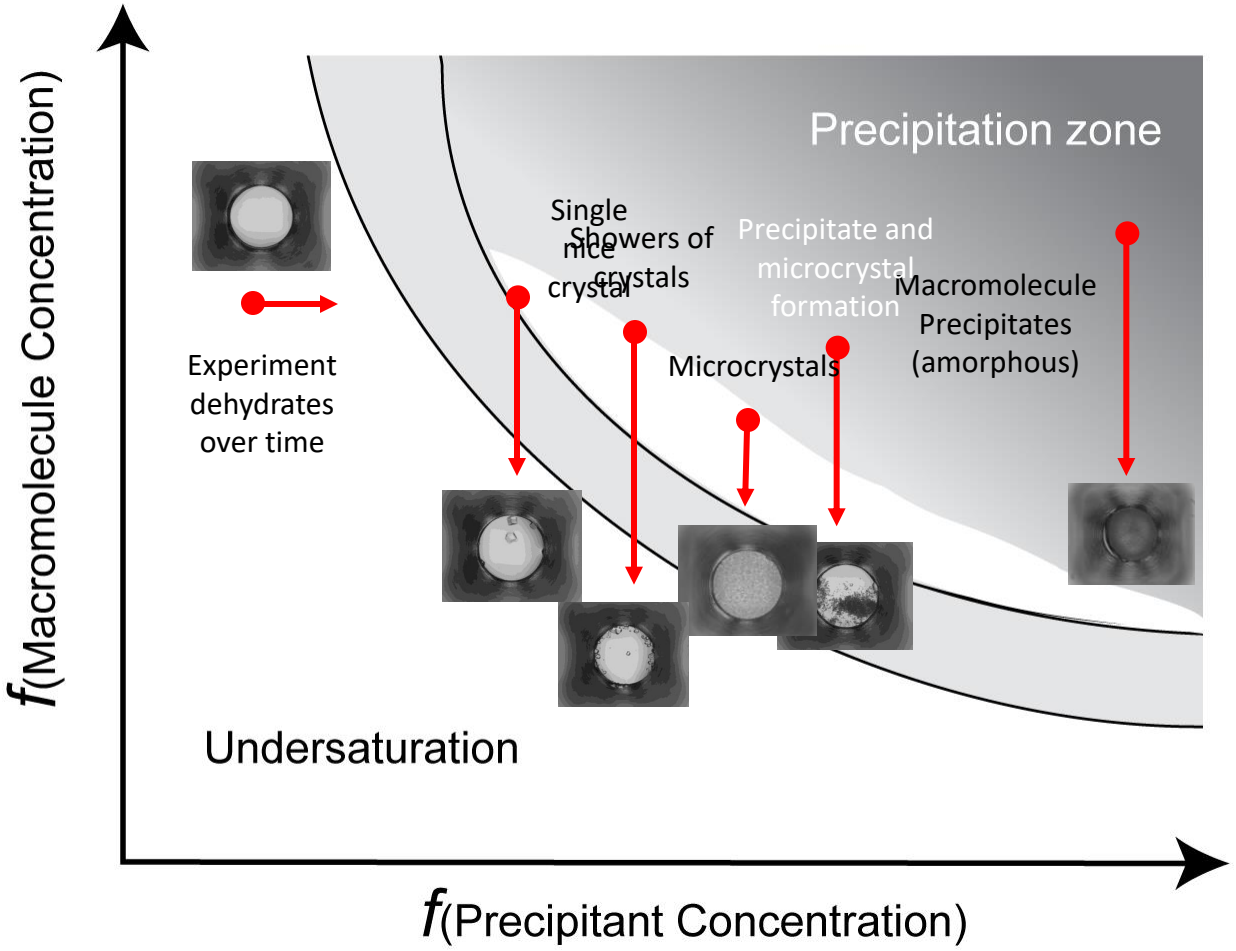
Simplified phase diagram for crystallization



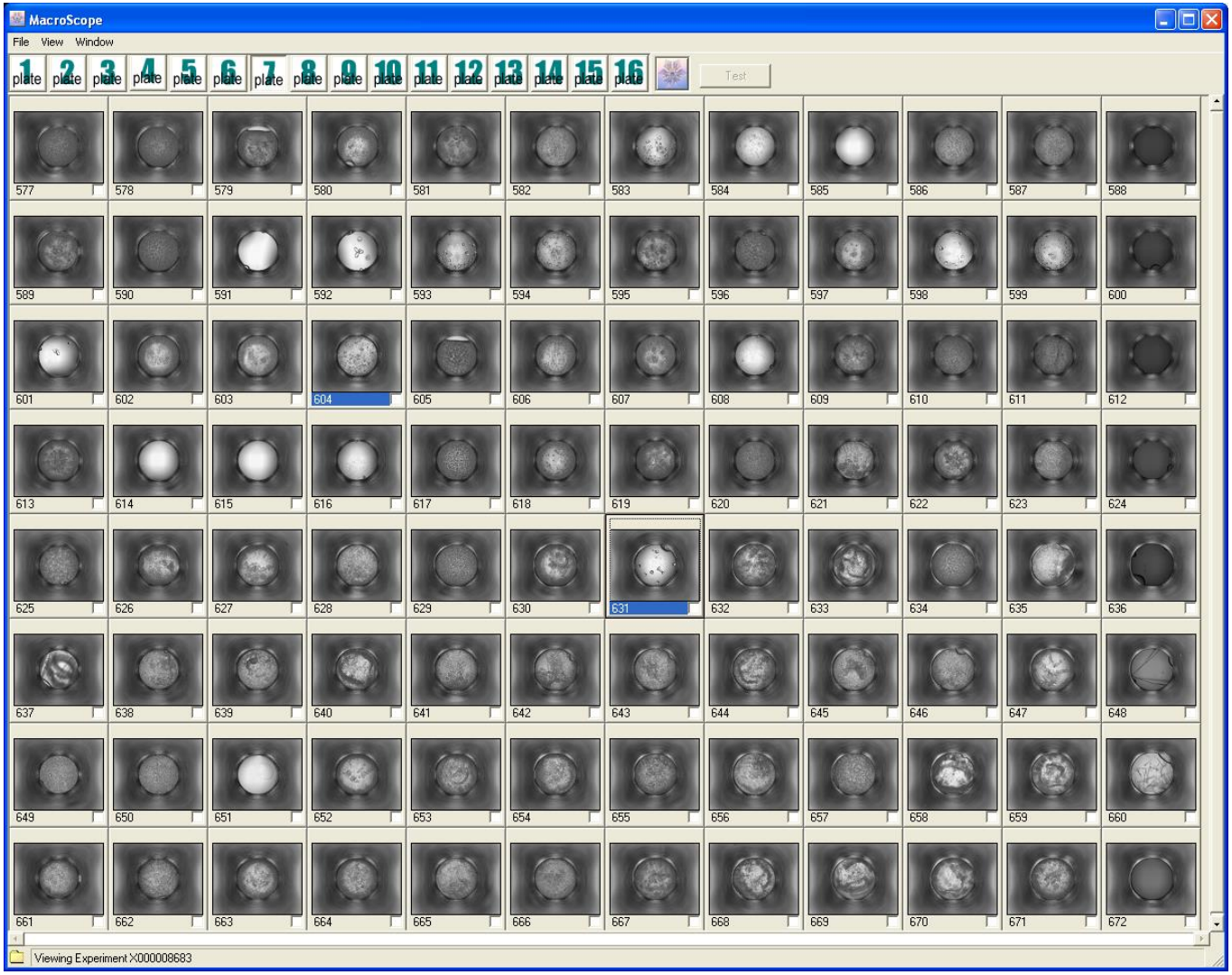
Different techniques traverse the crystallization space differently.



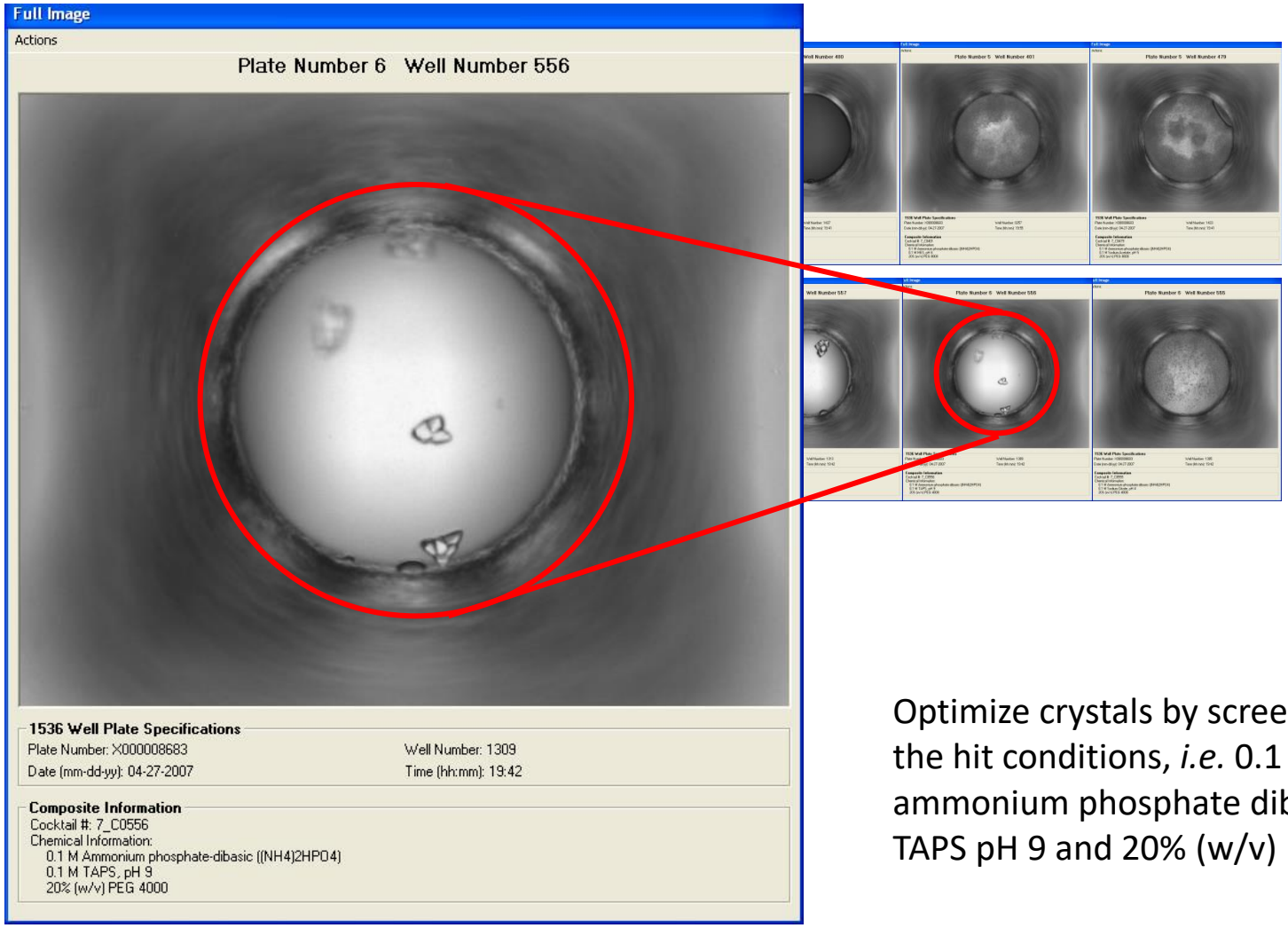
What results can we expect to see?



What do we actually see?

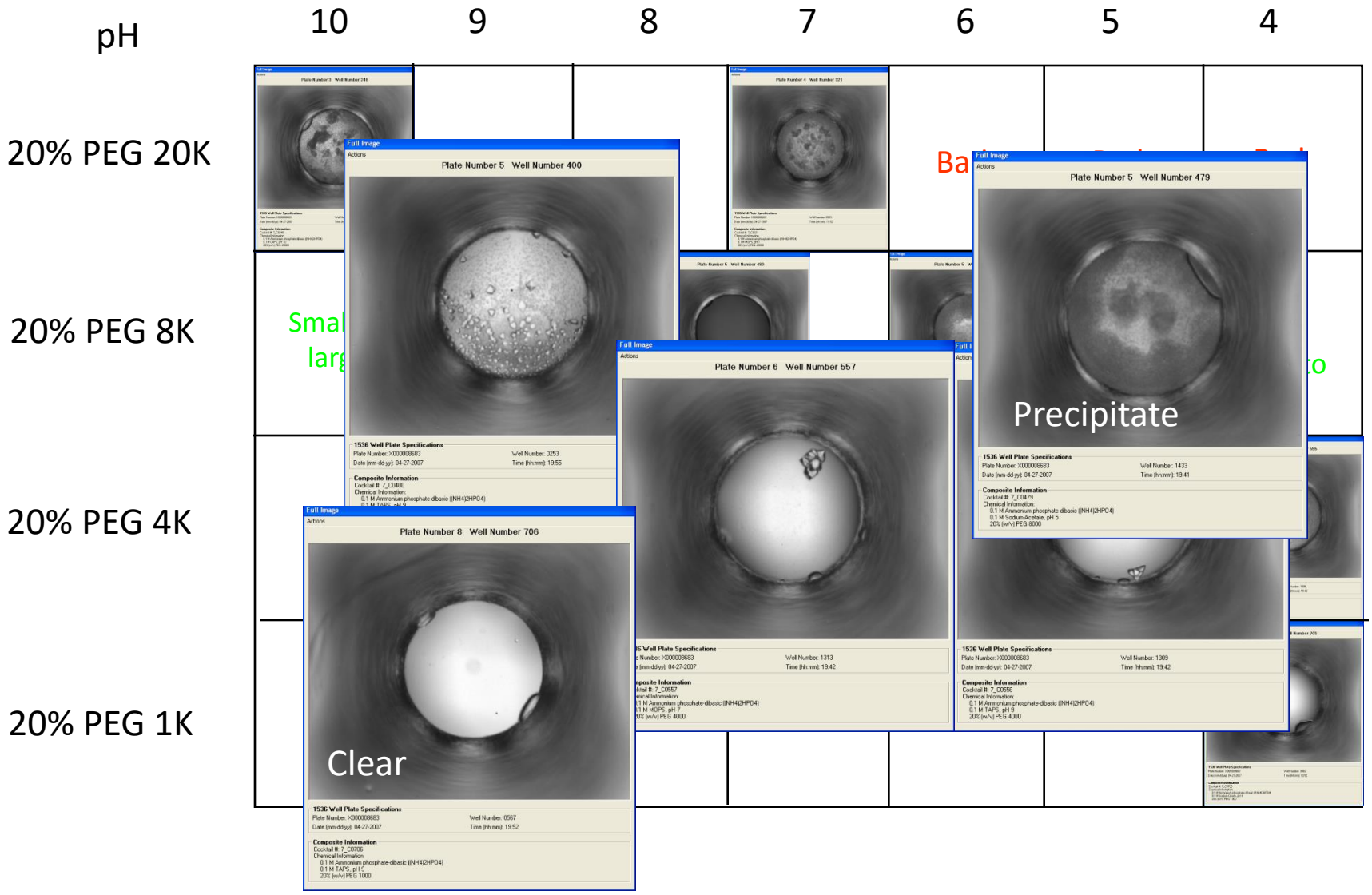


What do we actually see?



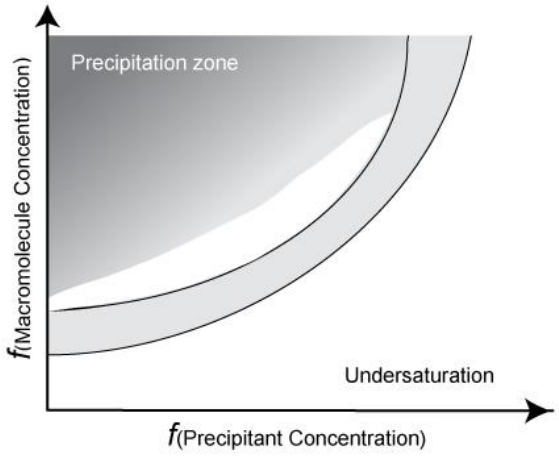
Optimize crystals by screening around the hit conditions, *i.e.* 0.1 M ammonium phosphate dibasic, 0.1 TAPS pH 9 and 20% (w/v) PEG 4000

If we plot the results in chemical space the road becomes clear

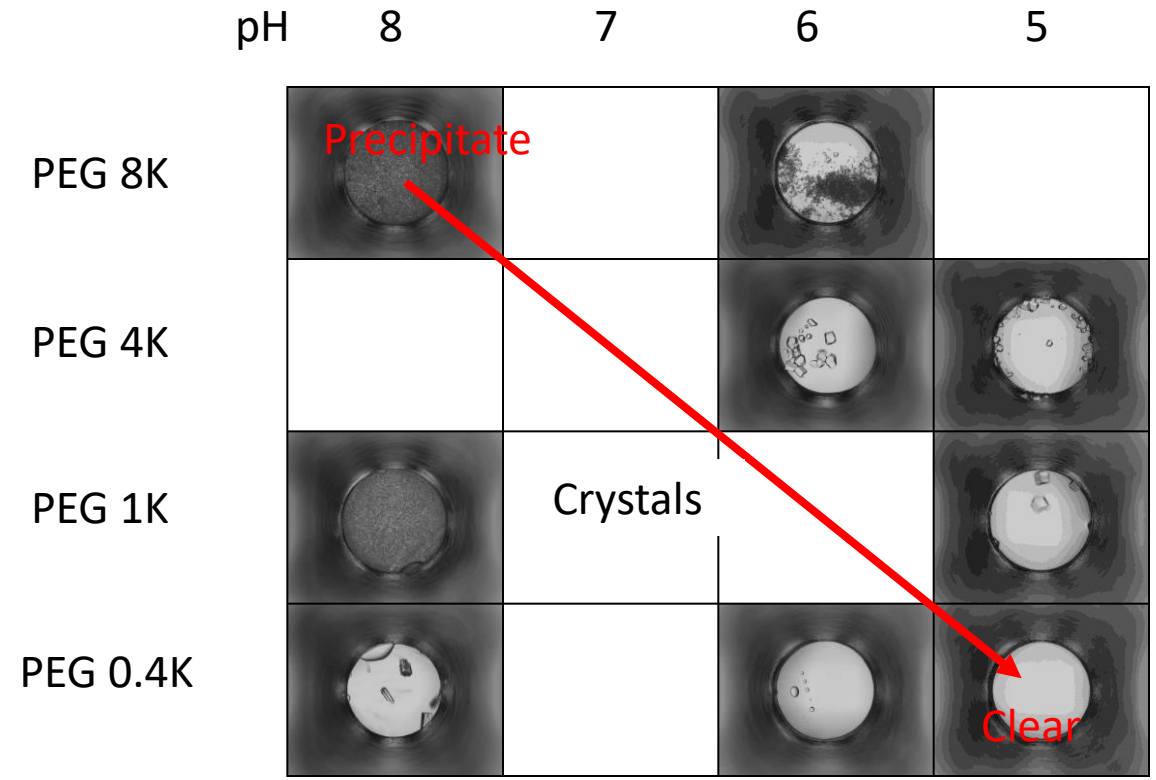


Chemical space provides a vector for optimization

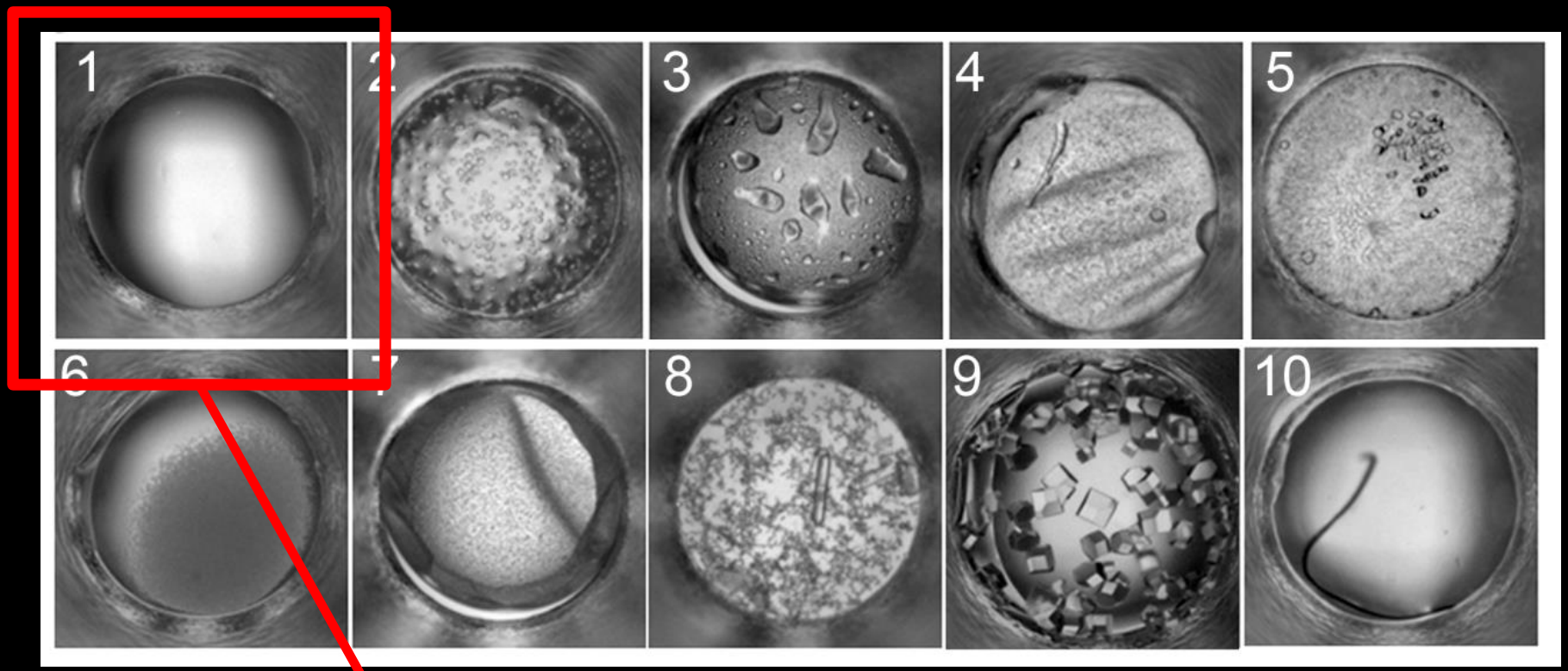
In this case the path from precipitate through crystals to clear is obvious. The phase diagram is reversed. Also clear are the number of chemical conditions that have not been sampled.



Ubiquitin, 40% PEG, 0.1M zinc acetate

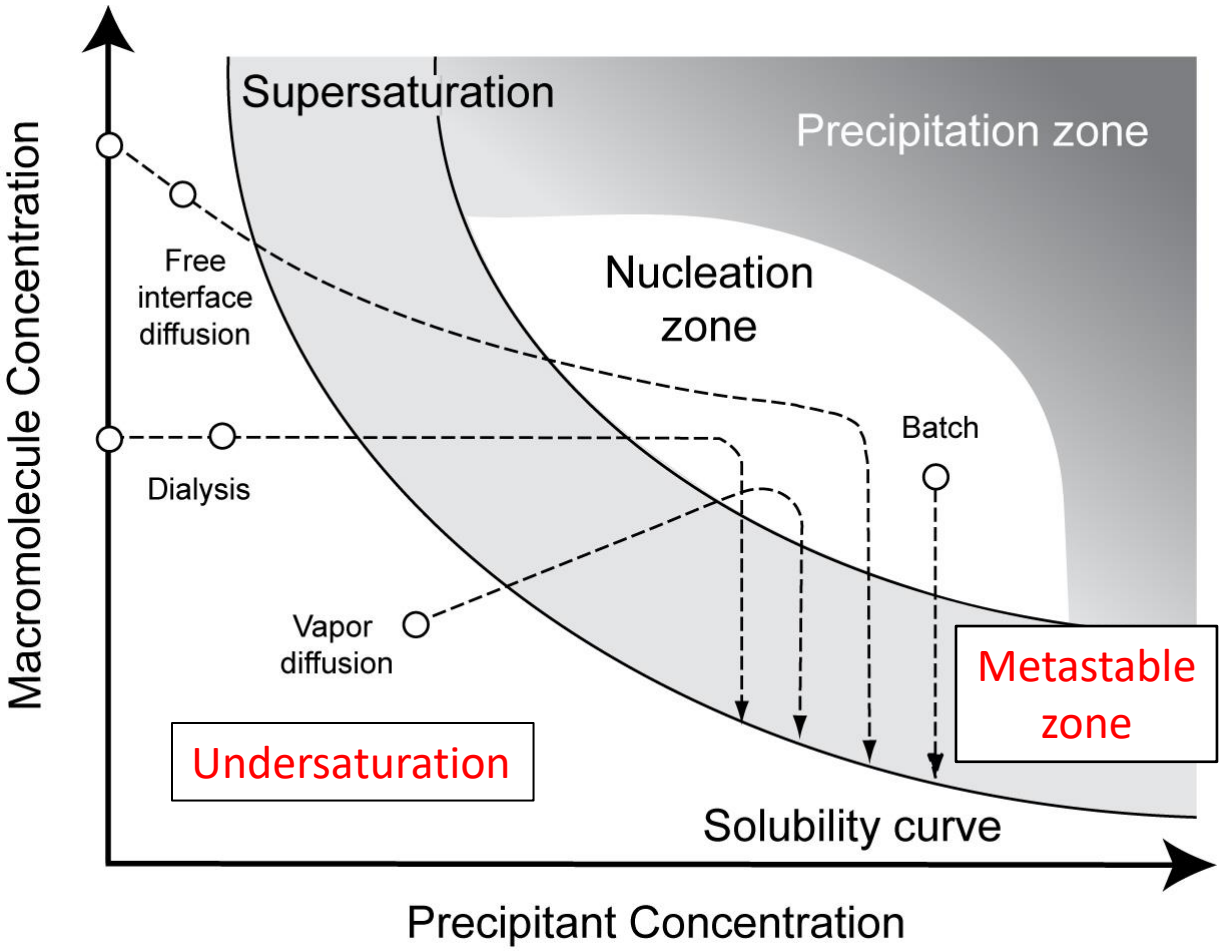


Outcomes



Clear

Simplified phase diagram for crystallization



Undersaturation and Metastable Zones

A clear drop in the undersaturated zone looks identical to a clear drop in the metastable zone.

But the two are decidedly different thermodynamic states. The former provides a boundary for crystallization optimization while the latter a central point for optimization.
Clear drops in isolation provide limited information

To distinguish undersaturated solutions from metastable solutions those conditions that are undersaturated will largely slow clear drops in chemically related experiments, while those in metastable, precipitate or even crystals in closely related chemical conditions.

Never consider a single crystallization screening result in isolation

If you are in the Metastable Zone

The solution conditions are at equilibrium with a single morphology of a crystalline phase.

Notwithstanding any chemical or physical changes the solution will remain stable with zero probability of nucleation and an infinite induction time for nucleation.

The metastable zone can be defined by its borders. The lower border is a loci of points called the *solubility curve,* painstakingly measured for a limited number of proteins. The upper boundary of the metastable zone is set by a limit where the probability of a nucleation event is certain with zero nucleation lag time.

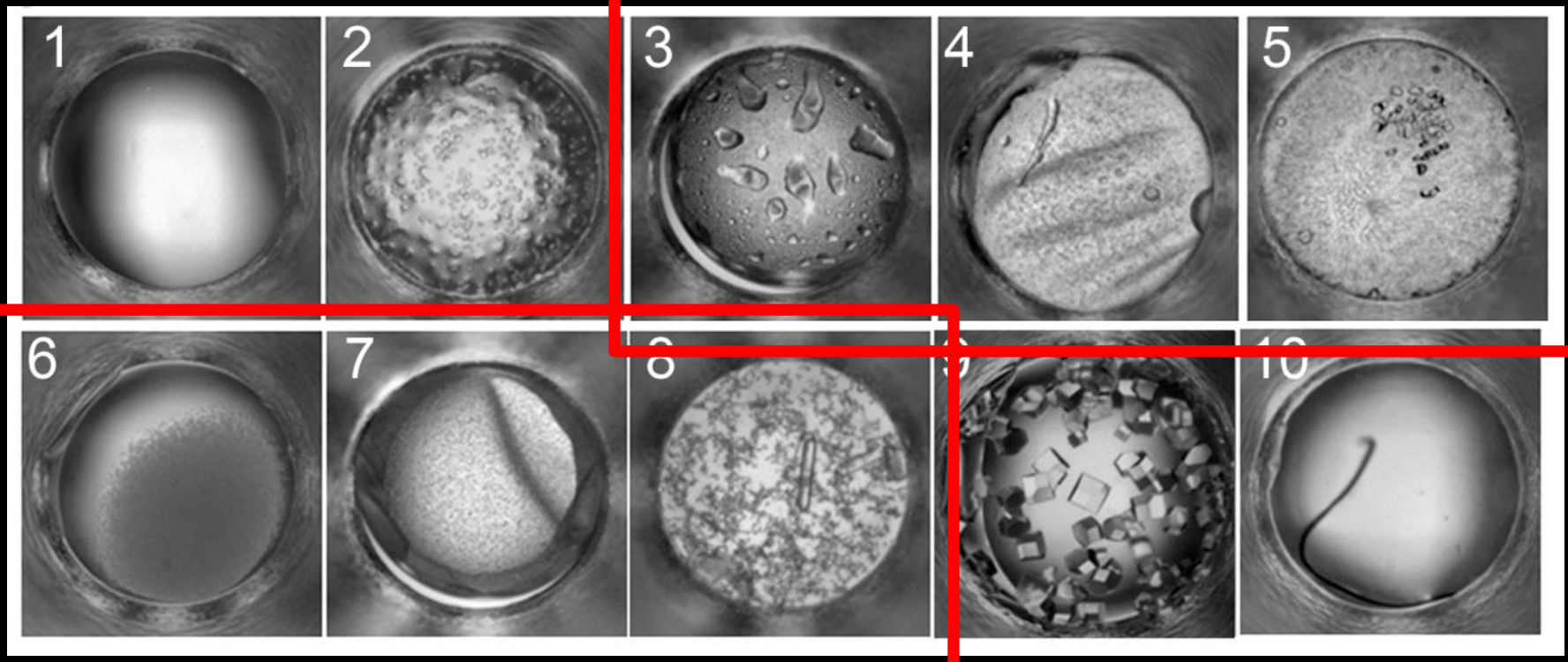
Crossing the upper boundary you would experimentally observe instantaneous, spontaneous, homogeneous nucleation. This is defined as the *supersolubility curve*.

If you are in the Metastable Zone

If a crystallization experiment appears clear, and the chemical conditions are very similar to conditions in the labile zone, then it is likely that the drop is at or near metastable supersaturation.

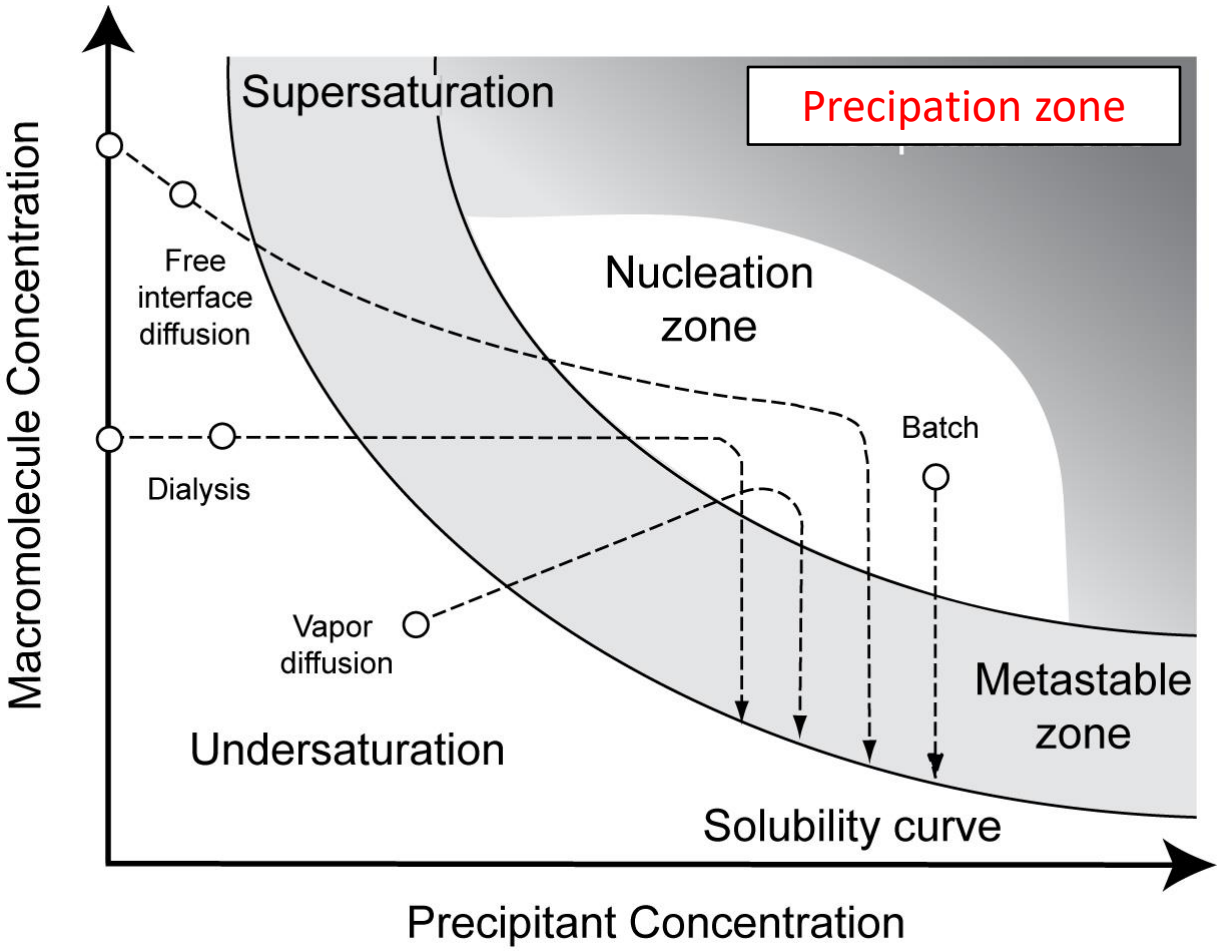
Seeding methods can be applied very effectively to crystallization experiments falling in this zone.

Outcomes



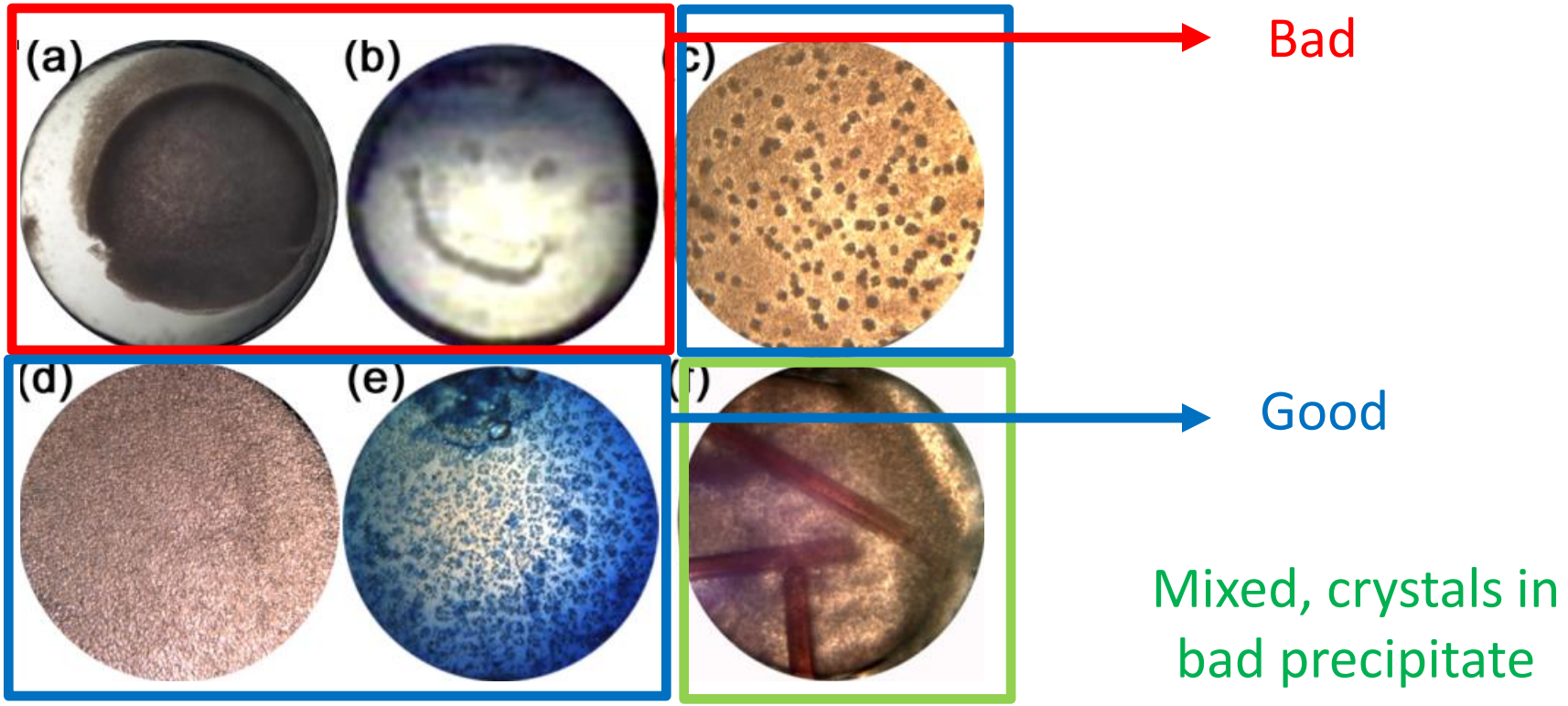
Precipitate (sometimes with crystals)

Simplified phase diagram for crystallization



Precipitate

Precipitation can be of two forms, bad (typically amorphous) or good (microcrystalline).



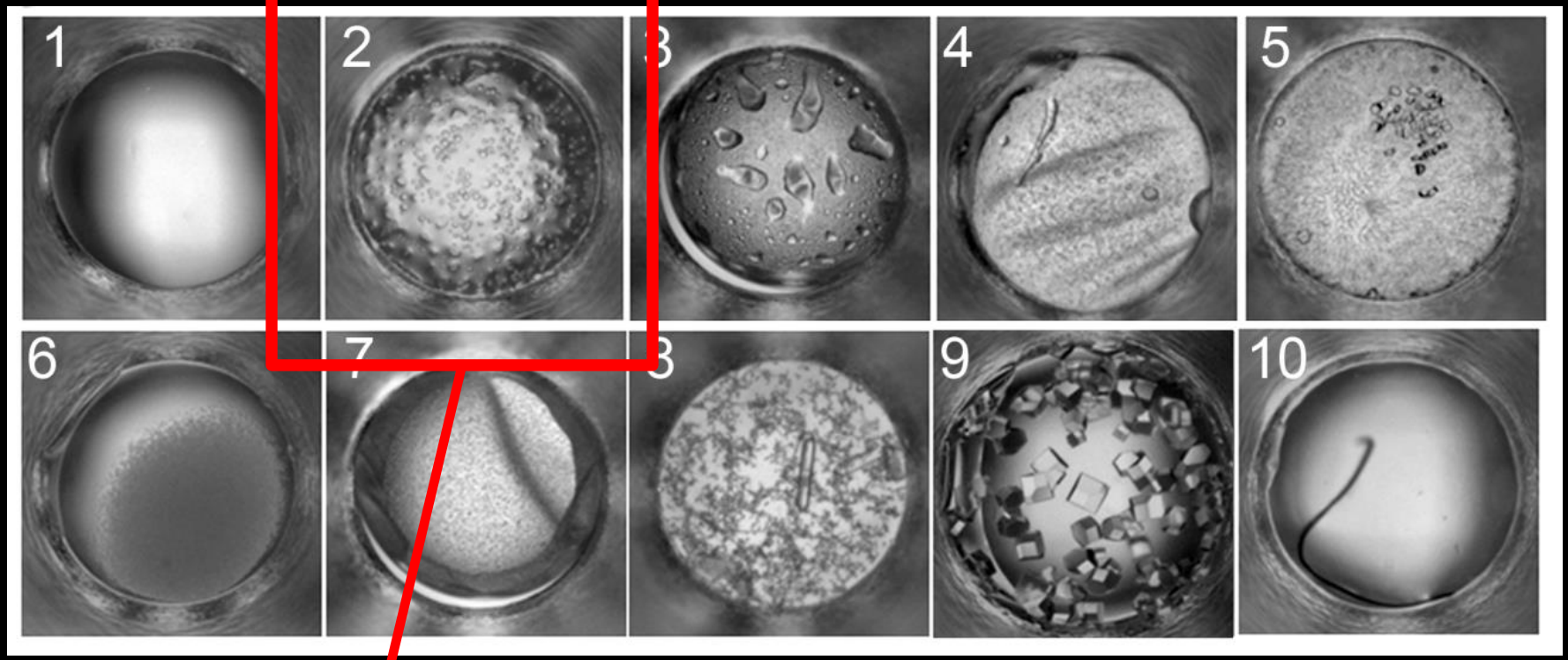
Precipitate

Precipitation indicates conditions where supersaturation is many times beyond the level required for growth of single crystals.

Amorphous precipitate can be brownish in color, frequently has a skin on the drop, will not re-dissolve, will not absorb dye and will not act as a successful seed. This type of precipitation occurs due to non-native protein aggregation.

Crystalline precipitation is protein aggregation where the chemical environment permits the native conformation to remain intact *i.e.* those likely to lead to a crystal with minimal optimization effort. Crystalline precipitates are patterned, can have a sandy appearance, typically but not always show bi-refringence (depending on the orientation and symmetry of the crystals), will re-dissolve, will absorb dye and can successfully act as seeds.

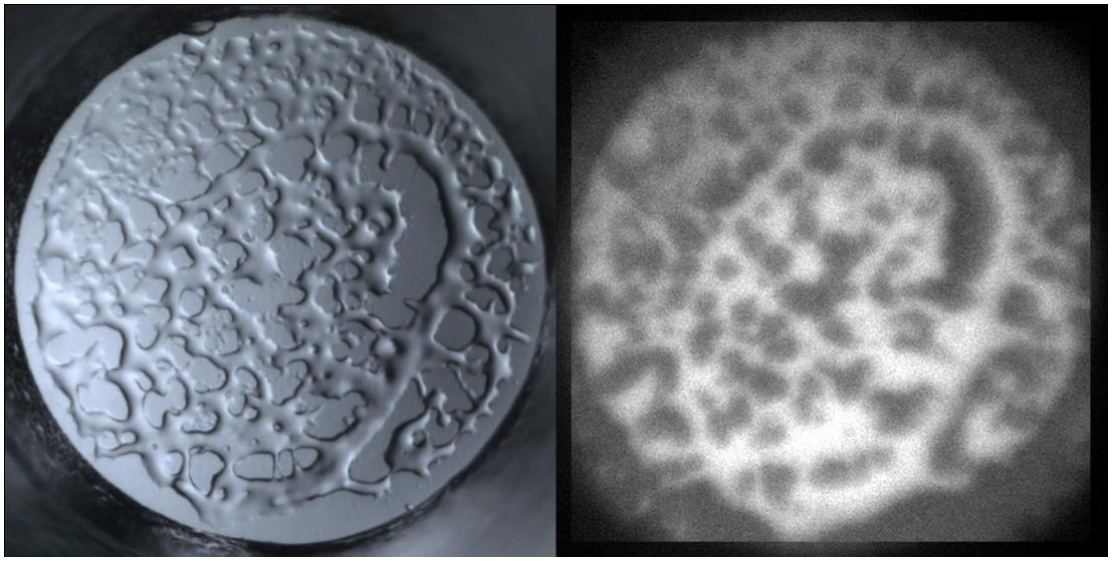
Outcomes



Liquid-Liquid phase separation

Liquid-Liquid phase separation

This is not shown on the typical phase diagram (it is lost in the simplified form).



Liquid-liquid phase separation is often seen as drops within drops, cloud-like patterns of liquid within the drop, or a similar appearance to a shaken bottle of oil and vinegar, the drop can even have a dimpled appearance similar to the surface of a golf ball

Liquid-Liquid phase separation

The formation of immiscible liquid-liquid (L-L) phase separation in the metastable region of the phase diagram occurs where there are short range, and/or highly anisotropic interactions between protein molecules.

If one phase is protein-rich and the other protein-poor (use UV observation to check) , then the system is very close to conditions that have the potential to produce crystals.

Temperature (and supersaturation) can drive crystallizaton. Crystals will sometimes form from the dense liquid phase without intervention.

As is the case with metastable conditions, this protein-rich immiscible liquid phase can be used for seeding

Liquid-Liquid phase separation - Temperature

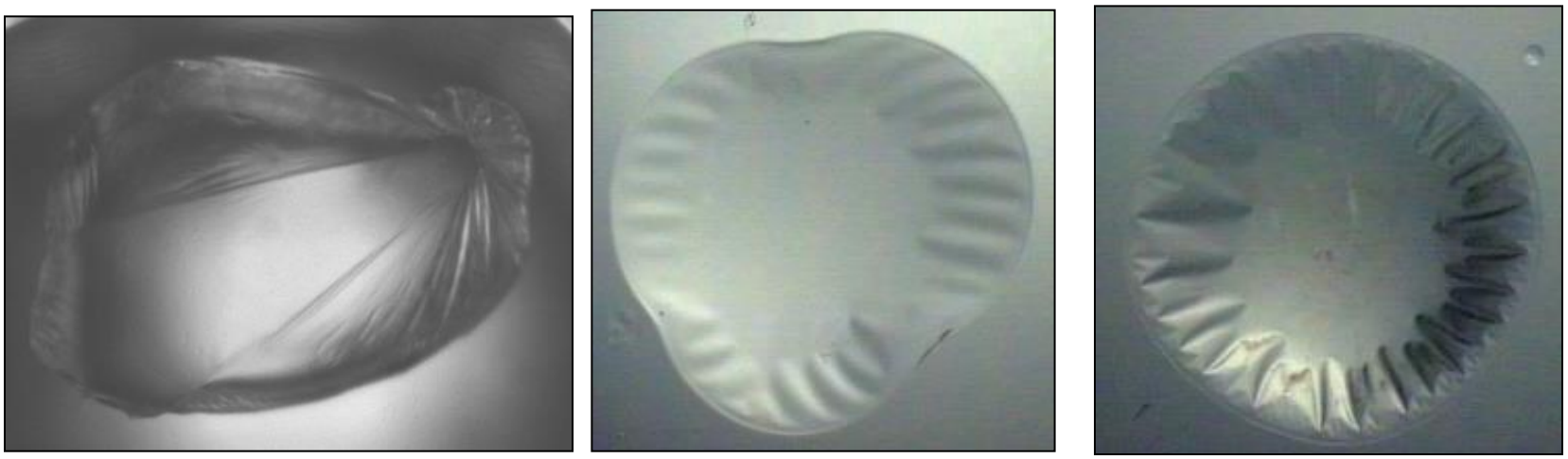
Mcromolecular solubility is dictated by the the protein and its chemical environment.The same macromolecue can have increased solubility at higher temperatures in one chemical environment, or higher solubility at lower temperatures in a different chemical environment.

If the protein/solvent is more soluble at higher temperatures and liquid-liquid phase separation is seen, then decreasing the temperature will drive the system towards a higher level of supersaturation and vice versa.

In these cases driving the system towards a higher level of supersaturation (by lowering or raising temperature), increases the attraction between protein molecules and promotes crystallization.

Skin formation

Again, this is not shown on the typical phase diagram.

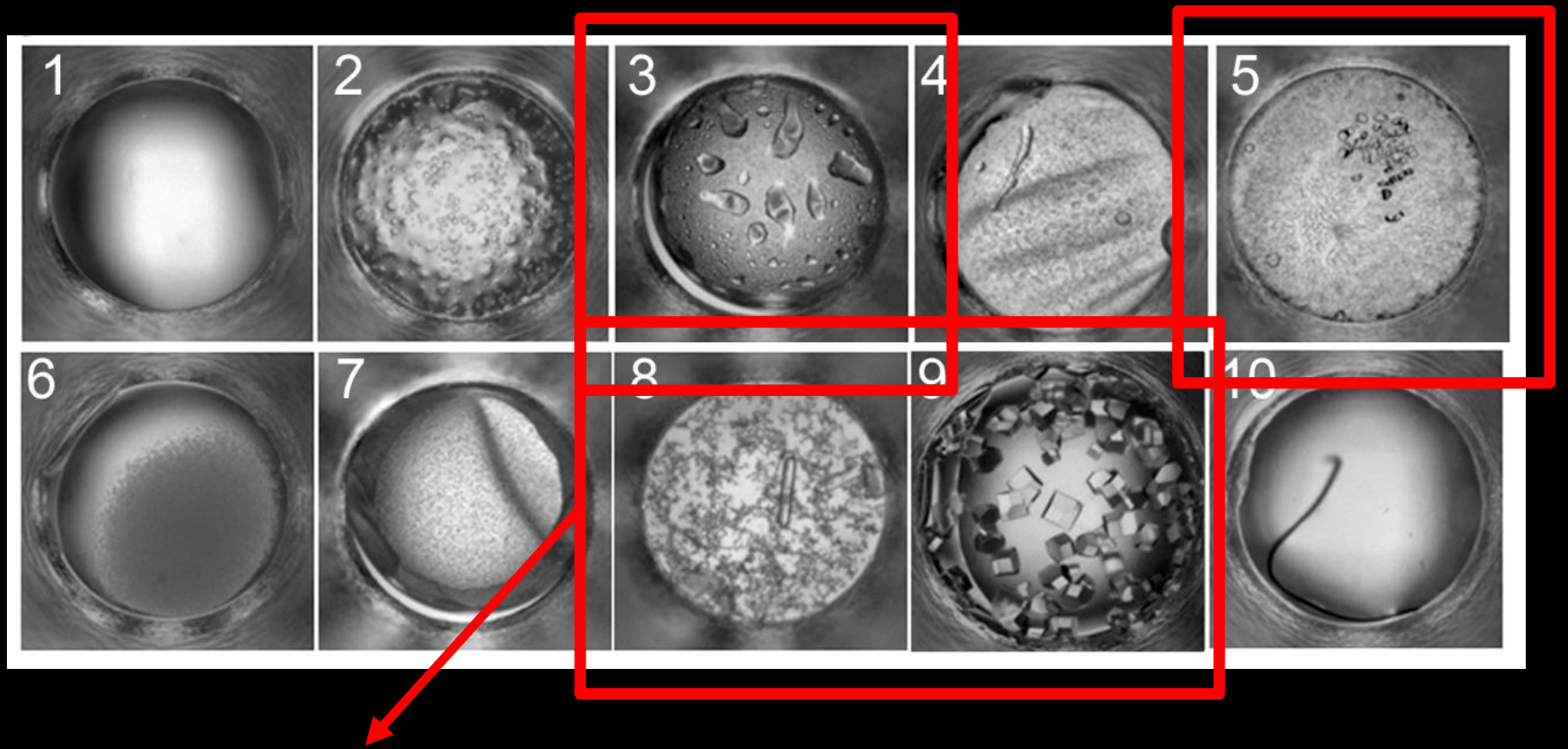


Skin formation is a thin layer of denatured protein that can be removed from the drop. The skin is a form of interfacial adsorption of the protein onto the interface whether it is solution/oil or solution/air, or solution/surface such as the plastic and glass materials that typically support the protein drop.

Avoiding skin formation

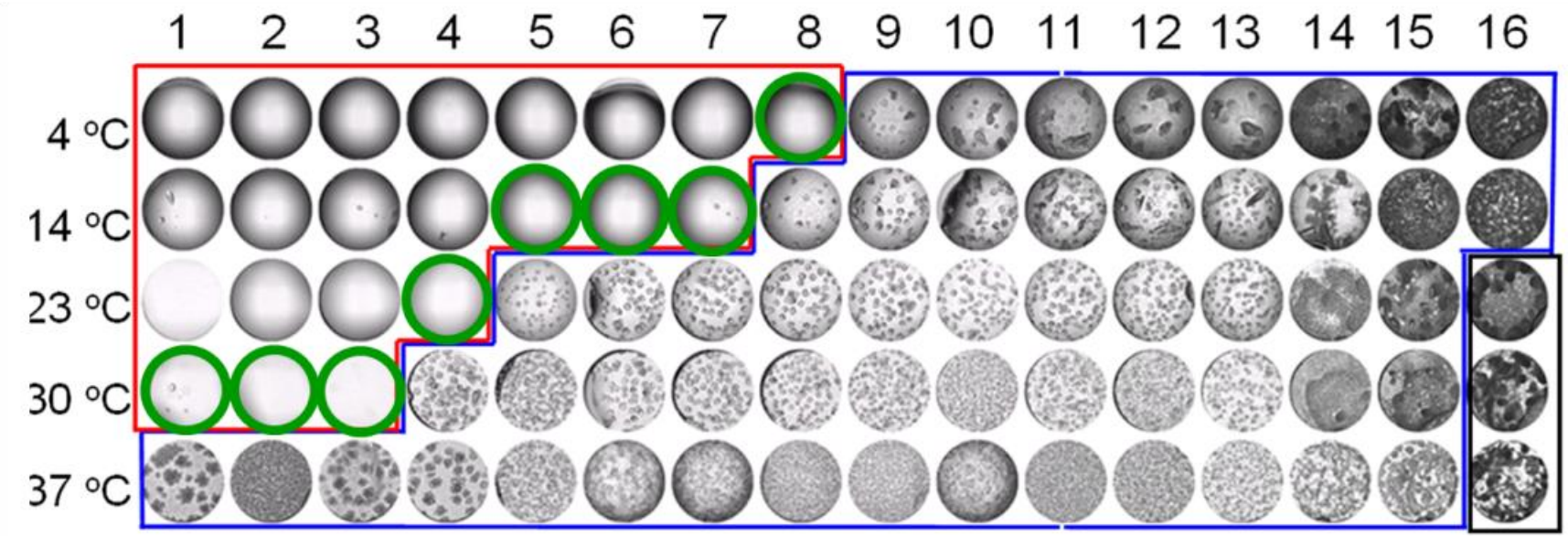
- There is a difference between air-water and air-oil interfacial adsorption phenomena.
- Change from vapor diffusion to batch can help.
- The protein may be denaturing – cool conditions.
- Oil layered over the top can help (slows the sample degradation processes (e.g. if oxidizing) and changees interfacial properties).
- Silicon grease and Fluorinert have been deposited onto a surface to float the crystallization drop to prevent crystals from adhering. While intended to ease removal of the fragile crystals that form on the substrate's surface, these protocols also change interfacial properties which can affect crystallization.

Outcomes

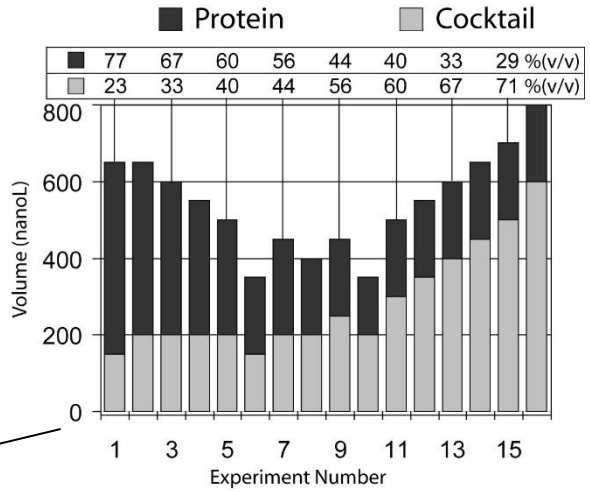


Crystals – success?

Optimization

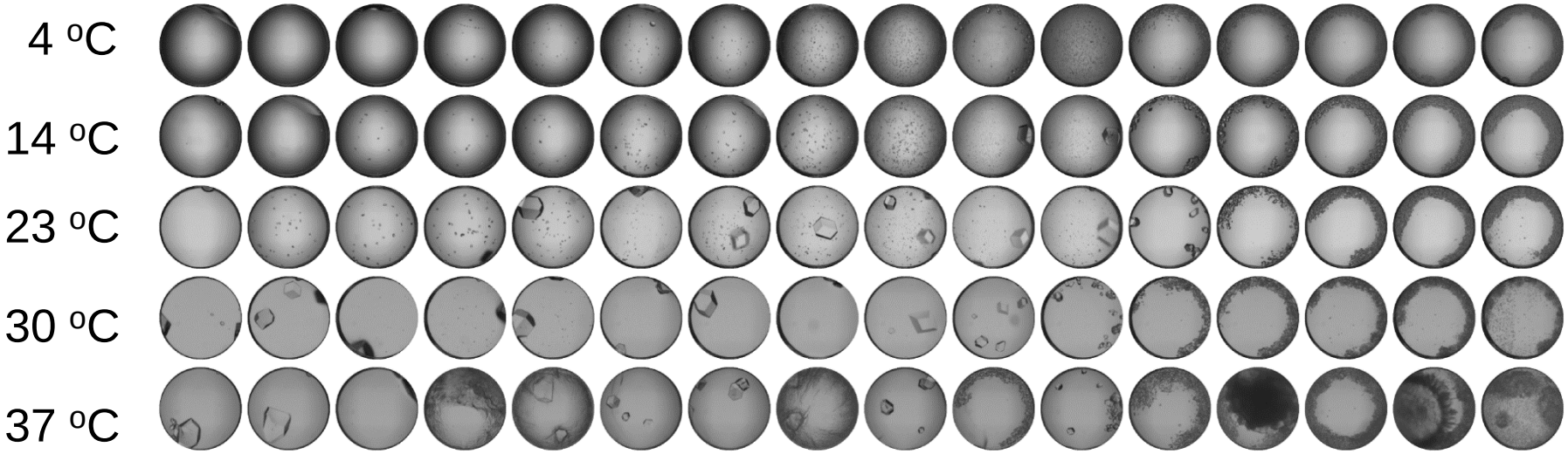


Varying the protein:precipitant ratio as a function of temperature can reveal the phase diagram and optimization strategies.

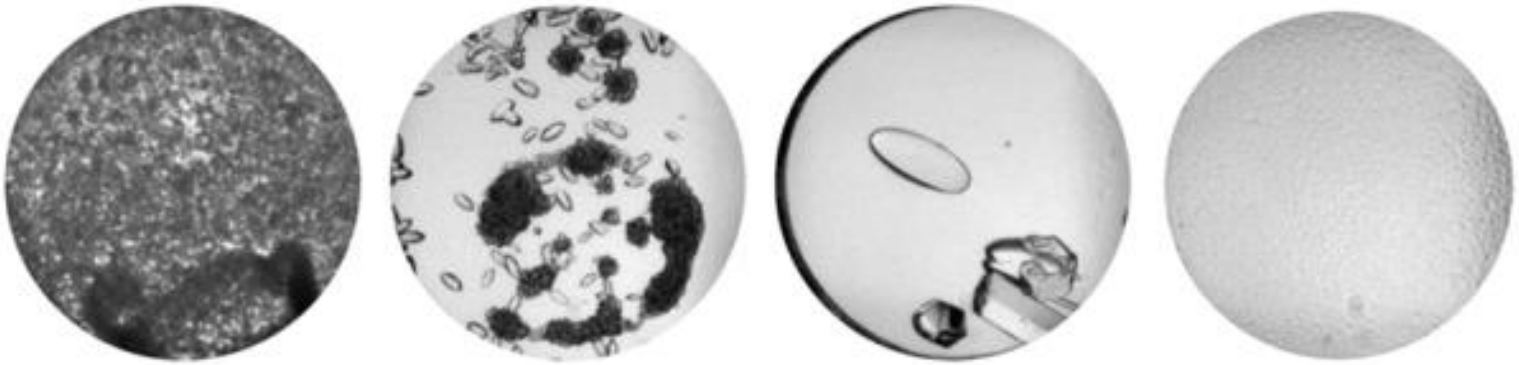


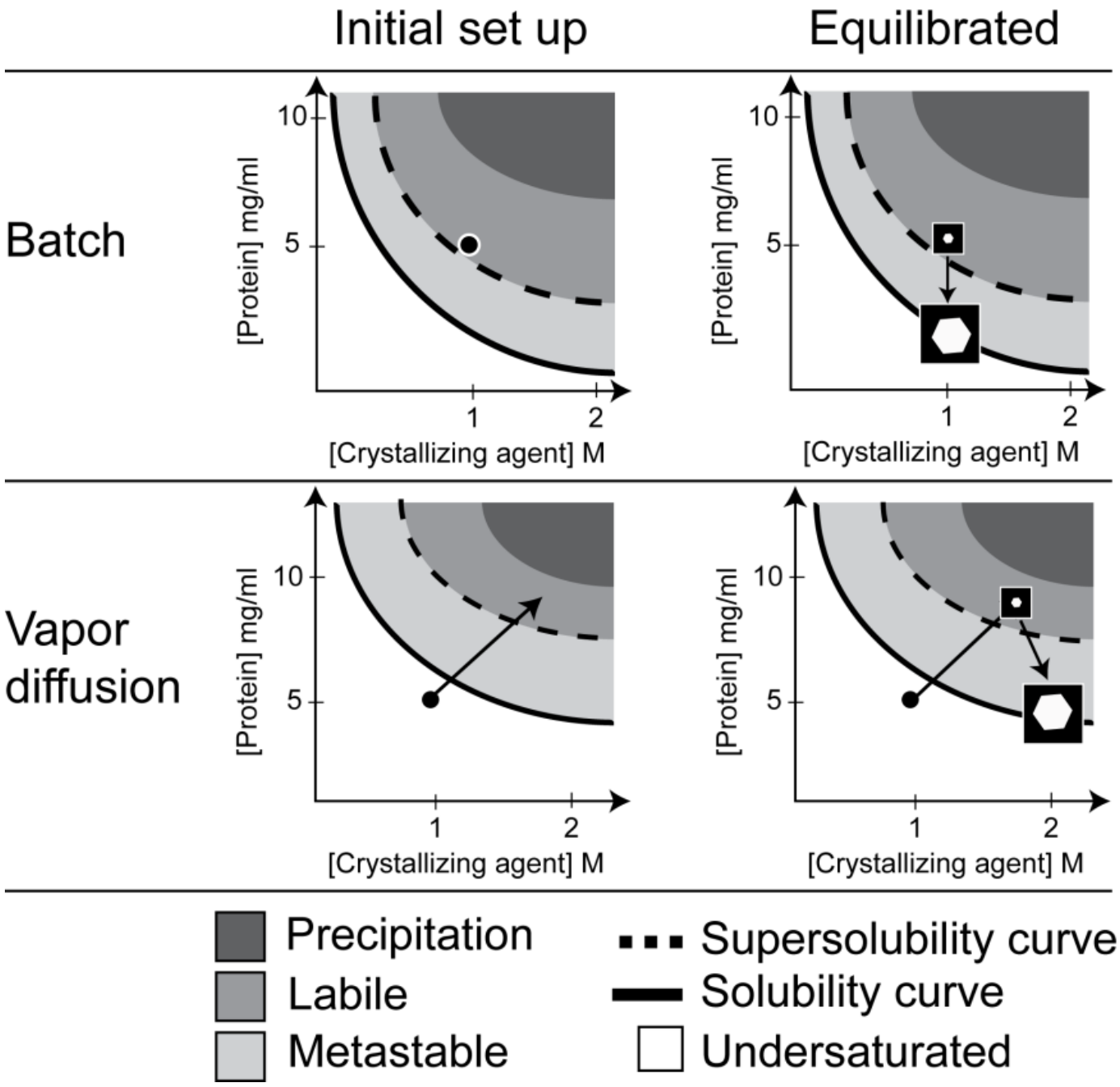
$V_{sample} > V_{cocktail}$

$V_{sample} > V_{cocktail}$



Optimization – try pH




Initial set up
Equilibrated
Batch
Vapor diffusion
[Protein] mg/ml
10
5
1
2
[Crystallizing agent] M
Precipitation
Labile
Metastable
Supersolubility curve
Solubility curve
Undersaturated

Time

The time required for the first observable crystals to form is important.

If crystals appear immediately or almost immediately then the system is at labile supersaturation from the onset. If crystals are observed shortly after the initial set up, equilibration was not required to drive the system to a state of labile supersaturation. Regardless of the crystallization method, the experiment is essentially a batch experiment.

If there is difficulty reproducing the crystals, experiments set up under seemingly identical conditions produce outcomes that are often microcrystalline, or otherwise unsuitable for structural solution, then rapid onset of supersaturation may be occurring.

Decreasing the starting concentration of the precipitating agent, or the protein can be beneficial, especially for vapor-diffusion experiments, where this would provide a starting point for the experiment that is not sufficiently supersaturated for crystallization at the onset, but rather approaches supersaturation slowly.

Delayed Crystallization

When a crystallization event takes several months, it can be caused by chemical or physical changes to the protein e.g. *in situ* proteolysis, chemical changes such as the loss of bound metal ion or co-factor, or a change in the protein's conformation.

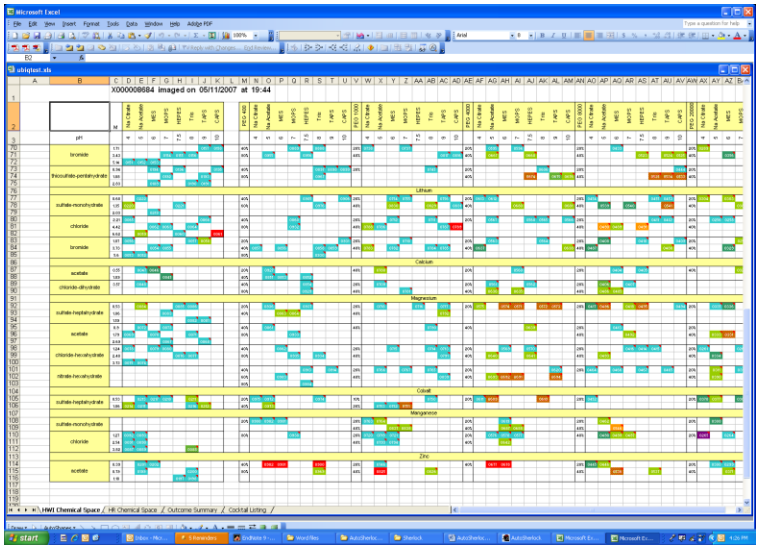
The experiment drop can simply become more concentrated with respect to non-volatile components, as volatile components evaporate the solutes that remain in increase in relative concentration. This will, under the appropriate biochemical conditions, lead to supersaturation, nucleation and crystal growth.

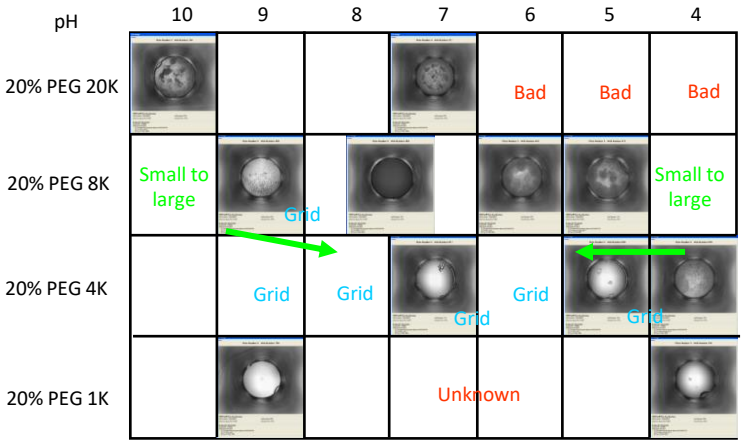
Most plastics, especially polystyrene, the material from which many crystallization plates are fabricated, are water permeable. No matter how good a seal is used on the plate, unless stored in a humid environment, water will slowly evaporate through the plastic plate, or plastic seal, to simultaneously concentrate the drop and reservoir solutions. The relative change in concentration per unit volume of water loss will be more significant at lower volumes. This should not be considered detrimental, but it is something that one should be aware of when trying to reproduce crystal hits from plates that have been incubated at room temperature for extended periods of time. The evaporative loss is mitigated at lower temperatures, *e.g*. 4^{o}C.

In some cases amorphous precipitate may be observed, followed over time by a small crystal which slowly grows while the precipitate recedes. This is described by Ostwald's rule of stages "When leaving a given state and in transforming to another state, the state which is sought out is not the thermodynamically stable one, but the state nearest in stability to the original state." This means that the least soluble solid state will be the first to come out of solution. The next to appear is not necessarily the most thermodynamically stable, but that closest in energy to the first. This process continues, with a series of intermediate metastable forms, whose appearance is dependent on kinetics and not solely thermodynamics, until the formation of the most thermodynamically stable state.

Another common example is when different morphologies of a crystalline protein are observed in a single drop where they can co-exist for some time. However over time, one crystal form, the most thermodynamically stable form, will increase in size at the expense of the other. Ostwald's rule of stages is not the same as Ostwald ripening. Ostwald's rule of stages transitions between different states to decrease the free energy of the system, while Ostwald ripening will decrease the surface free energy of a system of single small crystals through mass-controlled transport to larger crystals of the same form.

Chemical space mapping





research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The application and use of chemical space mapping to interpret crystallization screening results

Edward H. Snell,[a,b]* Ray M. Nagel,[a] Ann Wojtaszcyk,[a] Hugh O'Neill,[c] Jennifer L. Wolfley[a] and Joseph R. Luft[a,b]

[a]Hauptman–Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, NY 14203, USA, [b]Department of Structural Biology, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA, and [c]Center for Structural Molecular Biology, Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Correspondence e-mail: esnell@hwi.buffalo.edu

Received 17 June 2008
Accepted 7 October 2008

Macromolecular crystallization screening is an empirical process. It often begins by setting up experiments with a number of chemically diverse cocktails designed to sample chemical space known to promote crystallization. Where a potential crystal is seen a refined screen is set up, optimizing around that condition. By using an incomplete factorial sampling of chemical space to formulate the cocktails and presenting the results graphically, it is possible to readily identify trends relevant to crystallization, coarsely sample the phase diagram and help guide the optimization process. In this paper, chemical space mapping is applied to both single macromolecules and to a diverse set of macromolecules in order to illustrate how visual information is more readily understood and assimilated than the same information presented textually.

computer programs

Journal of
Applied Crystallography
ISSN 0021-8898

Received 17 June 2008
Accepted 9 September 2008

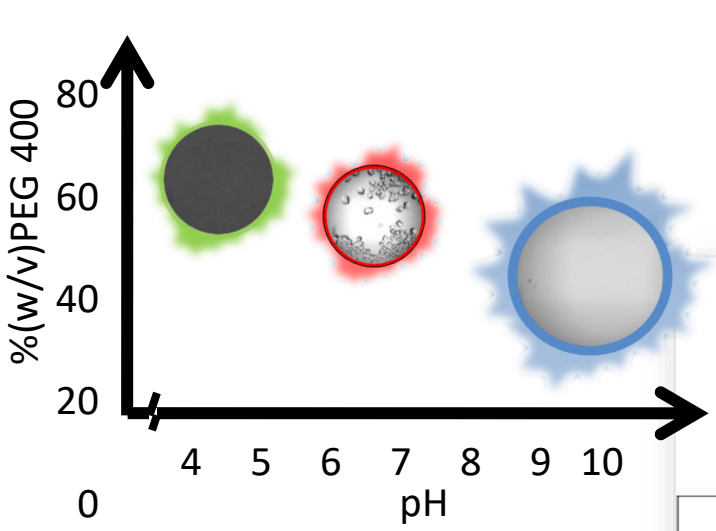
AutoSherlock: a program for effective crystallization data analysis

Raymond M. Nagel,[a] Joseph R. Luft[a,b] and Edward H. Snell[a,b]*

[a]Hauptman–Woodward Medical Research Institute, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA, and [b]Department of Structural Biology, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA. Correspondence e-mail: esnell@hwi.buffalo.edu

A program, *AutoSherlock*, has been developed to present crystallization screening results in terms of chemical space. This facilitates identification of lead conditions, rational interpretation of results and directions for the optimization of crystallization conditions.

Chemical space mapping (analysis)



Un-sampled
Crystals
Clear
Precipitate

		PEG 400	Na Citrate	Na Acetate	MES	Bis-Tris Propane	HEPES	Tris	TAPS	CAPS
			Lithium							
pH			4.2	5	6	7	7.5	8	9	10
bromide		20%								●
		40%	●		●			●	●	
		80%						●		

Chemical/Molecular Fingerprints (what you can do by looking at a complete crystallization screening data set)

Molecular Fingerprints

Molecular fingerprints are representations of chemical structures designed to capture molecular activity.

We use atomic properties and a SMILES string to capture six components:

1. Atomic number
2. Number of directly-bonded neighbors
3. Number of attached hydrogens
4. The atomic charge
5. The atomic mass
6. If the atom is contained in a ring

These components are calculated for the whole molecule in an iterative manner starting from an arbitrary non-hydrogen.

Example:
Sodium chloride, NaCl

Sodium [11,0,0,1,22.99,0]
Chlorine [17,0,0,-1,35.45,0]

Starting from Na two, properties are associated with Na and encoded by: (3,855,292,234,1) and (3,737,048,253, 1)*

One property is associated with Cl and encoded by: (2,096,516,726,1)

This information is stored in single integer with bits 3,855,292,234, 3,737,048,253 and 2,096,516,726 set to on.

* Rodgers and Hahn, J. Chem. Inf. Model. 2010, 50, 742-754

Cocktail Fingerprints

Cocktail fingerprints combine the molecular fingerprints and account for the molarity of each in the crystallization cocktail.

For example, consider a very simple example: 0.1 M sodium chloride and 0.1 M ammonium sulfate

Molecular fingerprint: Sodium chloride [(3855292234, 1),(3737048253, 1),(2096516726, 1)]
Ammonium chloride [(847680145, 1), (3855292234, 1),(2214760707, 1)]

Bit (3855292234, 1) is common in both so we set the bit count to 2 and multiply by the molar concentration

Cocktail fingerprint: [(3855292234, 0.2),(3737048253, 0.1),(2096516726, 0.1)
(847680145, 1),(2214760707, 0.1)]

The bits are stored in a single 64 bit number with the bit counts stored in a sequential array

Comparing Cocktail Fingerprints

Take a real example of two crystallization screening cocktails as stored in our database

Cocktail	Component	conc	unit	SMILES	MW	Density (g/cm^3)
C1249 pH 4.6	calcium chloride dihydrate	0.02	M	[Ca+2].[Cl-].[Cl-].O.O	147.0146	
	sodium acetate trihydrate	0.1	M	[Na+].[O-]C(=O)C.O.O.O	136.0796	
	mpd	30	%(v/v)	CC(O)CC(C)(C)O	118.1742	0.9254
C0160 pH 7.5	sodium chloride	4.48	M	[Na+].[Cl-]	58.4428	
	hepes	0.1	M	[O-]S(=O)(=O)CCN1CC[NH+](CC1)CCO	238.3045	

First convert all concentrations to molarity

Cocktail C1249 contains 30% (v/v) MPD. This is converted to 2.349 M. PEGs are more problematic as they can be polydispersive in which case the average molecular weight is used.

The cocktail fingerprint is calculated using the molecular fingerprint for each component and its molar concentration

$$F_k = \sum_{i=1}^{n} f_{ik}[c_i]$$

Where F_k is the cocktail fingerprint, i is the number of components, f the molecular fingerprint and c the concentration

An example of two cocktail fingerprints

```
C1249 = [(2245273601,2.35),(2214760707,0.02),(3537123720,4.70),(864942730,0.10),
 (1614748561,2.35),(786100370,2.35),(864666390,0.34),(3537119515,2.35),
 (3925650716,0.02),(2246728737,7.15),(864662311,4.70),(1582611257,2.35),
 (3737048253,0.10),(3855292234,0.04),(864942795,0.10),(2245384272,2.35),
 (3992738647,2.35),(1510323402,0.10),(248253150,2.35),(1542633699,2.35),
 (3219326737,0.10),(2246699815,0.10),(2355142638,2.35),(2245277810,2.35),
 (1542631284,2.35),(2096516726,0.10),(3545365497,0.10),(1510328189,0.10)]
C0160 = [(864942730,0.20),(951748626,0.10),(2143075994,0.10),(2227993885,0.10),
 (2968968094,0.40),(192851103,0.10),(2092489639,0.10),(2604889258,0.10),
 (2880892204,0.10),(1535166686,0.10),(4226502584,0.20),(825302073,0.10),
 (3855292234,4.48),(1412710081,0.20),(2828037323,0.10),(2228063684,0.20),
 (569967222,0.10),(2105180129,0.10),(2803848648,0.20),(4055698890,0.10),
 (864942795,0.10),(2808066764,0.20),(2245384272,0.40),(4023654873,0.10),
 (3336755162,0.10),(999334238,0.10),(1789200865,0.10),(864662311,0.10),
 (3737048253,4.48),(2096516726,4.48),(2257970297,0.10),(1634606847,0.10)]
```

Each is encoded in a single hashed number.

Comparing Cocktail Fingerprints

The Bray-Curtis dissimilarity measure is used to compute the dissimilarity.

$$BC(F_i, F_j) = \sum_k | F_{ik} - F_{jk} | / \sum_k | F_{ik} + F_{jk} |$$

This pH is incorporated along with the ability to weight individual components and the Cocktail Dissimilarity coefficient calculated.

$$CD_{coeff} = \frac{1}{sum(w)} \left(\left(\frac{| E(pH_i) - E(pH_j)}{14} \right) w_1 + BC(F_i, F_j) w_2 \right)$$

The Cocktail Similarity coefficient given by:

$$CS_{coeff} = 1 - CD_{coeff}$$

A real example with our
1,536 condition screen

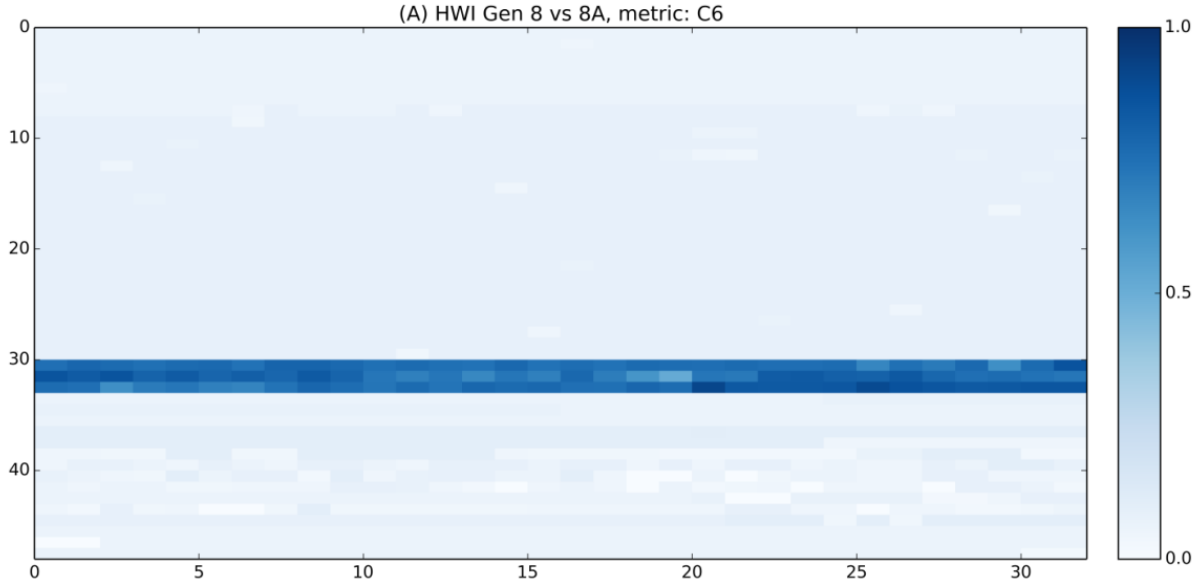
Cocktail similarity measures are not new.

We build on the original work by Janet Newman's in Melbourne, Australia who originated the concept of a similarity measure (termed C6) within crystallization to compare individual cocktails and different screening kits. (Newman J, Fazio VJ, Lawson B, Peat TS (2010) The C6 Web Tool: A Resource for the Rational Selection of Crystallization Conditions. Crystal Growth & Design 10: 2785-2792).

Our internal 1,536 screens are reformatted on a yearly basis to remove any conditions that produce salt crystals, to incorporate the latest screening developments, and building on internal research into crystallization processes.

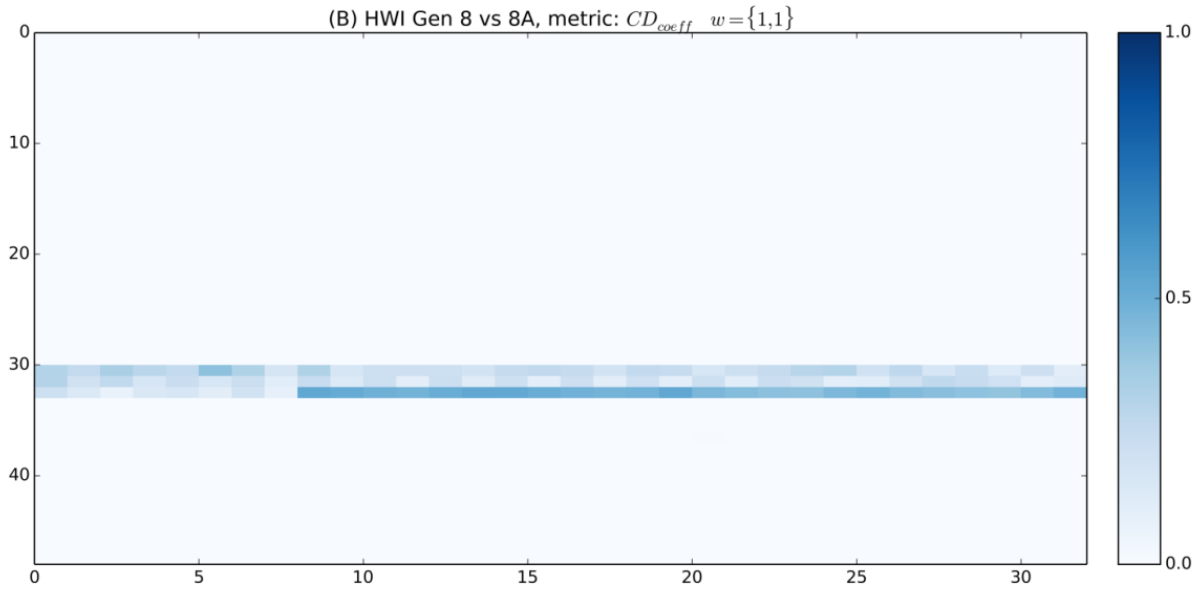
In this example we apply both the C6 and our new similarity measure to two generations of screen where 96 conditions have been replaced with a new commercially available screen/

The C6 metric color coded according to dissimilarity (0 is identical, 1 is most dissimilar)



The new dissimilarity metric.

Note that the only change in the screen was replacing 96 conditions



Clustering then using a hierarchal display

The Dissimilarity Measure Over the Whole Screen

Aspects of the screen design are clearly seen

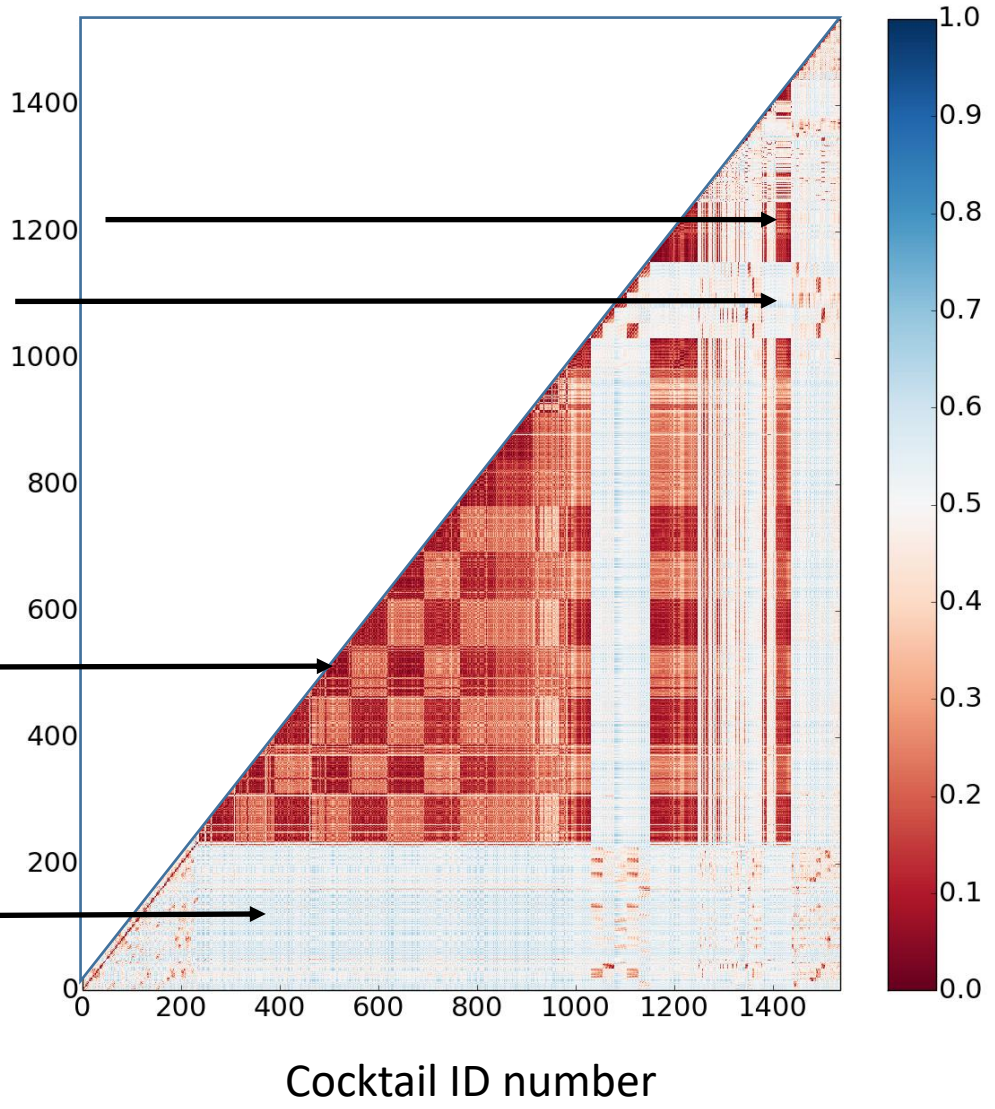
Hampton Research PEG/Ion screen

Hampton Research Silver Bullets

PEG based conditions sampling different molecular weight PEGS at two concentrations

Salt based screens

The scale is normalized to the most dissimilar chemical conditions

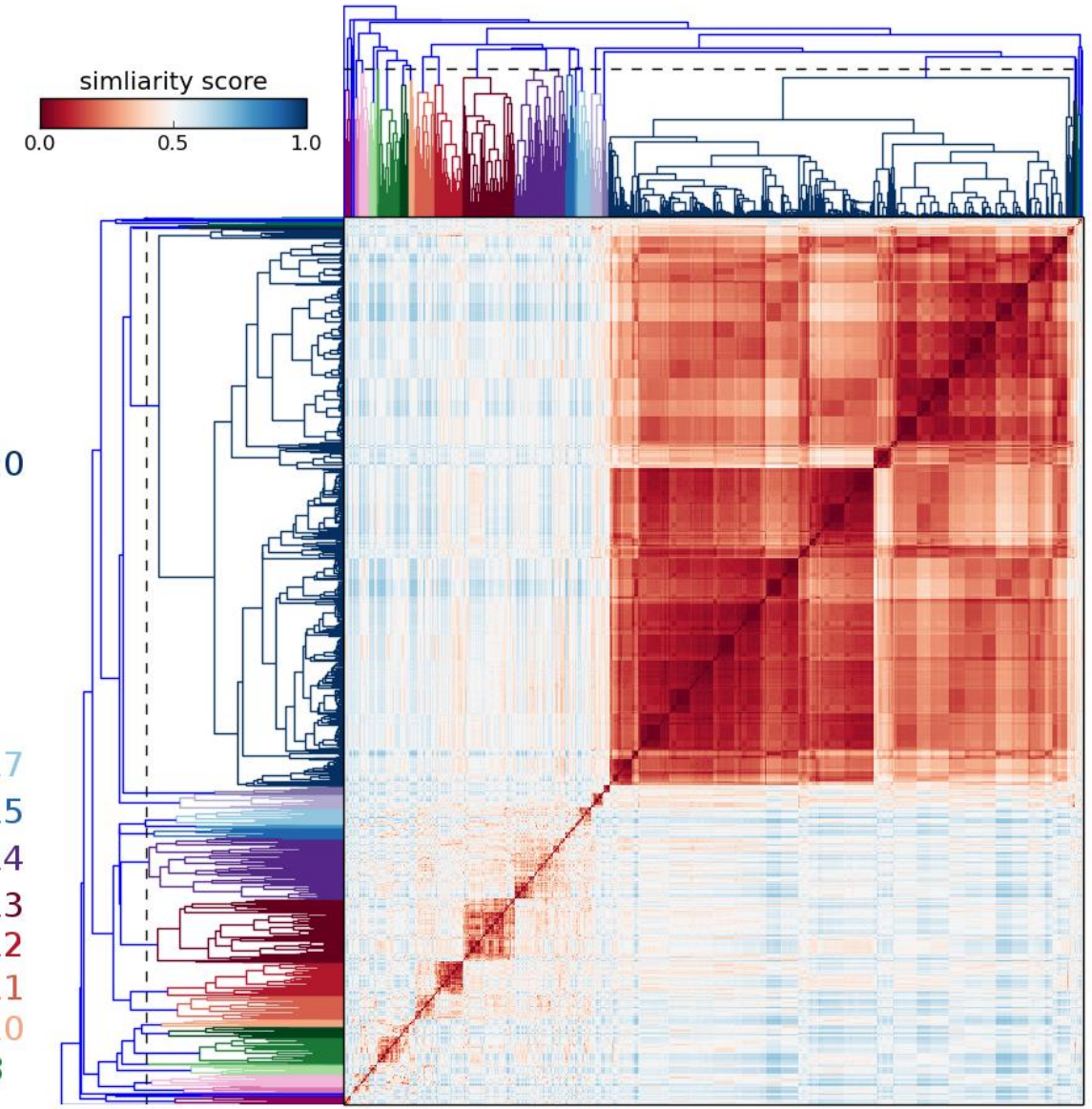


Automatic Clustering of the Results

Hierarchical clustering using a default max cophenetic distance cutoff of one standard deviation identified 28 clusters.

PEG based conditions →

Salts with different anions and cations →



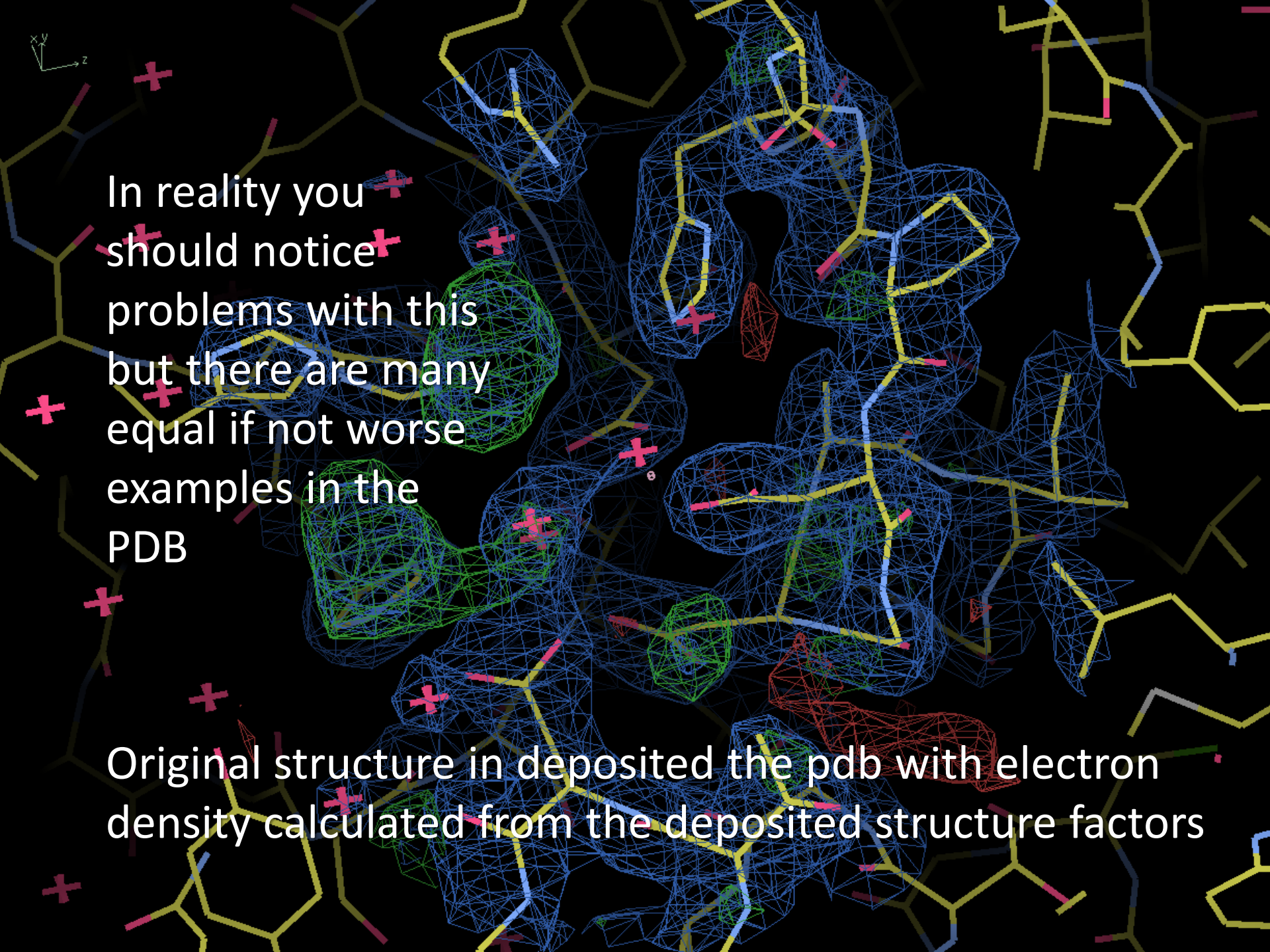
So how do we make use of it?

Cocktail similarity measures are not new.

BfR192, is a 343 residue protein with a molecular weight of 39.77 kDa. For crystallization screening the protein was prepared at 7.4 mg/ml in a 5 mM DTT, 100 mM NaCl, 10 mM Tris-HCl, pH 7.5, 0.02% NaN_3 buffer.

Several potential crystallization conditions for BfR192 SelMet labeled protein were identified

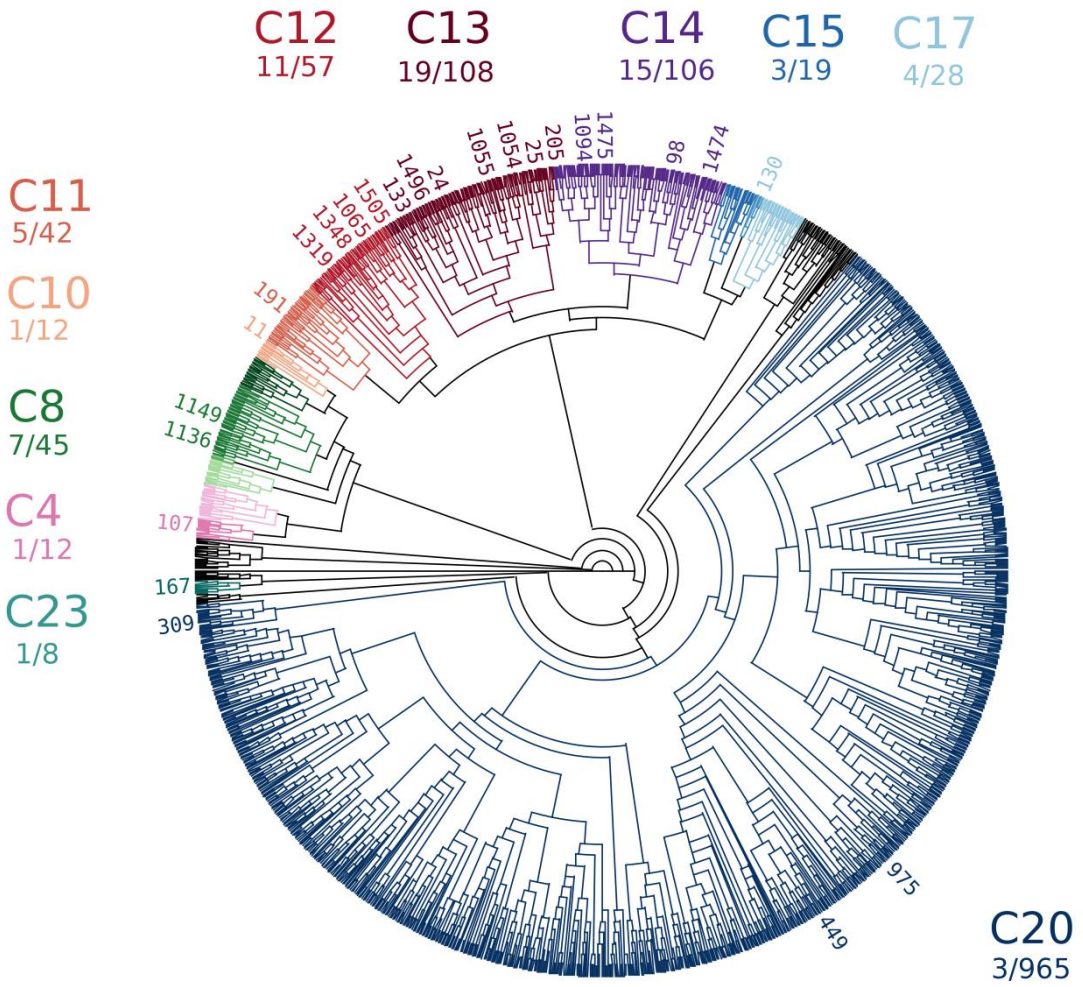
The optimized conditions for crystallization combined 5µl of the protein at 7.4 mg/ml concentration was mixed with the precipitant containing 320mM potassium acetate, 100 mM sodium acetate, pH 6.5 in 1:1 ratio. Crystals appeared in one week.


In reality you should notice problems with this but there are many equal if not worse examples in the PDB
Original structure in deposited the pdb with electron density calculated from the deposited structure factors

Overlaying Crystal Hits on the Cocktail Clustering

Conditions showing crystal hits are given for each cluster along with the total number of cocktails in that cluster.

A selection of cocktails that showed hits are listed on the outside of the dendogram. For clarity not all hits are shown

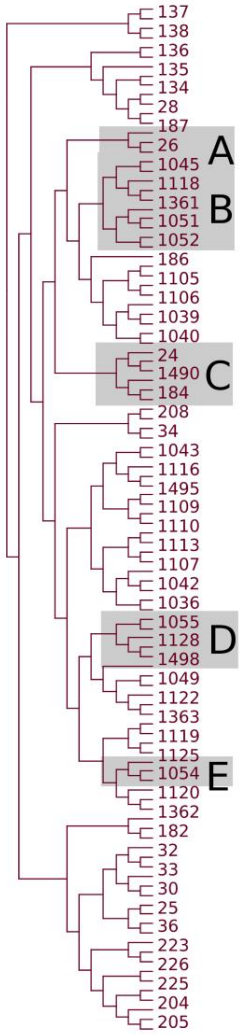


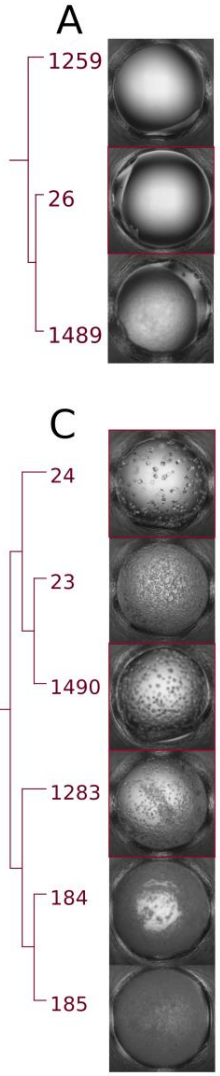
Cluster 20, PEG based, only 3 hits

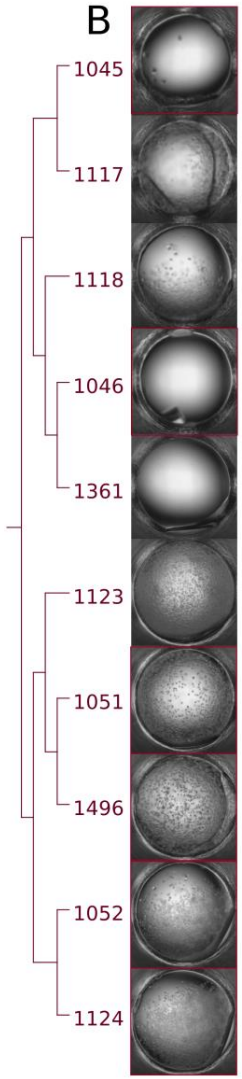
Cluster	Total	Hits	% hits	Sodium %	Potassium %	Phosphate %
All cocktails						
	1536	70	4.5	47	24	16
All crystal						
	70	70	100	70	27	30
Clusters with crystals						
C13	108	19	17.6	73	72	100
C14	106	15	14.2	65	21	0
C12	57	11	19.3	16	2	0
C8	45					
C11	42					
C17	28					
C20	965					
C15	19					
C23	8					
C4	12	1	8.3	83	25	0
C10	12	1	8.3	75	25	0

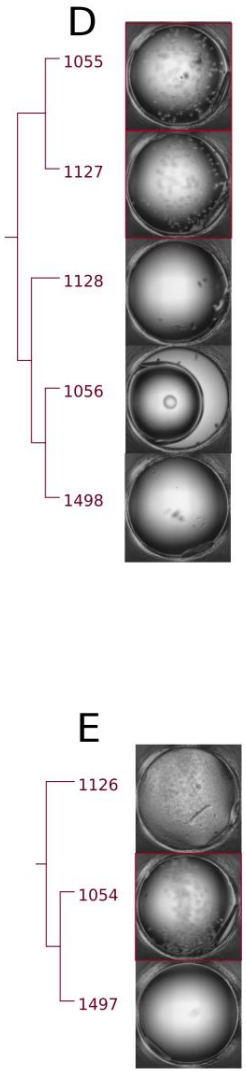
Cluster 13 proved interesting in that sodium is present in 73% of the conditions versus 47% for the 1536 condition screen overall, potassium is present in 72% of the conditions verses 24% overall and finally phosphate is present in 100% of the conditions versus 16% overall. This suggests a strong influence of these components in crystallization in this cluster.

Zoom in on Cluster 13



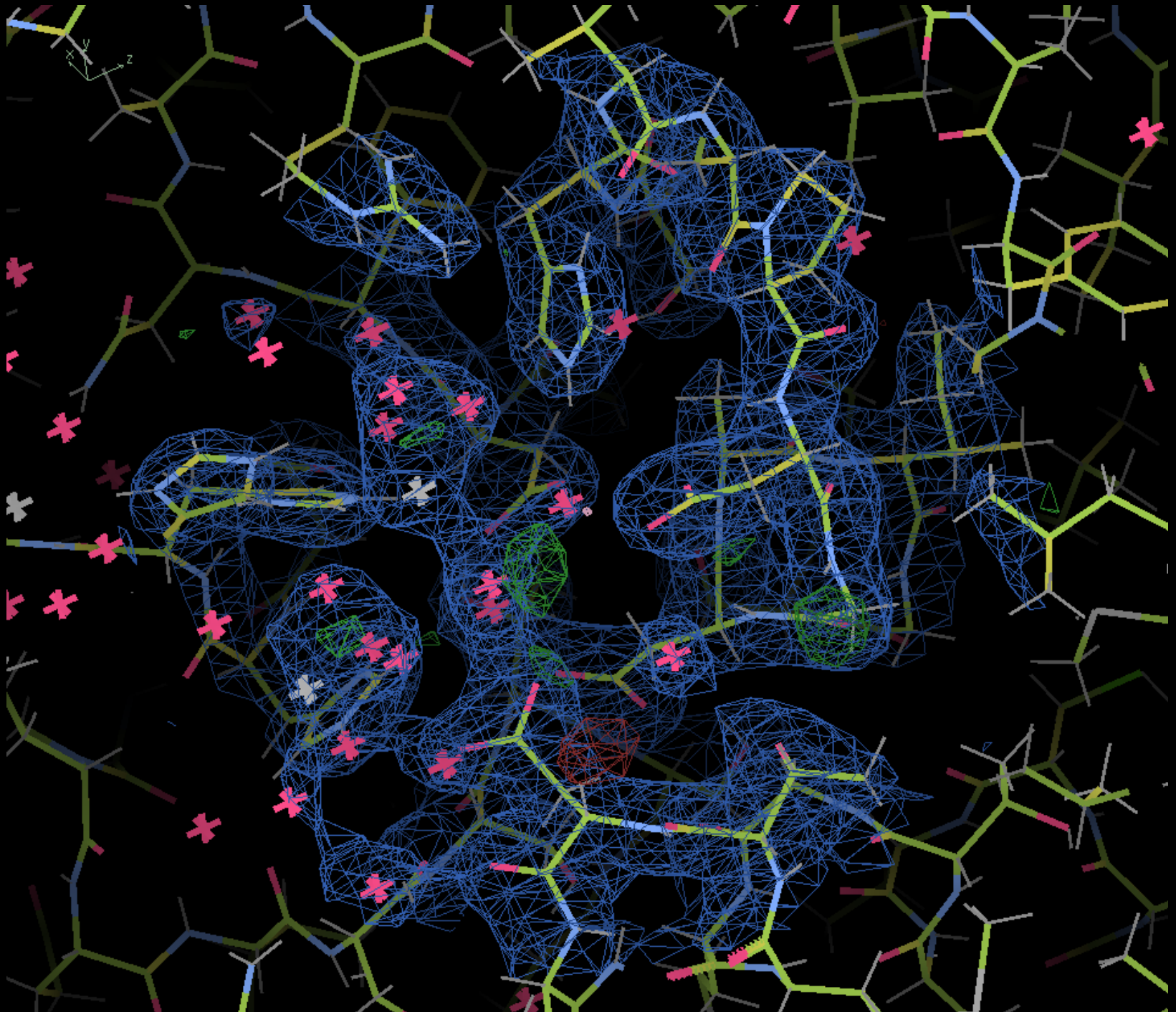




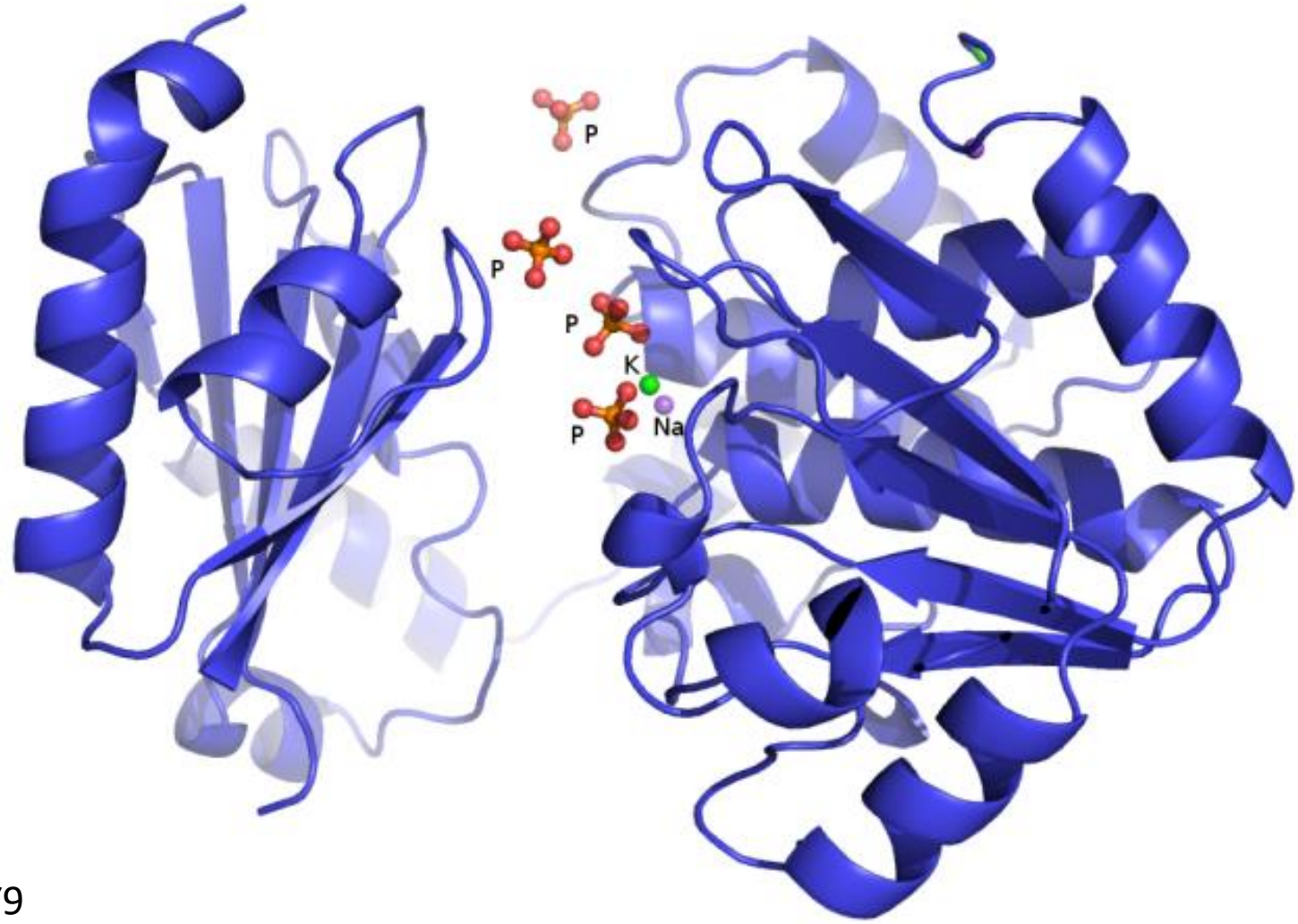


Identifies a pipette error

Clustering samples the phase diagram



A Revised Structure Illustrating Mechanism



PDB 4PY9

Biological implication of the phosphates identified

- The structure consists of two domains (N-terminal domain; residues 2 -212 and C-terminal domain residues 217-343) which are connected by a short loop – seen in the initial structure

- The N-terminal domain contains the DHH (Asp224-His225-His226) motif and the C-terminal domain contains a glycine-rich (GGGH-Gly308-Gly309-Gly310-His311) phosphate binding motif – seen but not identified in the initial structure.
- Three of the phosphates (presumably carried with the protein), and the potassium and the sodium ion are bound in the cl[illegible] between the two domains
- The phosphate ions interact wi[illegible]7.
- The location of the phosphate [illegible] might anchor in this pocket.
- The putative active site has fea[illegible]es which are involved in binding t[illegible]
- The possible roles of the active[illegible]on and polarization of the phosph[illegible] nucleophilic attack.
- The space around the phospha[illegible]s.

The important point here is not the details of the new information but that this information was obtained after the correct ligands were identified. Potential function and mechanism was revealed. While on could argue that these could have been identified earlier many examples in the PDB have ambiguous atoms – we have explored only a small sample of structures and seen problems in many of them.

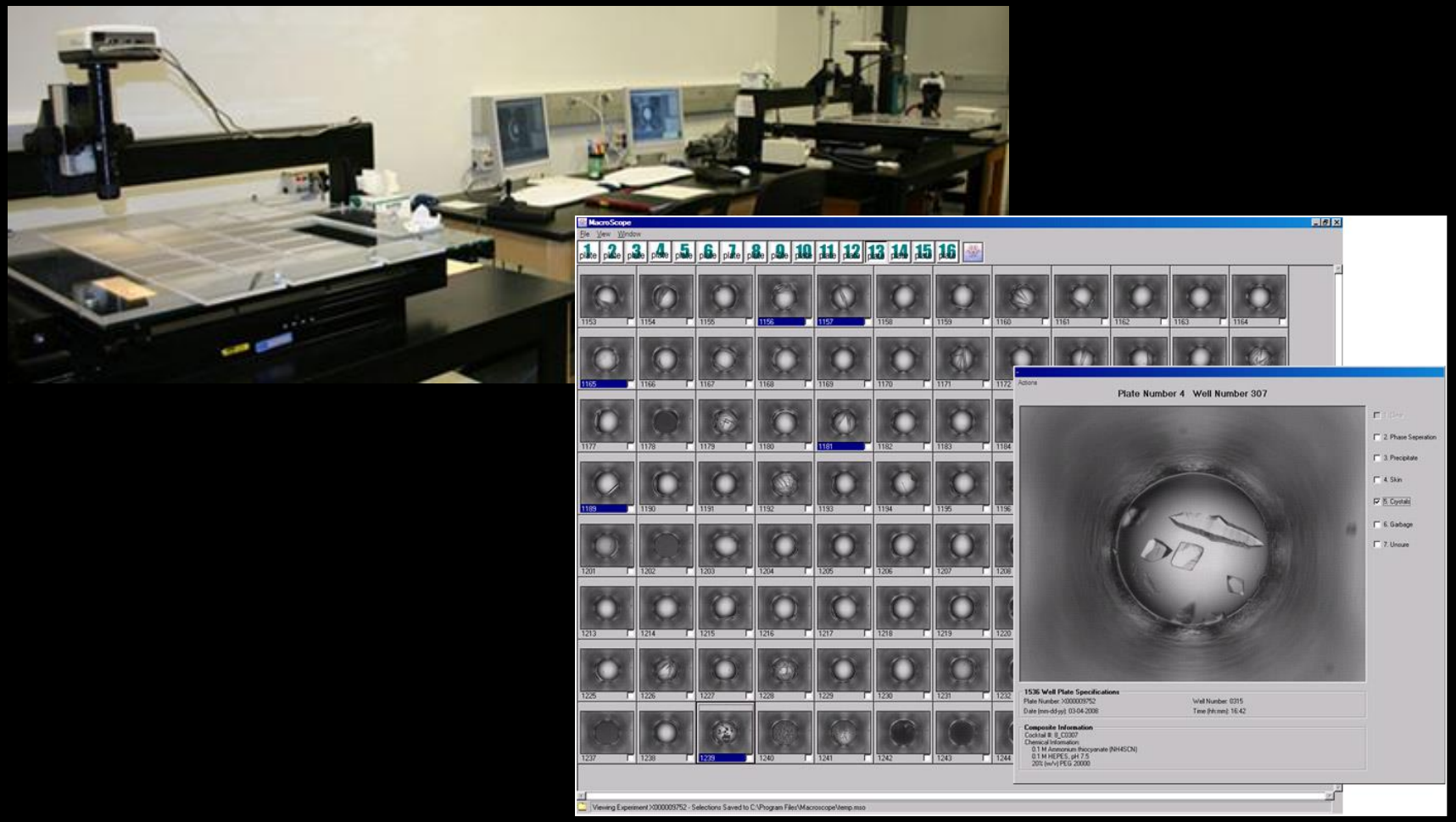
Other applications

- The code used to evaluate the CD_{coeff} is open source and freely available at http://ubccr.github.io/cockatoo/ or directly from the authors.
- Common chemical trends can be identified for optimization.
- The method can be applied with any crystallization screen, not just ours.
- It can also be used to design a screen where the clustering is equally spaced sampling the widest amount of chemical space with the minimum number of experiments.
- Other fingerprint definitions are available, e.g. activity. The fingerprints can be refined against outcome to determine how chemistry influences crystallization
- Comparing chemistry to outcome: The development of a chemical distance metric, coupled with clustering and hierarchal visualization applied to macromolecular crystallography. Bruno, Ruby, Luft, Grant, Seetharaman, Montelione, Hunt and Snell. PLOS One in press.

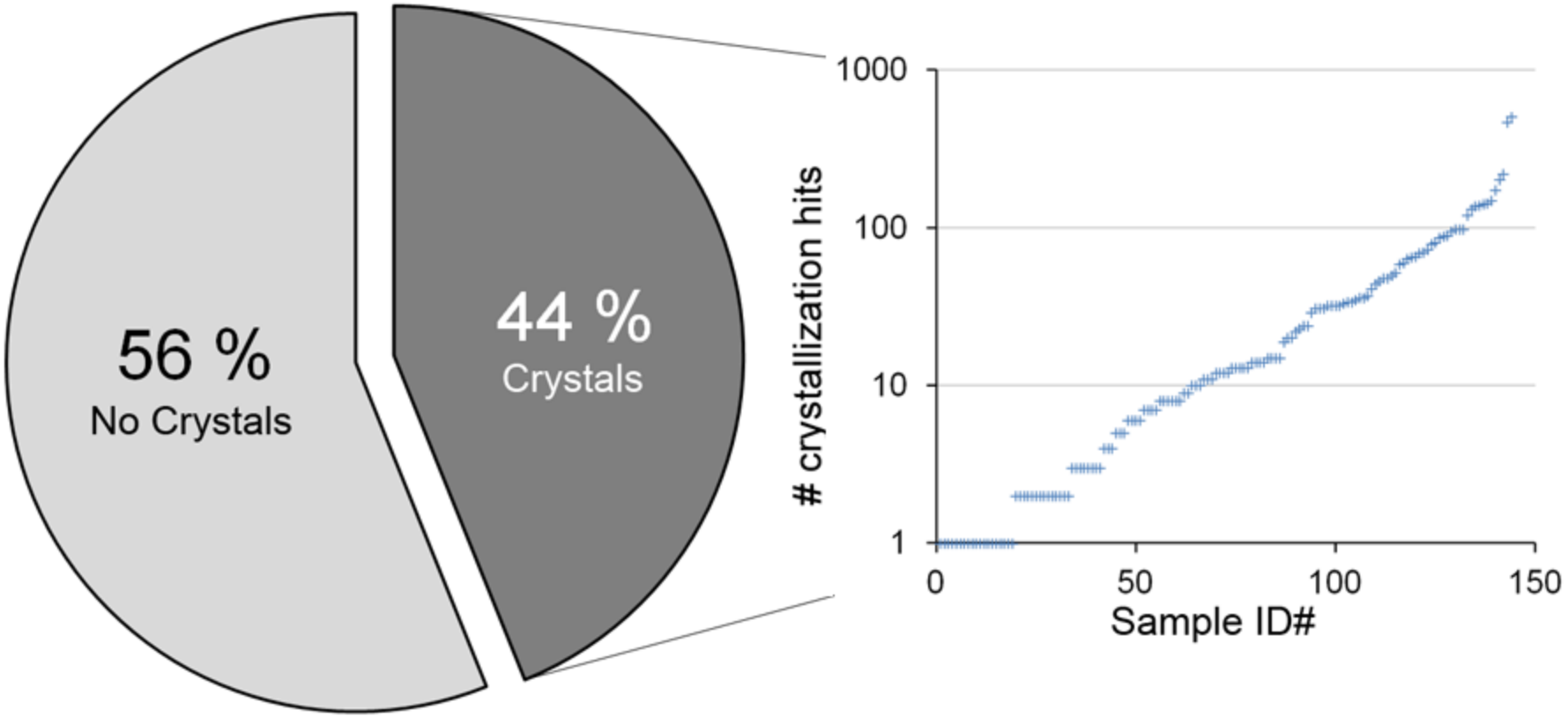
Summary

- By building on an existing chemical similarity metric and extending it to include all the components of the cocktail and the additional parameters of stoichiometry and chemical structure cocktails used for crystallization can automatically be clustered.
- The clustering can then be displayed as a hierarchal tree or dendogram.
- Overlying crystallization screening outcome on the dendogram can reveal details in an easily interpretable visible manner that drive further optimization
- The same overlay can also provide biological information that is otherwise missed.
- It can correct information that was missed or provide new information ‘fingerprinting’ the protein.
- It is quick – this analysis can be rapidly run on any result from the HWI screening laboratory.

Imaging technology based on standard microscopy – requires visible crystals



And it's still pretty good at indentifying them



Based on data from NESG 2013, 144/328 targets with one or more verified hits

Can we do better?

If you don’t see a crystal it didn’t crystallize?

A major advance in imaging technology can identify submicron crystals

Using SONICC and UV-TPEF we can observe and verify protein crystals < 1 micron in size.

~80% of proteins in PDB low-symmetry generate SHG

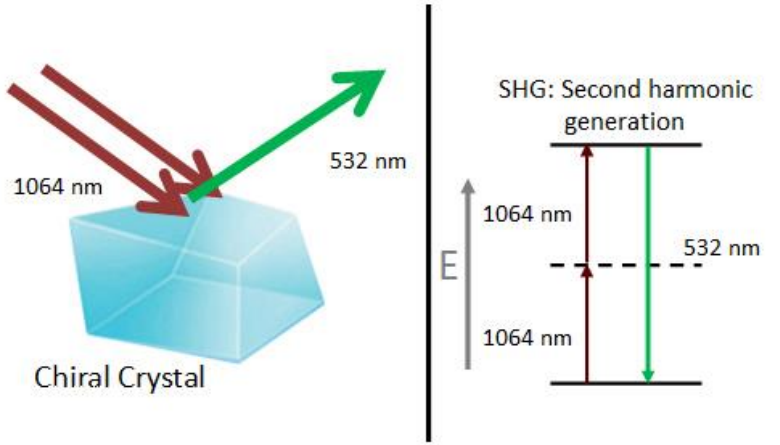


Figure 1. Two photons of IR (1064 nm) interact with a chiral crystal to generate SHG (532 nm).

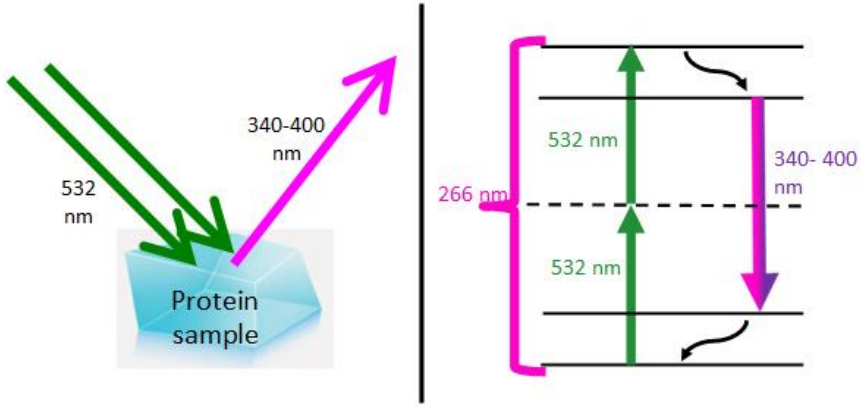


Figure 2. Depiction of UV-TPEF where two photons of green interact with a protein sample to generate UV excited fluorescence

http://www.formulatrix.com/products/protein-crystallography-tools/sonicc/how.html

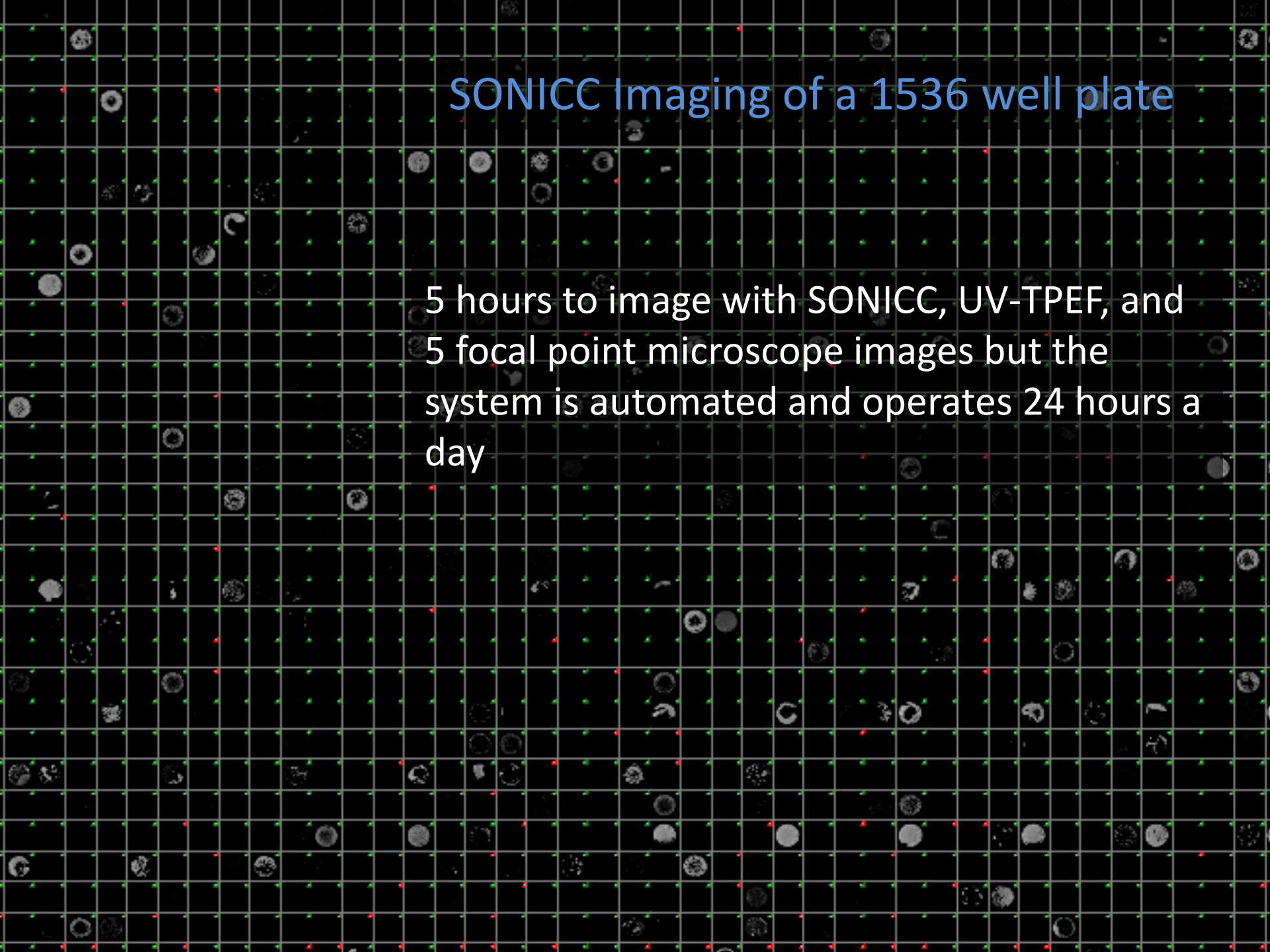
SONICC and UV-TPEF are well described elsewhere

Second-Order Nonlinear Optical Imaging of Chiral Crystals. David J. Kissick, Debbie Wanapun, and Garth J. Simpson. *Annu Rev Anal Chem*. 2011 ; 4: 419–437.

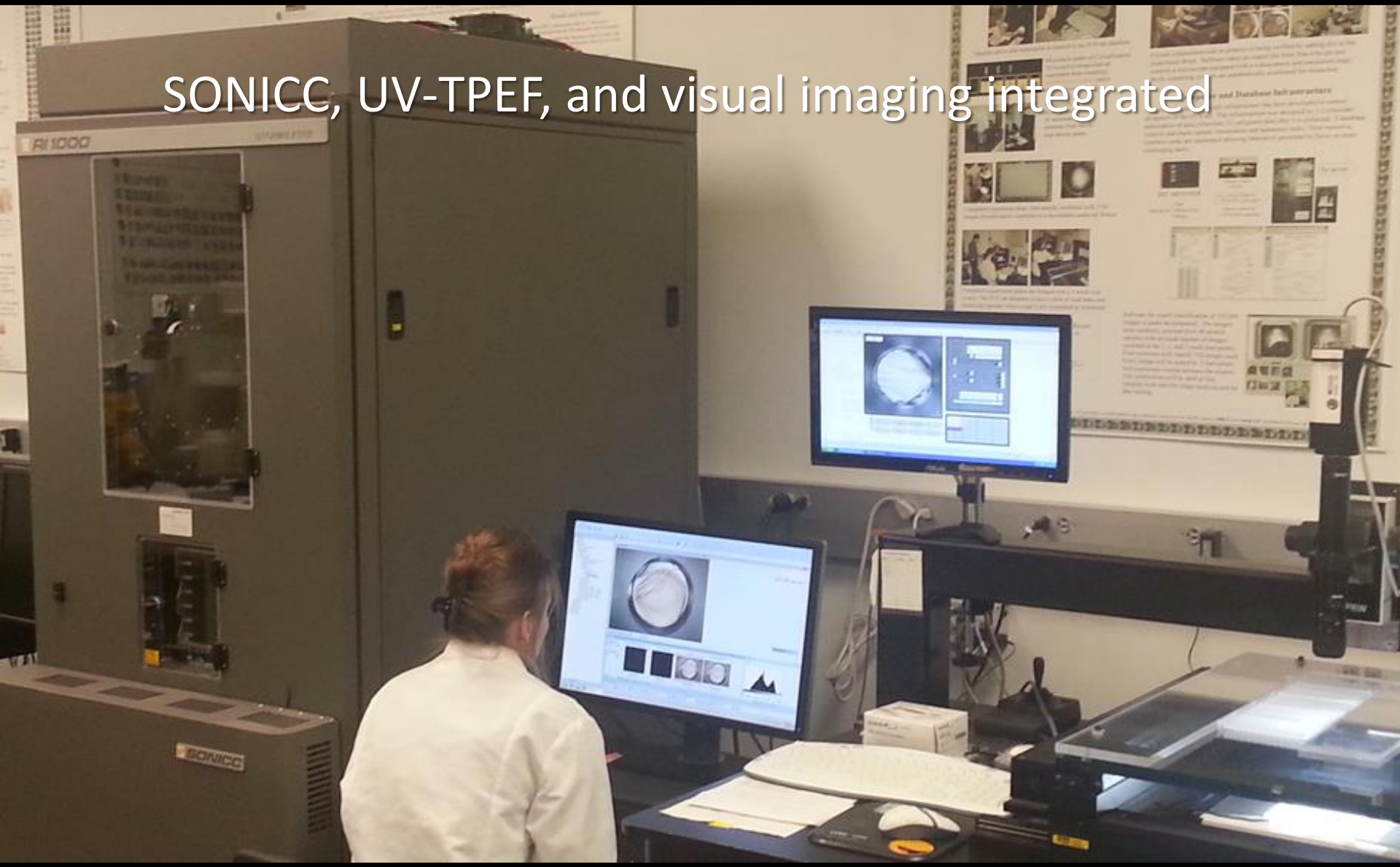
Two-photon fluorescence imaging of impurity distributions in protein crystals. Caylor, C. L., Dobrianov, I., Kimmer, C., Thorne, R. E., Zipfel, W. & Webb, W. W. (1999). Phys. Rev. E, 59, R3831–R3834

This is a talk about their application

SONICC Imaging of a 1536 well plate



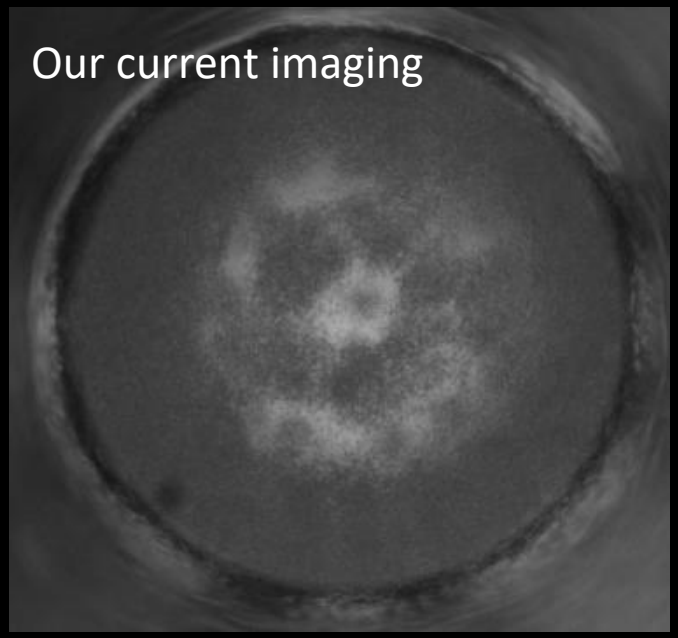
5 hours to image with SONICC, UV-TPEF, and 5 focal point microscope images but the system is automated and operates 24 hours a day

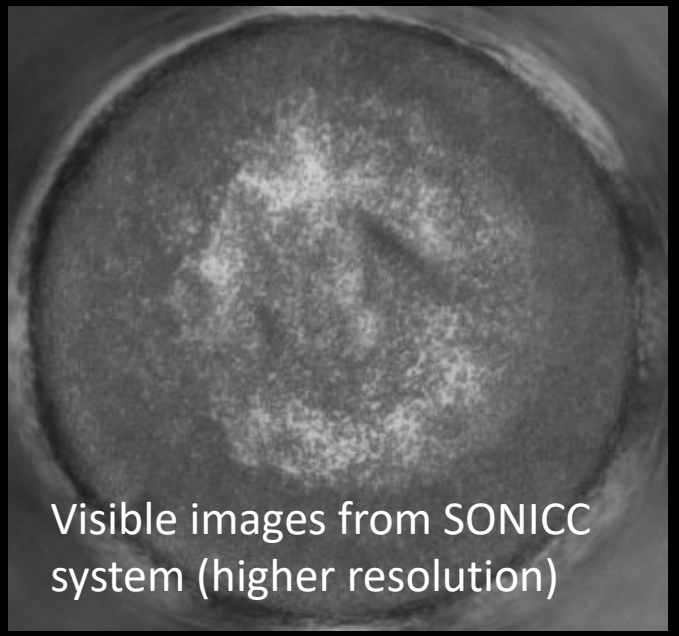

SONICC, UV-TPEF, and visual imaging integrated
RI 1000
SONICC

One protein in detail to lay out the experiment

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500 produced the following:



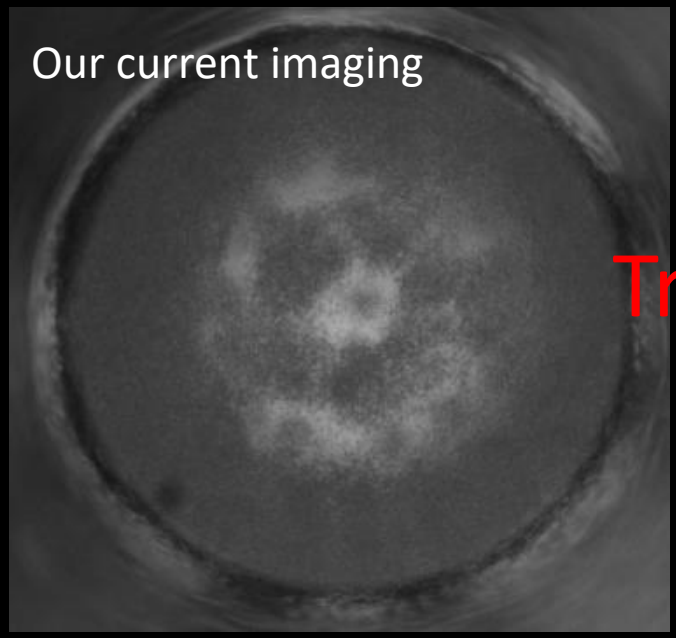


Immediately after the protein is added to the cocktail

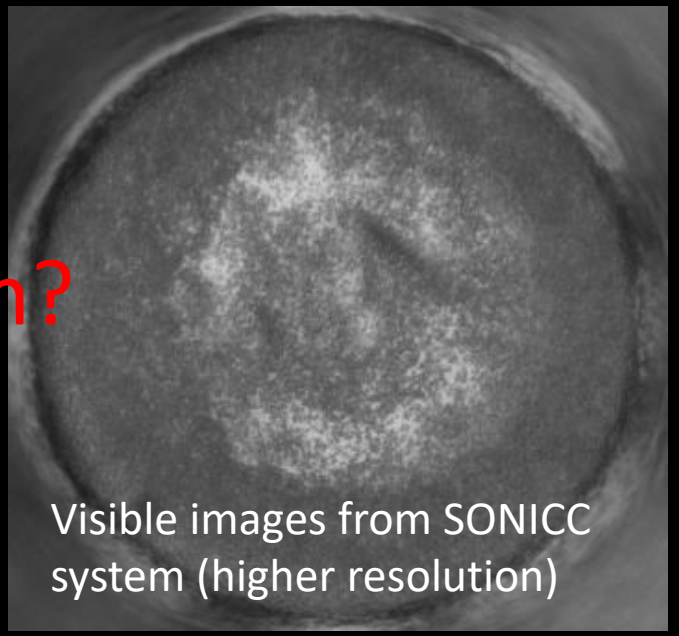
One protein tested

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500 produced the following:



Try again?

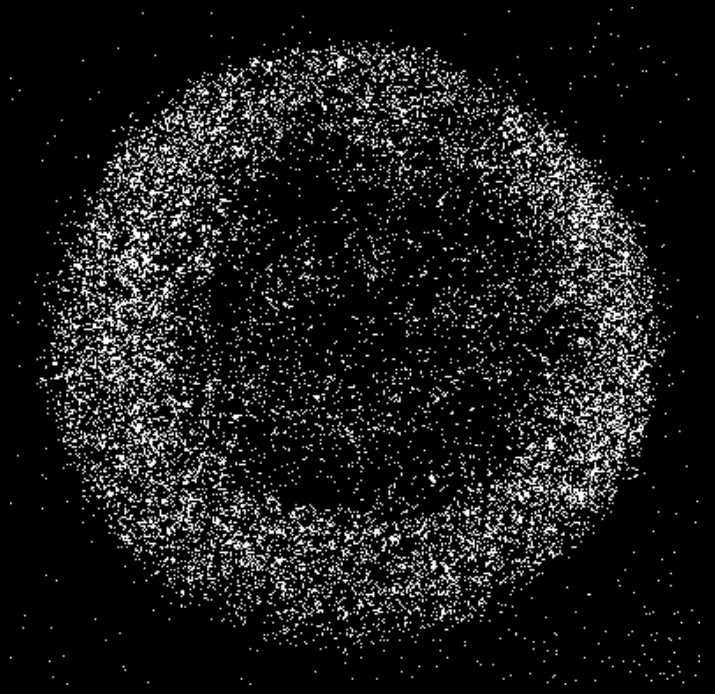


Immediately after the protein is added to the cocktail

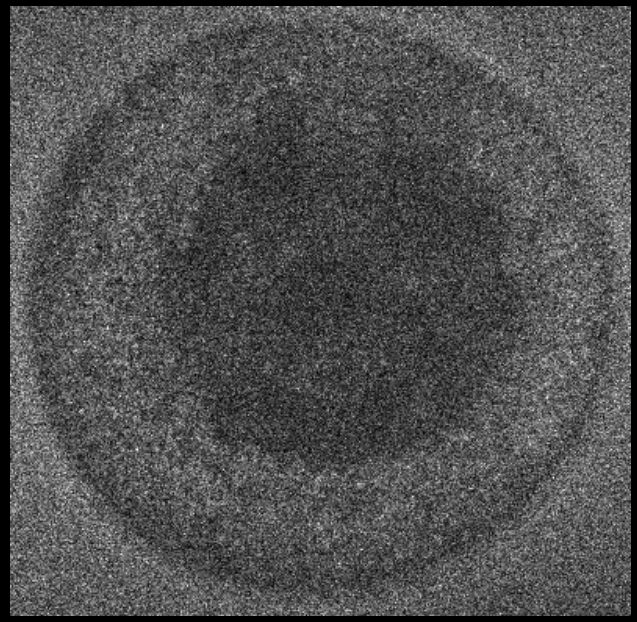
Initial use of SONICC and UV imaging

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500 produced the following:



SONICC SHG image

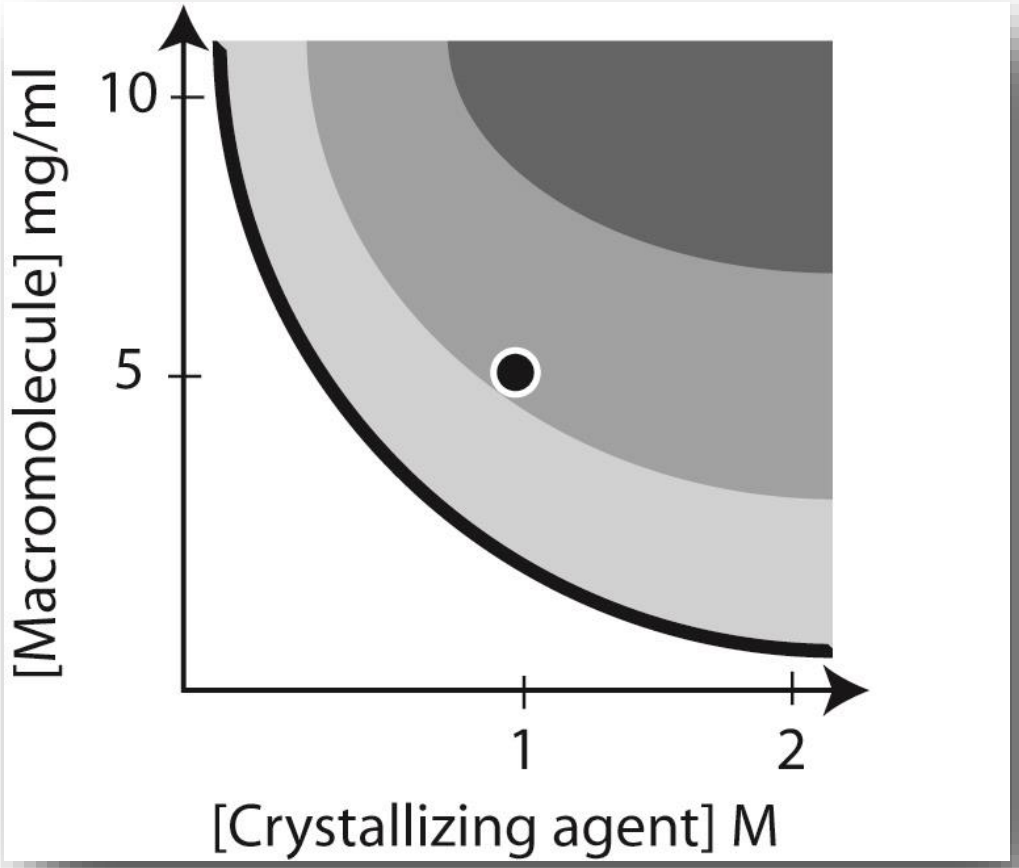


UV-TPEF image

The experiment design

- Use SONICC to review 1536 crystallization assay
- Identify micro-crystals missed visually
- Select 5 *chemically-diverse* cocktails with the micro-crystals
- Set up Drop/Volume Ratio Optimization
- 1 x [Protein]
- ½ x [Protein]
- Analyze results at 2 and 4 weeks

Batch experiments require labile supersaturation at their onset

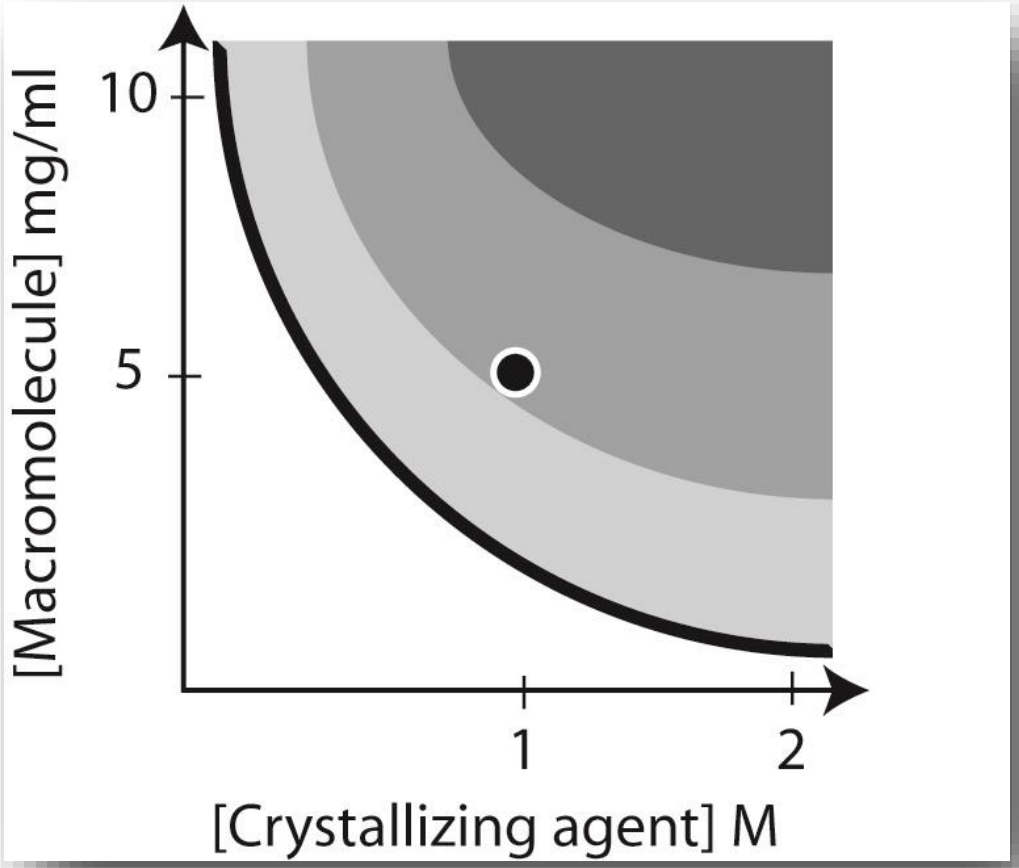


Precipitation

Nucleation (labile)

Growth (metastable)

The location of the experiment in the labile zone determines outcome

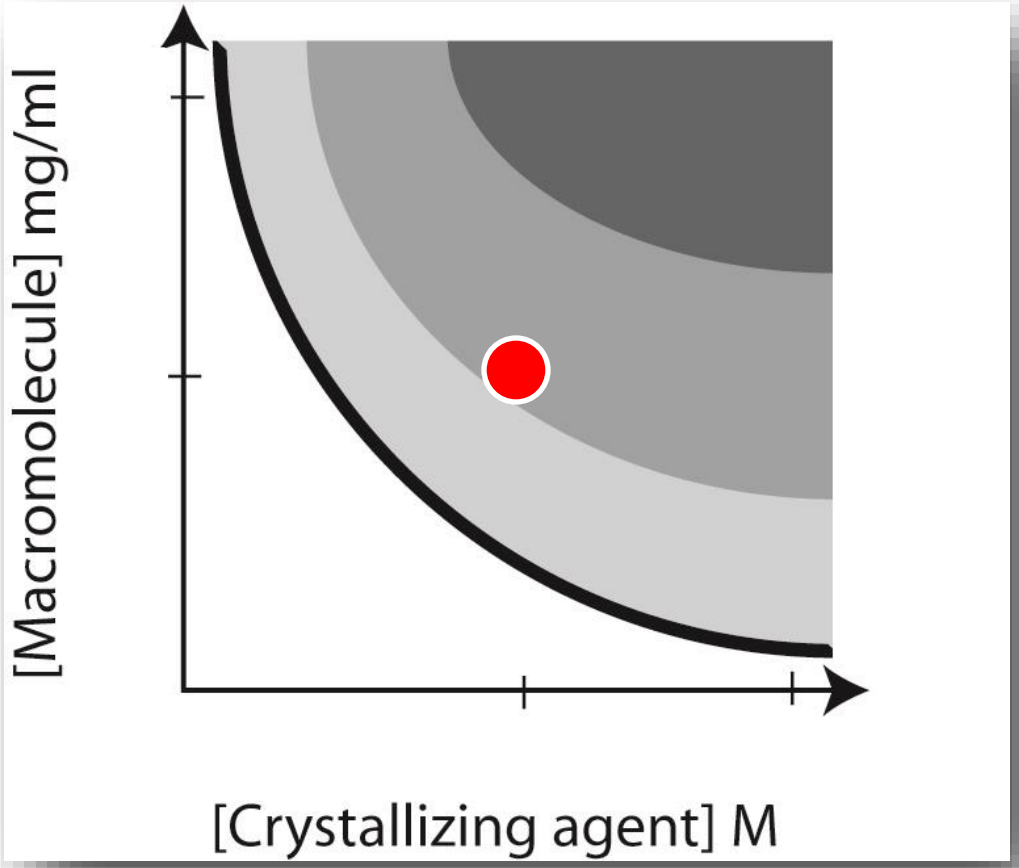


Precipitation

Nucleation (labile)

Growth (metastable)

The location of the experiment in the labile zone determines outcome

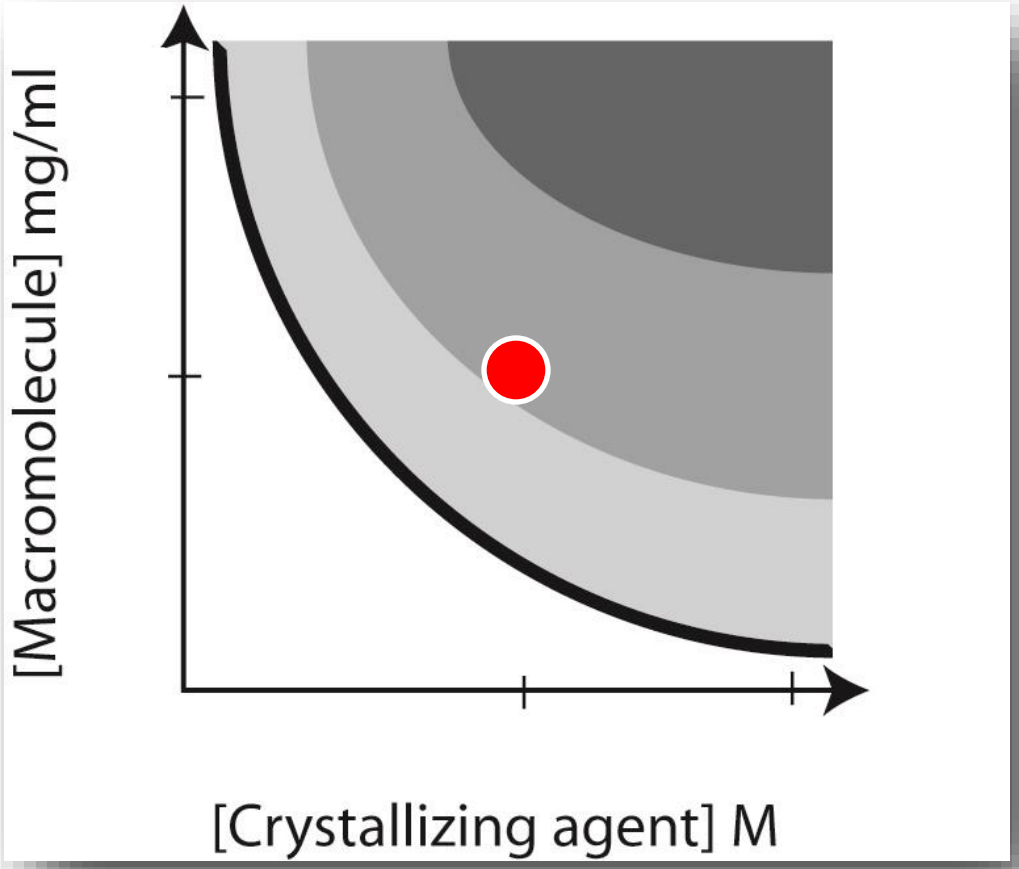


Precipitation

Nucleation (labile)

Growth (metastable)

In the labile zone
closer to the metastable zone

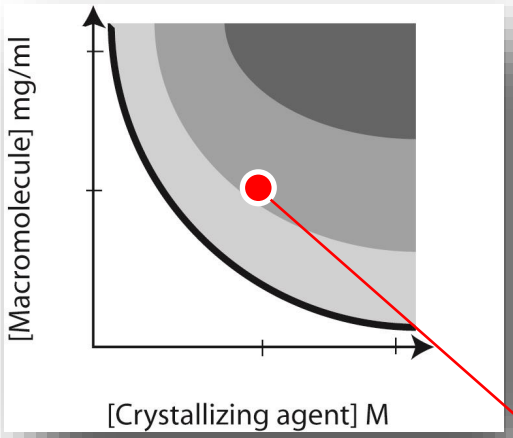


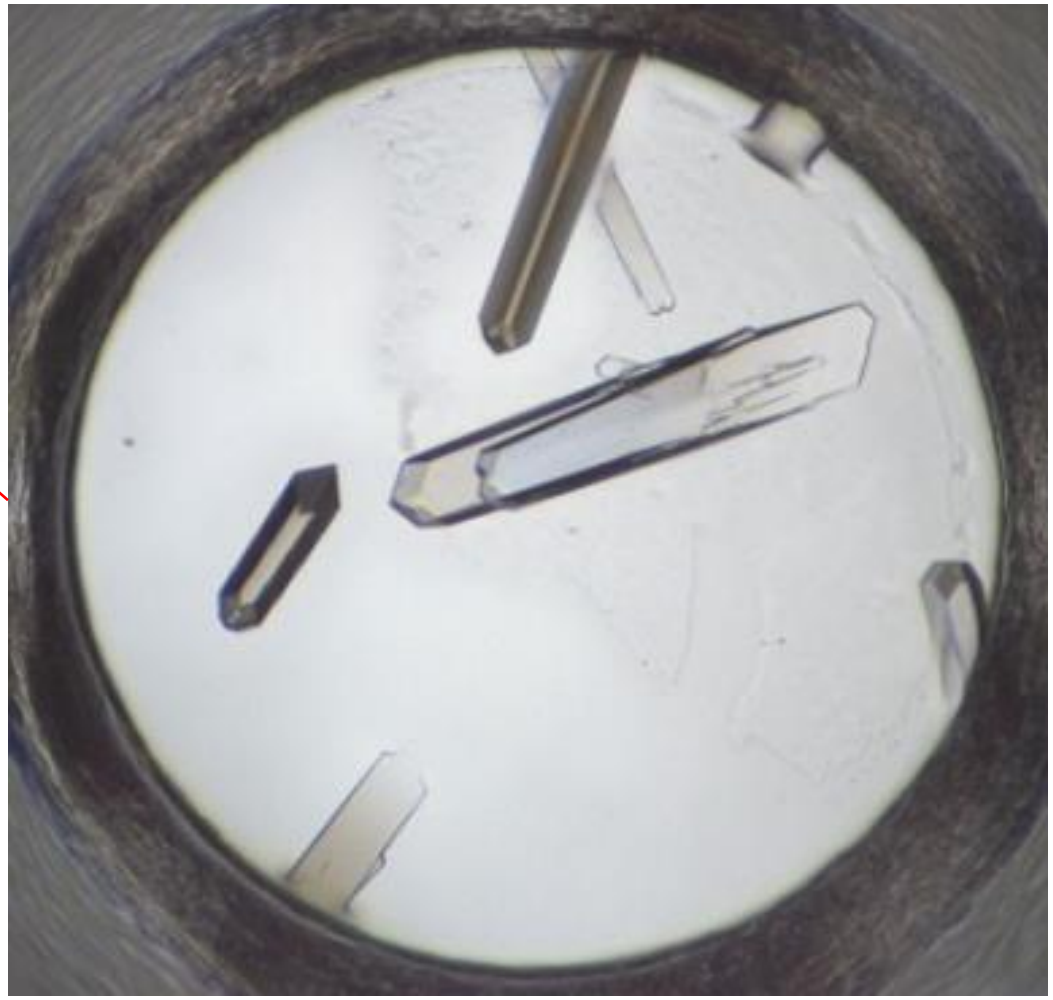
Precipitation

Nucleation
(labile)

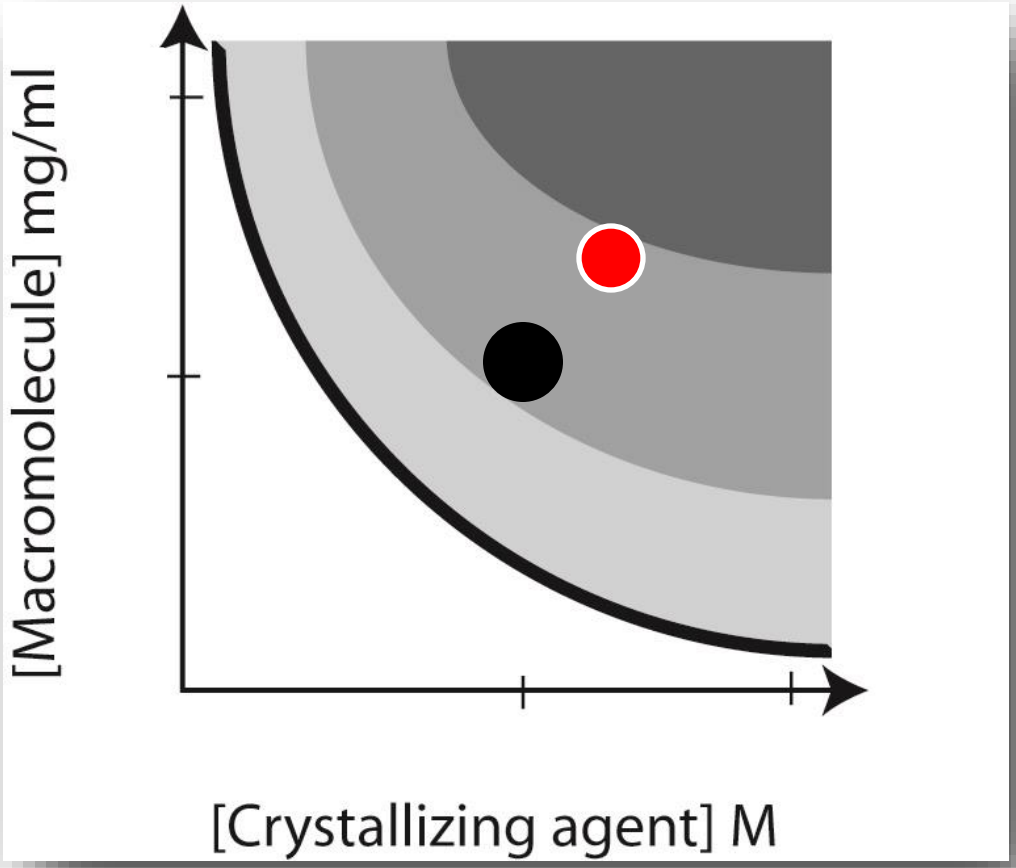
Growth
(metastable)

In the labile zone closer to the metastable zone, less nucleation





In the labile zone
closer to the precipitation zone

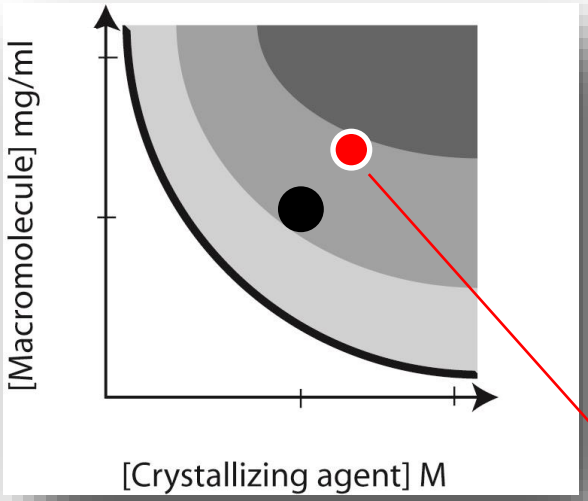


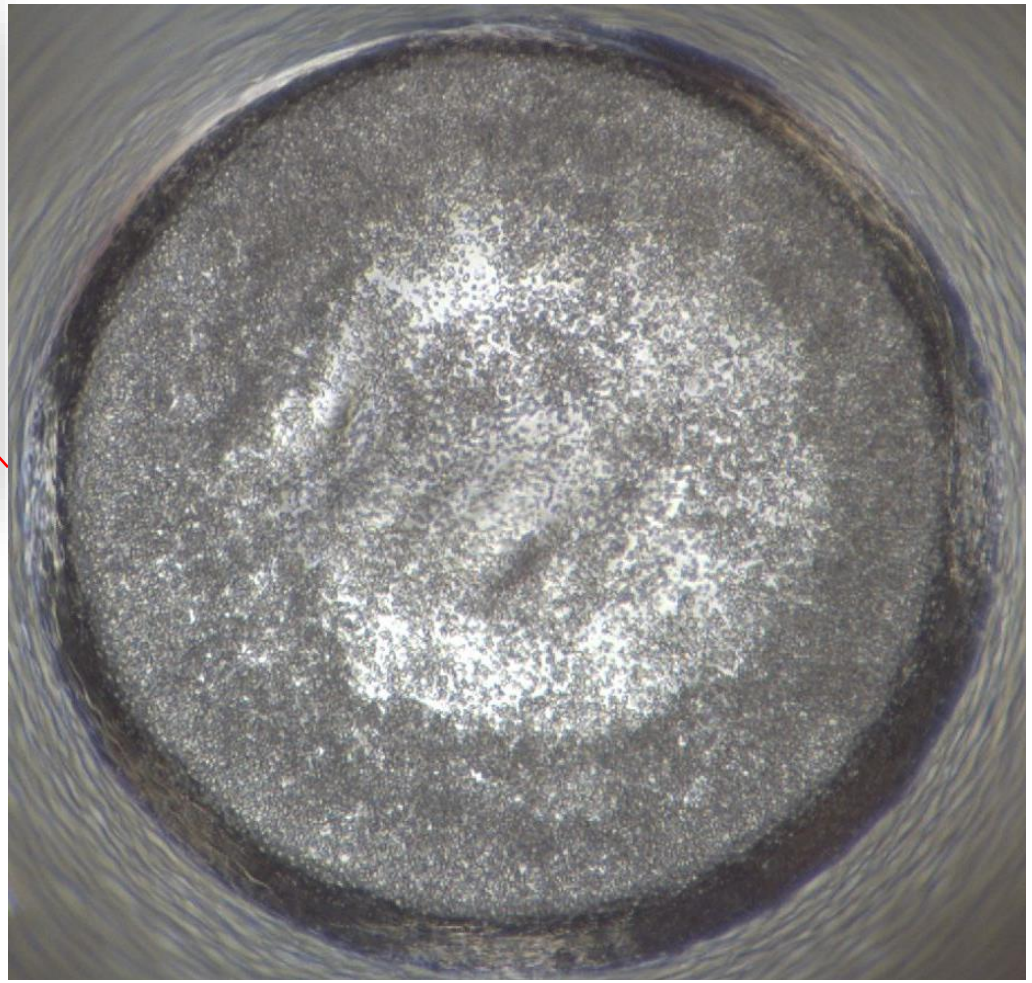
Precipitation

Nucleation
(labile)

Growth
(metastable)

In the labile zone closer to the precipitation zone, more nucleation

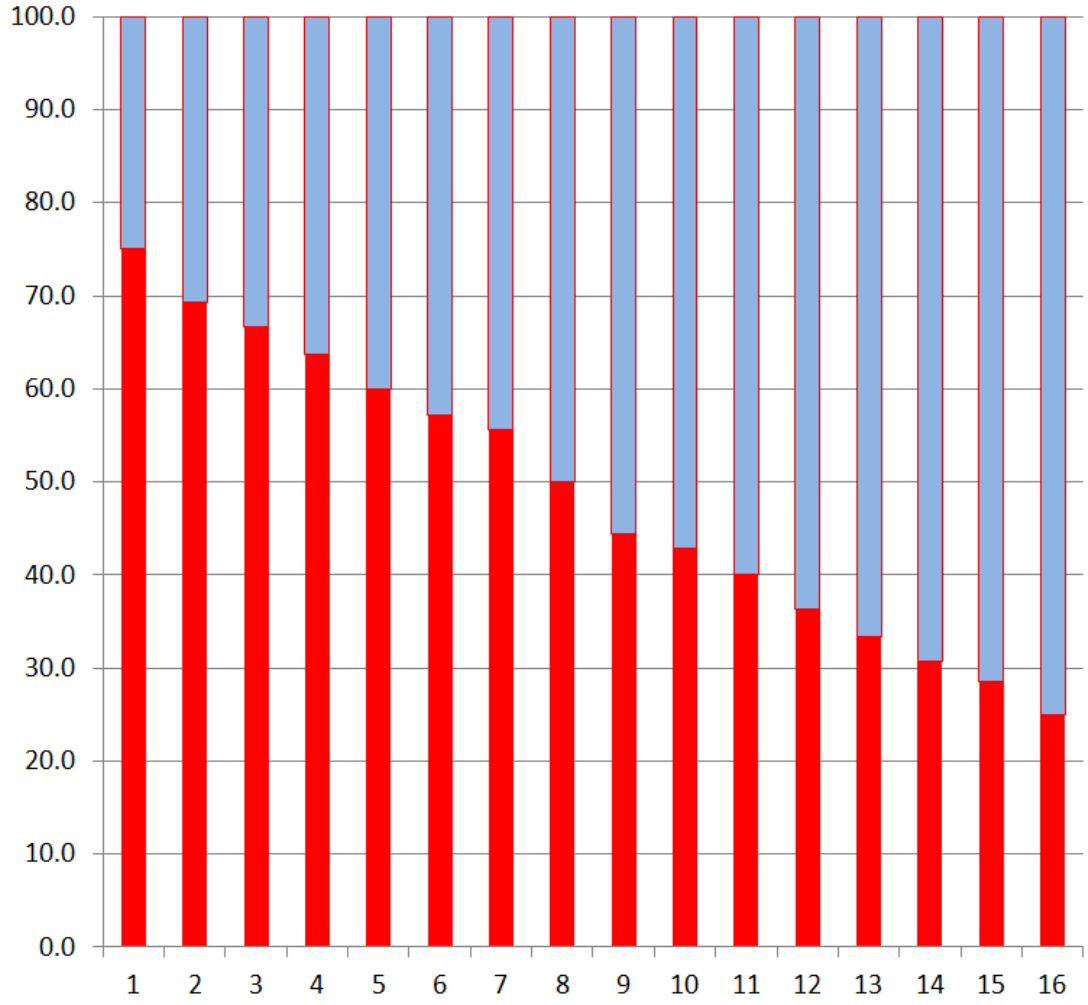




Optimization using Drop Volume Ratio

%(v/v) Protein

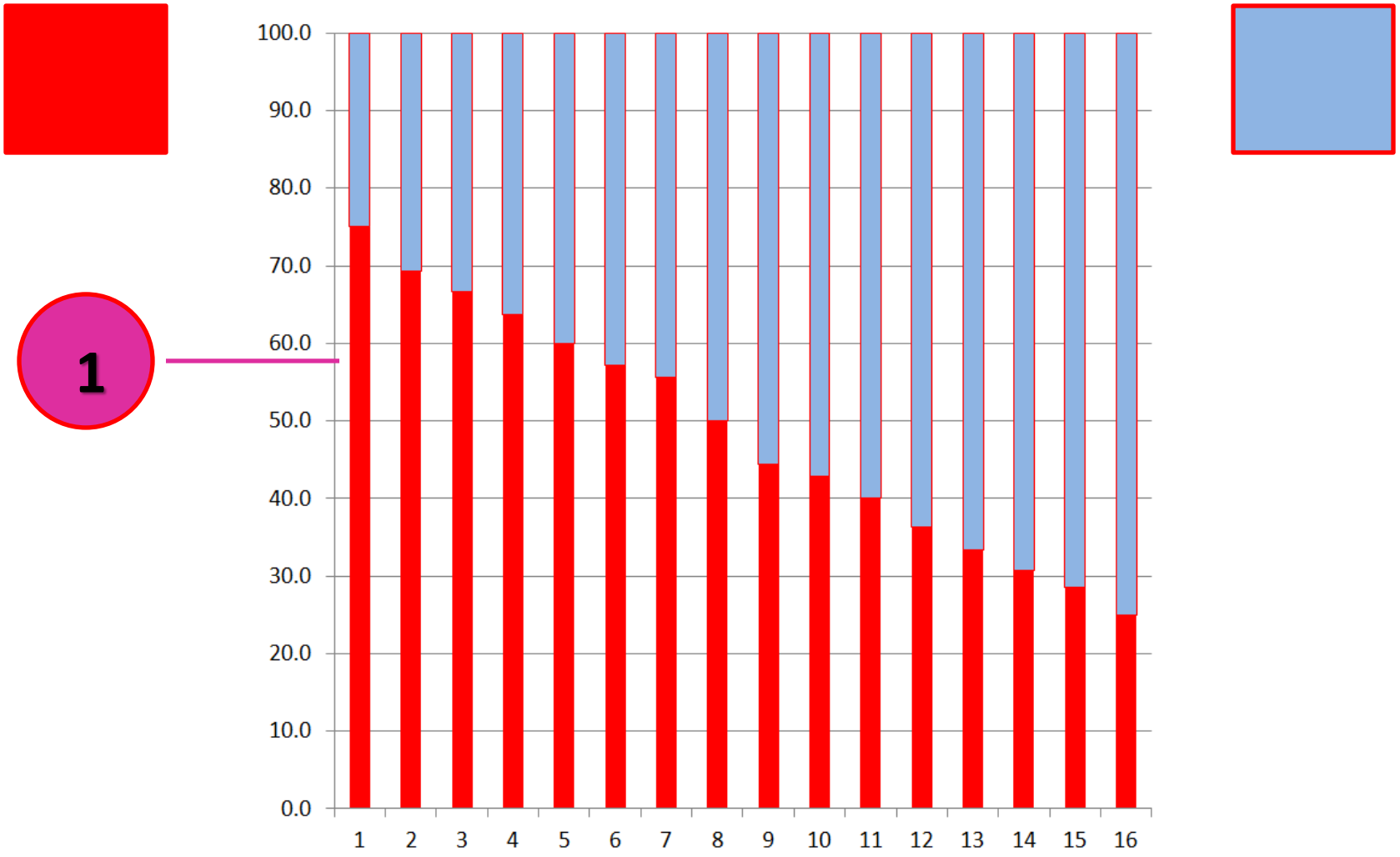
%(v/v) Cocktail



More Protein, Less Cocktail

%(v/v) Protein

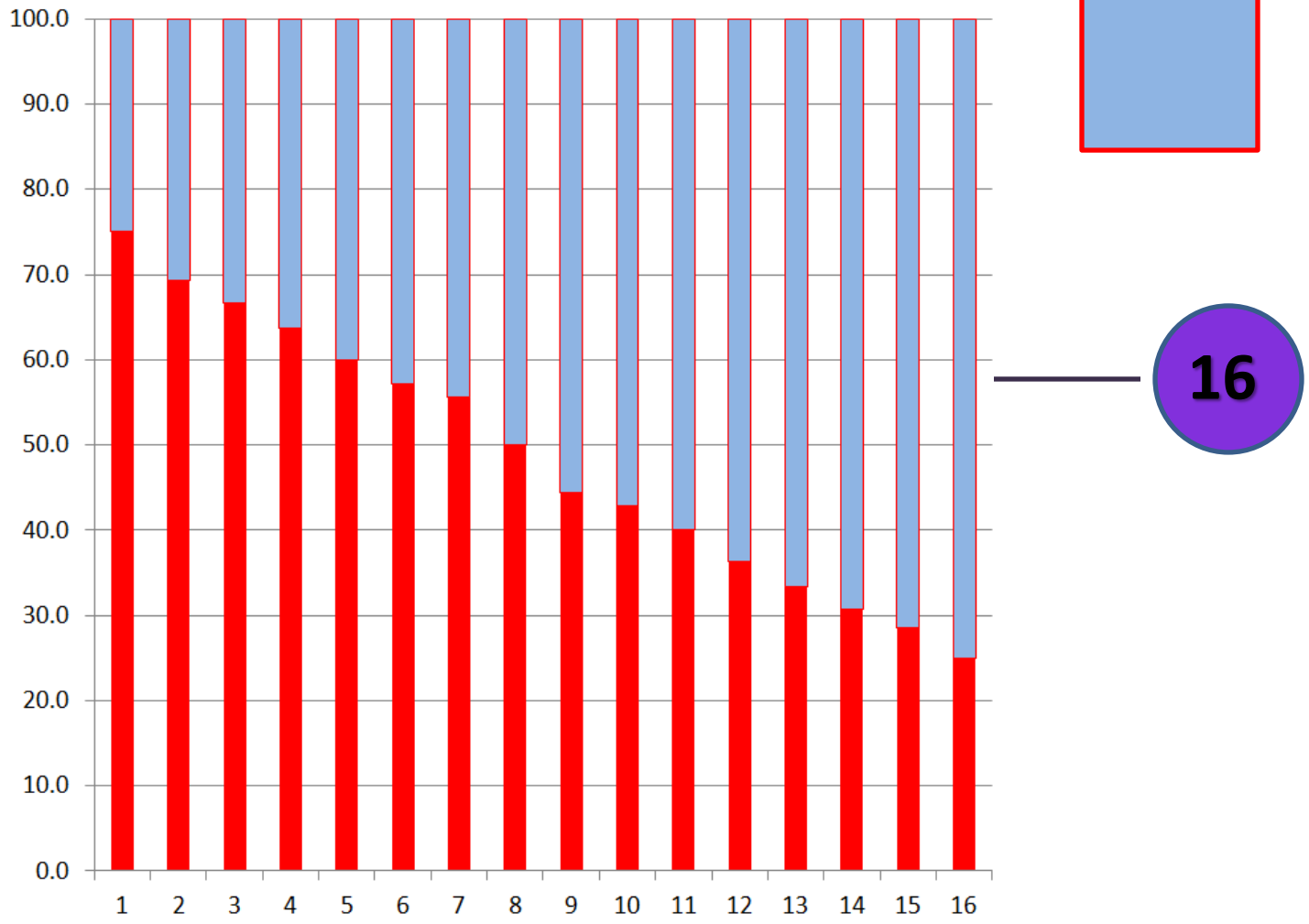
%(v/v) Cocktail



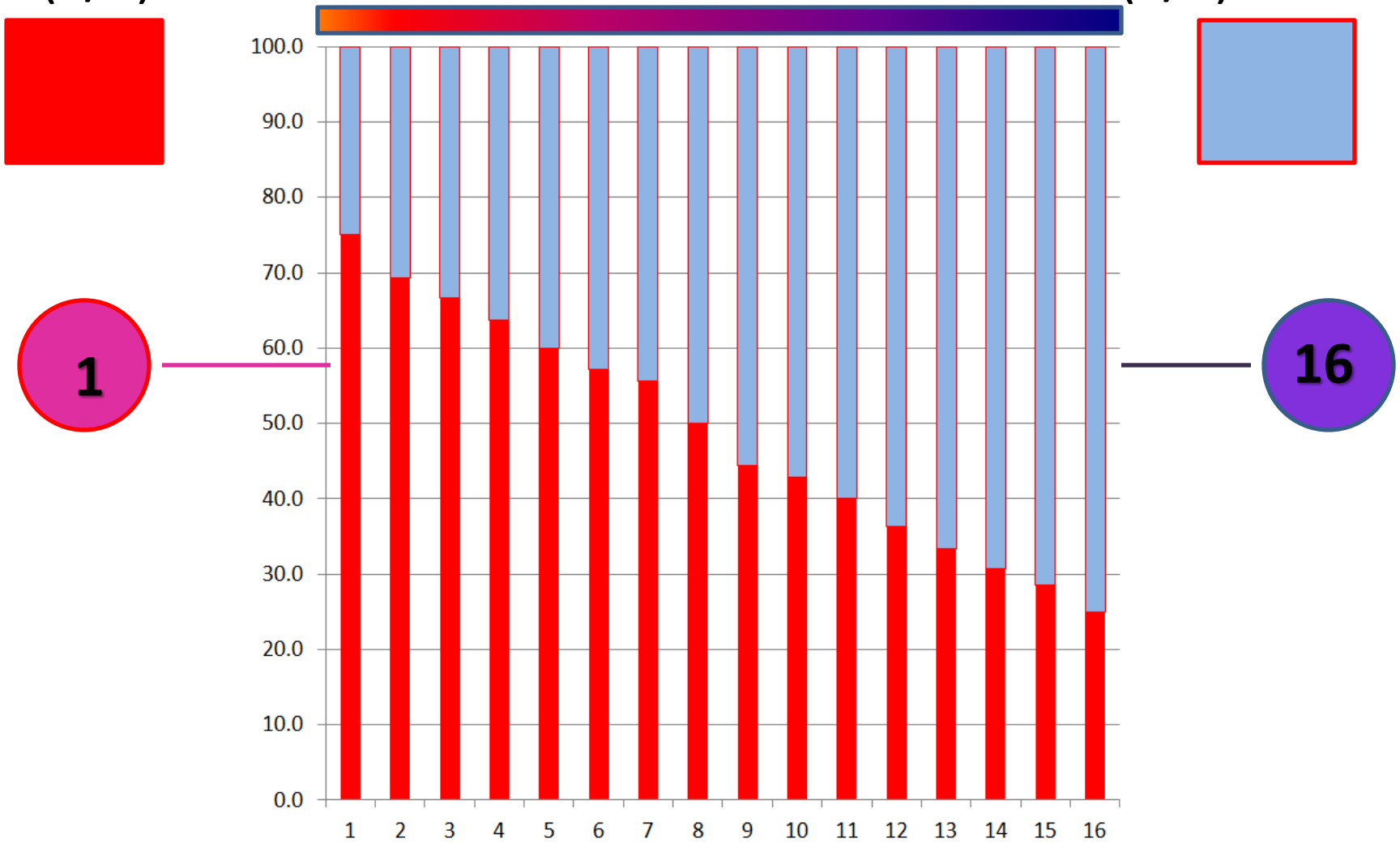
Less Protein, More Cocktail

%(v/v) Protein

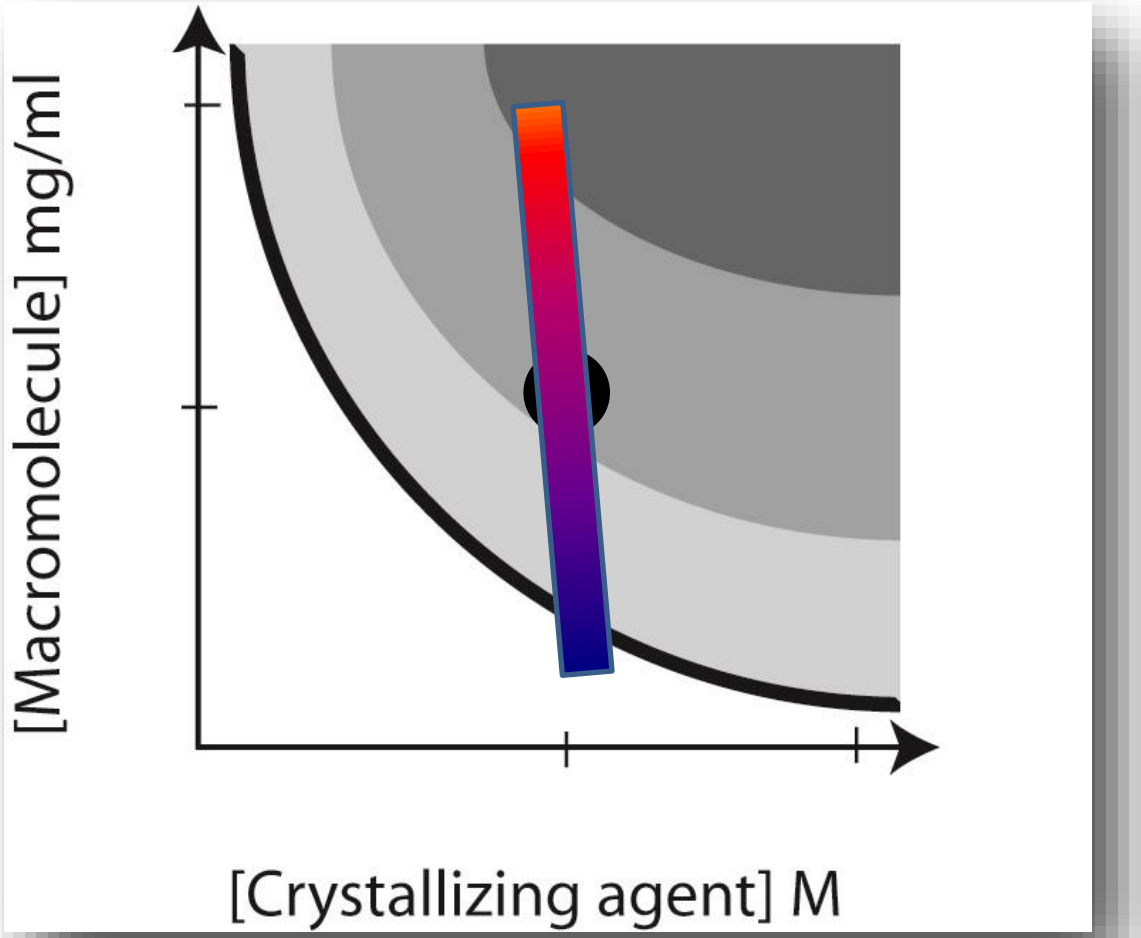
%(v/v) Cocktail



Drop Volume Ratio Optimization

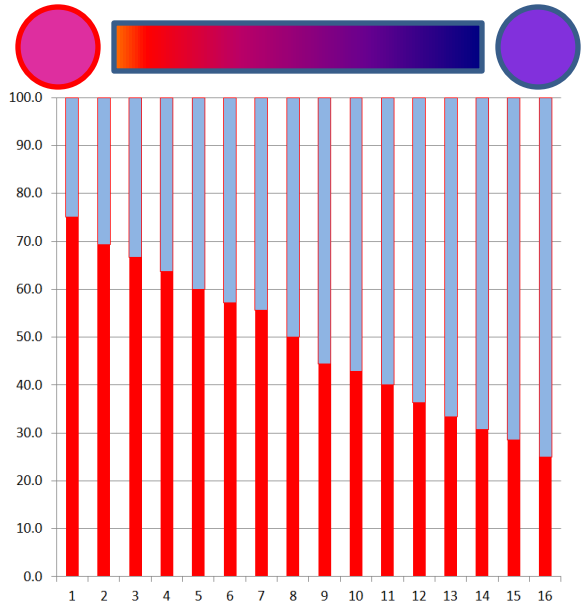


Drop volume ratio can be used to sample a path of supersaturation

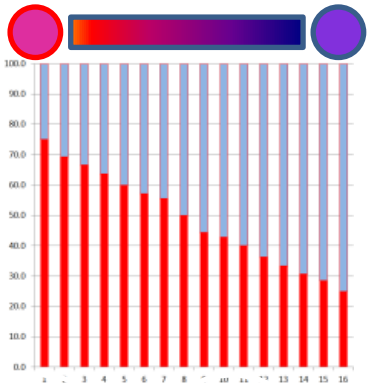


A different format for the same drop volume ratio data

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

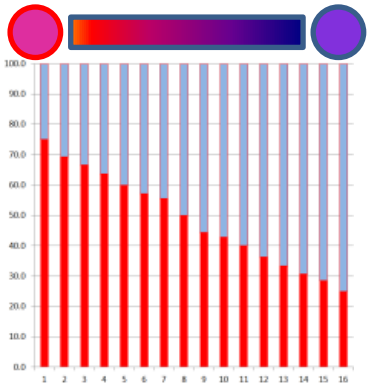


Two sets of 16 experiments



1	2	3	4	1	2	3	4
5	6	7	8	5	6	7	8
9	10	11	12	9	10	11	12
13	14	15	16	13	14	15	16

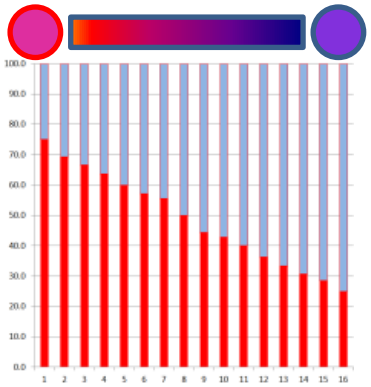
One set of 16 at $[protein]_{init}$



1	2	3	4	1	2	3	4
5	6	7	8	5	6	7	8
9	10	11	12	9	10	11	12
13	14	15	16	13	14	15	16

1x [Protein]

One set of 16 at ½ $[protein]_{init}$



1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

1x [Protein]

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

½ x [Protein]

Six proteins used in the study

Plate	Protein Name	[Protein] (mg/ml)	# trp	MW (kDa)
X14163	1	9.47	1	10.2
X14164	2	10.1	1	13.7
X14176	3	40	NA	47.9
X14222	4	9	3	18.8
X14223	5	7.72	3	29.4
X14224	6	11.4	5	40.6

These represent protein under active study and their names, while known to us, are obscured until the structural efforts are successful or have been exhausted

Four proteins had visible crystals in the 1536 screen, two did not

Plate	Protein Name	[Protein] (mg/ml)	# trp	MW (kDa)	
X14163	1	9	1	10.2	Visible crystals
X14164	2	10	1	13.7	
X14176	3	40	NA	47.9	Not
X14222	4	9	3	18.8	Visible crystals
X14223	5	8	3	29.4	
X14224	6	11	5	40.6	Not

The experiment

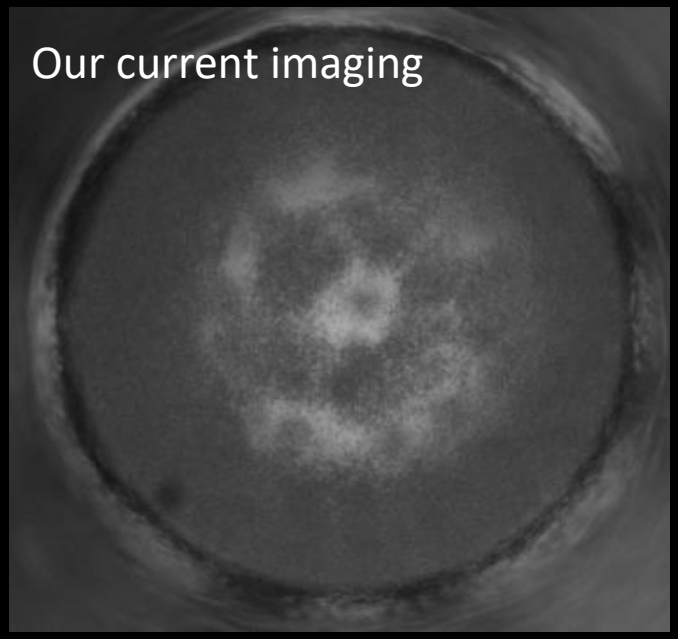
- For the samples that displayed visible crystal hits SONICC and UV-TPEF images were studied.
- Those that showed hits in SONICC and/or UV-TPEF but NOT visibly defined crystals were used for DVR optimization.
- Five chemically diverse as possible conditions were set up in the DVR optimization.

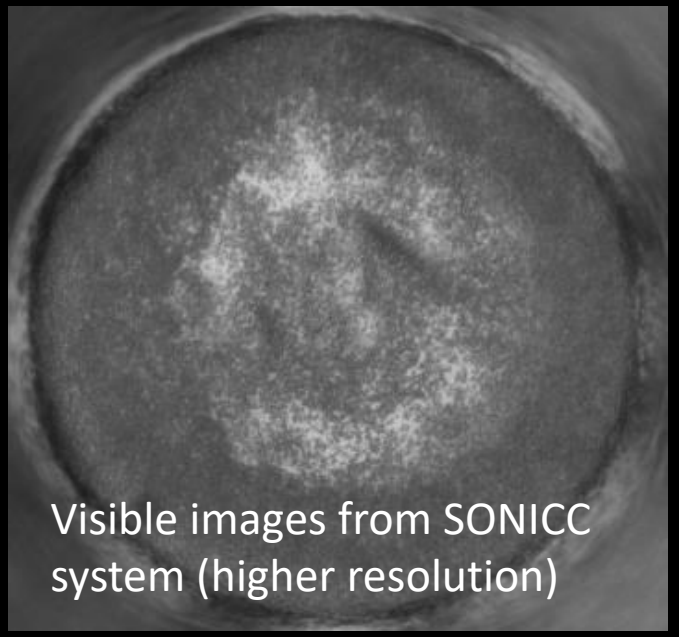
- For the samples that displayed no visible hits those that showed hits in SONICC and/or UV-TPEF were used for DVR/T optimization
- Five chemically diverse as possible conditions were set up in the DVR optimization.

Try again

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500 produced the following:



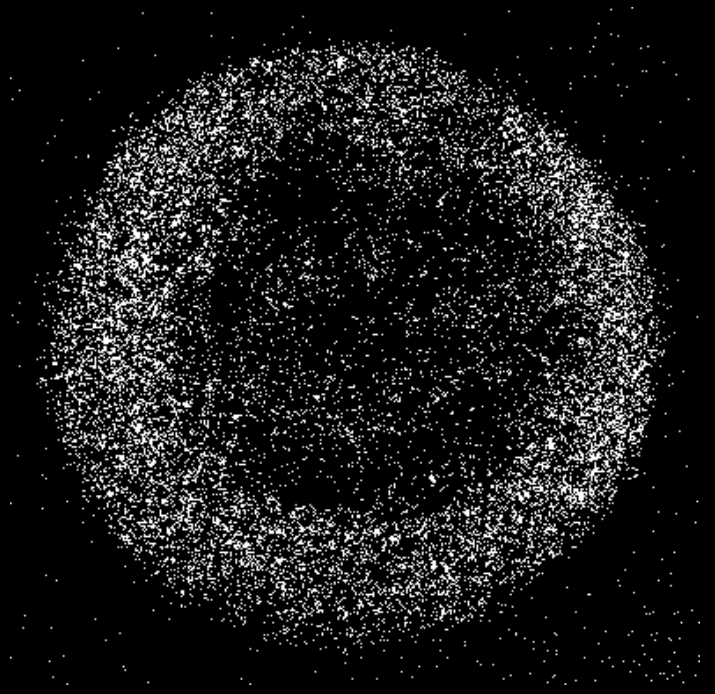


Immediately after the protein is added to the cocktail

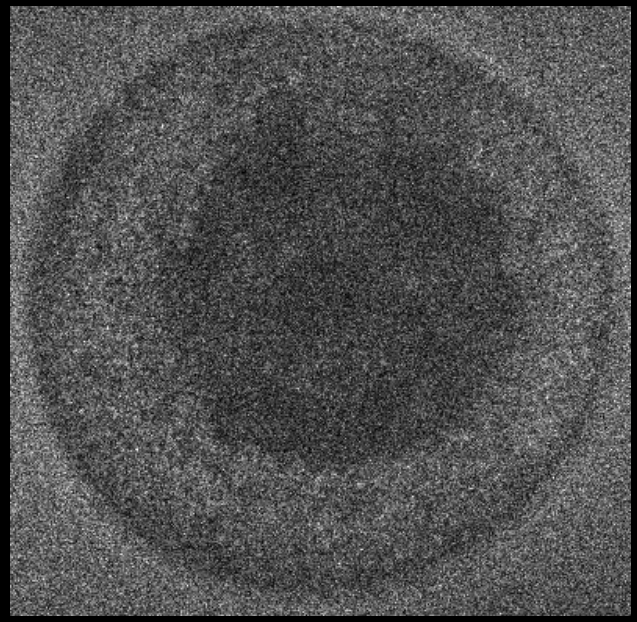
Initial use of SONICC and UV imaging

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500 produced the following:



SONICC SHG image



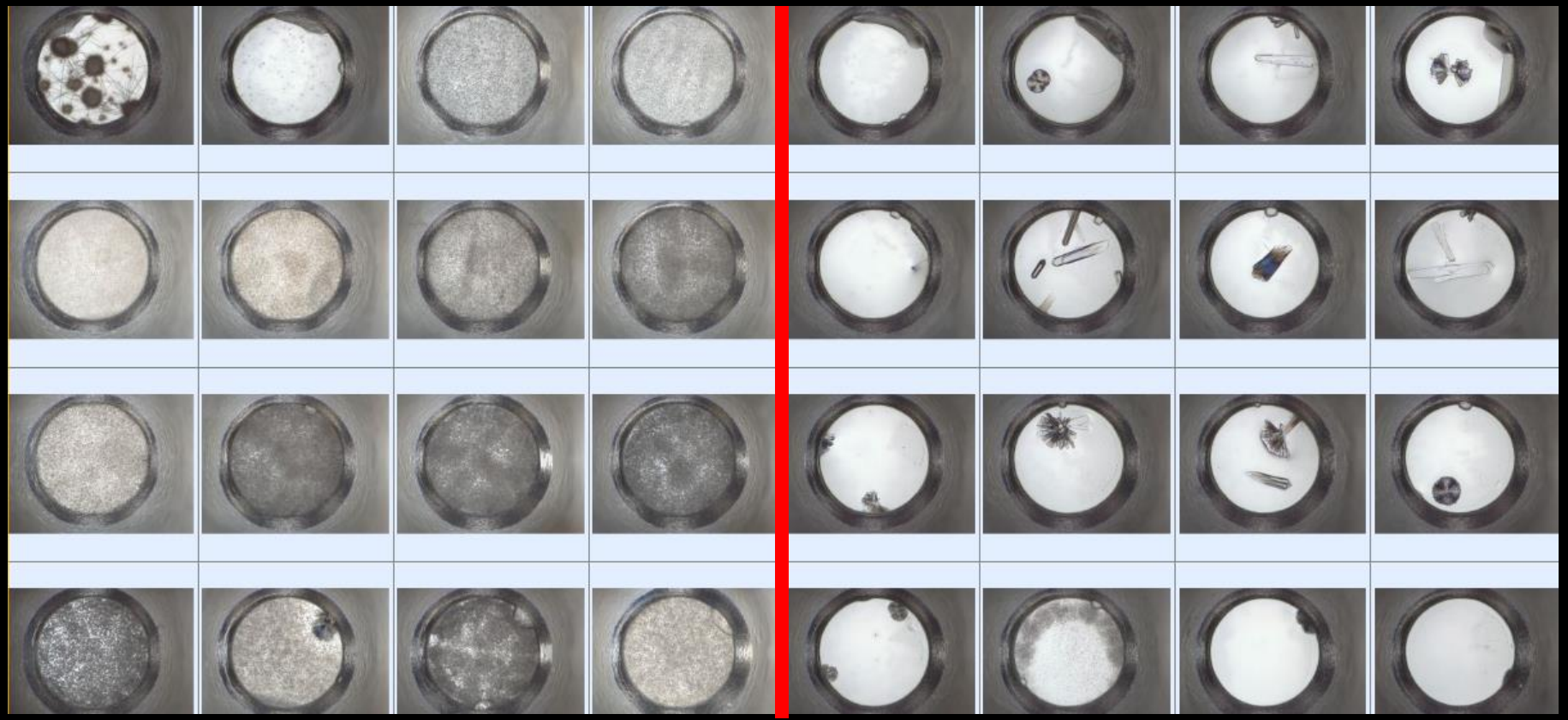
UV-TPEF image

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500

X14163- Full[P]- 10mg/ml

X14163- [P]/2- 5mg/ml



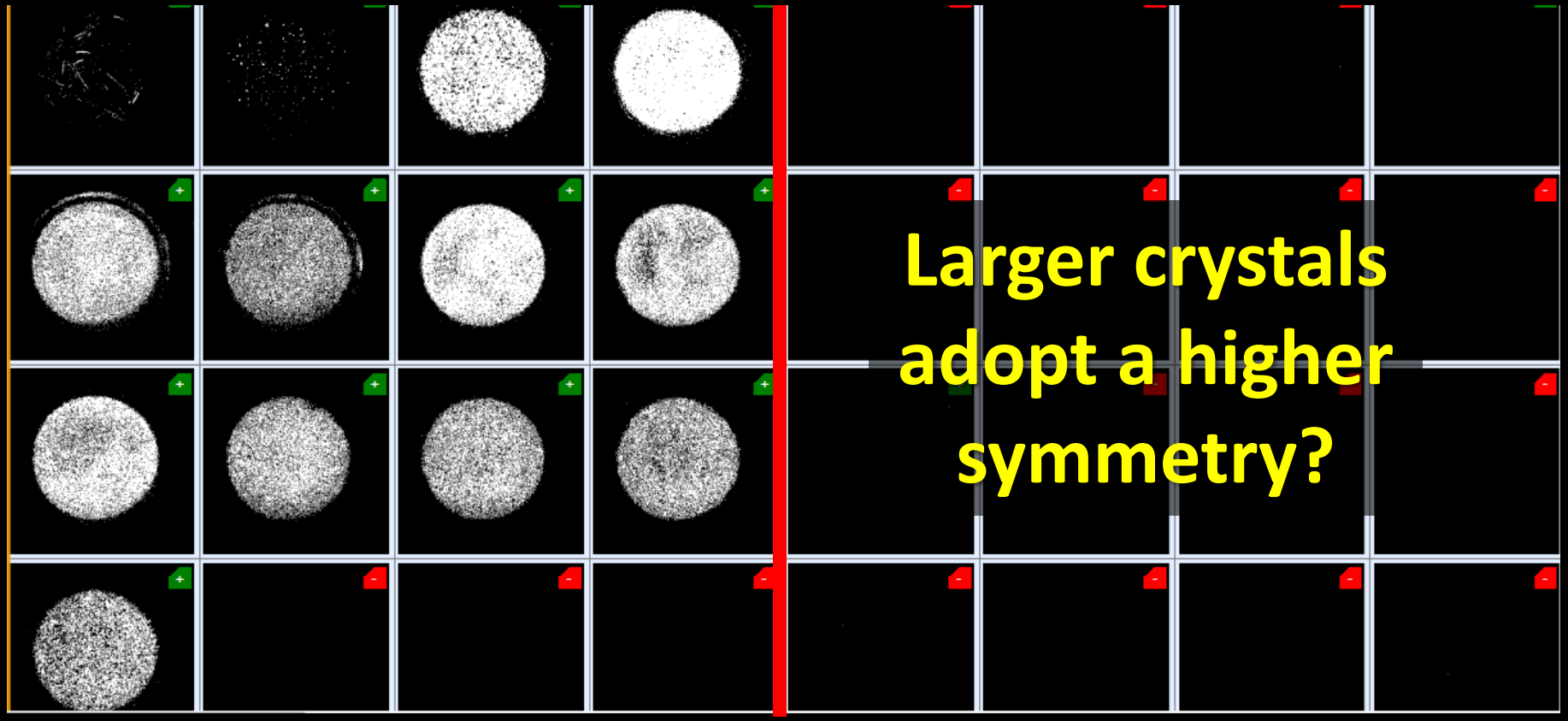
Visible at 4wk

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500

X14163- Full[P]- 10mg/ml

X14163- [P]/2- 5mg/ml



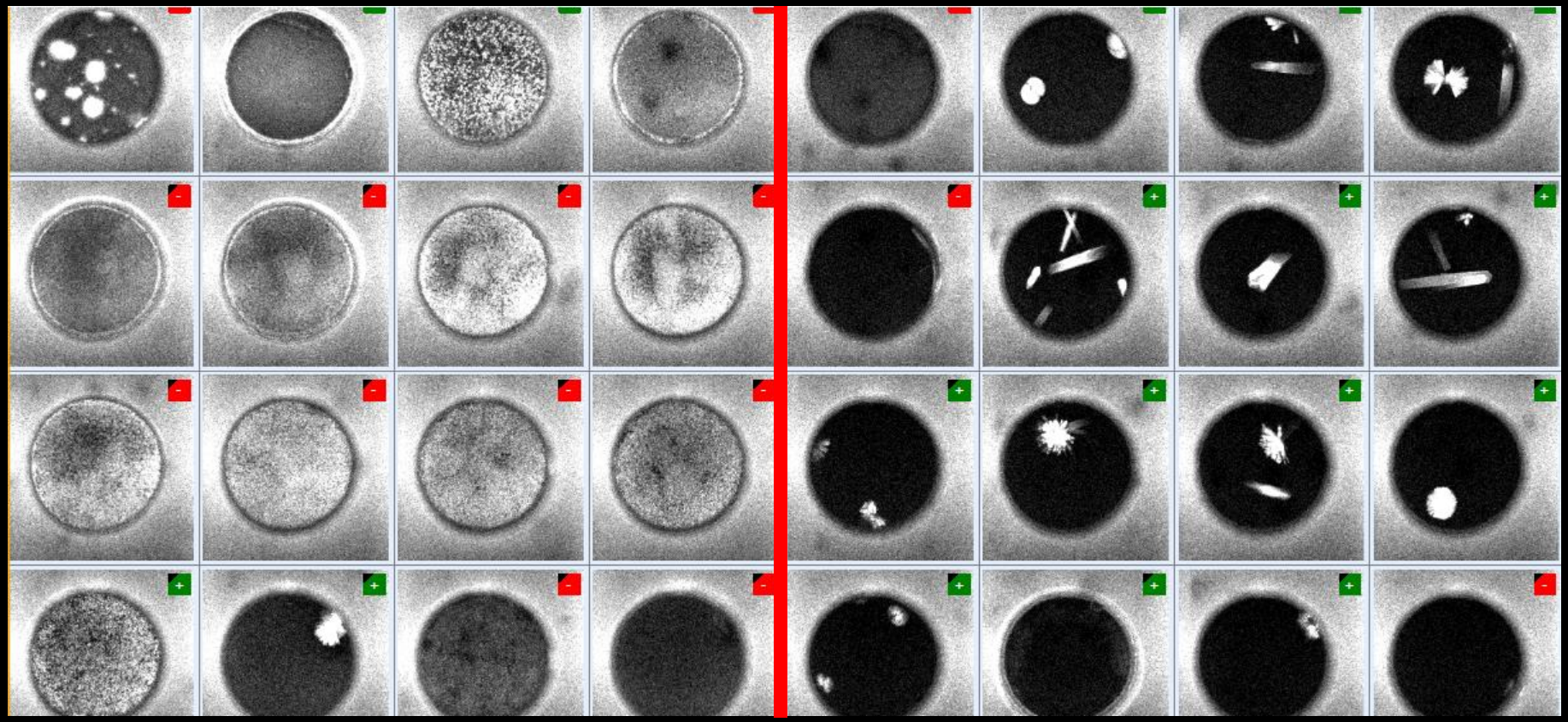
SHG at 4wk

Protein 1, part of the pyruvate dehydrogenase protein complex

Hampton Research PEGRx HT-F4, 4% (v/v) 2-Methyl-2,4 pentanediol, 0.1 M Citric Acid pH=3.5 20% (w/v) PEG 1500

X14163- Full[P]- 10mg/ml

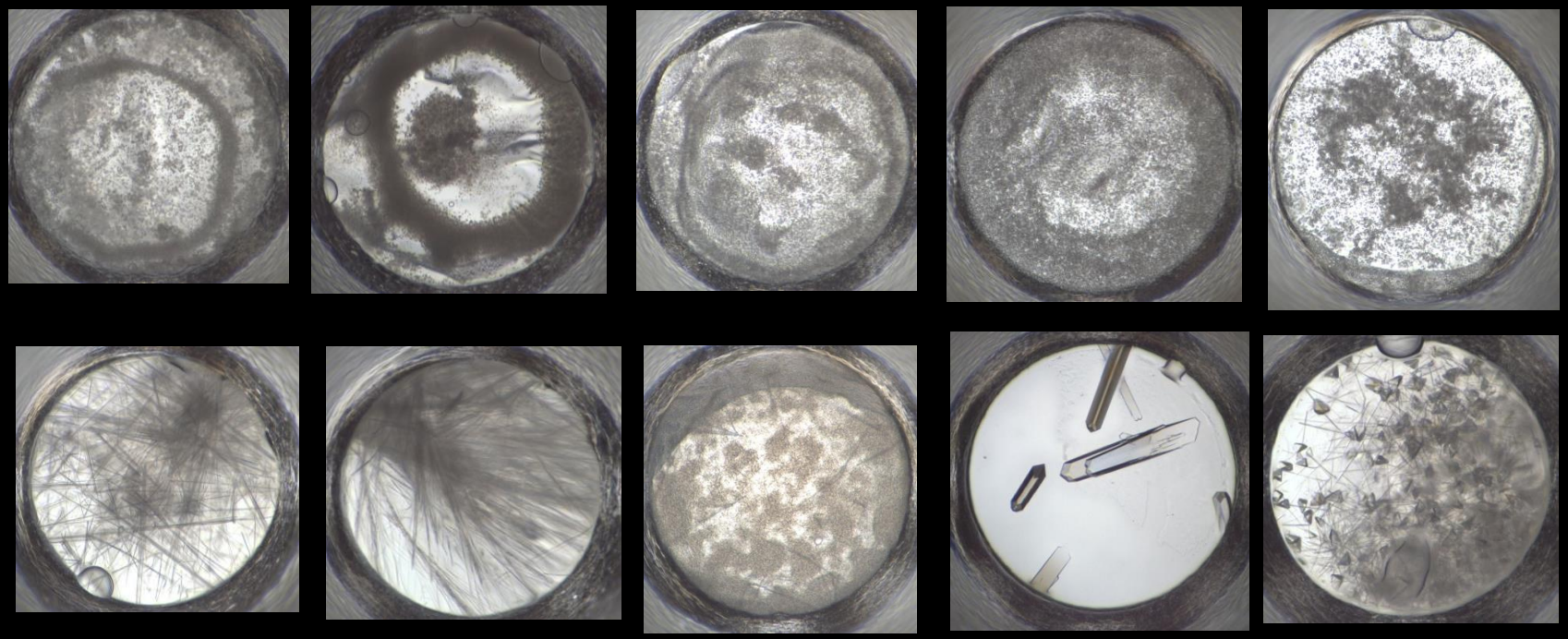
X14163- [P]/2- 5mg/ml



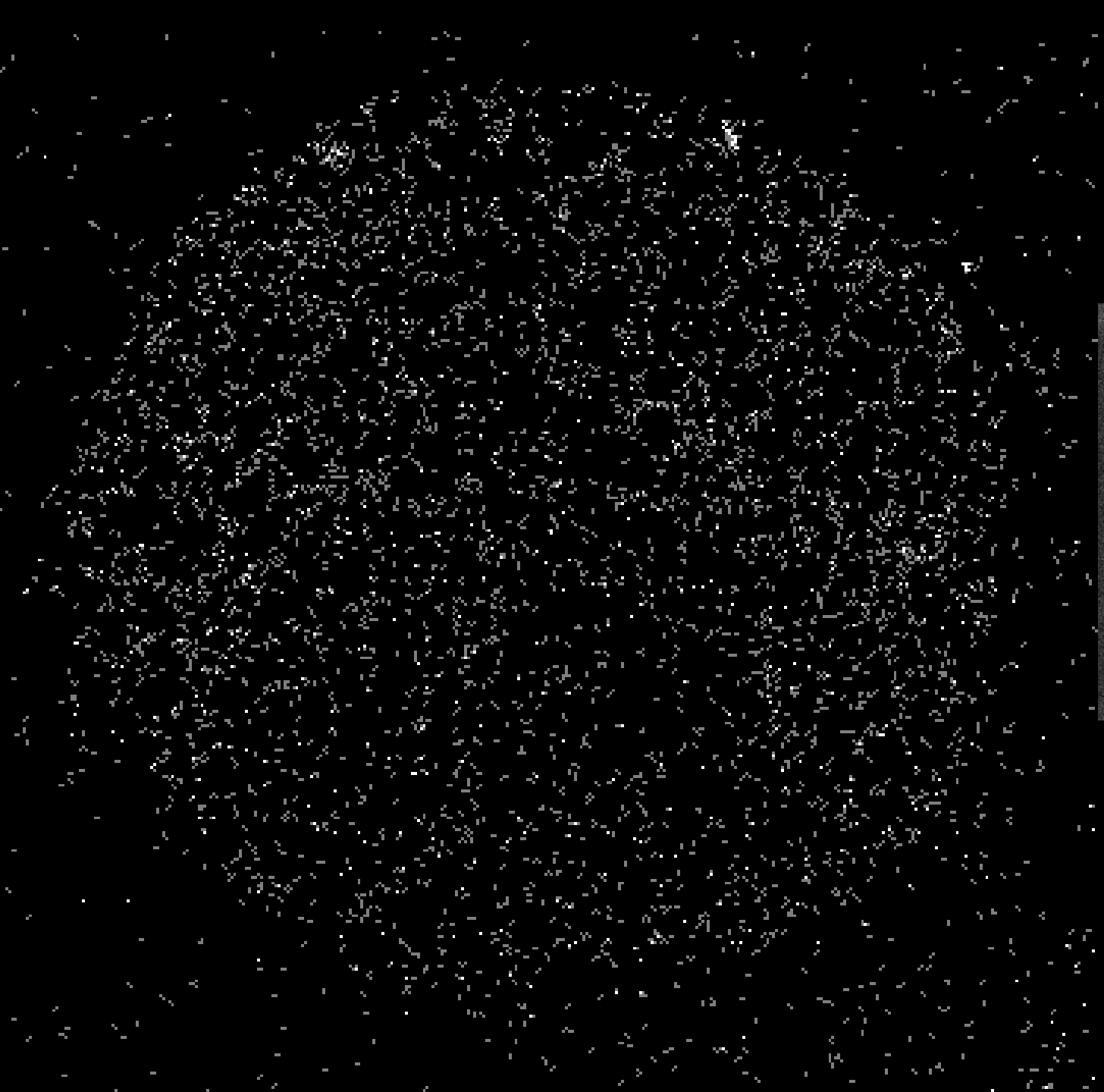
UV-TPEF at 4wk

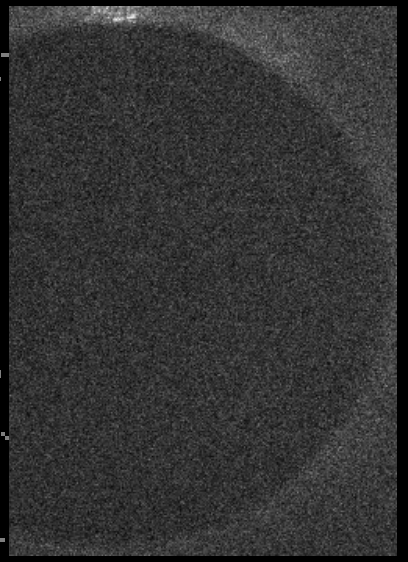
Protein 1 (crystals identified visually in other conditions)

Visual image where SHG/UV-TEV signal detected



Best optimized condition





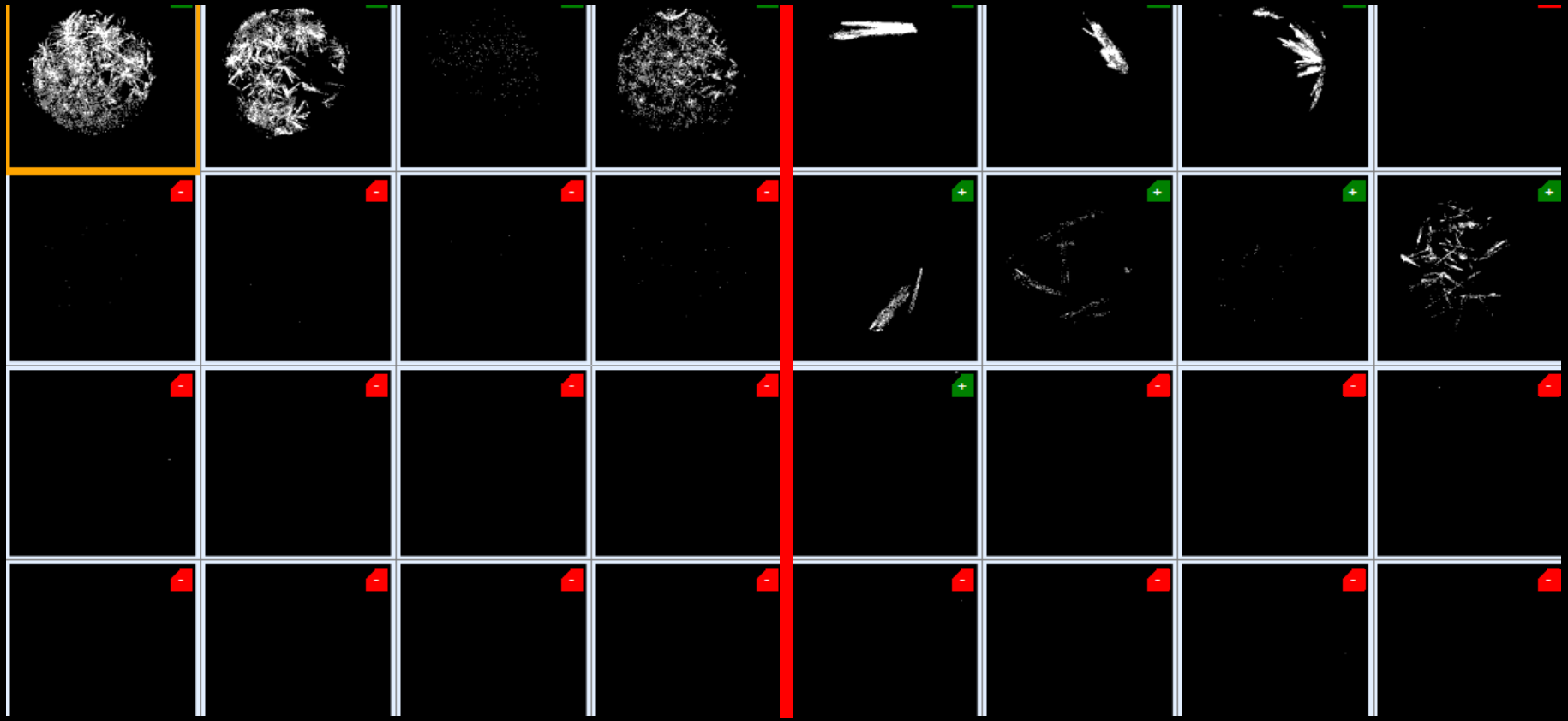
UV-TPEF

Hampton Research Ionic Liquids
5%(w/v) 1-Butyl-3-methyimidazolium dicyanamide

Protein 2

X14164- Full[P]-10 mg/ml

X14164- [P]/2- 5 mg/ml



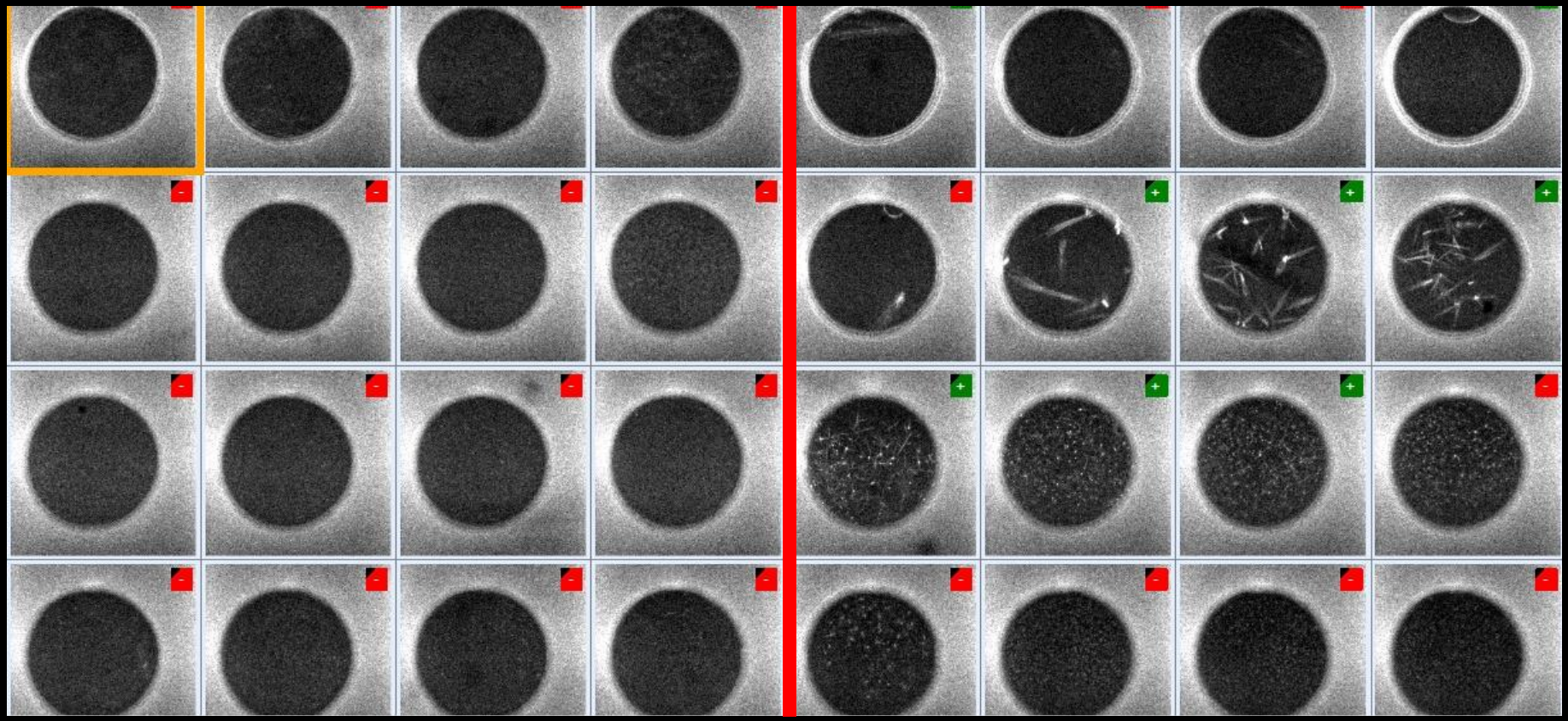
SHG at 4wk

Hampton Research Ionic Liquids
5%(w/v) 1-Butyl-3-methyimidazolium dicyanamide

Protein 2

X14164- Full[P]-10 mg/ml

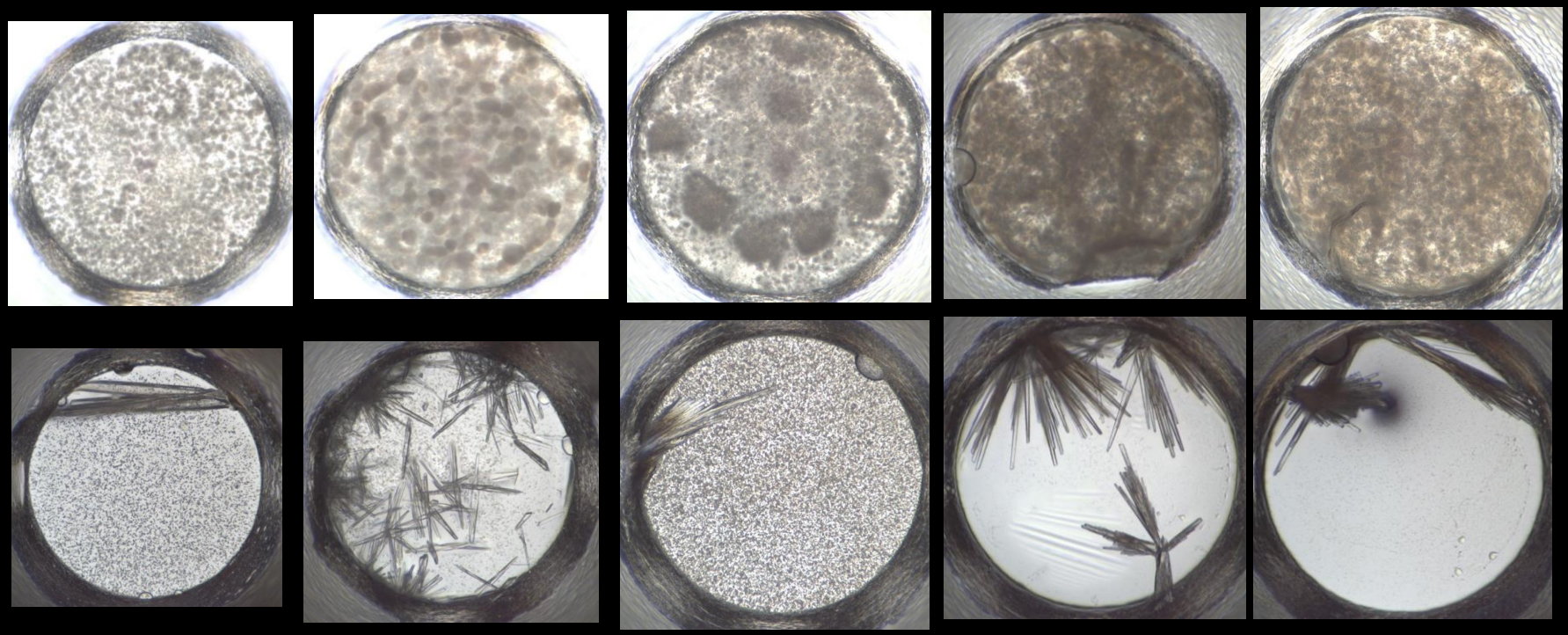
X14164- [P]/2- 5 mg/ml



UV-TPEF at 4wk

Protein 2 (crystals identified visually in other conditions)

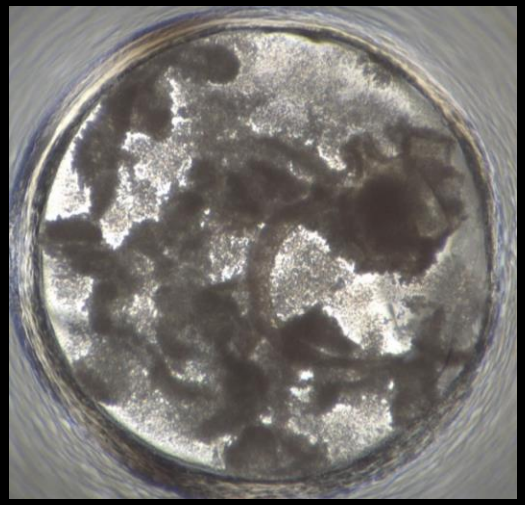
Visual image where SHG/UV-TEV signal detected



Best optimized condition

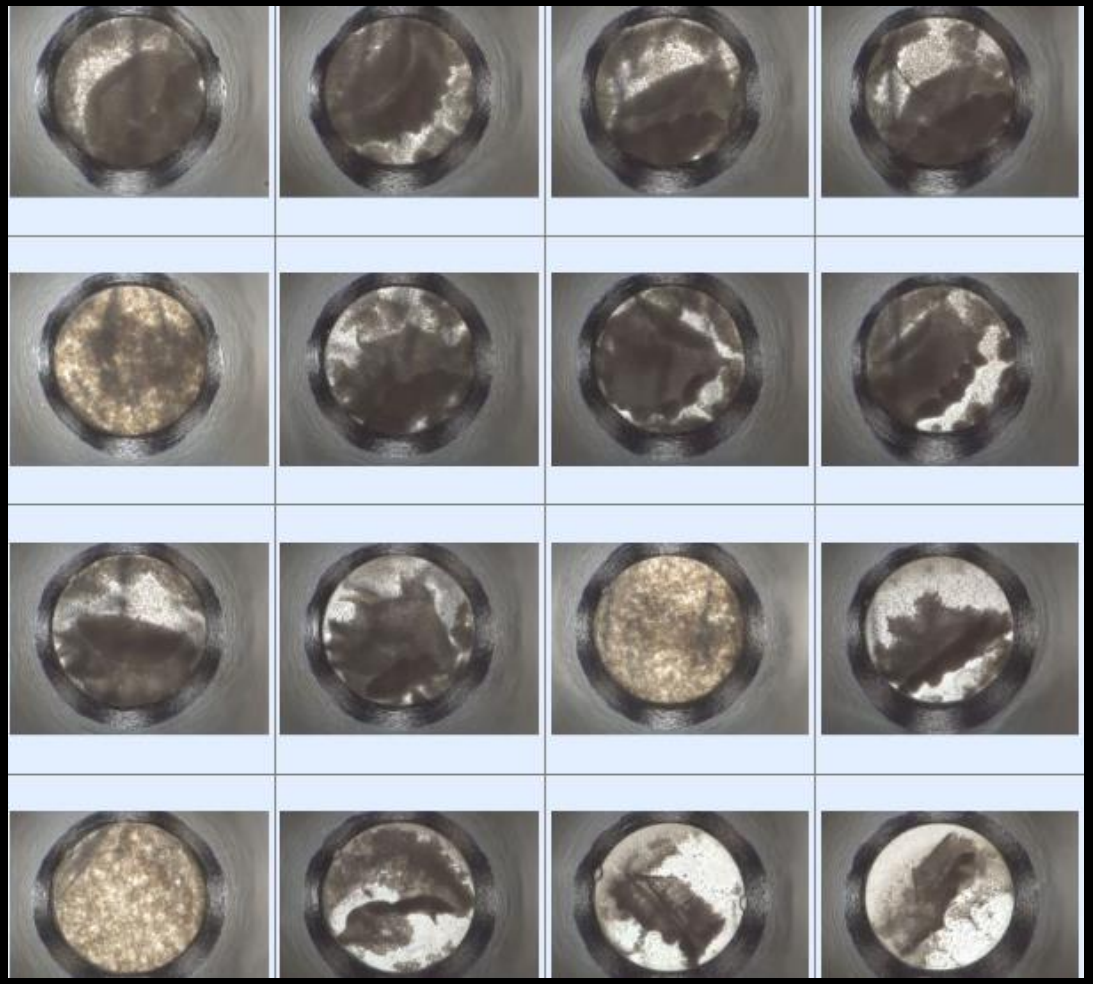
Protein 3 (no crystals identified visually in any conditions)

Visual image where SHG/UV-TEV signal detected



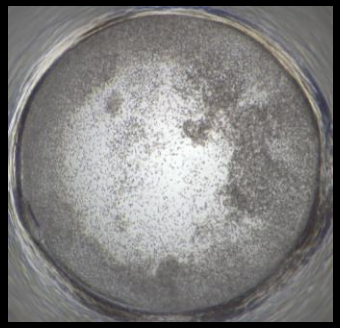
Initial hit

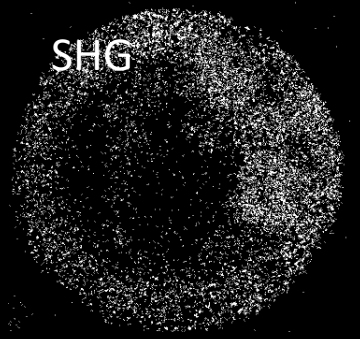
Optimization

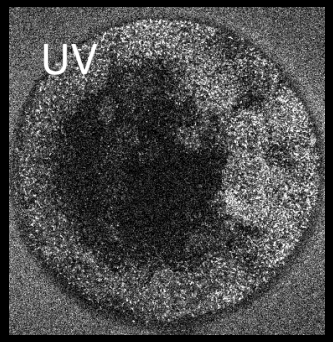


Protein 4 (crystals identified visually in other conditions)

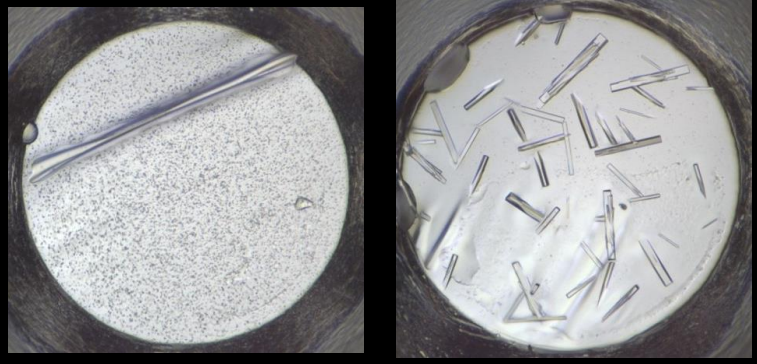
Visual image

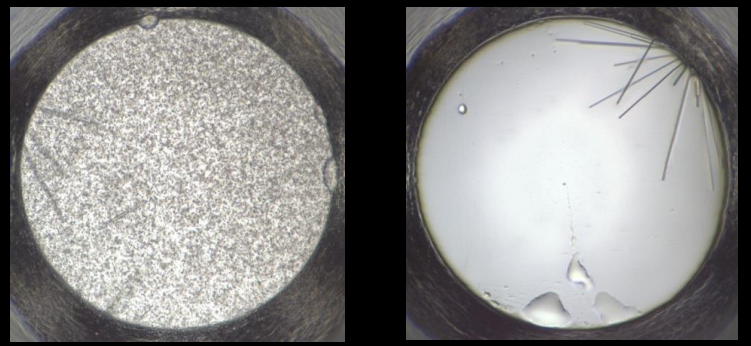






Optimization





Crystals in other 4 optimized conditons

Summary of the results

Protein	Macrocrystals in visible images	Microcrystals from SHG/UV-TPEF hits	Crystals out of 5 different trials	Success rate %
1	Yes	Yes	5	100
2	Yes	Yes	5	100
3	No	No	0	0
4	Yes	Yes	5	100
5	Yes	Yes	3	60
6	No	No	0	0

Summary of the results

Protein	Macrocrystals in visible images	Microcrystals from SHG/UV-TPEF hits	Crystals out of 5 different trials	Success rate %
1	Yes	Yes	5	100
2	Yes	Yes	5	100
3	No	No	0	0
4	Yes	Yes	5	100
5	Yes	Yes	3	60
6	No	No	0	0

Summary

Can we detect sub-micron crystals? Yes

Can we produce macrocrystals from these? Yes

Can we identify potential crystallization conditions missed visually?

Yes, SONICC/UV-TPEV identified microcrystals that were missed and produced optimized ‘macro’ crystals however, the SONICC/UV-TPEV imaging took place after the visual imaging and it is possible that other crystals could have appeared.

Questions still to be answered

Are nanocrystals produced where macrocrystal cannot be produced?

The answer is not available yet. To date over 460 proteins have been examined each in 1536 potential crystallization conditions with SONICC and UV-TPEV in our laboratory. The number of times a hit appears in a plate that has no other macrocrystal hits is almost zero from composite images (combined slices through the drop).

However, theoretically SONICC/UV-TPEF can visualize crystals down to 0.2 µm, a detailed visual analysis of each slice is ongoing but takes time. The first 9 with no macro crystals show no hits. Adapting an automated analysis approach may help but so far (with very preliminary data) it does not look promising.

Questions still to be answered

Can we go the opposite way, making a macrocrystal an naocrystal e.g. for time resolved studies or to get a few perfect domains?

- Yes in some cases. To achieve a large number of small crystals we need an initial crystallization condition and lots of nucleation with slow growth.
- Classically nucleation rate is given by $I=K\exp(-\Delta G/k_BT)$ where K, ΔG and k_B are constants for a given condition and T is temperature). Lowering temperature increases nucleation. ΔG can also be varied with supersaturation, increasing supersaturation increases nucleation.
- Crystal growth following nucleation is given by $\Delta\mu=k_BT\ln(1+S)$. Rate can be slowed by decreasing temperature and by decreasing supersaturation, S.
- However, not all proteins follow these rules and growth by a drop condensation phase has been reported which requires a minimum size of drop

Thank you and questions?

esnell@hwi.buffalo.edu