Crystals - how quaint! High-throughput developments for structural biology.

Edward H. Snell,

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

An introduction to the 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 for the general biomedical community and two Protein Structure Initiative large-scale structure production centers (NESG, Montelione, PI; SGPP/MSGPP, Hol, PI) and one PSI specialized PSI-2 center (CHTSB, DeTitta, PI).

The HTS lab screens samples against an incomplete factorial screen of two categories of crystallizing agents:

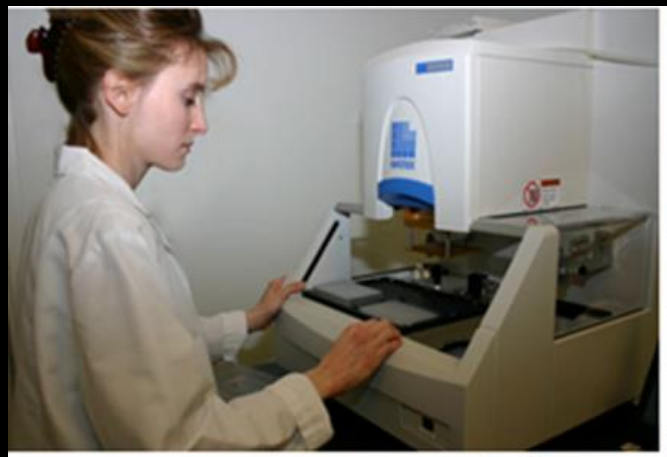
1. buffered (4<pH< 10), highly concentrated salts (35 salts total, sampling 18 different cations and 20 anions) – 229 conditions.
2. PEG/salt/buffer solutions (eight buffers (4<pH< 10), six molecular weight PEGs at three concentrations, and 35 salts at fixed 200 mM concentration) – 721 conditions.

Added to this is a screen of some 586 conditions encompassing screens commercially available from Hampton Research.

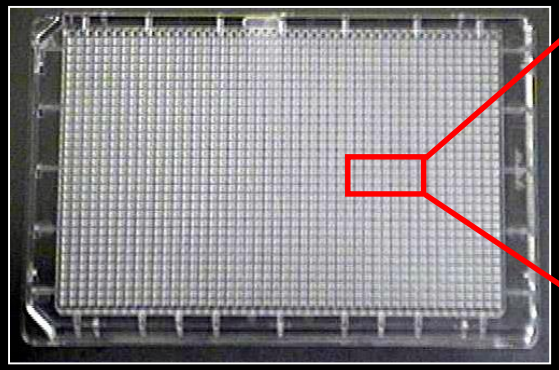
The crystallization method used is micro-batch under oil with 200 nl of protein solution being added to 200 nl of precipitant cocktail in each well of a 1536 well plate.

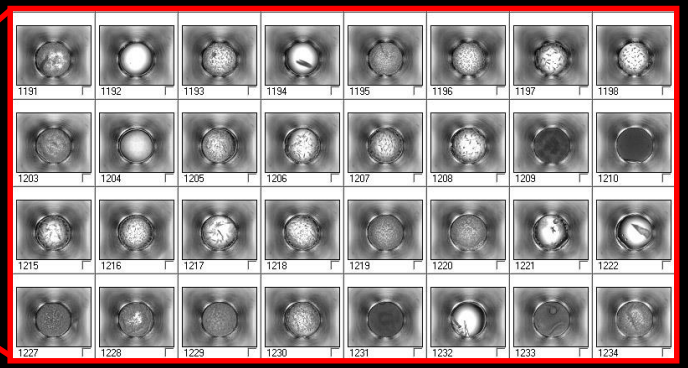
Wells are imaged before filling, immediately after filling then weekly for six weeks duration with images available immediately on a secure ftp server.

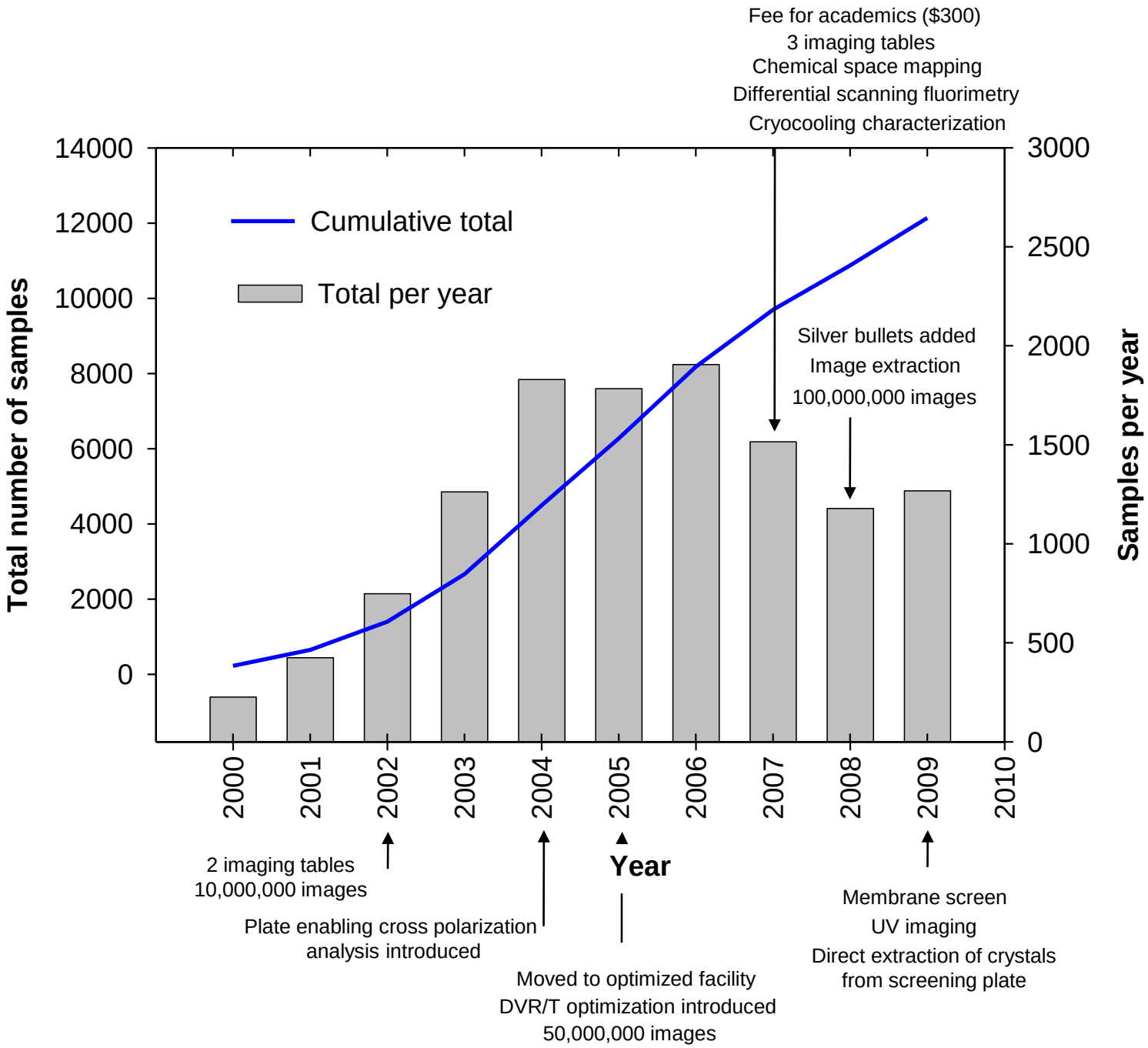
The HTSlab has investigated the crystallization properties of over 12,500 individual proteins archiving over 115,000,000 images of crystallization experiments.



The staff, instrumentation and crystallization plate used






Fee for academics ($300)
3 imaging tables
Chemical space mapping
Differential scanning fluorimetry
Cryocooling characterization
Cumulative total
Total per year
Silver bullets added
Image extraction
100,000,000 images
Total number of samples
Samples per year
14000
12000
10000
8000
6000
4000
2000
0
3000
2500
2000
1500
1000
500
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Year
2 imaging tables
10,000,000 images
Plate enabling cross polarization
analysis introduced
Moved to optimized facility
DVR/T optimization introduced
50,000,000 images
Membrane screen
UV imaging
Direct extraction of crystals
from screening plate

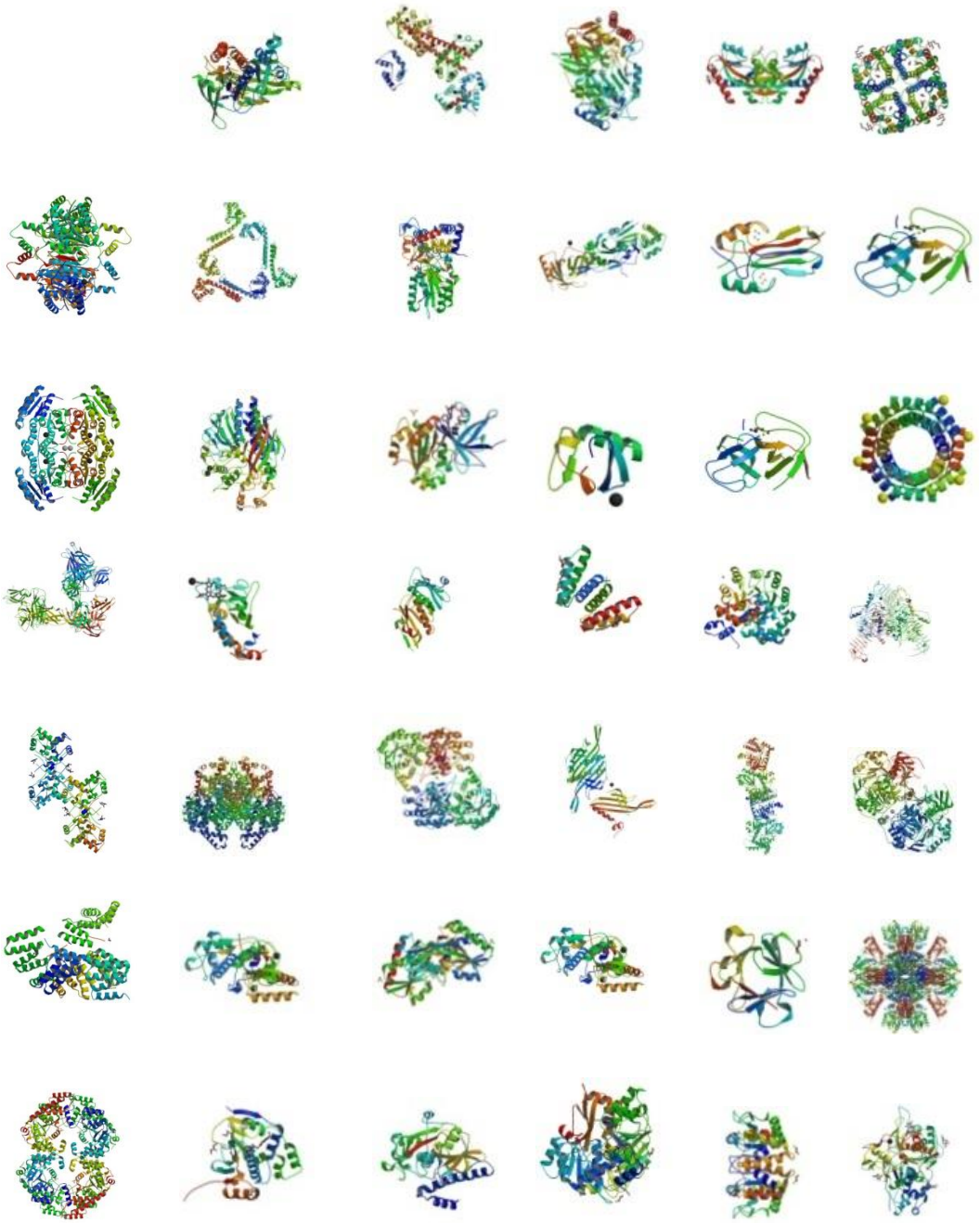
Born in Buffalo

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.



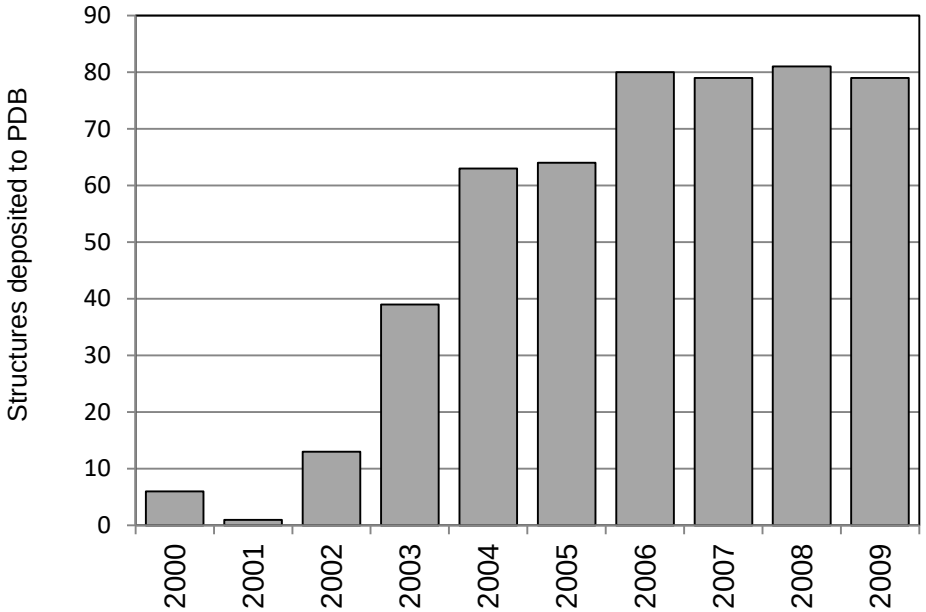
Where success is tracked.

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.

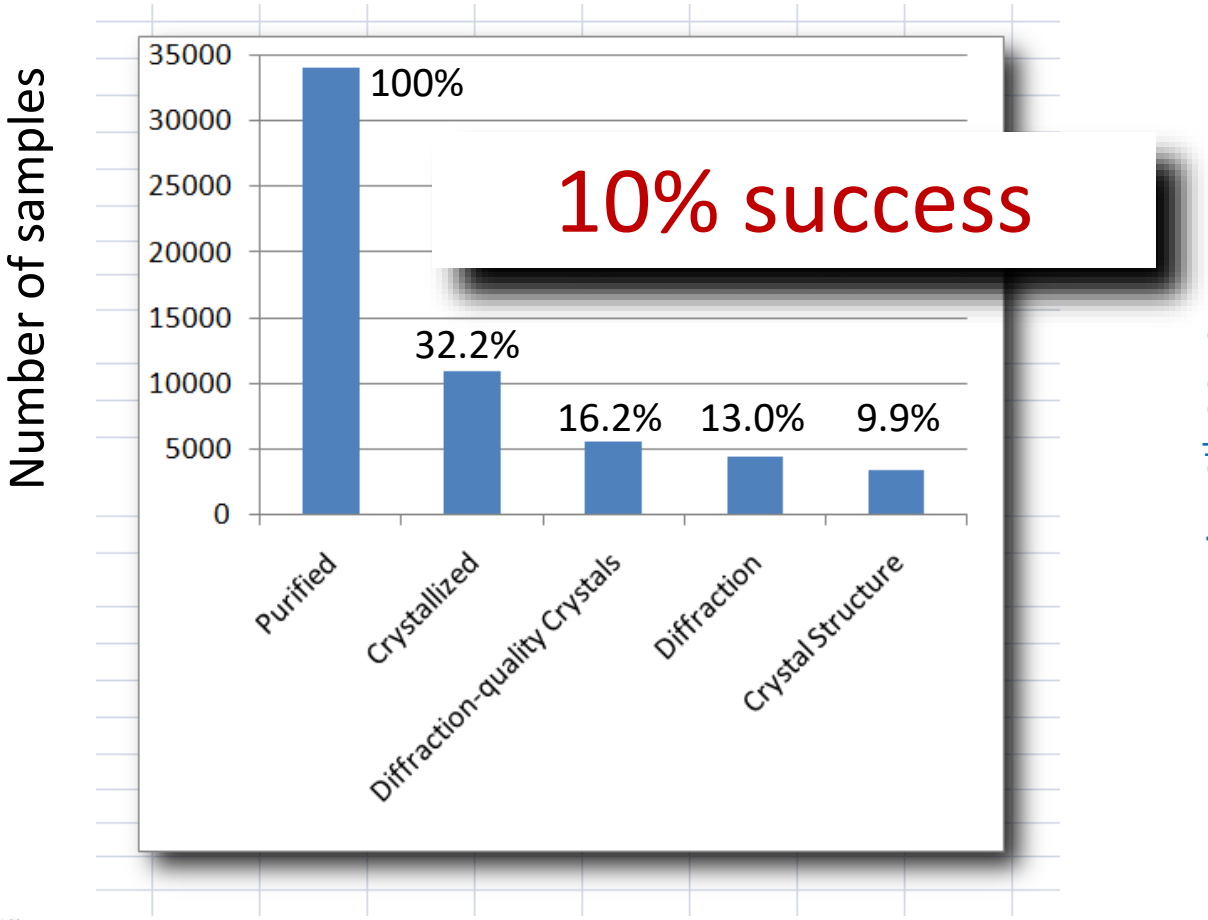
The graph demonstrates the ramp up of operations with maximum success reached from 2006 onward.

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.



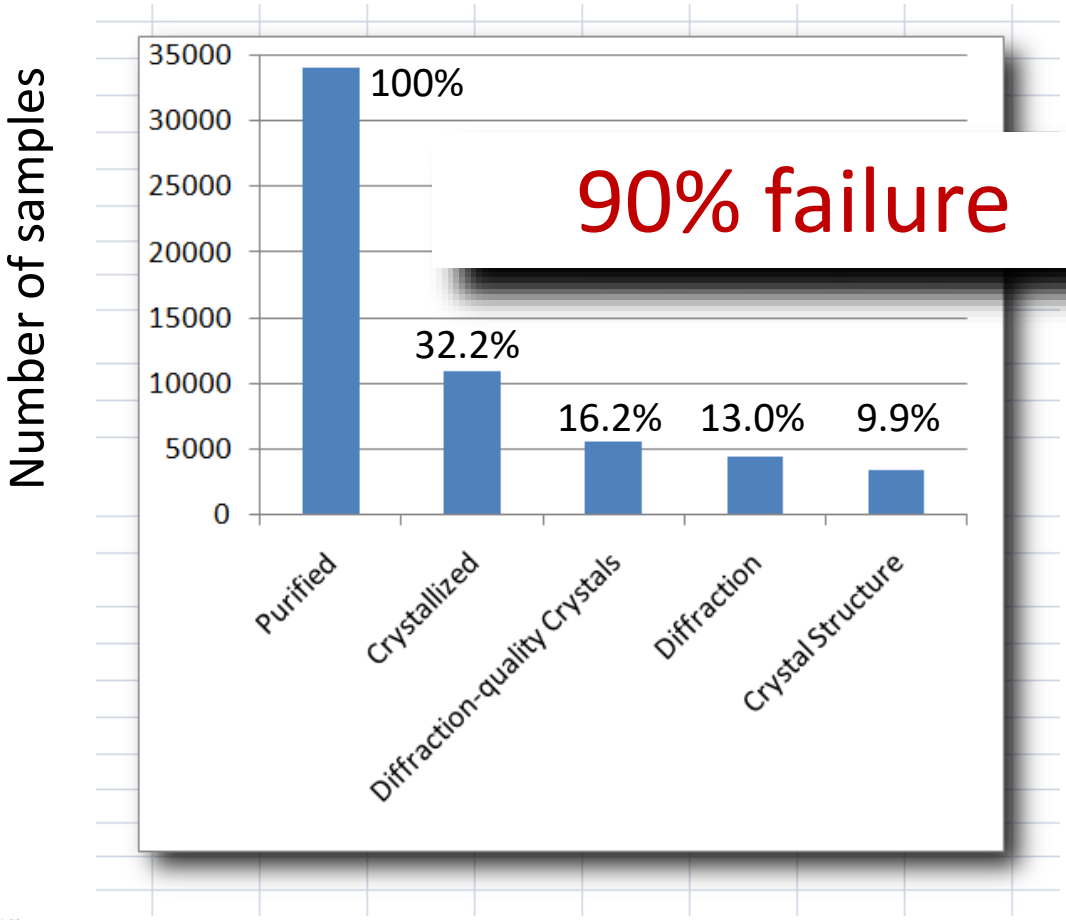
Only 9.9% of Protein Structure Initiative samples produced crystal structures.



Bioinformatics. 2004 Nov 1;20(16):2860-2.

Feb 16th 2010

90.1% of the Protein Structure Initiative samples failed to provide structures.



Feb 16th 2010,

Bioinformatics. 2004 Nov 1;20(16):2860-2.

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Is it the way we are crystallizing?

- Possibly but only subtle differences in results from vapor diffusion, batch, dialysis etc.
- More serious differences result from choice of temperature, pH etc.
- Evidence from DVR/T results.

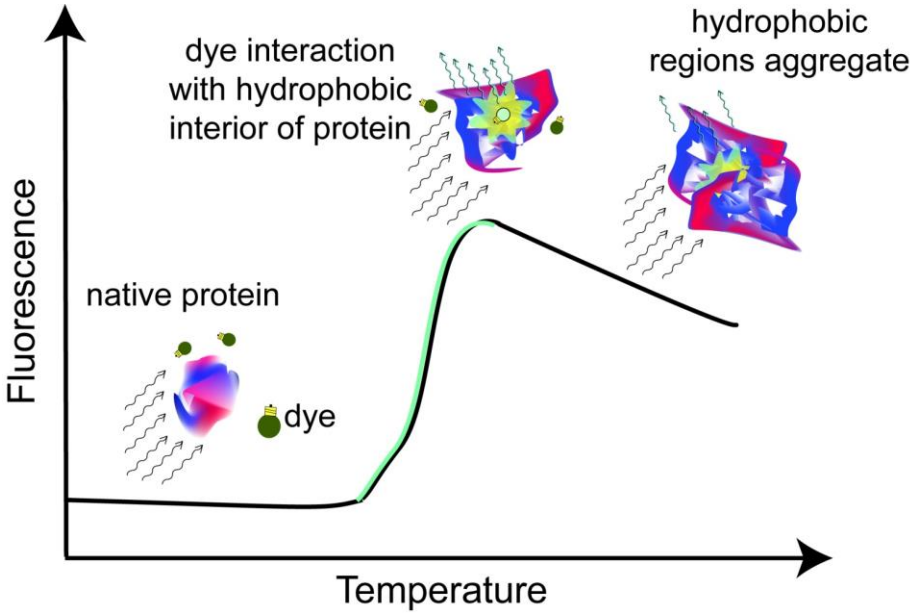
Is it the biochemistry?

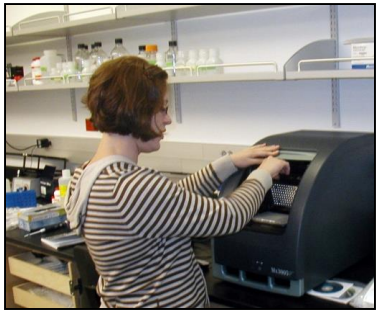
Did we sample to broad or too fine?

- Ask the sample:
  - What questions?
    - Response to biochemical conditions
  - What techniques?
    - Need a sensitive technique
    - Need a technique that requires minimal sample
    - Need a technique that provides an answer quickly

Ask the sample – What technique?

- Theromfluor®:
  - Mix sample with buffer, add Sypro Orange
  - Measure fluorescence signal as a function of temperature
  - Sample: 2 μL of 75 μM macromolecule per condition.





Ask the sample – What questions?

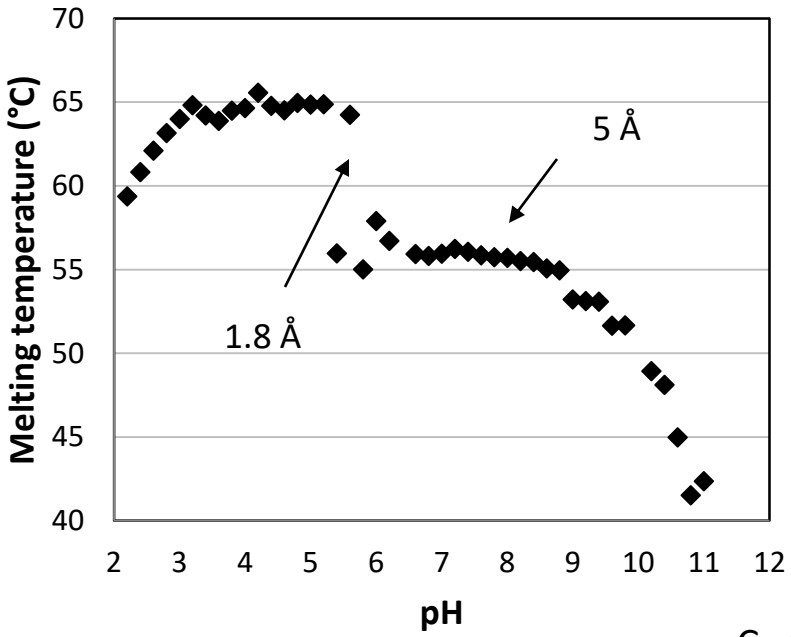
- What is important for the sample:
  - Response to pH
  - Response to Hofmeister series salts
  - Response to presence of sugars
  - Response to reducing agents
  - Proteolysis

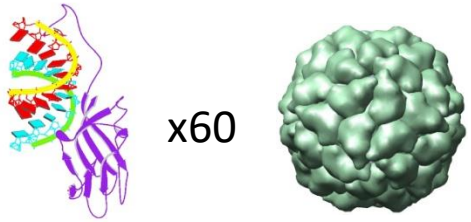
Ask the sample – What questions?

- What is important for the sample:
  - Response to pH
  - Response to Hofmeister series salts
  - Response to presence of sugars
  - Response to reducing agents
  - Proteolysis

All the following data was recorded by Elizabeth Snell

Satellite tobacco mosaic virus





Satellite tobacco mosaic virus (STMV) can undergo at least two physical transitions that significantly alter its mechanical and structural characteristics. At high pH the 17-nm STMV particles expand radially by about 5 Å to yield particles having diameters of about 18 nm...

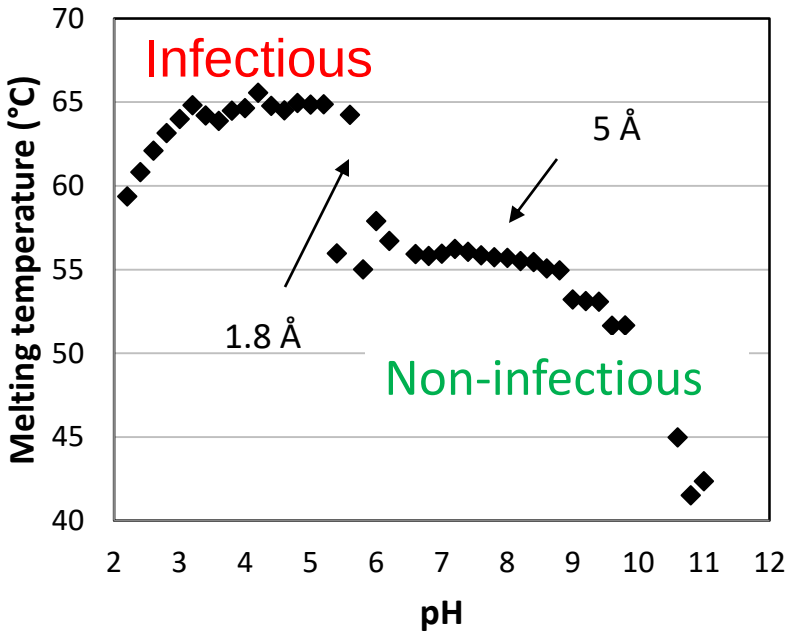
…While the native 17-nm particles crystallize as orthorhombic or monoclinic crystals which diffract to high resolution (1.8 Å), the enlarged 18-nm particles crystallize in a cubic form which diffracts to no better than 5 Å.

Kuznetsov, Larson, Day, Greenwood, and McPherson. Virology 284, 223-234 (2001).

Currently no data in the literature supports the prediction of crystallization conditions from T_m values. only the identification of ligands that stabilize macromolecules to improve crystallization outcomes

Higher melting temperature does not indicate better diffraction.

Interesting Aside

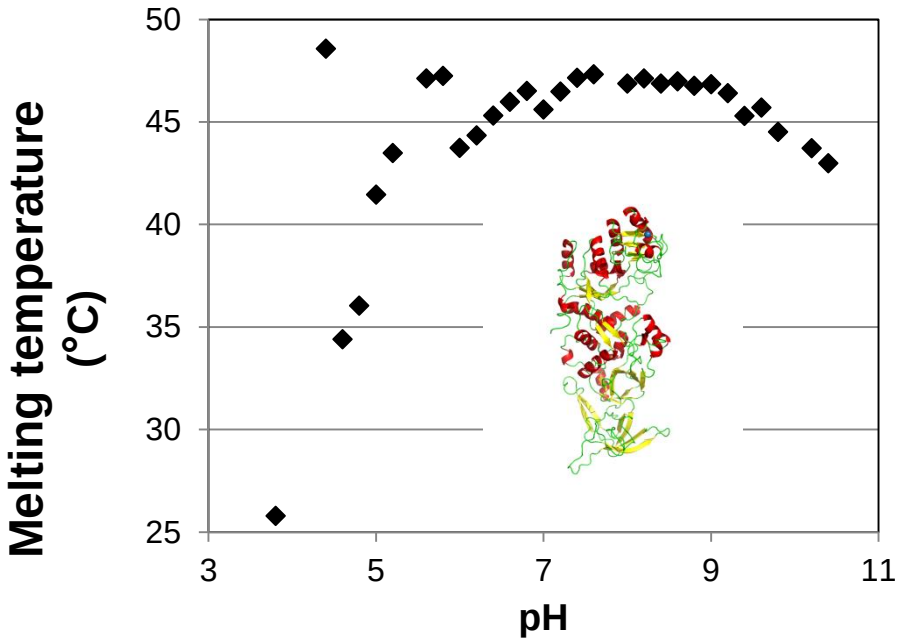


We know where to 'trap' virus particles to look at their dynamic mechanism – a whole new talk.

"In the life-cycles of viruses, dramatic morphological changes in their capsid structure are needed *to allow them to carry out the diverse set of functions required* ***for replication****. All virus capsids must form readily, have structural integrity, and have the proper biological trigger in order to be infectious."* *Canady et al., Journal of Molecular Biology, 299 573-584 (2000)*

We have an assay to determine if a virus particle is functional and to develop lead drug candidates – i.e. mix a quantity of potential therapeutic compounds and look for a lack of shift in melting temperature across the pH range (or other conditions) of interest

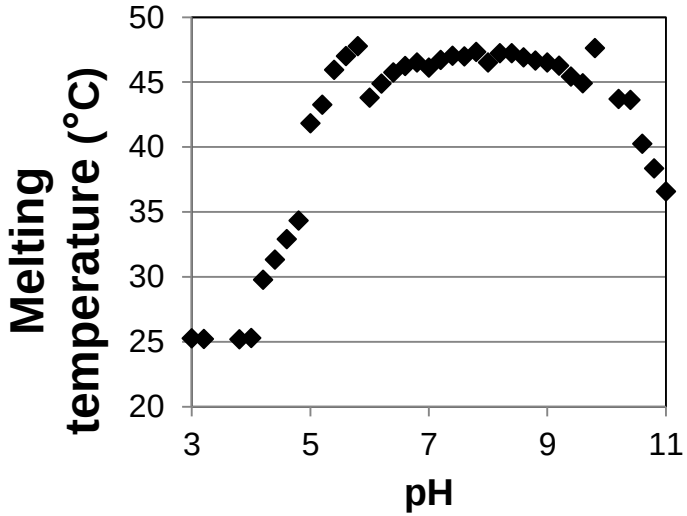
Gln-4

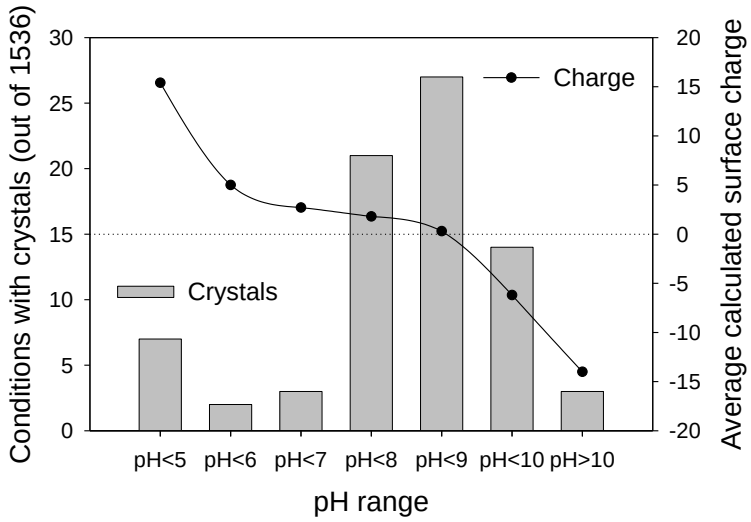


Preferred crystallization pH – 7

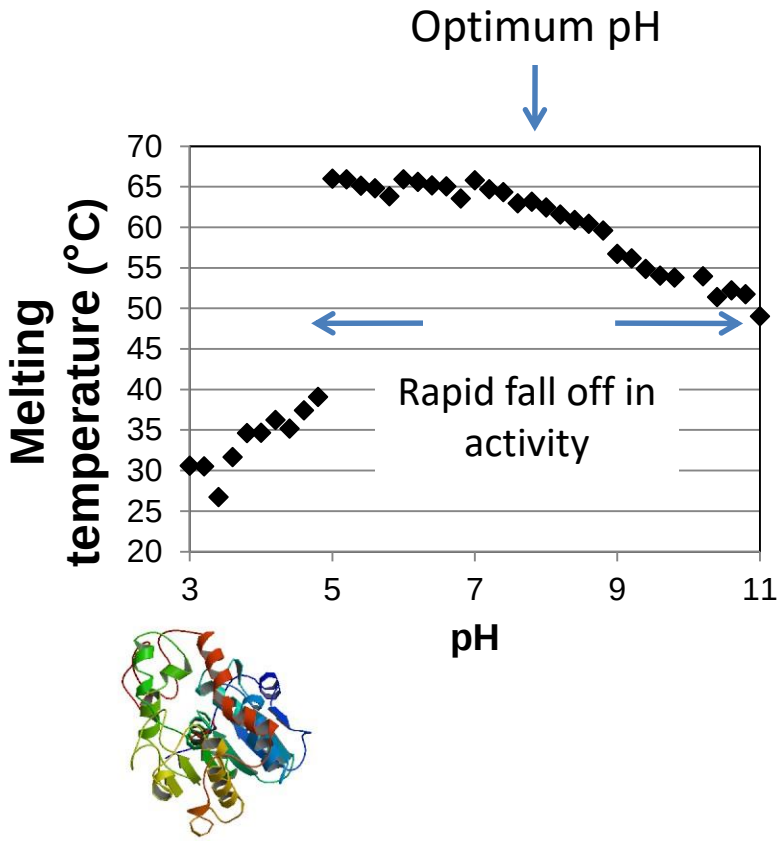
2.3A data collected on N-terminal arm

N-terminal arm





Lipase

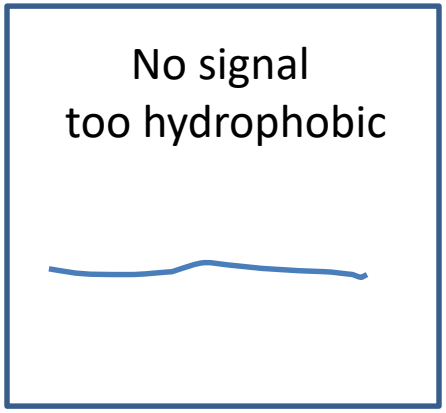


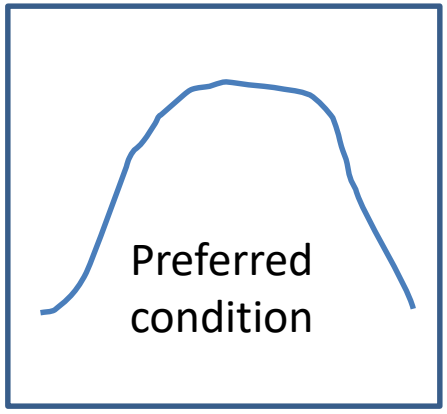
Structure	PO_4 buffer, pH 4.2 (%)	KAc buffer, pH 5.0 (%)	PO_4 buffer, pH 6.0 (%)	Cacod buffer, pH 7.0[a] (%)	PO_4 buffer, pH 9.0 (%)
α-Helix	32	38	37	38	30
β-Strand	21	20	22	20	25
Turns	15	14	11	18	15
Other	31	28	30	24	30

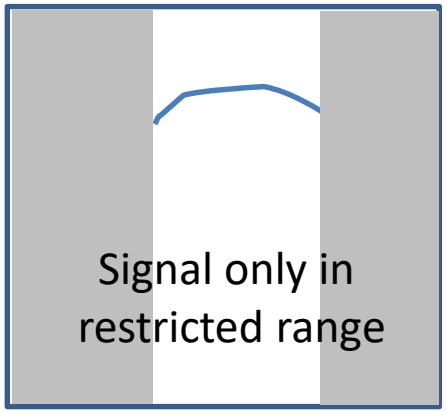
McCabe at al. Enzyme and Microbial Technology,36,70-74 (2005).

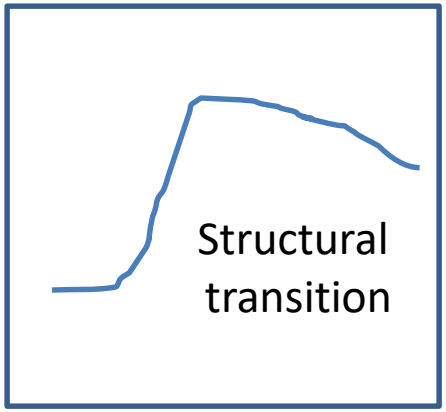
The pH screen has identified a structural transition. This is in agreement with CD data. Our structural knowledge is of the low pH form.

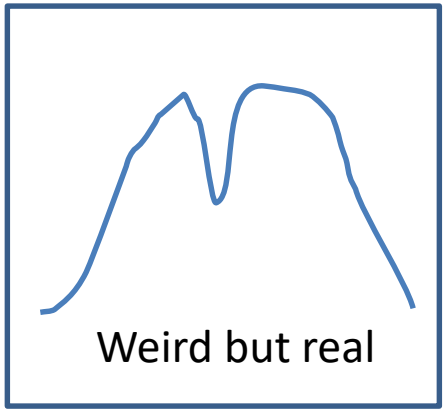
What signatures have been seen?





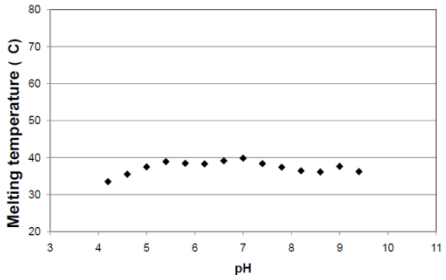




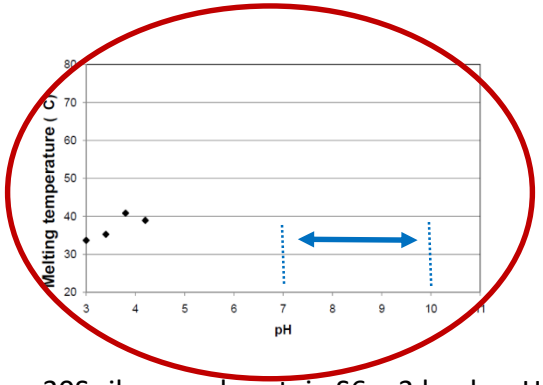


Samples to date.

What signatures have been seen?

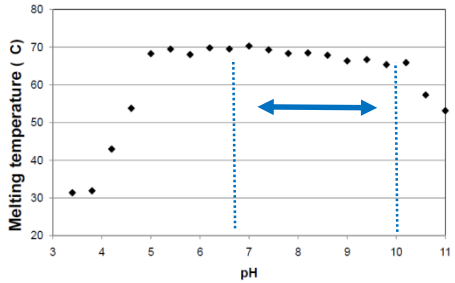


A Possible ATP-dependent DNA helicase RecG-related protein – no crystallization leads.

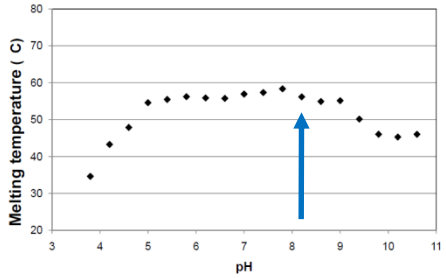


30S ribosomal protein S6e, 3 leads, pH 7-10, nitrate and sulfate

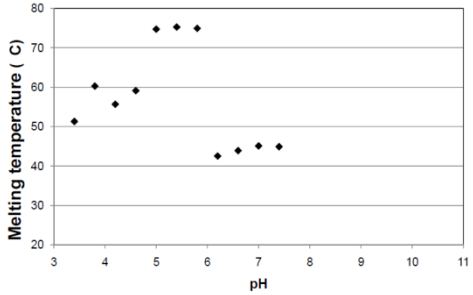
Candidate to test for salt?



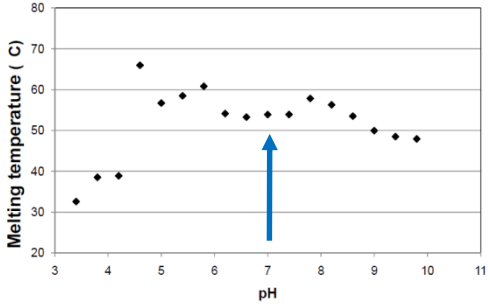
Putative diphthamide synthesis protein, 17 leads, pH 6.8-10



Putative uncharacterized protein– 1 crystallization lead, pH 8.2.

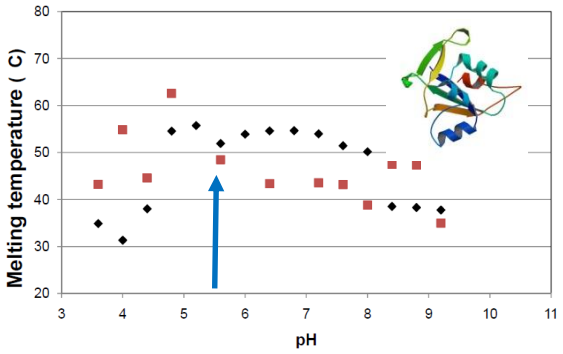


Phage integrase – no leads



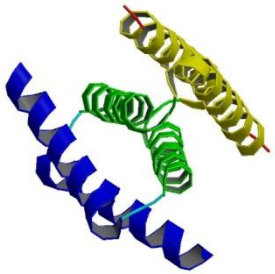
ATP-dependent DNA ligase – 1 lead, pH 7

What signatures have been seen?

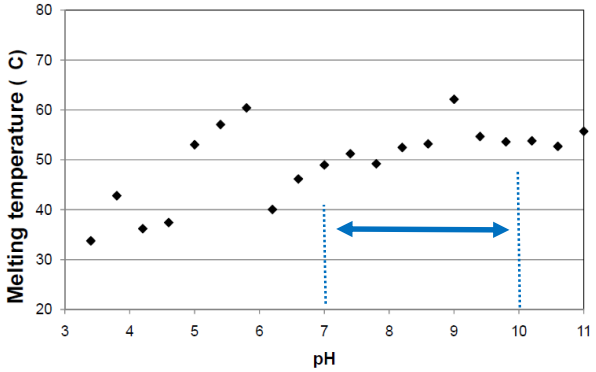


Hypothetical Protein from Caulobacter Crescentus – crystallization pH 5.6

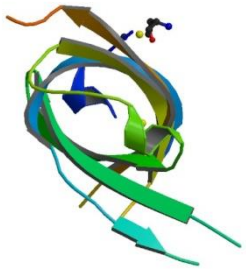
No Thermofluor® signal



Protein CC0527 (V27M / L66M double mutant) from Caulobacter crescentus .

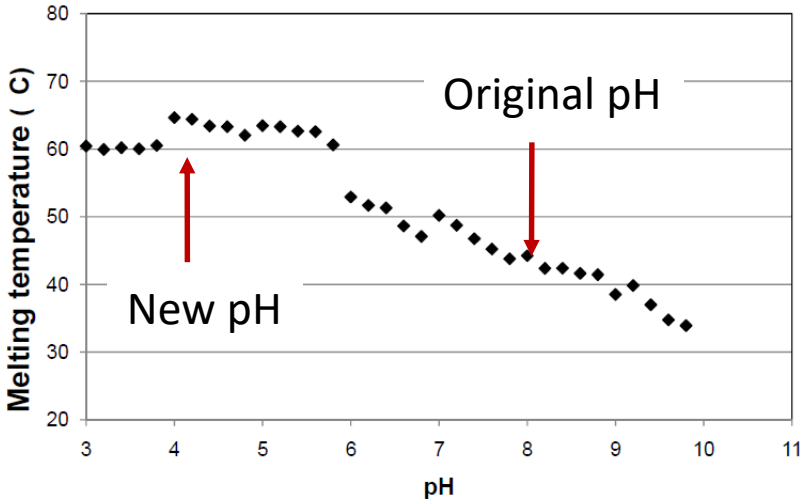


Tyrosine-protein kinase Tec – 3 leads, pH 7, 9 and 10



domain of replication protein A from Methanococcus maripaludis

An aside: Application to a protein from the Schultz lab



Influenza A Matrix protein (InfAM1).

InfAM1 is a protein that has substantial loss of protein during the first step of purification. Preliminary optimization of the purification protocol involved extensive screening of buffer conditions.
Acceptable, but non-optimal, conditions that gave decent protein yield were 50 mM Hepes pH 7.5, 150 mM NaCl, 5 mM DTT, 10%.

The protein under these conditions precipitates over time, and will not concentrate beyond 1.4 mg/mL

A pH scan was performed showing that InfAM1 favors acidic pHs with the Tm highest pH 4.0 with a Tm of 64.6 C.

At higher pHs the Tm declines significantly, at pH 7.5 that we were working at, the Tm is 20 degrees

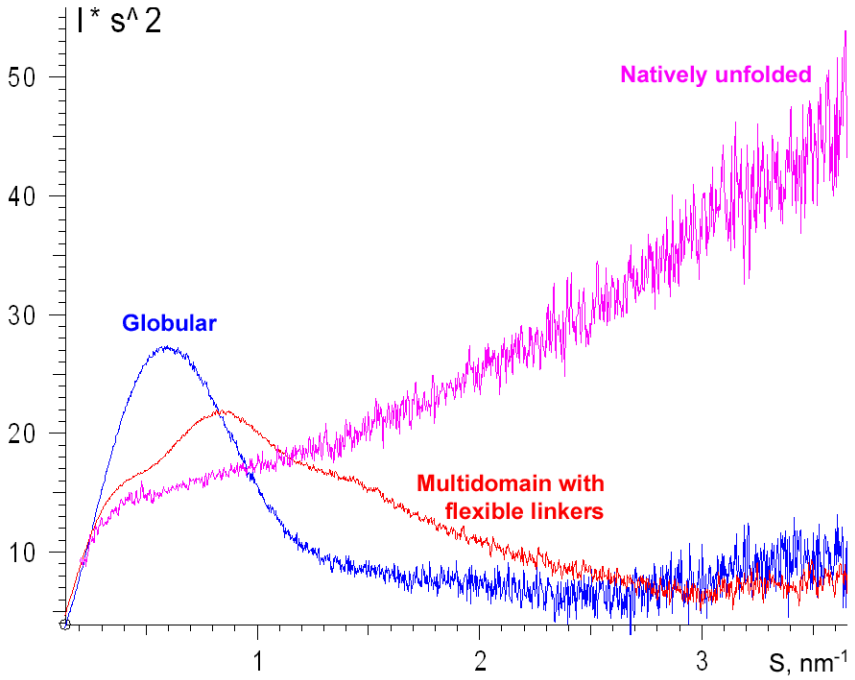
Changing to an acidic pH the Shultz lab have not observed anymore precipitation problems and were able to concentrate the protein beyond 6 mg/mL

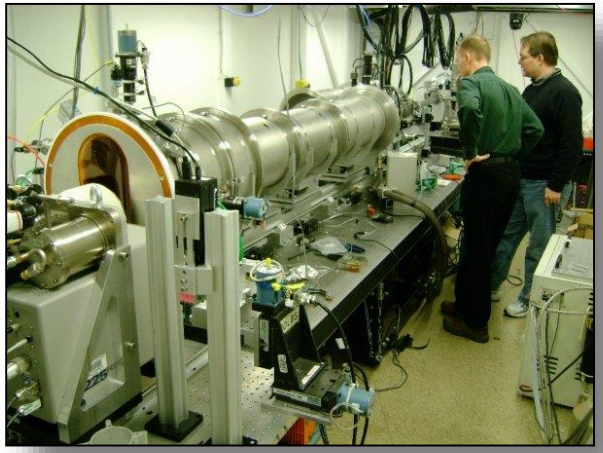
Thanks to L. Wayne Schultz and Paige (Pei) Chun Hang

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

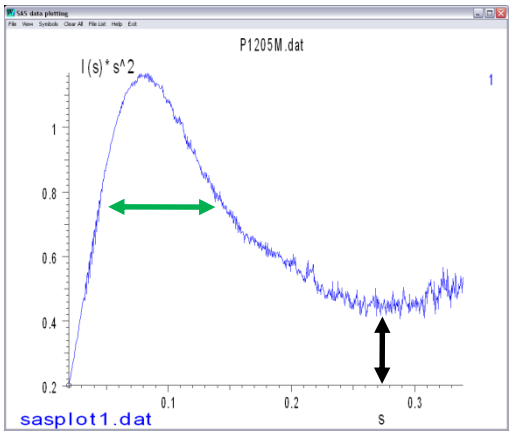
Look at the sample in solution.

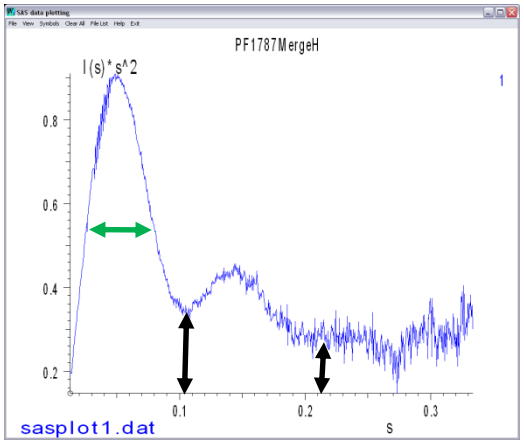


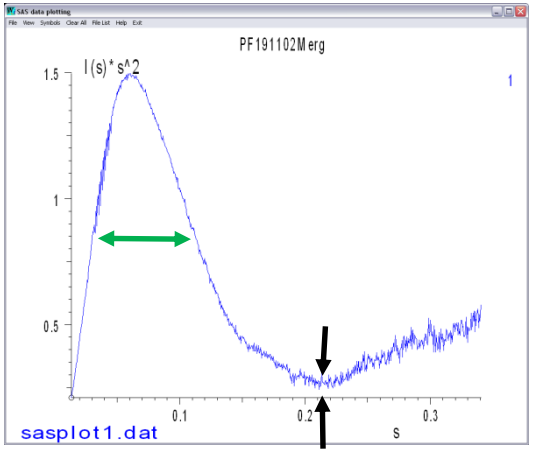


SAXS sensitive to aggregation (raw data), multimer state (intercept, radius of gyration) and the 'globularity' of the sample (Kratky plot).

Look at the sample in solution.







The Kratky plot indicates 'globularity' – We propose, for well folded samples, that the crystallizability in any condition is related to the full width at half maximum of the initial peak and the height above the axis of the second turning point.

We plan to test this linking the Thermofluor® assay with SAXS as a function of pH to identify conditions that plateau in the Thermofluor® where SAXS indicates the globularity is maximized and the radius of gyration in minimized.

How are we doing this?

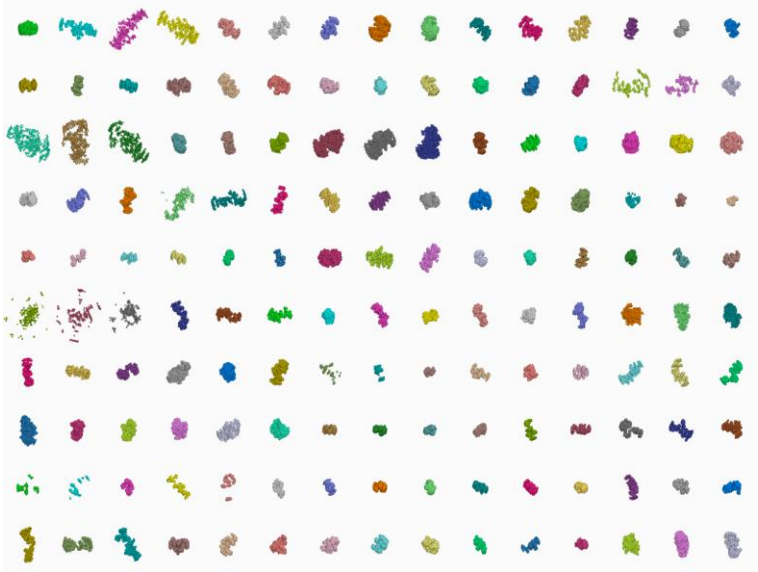
- Developing high-throughput SAXS methods in collaboration with SSRL.
- Current protocol
  - 3 concentrations, 8 x 3s exposures at each, 24s of beamtime
  - 12 minutes per sample (most time spent cleaning and liquid handling)
  - 5 samples per hour
  - 24 samples automatically collected in 4.8 hours
  - Potential of 360 samples every beamtime
- Actual experience
  - Occasional beamdump, loading error etc.
  - Realistic ~250 samples per beamtime.
  - Studying NESG samples (and others) ~300 NESG samples
  - 3,000 in the freezer – 10 beamtimes to complete current stock

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Statistics from NESG samples run to date

	NESG	%
Total number of samples	400	
Total processed to date	145	36%
Well folded	101	69% of processed
Aggregated	21	14% of processed
Other	18	12% of processed
Crystal structure	17*	12% of processed
	Others	
	50	Various stages

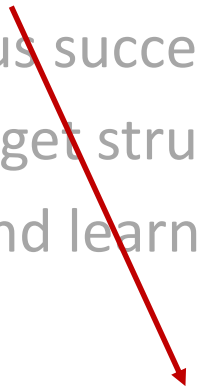


*As of December 2009

12% of submitted samples in this batch produced crystal structures yet 69% are globular and well folded. 14% are aggregated which may be a result of freeze/thaw cycles. The other 12% represent natively unfolded samples or experimental or practical problems.

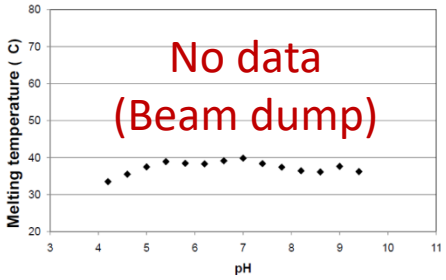
Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

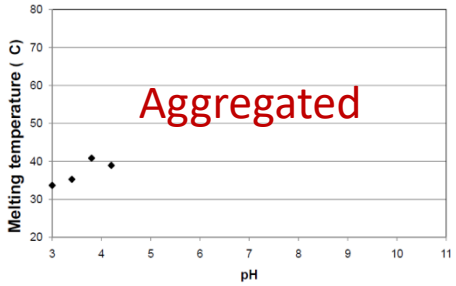


SAXS analysis suggests we should be able to crystallize at least 69% of our samples. Most of the failure is not a sample problem.

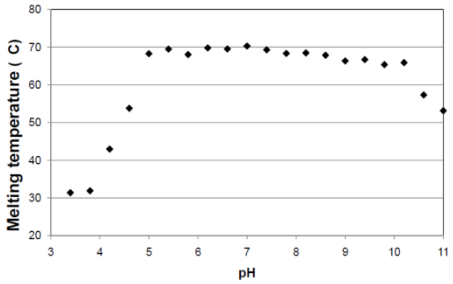
What does SAXS tell us?



A Possible ATP-dependent DNA helicase RecG-related protein – no crystallization leads.

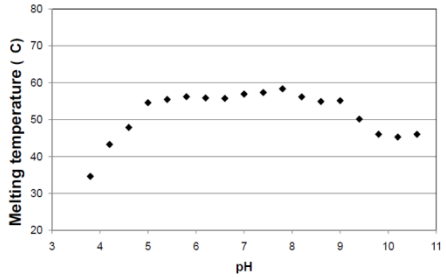


30S ribosomal protein S6e, 3 leads, pH 7-10, nitrate and sulfate



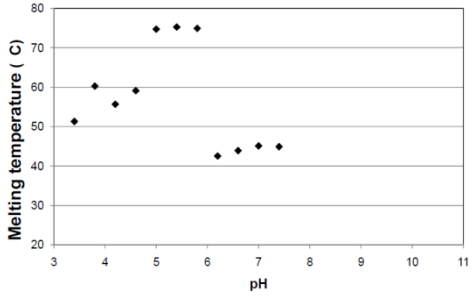
Data to be reprocessed

Putative diphthamide synthesis protein, 17 leads, pH 6.8-10



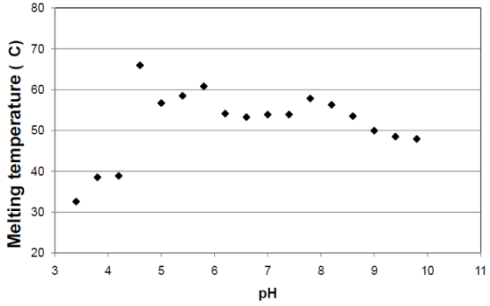
Nice globular protein

Putative uncharacterized protein– 1 crystallization lead, pH 8.2.



Nice globular protein

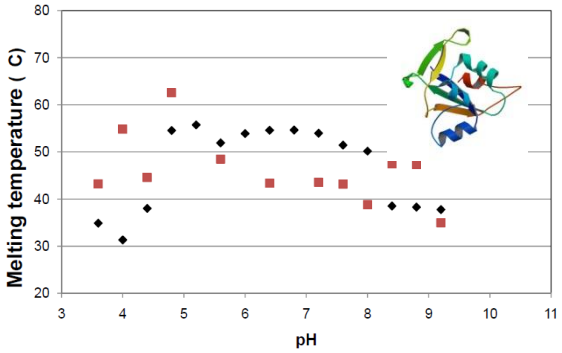
Phage integrase – no leads



Nice globular protein

ATP-dependent DNA ligase – 1 lead, pH 7

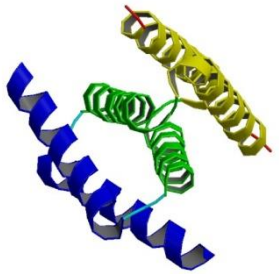
What does SAXS tell us?



Nice globular protein

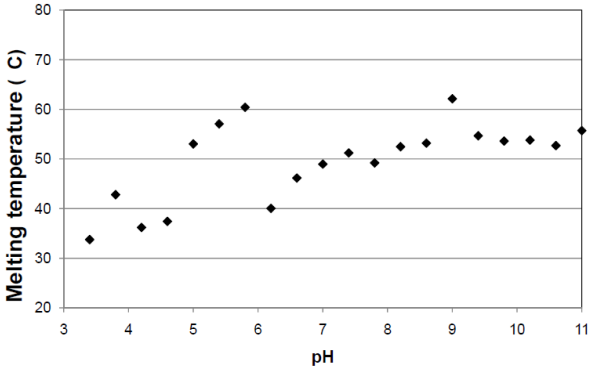
Hypothetical Protein from Caulobacter Crescentus – crystallization pH 5.6

No Thermofluor® signal



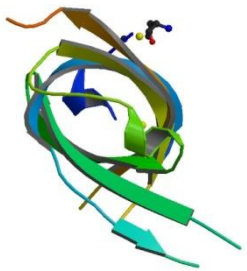
Nice globular protein

Protein CC0527 (V27M / L66M double mutant) from Caulobacter crescentus (-0.032)



Nice globular protein

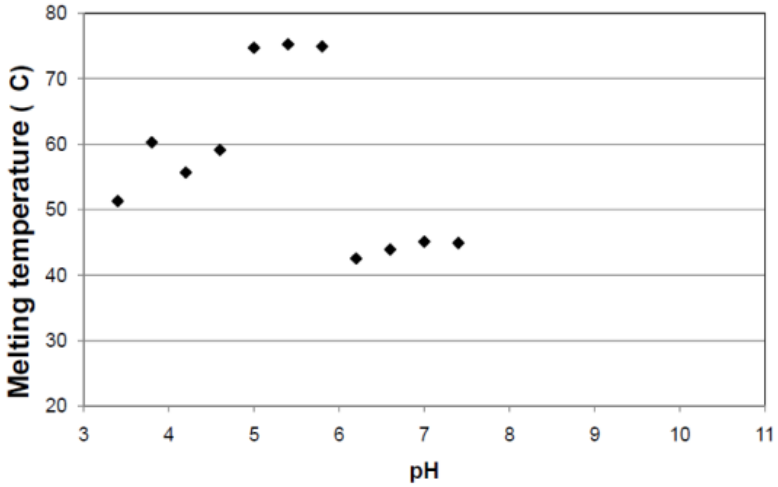
Tyrosine-protein kinase Tec – 3 leads, pH 7, 9 and 10



Nice globular protein

domain of replication protein A from Methanococcus maripaludis (-0.424)

What would we like SAXS to tell us?



Nice globular protein

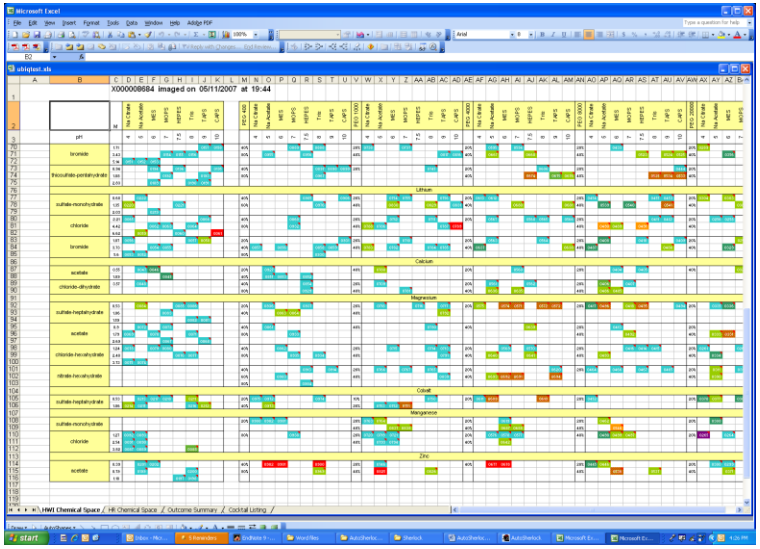
Phage integrase – no leads

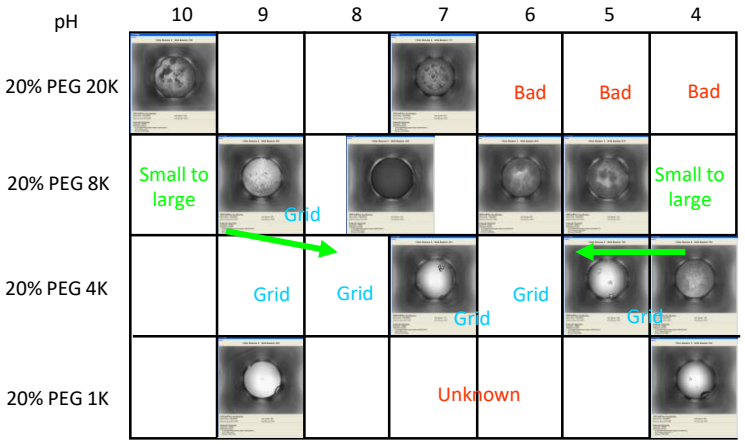
- A large structural change occurs at pH 6.
- Is this structurally meaningful or an aggregation effect.
- SAXS as a function of pH

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

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

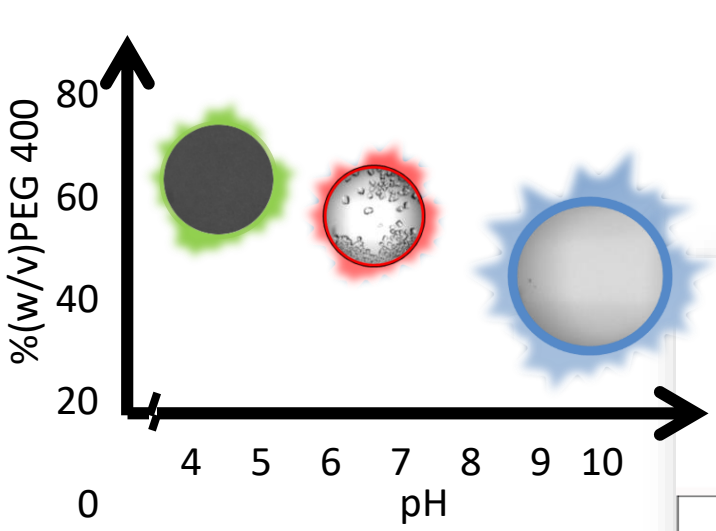
AutoSherlock: a program for effective crystallization data analysis

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)

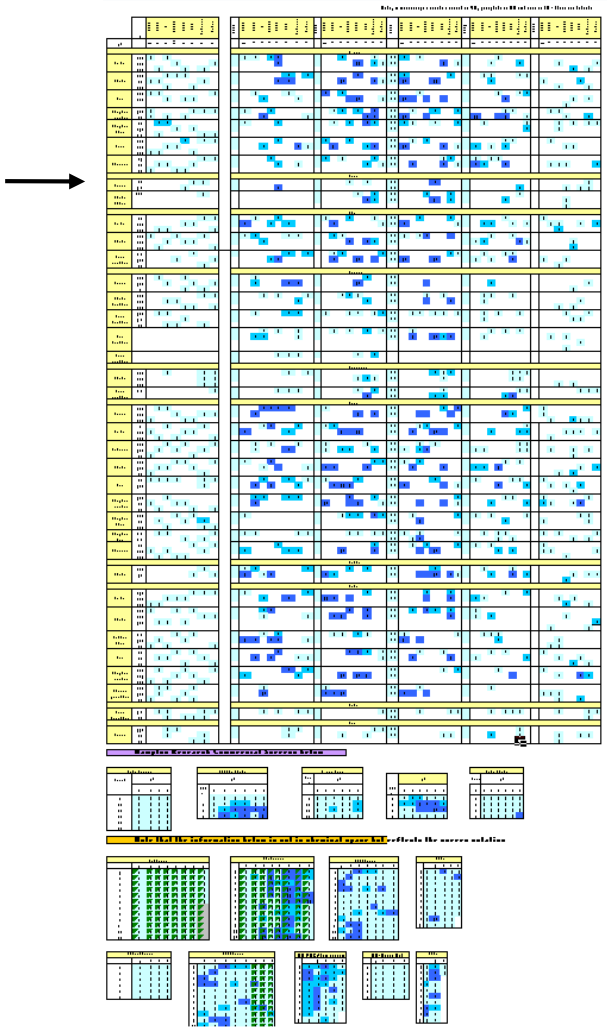


Un-sampled
Crystals
Clear
Precipitate

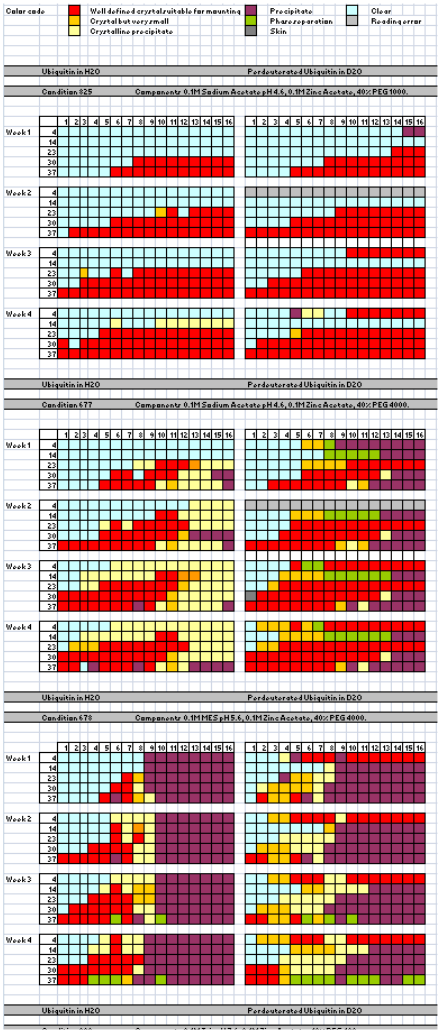
		PEG 400	Na Citrate	Na Acetate	MES	Bis-Tris Propane	HEPES	Tris	TAPS	CAPS
			Lithium							
pH			4.2	5	6	7	7.5	8	9	10
bromide		20%								●
		40%	●		●			●	●	
		80%						●		

Chemical space mapping

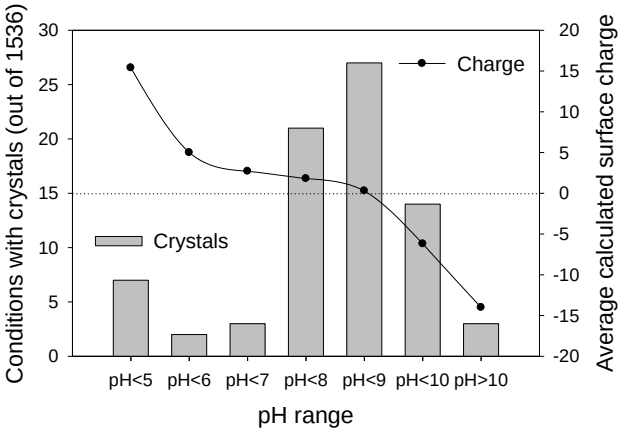
Analysis as a function of the entire cocktail screen and multiple proteins



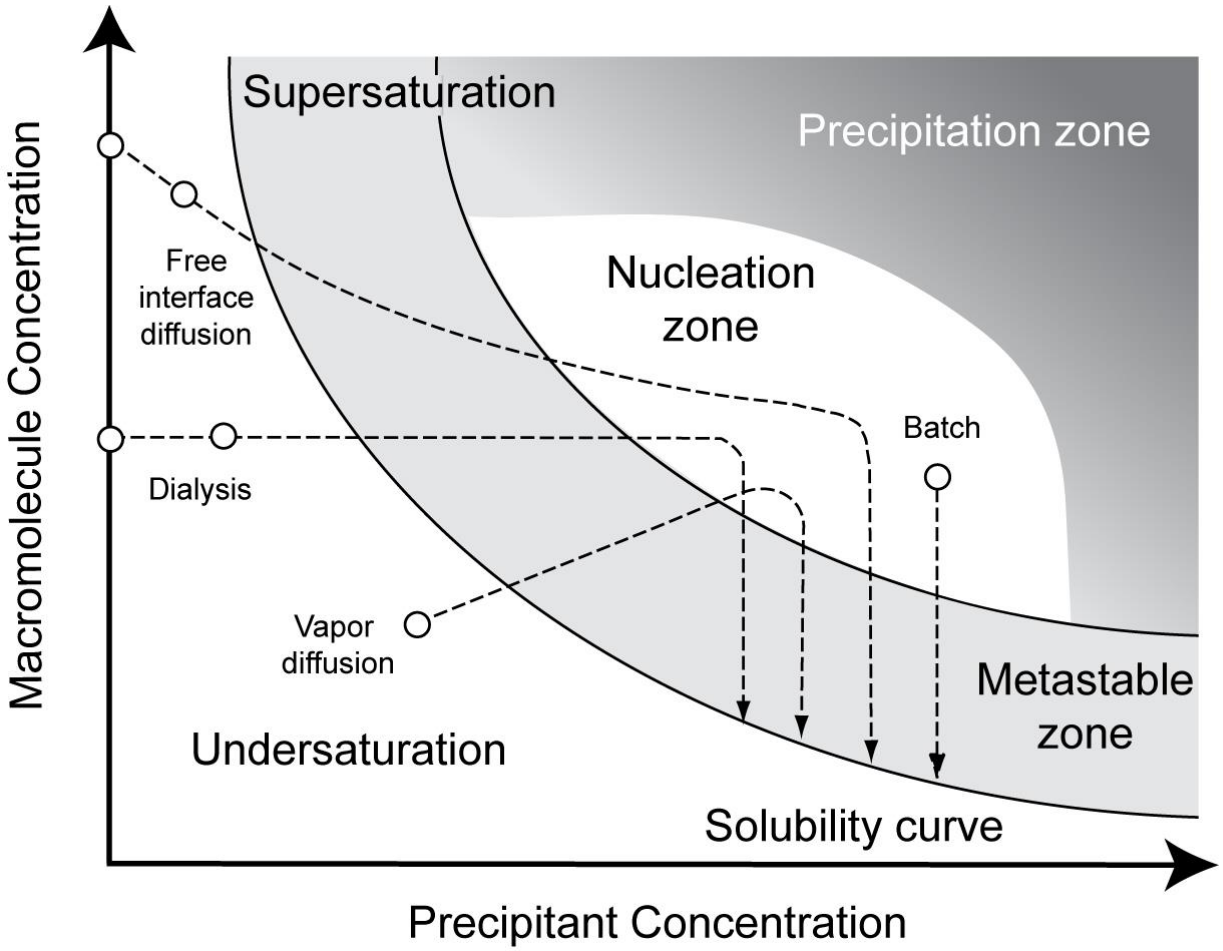
Or as a variable of time



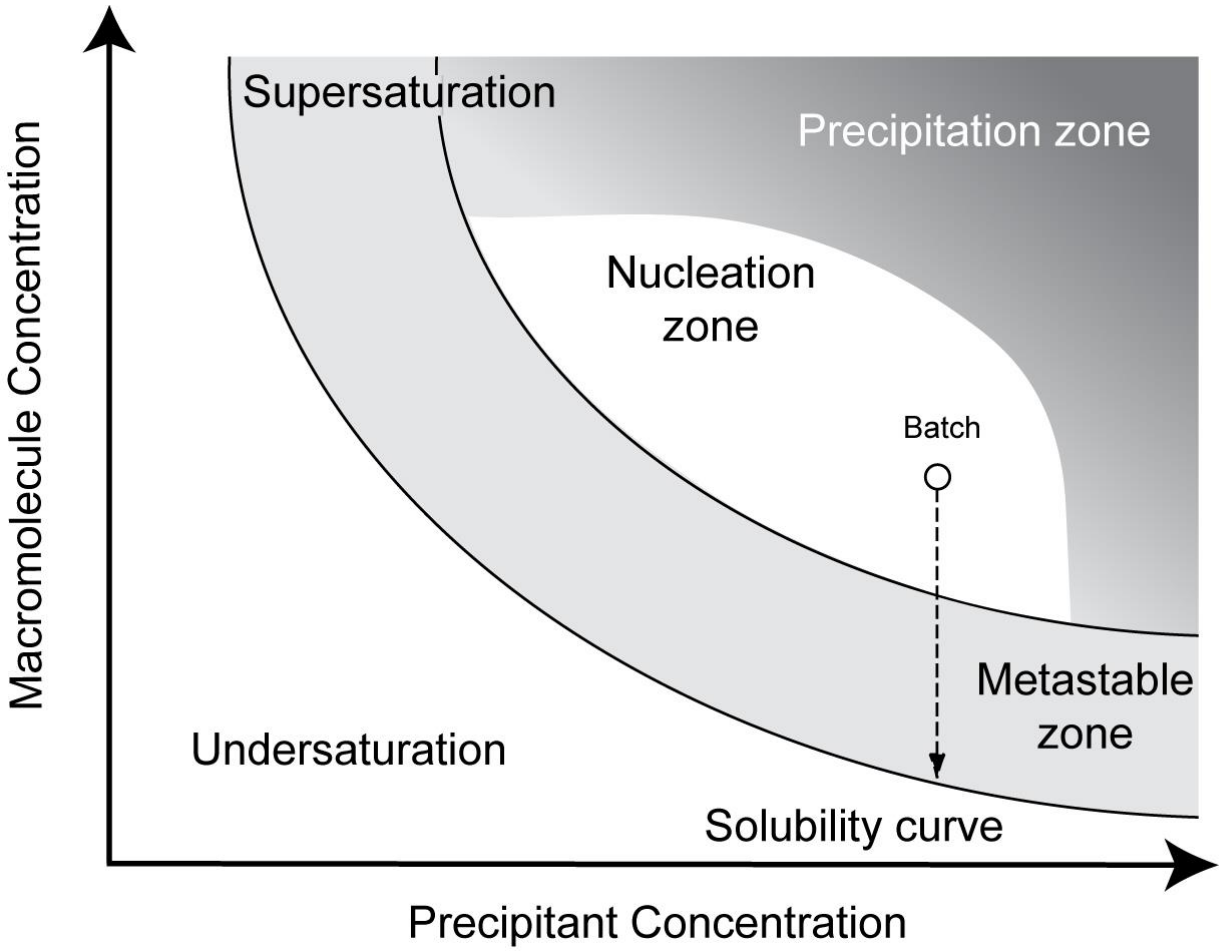
Or a single protein and a variation on biochemical conditions e.g. pH



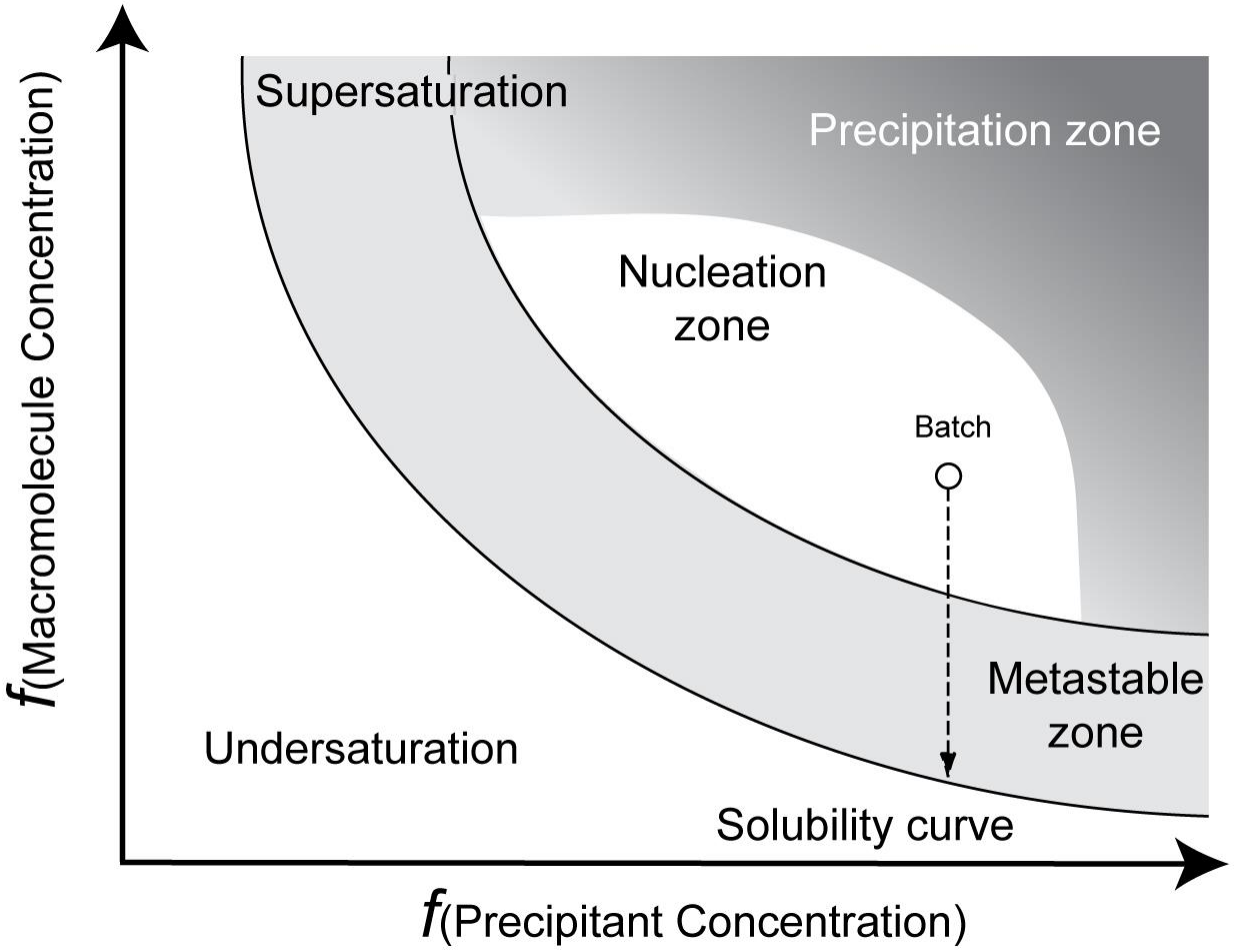
Simplified phase diagram for crystallization



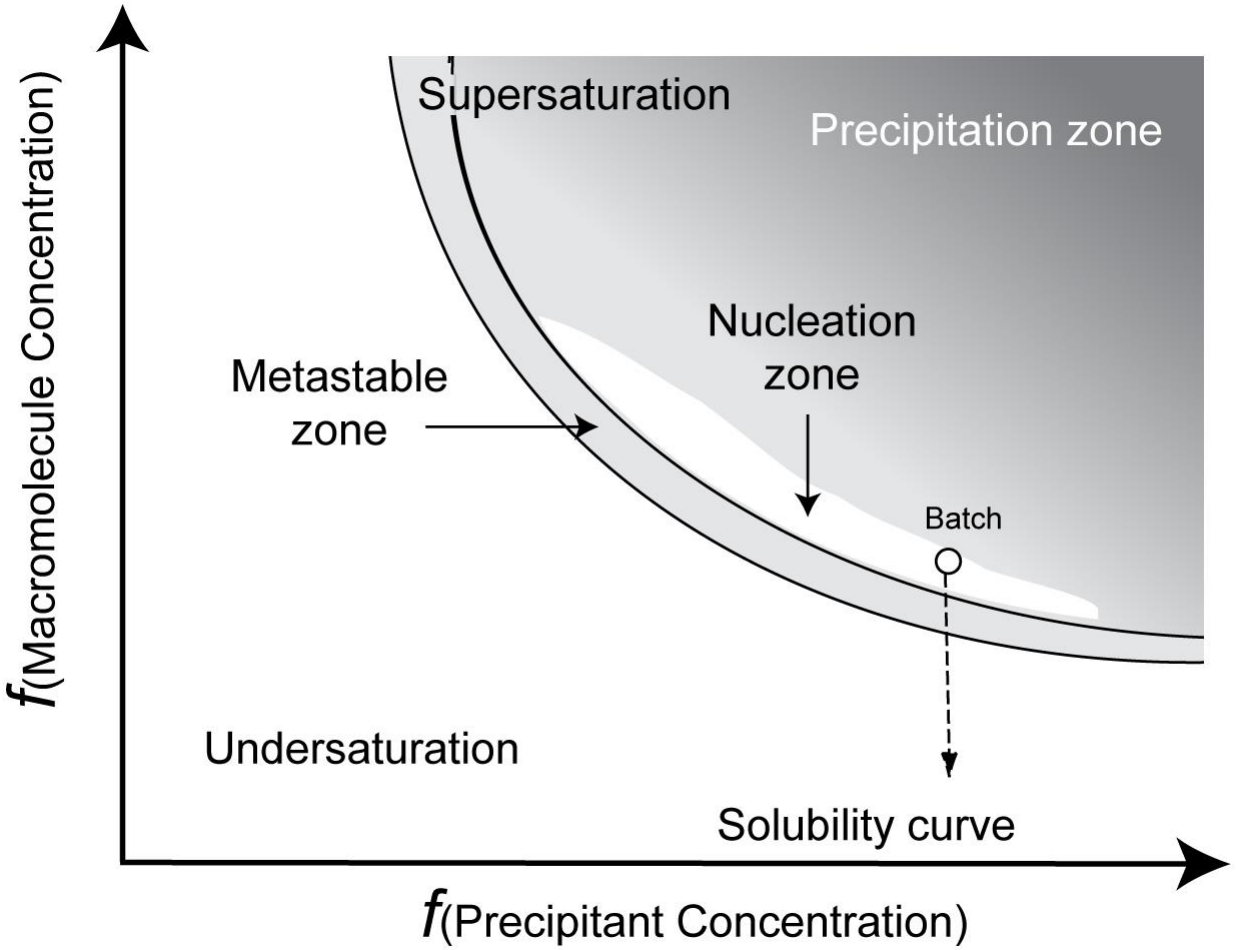
Even simpler phase diagram for crystallization



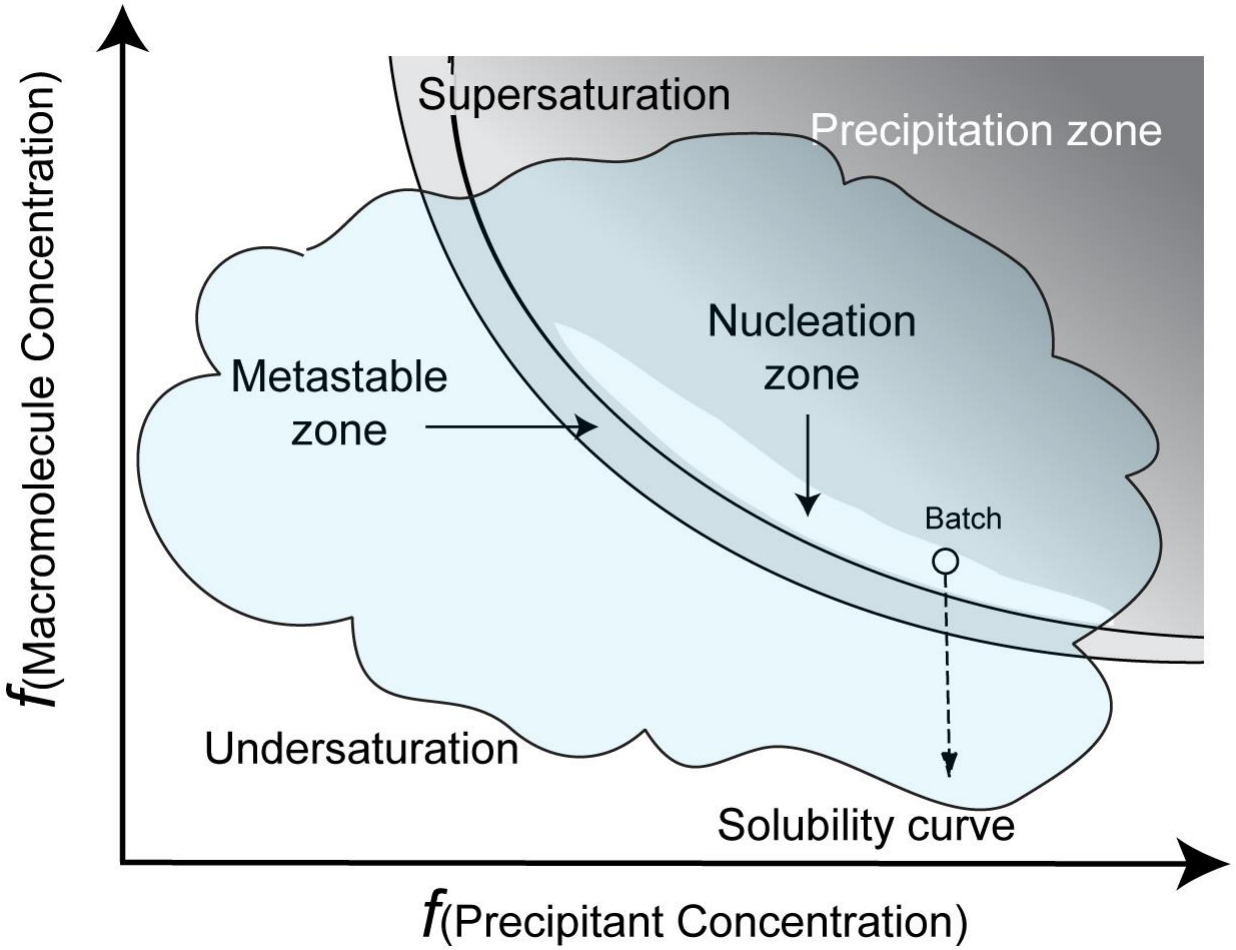
Start to throw some reality into the equation



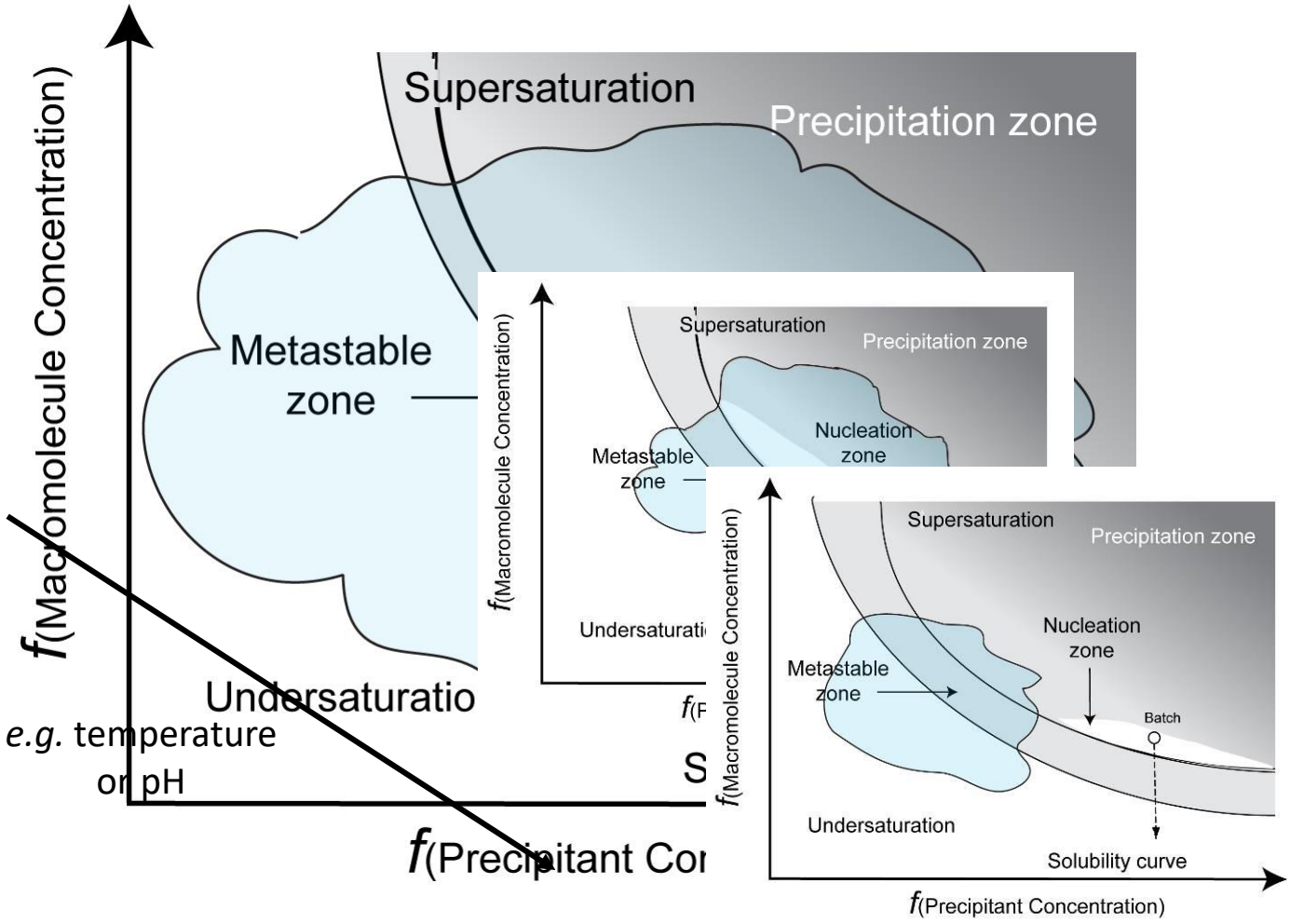
And reduce the chances of crystallization a little



Add the experimental space we sample



And the fact that it's not just two dimensions



Lets introduce a typical crystallographer …

Wile E. Coyote (Genius)

Overconfidentii Vulgaris

(Cristali Coltivatore Optimista)

And the crystal of interest …

Road Runner
(Beep beep)

Disappearialis Quickius

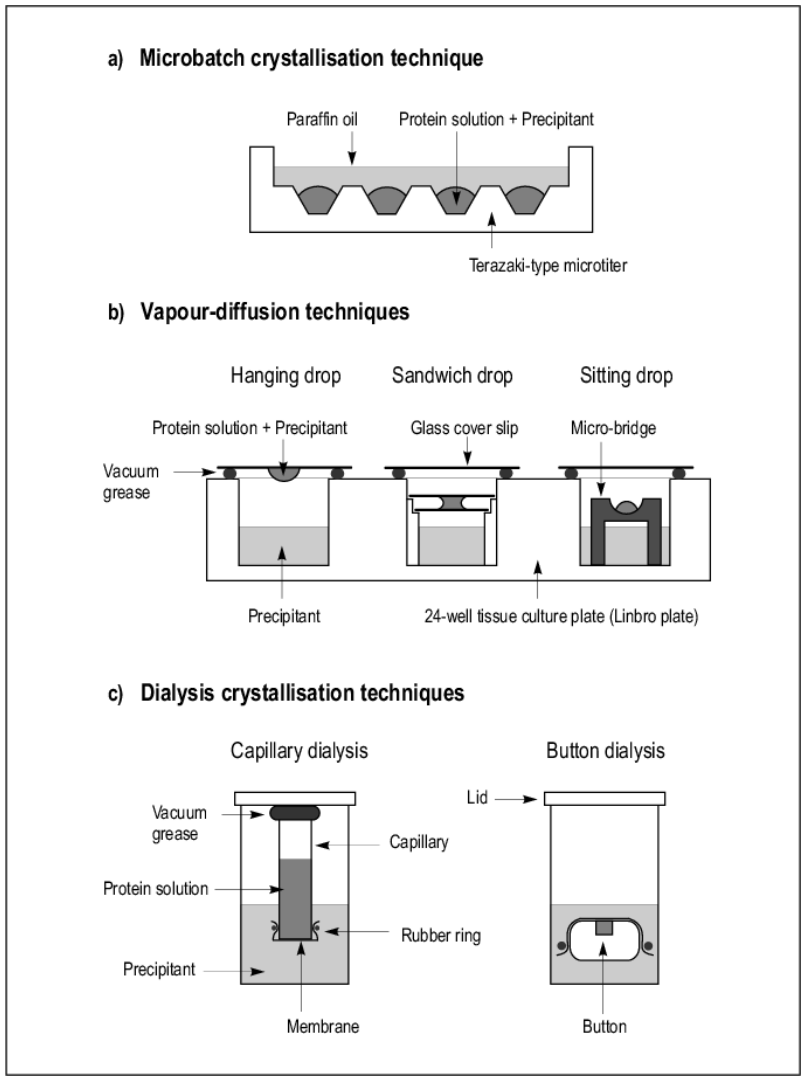
(Cristallio Perfetto)

And how the rules of the crystallographer relate to crystallography …

1. **Road Runner cannot harm the Coyote except by going "Beep! Beep!"**
2. **No outside force can harm the Coyote - only his own ineptitude or the failure of Acme products.**
3. **The Coyote could stop anytime - If he was not a fanatic.**
4. **No dialogue ever, except "Beep! Beep!"**
5. **Road Runner must stay on the road - for no other reason than that he's a roadrunner.**
6. **All action must be confined to the natural environment of the two characters -- the southwest American desert.**
7. **All tools, weapons, or mechanical conveniences must be obtained from the Acme Corporation.**
8. **Whenever possible, make gravity the Coyote's greatest enemy.**
9. **The Coyote is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the Coyote.**

1. **The crystal cannot harm the crystal grower except by not diffracting.**
2. **No outside force can harm the crystal grower - only his own ineptitude or the failure of Hampton research products.**
3. **The crystal grower could stop anytime - If they were not a fanatic.**
4. **No dialogue ever from the crystal.**
5. **The crystal will be on the path between precipitate and clear - for no other reason than it's a crystal.**
6. **All reactions must be confined to the natural environment of the crystal.**
7. **All tools, weapons, or mechanical conveniences must be obtained from Hampton Research.**
8. **Whenever possible, make salt crystals the crystal grower's greatest enemy.**
9. **The crystal grower is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the crystal grower.**

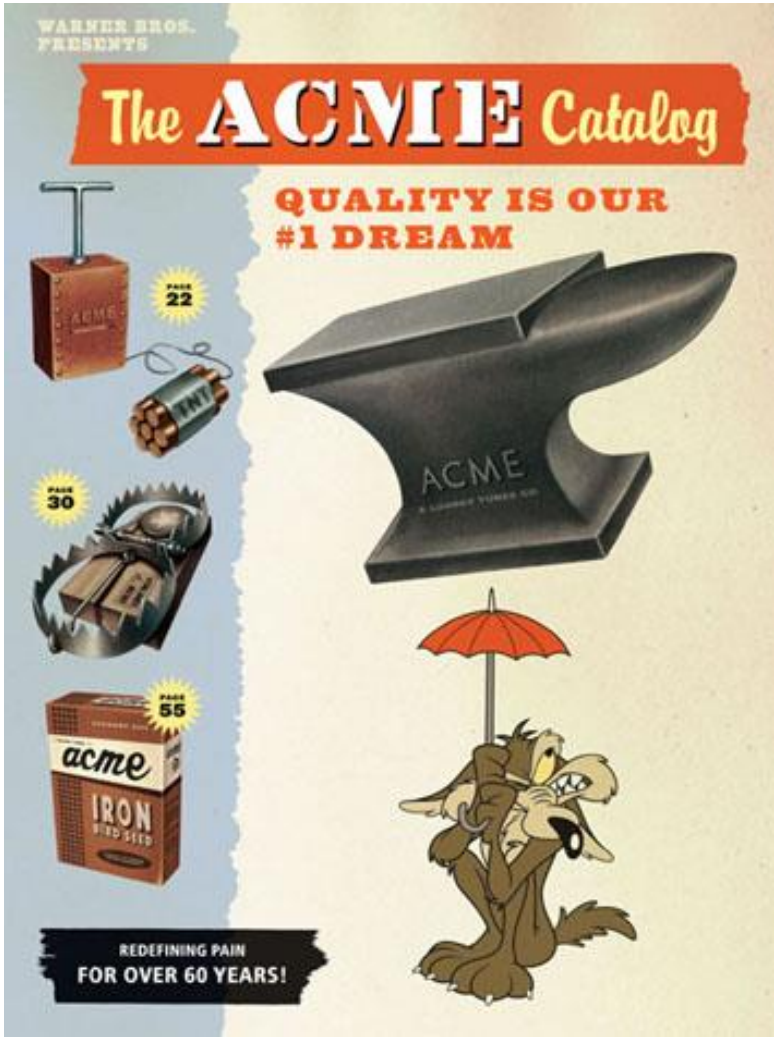
Crystallizing Macromolecules



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.

Catching Road Runners



Growing Crystals

HAMPTON RESEARCH

Molecular Dimensions Limited

Jena Bioscience

QIAGEN

Crystallization is complex

How do we grow crystals?

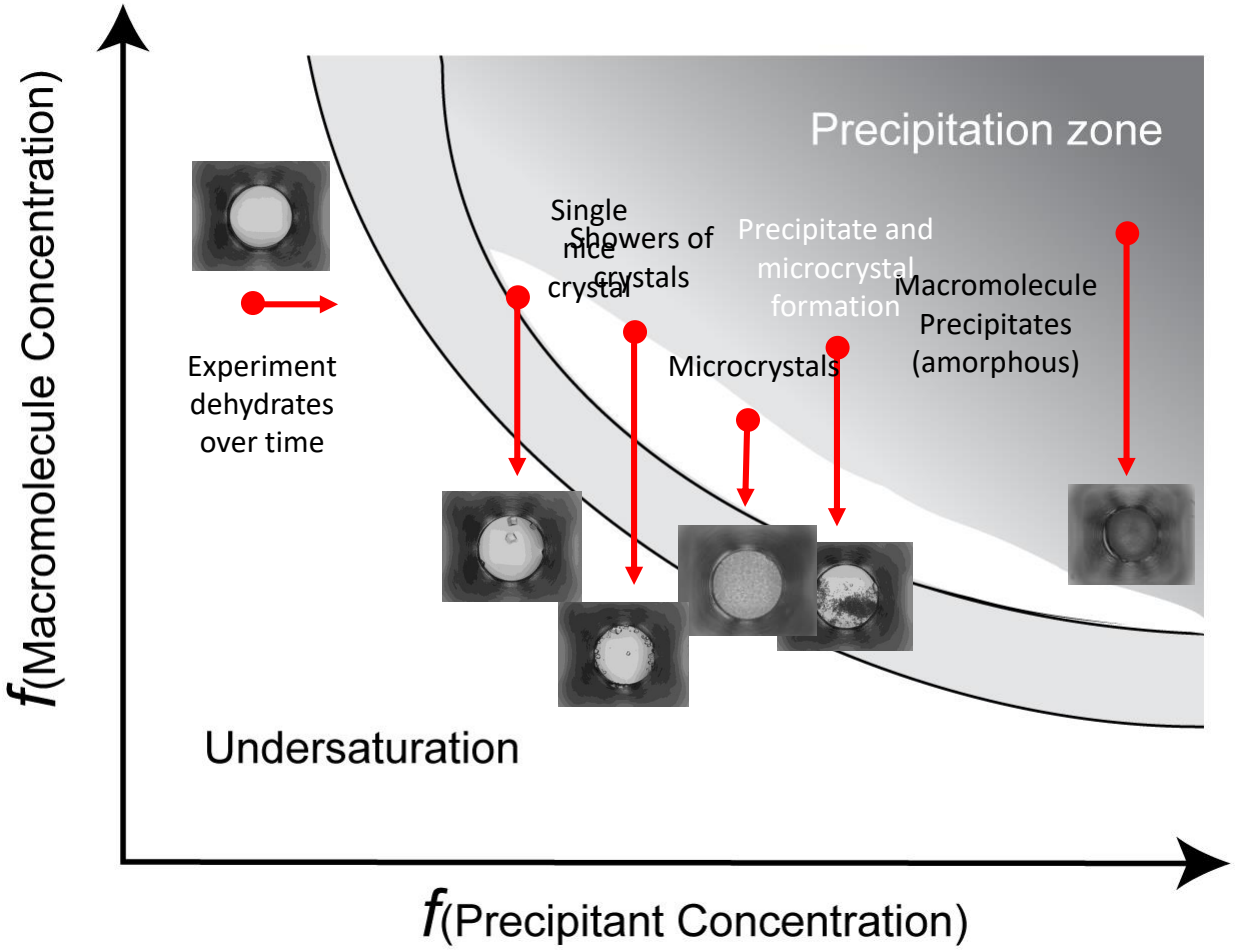
- Multiple guess?
- Intelligent design?

Set up many small scale experiments in conditions likely to be favorable for crystallization

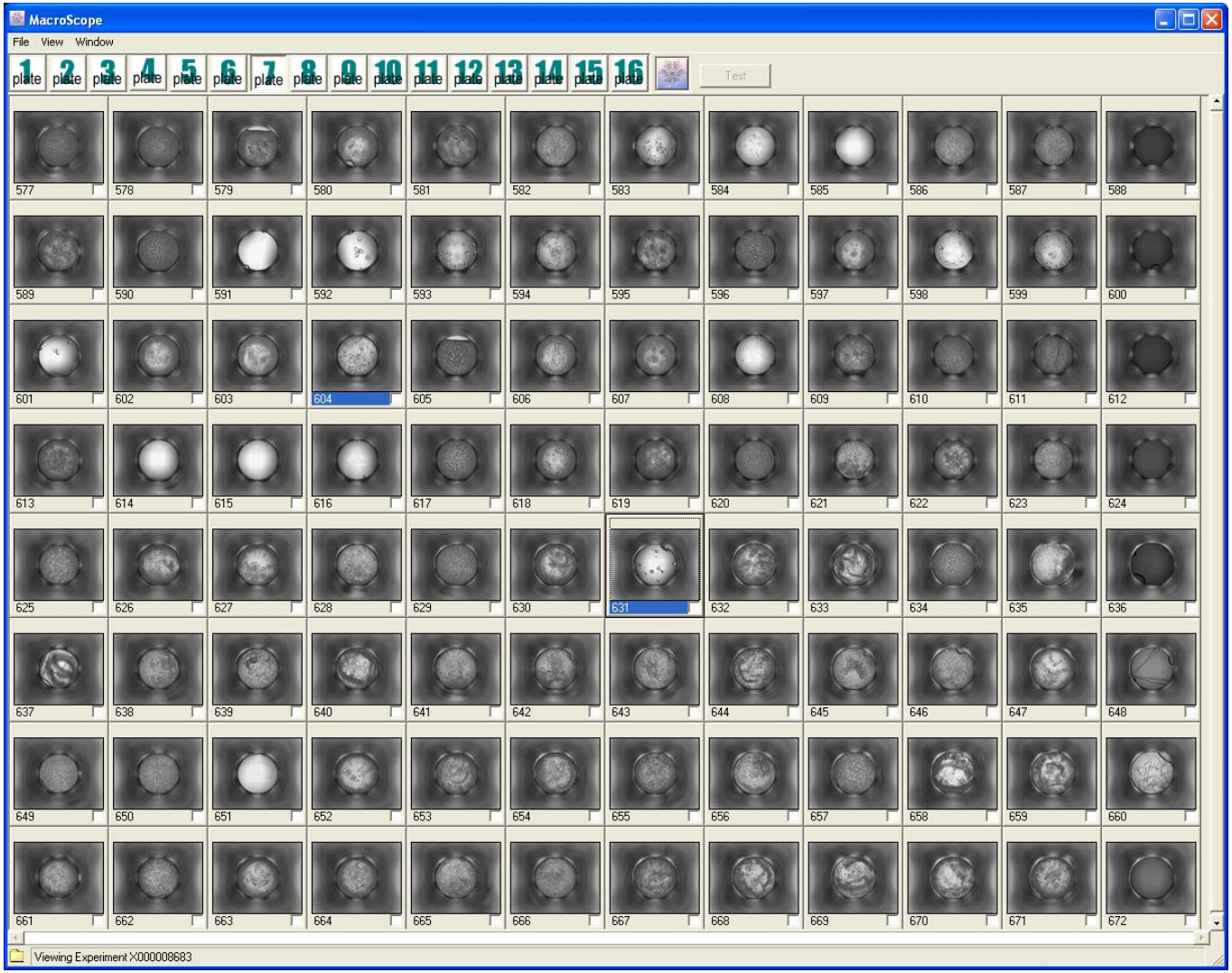
- Limited by amount of sample, time and effort.
- How many conditions is optimum? Divergent views (we'll return to this later)

Lets do the experiment

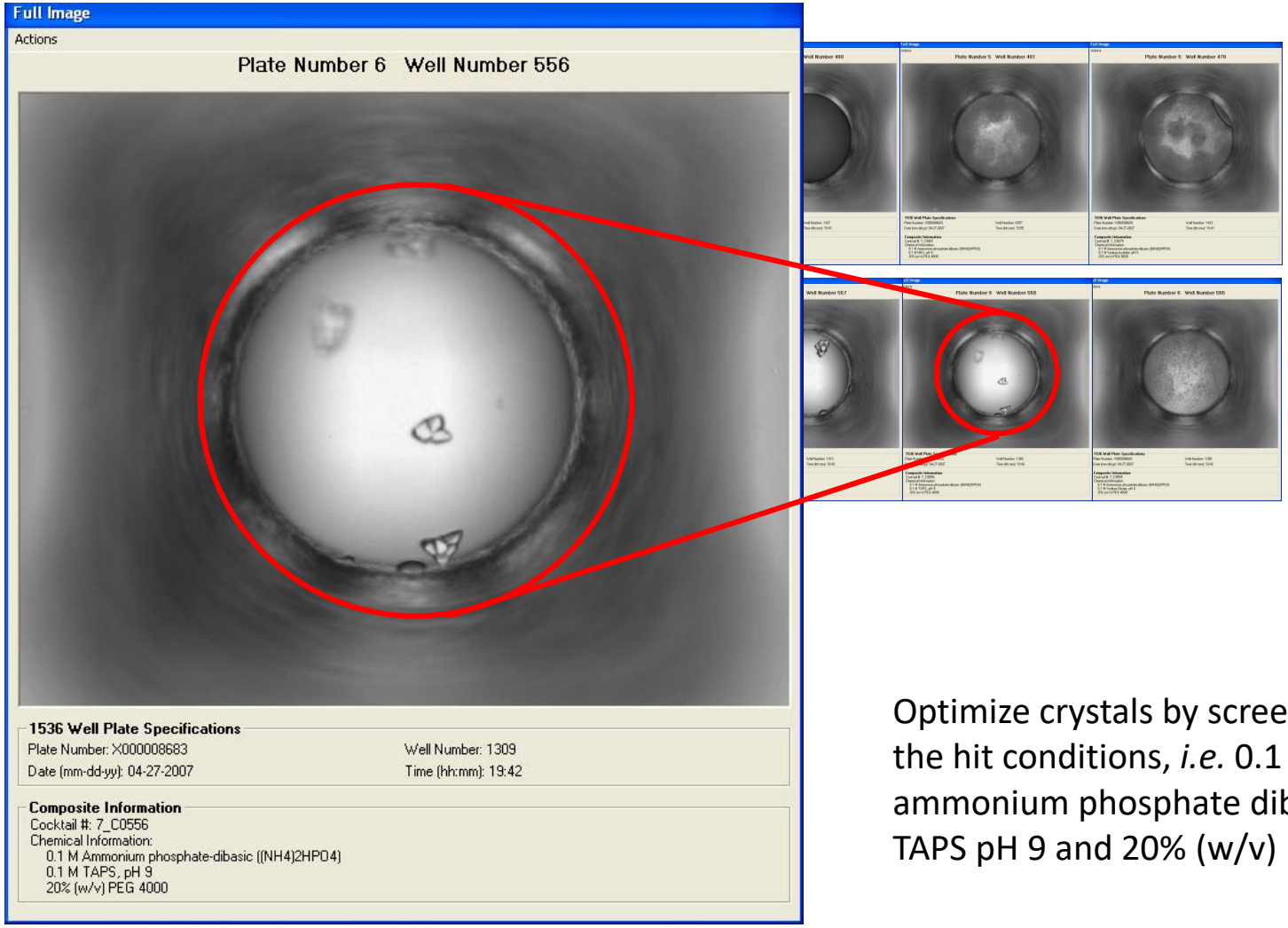
What results can we expect to see?



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

Remember how the rules of the crystallographer relate to crystallography …

1. **Road Runner cannot harm the Coyote except by going "Beep! Beep!"**
2. **No outside force can harm the Coyote - only his own ineptitude or the failure of Acme products.**
3. **The Coyote could stop anytime - If he was not a fanatic.**
4. **No dialogue ever, except "Beep! Beep!"**
5. **Road Runner must stay on the road - for no other reason than that he's a roadrunner.**
6. **All action must be confined to the natural environment of the two characters -- the southwest American desert.**
7. **All tools, weapons, or mechanical conveniences must be obtained from the Acme Corporation.**
8. **Whenever possible, make gravity the Coyote's greatest enemy.**
9. **The Coyote is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the Coyote.**

1. **The crystal cannot harm the crystal grower except by not diffracting.**
2. **No outside force can harm the crystal grower - only his own ineptitude or the failure of Hampton research products.**
3. **The crystal grower could stop anytime - If they were not a fanatic.**
4. **No dialogue ever from the crystal.**
5. **The crystal will be on the path between precipitate and clear - for no other reason than it's a crystal.**
6. **All reactions must be confined to the natural environment of the crystal.**
7. **All tools, weapons, or mechanical conveniences must be obtained from Hampton Research.**
8. **Whenever possible, make salt crystals the crystal grower's greatest enemy.**
9. **The crystal grower is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the crystal grower.**

Remember how the rules of the crystallographer relate to crystallography …

1. **Road Runner cannot harm the Coyote except by going "Beep! Beep!"**
2. **No outside force can harm the Coyote - only his own ineptitude or the failure of Acme products.**
3. **The Coyote could stop anytime - If he was not a fanatic.**
4. **No dialogue ever, except "Beep! Beep!"**
5. **Road Runner must stay on the road - for no other reason than that he's a roadrunner.**
6. **All action must be confined to the natural environment of the two characters -- the southwest American desert.**
7. **All tools, weapons, or mechanical conveniences must be obtained from the Acme Corporation.**
8. **Whenever possible, make gravity the Coyote's greatest enemy.**
9. **The Coyote is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the Coyote.**

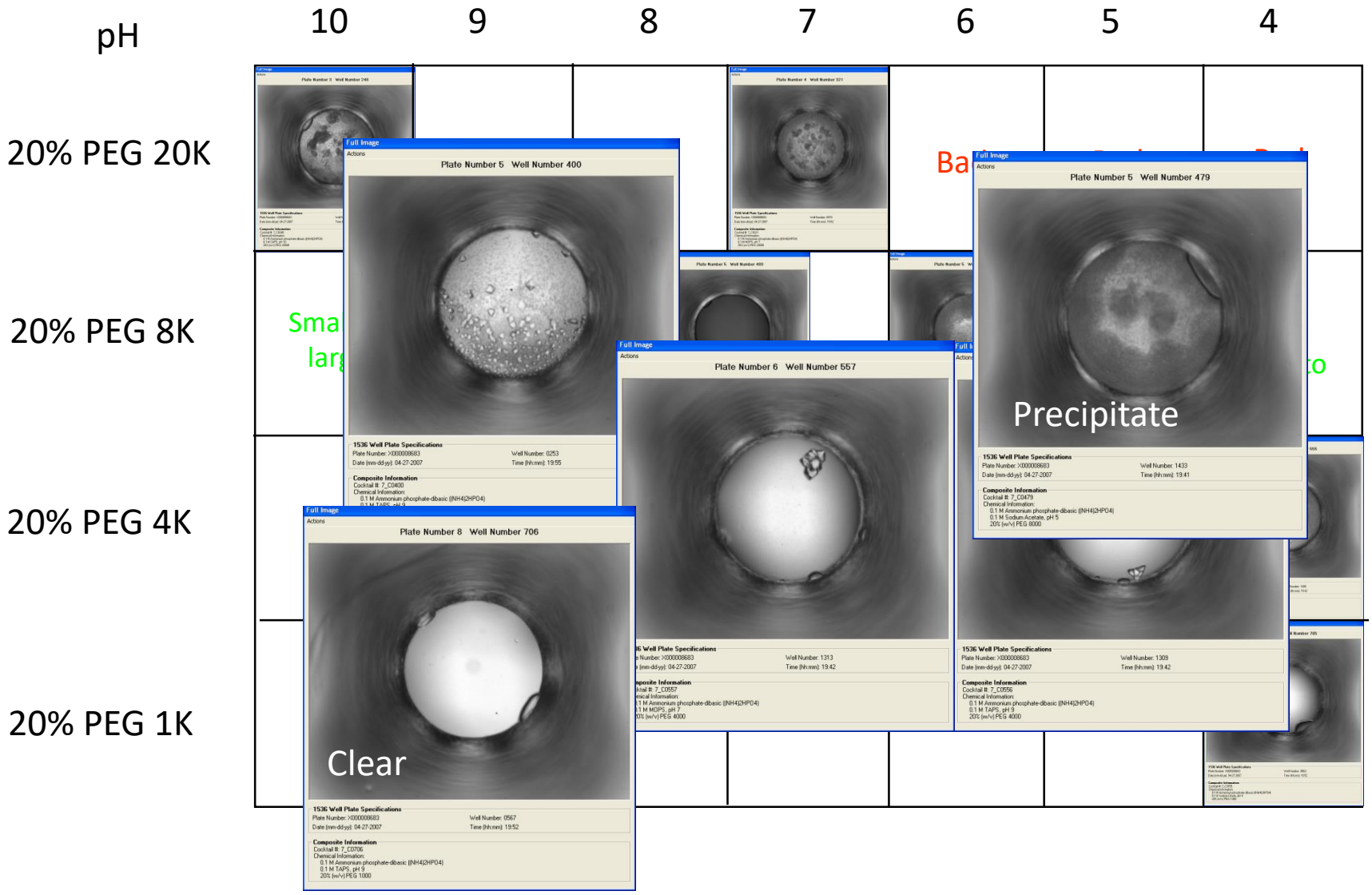
1. **The crystal cannot harm the crystal grower except by not diffracting.**
2. **No outside force can harm the crystal grower - only his own ineptitude or the failure of Hampton research products.**
3. **The crystal grower could stop anytime - If they were not a fanatic.**
4. **No dialogue ever from the crystal.**
5. **The crystal will be on the path between precipitate and clear - for no other reason than it's a crystal.**
6. **All reactions must be confined to the natural environment of the crystal.**
7. **All tools, weapons, or mechanical conveniences must be obtained from Hampton Research.**
8. **Whenever possible, make salt crystals the crystal grower's greatest enemy.**
9. **The crystal grower is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the crystal grower.**

Remember how the rules of the crystallographer relate to crystallography …

1. **Road Runner cannot harm the Coyote except by going "Beep! Beep!"**
2. **No outside force can harm the Coyote - only his own ineptitude or the failure of Acme products.**
3. **The Coyote could stop anytime - If he was not a fanatic.**
4. **No dialogue ever, except "Beep! Beep!"**
5. **Road Runner must stay on the road - for no other reason than that he's a roadrunner.**
6. **All action must be confined to the natural environment of the two characters -- the southwest American desert.**
7. **All tools, weapons, or mechanical conveniences must be obtained from the Acme Corporation.**
8. **Whenever possible, make gravity the Coyote's greatest enemy.**
9. **The Coyote is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the Coyote.**

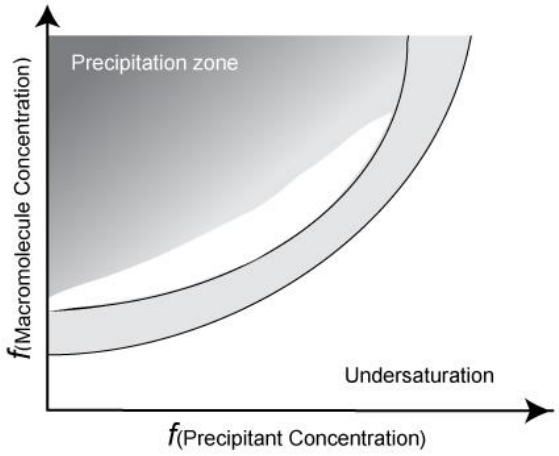
1. **The crystal cannot harm the crystal grower except by not diffracting.**
2. **No outside force can harm the crystal grower - only his own ineptitude or the failure of Hampton research products.**
3. **The crystal grower could stop anytime - If they were not a fanatic.**
4. **No dialogue ever from the crystal.**
5. **The crystal will be on the path between precipitate and clear - for no other reason than it's a crystal.**
6. **All reactions must be confined to the natural environment of the crystal.**
7. **All tools, weapons, or mechanical conveniences must be obtained from Hampton Research.**
8. **Whenever possible, make salt crystals the crystal grower's greatest enemy.**
9. **The crystal grower is always more humiliated than harmed by his failures.**
10. **The audience's sympathy must remain with the crystal grower.**

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

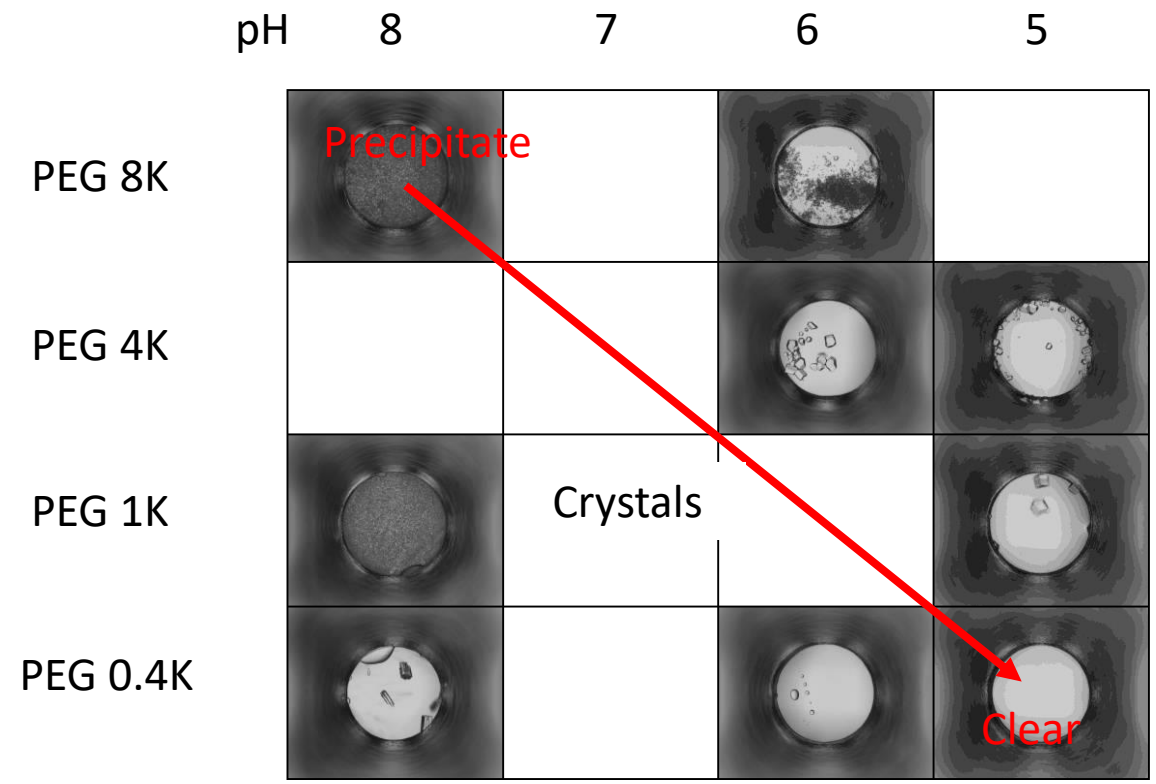


Chemical space provides a vector for optimization

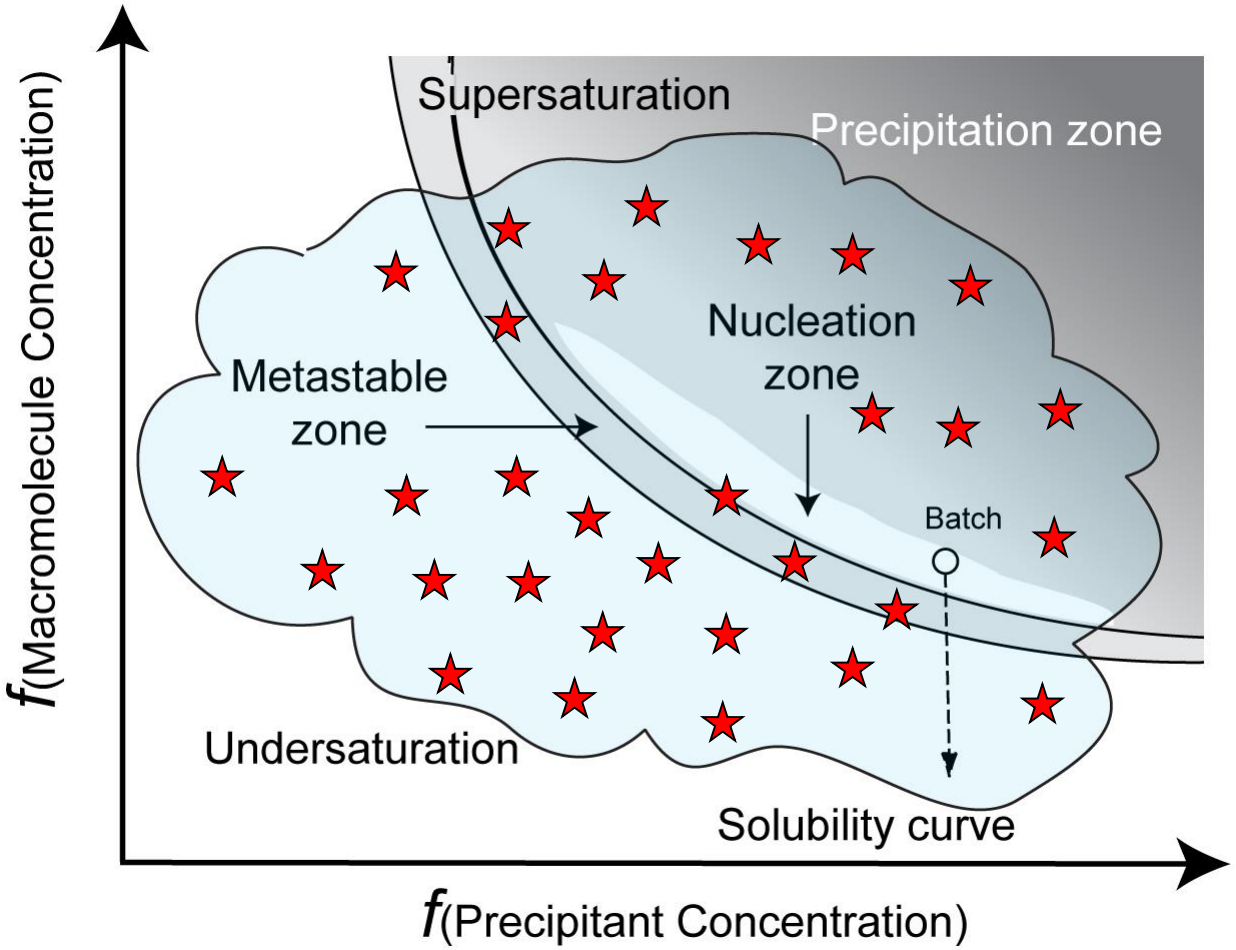
In this case the path from precipitate through crystals to clear is obvious. The phase diagram is reversed. Also clear are the number of chemical conditions that have not been sampled.



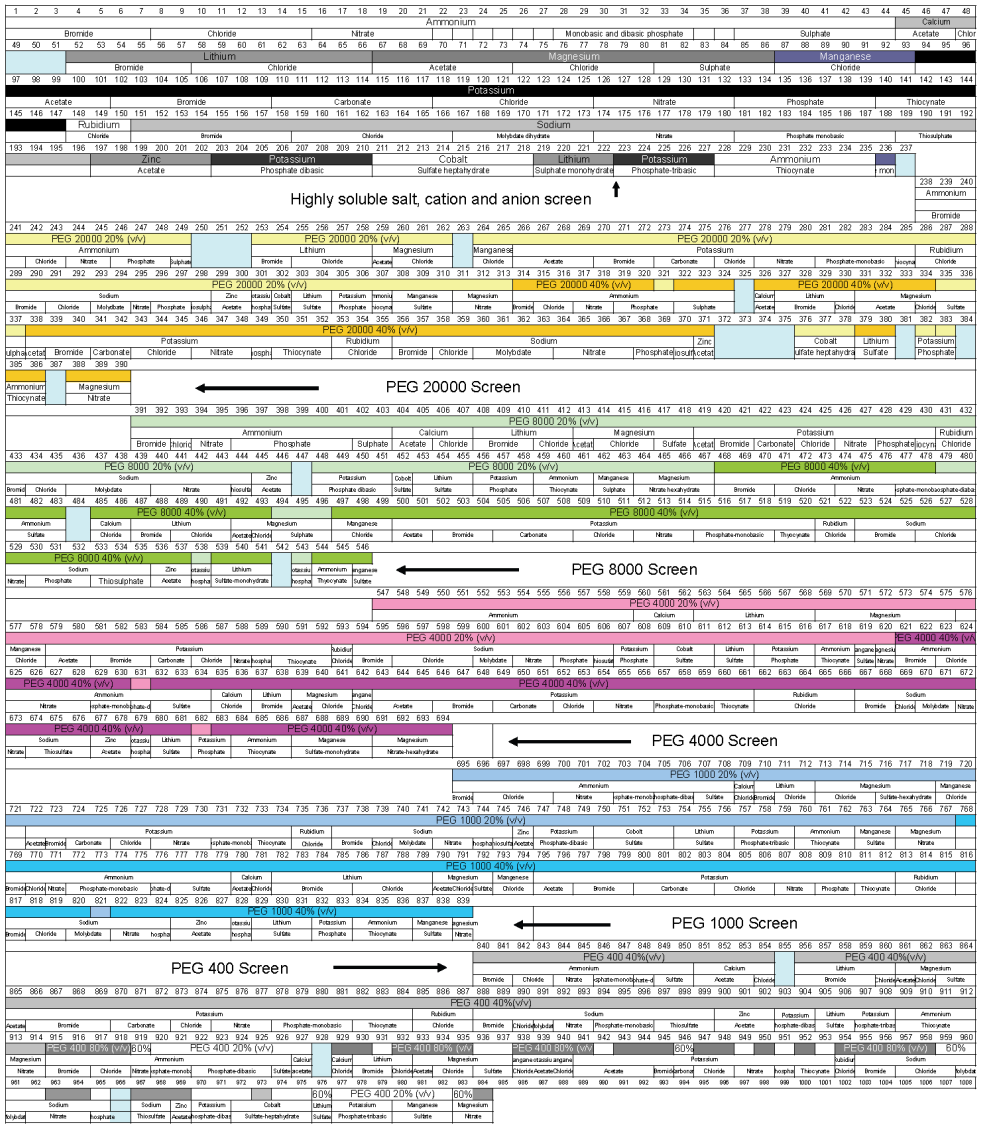
Ubiquitin, 40% PEG, 0.1M zinc acetate



It also illustrates the space we do not sample



We only sample discrete points within the sampling space



The HWI crystallization cocktail screen.

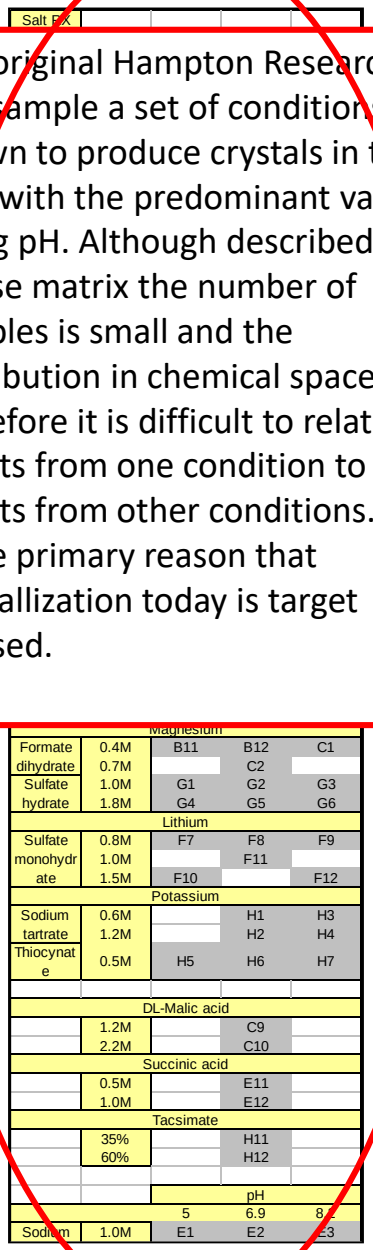
The 1536 diverse chemical cocktails (Luft et al., 2003). The 984 in-house conditions comprise a incomplete factorial sampling of 36 salts, eight buffers, and 5 different PEGs.

The remainder of 1536 cocktails are comprised of commercial screens available from Hampton Research. Specifically, in order of use; the Natrix Screen, Quick Screen, Nucleic Acid Screen, Sodium Malonate Grid, PEG/Ion, PEG 6000 Grid, Ammonium Sulfate Grid, Sodium Chloride Grid, HT Screen, Index and the SaltRx screen.

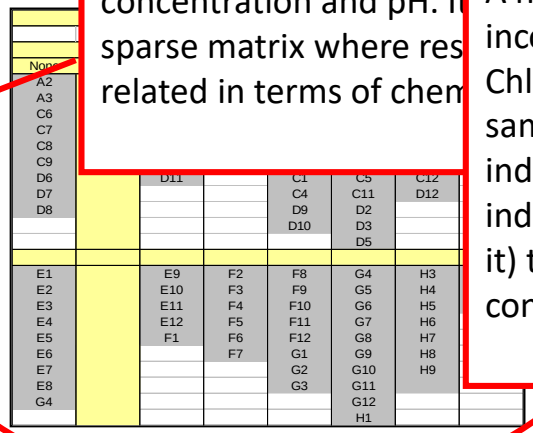
The Commercial Screens in the HWI crystallization cocktails

The commercial screens incorporate several distinct mechanisms of sampling the crystallization space. Examples are shown here.

The original Hampton Research 1+2 sample a set of conditions known to produce crystals in the past with the predominant variable being pH. Although described as a sparse matrix the number of samples is small and the distribution in chemical space wide therefore it is difficult to relate results from one condition to results from other conditions. This is the primary reason that crystallization today is target focused.



The SaltRx screen samples 22 crystallization salts with varying concentration and pH. It sparse matrix where res related in terms of chem



A number of Grid screens are incorporated, in this case Sodium Chloride. These provide a fine sampling of a small subset of individual conditions and serve to indicate the sensitivity (or lack of it) to small changes in precipitant conditions.

Sodium Chloride						
Conc	pH					
(M)	4	5	6	7	8	9
1	A1	A2	A3	A4	A5	A6
2	B1	B2	B3	B4	B5	B6
3	C1	C2	C3	C4	C5	C6
4	D1	D2	D3	D4	D5	D6

A special case – The Hampton Research Index Screen

Hampton Research Index Screen

Note, the HT screen is not a convential screen as such. It is designed to sample a range of reagents and provide an indication of the appropiate chemical area and variables that would be appropiate for crystallization and should be used in this manner.

pH	Ammonium Sulfate 2.0M	Sodium chloride 3.0M		Magnesium formate dihydrate			Sodium phosphate		Neutralized organic acids (ph 7.0)		High supersaturation salt and low polymer		Low ionic strength systems			Non-volatile organics	
pH				0.3M	0.5M		pH				pH		pH			pH	
3.5	A1	A7					5.6	B5		B9	5.5	C8	3.5	D4		5.5	D12
4.5	A2	A8					6.9	B6		B10	6.5	C6	4.5	D5			E2
5.5	A3	A9		B1			8.2	B7		B11	8.5	C7	5.5	D6		6.5	E1
6.5	A3	A10			B2					B12	7	C9	6.5	D7			E3
7.5	A5	A11		B3						C1		C10		D10		6.5	E6
8.5	A6	A12			B4					C2		C11		D11			E9
										C3		C12	7	D2			E10
	Classic salt versus pH									C4				D3		7.5	E4
										C5			7.5	D8			E7
	Hits here indicate that a variation of salt concentration and pH in a grid screen has a strong potential for crystallization												8.5	D9			E8
																	E11
																8.5	E5
																	E12

PEGs and Salts as a function of pH								PEG 3350 and salts											
	3.35K					10K	3.35K												
pH	Ammonium sulfate	Sodium chloride	Lithium sulfate monohydrate	Ammonium acetate	Magnesium Chloride hexahydrate	Ammonium acetate	Mixed chloridehydrates	%	Potassium sodium tartrate tetrahydrate	Sodium malonate pH 7.0	Ammonium citrate tribasic pH 7.0	Succinic acid pH 7.0	Sodium formate	DL-Malic acid pH 7.0	Magbesium formate dihydrate	Zinc acetate dihydrate	Sodium citrate tribasic dihydrate	Potassium thiocyanate	Potassium bromide
5.5	F6	F10	G2	G6	G10	F5		15				H5			H8				
6.5	F7	F11	G3	G7	G11			20	H2	H3	H4		H6	H7		H9	H10		
7.5	F8	F12	G4	G8	G12		F4	25											
8.5	F9	G1	G5	G9	H1			30										H11	H12

Coarse test for chemical conditions likely to produce crystallization

Sherlock and Watson.

"We approached the case, you remember, with an absolutely blank mind, which is always an advantage. We had formed no theories. We were simply there to observe and to draw inferences from our observations"

Sherlock Holmes to Dr. Watson

I never get your limits, Watson. There are unexplored possibilities about you.

Sherlock Holmes on Dr. Watson.

Two pieces of related software under development;

- Sherlock to look at the individual 'crime', *i.e.* examine results from a single macromolecule
- Watson to tell the complete story, *i.e.* look at trends from many experiments.

Sherlock and Watson.

"We approached the case, you remember, with an absolutely blank mind, which is always an advantage. We had formed no theories. We were simply there to observe and to draw inferences from our observations"

Sherlock Holmes to Dr. Watson

I never get your limits, Watson. There are unexplored possibilities about you.

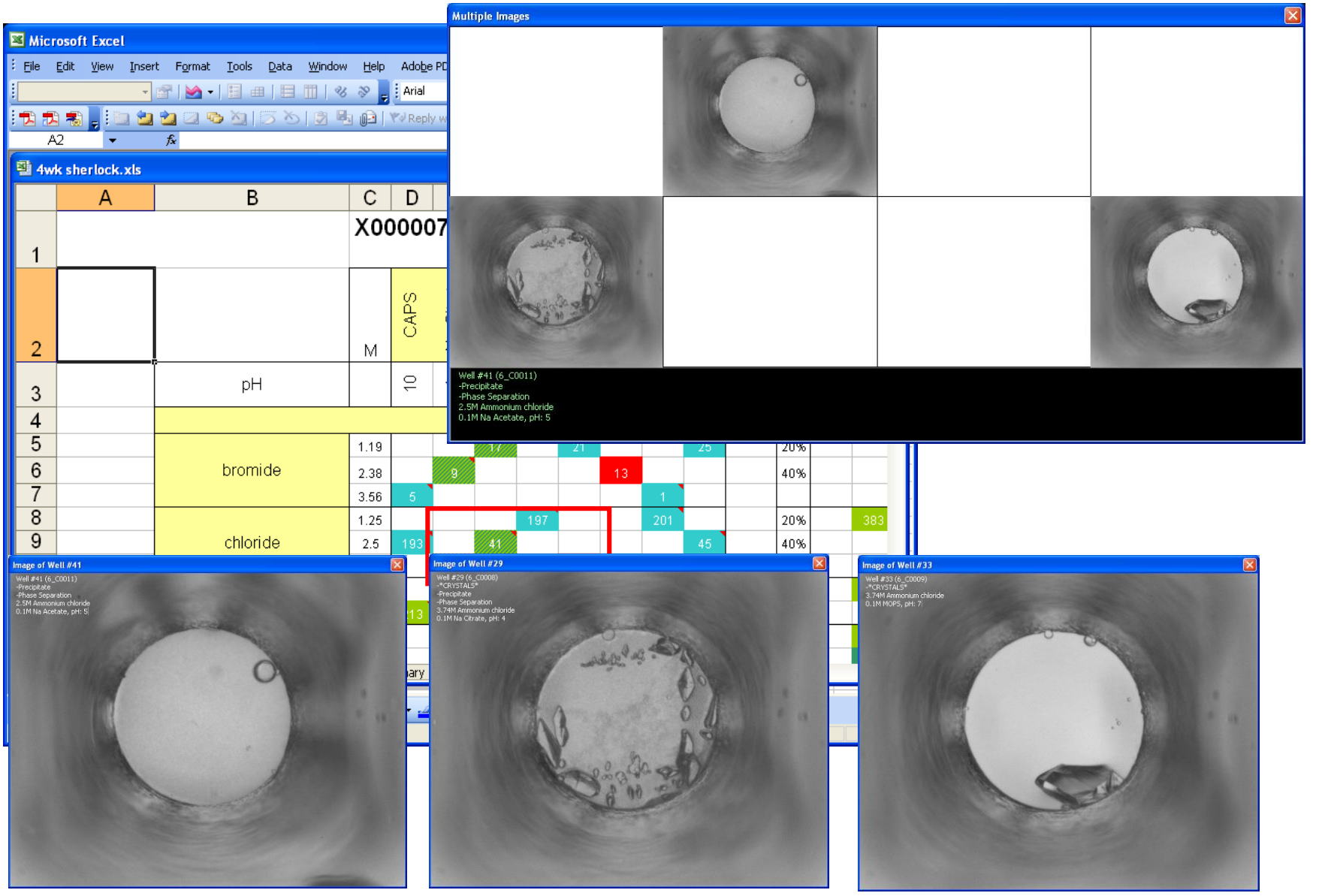
Sherlock Holmes on Dr. Watson.

Two pieces of related software under development;

- Sherlock to look at the individual 'crime', *i.e.* examine results from a single macromolecule
- Watson to tell the complete story, *i.e.* look at trends from many experiments.

Decreasing pH leads to crystallization. A large area of space along the crystallization pathway remains un-sampled. There are clear areas to pursue optimization.

Decreasing PEG % leads to crystallization. Again a large area of space along the crystallization pathway remains un-sampled. There are clear areas to pursue optimization.


Microsoft Excel
File Edit View Insert Format Tools Data Window Help Adobe PD
Arial
A2
4wk sherlock.xls
A B C D
X000007
M
CAPS
pH
10
bromide
chloride
1.19
2.38
3.56
1.25
2.5
20%
40%
Multiple Images
Well #41 (6_C0011)
-Precipitate
-Phase Separation
2.5M Ammonium chloride
0.1M Na Acetate, pH: 5
Image of Well #41
Image of Well #29
Well #29 (6_C0008)
-*CRYSTALS*
-Precipitate
-Phase Separation
3.74M Ammonium chloride
0.1M Na Citrate, pH: 4
Image of Well #33
Well #33 (6_C0009)
-*CRYSTALS*
3.74M Ammonium chloride
0.1M MOPS, pH: 7

Sherlock and Watson.

"We approached the case, you remember, with an absolutely blank mind, which is always an advantage. We had formed no theories. We were simply there to observe and to draw inferences from our observations"

Sherlock Holmes to Dr. Watson

I never get your limits, Watson. There are unexplored possibilities about you.

Sherlock Holmes on Dr. Watson.

Two pieces of related software under development;

- Sherlock to look at the individual 'crime', *i.e.* examine results from a single macromolecule
- Watson to tell the complete story, *i.e.* look at trends from many experiments.

Sherlock and Watson.

"We approached the case, you remember, with an absolutely blank mind, which is always an advantage. We had formed no theories. We were simply there to observe and to draw inferences from our observations"

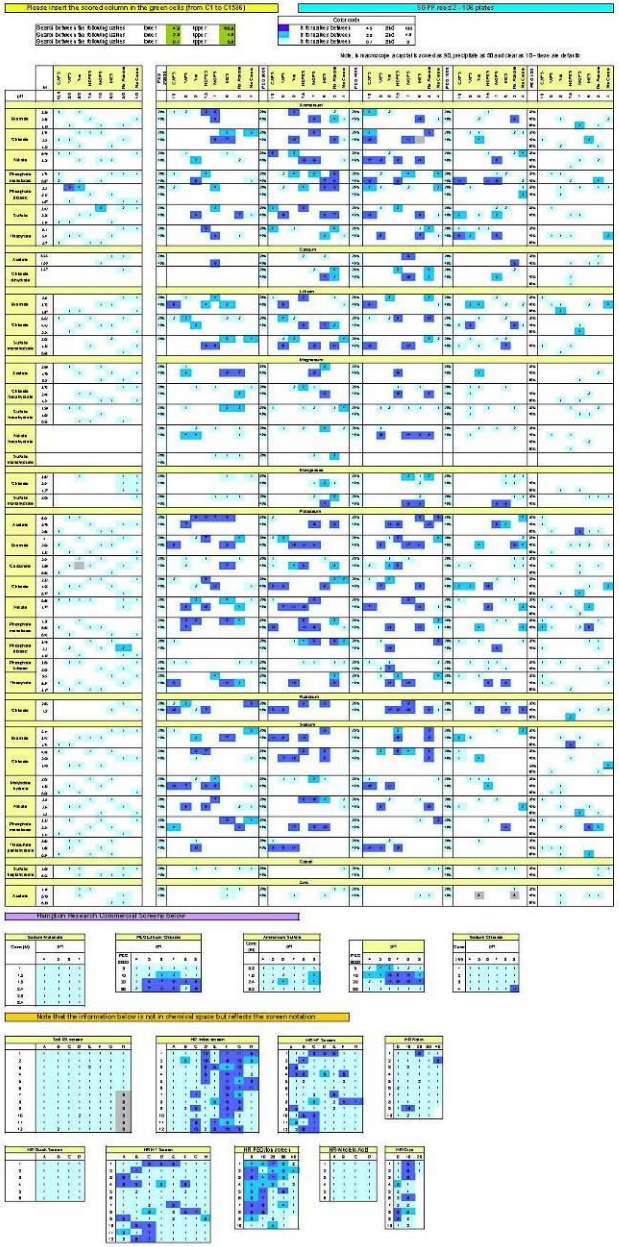
Sherlock Holmes to Dr. Watson

I never get your limits, Watson. There are unexplored possibilities about you.

Sherlock Holmes on Dr. Watson.

Two pieces of related software under development;

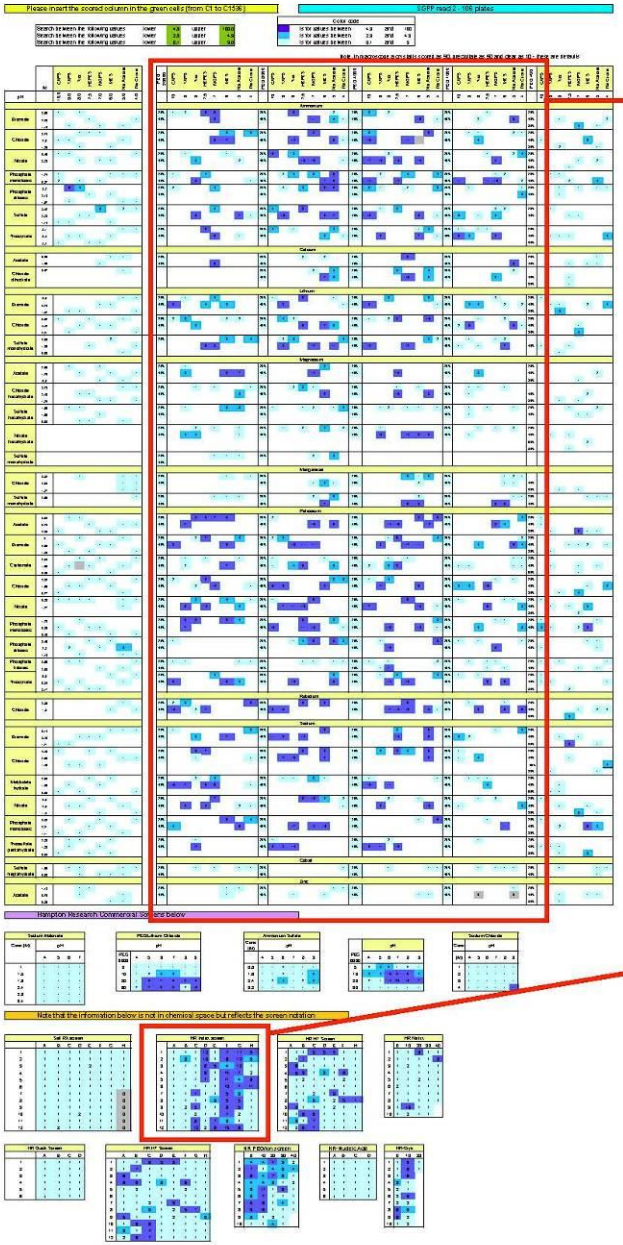
- Sherlock to look at the individual 'crime', *i.e.* examine results from a single macromolecule
- Watson to tell the complete story, *i.e.* look at trends from many experiments.



Watson looks at the complete picture rather than an single 1536 screen of one macromolecule.

In this case crystallization results for 106 macromolecues are shown several weeks into their growth. Only those samples showing crystals are tabulated here..

Dark blue indicates 5 or more crystal hits in the 106 conditions, medium blue is 3-4 hits and light blue is 1-2. Grey shows other conditions sampled (only 3 conditions) while white spaces are un-sampled regions of the incomplete factorial.



PEG 20K to 1K component of the screen

The index screen is working, it shows that PEG is a good chemical area to promote crystallization

Hampton Index Screen

Sherlock and Watson – Current Status

- Sherlock is currently being tested in the High-Throughput laboratory. The aim is to release it to external users as a beta version in the near future.
- There are several possible representations of chemical space available, only one was shown here.
- Currently it requires manual scoring of images. Developments in automated image analysis look very promising and there is near certainty that we can automatically score clear and precipitate images leaving a much smaller number of images to visually examine. Other research is underway to automatically score these as well.

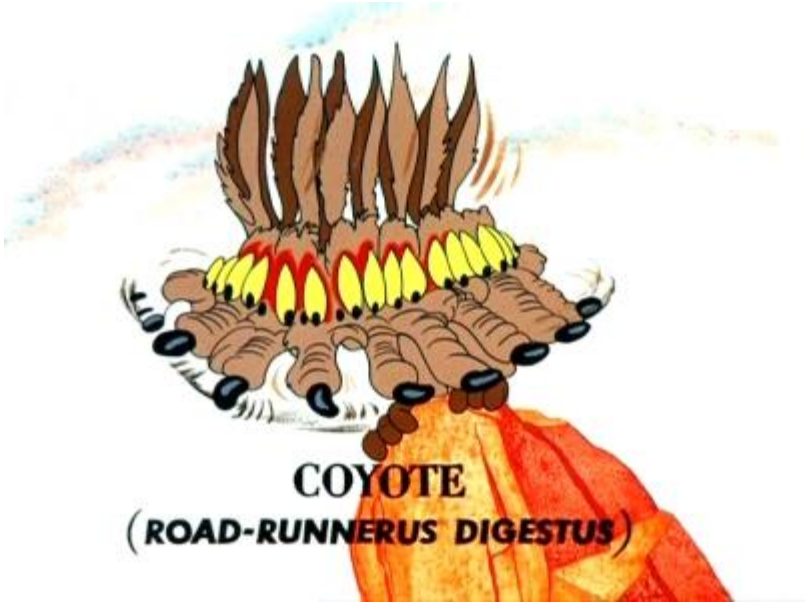
- Watson is under development and at present is only being used by a limited number of testers to analyze the performance of the HWI cocktails and commercial screens used in the laboratory.

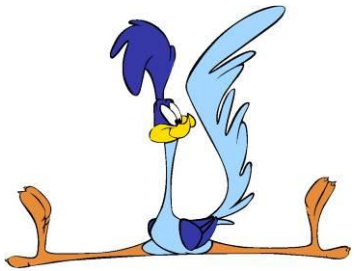
Future work

- To automatically flag patterns that may indicate potential regions for further exploration if a crystallization hit does not occur. For example, two results showing clear and precipitate separated by a long un-sampled chemically sensitive pathway.
- To produce separate programs for other screens.
- To incorporate time or temperature resolved data, predict the best optimization strategies or aid the interpretation of current optimization techniques such as Drop Volume Ratio/Temperature (DVR/T) Luft et al., 2007.

How many samples?

In using chemical space mapping to analyze a number of samples it has become clear that 1536 is a good number of experiments to try. It enables a wide range of chemical space to be investigated with sufficient detail to identify common regions for crystallization together with diversely separated regions where different crystal forms may result





It is important to investigate not a single hit but as many hits as you have sample. Visual observation only indicates a crystal, not that it diffracts well or even if it is a macromolecular crystal rather than salt or PEG. Spreading the effort among many hits is better than focusing exclusively on one.

Summary

- No experiment should be considered in isolation.
- In crystallization screening when you have a sparse matrix, incomplete factorial or any other designed sampling of chemical space the results build up a picture of the crystallization landscape.
- An experiment with no crystallization hits that which generates both precipitate and clear conditions is promising when those conditions are separated by an un-sampled chemically sensible direction.
- You should know what crystallization conditions you examined but more importantly how those relate to those that were not sampled.
- Optimize as many samples as you can.
- Check with X-rays as soon as possible.
- The axis of crystallization space have a complex relationship with those in chemical space. We have a limited understanding of those relationships and hopefully Watson will reveal a better story from the >9000 cases we currently have.
- There are many more variables to explore!

One portal website - Xtuition

“Wikicrystal”	Crystallization tips, theory and observations.
“Wikinot”	Advice on what to do if no crystal.
“Now you see it”	SAXS derived molecular envelope for all samples.
Expert system	Query of crystallization database and results for NESG data.
Xtuition	Automated hypothesis tester linking public and private data.

Phase 2: Fuzzy link to non-PSI data

Making use of image analysis

Machine Classifications

Human Classifications	A	B	C	D	E	F	G	H	I	J	TRUTH DATA	
A	3934	354	1039	280	88	1	4	268	1174	21	7163	A- Crystal
B	578	433	281	117	51	14	0	421	94	2	1991	B- Crystal/Phase
C	1016	153	2972	1721	296	23	2	211	69	0	6463	C- Crystal/Precip
D	397	49	1325	24547	987	52	4	1213	810	27	29411	D- Precip
E	120	24	206	1201	2557	5	3	98	29	8	4251	E- Precip/Skin
F	19	13	101	199	38	18	1	49	1	0	439	F- Precip/Phase
G	7	2	4	2	16	0	11	24	1	0	67	G- Phase/Skin
H	422	115	77	274	73	29	2	3721	1229	32	5974	H- Phase
I	101	1	12	128	9	0	1	123	28482	174	29031	I- Clear
J	19	1	0	33	4	0	1	4	163	246	471	J- Garbage

83% correct

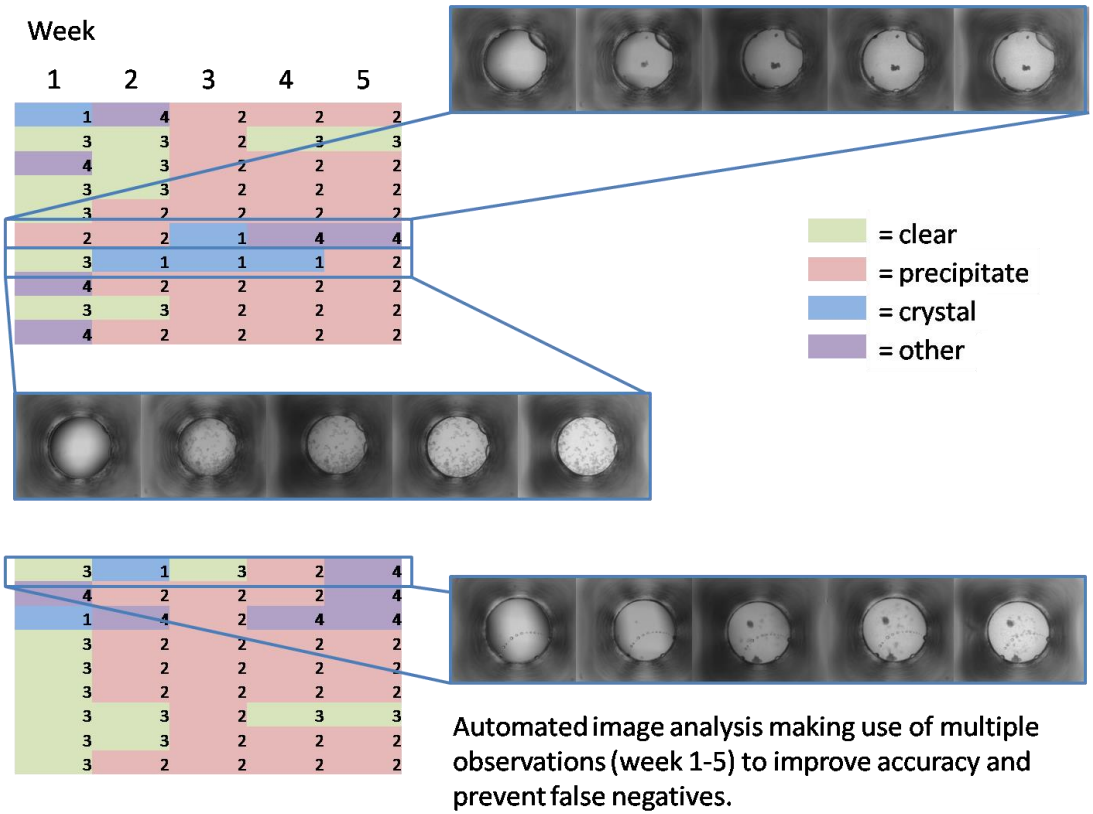
98% correct

Better than a human!

Unfortunately it takes 38 minutes per image! However, the code has been rewritten for a GPU system and tests on an early process indicate 6 minutes per image.

For a single plate, 40 days to do the mage analysis, 7 days on a GPU system

Image analysis with two three way classifiers (faster)



Automated image analysis making use of multiple observations (week 1-5) to improve accuracy and prevent false negatives.

Analysis under an hour on a standard desktop computer. By combining two 3-way classifiers with a time component the accuracy in finding crystal hits is improved.

By incorporating chemical knowledge we plan to improve the classification further by comparing chemically related results.

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

A solution to structure

- NMR chemical shift measurements.
- SAXS data and envelope calculation.
- Homology modeling
- Filter decoy set based on envelope and chemical shifts.

- Testing under way.
- Rosetta being adapted to use SAXS data.
- SAXS data collected on 20 samples where we also have chemical shift data and a crystal structure.
- If successful we will expand the process to other systems where we have chemical shift data, SAXS data but no structure.
- (Rosetta – painful to set up but fun to run, several thousand models this morning)

Why failure?

- Is it the way we are crystallizing?
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Goals

- To predict failure and success.
- To develop a sample specific structural determination strategy
- To mine data and test crystal growth hypothesis
- To provide expert advice automatically based on limited outcomes
- To significantly improve the 9.9% problem

How?

- Link crystallization outcomes to chemistry.
- Classify samples by SAXS (aggregation state, globularity, envelope).
- Link Thermofluor® based analysis to optimize conditions, look for dynamics etc.
- Feedback to SAXS to minimize Rg/globularity
- Use results to drive further crystallization.

Coming Soon

Formulation robot for solution making and an automated imaging system

Sample specific crystallization strategies

Hopefully later

An in-house SAXS system

Why failure?

- Is it the way we are crystallizing? Possibly
- Is it the sample?
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Why failure?

- Is it the way we are crystallizing? Possibly
- Is it the sample? Probably not
- Are we just going to have to live with it?
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Why failure?

- Is it the way we are crystallizing? Possibly
- Is it the sample? Probably not
- Are we just going to have to live with it? No
- Can we learn from our previous successes and failures?
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Why failure?

- Is it the way we are crystallizing? Possibly
- Is it the sample? Probably not
- Are we just going to have to live with it? No
- Can we learn from our previous successes and failures? Maybe
- Can we use other methods to get structural information?
- Can we combine everything and learn anything useful?

Why failure?

- Is it the way we are crystallizing? Possibly
- Is it the sample? Probably not
- Are we just going to have to live with it? No
- Can we learn from our previous successes and failures? Maybe
- Can we use other methods to get structural information? Yes
- Can we combine everything and learn anything useful?

Why failure?

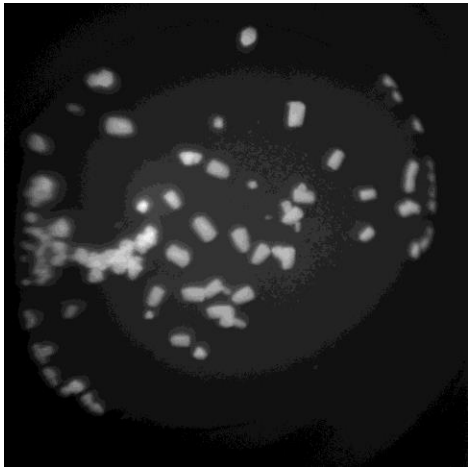
- Is it the way we are crystallizing? Possibly
- Is it the sample? Probably not
- Are we just going to have to live with it? No
- Can we learn from our previous successes and failures? Maybe
- Can we use other methods to get structural information? Yes
- Can we combine everything and learn anything useful?

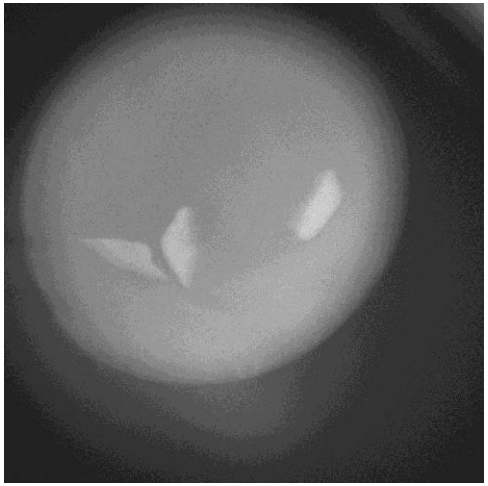
This is where we are heading

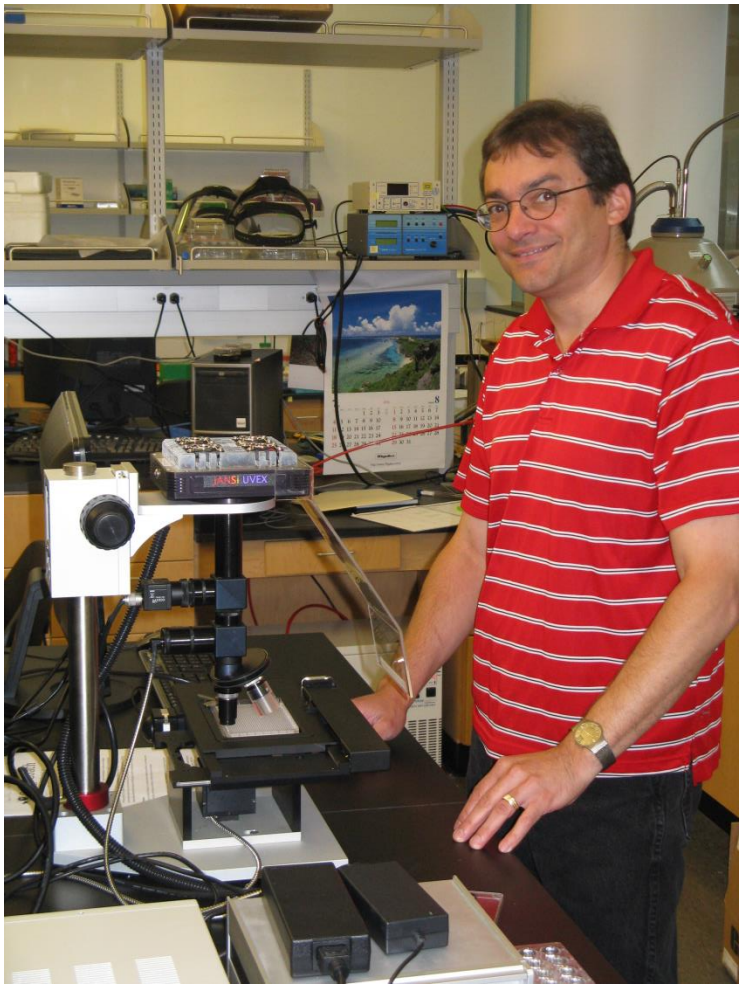
Acknowledgements

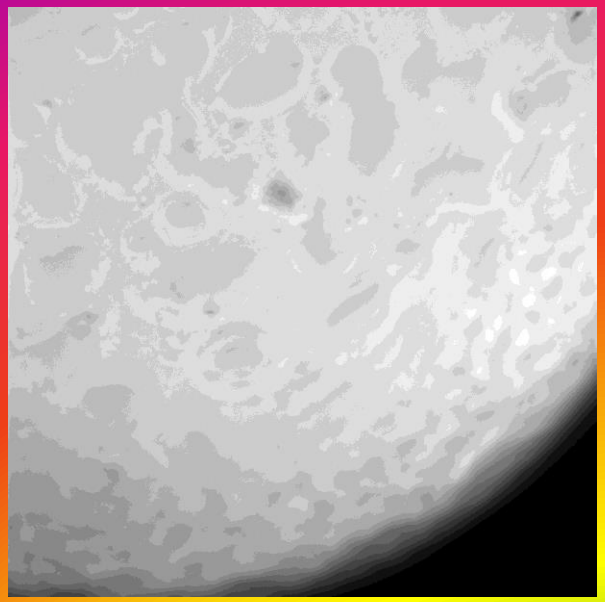
- Hauptman-Woodward
  - Joe Luft
  - Tom Grant
  - Elizabeth Snell
  - Jen Wolfley
  - Angela Lauricella
  - Tina Veatch
  - Stephen Potter
  - Wayne Schultz
  - Paige Hang
- SSRL
  - Hiro Tsuruta
  - Anne Martel
- NESG
  - Guy Montelione
  - (and others on related projects)
- CCR University of Buffalo
  - Steve Gallo
  - Andrew Bruno
- Ontario Cancer Institute
  - Igor Jurisica
  - Christian Cumbaa
- University of Washington
  - David Baker
  - Dominik Gront
- Funding
  - Oishei Foundation
  - Goode Foundation
  - NIH

UV imaging – is it protein?

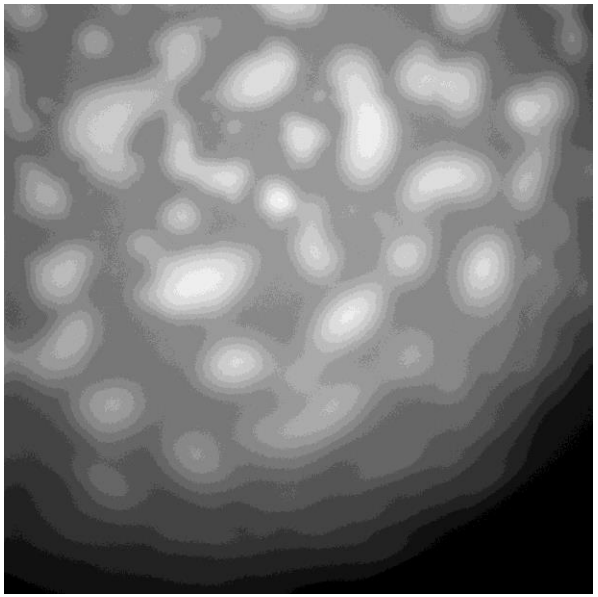


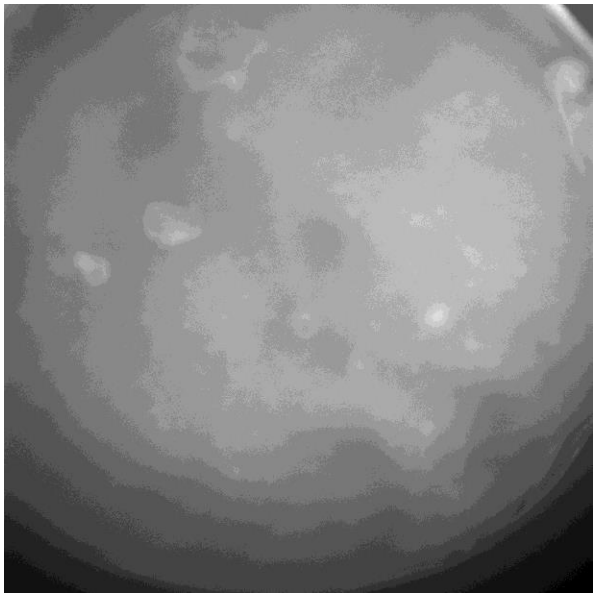






Visible

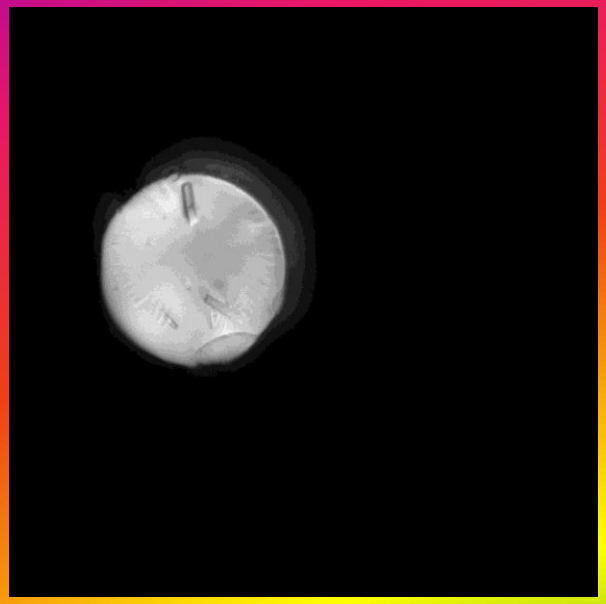


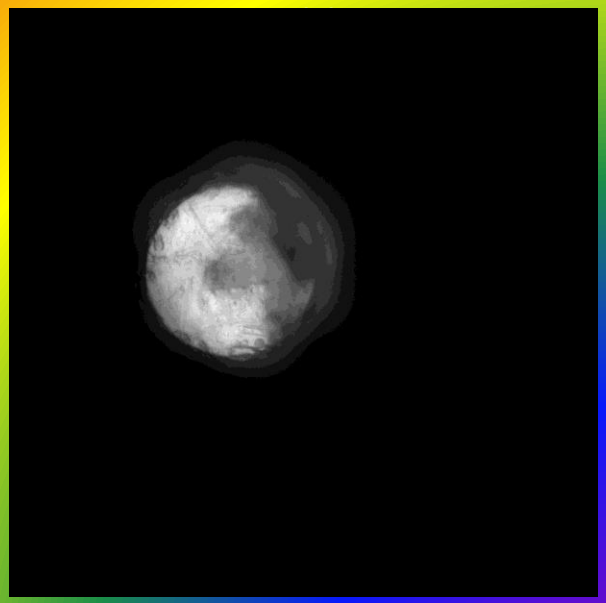


UV

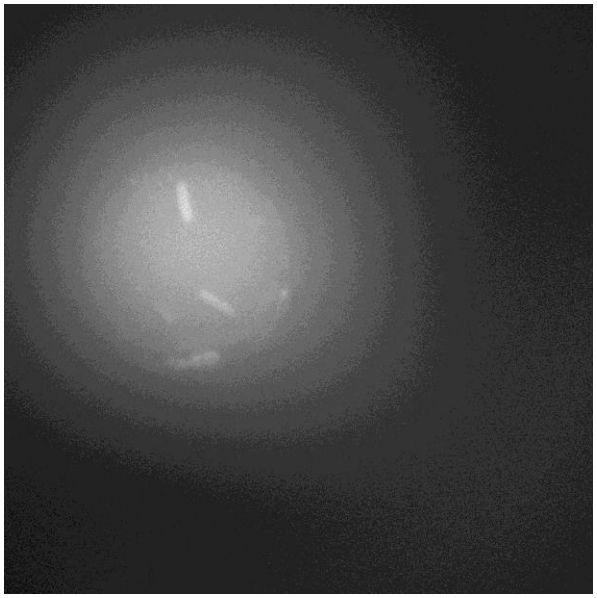
Protein phase

Protein crystal

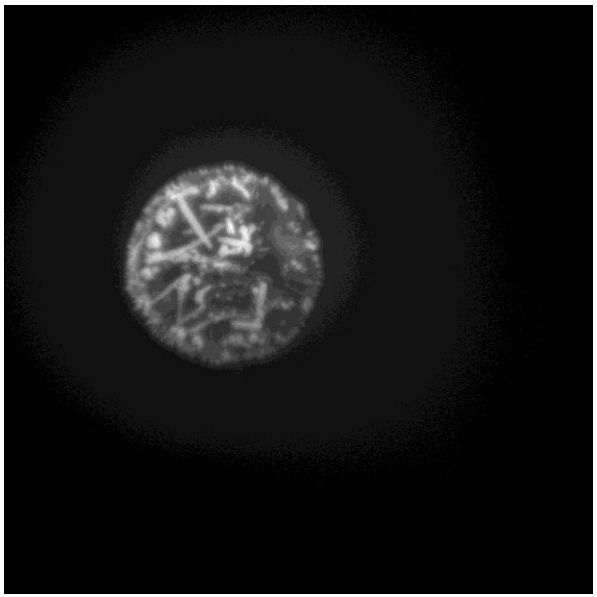




Visible

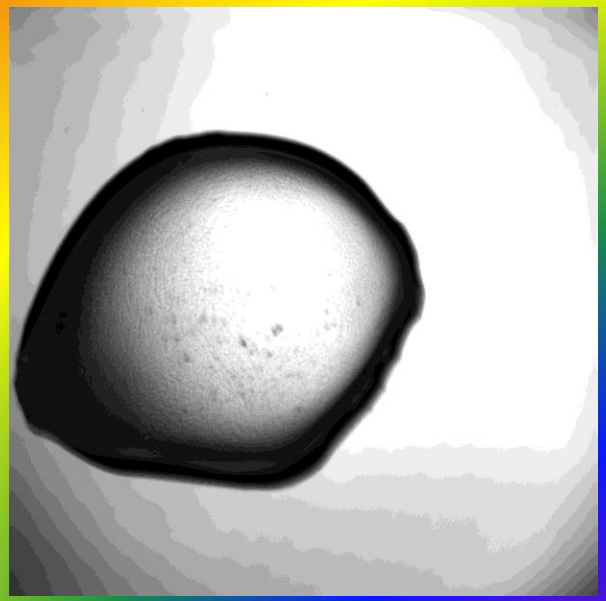


Protein crystal

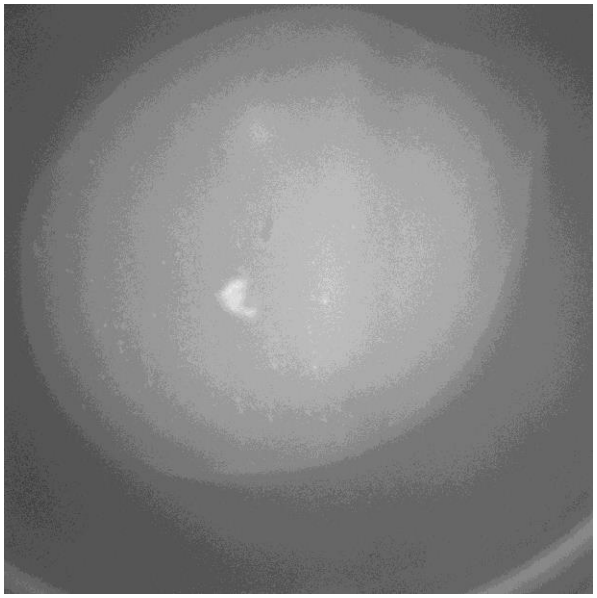


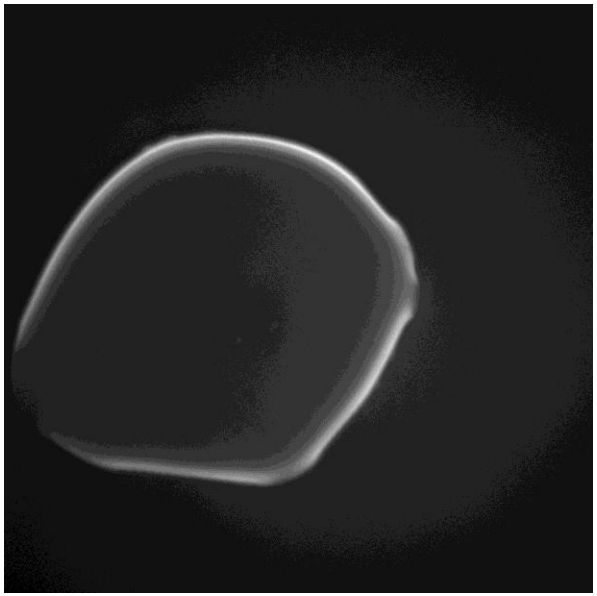
Protein crystal

UV



Visible

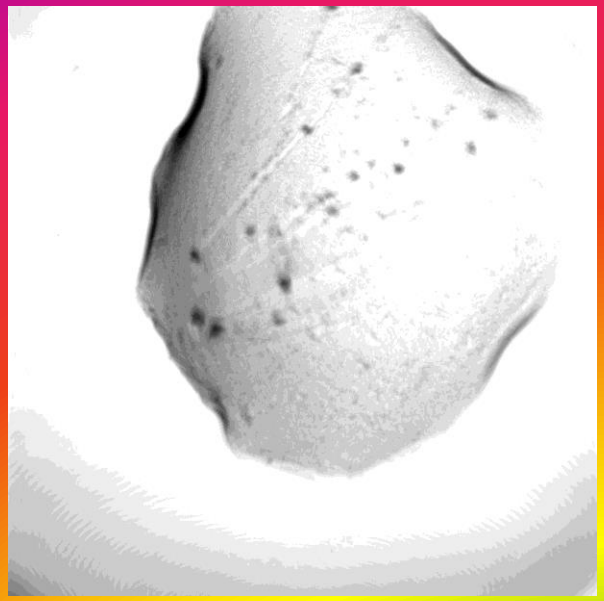


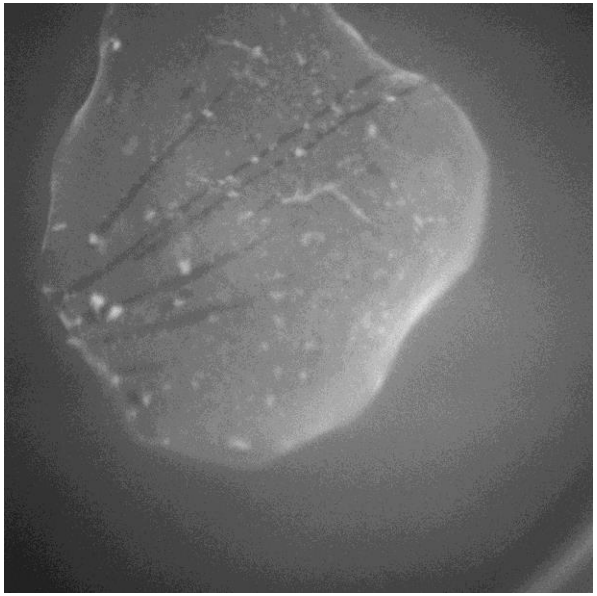


UV

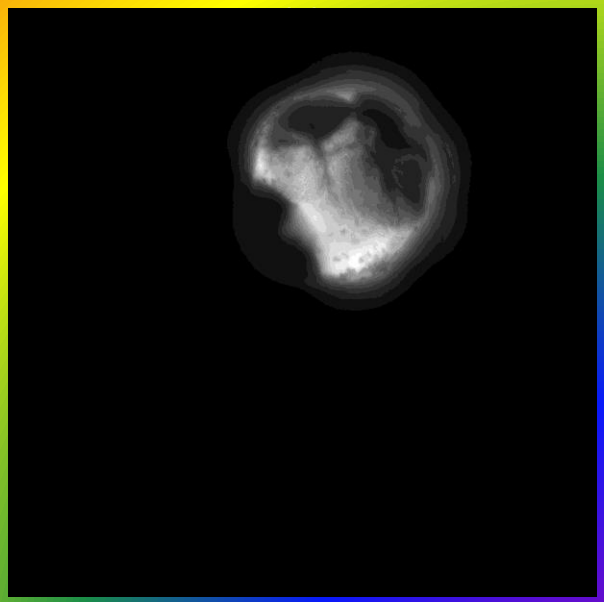
Protein crystal

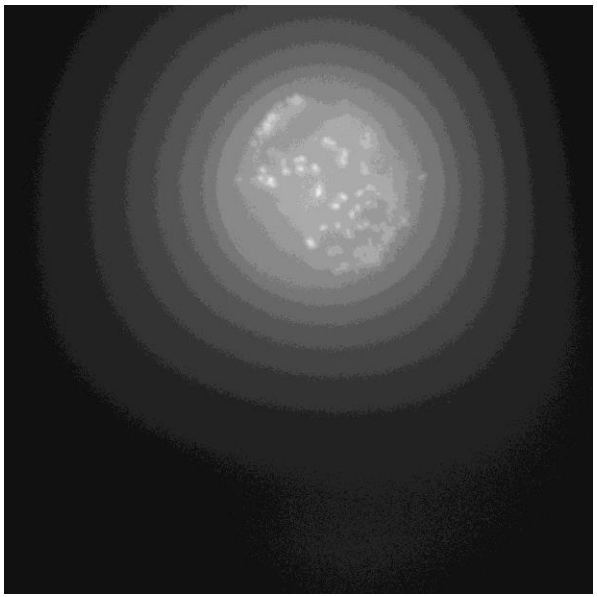
Salt crystals





Protein crystals





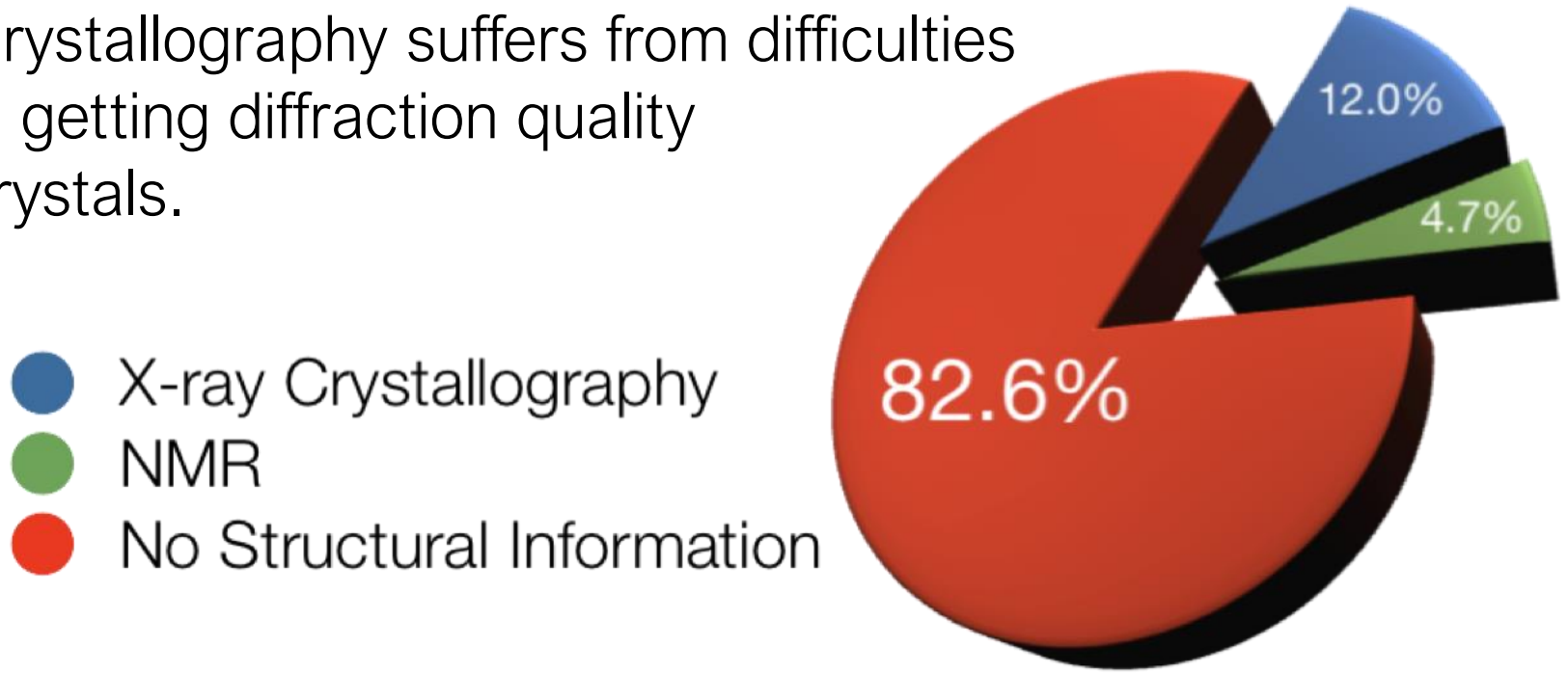
Protein crystals

Visible

UV

High-Throughput Structure Success

- According to TargetDB, 82.6% of soluble, purified targets provide no structural information.
- NMR - limited by protein size. Less than ~35 kDa.
- Crystallography suffers from difficulties in getting diffraction quality crystals.



High-Throughput SAXS Success

- ~350 targets submitted for SAXS data collection.
- 25% failed due to sample handling / instrumentation error
- Of remaining 260 targets, 23% suffered from concentration effects and/or aggregation.
- 2 were natively unfolded and not used for envelope reconstruction.
- 77% of 260 targets successfully gave structural information

High-Throughput SAXS Success