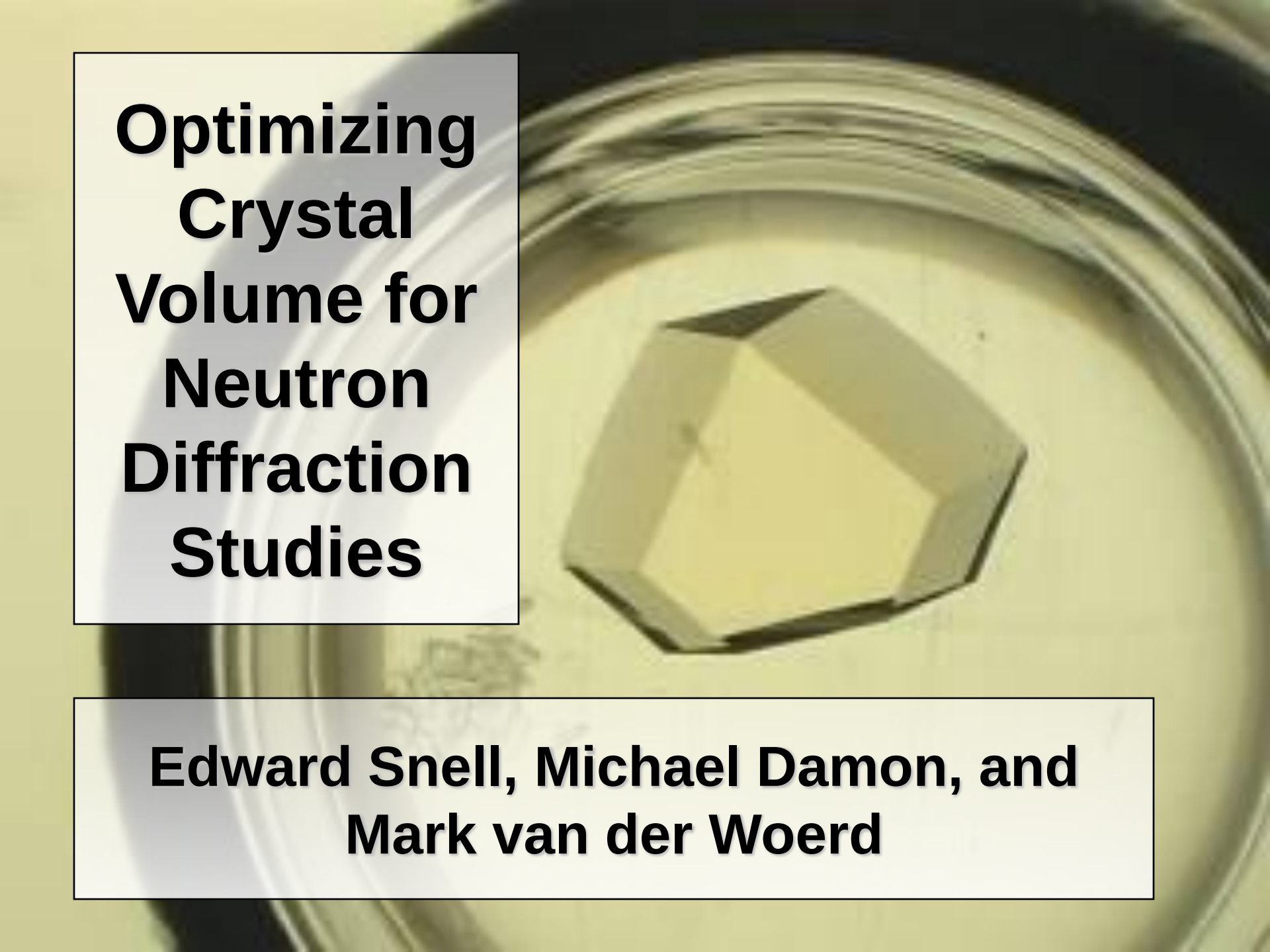
Optimizing Crystal Volume for Neutron Diffraction Studies



Edward Snell, Michael Damon, and Mark van der Woerd

Why make use of neutrons?

- For neutron diffraction the scattering amplitudes vary from element to element in a non systematic way - atoms of similar atomic mass can be easily distinguished.
- The scattering amplitude of hydrogen is of the same order of magnitude as the amplitudes of other atoms typically found in biological molecules - hydrogen atoms can be seen thereby;
  - revealing whether a particular acidic group is dissociated or has a hydrogen atom bound to it,
  - discriminating between water and hydroxyl anion in the active site of an enzyme,
  - determining the orientation of a water molecule etc.
- Deuterium and hydrogen have opposite sign scattering amplitudes enabling contrast matching techniques.
- Radiation damage is not a concern.

Statement of the problem:

Neutron sources have low fluxes:

- For example, the LADI (Laue Diffractometer) experimental station at Insitiute Laue Langevin has a flux of $3x10^7$ neutrons cm^{-2} s^{-1} for a partially monochromatised beam (l=3.5 A, $\delta\lambda/\lambda$=20%). A monochromatic beam from a wiggler source on a synchrotron has 10 orders of magnitude greater flux.

Neutrons are weakly scattered

- Neutrons are electrically neutral and interact weakly with matter, they are scattered by the nucleus and unpaired electrons.

Solutions to the problem

	Cost	Time	Probability of success
More neutrons			
Source intensity	Very expensive (new source or possibly use focusing optics)	Long-term	Certain
Source distance	Expensive (new station)	Medium-term	Certain
Exposure time	Inexpensive but reduces experimental throughput	Immediate	Certain
Better detection			
New and improved detector technlogy	Expensive	Medium-term	Good
Improve signal-to-noise			
Deuteration	Moderately expensive	Short-term	Good
Diffracting volume			
Grow larger crystals	Relatively inexpensive	Short-term	Good

Simplest solution: Bigger Crystals

The growth condition has a number of variables, e.g. protein concentration, precipitant concentration and temperature. Changing these variables changes the outcome of the experiment.

The initial condition is known, crystals already exist.

Conditions where no crystals are produced are known.

The goal is to grow a “few, large” crystals. This leads to two quantifiable metrics, crystal number and crystal size.

By changing the experimental variables within the known area of crystallization the experiment can be optimized to produce a few large crystals – the trick is doing this efficiently.

Microgravity routinely provides larger crystals but microgravity opportunities are limited so we need a way to grow bigger crystal on the earth

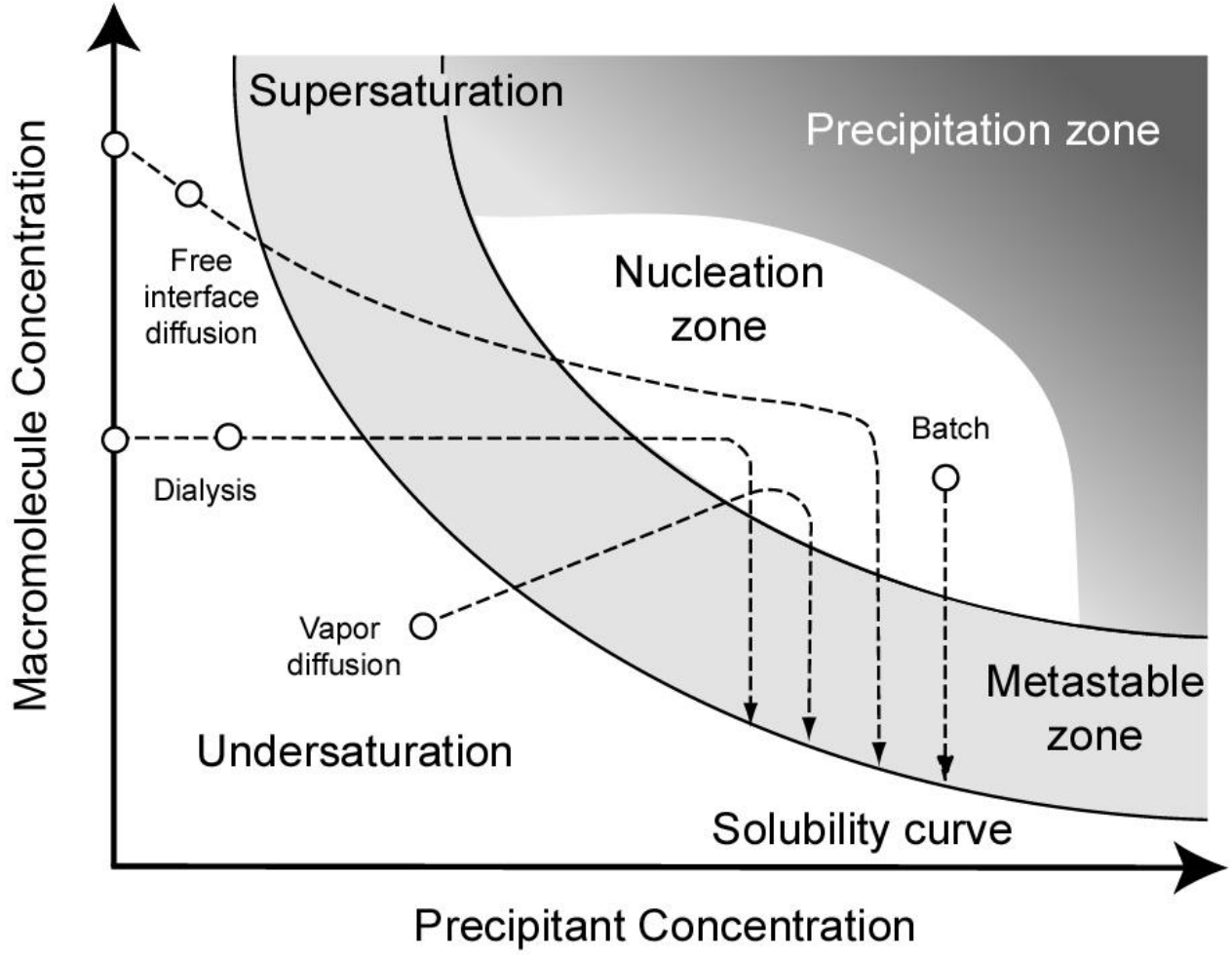
Design the experiment

Recognition of and statement of the problem.	→	Grow a few, large crystals
Choice of factors, levels and ranges.	→	Start experiment with known crystallization conditions and use range where crystals occur
Selection of the response variable.	→	Number and size (volume)
Choice of experimental design.	→	Response surface method as conditions only need optimizing
Performing the experiment.	→	Keep it simple
Statistical analysis of the data.	→	Keep it simple
Resolve the problem.	→	Understand the crystallization space

Experimental: **Keep it simple**

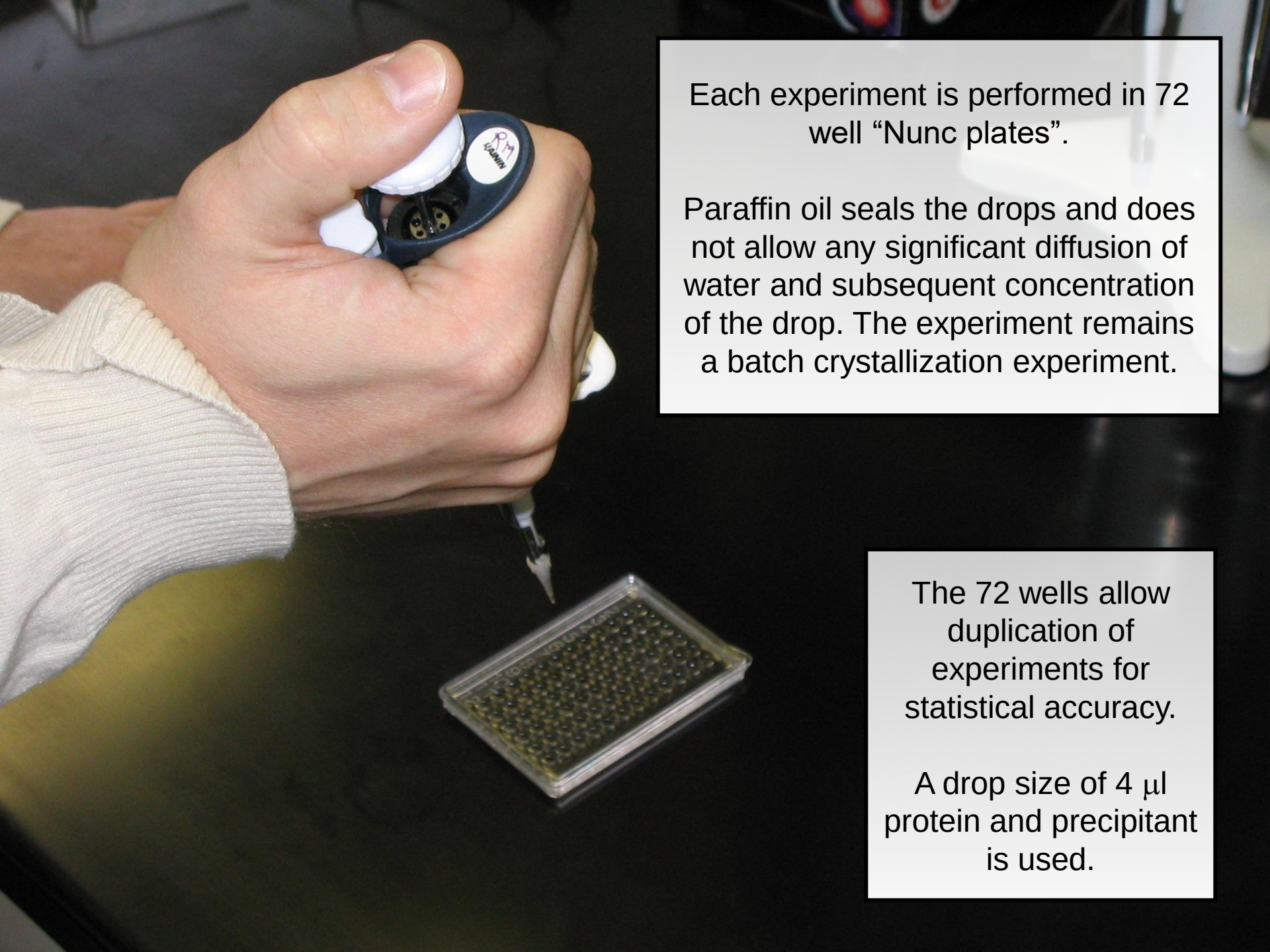
- Use the least complex method of growth
  - reduce the number of variables
  - allow the result to be scaled up
  - understand the crystallization space
- Use a quantitative, repeatable measure
  - avoid qualitative descriptions
  - allow mathematical analysis
- Use the minimum number of experiments
  - but have enough to maintain statistical validity
- Use the minimum amount of sample
  - but no more than the minimum

Method of Growth



The batch method is the simplest method of crystallization.

We use the microbatch technique and a highly efficient crystallization robot for experiment setup



Each experiment is performed in 72 well “Nunc plates”.

Paraffin oil seals the drops and does not allow any significant diffusion of water and subsequent concentration of the drop. The experiment remains a batch crystallization experiment.

The 72 wells allow duplication of experiments for statistical accuracy.

A drop size of 4 μl protein and precipitant is used.

Case Study: Xylose Isomerase

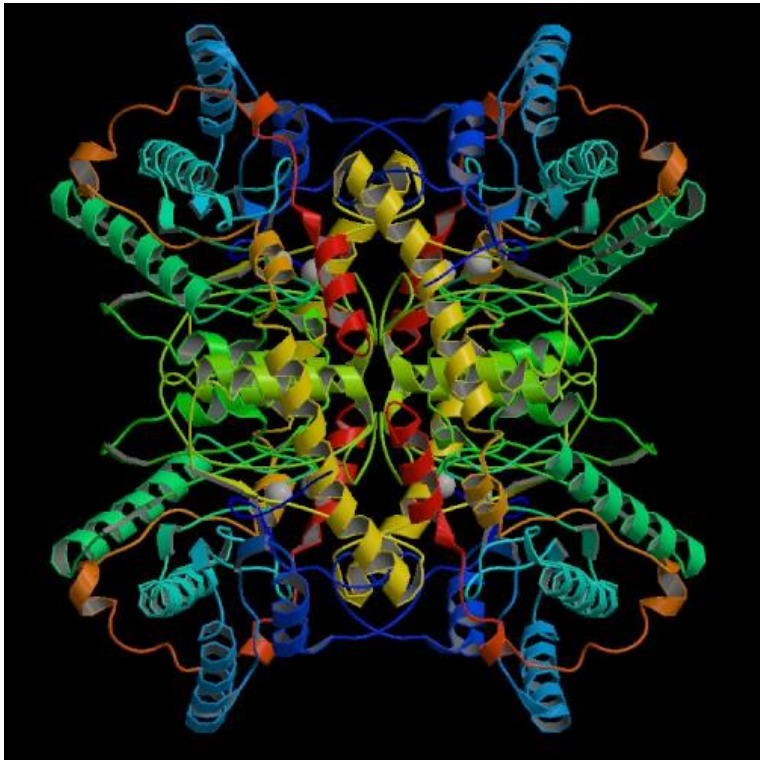
Enzymatic mechanism is a transfer of one H atom from one C atom of the substrate to an adjacent C atom.

Three mechanisms have been proposed – a base-catalyzed proton transfer, a simple hydride shift or a hydride shift mediated by a metal ion.

X-ray data, to date, has not revealed the exact mechanism.

Xylose isomerase is an important industrial catalyst for the production of fructose.

It has a molecular weight of ~ 160 KDa and can crystallize in the I222 space group – useful for Laue studies as every second reflection is systematically absent.

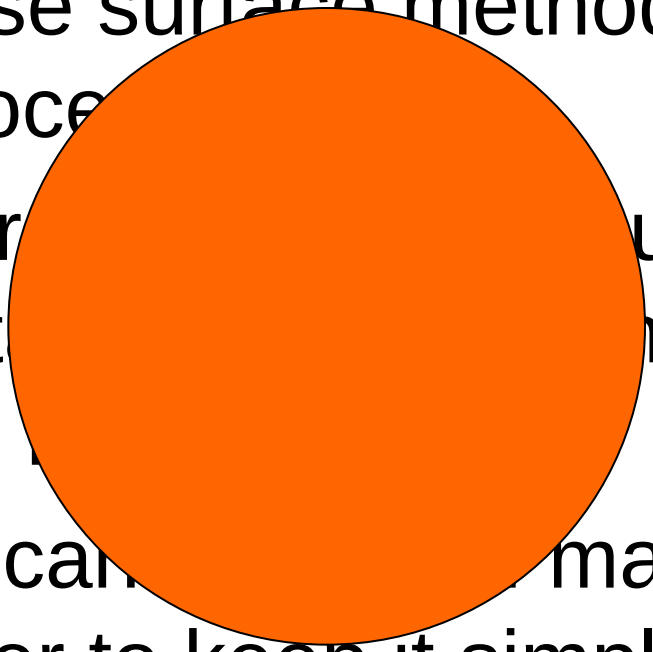


Design of experiment

- The goal is to maximize crystal volume and minimize crystal number.
- Response surface methods offer a simple method to achieve this goal.
- A response surface is a plot of a function derived from the measured response to variables of interest.
- Crystallization conditions (variables) are already known:
  - the function fitted to the response surface will be a second order function allowing for non-linear interactions between the experimental variables.
  - If crystallization conditions were not adequate a first order fit would be used to move the conditions along the path of steepest ascent to the optimum.
- A second order model requires an appropriate experimental design.
  - The most common design is the central composite design. We do not know exactly where the optimum will be so a spherical central composite design is used providing equal precision of estimation in all directions.
- A spherical central composite design for two variables:
  - Has nine individual experimental points were one point is the center of the design, four points are on a circle surrounding the optimum and four points are outliers at the maximum and minimum values of the variables of interest

Itterative process

- The response surface method is an itterative proce
- A coarse scr und the known cryst nd conditions surrounding
- The screen can many variables but it is better to keep it simple initially.



Analysis of experiment

The results are fit to the model:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum \sum_{i<j=2}^{k} \beta_{ij} x_i x_j + \varepsilon$$

Where y is the response, x are the variables, b are constants and e is the error or noise in the model.

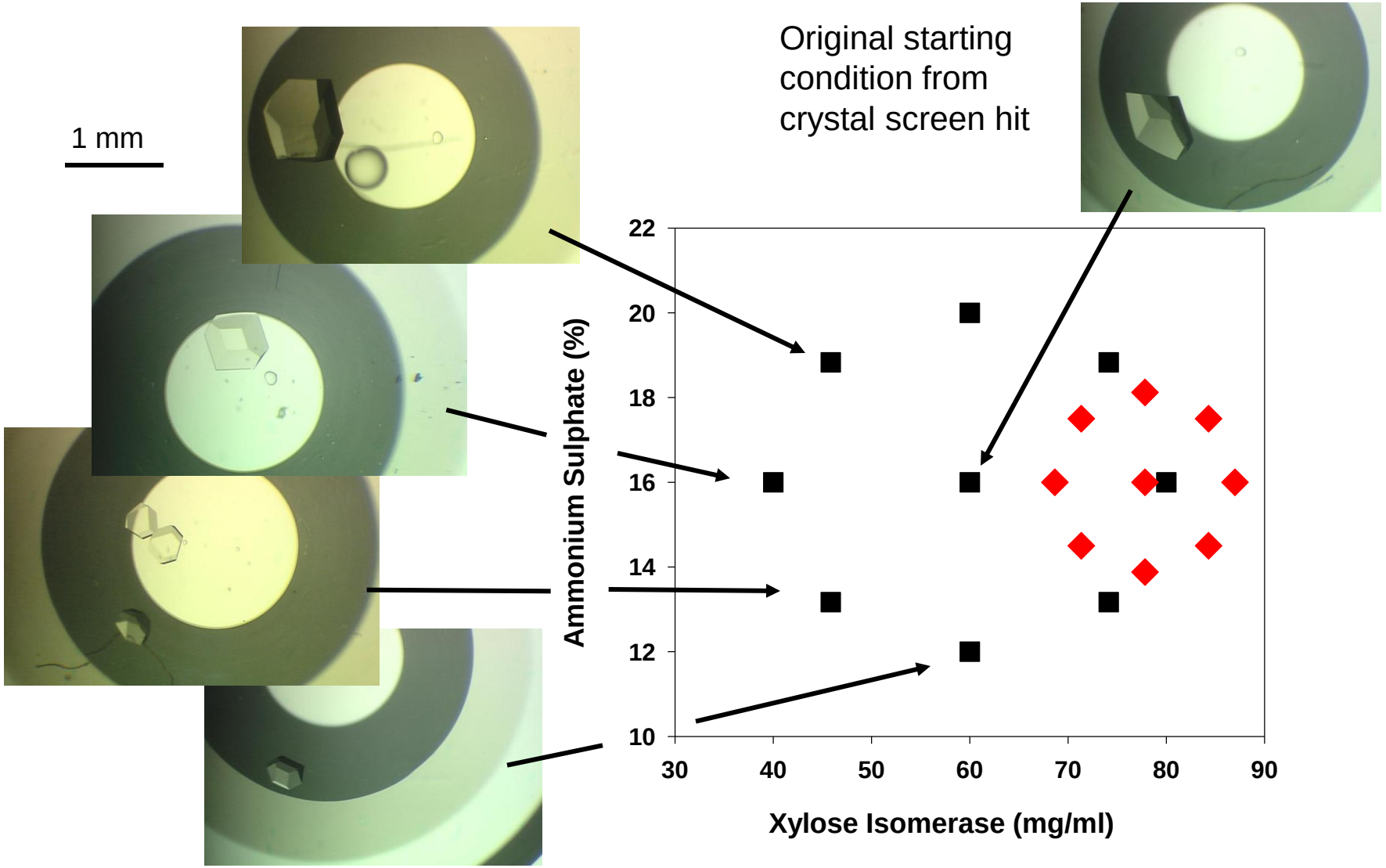
The method of least squares is used to fit the model. This also provides a number of checks for the validity of the model.

A linear model can be used for the coarse screen and the method of steepest ascent used to choose a central point for fine screen optimization.

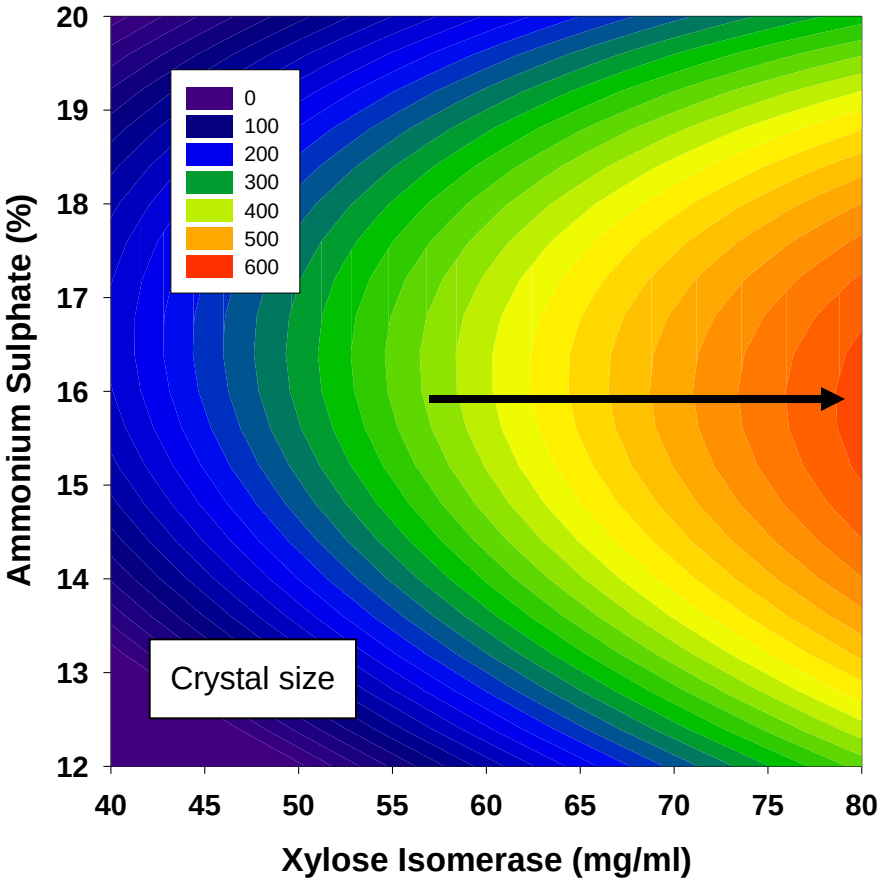
Experimental detail

- Initially 72 experiments were setup
  - 8 replicates at each experimental condition.
- Crystals were grown at 4 temperatures
  - 14,18, 22, 26.
- The crystals were analyzed
  - Size (largest dimension)
  - Number (count up to 100, estimate greater than that).
- The experiment was optimized
  - The origin was moved.
  - Range was decreased.
- Analysis took place
  - Number was not a good metric
  - Size was a good metric

Coarse screen (with fine screen shown in red)



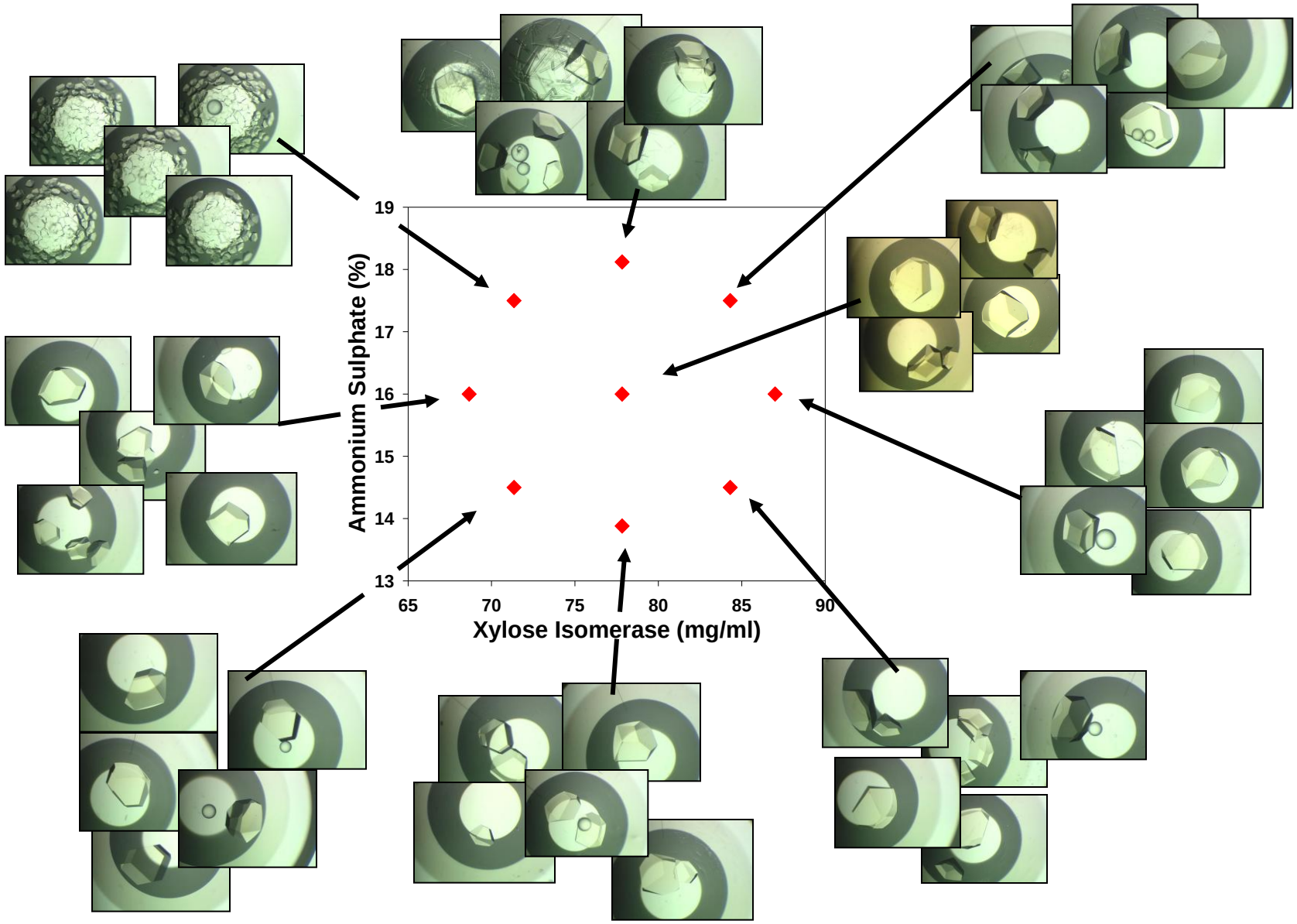
Fit to coarse screen



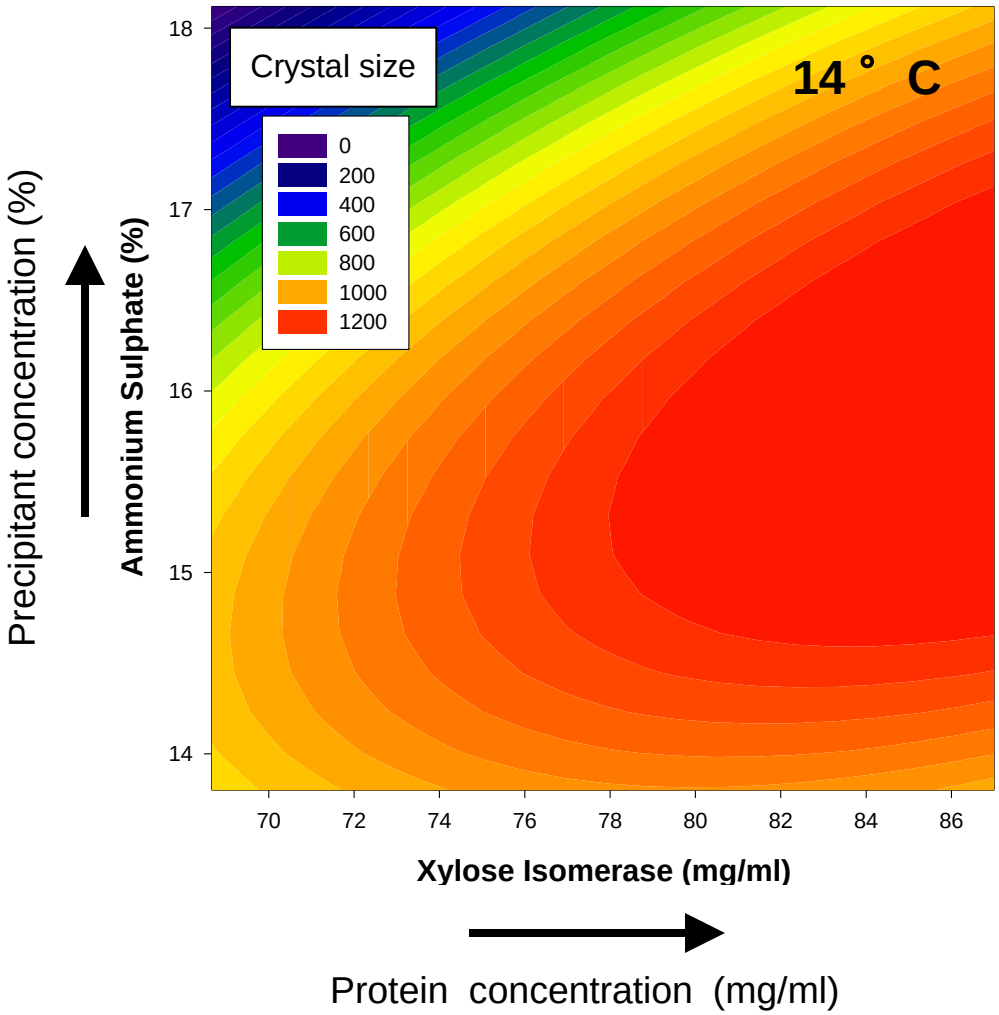
Use coarse screen to move new center of optimization experiments to the largest crystal region.

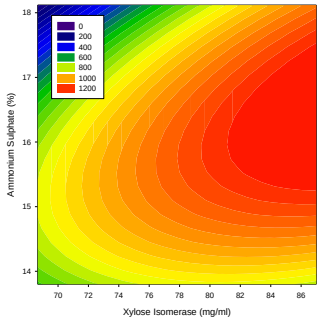
Note, the highest peak seen on this graph is 600 μm

Experimental design for fine tuning:

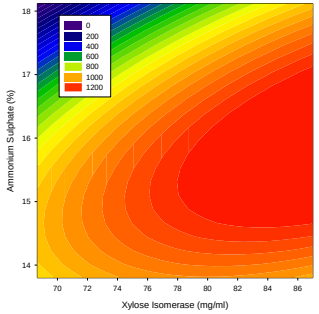


Crystal size in the experimental space sampled

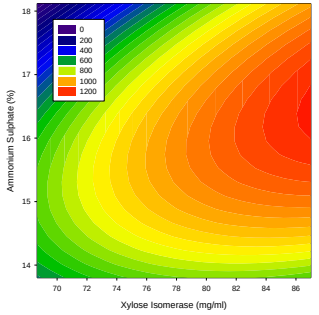




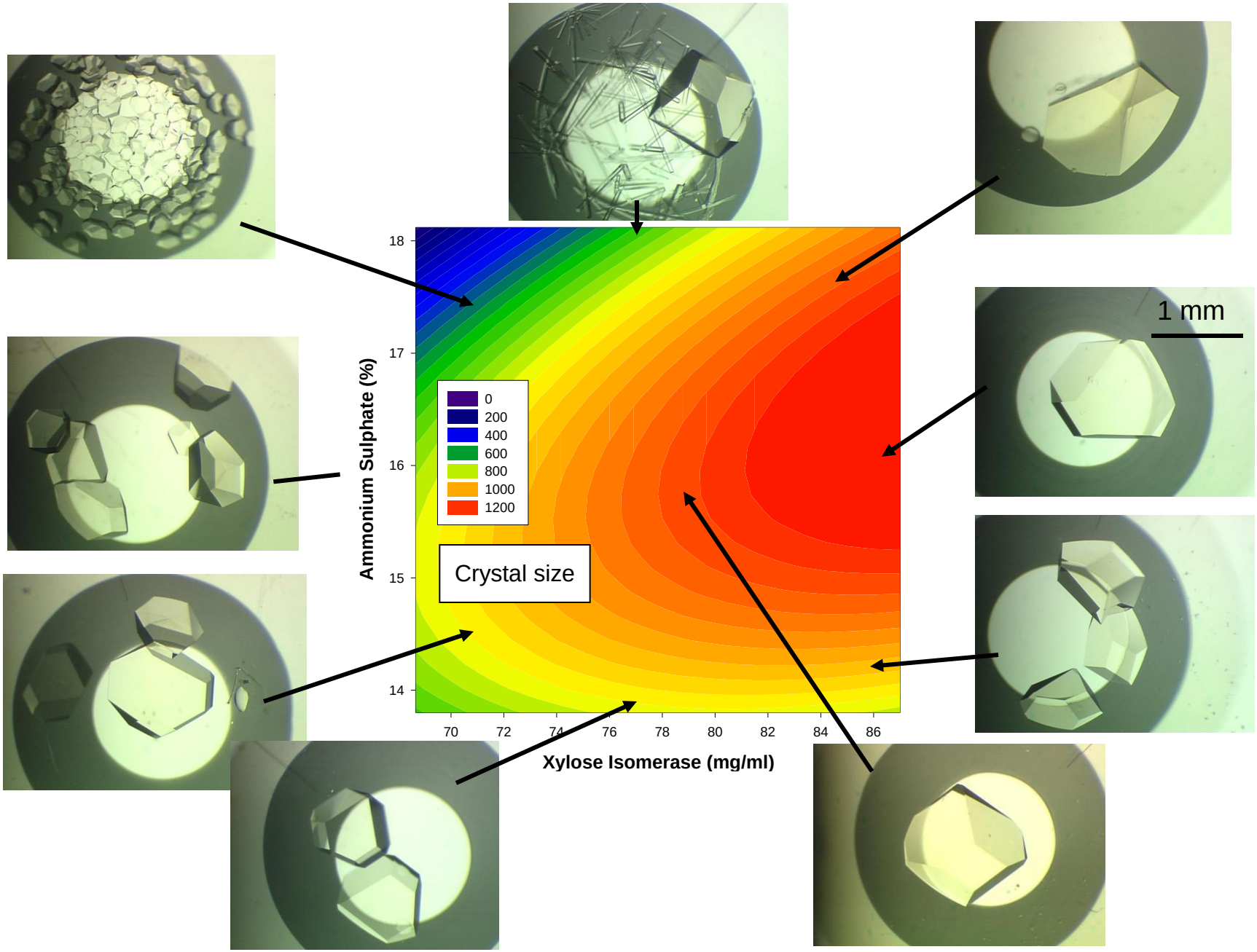
18 ° C



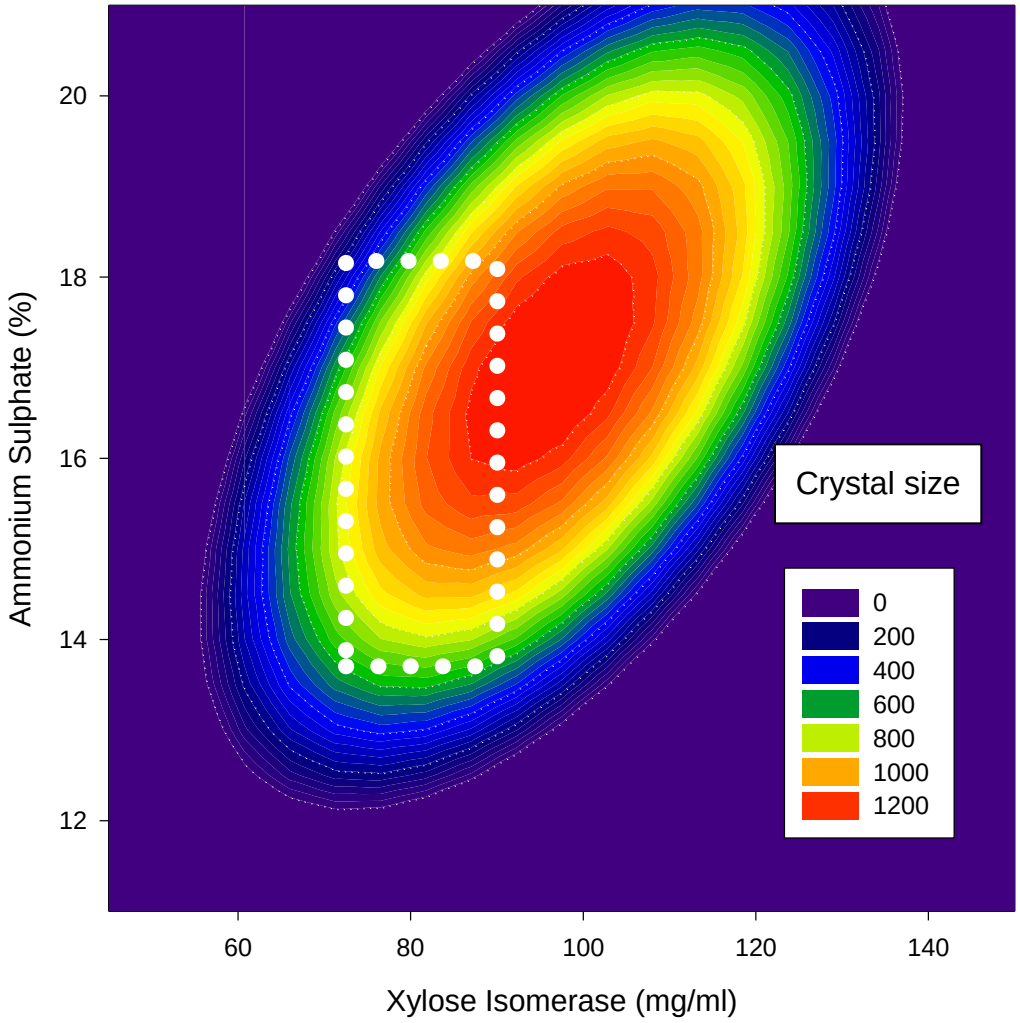
22 ° C



26 ° C


Ammonium Sulphate (%)
18
17
16
15
14
70
72
74
76
78
80
82
84
86
Xylose Isomerase (mg/ml)
0
200
400
600
800
1000
1200
Crystal size
1 mm

Plot of the predicted model



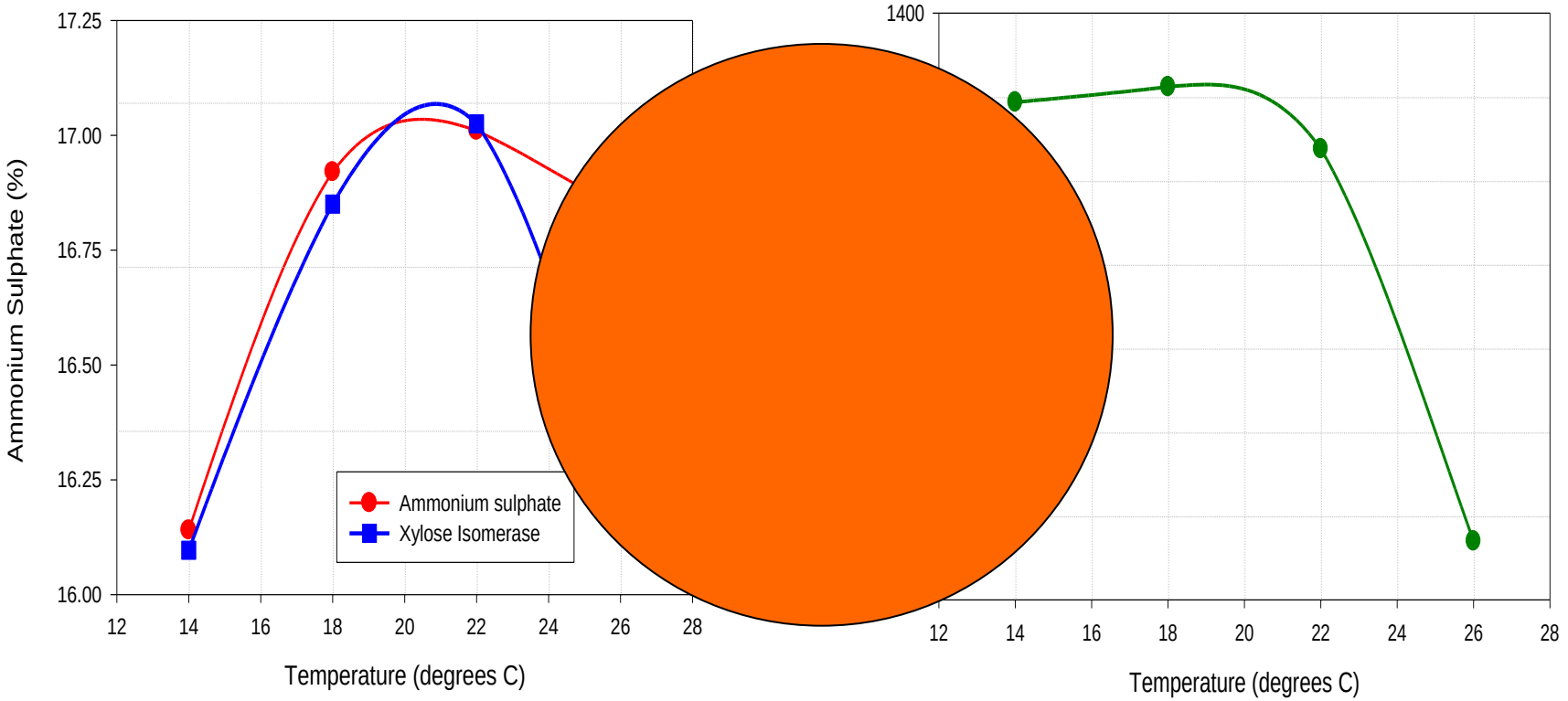
A limited crystallization space is sampled.

Extrapolating the second order model over a wider precipitant and protein range produces a response surface.

The peak is wide in this case.

Taking the peak for each temperature we can profile the conditions to produce the largest crystals.

Maximum predicted crystal size

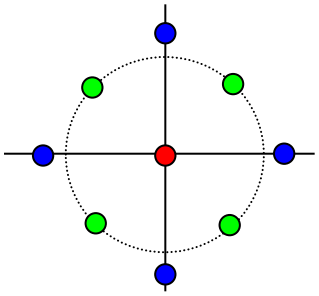


Gives an understanding of the crystallization space in terms of precipitant, protein and temperature.

How many experiments are needed?

Experimental	Total experiments (at each temperature)	72	36	26	26	18
	Replicates of factorial points	8	4	2	4	2
	Replicates of axial points	8	4	4	2	2
	Replicates of center point	8	4	2	2	2
Results	Model significance (F value)	40.67	17.18	14.00	13.79	5.84
	Lack of fit (F value)	17.27	12.78	8.48	8.85	4.64
	Prediction power	23.38	15.27	14.32	13.37	8.66
	Adjusted R square	0.659	0.613	0.621	0.617	0.488
	Predicted R square	0.639	0.563	0.539	0.539	0.336

Central composite design

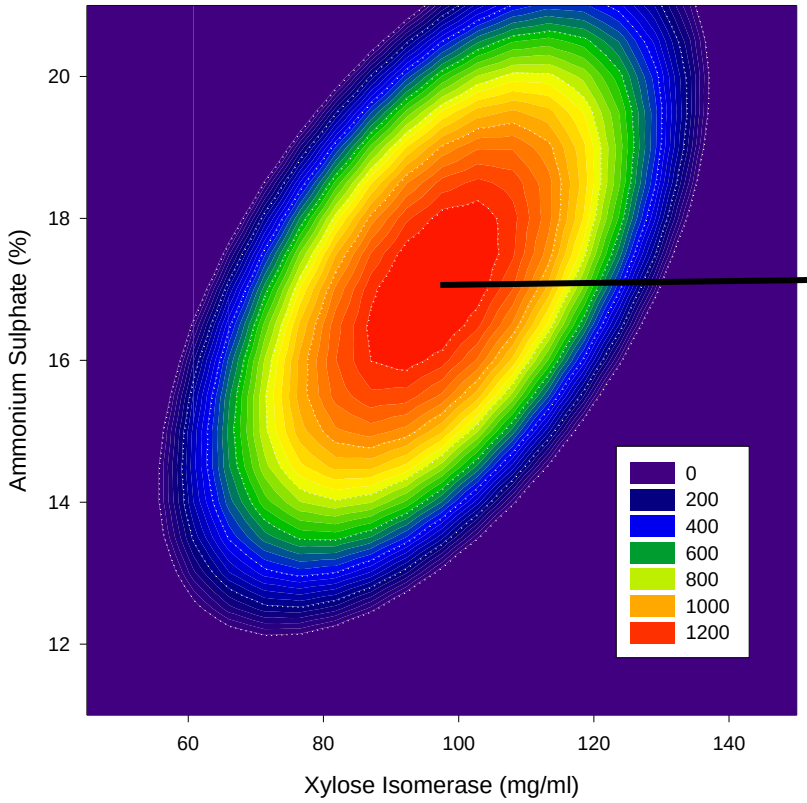


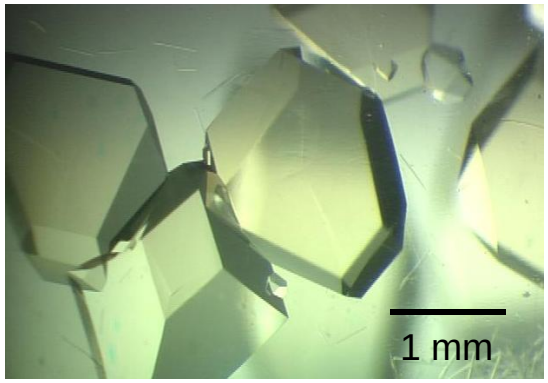
- Model significance in each case is high – there is a 0.01% chance that the fit could occur due to noise.
- The lack of fit is also high – the model is not a perfect description of the process.
- The prediction power is a measure of how well the model can be used to navigate the crystallization space (>4 is good).
- Adjusted R square is a maximum of 1.00 for a model that explains 100% of the data (>60% is good given the imperfect model).
- Predicted R square is a maximum of 100% for a model that predicts 100% of the data.

Answer – the more the better but even 18 provides useful predicting power

Making use of the model

The batch method can be easily scaled up.





Crystals grown in a PCR tube by the batch method.

Note: In this case the optimum region is large enough that substituting D2O for water still produces similar large crystals

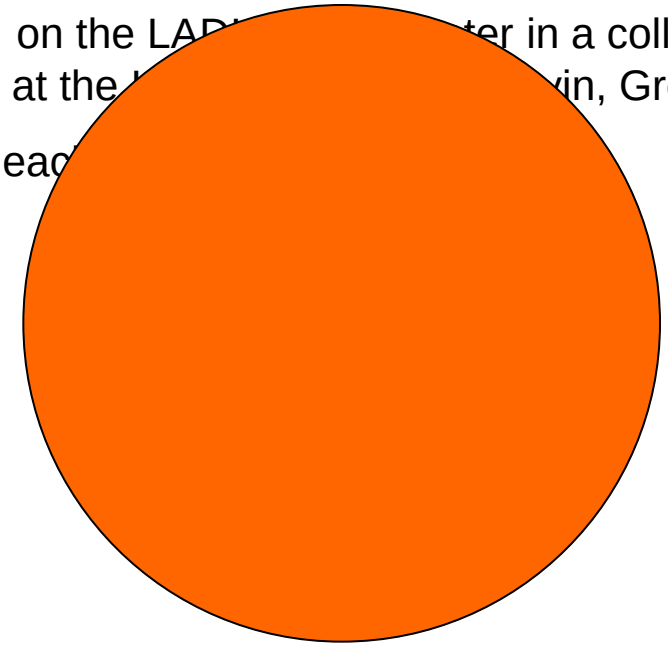
18C, 16.87% Ammonium Sulphate, 95 mg/ml Xylose isomerase

Neutron data – collected on the ILL LADI Line

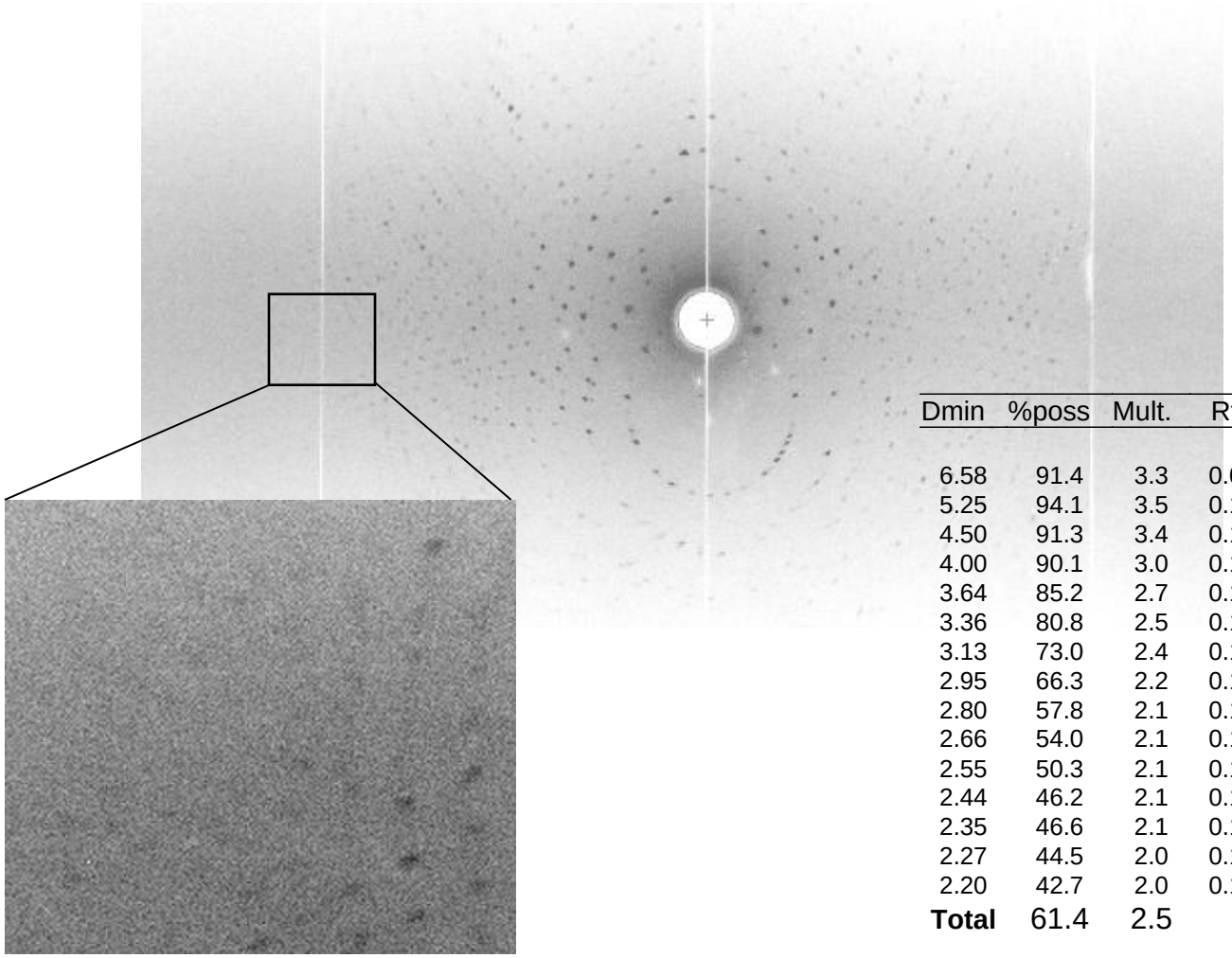
Large crystals were grown with (a) no duteration, (b) exchange with D2O and (c) in solutions made with D2O rather than H2O.

Laue date was collected on the LAD[illegible]ter in a collaboration with Dean Myles and Flora Meilleur at the [illegible]vin, Grenoble France.

Data was collected from eac[illegible]



Neutron data – collected on the ILL LADI Line

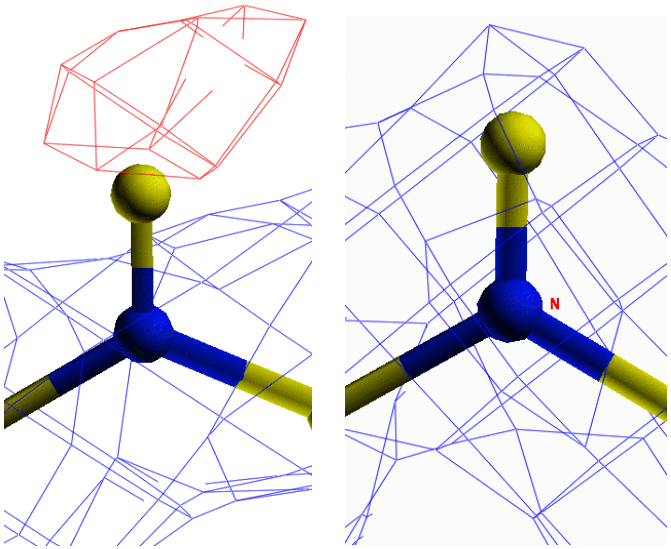


Dmin	%poss	Mult.	Rfac	Rcum	I/sigma
6.58	91.4	3.3	0.093	0.093	6.3
5.25	94.1	3.5	0.100	0.096	6.3
4.50	91.3	3.4	0.105	0.099	5.5
4.00	90.1	3.0	0.107	0.101	5.4
3.64	85.2	2.7	0.111	0.103	5.5
3.36	80.8	2.5	0.113	0.104	5.5
3.13	73.0	2.4	0.116	0.105	5.7
2.95	66.3	2.2	0.125	0.106	5.4
2.80	57.8	2.1	0.131	0.107	5.4
2.66	54.0	2.1	0.137	0.108	4.9
2.55	50.3	2.1	0.141	0.109	4.9
2.44	46.2	2.1	0.149	0.110	4.5
2.35	46.6	2.1	0.151	0.111	4.6
2.27	44.5	2.0	0.154	0.112	4.6
2.20	42.7	2.0	0.164	0.113	4.3
Total	61.4	2.5		0.113	5.1

I222 (a = 92.8 Å, b = 98.4 Å, c = 101.5 Å)

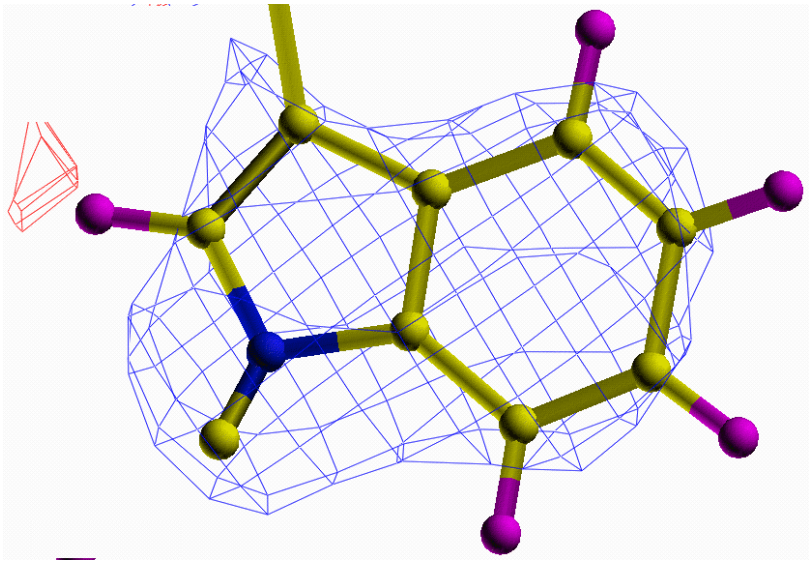
Preliminary neutron density results

All neutron density: Blue is 2fo-fc>0 Red is 2fo-fc<0



13 37

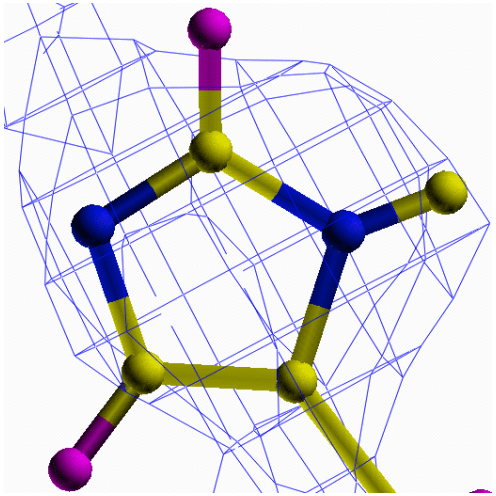
Deuterium exchange on the Ca backbone, 13 no exchange, 37 exchange likely



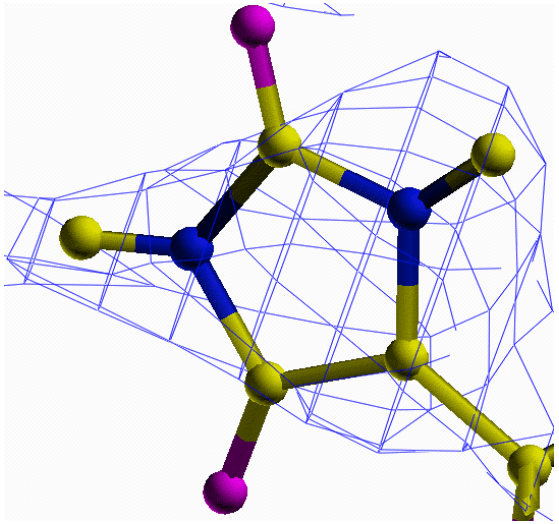
Trp 137 shows positive density for the nitrogen bound deuterium, negative for the carbon bound hydrogen

Histidines at the active site with different

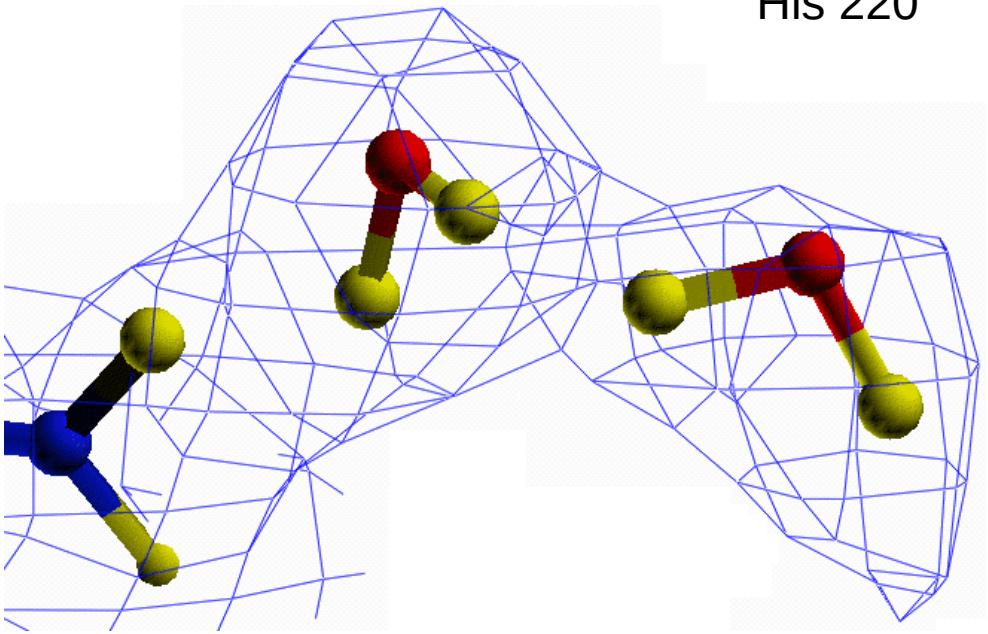
protonation states



His 220



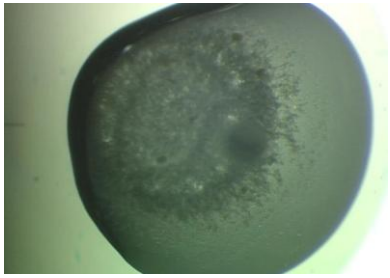
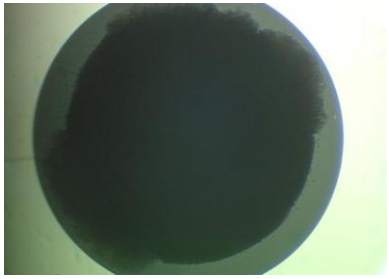
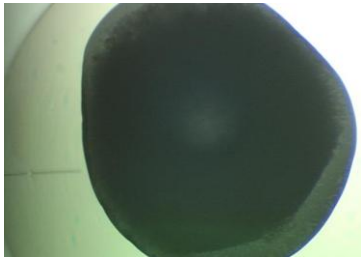
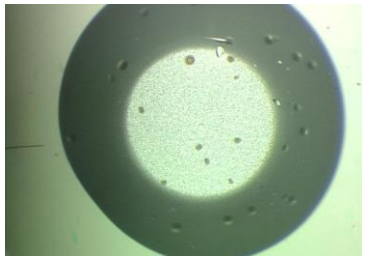
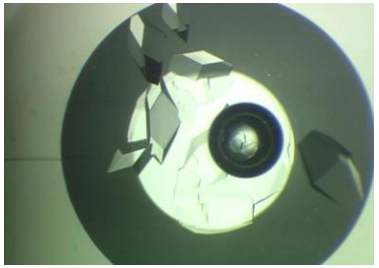
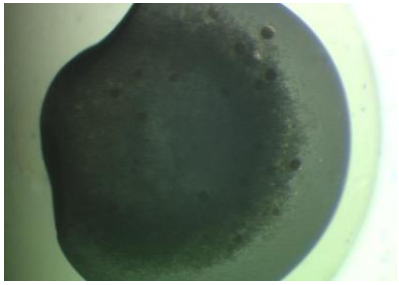
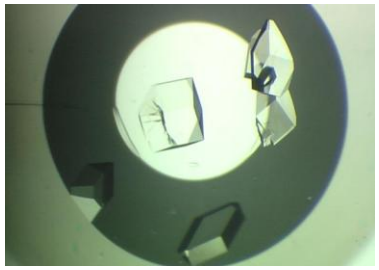
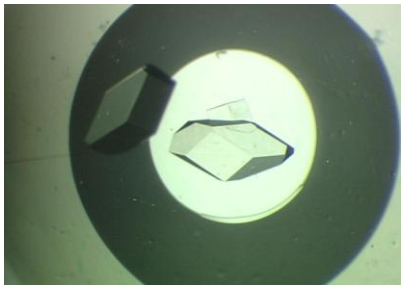
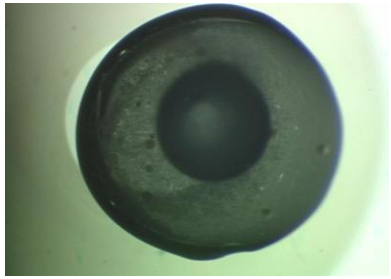
His 54



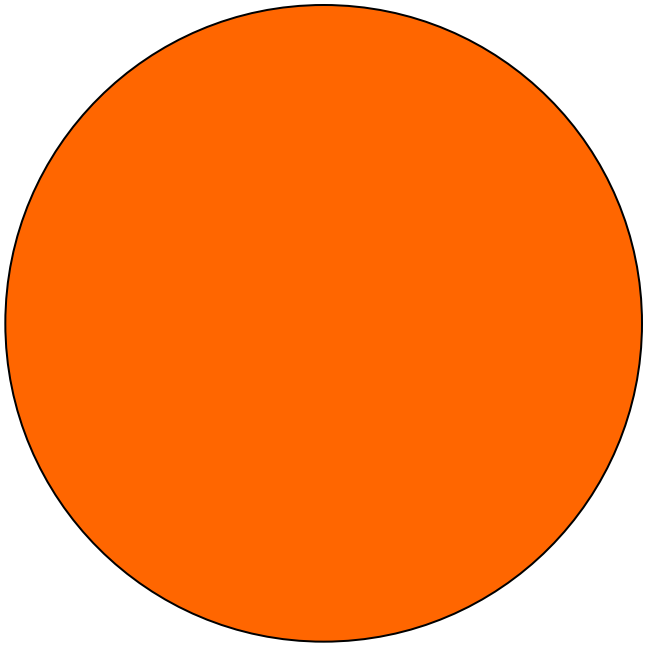
Water molecules seen as D_2O

Flora Meilleur, ILL.

An example with Lysozyme

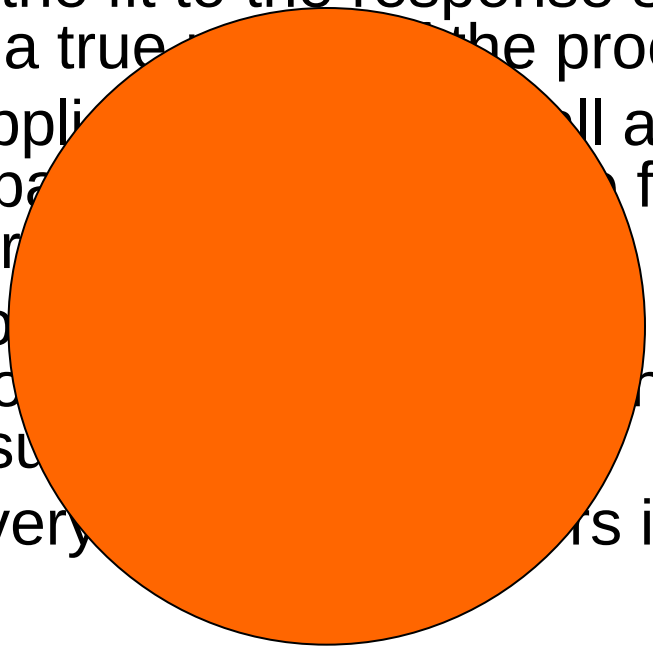










Other systems under study



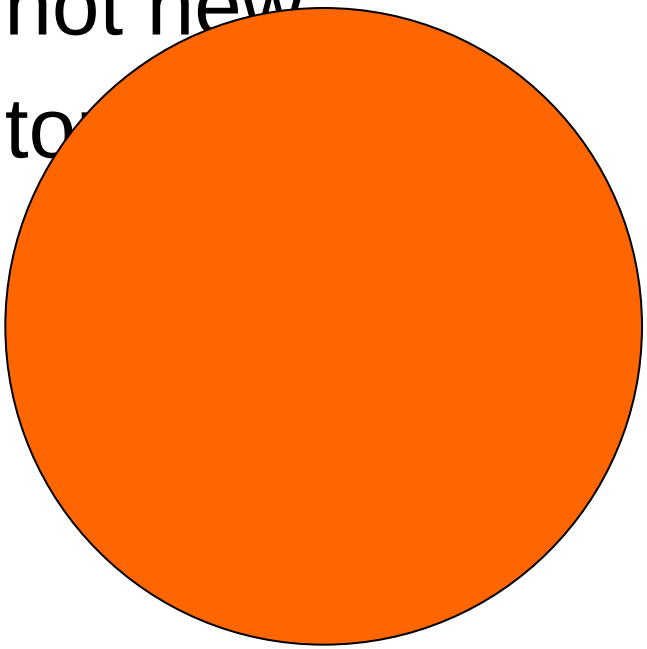
Caution!

- Although the method predicts the region to grow the largest crystals the fit to the response surface should not be regarded as a true [illegible] the process.
- The fit is only appli[illegible]ll area of the crystallization spa[illegible] fit is extrapolated the accuracy is r[illegible]
- The method is o[illegible] are initial crystallization co[illegible]nditions give reproducible resu[illegible]
- The method is very [illegible]rs in the sample measurement.



Summary

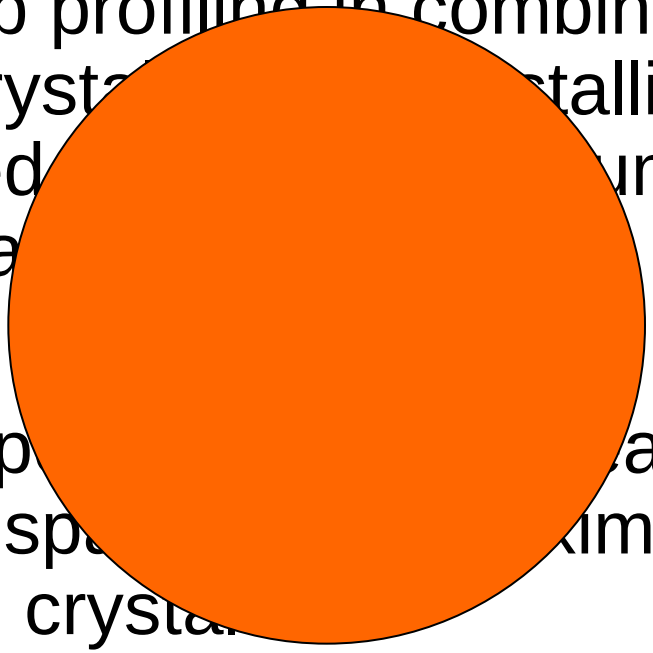
- Techniques not new
- Suited to auto



Conclusion

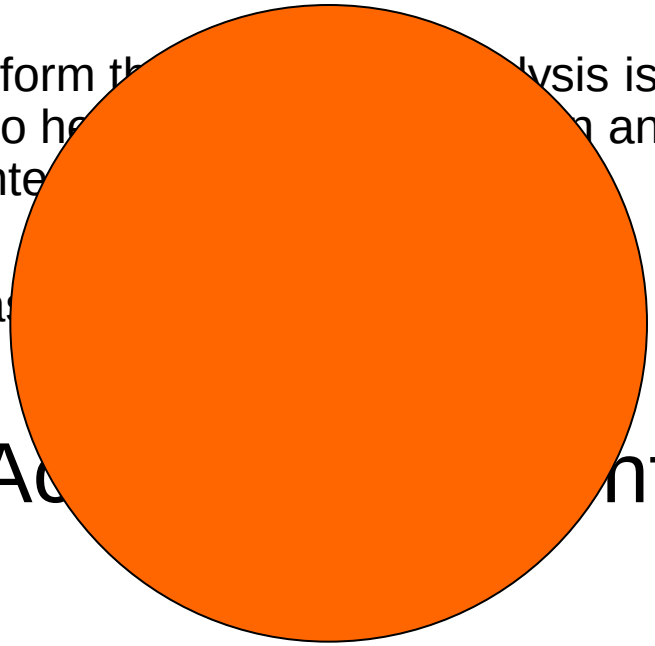
- By using design of experiment techniques, and response map profiling in combination with microbatch crysta stallization space can be profiled umber of experiments a

- The peak resp a of experimental sp imizes the volume of the cryst



Happy to Help

- The design of experiment and response surface analysis described are simple techniques.
- The time taken to perform th lysis is minimal and we would be more than happy to he and analysis for any groups that may be inte
- Eddie.Snell@msfc.na



Ac nts

Funded by NASA grants
Dr. Russell Judge, Abbott Laboratories.
Ricardo Lierdi
Richard Kephart, Dean Myles, Flora