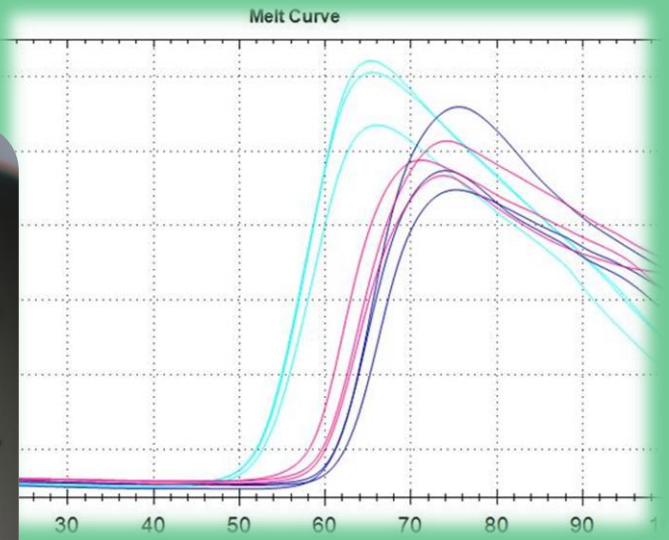


Data for Crystallisation – Answers are in the distance

Janet Newman, CSIRO Collaborative Crystallisation Centre

October, 2018

Growing crystals

For crystallisation:

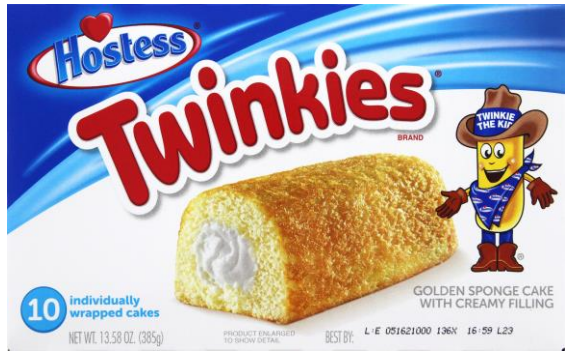
- Need supersaturation
- Need nucleation

For proteins

- Need gentle methods

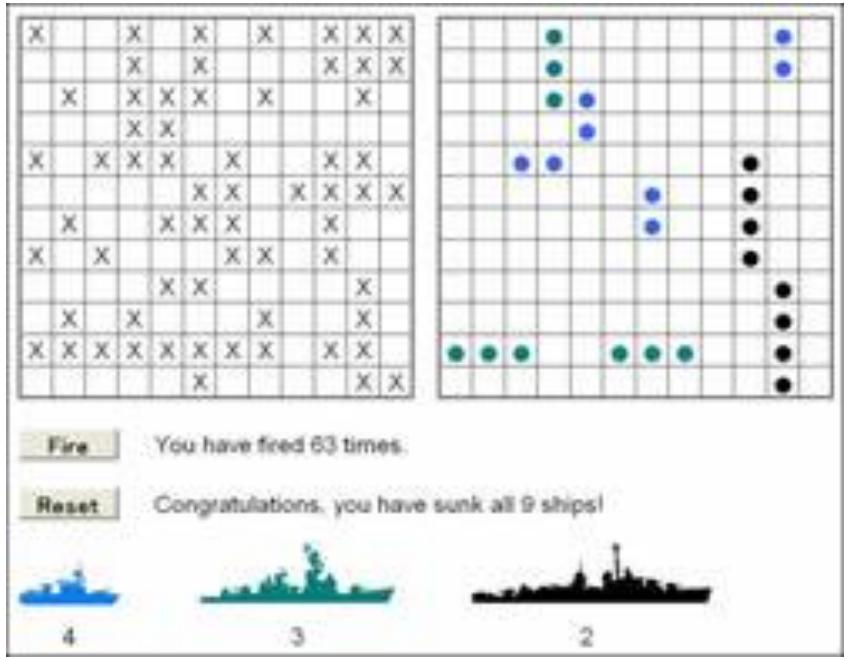
Data, information, knowledge

- Need data (need to clean it up)
- Data – understanding it – gives information
- Information – understanding it – gives knowledge

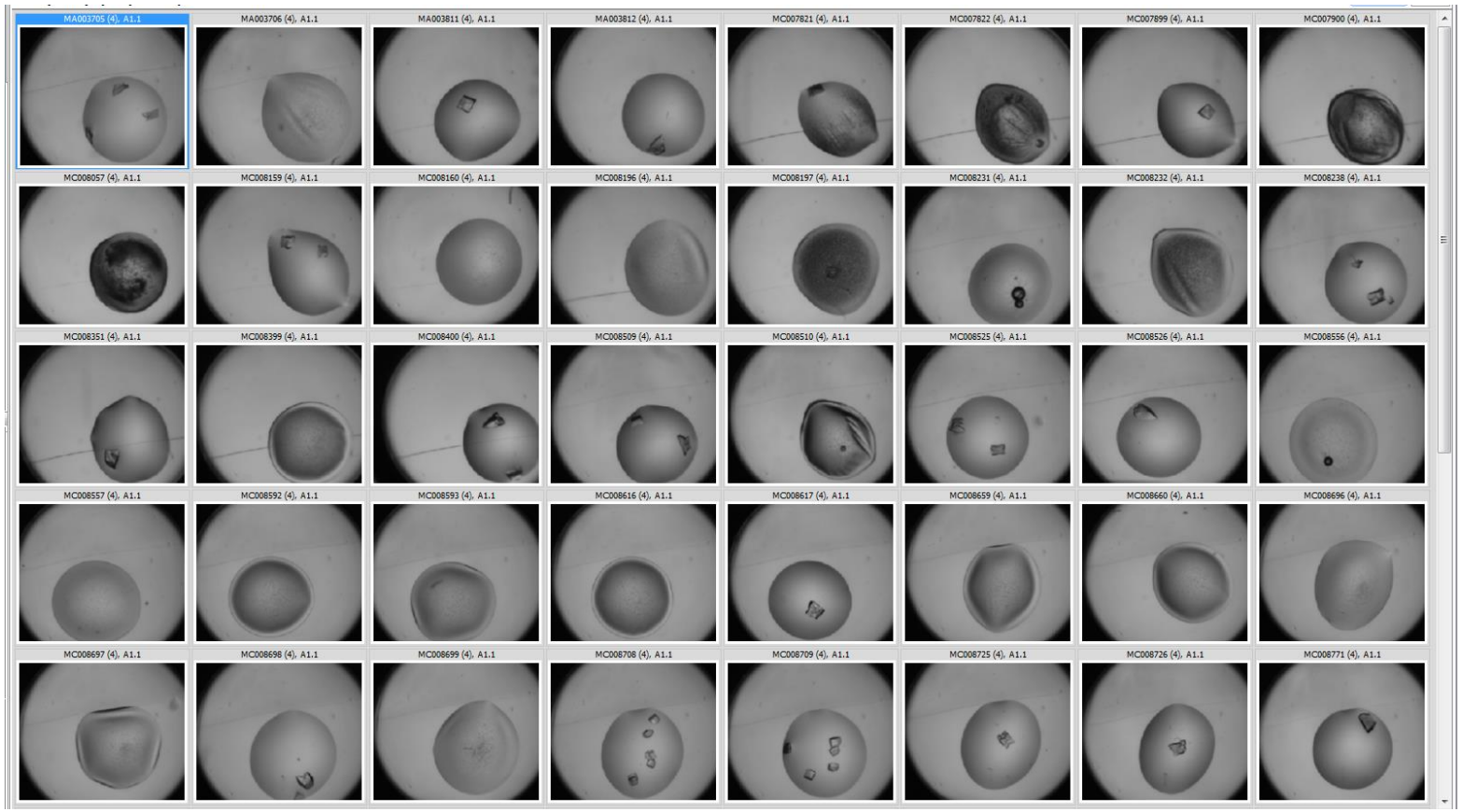


Playing battleships – looking for ships

- 20 billion point grid
- <nobody> knows where the ships are
- Not all the ammunition is live



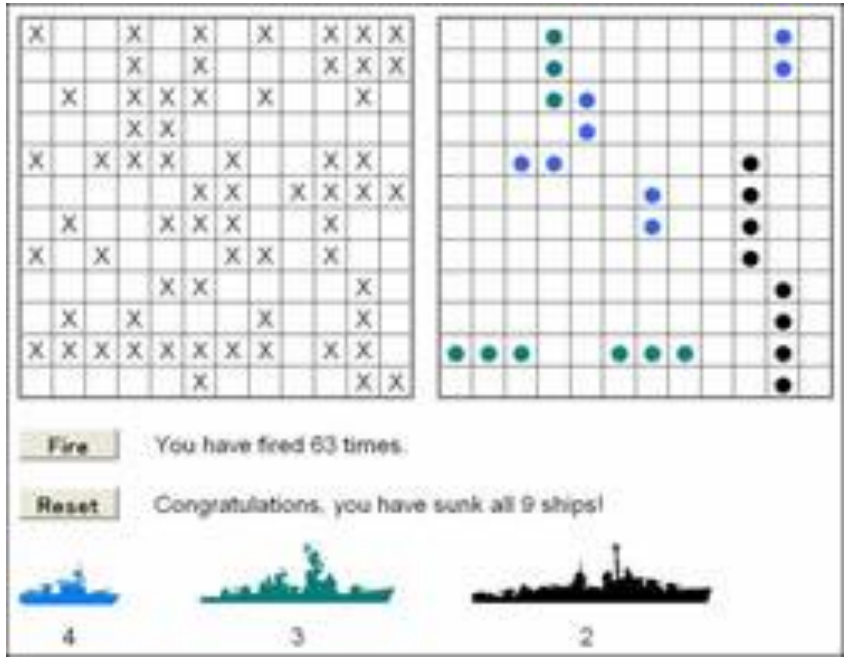
Crystallisation is *stochastic*



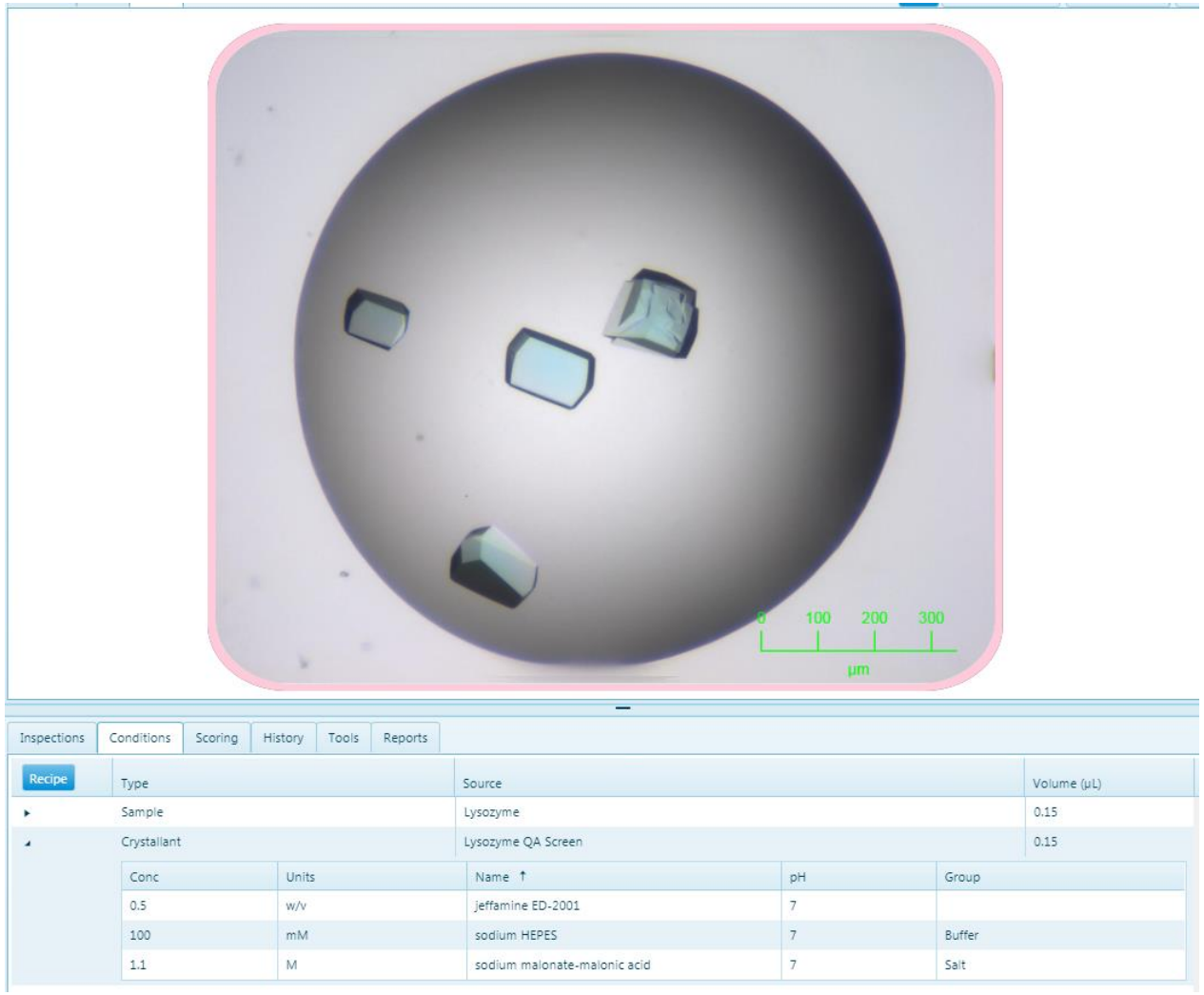
4th inspection of drop A1:1 from lysozyme QC plates with 0.3 µL drops set up after Jan 2013

Playing Battleships – sinking the ship

- How big is the ship?
- How close are the gridpoints?
- How much can we trust the hit?

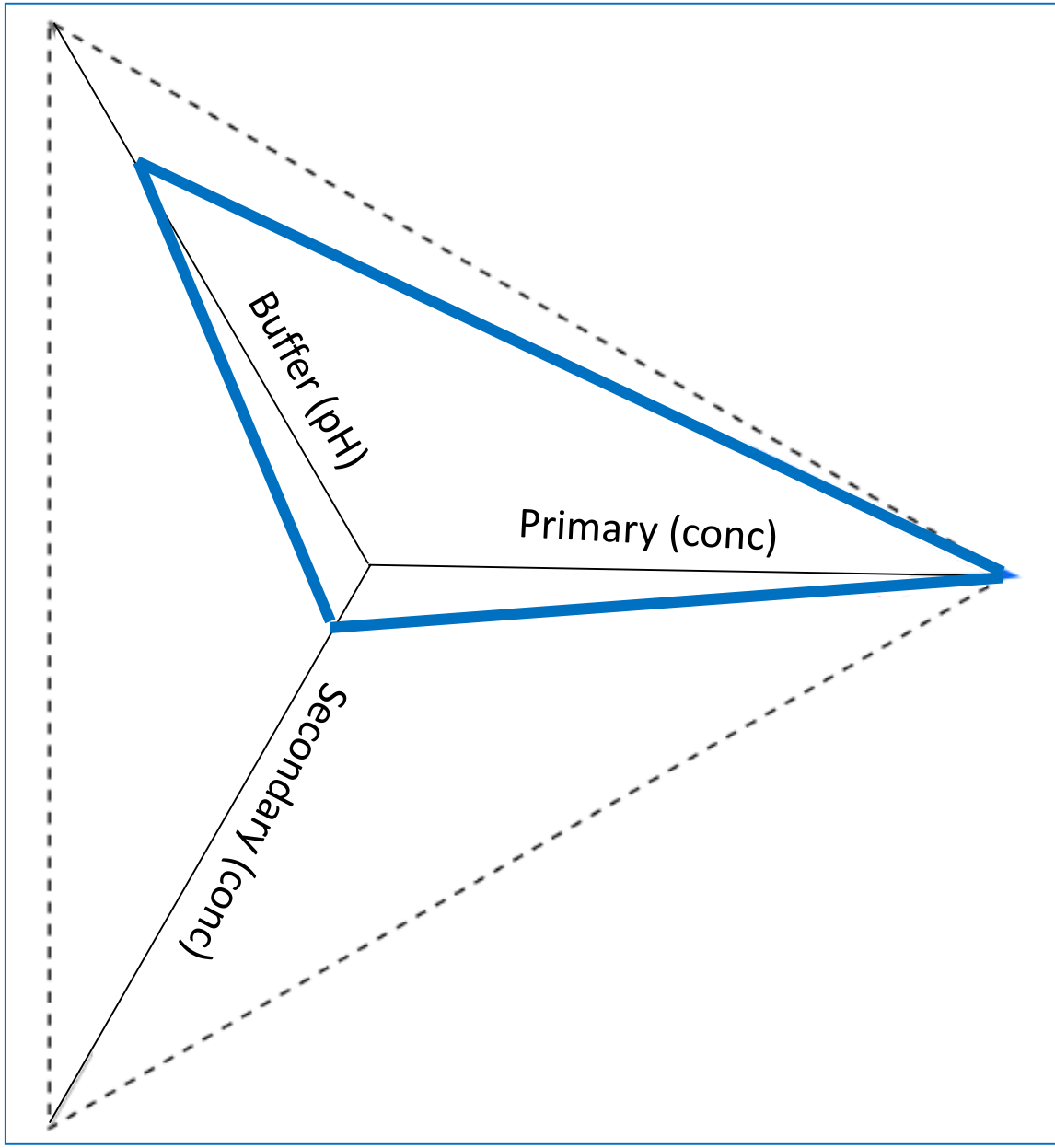


A crystallisation experiment

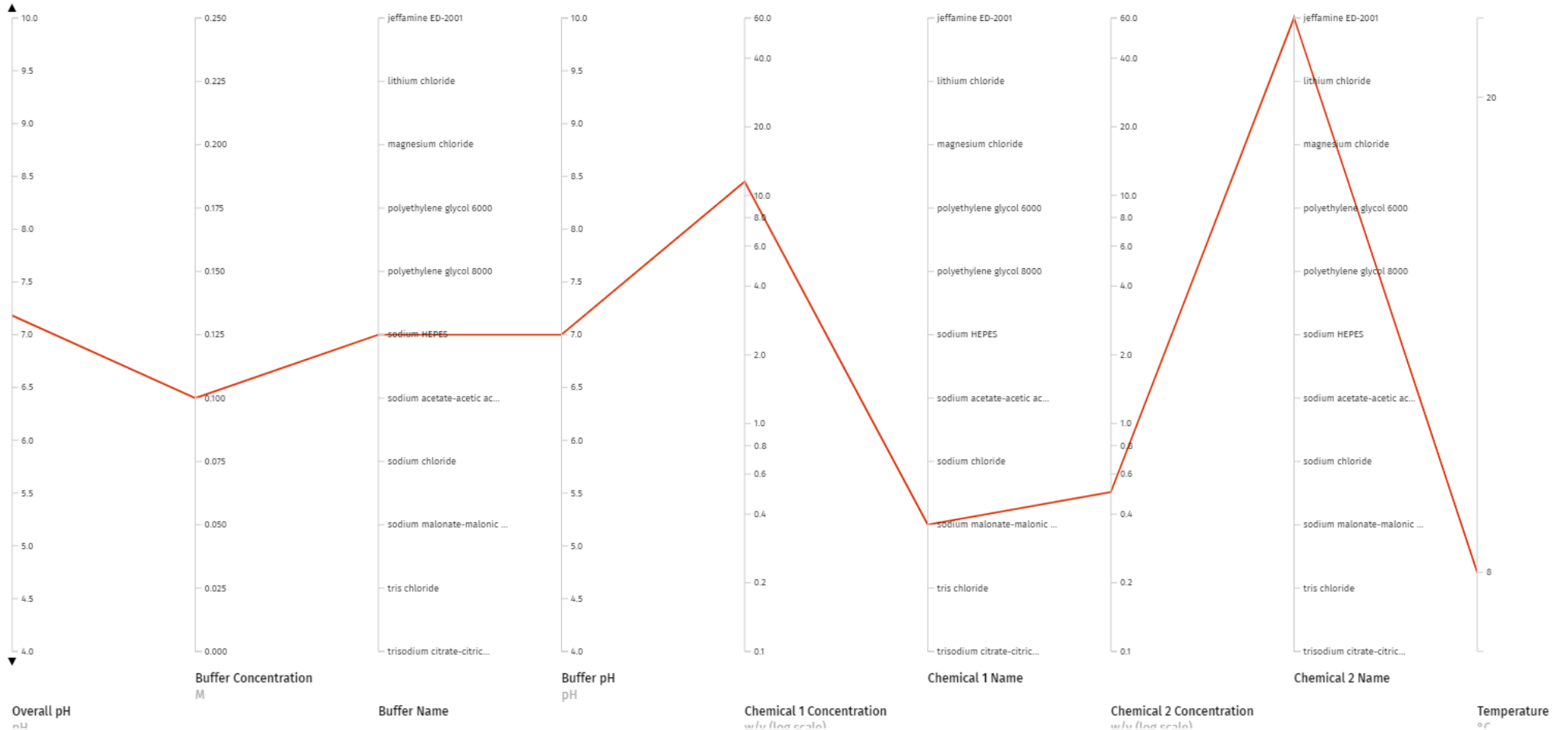


Spiderview

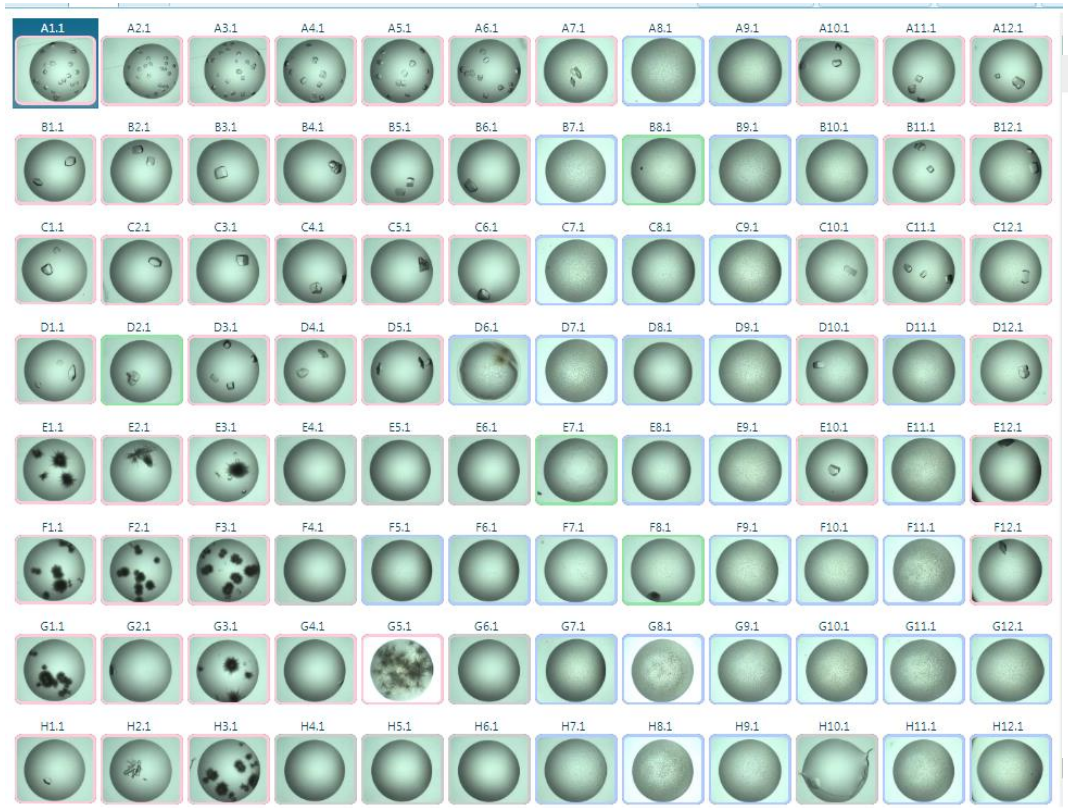
- Each axis is a chemical
- Distance along axis is concentration or pH
- Positive X-axis is primary factor
- Clockwise for concentration factors
- Anticlockwise for buffering factors



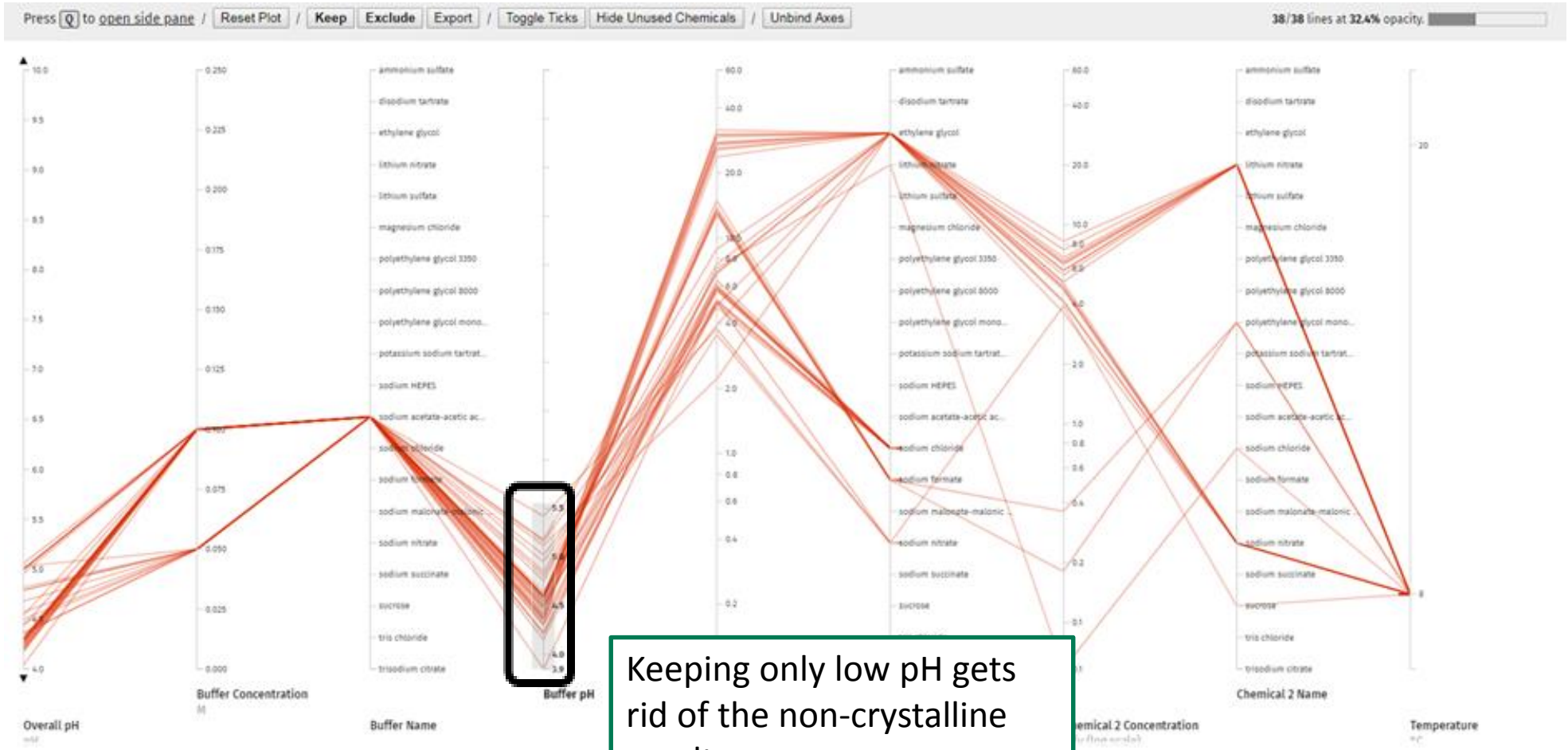
Parallel axis plots



Many crystallisation experiments



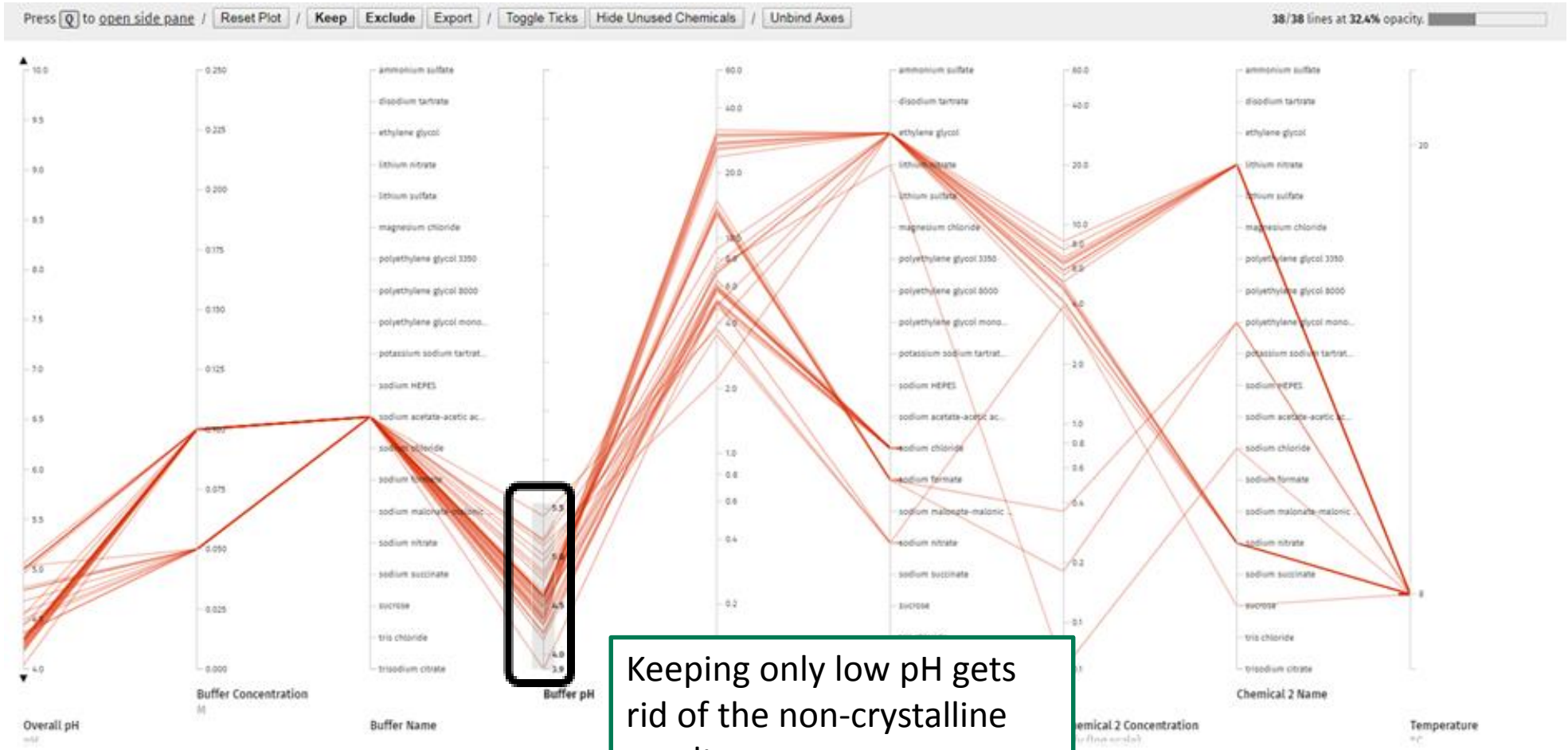
Hypothesizing



Keeping only low pH gets rid of the non-crystalline results

Hypothesizing

Hypothesizing



Keeping only low pH gets rid of the non-crystalline results

Or maybe?

Removing the PEG conditions gets rid of MOST of the non-crystalline results, and keeps more crystal results.

Or maybe?

crystal results.

Or maybe?

Removing the PEG conditions gets rid of MOST of the non-crystalline results, and keeps more crystal results.

What needs to be in place?

- Description of the experiment
  - Complete
  - Consistent
- Description of the results
  - Complete
  - Consistent

Complete-ish: Crystallisation cocktail~~, protein formulation, PTM, geometry, cryo, freezing method etc~~

Consistent and unambiguous naming

Complete-ish: Clear, Precipitate, Crystal, Other (OR not AND~~, skin, oil etc~~)

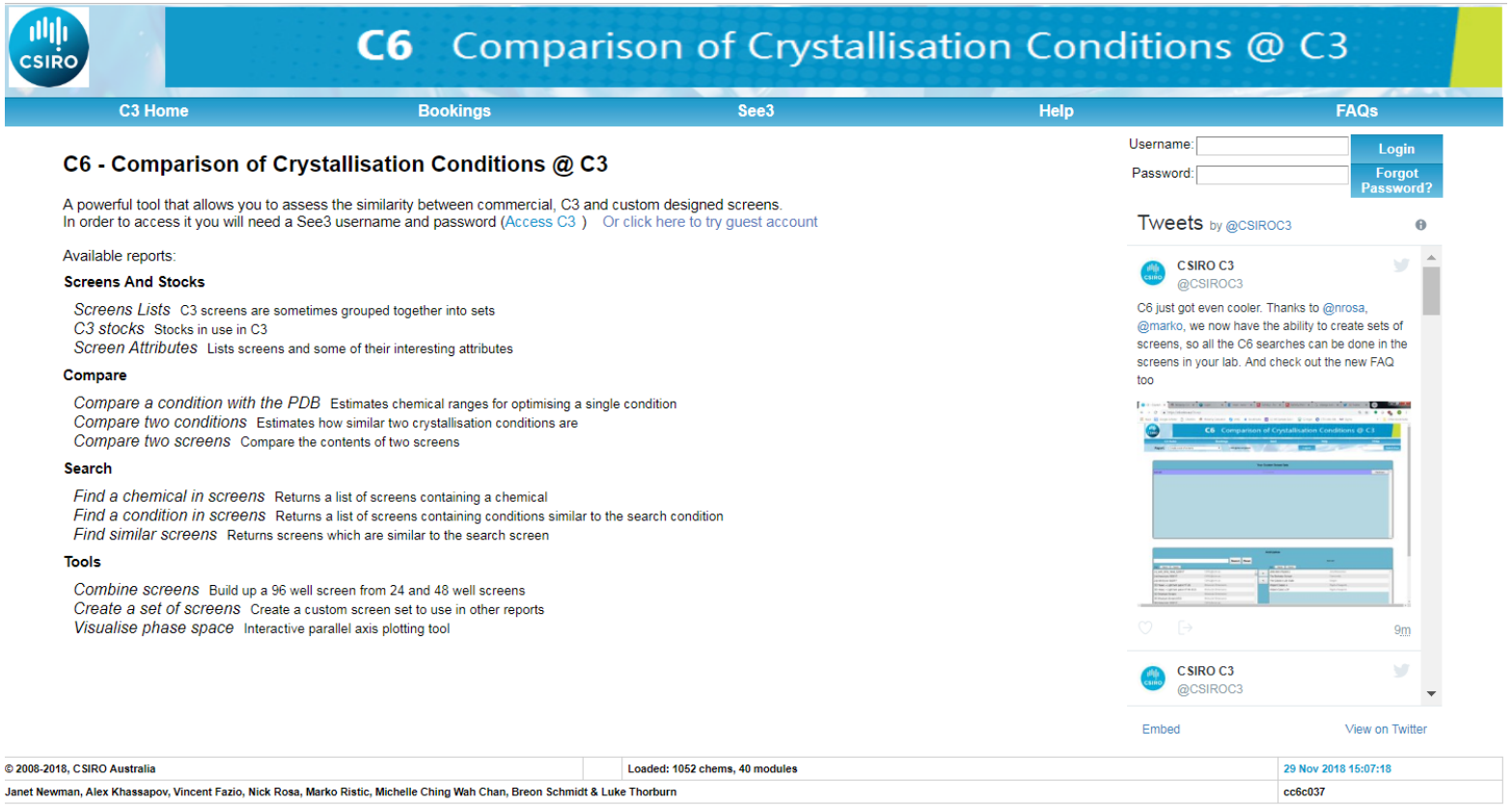
MARCO

Consistent and unambiguous naming

- Consider two conditions:
  - 100mM citric acid pH5
  - 200mM MgSO4•7H2O
  - 25w/v MPEG 5000
  - 25% PEG MME 5K
  - 0.1M sodium citrate pH 5.0
  - 0.2M magnesium sulfate
- **Data standards**
  - **Naming**
  - **Spelling**
  - **Units**
  - **Ordering**

Given clean data

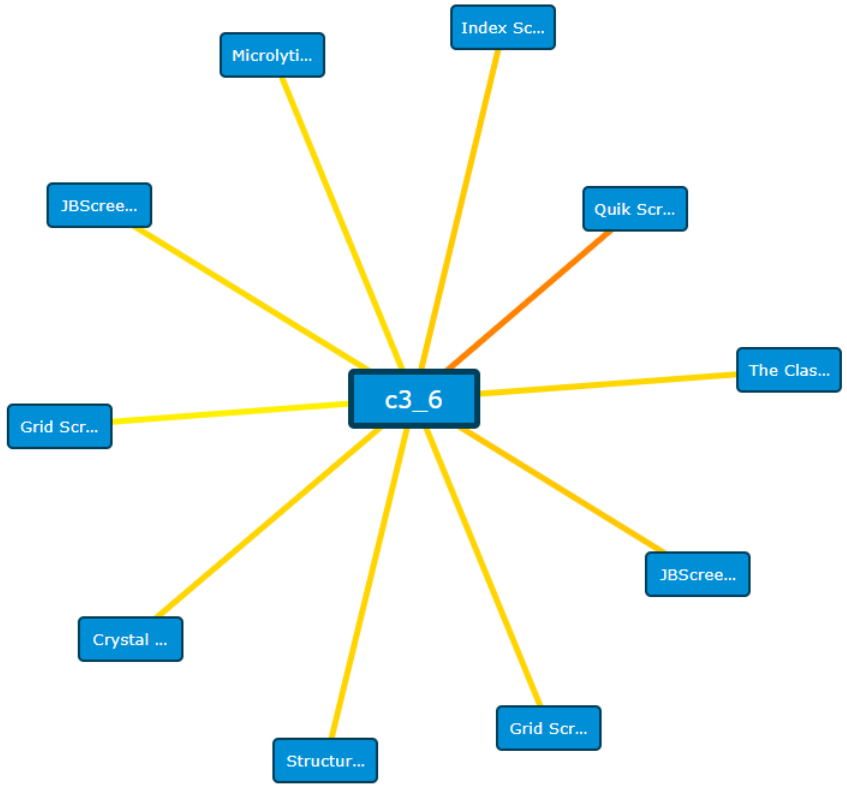
- Search
- Sort
- Compare



C6 (Comparison of Crystallisation Conditions @ C3)

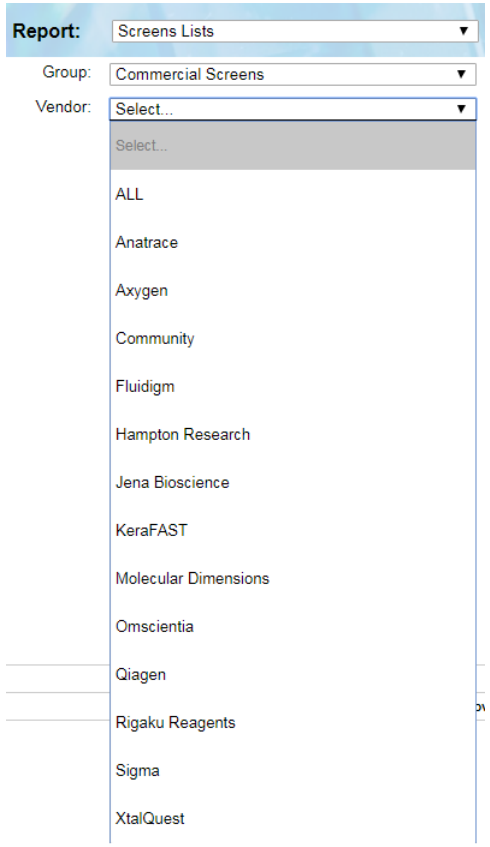
Four categories of reports

- Lists
- Compare
- Search
- Tools



Lists

- Screens
- Stocks
- Attributes



Group: Commercial Screens

Vendor: Anatrace

Screens (COMMERCIAL Anatrace) Group Stats			Export: CSV-ROW CSV-CELL TEXT XML MMCIF RECIPE	
Screen Name(a↑)	Screen Owner	Create date	Barcodes	Comments
Analytic Crystallizer	Anatrace	-		
Microlytic MCSG1	Anatrace	-		
Microlytic MCSG2	Anatrace	-		
Microlytic MCSG3	Anatrace	-		
Microlytic MCSG4	Anatrace	-		
Microlytic PurePegs	Anatrace	-		
Microlytic SuperCombi	Anatrace	-		
Microlytic Top96	Anatrace	-		
PurePEG48	Anatrace	-		
XZ Screen	Anatrace	-		

1-propanol	propan(1)ol (80% v/v)	80	v/v
2,2,2-trifluoroethanol	2,2,2-trifluoroethanol (40% v/v)	40	v/v
2,5-hexanediol	2,5-hexanediol (50% v/v)	50	v/v
2-ethoxyethanol	ethoxyethanol (100% w/v)	100	v/v
2-methyl-2,4-pentanediol	2-methyl-2,4-pentanediol (100% v/v)	100	v/v
2-propanol	propan(2)ol (80% v/v)	80	v/v
6-aminocaproic acid	6-aminocaproic acid (30% w/v)	30	w/v
acetate-ADA-bicine	acetate-ADA-bicine pH 4.0 (100% v/v)	100	v/v
acetate-ADA-bicine	acetate-ADA-bicine pH 9.0 (100% v/v)	100	v/v
acetic acid	acetic acid (1M)	1	M
ADA	ADA pH 5.5 (0.5M)	0.5	M
ADA	ADA pH 7.5 (0.5M)	0.5	M
alanine	L-alanine (1M)	1	M
ammonium acetate	ammonium acetate (5M)	5	M
ammonium chloride	ammonium chloride (1M)	1	M
ammonium chloride	ammonium chloride (3M)	3	M
ammonium chloride	ammonium chloride (5M)	5	M
ammonium dihydrogen phosphate	ammonium dihydrogen phosphate (2.5M)	2.5	M
ammonium fluoride	ammonium fluoride (1M)	1	M

Same screen, two ways

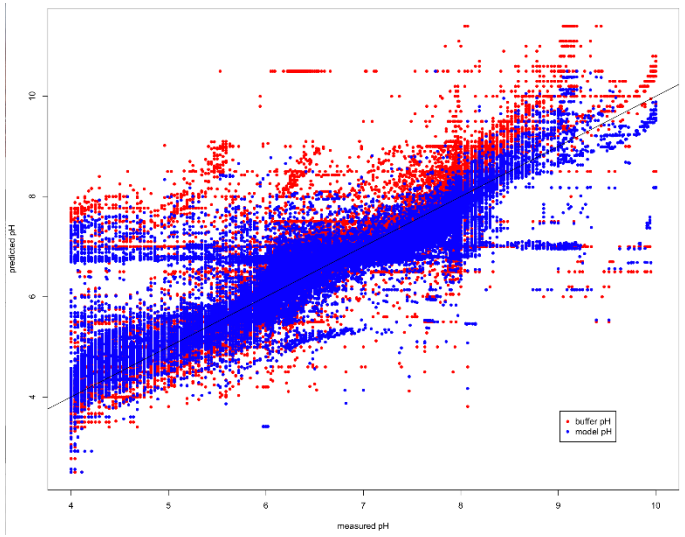
Screen (click for **stats**)	Shotgun
Export as	CSV-ROW CSV-CELL TEXT XML MMCIF RECIPE
Owner	c3@csiro.au
Create date	2018-10-26
Description	This is a collection of the 96 most successful commercial conditions as determined by mining the PDB
Barcodes	Z0002230, Z0002229, Z0002218, Z0002217
Well count	96
Number of distinct chemicals	50
Internal diversity (mean dist)	0.912

NB: Buffers are shown in *italic* font			
Well(a↑)	**Concentration Units Chemical, pH**	**pH**	**Recipe**
A1 (1)	0.2 M magnesium chloride; 30 w/v polyethylene glycol 4000; *0.1 M tris chloride, pH=8.5;*	8.1	> A1 200.0 ul of: magnesium chloride (1M) 600.0 ul of: poly(ethylene glycol) 4000 (50% w/v) 21.2 ul of: tris chloride pH 7.0 (1M) 78.8 ul of: tris chloride pH 9.0 (1M) 100.0 ul of: H2O
A2 (2)	2 M ammonium sulfate;	5.7	> A2 571.4 ul of: ammonium sulfate (3.5M) 428.6 ul of: H2O
A3 (3)	20 w/v polyethylene glycol 3350; 0.2 M sodium acetate;	7.3	> A3 400.0 ul of: poly(ethylene glycol) 3350 (50% w/v) 200.0 ul of: sodium acetate (1M) 400.0 ul of: H2O
A4 (4)	2 M ammonium sulfate; *0.1 M tris chloride, pH=8.5;*	7.9	> A4 21.2 ul of: tris chloride pH 7.0 (1M) 571.4 ul of: ammonium sulfate (3.5M) 78.8 ul of: tris chloride pH 9.0 (1M) 328.6 ul of: H2O
A5 (5)	20 w/v polyethylene glycol 3350; 0.2 M trisodium citrate;	8.1	> A5 400.0 ul of: poly(ethylene glycol) 3350 (50% w/v) 125.0 ul of: (tri)sodium citrate (1.6M) 475.0 ul of: H2O

Number of conditions	Name of chemical	Units	Concentration					pH				
			Mean	Mode	Min	Max	StdDev	Mean	Mode	Min	Max	StdDev
42	polyethylene glycol 3350	w/v	22.3	20	20	25	2.5					
22	ammonium sulfate	M	1	0.2	0.2	2	0.749					
16	sodium HEPES	M	0.1	0.1	0.1	0.1	0	7.5	7.5	7.5	7.5	0
14	bis-tris chloride	M	0.1	0.1	0.1	0.1	0	6	5.5	5.5	6.5	0.5
10	sodium acetate	M	0.19	0.2	0.1	0.2	0.064					
10	polyethylene glycol 8000	w/v	21.8	20	10	30	7.5					
8	magnesium chloride	M	0.2	0.2	0.2	0.2	0					
8	polyethylene glycol 4000	w/v	25.4	30	8	30	9.4					
8	tris chloride	M	0.1	0.1	0.1	0.1	0	8.5	8.5	8.5	8.5	0
6	lithium sulfate	M	0.55	0.2	0.2	1.5	0.64					
6	sodium cacodylate	M	0.1	0.1	0.1	0.1	0	6.5	6.5	6.5	6.5	0
6	sodium acetate-acetic acid	M	0.1	0.1	0.1	0.1	0	4.6	4.6	4.6	4.6	0
5	sodium MES	M	0.1	0.1	0.1	0.1	0	6.4	6.5	6	6.5	0.292
4	trisodium citrate	M	1.1	0.2	0.2	1.6	0.536					

Attributes

- By screen
- pH, pH range
- % PEG
- % Salt
- Internal diversity



C3 Home | Bookings | See3 | Help

Report: Screen Attributes

CSHL@csiro.au

Group: My Screen Sets

Screen Set: test set

Table of Screen attributes Right click and 'Save link as...' to download a csv file

Internal diversity estimates the difference between the conditions within a screen: 0 if all conditions are identical, 1 if all conditions are completely different.
% PEG gives percentage of conditions where the primary factor is a polyethylene glycol
% Salt gives percentage of conditions where the primary factor is a salt

Screen Name	Owner Name	Internal Diversity	% PEG	% Salt	Av. pH	Wells within Av. pH +/- 1	Distinct chemicals
The Classics Lite Suite	Qiagen	0.95	34	35	6.8	37	55
Wizard Classic 4	Rigaku Reagents	0.961	21	54	6.7	18	52
JBScreen Wizard 2	Jena Bioscience	0.935	38	40	6.8	24	30
Wizard Cubic LCP	Rigaku Reagents	0.894	26	0	6.6	52	17
The Berkeley Screen	Community	0.928	14	69	6.7	42	40

Compare

- condition to the PDB
- 2 conditions
- 2 screens

$$D_{ij} = 1 \text{ (no species in common)}$$

$$D_{ij} = \frac{1}{(T+3)}\left(\left(\left(\sum_{t=1}^{T}\frac{|[s_{ti}]-[s_{tj}]|}{\max[s_t]}\right)+\left(\frac{|E(\mathrm{pH}_i)-E(\mathrm{pH}_j)|}{\mathrm{gul(pH)}-\mathrm{gll(pH)}}\right)\right)+\min\left(1,\left[\left(\frac{|[\mathrm{ion_i}]-[\mathrm{ion}_j]|}{(\max[\mathrm{ion}_i]+\max[\mathrm{ion}_j])}\right)+0.3\right]\right)+\min\left(1,\left[\left(\frac{|[\mathrm{PEG}_i]-[\mathrm{PEG}_j]|}{(\max[\mathrm{PEG}_i]+\max[\mathrm{PEG}_j])}\right)+0.2\right]\right)\right)$$

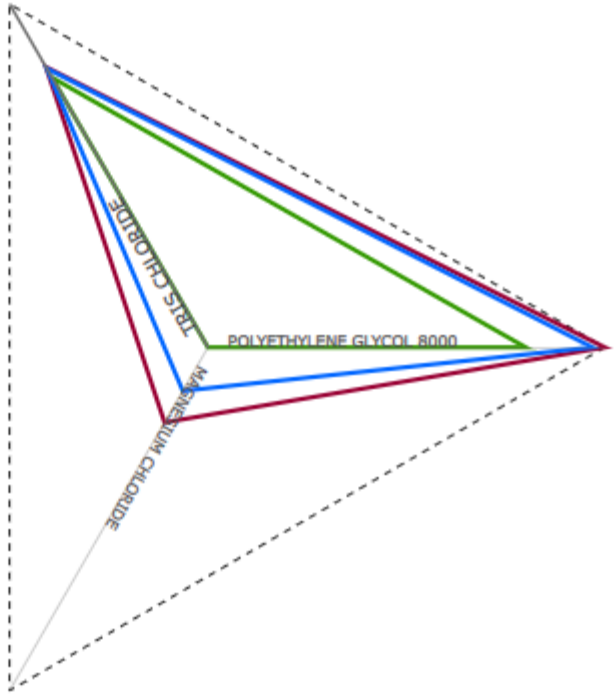
$$\mathrm{score}_{s,t} = \frac{1}{2}\left(\frac{1}{\mathrm{cond}_s}\sum_{i=1}^{\mathrm{cond}_s}\min_{j\in(1,\,\mathrm{cond}_t)}(D(c_{si},c_{tj}))+\frac{1}{\mathrm{cond}_t}\sum_{i=1}^{\mathrm{cond}_t}\min_{j\in(1,\,\mathrm{cond}_s)}(D(c_{ti},c_{sj}))\right)$$

A condition to the PDB

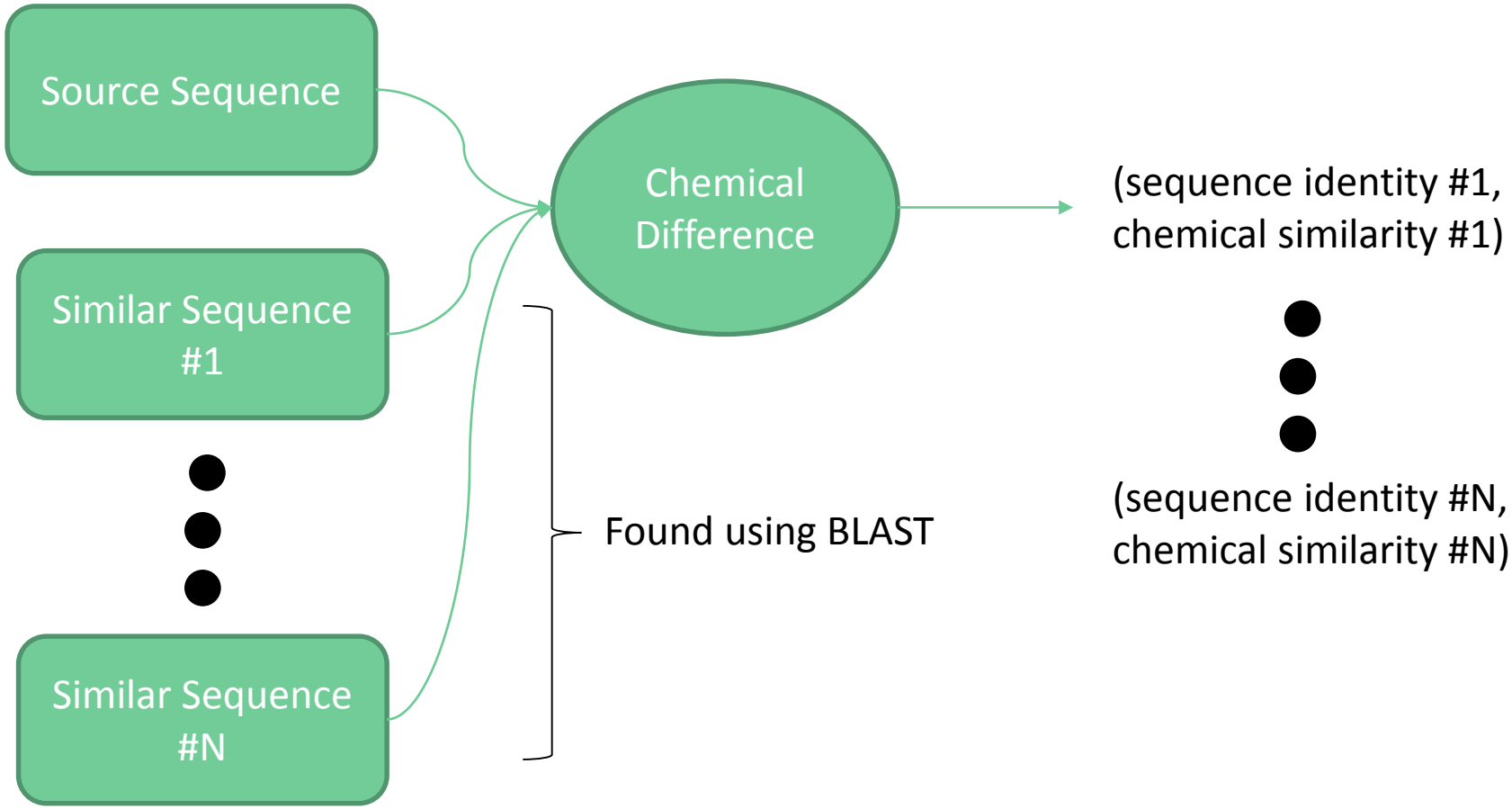
Lysozyme QA Screen (Well: A7)				
Chemical	**Concentration**	**Units**	**pH**	**Orig. Name**
magnesium chloride	0.2	M	-	magnesium chloride
tris chloride	0.1	M	8.5	tris chloride
polyethylene glycol 8000	20	w/v	-	polyethylene glycol 8000

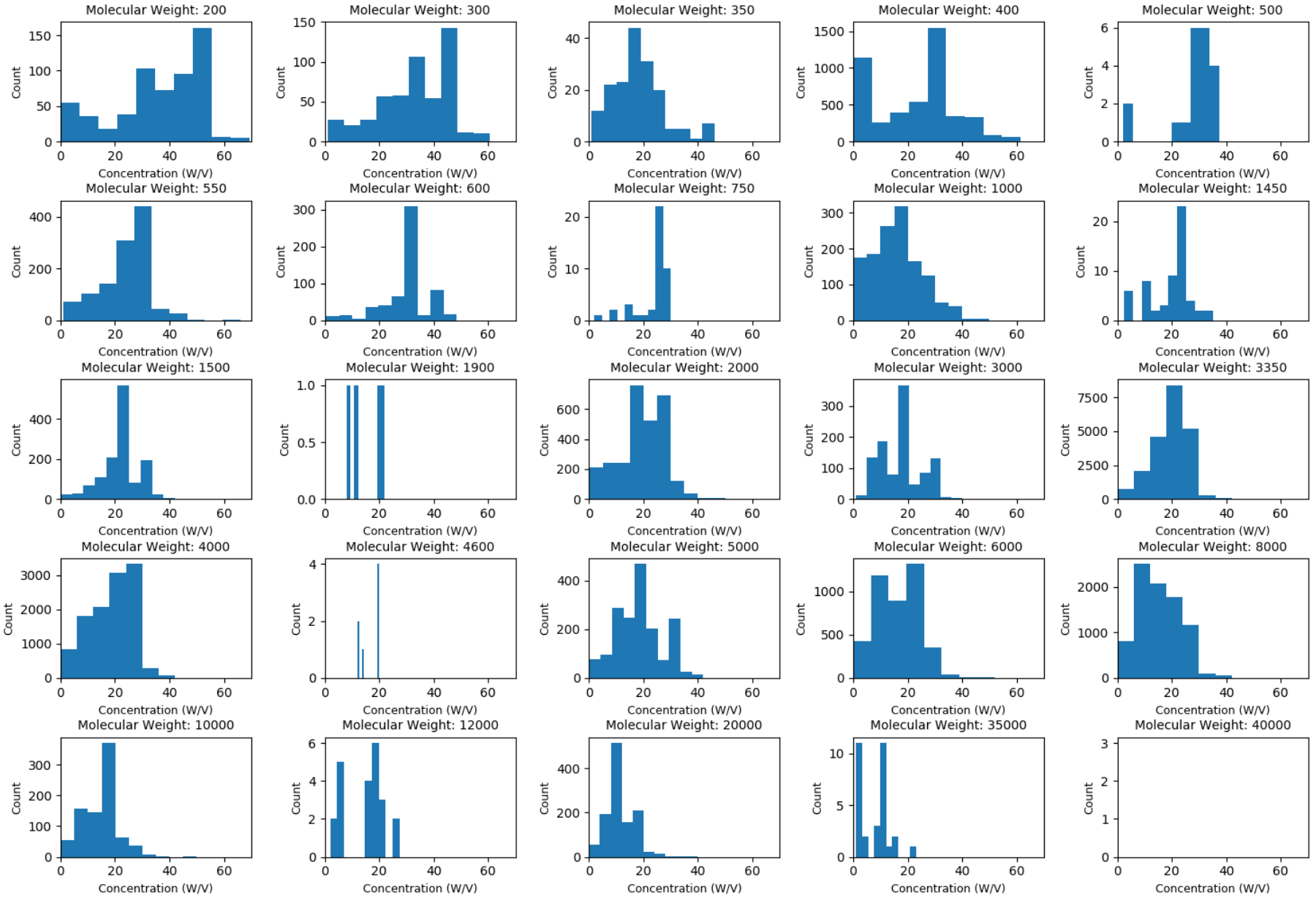
PDB Comparison table				
Chemical	**Value**	**Lower Value †**	**Upper Value †**	**Variable**
polyethylene glycol 8000	20	8	23.7	w/v
magnesium chloride	0.2	0	0.331	M
tris chloride	8.5	7.4	8.7	pH

- these values are extracted from the crystallisation conditions
eposited in the PDB (Protein Data Bank) www.wwpdb.org

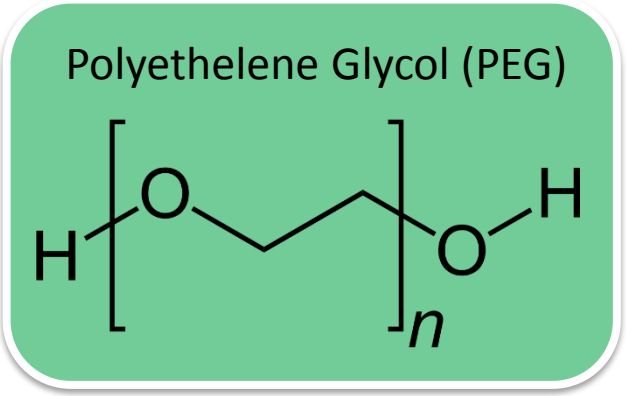


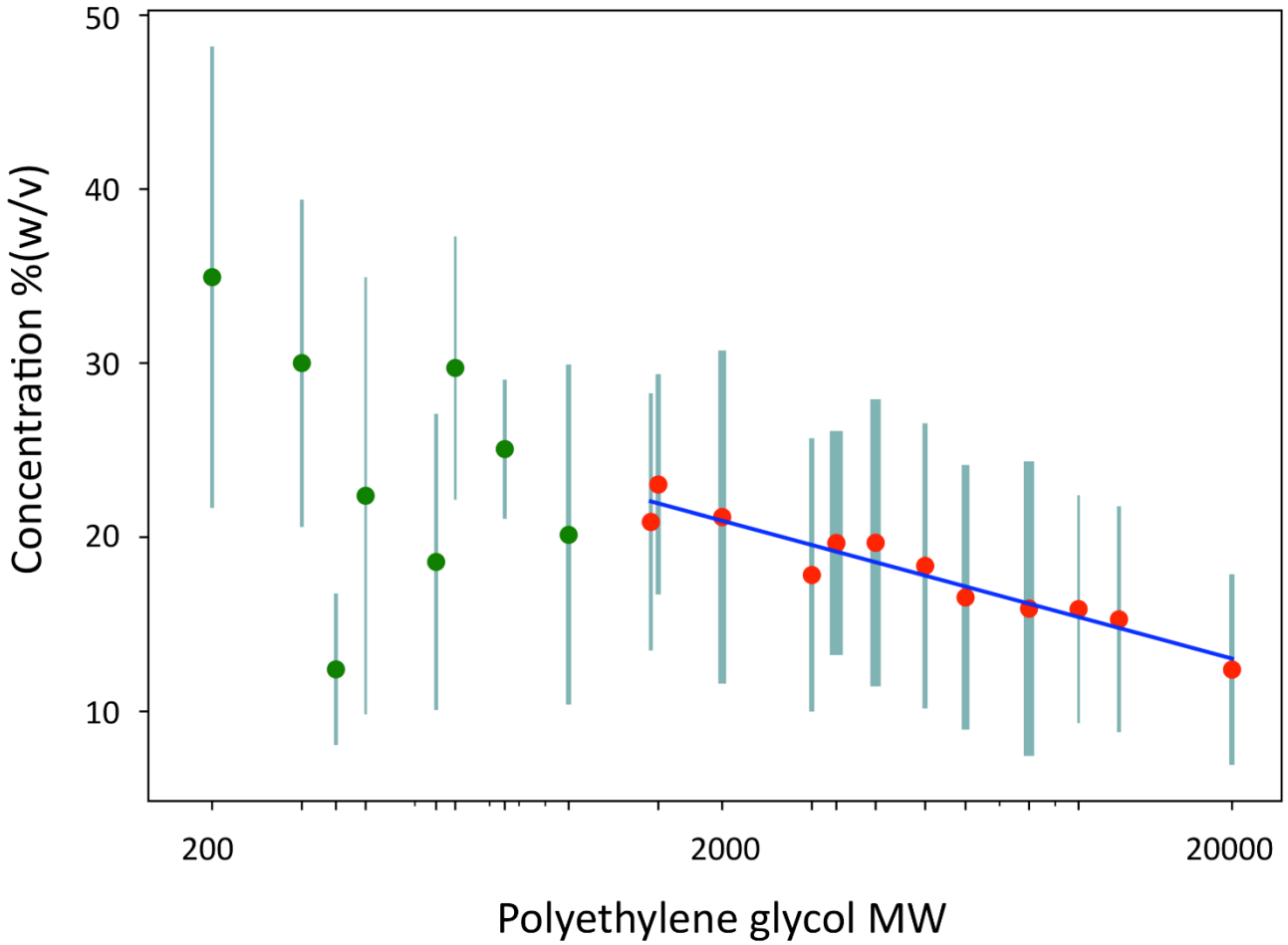
BLASTing


Source Sequence
Similar Sequence #1
Similar Sequence #N
Chemical Difference
Found using BLAST
(sequence identity #1, chemical similarity #1)
(sequence identity #N, chemical similarity #N)

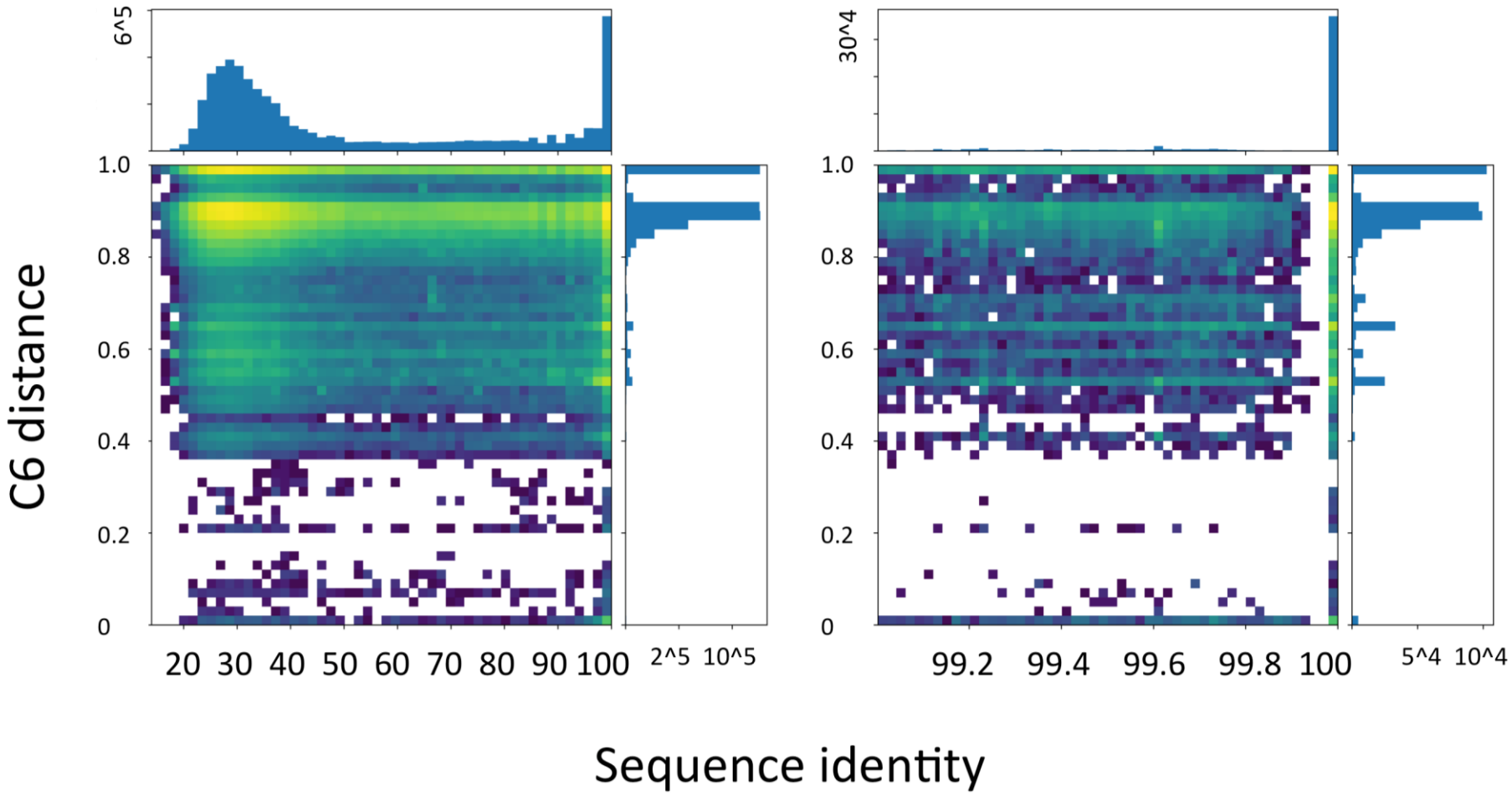

Molecular Weight: 200
Molecular Weight: 300
Molecular Weight: 350
Molecular Weight: 400
Molecular Weight: 500
Molecular Weight: 550
Molecular Weight: 600
Molecular Weight: 750
Molecular Weight: 1000
Molecular Weight: 1450
Molecular Weight: 1500
Molecular Weight: 1900
Molecular Weight: 2000
Molecular Weight: 3000
Molecular Weight: 3350
Molecular Weight: 4000
Molecular Weight: 4600
Molecular Weight: 5000
Molecular Weight: 6000
Molecular Weight: 8000
Molecular Weight: 10000
Molecular Weight: 12000
Molecular Weight: 20000
Molecular Weight: 35000
Molecular Weight: 40000
Count
Concentration (W/V)

BLASTing: PEG Eq

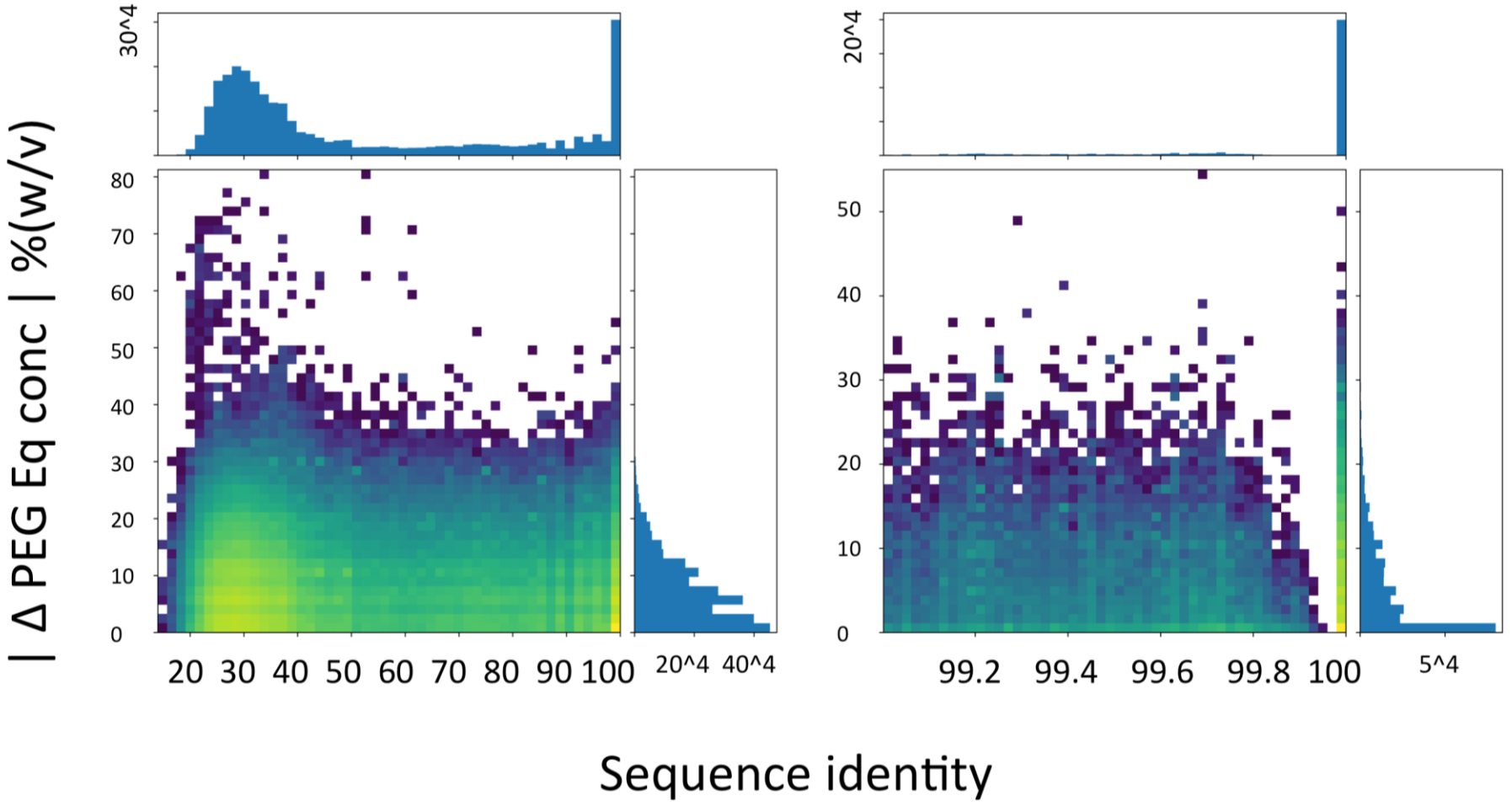




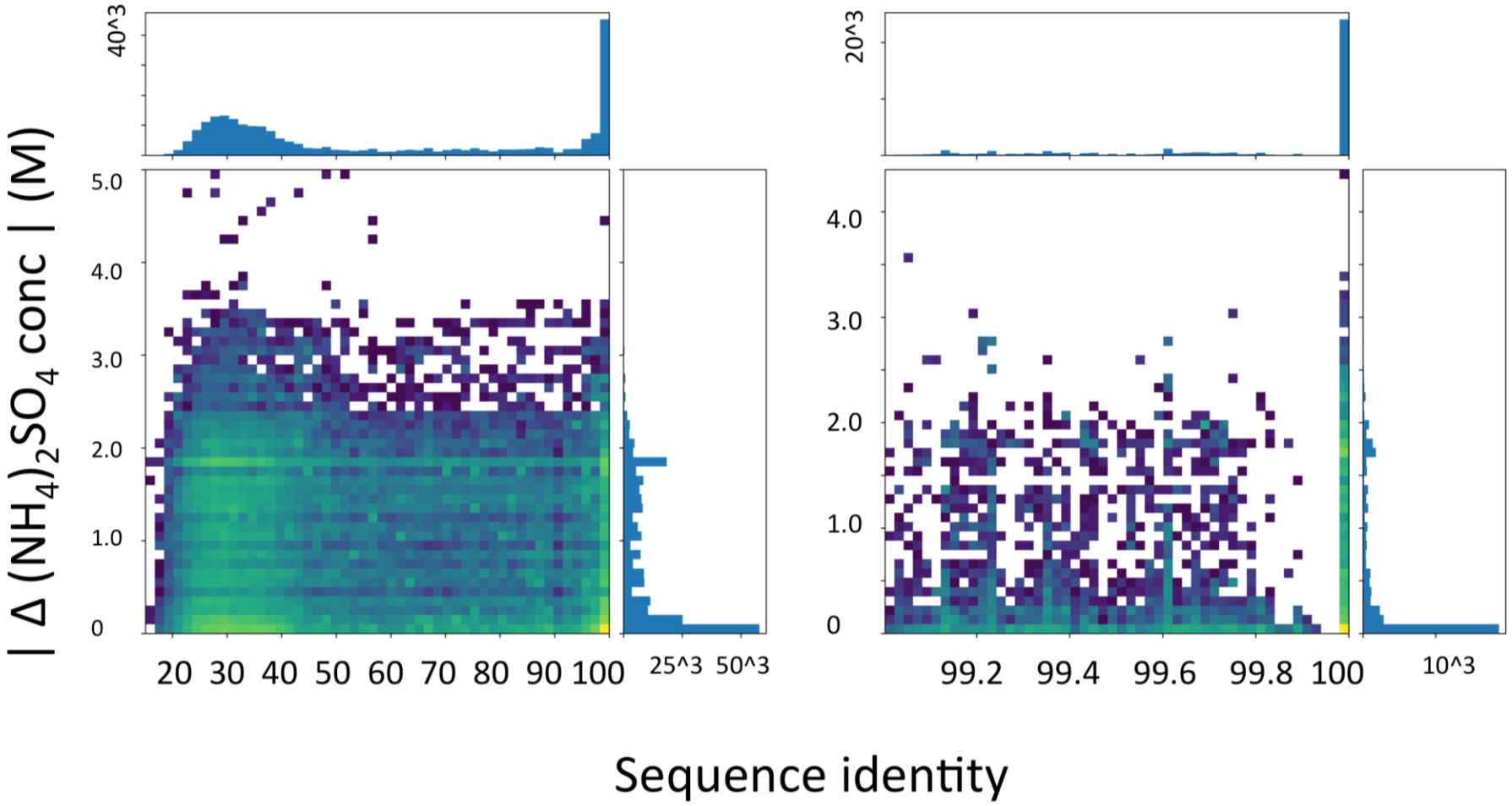
BLASTing: C6 (5.5 million pairs)



BLASTing: PEG Eq (2.3 million pairs)



BLASTing: $(NH_4)_2SO_4$ (0.25 million pairs)



Soooo…

- No correlation between sequence identity and chemical score (C6 or PEG Eq).
- Attempting to use sequence similarity to determine crystallisation conditions entry point does not appear likely to be a useful endeavour
- PEG Eq could provide a new way of comparing crystallisation conditions which is useful for optimisation

Acknowledgements

- Alex Khassapov
- Nick Rosa
- Marko Ristic
- Luke Thorburn
- Breon Schmidt
- Michelle Chan
- Vincent Fazio
- Brian Lawson
- Gabi Abrahams