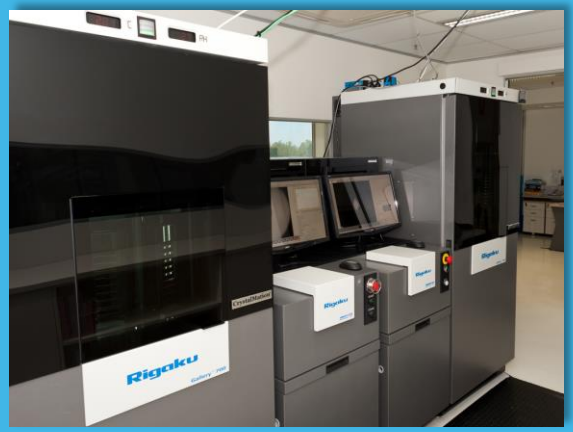


Crystallisation Conditions

The good, the bad and those with special challenges

Recent Advances in Macromolecular Crystallisation | Janet Newman

09 September, 2013

Setting the scene

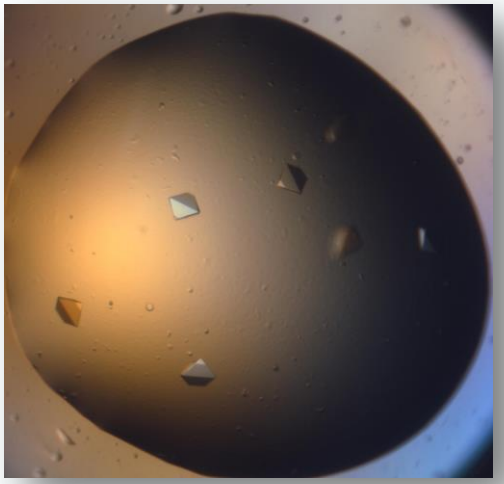
So how many crystallisation conditions are there?

Possible (20 ^9?)

Commercially available (16 ^3)

Set up each year (10-100 ^6 ???)

Ones that we know work? (50 ^3)



Reality Check: ***Crystals aren't interesting***

What is the most effective way of producing the biochemical / chemical knowledge that comes from X-ray diffraction analyses of proteins?

Some definitions

Chemical (noun)

- The name of a chemical. (sodium chloride *but not* sodium phosphate)

Factor (noun)

- A chemical with an associated concentration, unit [and pH]

Condition (noun)

- A collection of factors found in one solution

Kit (noun)

- A collection of conditions

Screen (verb)

Scry (verb)

Strategy (noun, rarely used)

Screening

Goal:

To identify those factors which are positively and negatively correlated with crystal formation

Approach:

Set up a number of balanced conditions

Requires:

Knowing what an ‘almost’ crystal looks like.

Scrying

Goal:

To identify a well diffracting crystal from initial experiments

Approach:

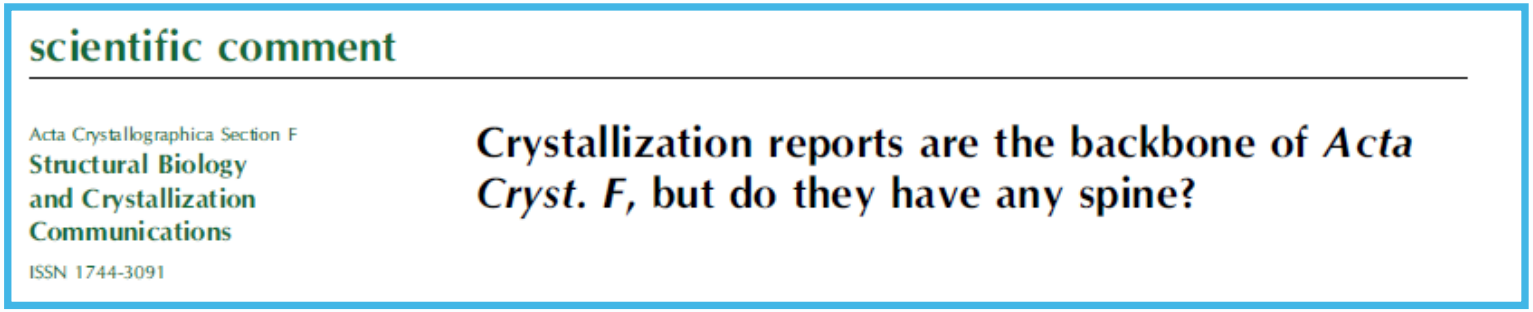
Set up lots of conditions, in replication (matrix seeding?)

Requires:

Luck

Scrying - does it work?

Look at the Literature



75% of systems required (crystal) optimisation

We all secretly scry -

How many times do we use the balanced, rigorous DOE-based screen which tests the top n factors at three levels in a pairwise manner?

(that screen doesn't exist commercially)

Are all conditions equally likely?

No

Which are the best conditions?

In a world of rational, knowledgeable crystallisers, (*homo crystallus)*

- Ask the vendors which are the most popular kits

(and are crystallisers rational?)

Even so, (hopefully) the more successful kits continue to be bought

- Are the kits that are bought a good representation of the conditions that are set up?
- How long does it take to trial out a new kit? (who does this testing?)

Which are the best conditions?

Look at all the conditions that have produced well diffracting crystals.

Look in data repositories for any data at all

BMCD, PSI, **PDB**

→Cleanup required before analysis

Cleanup spelling, format

- Consider a condition:
  - *100mM citric acid pH5*
  - *200mM MgSO4•7H2O*
  - *25w/v MPEG 5000*
- Consider another
  - *25% PEG MME 5K*
  - *0.1M sodium citrate pH 5.0*
  - *0.2M magnesium sulfate*

AMMONIUM SULFATE
AMMONIUM SULPHATE
AMMONIUM 2(S04)
AMMONIUMSULFATE
AMMONIUMSULFAT
AMMONIUMSULFA
AMMONIUM SLUFATE
AMONNIUM SULFATE
AMMONIUM SULF
AMMMONIUMSULFAT
AMMONIUM SULFACE
AMMONIUM SUPLHATE
AMM.SULFATE
AMM.SUL
AMM.SULPH.
AMMON. SULPH.
(NH4)2(SO4)
(NH4)2(S04)
(NH4) 2SO4
NH4-SO4
NH4SO4
(NH)SO4
NH4/SO4
SULFATE(NH4)
NH2 SULFATE
AMSO4
AMMSO4
AS
AMS
A2S
A/S
A.S.

Classes, rules

Sodium citrate pH 5.6

Citric acid pH 5.6

} Neither exist

Tri-sodium citrate (pH ≈ 8.1)

Citric acid (pH ≈ 1.5)

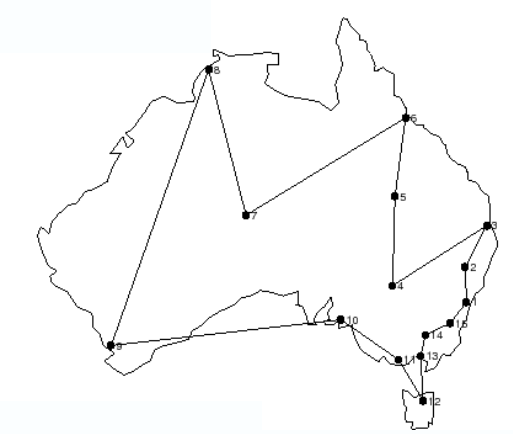
Tri-sodium citrate – citric acid pH 5.6

(pK_a citrate = 3.1, 4.7, 5.4)

Citrate buffer class

Citrate non-buffer class

Maps and measurement



Some concepts

Two conditions may be identical, or not

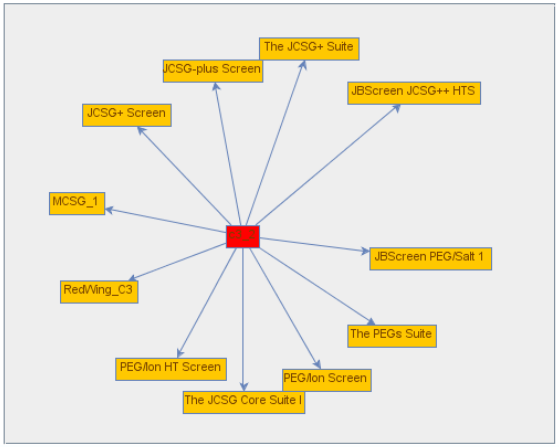
(sometimes hard to tell – or ambiguous – (e.g. buffers)

Two conditions may be similar

Given three conditions, two may be more similar to each other than to the other

Need a *quantifiable* measurement of distance

(http: c6.csiro.au)



Crystallisation space

Different sets of conditions:

Set 1. Those conditions which are set up in experiments (?)

Set 1a. Those conditions which are commercially available (15,810)

Set 2. Those conditions which are successful (?)

Set 2a. Those conditions which are found in the PDB (***53,937***)

Assumptions: yep.

Does input space look like *output space*?

Factor Average

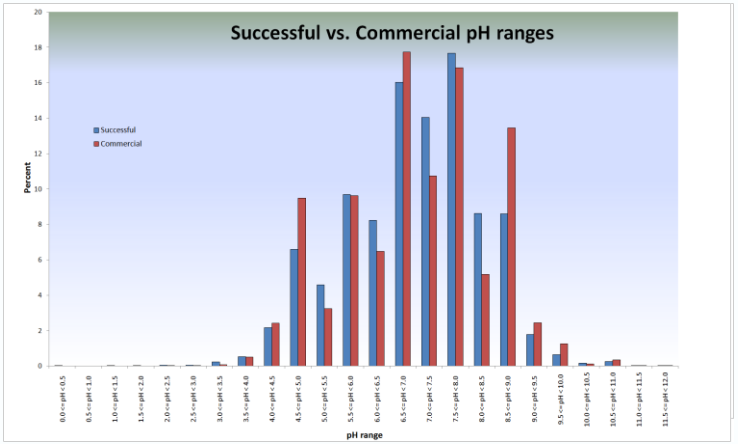
- 2.53 — ***2.81***

Distinct chemicals

- 793 (505 not SB, SBB) — ***596***

Top chemicals

- sodium acetate — ***ammonium sulfate***
- ammonium sulfate — ***PEG 3350***
- PEG 3350 — ***PEG 4000***
- HEPES — ***sodium chloride***
- PEG 4000 — ***tris buffer class***



Reality check

Largest cluster of identical sequences in PDB

320 members

162 have a condition (REMARK280)

- ACETATE NON-BUFFER CLASS occurs 40.74%
- ACETATE BUFFER CLASS occurs 46.30%
- SODIUM CHLORIDE occurs 82.72%

Mapping PDB to commercial

14% PDB conditions match a condition from commercial space (7,391 of 53,937 PDB conditions)

(1,016 of 7,754 distinct commercial conditions - 13%)

34% are within 0.1 similarity units of a commercial condition

(18,188 of 53,937 PDB conditions)

(1,712 of 7,754 distinct commercial conditions – 22%)

Mapping NRPDB to commercial

14% PDB conditions match a condition from commercial space (1380 of 10111 NRPDB conditions)

(493 of 7,754 distinct commercial conditions - 6%)

30% are within 0.1 similarity units of a commercial condition

(3021 of 10111 PDB conditions)

(886 of 7,754 distinct commercial conditions – 11%)

Top 10 list (1)

Top 10 most common commercial conditions in the whole PDB

- 2.0 M ammonium sulfate (134)
- 20.0 w/v polyethylene glycol 3350; 0.2 M citrate non-buffer class (115)
- 20.0 w/v polyethylene glycol 3350; 0.2 M acetate non-buffer class; (99)
- 30.0 w/v polyethylene glycol 4000; 0.1 M tris buffer class, pH 8.5; 0.2 M magnesium chloride (95)
- 2.0 M ammonium sulfate; 0.1 M tris buffer class, pH 8.5 (82)
- 4.0 M sodium formate (82)
- 2.0 w/v polyethylene glycol 400; 2.0 M ammonium sulfate; 0.1 M hepes buffer class, pH 7.5 (76)
- 1.4 M citrate non-buffer class; 0.1 M hepes buffer class, pH 7.5 (67)
- 1.6 M citrate non-buffer class; (65)
- 20.0 w/v polyethylene glycol 3350; 0.2 M calcium chloride (64)

Top 10 list (2)

Top 10 most common commercial conditions in the NRPDB

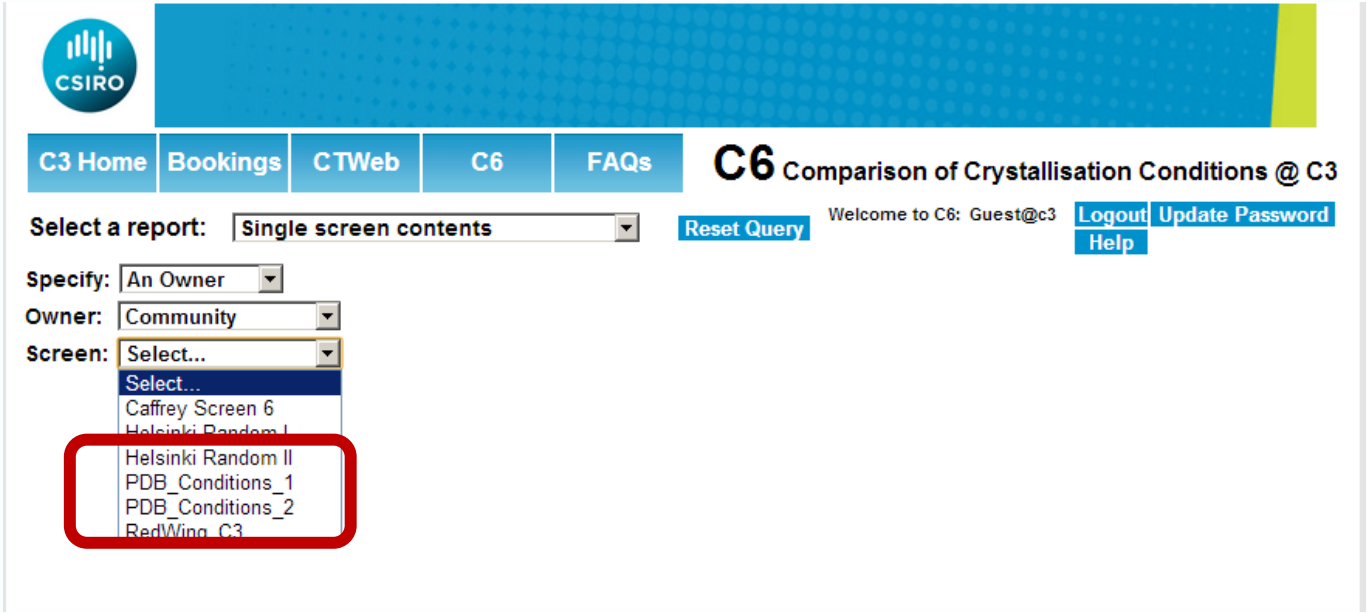
- 20.0 w/v polyethylene glycol 3350; 0.2 M acetate non-buffer class (25)
- 30.0 w/v polyethylene glycol 4000; 0.1 M tris buffer class, pH 8.5; 0.2 M magnesium chloride (21)
- 1.4 M citrate non-buffer class; 0.1 M hepes buffer class, pH 7.5 (20)
- 0.2 M acetate non-buffer class; 0.1 M tris buffer class, pH 8.5; 30.0 w/v polyethylene glycol 4000 (17)
- 20.0 w/v polyethylene glycol 3350; 0.2 M citrate non-buffer class (16)
- 0.2 M acetate non-buffer class; 30.0 w/v polyethylene glycol 4000; 0.1 M citrate buffer class, pH 5.6 (16)
- 2.0 M ammonium sulfate; 0.1 M tris buffer class, pH 8.5 (15)
- 2.0 M ammonium sulfate (14)
- 0.1 M hepes buffer class, pH 7.5; 20.0 w/v polyethylene glycol 4000; 10.0 w/v 2-propanol (13)

Community screens

Screens which are potentially useful

- But not really worth publishing

192 condition of exact matches are available as 2 x 96 condition sets



Summary

New words for new concepts – Screening/scrying

Data Standards are our friends

Quantify distance (http://c6.csiro.au)

PDB *vs*. commercial

Repository for community screens

Acknowledgements

Vincent Fazio

Tom Peat

Thank you