Brilliance, rate, and accessibility – Opportunities for structural biology with an XFEL in your back pocket

Edward Snell,
Hauptman Woodward Medical Research Institute
and SUNY Buffalo

CXFELs provide synchrotron like capabilities

Identity	Beamline	Type	Flux (ph/s)	Beamsize (μm)	Flux density (ph/μm^2/s)	Dose rate kGy/s	Crystal Lifetime (m)	Site lifetime (s)
CXLS			**$5x10^{10}$**	**20**	**$1.6x10^{8}$**	**116**	**4.3**	**17**
ALS	8.3.1	MAD	$1x10^{12}$	60x80	$2.1x10^{8}$	154	3.3	13
SSRL	12.2	MAD	$4x10^{12}$	100x200	$2.0x10^{8}$	96	5.2	21
CHESS	F1	Mono	$3x10^{11}$	100	$3.0x10^{7}$	14	35	138
APS	23-ID-D	MAD	$1x10^{13}$	75x25	$5.3x10^{9}$	3,001	0.16	0.66
NSLS-II	17-ID-2	MAD	$2x10^{12}$	1x1	$1.3x10^{12}$	62,600	$8x10^{-4}$	$5.3x10^{-5}$
Diamond	I04	MAD	$3x10^{11}$	10x5	$6.0x10^{9}$	3,670	0.14	0.54
ESRF	MASSIF_3	Mono	$1.5x10^{13}$	15	$6.7x10^{10}$	30,600	$1.6x10^{-2}$	$1.1x10^{-3}$
PETRA3	P14_MX2	SAD	$5.0x10^{12}$	10x5	$1.0x10^{11}$	101,000	$5.0x10^{-3}$	0.02

The CXFEL performance in terms of flux is comparable with workhorse synchrotron beamlines that perform structural biology studies.

CXFEL near term information from William Graves, other data from James Holton (http://bl831.als.lbl.gov/damage_rates.pdf).

What does this mean from a structural biology perspective?

Brilliance – this enables solution scattering and diffraction resolution similar to a workhorse structural beamline

Rate – With a pulses up to 1,000 times a second at $1x10^8$ photons, measurements are possible within each pulse.

- This allows you to do things '*differently*' to a synchrotron

Accessibility – Unlike XFEL facilities, a CXFEL could be a regional or even local facility greatly expanding access and availability for a wider range of studies.

- This allows you to do '*things*'.

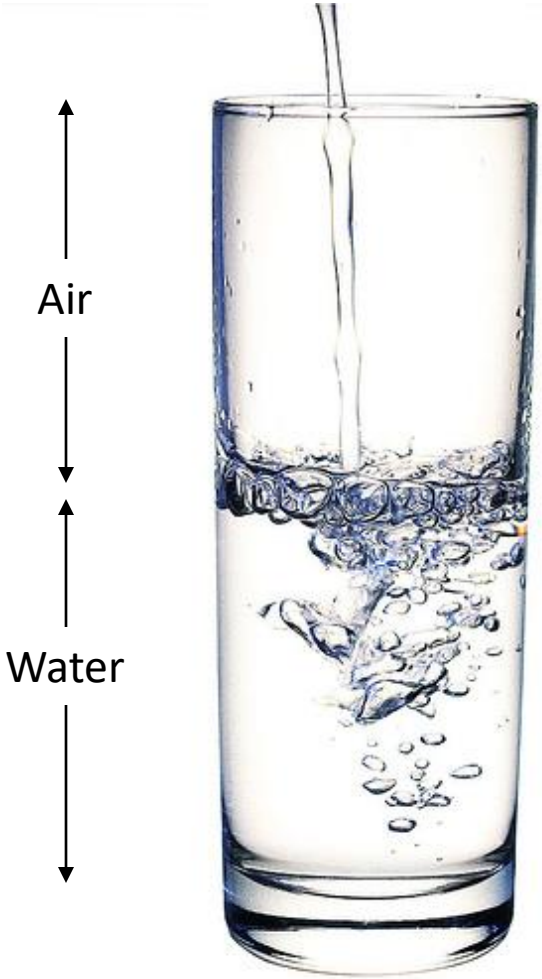
Potential impact in all areas of structural biology

- Crystallization – SAXS based profiling of crystallization conditions.
- Optimization – Diffraction based characterization of outcome.
- Structure – In situ diffraction.
- Ligand binding – Signals in molecular replacement data.
- Ligand complex structures – In situ diffraction.

Potential impact in all areas of structural biology

- Crystallization – SAXS based profiling of crystallization conditions.
- Optimization – Diffraction based characterization of outcome.
- Structure – In situ diffraction.
- Ligand binding – Signals in molecular replacement data.
- Ligand complex structures – In situ diffraction.

Pessimists, Optimists, and Crystallographers



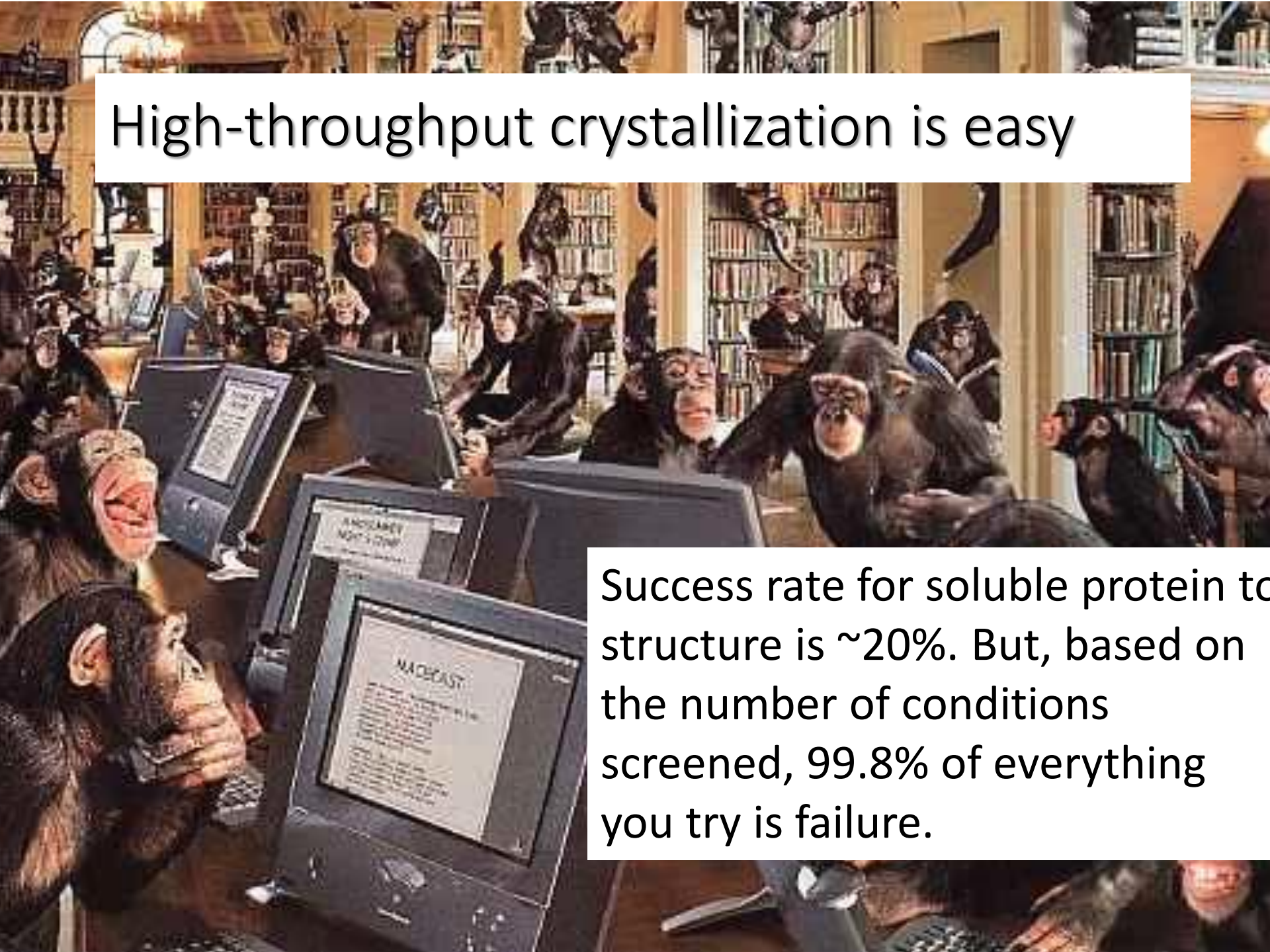
Consider a glass of water

Pessimist
(the glass is half empty)

Optimist
(the glass is half full)

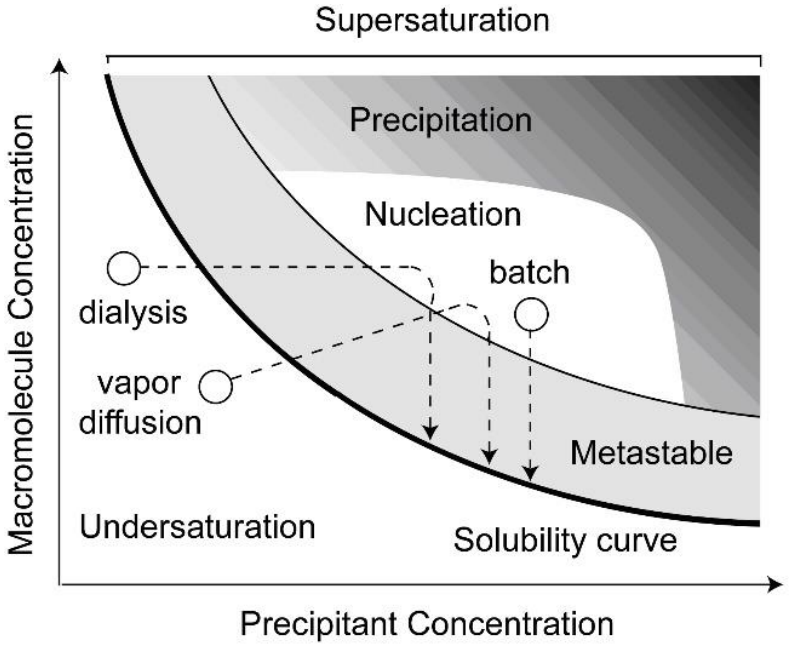
Crystallographer
(the glass is completely full)

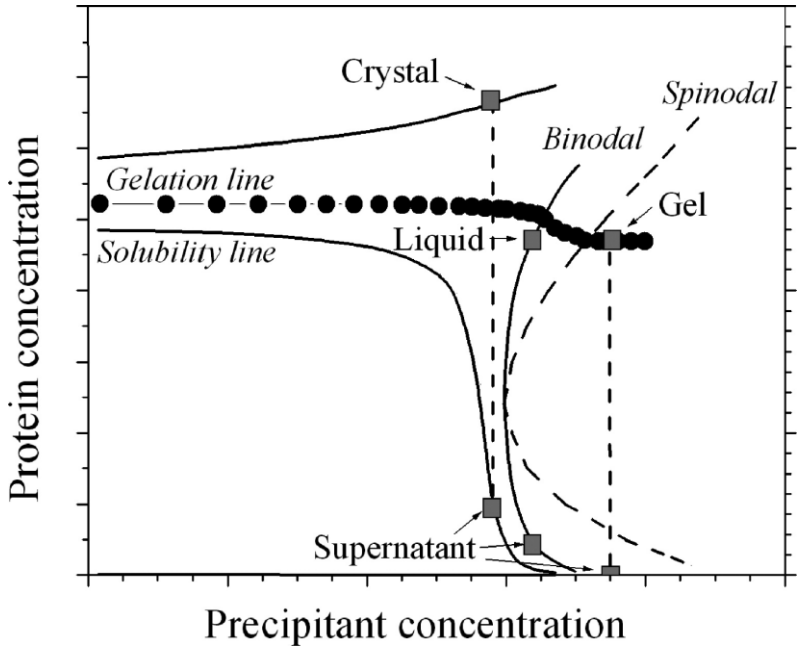
Fantasy


High-throughput crystallization is easy
Success rate for soluble protein to
structure is ~20%. But, based on
the number of conditions
screened, 99.8% of everything
you try is failure.

Crystallization theory is well established

We understand the fundamental theory around the protein phase diagram and the most efficient method to probe it is chemically screening different conditions

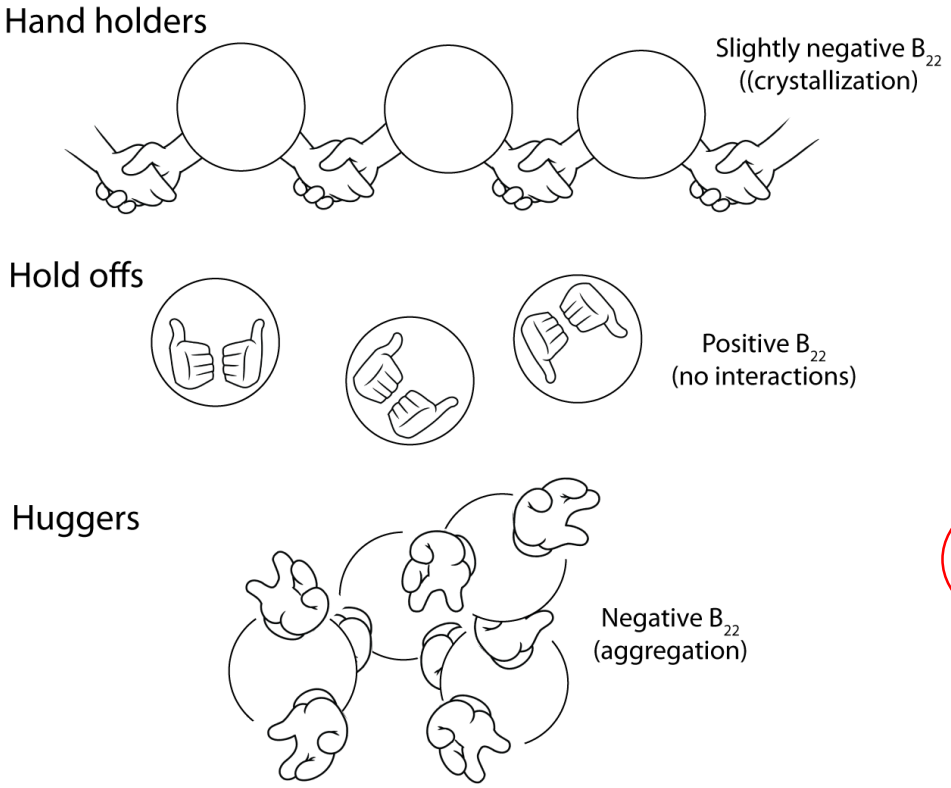




(Dumetz et al., 2009)

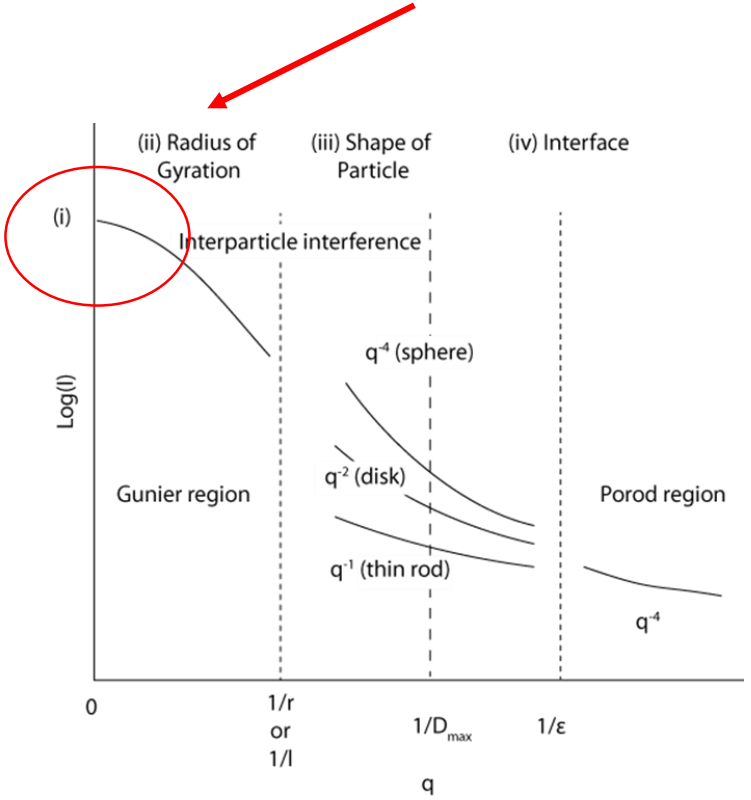
The driving force in outcome is to establish interactions between individual molecules of the protein of interest.

Proteins can be hand holders, hold offs, and huggers



The interaction is measured using the second viral coefficient or B_{22}, in essence a measure of how social the protein molecules are.

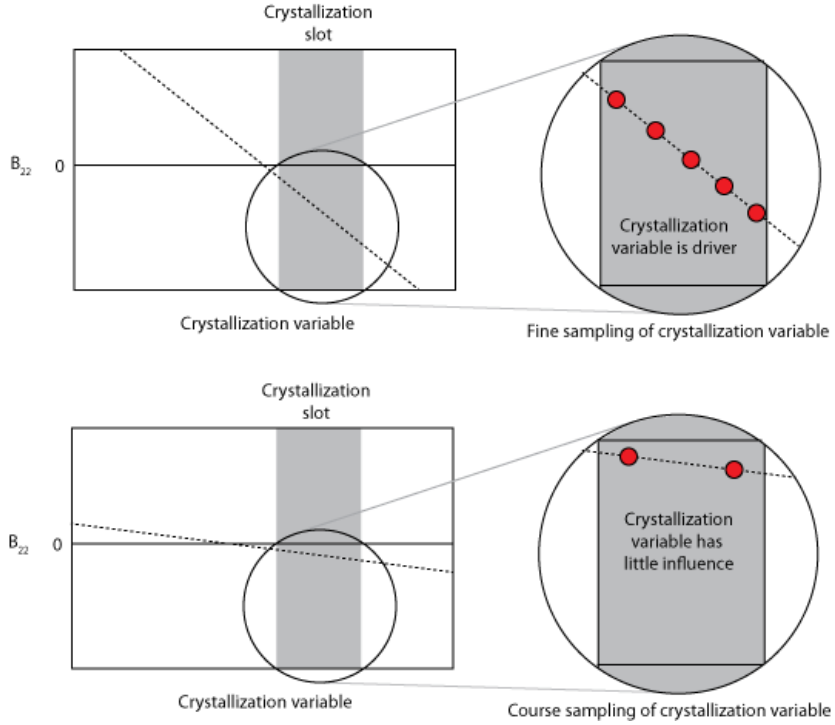
SAXS can measure interparticle interactions efficiently.
Low resolution (high signal) part of SAXS scattering



B_{22} measurements can be used to drive rational crystallization

B22 values in the range from -1x10-4 to -8x10-4 mol ml g^{-2}, known as the 'crystallization slot', have slightly attractive intermolecular interactions and have been correlated with crystallization (George and Wilson 1994, Wilson and DeLucas, 2014).

They have been used to successfully crystallize, Bovine Pancreatic Trypsin Inhibitor and Urate Oxidase, and rationally drive the process. As B22 became more negative nucleation rate increased and smaller crystals resulted.

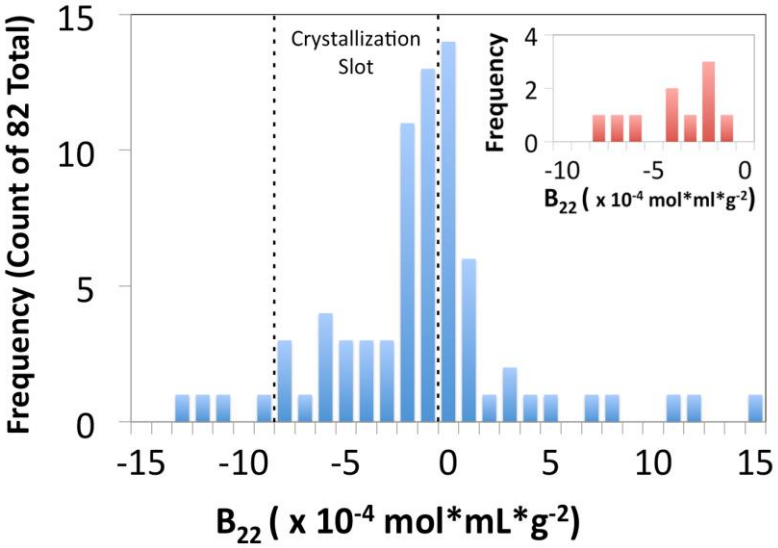


Experimentally measuring B22 with SAXS:

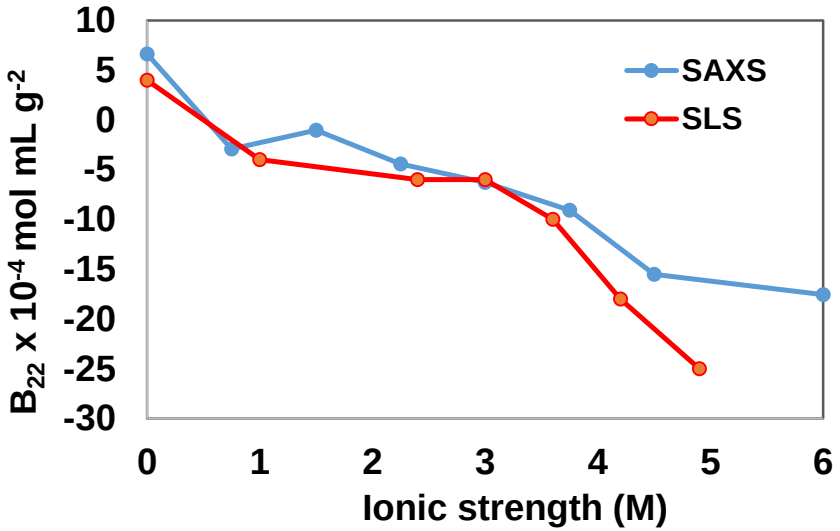
Three solutions

1. Protein in buffer
2. Buffer
3. Crystallization variable in buffer

A concentration series is run based on the crystallization variable rather than the protein concentration



B_{22} values from multiple native proteins in buffer solution and the B_{22} values determined by George and Wilson from light scattering.



A comparison of our SAXS B_{22} measurements for $(NH_4)_2SO_4$ and lysozyme together with light scattering results showing that SAXS B_{22} values are consistent with existing measurements.

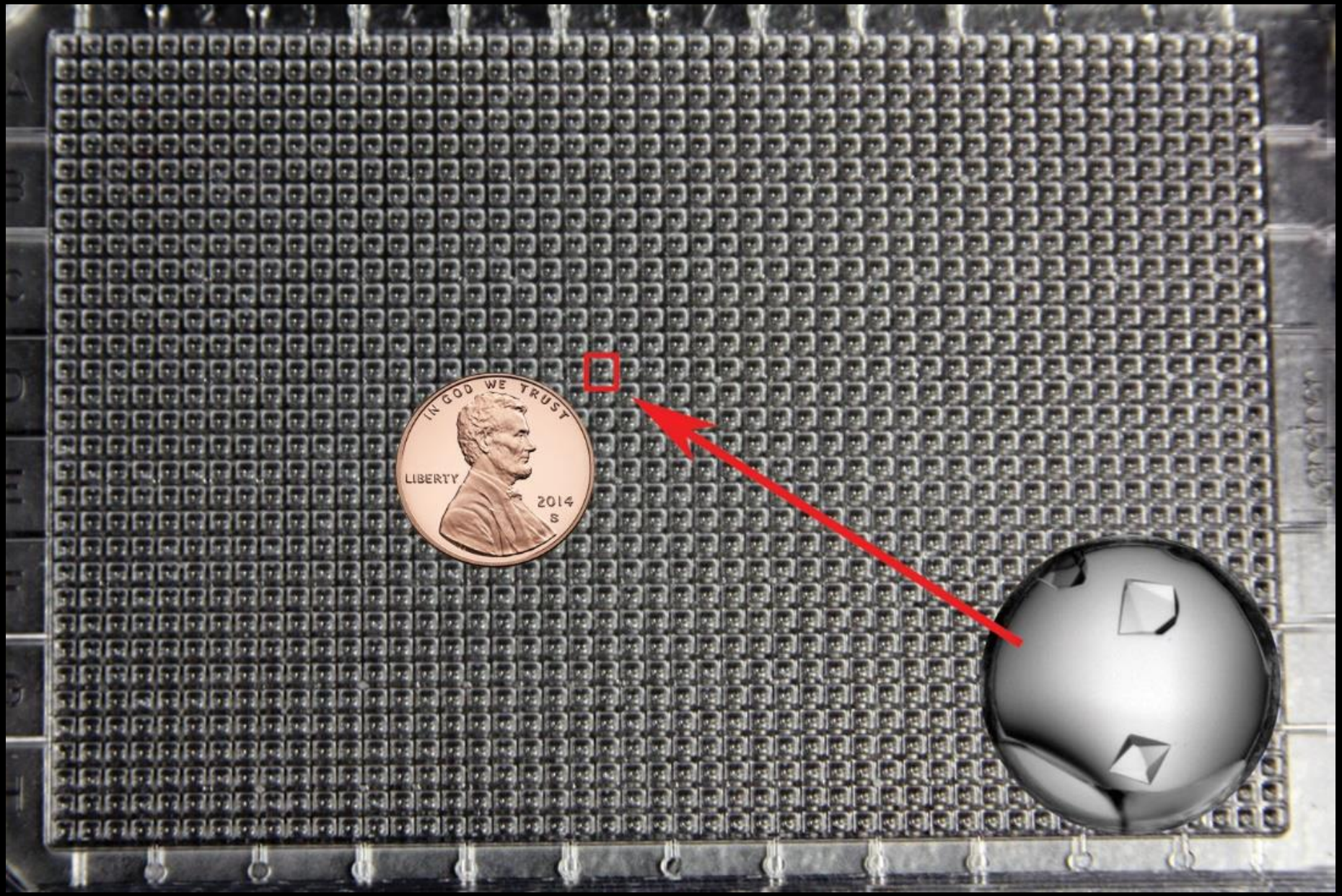
Proof of technique with data gathered on beamline 4-2 at SSRL. In one experiment 82 proteins were screened to determine B_{22} in buffer, in another lysozyme was studied to compare with B_{22} measurements made by light scattering.

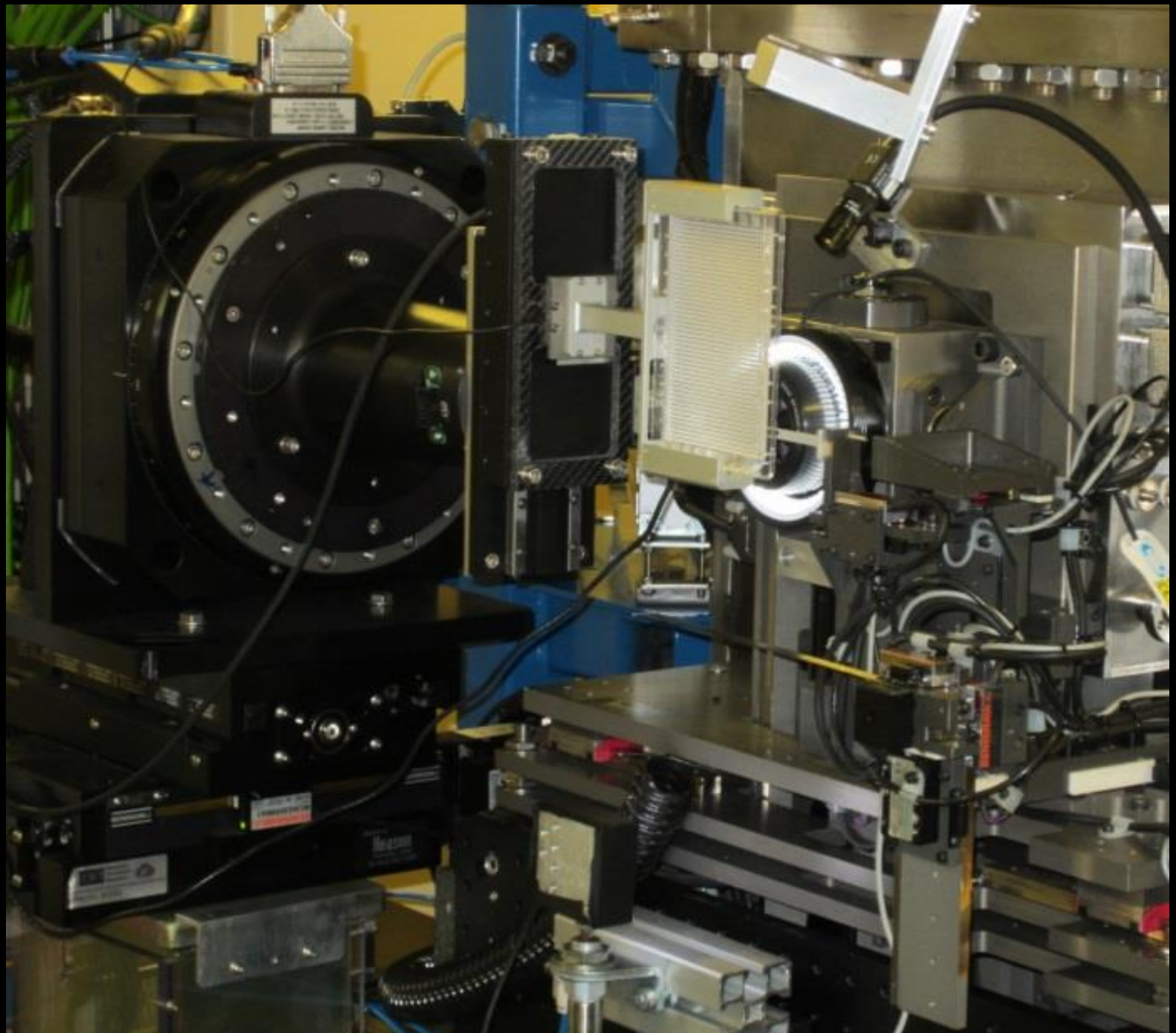
SAXS methods were rapid, required minimal protein concentration, and replicated results from static light scattering.

SAXS can be used to probe the dimensions of the crystallization window and from that determine the sampling fidelity for crystallization screening.

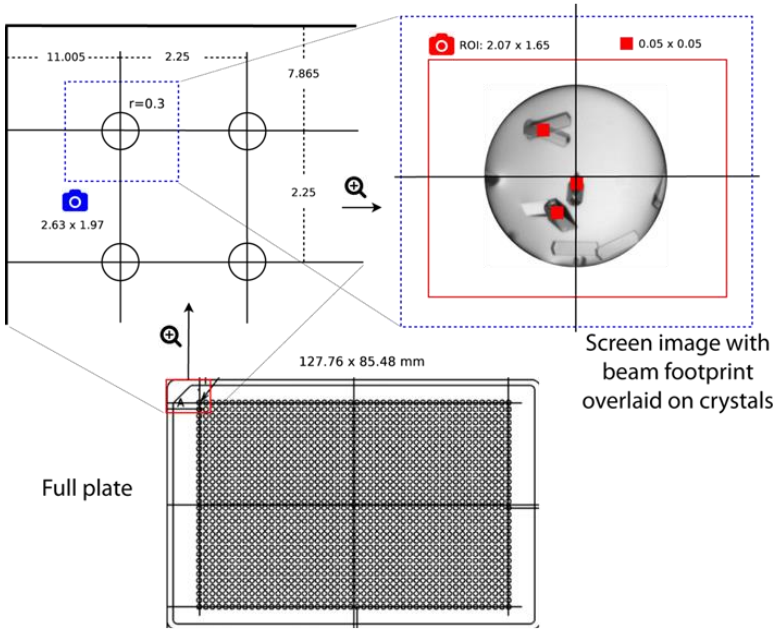
Potential impact in all areas of structural biology

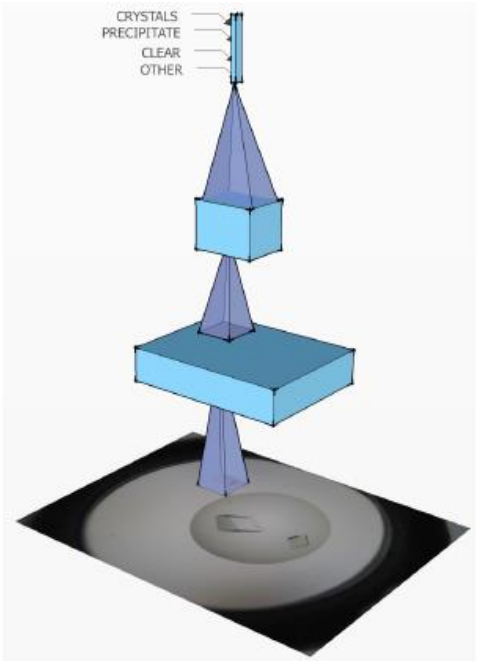
- Crystallization – SAXS based profiling of crystallization conditions.
- Optimization – Diffraction based characterization of outcome.
- Structure – In situ diffraction
- Ligand binding – Signals in molecular replacement data
- Ligand complex structures – In situ diffraction


IN GOD WE TRUST
LIBERTY
2014
S



Adding adjacent developments





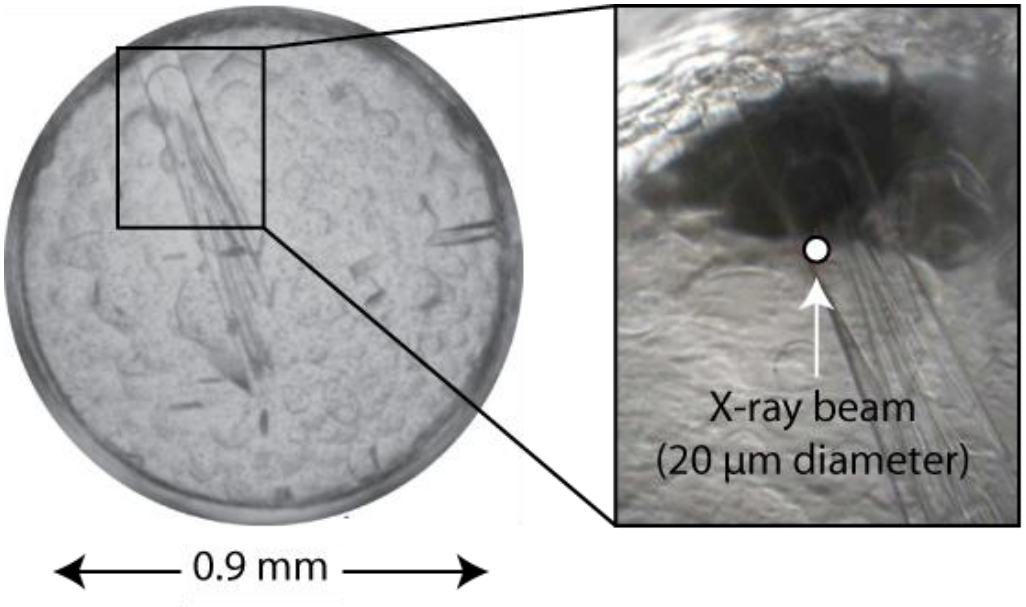
Deep convolutional neural network used to analyze ~500,000 images from 5 sources – collaboration with Google Brain.

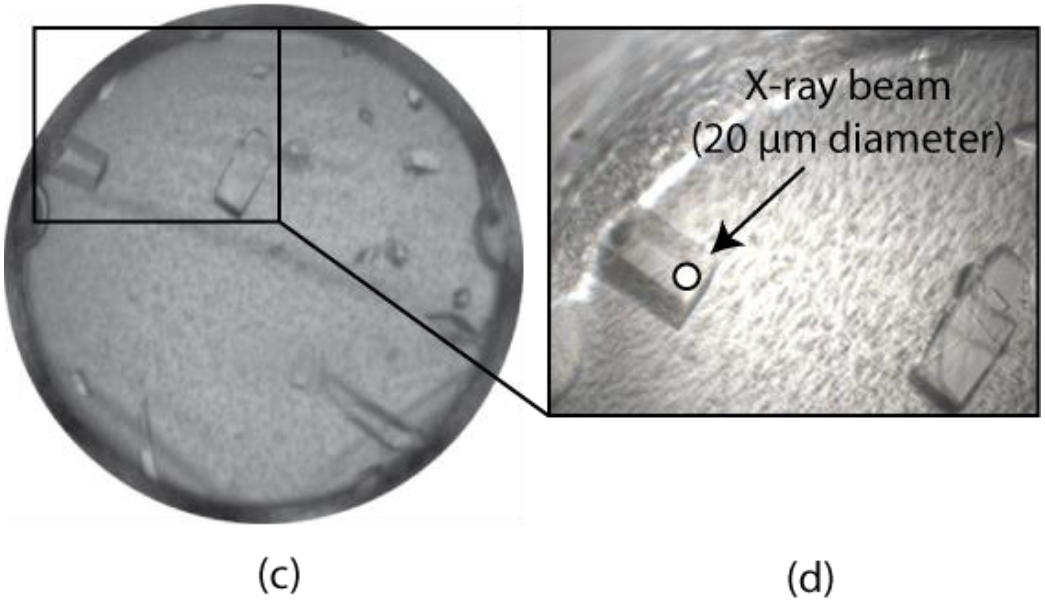
Bruno et al, PlosONE submitted.

Reliable and automatic identification of crystals

Haptic interfaces and web services as a screen to beam interface in crystallography. Bruno et al., J. Appl. Cryst, 2016.

True label	Predictions			
	Crystals	Precipitate	Clear	Other
Crystals	**91.0%**	5.8%	1.7%	1.5%
Precipitate	0.8%	**96.1%**	2.3%	0.7%
Clear	0.2%	1.8%	**97.9%**	0.2%
Other	4.8%	19.7%	5.9%	**69.6%**



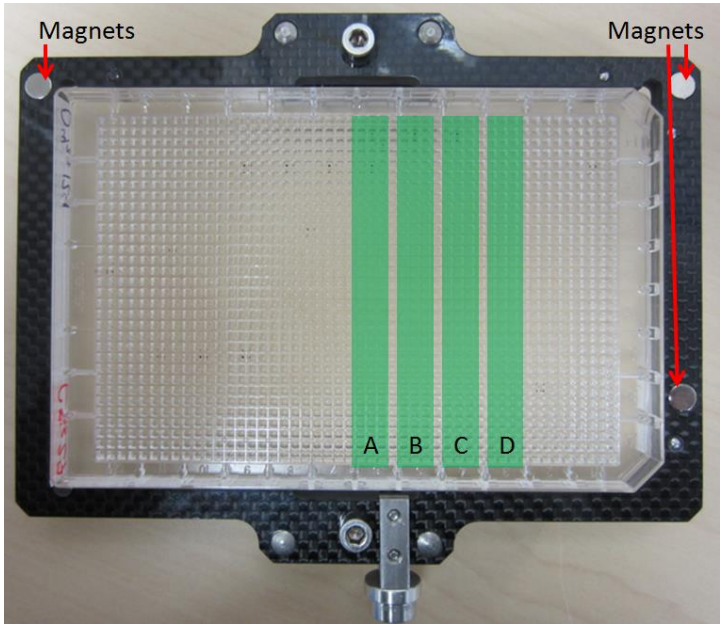


Crystal plates shipped by FedEx (Diamond and NSLS) and suitcase (Diamond)

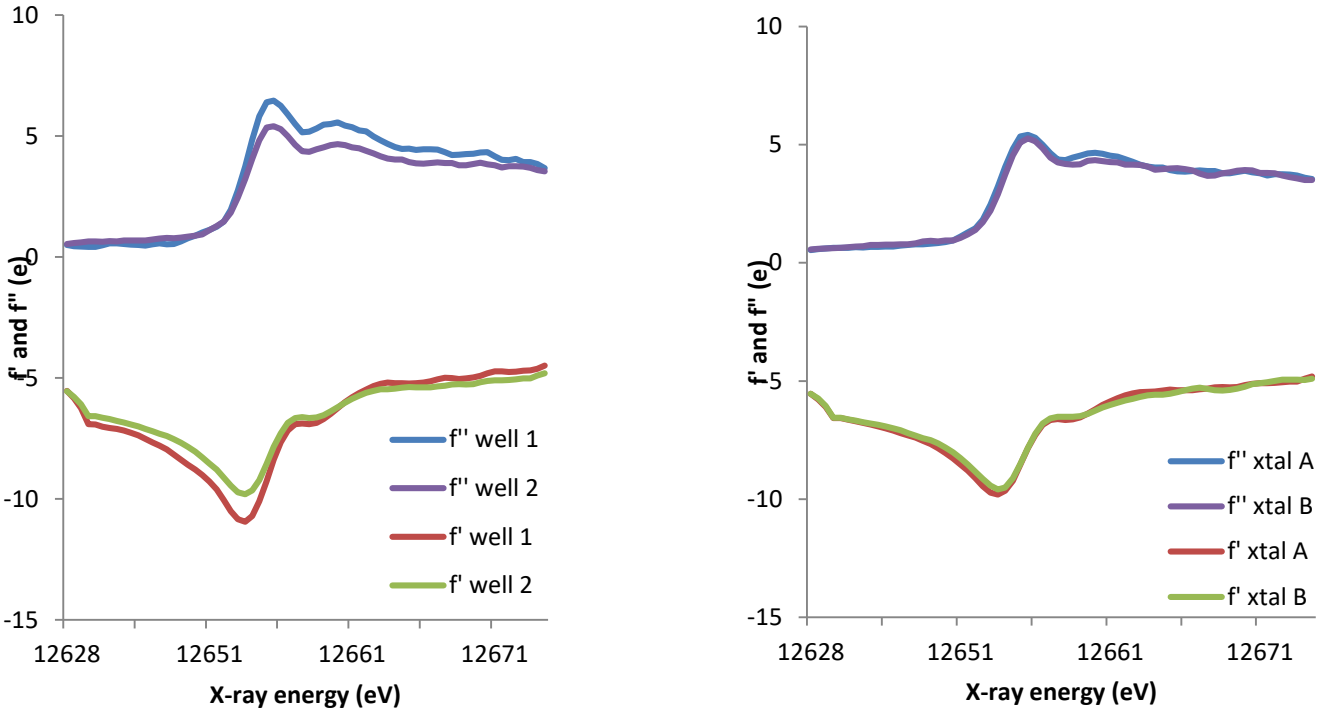
Crystals remained in place and diffracted.

In development with SSRL and NSLS-II using SAM and G-ROB robots respectively.

Software link from screening to positioning.



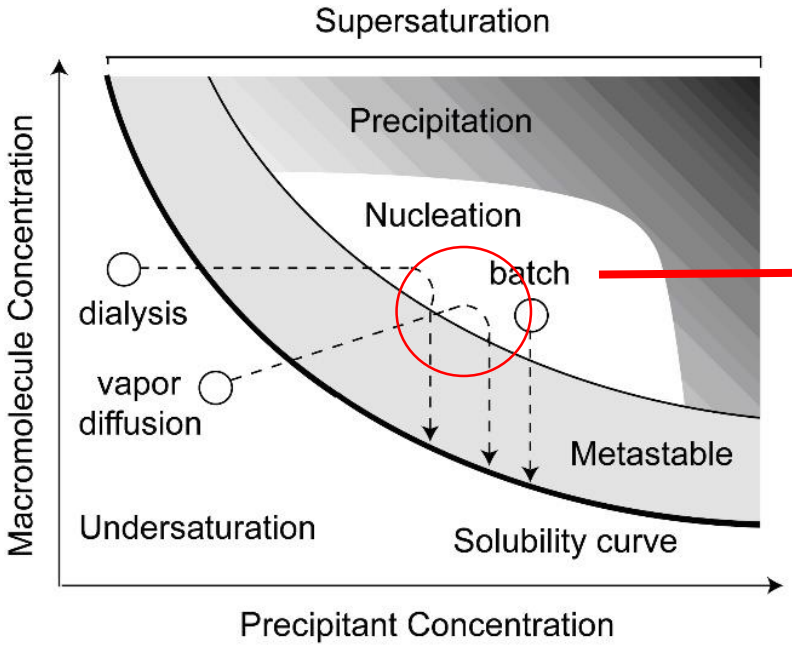
In situ MAD is possible

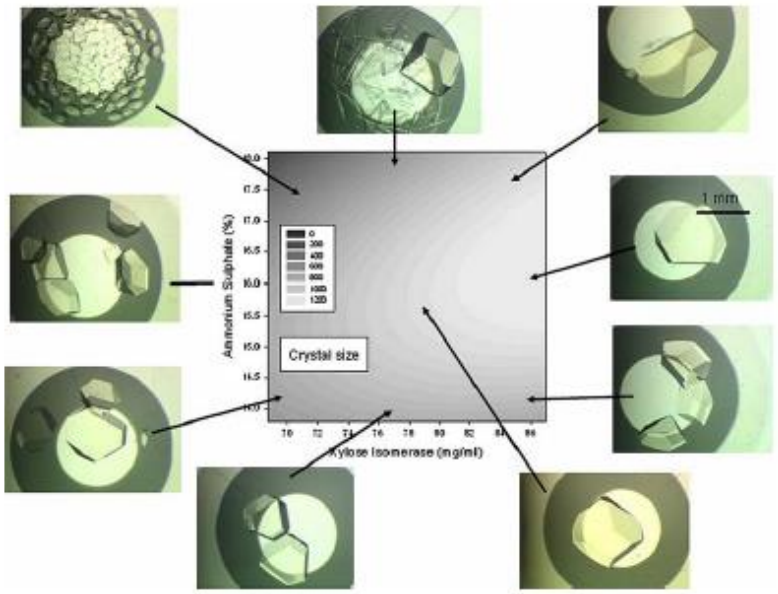


***In situ* fluorescence scans of crystals**. Top left is an overlay of two fluorescence scans from two separate crystals grown in the same well (condition B). Top right is an overlay of fluorescence scans of crystals grown in the same condition (B), but in different wells.

But not with 3% bandwidth? SAD or MR methods needed or wait until CXFEL.

X-ray based optimization is also possible





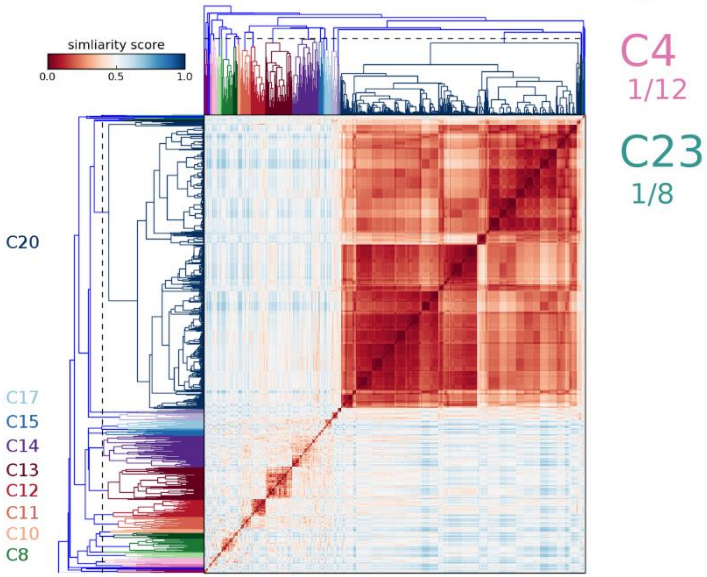
Response surface optimization methods can be used successfully for quantitative metrics (e.g. crystal volume).

This can be used to understand conditions that promote the best diffraction rather than nicest looking crystals.

Rapid X-ray analysis can provide such a metric allowing an understanding of the impact of the pathway through the nucleation and metastable zone and potentially strategies to improve diffraction.

New information representations

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



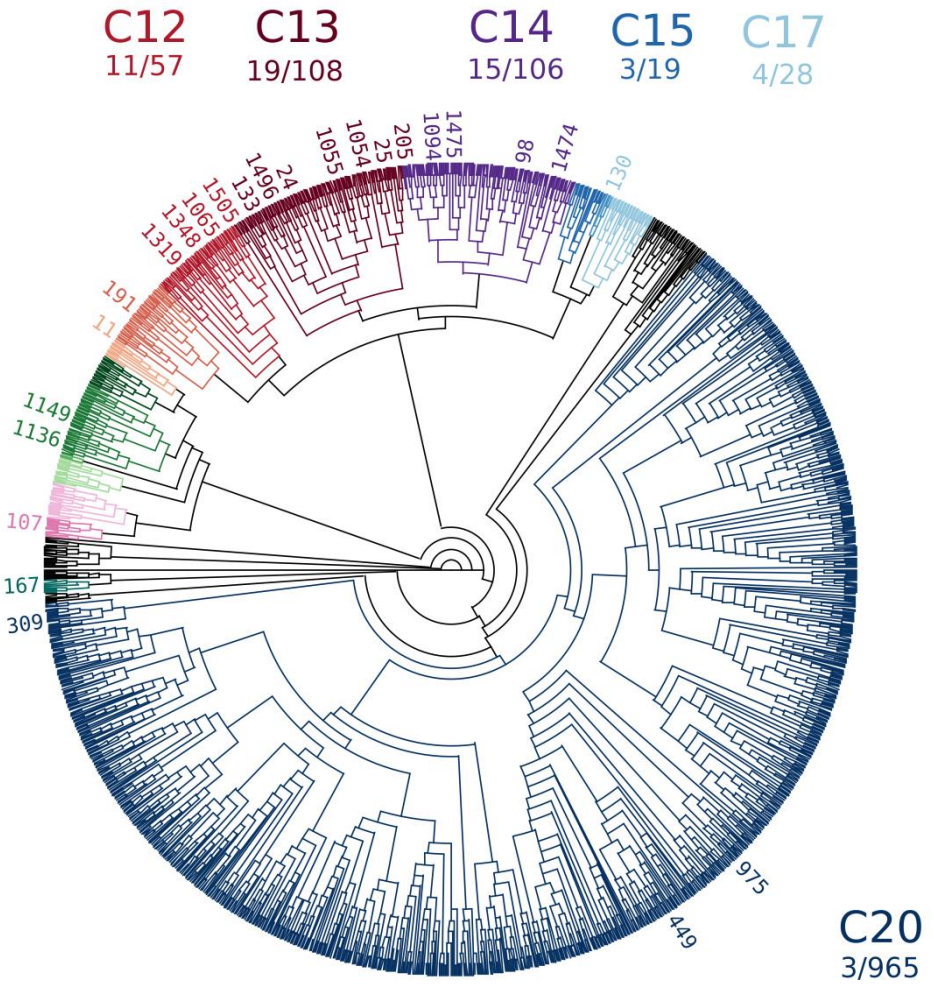
C11
5/42

C10
1/12

C8
7/45

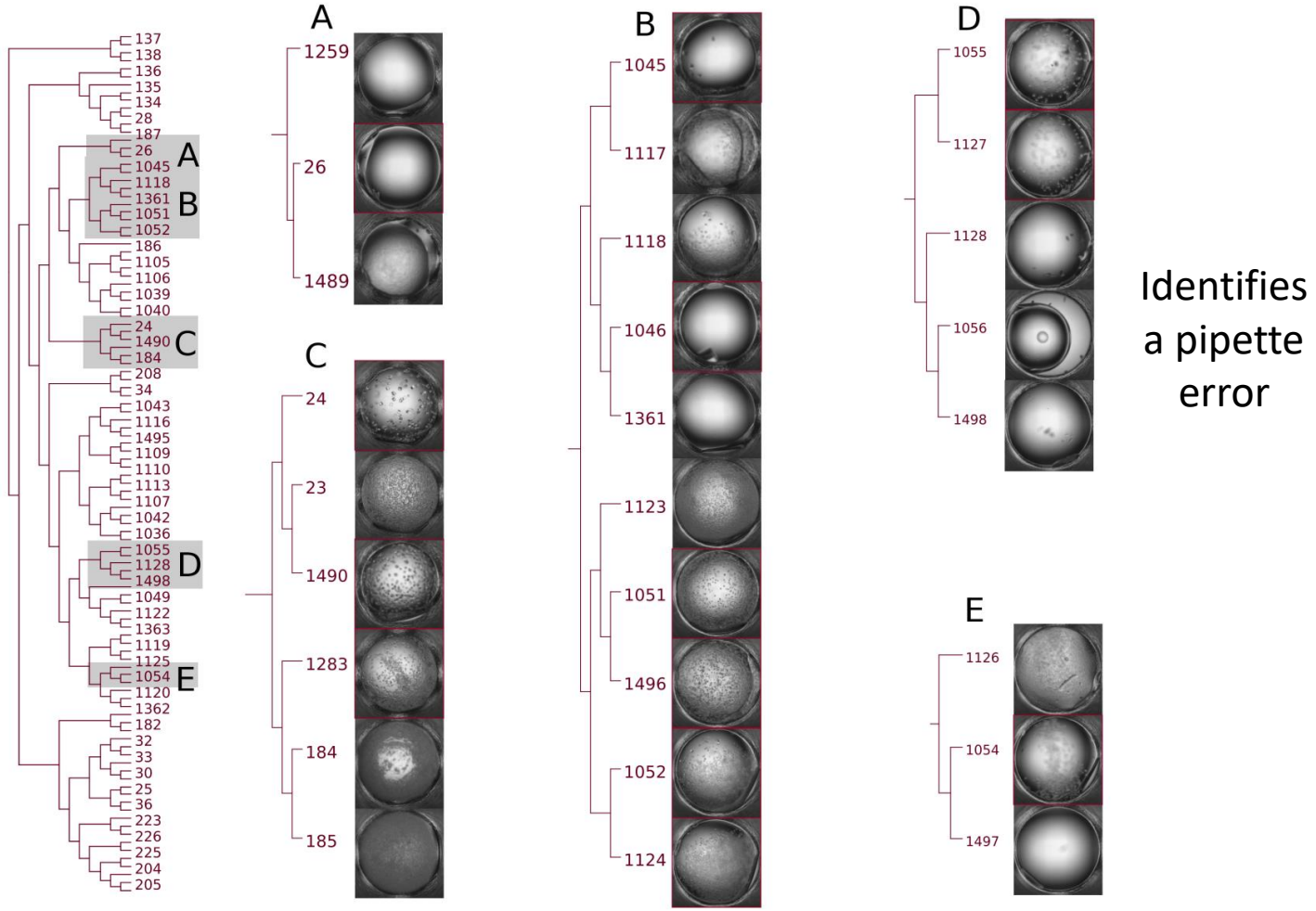
C4
1/12

C23
1/8



Cluster 20, PEG based, only 3 hits

Potential to understand phase diagram in terms of X-ray diffraction properties



Clustering samples the phase diagram

Potential impact in all areas of structural biology

- Crystallization – SAXS based profiling of crystallization conditions.
- Optimization – Diffraction based characterization of outcome.
- Structure – In situ diffraction
- Ligand binding – Signals in molecular replacement data
- Ligand complex structures – In situ diffraction

Ligand binding

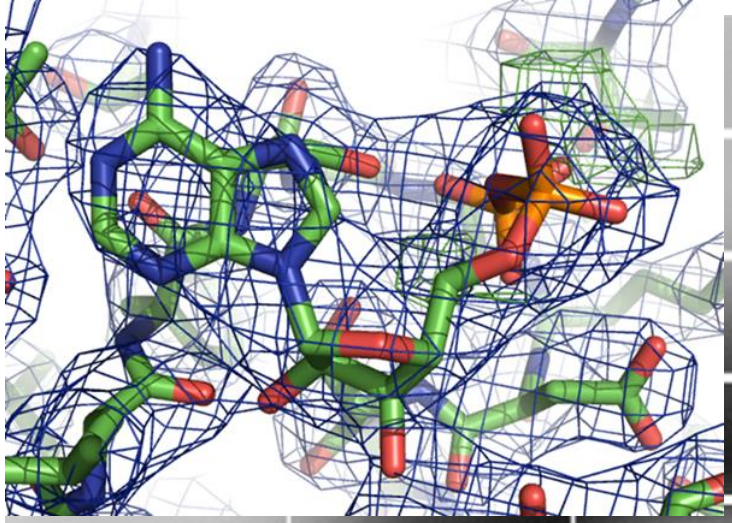
Potentially possible to do MR from a single image and provide information in if and possibly how a ligand is bound.

From “what if” conversations with Randy Read about what you can do with incomplete X-ray data:

- *… the likelihood approach should be less sensitive to systematic incompleteness than Patterson based methods, because there shouldn't be Fourier artefacts -- it's considering the information from one reflection at a time. … The current version of Phaser now uses this to predict what resolution limit you can get away with for the initial search … I would expect that it doesn't matter too much where the reflections come from (all else being equal, but of course if you have a poor model then adding a poorly-predicted high resolution reflection isn't the same as including a well-predicted low resolution reflection).*
- … With a good model, you get a very clear solution even restricting the resolution to 5A (<600 reflections) …
- … if all you want to do is rigid-body refinement, without needing a clear solution from the full molecular replacement search, you might be alright with even less data.

Potential impact in all areas of structural biology

- Crystallization – SAXS based profiling of crystallization conditions.
- Optimization – Diffraction based characterization of outcome.
- Structure – In situ diffraction
- Ligand binding – Signals in molecular replacement data
- Ligand complex structures – In situ diffraction



Minimal background from plate and oil

Diffraction to 2.3A from plate

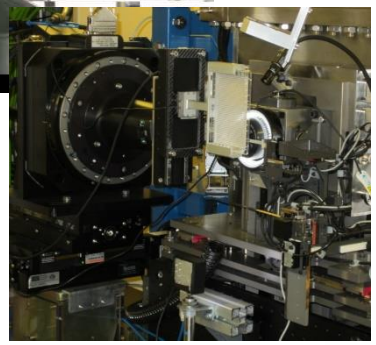
On a microfocus system, multiple crystals can be shot individually within each well and data is sufficient to locate ligands.

Ligand bound in N-type ATP Pyrophosphatase. Above is the structure solved for the N-type ATP Pyrophosphatase from P. furiousus in which cocrystallized with ADP. This structure was solved by molecular replacement with the published ATP bound structure (PDB ID: 3RK1) with ATP removed from the search model

What is needed to make this possible beyond a compact X-ray source?

Cannot use methods that are suitable for single of few proteins – dead beamtime

Sample prep



Sample handling

Data analysis

People

Instrumentation considerations

- **Crystallization** – SAXS based profiling of crystallization conditions – rapid motion, low absorption, easy setup, water impermeable, reproducible sample chamber (rapid).
- **Optimization** – Diffraction based characterization of outcome – rapid motion, low absorption, easy setup, water impermeable, not impeding diffraction cone, bandwidth variation to estimate mosaicity (rapid).
- **Structure** – In situ diffraction – as above but also potential motion stage needed for phi oscillation or use of wider bandwidth and therefore Ewald sphere sampling. Difficulty phasing, reduce bandwith but at expense of brilliance.
- **Ligand binding** – Signals in molecular replacement data – as above, possibly single images (rapid).
- **Ligand complex structures** – In situ diffraction – same considerations as structure.

Experimental design considerations

- **Crystallization** – SAXS based profiling of crystallization conditions – sampling variables (screen to use).
- **Optimization** – Diffraction based characterization of outcome – how to define optimum, volume impact.
- **Structure** – In situ diffraction – ideal bandwidth, multiple shots or crystals, minimizing dead time.
- **Ligand binding** – Signals in molecular replacement data – as above.
- **Ligand complex structures** – In situ diffraction – same considerations as structure.

Why would you do this? Beyond the science it could be self supporting?

- This has the potential to be a bread an butter use of the compact X-ray source.
- Could identify crystallization driving forces and if crystallization is likely to fail. Information that is valuable.
- Could enable crystallization and focus on conditions that have best diffraction properties. Information that is valuable.
- Rapid ligand screening with binding information and potential structure. Information that is valuable.

- SBS format – ~11,000 mm^2, (~128 x 85 mm), translation of 10 mm/s easy, 2 minutes per plate plus change overhead. 20 plates per hour, 480 per day. If information on plate is worth $100, $48K per day, $4M per year if run for once a week, once a month.
- Are there enough people that want this, is this too much demand on the instrument?

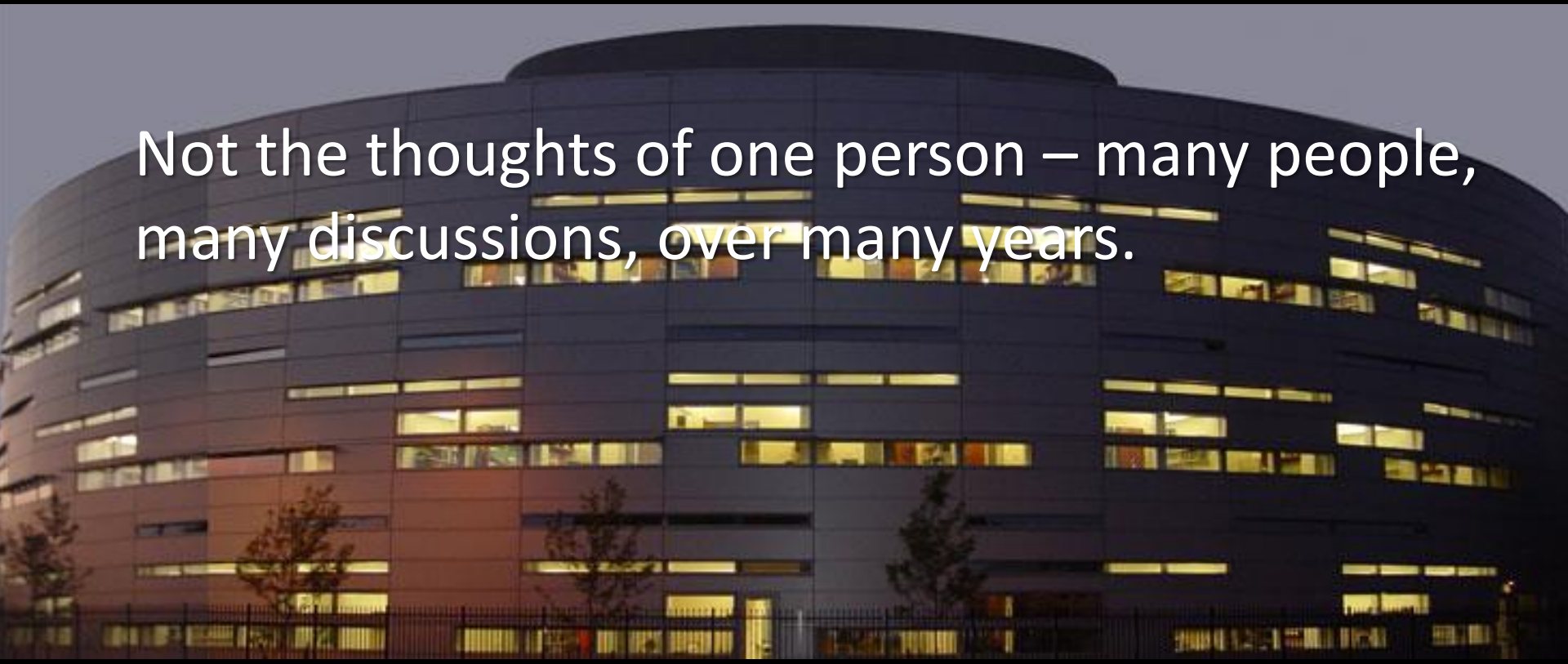
Got a protein?

Get a crystal™

500 µl protein at a ~10 mg/ml, setup against almost every Hampton screen and an incomplete factorial sampling of chemical space, visual images weekly over 6 weeks, SONICC and UV verification, remote data access. Automated optimization also available.

Details at: ***GetACrystal.org***

Thank you and questions?



Not the thoughts of one person – many people, many discussions, over many years.

esnell@hwi.buffalo.edu