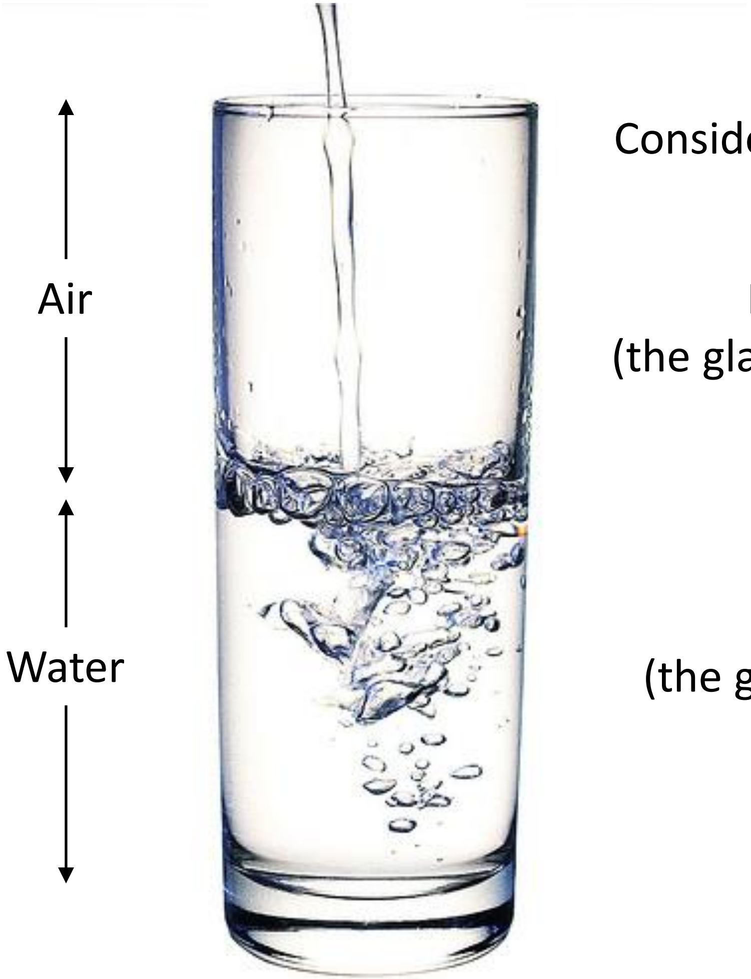
Small Angle X-ray Scattering as a Complement to X-ray Crystallography

Edward H. Snell

Pessimists, Optimists, and Crystallographers



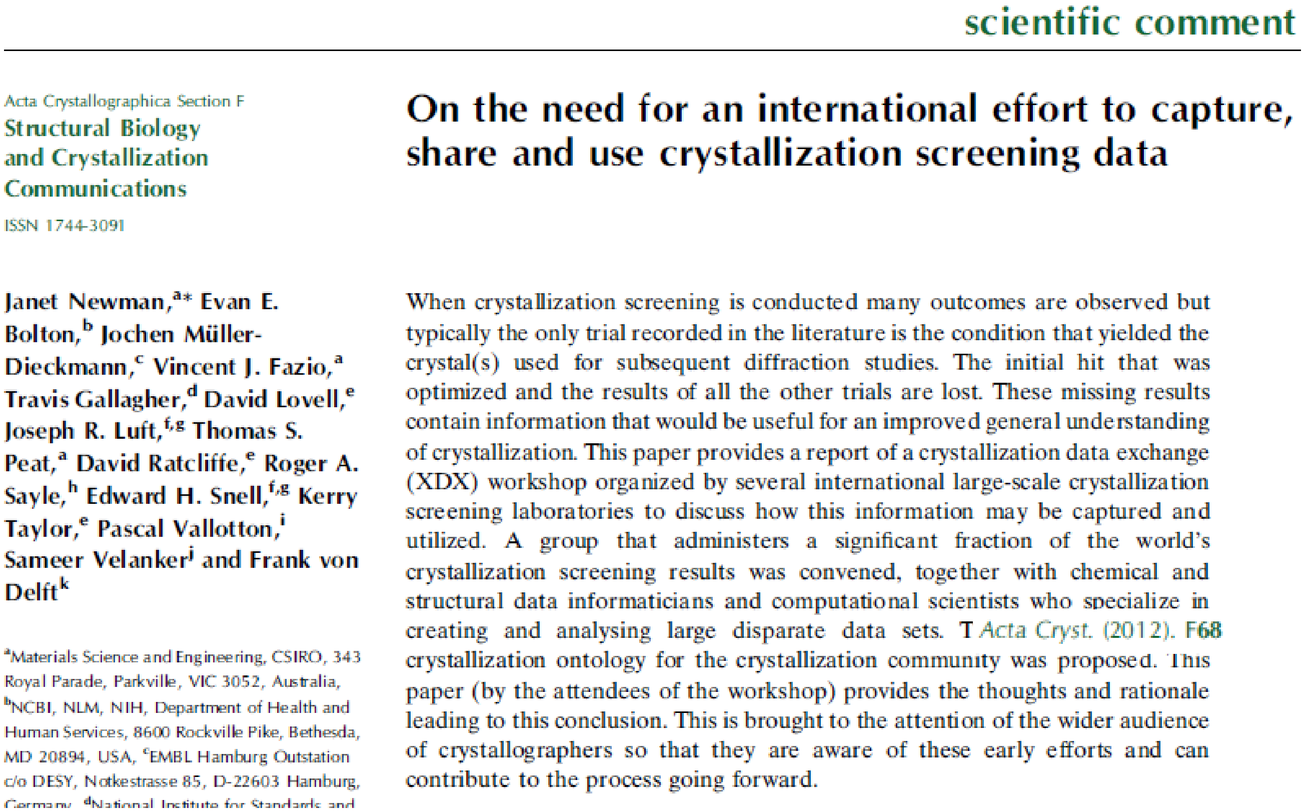
Consider a glass of water

Pessimist
(the glass is half empty)

Crystallographer
(the glass is completely full)

Optimist
(the glass is half full)

Only approximately 11% of the proteins we target for crystallography yield a crystallographic structure.



Acta Crystallographica Section F
Structural Biology
and Crystallization
Communications

ISSN 1744-3091

On the need for an international effort to capture, share and use crystallization screening data

Janet Newman,[a]* Evan E. Bolton,[b] Jochen Müller-Dieckmann,[c] Vincent J. Fazio,[a] Travis Gallagher,[d] David Lovell,[e] Joseph R. Luft,[f,g] Thomas S. Peat,[a] David Ratcliffe,[e] Roger A. Sayle,[h] Edward H. Snell,[f,g] Kerry Taylor,[e] Pascal Vallotton,[i] Sameer Velanker[j] and Frank von Delft[k]

[a]Materials Science and Engineering, CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia, [b]NCBI, NLM, NIH, Department of Health and Human Services, 8600 Rockville Pike, Bethesda, MD 20894, USA, [c]EMBL Hamburg Outstation c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, [d]National Institute for Standards and

When crystallization screening is conducted many outcomes are observed but typically the only trial recorded in the literature is the condition that yielded the crystal(s) used for subsequent diffraction studies. The initial hit that was optimized and the results of all the other trials are lost. These missing results contain information that would be useful for an improved general understanding of crystallization. This paper provides a report of a crystallization data exchange (XDX) workshop organized by several international large-scale crystallization screening laboratories to discuss how this information may be captured and utilized. A group that administers a significant fraction of the world's crystallization screening results was convened, together with chemical and structural data informaticians and computational scientists who specialize in creating and analysing large disparate data sets. T Acta Cryst. (2012). F68 crystallization ontology for the crystallization community was proposed. This paper (by the attendees of the workshop) provides the thoughts and rationale leading to this conclusion. This is brought to the attention of the wider audience of crystallographers so that they are aware of these early efforts and can contribute to the process going forward.

At least 99.8% of crystallization experiments produce an outcome other than crystallization.

Fantasy

High-throughput Crystallization Screening at the Hauptman-Woodward Medical Research institute

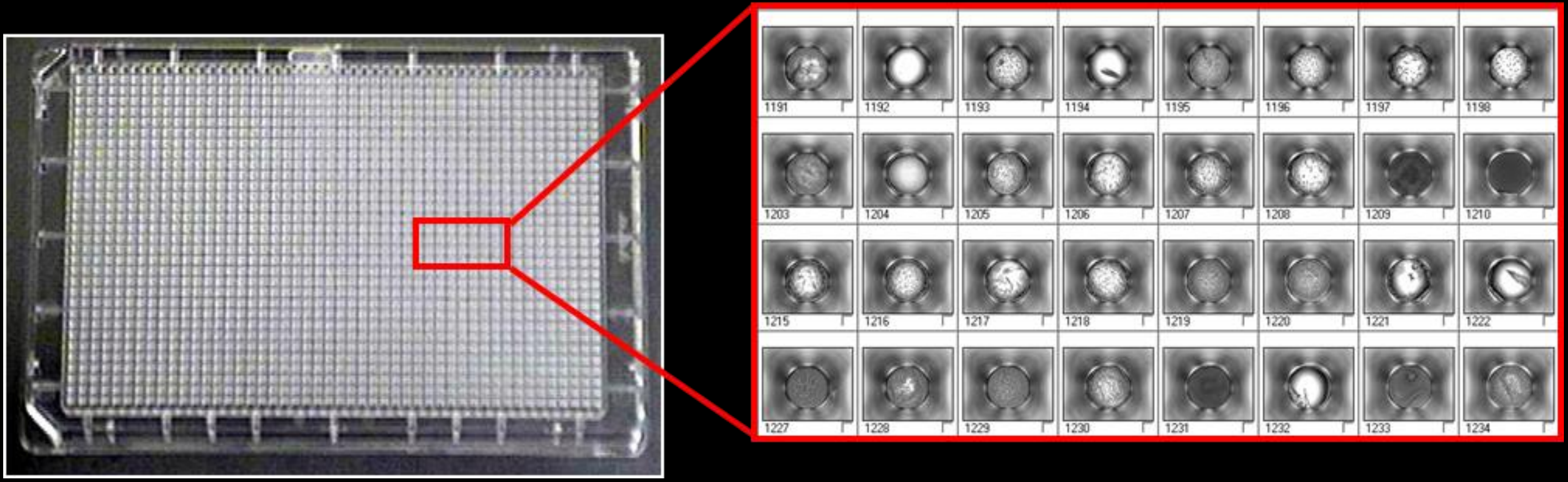
The Crystallization Screening laboratory at the Hauptman-Woodward Medical Research Institute

Since February of 2000 the High Throughput Search (HTS) laboratory has been screening potential crystallization conditions as a high-throughput service

The HTS lab screens samples against three types of cocktails:

1. Buffered salt solutions varying pH, anion and cation and salt concentrations
2. Buffered PEG and salt, varying pH, PEG molecular weight and concentration and anion and cation type
3. Almost the entire Hampton Research Screening catalog.

The HTSlab has investigated the crystallization properties of over 15,000 individual proteins archiving approximately 140 million images of crystallization experiments.

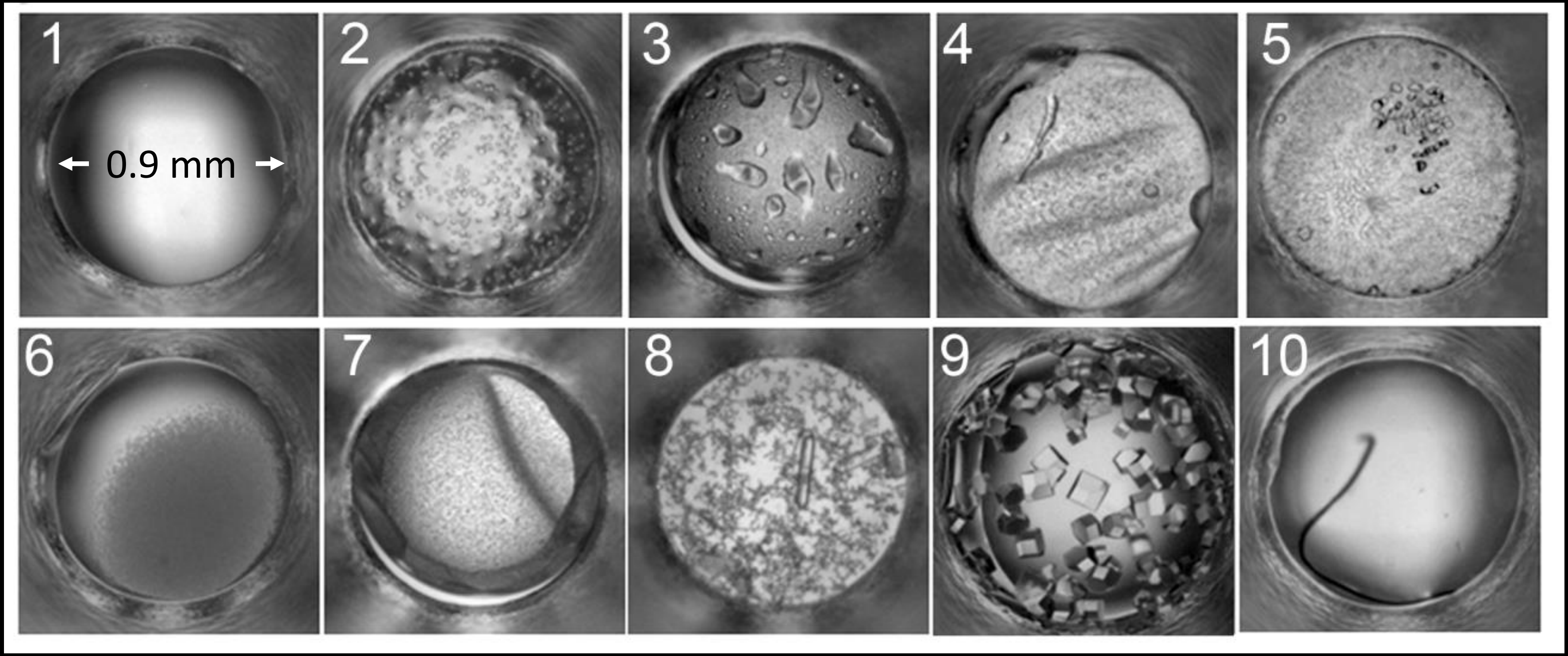


The crystallization method used is micro-batch under oil with 200 nl of protein solution being added to 200 nl of precipitant cocktail in each well of a 1536 well plate.

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

Several software utilities for viewing and analyzing data are available.

Outcomes



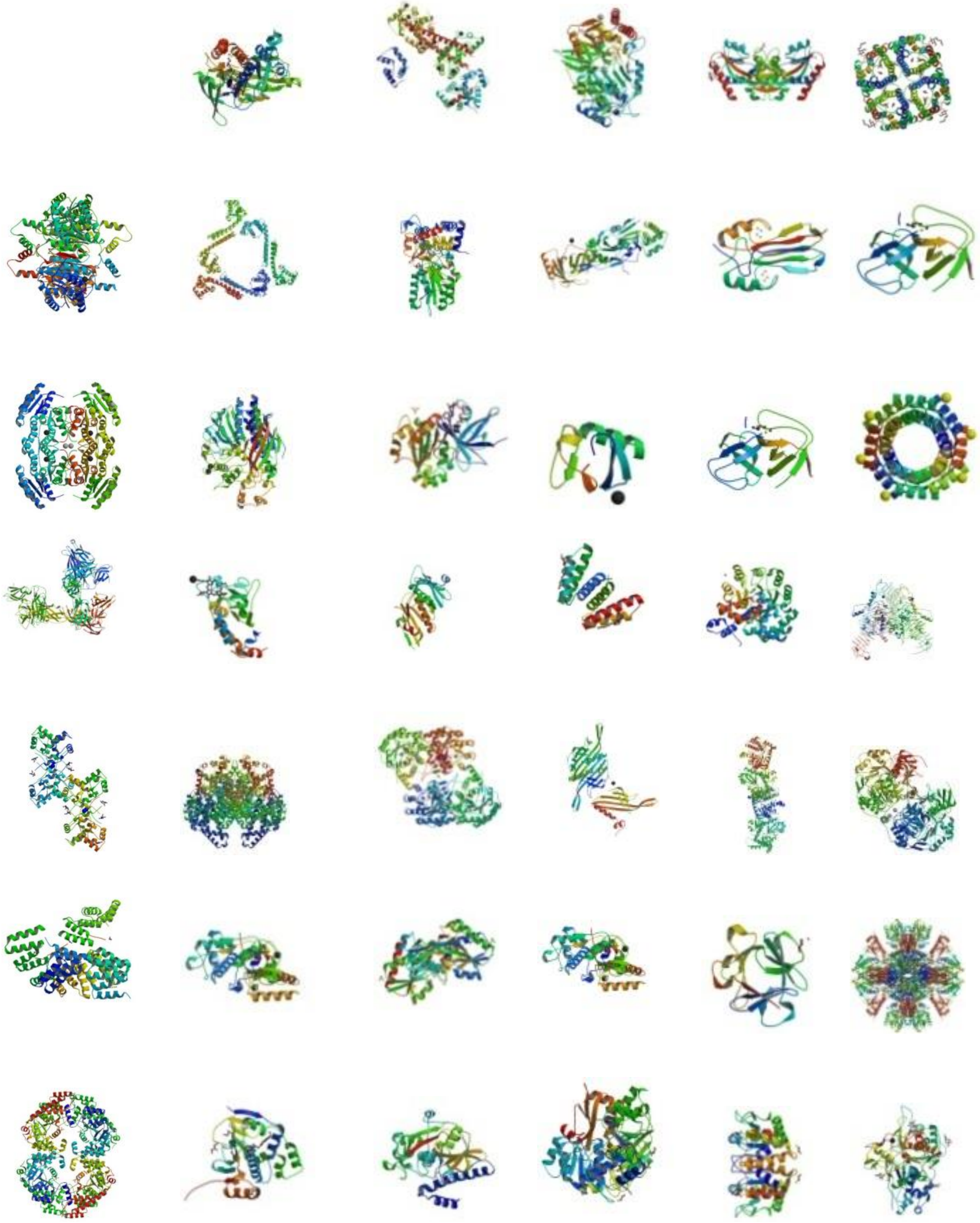
Born in Buffalo

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.



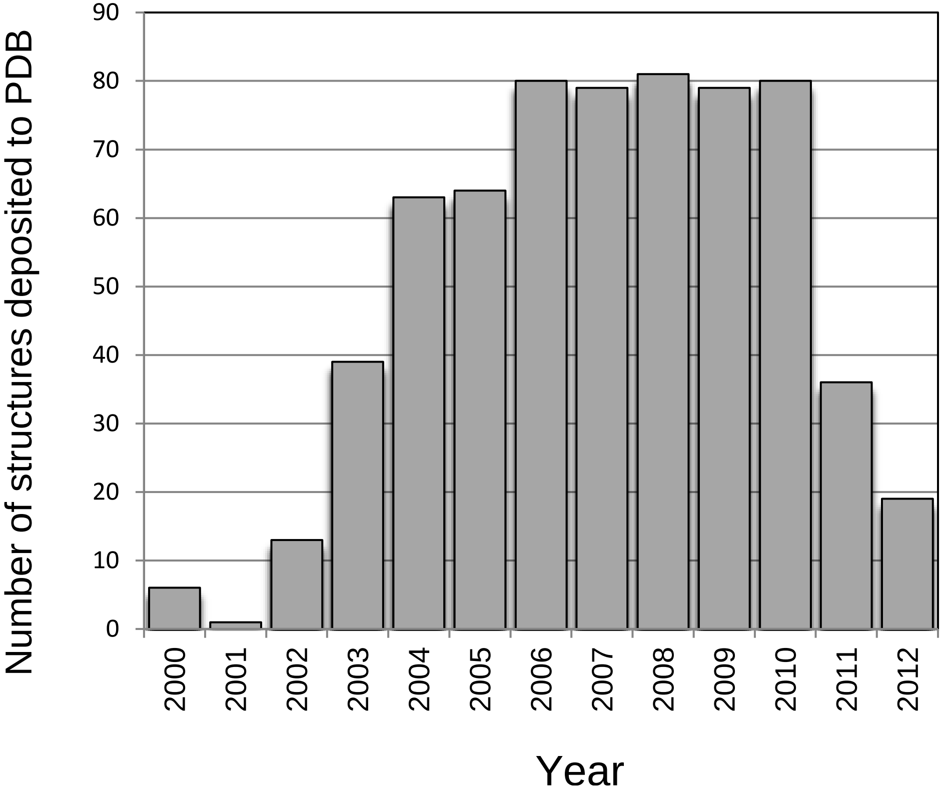
Where success is tracked.

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.



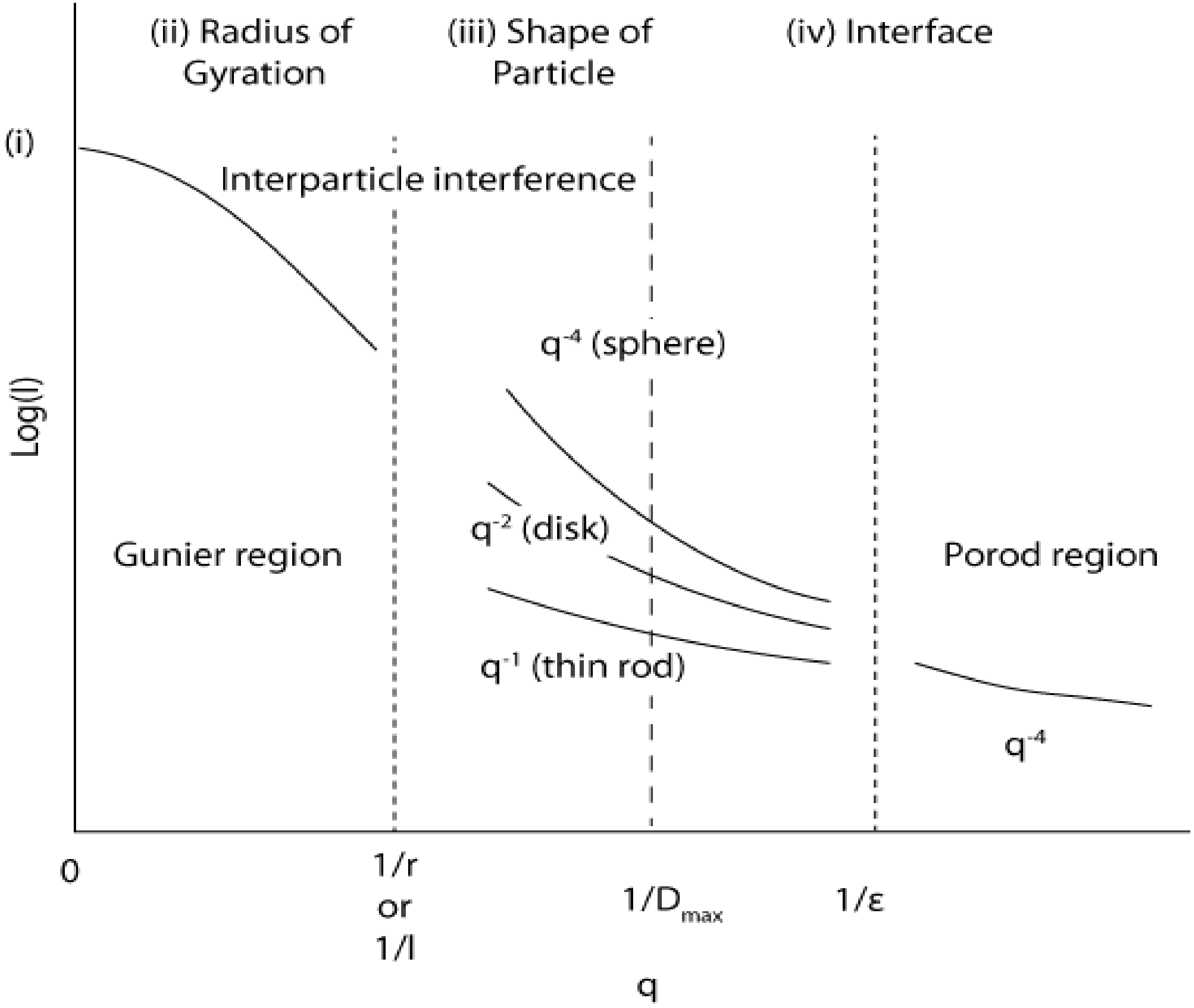
In 2011 we switched to PSI Biology – More difficult targets

High throughput

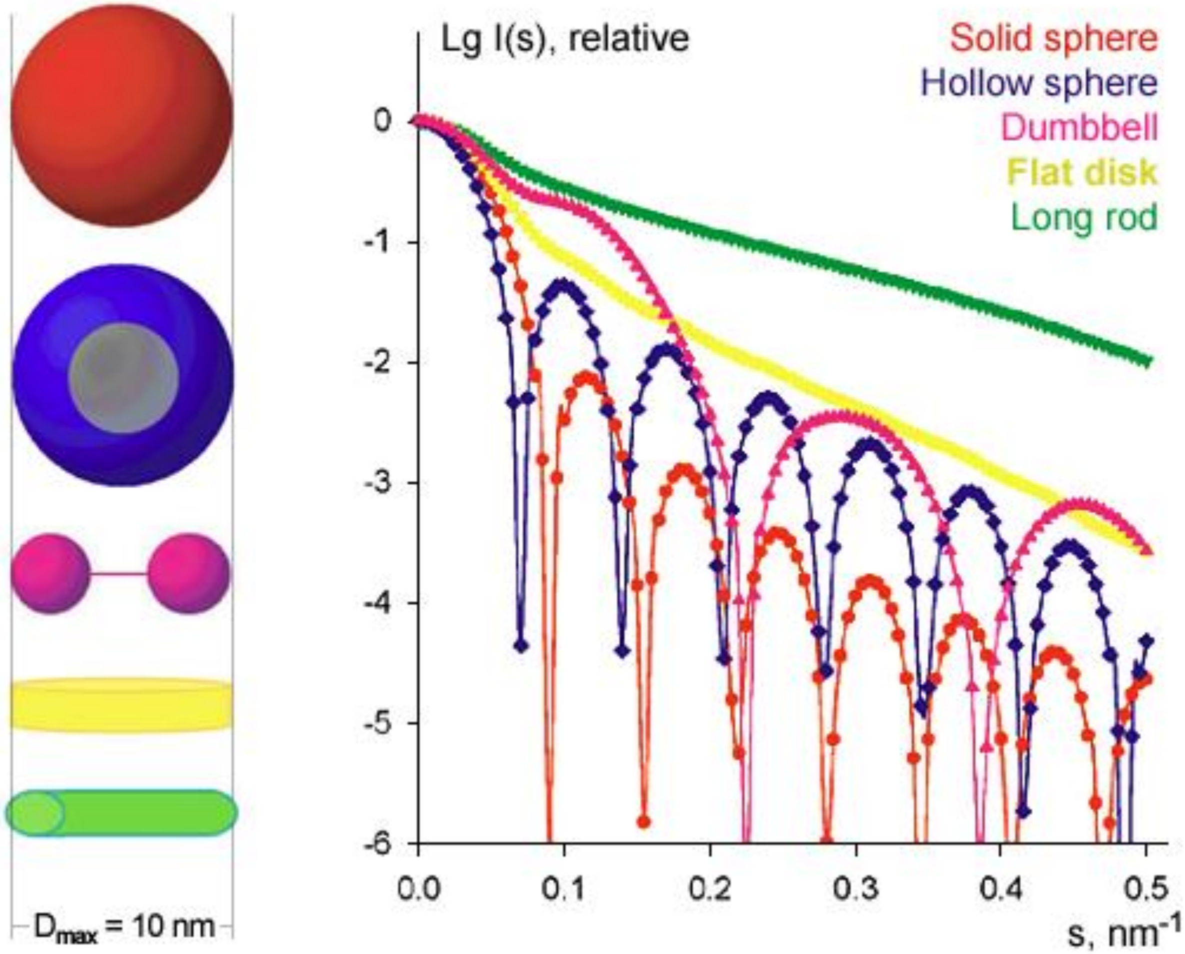
- At our high-throughput crystallization facility we have run ~16,000 different proteins.
- Crystals result in about 50% of cases.
- Where we track results (PSI samples, ~4,000) about 50% of samples that give crystals go on to a PDB deposition (25% of total).

75% of samples do not give structures - Frustration

- All our samples are in solution.
- So since 2007 we have been developing high-throughput strategies to take the remaining dregs of crystallization samples from NESG (~60 microL) and gathering SAXS data.
- To date, SAXS data from over 1,000 different proteins (at least 3 concentrations each)

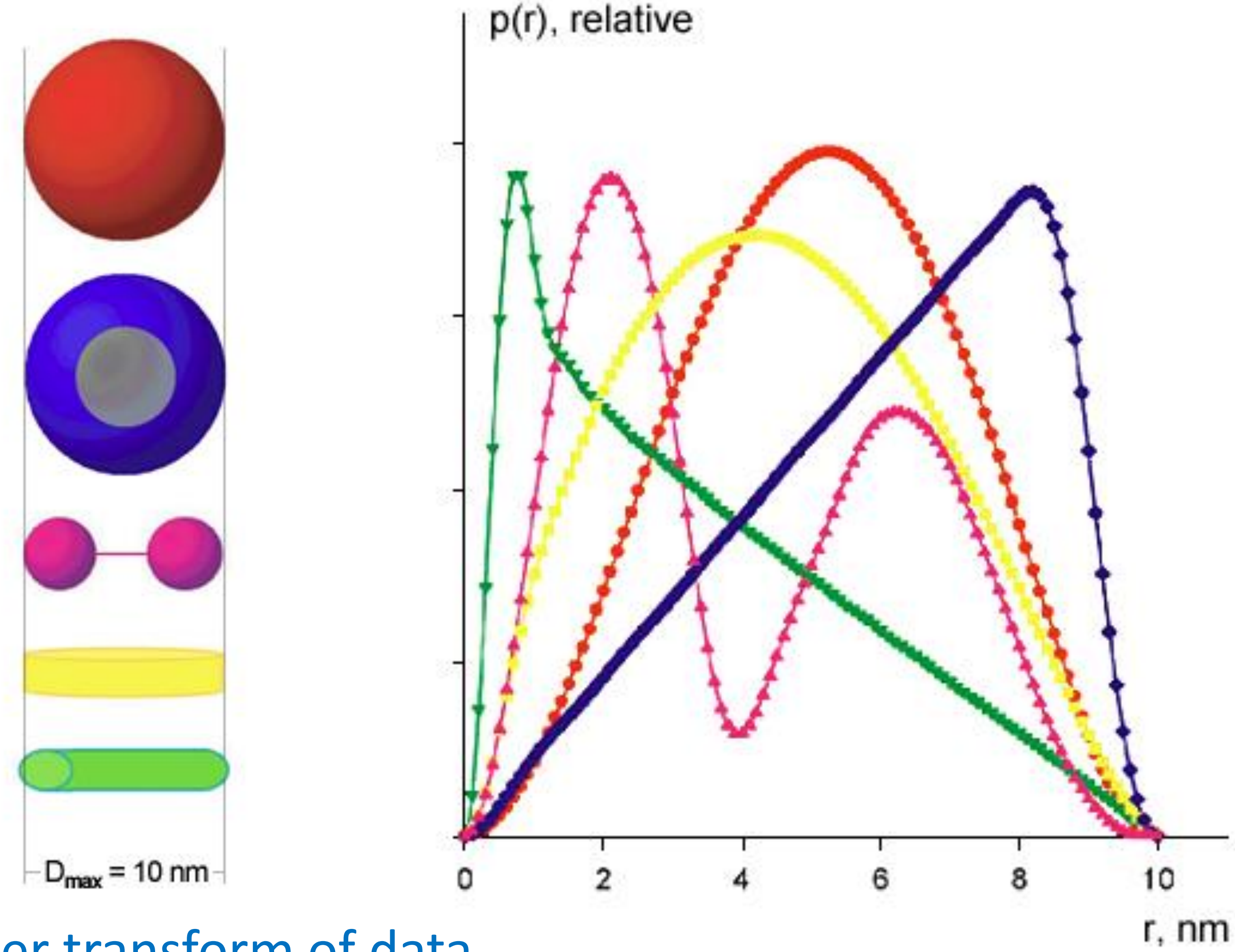

(ii) Radius of Gyration
(iii) Shape of Particle
(iv) Interface
(i)
Interparticle interference
Log(I)
q^{-4} (sphere)
q^{-2} (disk)
q^{-1} (thin rod)
Gunier region
Porod region
q^{-4}
0
1/r or 1/l
$1/D_{max}$
1/ε
q

Data



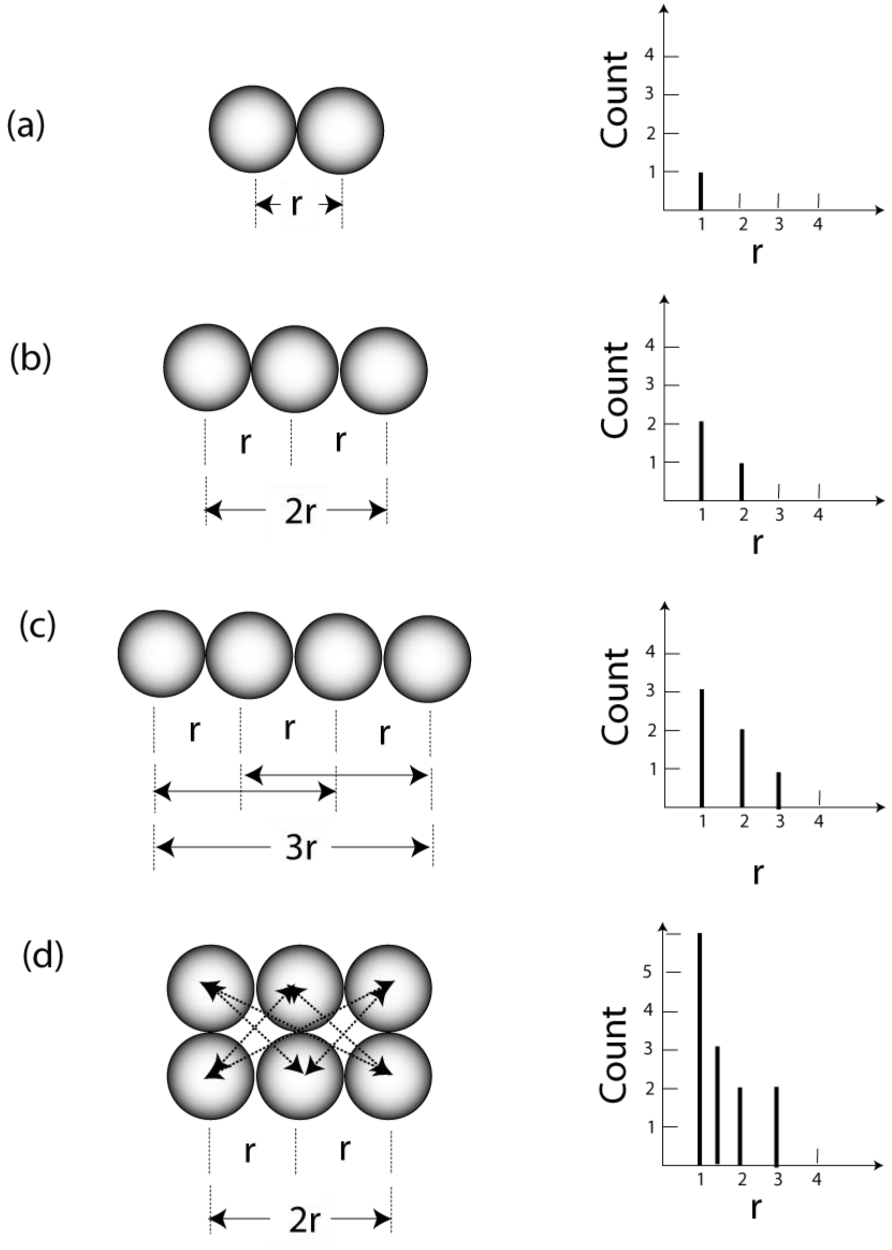
From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)

Pair distribution function

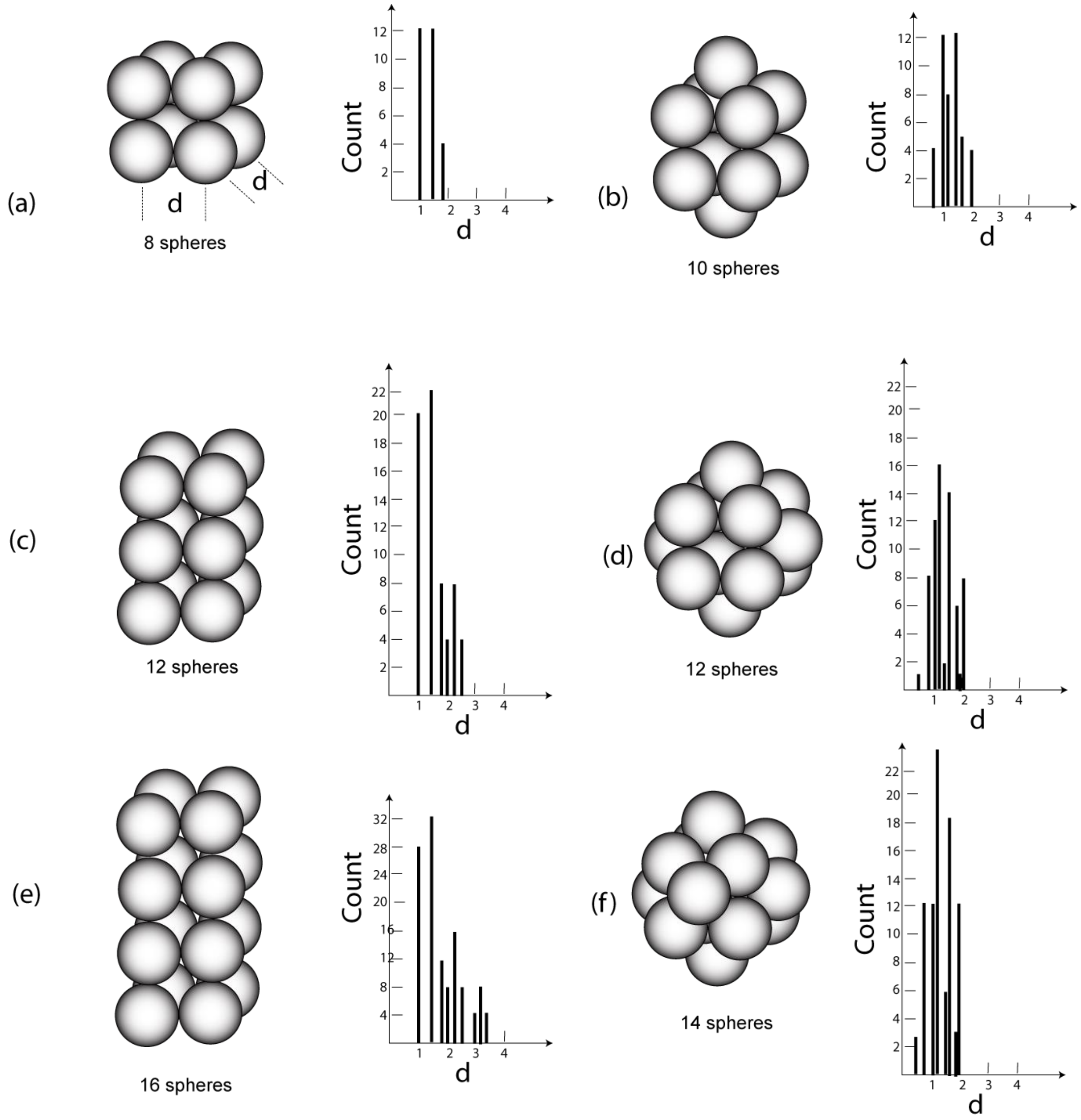


Fourier transform of data.

From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)



P(r) plot is simply the histogram of interatomic scattering



P(r) plot is simply the histogram of interatomic scattering

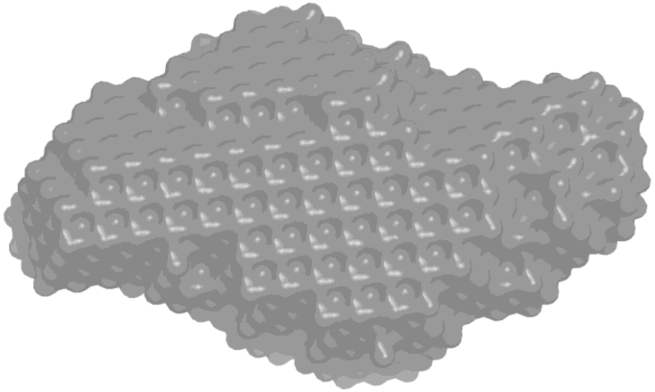
Larger compact molecules have a high distribution at lower angle (consider detector distance etc.)

SAXS can determine *ab initio* molecular envelopes

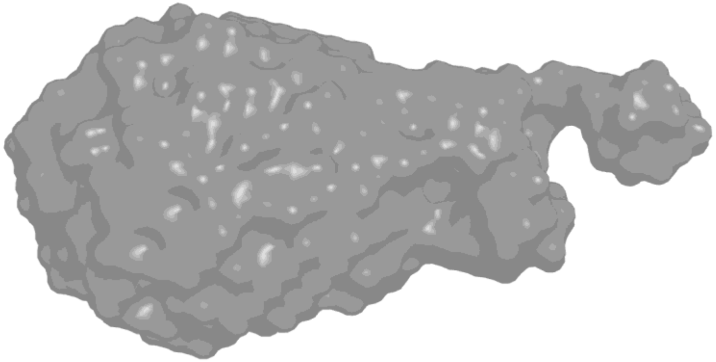
But keep in mind, it is possible to accurately predict scatting from a given model.

(many applications do not need an envelope to test a hypothesis)

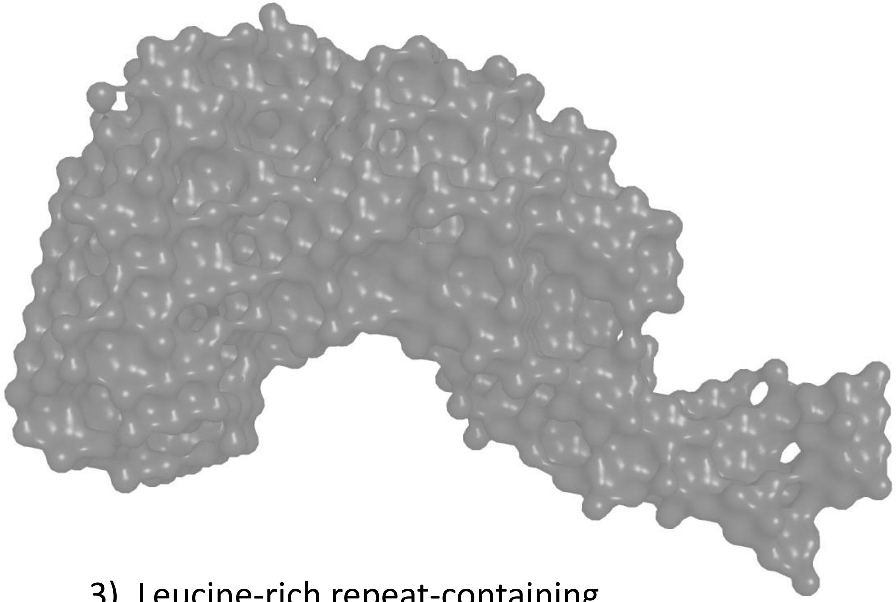
Ab intio envelopes



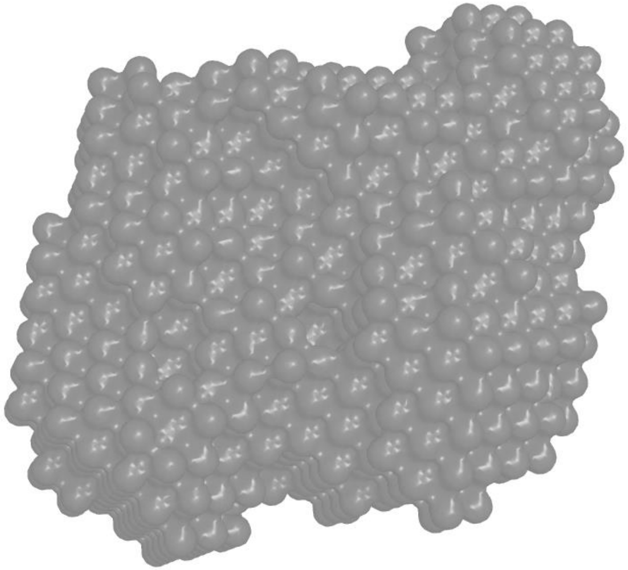
1). alr0221 protein from Nostoc (18.6 kDa)



2). C-terminal domain of a chitobiase (17.9 kDa)



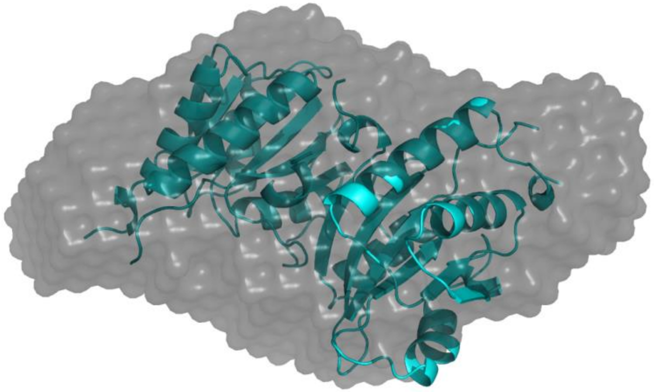
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



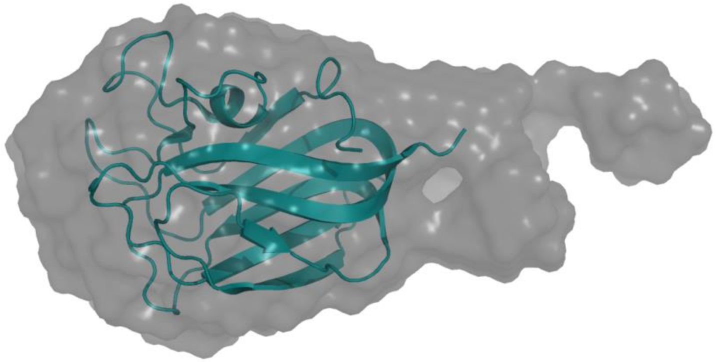
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

These are compatible with structural data

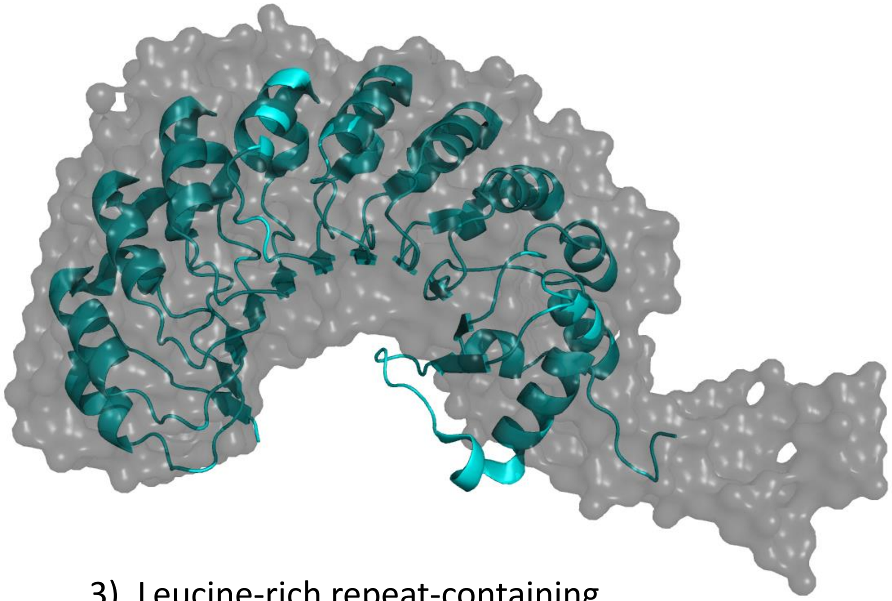
Overlaid with subsequent X-ray structures



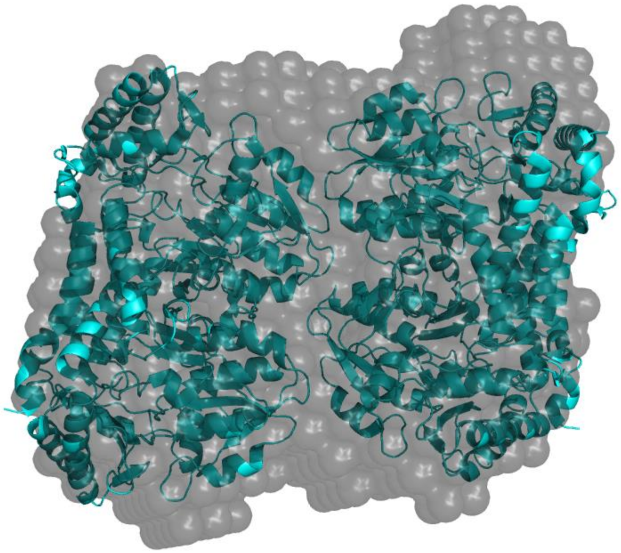
1). alr0221 protein from Nostoc (18.6 kDa)



2). C-terminal domain of a chitobiase (17.9 kDa)



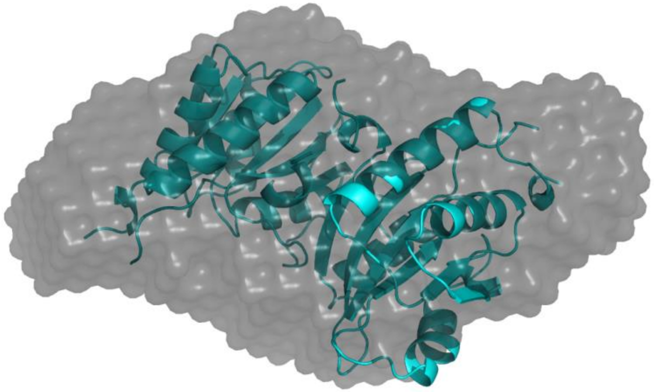
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



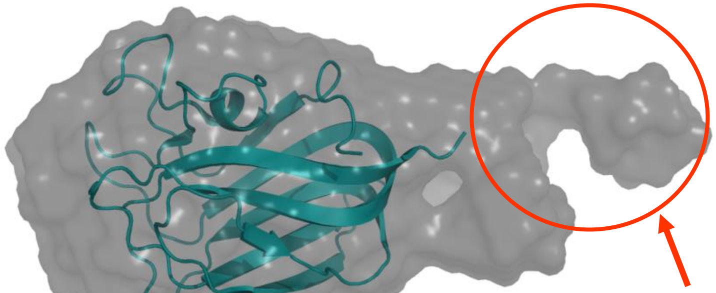
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

And provide extra information on residues present in the construct but structurally undefined

And data on what was missing …

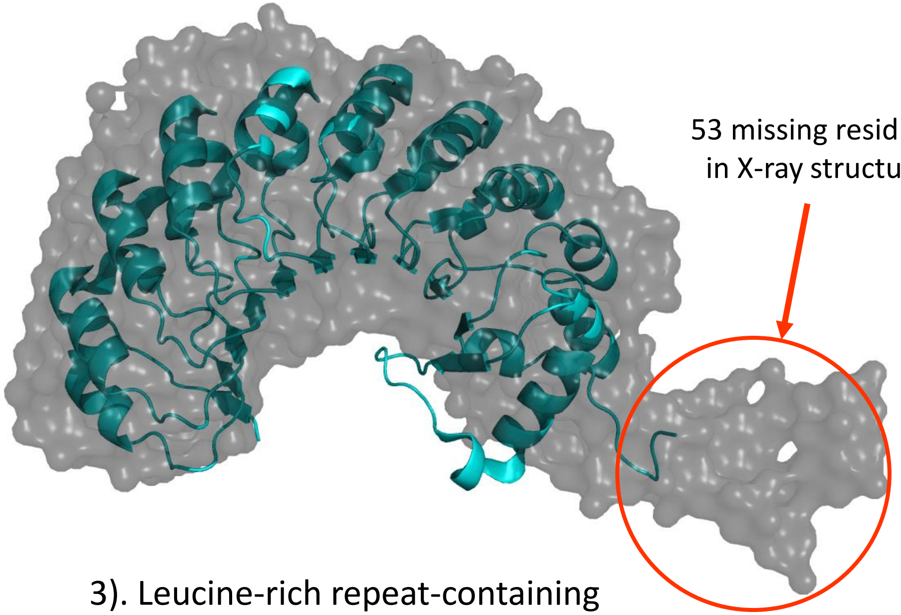


1). alr0221 protein from Nostoc (18.6 kDa)



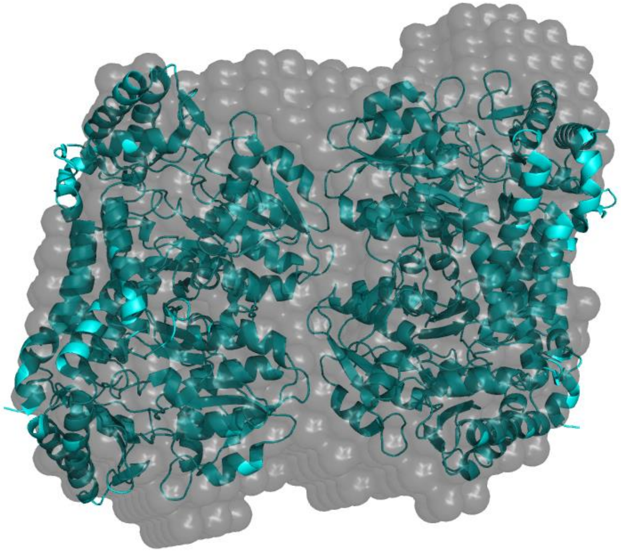
12 missing residues
in X-ray structure

2). C-terminal domain of a chitobiase (17.9 kDa)



53 missing residues
in X-ray structure

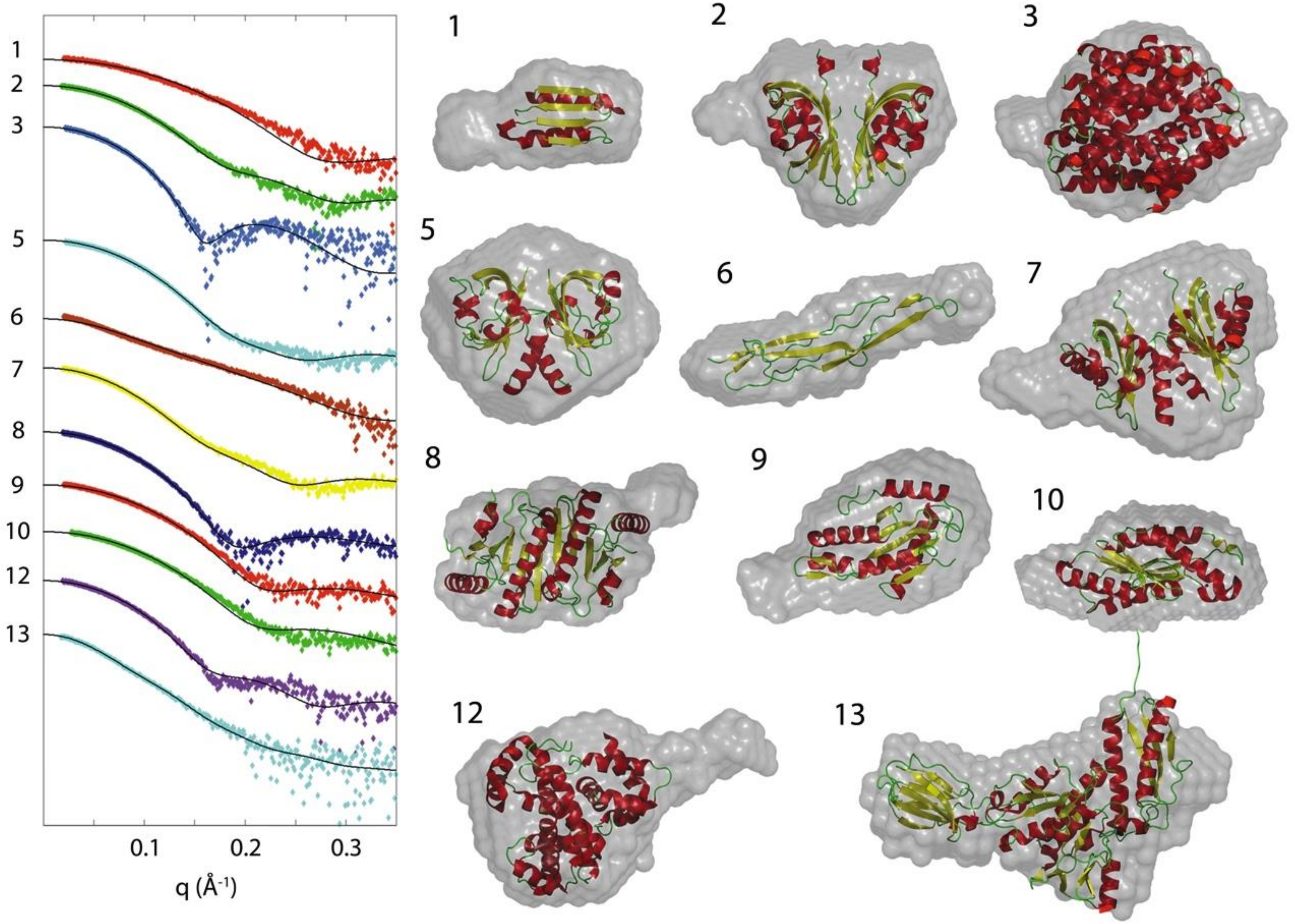
3). Leucine-rich repeat-containing
protein LegL7 (39 kDa)

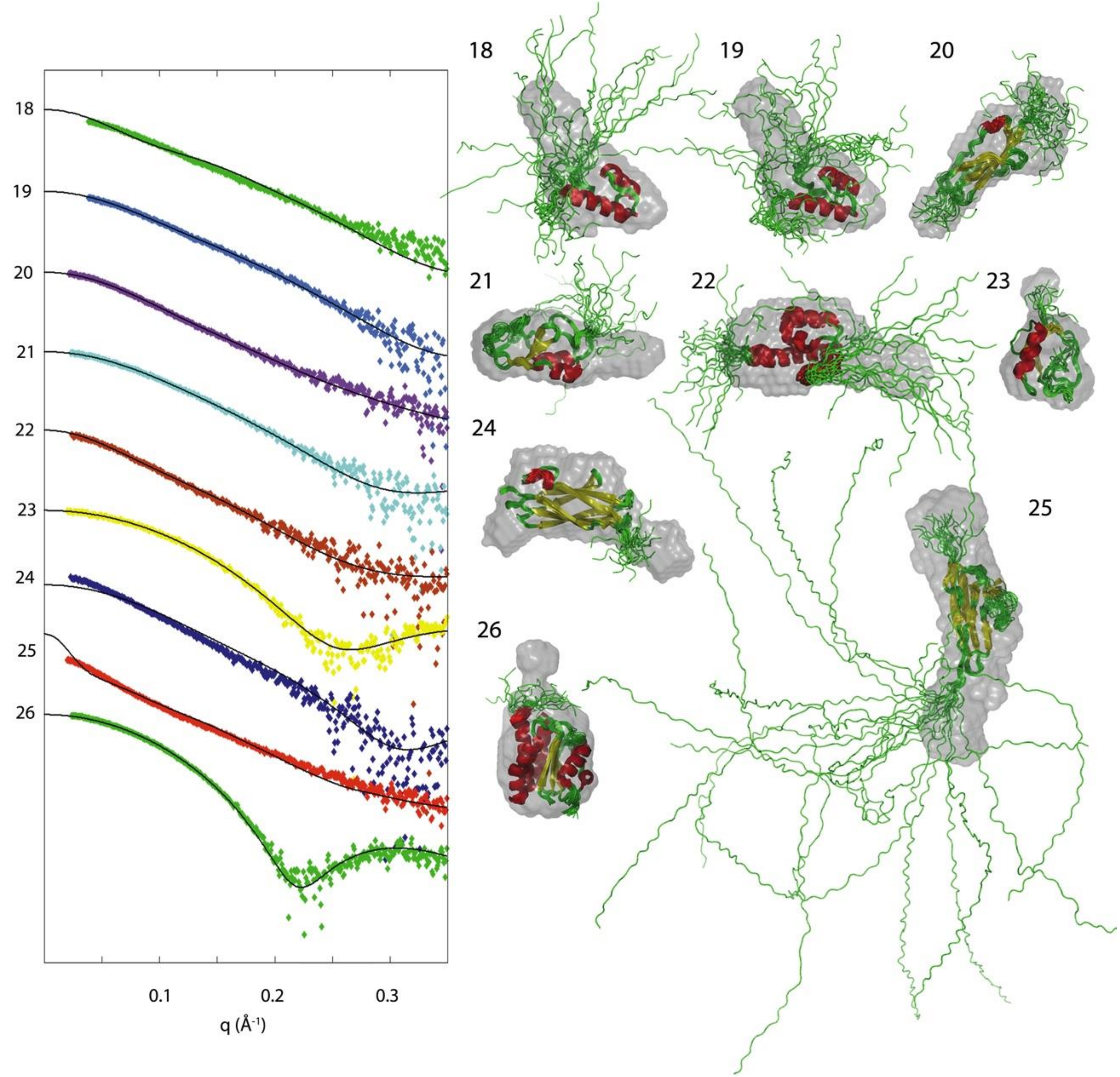


4). *E. Coli.* Cystine desulfurase
activator complex (170 kDa)

#	Name	NESG ID	PDB	Ref	State	Conc	MW	Res
Samples where crystallographic structures were available								
1	Domain of unknown function	DhR2A	3HZ7	16	M	6.9	9523	87
2	Diguanylate cyclase with PAS/PAC sensor	MqR66C	3H9W	17	D	8.2	13,611	210
3	Nmul_A1745 protein from *Nitrosospira multiformis*	NmR72	3LMF	18	T	6.9	14,069	484
4	Domain of unknown function	DhR85C	3MJQ	19	D	10.7	14,609	252
5	Sensory box/GGDEF family protein	SoR288B	3MFX	20	D	9.1	14,779	258
6	MucBP domain of the adhesion protein PEPE_0118	PtR41A	3LYY	21	M	9.5	14,300	131
7	Sensory box/GGDEF domain protein	CsR222B	3LYX	22	D	12.7	15,341	248
8	HIT family hydrolase	VfR176	3I24	23	D	11.0	17,089	298
9	EAL/GGDEF domain protein	McR174C	3ICL	24	M	5.0	18,738	171
10	Diguanylate cyclase	MqR89A	3IGN	25	M	7.5	20,256	177
11	Putative NADPH-quinone reductase	PtR24A	3HA2	26	D	9.5	20,509	354
12	MmoQ (response regulator)	McR175G	3LJX	27	M	8.8	32,032	288
13	Putative uncharacterized protein	DhR18	3HXL	28	M	9.6	48,519	446
Samples where multiple constructs and crystallographic structures were available								
14	Putative hydrogenase	PfR246A (78–226)	3LRX	29	D	11.4	17,701	316
15		PfR246A (83–218)	3LYU	30	D	8.4	16,321	284
16	Alr3790 protein	NsR437I	3HIX	31	M	5.3	11,760	105
17		NsR437H	3HIX	31	M	6.5	15,700	141
Samples where NMR structures were available								
18	MKL/myocardinlike protein 1	HR4547E	2KW9 (NMR)	32	D	10.4	8276	75
19	MKL/myocardinlike protein 1	HR4547E	2KVU (NMR)	33	D	10.4	8276	75
20	Putative peptidoglycan bound protein (LPXTG motif)	LmR64B	2KVZ (NMR)	34	M	5.0	9712	85
21	E3 ubiquitin-protein ligase Praja1	HR4710B	2L0B (NMR)	35	M/D	5.6	10,297	91
22	Transcription factor NF-E2 45 kDa subunit	HR4653B	2KZ5 (NMR)	36	M	10.0	10,623	91
23	YlbL protein	GtR34C	2KL1 (NMR)	37	M	11.0	10,661	94
24	Cell surface protein	MvR254A	2L0D (NMR)	38	Tri	5.9	12,385	114
25	Domain of unknown function	MaR143A	2KZW (NMR)	39	M	6.6	16,312	145
26	N-terminal domain of protein PG_0361 from *P. gingivalis*	PgR37A	2KW7 (NMR)	40	M	12.9	17,485	157
Samples where both crystallographic and NMR structures were available								
27	GTP pyrophosphokinase	CtR148A	2KO1 (NMR)	41	D	8.0	10,042	176
			3IBW	42	T	8.0	10,042	176
28	Lin0431 protein	LkR112	2KPP (NMR)	43	M/Hep	6.3	12,747	114
			3LD7	44	M	6.3	12,747	100

Comparing X-ray structures

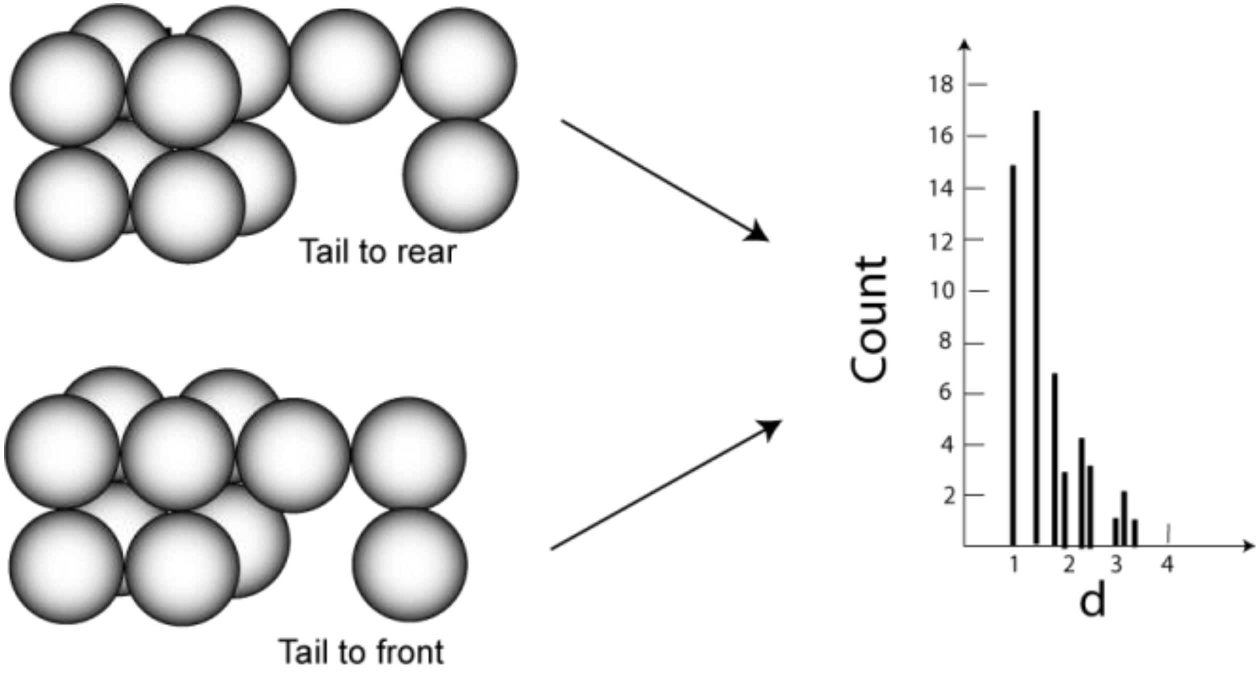




Comparing NMR structures

20 lowest energy Conformations shown

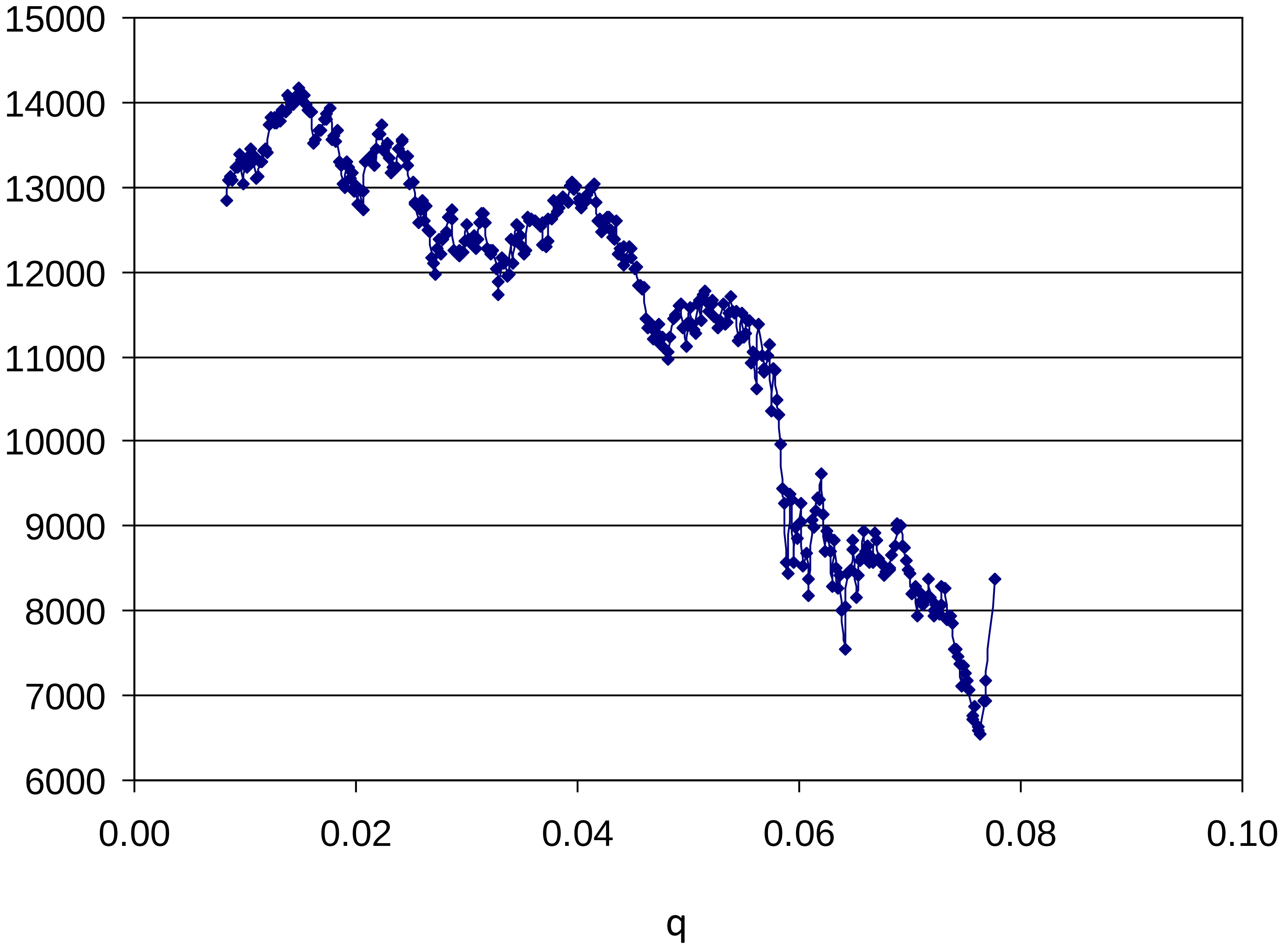
What can possibly go wrong?



Sometimes a unique reconstruction is not available.

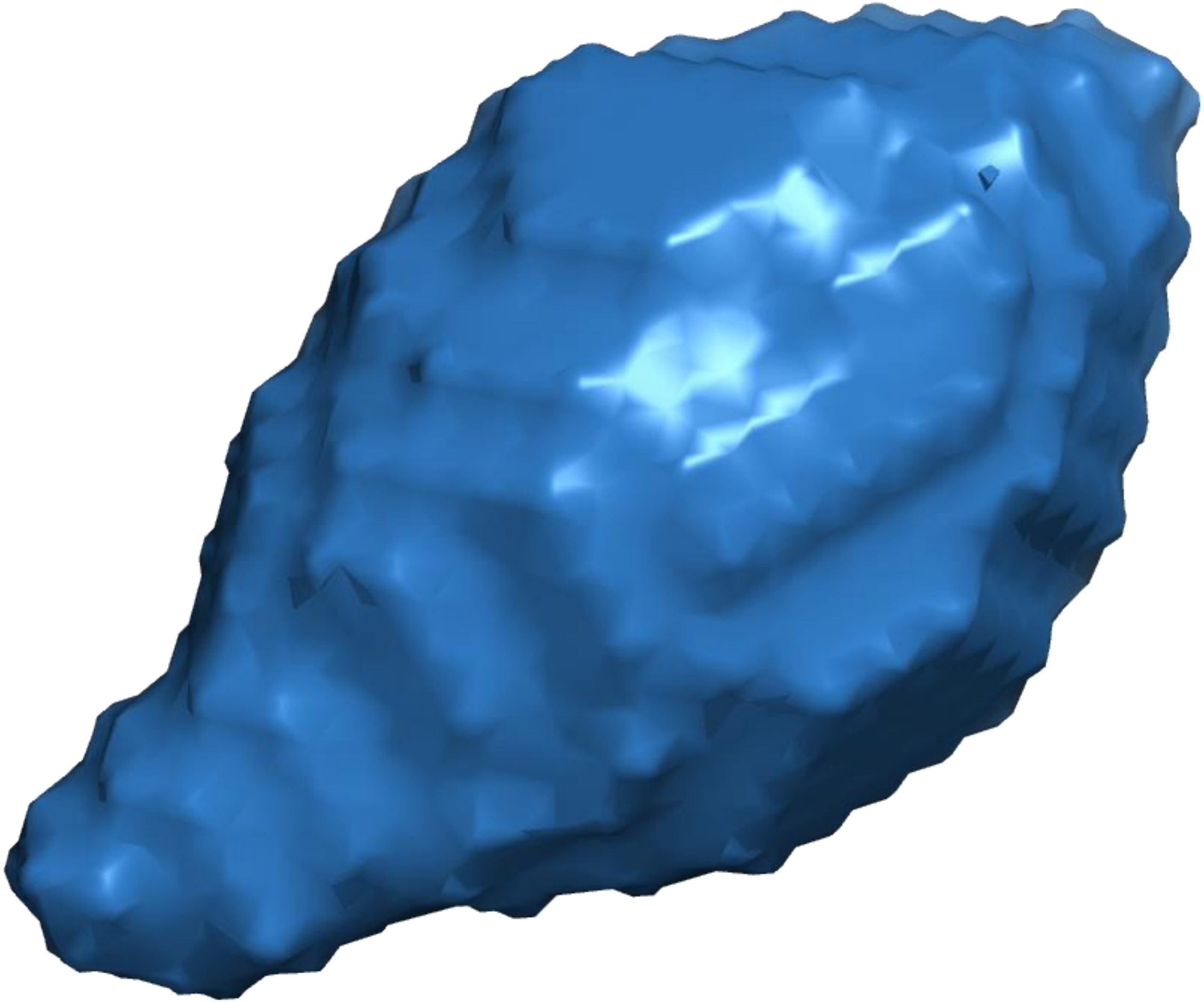
Garbage in, Garbage out

Lets take some ‘*scattering*’ data

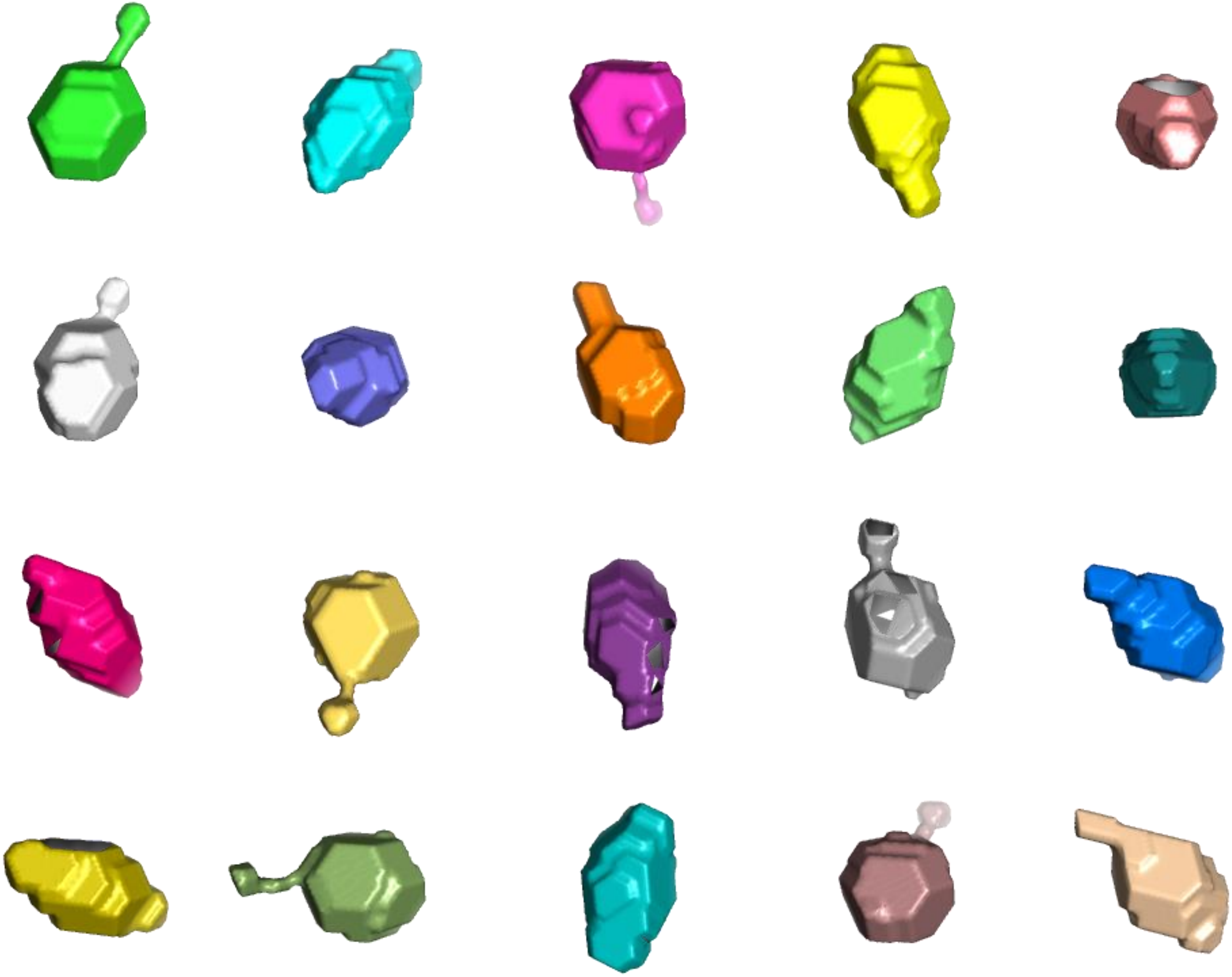


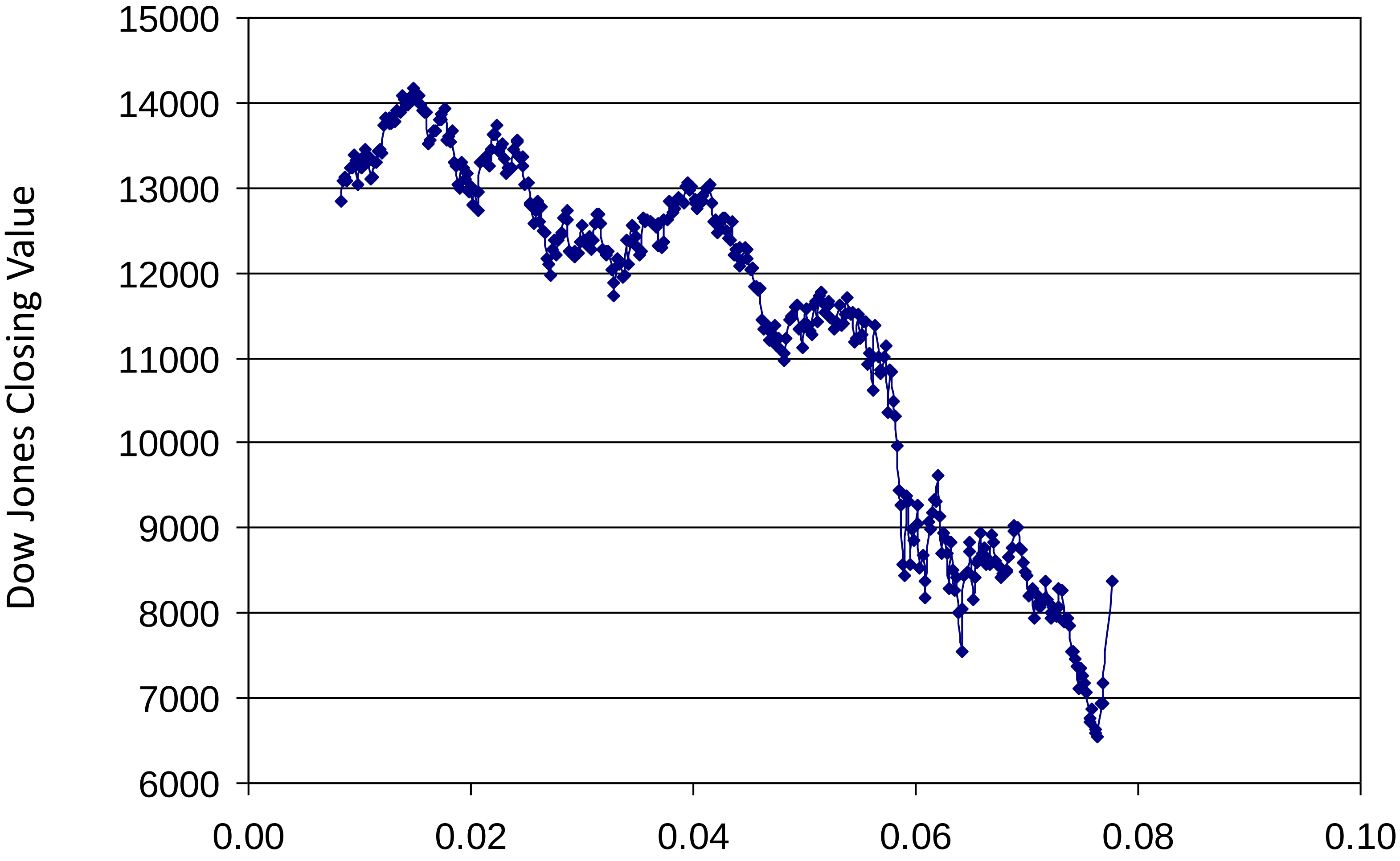
Envelope Reconstruction

- Produce 10-20 *ab initio* reconstructions
- Determine the most probable model, i.e. the least different from the rest and align all to this.
- Estimate the similarity of the models using the Normalized Spatial Discrepancy (NSD)
  - Average NSD ~ 0.5 implies good stability of solution
  - Average NSD ~ 0.7-0.9 implies fair stability
  - Average NSD > 1.0 implies poor stability.
- NSD can yield some idea of flexibility or possible oligomeric mixtures.
- DAMAVER can be used to select the most populated volume from all reconstructions



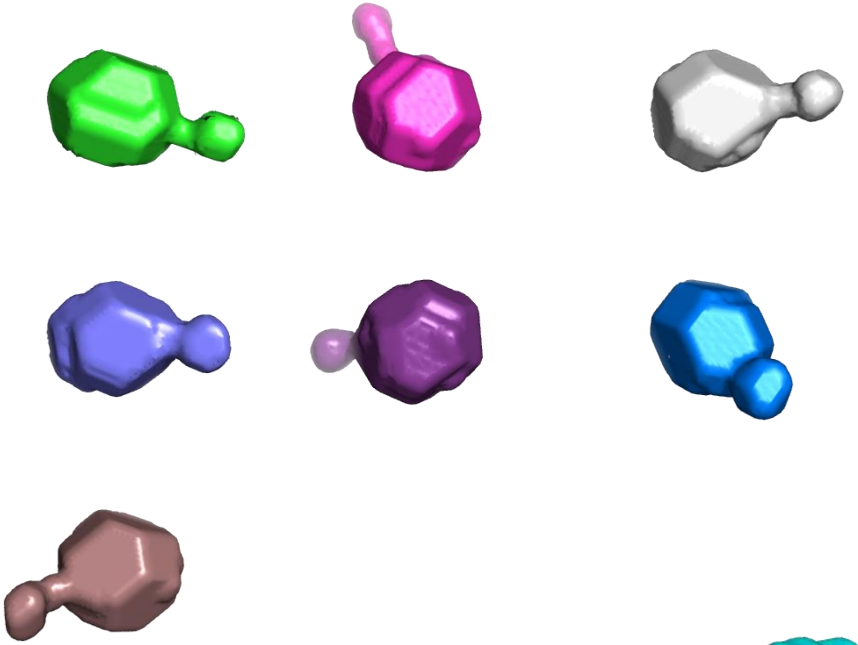
NSD = 0.613, 20 reconstructions





Aug 2007

March 2009

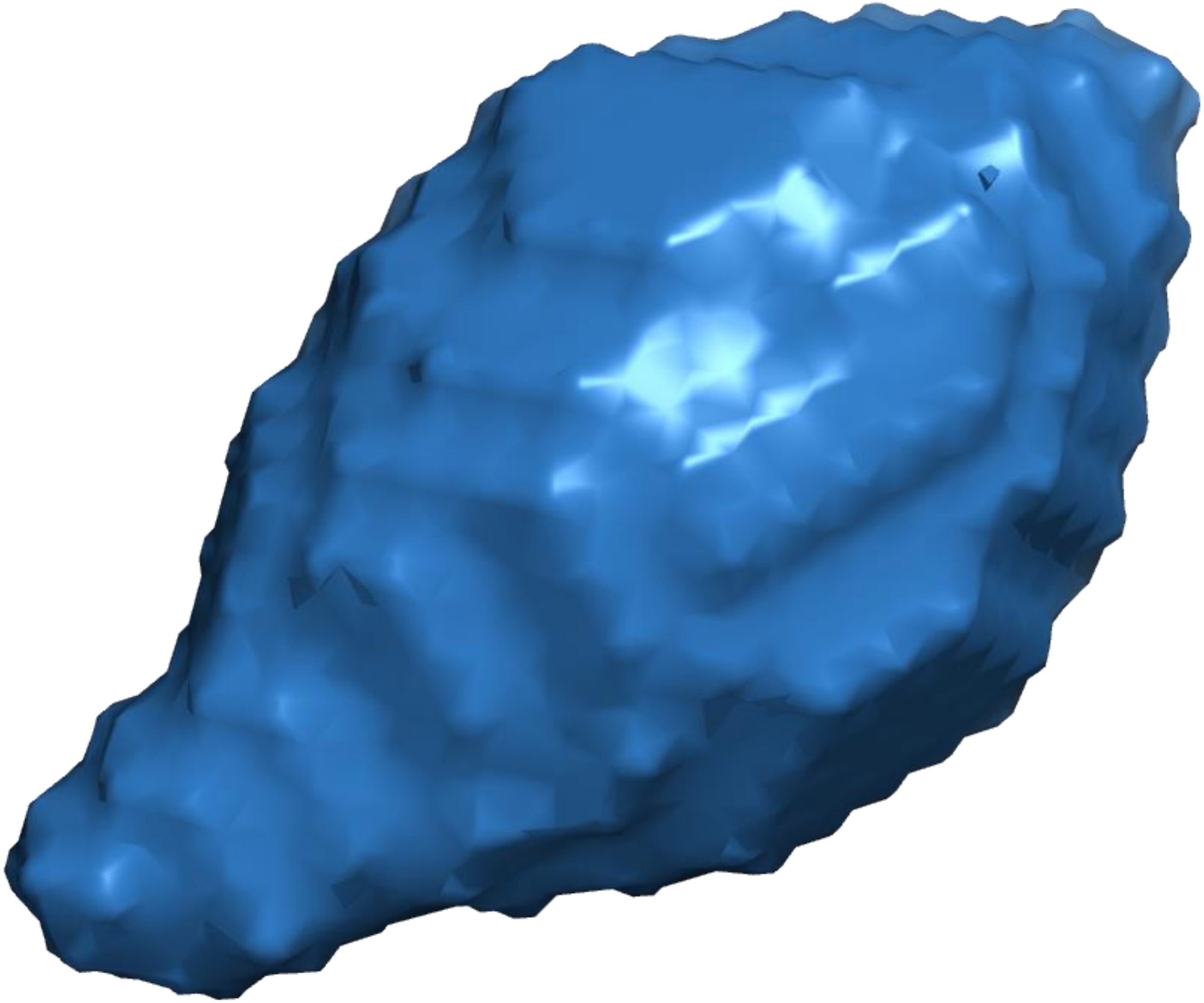


Actually two populations

Both are correct, i.e. they explain the scattering data

A Bull or a Bear market!

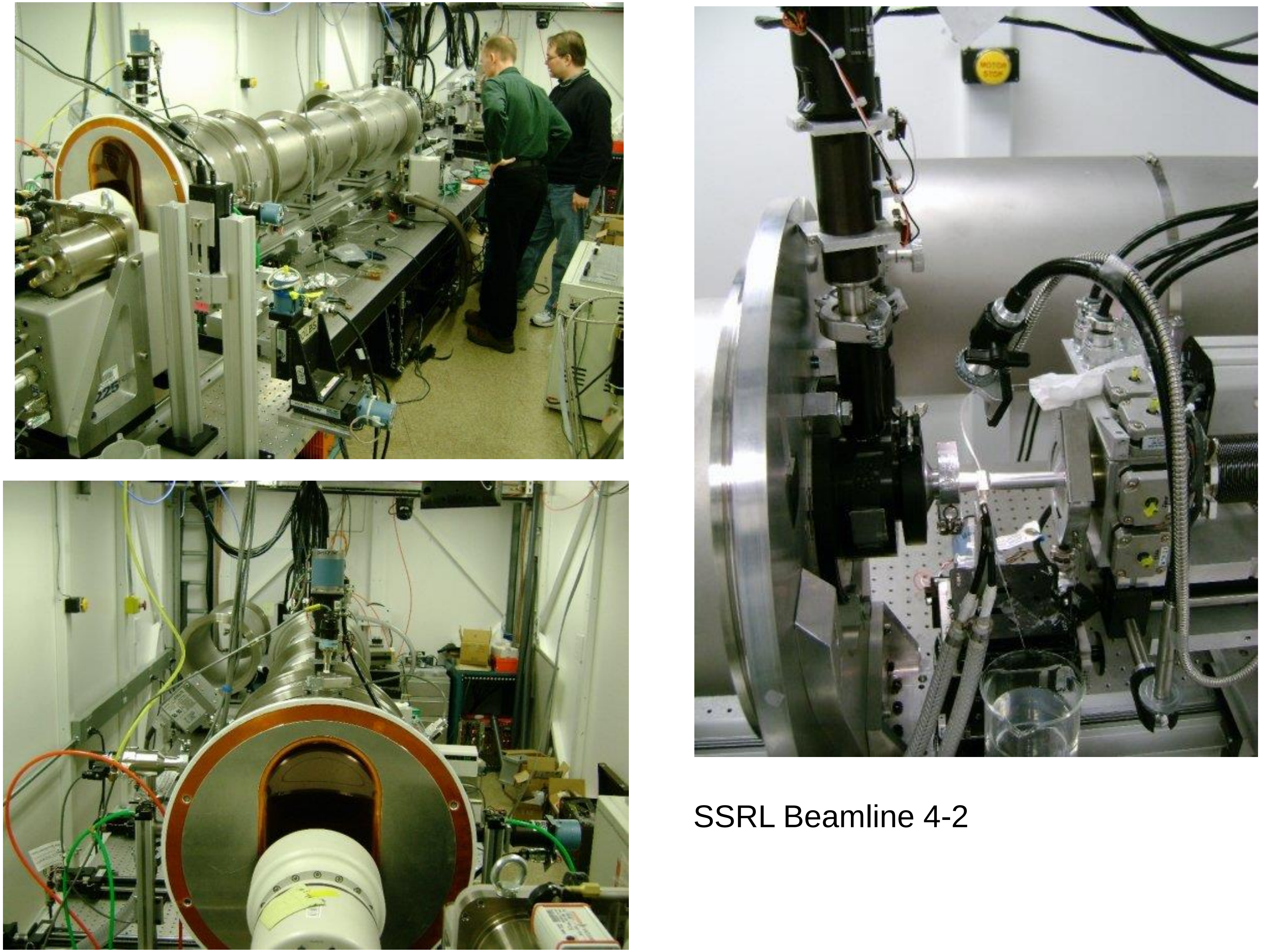
This is the molecular envelope of the recession, not a protein



NSD = 0.613, 20 reconstructions

An envelope can be calculated even if it’s not SAXS data

Now that you have been warned …
lets try high-throughput SAXS



SSRL Beamline 4-2

High throughput protocol

Up to 12 different PCR strips.

3-7 different concentrations per sample.

For high-throughput studies, 2 samples per strip, 24 samples in total

Start with buffer then lowest concentration first. End with buffer

8 exposures, 1-2s each dependent on sample molecular weight, buffer and concentration.

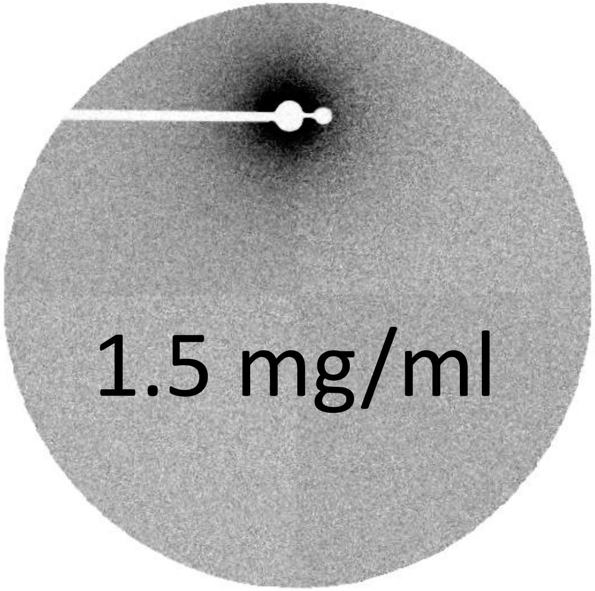
Oscillate sample to minimize radiation damage

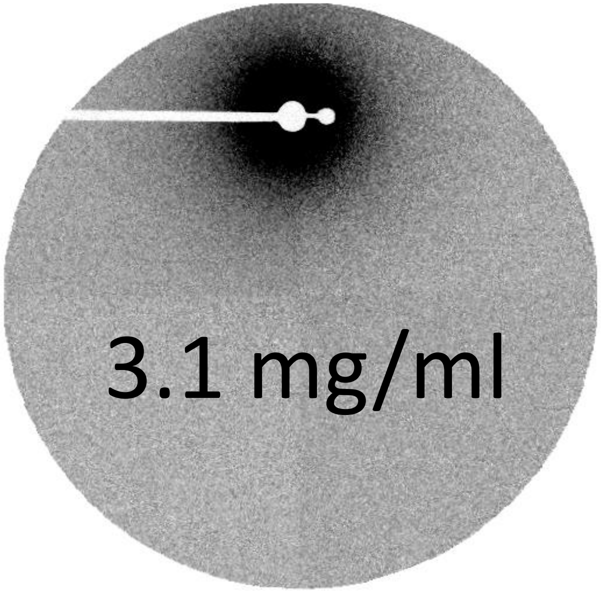
Repeat the buffer.

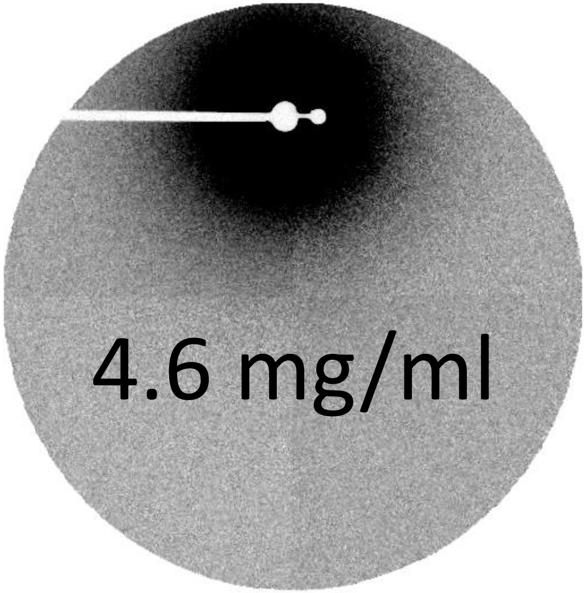
Load next sample

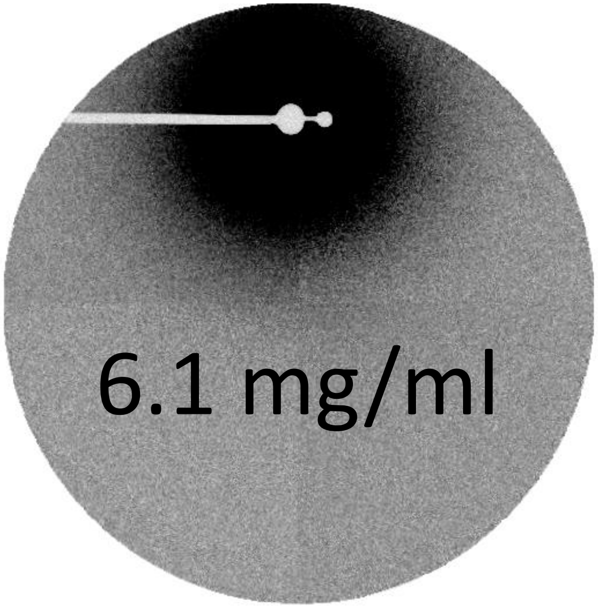
Time per concentration series – approximately 10 to 15 minutes. In high-throughput mode 24 samples in 3 to 4 hours.

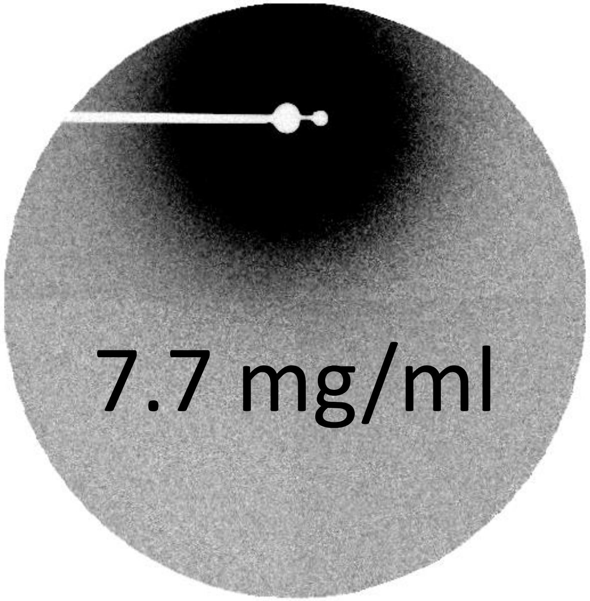
Enables two important things – eat and sleep!


1.5 mg/ml

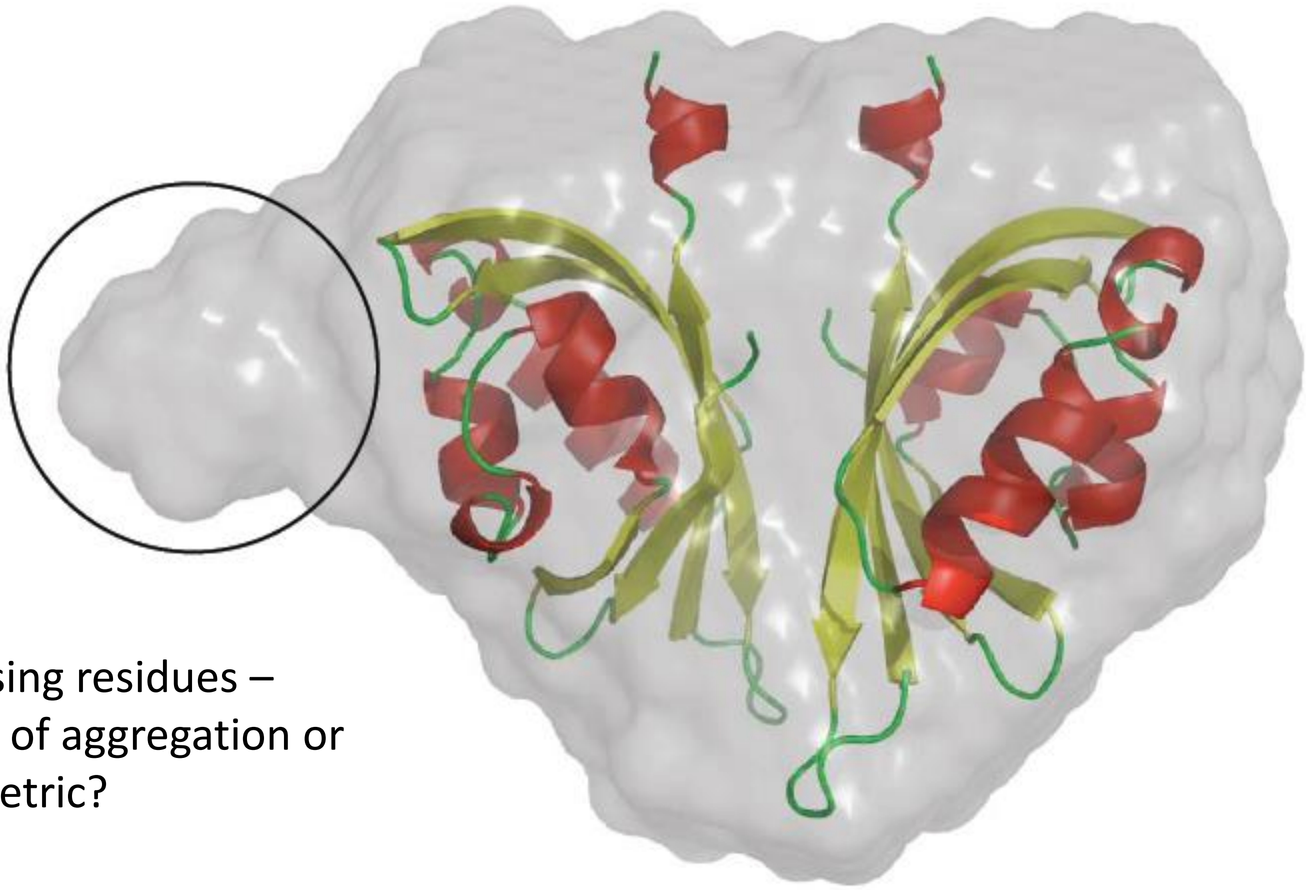

3.1 mg/ml


4.6 mg/ml


6.1 mg/ml


7.7 mg/ml

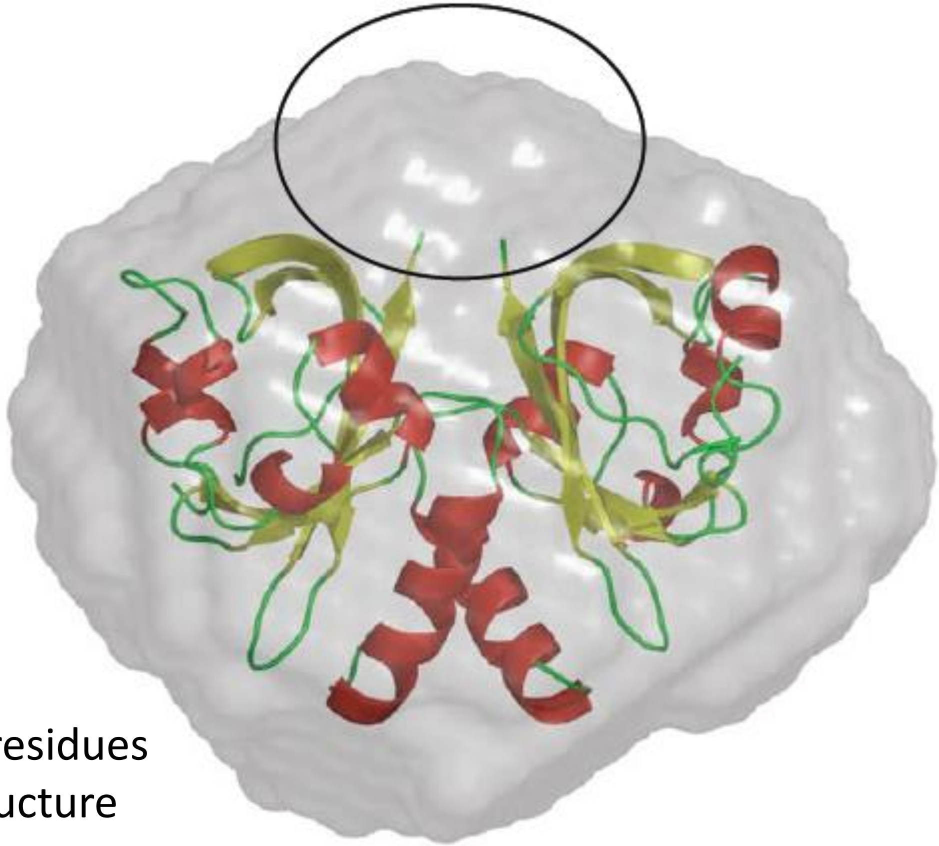
Diguanylate cyclase



12 missing residues – artifact of aggregation or asymmetric?

12 missing residues – artifact of aggregation or assymetric

Sensory Box/GGDEF Protein Family



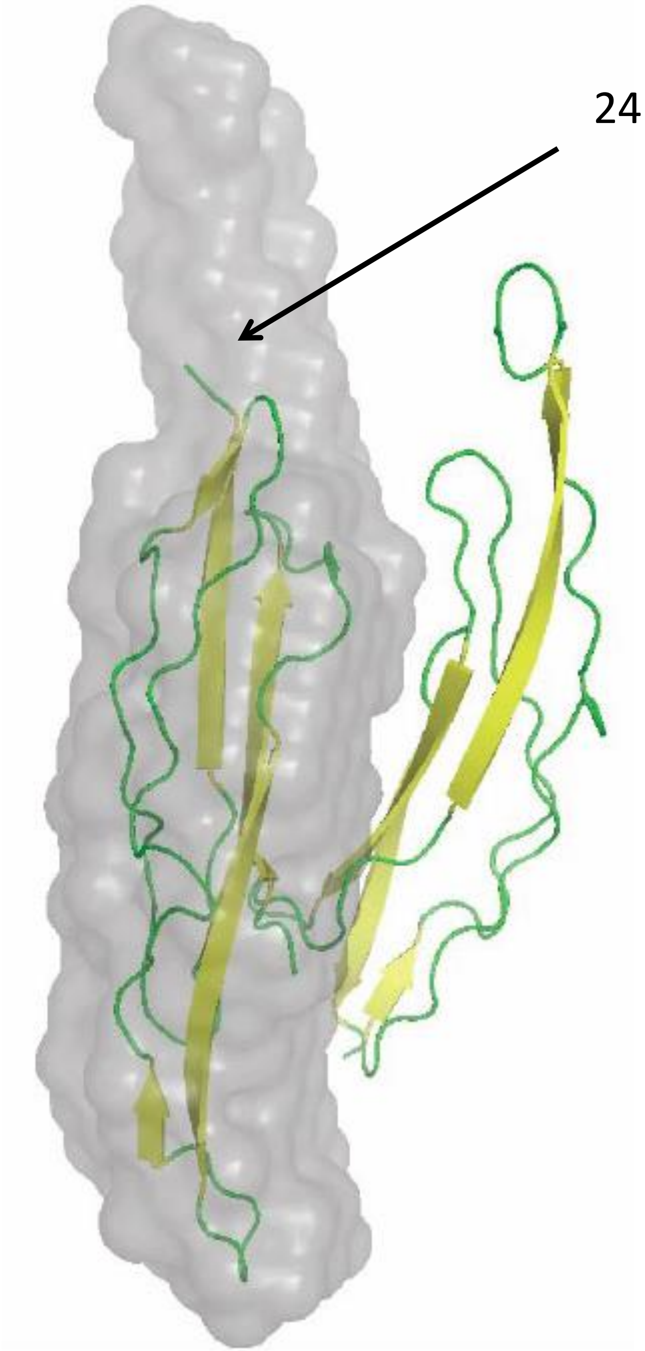
When a significant percentage of the residues are missing in a structure positioning within an envelope may be ambiguous – *a potato is a potato*.

MucBP Domain of PEPE_0118

Biological unit was thought to be a dimer from crystallography.

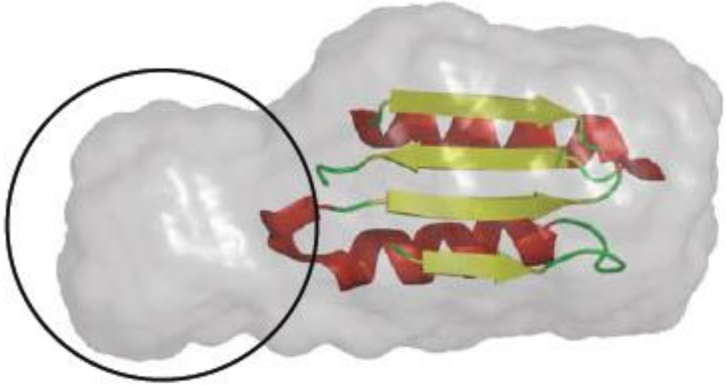
Solution state is not.

The biological state is not necessarily the solution or crystallographic state.

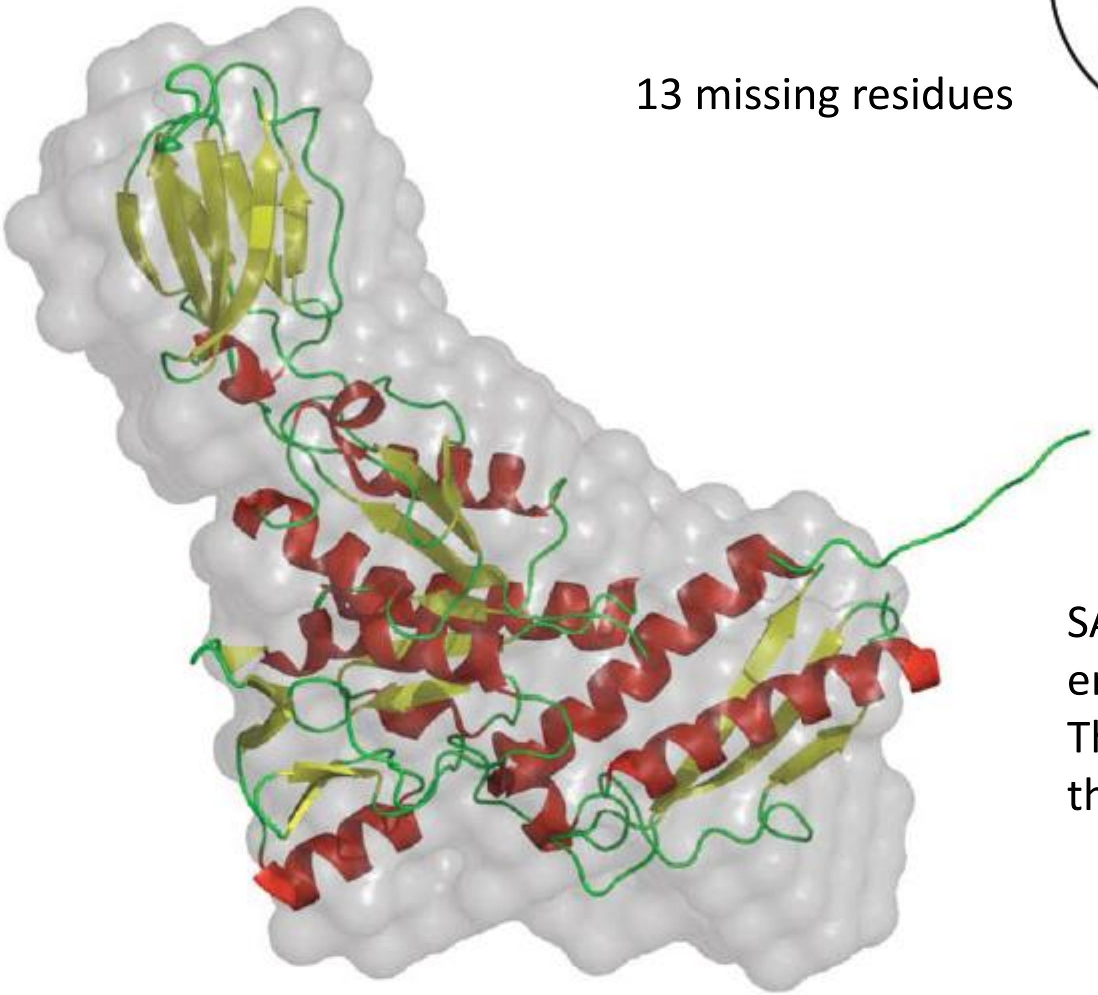


24 missing residues

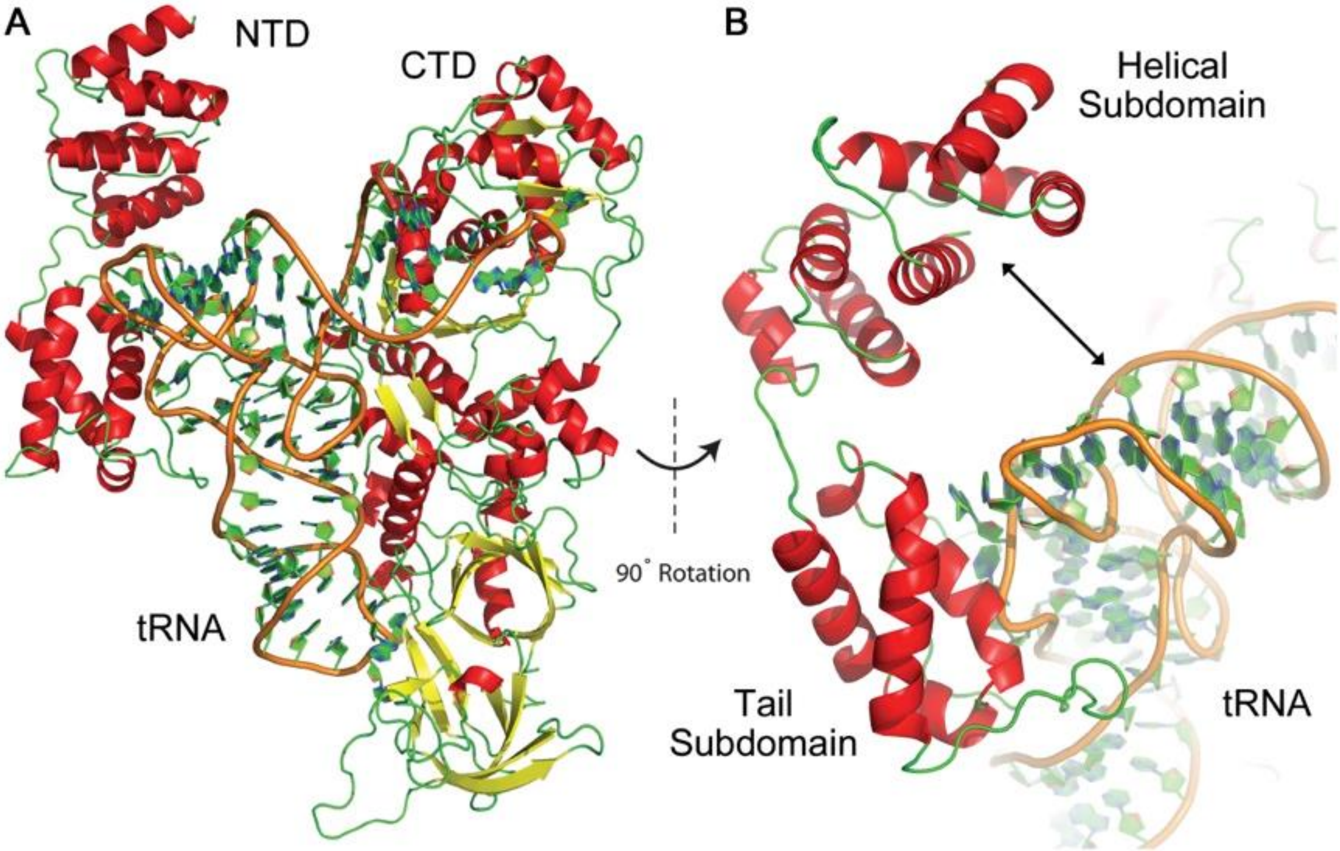
Size matters



13 missing residues



SAXS is not just about shape of the envelope but also it's overall size. The envelope produced reflects the size of the sample.



Homology Model of Full-length *Sc*GlnRS Bound to tRNA^{gln}. A. Full-length *Sc*GlnRS shown bound to tRNA^{gln}. B. Enlarged and rotated model showing gap between NTD helical subdomain and tRNA molecule.

What do we know?

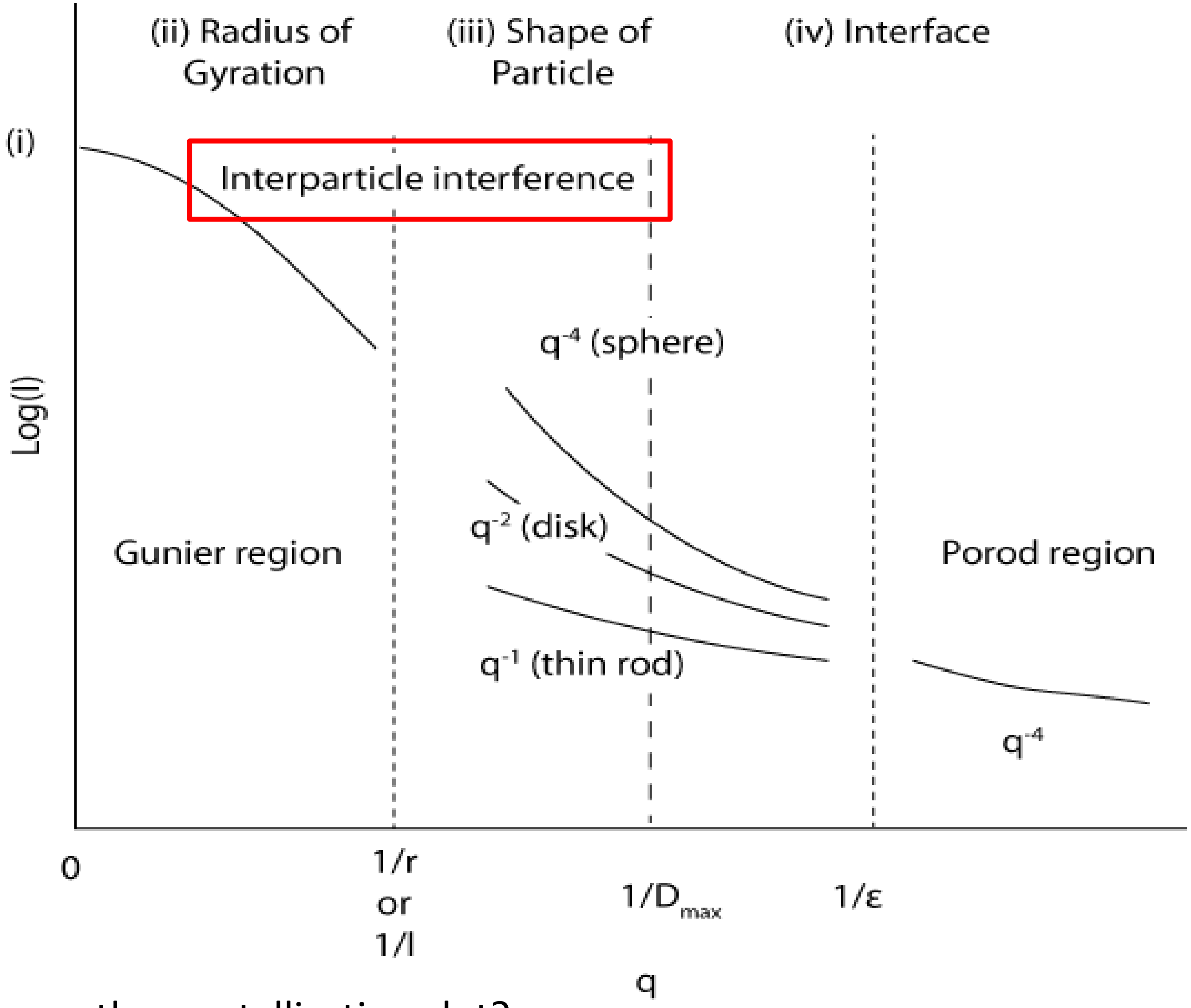
- SAXS characterizes the sample and can identify well folded samples from those that are natively unfolded.
- Similarly it can establish a degree of globularity and indicate how much disorder is present or if there may be multiple domains.
- It is sensitive to sample aggregation.
- It can produce a low resolution molecular envelope of the sample.
- Theoretically envelope is not a unique solution.

What we'd like to know?

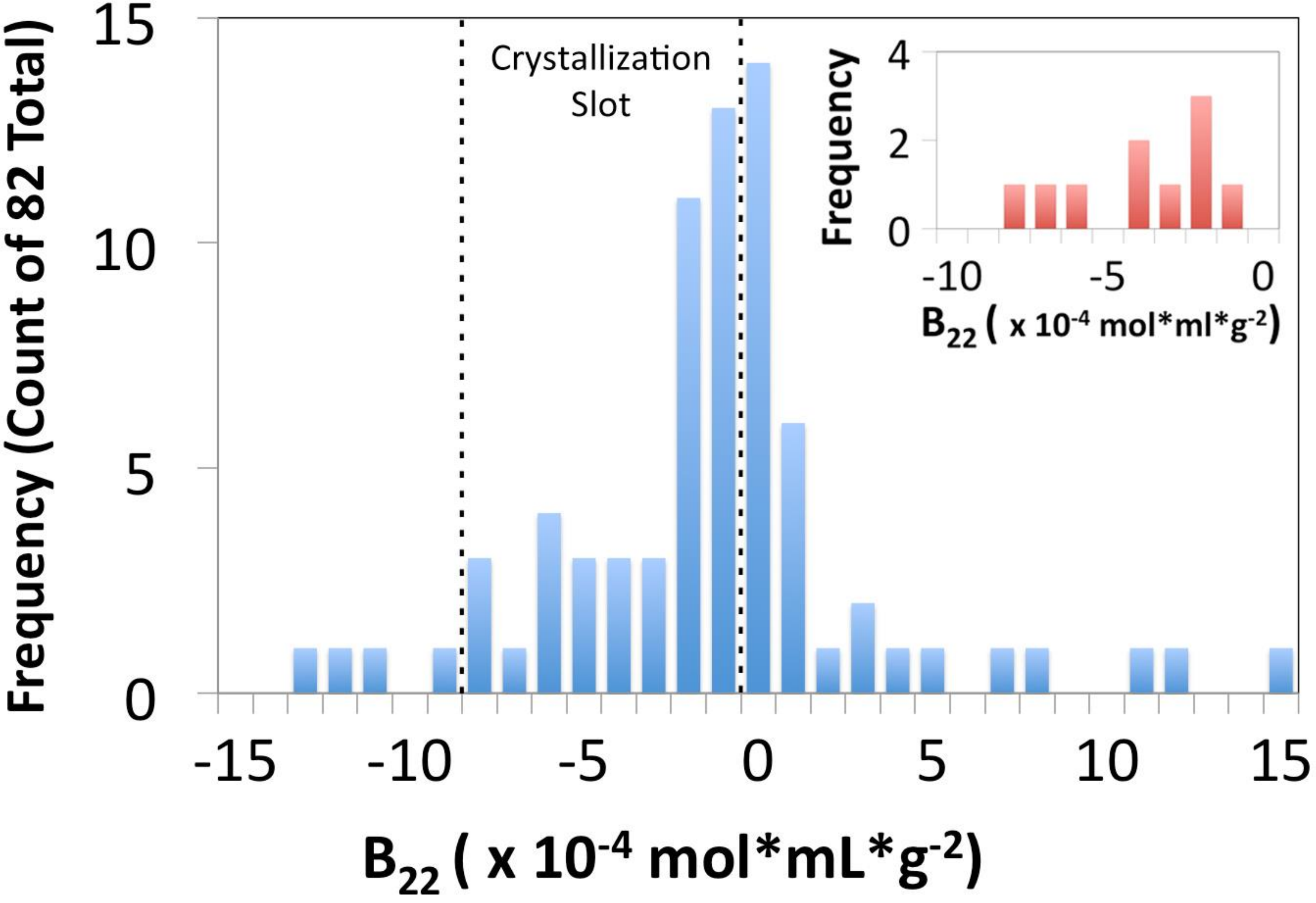
- How many of our samples that don't crystallize are 'bad'.
- How reliable is the molecular envelope – what degree of confidence can we put in it?

High-throughput is not useful for everyone but it has applications for some.

Why the interest in high-throughput for crystallization?



Can you measure the crystallization slot?


Frequency (Count of 82 Total)
15
10
5
0
Crystallization
Slot
-15
-10
-5
0
5
10
15
B$_{22}$ (x 10^{-4} mol*mL*g^{-2})
Frequency
4
2
0
-10
-5
0
B$_{22}$ (x 10^{-4} mol*ml*g^{-2})

Requirements for high-throughput data collection (creating an pipeline)

- High-throughput
  - Maximize number of samples
  - Minimize cleaning time
- Rapid analysis of data
  - The sample is monomodal
  - It does not aggregate
  - It does not repel
  - It is globular
  - It is stable
  - It does not suffer from radiation damage
- Rapid processing of data

Requirements for high-throughput data collection (creating a pipeline)

- High-throughput
  - Maximize number of samples
  - Minimize cleaning time
- Rapid analysis of data
  - The sample is monomodal
  - It does not aggregate
  - It does not repel
  - It is globular
  - It is stable
  - It does not suffer from radiation damage
- Rapid processing of data

Rapid analysis of data quality

- Radiation damage:
  - Ionizing radiation can cause biological macromolecules to form high molecular weight oligomers
  - These effects manifest themselves as changes in the Guinier plot, radius of gyration (R_g), maximum particle dimension (D_{max}), and forward scattering intensity (I(0)).
  - Compare changes in overall scattering profile, maximum particle dimension, Rg and I(0).
  - Collect series of short exposures (typically 10-20, 1s exposures) and compare them to see if statistically significant changes are occurring.
  - The t-statistic is used which describes the likelihood that a slope is significant i.e. that trends in SAXS parameters as a function of radiation are significant, and therefore indications of radiation damage are present.

Details in press, “The accurate assessment of small angle X-ray scattering data”, Thomas D. Grant, Joseph R. Luft, Lester G. Carter, Tsutomu Matsui, Thomas M. Weiss, Anne Martel, and Edward H. Snell, Acta Cryst D70, 2014

Rapid analysis of data quality

- Multiple Guinier regions:
  - Use the traditional region (1), one that takes account maximum particle dimension and potentially sparse sampling due to particle size (2) and finally, a region that allows for parasitic scatter and divergence in the beam if it does not exceed the first Shannon channel (i.e. does not affect the information content) (3).

Guinier Region 1

$$q < \frac{1.3}{R_g}$$

Guinier Region 2

$$\left[q_{min,G}, q_{max,G}\right] = \left[\frac{0.65}{R_g}, \frac{1.3}{R_g}\right]$$

Guinier Region 3

$$\left[q_{min,G}, q_{max,G}\right] = \left[\frac{\pi}{D_{max}}, \frac{1.3}{R_g}\right]$$

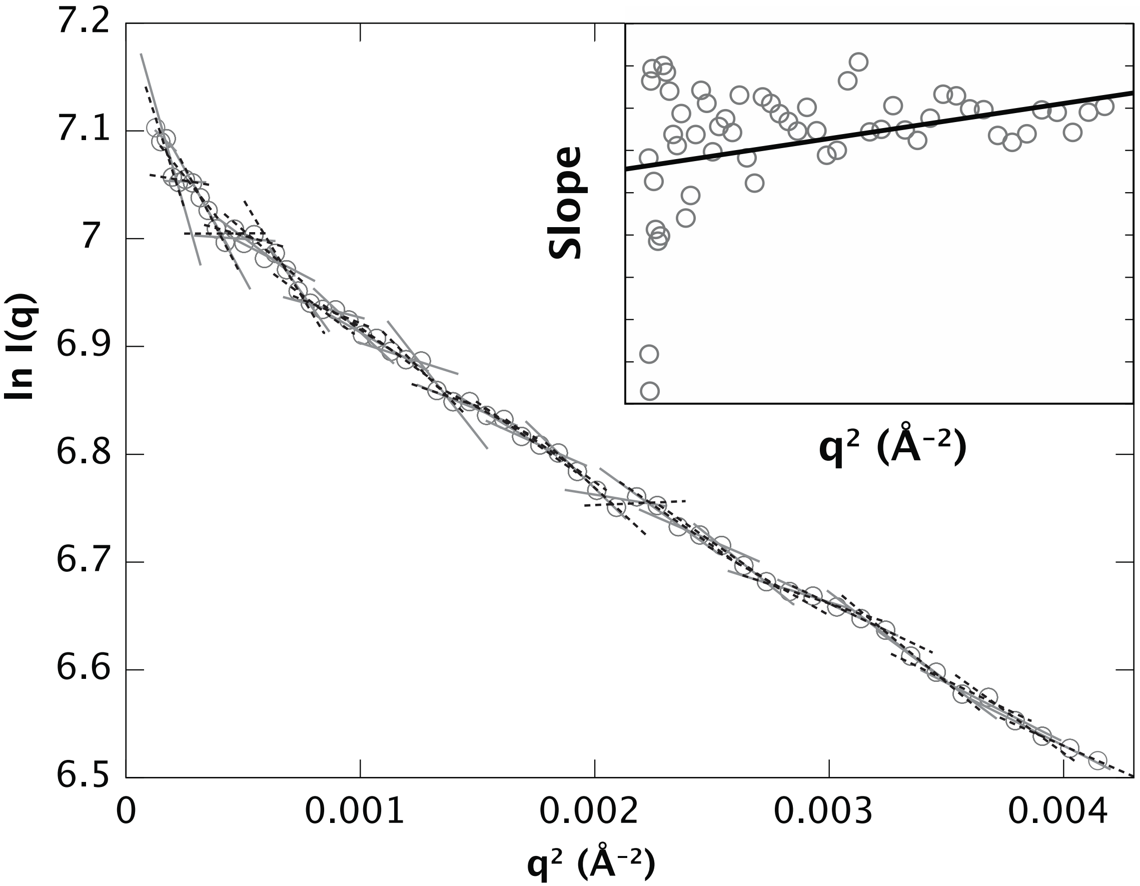
Rapid analysis of data quality

- Interparticle interactions:
  - Collect a minimum of three concentrations
  - Interactions also manifest themselves as changes in the Guinier plot, radius of gyration (R_g), maximum particle dimension (D_{max}), and forward scattering intensity (I(0)).
  - The data is scaled (non-trivial due to experimental errors in concentration and dilution).
  - Calculate concentration using previously collected standards.
  - Again the t-statistic is used which describes the likelihood that a slope is significant i.e. that trends in SAXS parameters as a function of radiation are significant, and therefore indications of radiation damage are present.

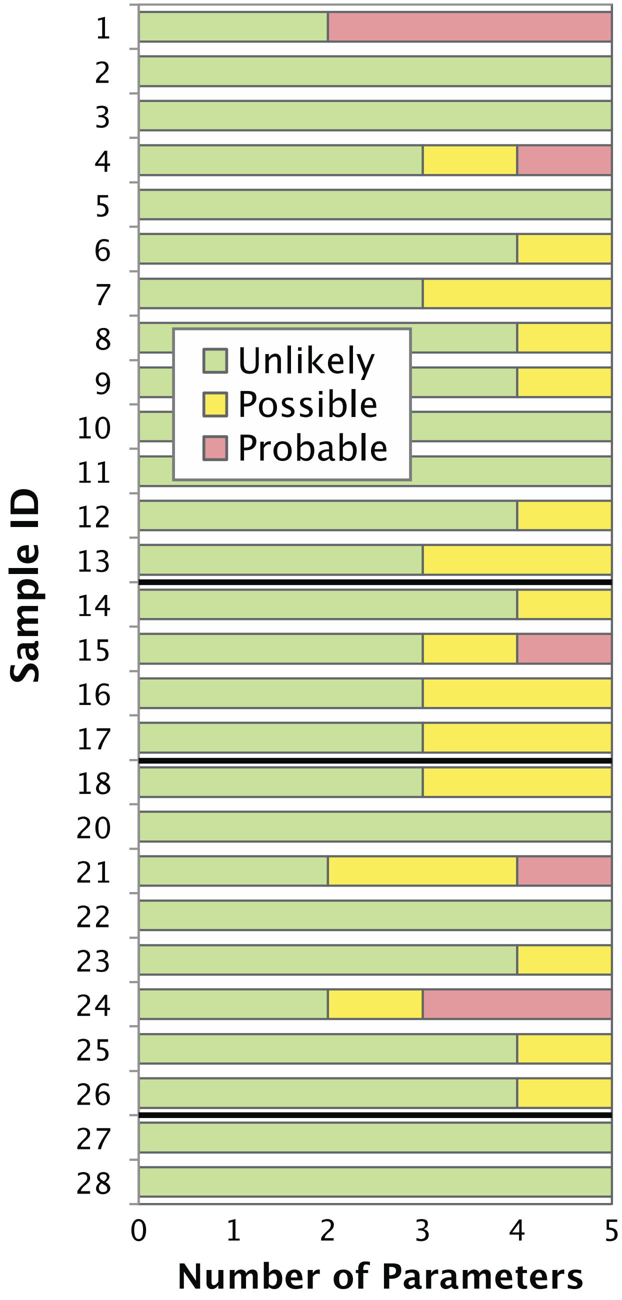
Rapid analysis of data quality

- Linearity in the Guinier region:
  - Analyze each concentration and each Gunier region
  - Apply least squares fit to each block of three data points and calculate slope, shift by one point and recalculate.
  - A linear regression is calculated through the set of slopes.
  - If linear the set of slopes should be constant.
  - A slope determines (a) if interactions are present and (b) if they are attractive or (c) repulsive.

  - Useful for determining the crystallization slot.



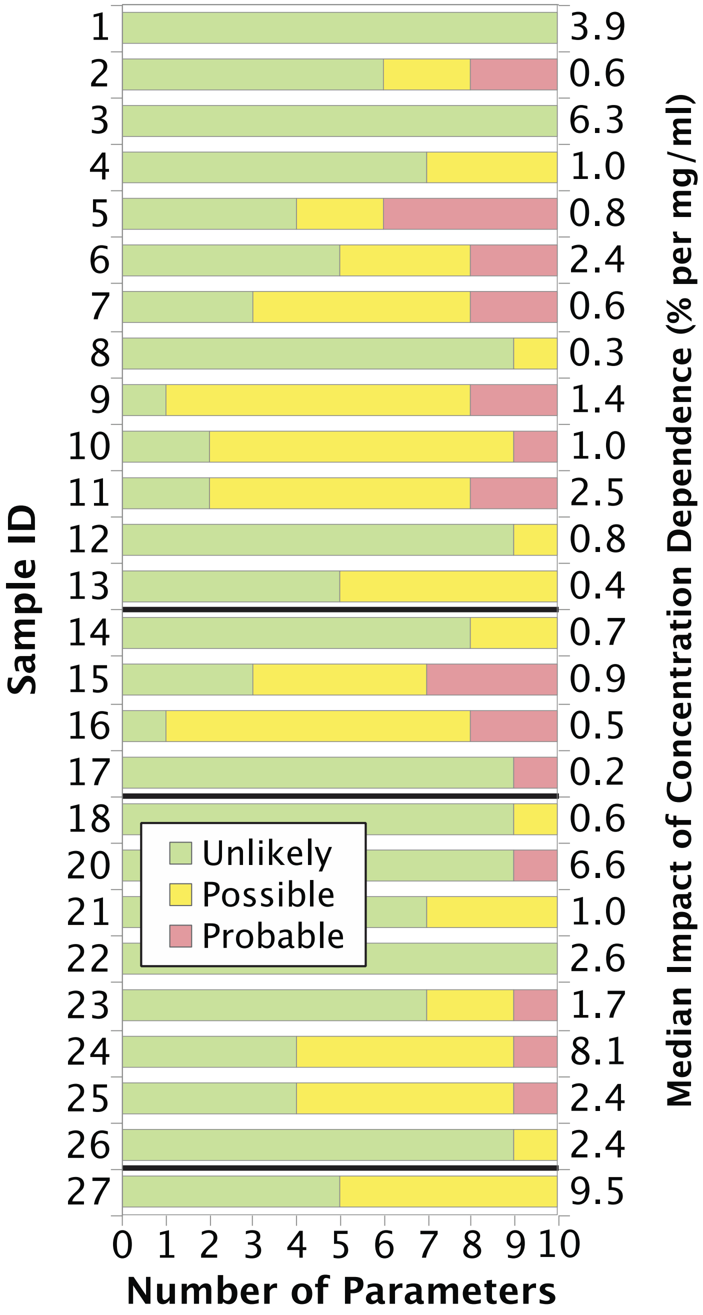
Detecting non-linearity in Guinier plots. A typical example of a Guinier plot for Guinier region 1 is shown. Data points are plotted as gray circles. The linear fit through each set of three data points is plotted with alternating solid gray and dashed black lines for clarity. A plot of the slope of each fit is shown in the inset. The set of slopes is fit with a linear regression, shown by the solid black line. Guinier regions that are linear will show a flat line with no dependence on q^2. Guinier regions that are non-linear will exhibit a dependence on q^2, detected using the t-statistic



A correlation Frequency plot is used to describe graphically the information Here, the sample ID is on the vertical axis, while the number of parameters with a given p-value is shown on the horizontal axis. The likelihood of a correlation being present is determined by the p-value, which is identified by color as unlikely (green, $p > 0.20$), possible (yellow, $0.05 < p \leq 0.20$), or probable (red, $p \leq 0.05$).

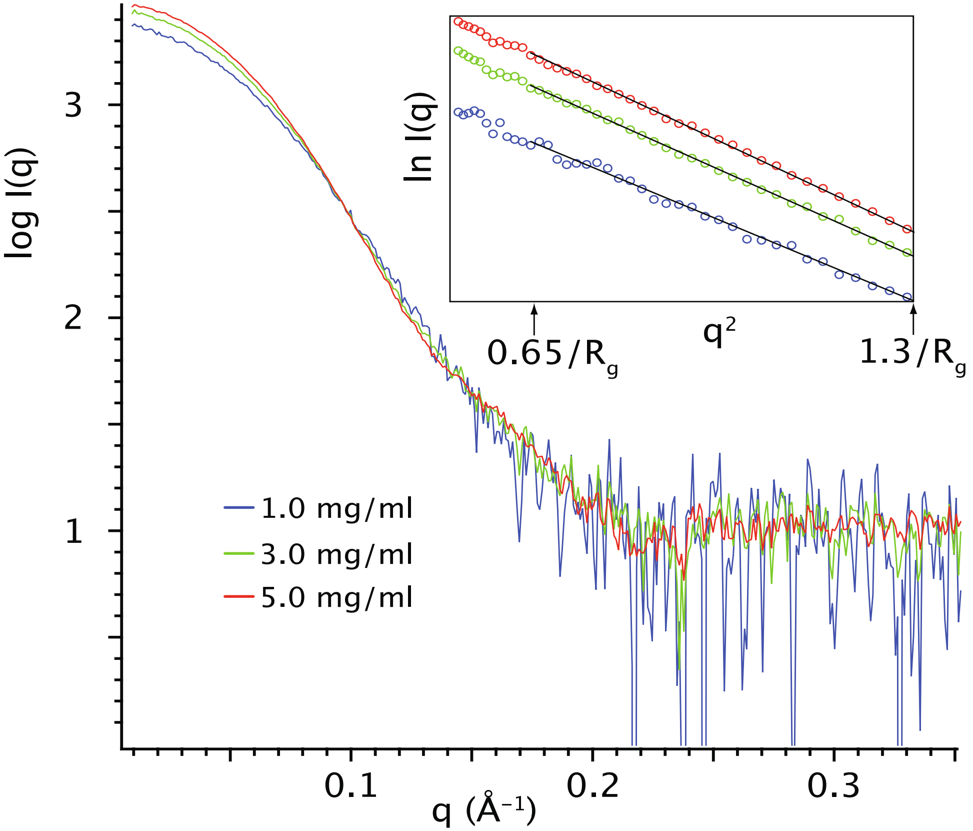
The plot shows Radiation Damage Analysis for the Highest Concentration of Each Sample. The number of SAXS parameters (out of 5 total) that were unlikely (green, $p > 0.20$), possibly (yellow, $0.05 < p \leq 0.20$), or probably (red, $p \leq 0.05$) affected by radiation damage is shown. Any exposures that were affected by radiation damage ($p < 0.05$) in any of the five parameters analyzed were rejected from averaging

Analysis of multiple images from one same sample



Correlation Frequency Plot for Concentration Dependence Analysis. The number of SAXS parameters (out of 10 total) that were unlikely (green, $p > 0.20$), possibly (yellow, $0.05 < p \leq 0.20$), or probably (red, $p \leq 0.05$) affected by concentration dependence is shown. For each sample, the absolute value of the slope of the linear regression for each of the ten parameters has been calculated as a percentage of the y-intercept of the regression. The median of these values is shown to the right of the chart to describe the typical impact that the concentration dependence has on the determination of SAXS parameters for each sample.

Analysis of multiple concentrations from one same sample



Concentration dependence detected for sample 11. Scattering profiles for the lowest (blue), middle (green), and highest (red) concentrations are shown after scaling. The increase in slope and intercept of the data at the low-q region as a function of concentration reflect an increase in the size of the particle. Inset: Guinier plots for each of the three concentrations. For clarity, only the linear fits to points in Guinier region 2 are shown by black solid lines. The upper and lower limits of Guinier region 2 are noted by black arrows and labeled

	Low Conc			Mid Conc			High Conc		
#	G1	G2	G3	G1	G2	G3	G1	G2	G3
1	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
2	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
3	🟩	🟩	🟩	🟥 +	🟨 -	🟩	🟩	🟩	🟩
4	🟥 +	🟩	🟩	🟩	🟩	🟨 +	🟩	🟩	🟩
5	🟥 +	🟩	🟩	🟩	🟨 -	🟩	🟥 +	🟨 -	🟥 -
6	🟥 +	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
7	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
8	🟨 +	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
9	🟥 +	🟩	🟩	🟨 +	🟨 +	🟩	🟥 +	🟩	🟩
10	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
11	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨 -
12	🟩	🟨 +	🟩	🟩	🟩	🟩	🟩	🟩	🟩
13	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟨 +	🟩
14	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
15	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
16	🟩	🟩	🟩	🟩	🟩	🟩	🟨 +	🟩	🟩
17	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
18	🟩	🟩	🟩	🟩	🟩	🟩	🟥 +	🟩	🟩
20	🟩	🟩	🟩	🟥 +	🟩	🟩	🟩	🟩	🟩
21	🟩	🟩	🟩	🟨 +	🟩	🟩	🟨 +	🟩	🟩
22	🟨 +	🟩	🟩	🟩	🟩	🟩	🟨 +	🟩	🟩
23	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
24	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
25	🟨 +	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
26	🟩	🟩	🟩	🟩	🟩	🟩	🟥 +	🟩	🟩
27	🟩	🟩	🟩	🟨 +	🟩	🟩	🟩	🟩	🟩
28	🟩	🟩	🟩	🟨 +	🟩	🟩			

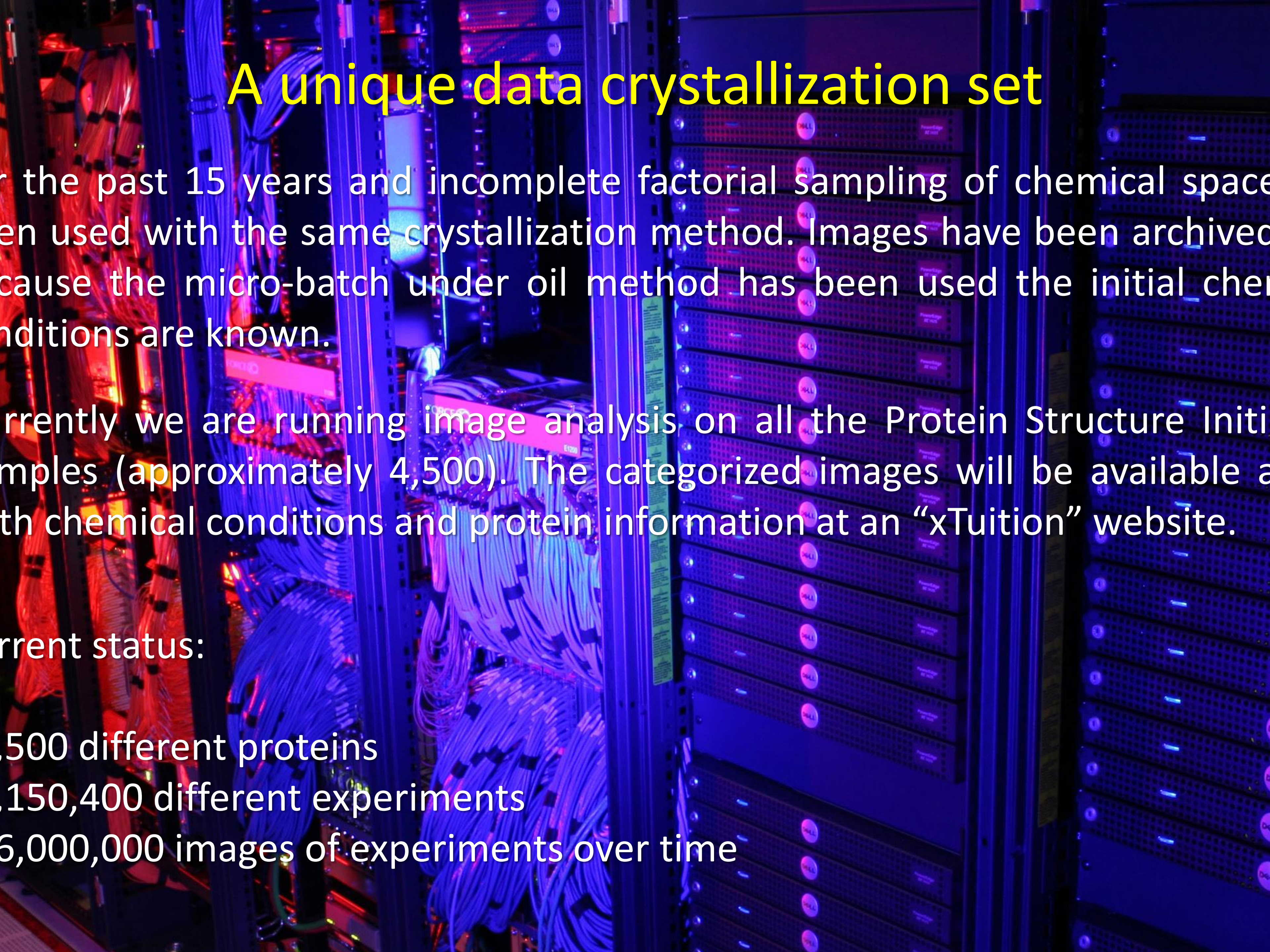
Nonlinearity evaluated for all three Guinier regions. Guinier regions that were unlikely (green, $p > 0.20$), possibly (yellow, $0.05 < p \leq 0.20$), or probably (red, $p \leq 0.05$) nonlinear are shown for each of the three Guinier regions (G1, G2, G3, see section 2.3.2 for details). Attractive forces are denoted as positive (+) and repulsive as negative (-).

SAXS data available

- Data from ~1000 samples
- Three concentrations each
- Analyzed as a function of quality (publishable)
- Metadata including concentrations, data collection characteristics.
- Will be used to compare against crystallization outcome (in progress)

Using the data?

- Oligomer determination
- Protein characterization (construct studies)
- Envelope determination
- Compare to structural homologs
- Priority of SAXS targets?



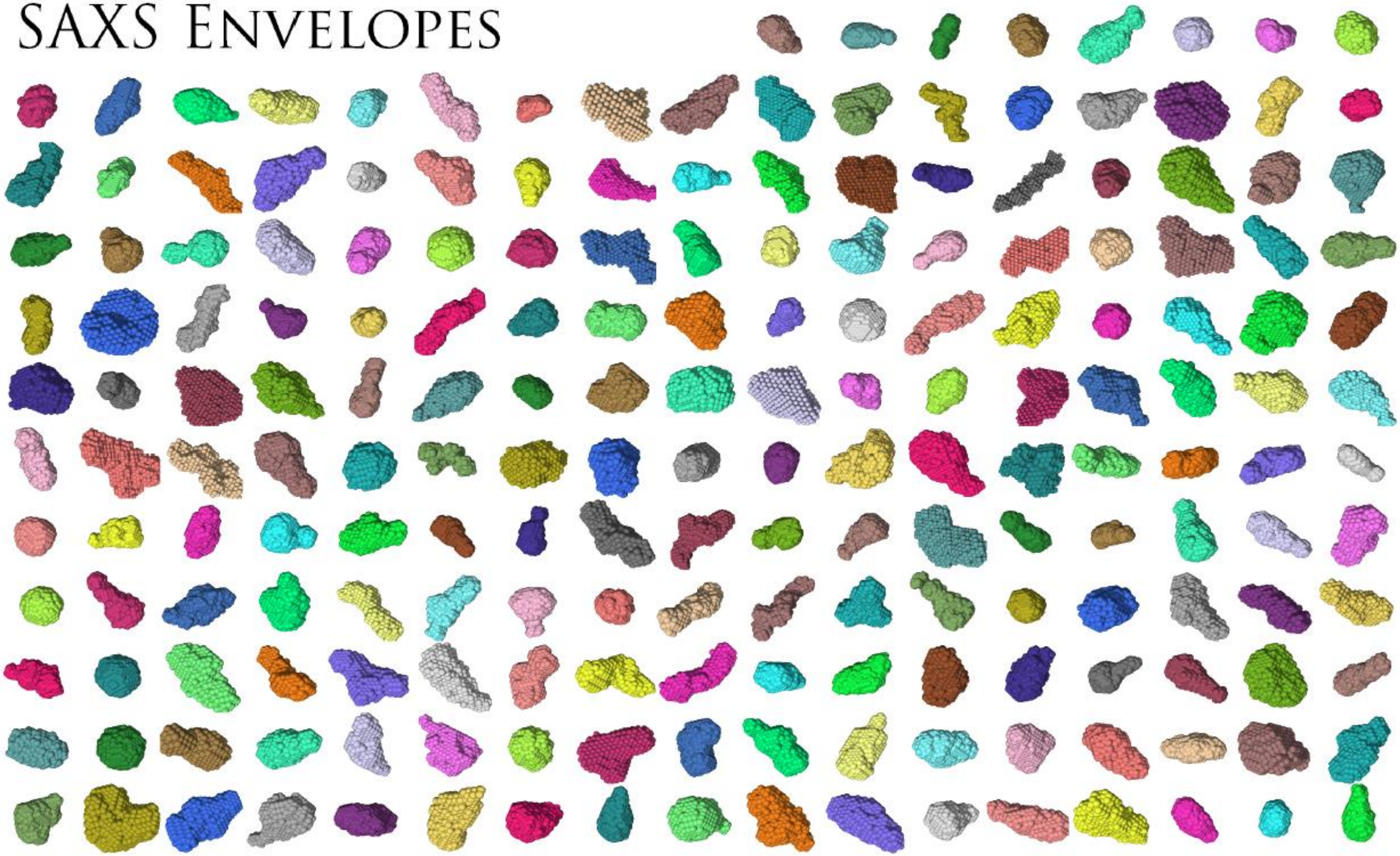
A unique data crystallization set

r the past 15 years and incomplete factorial sampling of chemical space en used with the same crystallization method. Images have been archived cause the micro-batch under oil method has been used the initial che nditions are known.

rrently we are running image analysis on all the Protein Structure Initi mples (approximately 4,500). The categorized images will be available a th chemical conditions and protein information at an “xTuition” website.

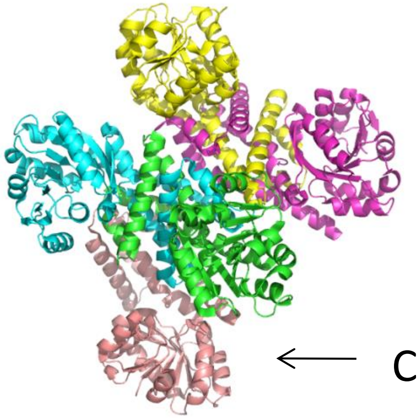
rrent status:

,500 different proteins
,150,400 different experiments
6,000,000 images of experiments over time



SAXS : the T-shirt (Tom Grant)

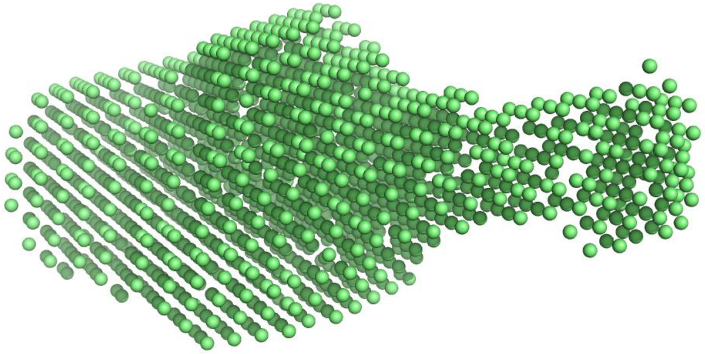
Another story



2izz from the PDB
(5 chains in PDB)

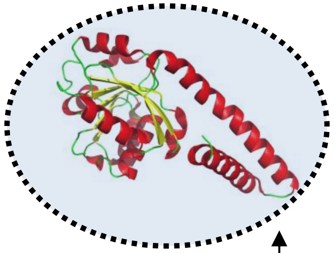
Crystal packing artifact

Solution envelope from BcR38B-21.20-SeMa-Gf (3gt0)

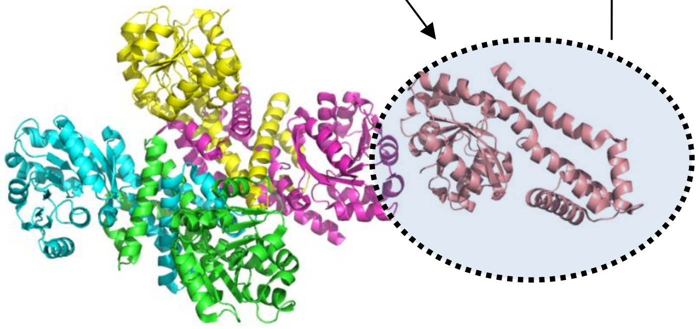


~165A

3gt0 from the PDB



Correct position for 5th chain



~165A

Biological unit based on 2izz and SAXS

Summary, the start of turning high-throughput crystallization to high output

The current success rate is 22%, *i.e.* 1 out of every 5 samples coming through the laboratory door lead to a structure deposited in the PDB.

Despite having soluble pure samples ~80% of the time we fail to obtain structure.

Small Angle X-ray Scattering (SAXS) provides a radius of gyration and characterization of the sample in terms of globular, domains with flexible linkers or natively unfolded. It can also provide a low resolution (15A) envelope of the structure.

We use 60 µl of sample (left over from crystallization screening) and run 3 concentrations at SSRL beamline 4-2. Each sample takes 10-15 minutes to run.

Out of 260 samples analyzed (from ~3,000 in the freezer) 77% gave good SAXS data and were well folded globular samples (compared to only 22% that crystallized). Out of the remainder 2 were natively unfolded.

High-throughput SAXS has applications in crystallization

Wrap-Up

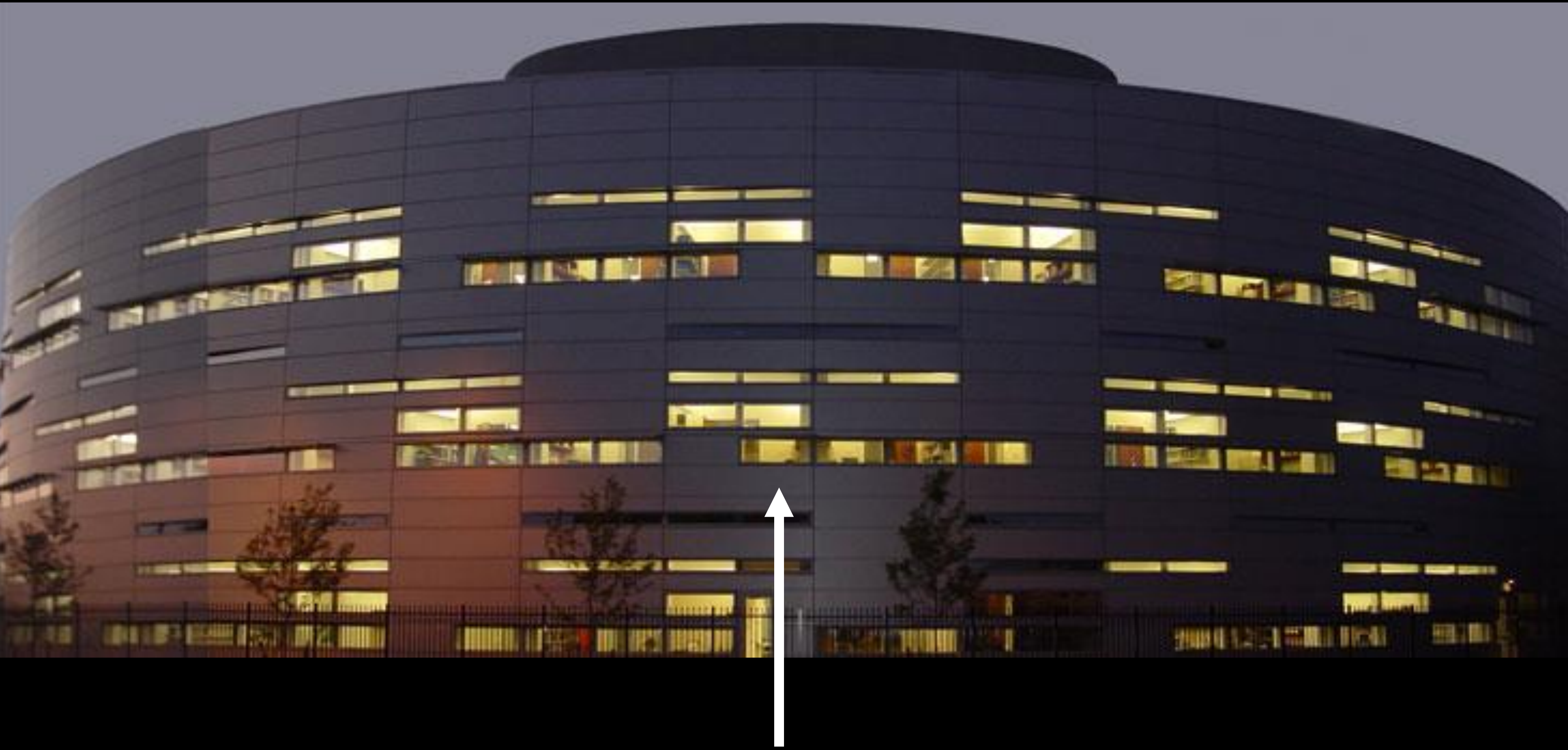
- Defining the question is fundamental to reliable conclusions.
- Many SAXS analyses require monodispersity, so make sure you've got good quality data before trying to draw those conclusions.
- SAXS "resolution" is ambiguous, not directly 2π/q. Resolution is really the ability to discriminate between models.
- While useful, don't read too much into envelopes. SAXS is not an appropriate method for placing short loops of residues or other such "high-resolution" structural questions.
- SAXS is a solution technique, so what's in solution is very important. Temperature, pH, or additives can alter your solution structure.
- Be sure to back up any conclusions you draw with other experimental evidence before publishing SAXS data.

Acknowledgements

Thomas Grant, Joseph Luft,
Hiro Tsuruta, Anne Martel,
Lester Carter, Tsutomu Matsui
and Thomas Weiss

Support and Funding: NIH, DoD and NSF

Questions?



esnell@hwi.buffalo.edu