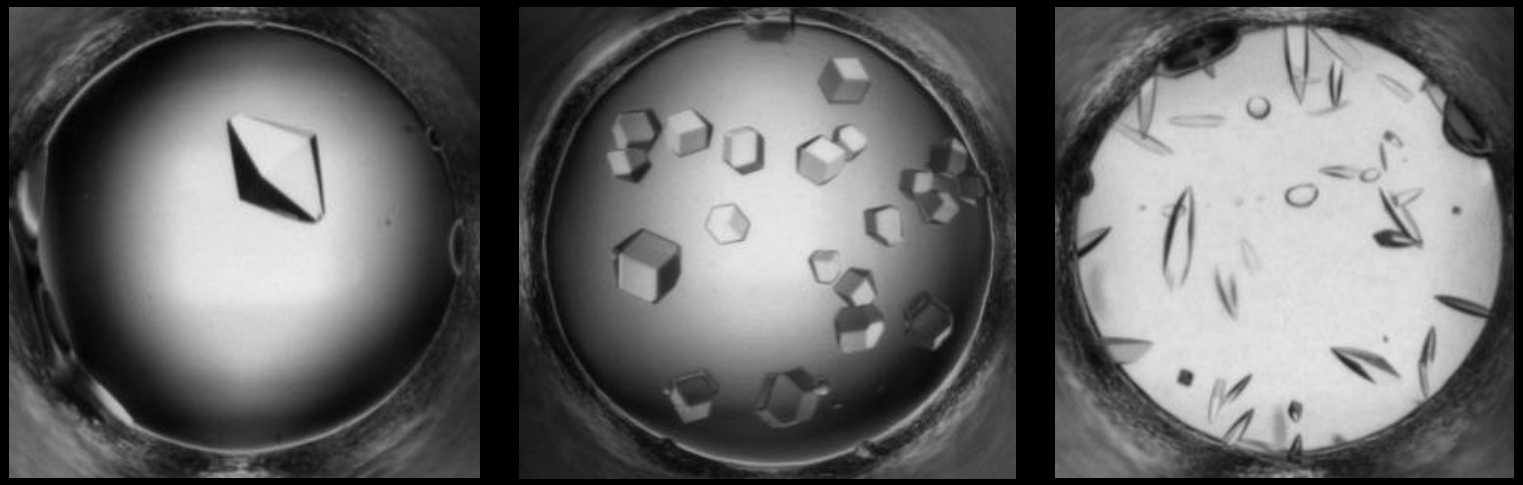
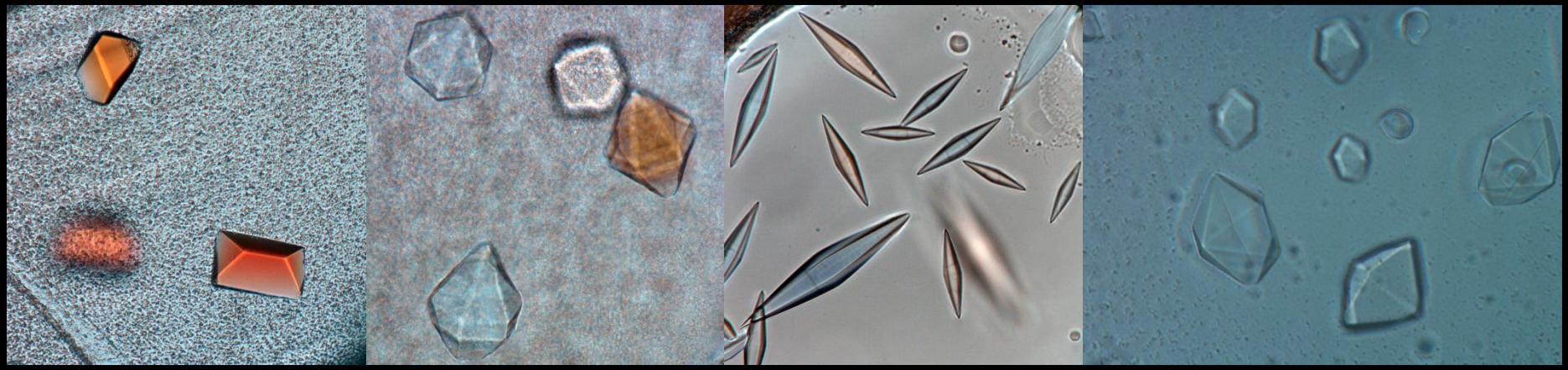
Small-Angle Scattering as a complementary technique in structural biology

Edward H. Snell

CEO Hauptman-Woodward Medical Research Institute



Crystallography Requires Crystals



No crystal …

No crystallography ….

No crystallographer ….

However …

- It is possible to get low resolution structural information from a protein or complex in solution.
- This can tell you about the foldedness and dynamics of the system.
- It can position known structural information in a complex.
- It can determine the area sampled by flexible regions not resolvable crystallographically.
- It is not limited to the chemistry where crystallization occurs.
- It can determine if gross structural changes occur.
- It can be used to provide information to guide crystallization
- New algorithms may enable direct electron density determination.

Introduction to Small Angle Solution Scattering (X-ray or Neutron)

(one of several complementary techniques)

SAXS Literature and Software

Reviews:

- Putnam et al, Q Rev Biophys. Aug 2007; 40(3): 191-285.
- Jacques and Trewhella, Protein Science 2010 Apr; 19(4): 642–657.
- Svergun et al, Oxford University Press 2013, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*

- Long list of software for SAS data analysis for biological and non-biological applications available at:

http://smallangle.org/content/software

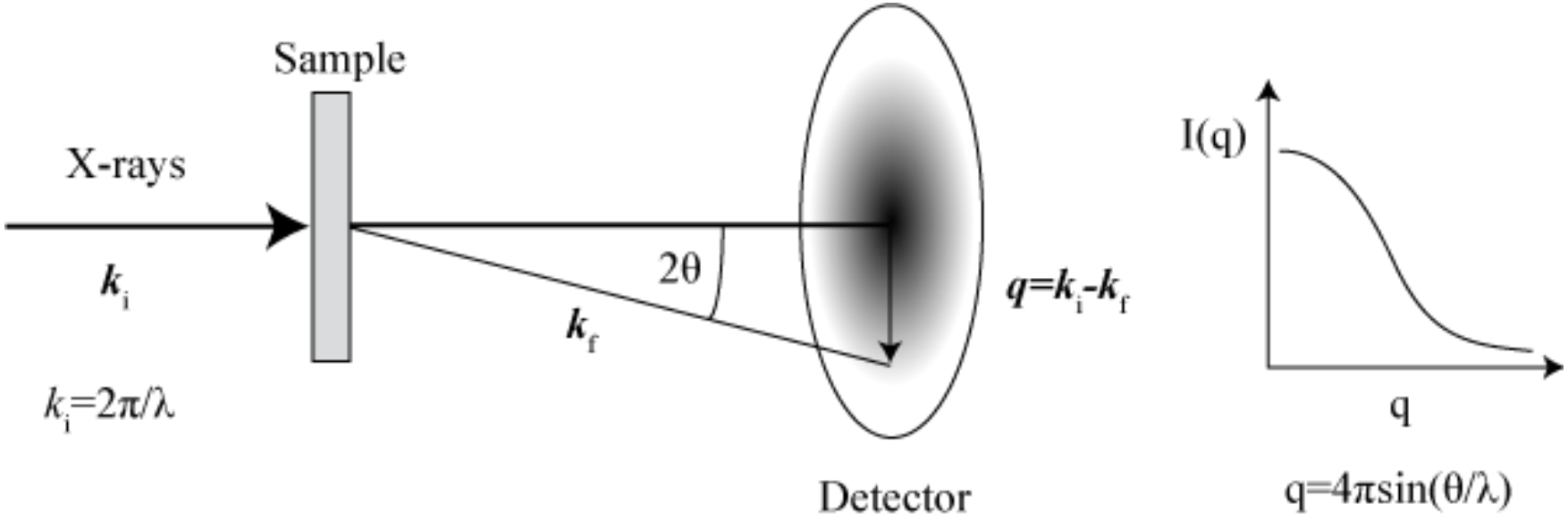
- Most common package for analysis and modeling of biological SAS data is ATSAS, however many other excellent software packages exist

Many illustrations in this talk are from Grant, Lattman, and Snell: Biological Small Angle Scattering: Theory and Practice, IUCr Monograph, Oxford University Press, to be published 2017

Small Angle Scattering

- A solution is illuminated with a parallel, monochromatic X-ray or neutron beam, and the scattered radiation is collected on a detector placed far back from the specimen. Because the solution is homogeneous and isotropic, the observed scattering pattern is circularly symmetric. The full pattern, a three-dimensional function in diffraction space, is spherically symmetric.
- The term solution scattering is applied to the general phenomenon, with the term small angle scattering reserved for the most common application in which observations are confined to radiation scattered within a small angular cone around the main beam.
- Both X-rays and neutrons are used with the terminology Small-Angle X-ray scattering (SAXS) and Small-Angle Neutron Scattering (SANS).
- For the most part, discussions on SAXS and SANS are interchangeable but each has specific advantages and disadvantages

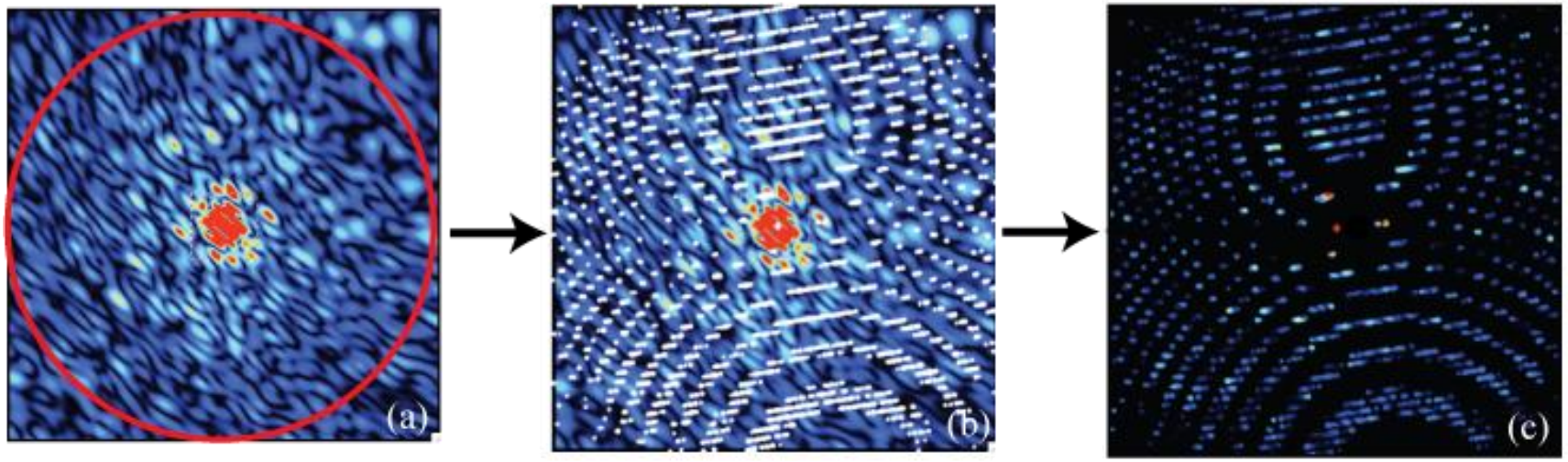
SAXS is everything behind the beamstop



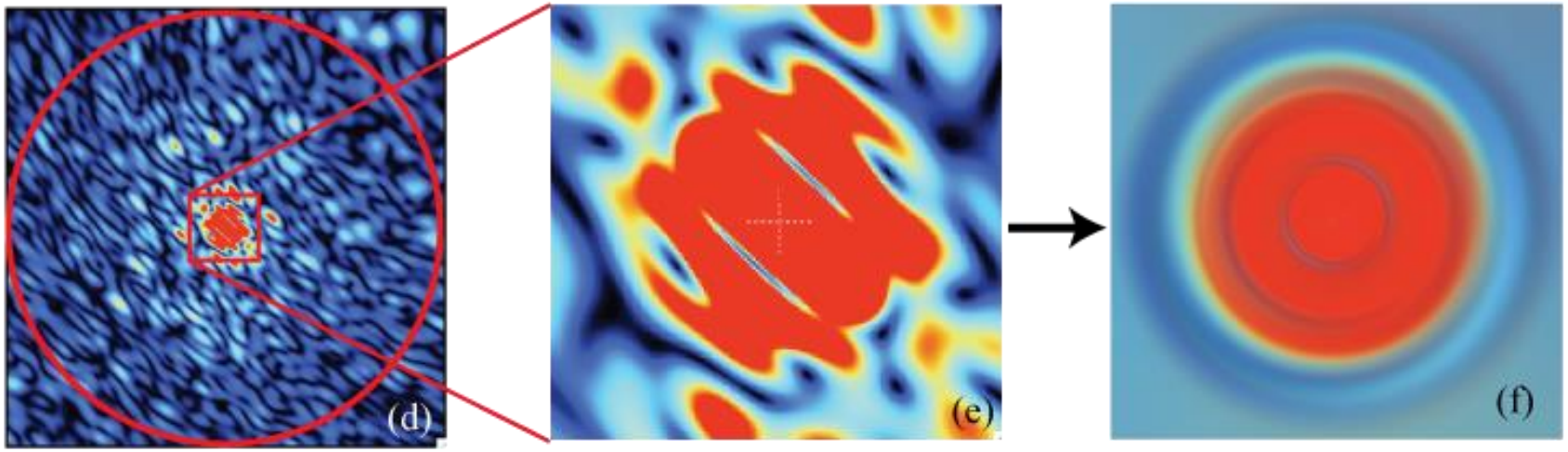
Experimentally, small-angle scattering is simple, practically it is very challenging

- Particles in solution tumble – spherically averaged intensity is recorded.
- Radial integration results in one dimensional SAXS profile.
- Larger particles scatter at smaller angles.
- Analysis of the 1D profile yields information about size and shape.

Single Crystal Diffraction



Small Angle Scattering

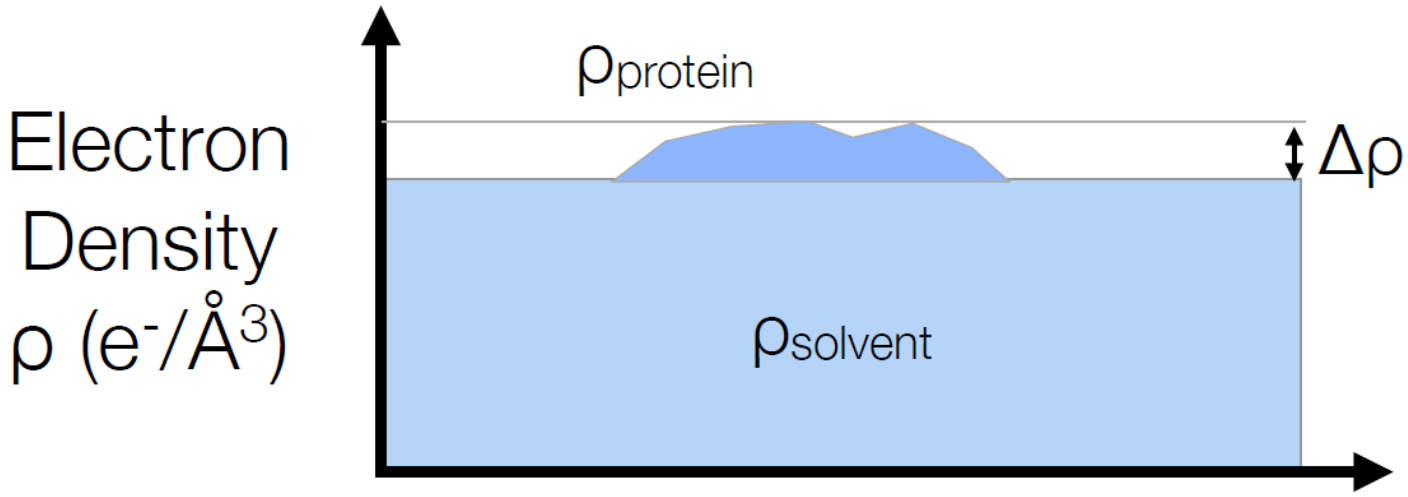


2Å

20Å

SAXS is a Contrast Technique

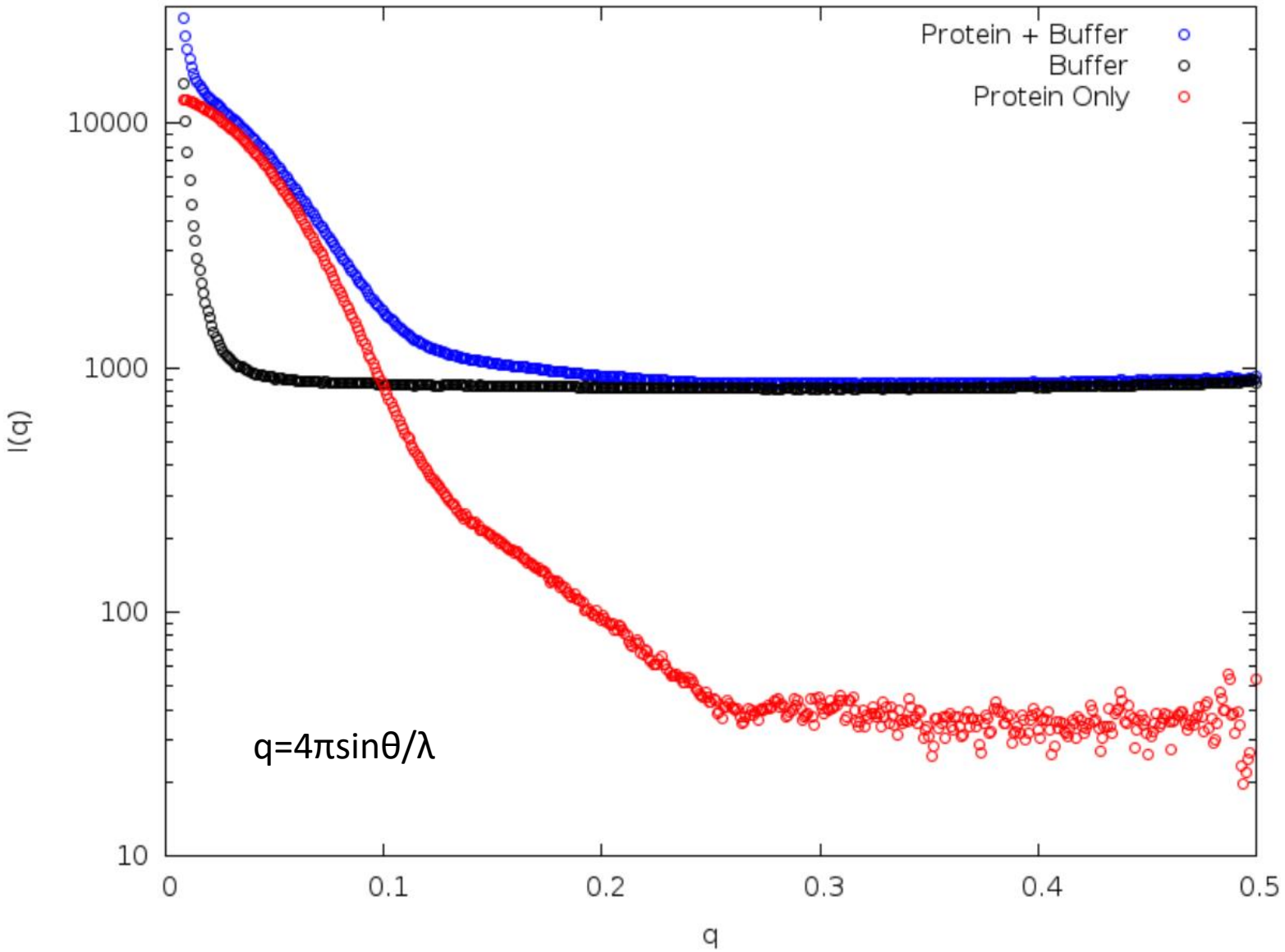
- SAXS is a contrast method, i.e. it depends on the square of the difference in the electron density between the molecule and the solvent



$$(\Delta\rho)^2 = (\rho_{protein} - \rho_{water})^2 = (0.44\text{-}0.33)^2 \approx$$ 10% above background

SAXS data (what you get from the beam)

SAXS data is the sample data with the buffer signal subtracted



SAXS consists of intensity due to the from factor and interparticle contributions

- Equation for scattering intensity:

$$I(q) = F(q) * S(q)$$

Experimental Intensity

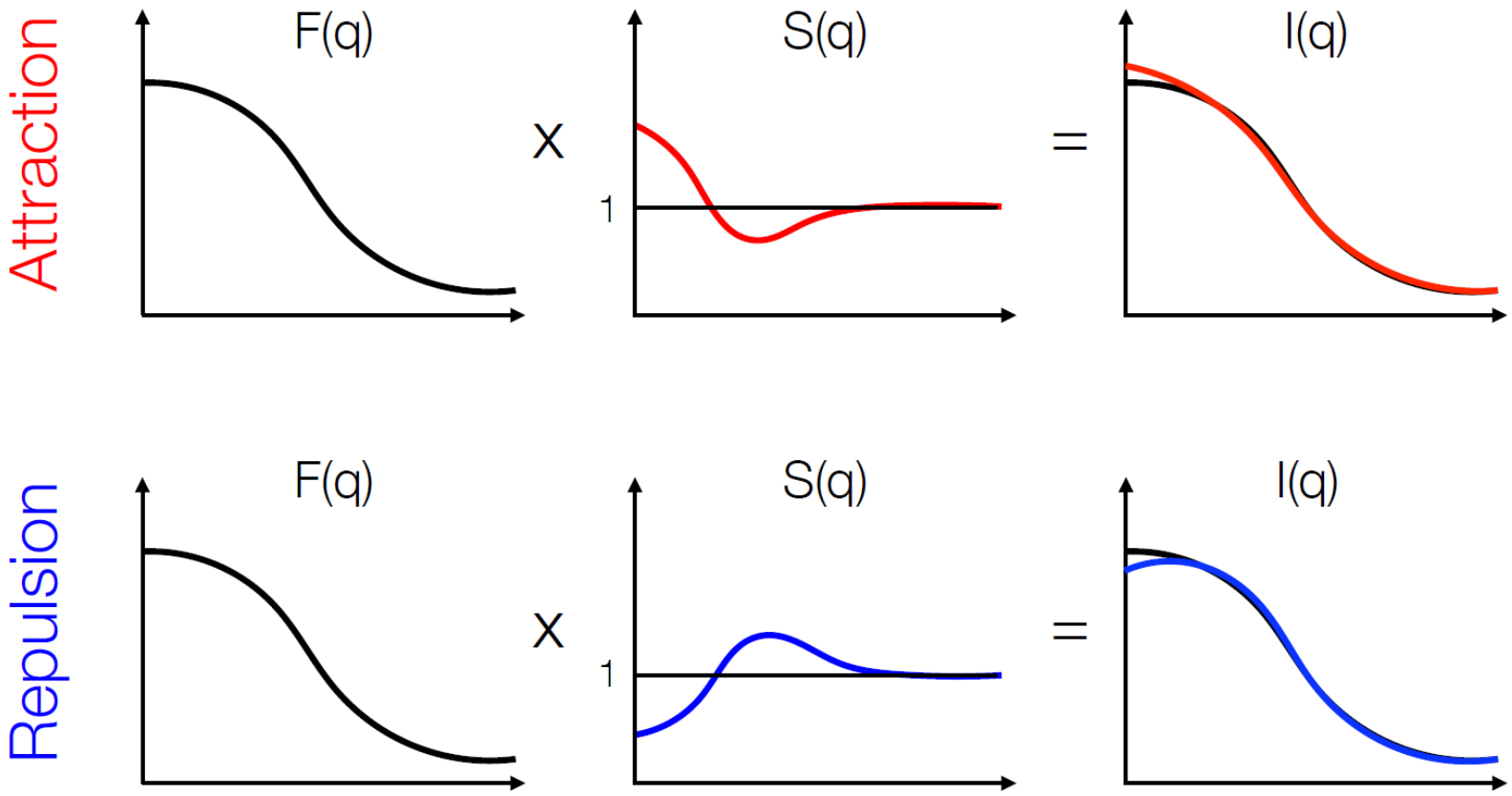
Form factor of particle

Structure factor of solution

- Form factor describes *intra*particle interactions, i.e. size and shape
- Structure factor describes *inter*particle interactions, i.e. repulsion/attraction
- Ideally a monodisperse solution for SAXS should have no interparticle interactions, i.e. S(q) = 1

Interparticle Interactions

S(q) ≠ 1 affects low q data most



Similar to data from light scattering and can be used in the same manner

Sample characterization: Guinier approximation

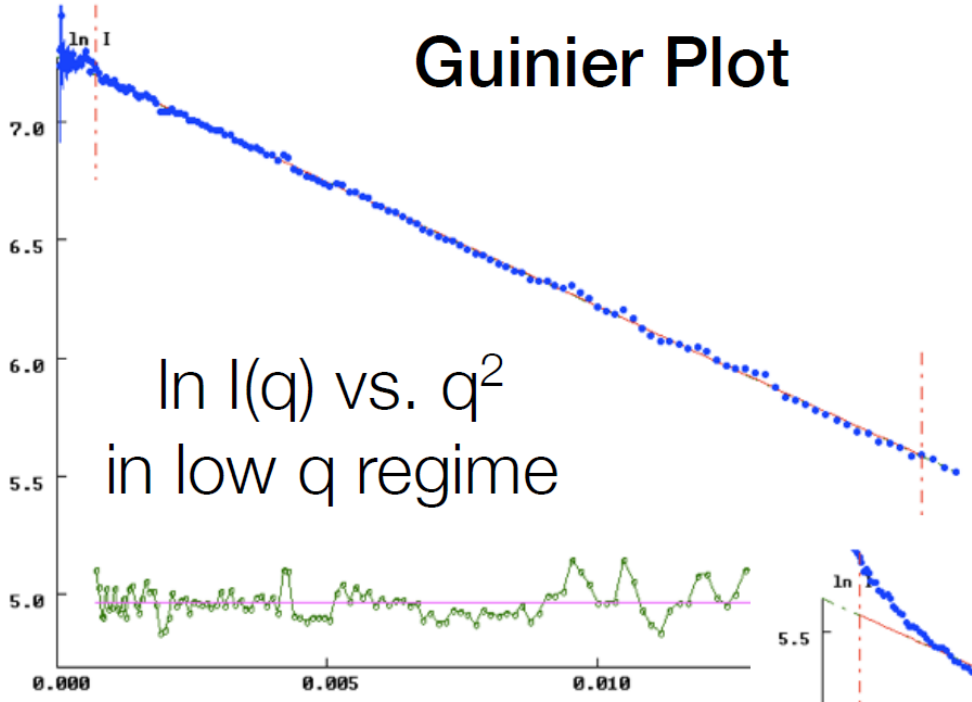
- Developed by André Guinier in 1939.
- As q → 0, intensity can be approximated by:

$$I(q) = I_0 e^{-q^2 R_g^2/3}$$

$$\ln I(q) = \ln I_0 - \frac{R_g^2}{3} q^2$$

y = b + m * x

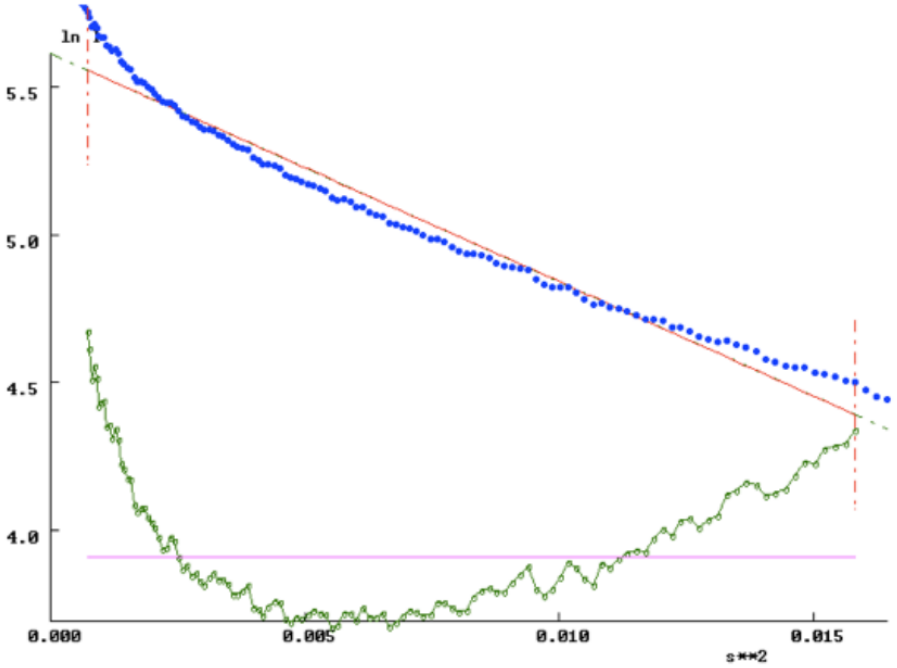
Guinier Plot



ln I(q) vs. q^2
in low q regime

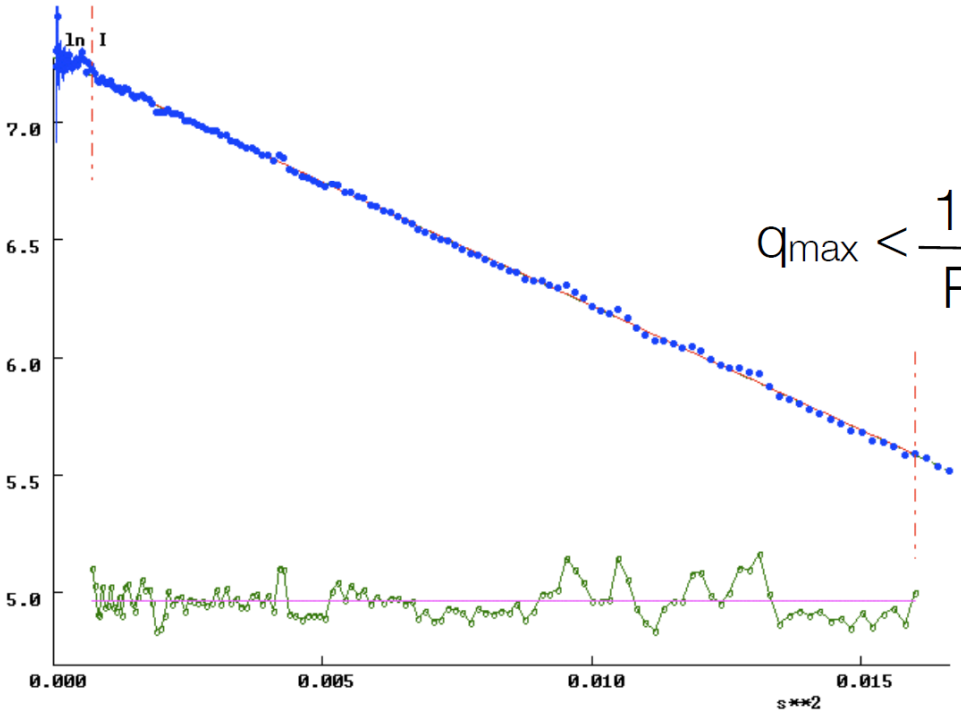
Straight line shows
no aggregation:
Can determine R_g
from slope of line,
I(0) from intercept

Curved line shows
attraction or aggregates
present:
No SAXS processing
should be done



Approximation only valid over a certain region of scattering space

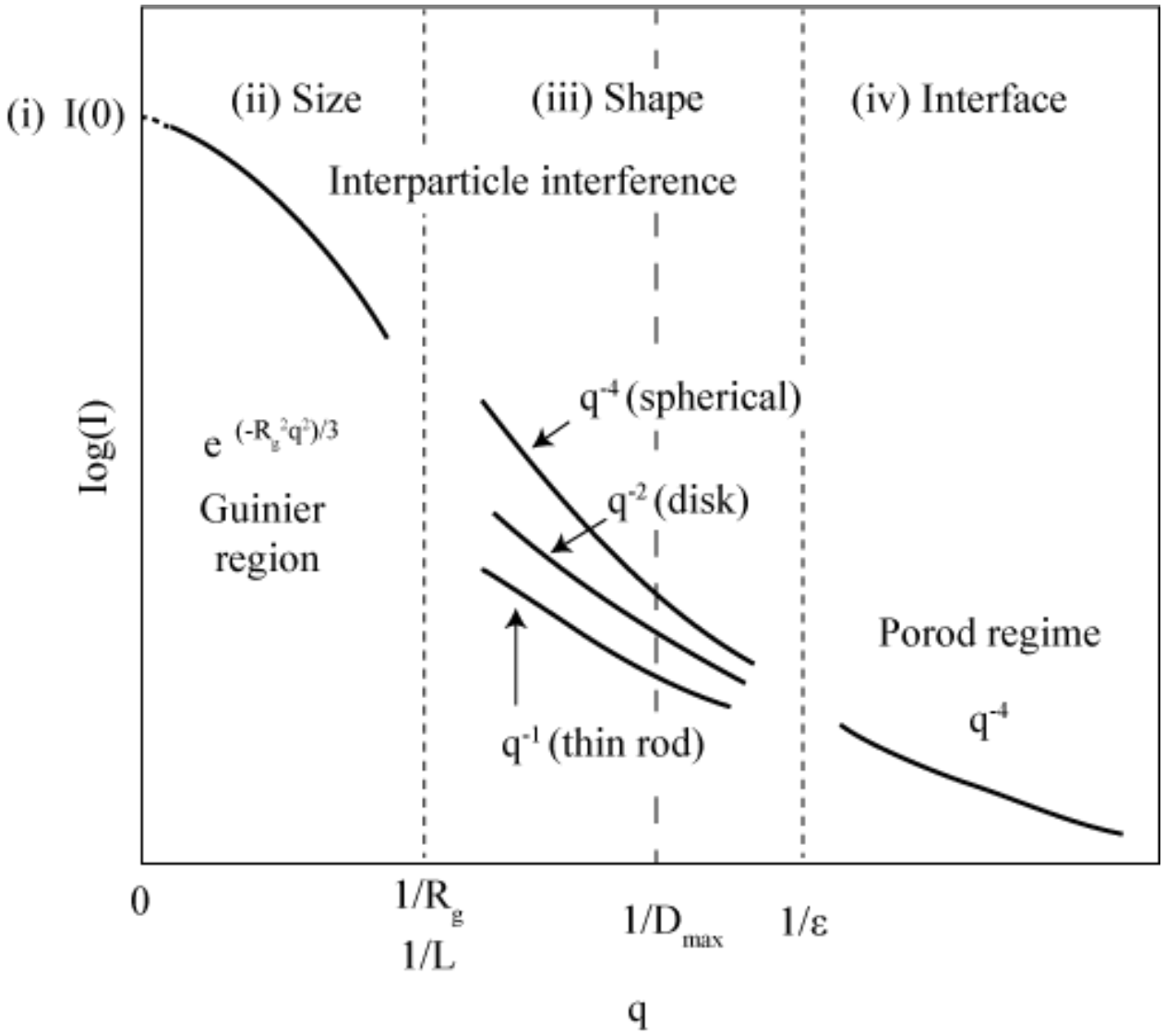
$$q_{min} < \frac{\pi}{D_{max}}$$



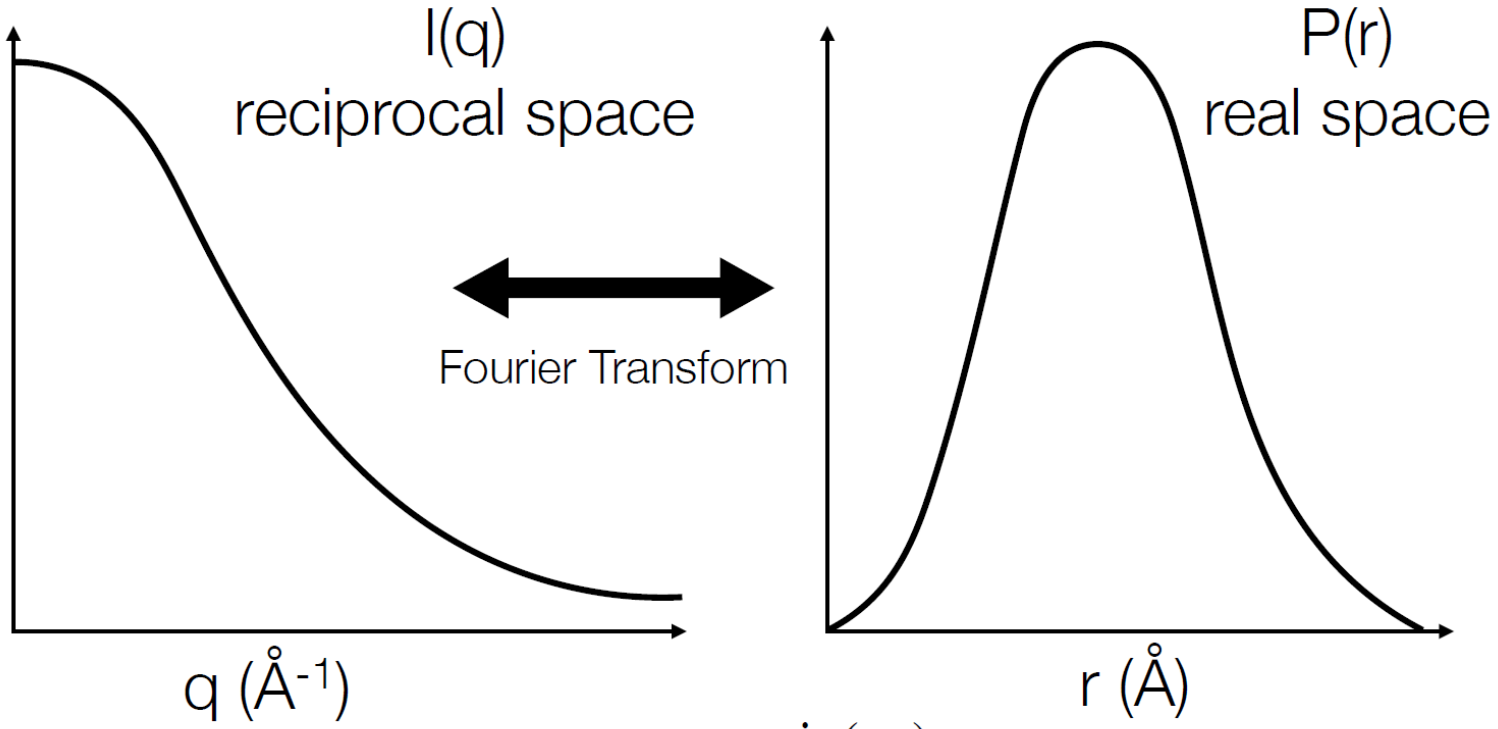
$$q_{max} < \frac{1.3}{R_g}$$

The Shape of the Scattering Curve is important

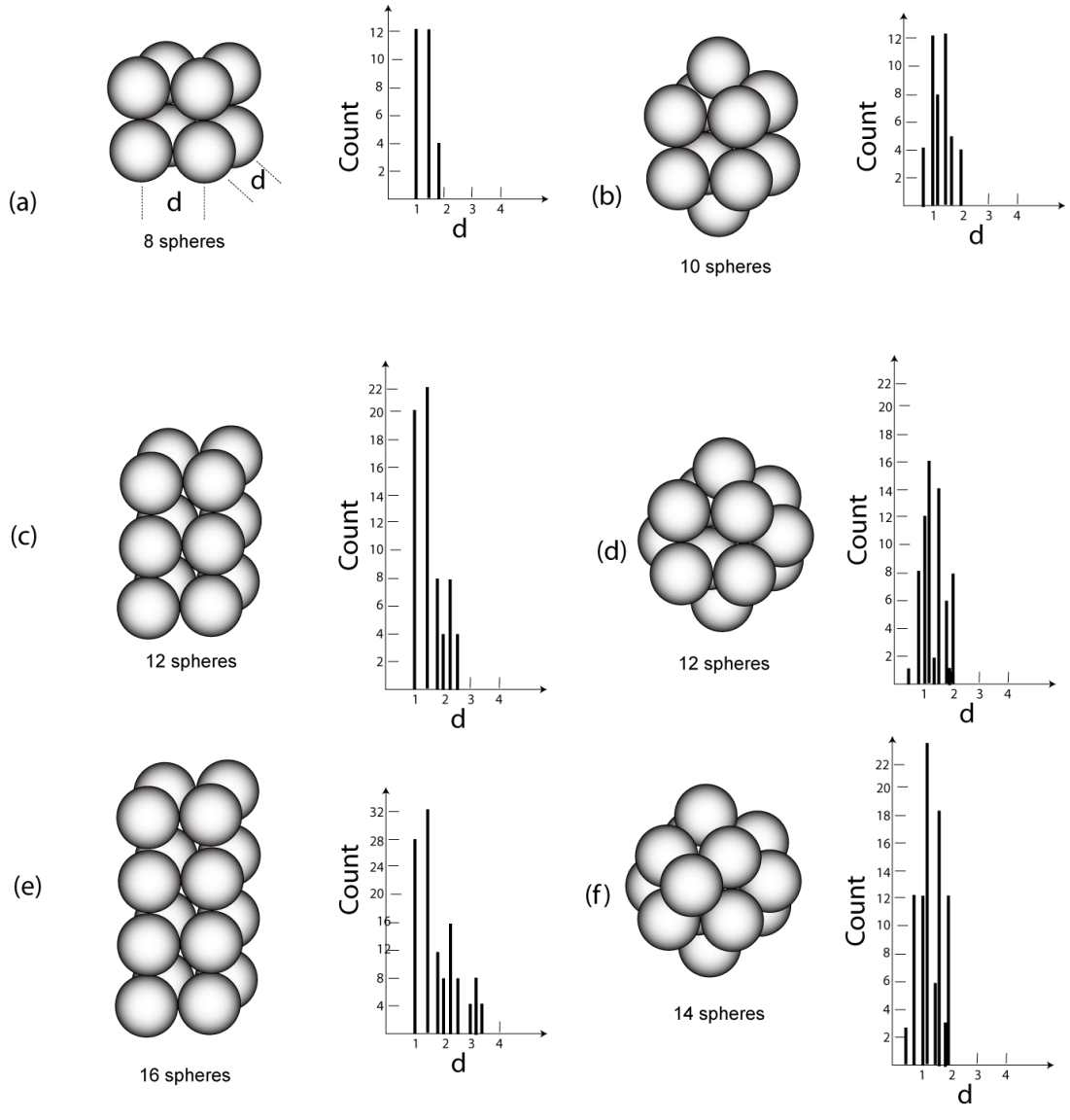
but not the absolute intensity



Scattering is in Fourier space, transform to real space



$$I(q) = \int p(r) \frac{\sin(qr)}{qr} dr$$

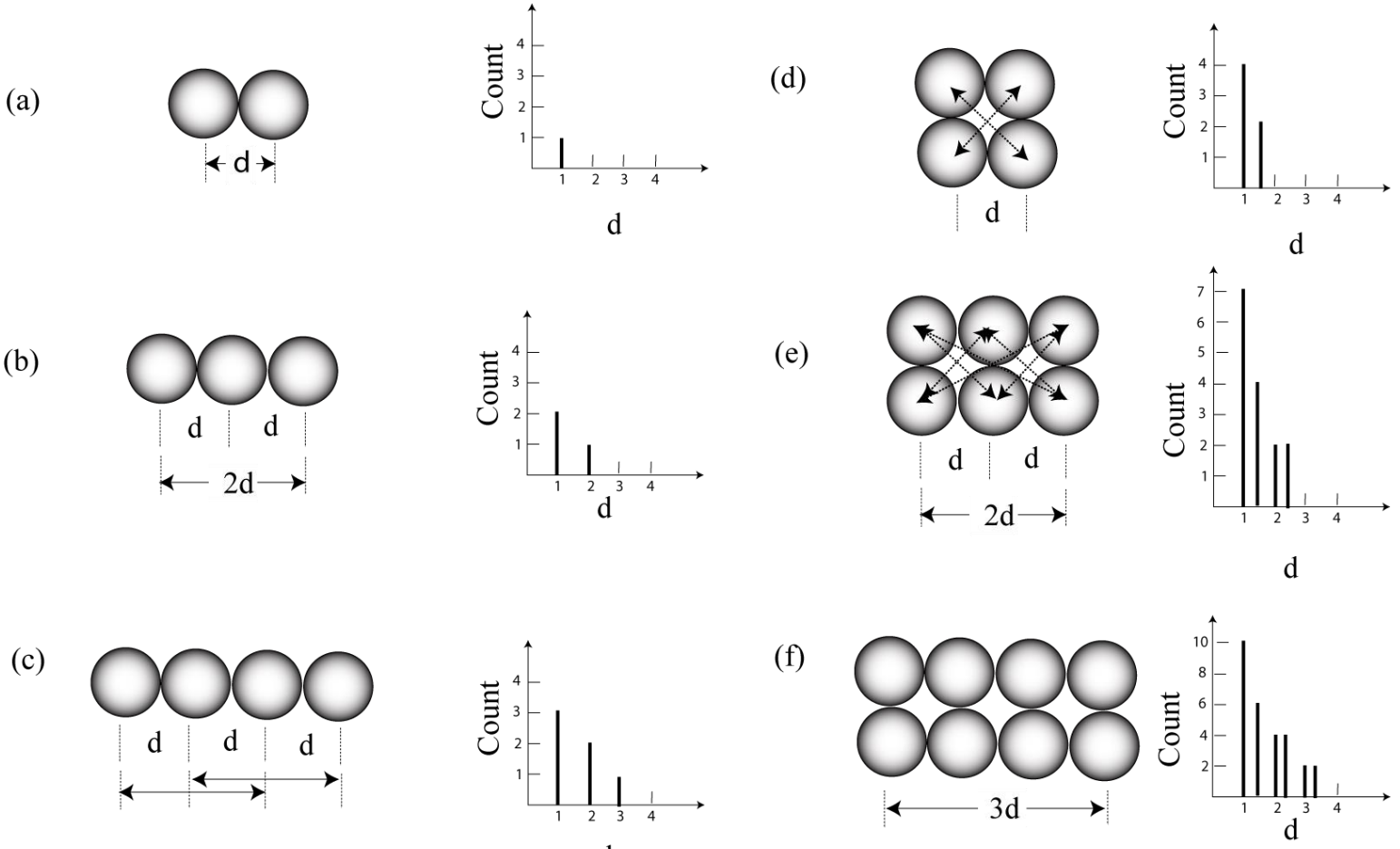


Fourier space yields frequency of interatomic scattering vectors as a function of the length of the vector

P(r) (**Pair distribution function**) plot is simply the histogram of interatomic scattering

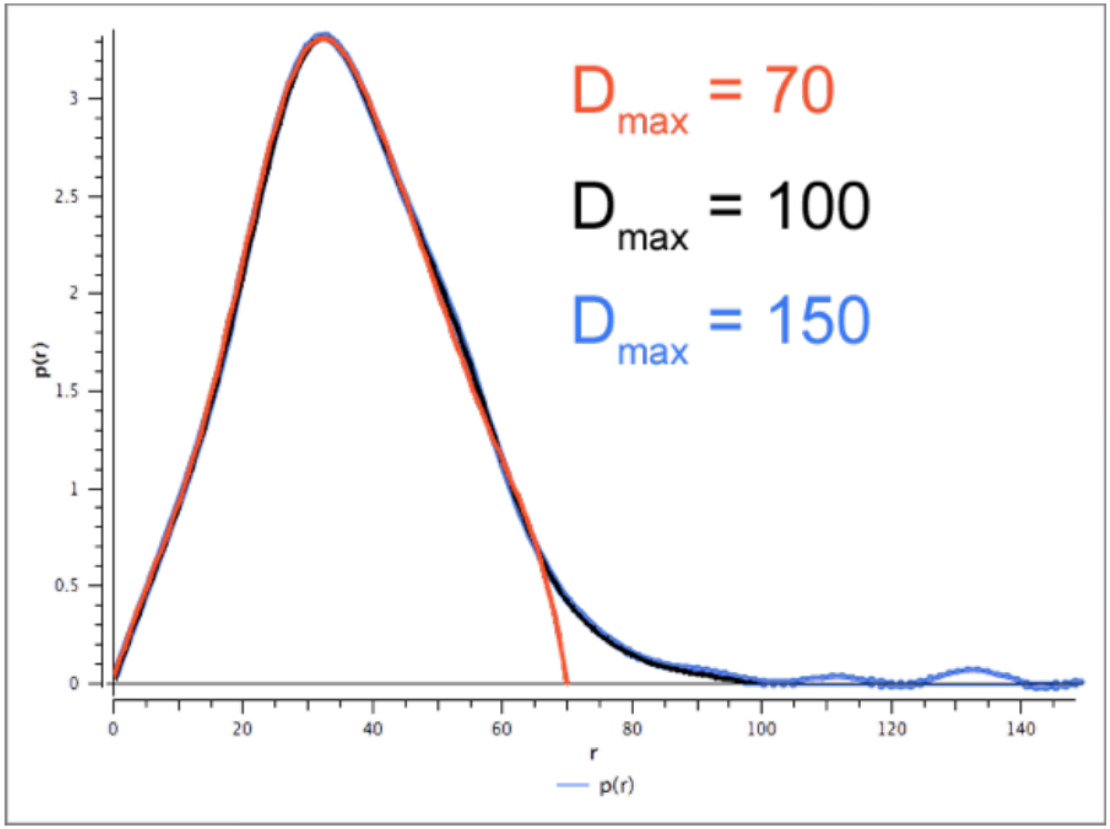
Larger compact molecules have a high distribution at lower angle (consider detector distance etc.)

Two-dimensional examples



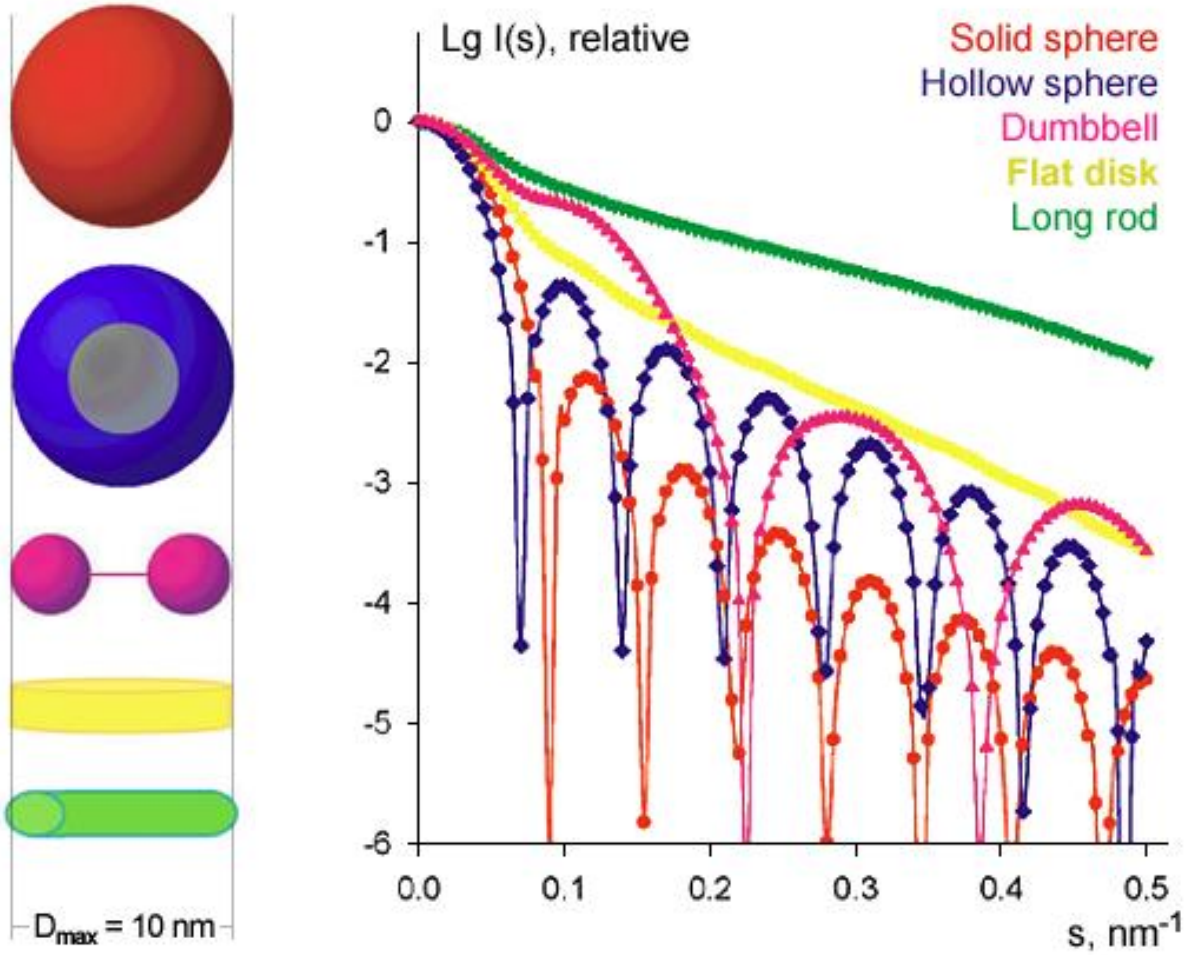
Pair distribution function is used to determine the maximum particle dimension

- Can be used to determine D_{max}
- P(r) should gradually fall to zero at D_{max}
- Underestimated D_{max} appears as abrupt, forced descent to zero
- Starting with large values should identify a decent estimate of D_{max}, given good quality data
- Errors in D_{max} can be large, (~10 - 20%) for good data



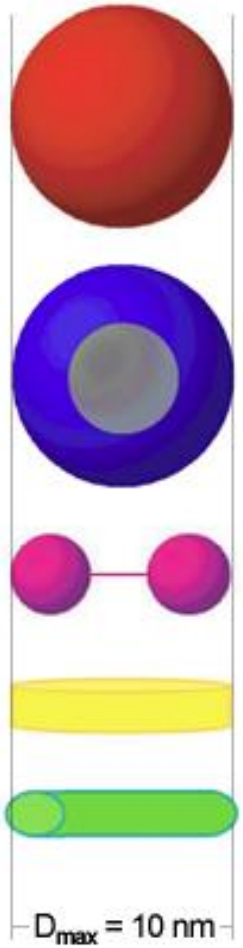
The maximum particle dimension is given by the distance between the furthest interatomic scattering

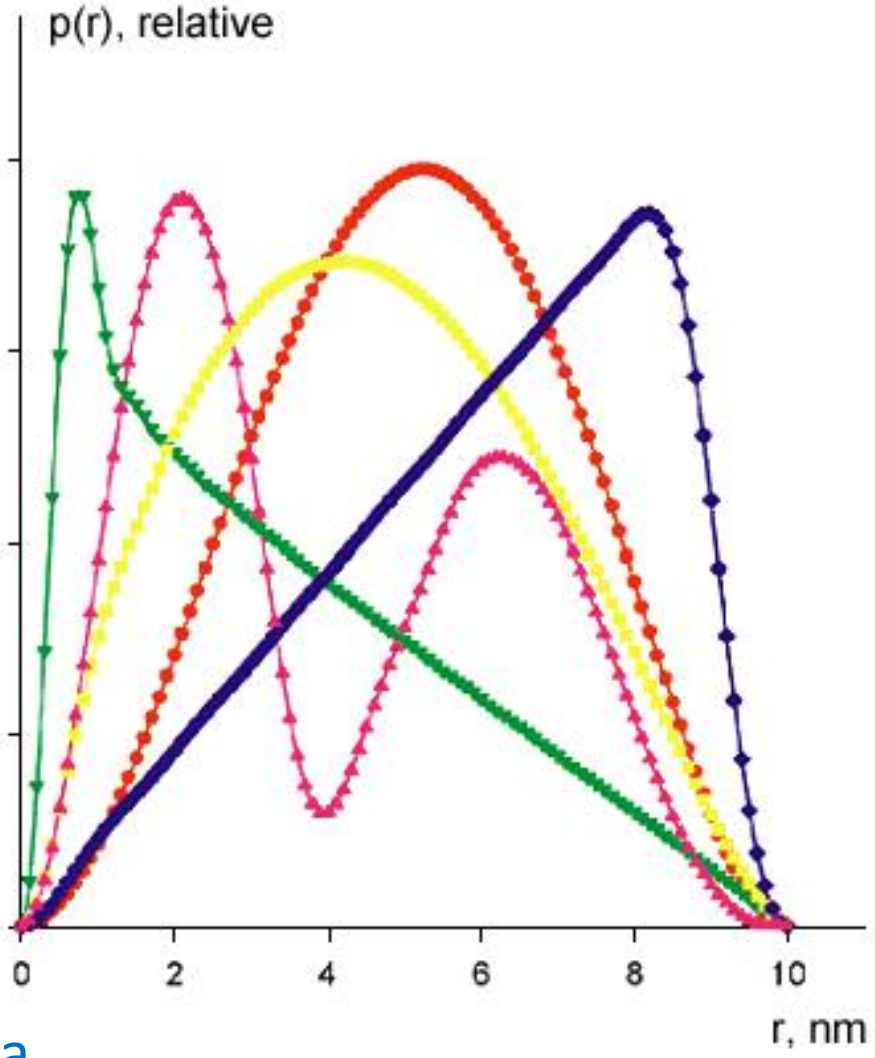
Data



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

Pair distribution function

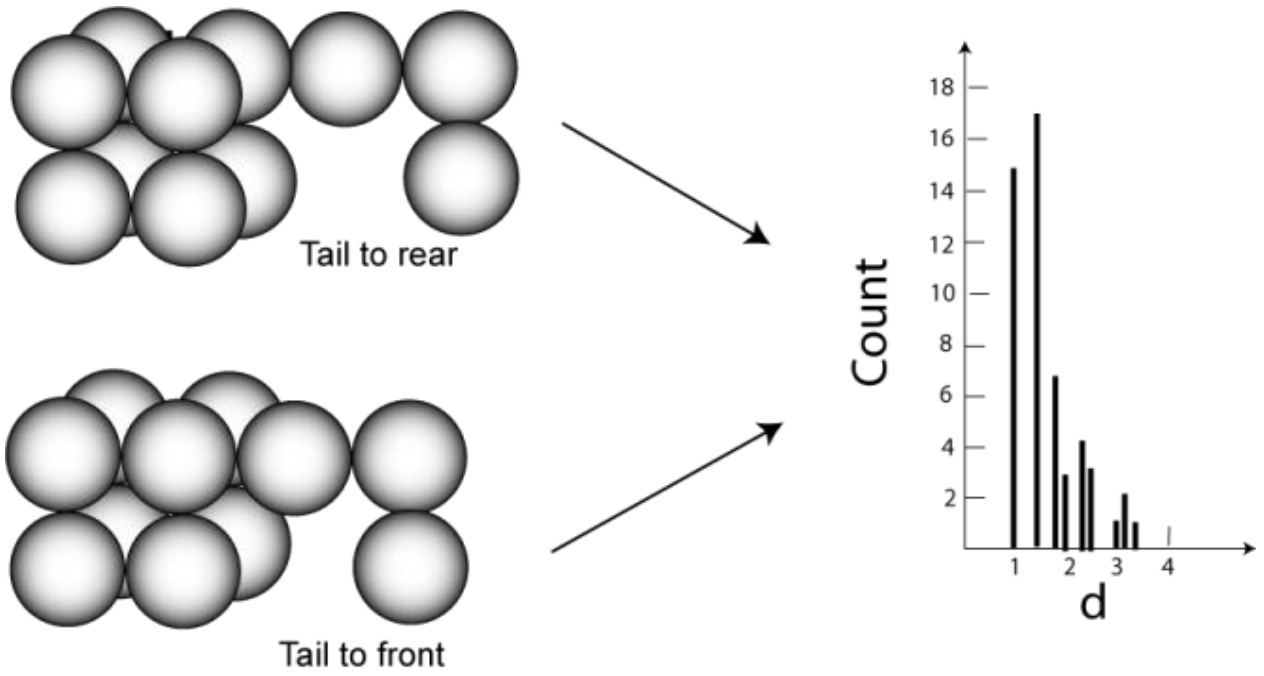




Fourier transform of data.

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

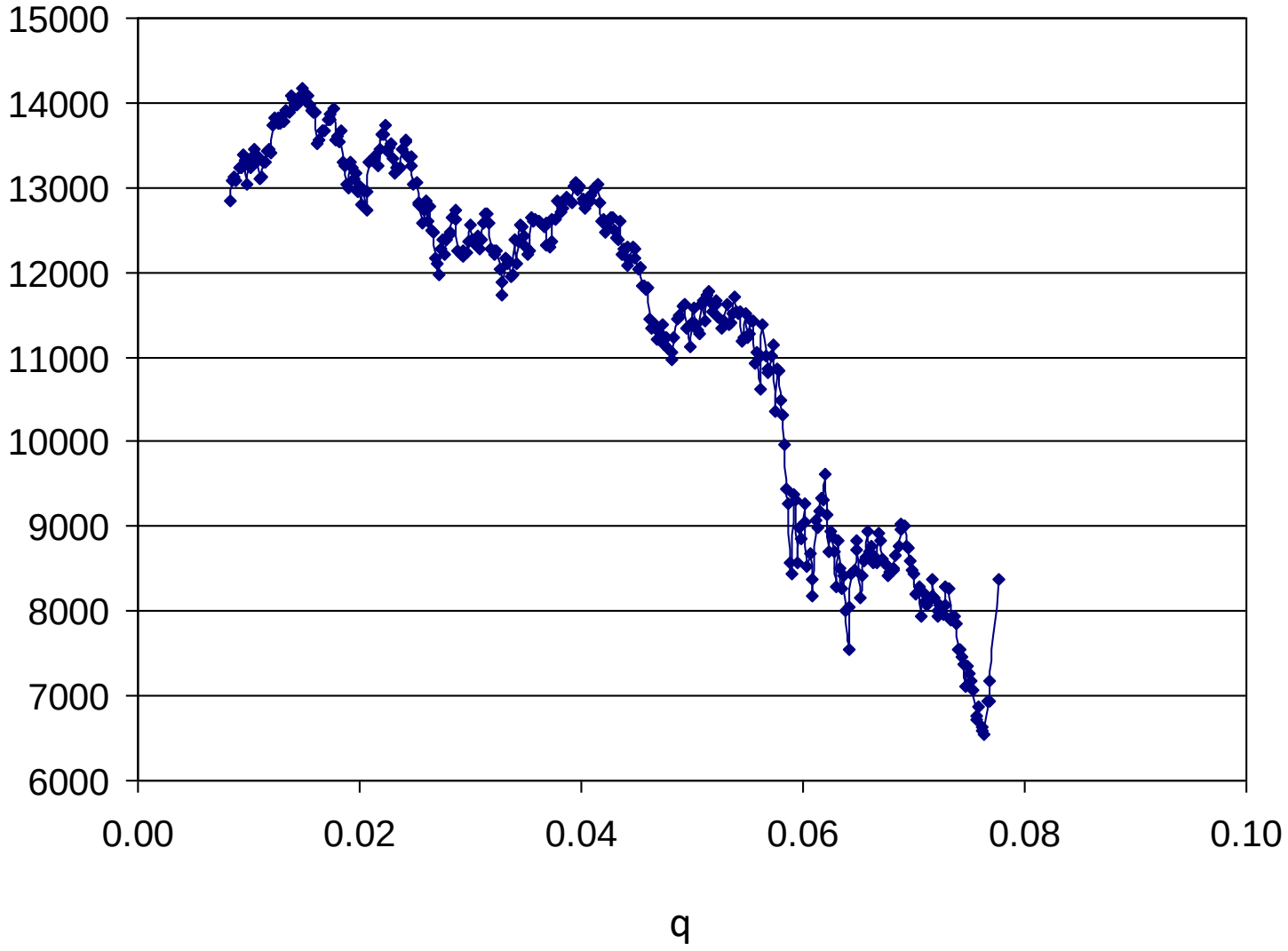
What can possibly go wrong?



Sometimes a unique reconstruction is not available.

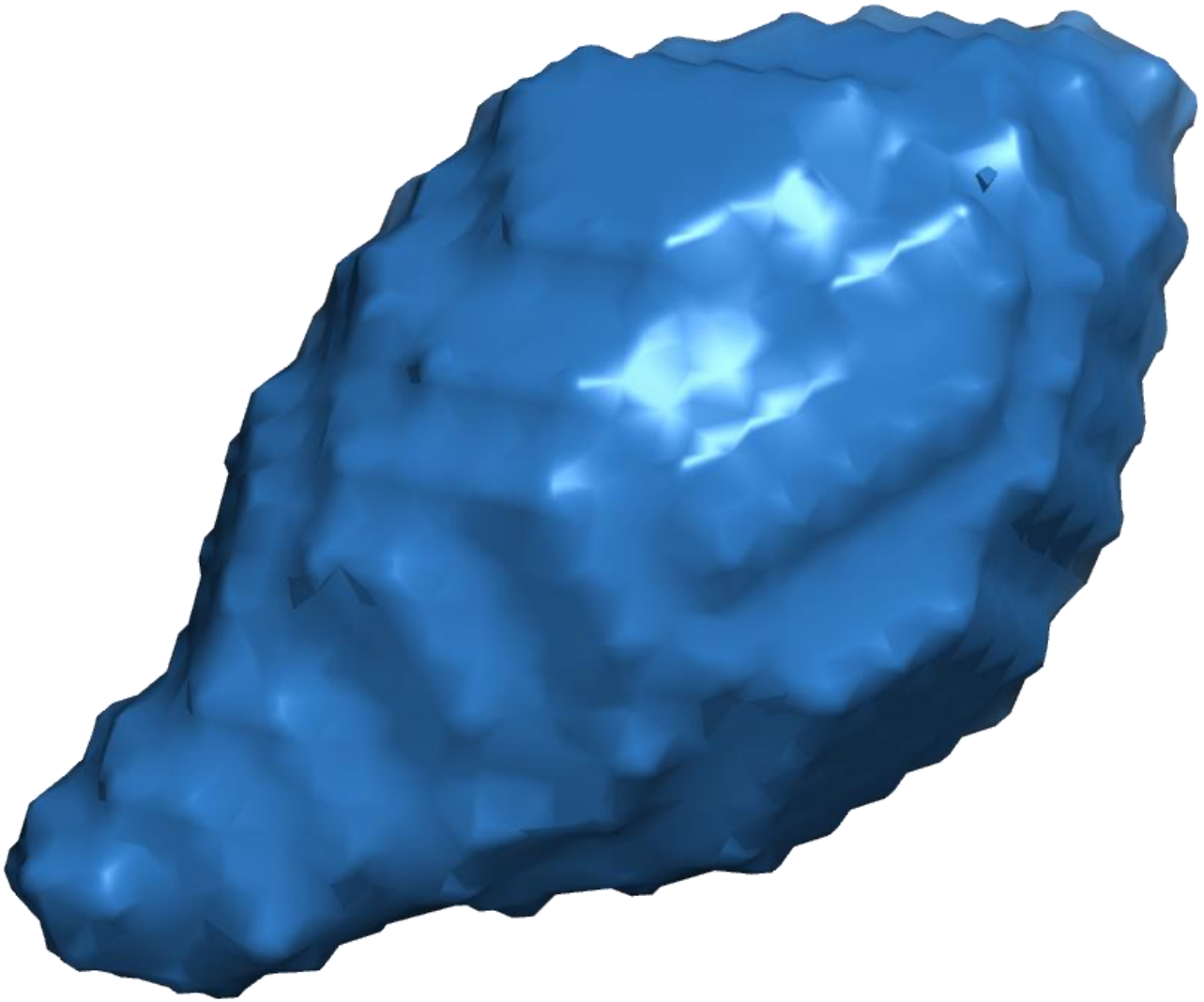
A limitation of the technique is that good or bad data can produce a result

Lets take some '*scattering*' data

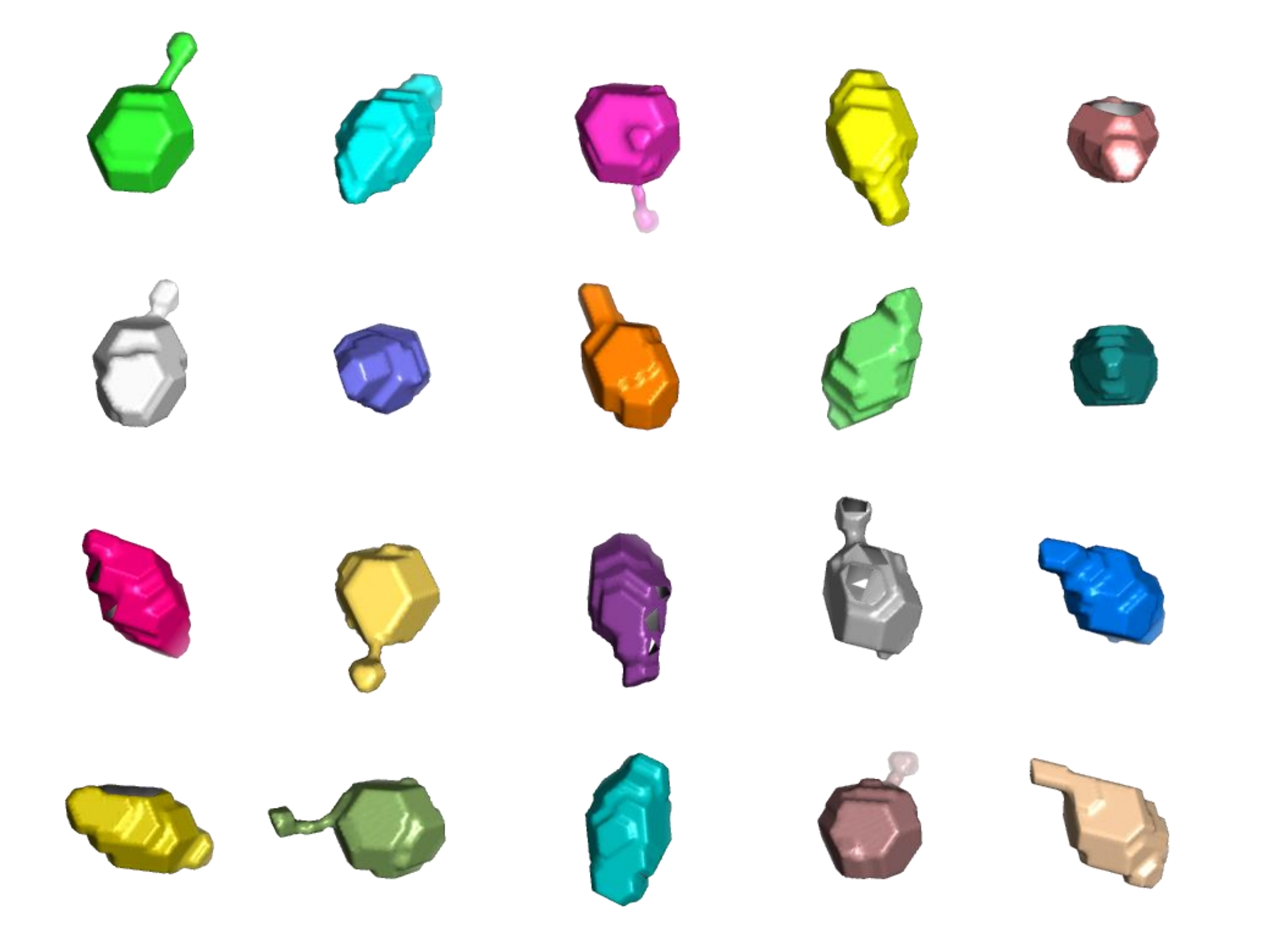


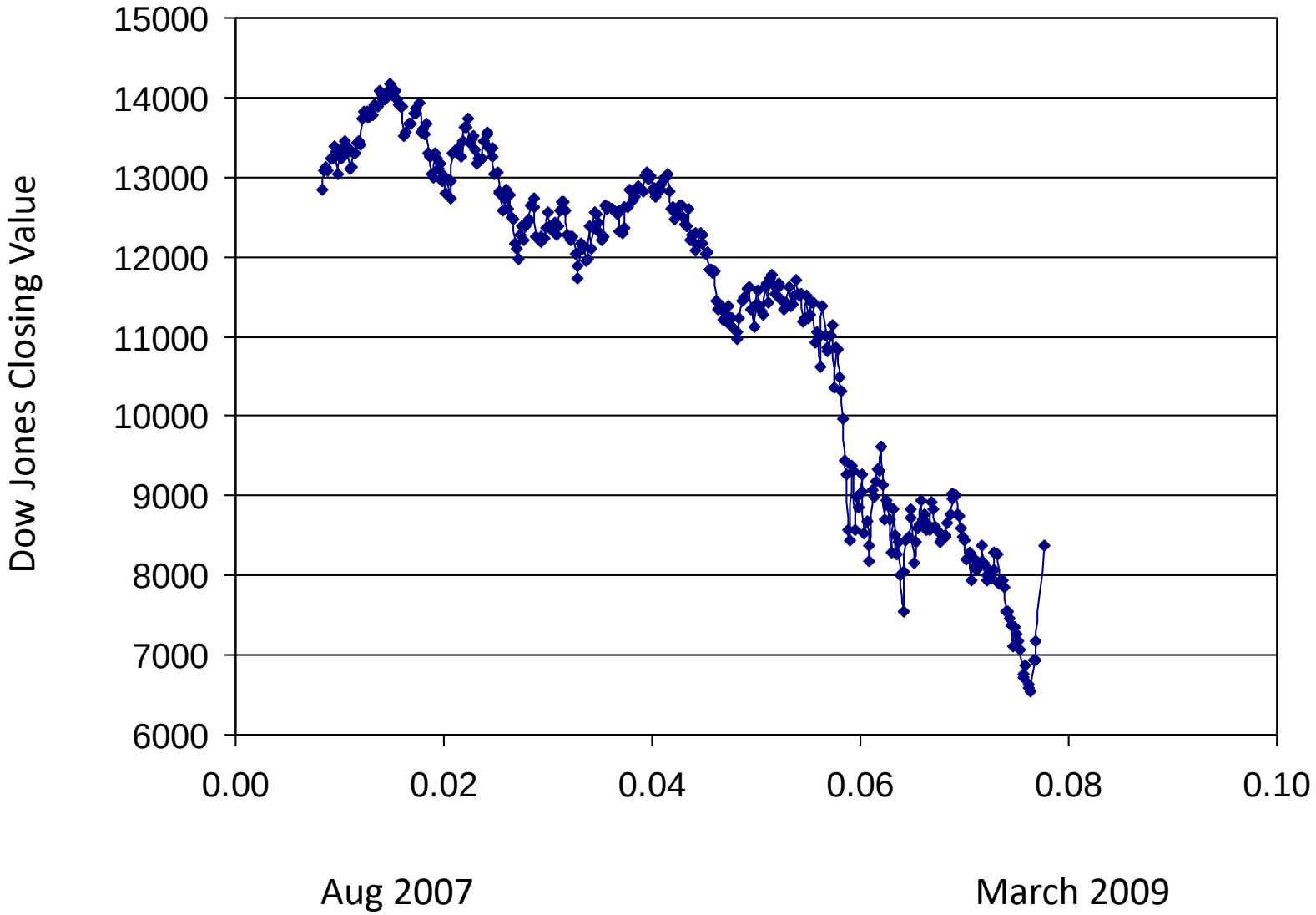
Envelope Reconstruction

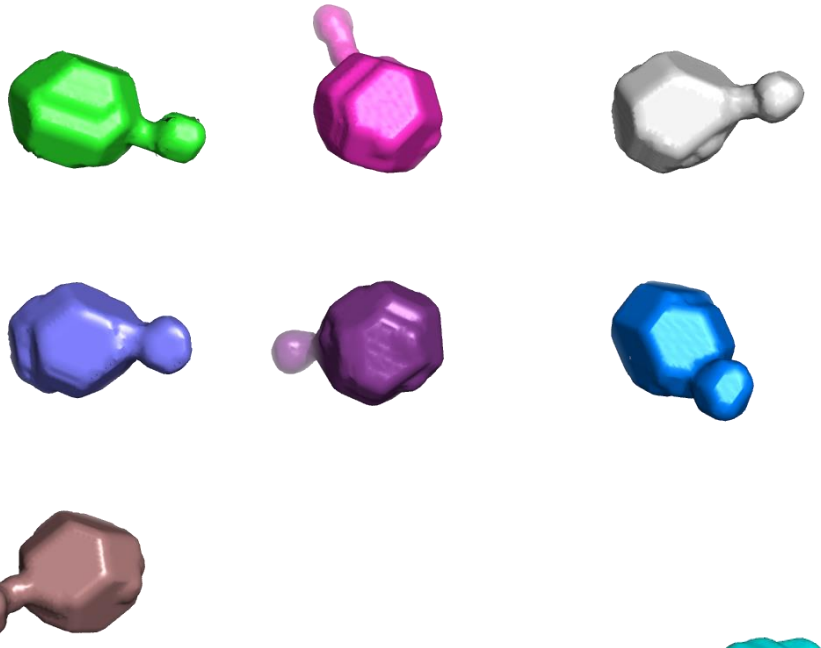
- Produce 10-20 *ab initio* reconstructions
- Determine the most probable model, i.e. the least different from the rest and align all to this.
- Estimate the similarity of the models using the Normalized Spatial Discrepancy (NSD)
  - Average NSD ~ 0.5 implies good stability of solution
  - Average NSD ~ 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



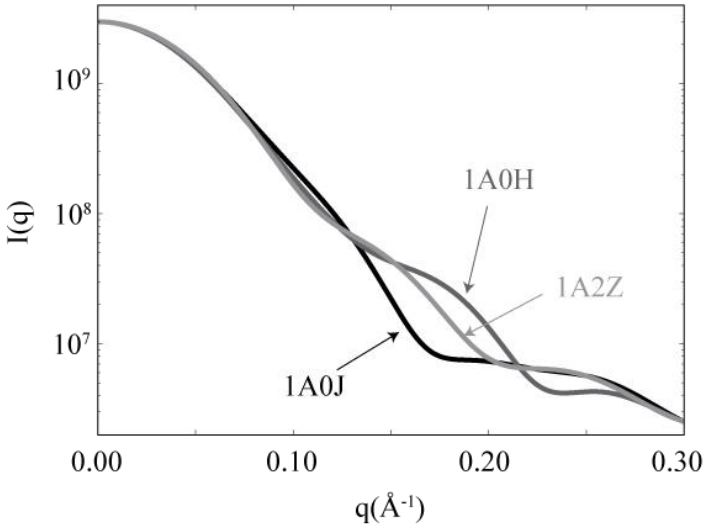

Dow Jones Closing Value
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
0.00
0.02
0.04
0.06
0.08
0.10
Aug 2007
March 2009

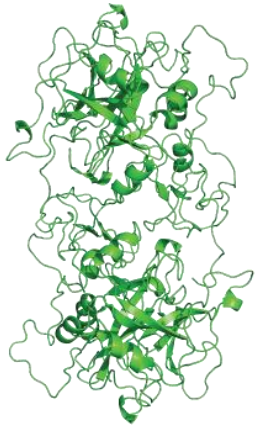


Actually two populations

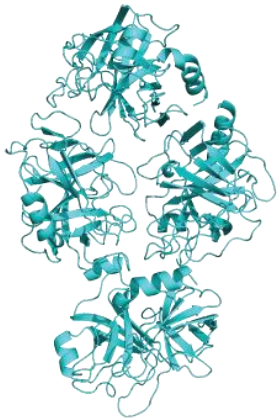
Both are correct, i.e. they explain the scattering data

A Bull or a Bear market!

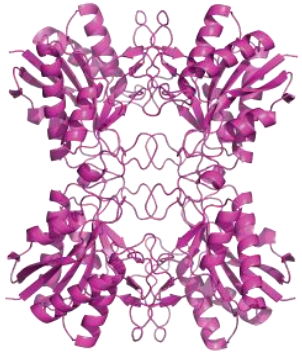




1A0H
MW: 96.8 kDa
RG: 31.5 Å



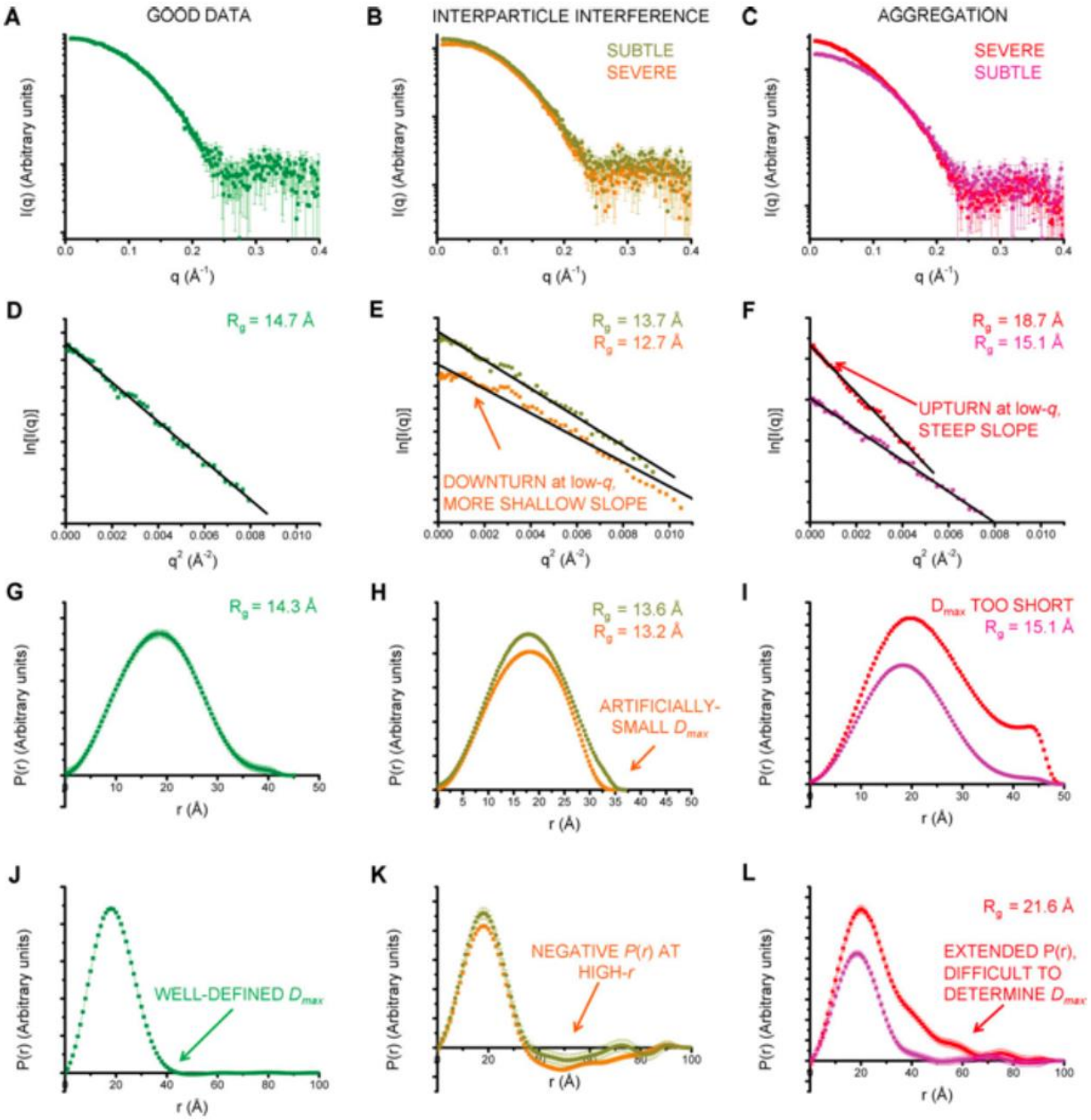
1A0J
MW: 96.6 kDa
RG: 32.4 Å



1A2Z
MW: 99.4 kDa
RG: 30.9 Å

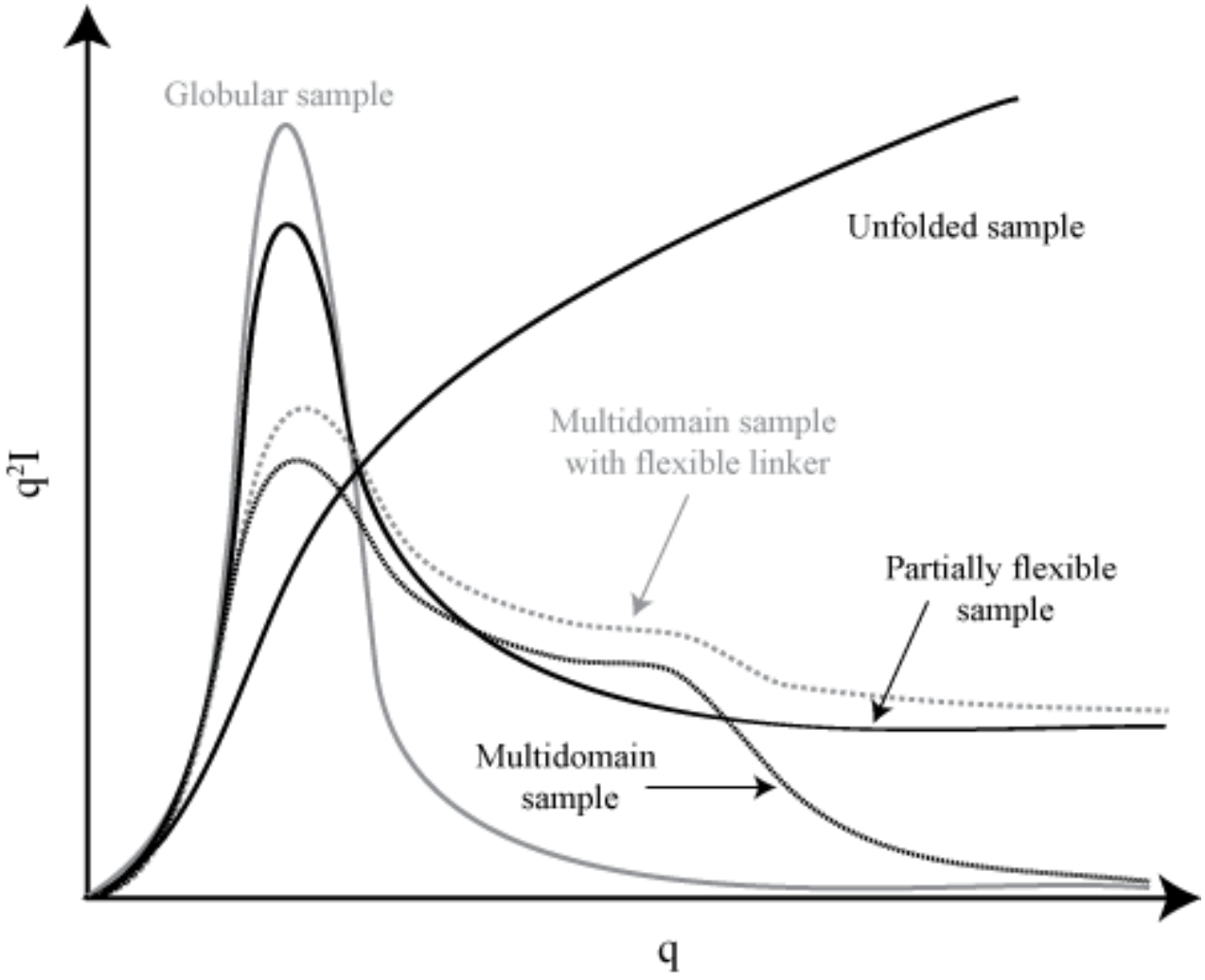
Quality in SAXS data

Sample quality greatly affects data analysis

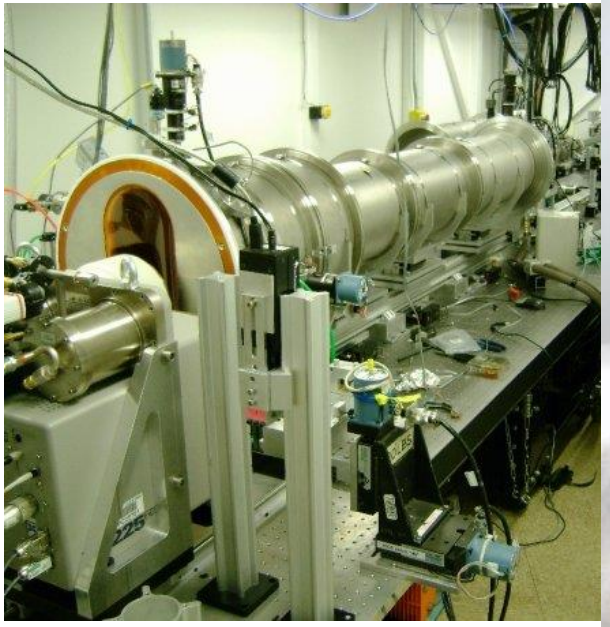


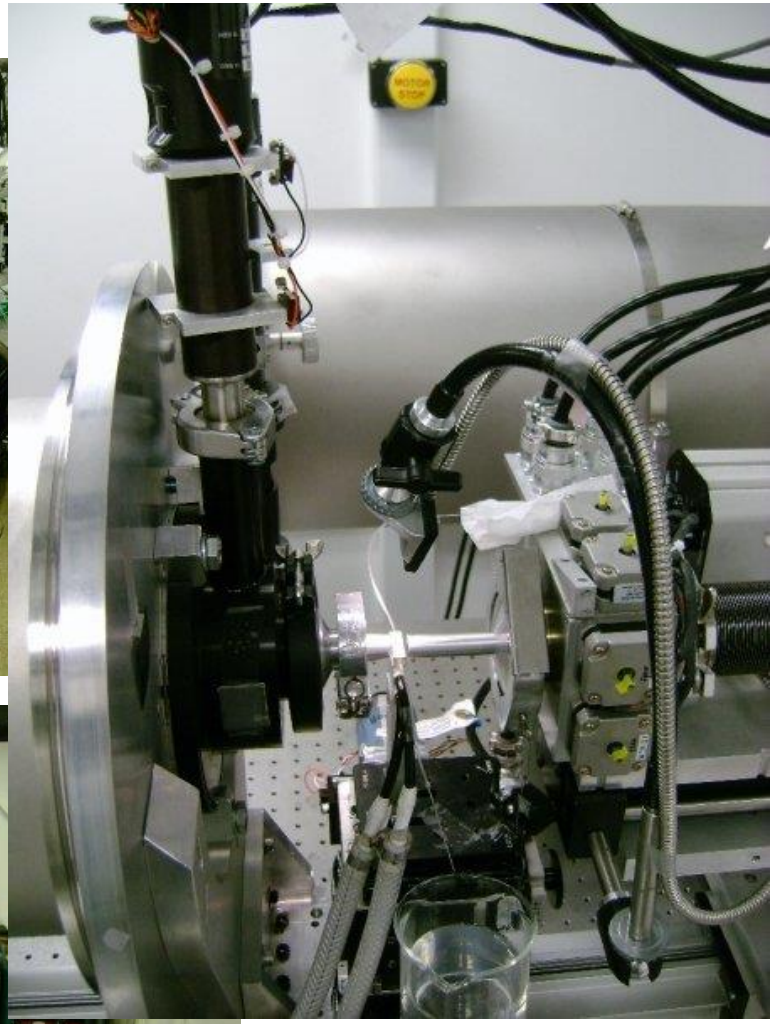
Jacques and Trewhella, 2010 *Protein Science* Review

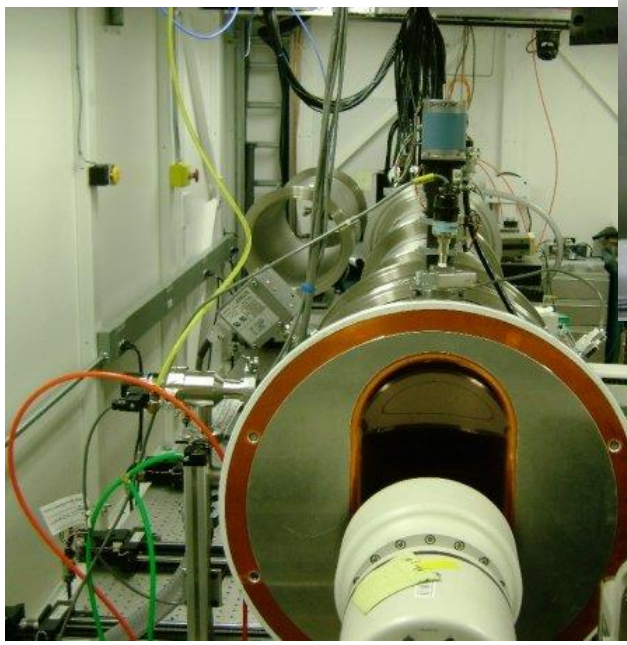
Characterization of samples from SAXS data


Globular sample
Unfolded sample
Multidomain sample
with flexible linker
Partially flexible
sample
Multidomain
sample
q²I
q

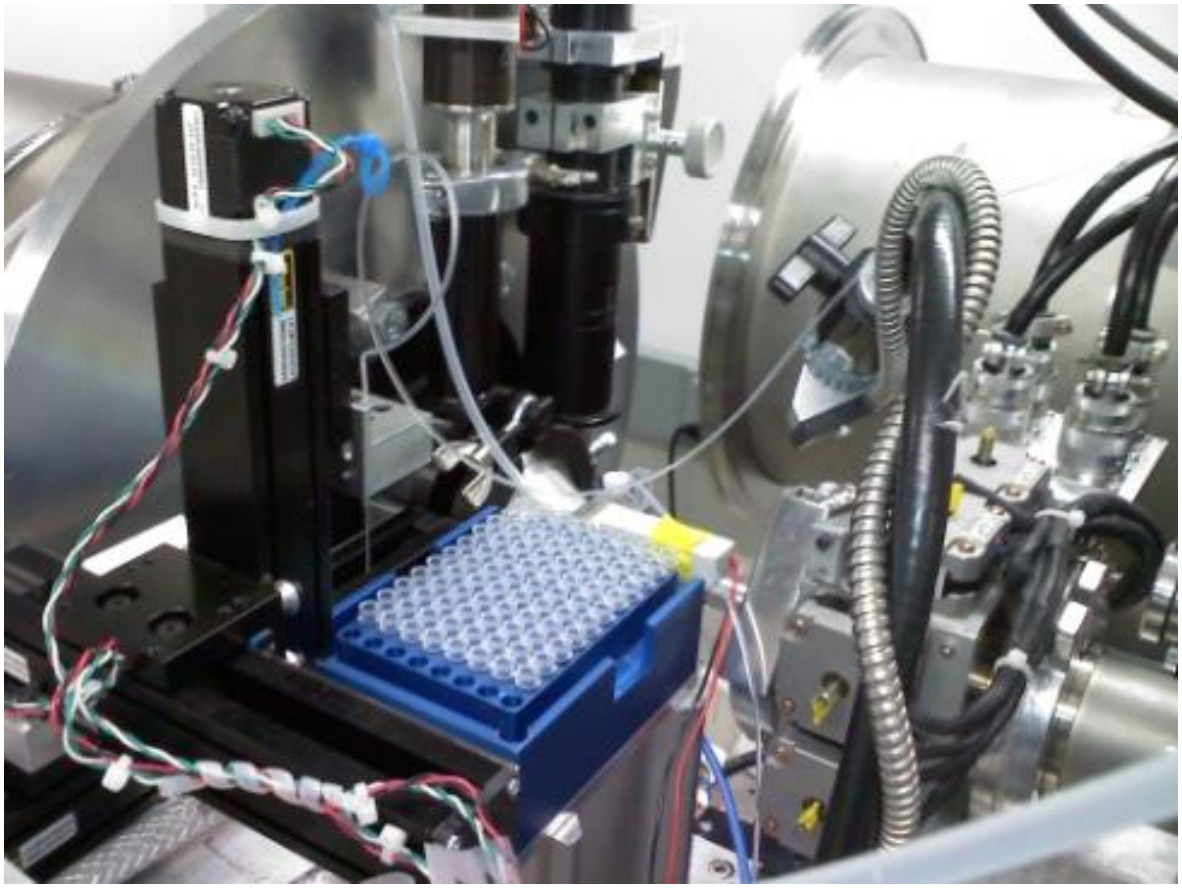
Practical SAXS data







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.

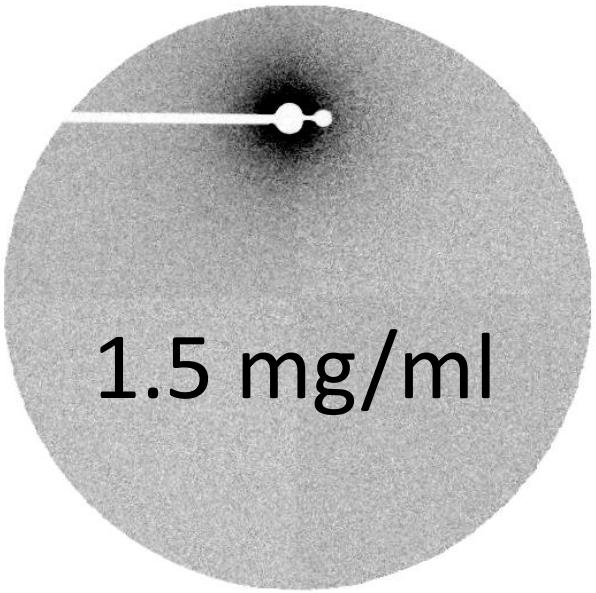
Oscillate sample to minimize radiation damage

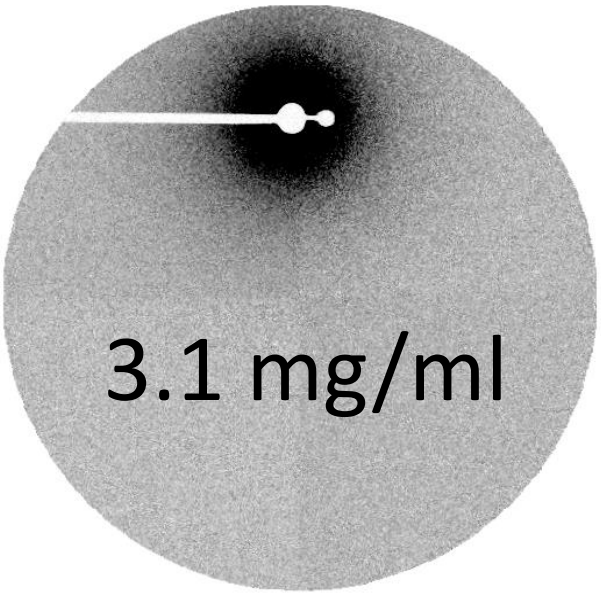
Repeat the buffer.

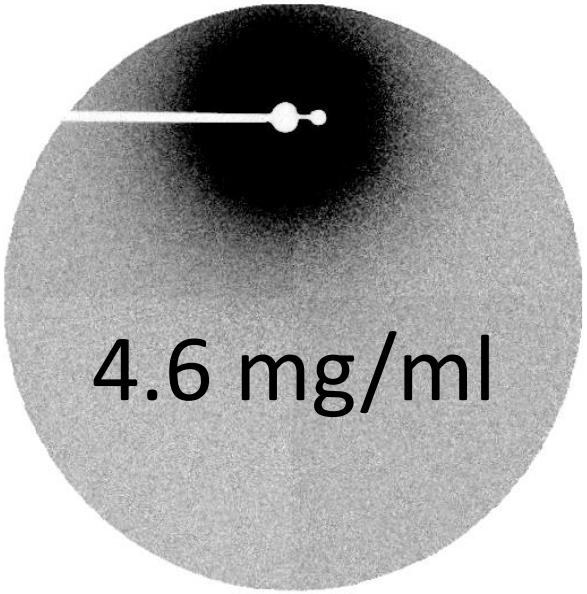
Load next sample

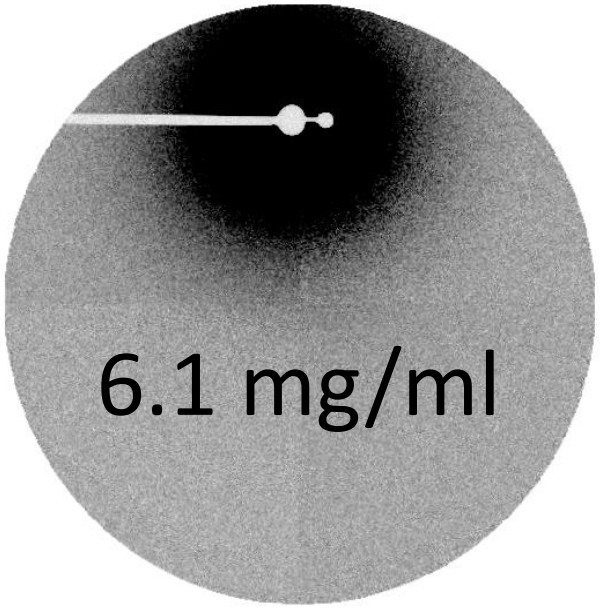
Time per concentration series – approximately 10 to 15 minutes. In high-throughput mode 24 samples in 3 to 4 hours.

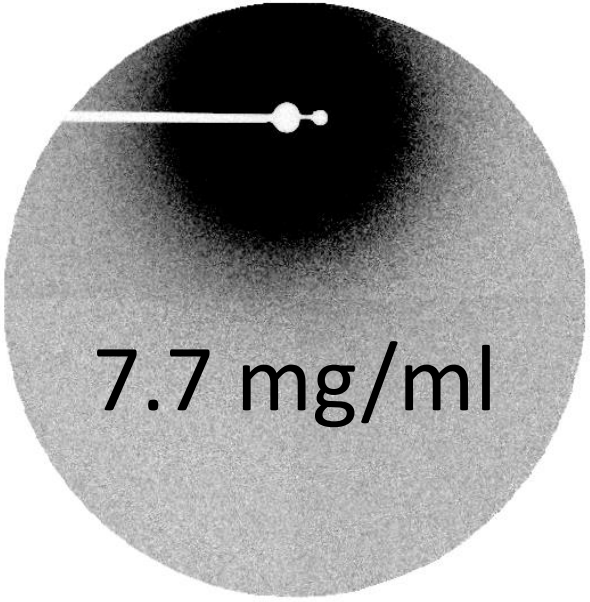
Enables two important things – eat and sleep!


1.5 mg/ml

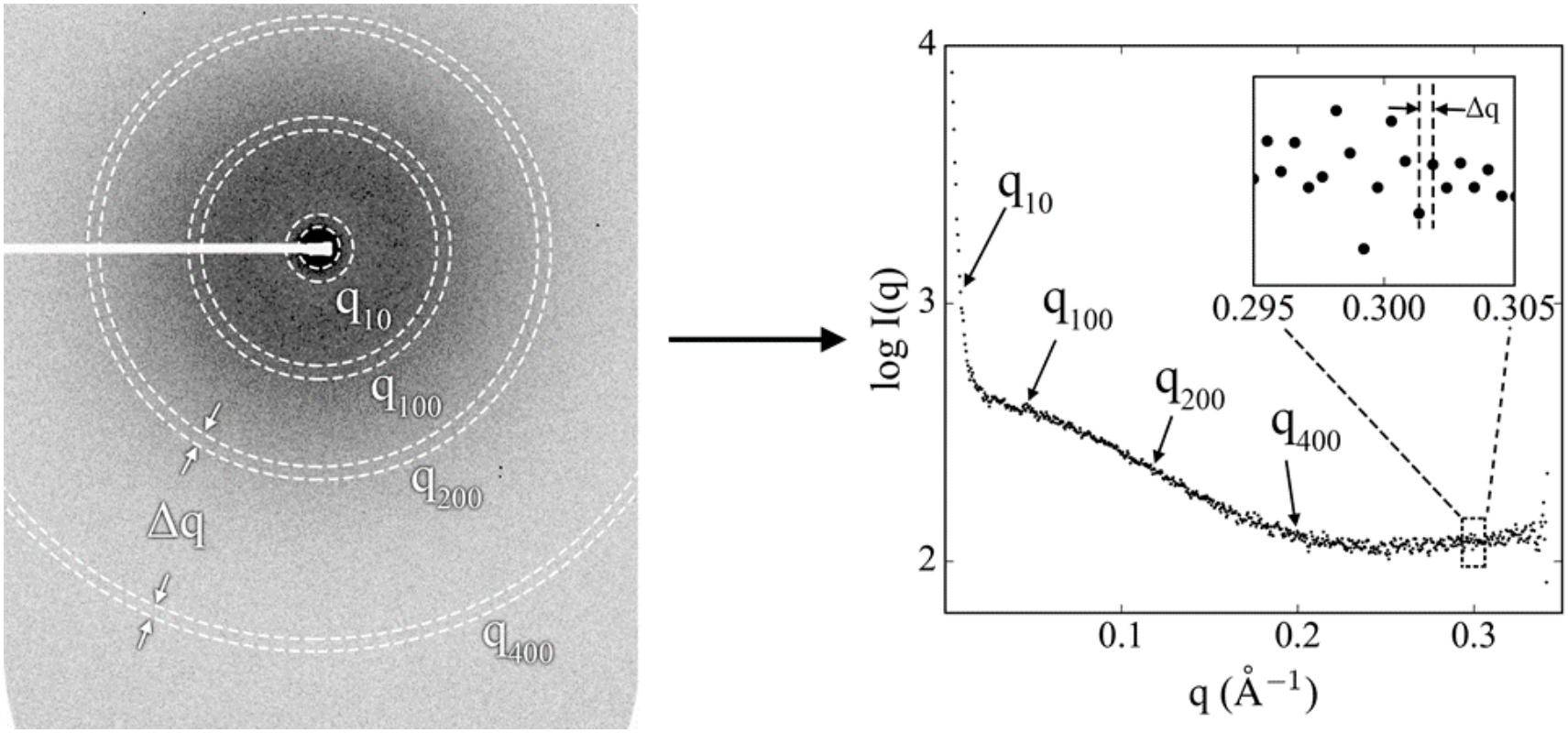

3.1 mg/ml

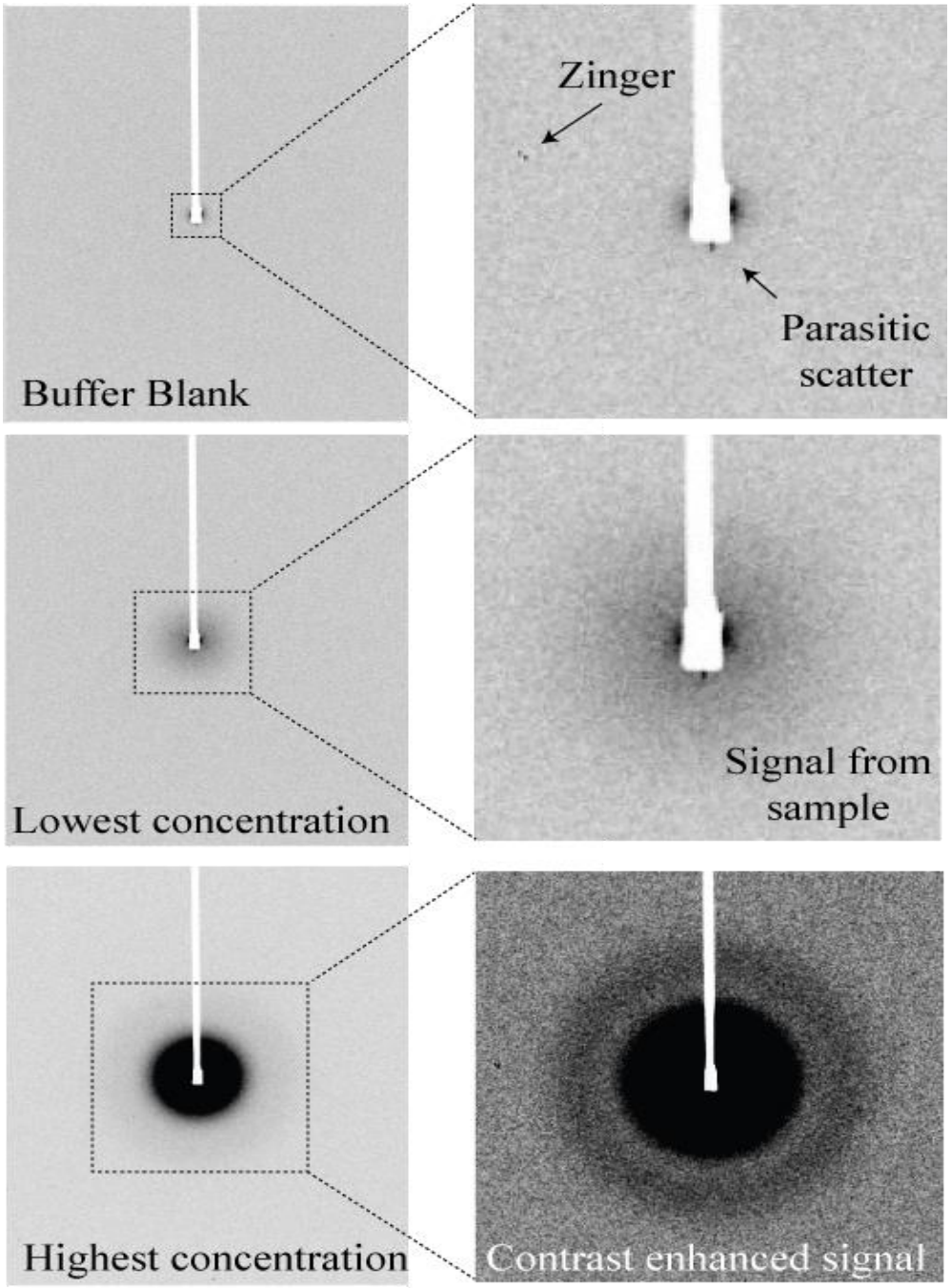

4.6 mg/ml


6.1 mg/ml

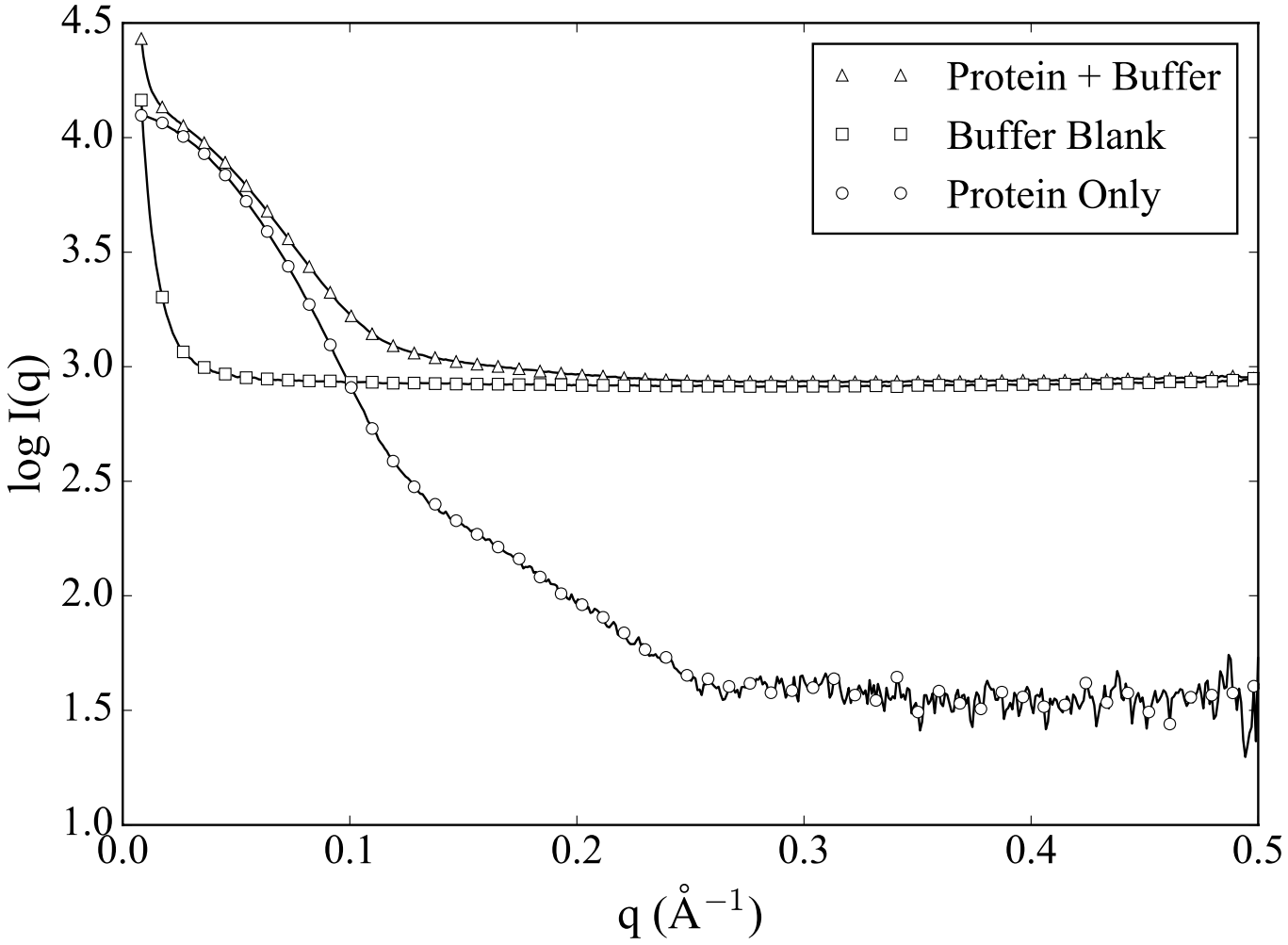

7.7 mg/ml

Radial integration with significant oversampling



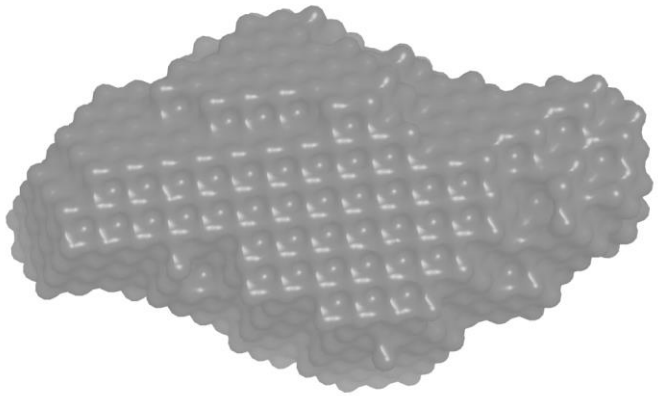


Small angle scattering data near the beamstop

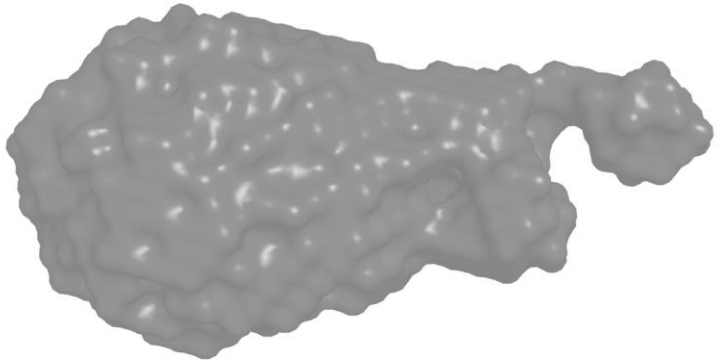

Protein + Buffer
Buffer Blank
Protein Only
log I(q)
q (Å^{-1})
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.0
0.1
0.2
0.3
0.4
0.5

SAXS can determine *ab initio* molecular envelopes

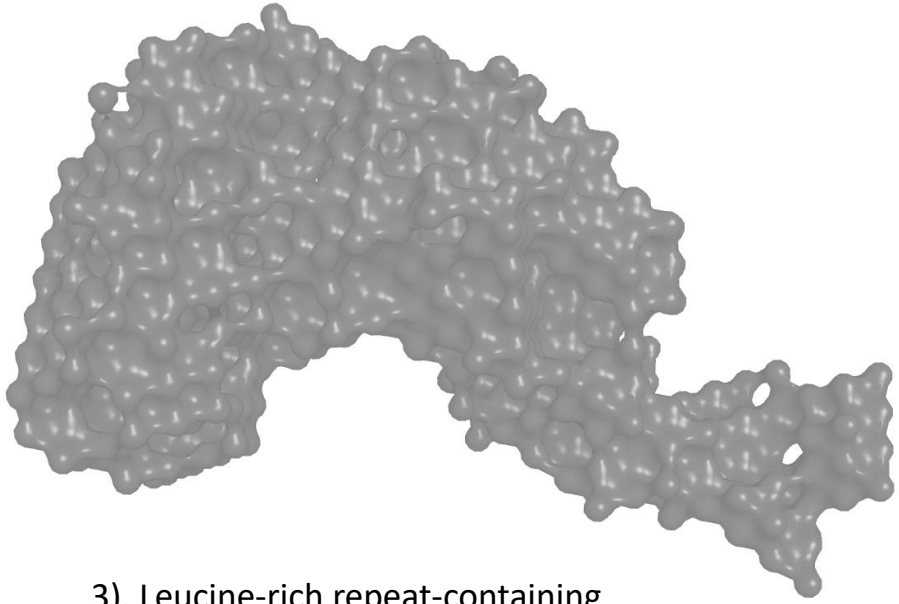
Ab intio envelopes



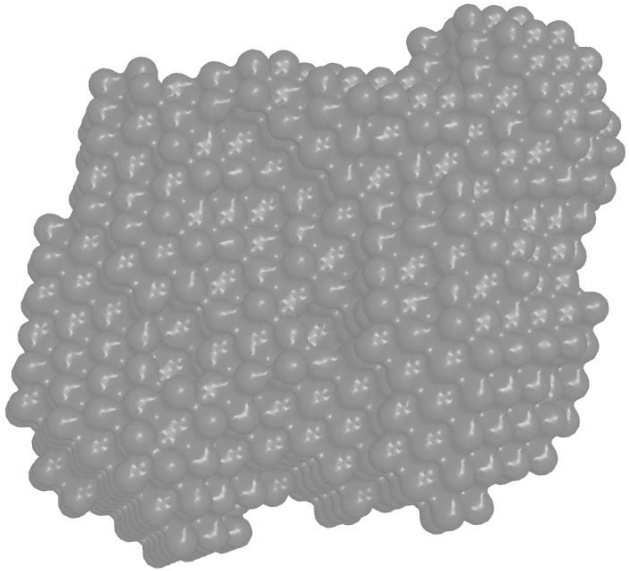
1). alr0221 protein from Nostoc (18.6 kDa)



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



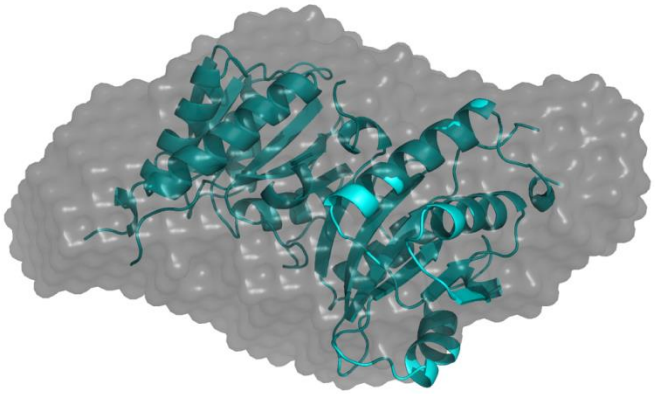
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



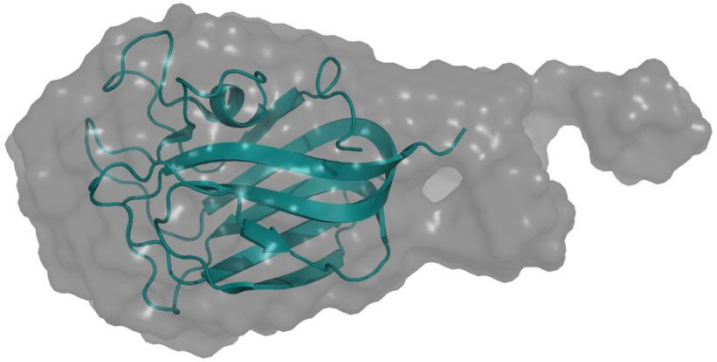
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

Ab initio envelopes are compatible with structural models

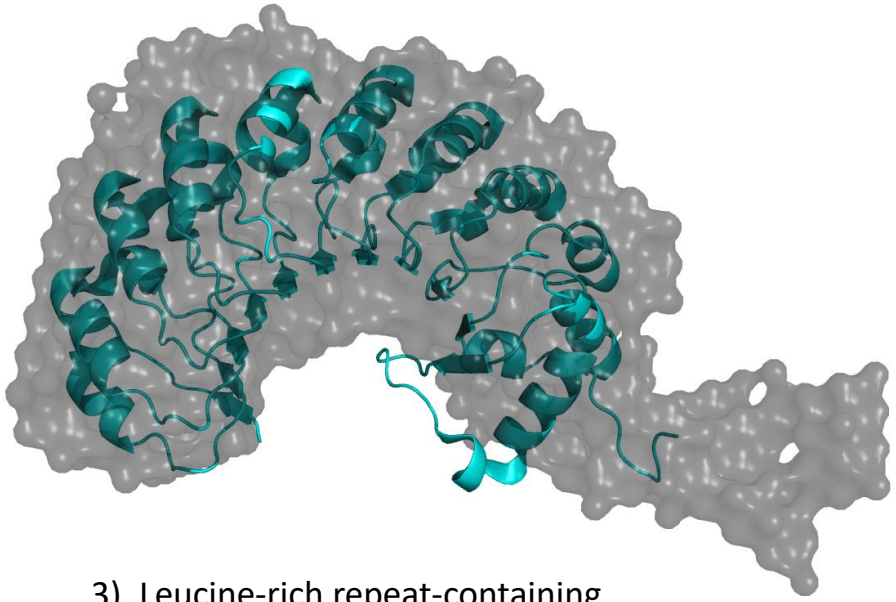
Overlaid with subsequent X-ray structures



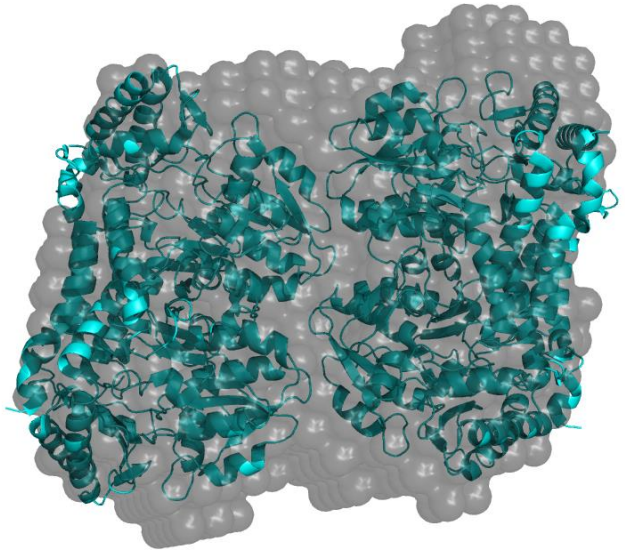
1). alr0221 protein from Nostoc (18.6 kDa)



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



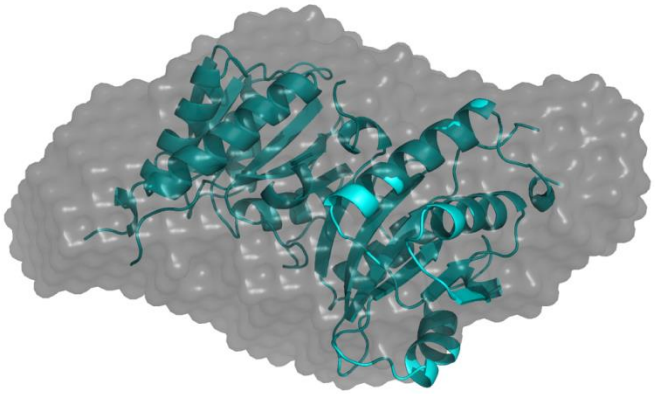
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



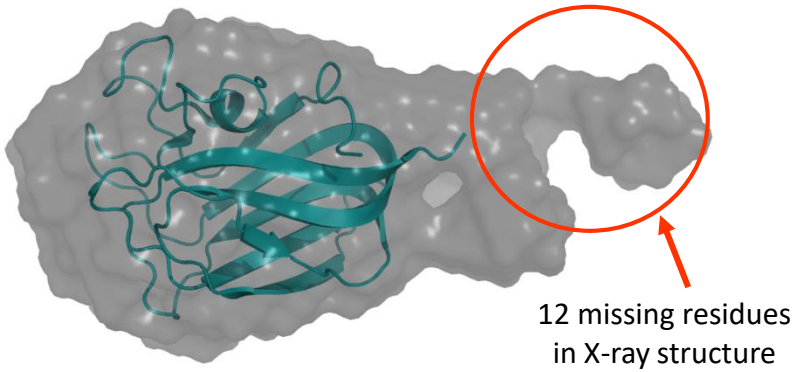
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

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

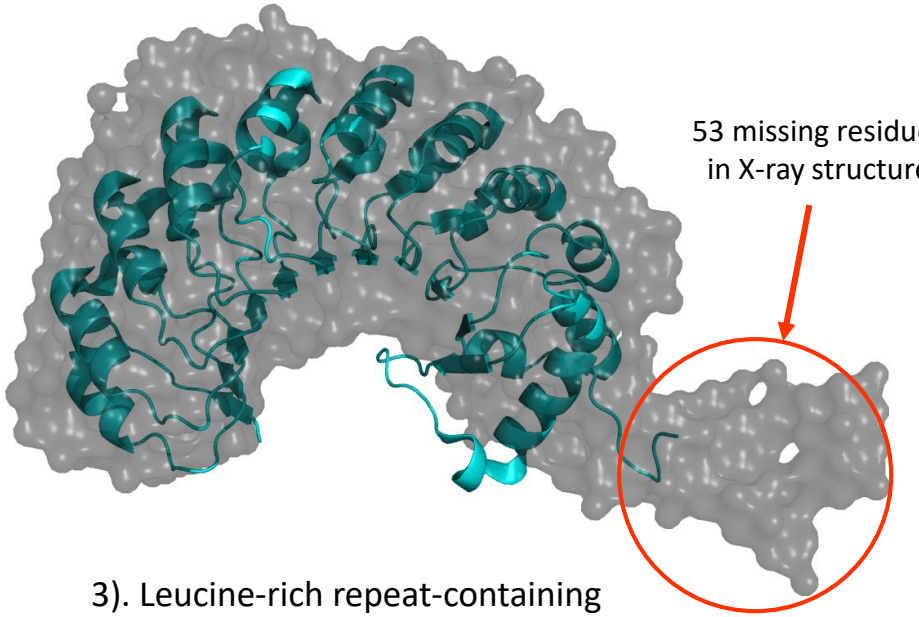
And data on what was missing …



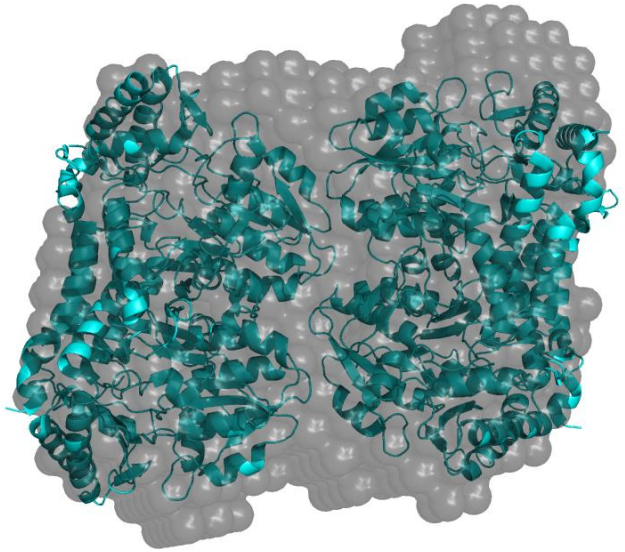
1). alr0221 protein from Nostoc (18.6 kDa)



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



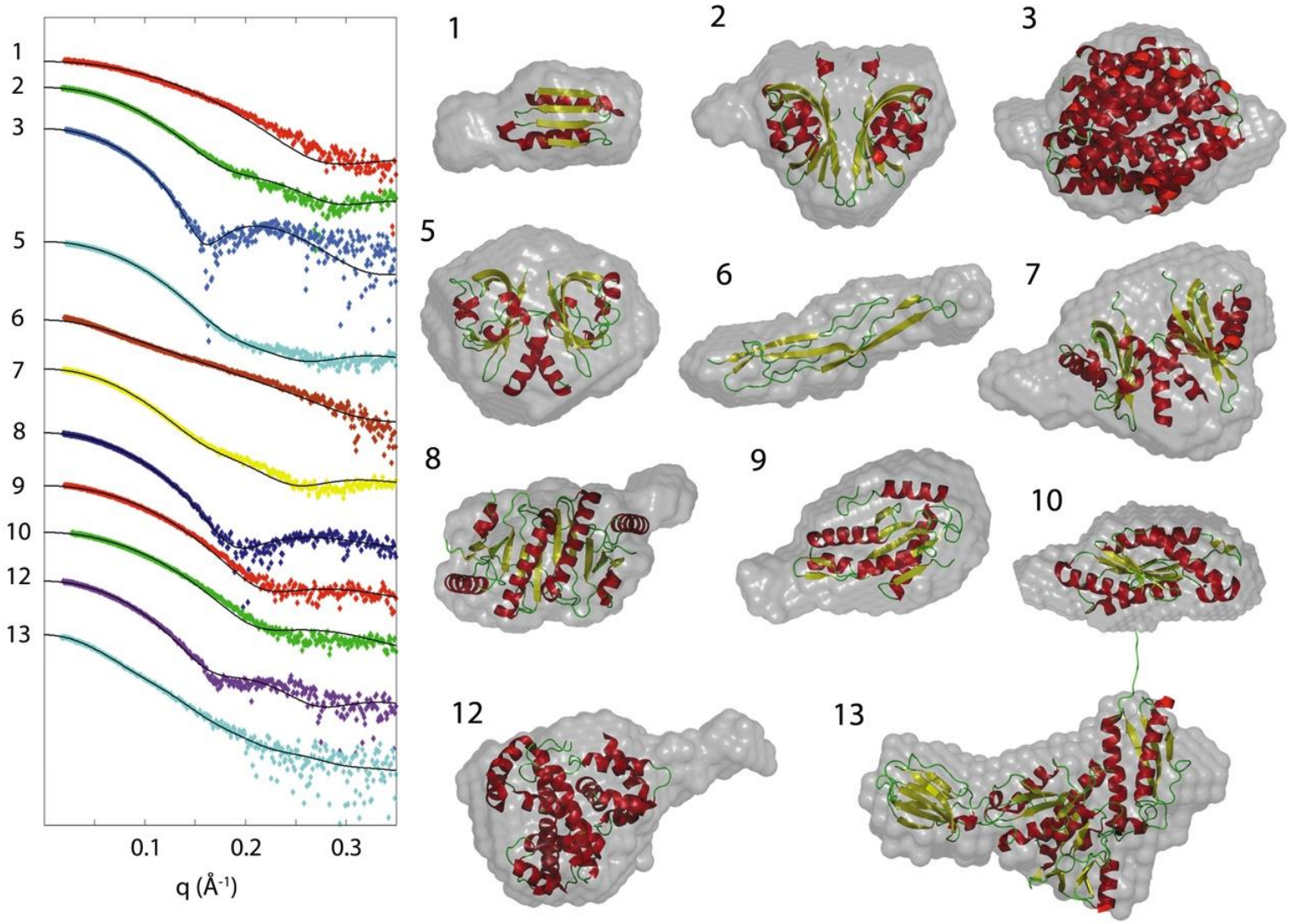
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

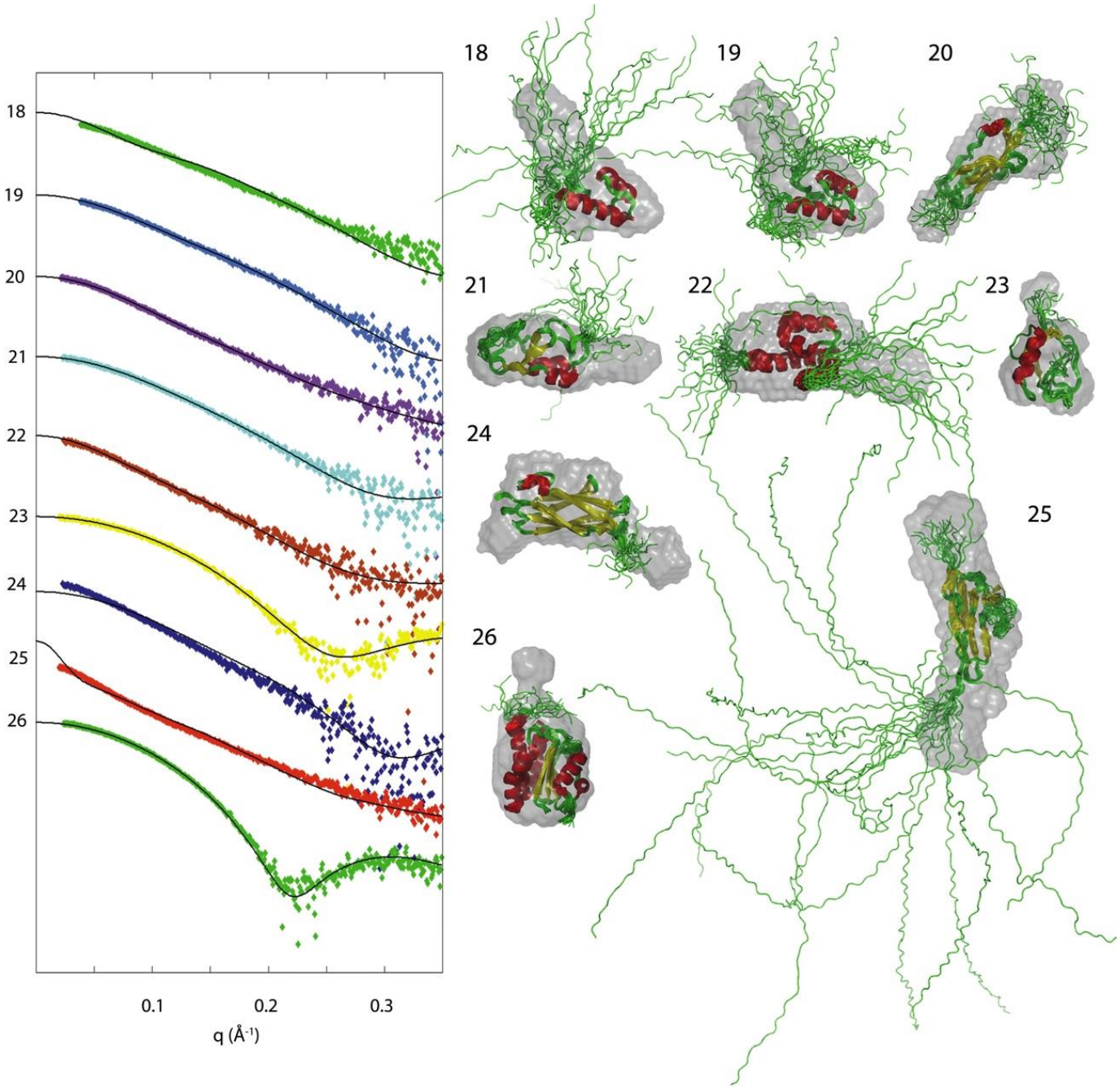


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

#	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

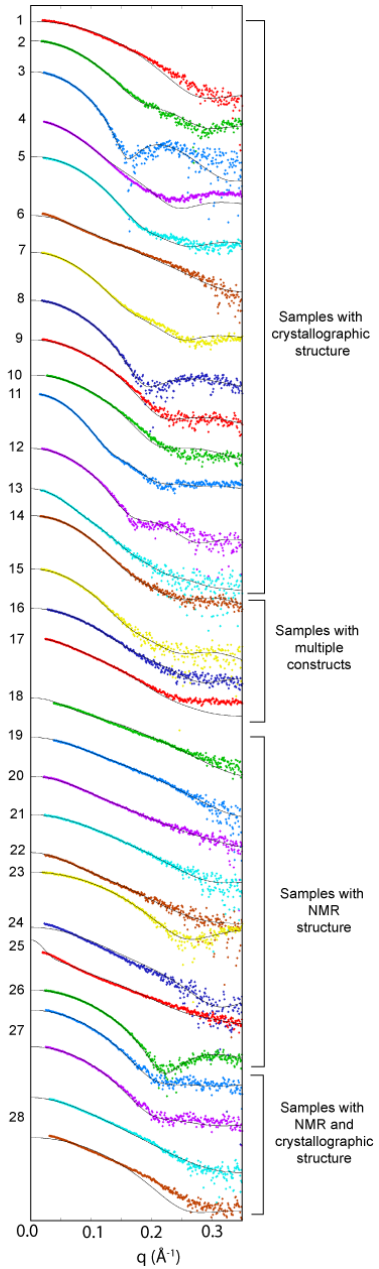
Comparing X-ray structures

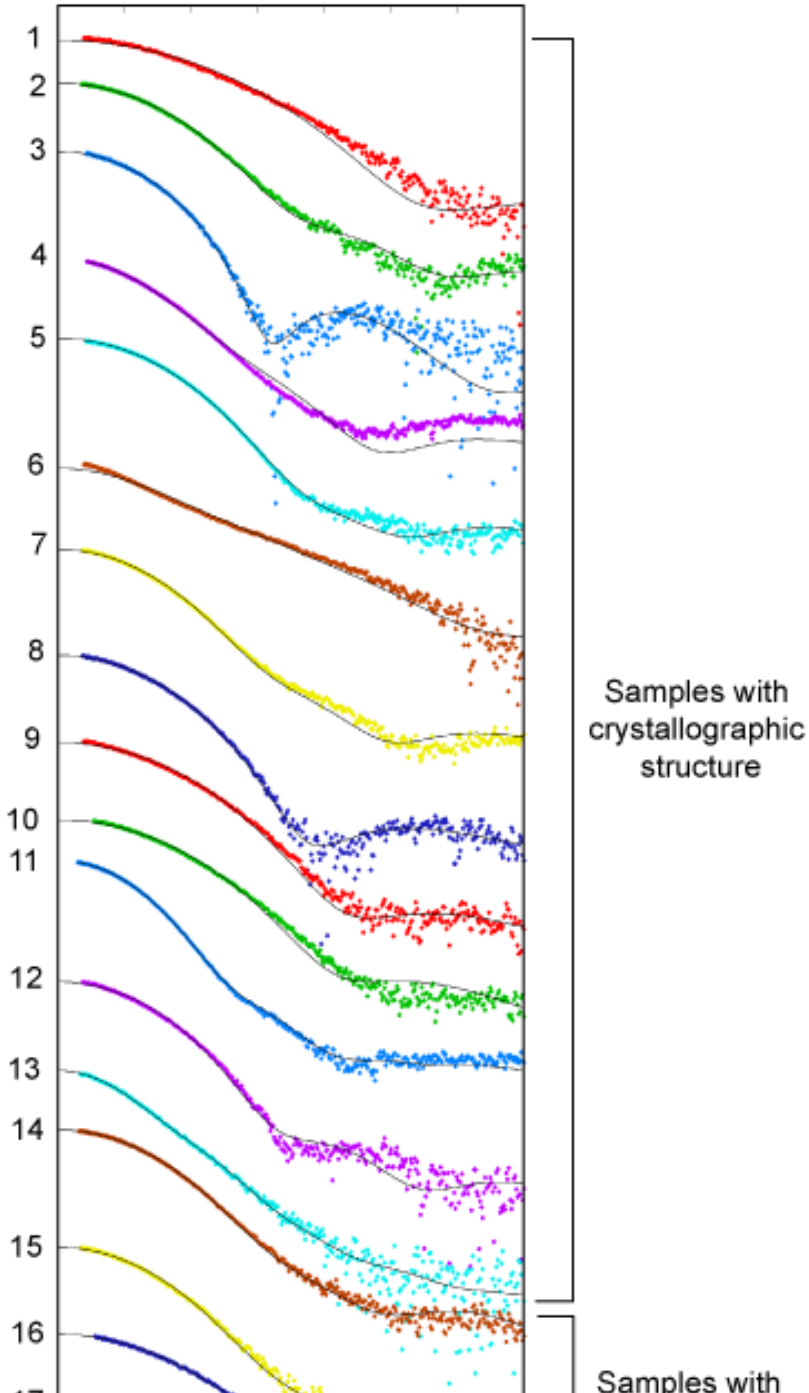




Comparing NMR structures

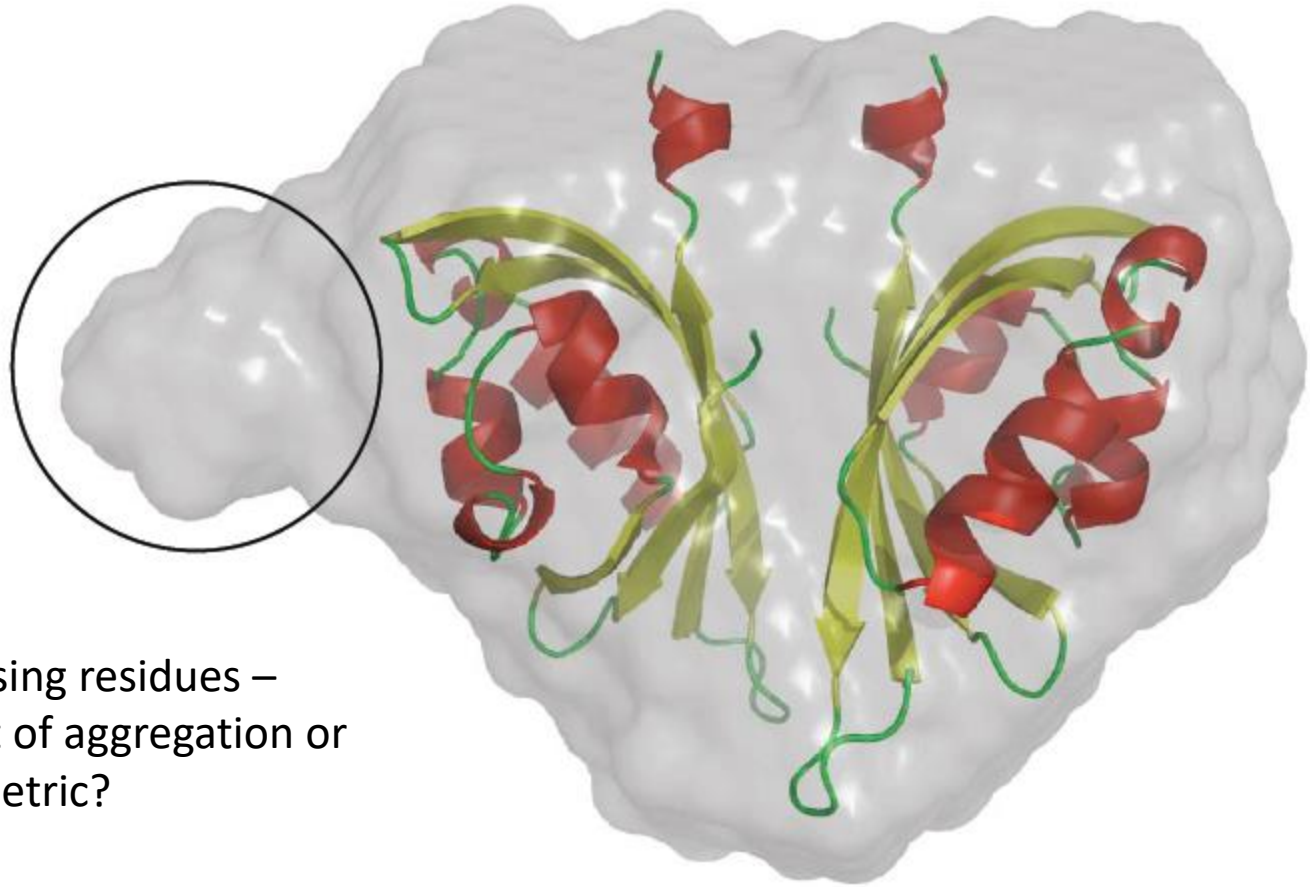
20 lowest energy Conformations shown


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
0.0
0.1
0.2
0.3
q (Å-1)
Samples with crystallographic structure
Samples with multiple constructs
Samples with NMR structure
Samples with NMR and crystallographic structure


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Samples with crystallographic structure
Samples with

SAXS may provide more questions

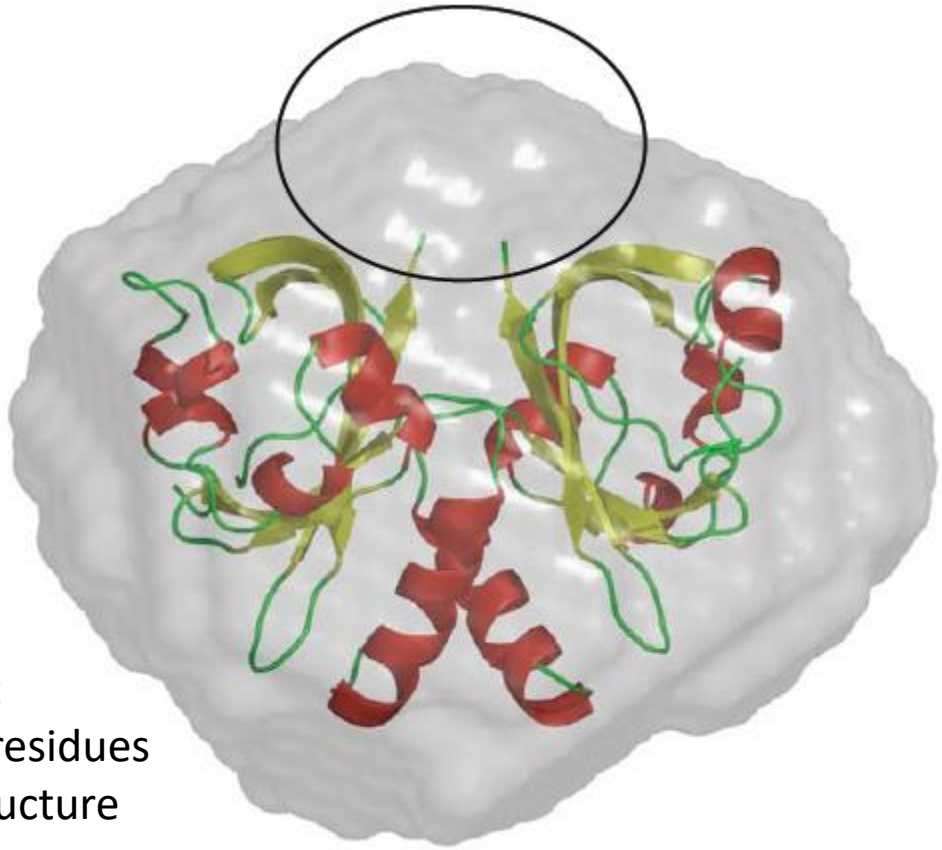
Diguanylate cyclase



12 missing residues – artifact of aggregation or asymmetric?

12 missing residues – artifact of aggregation or assymetric

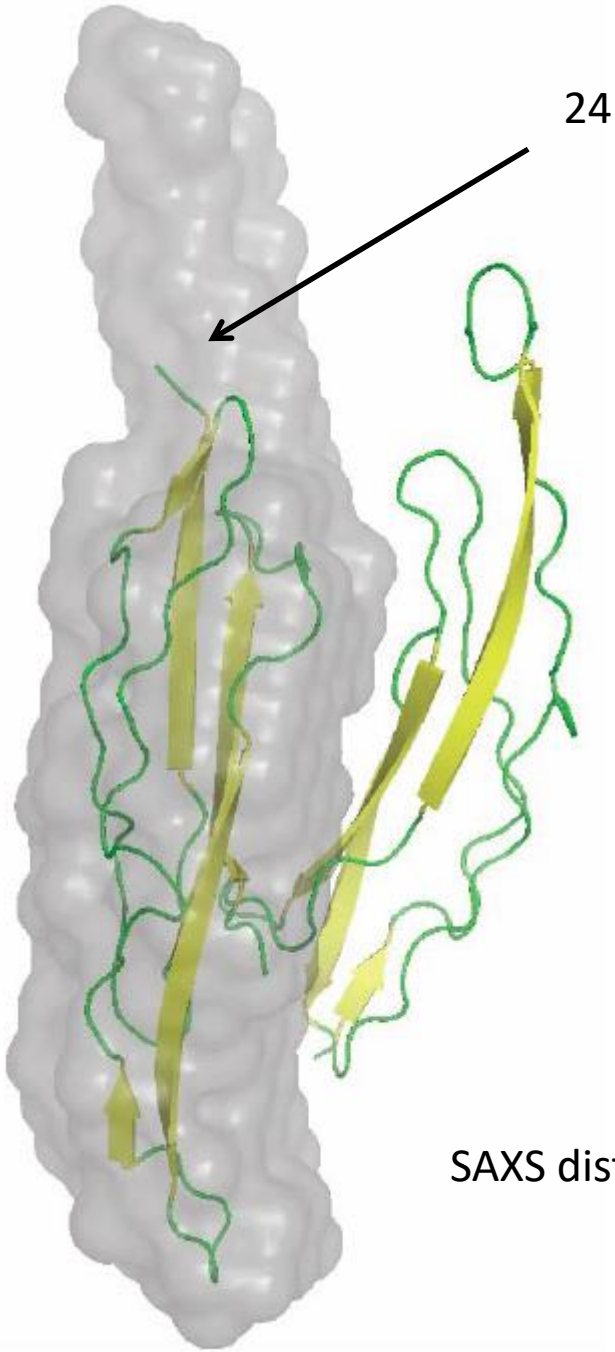
Sensory Box/GGDEF Protein Family



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

SAXS may be ambiguous

MucBP Domain of PEPE_0118



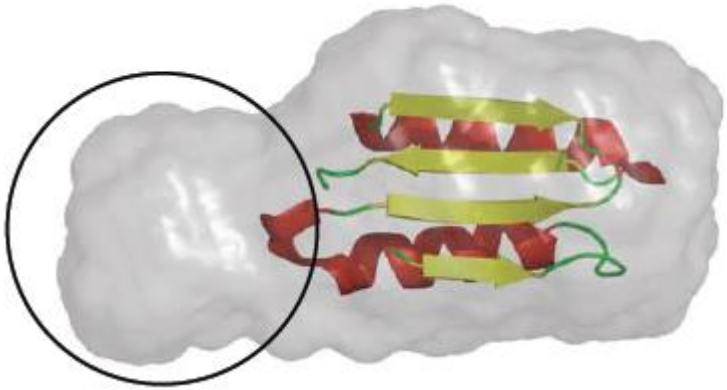
Biological unit was thought to be a dimer from crystallography.

Solution state is not.

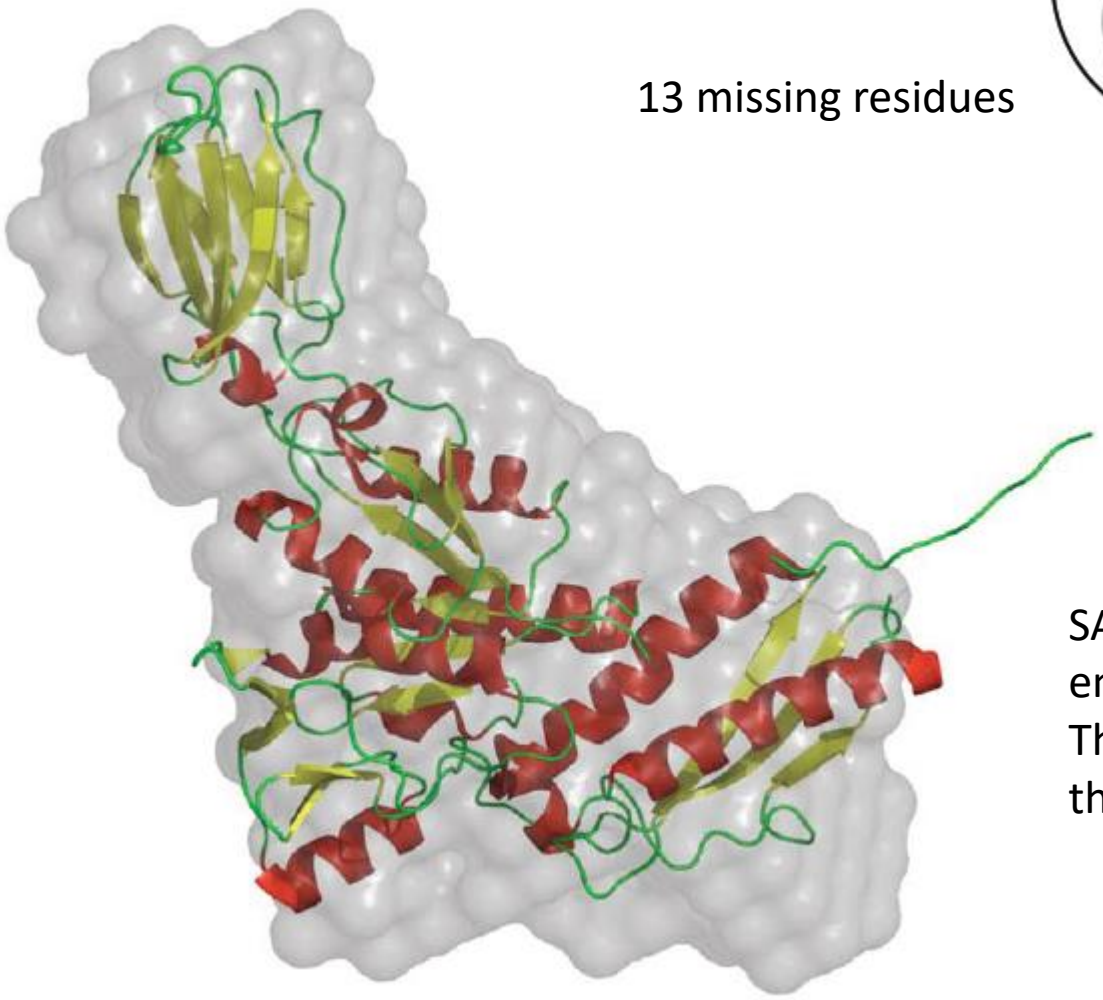
The biological state is not necessarily the solution or crystallographic state.

SAXS distinguishes solution states

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.

But ab initio shape reconstruction is the least useful capability for SAXS

- It is possible to accurately model a SAXS or SANS profile
- SAXS and SANS data provides
  - Molecular mass M
  - Radius of gyration R_g
  - Porod invariant Q
  - Particle volume V
  - Maximum particle dimension D_{max}
  - Particle surface area S
  - Correlation length l_c
  - Volume of correlation V_c
- SAXS can be used to test hypothesis ut not validate them.

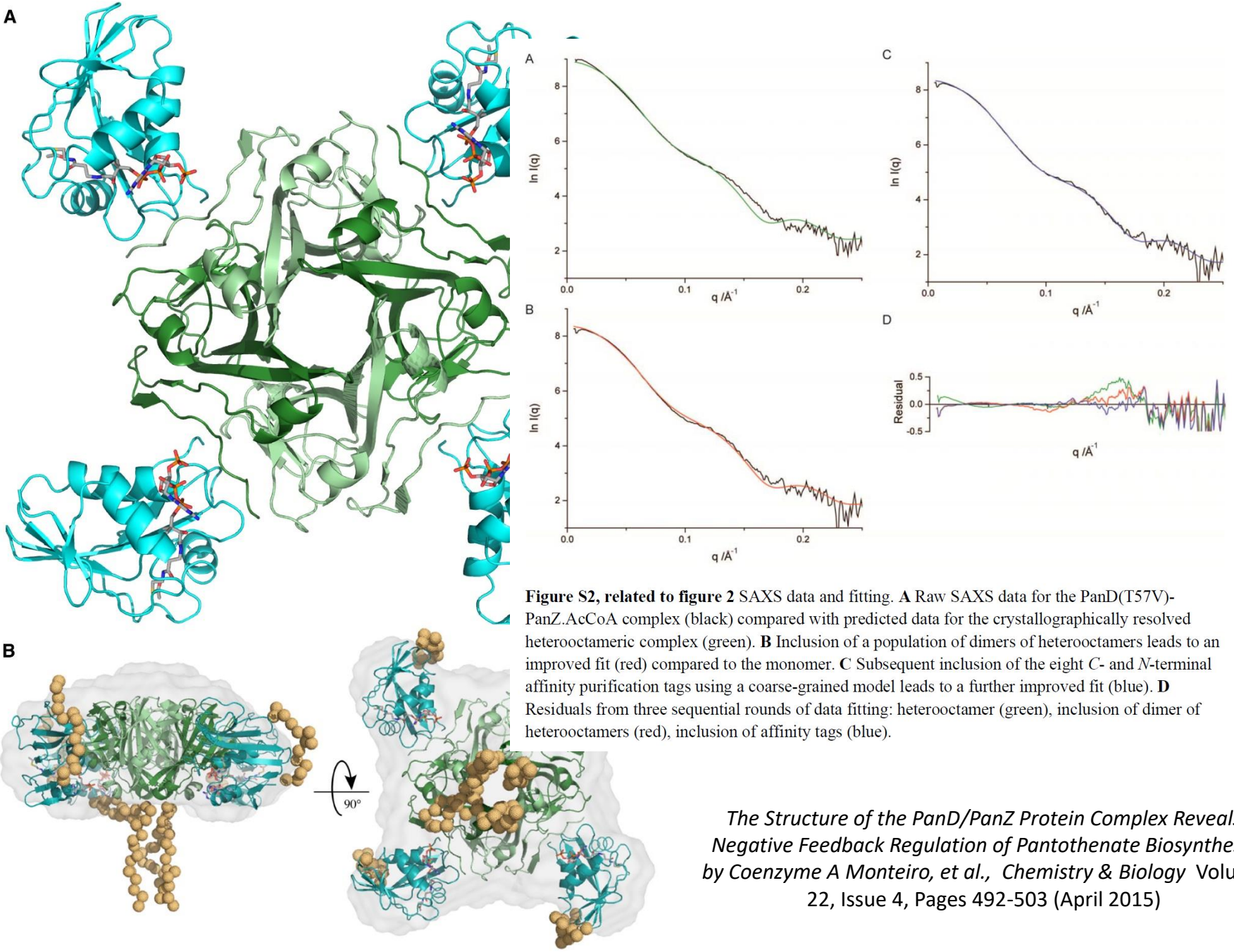


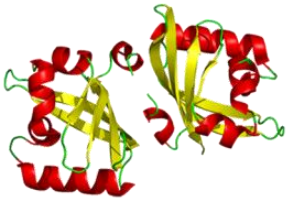
Figure S2, related to figure 2 SAXS data and fitting. **A** Raw SAXS data for the PanD(T57V)-PanZ.AcCoA complex (black) compared with predicted data for the crystallographically resolved heterooctameric complex (green). **B** Inclusion of a population of dimers of heterooctamers leads to an improved fit (red) compared to the monomer. **C** Subsequent inclusion of the eight *C*- and *N*-terminal affinity purification tags using a coarse-grained model leads to a further improved fit (blue). **D** Residuals from three sequential rounds of data fitting: heterooctamer (green), inclusion of dimer of heterooctamers (red), inclusion of affinity tags (blue).

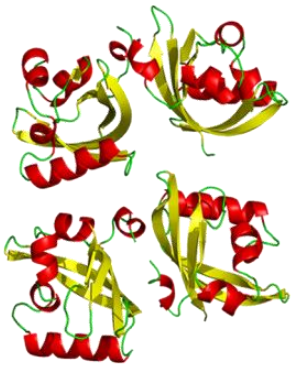
The Structure of the PanD/PanZ Protein Complex Reveals Negative Feedback Regulation of Pantothenate Biosynthesis by Coenzyme A Monteiro, et al., Chemistry & Biology Volume 22, Issue 4, Pages 492-503 (April 2015)

63% Dimer

37% Tetramer

4

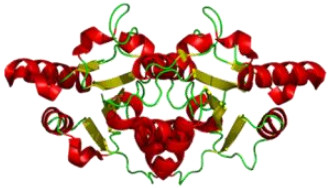


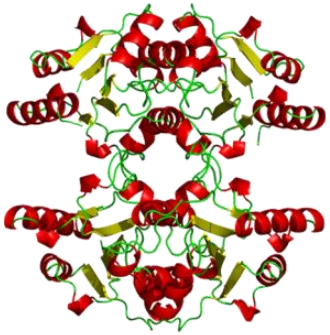


47% Dimer

53% Tetramer

11

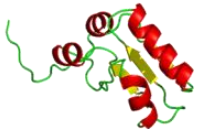


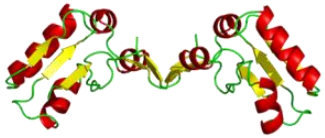


41% Monomer

59% Dimer

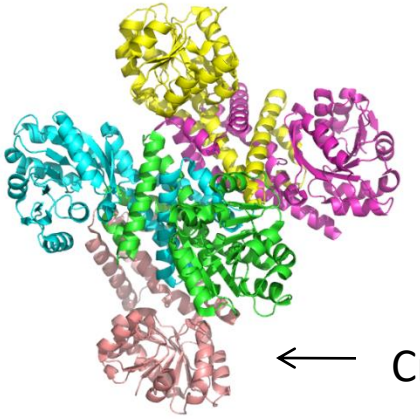
17





Identification of mixtures if you know the initial structure (another story)

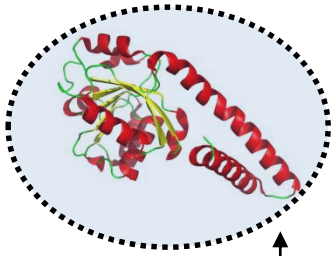
Another story



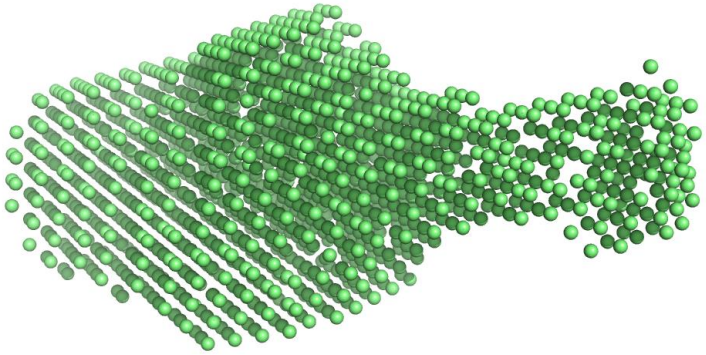
2izz from the PDB
(5 chains in PDB)

← Crystal packing artifact

3gt0 from the PDB

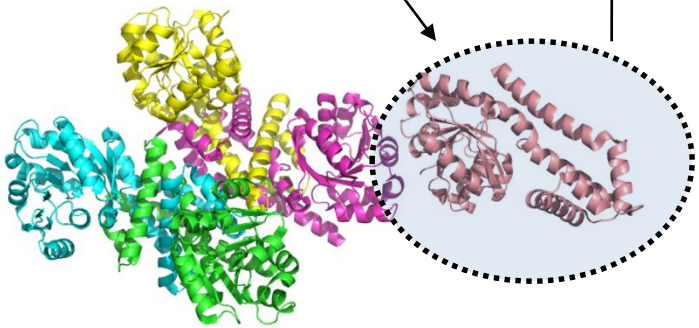


Solution envelope from BcR38B-21.20-SeMa-Gf (3gt0)



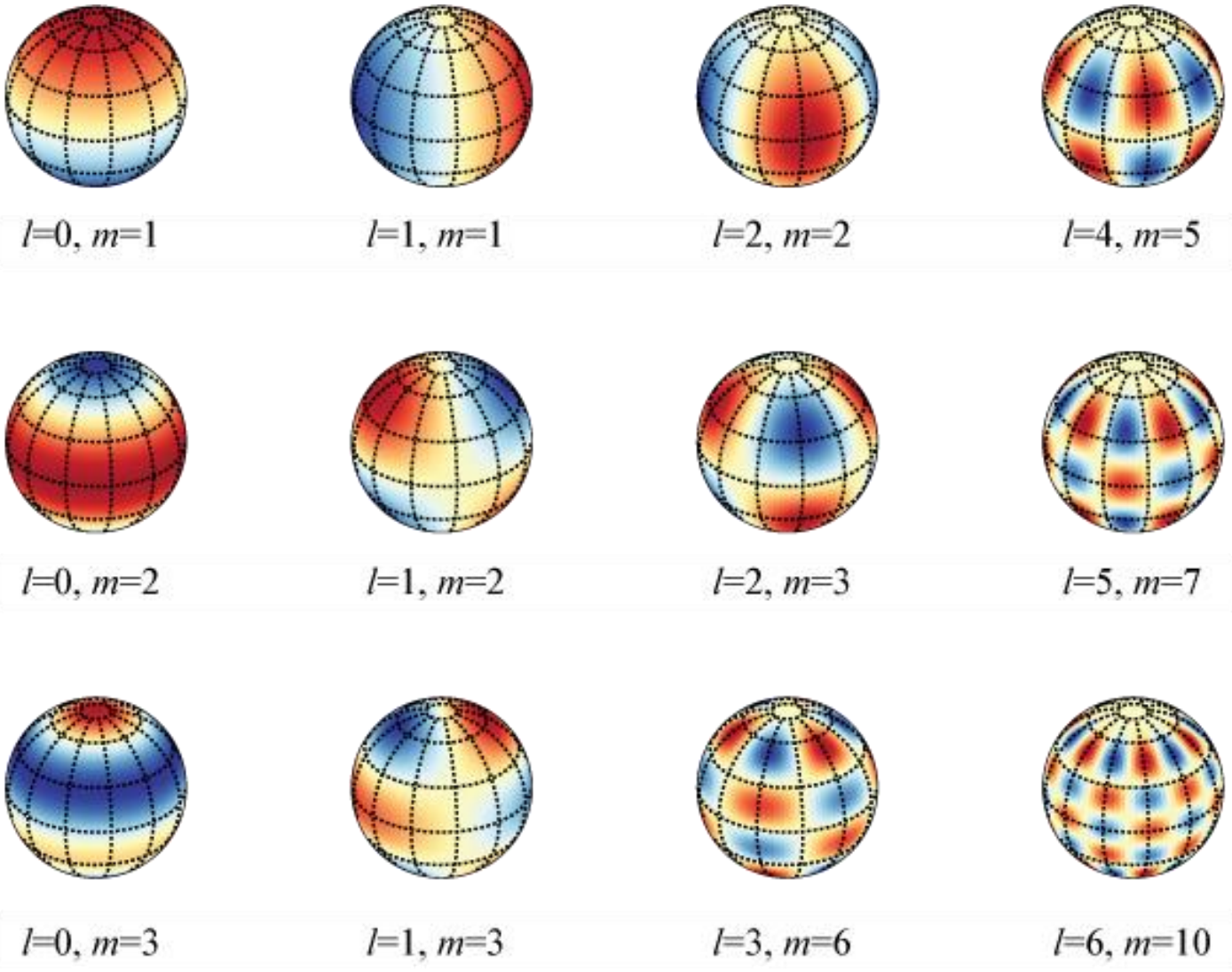
~165A

Correct position for 5th chain



~165A

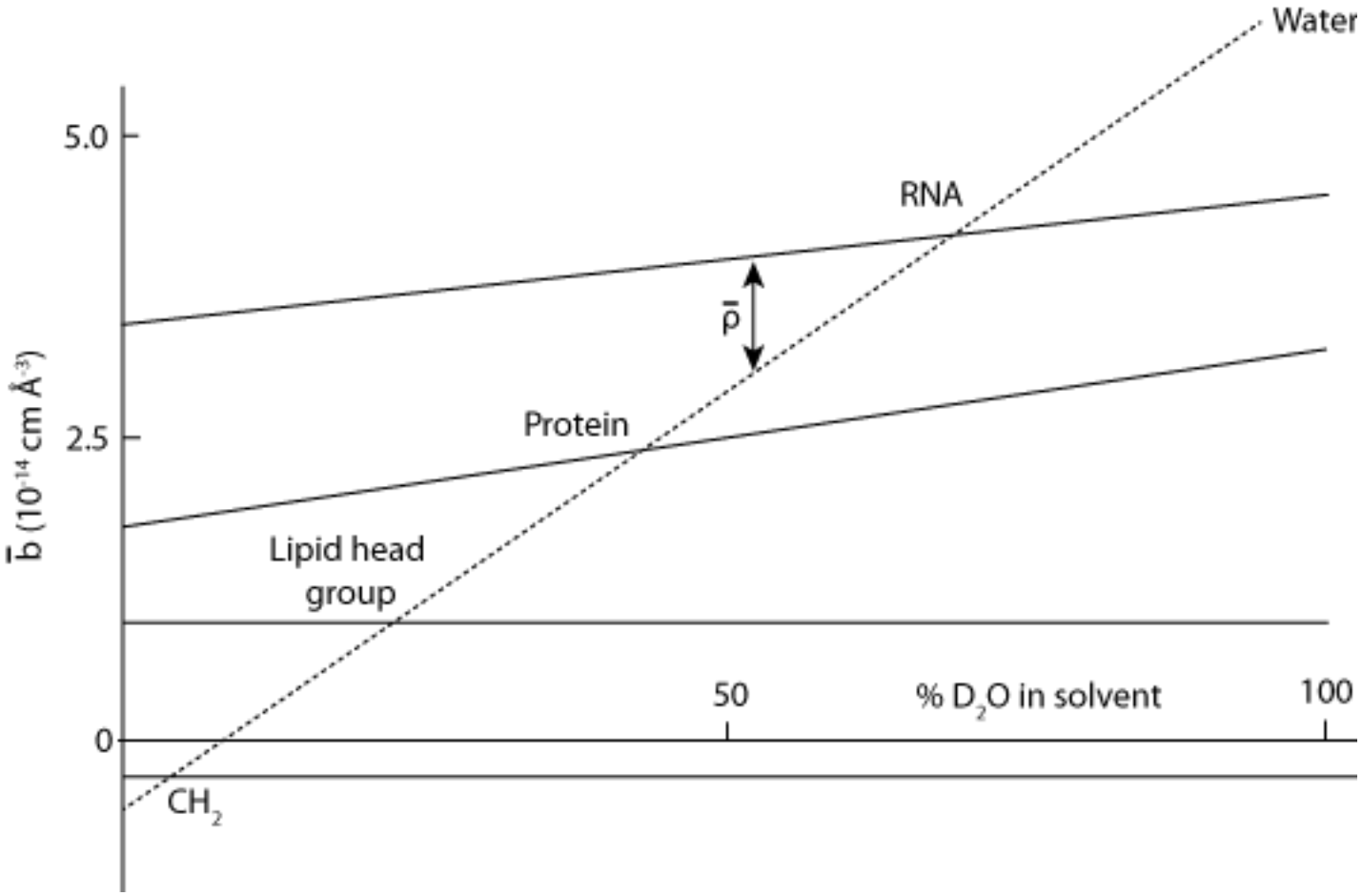
Biological unit based
on 2izz and SAXS

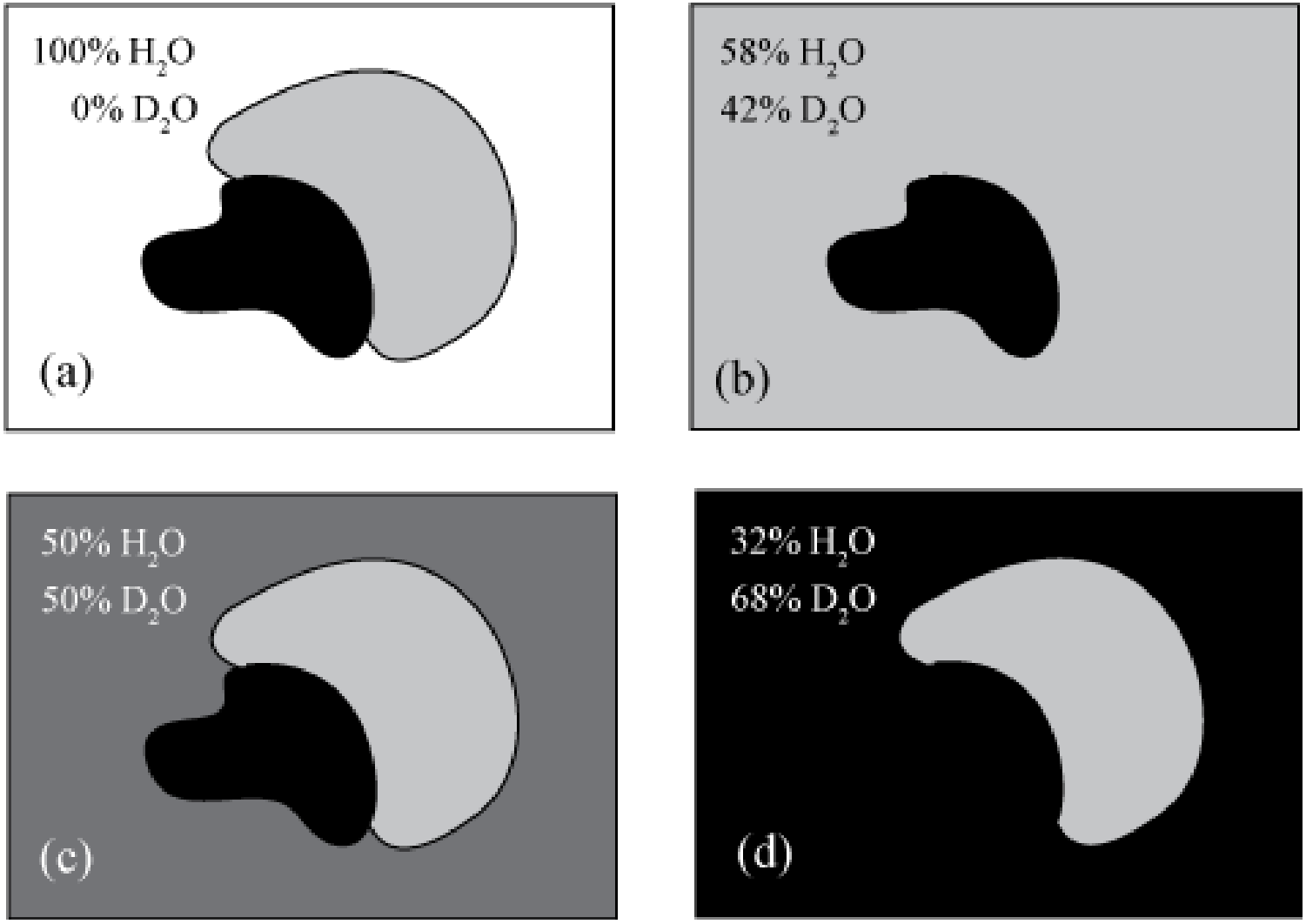

l=0, m=1
l=1, m=1
l=2, m=2
l=4, m=5
l=0, m=2
l=1, m=2
l=2, m=3
l=5, m=7
l=0, m=3
l=1, m=3
l=3, m=6
l=6, m=10

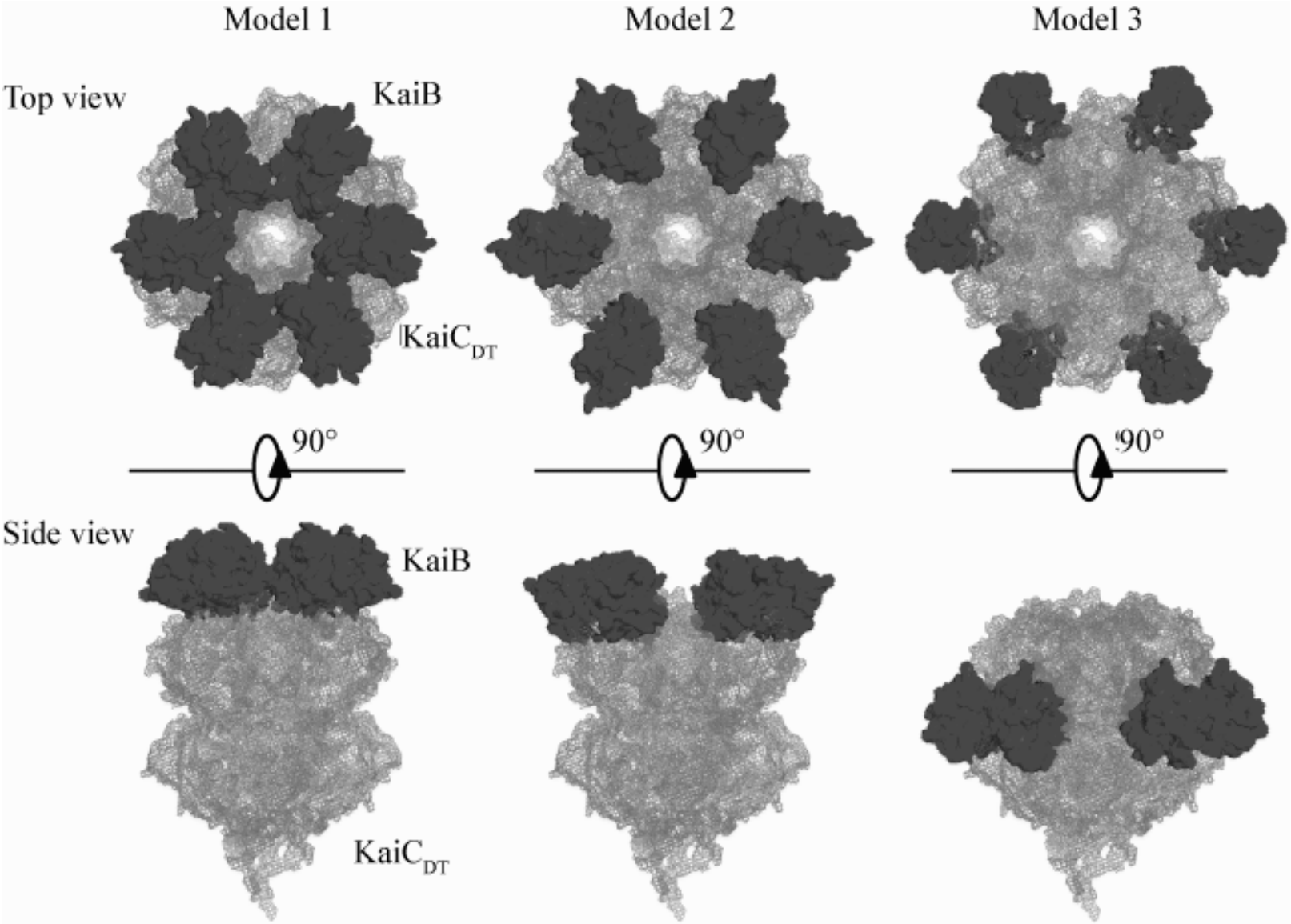
Small Angle Scattering with Neutrons

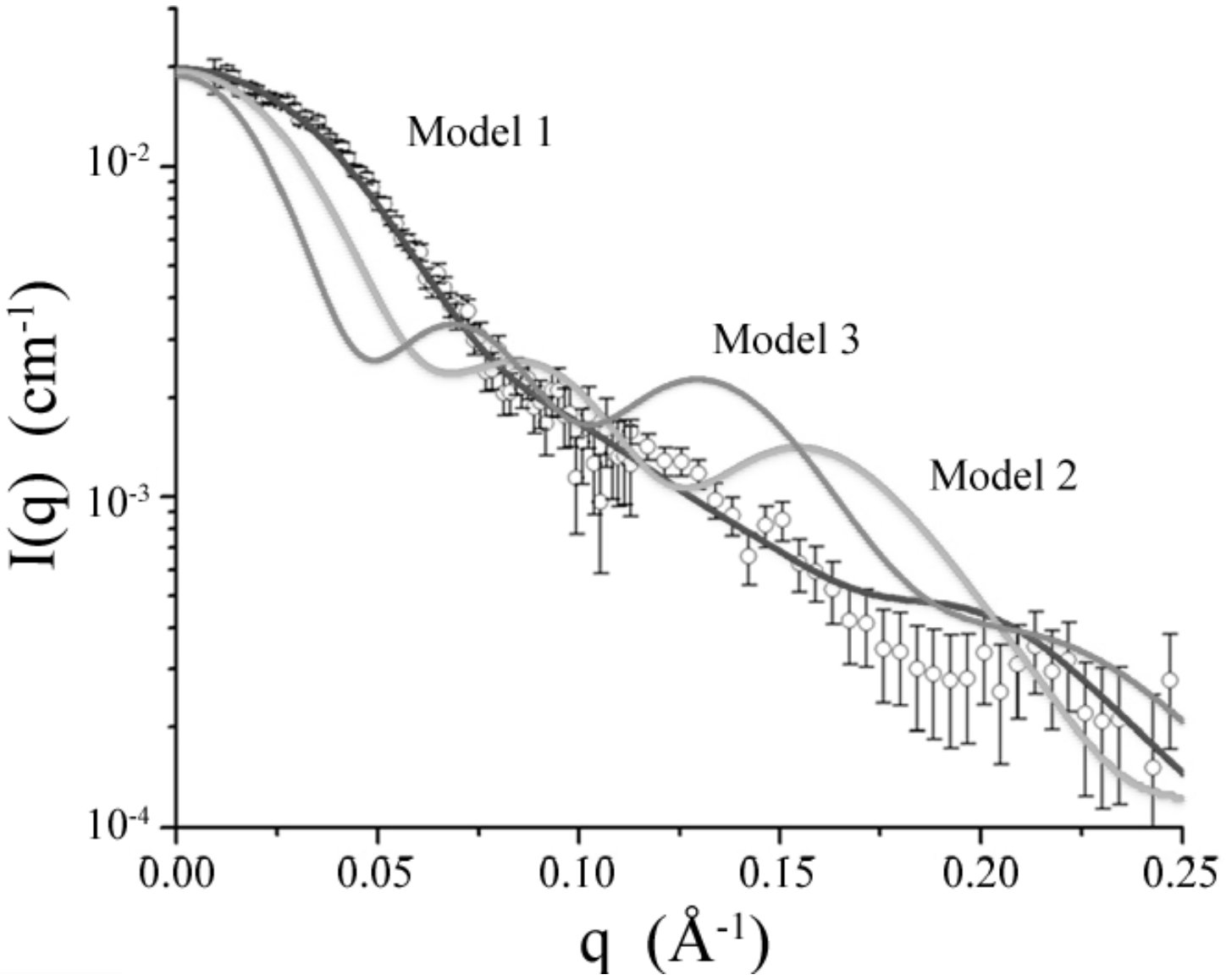

NORTH

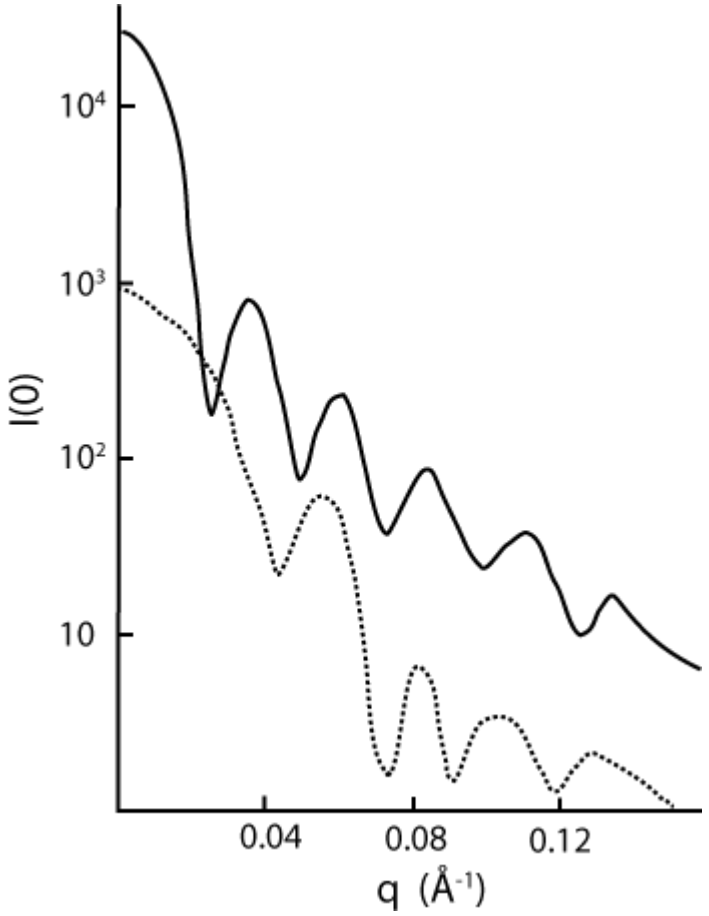
Contrast matching (more difficult in the X-ray case)




100% H_2O
0% D_2O
(a)
58% H_2O
42% D_2O
(b)
50% H_2O
50% D_2O
(c)
32% H_2O
68% D_2O
(d)


Model 1
Model 2
Model 3
Top view
KaiB
KaiC$_{DT}$
90°
90°
90°
Side view
KaiB
KaiC$_{DT}$


Model 1
Model 3
Model 2
I(q) (cm-1)
10-2
10-3
10-4
0.00
0.05
0.10
0.15
0.20
0.25
q (Å-1)

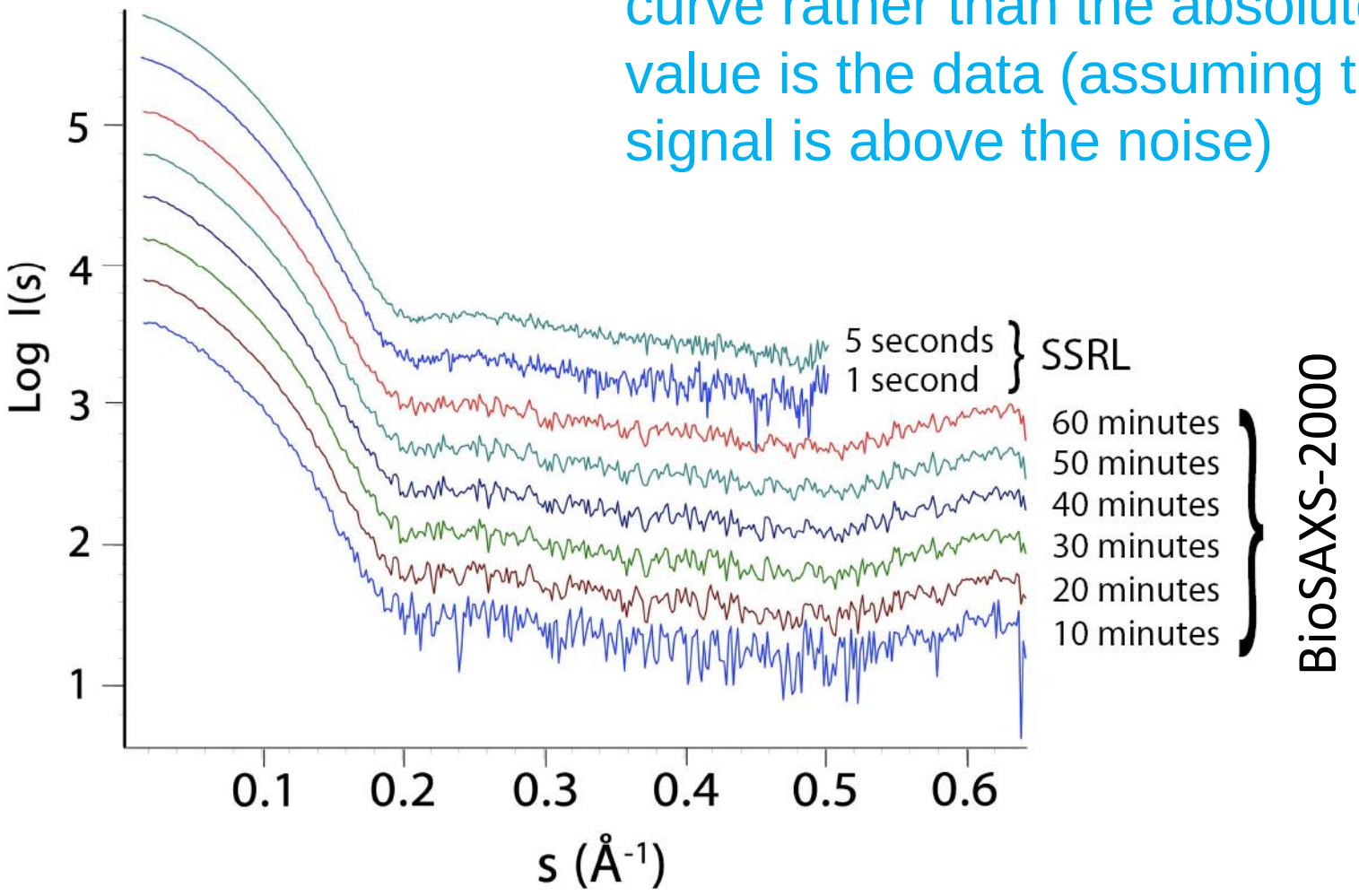


Scattering curve from Southern Bean Mottle virus in solutions of different D_2O content. The continuous line with 69.5% D2O and scattering mostly due to the protein shell and the dashed line with solvent content 42% D_2O and the scattering mostly by the nuclein acid (RNA). The subsidiary maxima are shifted to a larger q which indicate that the sphere that approximates the volume occupied by the RNA has a smaller diameter than the virus (Chauvin et al., 1976).

SAXS in the laboratory

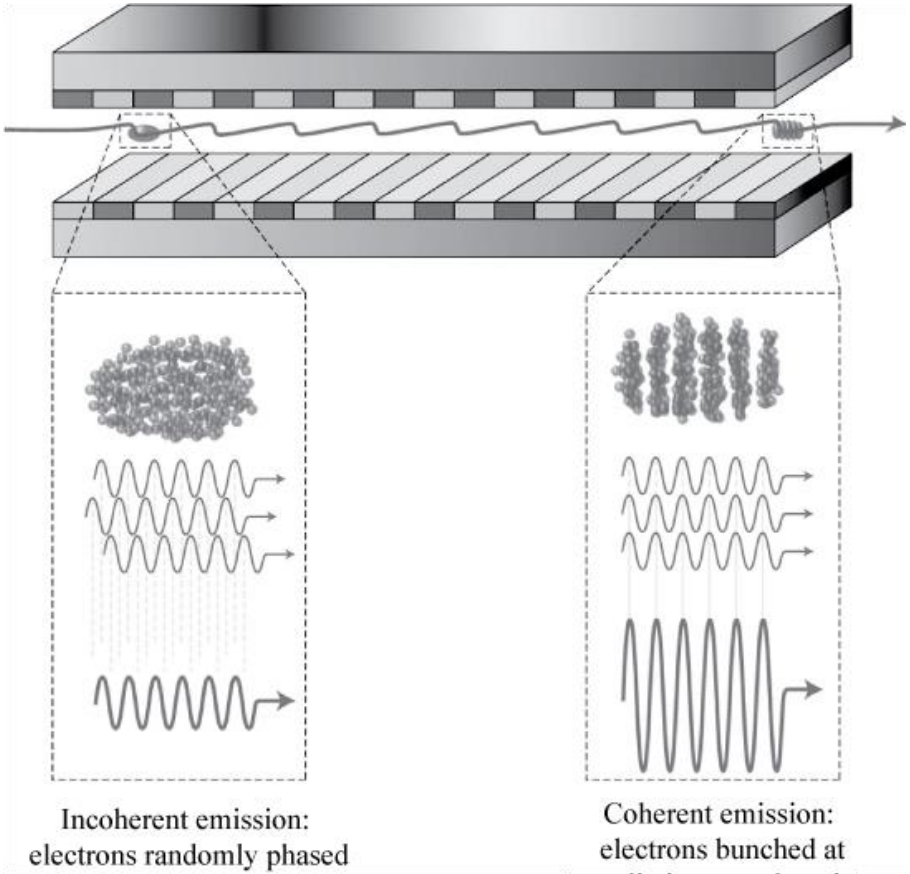

Rigaku
BioSAXS-2000
Automatic Sample Changer
DECTRIS
PILATUS
200K
EMO
Rigaku
FR-X
Rigaku
Rigaku

The shape of the scattering curve rather than the absolute value is the data (assuming the signal is above the noise)

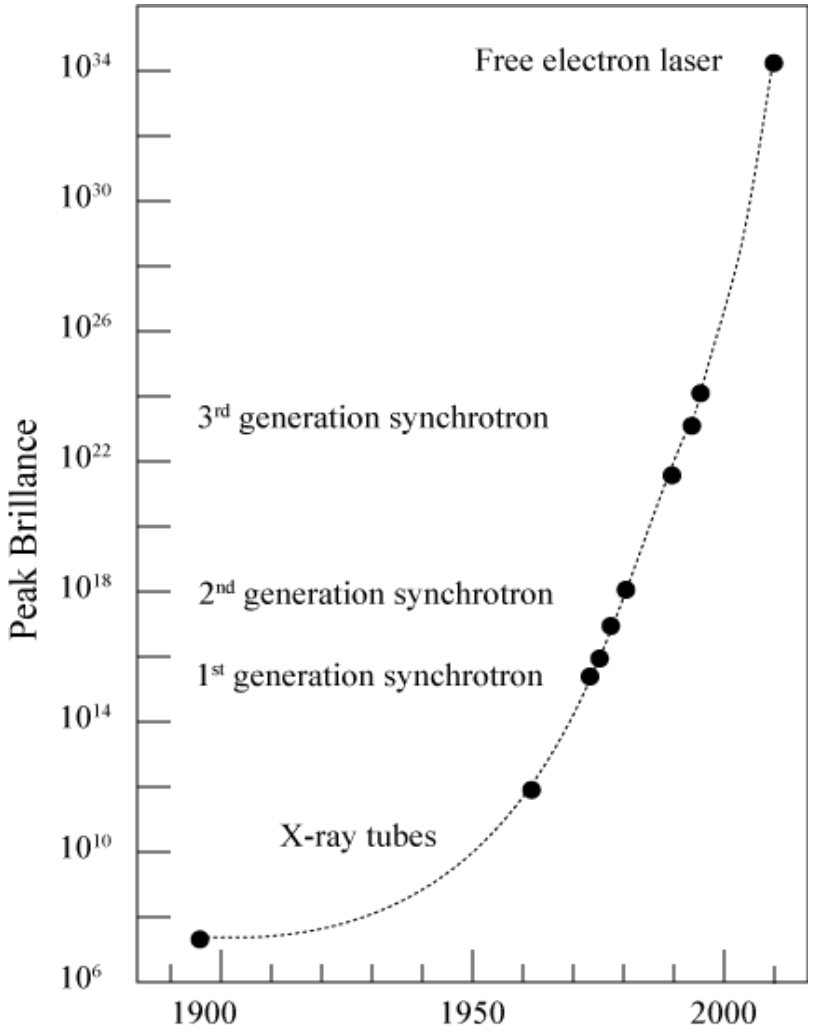


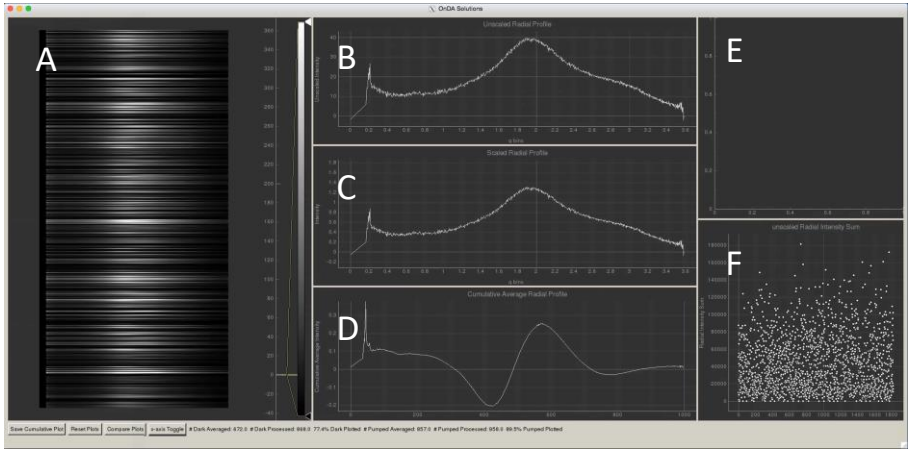
The near future

X-ray free electron lasers



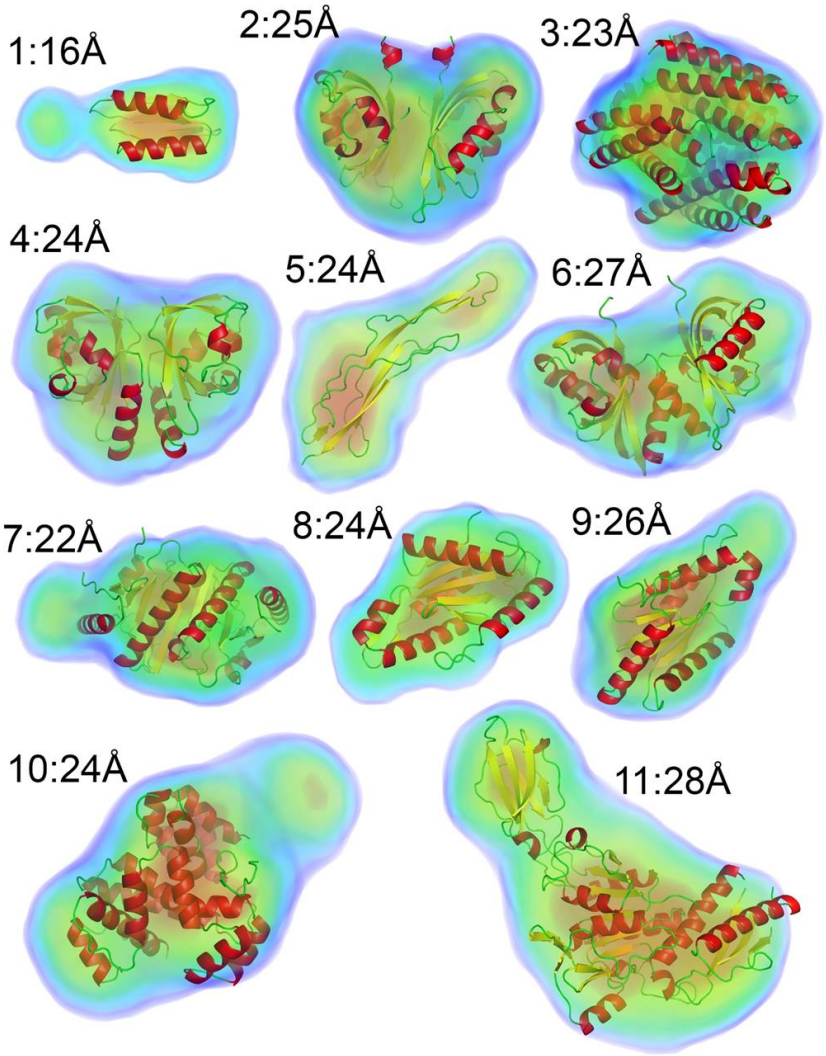
SAXS experiments capture short time points at low concentrations

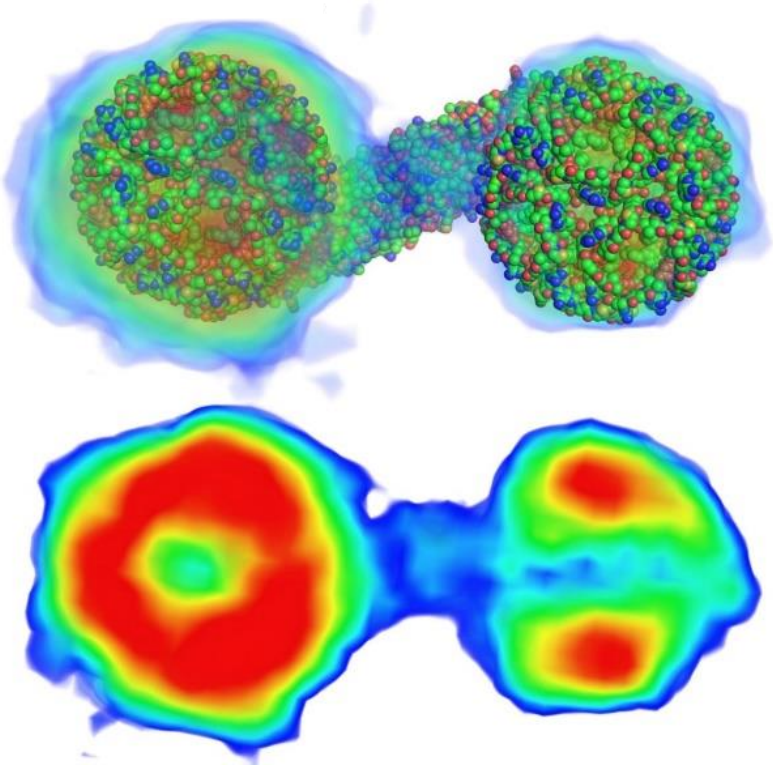




Time resolved studies and potentially extension of resolution

New algorithms – Direct electron density determination





Summary

- SAXS is a solution technique.
- It can characterize a sample to determine if crystallization should be attempted and the potential level of difficulty
- When other structural information is known it is a powerful complementary technique.
- It can reveal the solution oligomer and the spatial sampling of flexible regions.
- It’s easy to make mistakes with it and preparation is critical.

HAUPTMAN-WOODWARD
MEDICAL RESEARCH INSTITUTE

University at Buffalo

Acknowledgements

Edward H. Snell
Thomas D. Grant
Eatton (Ed) Lattman

Thank you and questions?

esnell@hwi.buffalo.edu