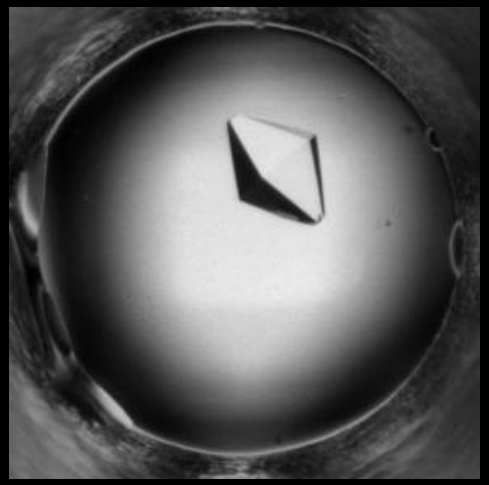
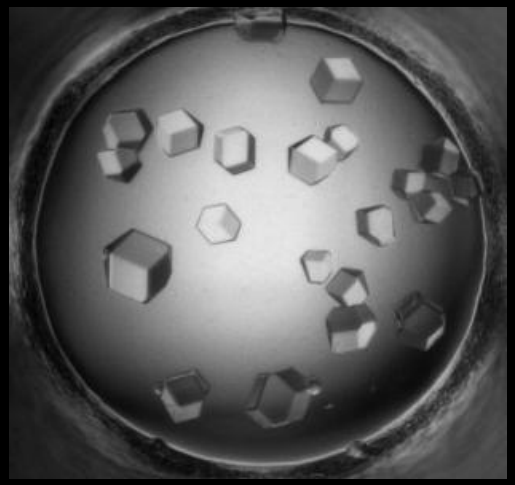
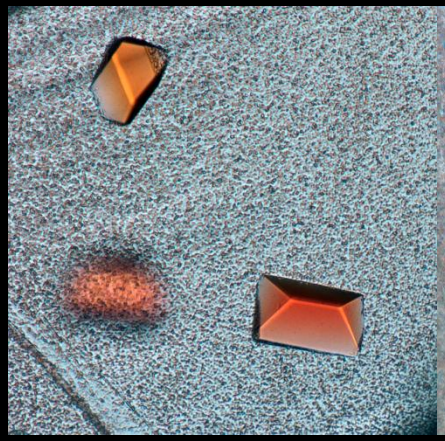
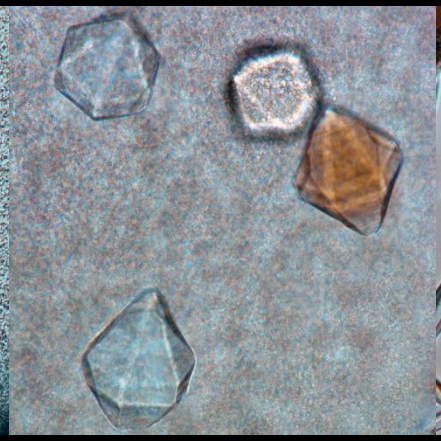
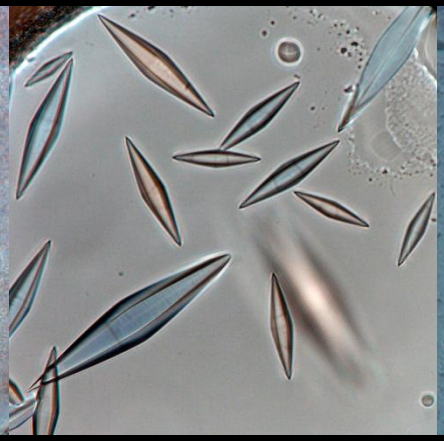
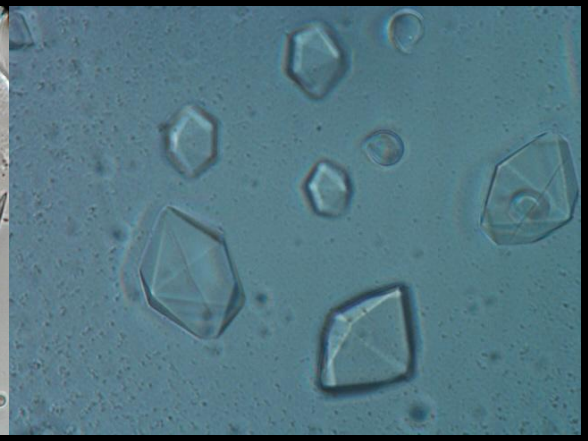
Small Angle Scattering as a Complementary Technique in Structural Biology

Edward H. Snell and Thomas D. Grant

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 foldeness and dynamics of the system (important for crystallization).
- 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

Structural Biology is not crystallography

- Low resolution structural information provides useful details.
- Foldeness and dynamics of the system can be important in mechanism.
- Complex formation is critical to mechanism
- Flexible regions can be critical to mechanism.
- Chemistry is critical to mechanism.
- Gross structural changes can be critical to mechanism.
- Crystal oligomer may not be biological oligomer.

Complementary techniques provide complementary information

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

(one of several complementary techniques)

SAXS Literature and Software

Reviews:

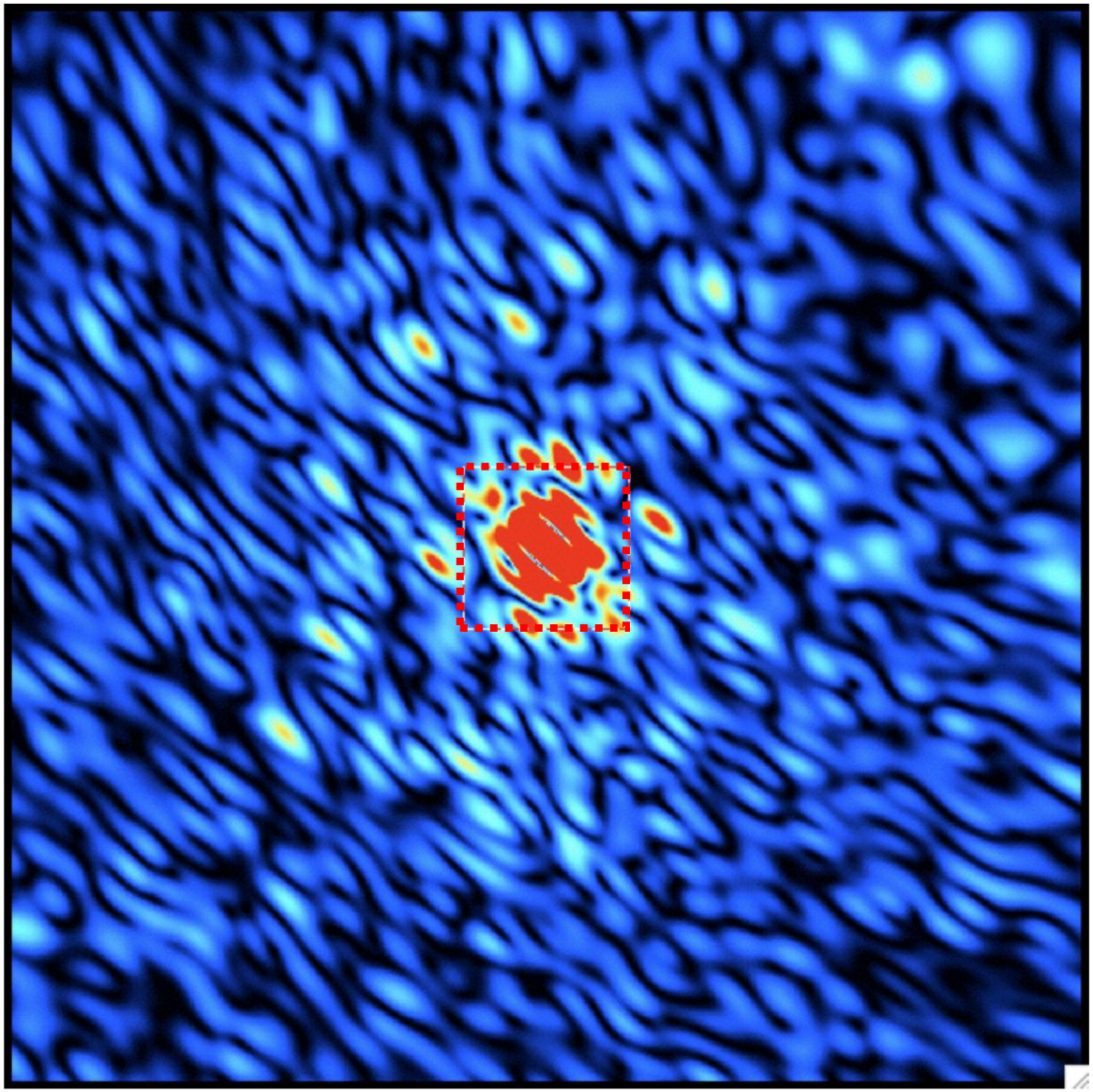
- Putnam et al, Q Rev Biophys. Aug 2007; 40(3): 191-285.
- Jacques and Trewhella, Protein Science 2010 Apr; 19(4): 642–657.
- Svergun et al, Oxford University Press 2013, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*

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

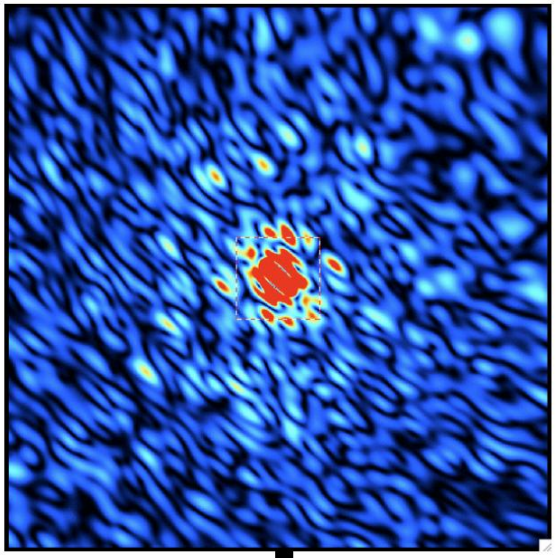
http://smallangle.org/content/software

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

Molecular Transform

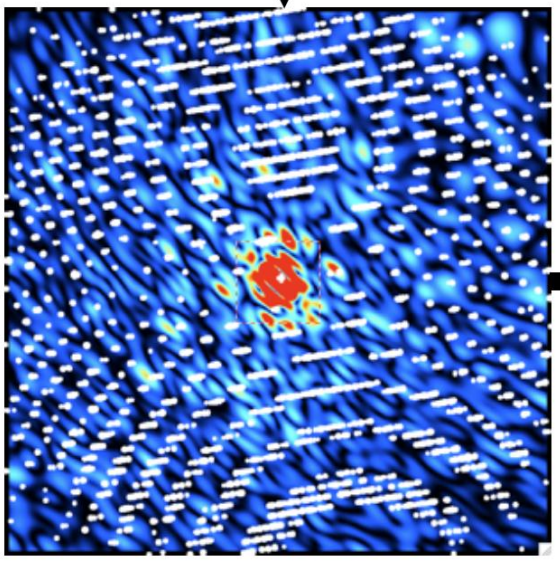


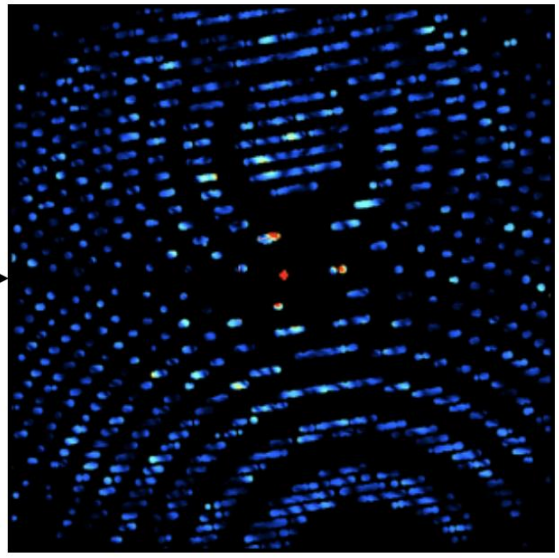
See Intro_to_SAXS.pdf at www.BioXFEL.org
Seminars by Thomas Grant



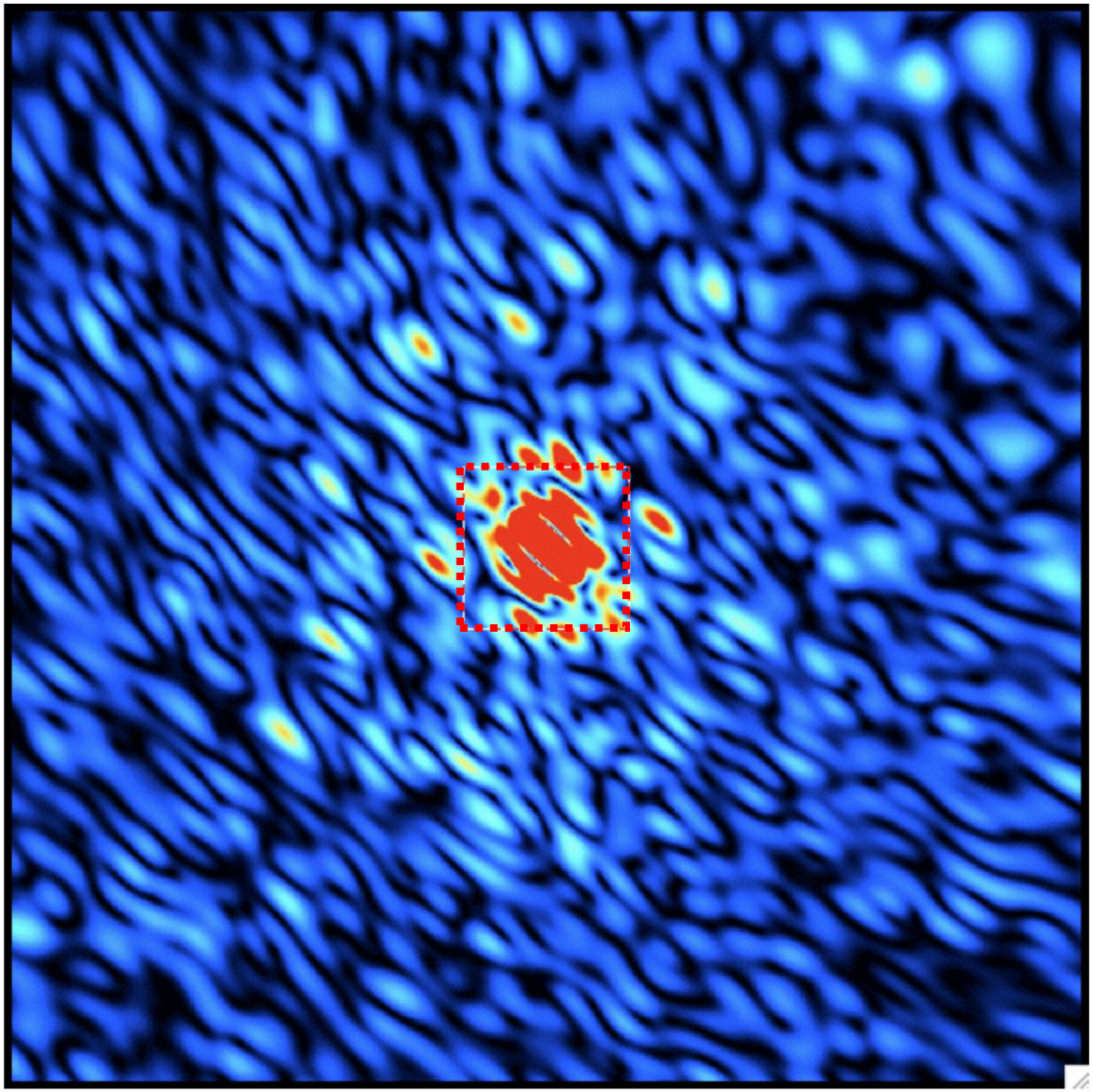
Molecular Transform

Bragg Sampling from
X-ray Crystallography



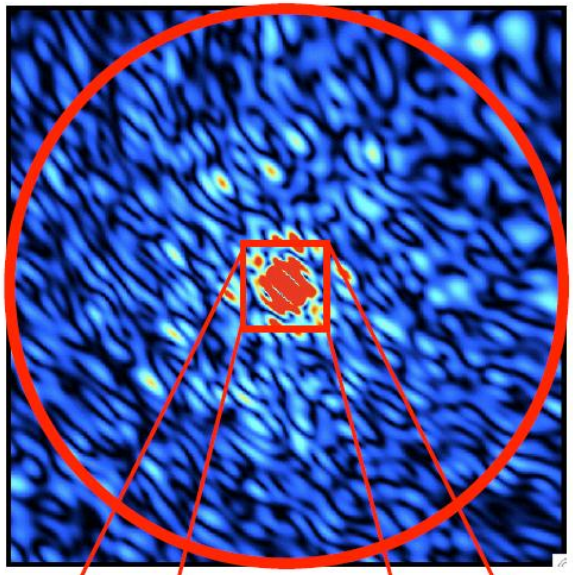


Molecular Transform

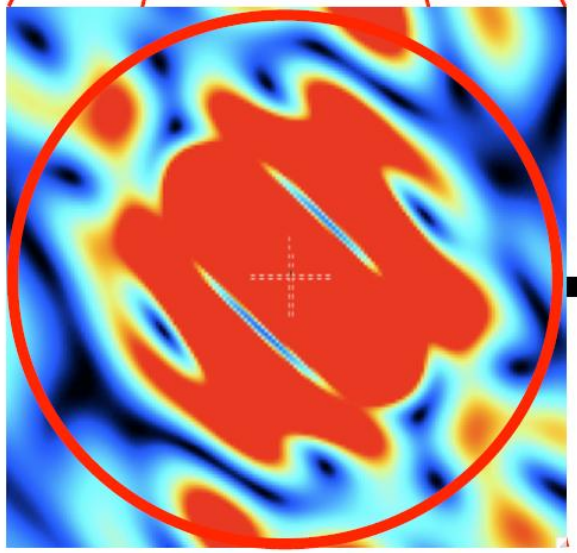


See Intro_to_SAXS.pdf at www.BioXFEL.org
Seminars by Thomas Grant

2Å

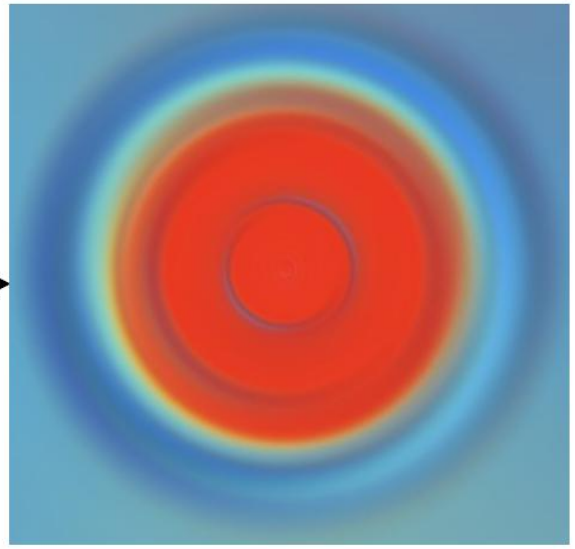


20Å

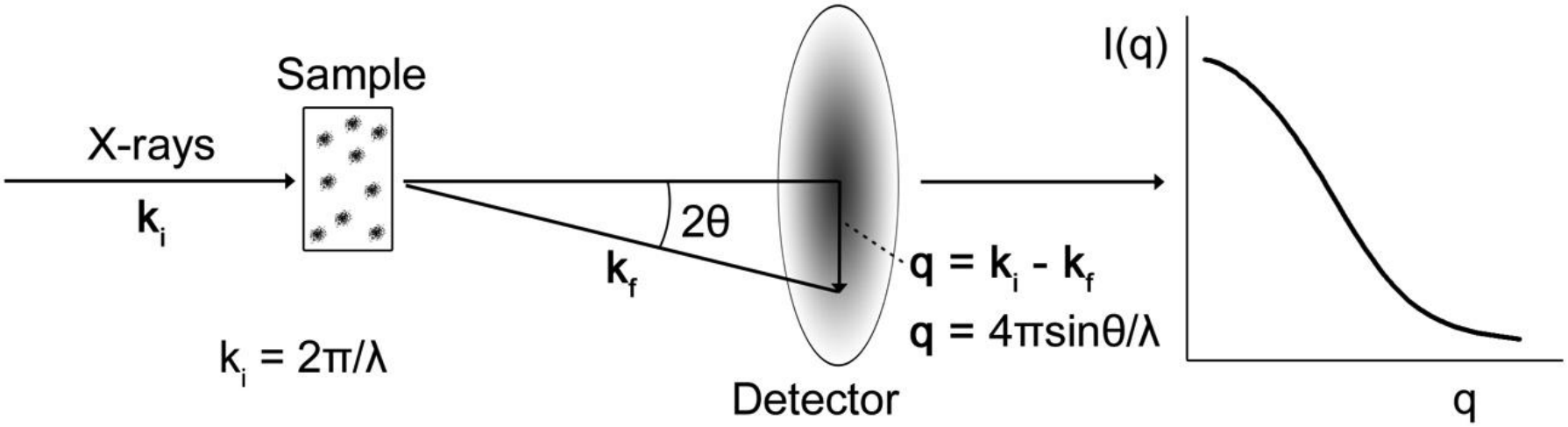


Molecular Transform

Spherical averaging from solution of tumbling molecules



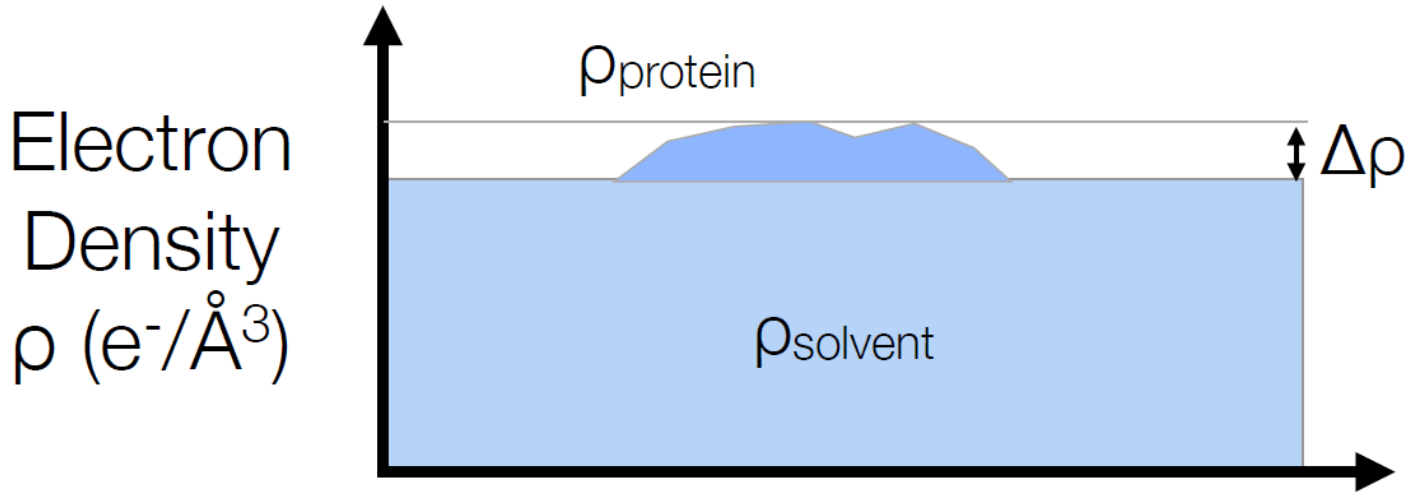
SAXS images everything behind the beamstop



- 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.

SAXS is a Contrast Technique

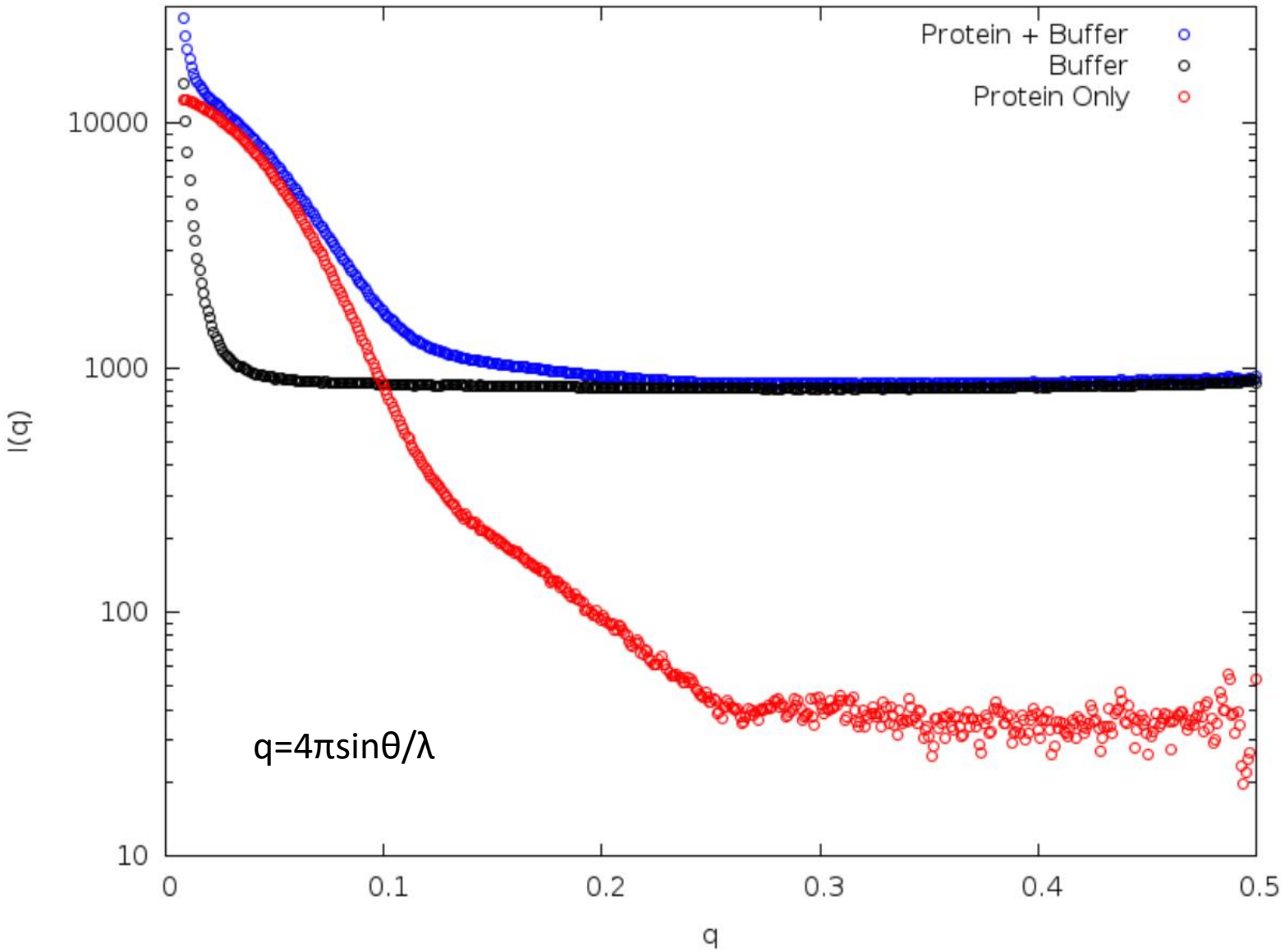
- SAXS is a contrast method, i.e. it depends on the square of the difference in the electron density between the molecule and the solvent



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

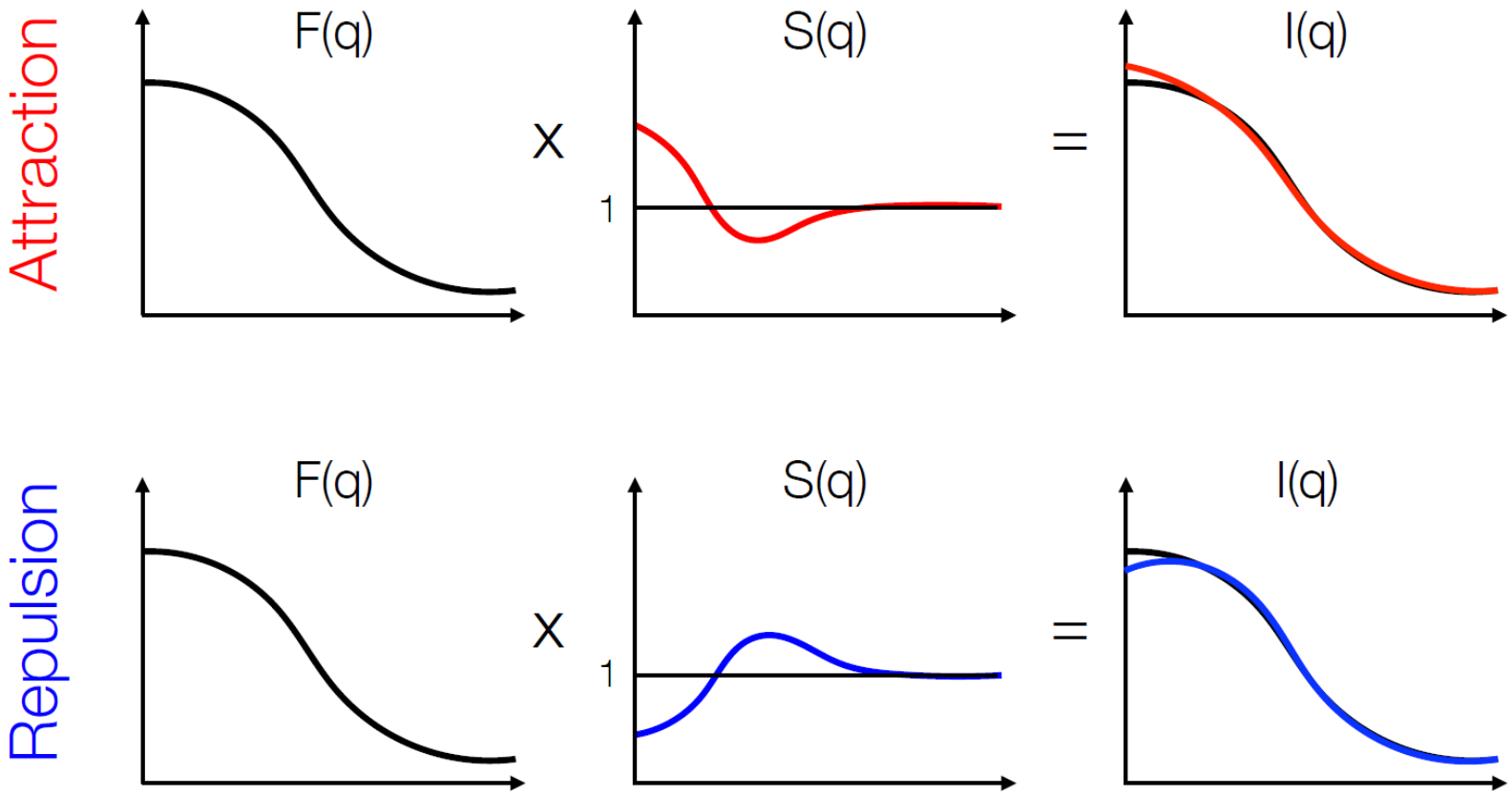
Form factor of particle

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

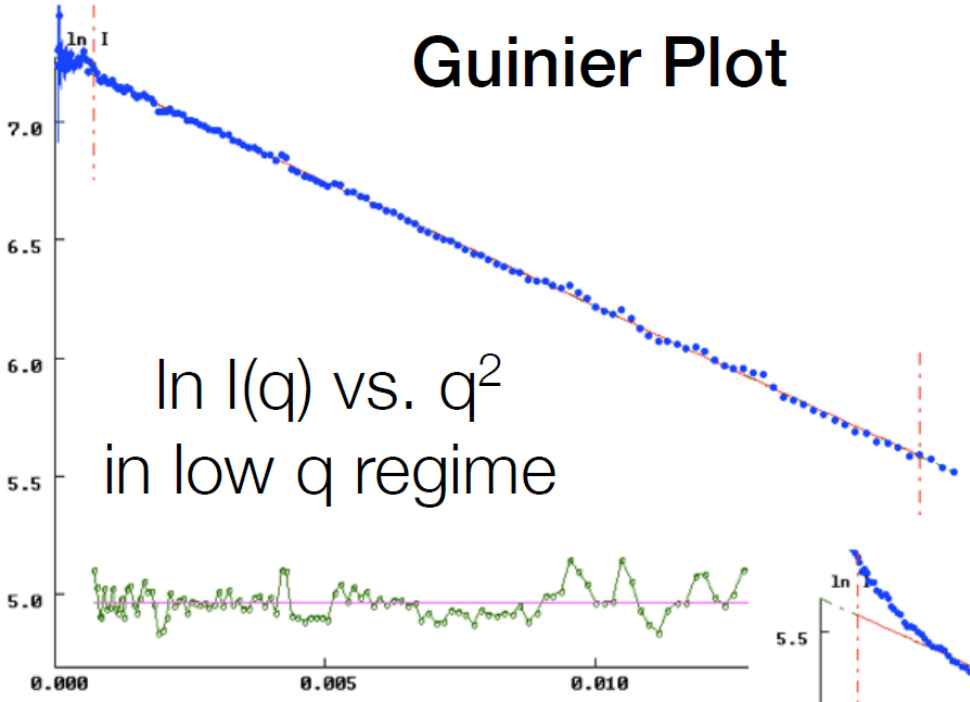
- Developed by André Guinier in 1939.
- As q → 0, intensity can be approximated by:

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

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

y = b + m * x

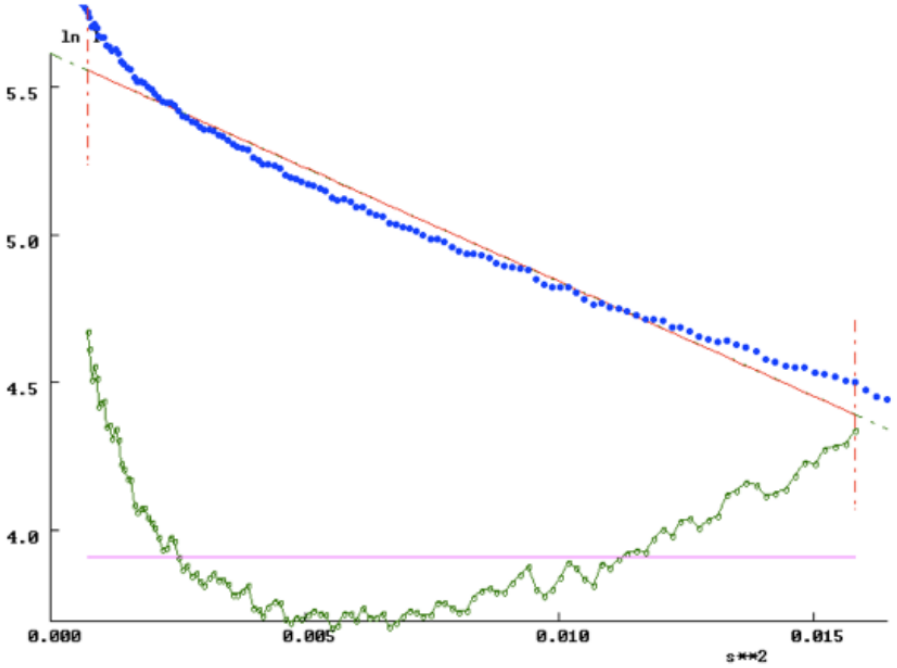
Guinier Plot



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

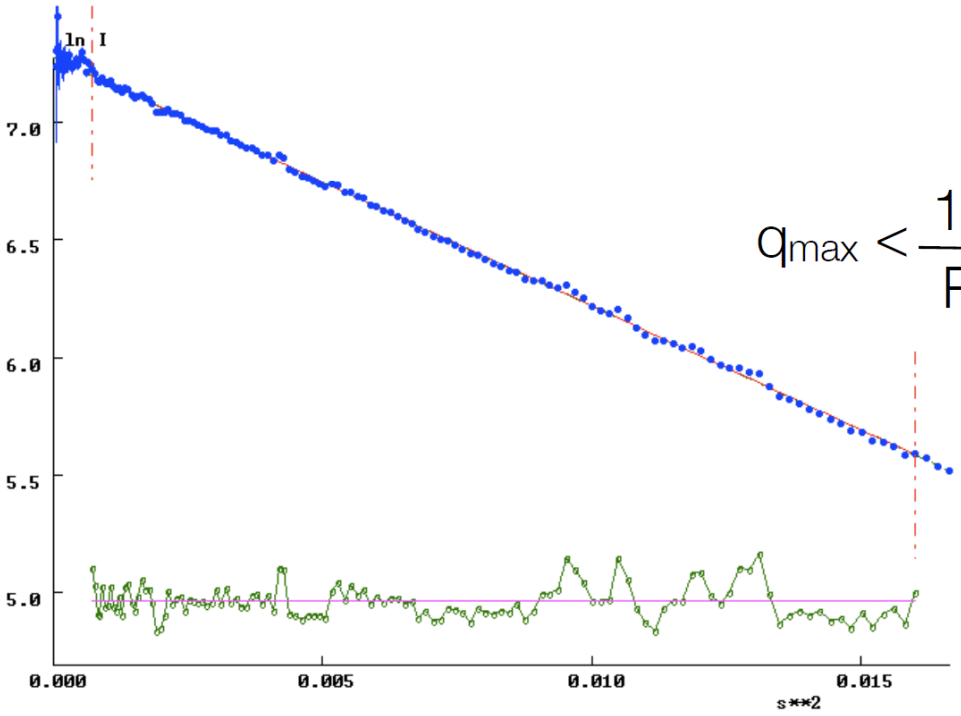
Straight line shows
no aggregation:
Can determine R_g
from slope of line,
I(0) from intercept

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

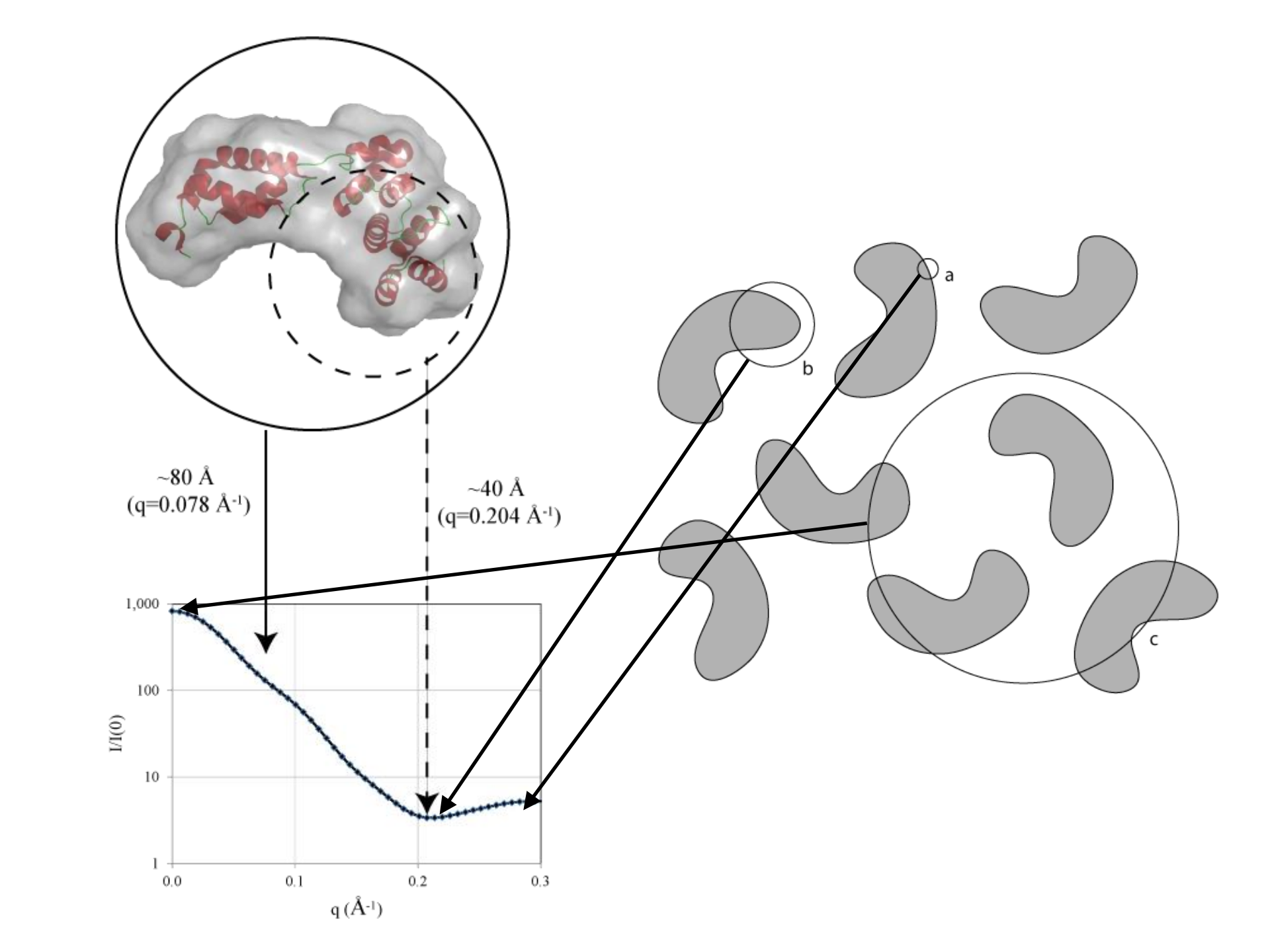


Approximation only valid over a certain region of scattering space

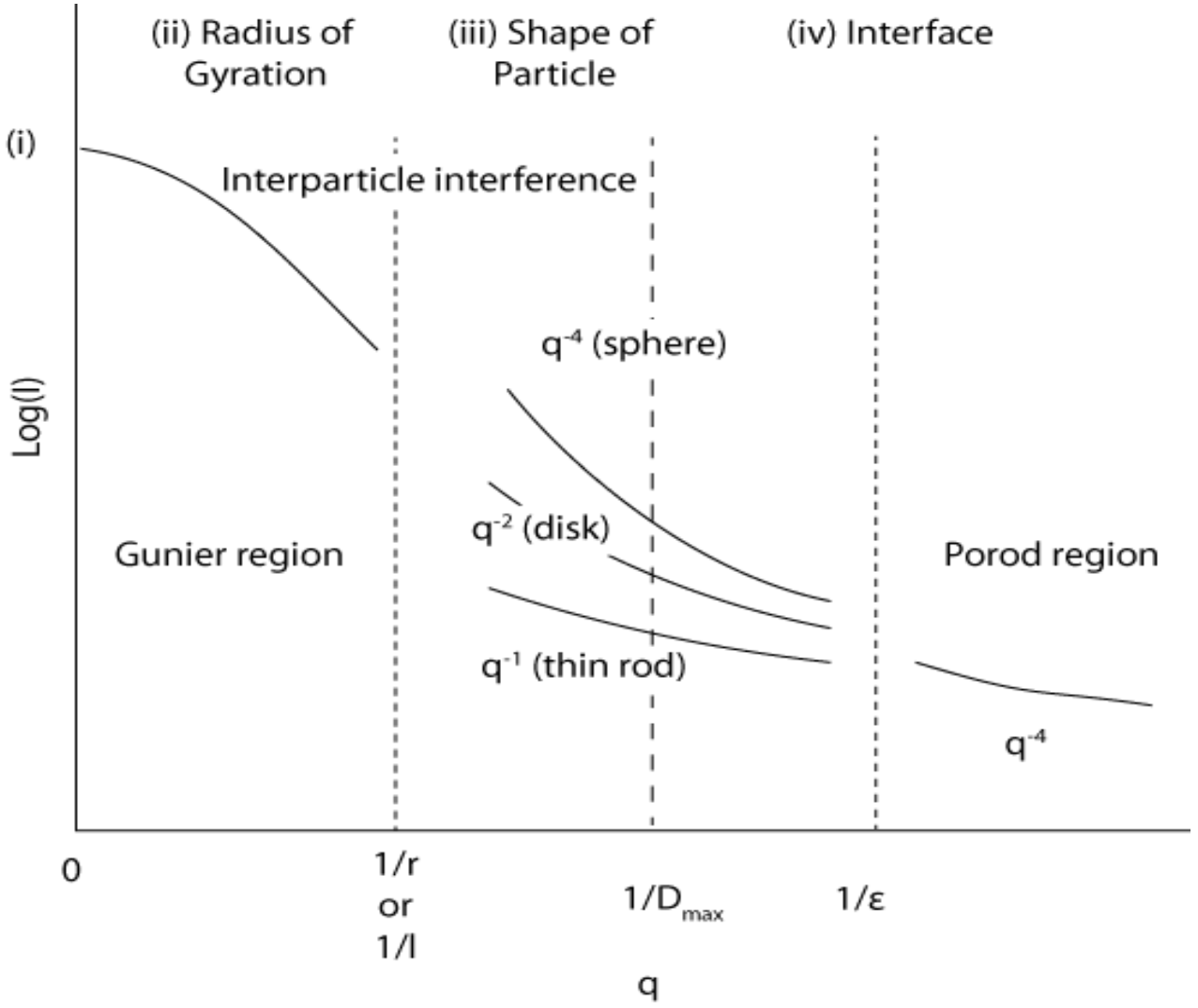
$$q_{min} < \frac{\pi}{D_{max}}$$



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

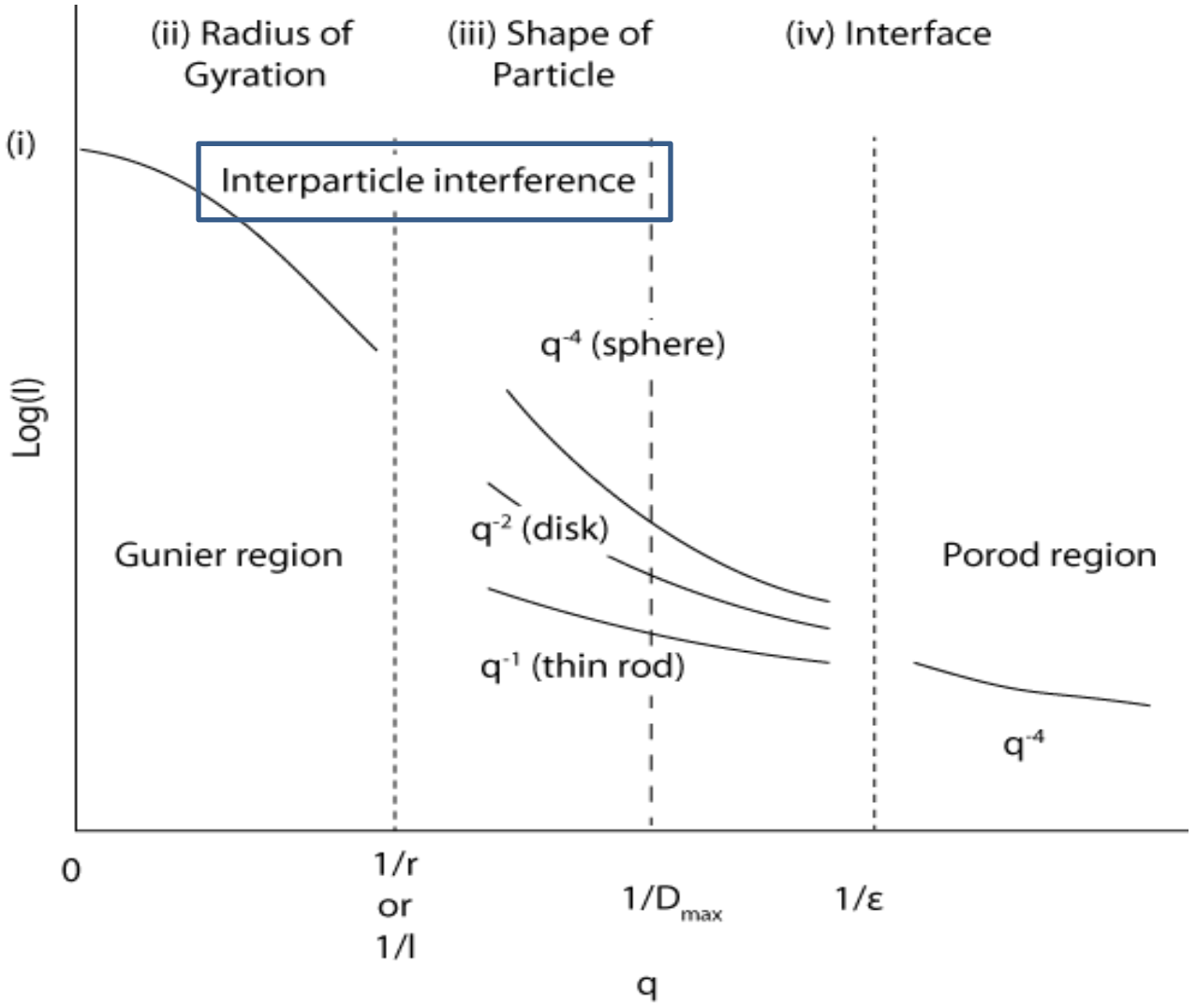

~80 Å
(q=0.078 Å-1)
~40 Å
(q=0.204 Å-1)
a
b
c
1,000
100
10
1
I/I(0)
0.0
0.1
0.2
0.3
q (Å-1)

The Shape of the Scattering Curve is important but not the absolute intensity

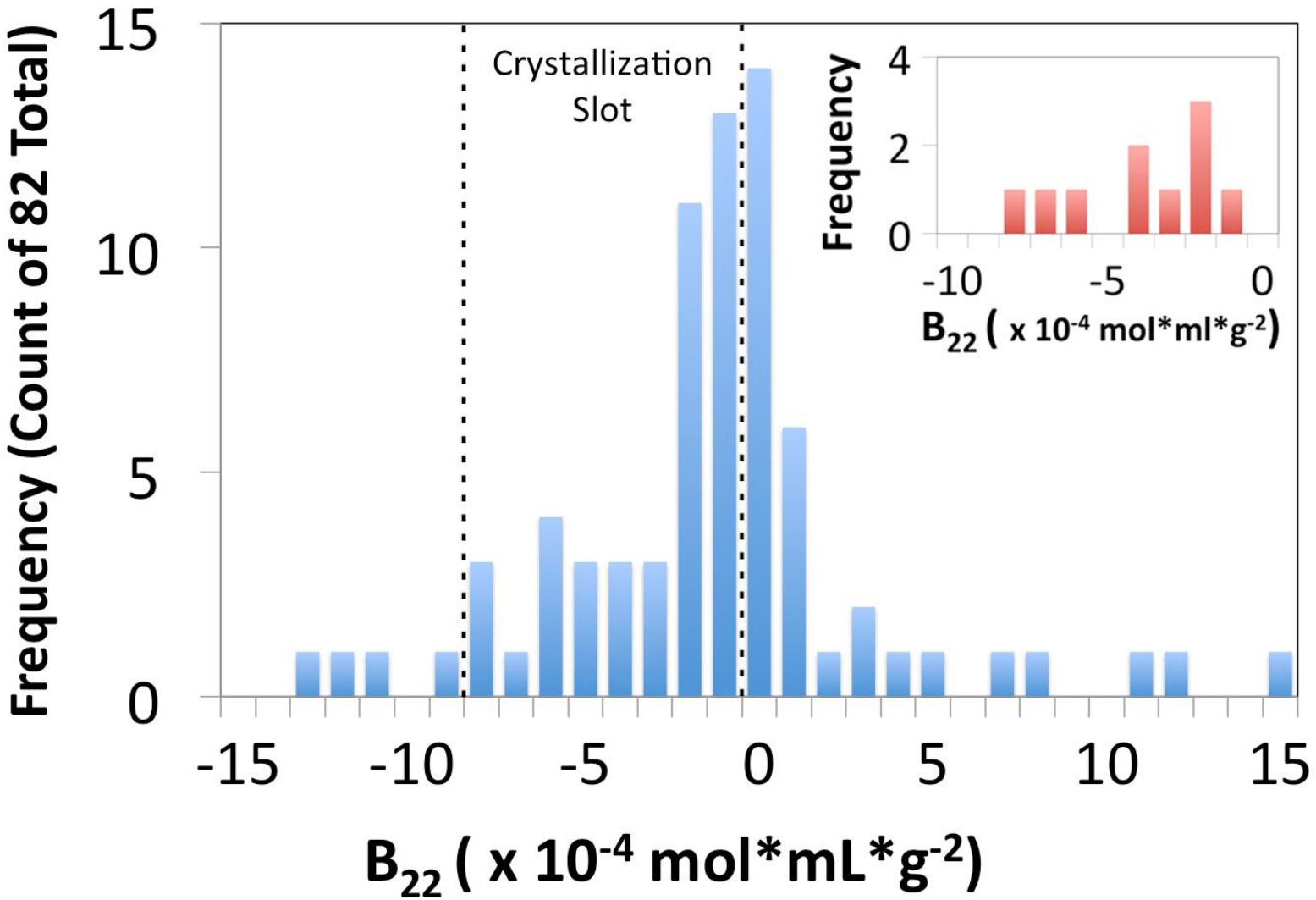


Only concentration information is contained in the intensity values

The Shape of the Scattering Curve is important but not the absolute intensity



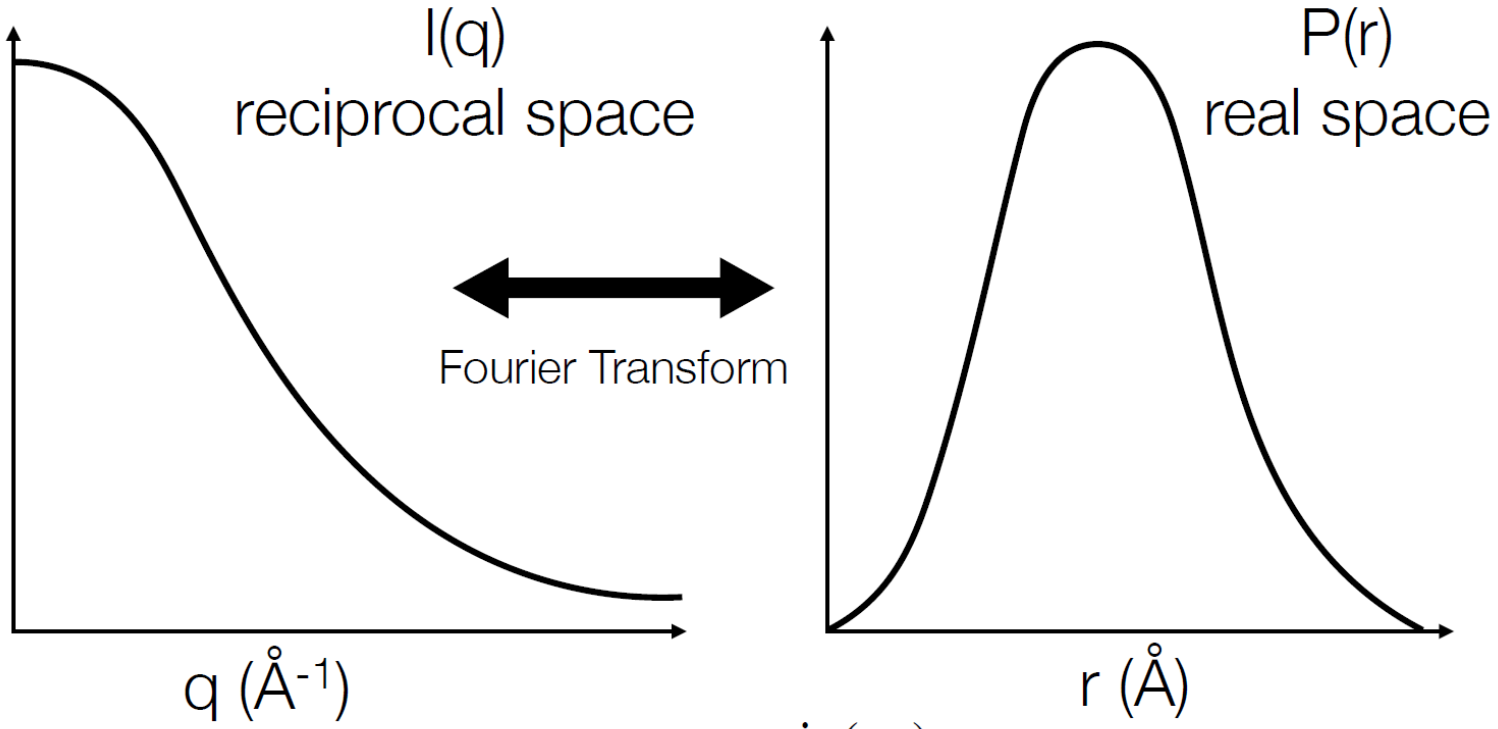
Only concentration information is contained in the intensity values


Frequency (Count of 82 Total)
15
10
5
0
Crystallization
Slot
-15
-10
-5
0
5
10
15
B22 (x 10-4 mol*mL*g-2)
Frequency
4
2
0
-10
-5
0
B22 (x 10-4 mol*ml*g-2)

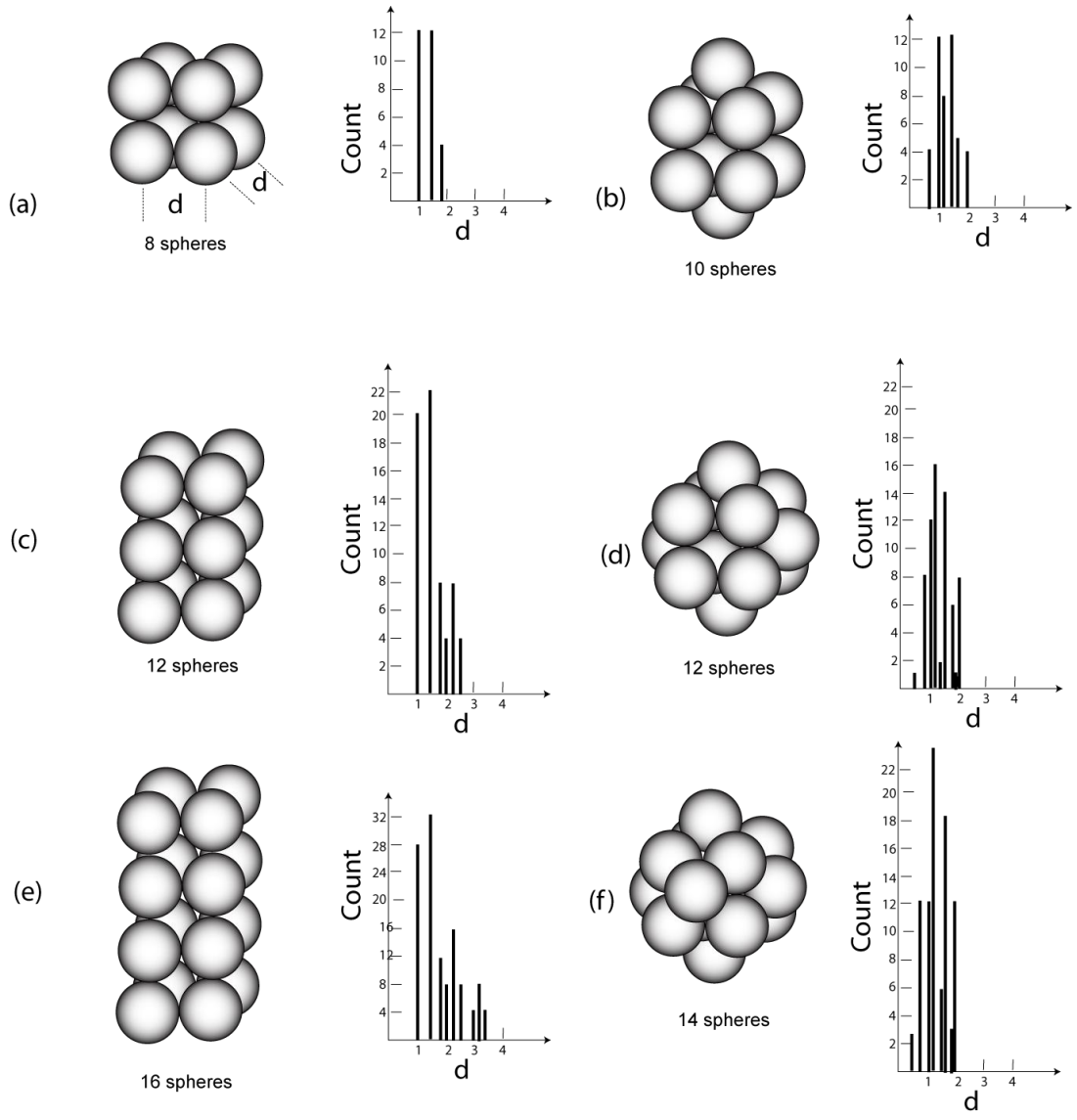
Transformation of SAXS data into Structural Information

(the useful stuff)

Scattering is in Fourier space, transform to real space



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

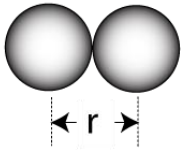


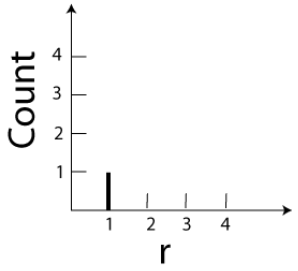
Fourier space yields frequency of interatomic scattering vectors as a function of the length of the vector

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

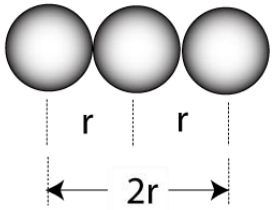
Larger compact molecules have a high distribution at lower angle (consider detector distance etc.)

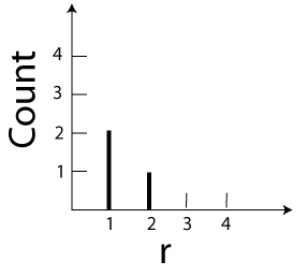
(a)



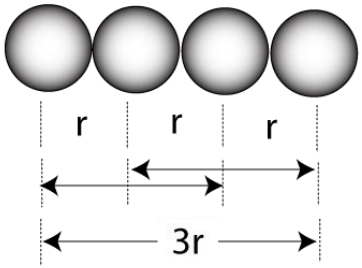


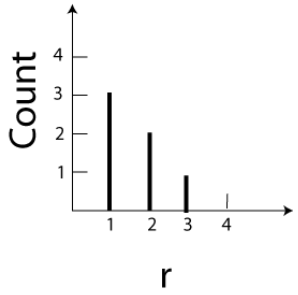
(b)



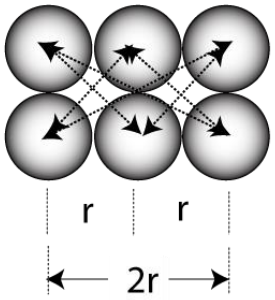


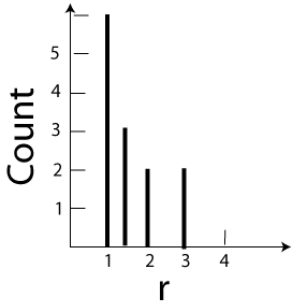
(c)





(d)

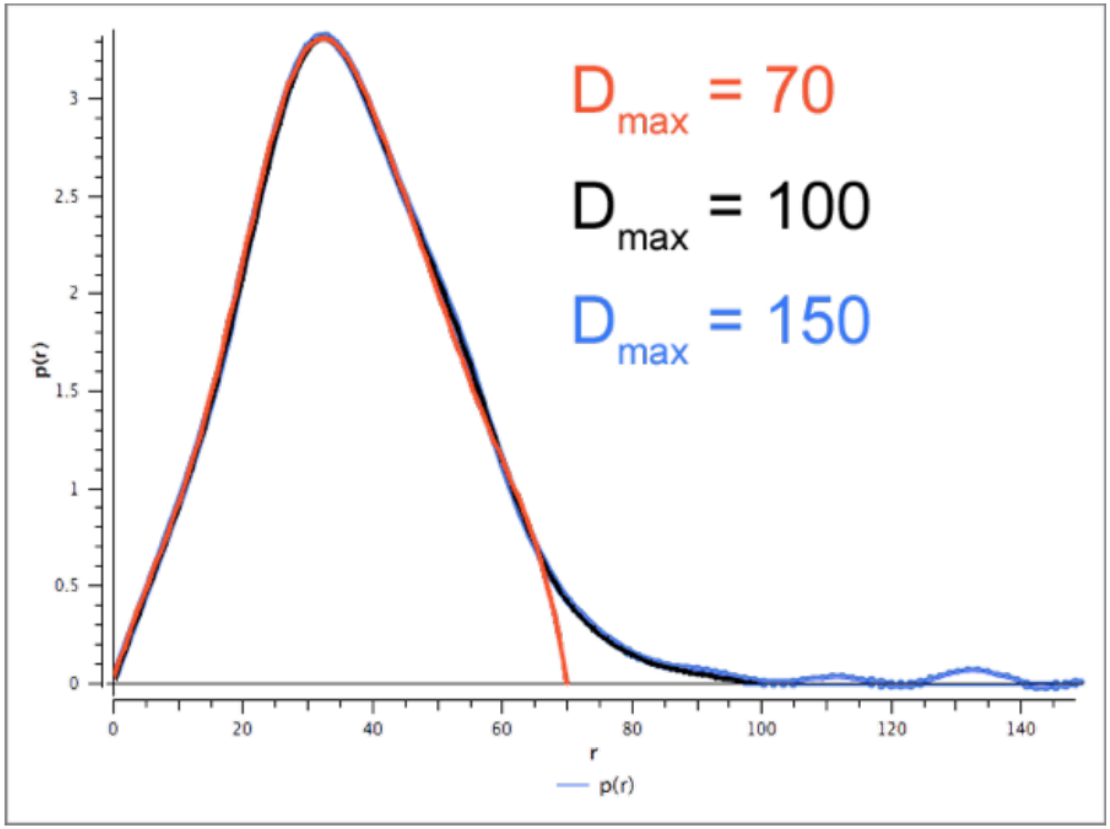




P(r) plot is simply the histogram of interatomic scattering

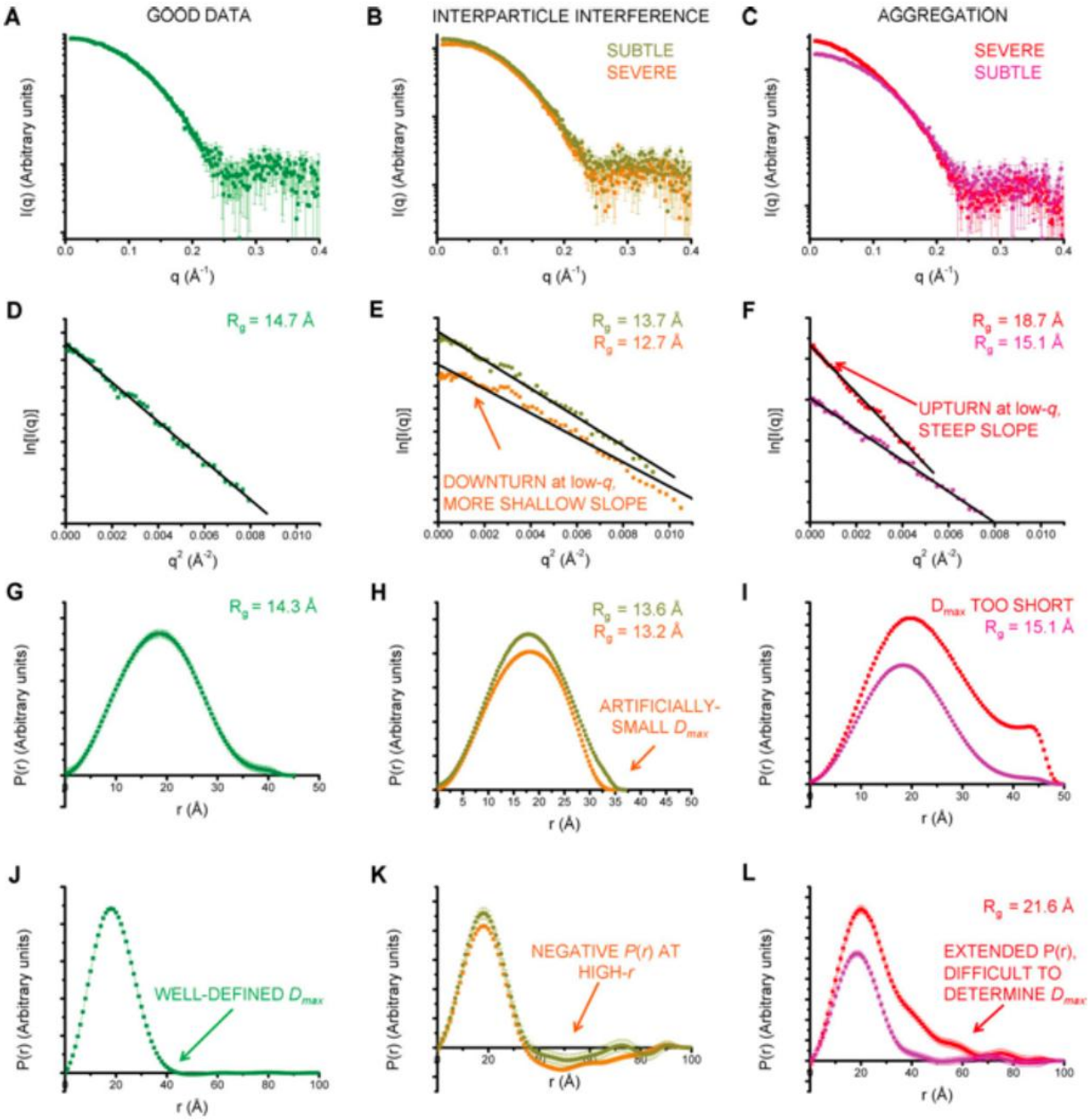
Pair distribution function is used to determine the maximum particle dimension

- 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



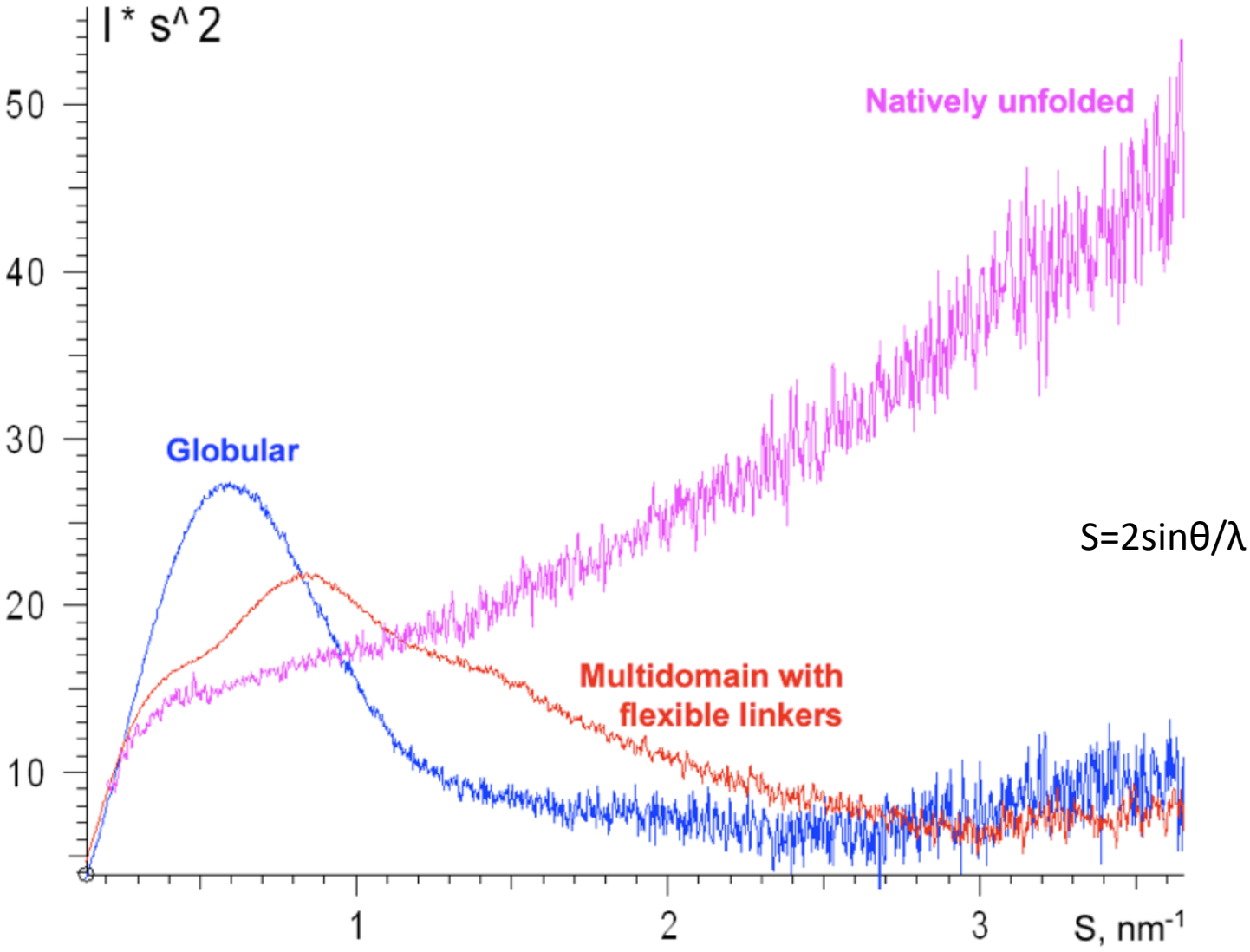
The maximum particle dimension is given by the distance between the furthest interatomic scattering

Sample quality greatly affects data analysis



Jacques and Trewhella, 2010 *Protein Science* Review

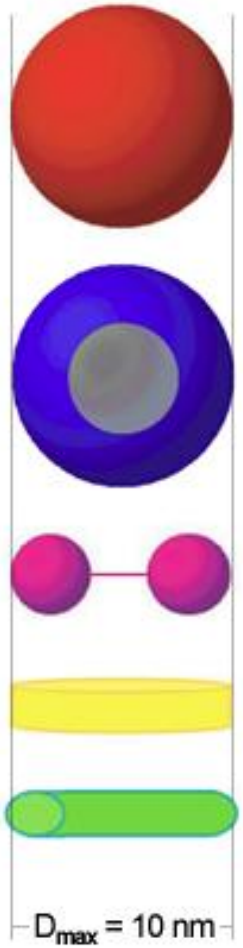
Kratky analysis reveals dynamics

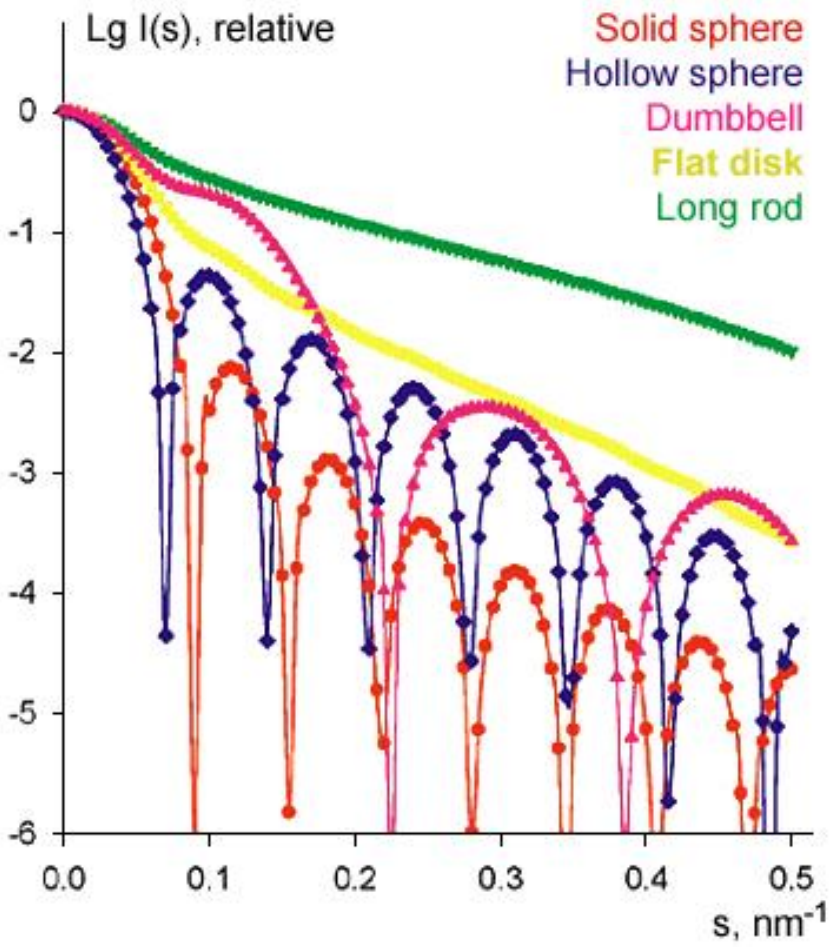


S=2sinθ/λ

http://www.saxier.org/forum/viewtopic.php?t=337

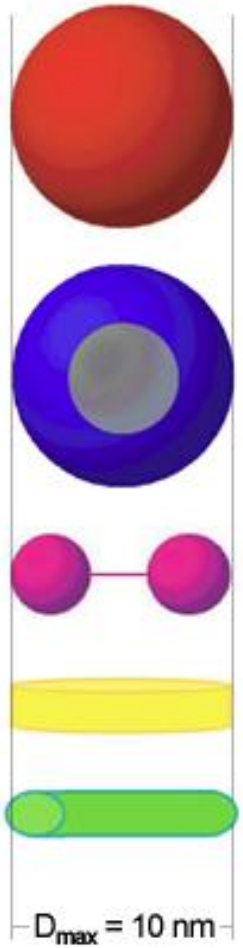
Data

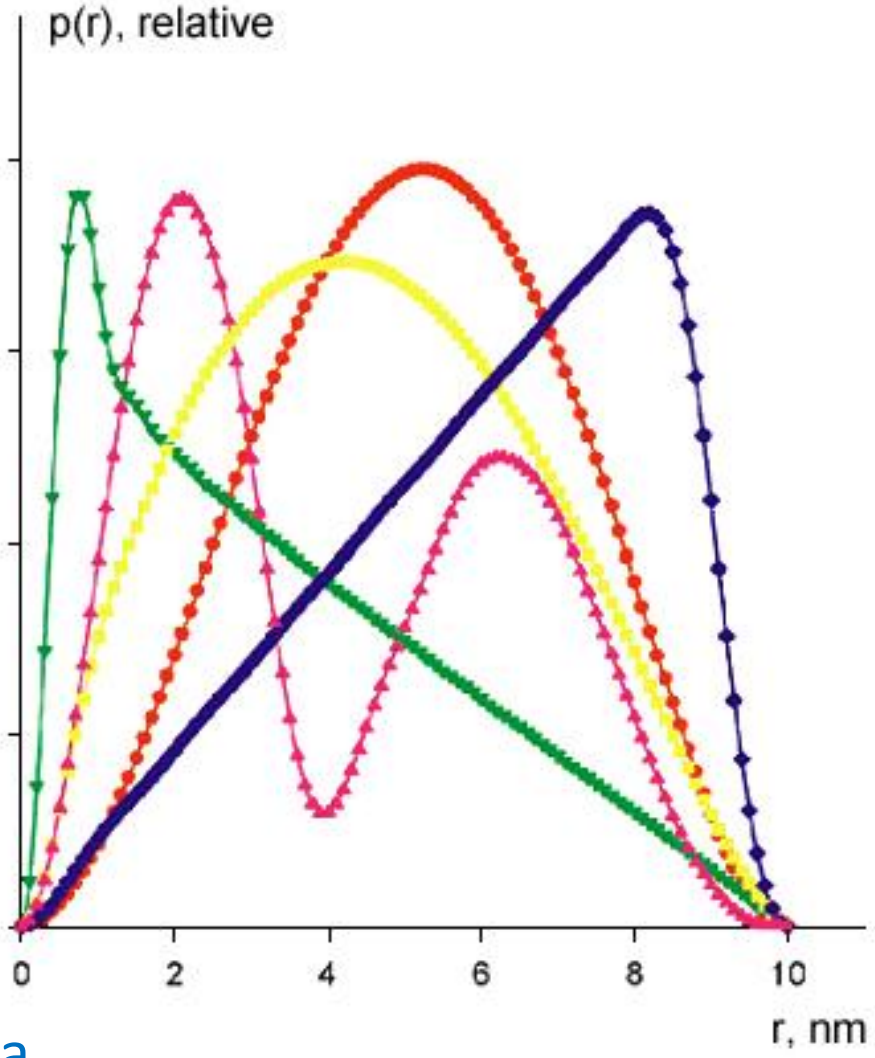




From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)

Pair distribution function

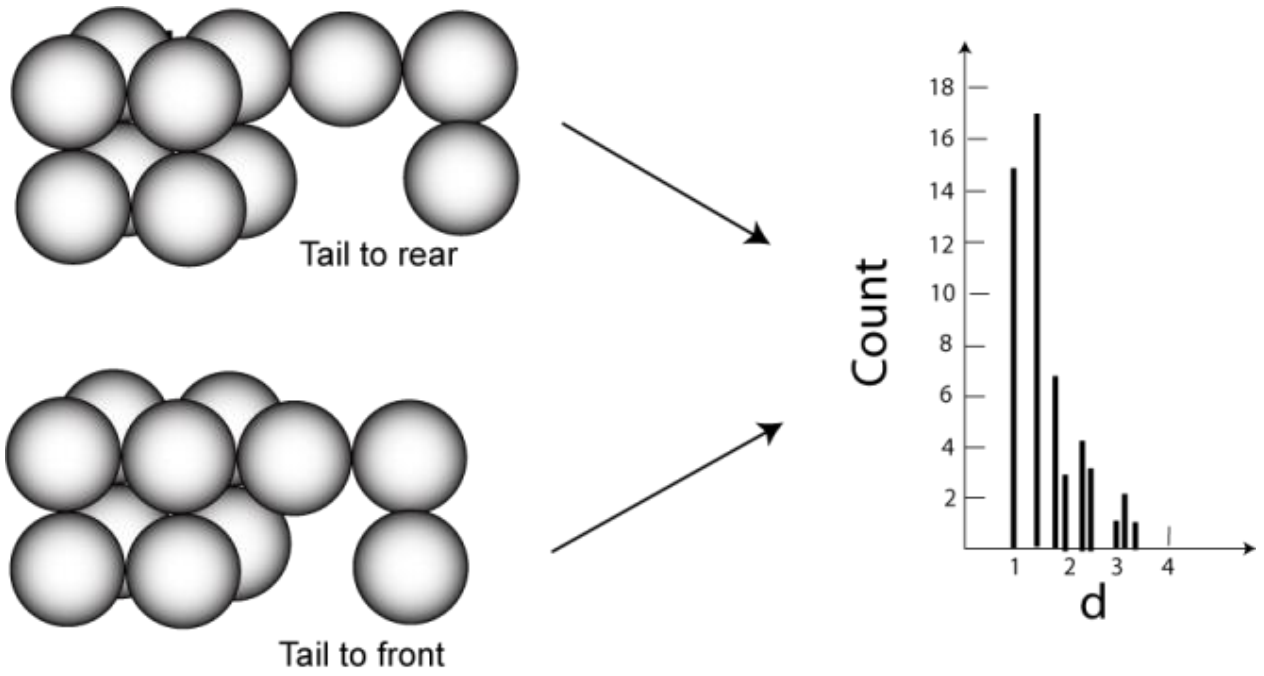




Fourier transform of data.

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

What can possibly go wrong?

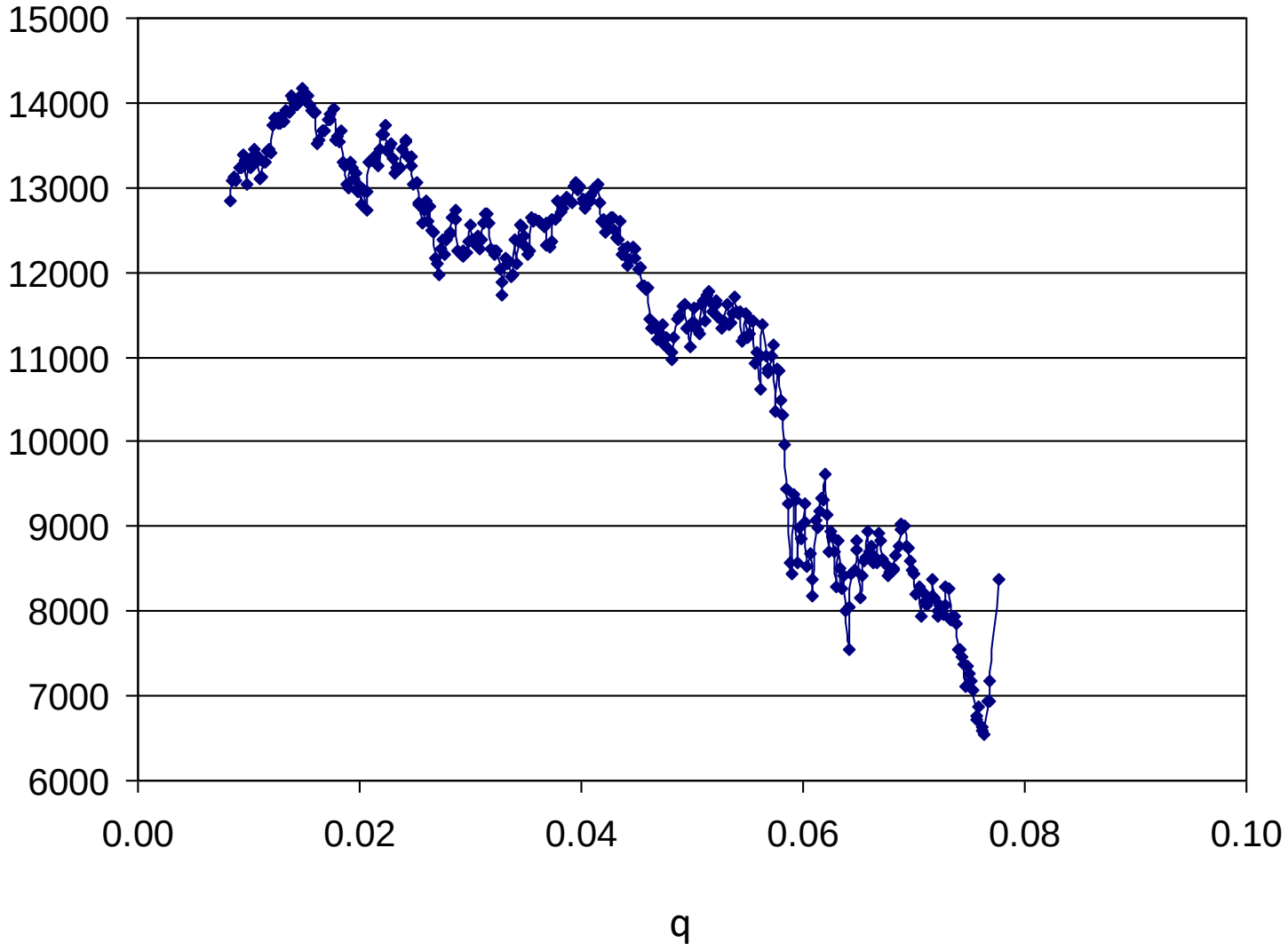


Sometimes a unique reconstruction is not available.

Garbage in, garbage out

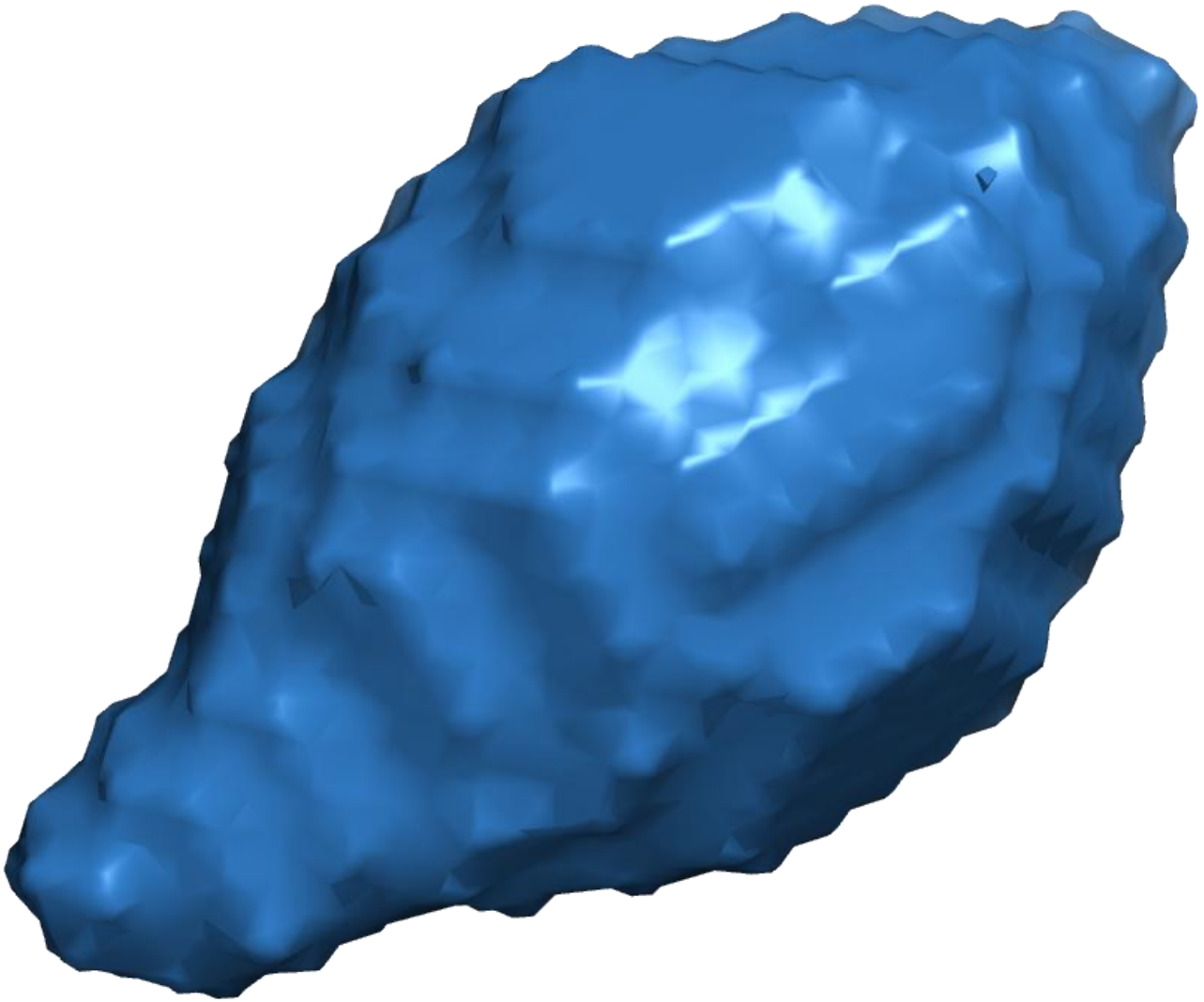
A limitation of the technique is that good or bad data can produce a result

Lets take some '*scattering*' data

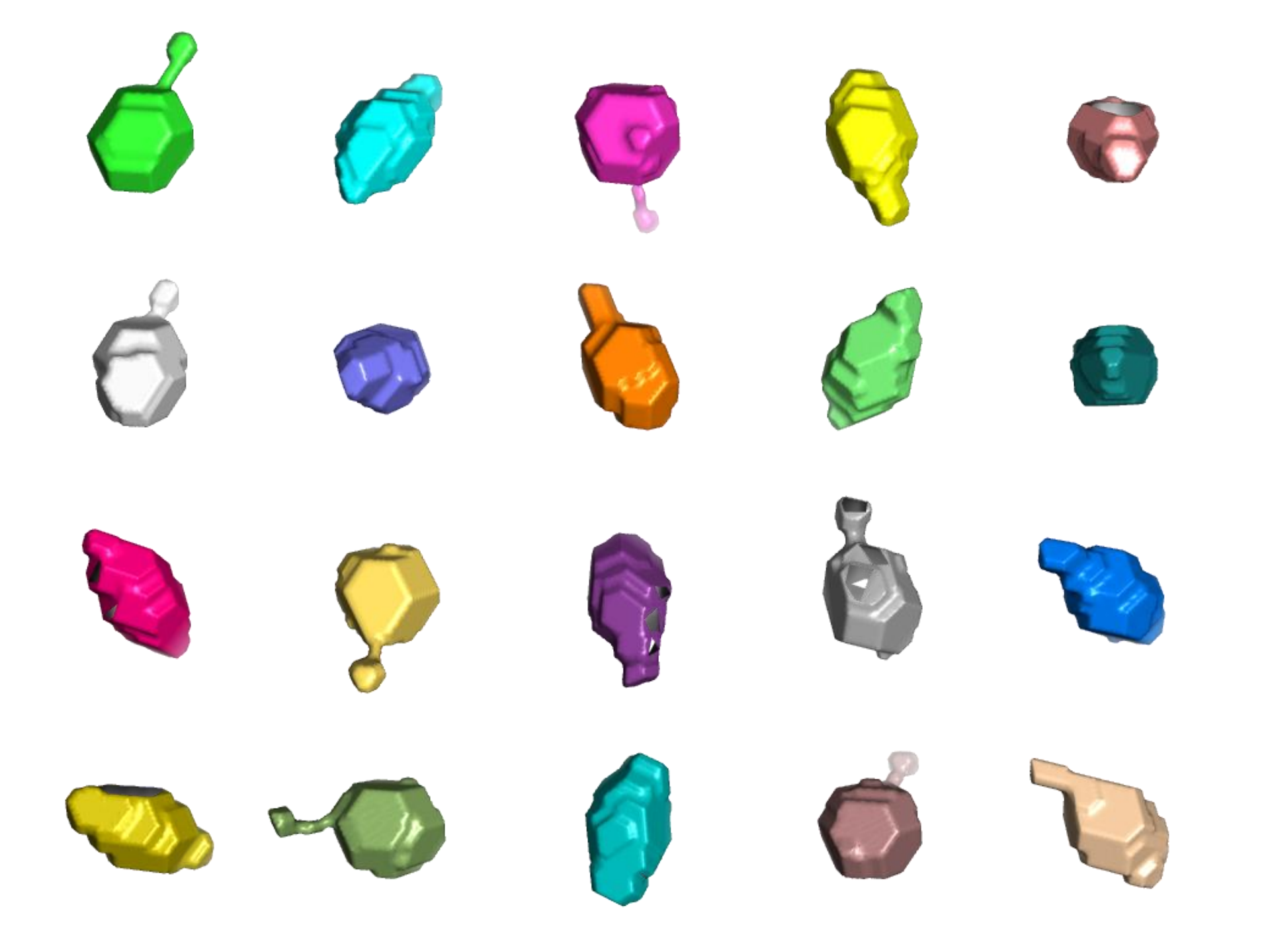


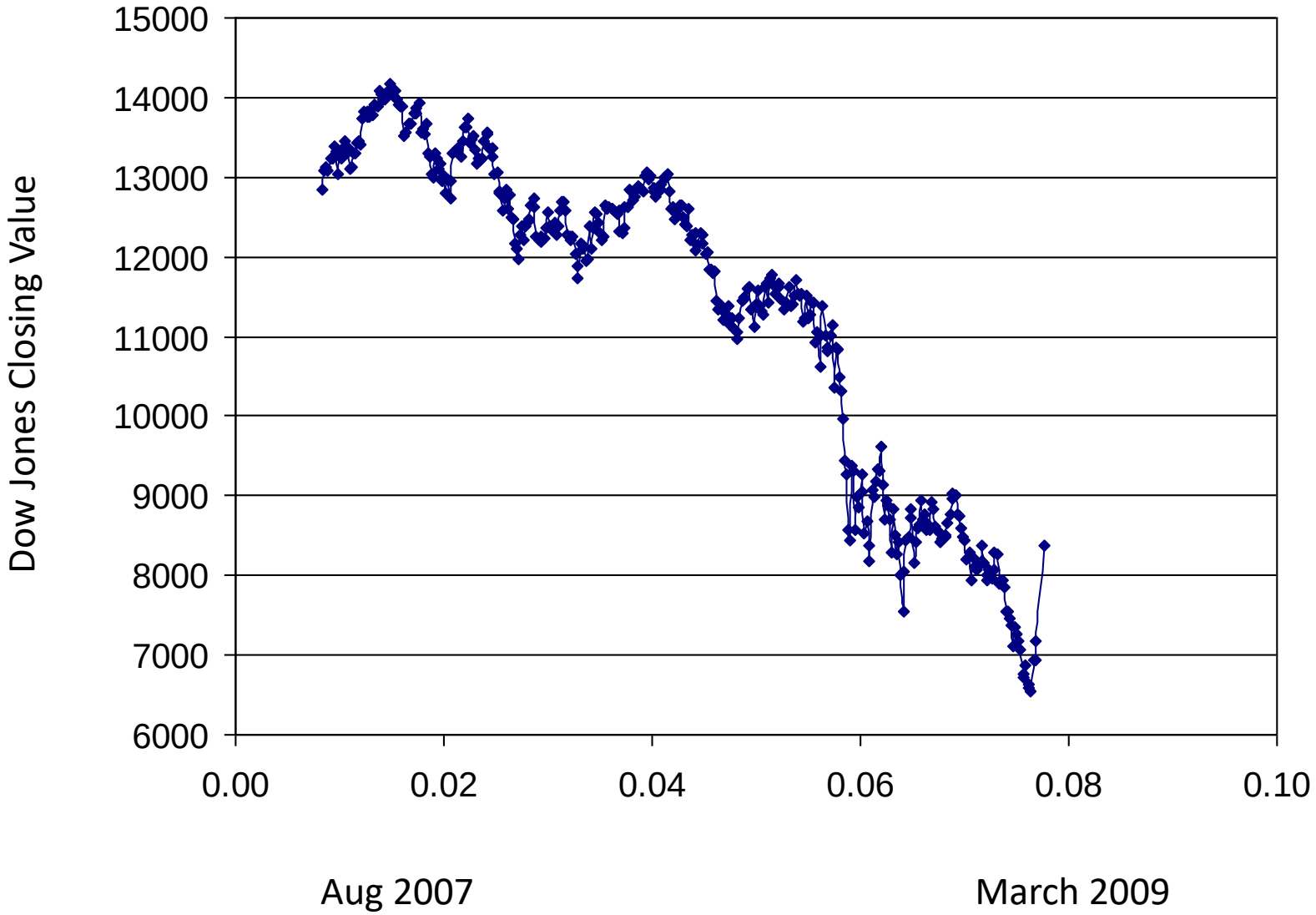
Envelope Reconstruction

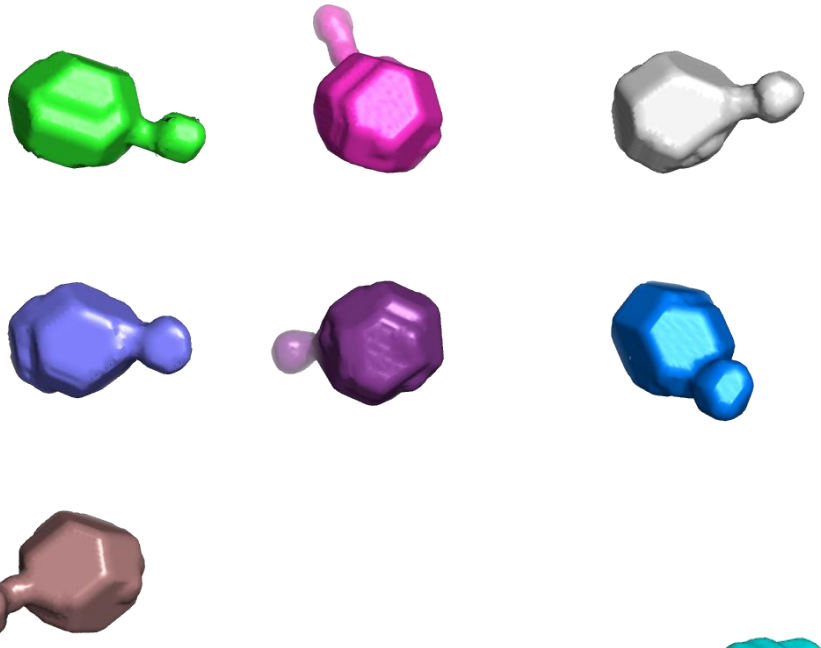
- Produce 10-20 *ab initio* reconstructions
- Determine the most probable model, i.e. the least different from the rest and align all to this.
- Estimate the similarity of the models using the Normalized Spatial Discrepancy (NSD)
  - Average NSD ~ 0.5 implies good stability of solution
  - Average NSD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

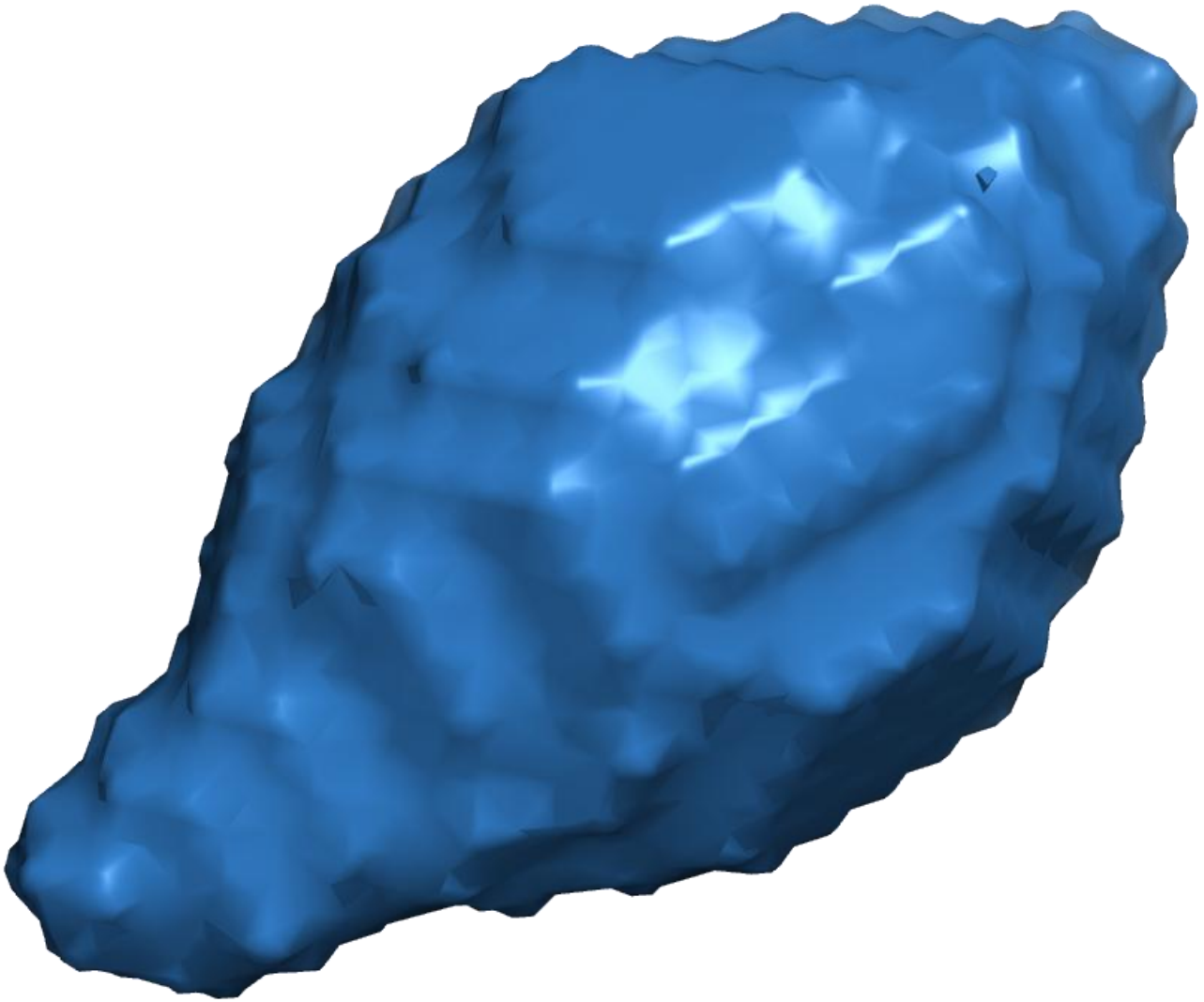


Actually two populations

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

A Bull or a Bear market!

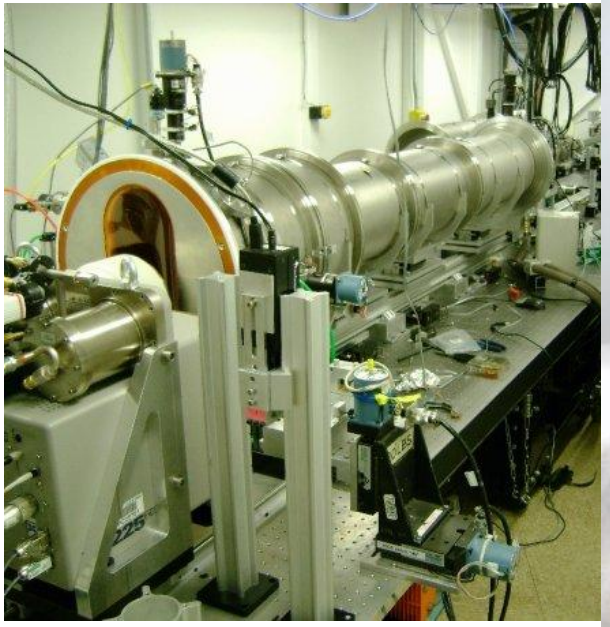
This is the molecular envelope of the recession, not a protein

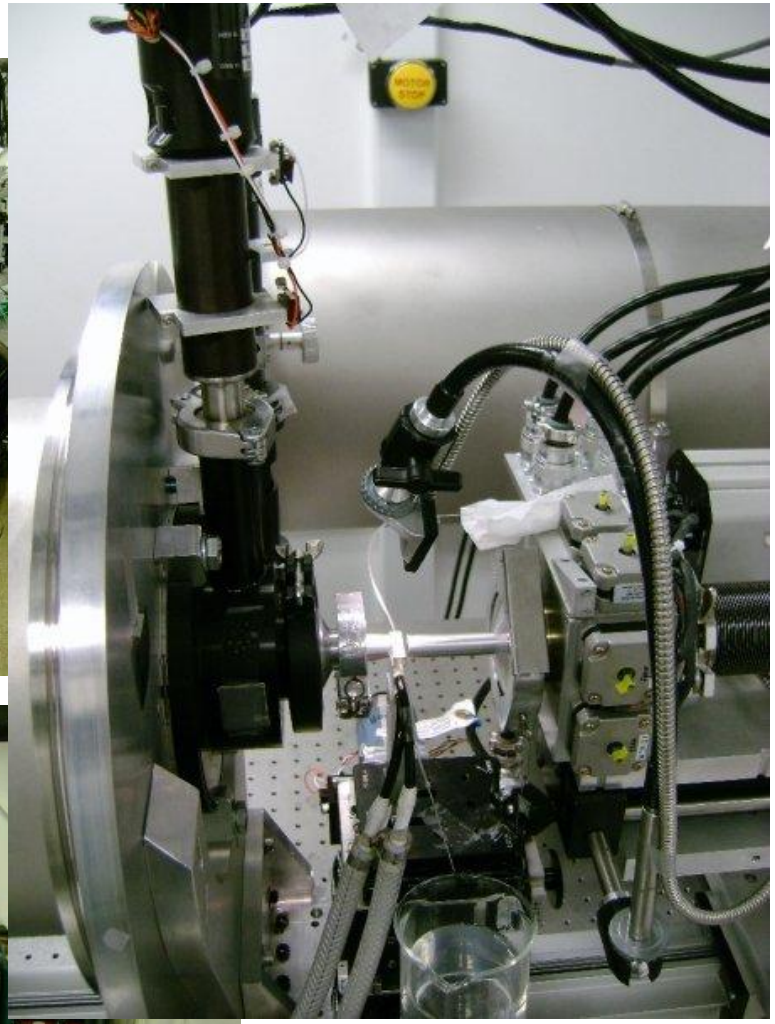


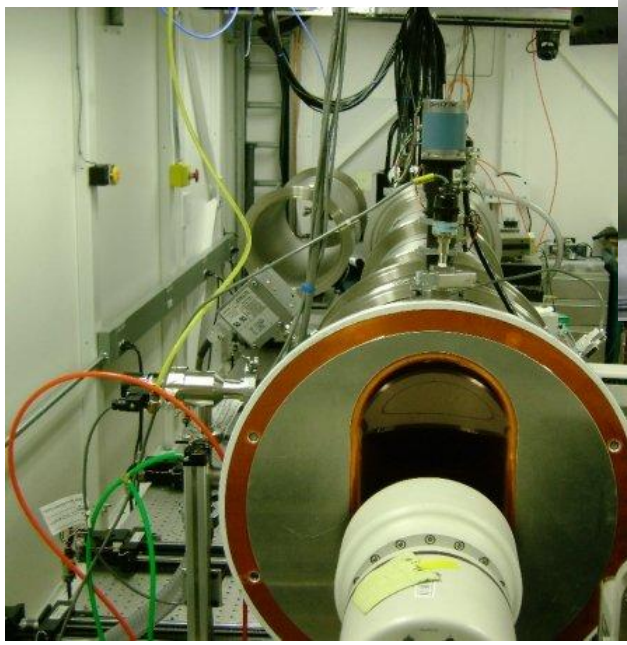
NSD = 0.613, 20 reconstructions

Warning, unlike a crystal structure (which requires a diffracting crystal) 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.

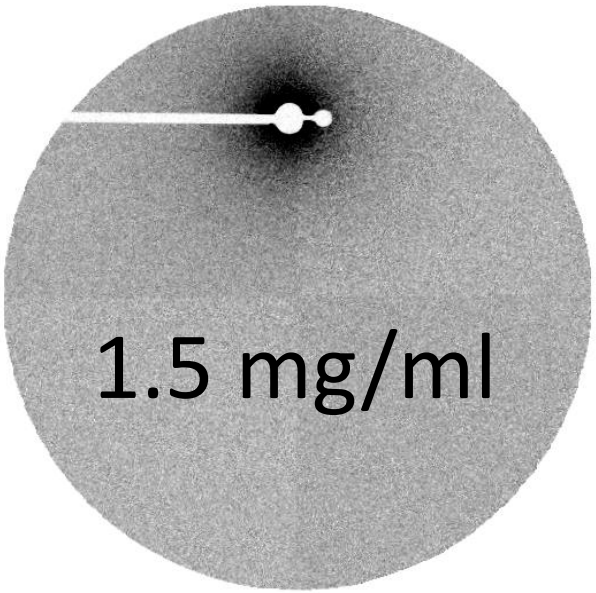
Oscillate sample to minimize radiation damage

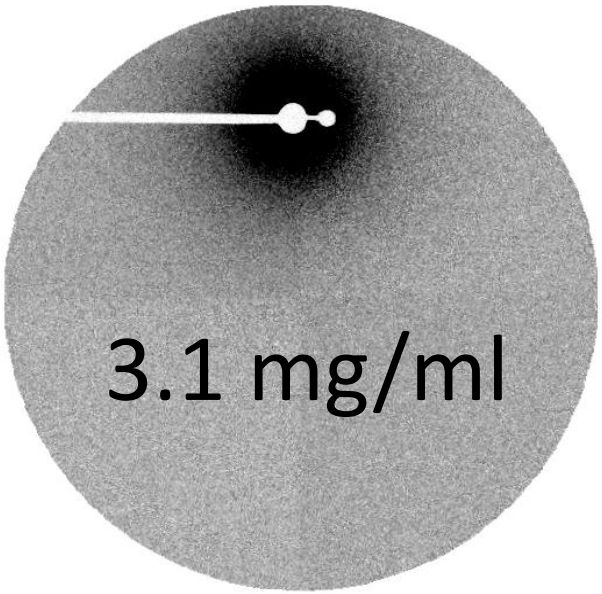
Repeat the buffer.

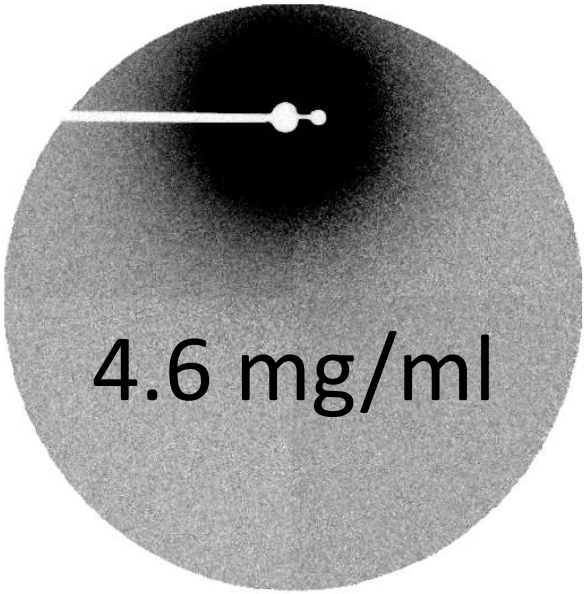
Load next sample

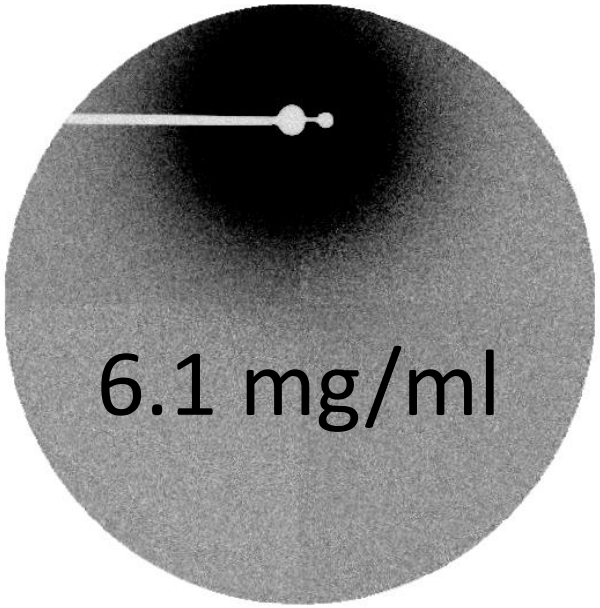
Time per concentration series – approximately 10 to 15 minutes. In high-throughput mode 24 samples in 3 to 4 hours.

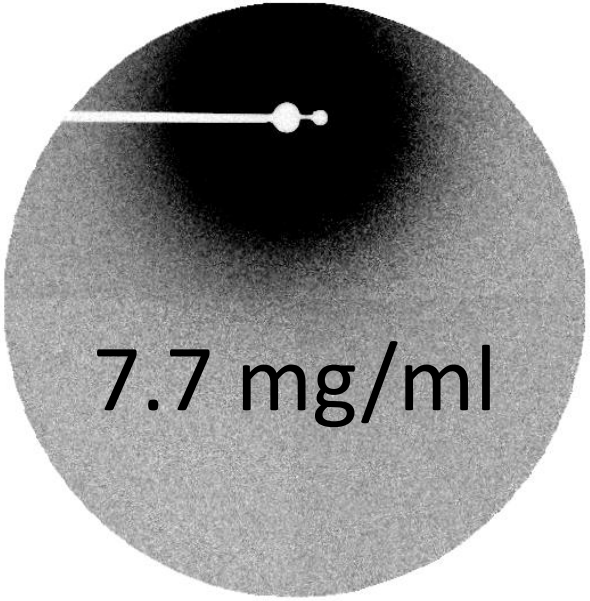
Enables two important things – eat and sleep!


1.5 mg/ml


3.1 mg/ml

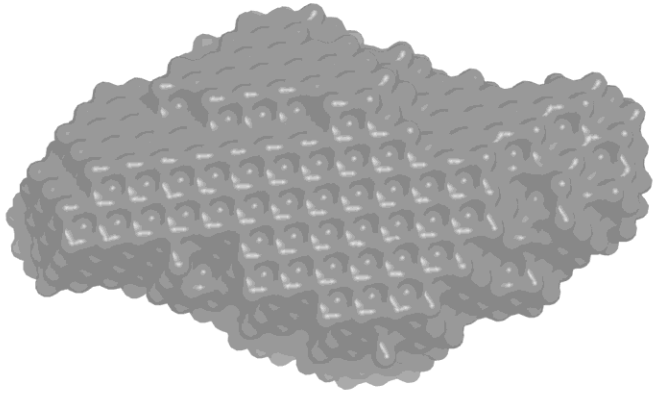

4.6 mg/ml


6.1 mg/ml

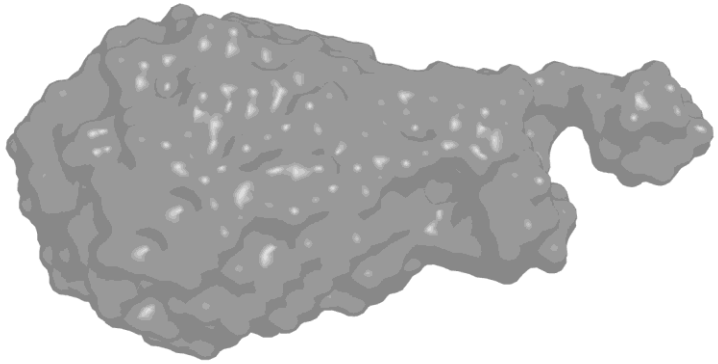

7.7 mg/ml

SAXS can determine *ab initio* molecular envelopes

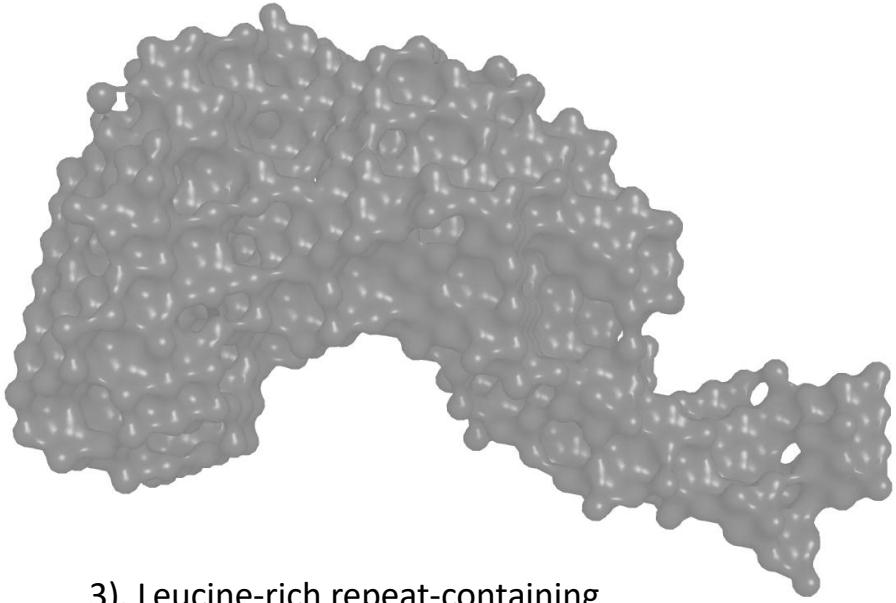
Ab intio envelopes



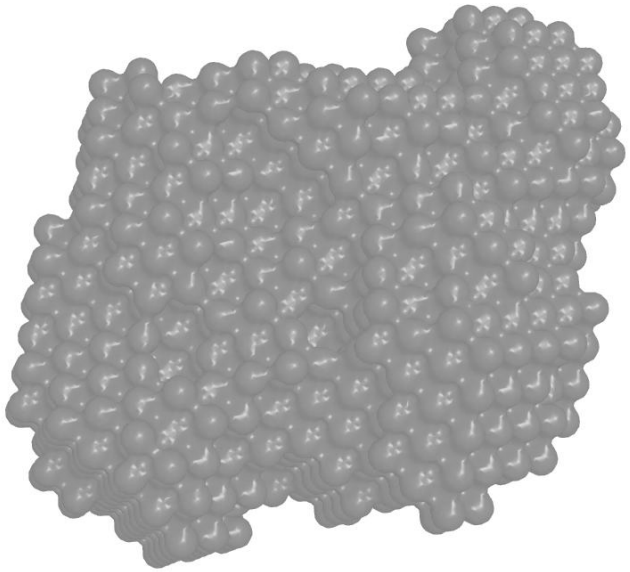
1). alr0221 protein from Nostoc (18.6 kDa)



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



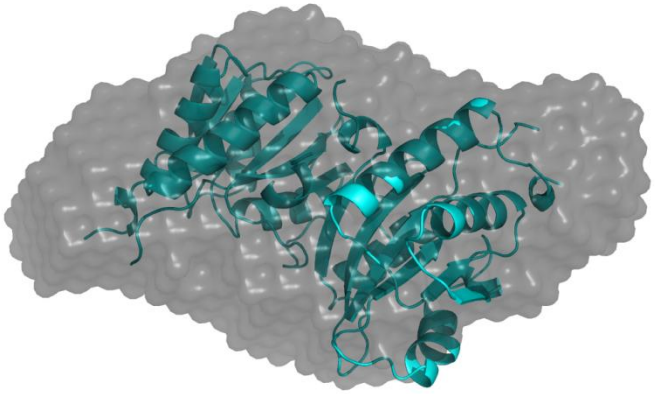
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



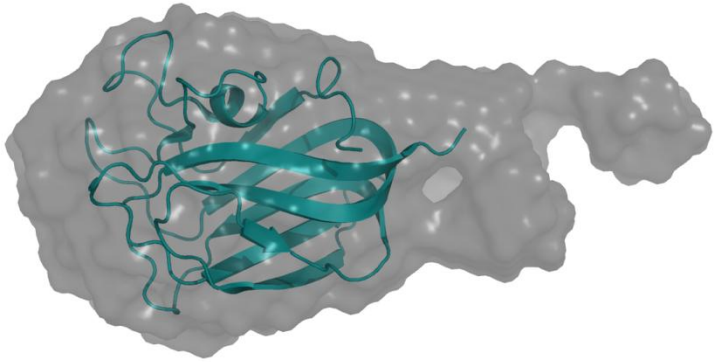
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

These are compatible with structural data

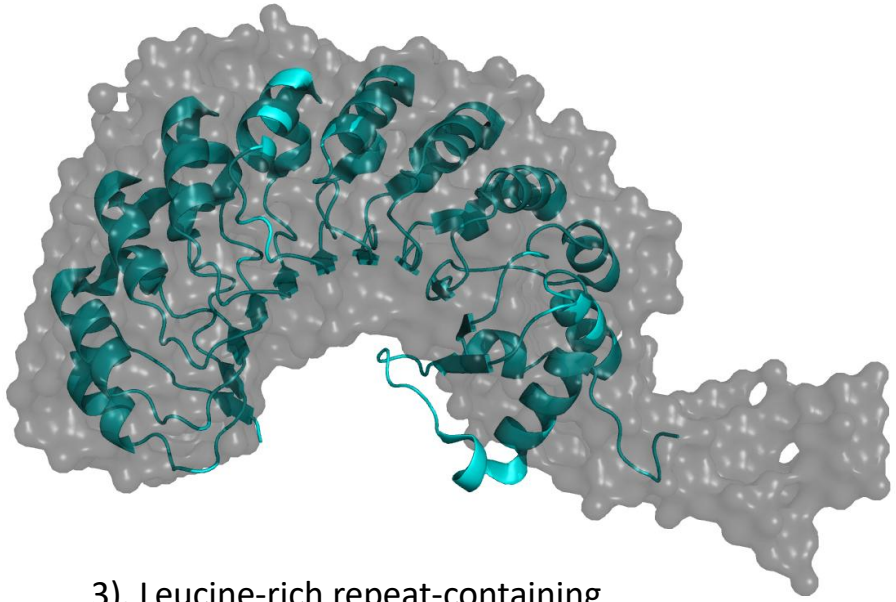
Overlaid with subsequent X-ray structures



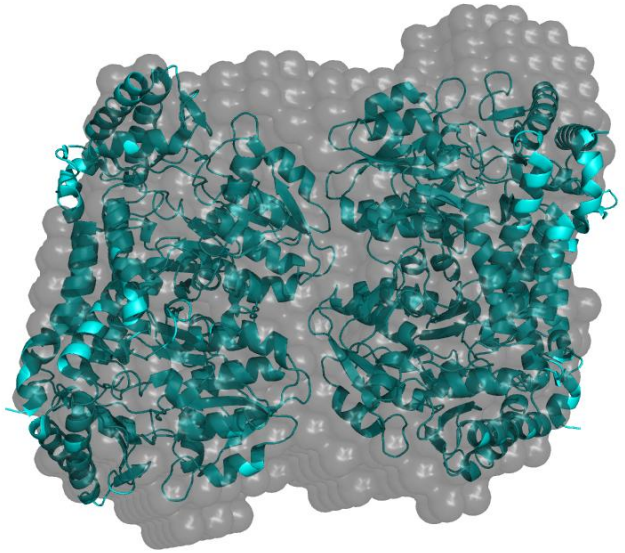
1). alr0221 protein from Nostoc (18.6 kDa)



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



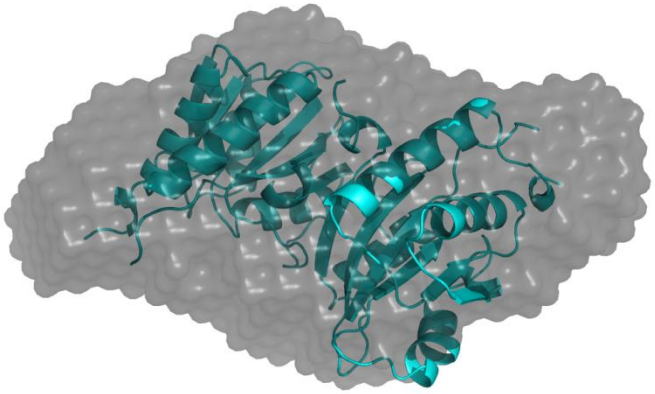
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



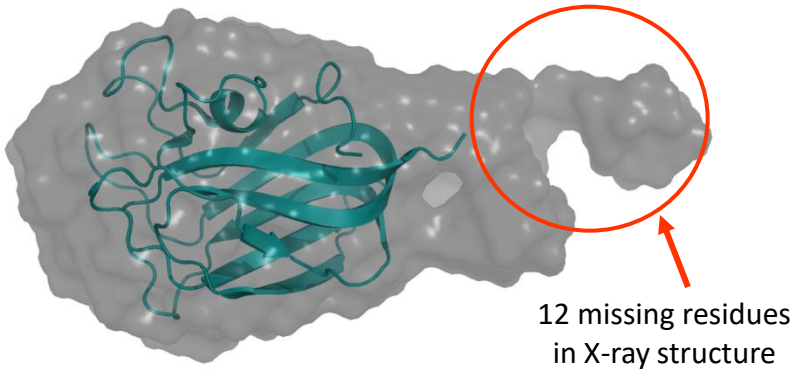
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

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

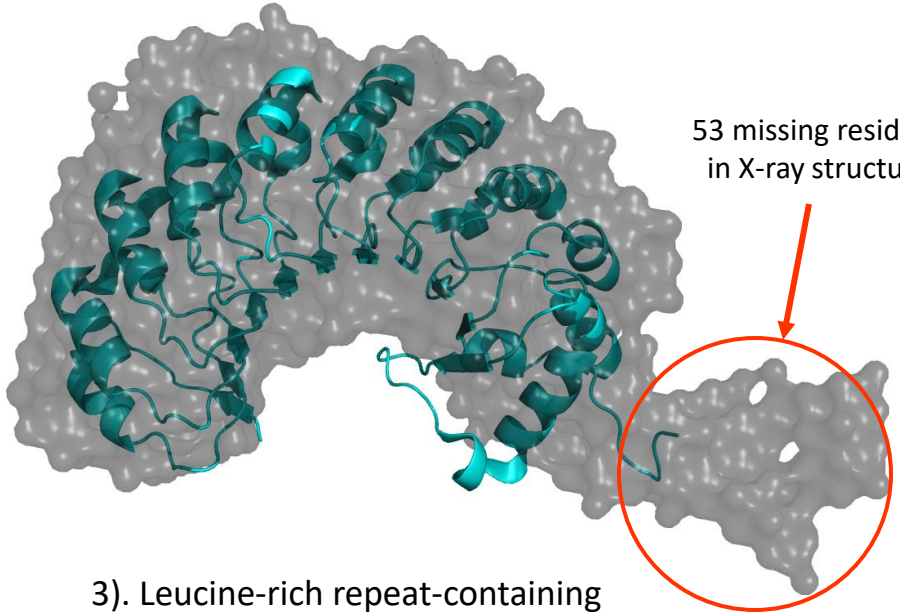
And data on what was missing …



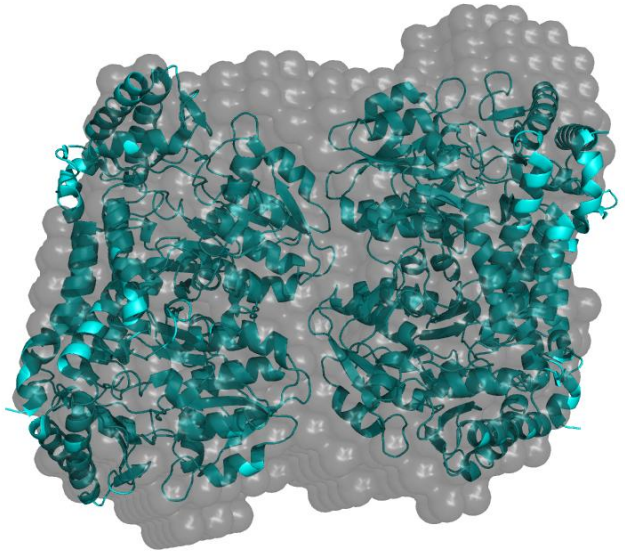
1). alr0221 protein from Nostoc (18.6 kDa)



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



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

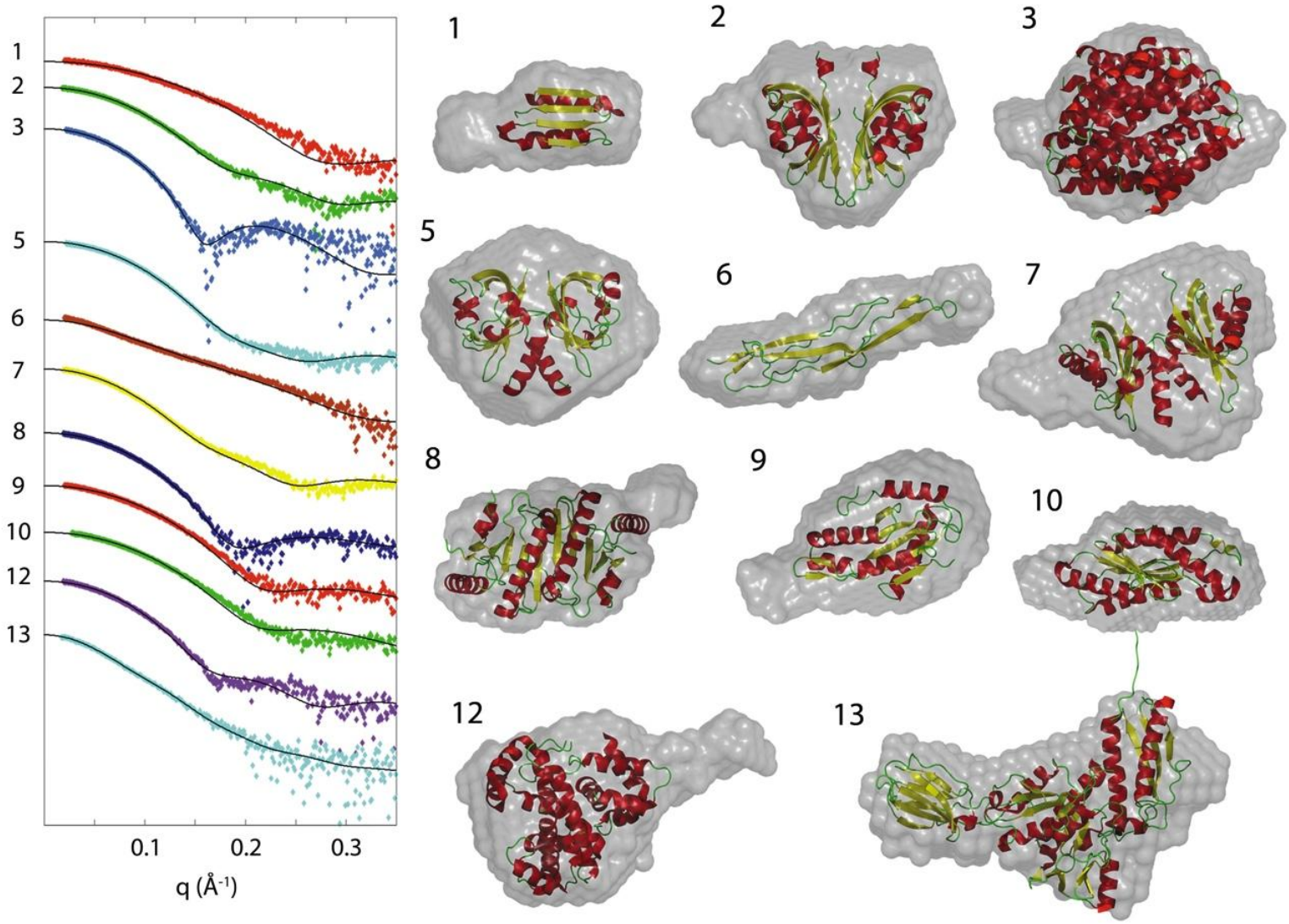


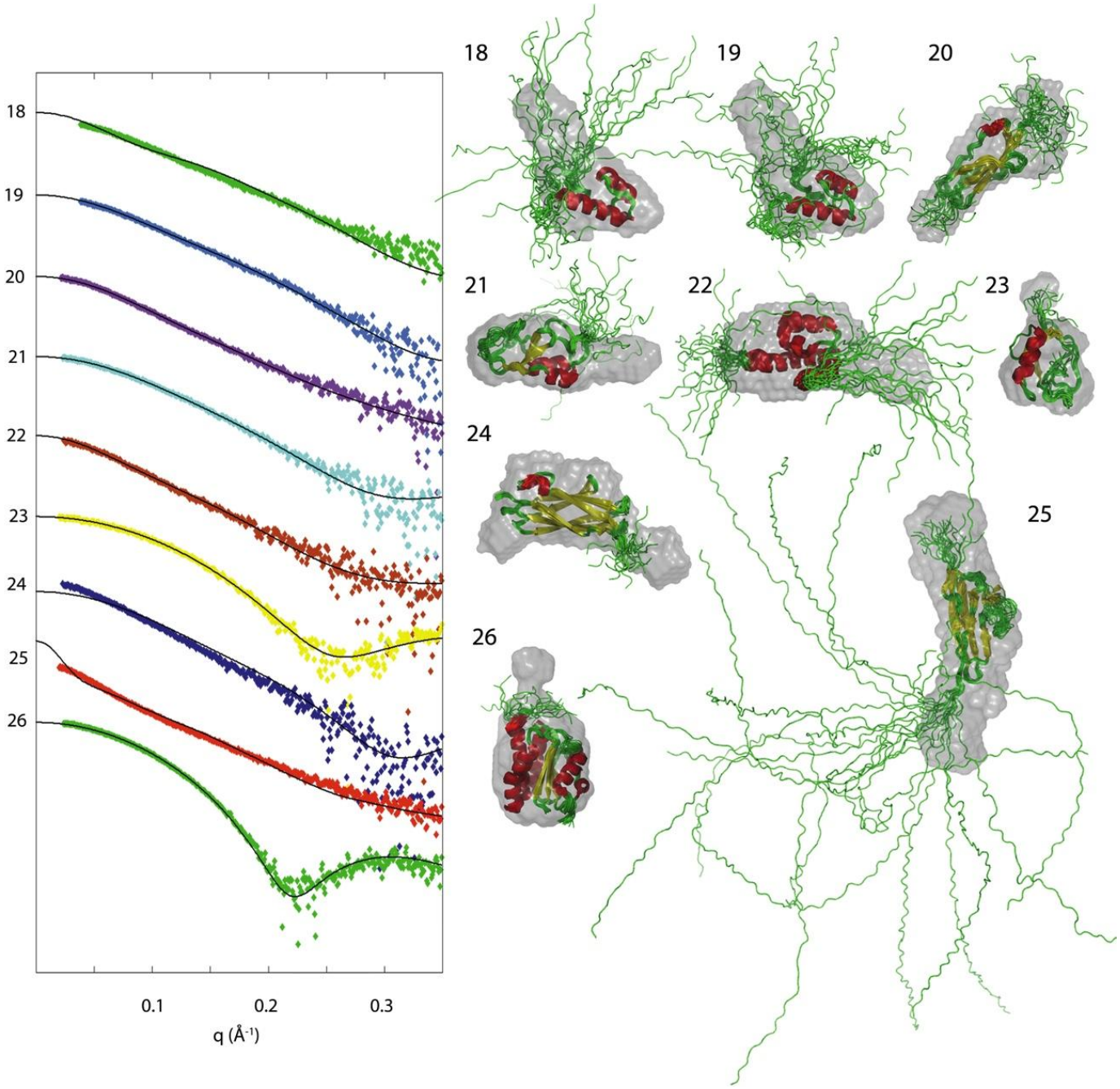
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

Increase the sample numbers

#	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

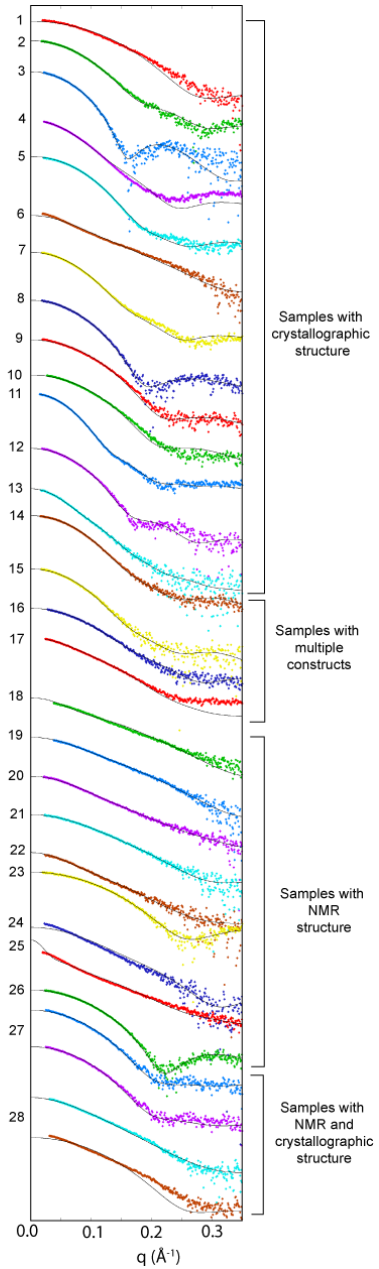
Comparing X-ray structures

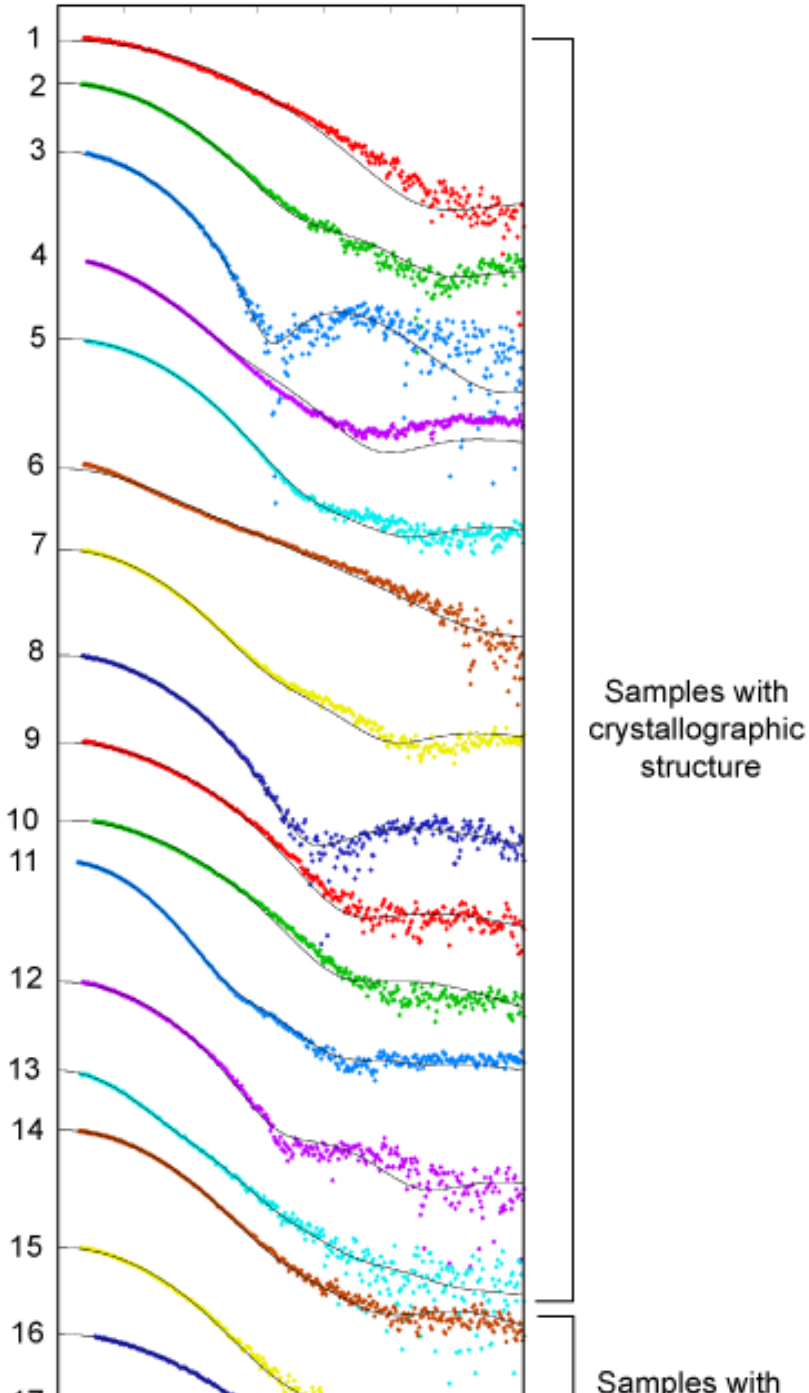




Comparing NMR structures

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

Radius of gyration

Maximum dimension

Molecular weight

Solution oligomer

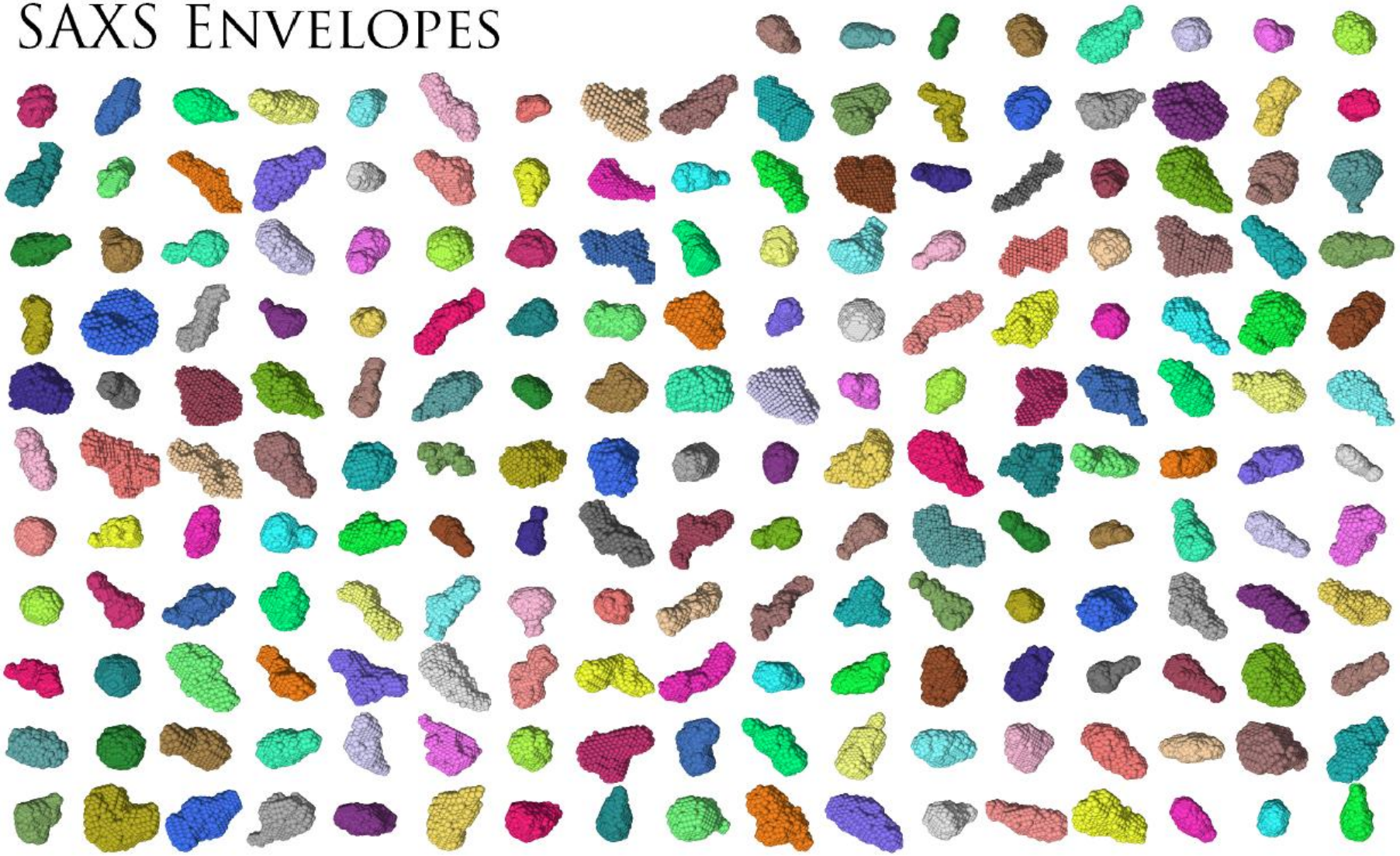
Agreement (or not) with models of structure

#	Residues observed	# Res missing	Rg structure	Dmax structure	Rg SAXS	ΔRg	Dmax SAXS	Δ dmax	Porod MW	MW Ratio	SAXS oligomer[1]	Oligomer Assign.	SAXS fit (χ)
Samples where crystallographic structures were available													
1	74	13	13.7	42.0	14.9	1.2	53.2	11.2	7827	0.8	M		4.2
2	198	12	16.6	67.0	19.8	3.2	67.4	0.4	24555	1.8	D	sym	2.6
3	436	48	22.4	62.3	23.2	0.8	75.3	13.0	50064	3.6	T	sym	1.6
4	214	38	23.3	81.2	23.6	0.3	82.7	1.5	37348	2.6	D/T*	PDB	2.6
5	224	34	19.9	57.6	19.8	-0.1	64.2	6.6	28828	2.0	D	PDB	2.2
6	107	24	19.6	76.3	21.5	1.9	82.0	5.7	11085	0.8	M		6.1
7	236	12	21.4	64.7	22.2	0.8	76.8	12.1	31410	2.0	D	PDB	3.8
8	286	12	20.5	63.1	21.1	0.6	71.4	8.3	34786	2.0	D	PDB	2.0
9	162	9	17.6	54.0	18.7	1.1	65.5	11.5	20468	1.1	M		3.7
10	165	12	17.5	58.0	18.5	1.0	65.8	7.8	19069	0.9	M		4.2
11	330	18	20.1	80.8	20.6	0.5	89.7	8.9	59937	2.9	D/T*	PDB/sym	1.4
12	252	36	21.3	61.5	22.5	1.2	81.9	20.4	37254	1.2	M		2.9
13	416	30	28.5	95.0	27.6	-0.9	98.5	3.5	40027	0.8	M		1.4
Samples where multiple constructs and crystallographic structures were available													
14	272	44	20.8	59.6	21.1	0.3	69.2	9.6	30670	1.9	D	PDB	1.9
15	258	26	21.1	81.8	22.8	1.7	79.7	-2.1	32657	1.9	D	PDB	1.8
16	93	12	18.0	59.5	18.2	0.2	64.7	5.2	15875	1.3	D2	PDB	1.7
17	93	48	20.4	75.0	20.8	0.4	73.0	-2.0	15920	1.0	D1	PDB	2.5
Samples where NMR structures were available													
18	75	0	22.5	122.4	16.8	-5.9	58.4	-64.0	6771	0.8	M		4.7
19	75	0	17.7	94.4	16.5	-1.2	58.4	-36.0	6771	0.8	M		1.4
20	85	0	19.0	80.8	18.7	-0.3	68.0	-12.8	9724	1.0	M		1.7
21	91	0	16.4	71.0	15.9	-0.5	59.6	-11.4	7862	0.8	M		1.5
22	91	0	22.3	123.1	19.6	-2.7	68.0	-55.1	10762	1.0	M		1.6
23	87	7	14.3	55.8	14.5	0.2	49.7	-6.1	8479	0.8	M		1.4
24	114	0	16.5	67.8	16.6	0.1	66.6	-1.2	12600	1.0	M		5.9
25	145	0	49.0	325.5	26.6	-22.4	94.7	-230.8	15386	0.9	M		7.4
26	157	0	19.8	67.5	17.5	-2.3	60.6	-6.9	15238	0.9	M		2.1
Samples where both crystallographic and NMR structures were available													
27*	176	0	18.0	66.7	19.1	1.1	68.3	1.6	22589	2.2	D	PDB	2.5
	158	18	18.1	52.5	19.0	0.9	68.3	15.8				PDB	2.4
28*	114	0	18.5	104.4	18.5	0.0	68.2	-36.2	10721	0.8	M		2.3
	87	13	14.8	44.1	18.4	3.6	68.2	24.1					7.4

Table 2. A summary of structural (crystallography and NMR) and SAXS results. The sample # refers to the identical number in Table 1. The number of unresolved residues in the structure (mainly crystallographic) is listed together with the R_g and D_{max} (in Å) determined from the available structure. The R_g and D_{max} from the SAXS data are shown together with the difference from the available structural information. The molecular weight (in Da) calculated from a Porod analysis is listed along with the ratio of this weight with that derived initially from mass spectrometry in table 1. Finally the SAXS determined oligomer, (Monomer, Dimer or Tetramer), the relationship to the available structure and the χ of the fit are listed. A special case is described below for samples 16 and 17. Further details are given in the text.

Increase the sample numbers even more

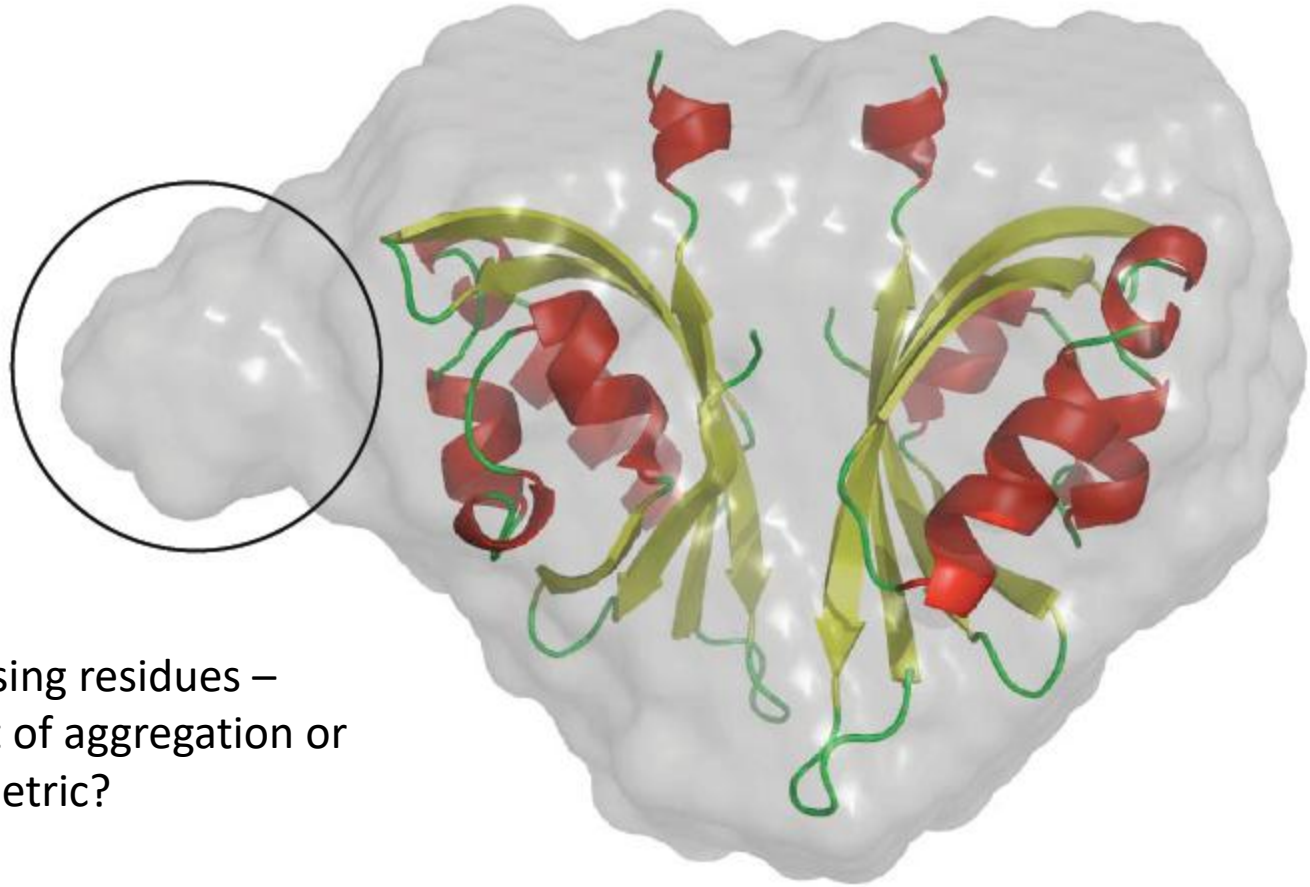
SAXS Envelopes



SAXS : the T-shirt (Tom Grant LLC)

SAXS may provide more questions

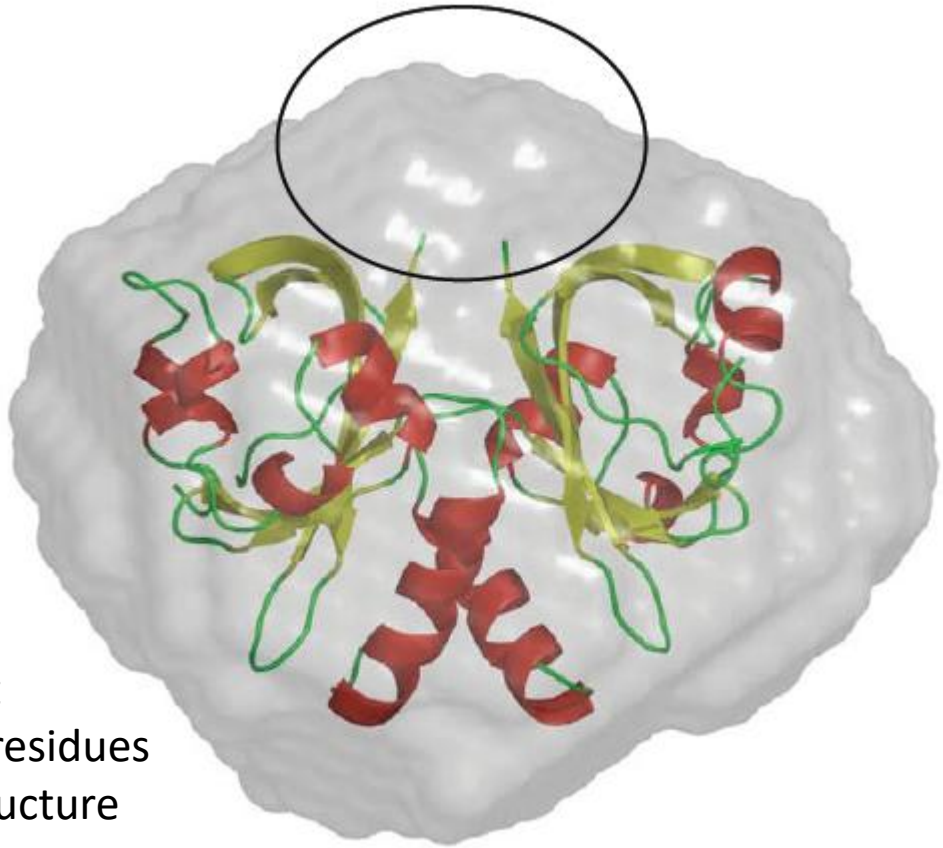
Diguanylate cyclase



12 missing residues – artifact of aggregation or asymmetric?

12 missing residues – artifact of aggregation or assymetric

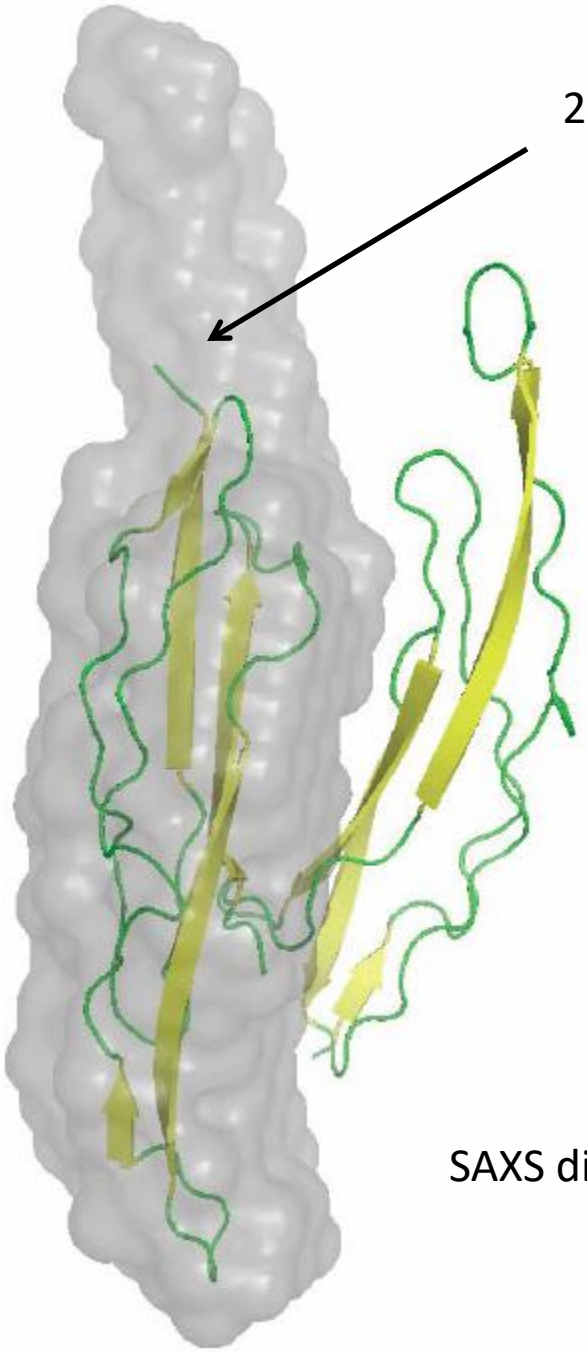
Sensory Box/GGDEF Protein Family



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



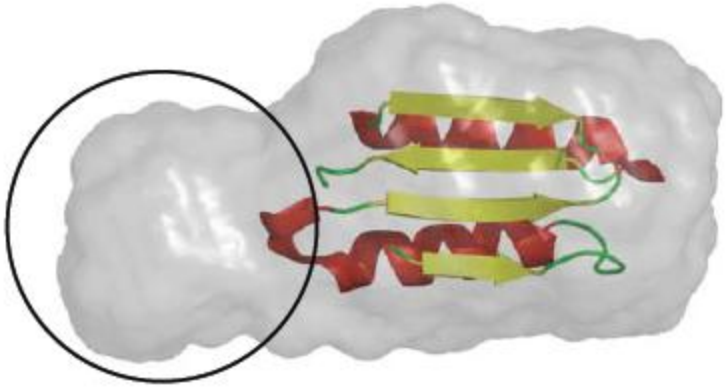
Biological unit was thought to be a dimer from crystallography.

Solution state is not.

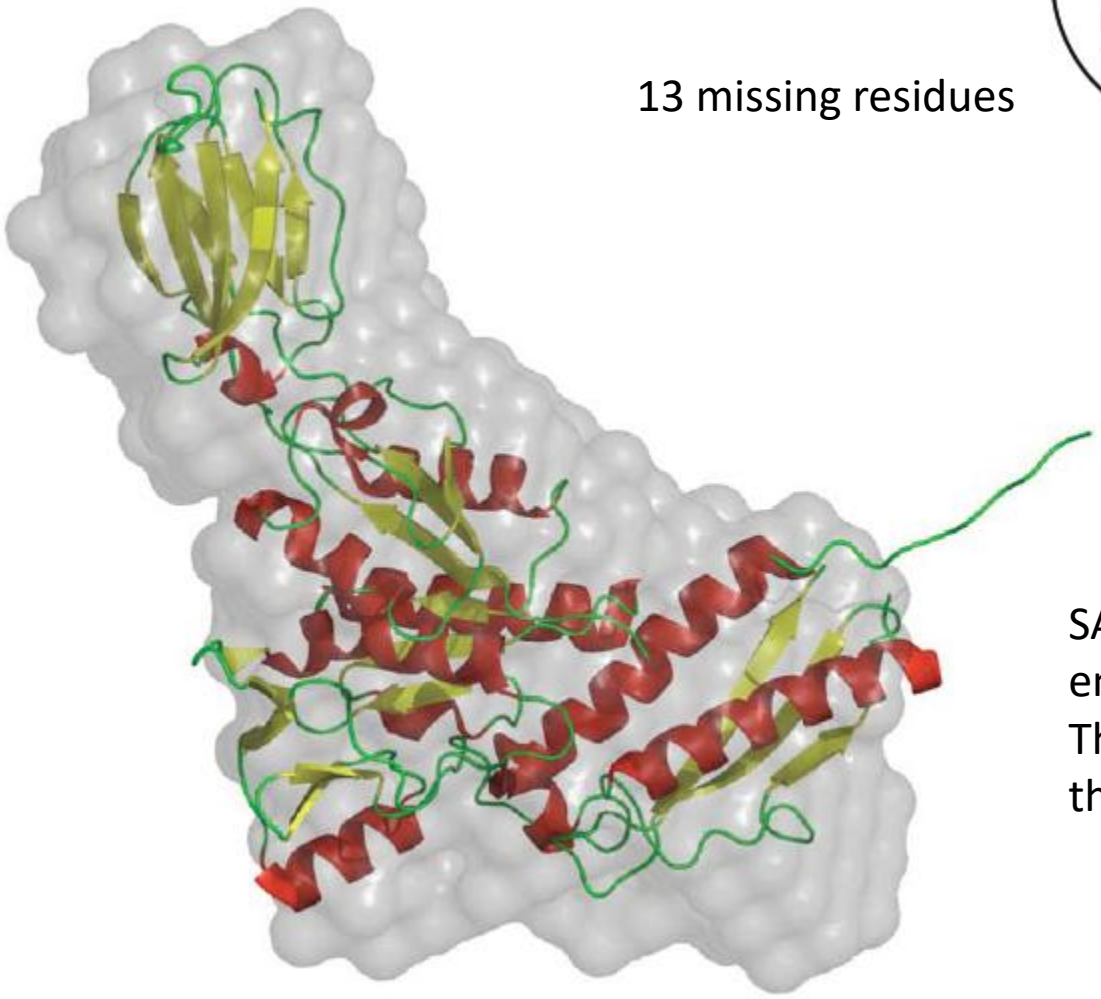
The biological state is not necessarily the solution or crystallographic state.

SAXS distinguishes solution states

Size matters



13 missing residues



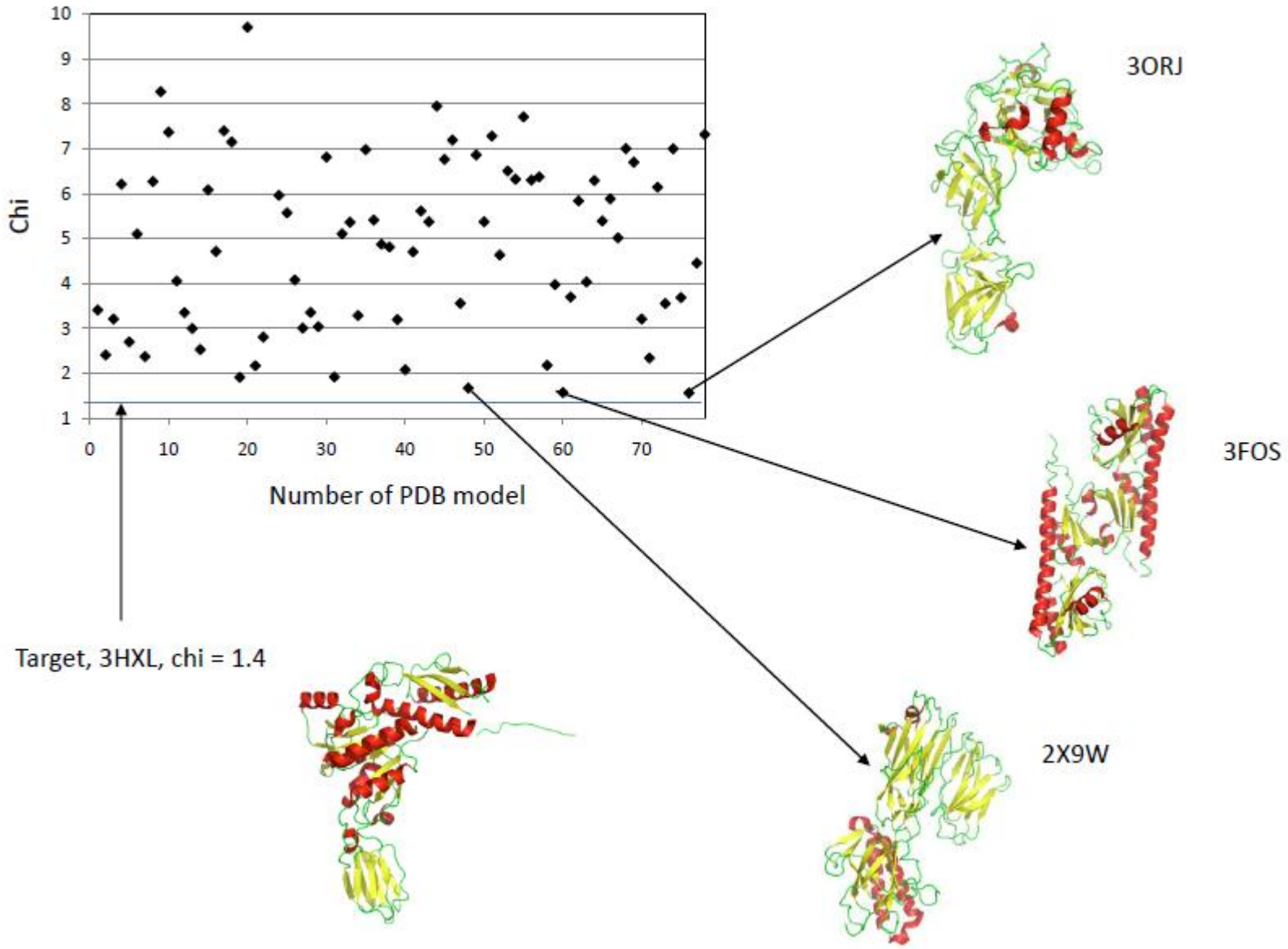
SAXS is not just about shape of the envelope but also it's overall size. The envelope produced reflects the size of the sample.

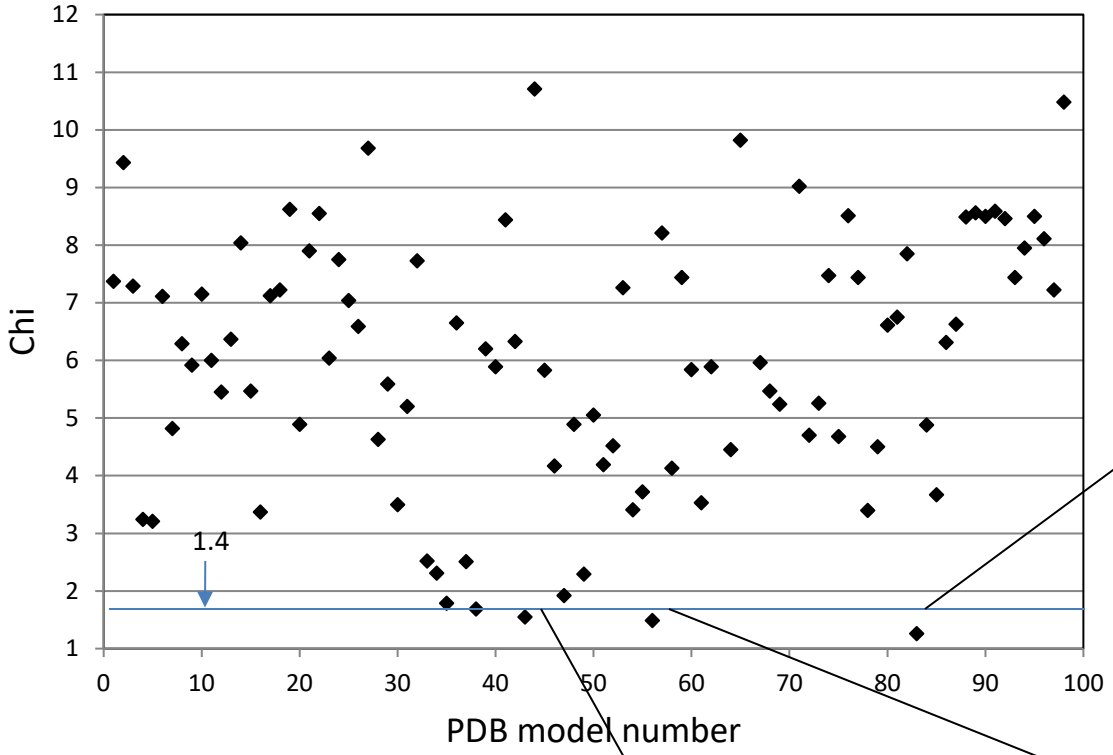
Calculation of the scattering curve is sensitive to the experimental scattering

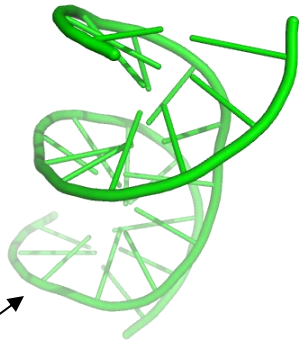
Sample

Data set	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	29	30
1	Samples where crystallographic structures were available																													
2	4.2	15.4	24.5		14.4	5.2	16.3	19.3	7.8	10.0		17.5	23.6	16.6	16.6	9.5		7.1	3.7	3.0	1.4	6.7	3.9	1.7	10.4	9.5	12.2	6.7	4.2	3.5
3	32.1	**2.6**	13.7		3.1	20.6	1.5	5.2	6.8	5.6		3.0	17.4	1.6	1.6	8.9		32.2	21.6	28.3	33.1	19.5	29.2	31.7	31.9	16.7	8.6	5.8	18.2	26.8
4	33.8	7.3	**1.6**		9.5	26.6	5.9	2.5	17.2	15.6		4.9	9.0	5.5	5.5	17.4		31.2	25.9	30.0	33.6	24.7	32.3	32.7	29.9	25.0	20.0	17.6	25.5	31.0
5	76.5	10.6	16.5	**2.6**	12.2	56.1	6.1	12.0	30.5	26.5		10.3	16.0	6.9	6.8	27.5		72.8	55.2	66.9	76.7	50.7	71.7	74.4	67.0	50.5	36.9	31.0	52.0	68.0
6	74.4	3.7	20.1		**2.2**	56.5	7.5	3.8	14.9	9.6		2.7	31.9	5.9	6.3	27.5		82.0	59.0	70.2	78.7	55.0	69.1	75.8	73.9	41.2	23.7	15.6	51.4	63.8
7	18.3	20.5	32.3		20.5	6.1	20.5	25.9	17.2	17.8		23.1	25.7	20.9	21.0	7.1		5.8	3.7	6.4	11.8	4.2	16.9	11.2	6.7	19.1	18.9	16.4	8.5	17.2
8	57.6	6.5	13.1		8.9	39.8	**3.8**	10.5	22.0	18.8		7.0	14.4	4.9	4.5	18.3		51.0	38.7	48.6	56.7	34.2	53.3	55.2	47.7	37.4	26.7	22.1	36.2	50.8
9	34.4	3.8	5.1		2.7	24.5	4.2	**2.0**	10.1	8.2		3.6	12.9	3.6	3.7	12.3		32.1	24.4	29.8	34.6	21.9	31.7	33.7	30.9	20.3	13.9	10.9	22.2	29.9
10	18.9	4.1	18.1		3.1	10.7	4.8	7.9	**3.7**	3.6		5.9	18.6	4.5	4.8	7.2		21.2	12.7	16.0	19.5	11.7	16.0	18.7	20.0	7.1	3.4	3.4	8.1	14.5
11	20.4	4.9	22.4		3.0	12.2	5.8	10.4	4.1	**4.2**		7.5	20.8	5.7	6.1	7.9		25.5	15.1	19.3	22.2	13.7	17.3	21.2	21.9	8.0	3.5	3.5	9.5	15.4
12	94.2	37.1	19.8		41.6	77.8	31.1	26.3	59.4	56.0	**3.0**	31.0	15.7	30.8	30.7	54.9		84.3	75.8	86.4	93.8	71.1	91.1	92.2	78.1	75.8	65.2	60.0	75.6	88.4
13	33.2	3.2	4.2		4.6	23.8	2.7	3.1	12.8	10.9		**2.9**	9.0	2.5	2.5	11.9		29.2	23.3	28.9	33.3	21.0	31.3	32.2	28.1	21.9	15.9	13.1	22.4	29.5
	26.4	9.3	7.9		10.4	19.2	7.5	8.0	15.1	14.1		8.3	**1.4**	7.7	7.6	11.1		20.0	18.0	22.4	25.7	15.8	25.2	25.0	18.5	20.1	16.6	15.2	18.6	24.3
	Samples where multiple constructs and crystallographic structures were available																													
14	41.6	3.5	9.1		4.6	28.0	1.7	6.0	13.0	10.6		3.9	13.9	**1.9**	1.8	12.0		37.7	27.7	35.4	41.6	24.6	38.3	40.2	37.5	25.1	16.8	13.2	25.4	36.1
15	19.3	2.5	4.1		2.7	12.8	1.7	3.0	6.8	5.8		2.5	5.5	1.9	**1.8**	5.2		16.3	12.4	16.2	19.1	10.9	18.0	18.5	16.4	12.2	8.4	6.9	11.9	16.9
16	8.9	3.8	12.5		3.2	4.3	4.5	7.1	3.0	2.8		5.6	12.0	4.7	4.8	**1.7**		10.5	5.8	6.5	8.5	5.3	7.3	8.1	8.6	4.2	2.5	2.6	2.9	6.5
17	11.8	9.7	21.2		9.2	3.4	10.1	14.4	7.6	7.5		12.2	18.1	10.3	10.4	2.5	**2.1**	10.8	5.3	5.6	9.8	3.7	10.0	9.7	8.6	8.8	7.8	7.1	2.3	9.3
	Samples where NMR structures were available																													
18	7.0	16.5	26.3		15.6	2.0	17.1	20.6	9.8	11.5		18.6	22.1	17.4	17.5	7.9		**4.7**	**1.4**	1.3	2.3	3.8	6.6	2.0	2.7	11.4	13.6	8.7	4.8	6.3
20	10.2	13.0	22.8		12.6	1.9	13.5	17.2	9.2	9.8		15.2	19.4	13.7	13.8	4.2		6.3	2.0	**1.7**	6.0	2.2	9.1	5.5	5.5	10.7	11.1	8.5	3.2	9.3
21	5.2	14.8	24.3		13.7	3.4	15.6	18.6	7.6	9.5		16.9	21.8	15.8	16.0	7.9		5.4	2.4	1.8	**1.5**	4.6	4.7	1.7	5.7	9.0	11.7	6.6	3.5	4.5
22	6.3	6.6	12.0		6.4	1.6	6.8	8.8	5.3	5.3		7.7	10.1	6.9	6.9	1.8		3.9	1.4	1.8	4.3	**1.6**	5.6	4.0	3.2	5.9	5.7	5.0	1.8	5.6
23	1.6	10.1	16.9		8.9	6.6	10.9	12.5	3.7	5.6		11.5	17.1	11.0	11.1	8.1		8.0	5.7	5.3	3.6	7.1	**1.4**	4.3	10.8	4.1	7.3	3.5	4.5	1.8
24	7.8	6.3	12.7		6.1	1.8	6.4	8.9	5.7	5.6		7.6	10.2	6.6	6.7	1.6		6.0	2.0	4.2	6.5	2.0	6.9	**5.9**	5.2	6.4	5.6	5.4	2.0	6.6
25	18.3	15.8	24.8		16.2	7.9	15.2	19.7	15.9	15.7		17.6	17.5	15.5	15.5	6.1		6.3	5.1	9.0	14.2	3.5	17.0	13.6	**7.4**	17.0	15.7	15.3	8.3	16.8
26	16.4	8.7	26.0		4.9	13.0	10.8	12.7	1.9	2.3		11.2	24.8	10.3	10.8	11.8		26.4	16.8	18.1	19.8	15.9	13.5	18.9	21.9	**2.1**	3.1	2.9	10.9	11.3
	Samples where both crystallographic and NMR structures were available																													
27	13.9	2.4	10.2		2.5	8.3	2.5	4.0	2.7	2.5		3.0	10.7	2.2	2.3	4.8		14.2	9.0	10.5	13.6	8.1	12.0	13.3	12.2	5.8	**2.5**	**2.4**	6.4	11.1
28	8.3	13.1	23.9		12.2	1.7	13.6	17.5	7.6	8.7		15.4	19.5	13.9	14.0	3.8		6.1	2.0	1.8	4.5	1.8	7.3	4.0	3.1	9.1	10.3	6.9	**2.3**	**7.4**
	18%	6%	11%	18%	15%	22%	5%	4%	6%	7%	5%	14%	7%	16%	10%	13%	52%	0%	0%	0%	0%	0%	8%	0%	0%	0%	0%	11%	0%	15%

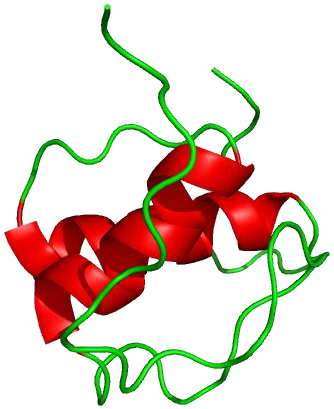
Comparing 100 nearest molecular weight PDB entries

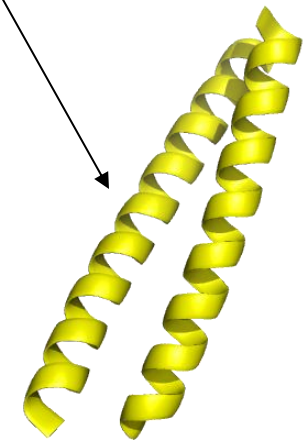


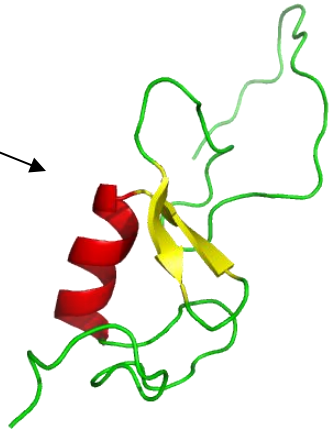




Actual structure







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?

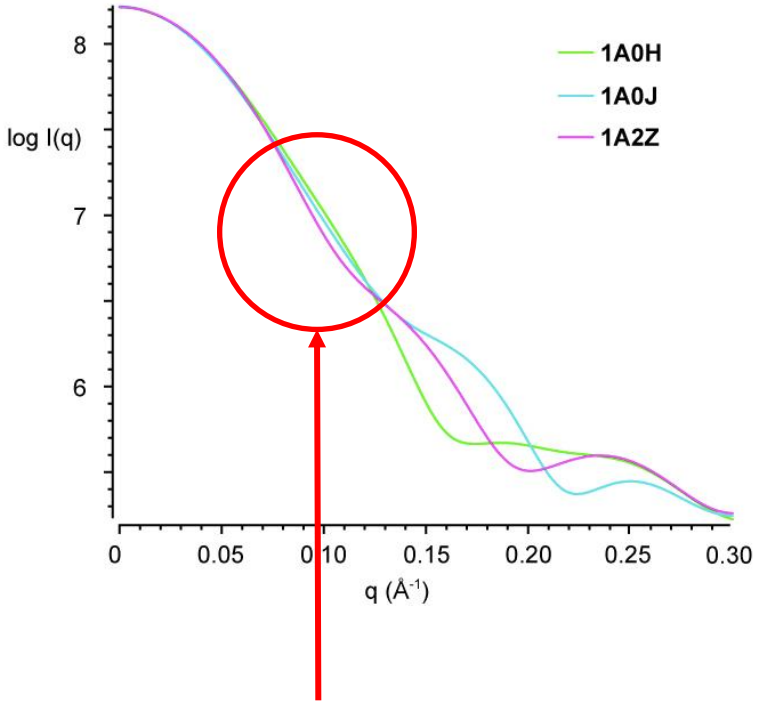
Most important point

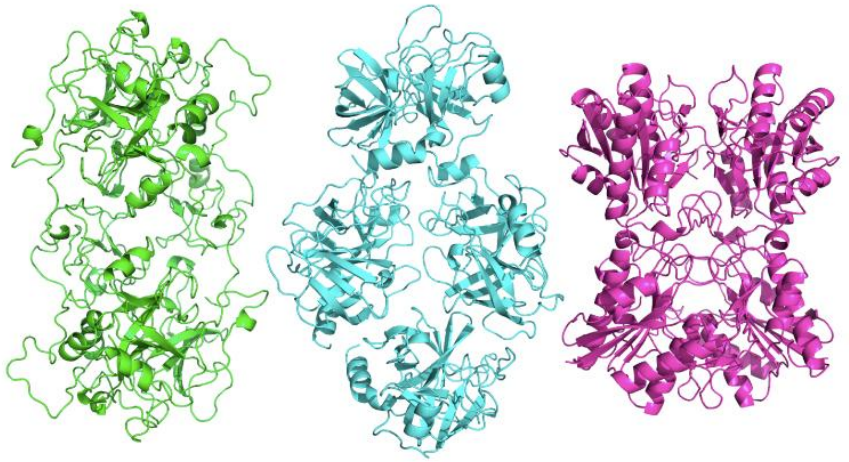
- While envelopes look good they are the least important feature of SAXS analysis and probably the least useful (unless you are trying to keep your friendly molecular biologists happy).
- The SAXS scattering profile can be accurately calculated given a structural model.
- The strength of SAXS lies in:
  - Being able to invalidate models
  - To generate hypothesis
  - To place known structural data
  - To characterize your protein
  - (and other stuff beyond basic uses)

Beyond the envelope

- testing models
- extending existing structure
- Placing known structural components
- Understanding mixtures
- (Distinguishing oligomers)
- (more complex dynamics studies)

Structures with very similar radius of gyrations can have very different scattering curves



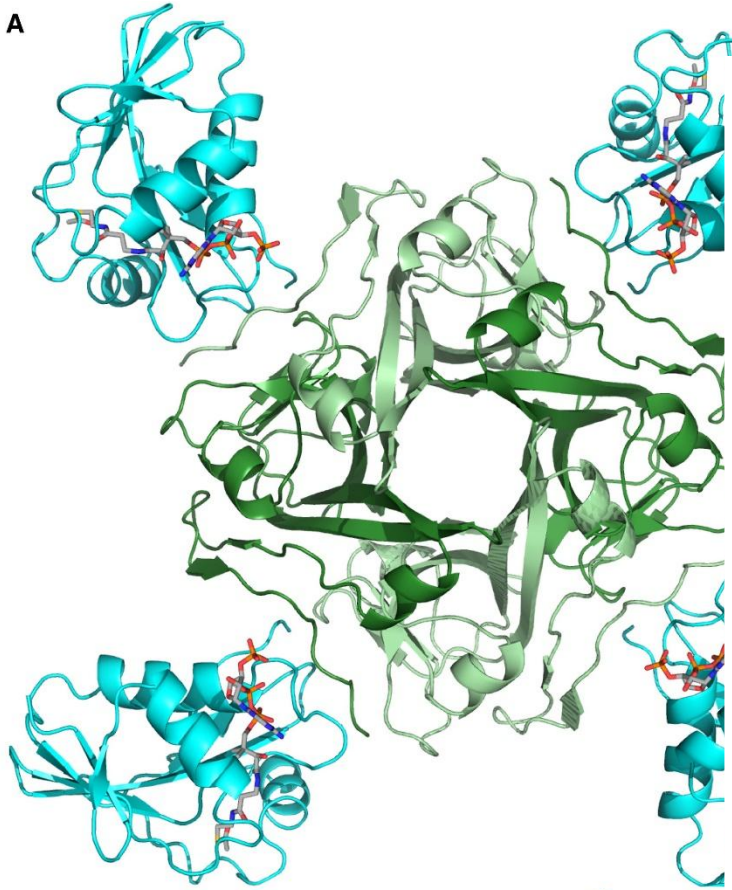


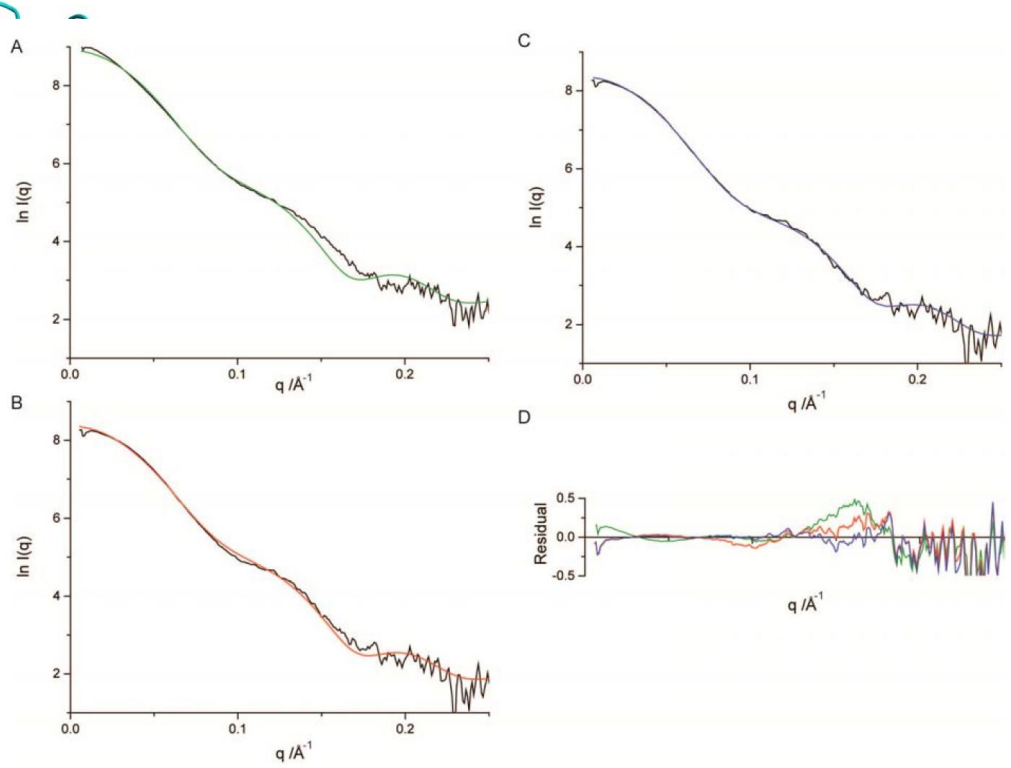
1A0H
MW: 96.8 kDa
Rg: 31.5 Å

1A0J
MW: 96.6 kDa
Rg: 32.4 Å

1A2Z
MW: 99.4 kDa
Rg: 30.9 Å

Note that even this is a significant difference in SAXS data





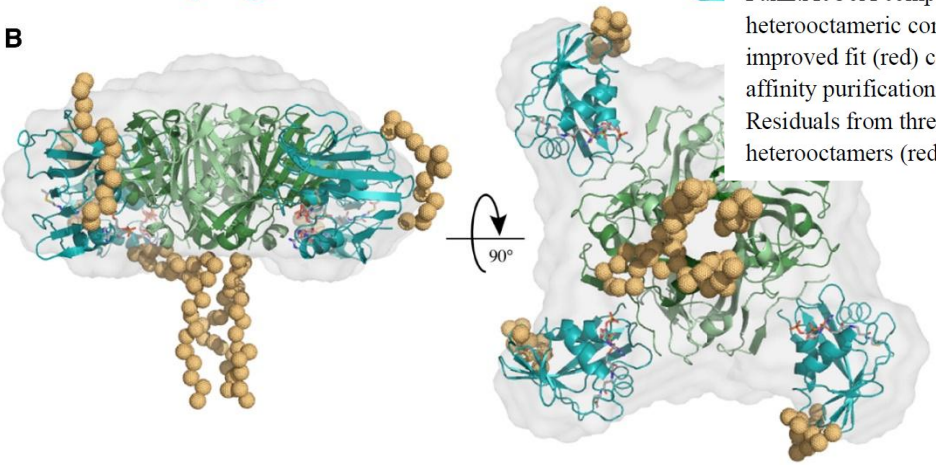
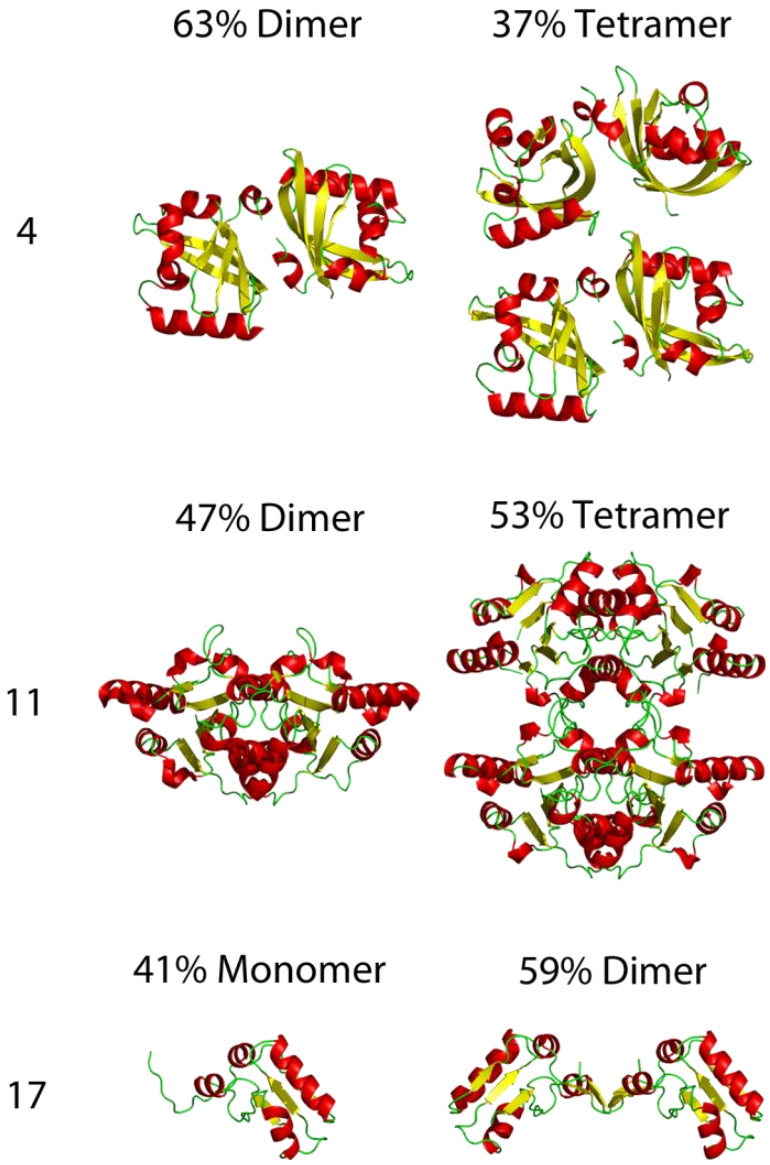


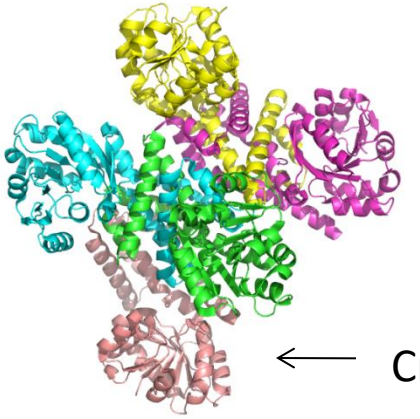
Figure S2, related to figure 2 SAXS data and fitting. **A** Raw SAXS data for the PanD(T57V)-PanZ.AcCoA complex (black) compared with predicted data for the crystallographically resolved heterooctameric complex (green). **B** Inclusion of a population of dimers of heterooctamers leads to an improved 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)



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

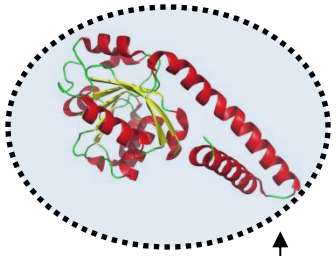
Another story



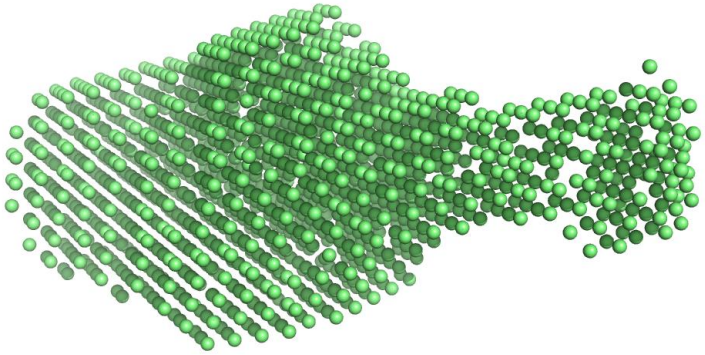
2izz from the PDB
(5 chains in PDB)

← Crystal packing artifact

3gt0 from the PDB

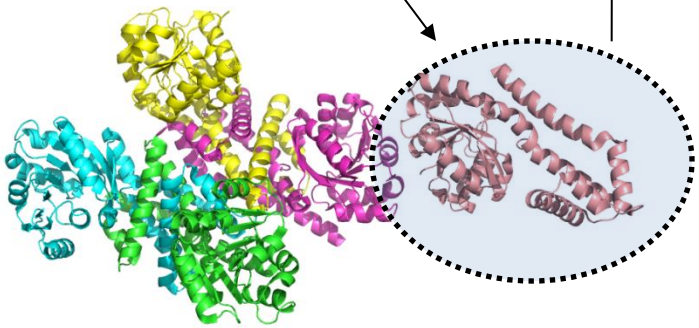


Solution envelope from BcR38B-21.20-SeMa-Gf (3gt0)



~165A

Correct position for 5th chain



~165A

Biological unit based
on 2izz and SAXS

How accurate is the information in a SAXS curve?

A SAXS curve is a continuous sampling of the molecular transform.

It contains a few (10-15 reflection equivalents) if we take the Fourier approach.

These are low resolution information.

However, these are continuously sampled so each distinct information point (Shannon channel) is extensively over sampled.

It’s low resolution information but it’s very accurate low resolution information.

An example of the use of SAXS with crystallography and molecular dynamics

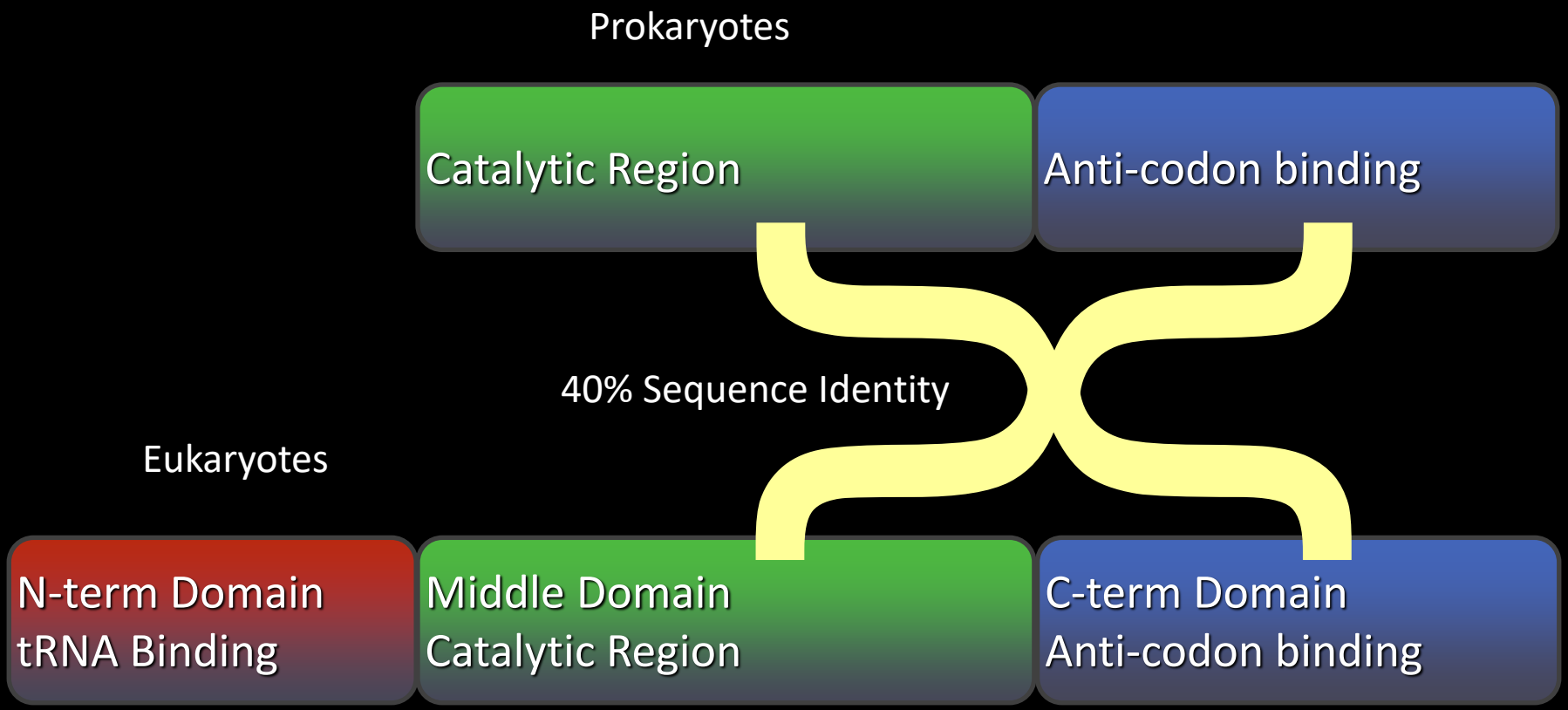
tRNA synthetase of Eukaryotes and Prokaryotes

- Most of our structural knowledge of tRNA synthetases comes from prokaryotes

Appended Domains

- Eukaryotic tRNA synthetases often carry appended domains not present in prokaryotic homologs
- These domains are known to bind RNA non-specifically
- Little is known about their function or structure

Glutamine tRNA Synthetase



Target

- Our target today is Glutaminyl tRNA synthetase (Gln4) from yeast *Saccharomyces cerevisiae*
- Yeast *Saccharomyces cerevisiae is* a well-established model system for understanding fundamental cellular processes of higher eukaryotic organisms.

- Many eukaryotic tRNA synthetases like Gln4 differ from their prokaryotic homologs by the attachment of an additional domain appended to their N or C-terminus, but it is unknown how these domains contribute to tRNA synthetase function, and why they are not found in prokaryotes

- The 228 amino acid N-terminal domain of Gln4 is among the best studied of these domains, but is structurally uncharacterized.

- The N-terminal domain appears to have non specific RNA binding.

- The role of a nonspecific RNA binding domain in the function of a highly specific RNA binding enzyme is baffling, but clearly crucial given its prevalence among tRNA

Crystallization/Data collection

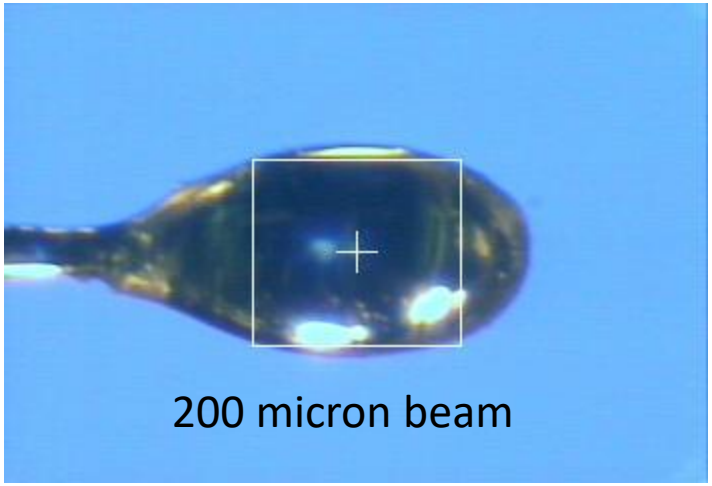
- Gln4 Screened against 1536 different biochemical conditions, ~1000 forming an incomplete factorial of chemical space and ~500 representing commercially available screens.
- Crystal leads seen, several were chosen based on ease of cryoprotection of the native hit.
- Crystals were optimized with a Drop Volume Ratio versus Temperature (DVR/T) technique.
- Cryoprotected and 'drop' shipped to SSRL by FedEx.

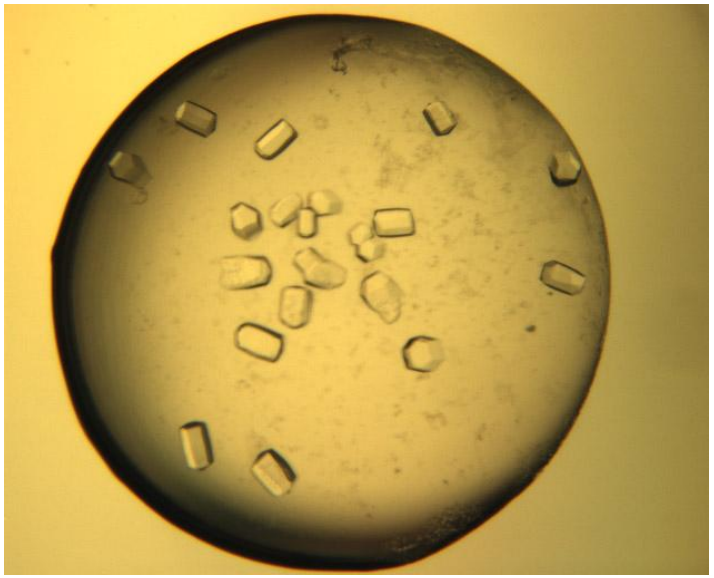
- Only 2 structures for related glutaminyl tRNA synthetases are available (~40% sequence homology), we had 228 extra residues (almost 40% more residues) therefore we expected problems in molecular replacement and didn't have a SeMet example.

- EXAFS data indicate Zinc present in the *E. coli.* Case (not seen in the X-ray structure). The zinc acts to stabilize the structure in a pseudo zinc finger motif.

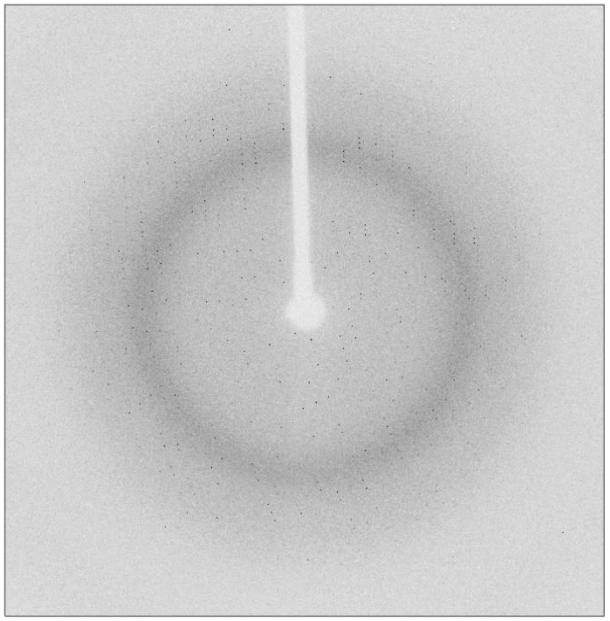
- We collected data remotely with an excitation scan to determine if Zinc was present.

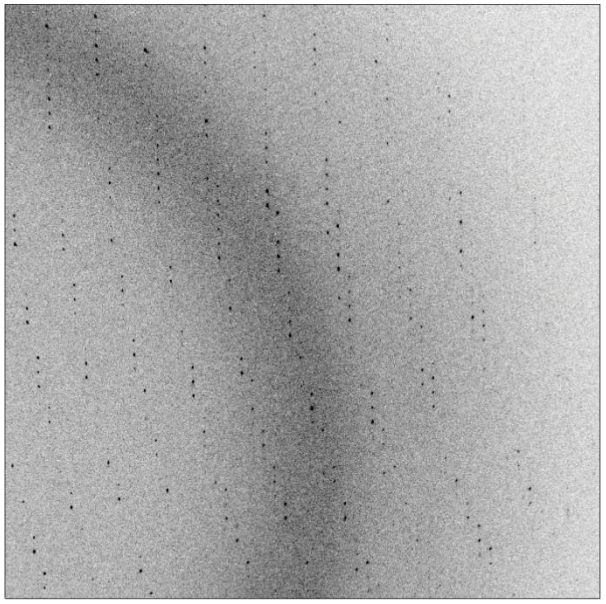
- It was!





80% PEG 400 in the
crystallization cocktail

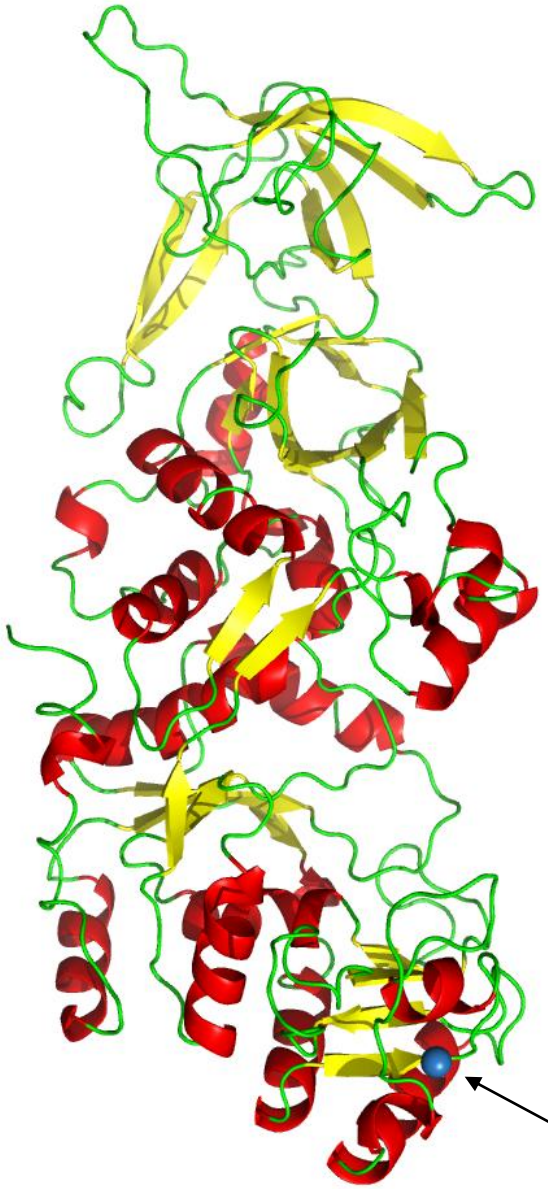




Data collection/Processing

- We used beamline 11-1 at SSRL with a Mar 325 CCD detector, 340 mm crystal to detector distance.
- We collected 200˚ of data, 0.4˚ per frame, 500 images, 3.7s per frame, wavelength 1.169 Å (as close as we could get to Zinc on the beamline used) (deliberately high redundancy for the anomalous signal).
- We indexed in P3121, a=b=176.75 Å, c=72.22 Å, α=β=90, gamma=120˚

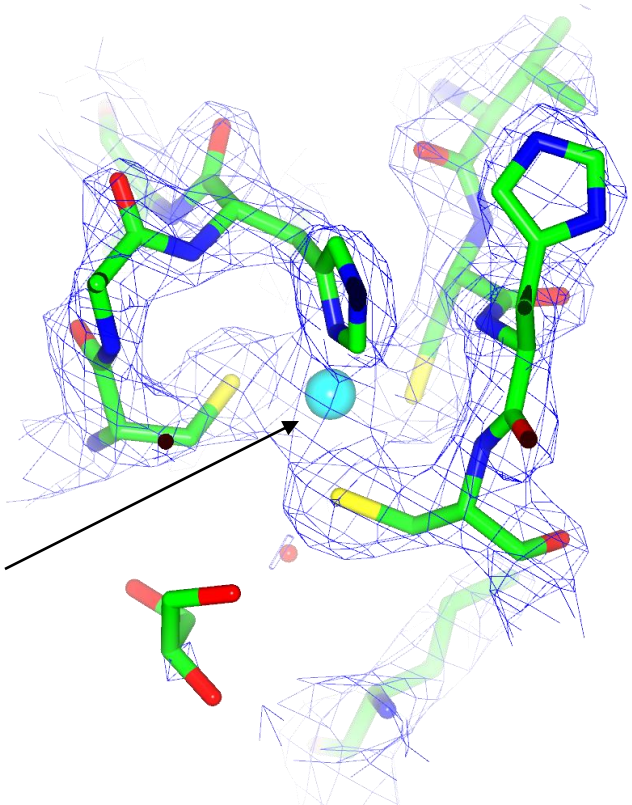
	Overall	Inner Shell	Outer Shell
Low resolution limit (Å)	40.00	40.00	2.64
High resolution limit (Å)	2.5	7.91	2.5
R_{merge}	0.104	0.036	0.743
R_{pim}	0.032	0.011	0.273
	3.2%	1.1%	27.3%
Total number of observations	508484	17694	51511
Total number unique	44752	1523	6332
Mean((I)/sd(I))	24.6	86.6	2.2
Completeness (%)	99.7	99.9	97.9
Multiplicity	11.4	11.6	8.1

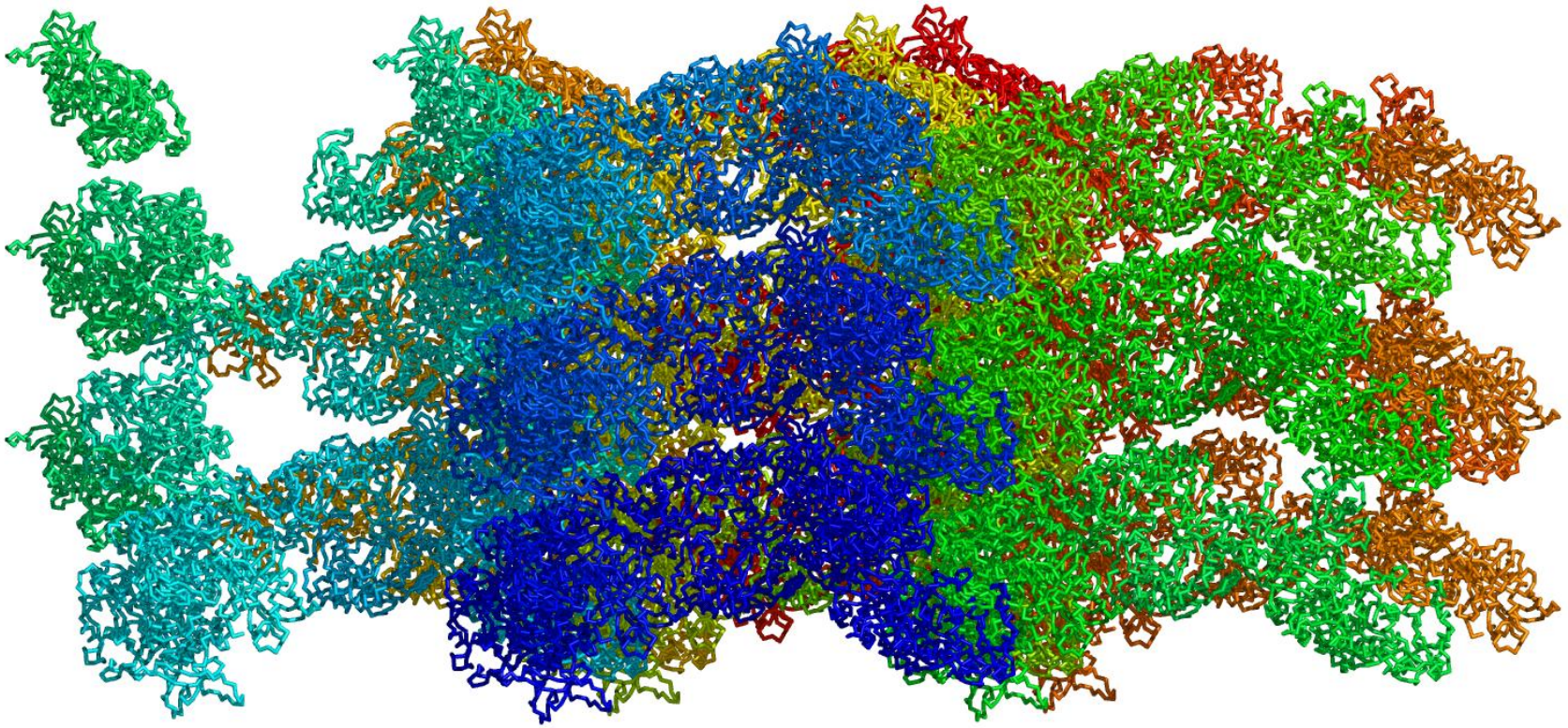


Zinc

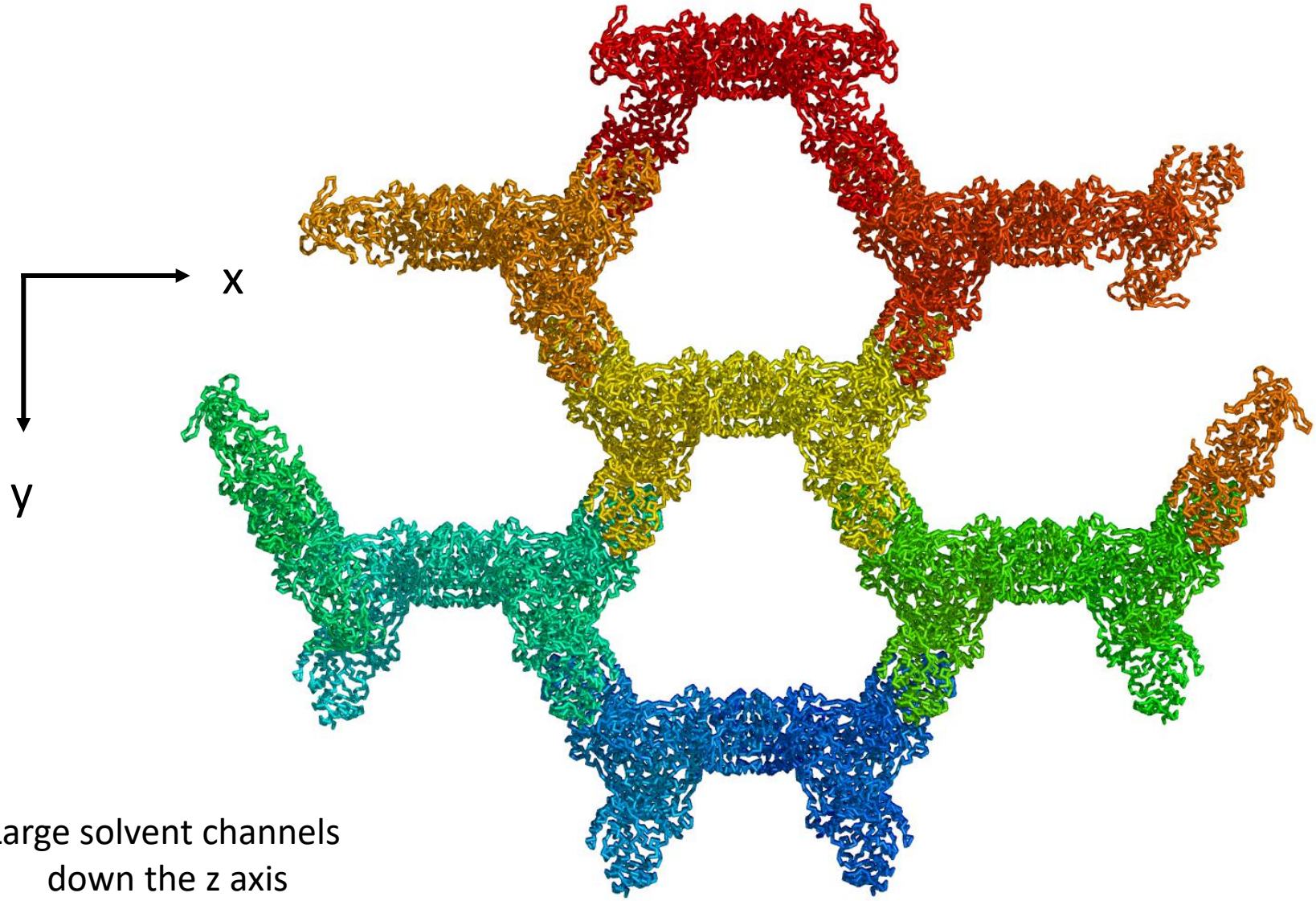
Structure solved (with help of the zinc signal) and refined with Phenix.

Overall R and R_{free} are 15.9 and 21.1% respectively.

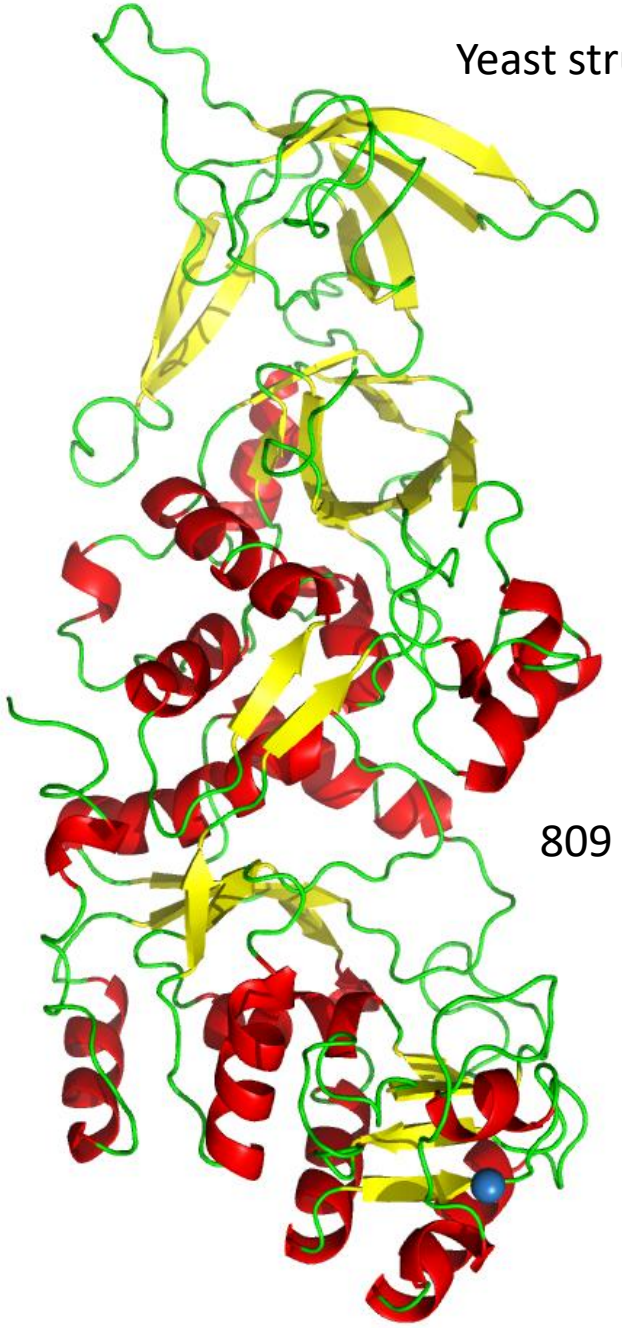




Tight packing in z and y



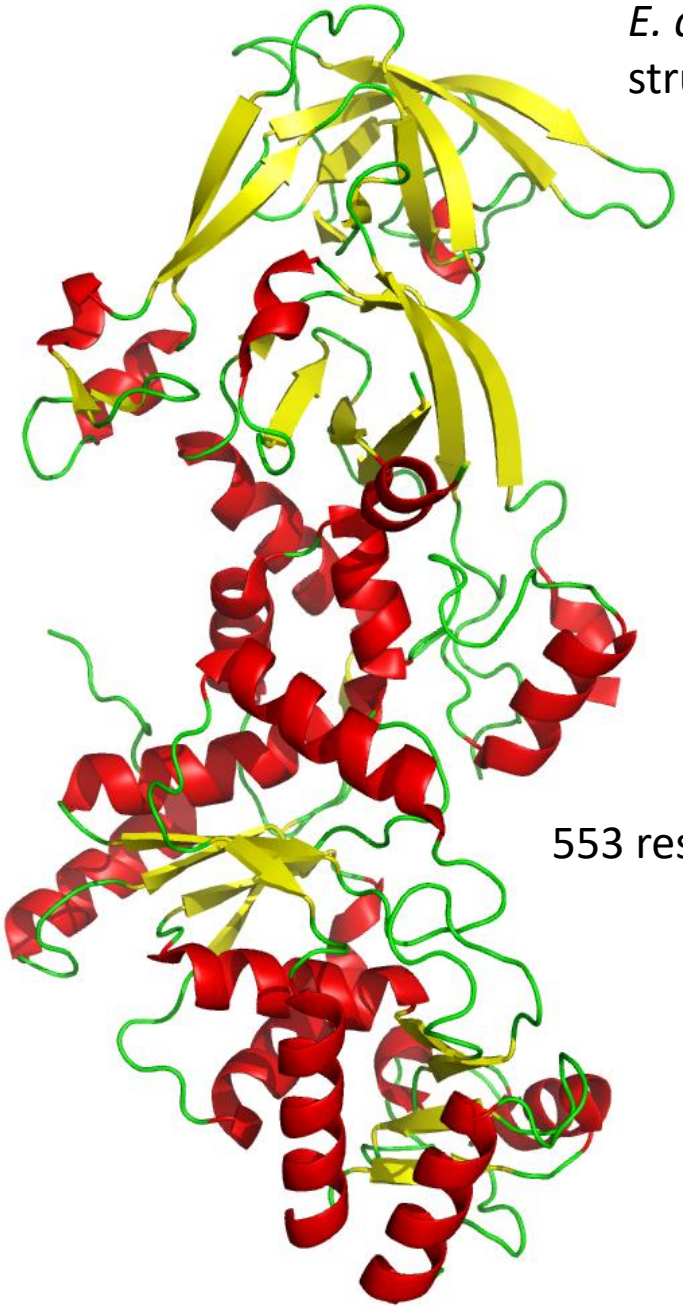
Large solvent channels
down the z axis



Yeast structure

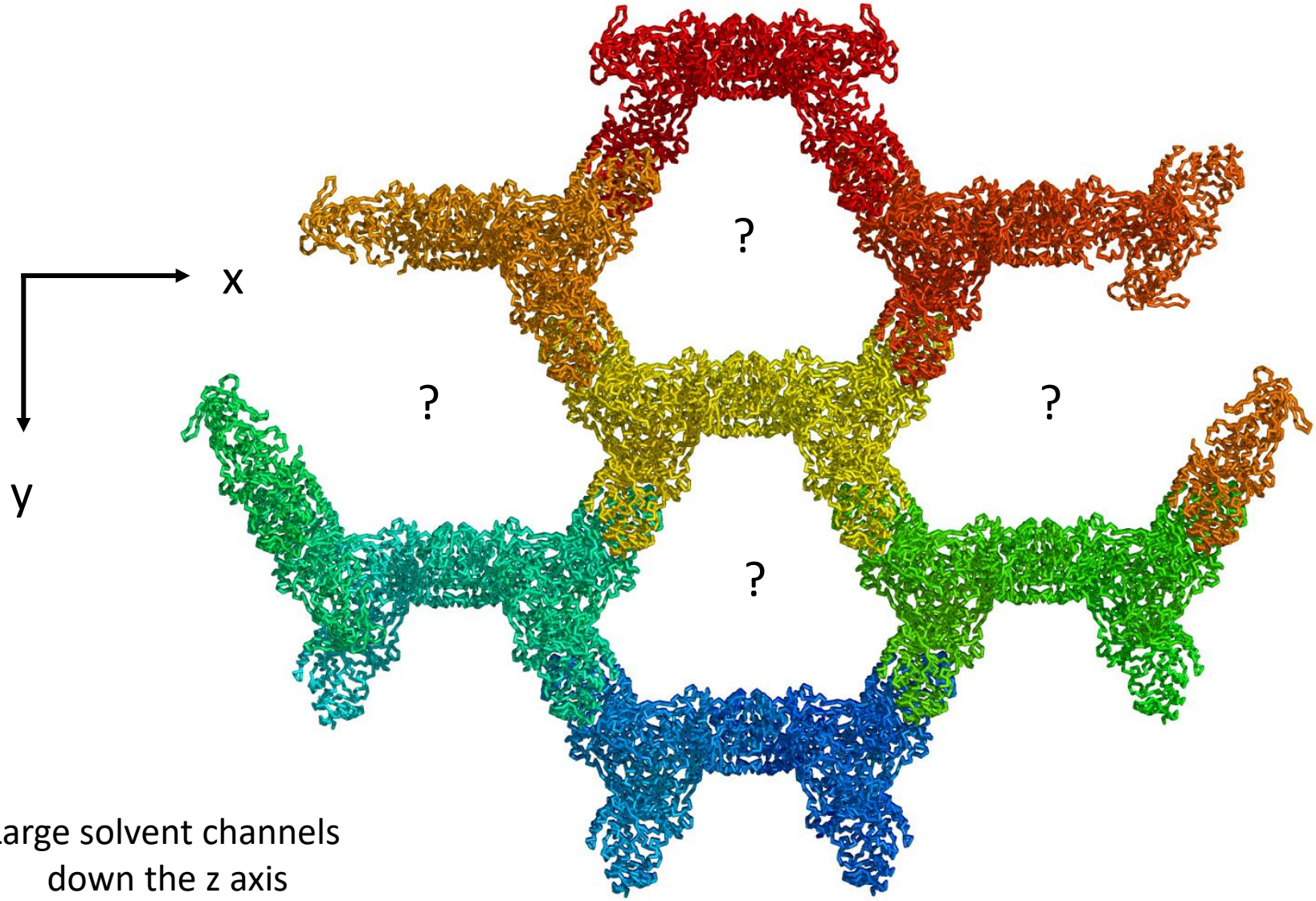
?

809 residues



E. coli. structure

553 residues

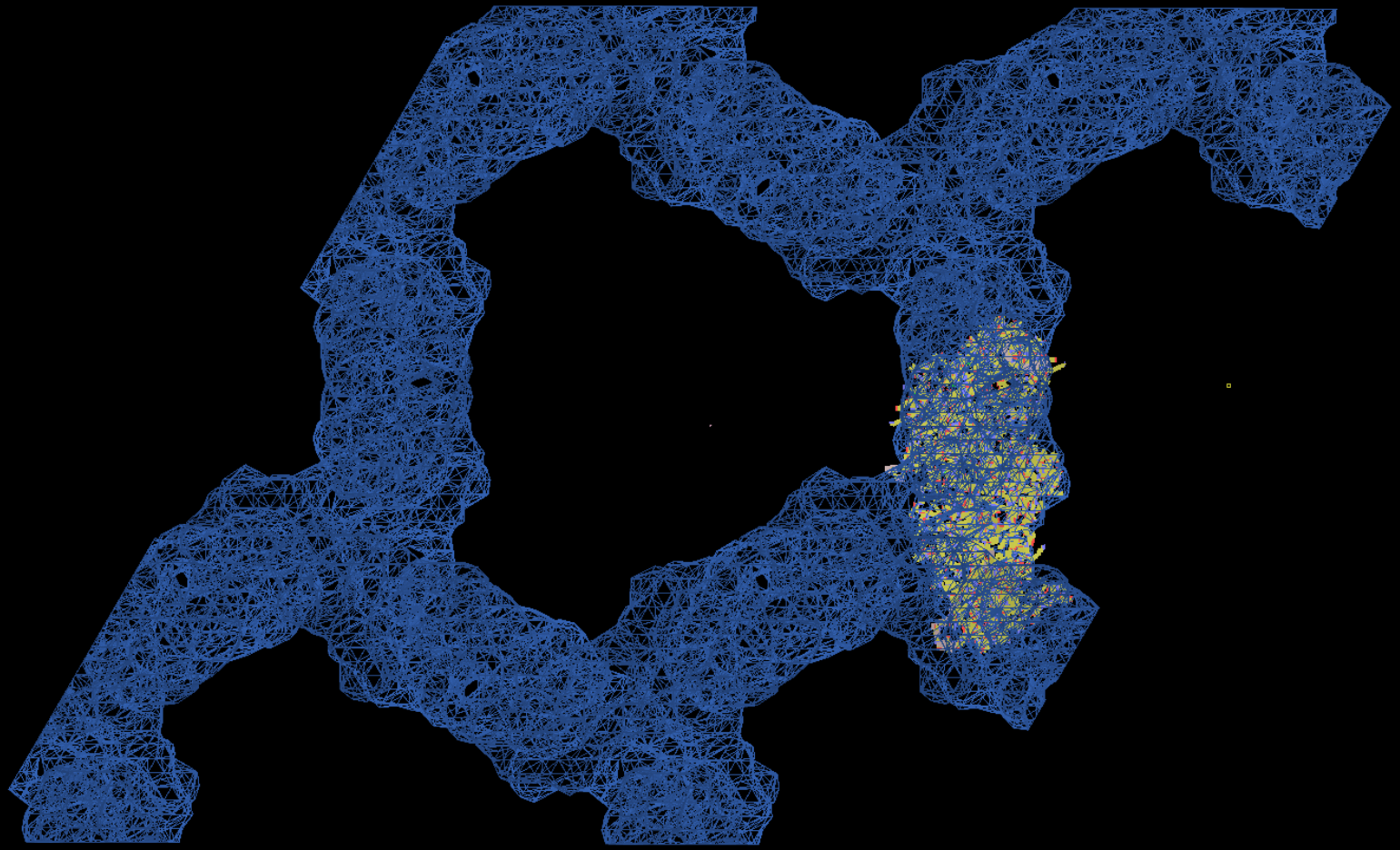


Large solvent channels
down the z axis

Missing residues

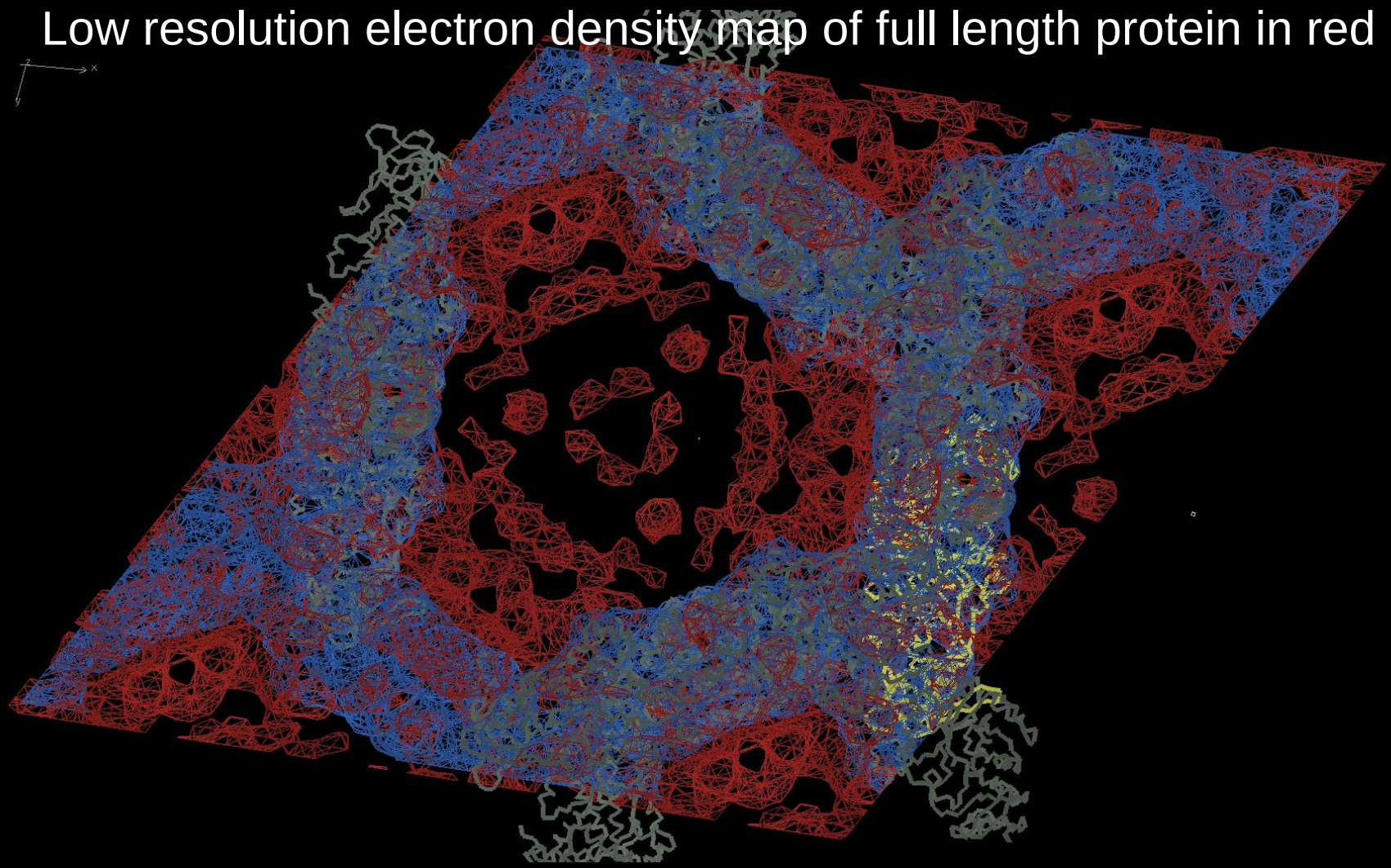
- There were 216 missing residues from the structure, 95% of the N-terminal domain.
- Where they in the mix to start with?.
- SDS PAGE gel on the remaining crystals indicated that the full length protein was present.
- For a more concrete answer the protein was re-expressed with a His tag attached to the N-terminal domain.
  - It was purified with a nickel affinity column.
  - It was crystallized and the structure solved, again with missing residues.
  - A western blot on the dissolved crystals confirmed the presence of the N-terminal domain His tag.
  - No protein degradation had taken place during crystallization.
- For the re-expressed protein the full N-terminal domain was present in the protein but not seen in the crystallographic structure.

Protein with N-terminal arm cleaved

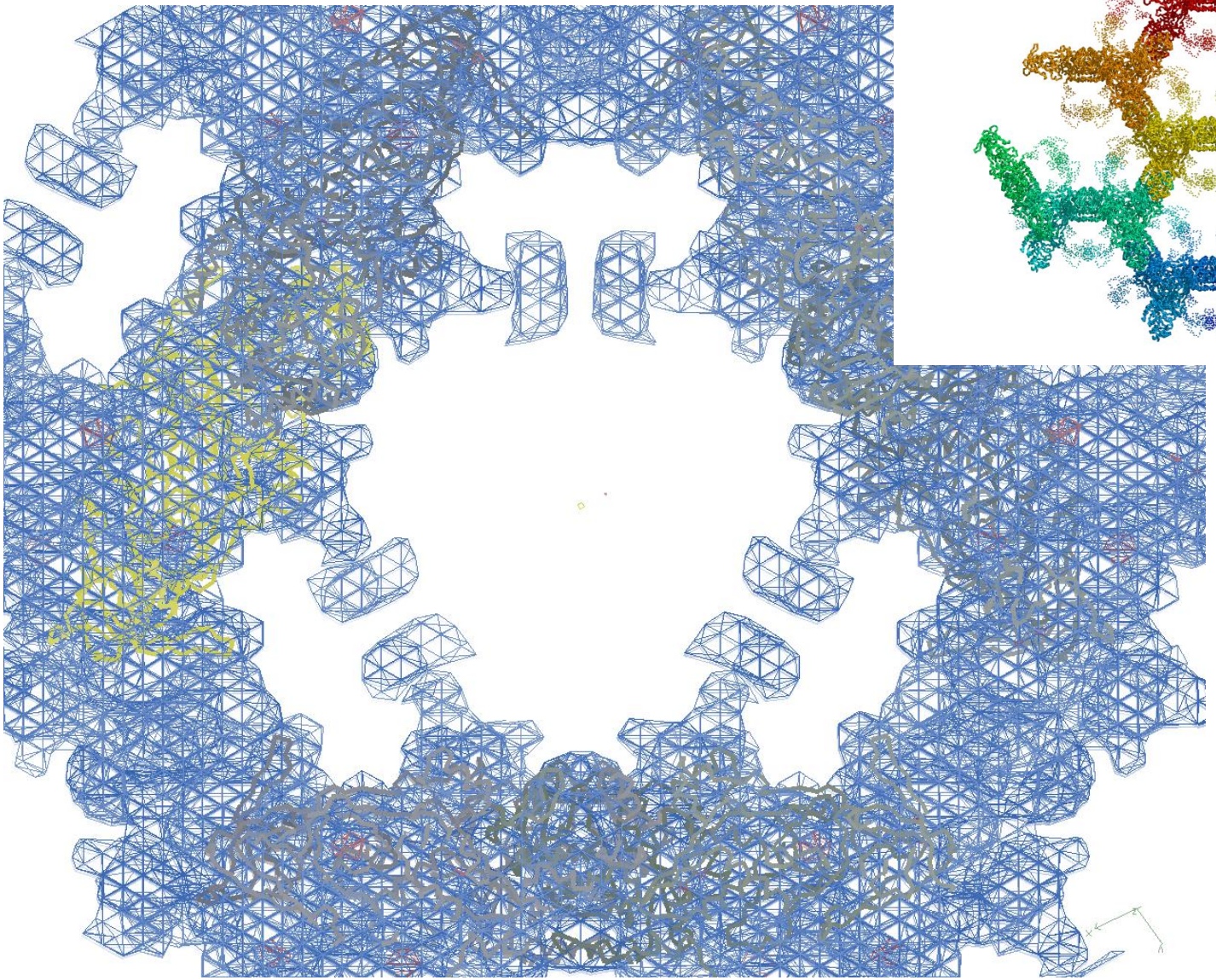


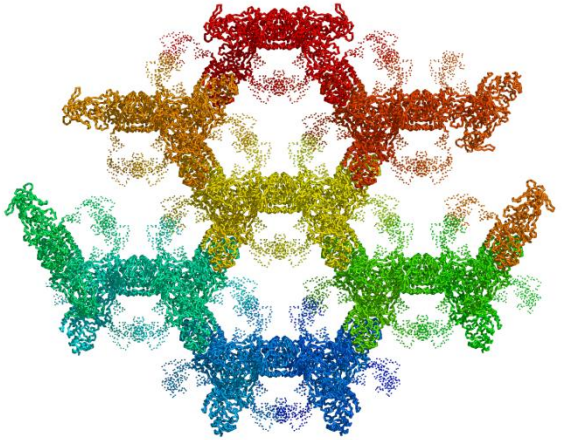
Crystallized, data truncated to 20A (data to 78A still plenty of reflections due to geometry and wavelength used purposely used for data collection)

Low resolution electron density map of full length protein in red

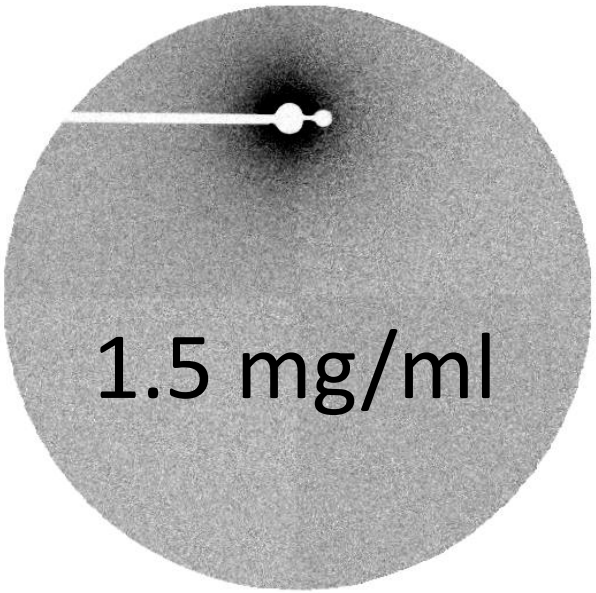


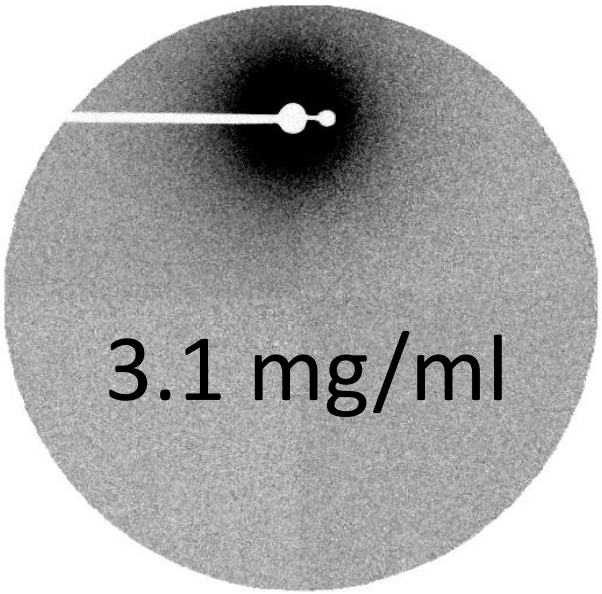
Data truncated to 20A (data to 78A still plenty of reflections due to geometry and wavelength used purposely used for data collection)

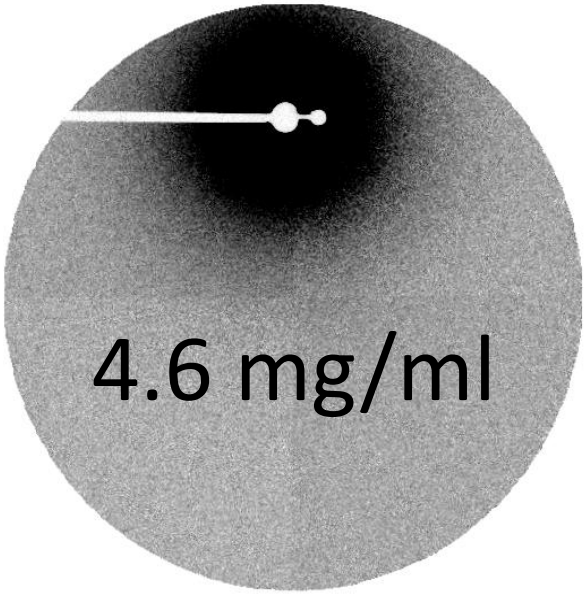


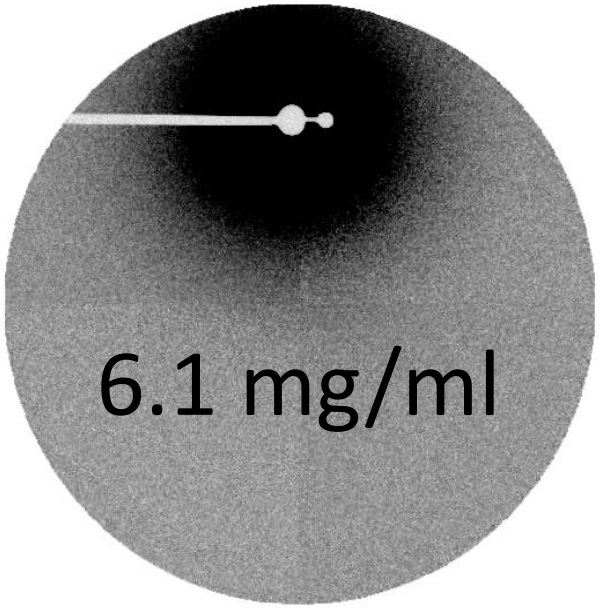


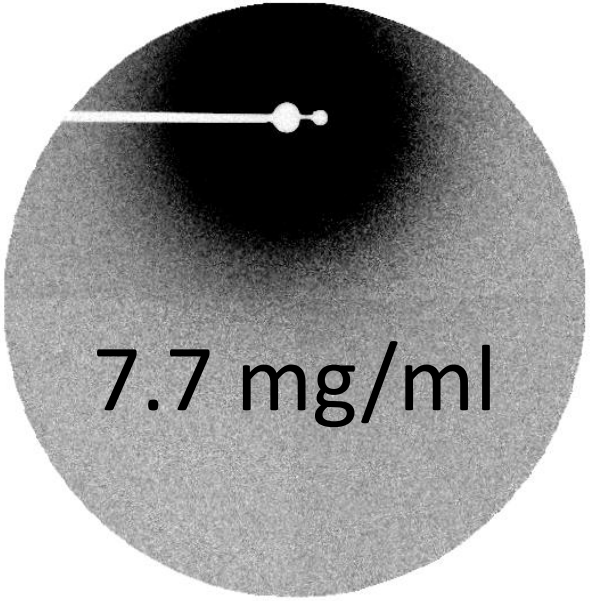
Low resolution truncation (15 Å) of the single crystal data, 1σ, real but not traceable?

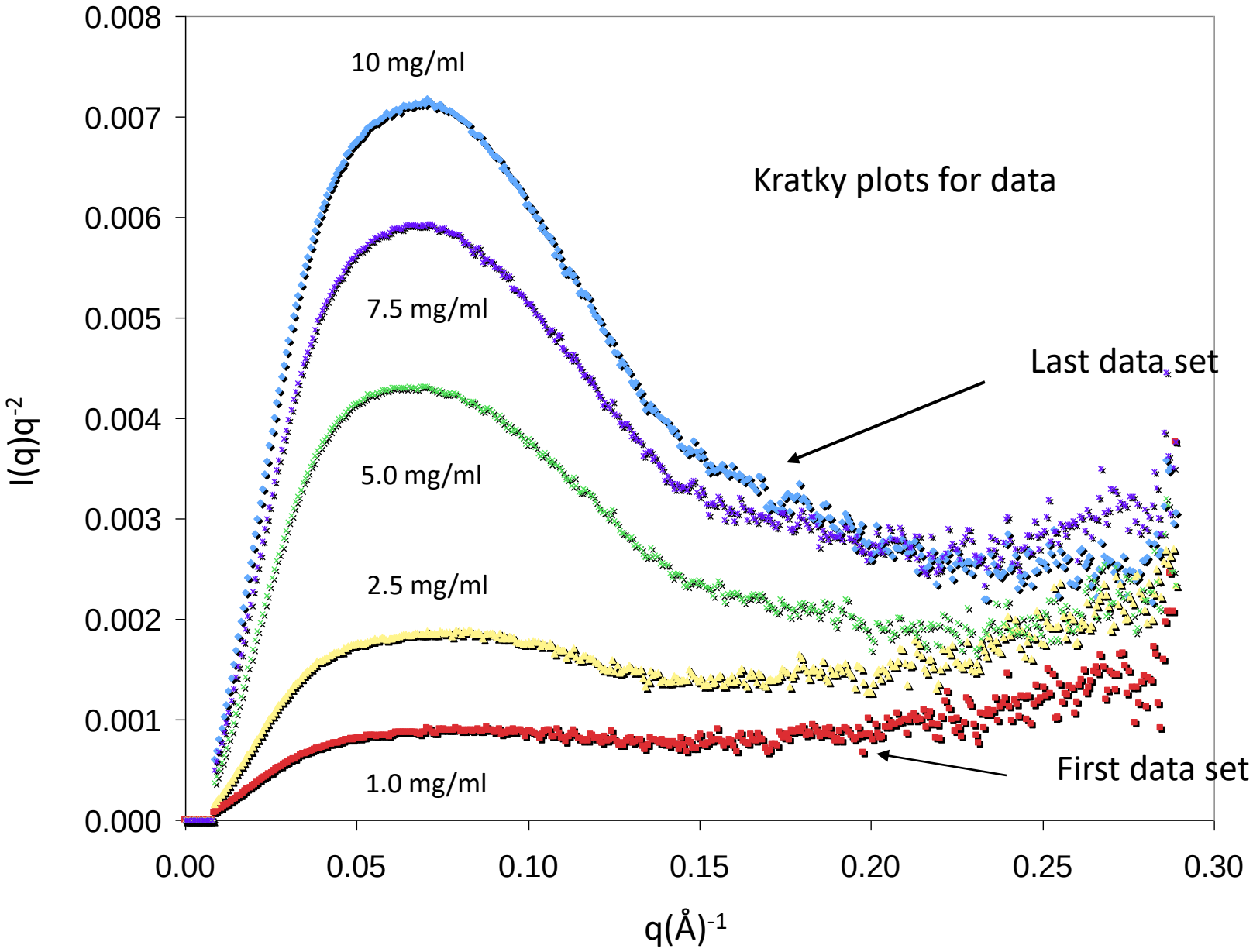

1.5 mg/ml

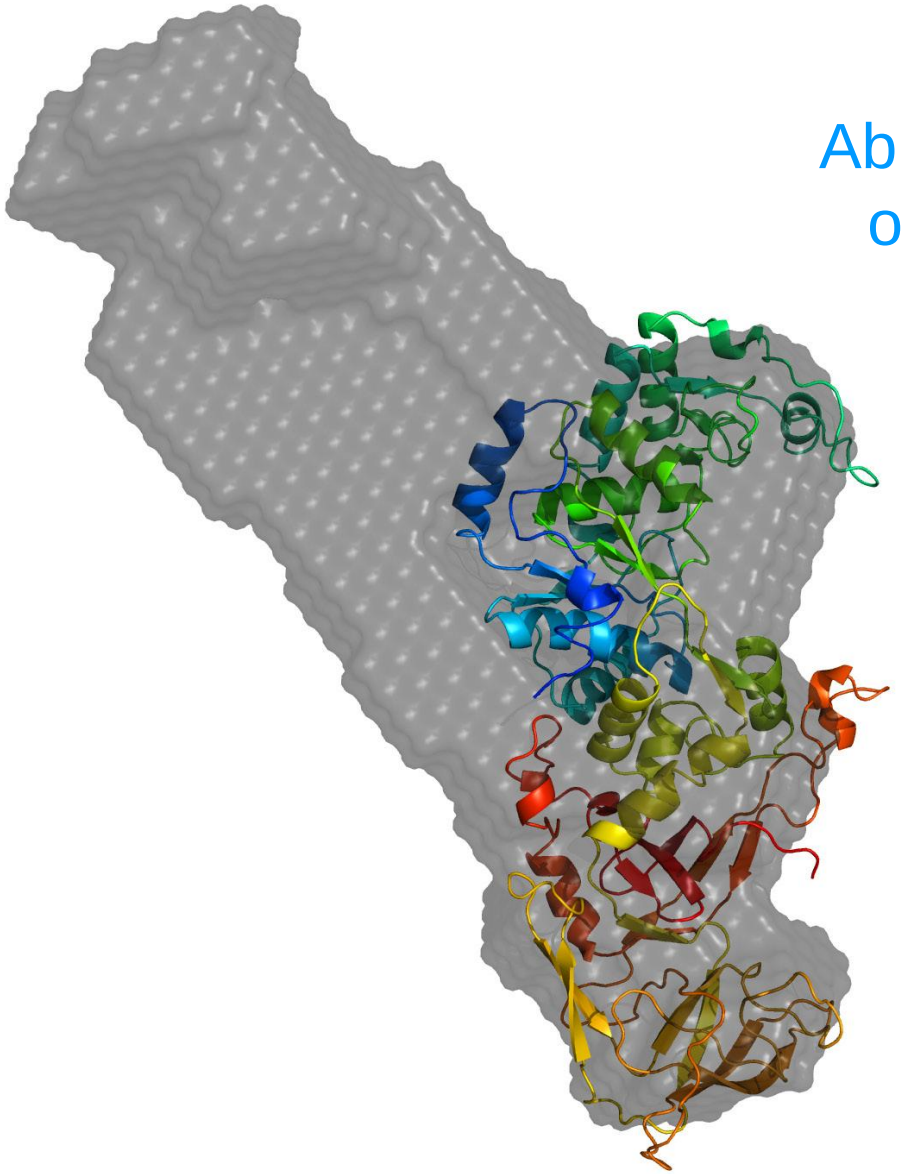

3.1 mg/ml


4.6 mg/ml


6.1 mg/ml

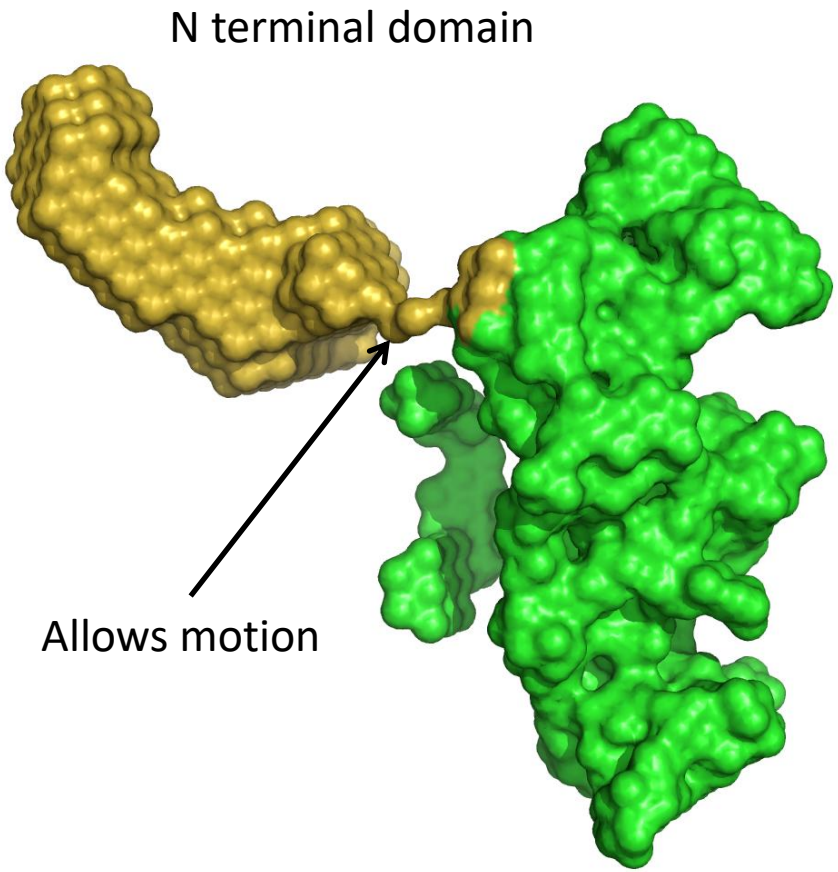

7.7 mg/ml


0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
I(q)q-2
10 mg/ml
7.5 mg/ml
5.0 mg/ml
2.5 mg/ml
1.0 mg/ml
Kratky plots for data
Last data set
First data set
0.00
0.05
0.10
0.15
0.20
0.25
0.30
q(Å)-1

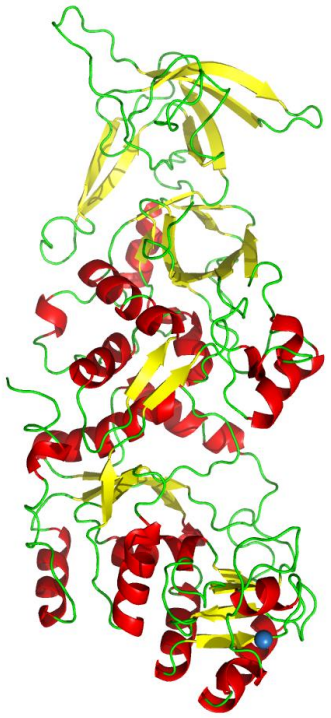


Ab initio structure overlaid on the crystallographic structure

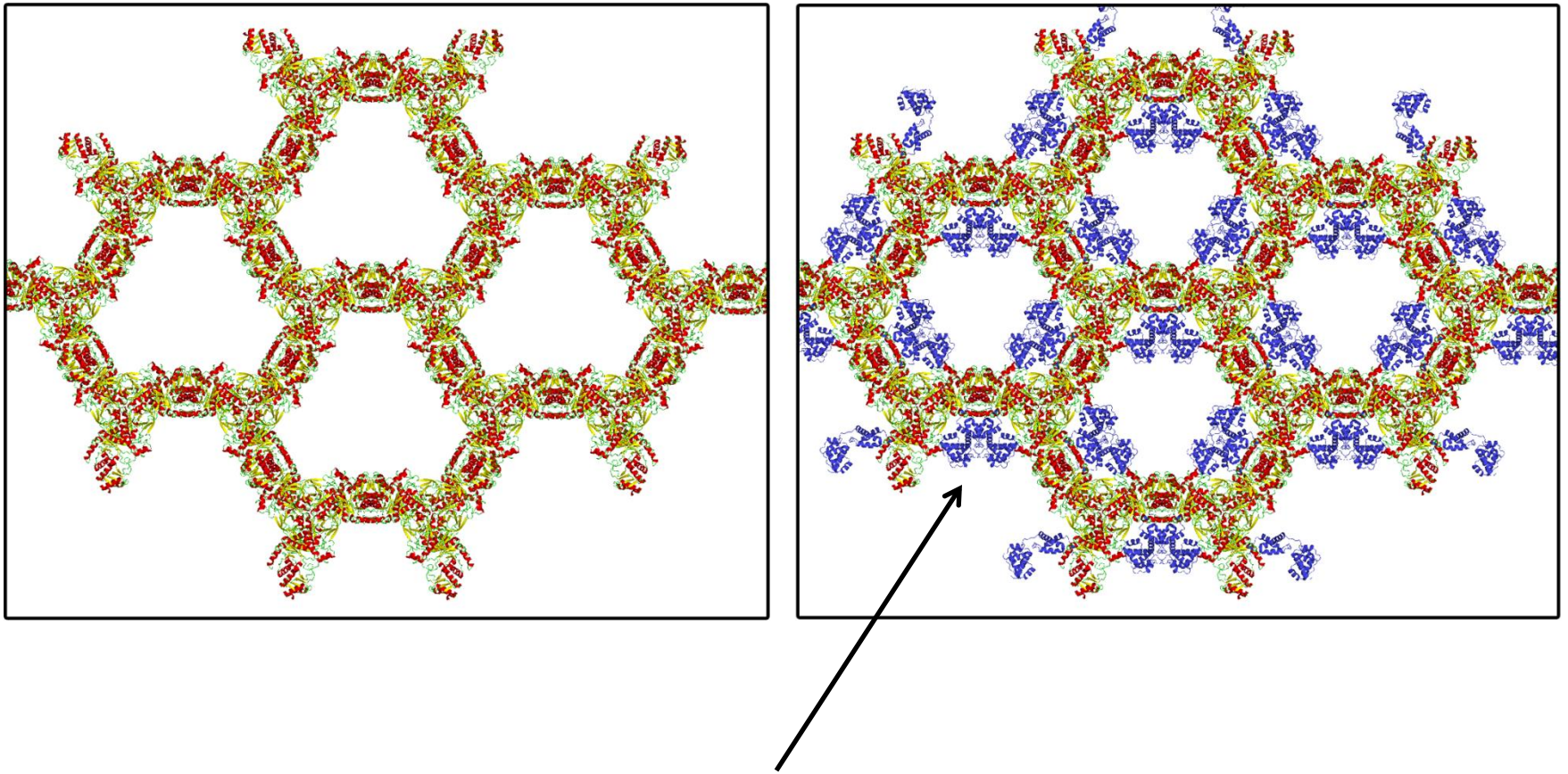
Envelope reconstruction using the crystallographic structure



C terminal domain

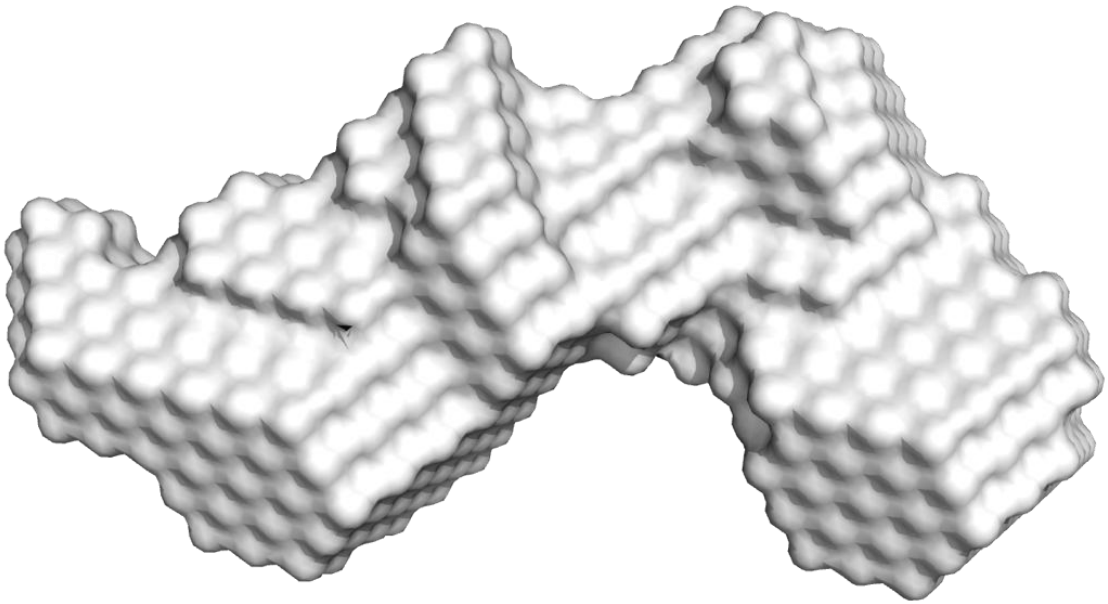


The crystal structure (which shows only the C-domain)



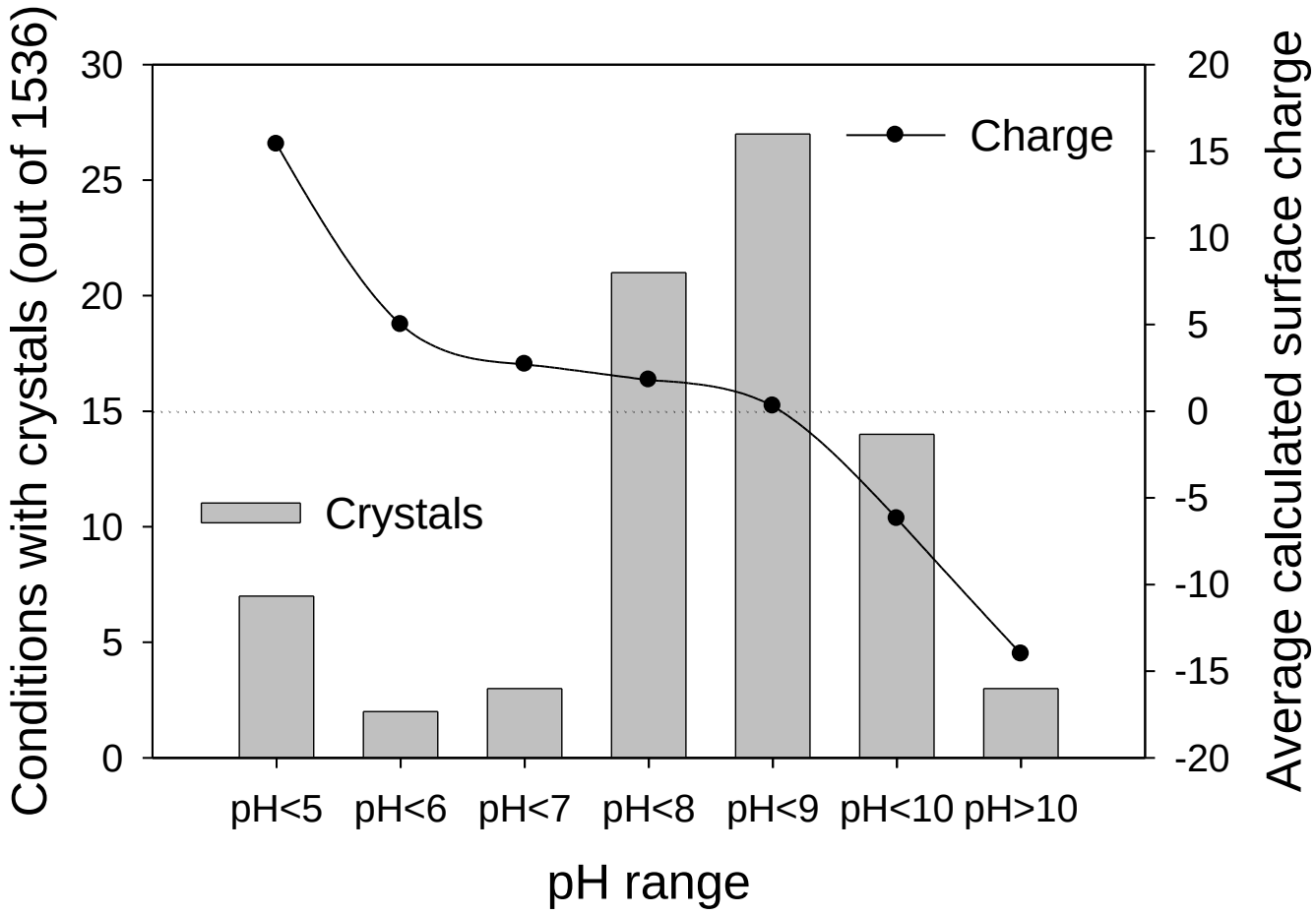
The N-terminal ‘arm’ is completely compatible with the crystal structure

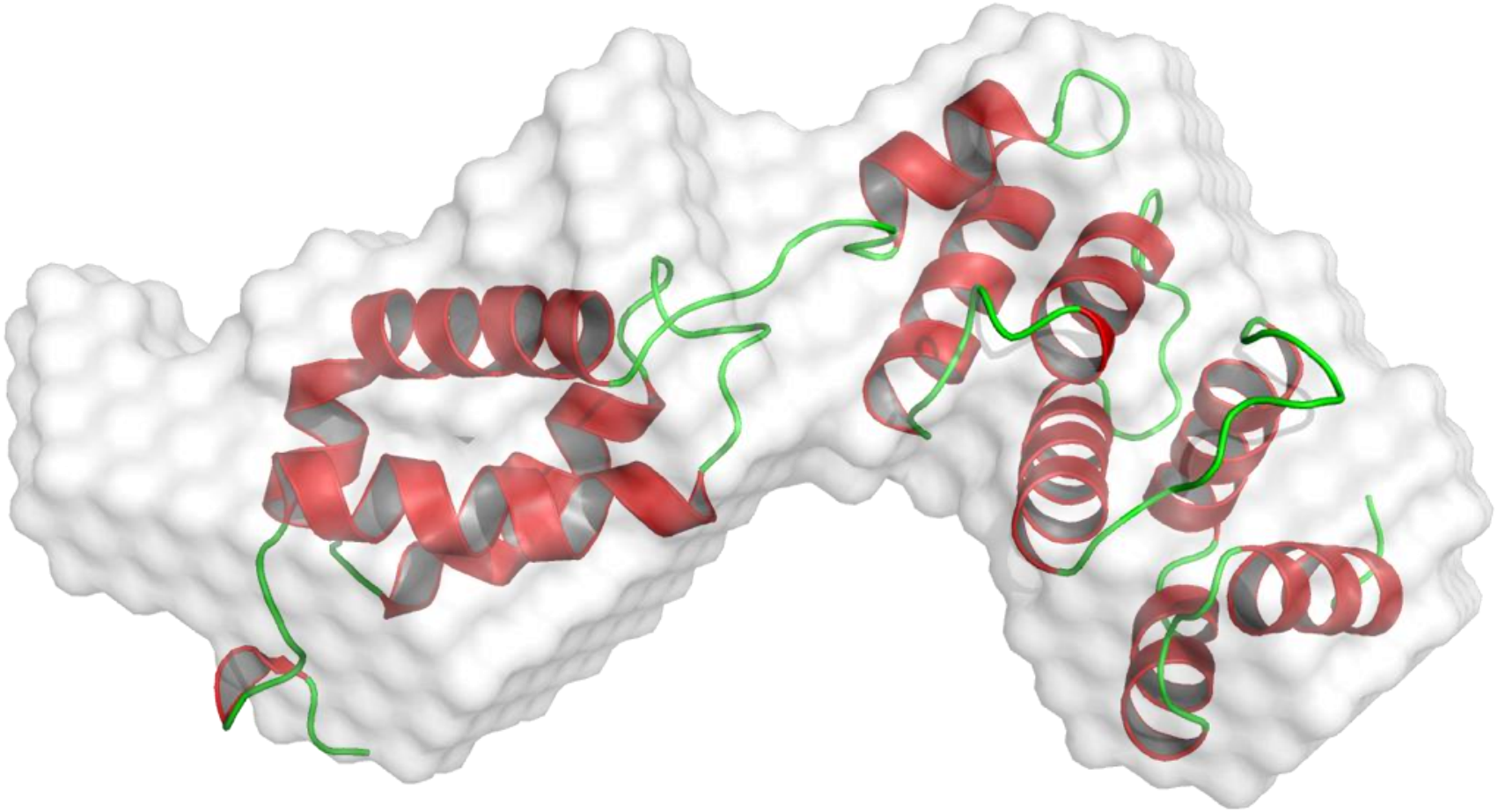
Envelope reconstruction of the N-terminal domain

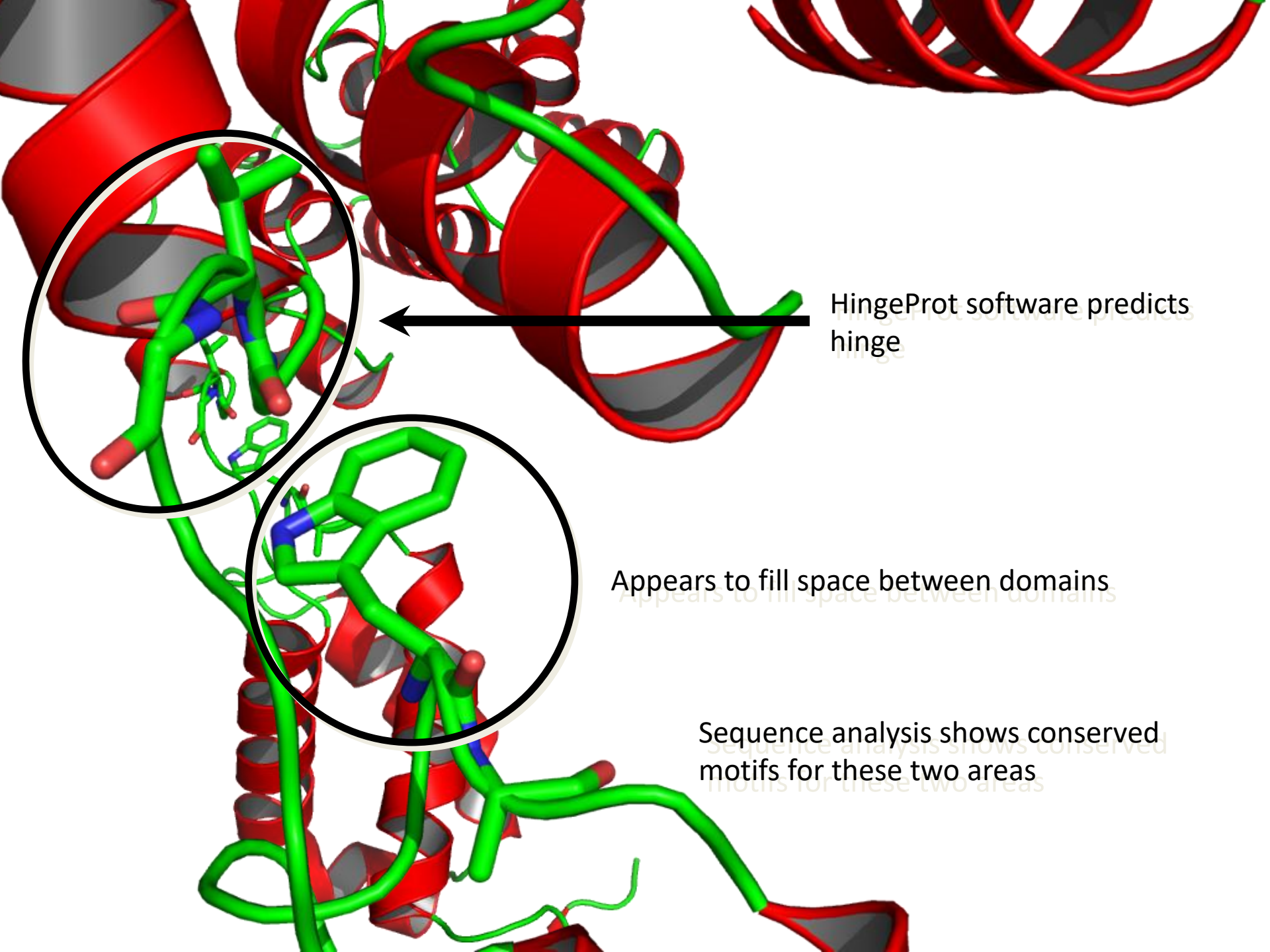


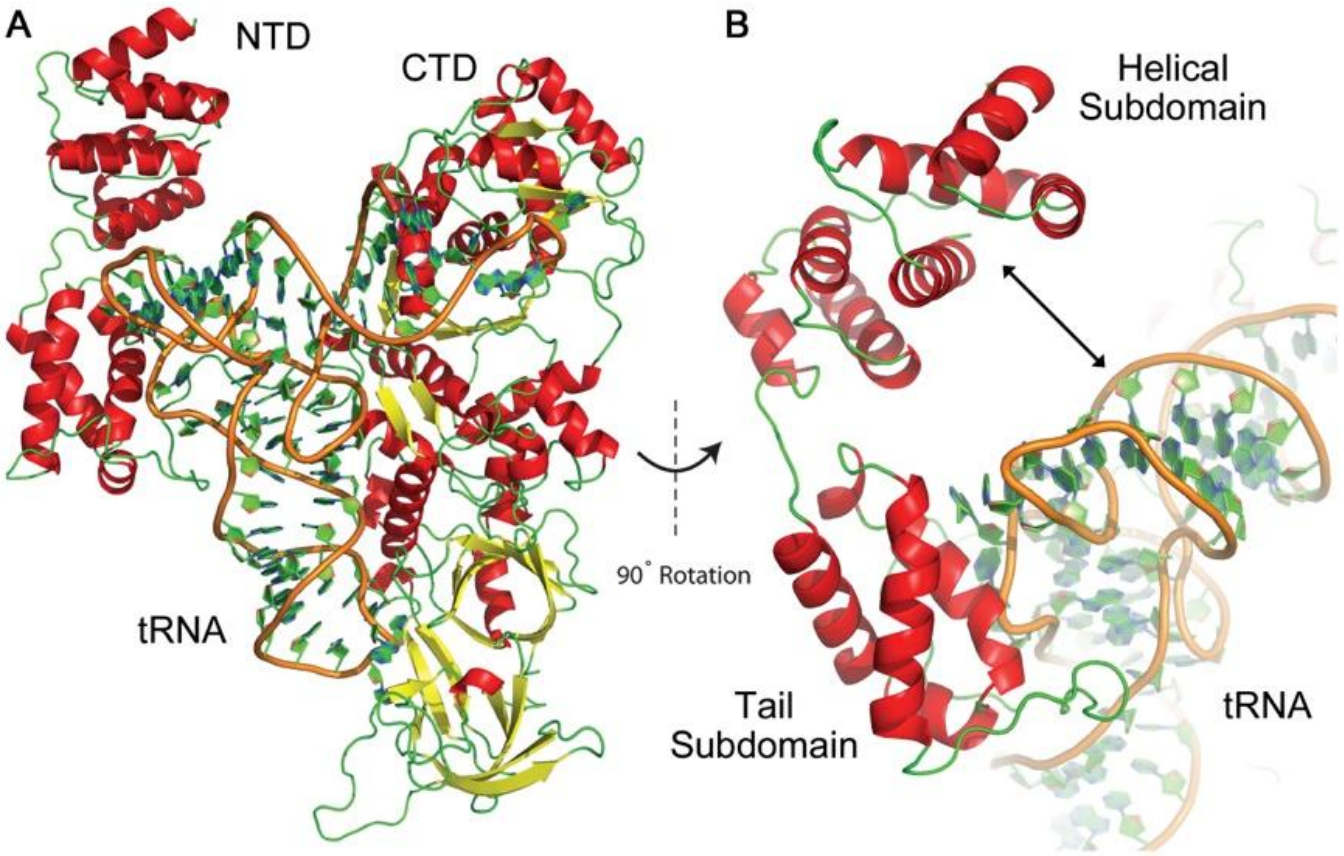
Back to crystallography

Crystallization trials of the N-terminal domain



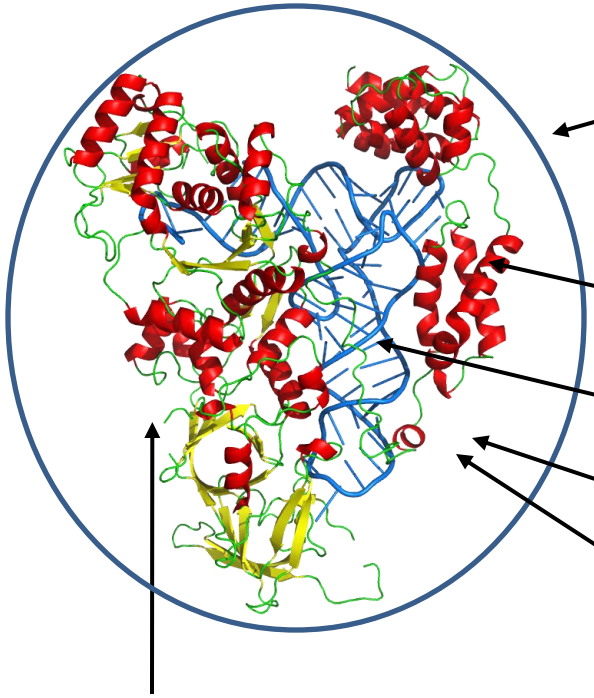



HingeProt software predicts hinge
Appears to fill space between domains
Sequence analysis shows conserved motifs for these two areas



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.

Eukaryotic Gln tRNA synthetase



SAXS data indicating a larger but well folded system in solution

A Sherlock analysis indicated a preferential pH

The truncated terminal was crystallized

It was extracted directly from the screening plate and X-rayed to give the structure.

tRNA was docked in

SAXS aided by sequence analysis identified a flexible region

Homology modeling (FREAD) gave the flexible region

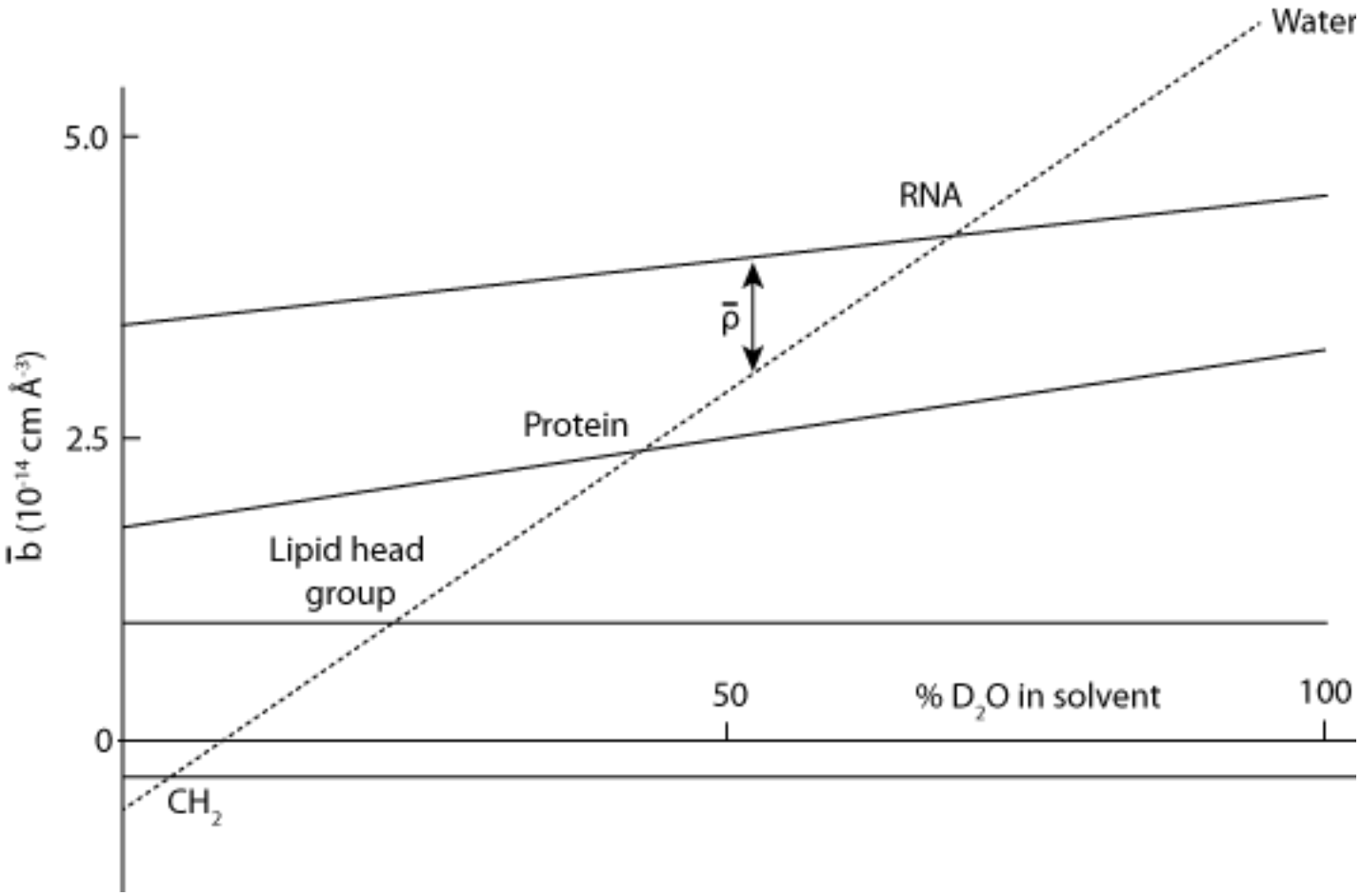
Crystallized the C-terminal in the standard screen, conditions chosen that were already known to be good cryo-conditions.

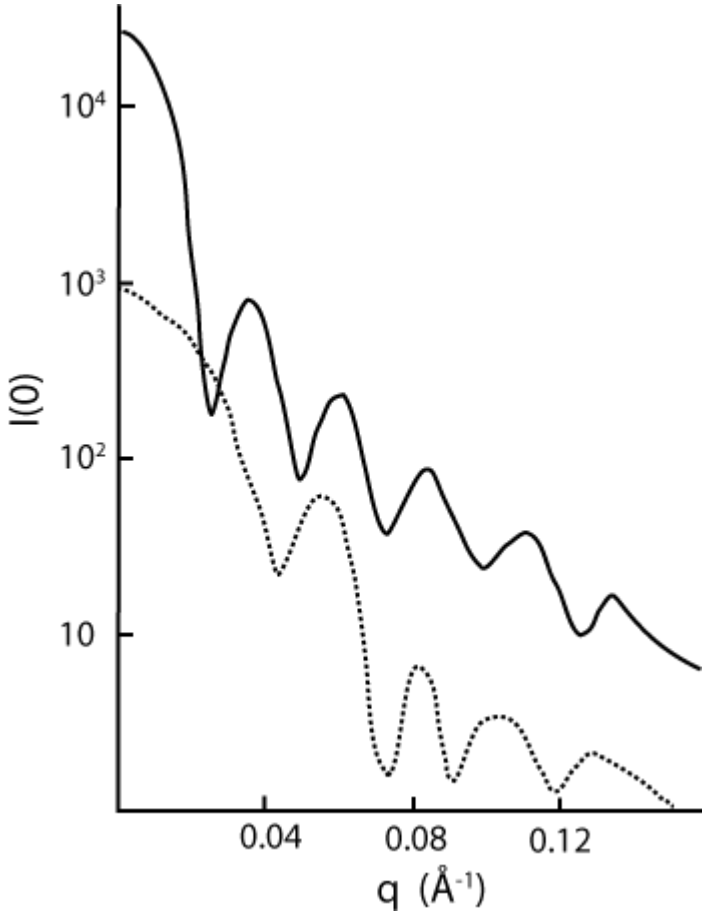
A combination of crystallography, SAXS, homology modeling and computational modeling was used to give the complete structure and tested by biochemical analysis.

Small Angle Scattering with Neutrons


NORTH

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

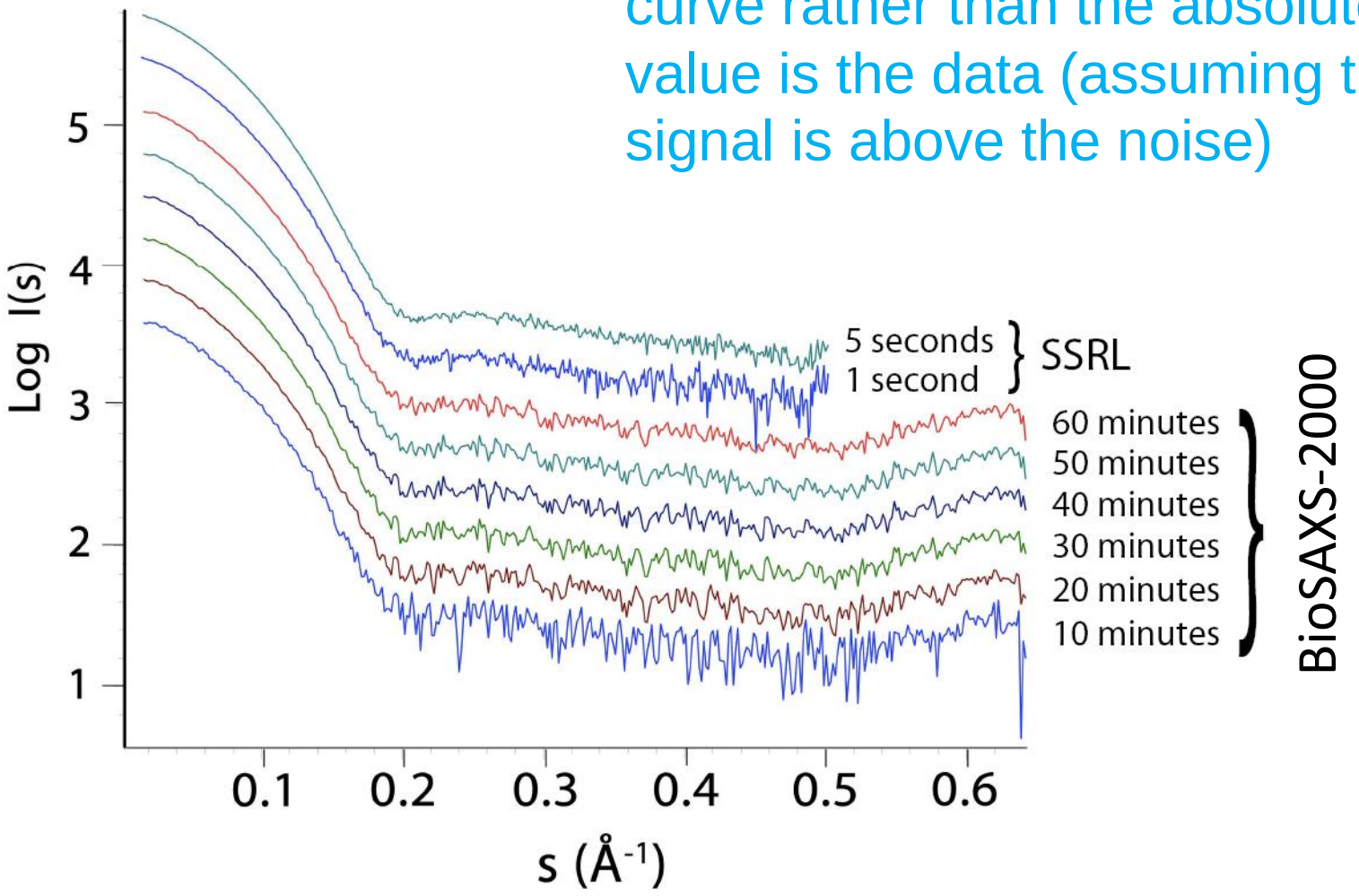




Scattering curve from Southern Bean Mottle virus in solutions of different D_2O content. The continuous line with 69.5% D2O and scattering mostly due to the protein shell and the dashed line with solvent content 42% D_2O and the scattering mostly by the nuclein acid (RNA). The subsidiary maxima are shifted 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

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



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.

Acknowledgements

Thomas Grant, Joseph Luft, Hiro Tsuruta, Anne Martel, Lester Carter, Tsutomu Matsui, the NESG, Eric Phizicky, Beth Grayhack, and Thomas Weiss

Support and Funding: NIH, DoD and NSF

Thank you and questions?

esnell@hwi.buffalo.edu