Small Angle X-ray Scattering (SAXS) as a Complementary Structural Biology Technique: Perils, Pitfalls and Potential.

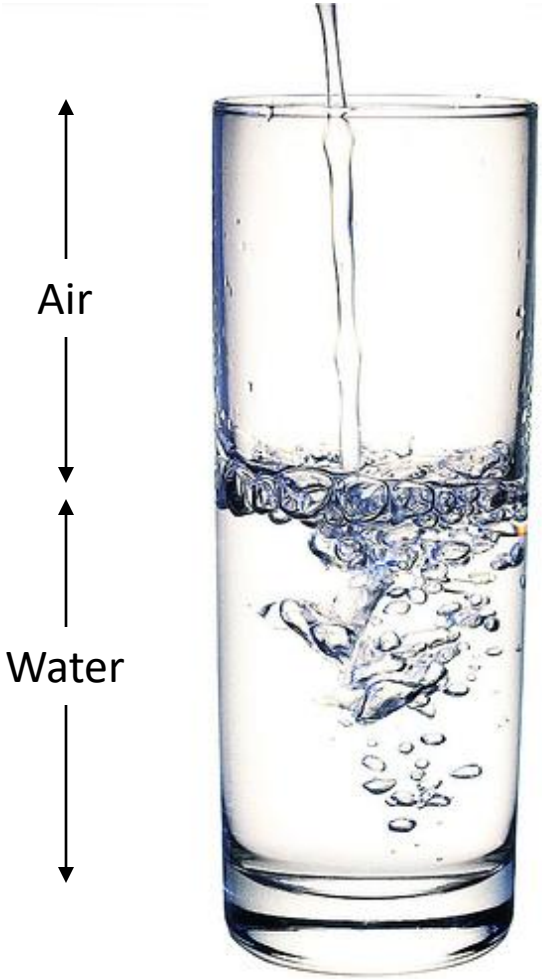
Edward H. Snell and many others

Hauptman Woodward Medical Research Institute, Buffalo, NY 14203, USA.

Outline

- Crystallization
- SAXS theory
- SAXS practice
- and Ptifalls
- Examples
- Complementary application – Perils, pitfalls and potential.

Pessimists, Optimists, and Crystallographers



Consider a glass of water

Pessimist
(the glass is half empty)

Optimist
(the glass is half full)

Crystallographer
(the glass is completely full)

Fantasy

The crystallization screening laboratory at the Hauptman-Woodward Medical Research Institute

Since February of 2000 the High Throughput Search (HTS) laboratory has been screening potential crystallization conditions for the general biomedical community and two Protein Structure Initiative large-scale structure production centers (NESG, Montelione, PI; SGPP/MSGPP, Hol, PI) and one PSI specialized PSI-2 center (CHTSB, DeTitta, PI).

The HTS lab screens samples against an incomplete factorial screen of two categories of crystallizing agents:

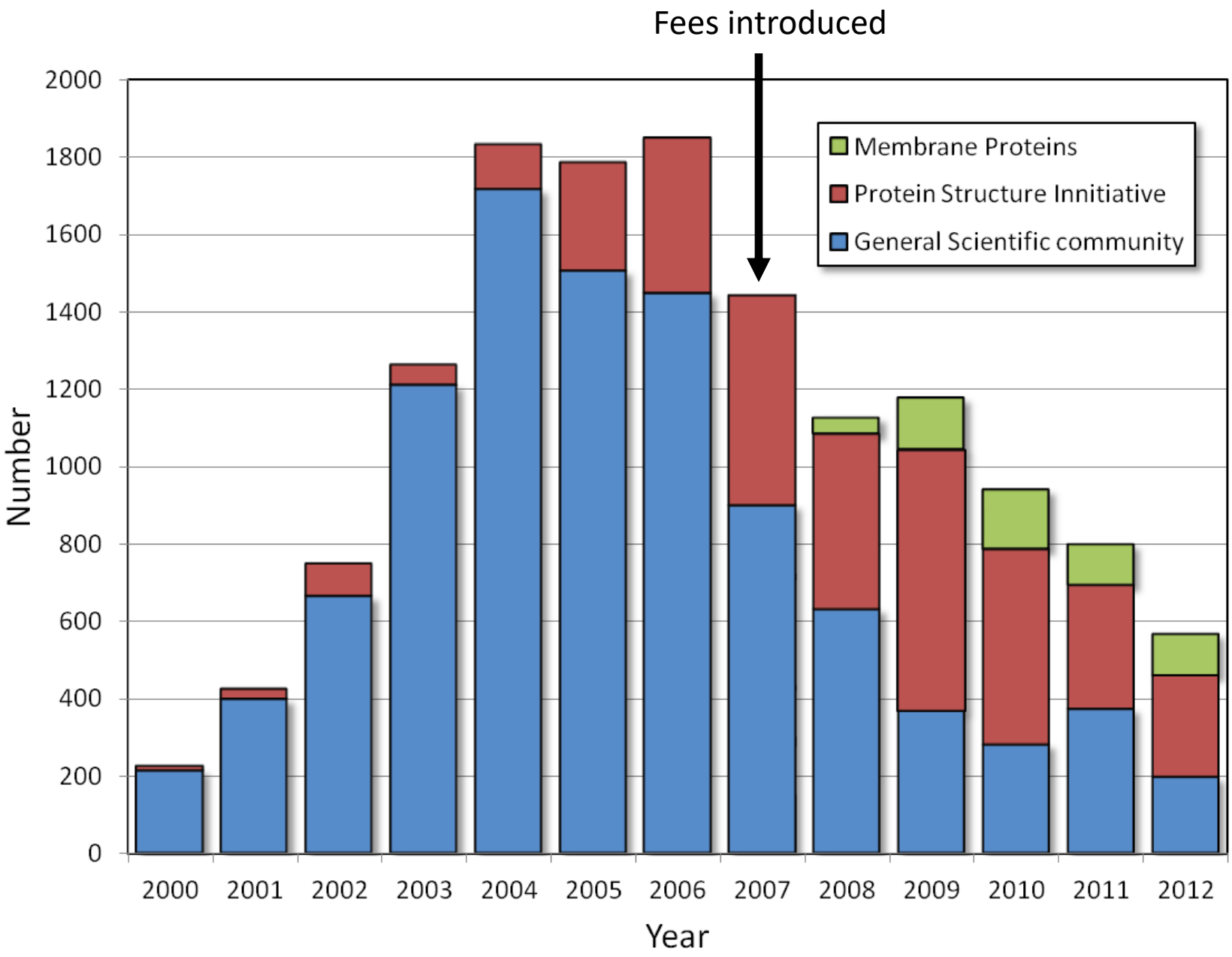
1. buffered (4<pH< 10), highly concentrated salts (35 salts total, sampling 18 different cations and 20 anions) – 229 conditions.
2. PEG/salt/buffer solutions (eight buffers (4<pH< 10), six molecular weight PEGs at three concentrations, and 35 salts at fixed 200 mM concentration) – 721 conditions.

Added to this is a screen of some 586 conditions encompassing screens commercially available from Hampton Research.

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

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

The HTSlab has investigated the crystallization properties of over 13,900 individual proteins archiving over 115,000,000 images of crystallization experiments.


Fees introduced
Membrane Proteins
Protein Structure Innitiative
General Scientific community
Number
Year
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012

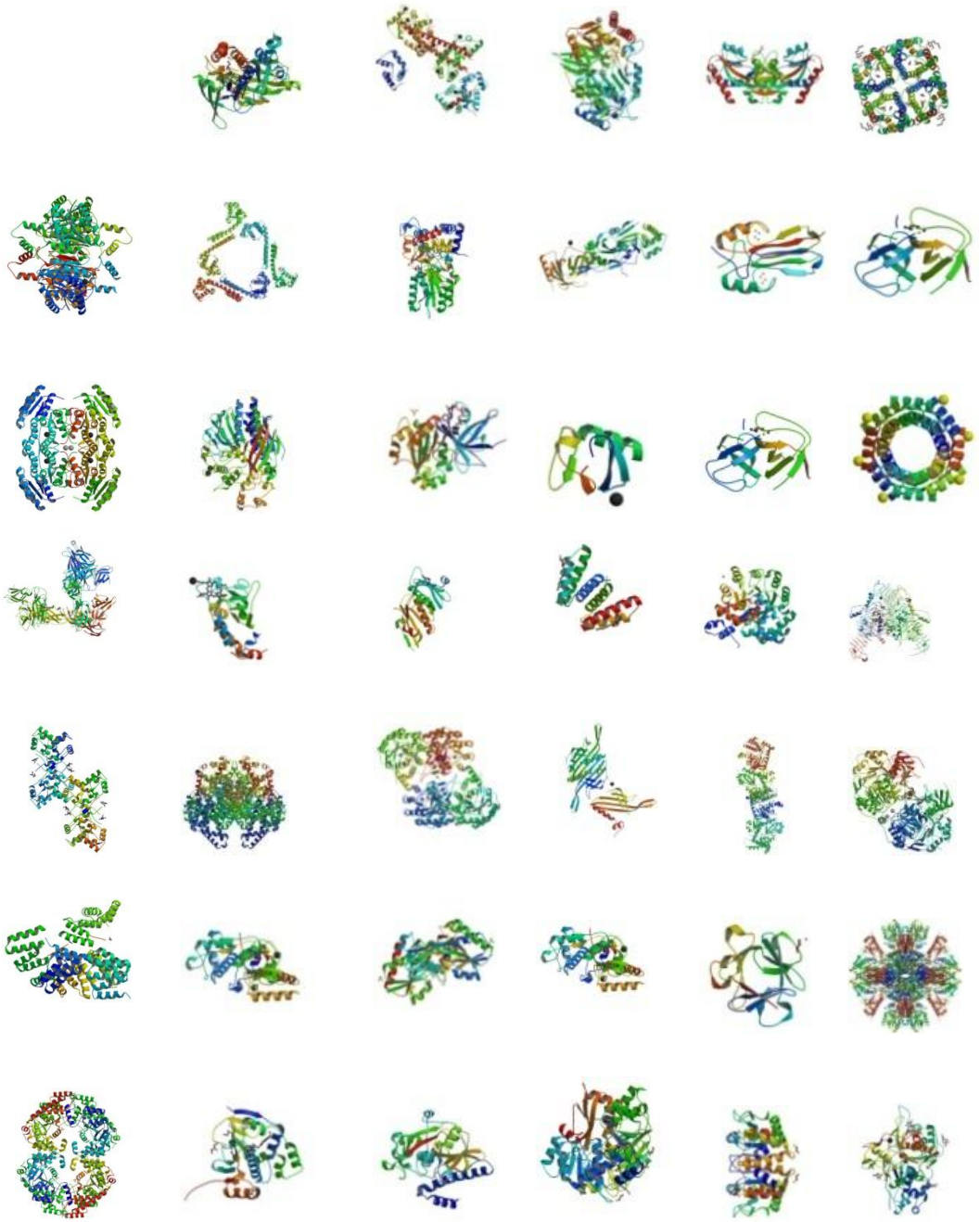
Born in Buffalo

Over 1,000 general biomedical laboratories world wide use the crystallization screening service with approximately 2,000 unique investigators.

Investigators are sent photographs of the results, analyze these images and perform their own optimization of any hits observed.

No information is released on targets. Progress is tracked by acknowledgements and citation searches. Currently no other metrics are used to measure success rates for the general biomedical community.

These images represent examples of structures from initial hits in the HTS laboratory.



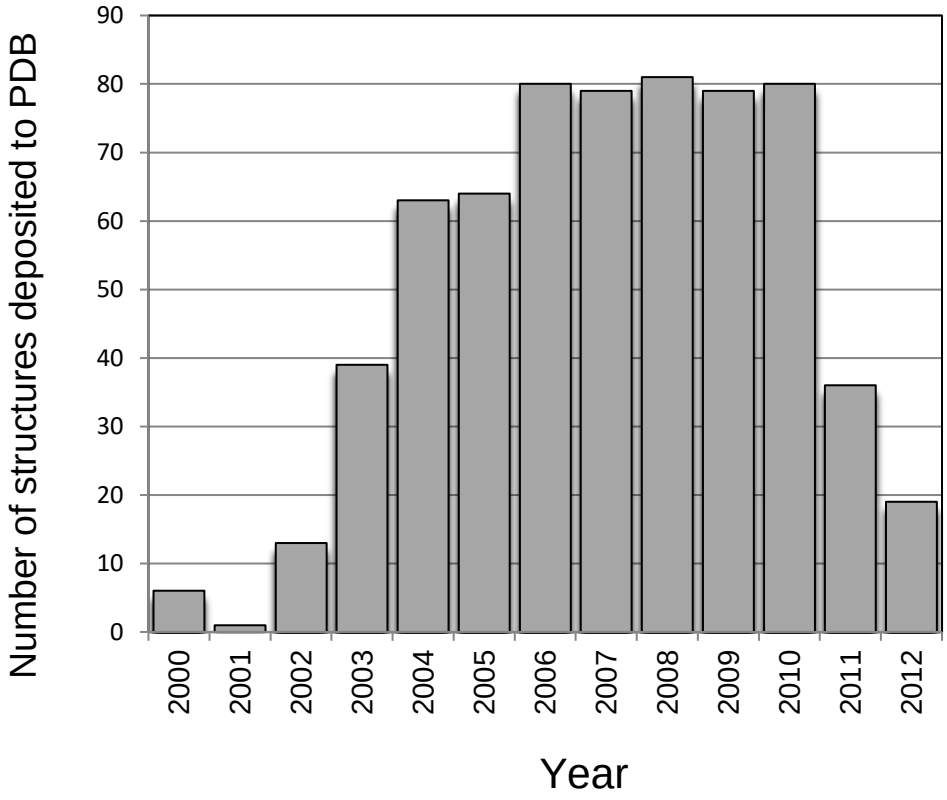
Where success is tracked.

For our Protein Structure Initiative partners both success and failure is tracked. In the case of NESG our initial screening hits enable on average 80 structures per year to be deposited to the PDB.

The graph demonstrates the ramp up of operations with maximum success reached from 2006 onward.

Our success rate from protein in the door to a crystallization hit leading to a PDB deposition is **22%**.

The NESG samples represent a special case in that they are well characterized beforehand – size exclusion chromatography, mass spec analysis and dynamic light scattering studies.



In 2011 we switched to PSI Biology – More difficult targets

Why Small Angle X-ray Scattering (SAXS)?

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

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

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

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

When crystallization screening is conducted many outcomes are observed but typically the only trial recorded in the literature is the condition that yielded the crystal(s) used for subsequent diffraction studies. The initial hit that was optimized and the results of all the other trials are lost. These missing results contain information that would be useful for an improved general understanding of crystallization. This paper provides a report of a crystallization data exchange (XDX) workshop organized by several international large-scale crystallization screening laboratories to discuss how this information may be captured and utilized. A group that administers a significant fraction of the world's crystallization screening results was convened, together with chemical and structural data informaticians and computational scientists who specialize in creating and analysing large disparate data sets. The development of a crystallization ontology for the crystallization community was proposed. This paper (by the attendees of the workshop) provides the thoughts and rationale leading to this conclusion. This is brought to the attention of the wider audience of crystallographers so that they are aware of these early efforts and can contribute to the process going forward.

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

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

There exists a large quantity of soluble purified protein that remains structurally uncharacterized.

Crystallization is hard

Making the protein is easier

Perils and Pitfalls

SAXS is even easier ……

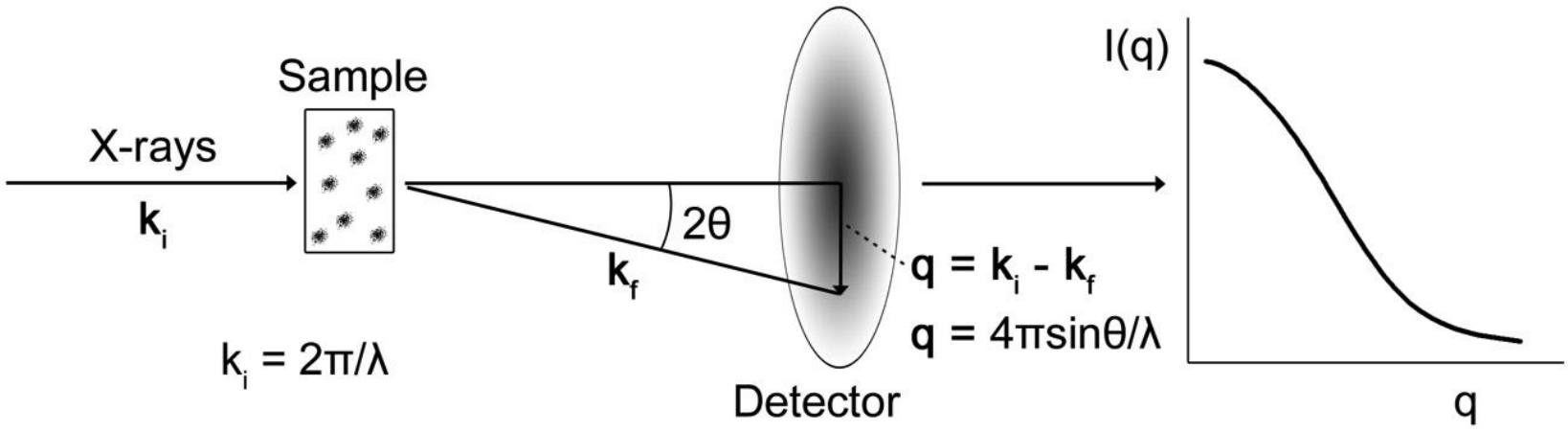
but

History of SAXS

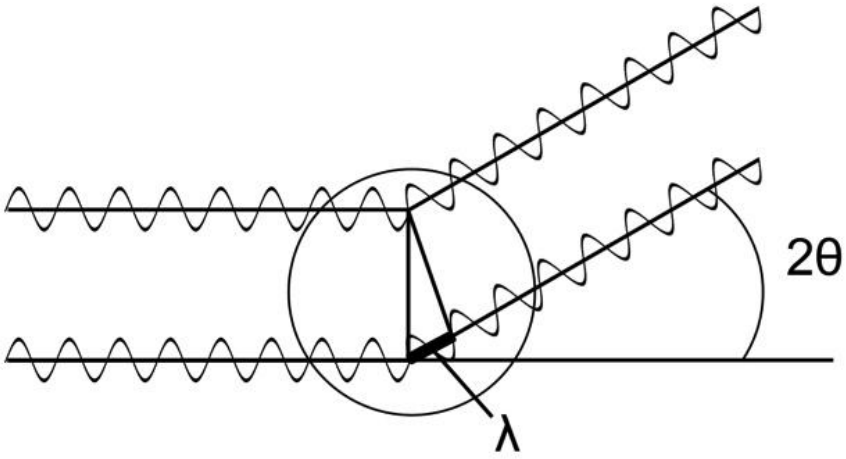
- In 1939 André Guinier found that X-ray scattering at the smallest angles was only present for heterogeneous solutions.
- He that the X-ray intensity was strongest at these angles for fine grains 10 to 100 nm in size and determined a method, to calculate the sizes of the particles from the scattering.
- SAXS began being used on biological macromolecules in the 1960s as a method to gain low-resolution structural information in the absence of crystals .
- The introduction of high-flux neutron sources enabled contrast variation studies using small angle neutron scattering (SANS) of perdeuterated solutions .
- Until the 1990s, only parameters about shape and size could be extracted from SAXS data including radius of gyration and particle volume,.
- Information about the 3D structure of a particle was limited to modeling estimations using simple geometrical bodies such as ellipsoids.

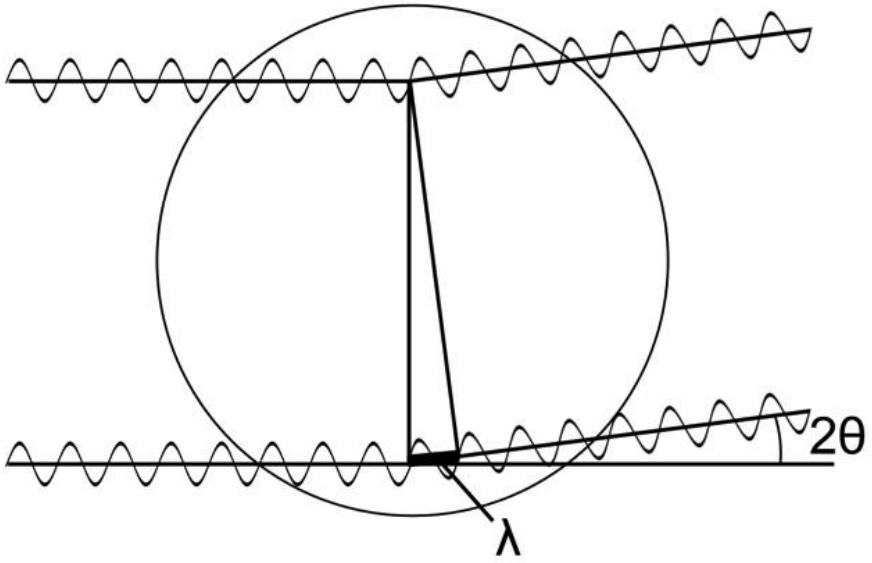
Developments in the last decade that have revolutionized SAXS

- Modern third-generation sources offer brilliance, i.e. flux on the sample and a highly parallel beam.
- Rapid readout noiseless detectors provide high-signal to noise (the SAXS signal is weak and has a high dynamic range)
- Computational algorithms have advanced (spherical harmonic approaches and more recently, molecular dynamics coupling to bead modeling).
- Computational power – thank the video gamers!

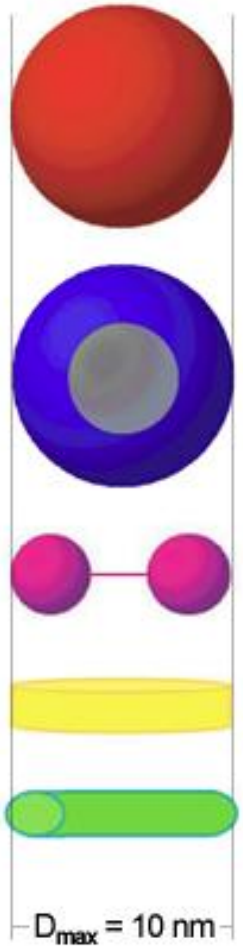


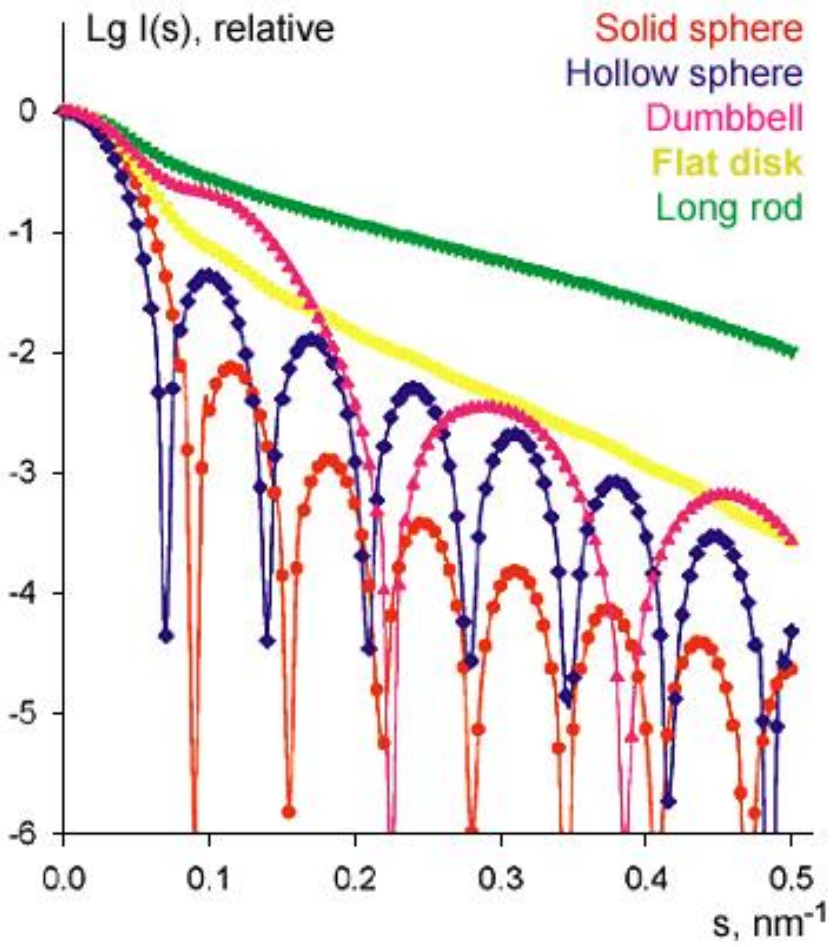
$$I(q) = \int 4\pi r^2 \cdot \bar{\rho}^2(r) \cdot \frac{\sin qr}{qr}\, dr$$


2θ
λ


2θ
λ

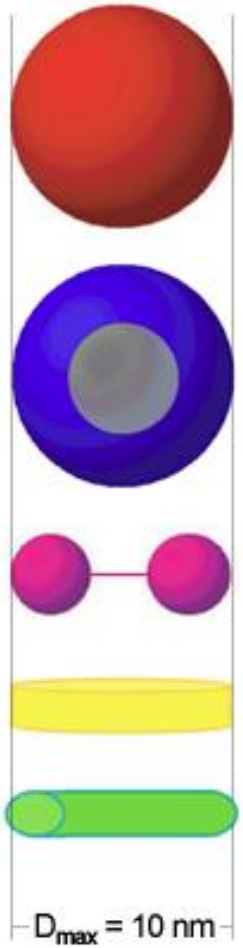
Data

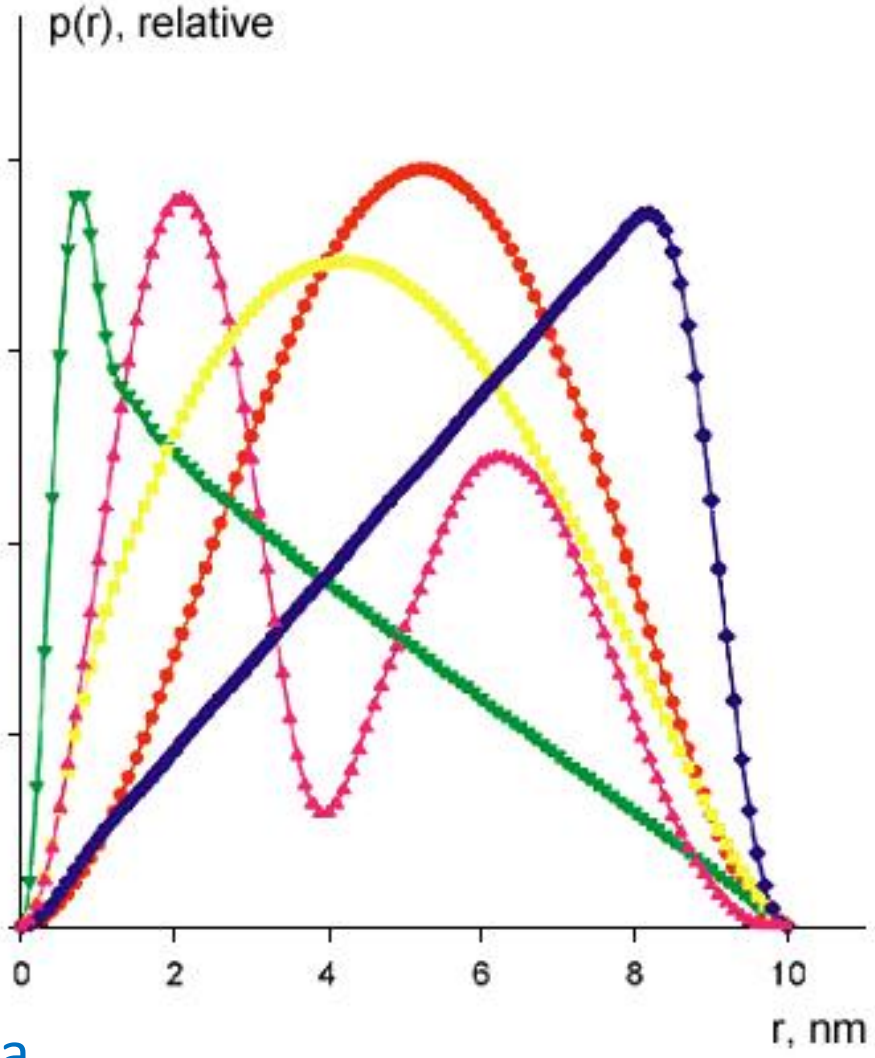




From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)

Pair distribution function

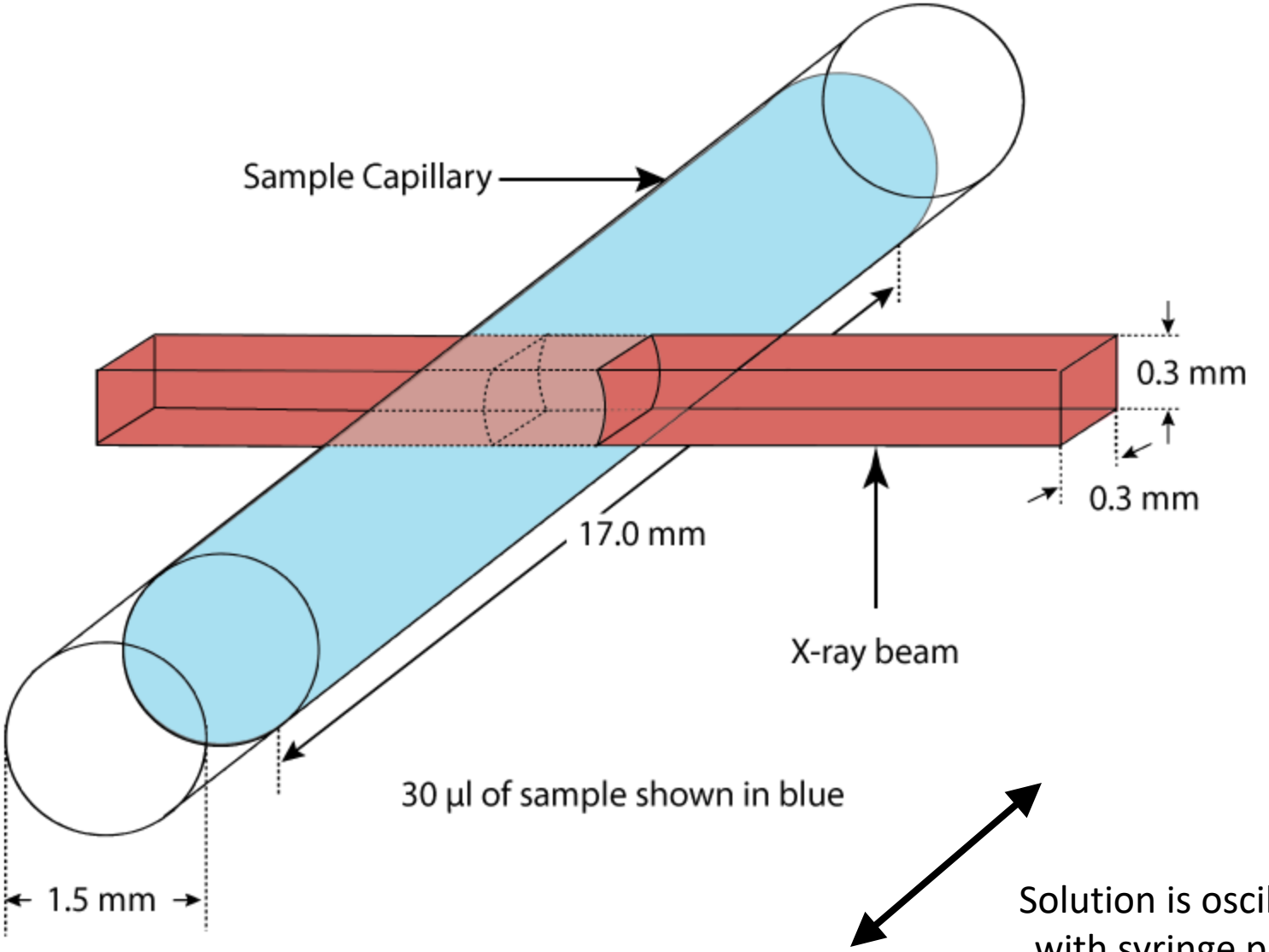


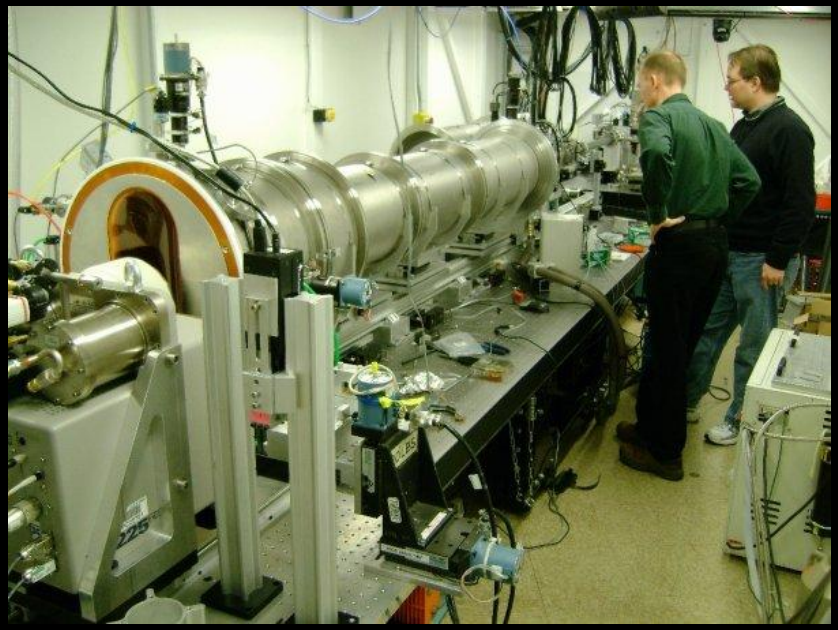


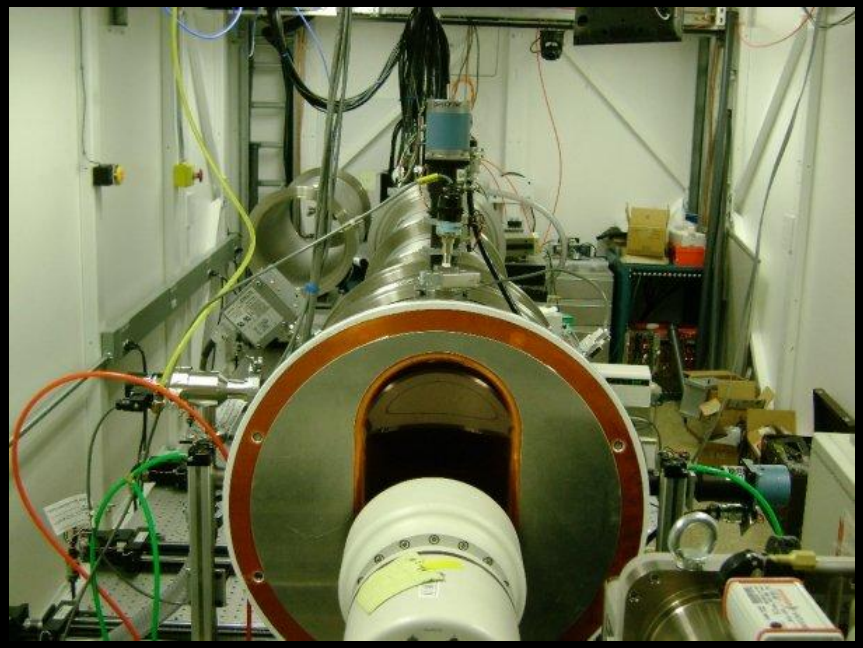
Fourier transform of data.

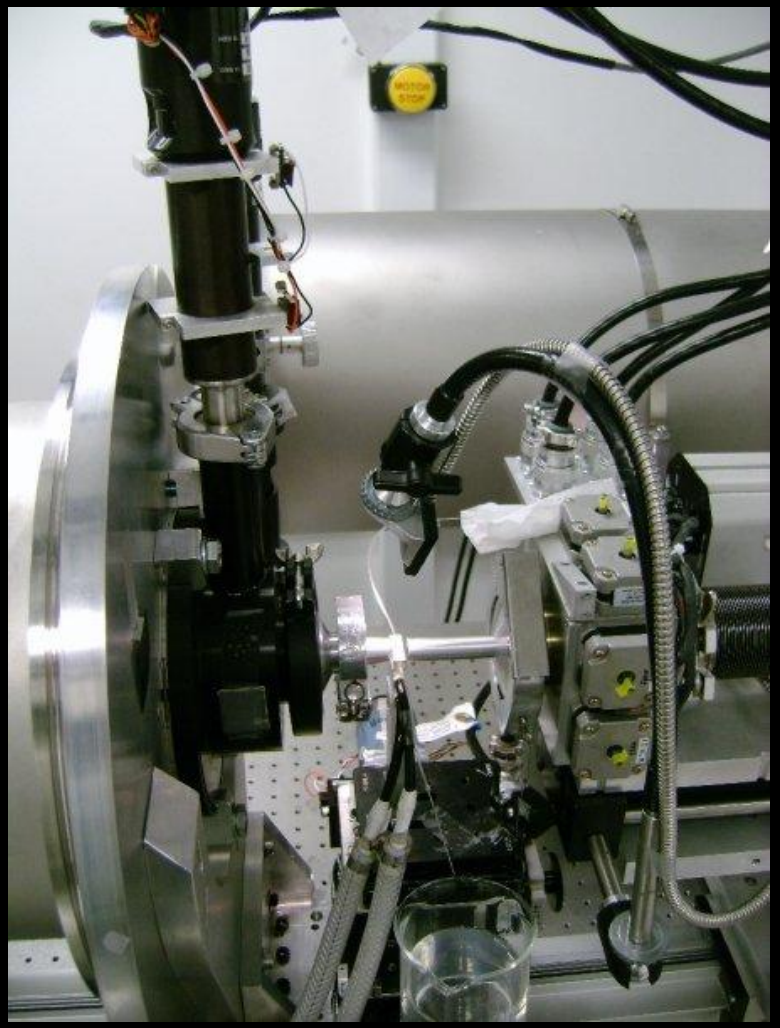
From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)

Experiment Setup









Beamline 4-2 SSRL

High throughput protocol

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.

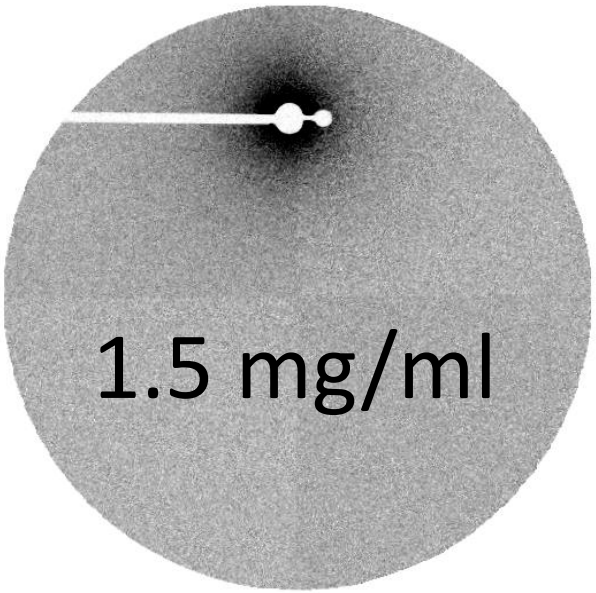
Oscillate sample to minimize radiation damage

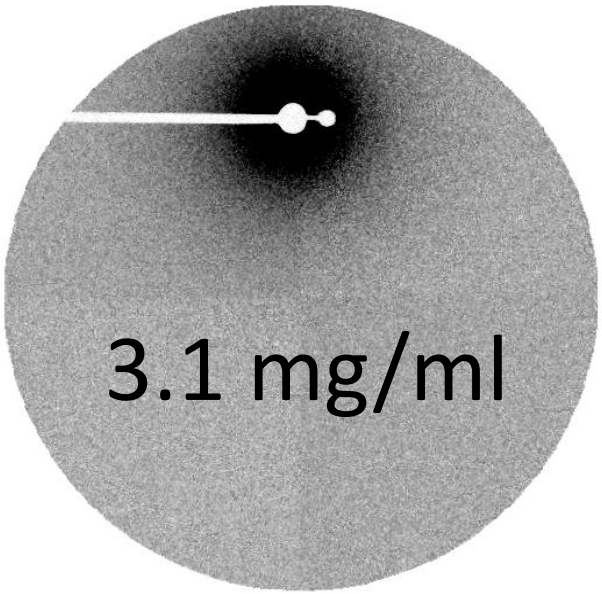
Repeat the buffer.

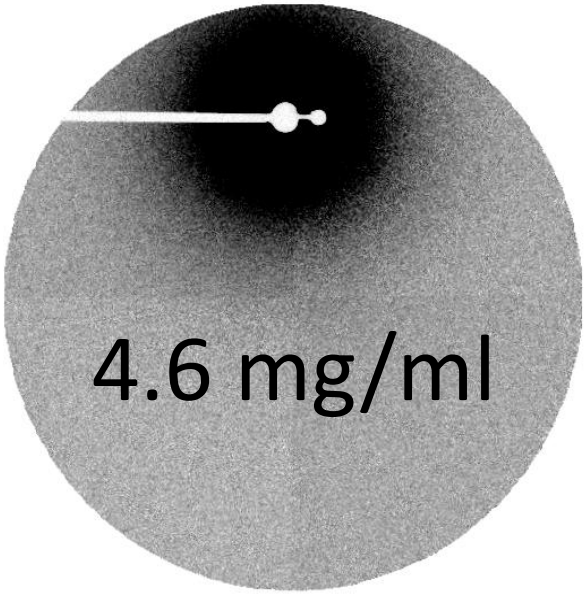
Load next sample

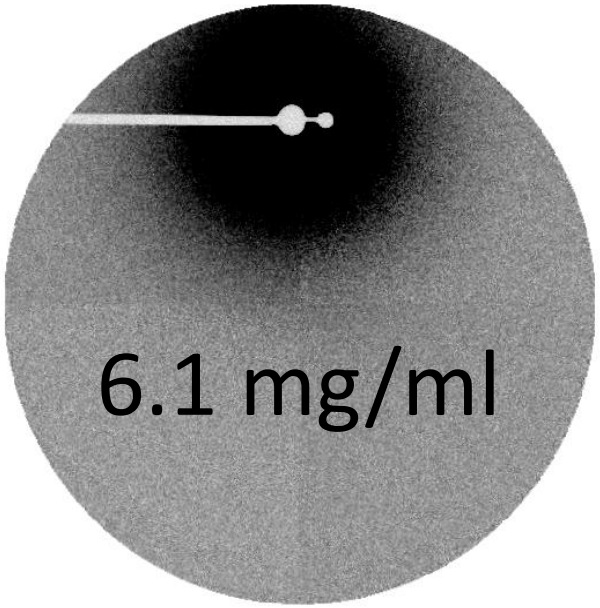
Time per concentration series – approximately 10 to 15 minutes. In high-throughput mode 24 samples in 3 to 4 hours.

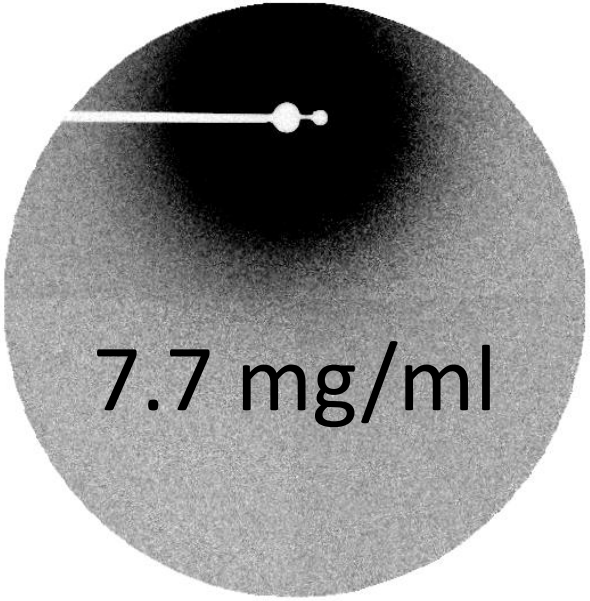
Enables two important things – eat and sleep!

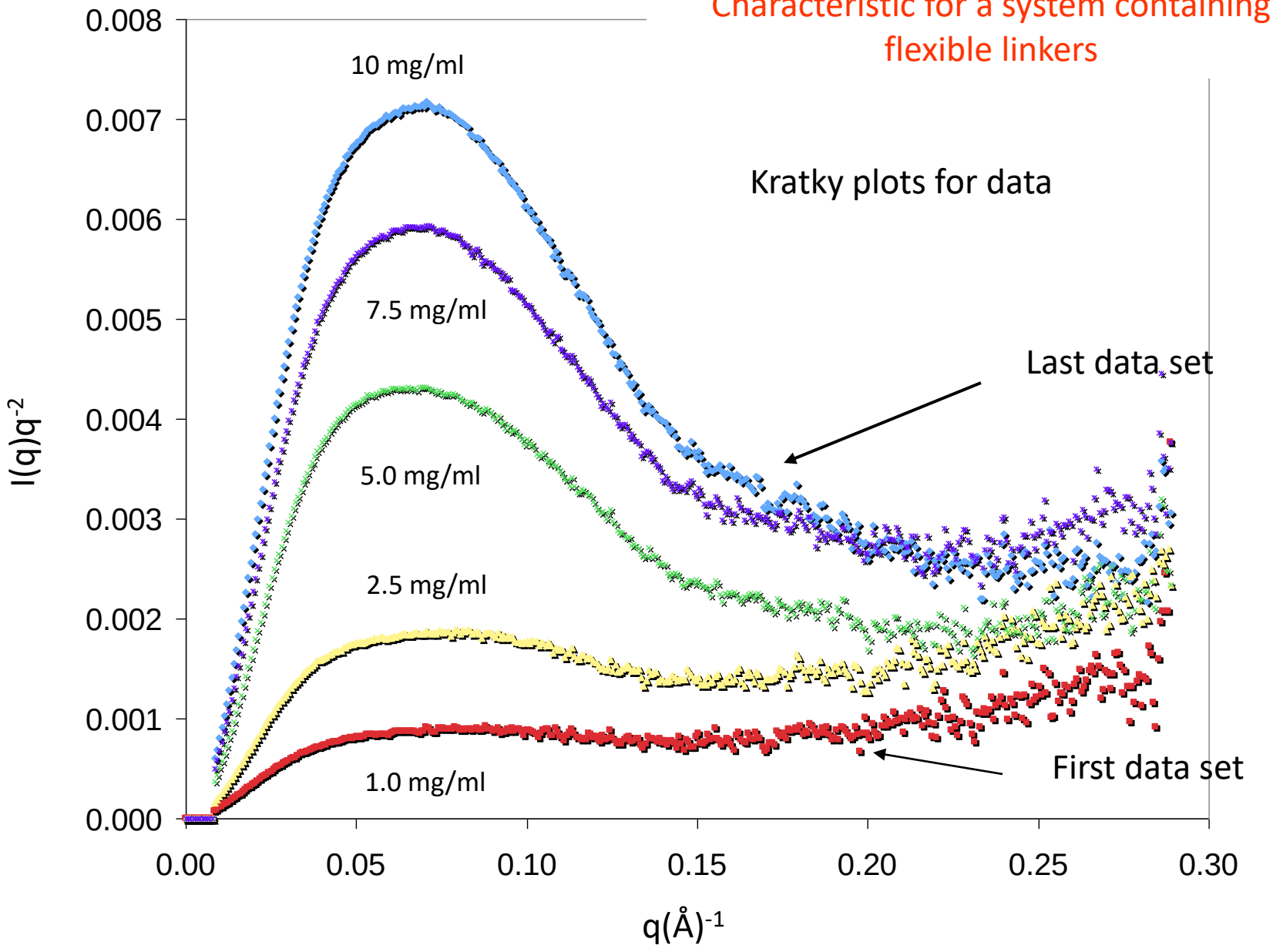

1.5 mg/ml


3.1 mg/ml


4.6 mg/ml


6.1 mg/ml


7.7 mg/ml

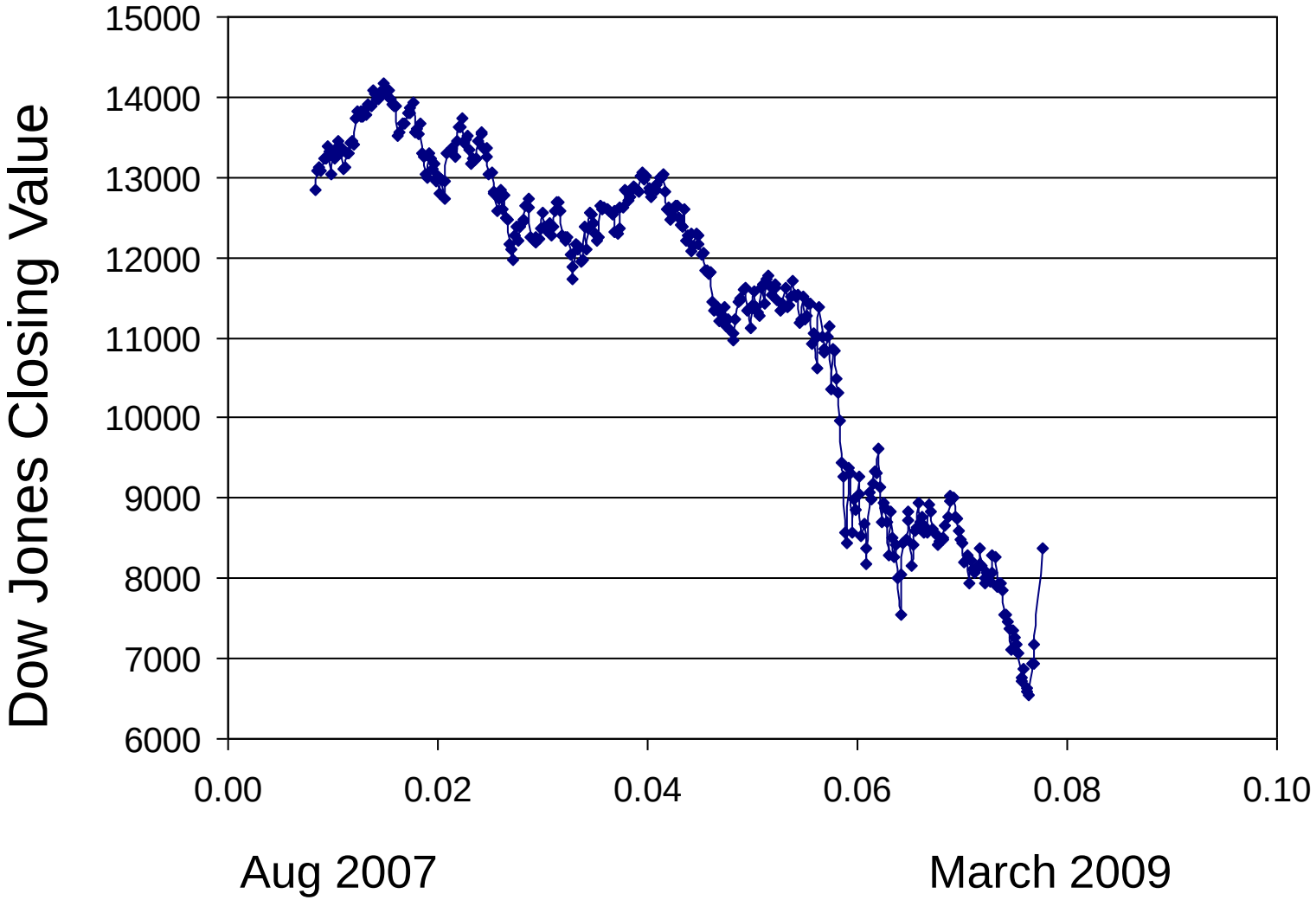

Characteristic for a system containing flexible linkers
Kratky plots for data
10 mg/ml
7.5 mg/ml
5.0 mg/ml
2.5 mg/ml
1.0 mg/ml
Last data set
First data set
I(q)q-2
q(Å)-1
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.00
0.05
0.10
0.15
0.20
0.25
0.30

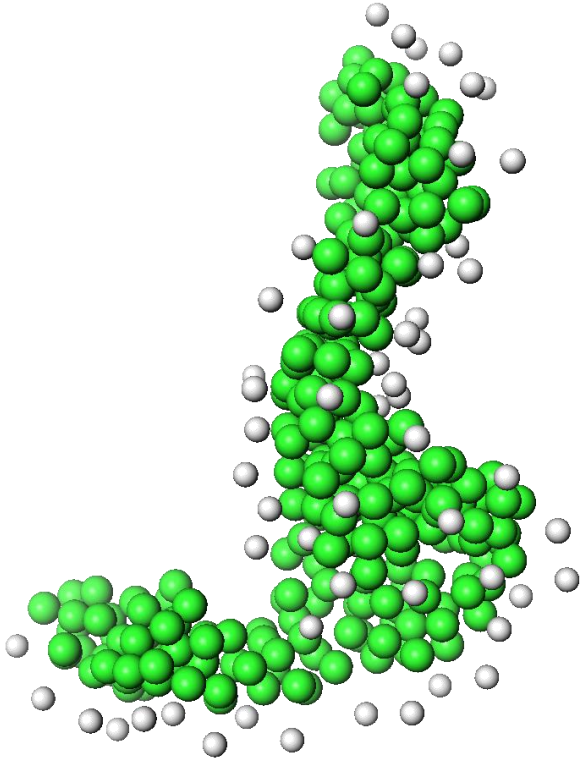
Warning – SAXS produces a scattering profile from which a three dimensional envelope can be reconstructed

It’s not necessarily the correct envelope

Warning – SAXS produces a scattering profile from which a three dimensional envelope can be reconstructed

It’s not necessarily the correct envelope


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

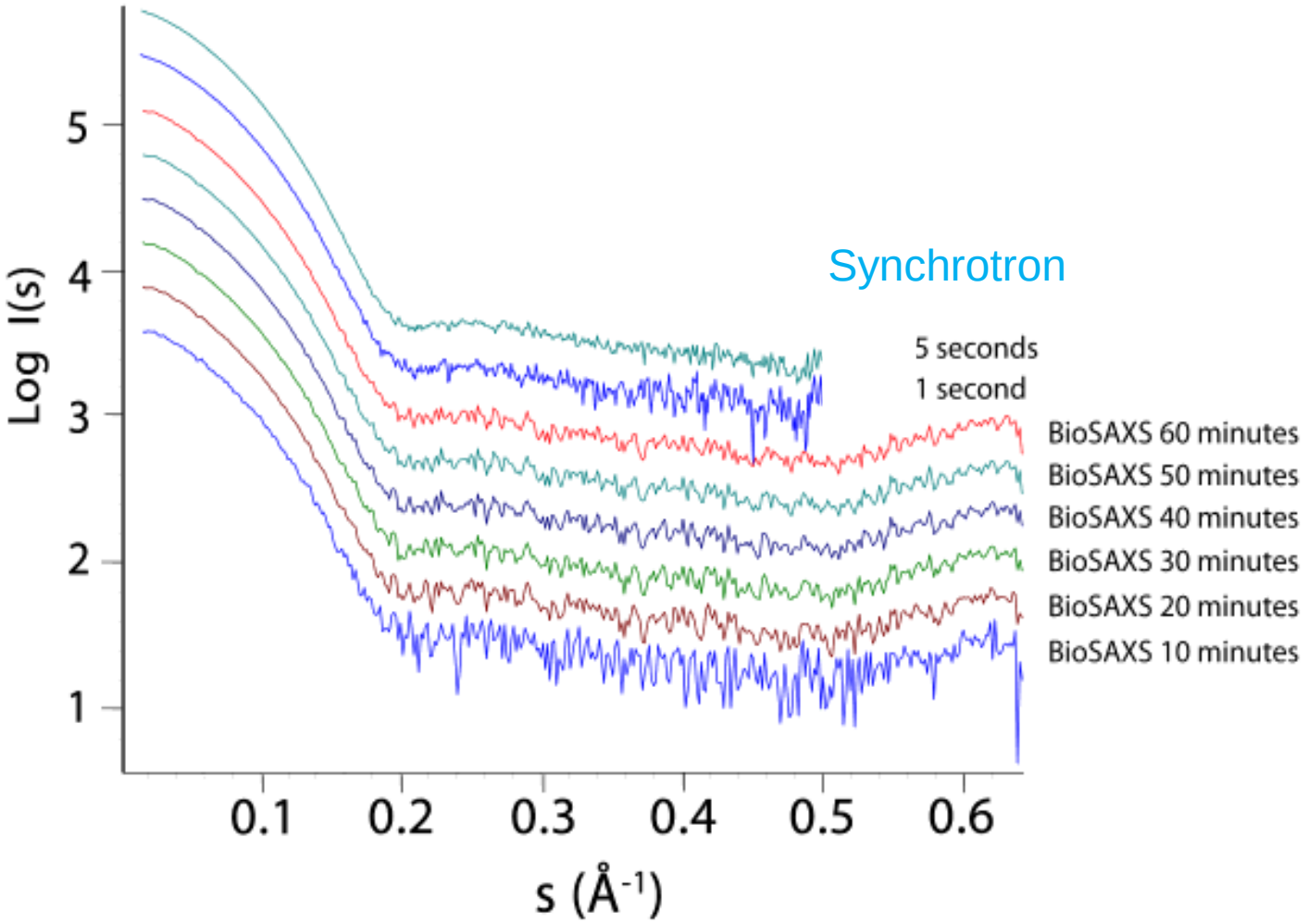


The scattering data from SAXS provides a 1D Fourier transform of the envelope of the particle.

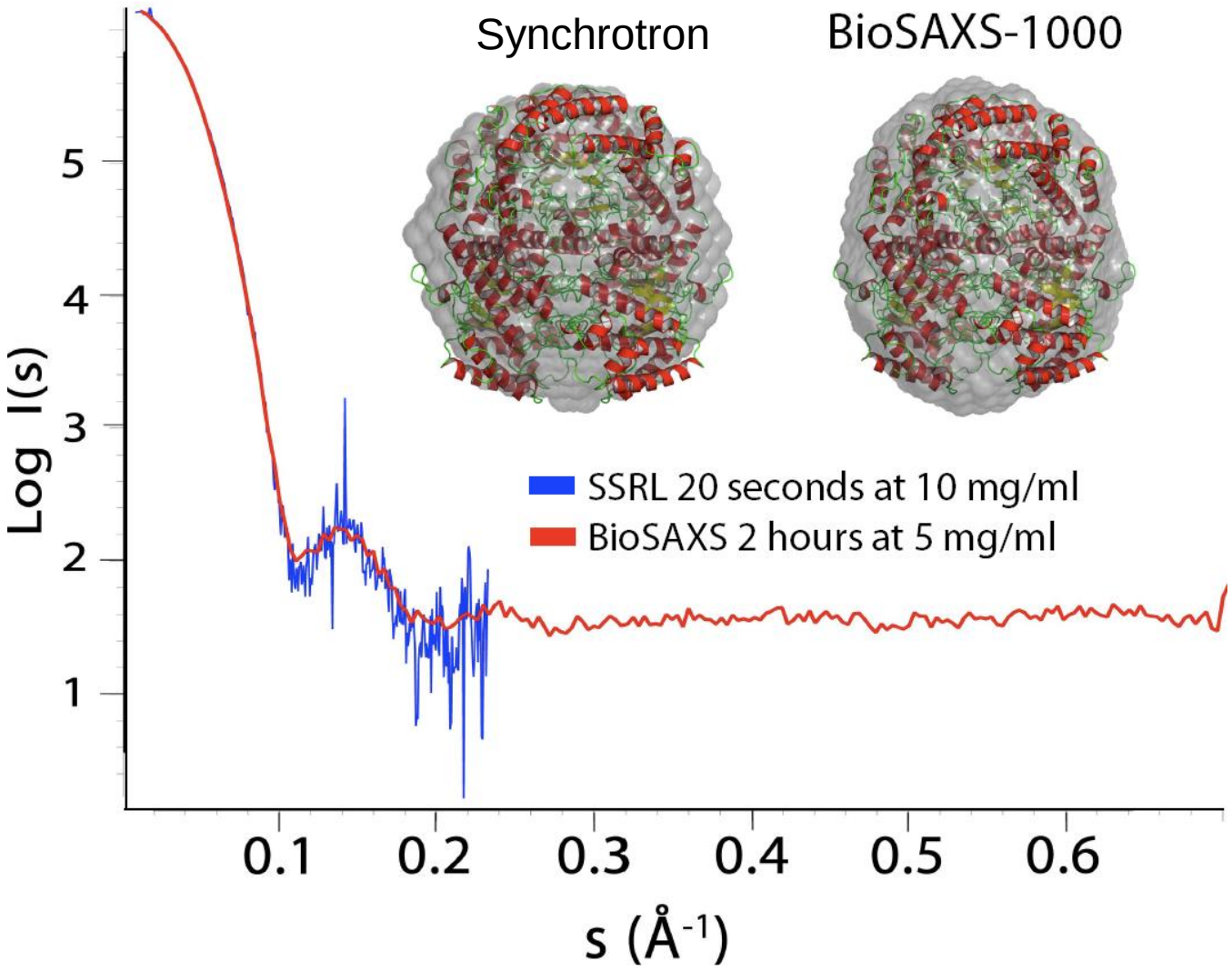
It’s possible to fit multiple envelopes to the data.

You will always get an envelope despite the data!

SAXS Information comes from shape and not intensity



Laboratory data scaled to synchrotron



Why is SAXS useful (beyond the fact you only need a solution)?

What can SAXS provide?

Radius of gyration

Maximum particle dimension

Oligomeric state and organization in solution

Amount of native flexibility or unfoldedness

Visualization of disordered regions not seen in X-ray crystallography

Low resolution molecular envelope

Characterization of mixtures

Requirements for Successful SAXS experiment

Requirements during data collection

- The sample is monomodal
- It does not aggregate
- It does not repel
- It is globular
- It is stable
- It does not suffer from radiation damage

Monomodal:

Calculate molecular weight from the SAXS data (two methods), compare to predicted and measured weight, look for oligomer values.

Aggregation:

Look for deviations from expected properties (Gunier plot), concentration dependence of intensity at lowest angles, upswing in raw data at lowest angles.

Repulsion:

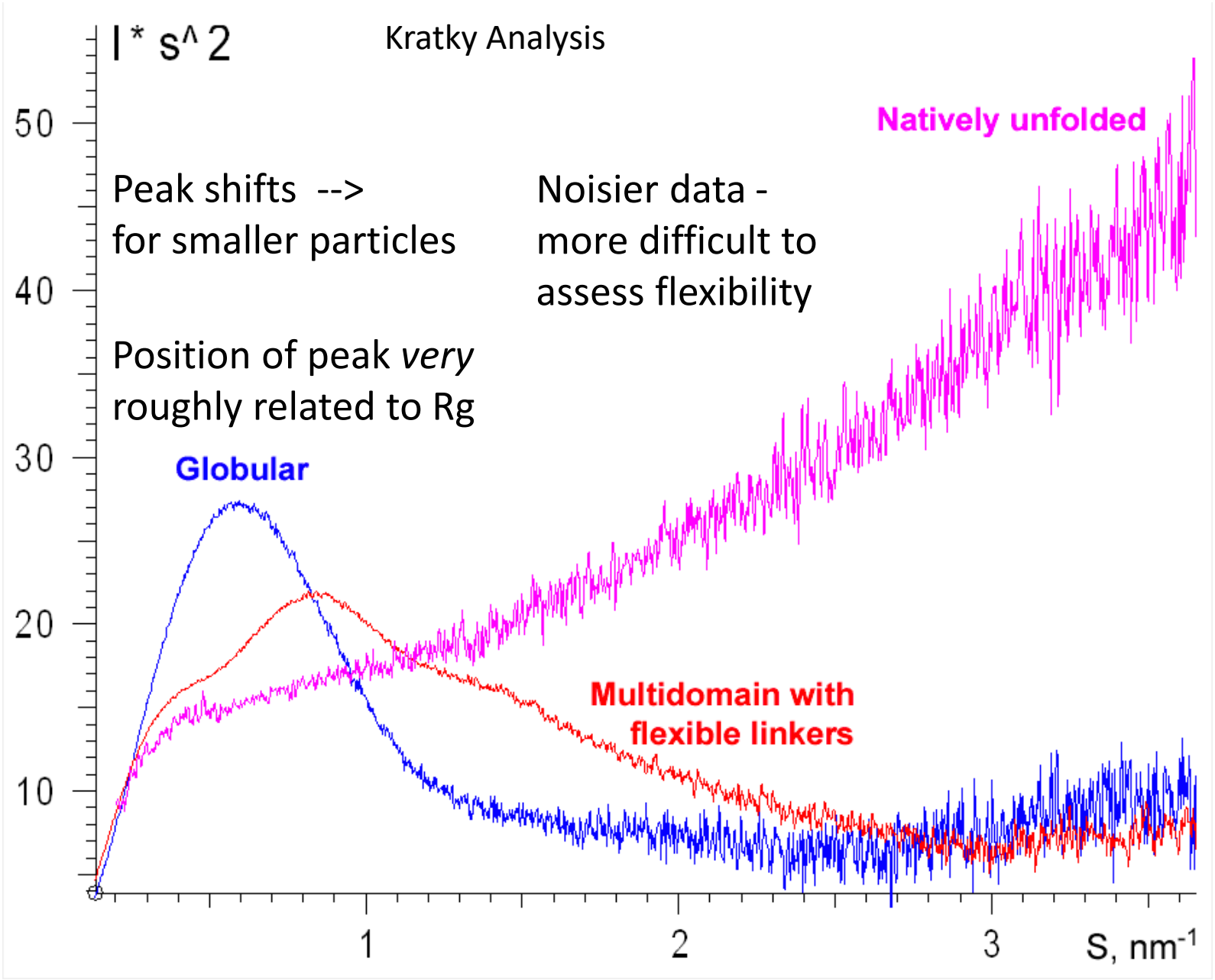
Concentration dependent downward trend in data as a function of concentration at the lowest angles. Can be corrected with dilute solutions.

Globular:

The globularity of the protein can be determined from the Kratky plot which shows if it is well folded, has flexible linker regions or is denatured (SAXS is a powerful technique to characterize the protein).

Stable:

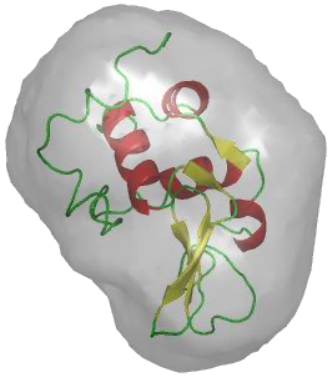
If a biochemical assay is available this can be used. In terms of data collection multiple exposures are taken over time and compared. In some cases this comparison takes place over multiple beamtimes.



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

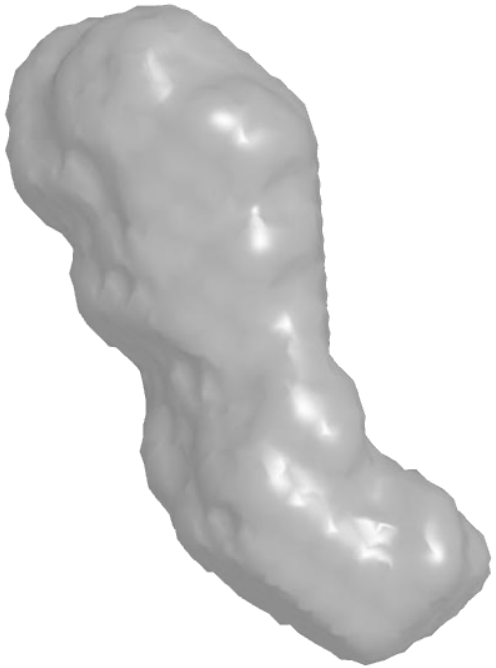
Lysozyme reconstruction

Lowest concentration with oscillation



First 10 exposures

Fact



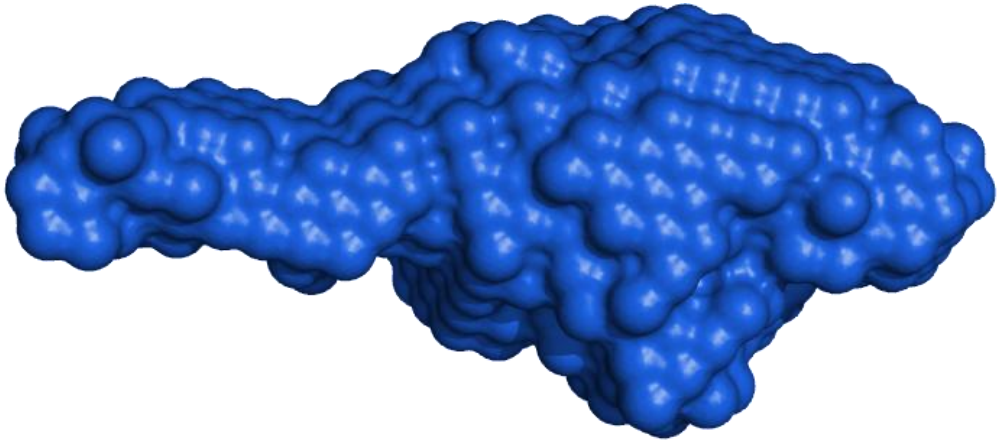
Exposures 89-99

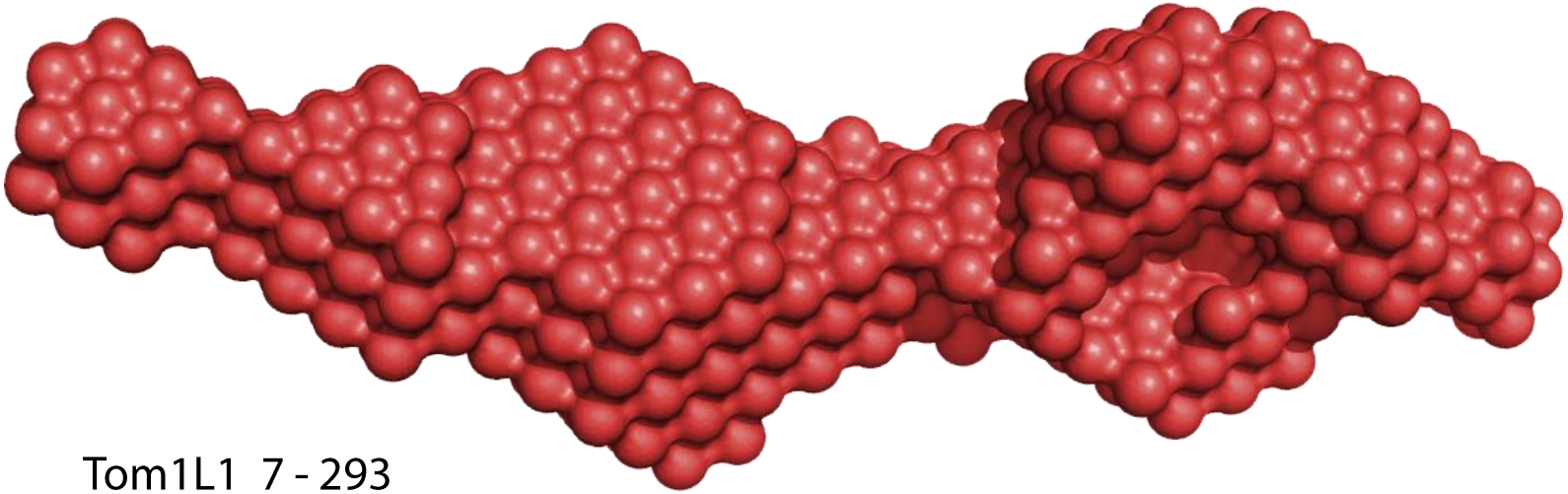
Fiction

Kratky plot indicates little to no unfolding. Increase of Rg appears to be coming from oligomerization.

Ab Initio SAXS Envelopes

Tom1L1 7 - 186



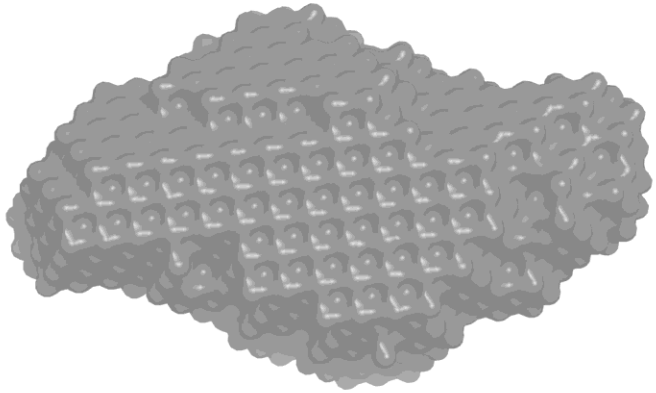


Tom1L1 7 - 293

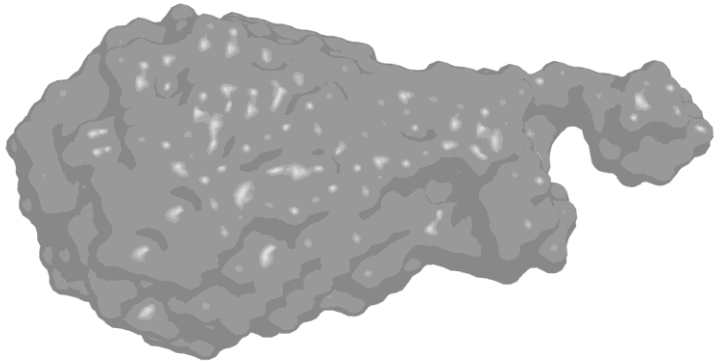
This is the only known structural information about TOM1L1 to date

Examples

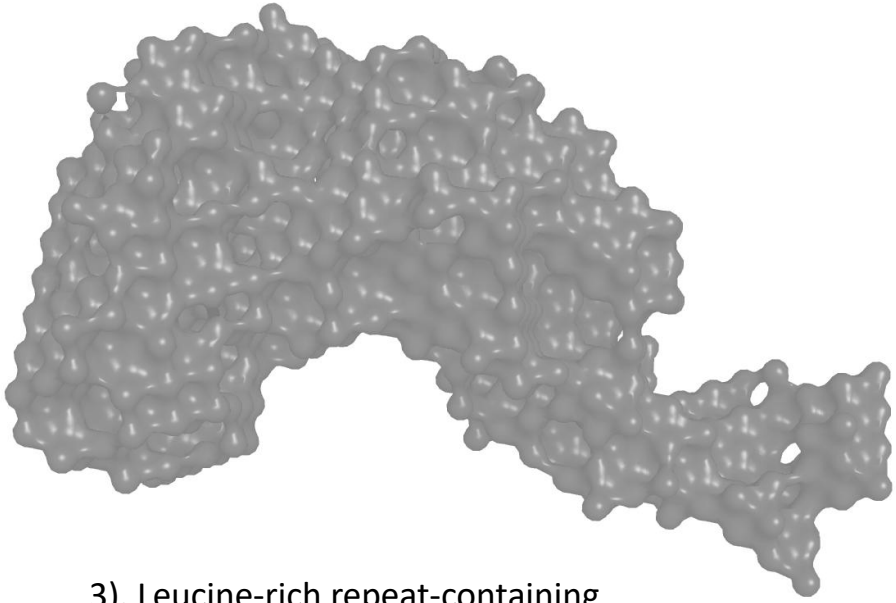
Ab intio envelopes



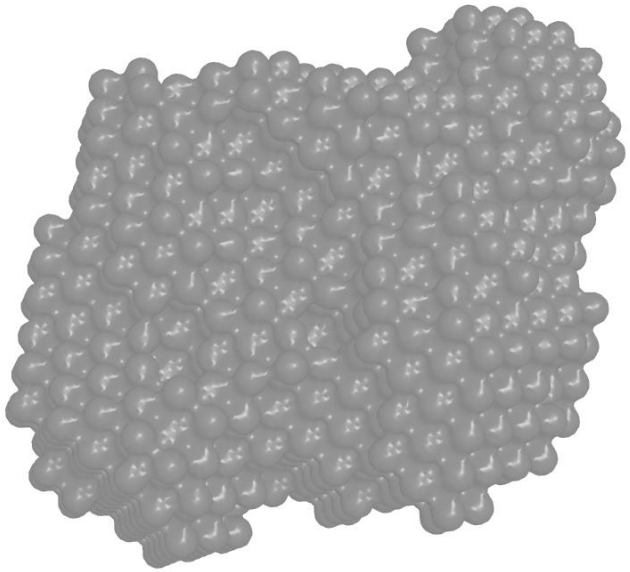
1). alr0221 protein from Nostoc (18.6 kDa)



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

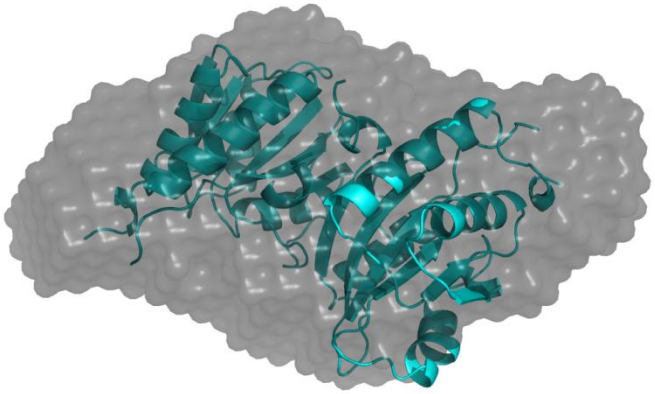


3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

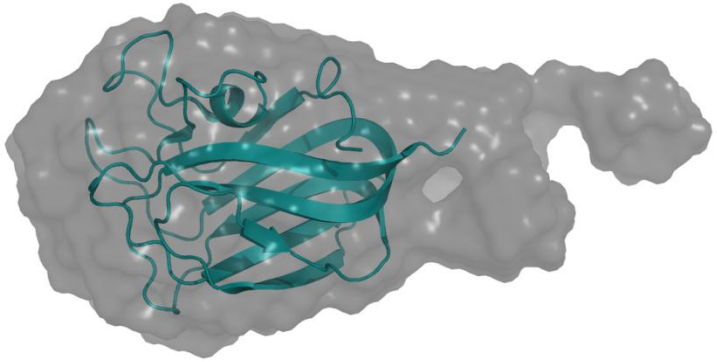


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

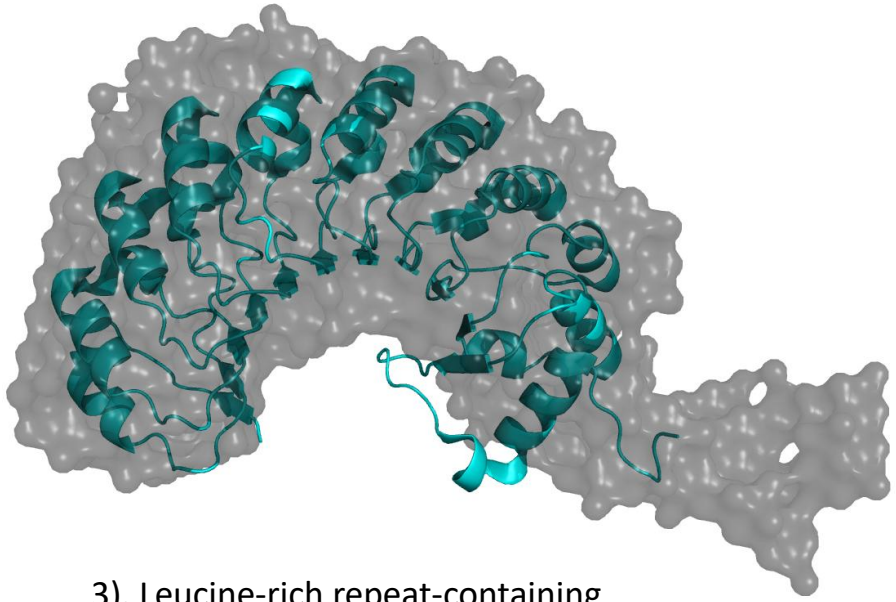
Overlaid with subsequent X-ray structures



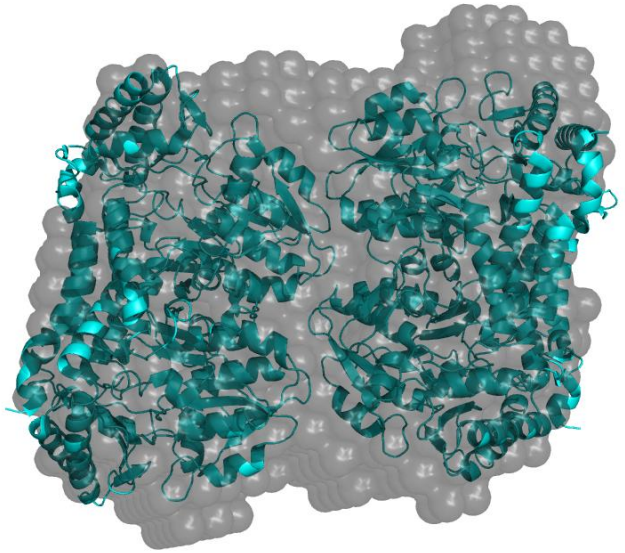
1). alr0221 protein from Nostoc (18.6 kDa)



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

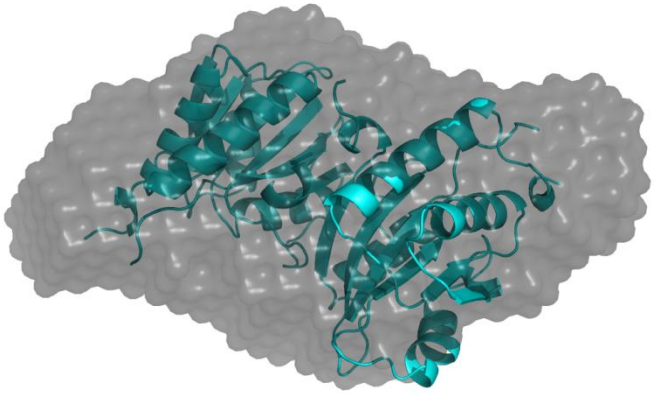


3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

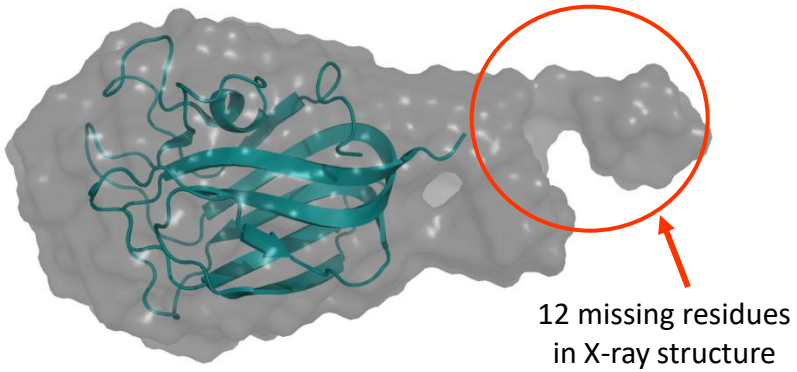


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

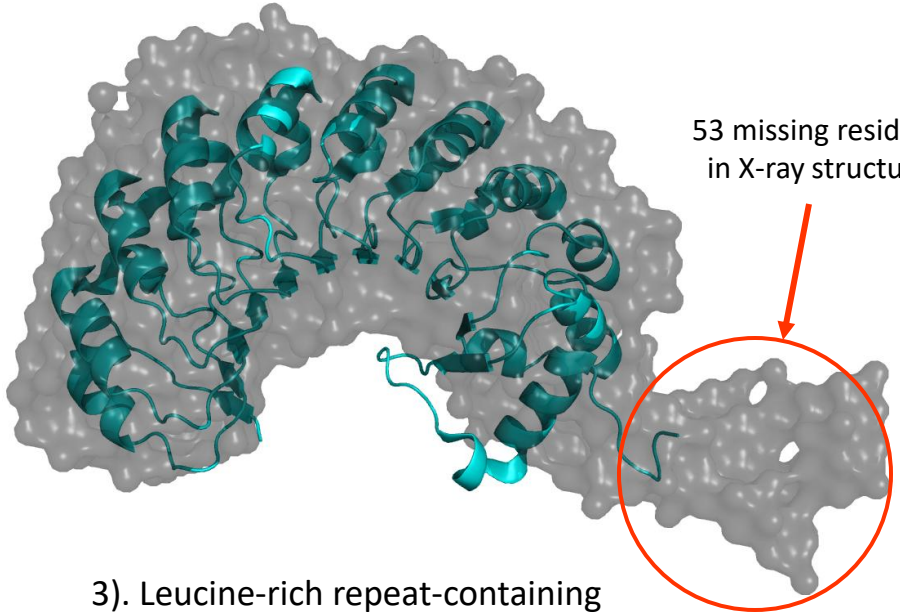
And data on what was missing …



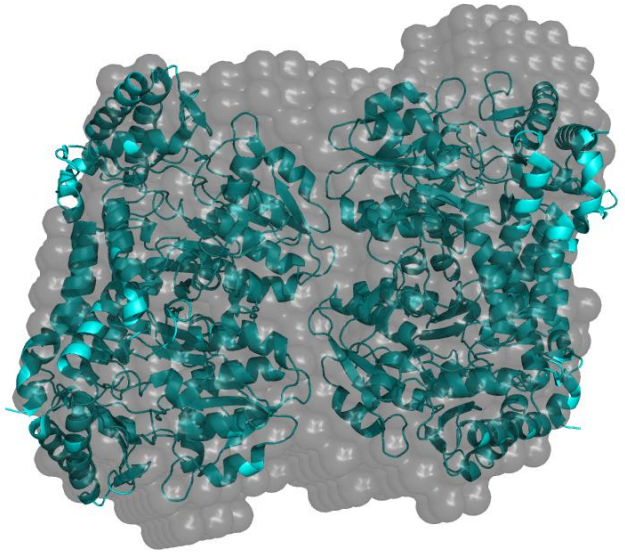
1). alr0221 protein from Nostoc (18.6 kDa)



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

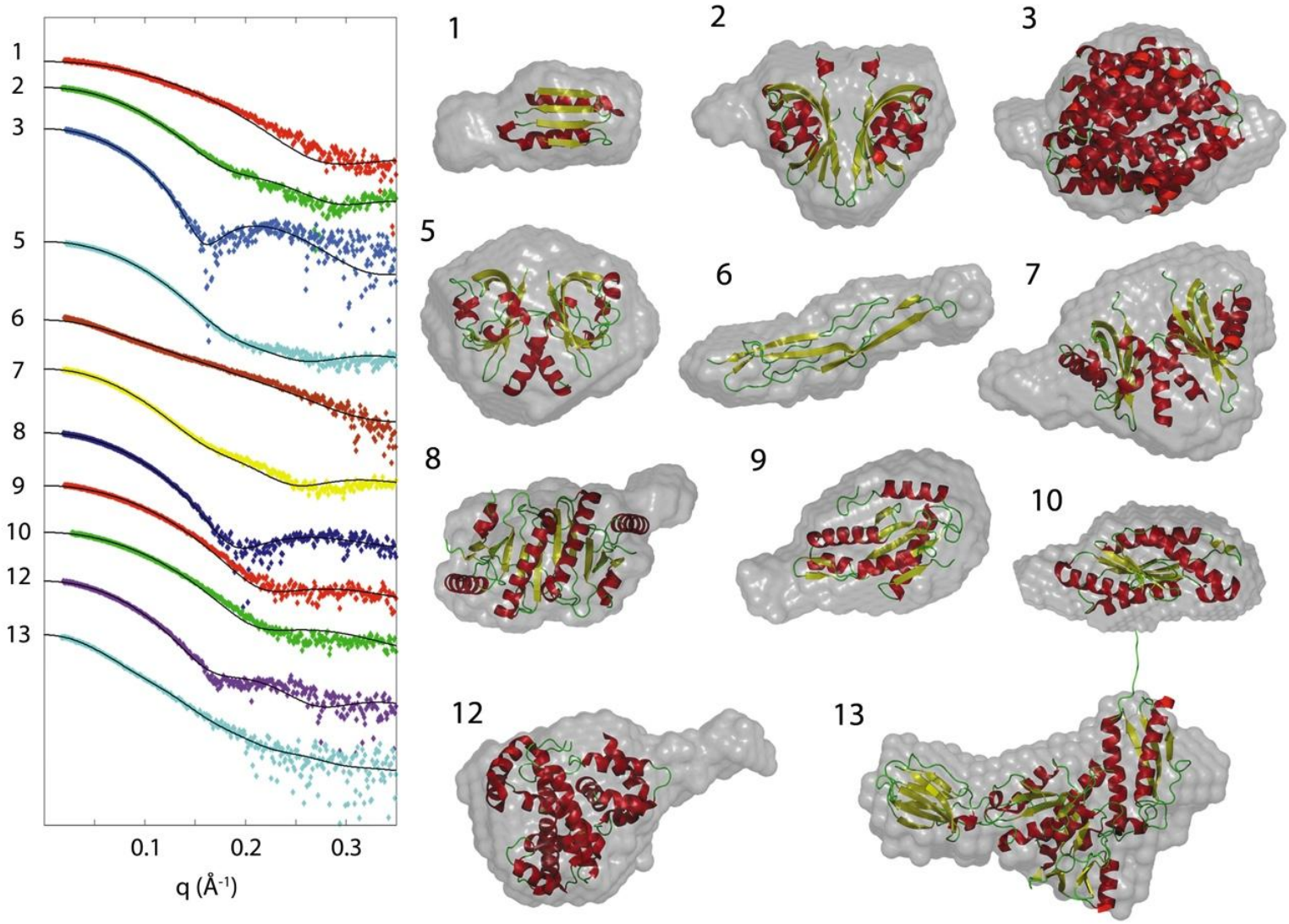


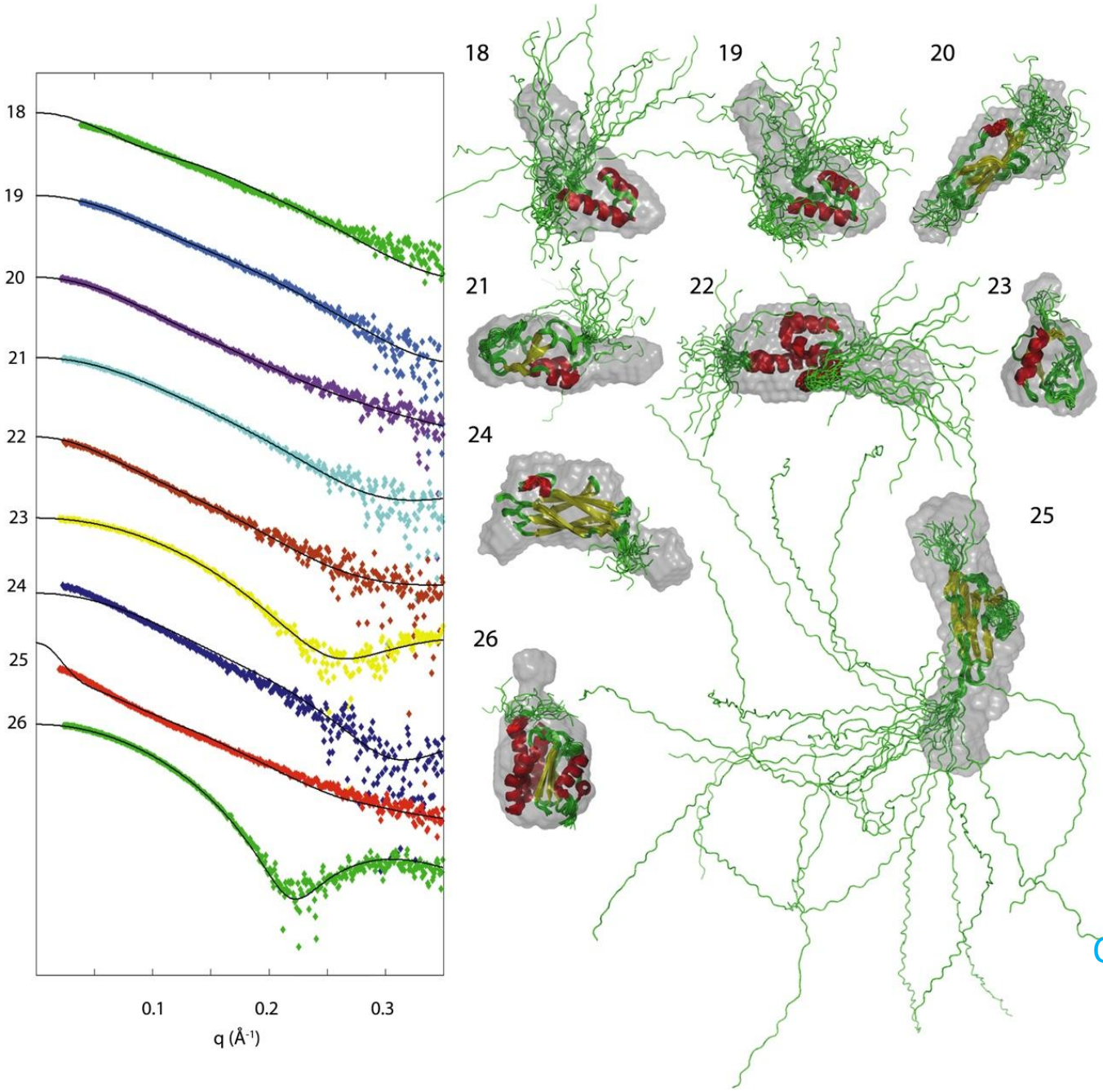
3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



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

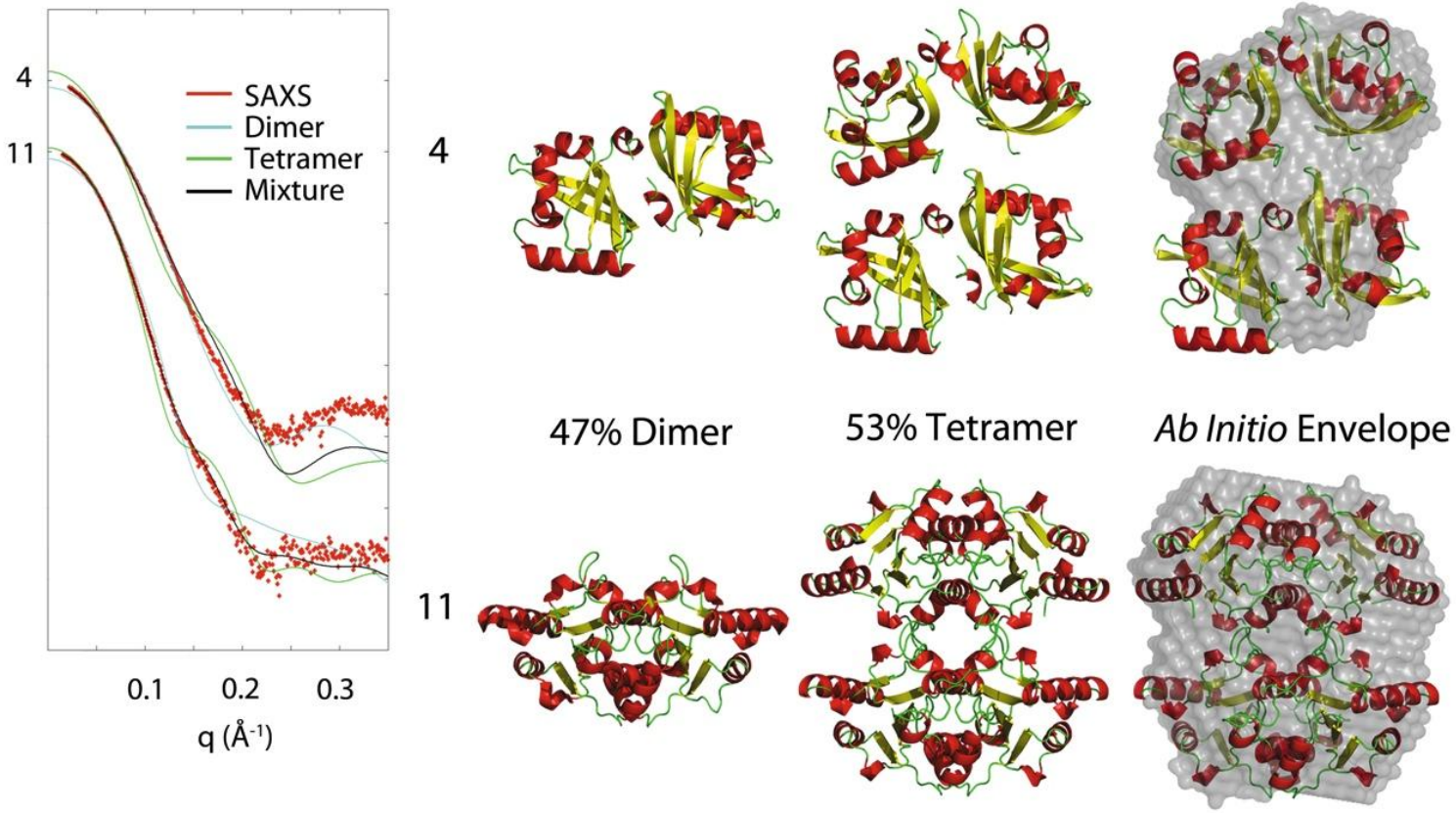
Comparing X-ray structures



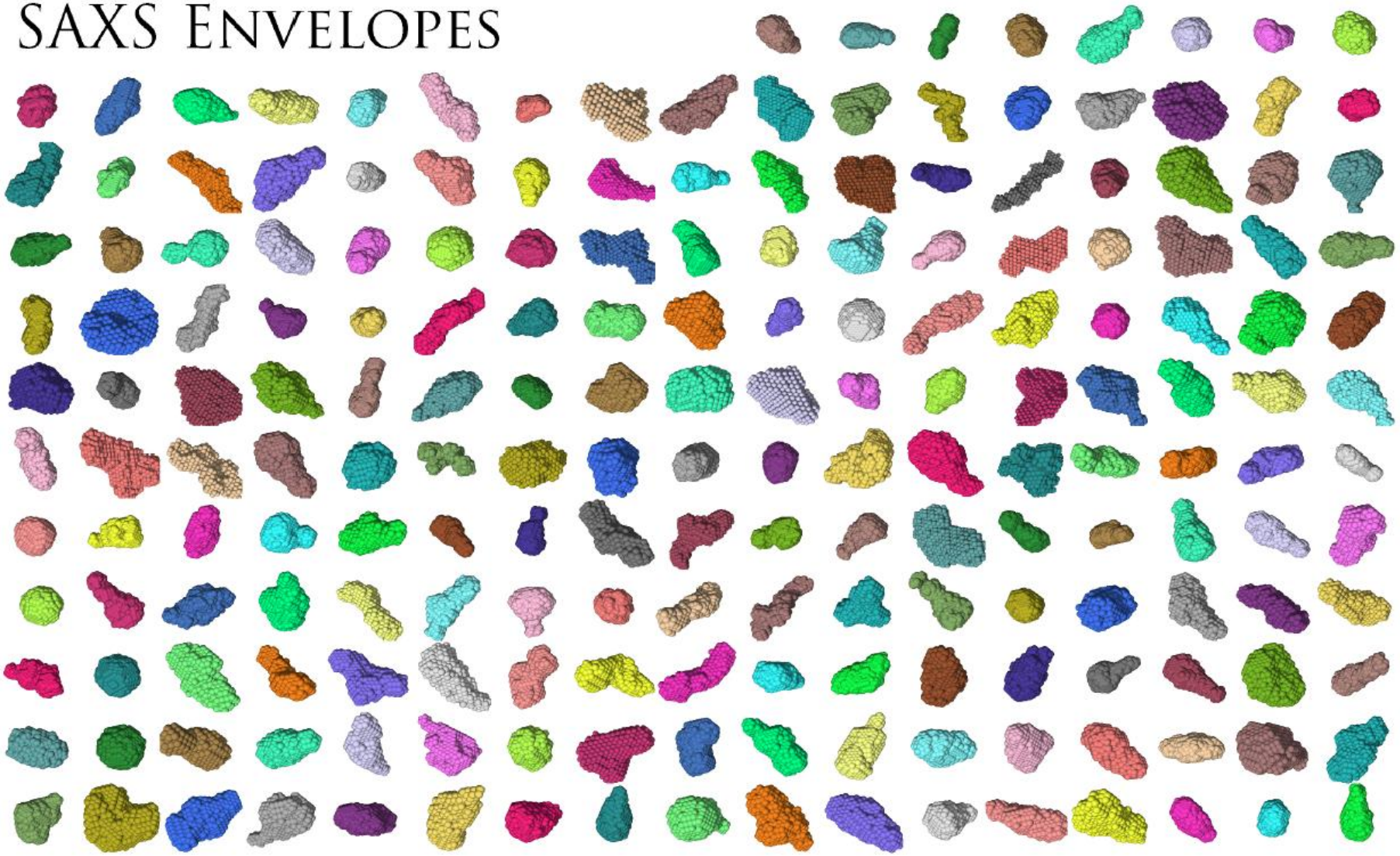


Comparing NMR structures

#	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


4
11
SAXS
Dimer
Tetramer
Mixture
0.1
0.2
0.3
q (Å-1)
63% Dimer
37% Tetramer
Ab Initio Envelope
4
47% Dimer
53% Tetramer
Ab Initio Envelope
11

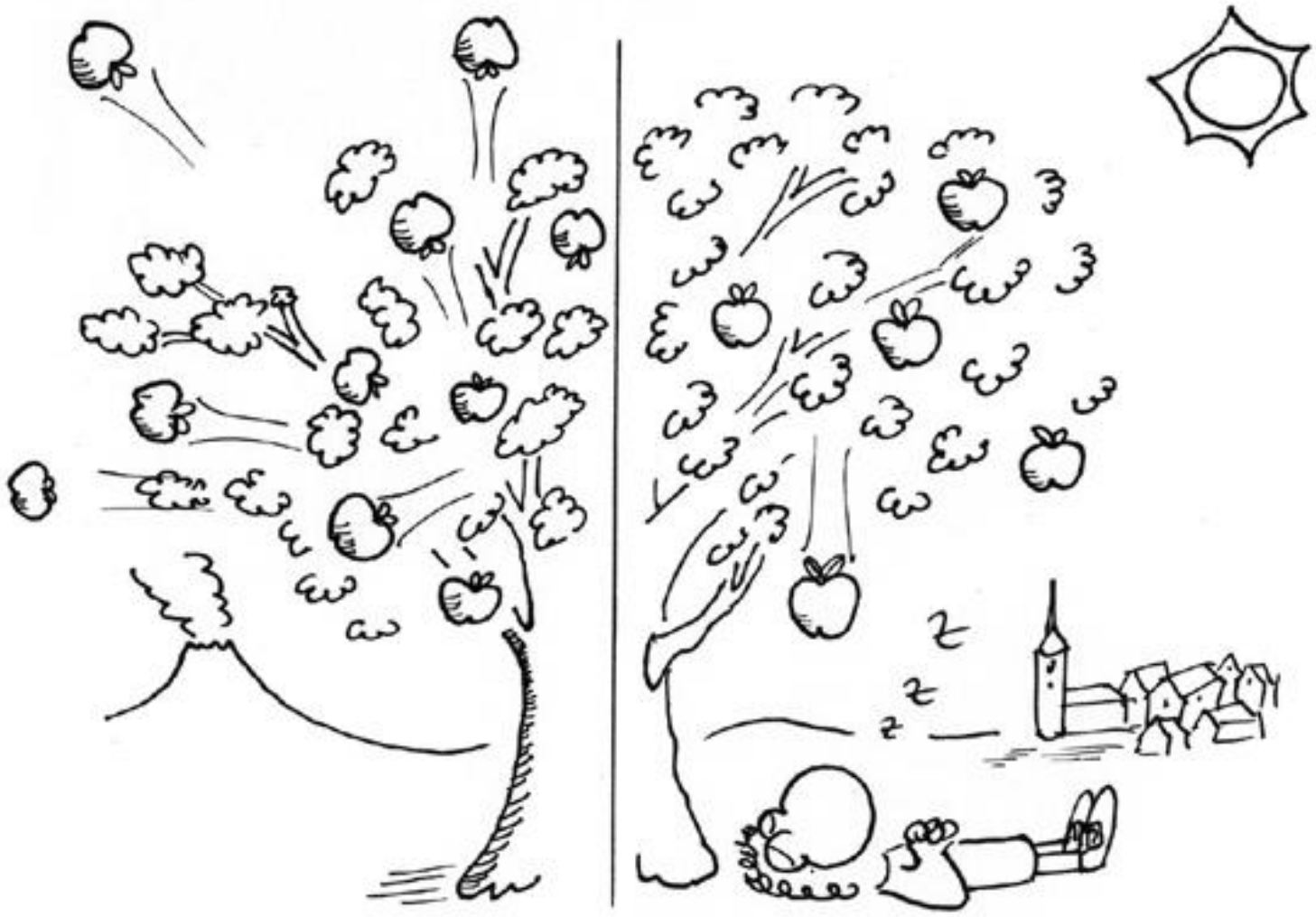
SAXS Envelopes



SAXS : the T-shirt (Tom Grant LLC)

A Biological Puzzle

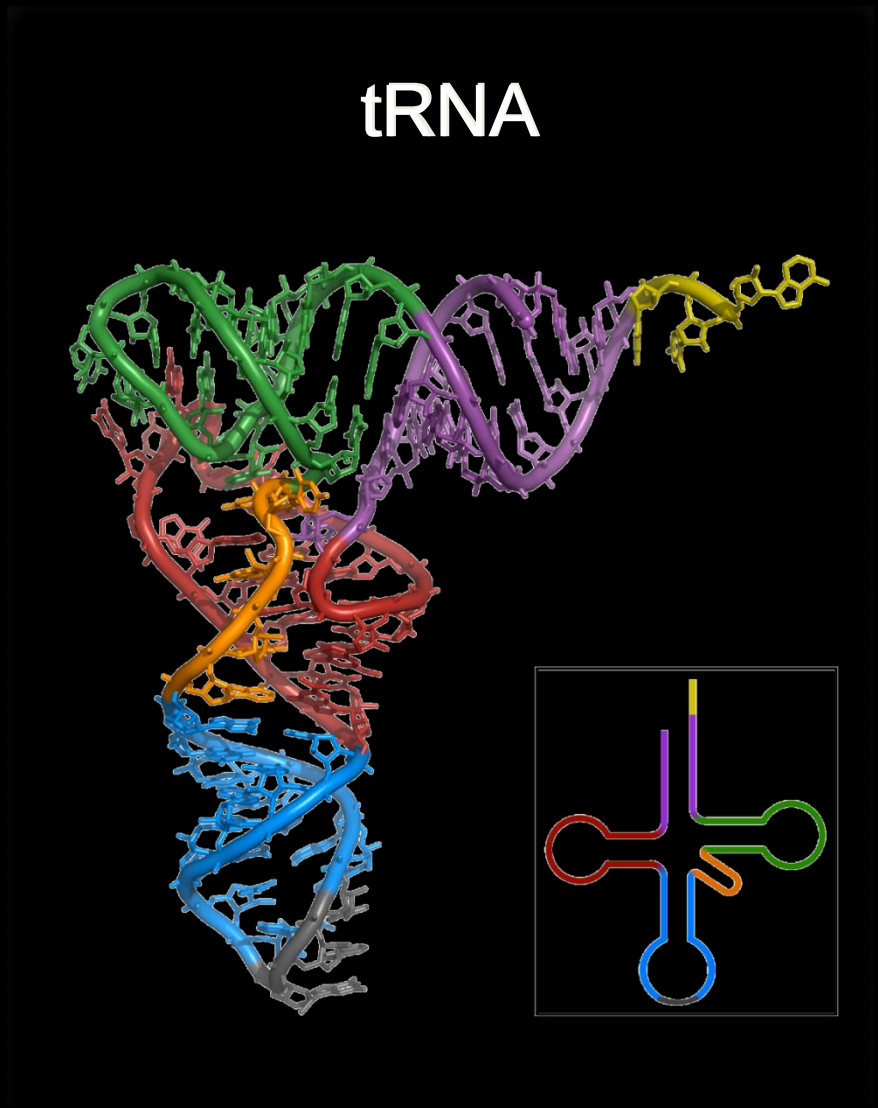
PHYSICS FOR BIOLOGISTS



A long time ago the apple trees used to shoot the apples in all directions. Only those that did it downward got reproduced. Then, after millions years of natural selection and evolution, gravity was finally discovered.

tRNA Synthetases

- Amino acids are attached to tRNA molecules which are then transferred to the ribosome for use in protein synthesis
- tRNA synthetases act as the “codebook” in the central dogma
- In most cases, one tRNA synthetase exists for each amino acid

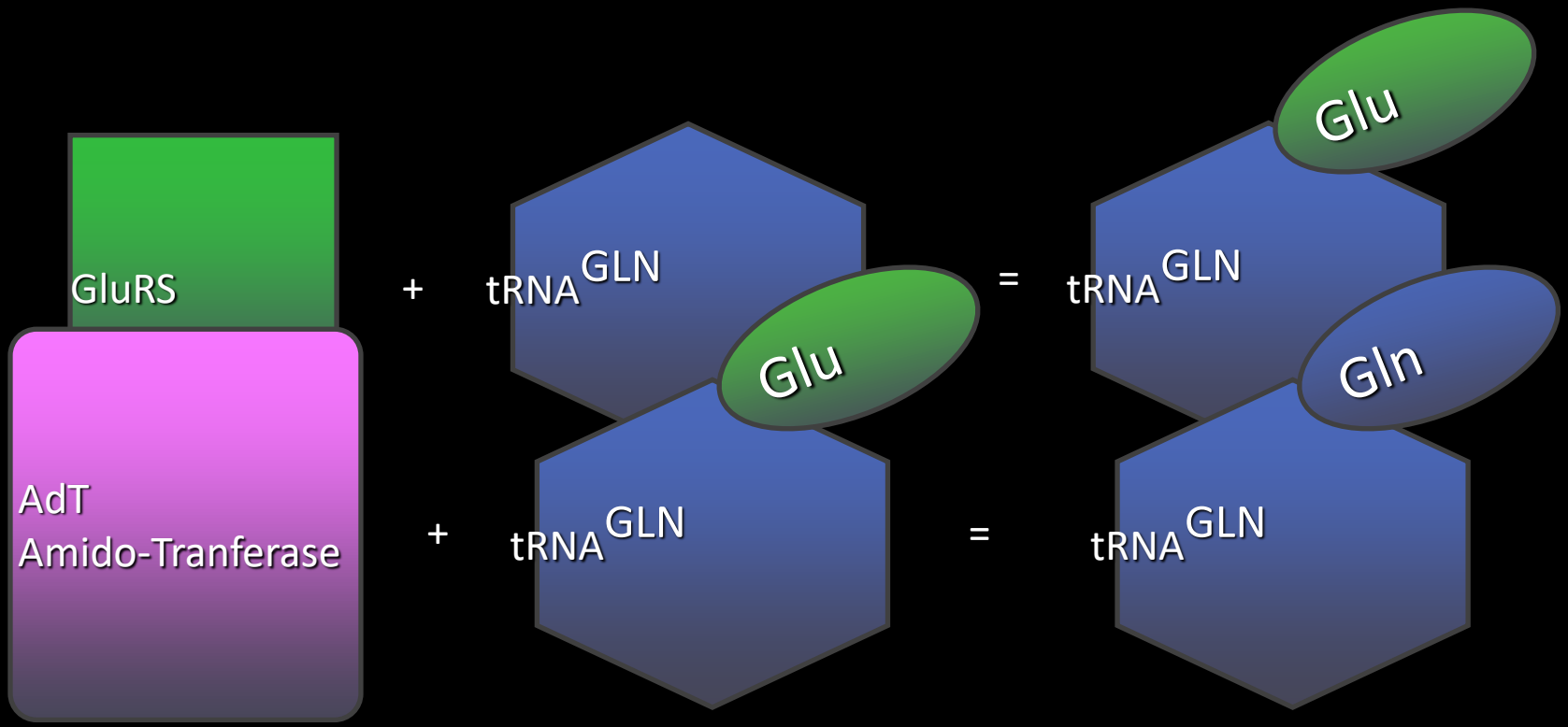


Two routes of gln-tRNA^{GLN} Formation

Direct Route: Eukaryotes and few bacteria

Two routes of gln-tRNA^{GLN} Formation

Indirect Route: Archaea and Most Bacteria



tRNA synthetase of Eukaryotes and Prokaryotes

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

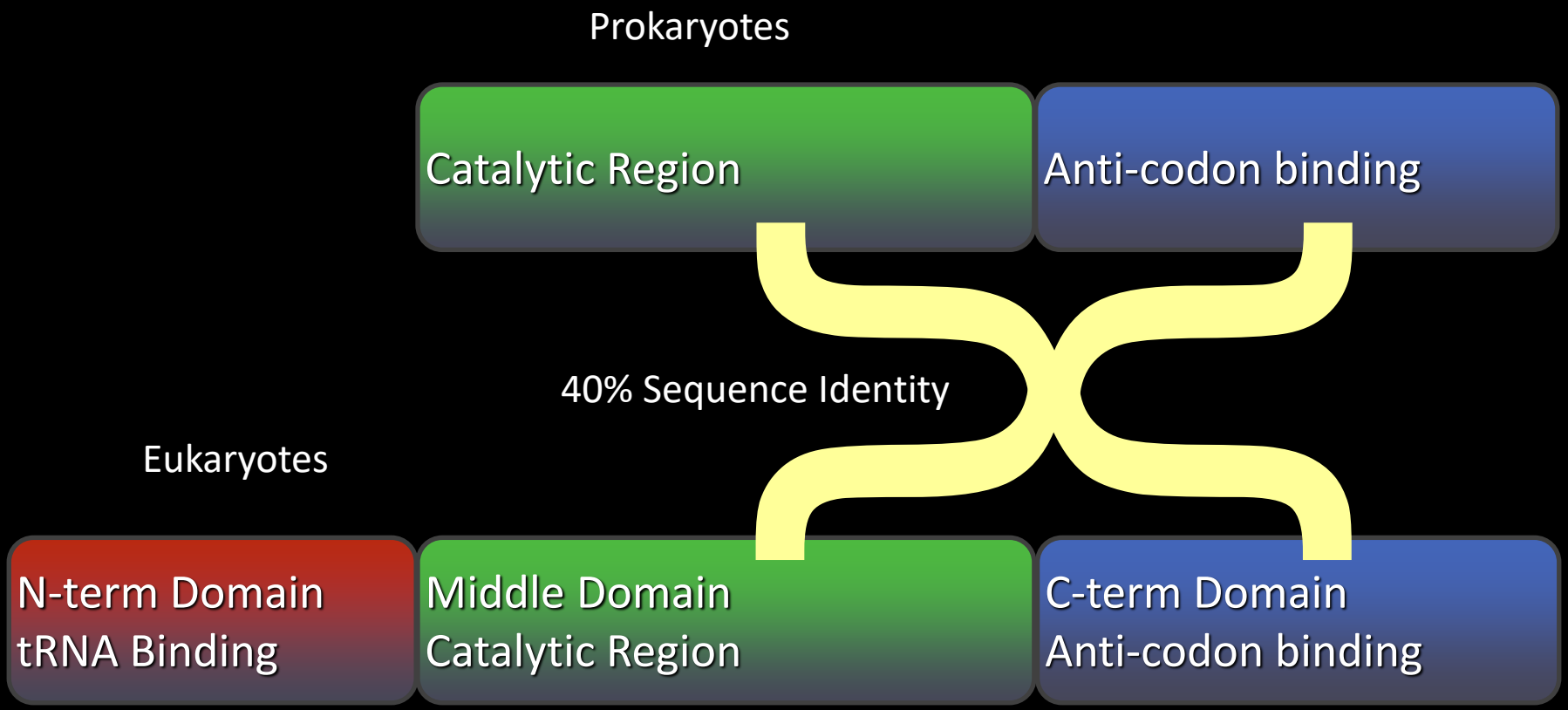
Appended Domains

- 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

The N-terminal domain (NTD)

- Eukaryotic tRNA synthetases are distinctly more complex than their prokaryotic homologs because they have progressively acquired and retained additional domains throughout evolution
- Like other eukaryotic GlnRS species, *Saccharomyces cerevisiae* Gln4 contains both a highly conserved C-terminal domain (CTD) with all of the known features of class I synthetases, as well as a less conserved appended N-terminal domain (NTD) with no obvious sequence homology to any known protein domain.
- While some appended domains are shared among synthetase families and are similar to domains in other proteins implicated in either nucleic acid binding or protein-protein interactions at least eight domains are uniquely associated with a single synthetase family, and neither their structures nor their roles are generally understood.
- The origin and function of the NTD in GlnRS are of particular interest.

Glutamine tRNA Synthetase



Target

- Our target is Glutaminyl tRNA synthetase (Gln4) from yeast *Saccharomyces cerevisiae*

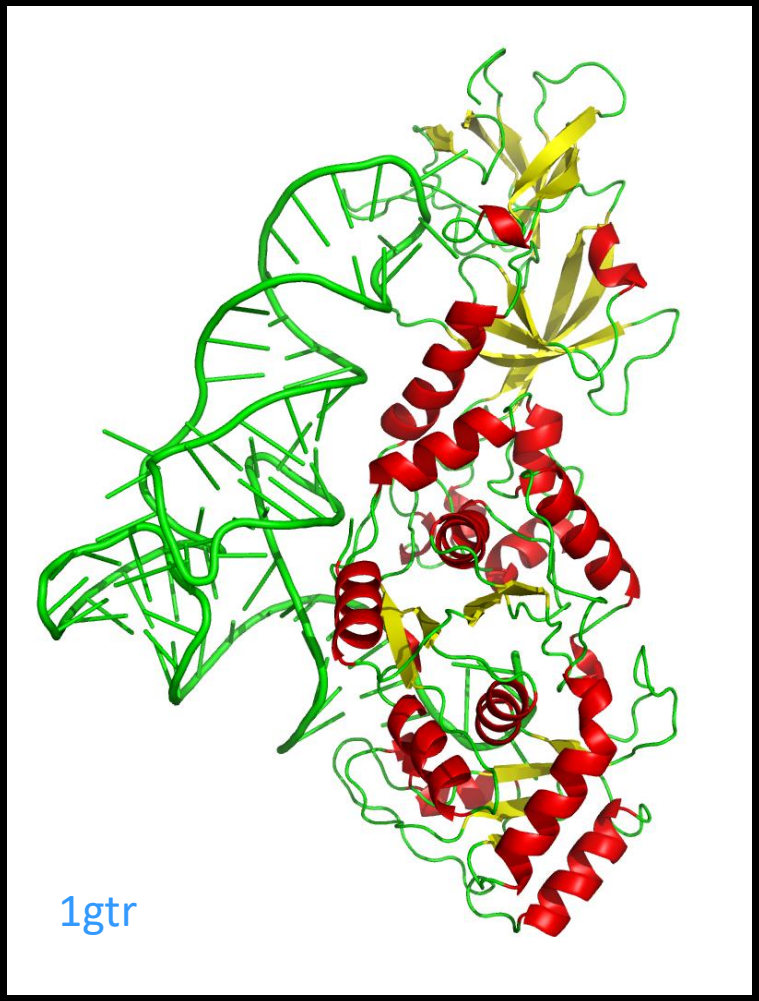
- Yeast *Saccharomyces cerevisiae is* a well-established model system for understanding fundamental cellular processes of higher eukaryotic organisms.

- 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



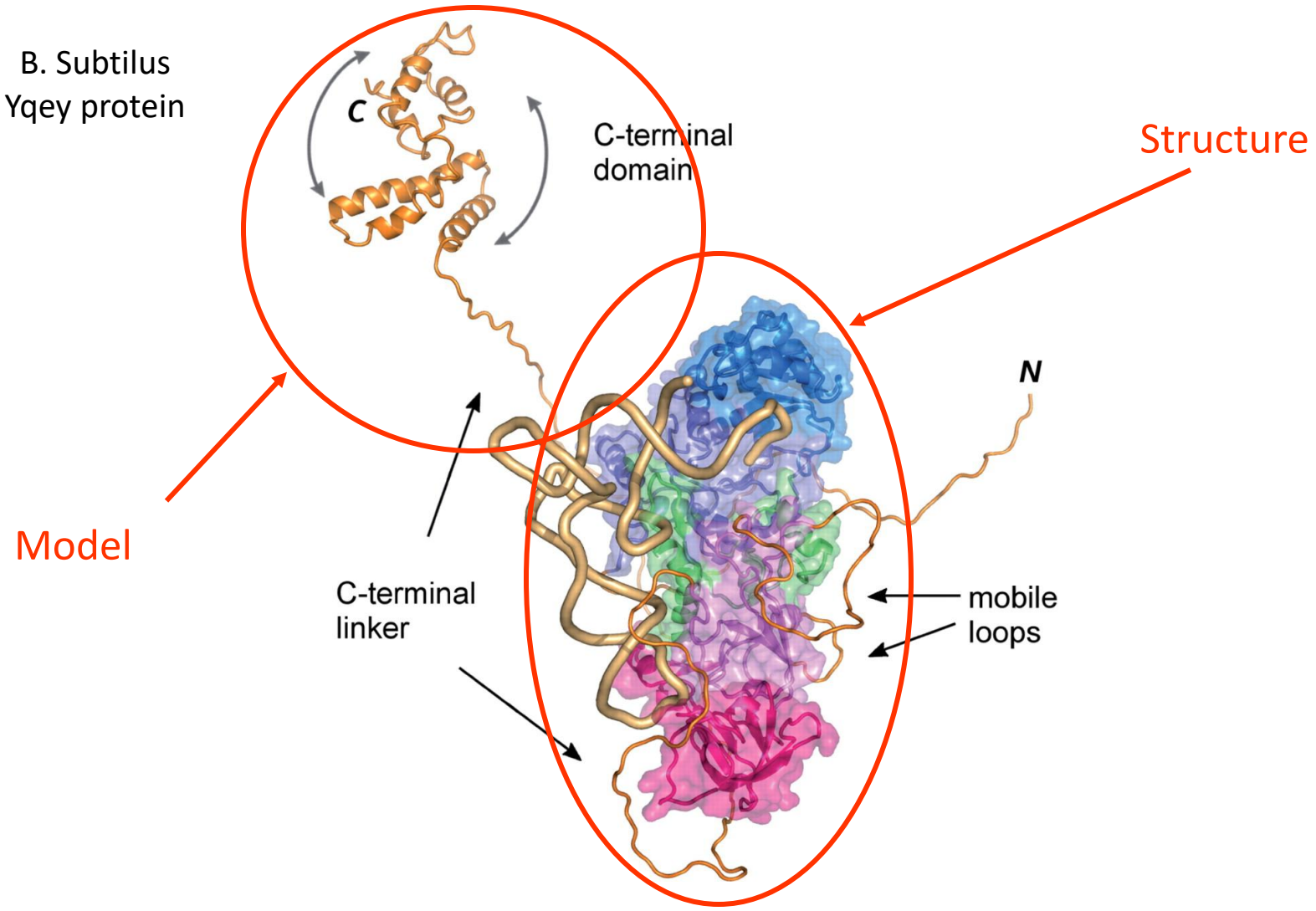
Structural model of *E. coli* glutaminyl-tRNA synthetase

These enzymes are not gentle with tRNA molecules. The enzyme firmly grips the anticodon, spreading the three bases widely apart for better recognition. At the other end, the enzyme unpairs one base at the beginning of the chain, seen curving upward here, and kinks the long acceptor end of the chain into a tight hairpin, seen here curving downward. This places the 2' hydroxyl on the last nucleotide in the active site, where ATP and the amino acid (not present in this structure) are bound.

Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. Rould, Perona, and Steitz **Journal:** (1991) Nature**352:** 213-218

Structures only known from *E.coli* and *D. radiodurans*

Model of *D. radiodurans* GlnRStRNA^{Gln} complex



Crystallography

Crystallization/Data collection

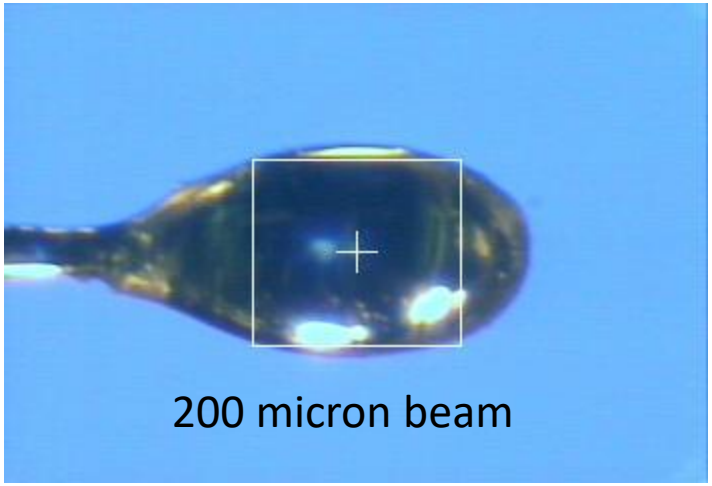
- Gln4 Screened against 1536 different biochemical conditions, ~1000 forming an incomplete factorial of chemical space and ~500 representing commercially available screens.
- Crystal leads seen, several were chosen based on ease of cryoprotection of the native hit.
- Crystals were optimized with a Drop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

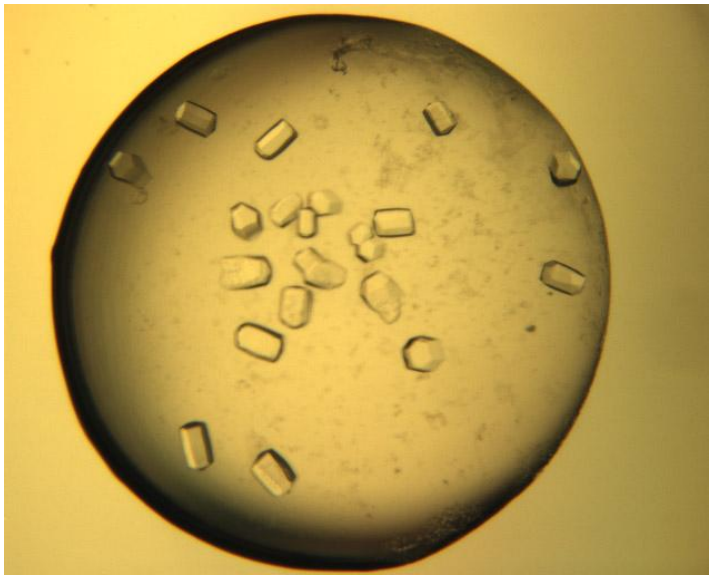
- Only 2 structures for related glutaminyl tRNA synthetases are available (~40% sequence homology), we had 228 extra residues (almost 40% more residues) therefore we expected problems in molecular replacement and didn't have a SeMet example.

- 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.

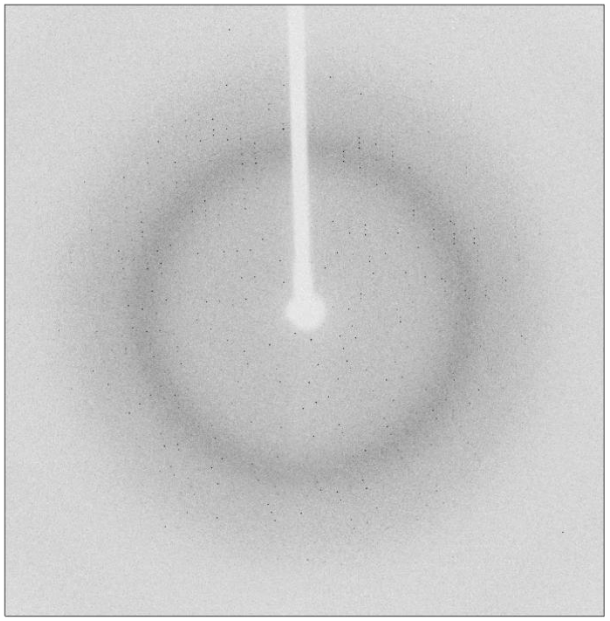
- We collected data remotely with an excitation scan to determine if Zinc was present.

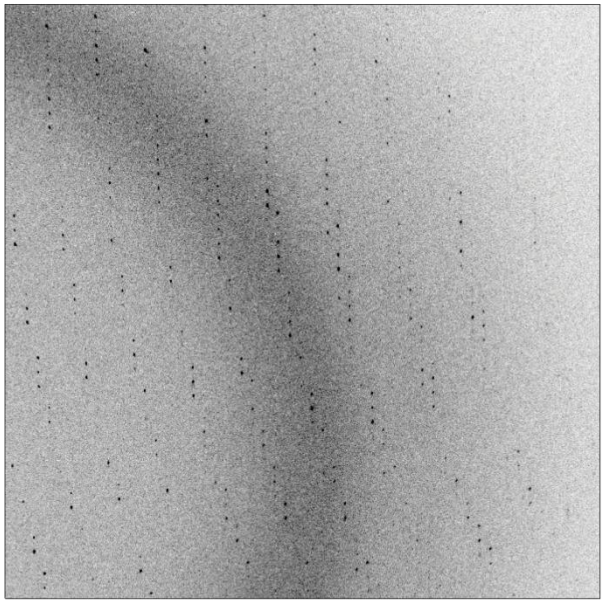
- It was!





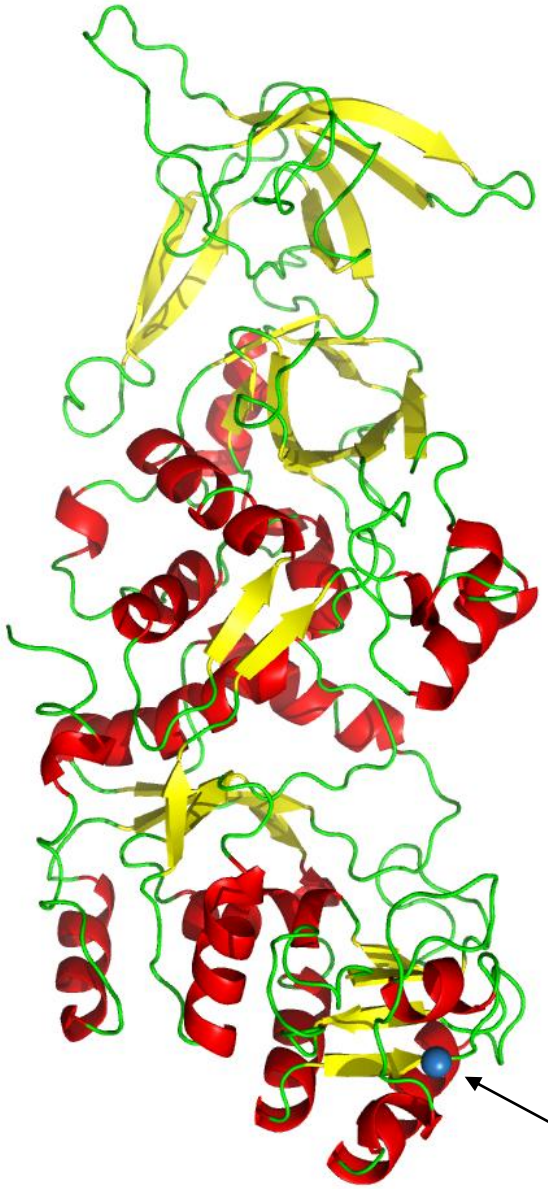
80% PEG 400 in the
crystallization cocktail



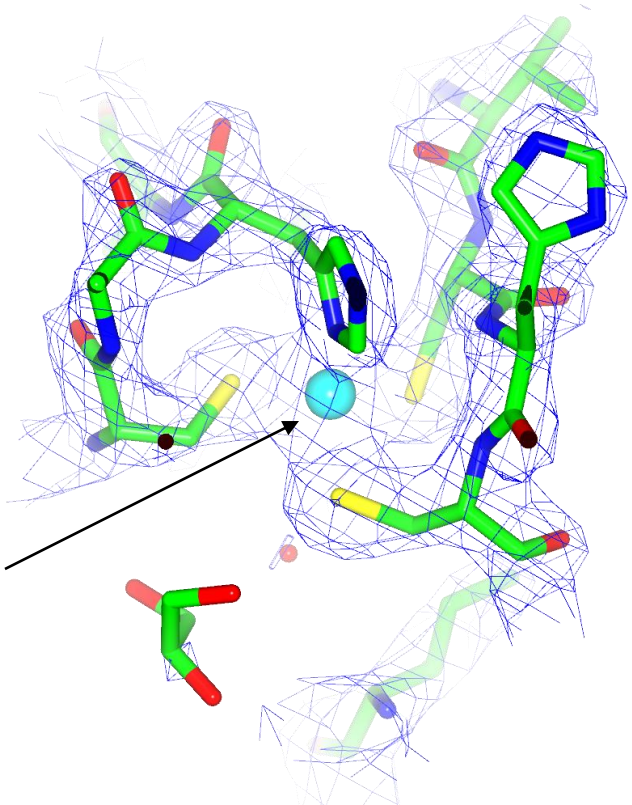


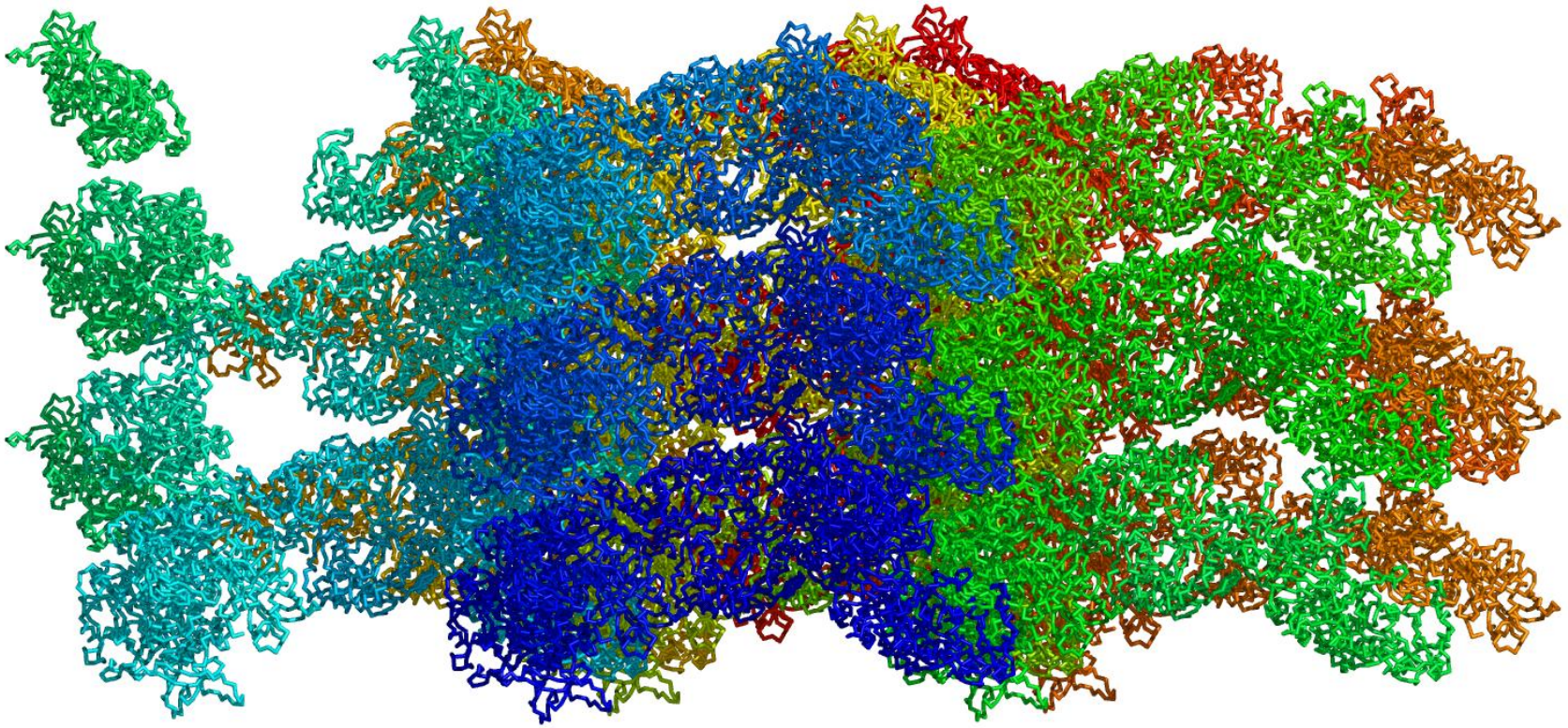
	*Sc*GlnRS
Data collection	
Beamline	SSRL BL 11-1
Wavelength (Å)	1.169
Space group	P 3_1 2 1
Cell dimensions	
a, *b*, *c* (Å)	176.611, 176.611, 72.1884
α, β, γ (°)	90, 90, 120
Resolution (Å) *	52.49 – 2.15 (2.23 – 2.15)
R_{sym} or R_{merge} *	0.068 (0.348)
Completeness (%) *	99.86 (99.84)
$I/\sigma I$ *	23.26 (2.98)
Unique reflections *	70276 (6963)
Redundancy *	11.2 (4.5)
Wilson B-factor (Å^2)	33.55
Refinement	
Resolution (Å)	52.49 – 2.15
R_{work}/ R_{free} *	0.1633/0.1826 (0.2232/0.2514)
No. atoms	10537
Protein	5043
Ligand/ion	75
Water	449
B-factors (Å^2)	
Protein	40.40
Ligand/ion	34.47
Water	44.90
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (º)	0.90
Ramachandran favored (%)	98.0
Ramachandran outliers (%)	0.17
Clashscore	6.55

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

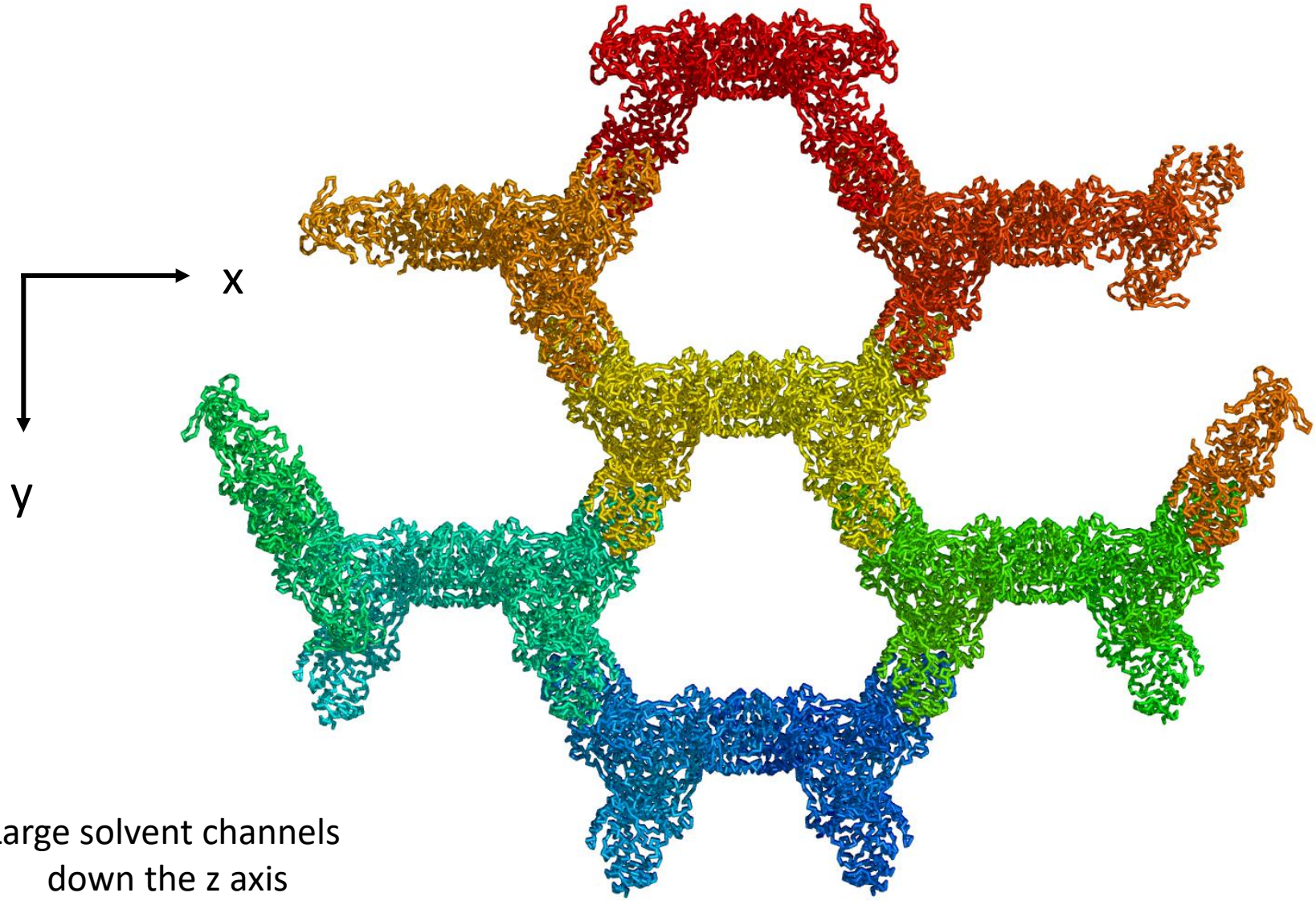


Zinc

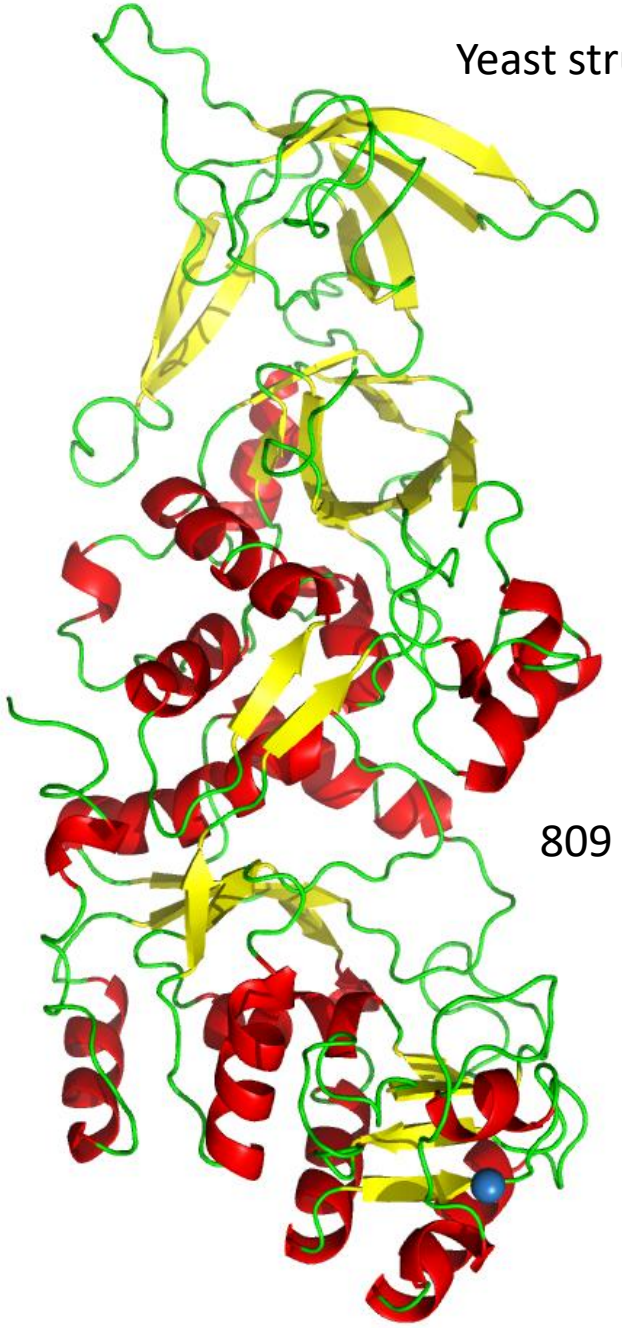




Tight packing in z and y



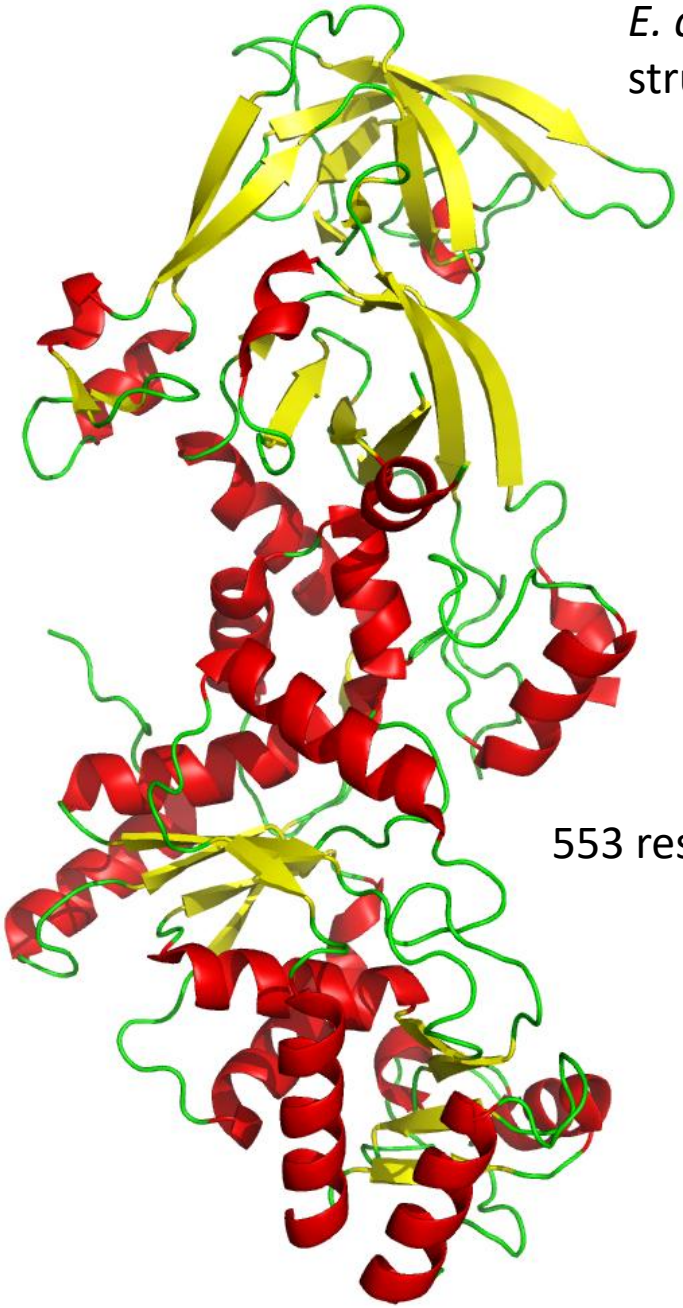
Large solvent channels
down the z axis



Yeast structure

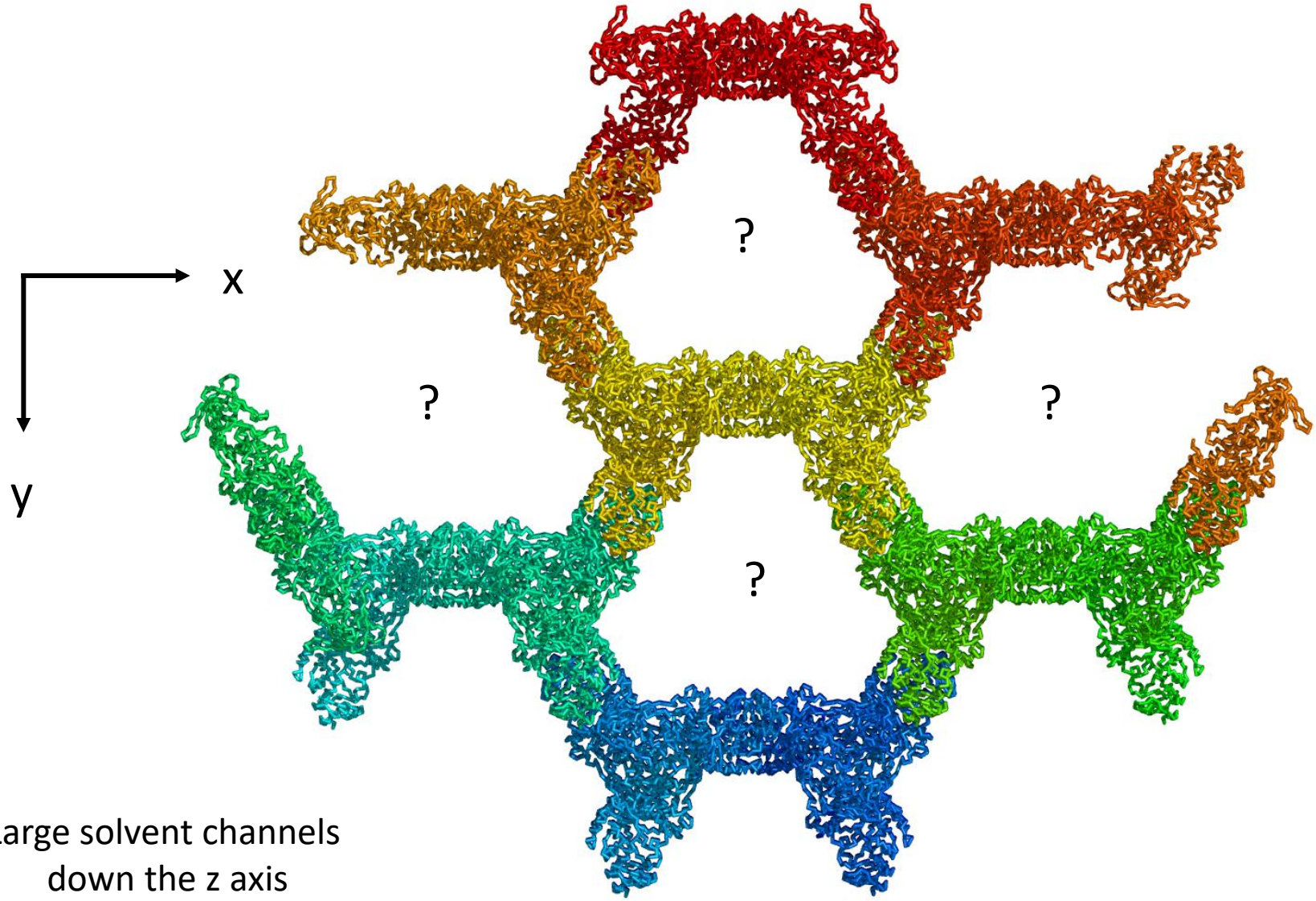
?

809 residues

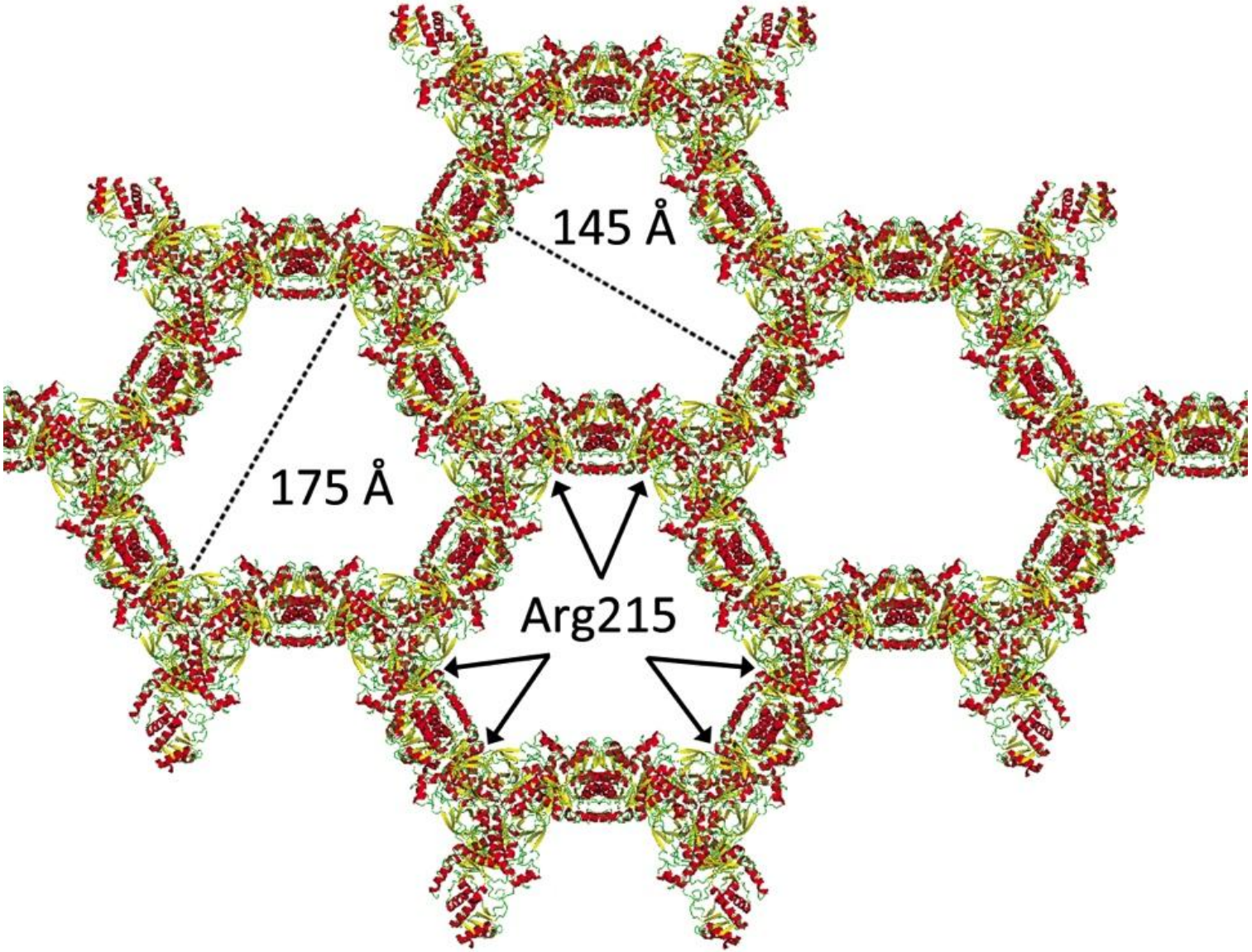


E. coli. structure

553 residues

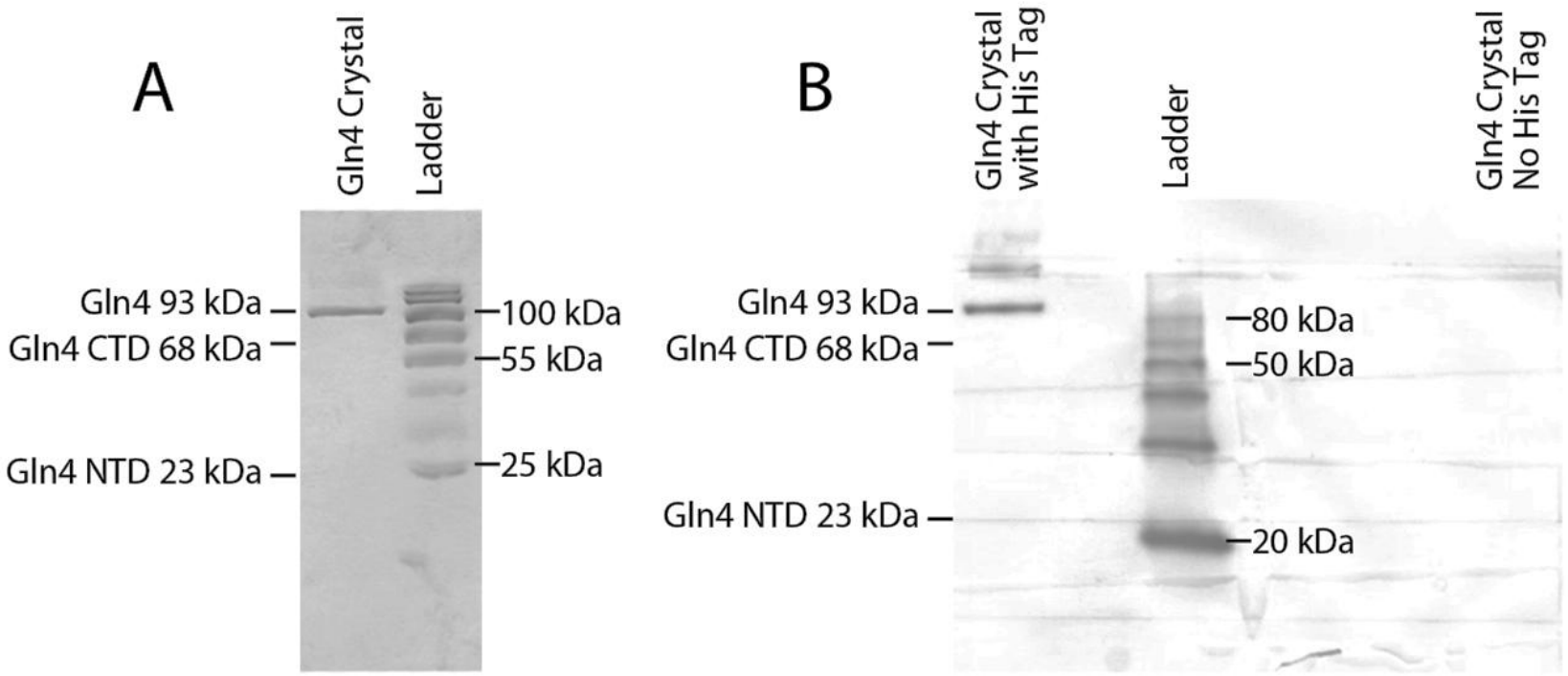


Large solvent channels
down the z axis

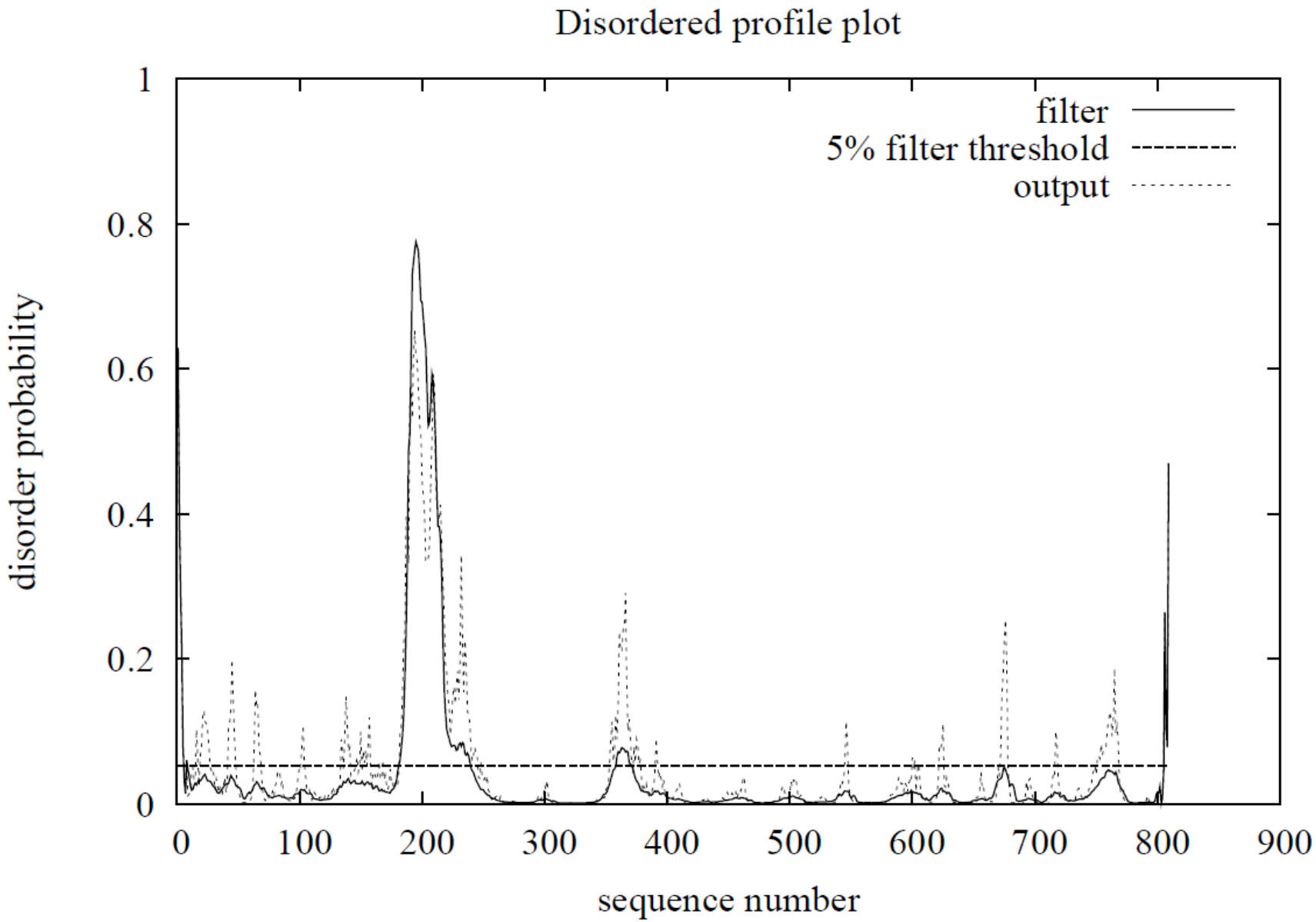

145 Å
175 Å
Arg215

Missing residues

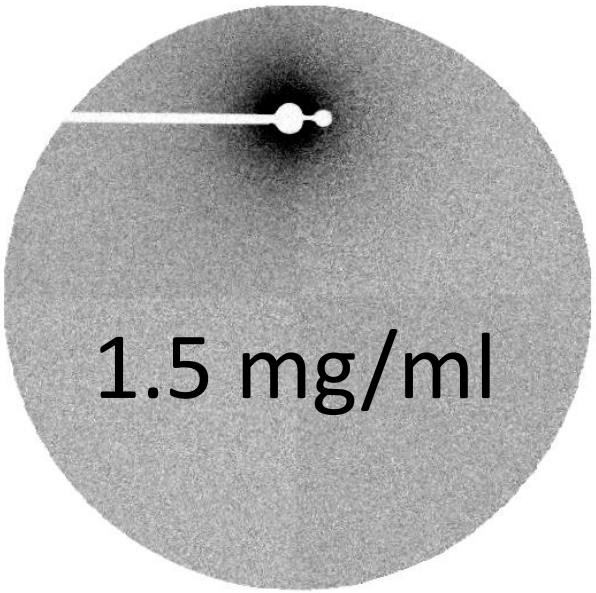
- There were 216 missing residues from the structure, 95% of the N-terminal domain.
- Where they in the mix to start with?
- SDS PAGE gel on the remaining crystals indicated that the full length protein was present.
- For a more concrete answer the protein was re-expressed with a His tag attached to 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.

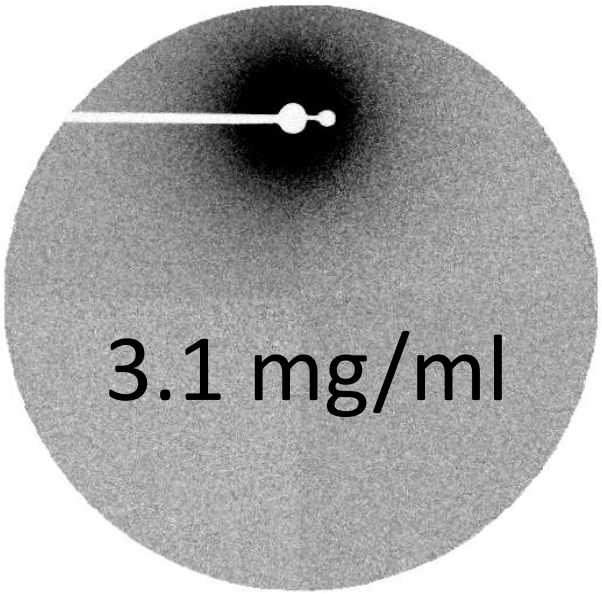


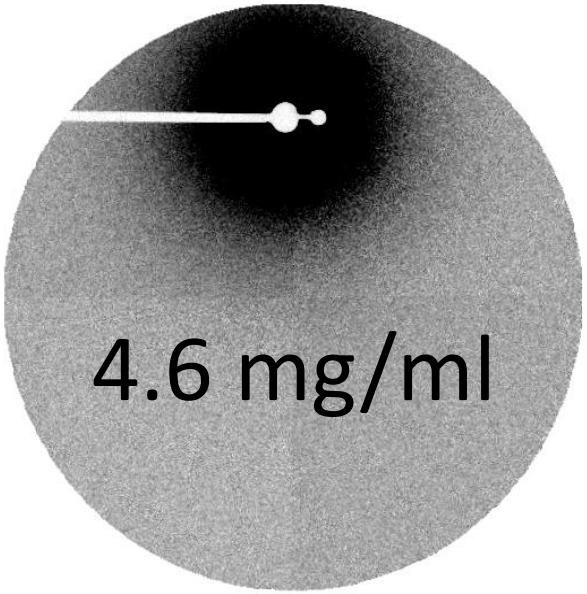
A. SDS PAGE gel showing dissolved Gln4 protein crystals is shown in the left lane, and the molecular weight ladder is shown in the right lane. Labels for the full-length protein, and both the NTD and CTD fragments are given. The presence of full-length Gln4 and absence of NTD and CTD fragments indicates that only the full-length protein is present in the crystal. B. Western blot using an anti-His antibody for crystals containing both His-tagged (left-most lane) and non-His-tagged (right-most lane) Gln4 protein. The molecular weight ladder is shown in the middle lane.

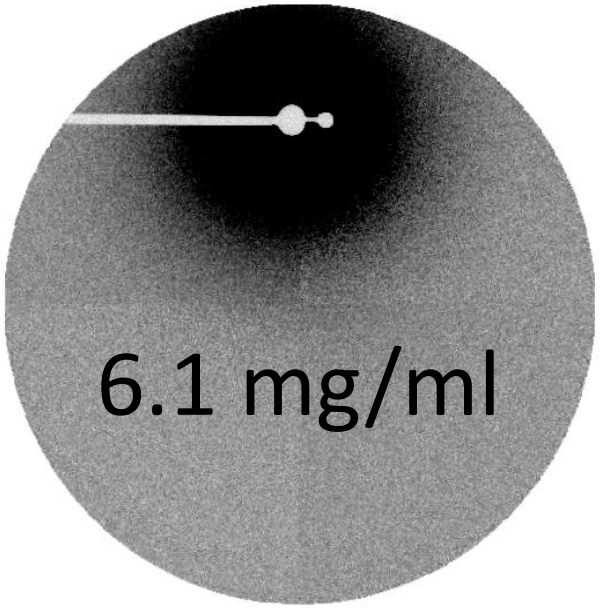

Disordered profile plot
filter
5% filter threshold
output
disorder probability
sequence number
0
0.2
0.4
0.6
0.8
1
0
100
200
300
400
500
600
700
800
900

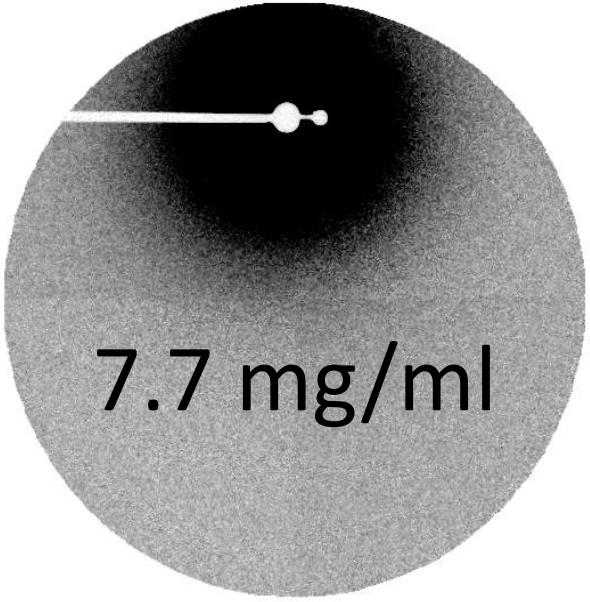
Back to SAXS

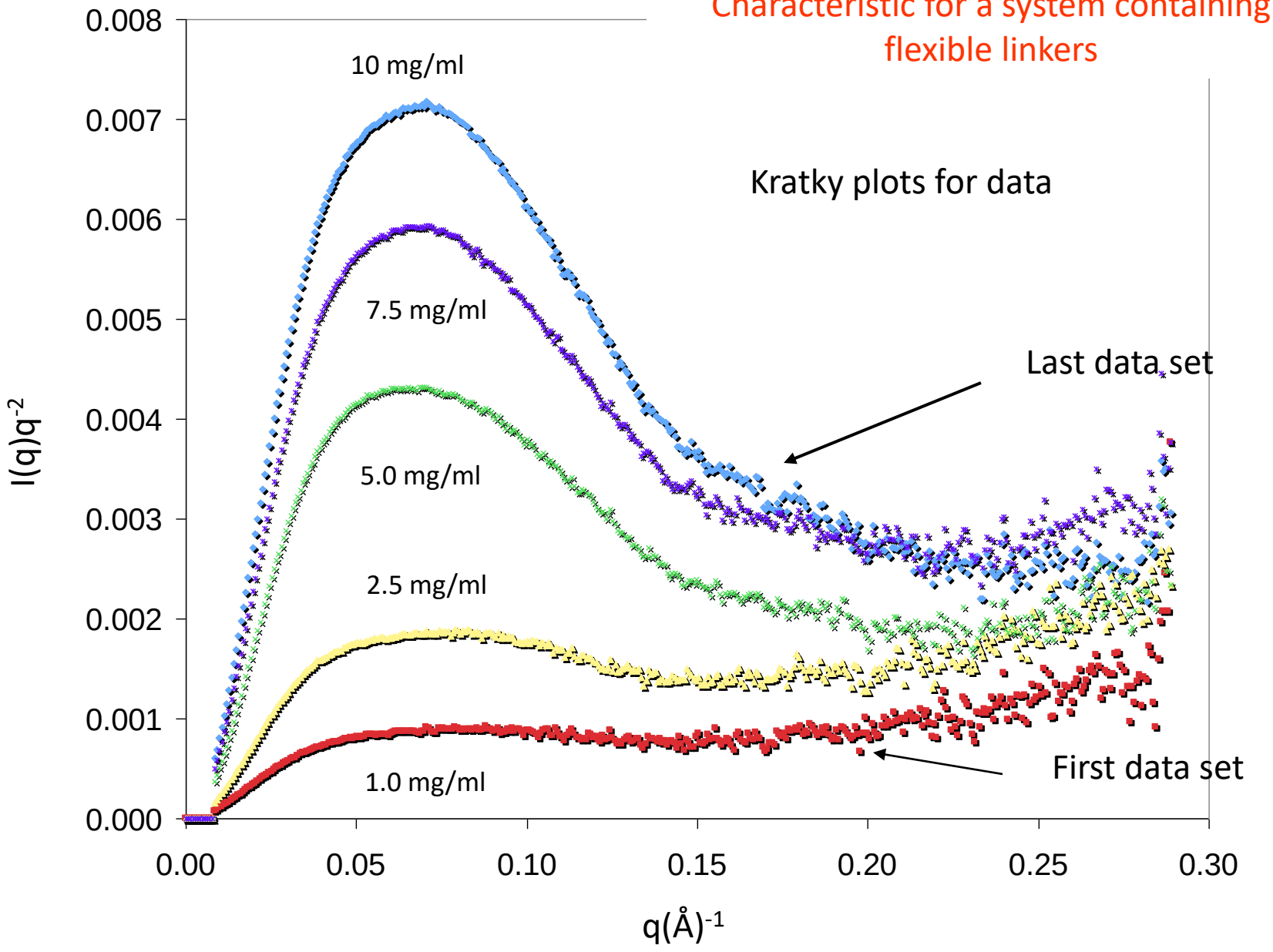

1.5 mg/ml


3.1 mg/ml

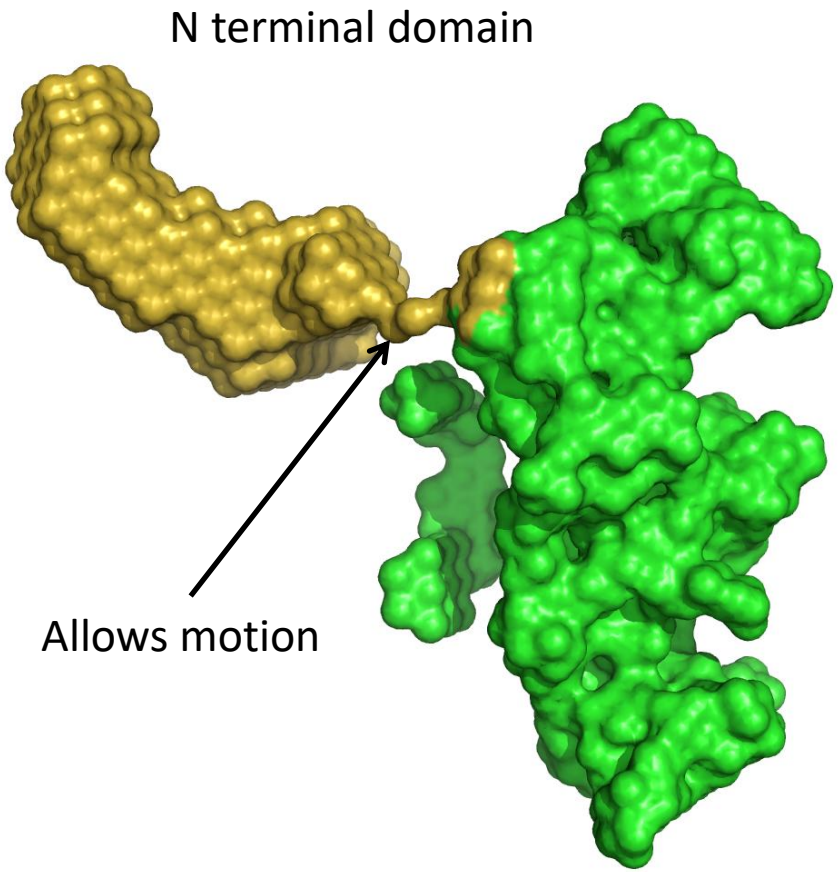

4.6 mg/ml


6.1 mg/ml

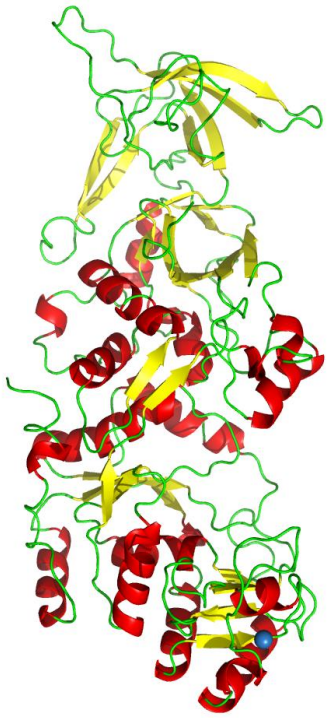

7.7 mg/ml


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

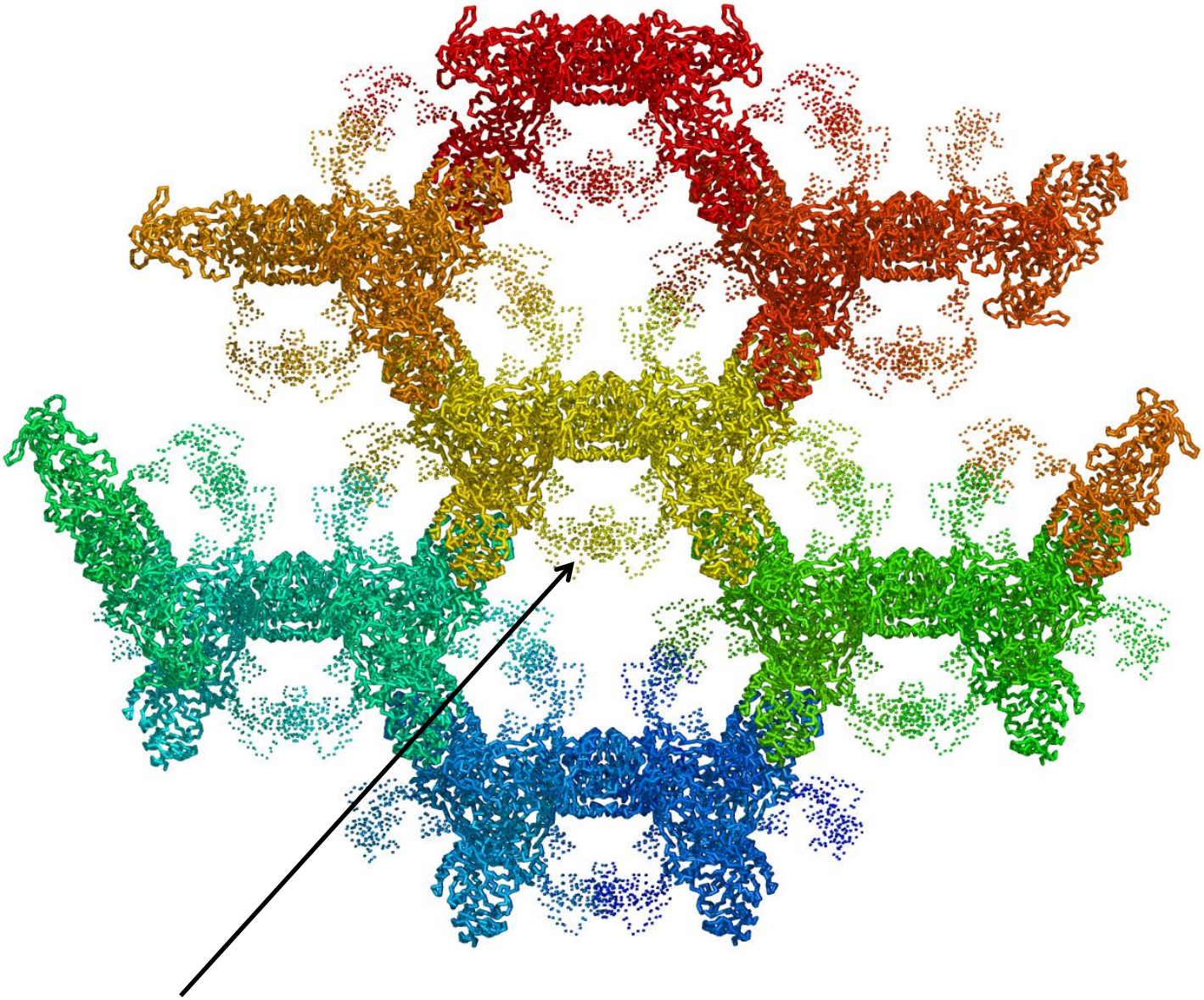
Envelope reconstruction using the crystallographic structure



C terminal domain



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



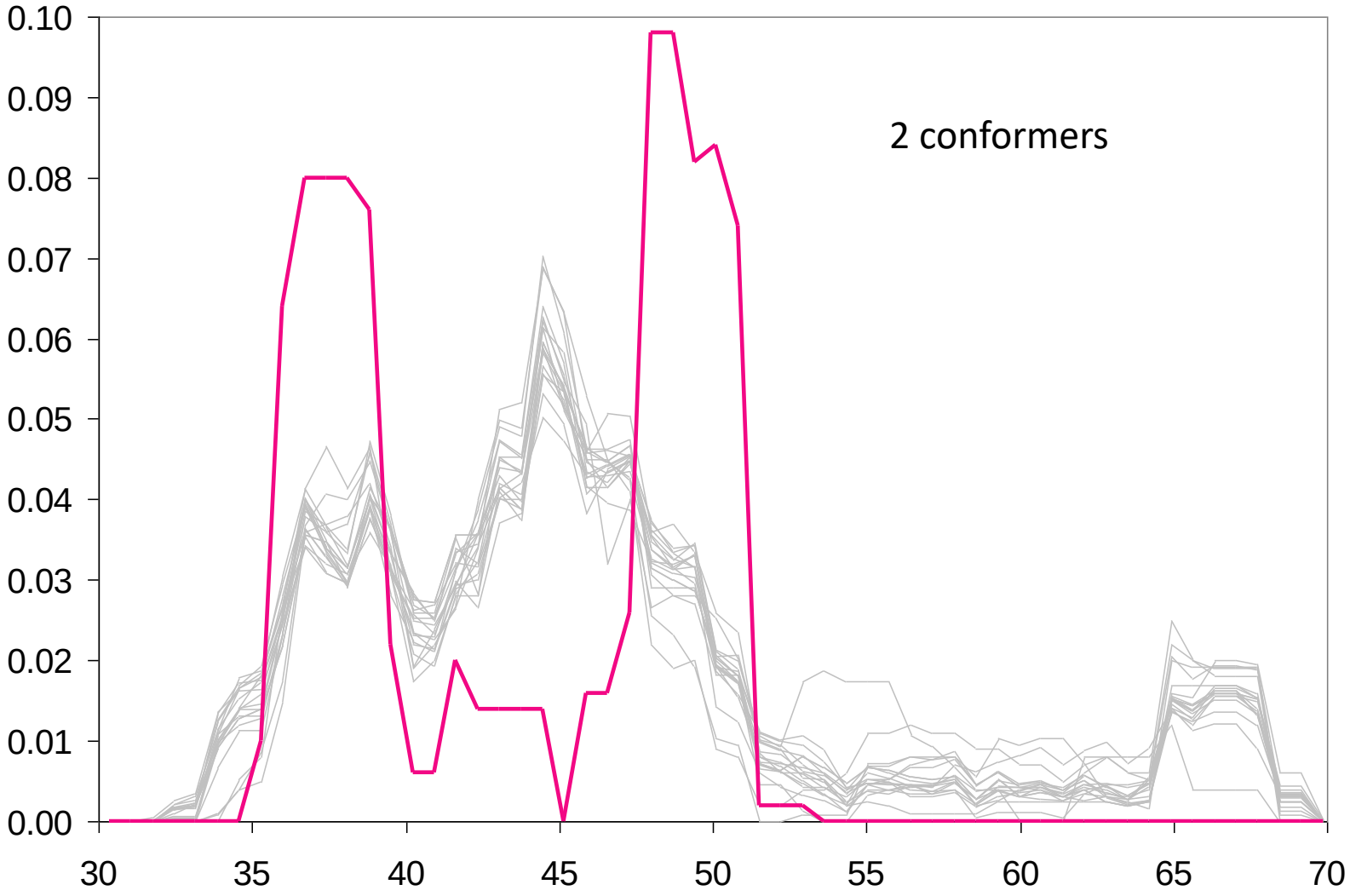
The N-terminal ‘arm’ is completely compatible with the crystal structure

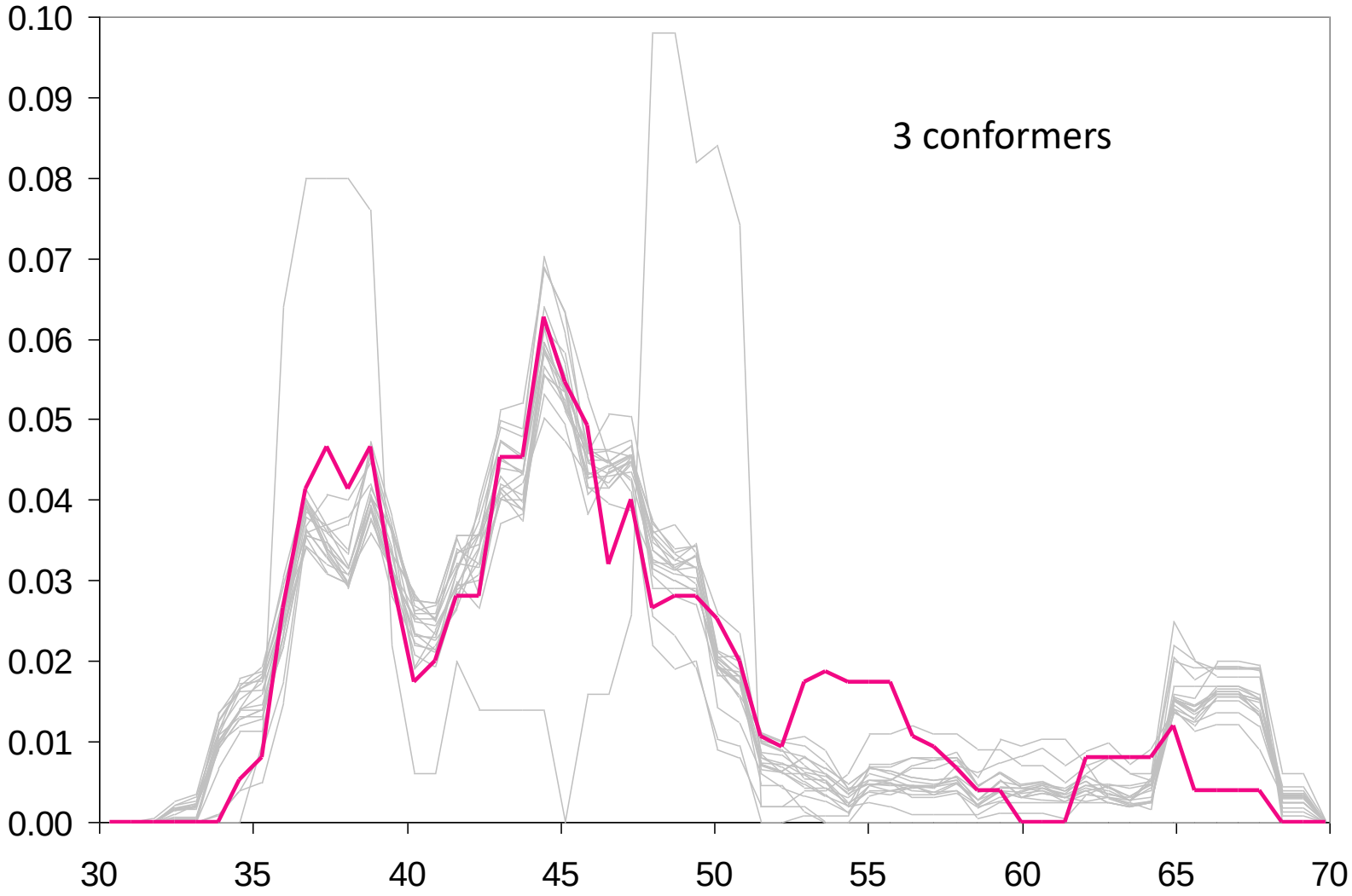
Wild but exciting Goose chase

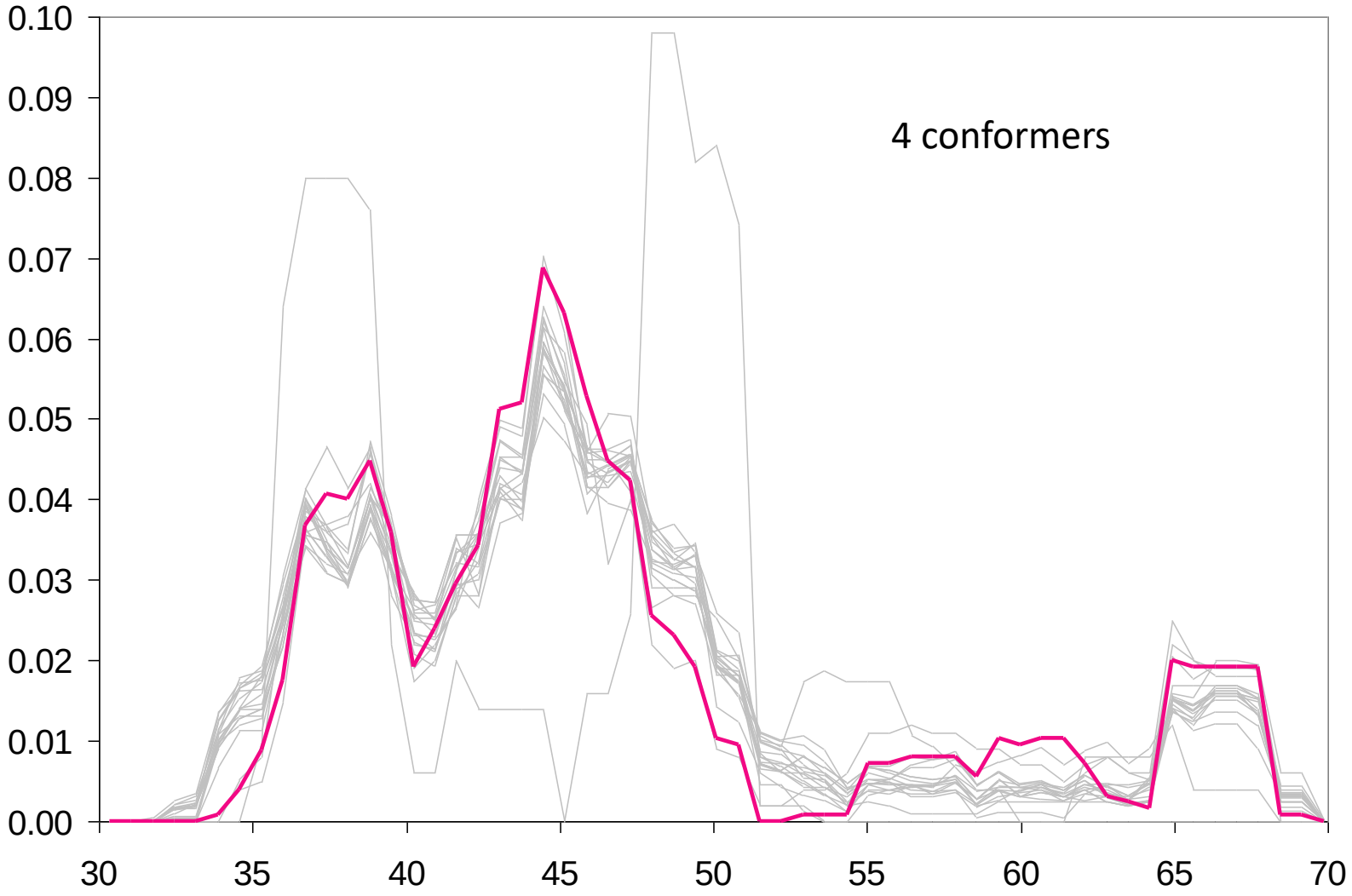
Ensemble optimization

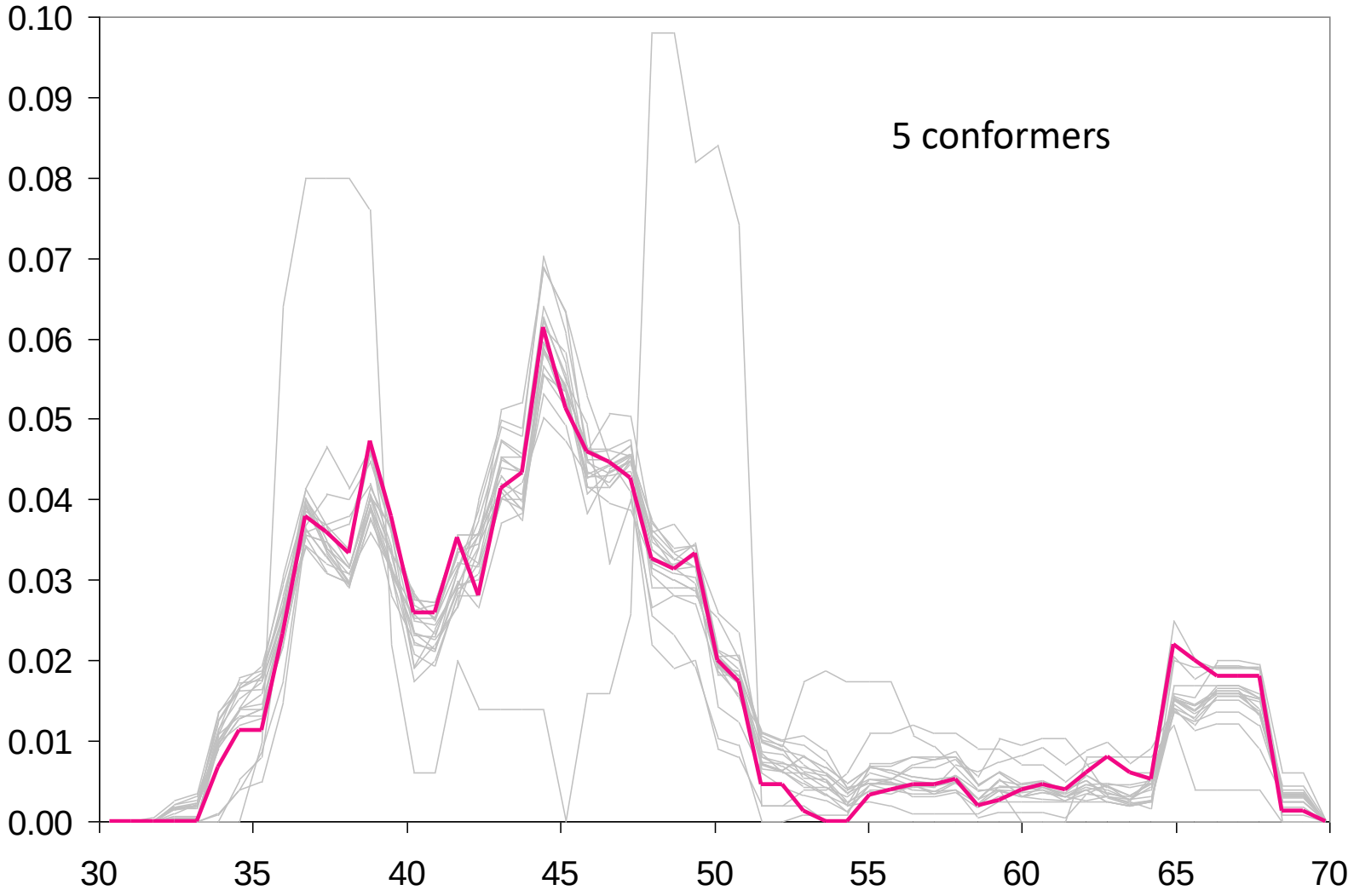
- The Ensemble Optimization Method (EOM) was used to assess the flexibility of the Gln4 N-terminal domain.
- RanCh (**Ran**dom **Ch**ain Generator) generated 10,000 conformers of the N-terminal sequence of Gln4 covering all possible configuration space.
- Sets of these conformers were binned to create ensembles.
- GAJOE (**G**enetic **A**lgorithm **J**udging **O**ptimization of **E**nsembles) optimized the ensembles by comparing the average scattering profile of their conformers to the experimental data.
- Plotting the *Rg* distribution for successive runs, each using an increasing number of conformers per ensemble, allows us to identify the optimal number of conformers that most accurately characterizes the system.
- Analysis of chi (an error indicator) shows an systematic decrease, converging at eight conformers in each ensemble.

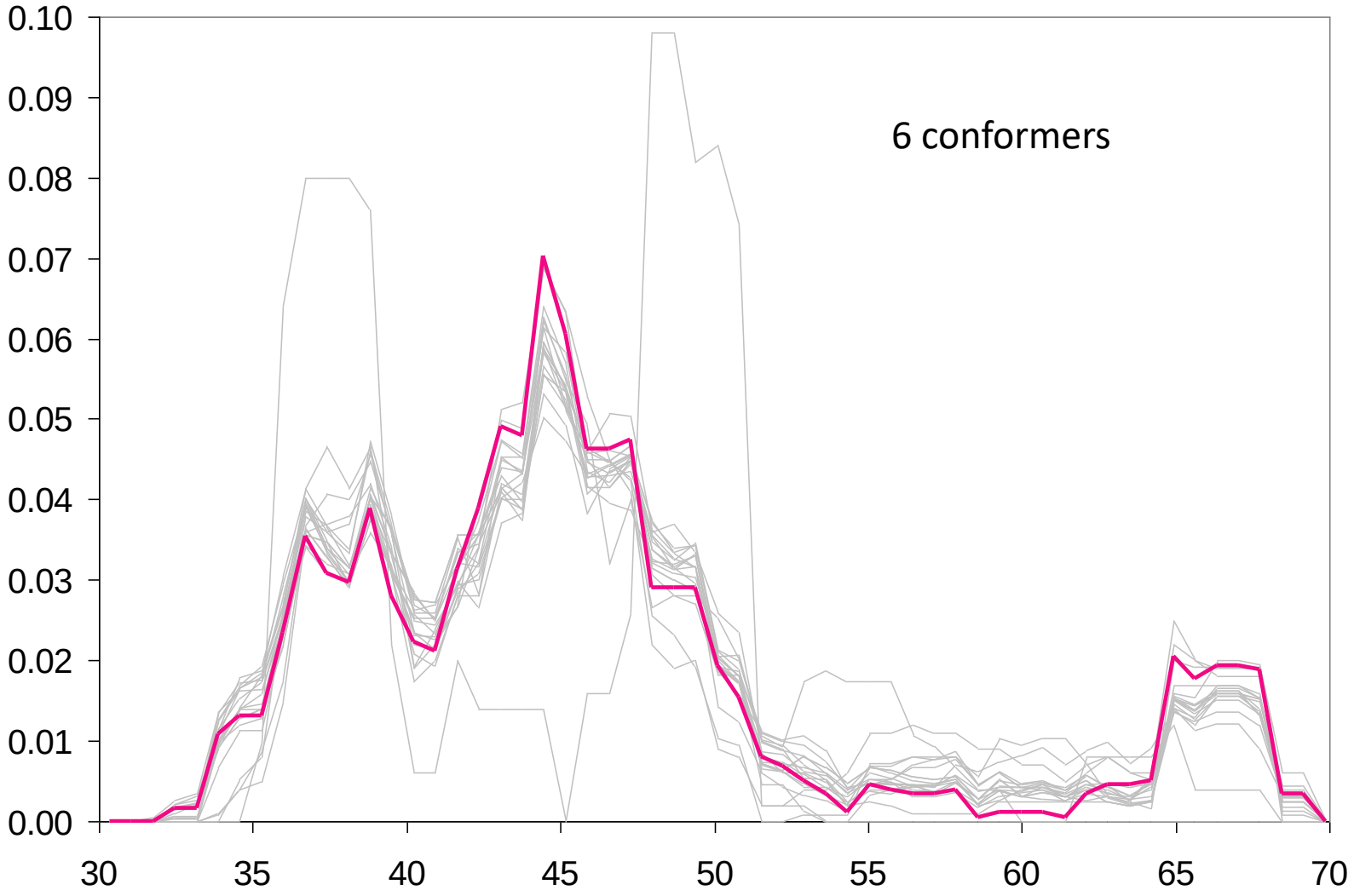
The convergence of the population distribution on distinct populations indicates that dynamic motion or different species are present when this is not the case the distribution is monomodal (confirmed by similar analyses on static systems).

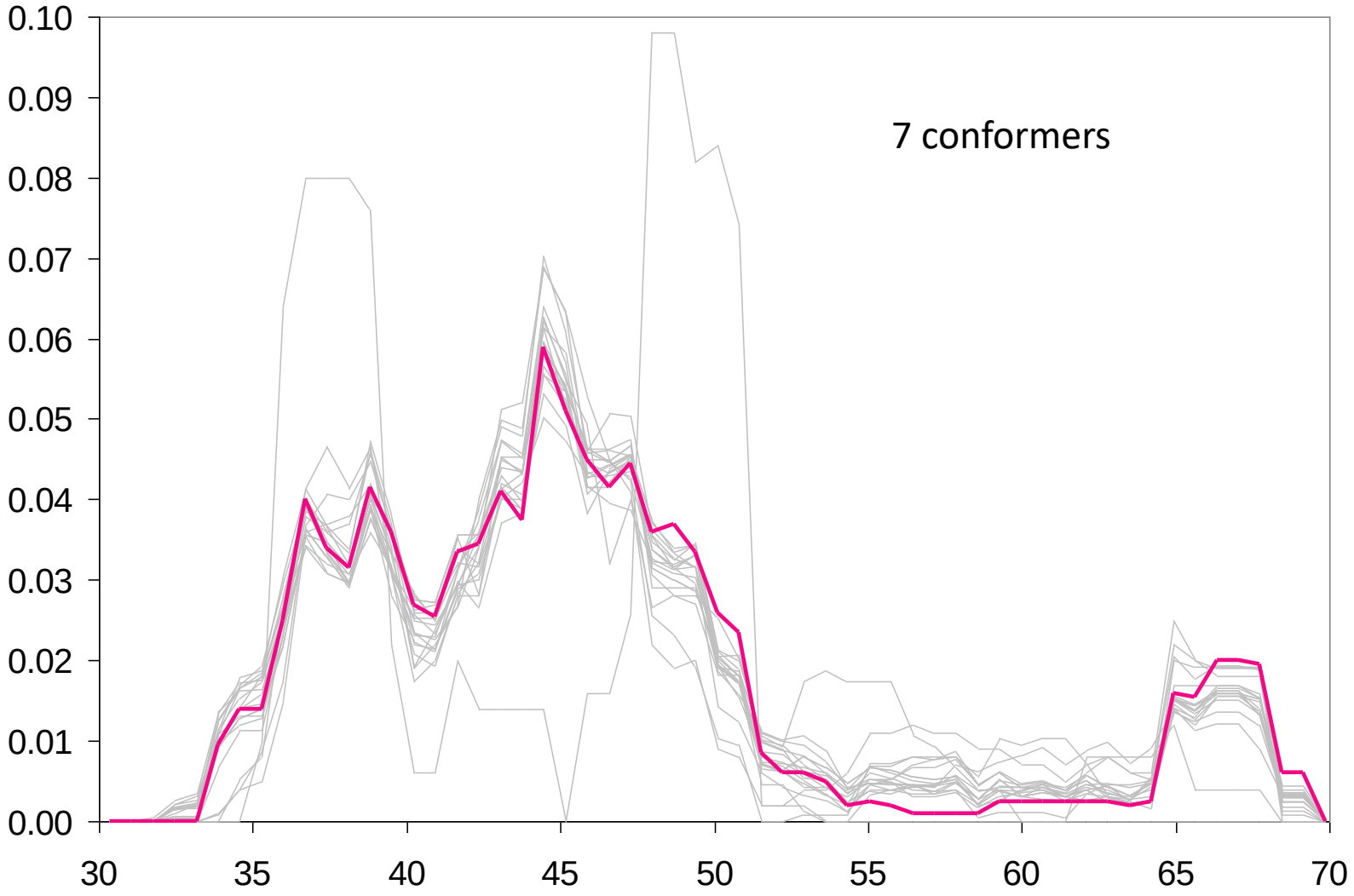

0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70
2 conformers

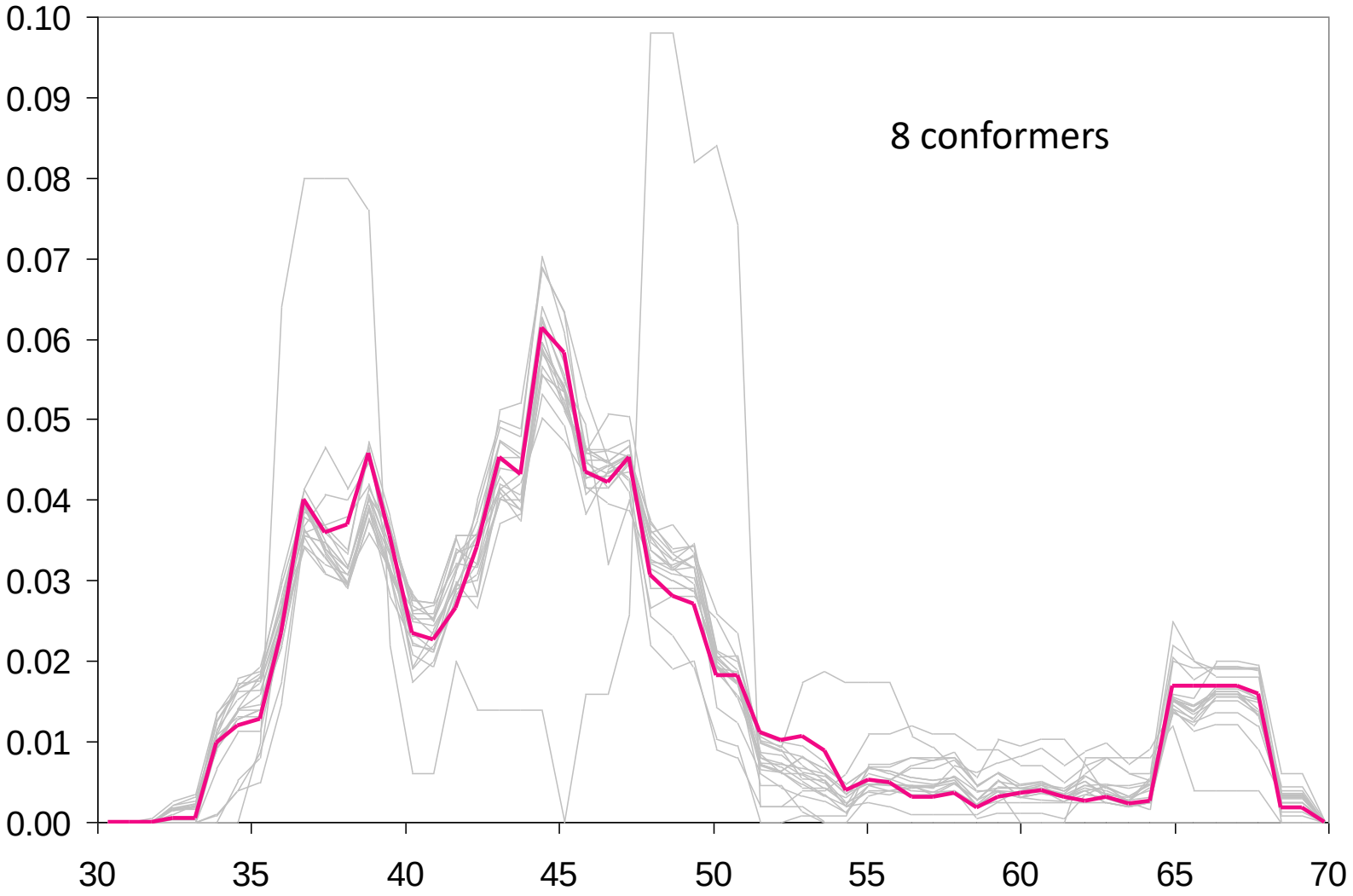

3 conformers
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70

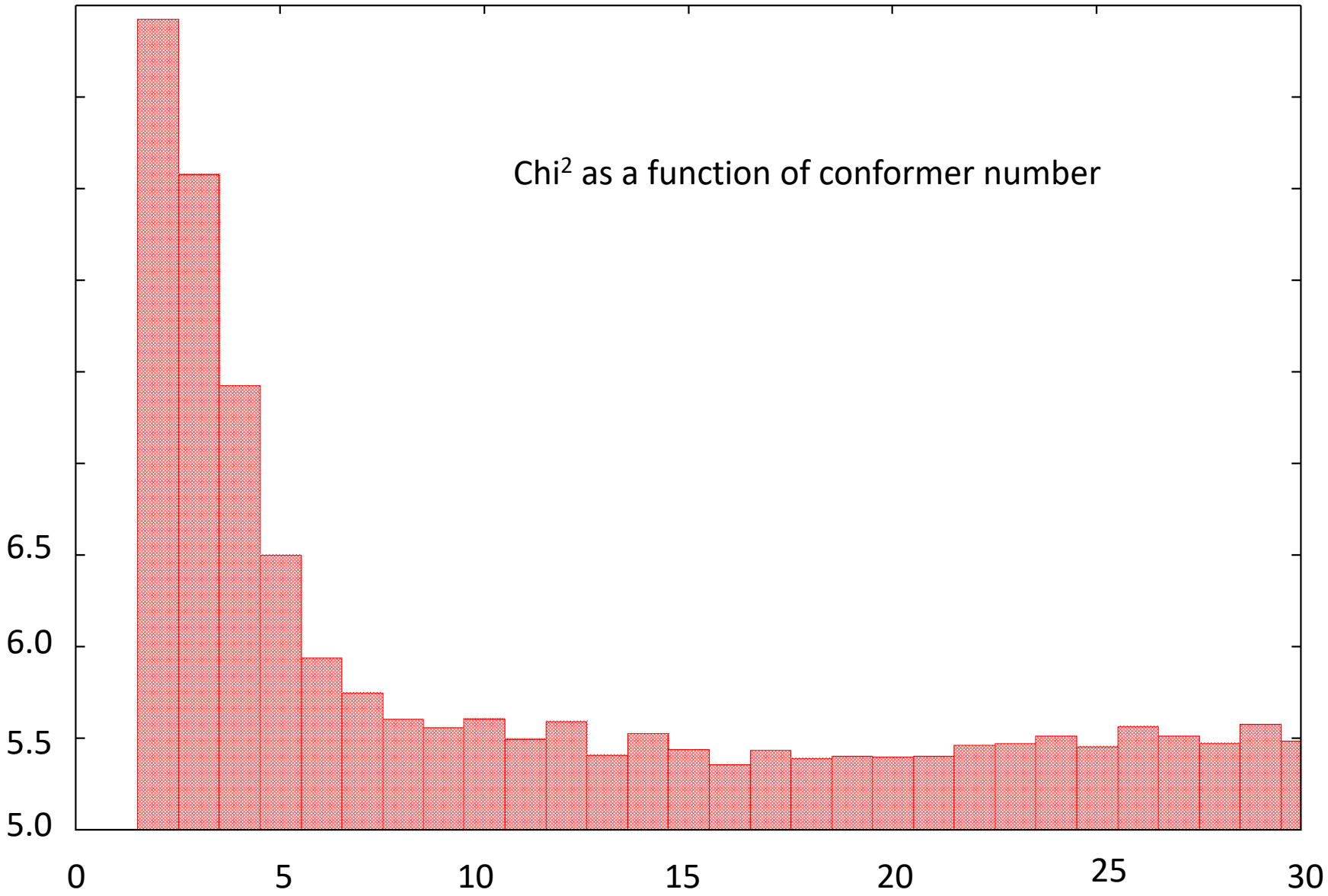

4 conformers
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70


5 conformers
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70


6 conformers
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70

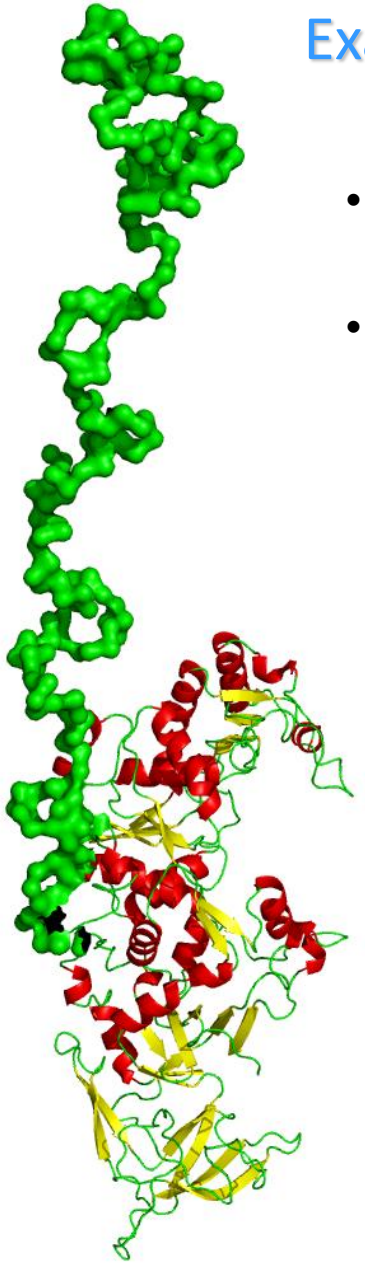

0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70
7 conformers

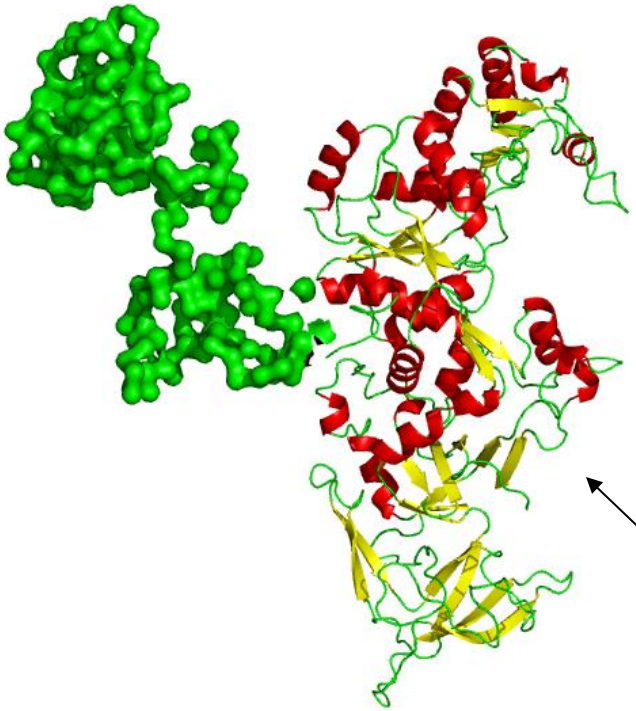

8 conformers
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
35
40
45
50
55
60
65
70

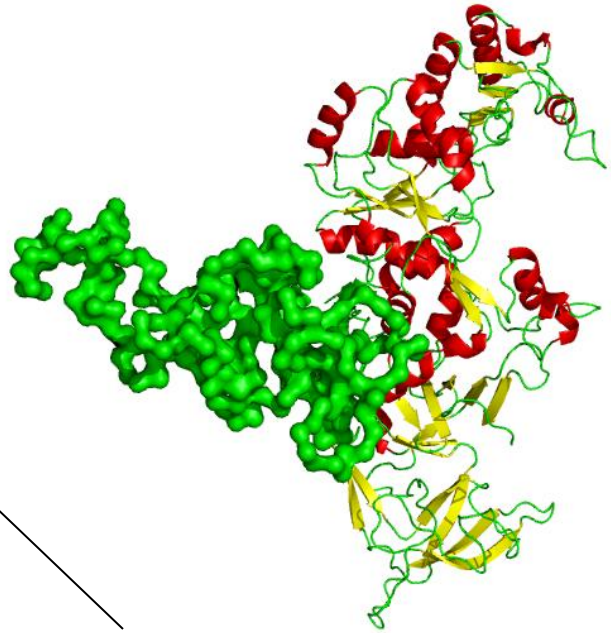

Chi2 as a function of conformer number
6.5
6.0
5.5
5.0
0
5
10
15
20
25
30

Example Conformers from the Ensemble optimization

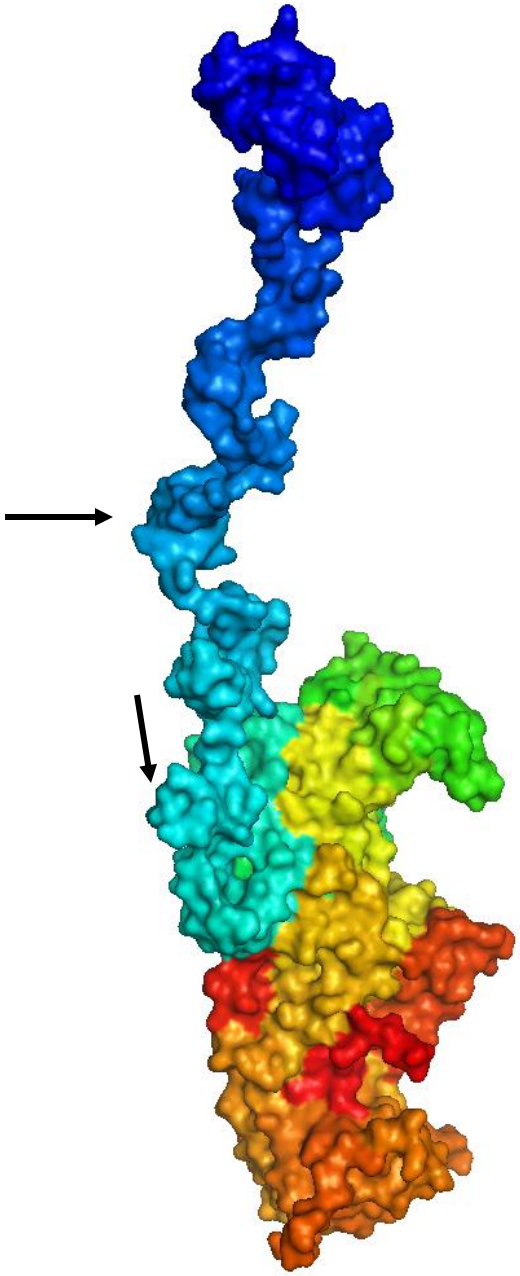
- Ensemble optimization told us that the SAXS data could be best explained with a minimum of 8 different conformers.
- The single *ab initio* model produced by traditional techniques represents the average conformation in solution.

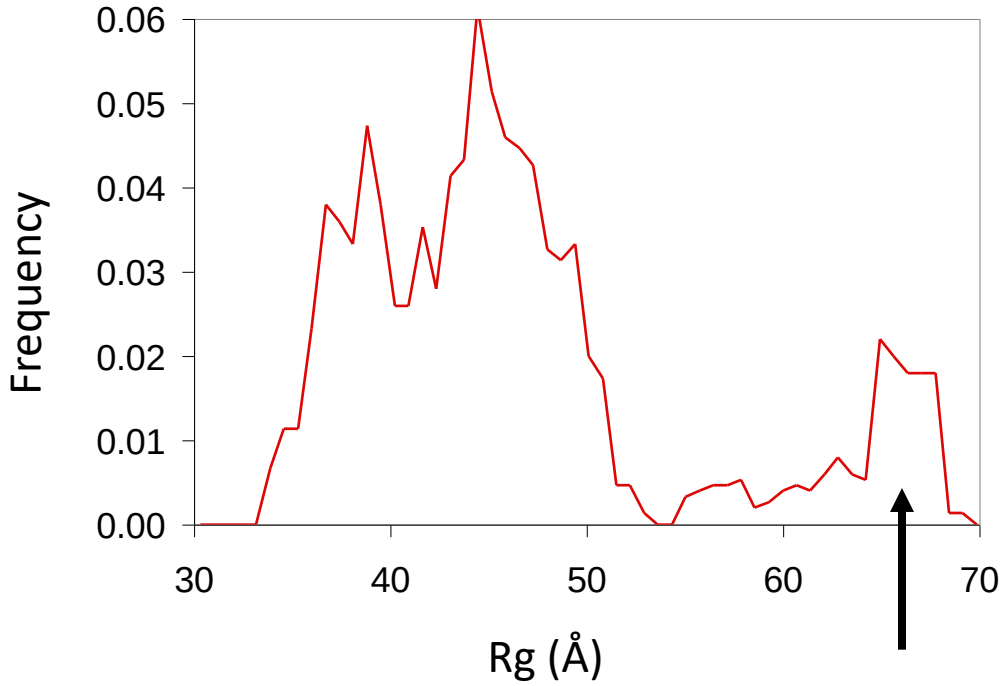


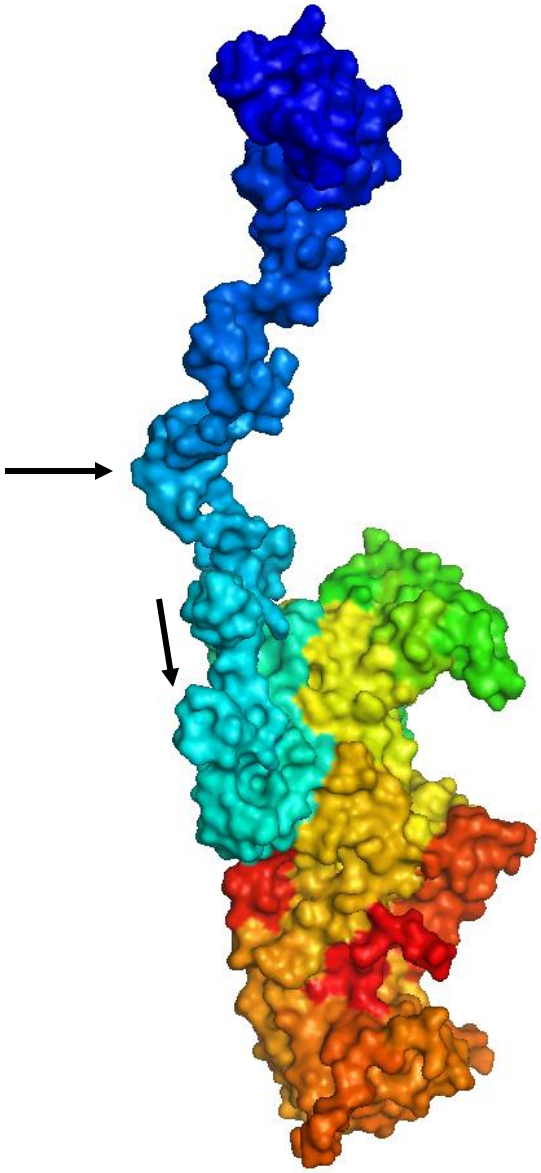


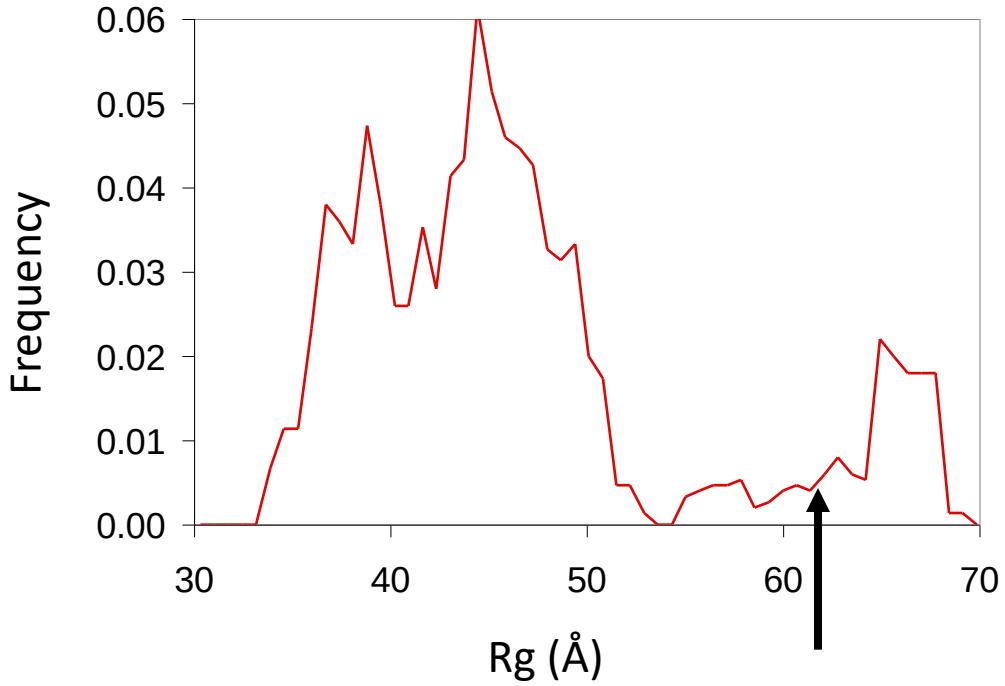


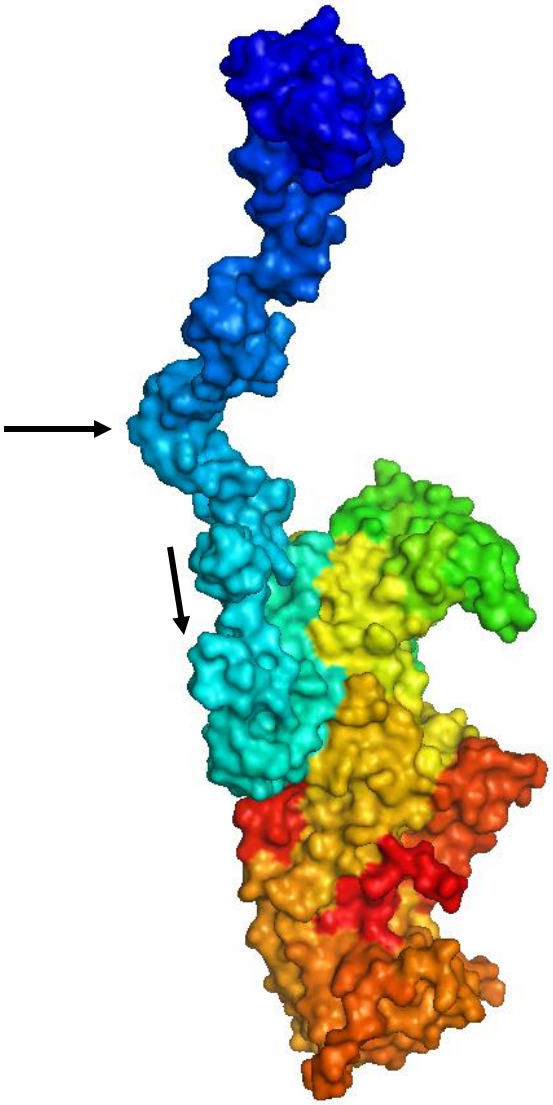
Crystallographic structure used

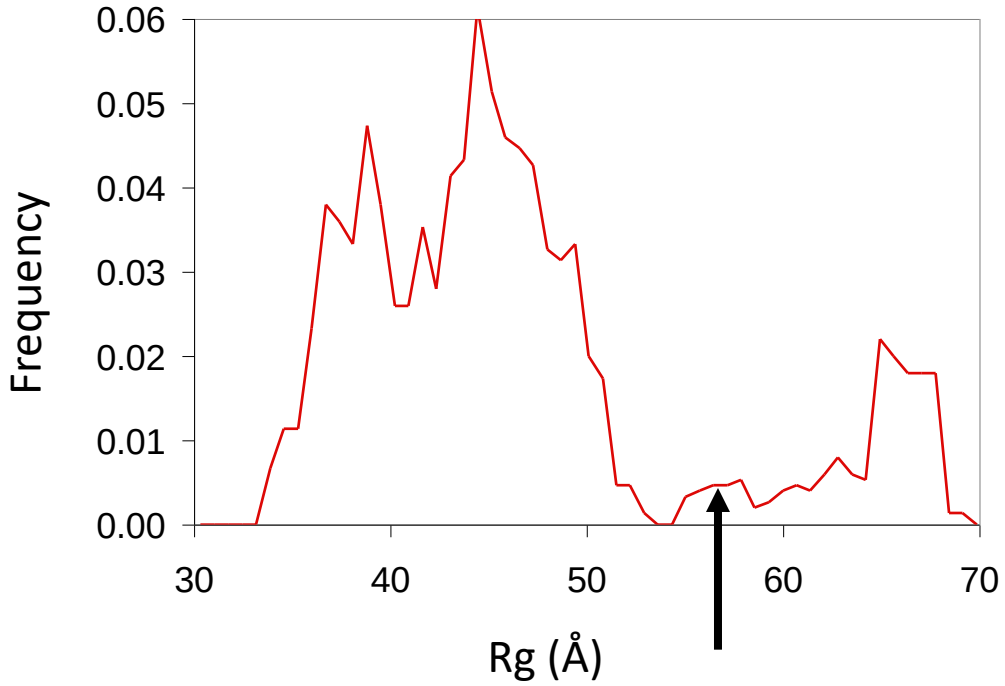


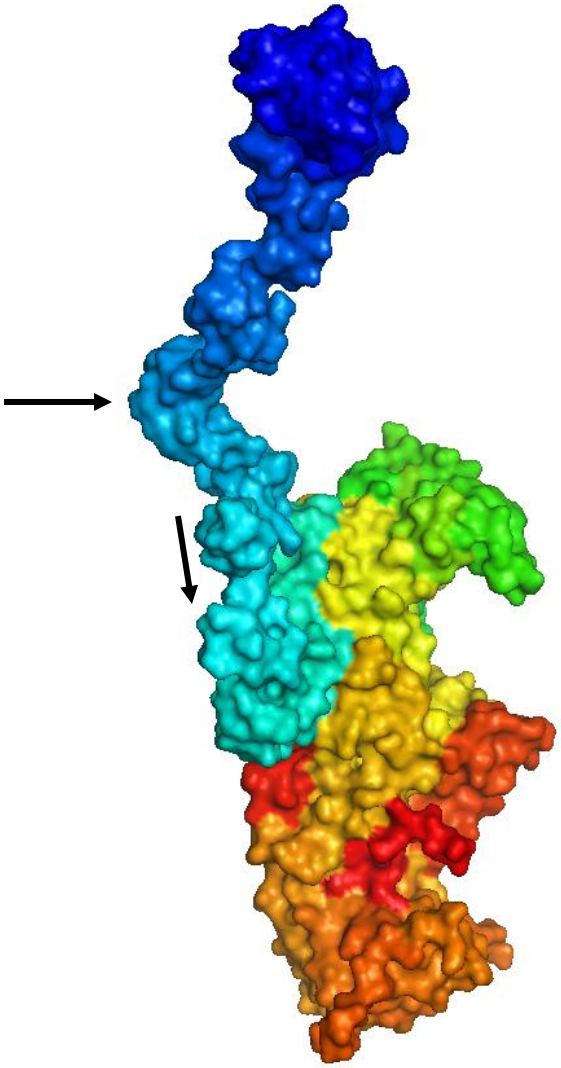

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

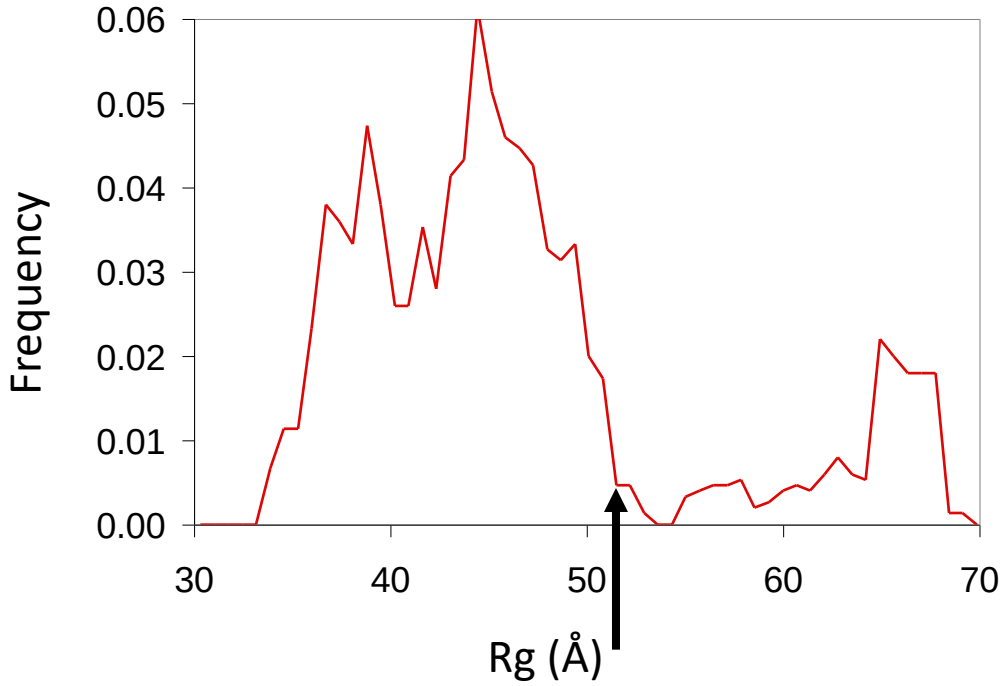


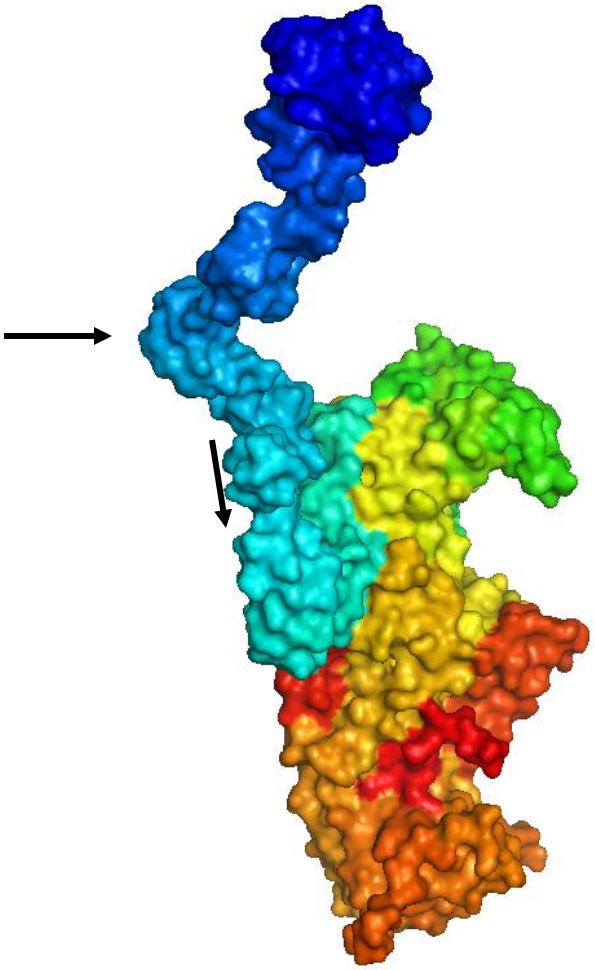

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

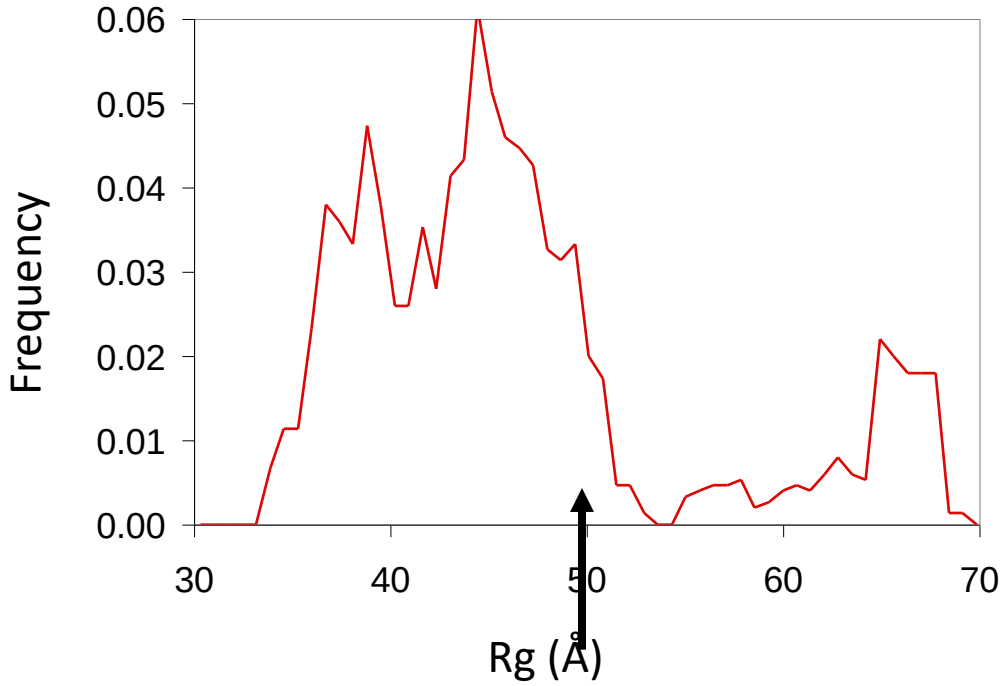


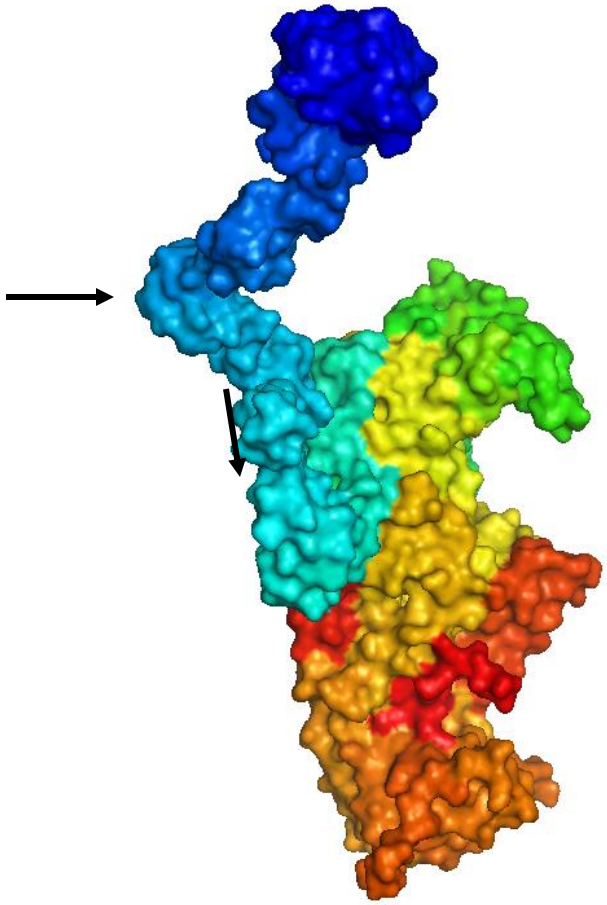

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

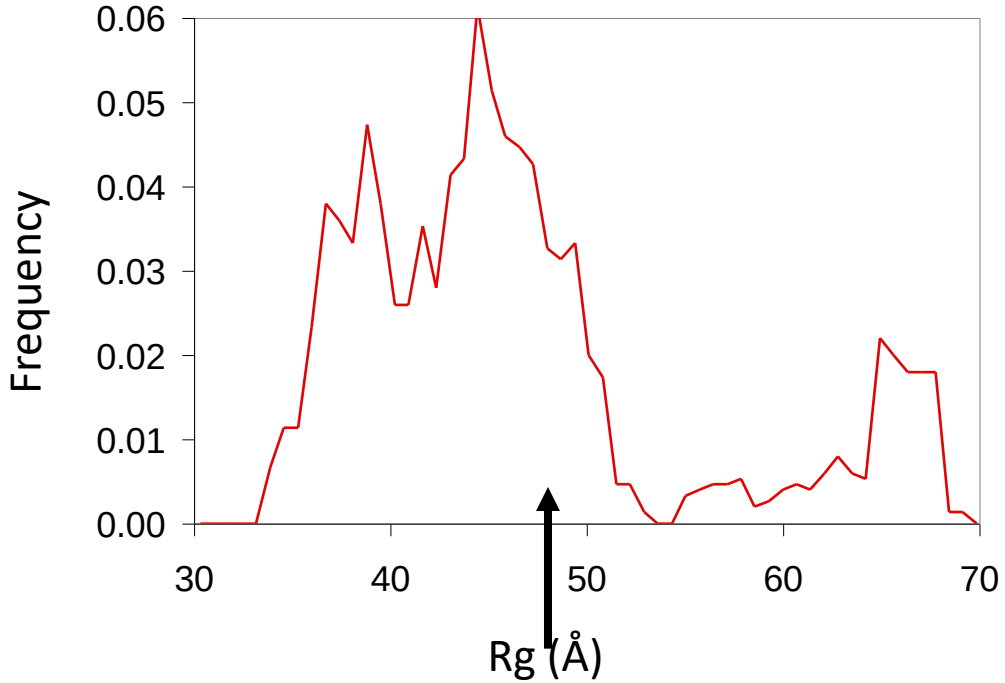


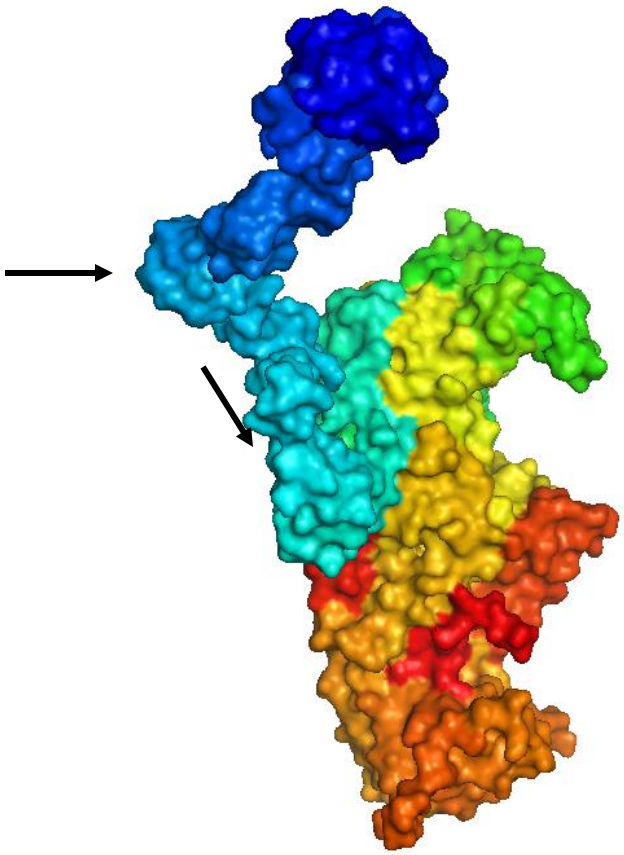

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

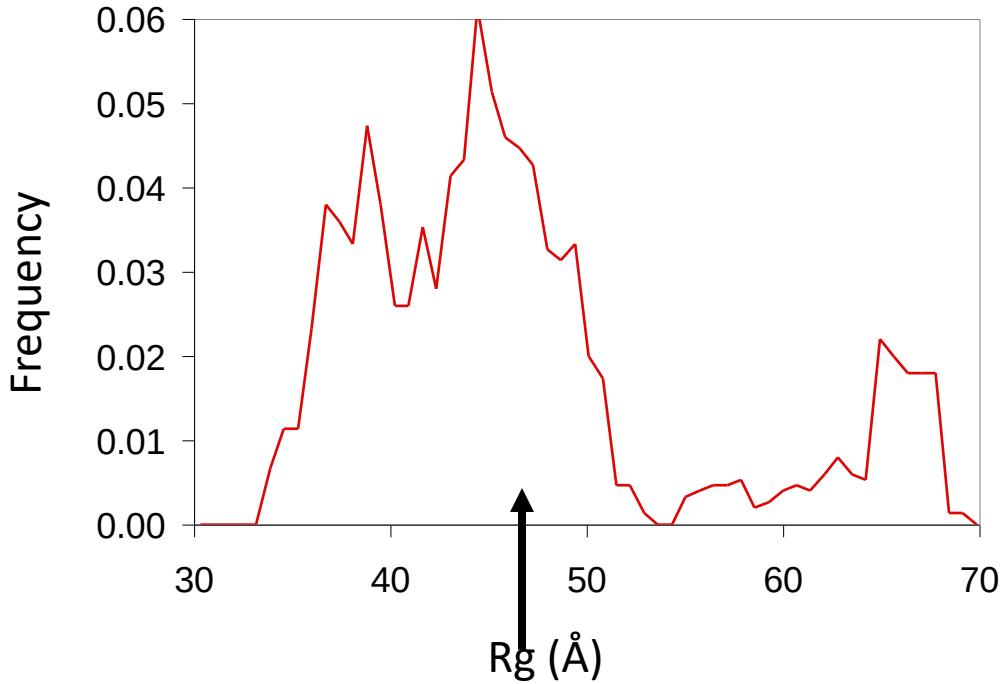


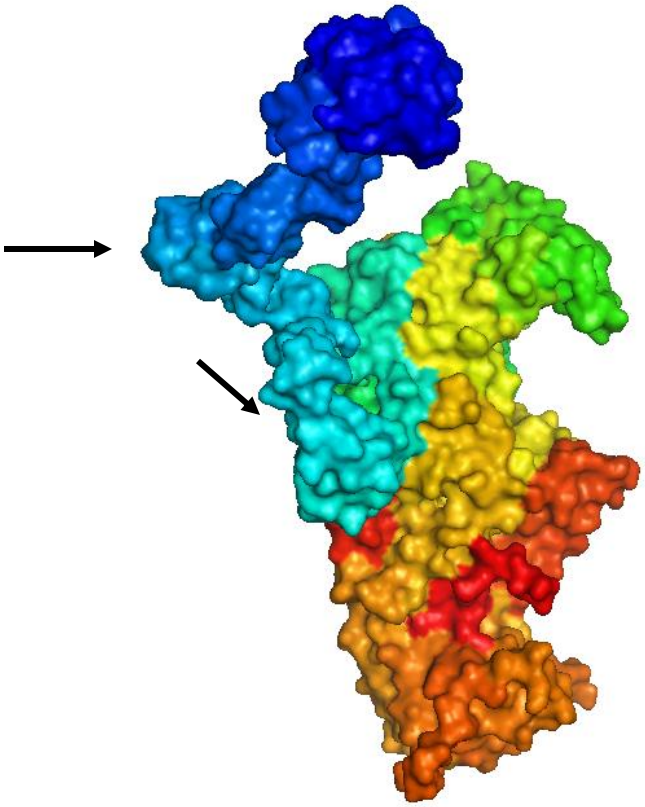

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

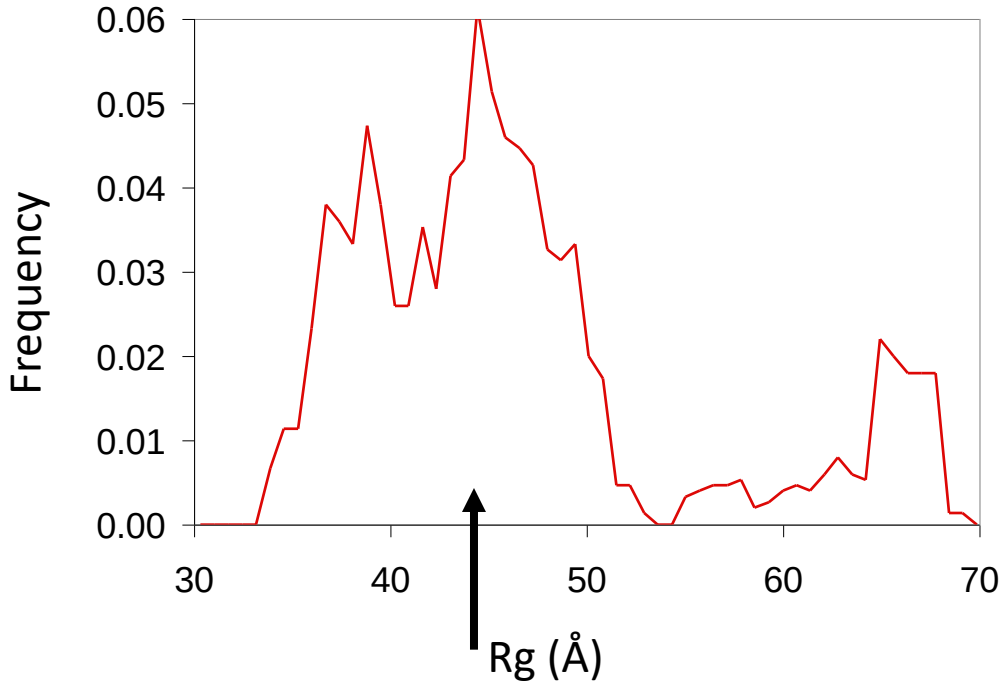


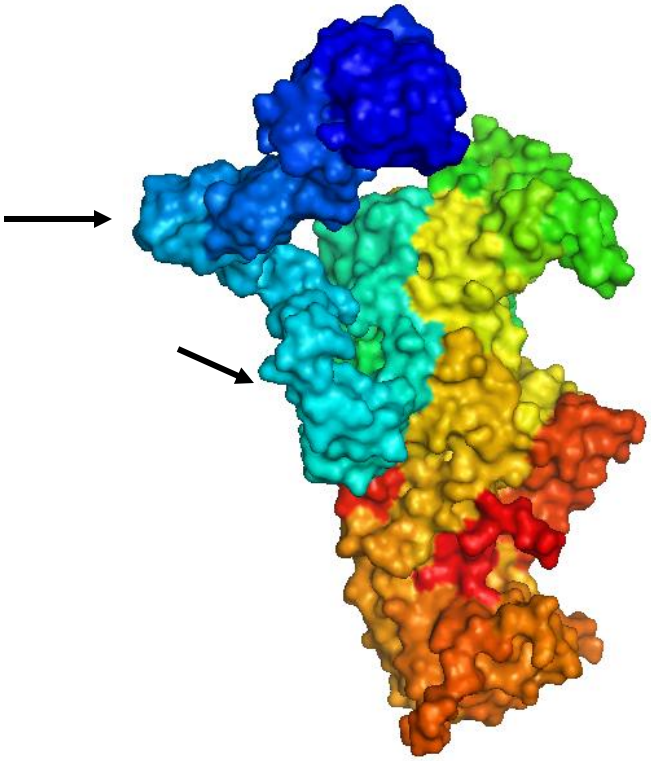

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

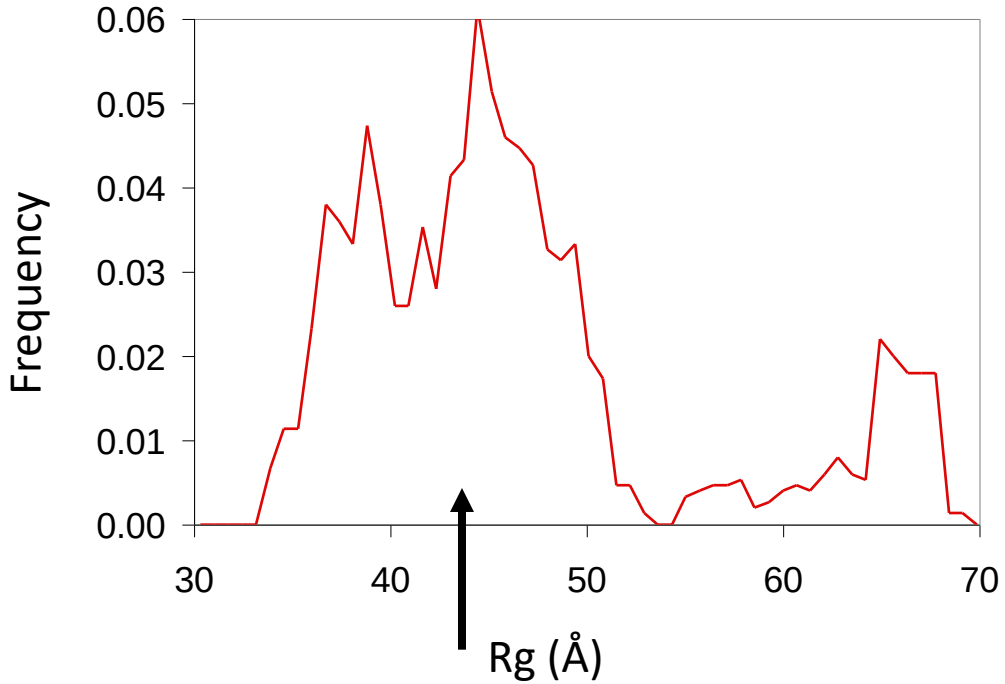


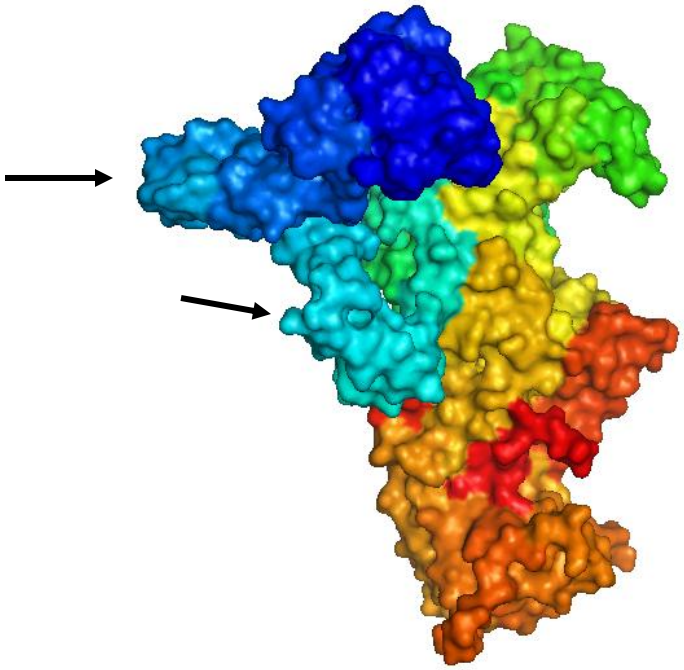

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

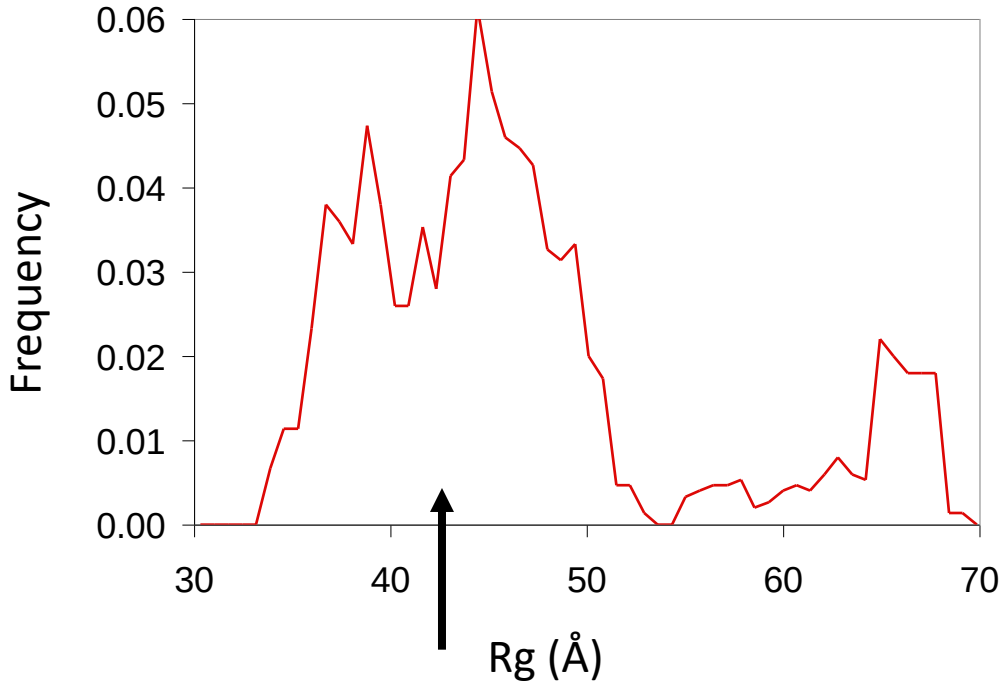


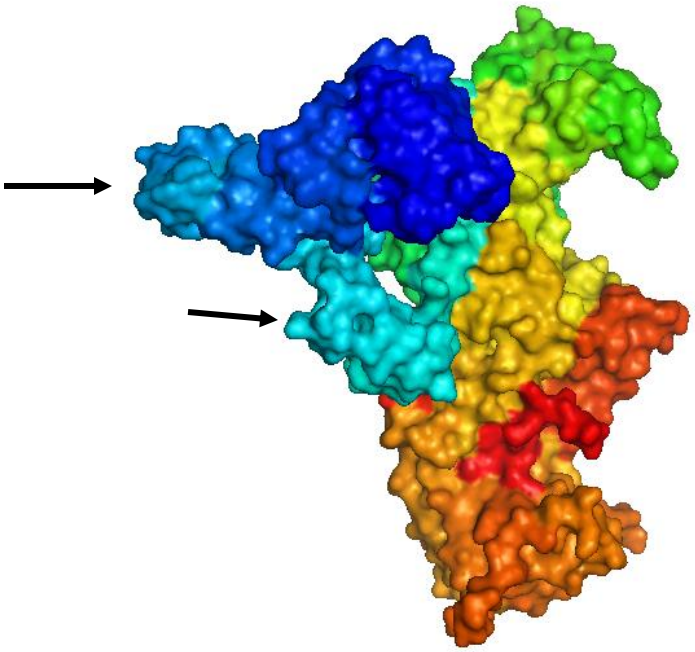

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

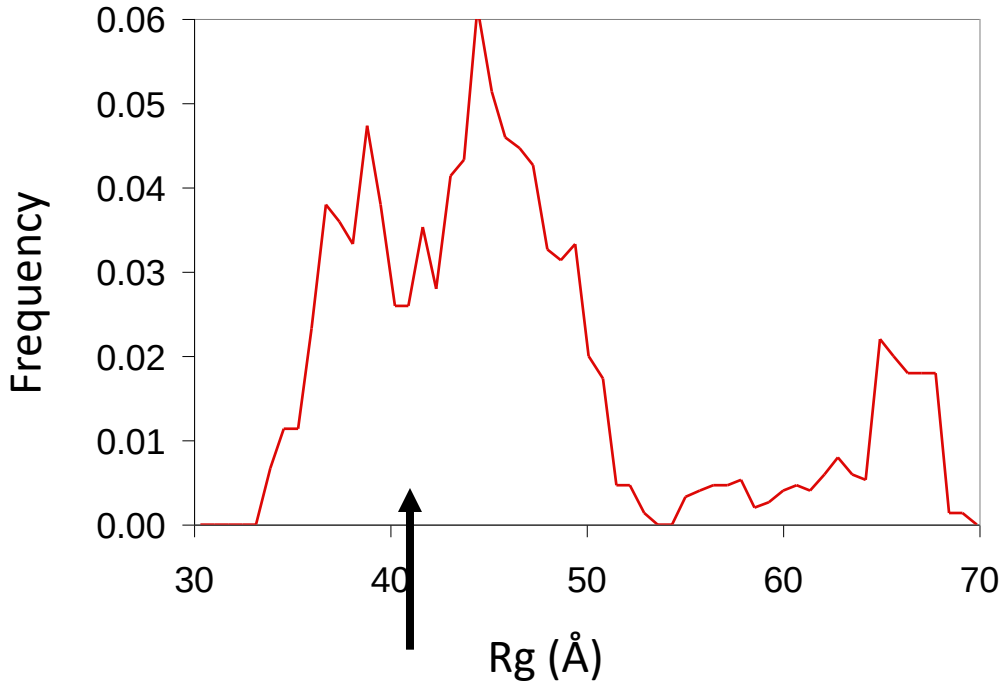


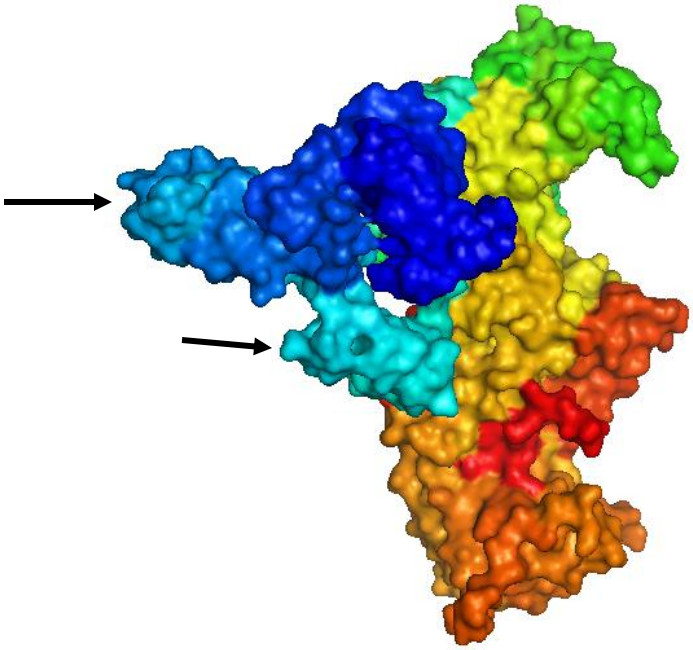

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

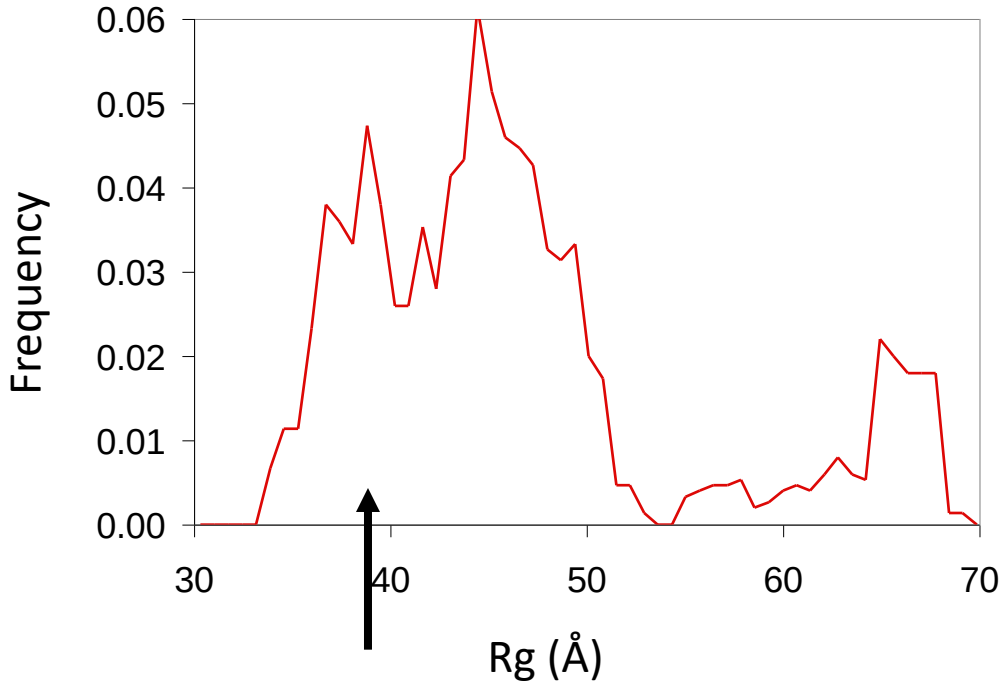


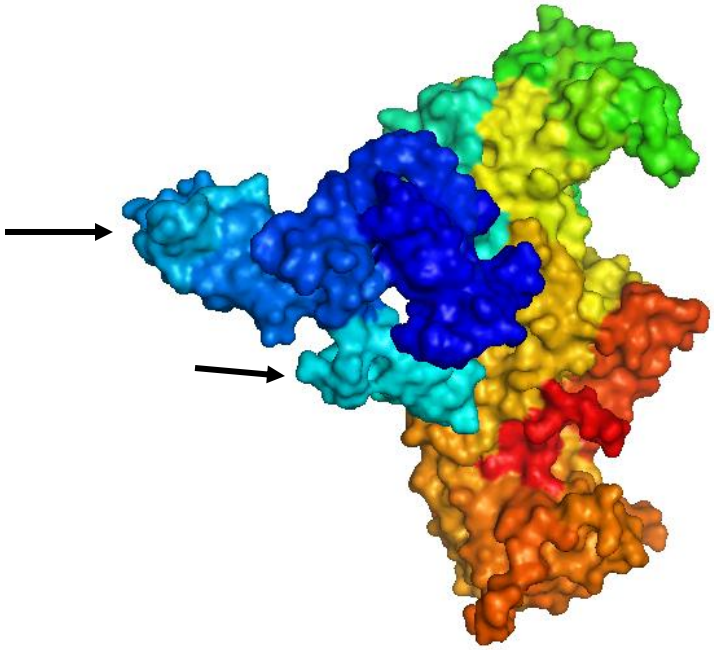

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

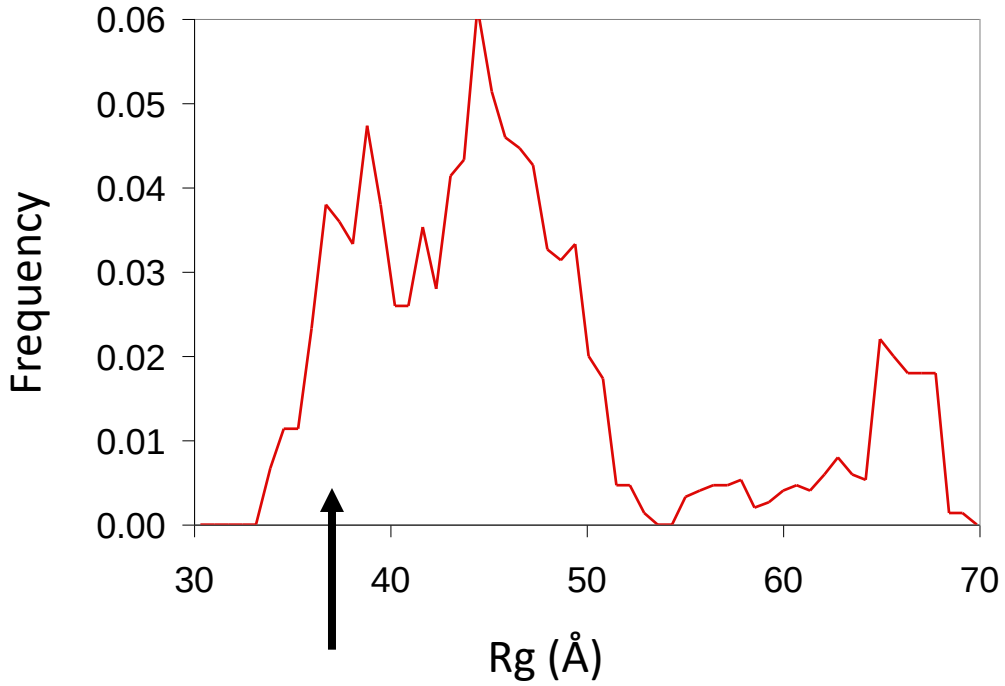


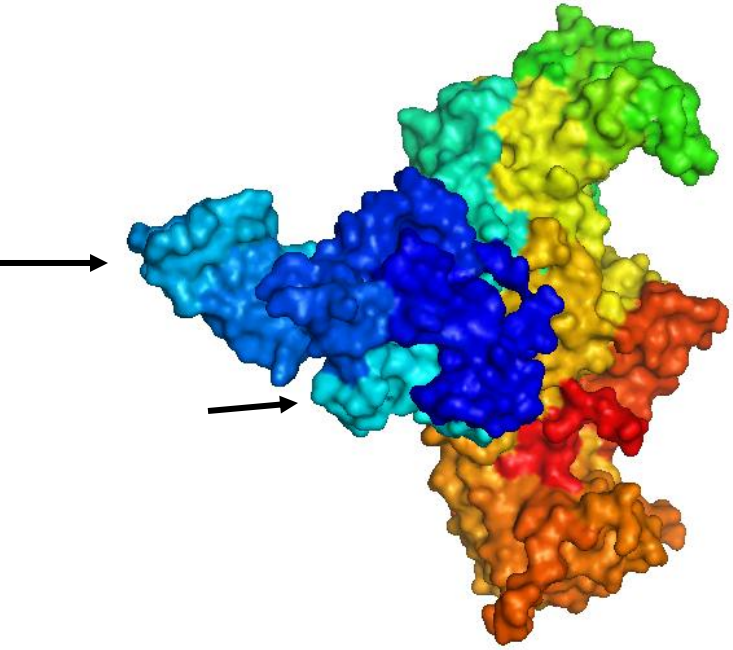

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

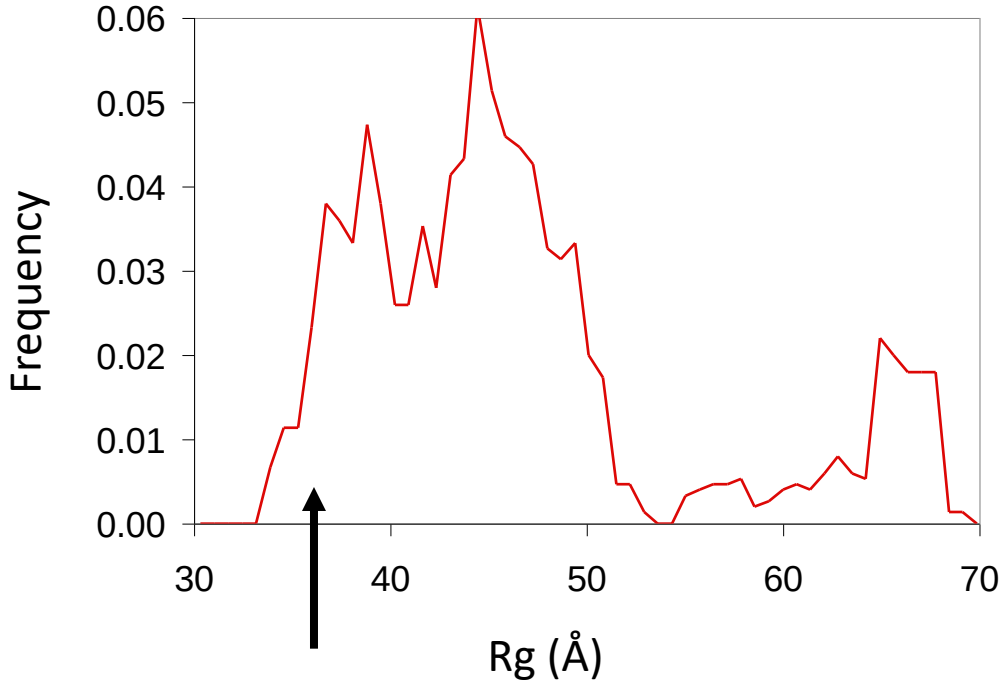


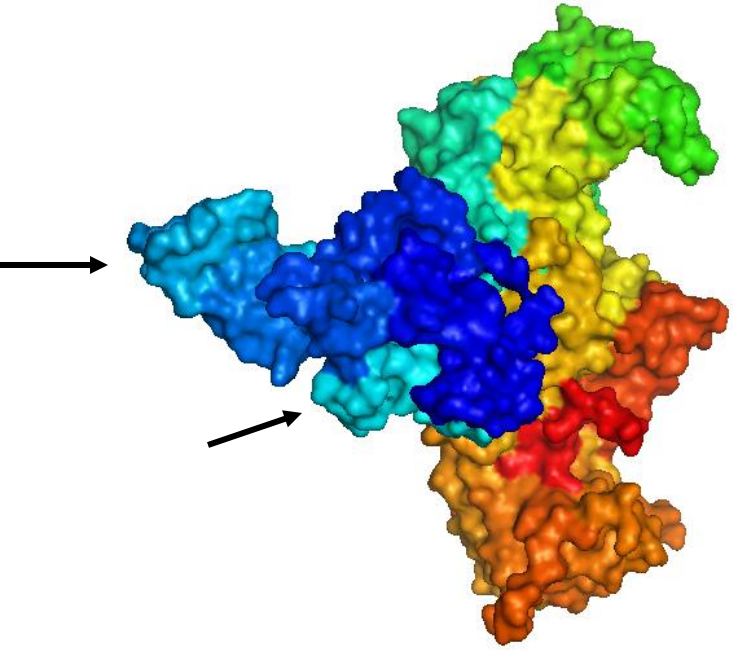

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

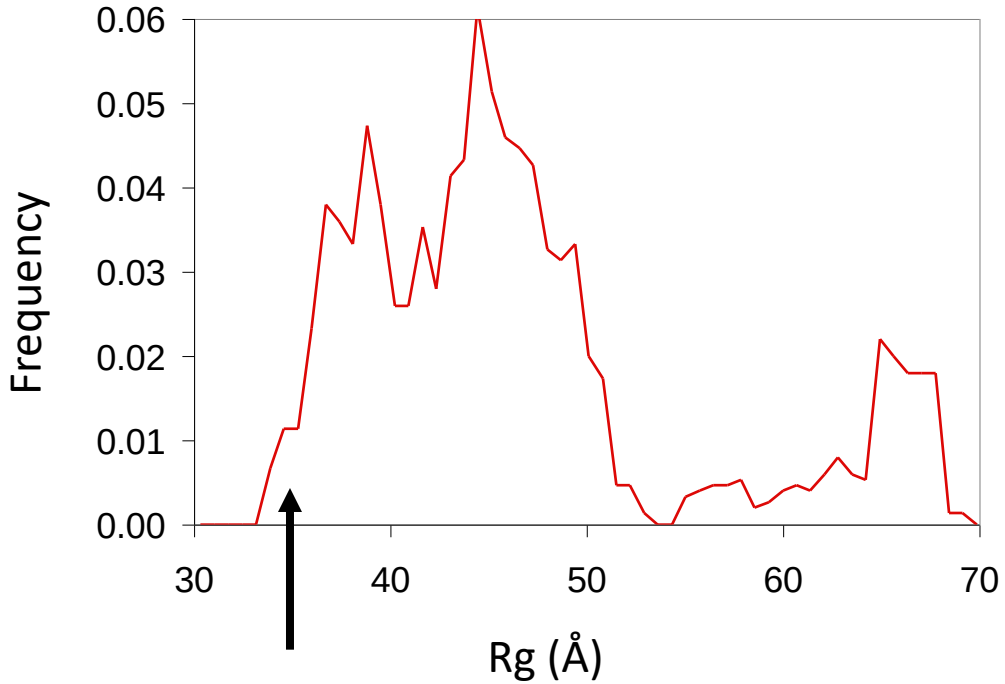



0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

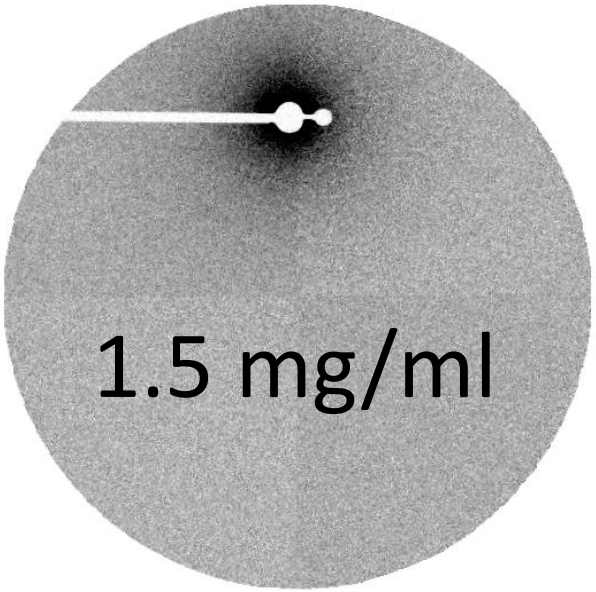


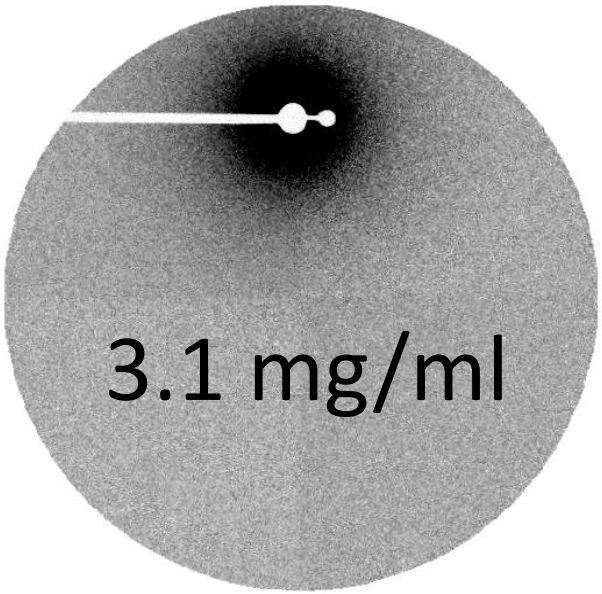

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

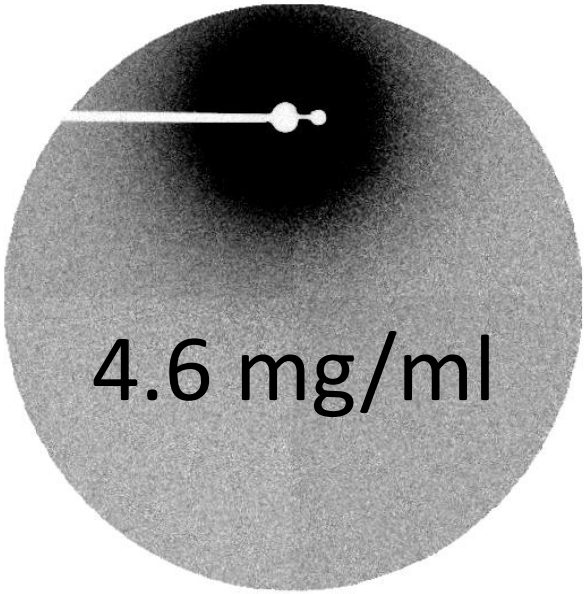


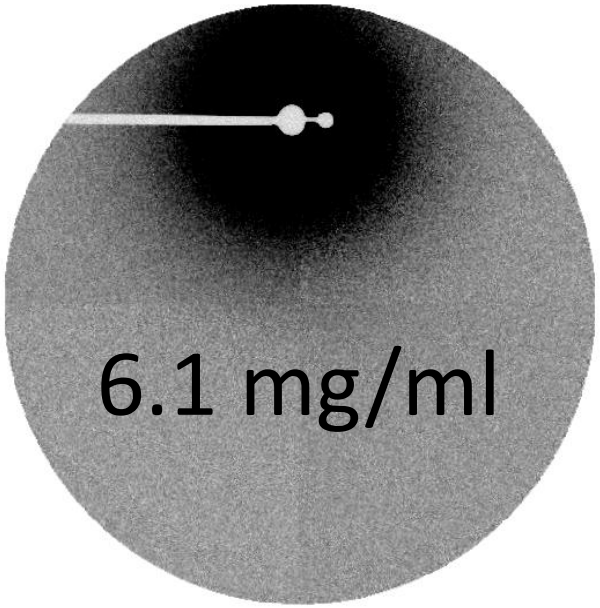

0.06
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
30
40
50
60
70
Rg (Å)

Really cool but wrong …

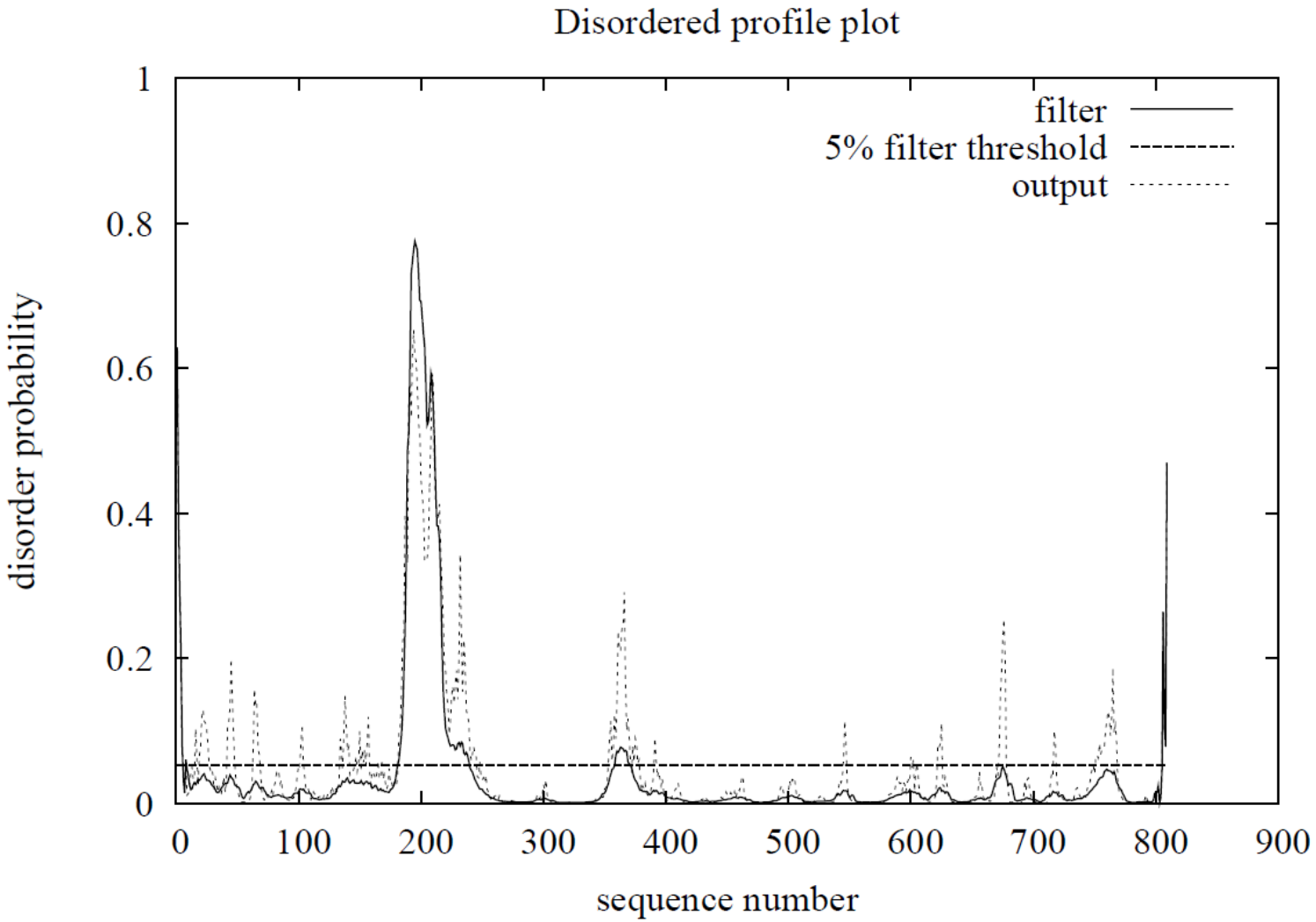

1.5 mg/ml


3.1 mg/ml


4.6 mg/ml

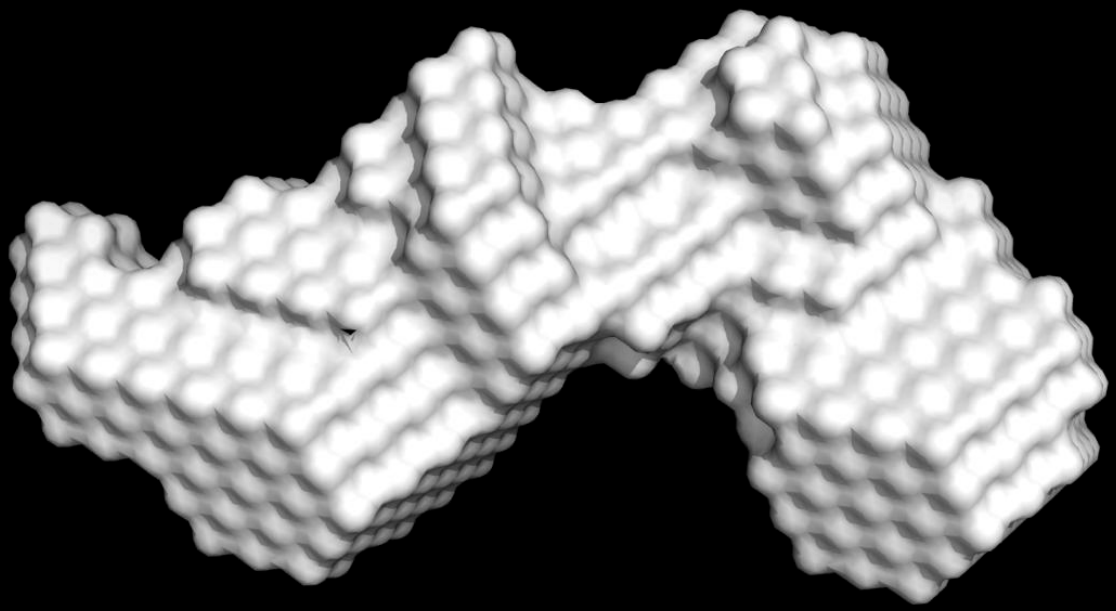

6.1 mg/ml

Aggregation in the highest concentration



Disorder Prediction Analysis of the Primary Sequence of ScGlnRS. The probability of disorder is shown on the y-axis and the residue number is shown on the x-axis. The linker connecting the N-terminal and C-terminal domains extends from residue 188 to 214. Disorder probability was calculated using DISOPRED2.

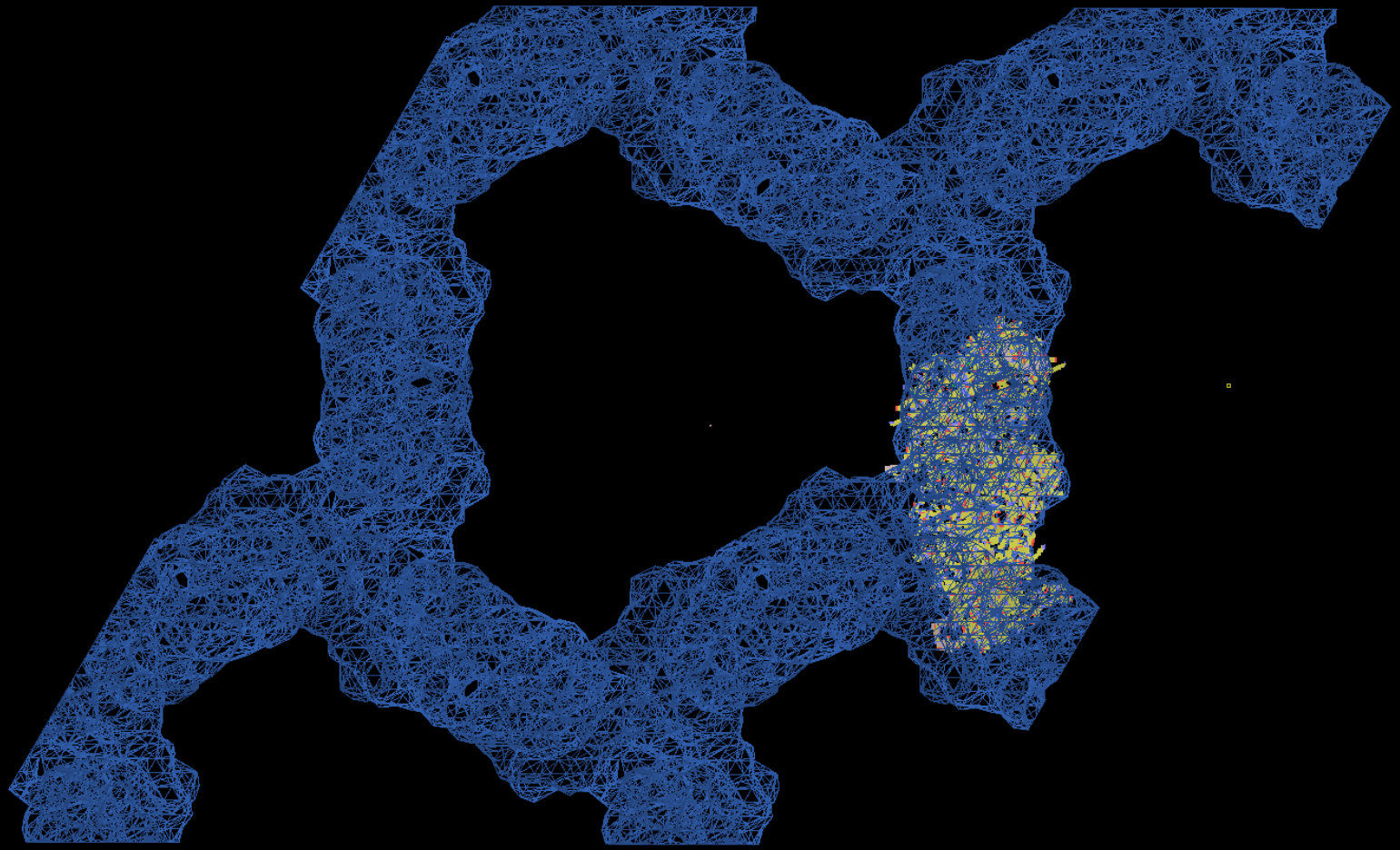
Envelope reconstruction of the N-terminal domain



Express N-terminal domain, C-terminal domain, tRNA, SAXS studies on all

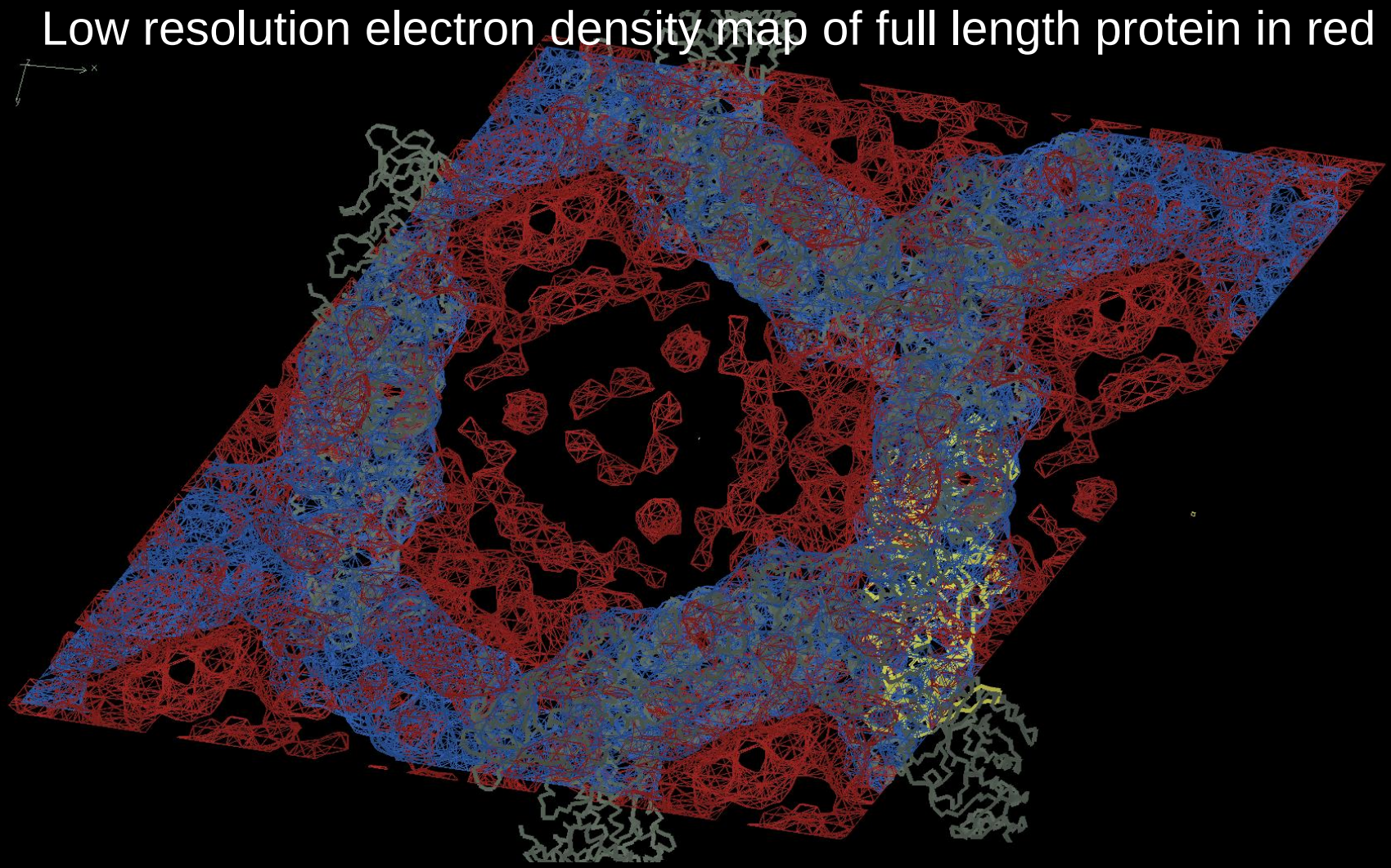
Check the crystallography again

Protein with N-terminal arm cleaved



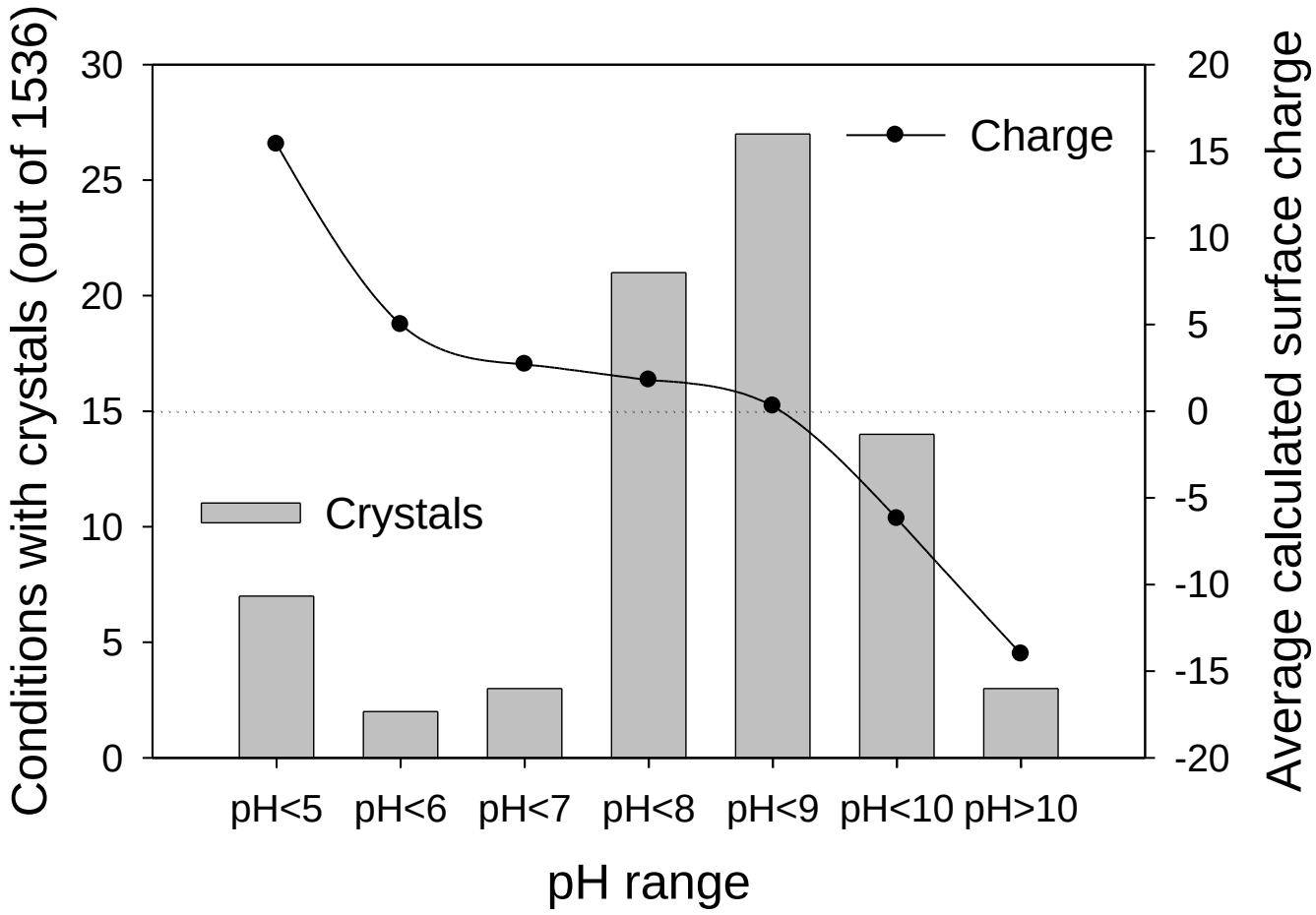
Crystallized, data truncated to 20A (data to 78A still plenty of reflections due to geometry and wavelength used purposely used for data collection)

Low resolution electron density map of full length protein in red

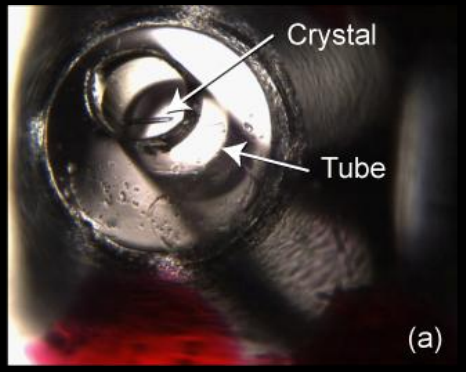


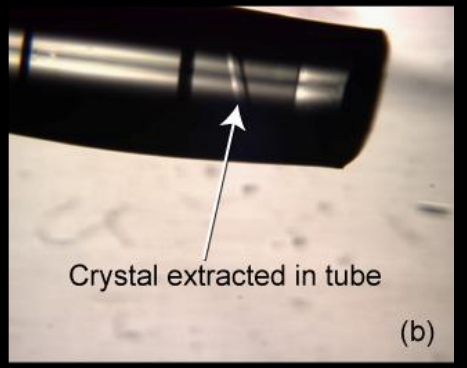
Data truncated to 20A (data to 78A still plenty of reflections due to geometry and wavelength used purposely used for data collection)

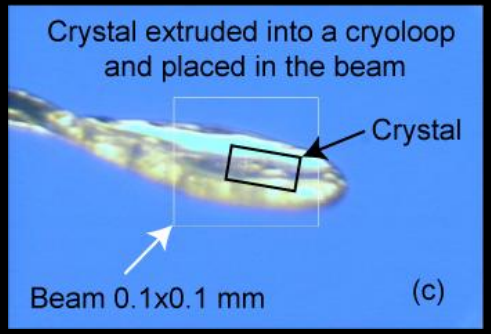
Crystallization trials of the N-terminal domain

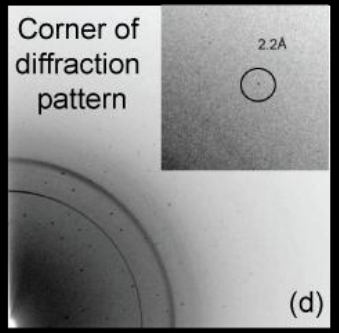


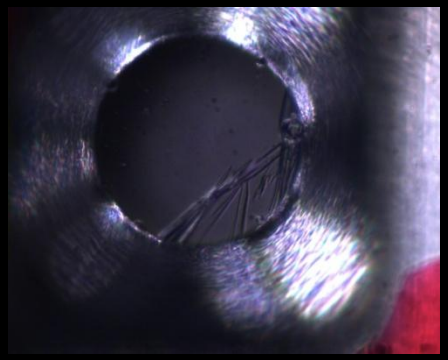
Does it diffract? Screening before the synchrotron

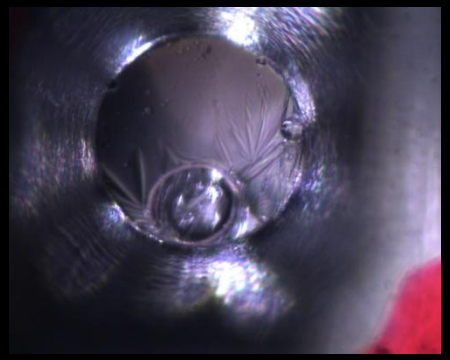


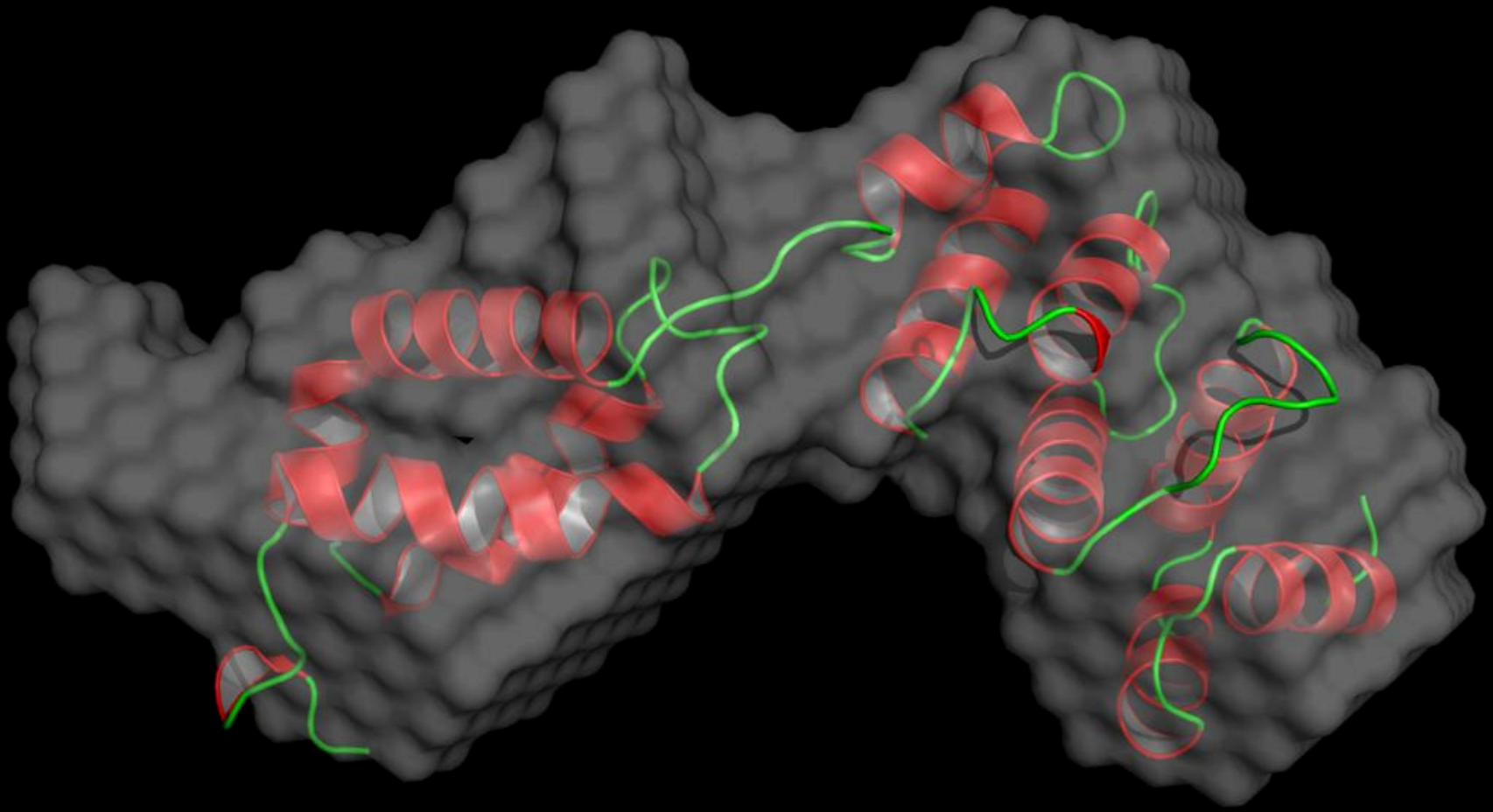


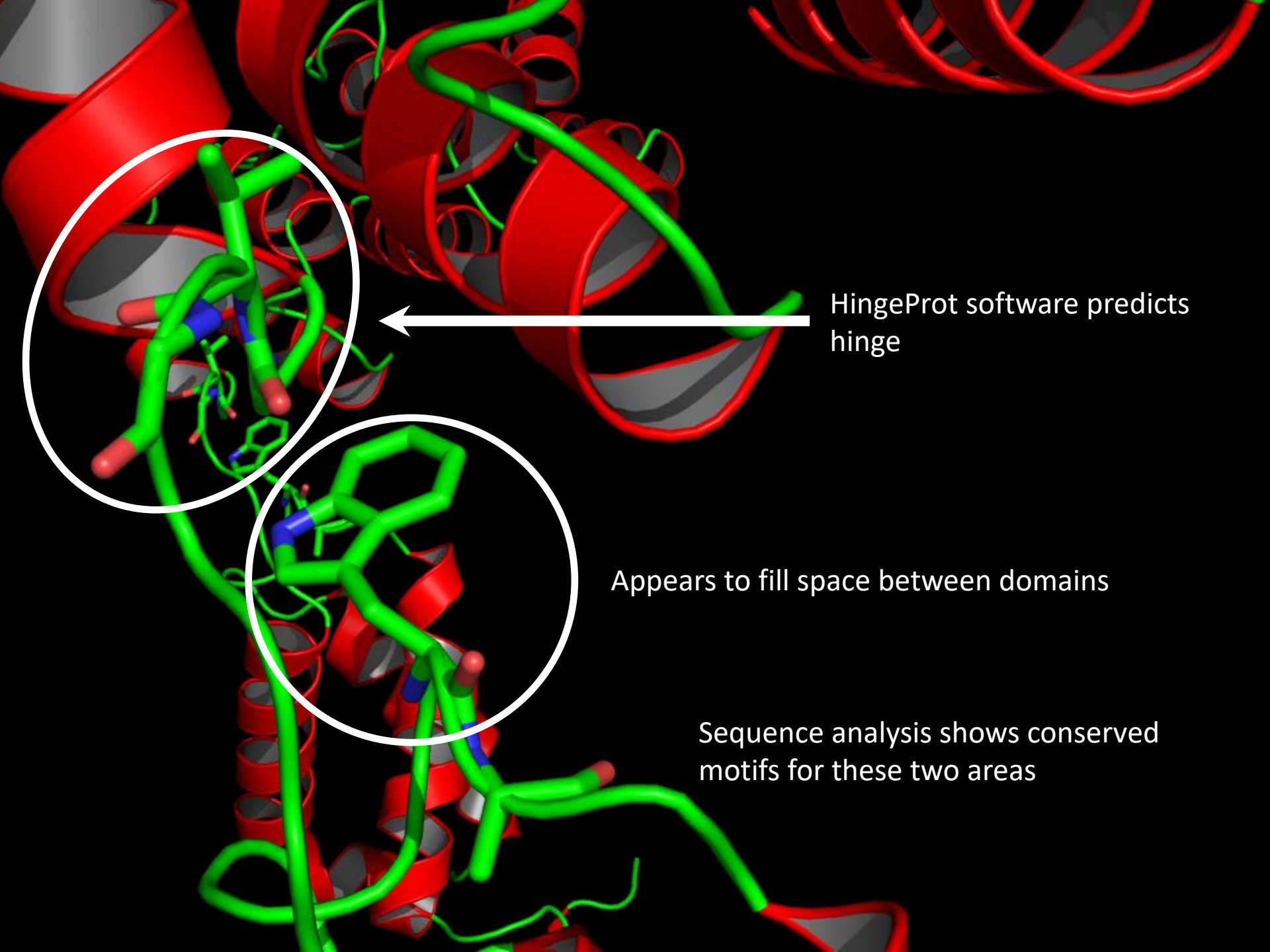










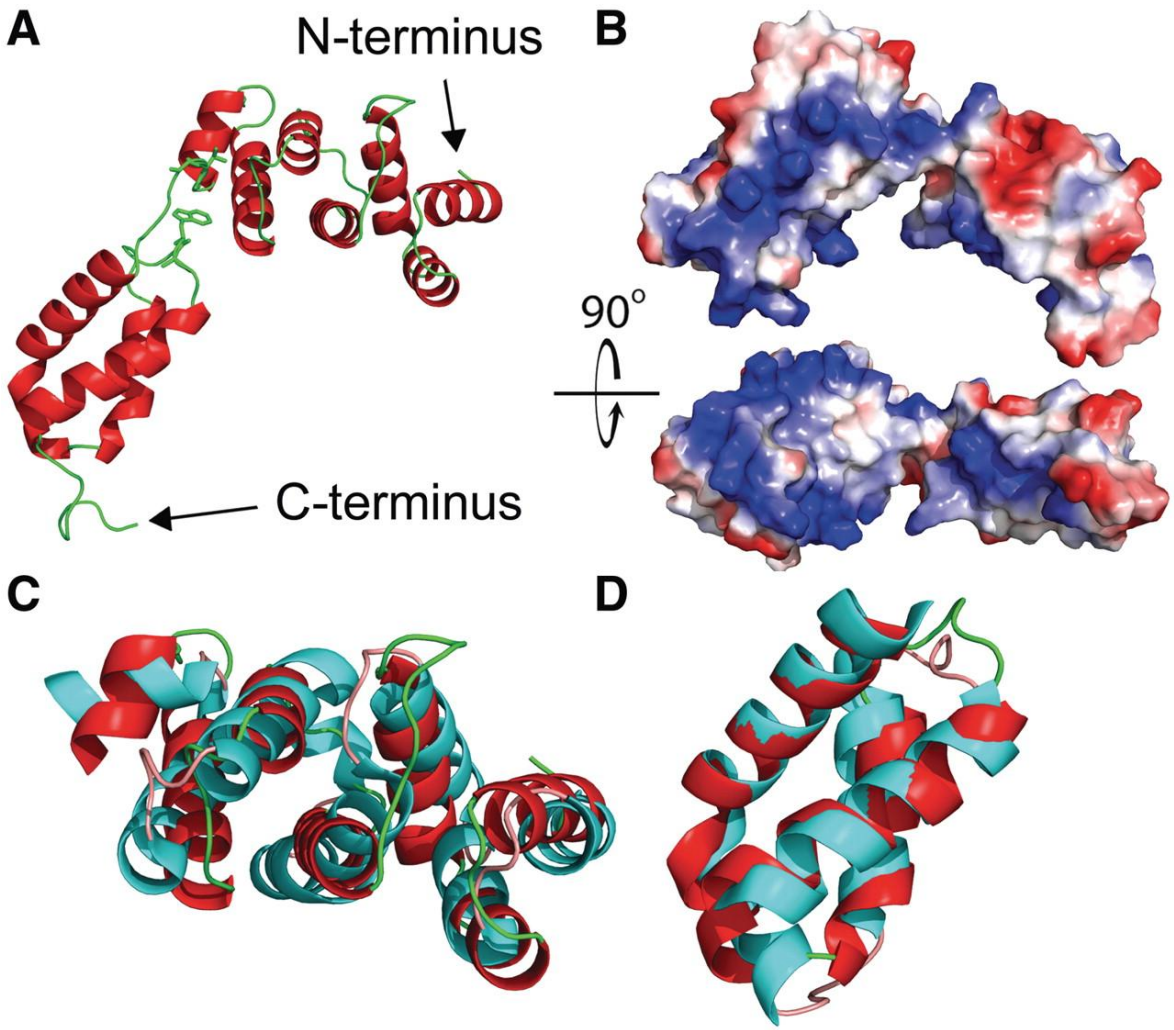

HingeProt software predicts hinge
Appears to fill space between domains
Sequence analysis shows conserved motifs for these two areas

Structural Homologs

- DALI search resulted in two hits of structurally similar molecules
- Combined with the SAXS this allowed us to position the N-terminal
- Due to the nature of the homologs we have a ‘big clue’ to the function of the N-terminal appended domain.
- SAXS studies of other species show a similar domain.
- Allowed us to better understand the evolutionary tree.

A blast search did not reveal structural homologs – having the structure of the N-terminal arm was critical.

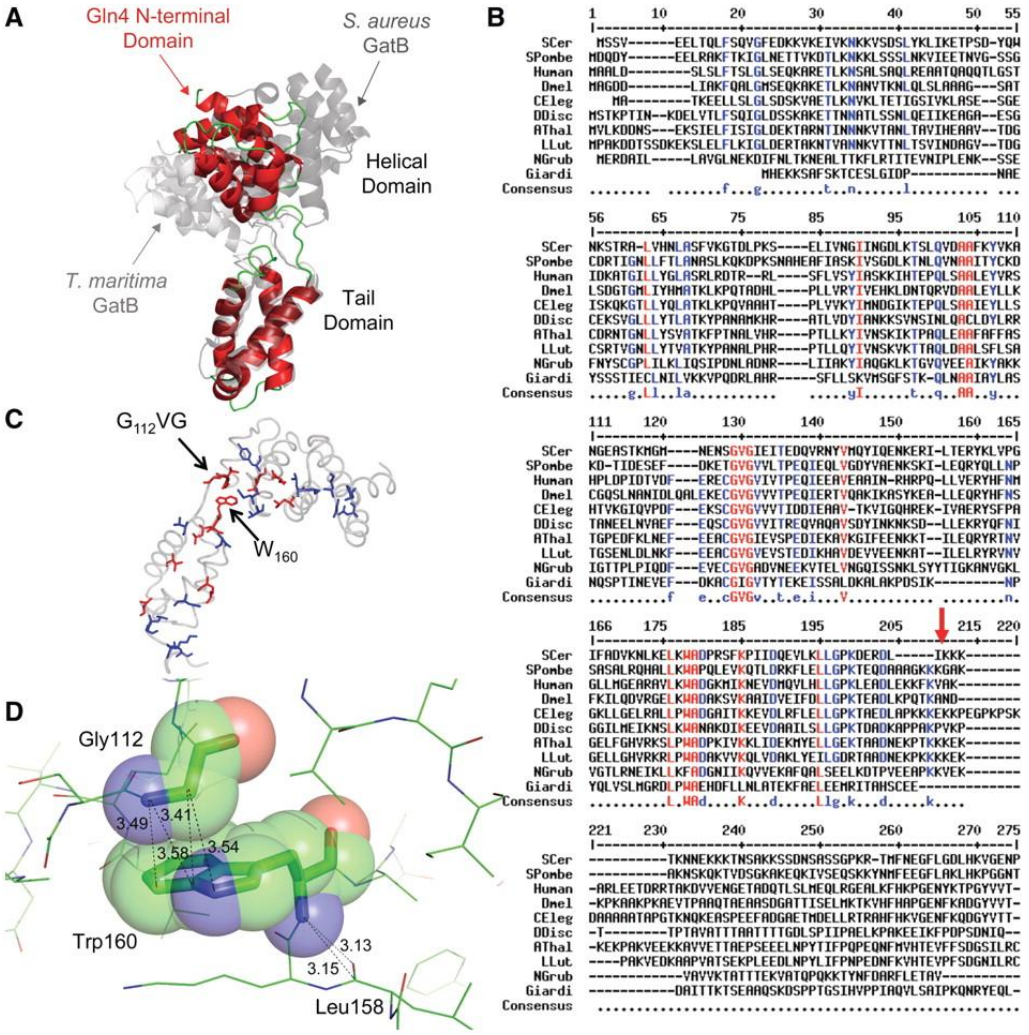
Structure of Gln4(1–187) with comparisons to domains in S. aureus GatB (PDB ID: 3IP4).



Grant T D et al. Nucl. Acids Res. 2011;nar.gkr1223

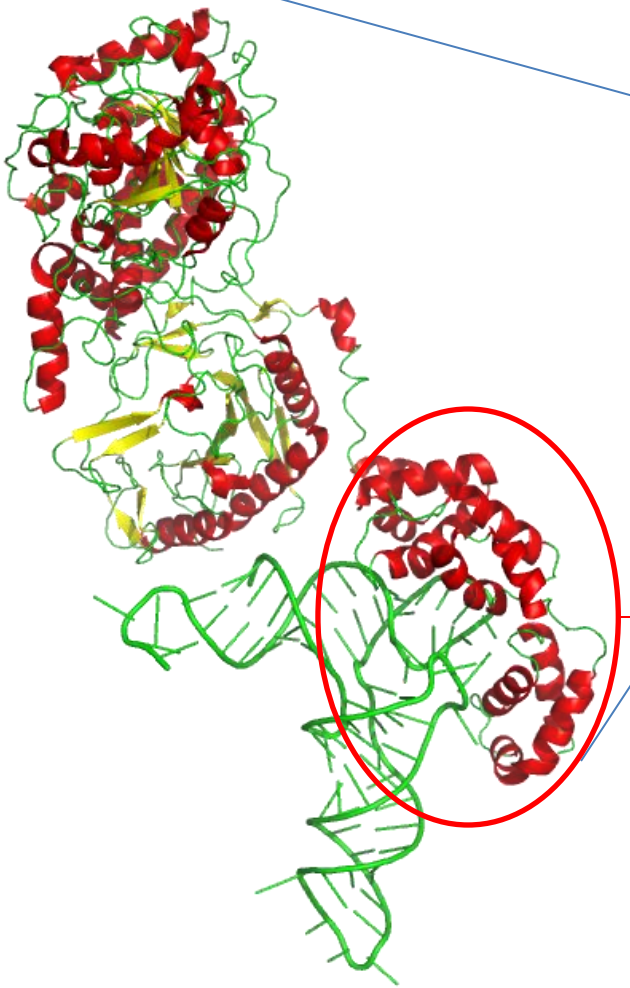
Nucleic Acids Research

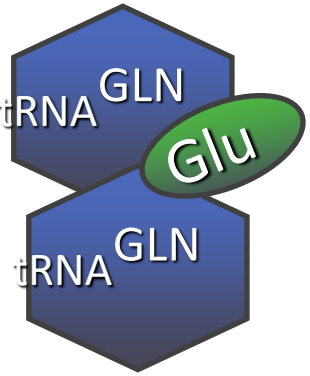
The linker between the two domains in Gln4(1–187) likely behaves as a hinge, is highly conserved and is important for tRNA binding.

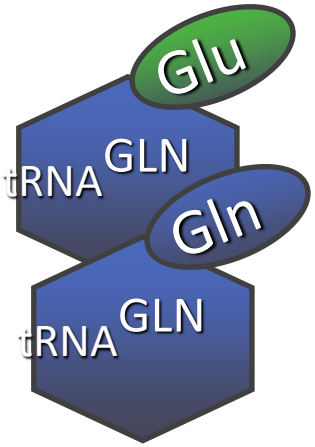


Grant T D et al. Nucl. Acids Res. 2011;nar.gkr1223

Nucleic Acids Research



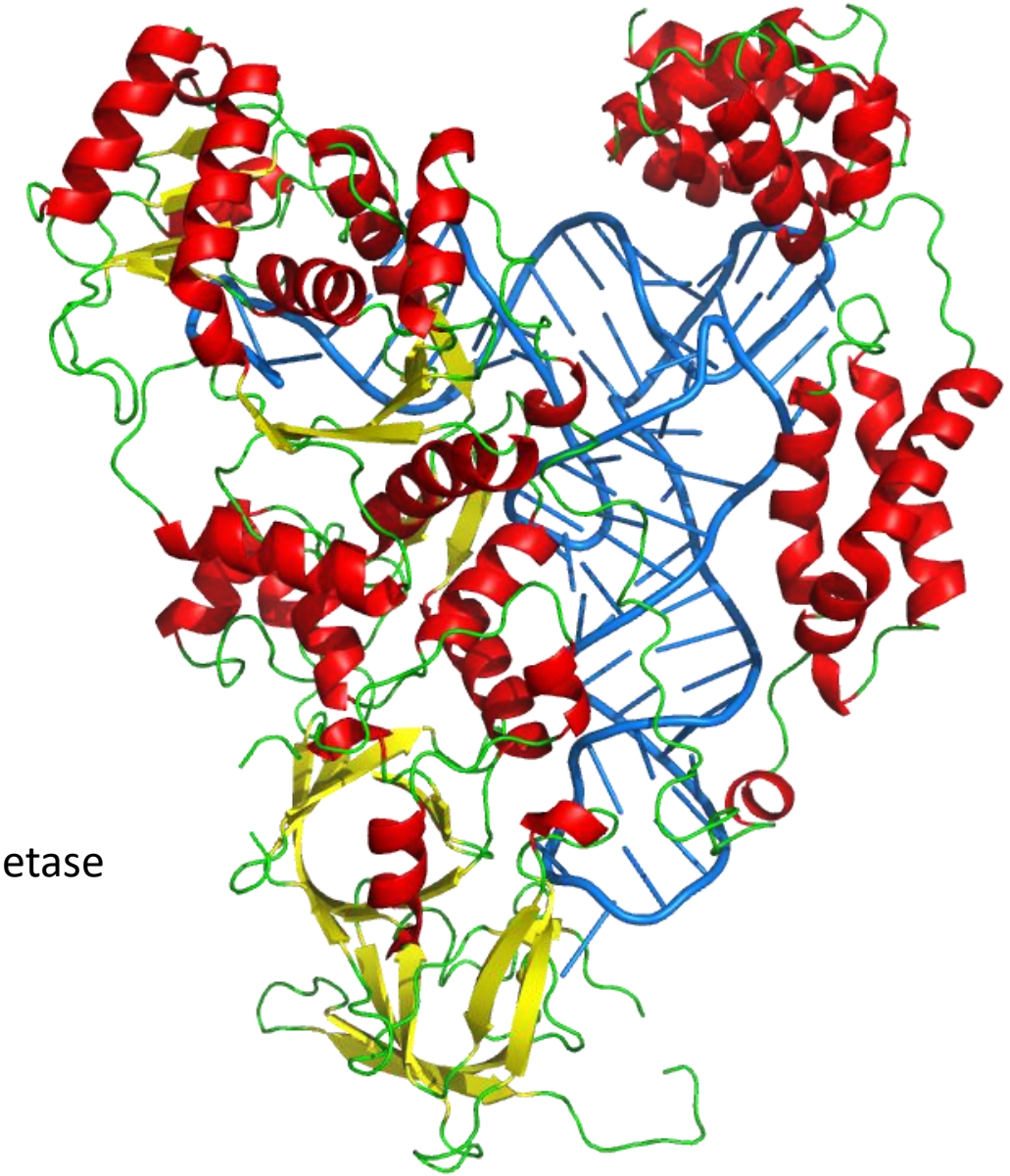


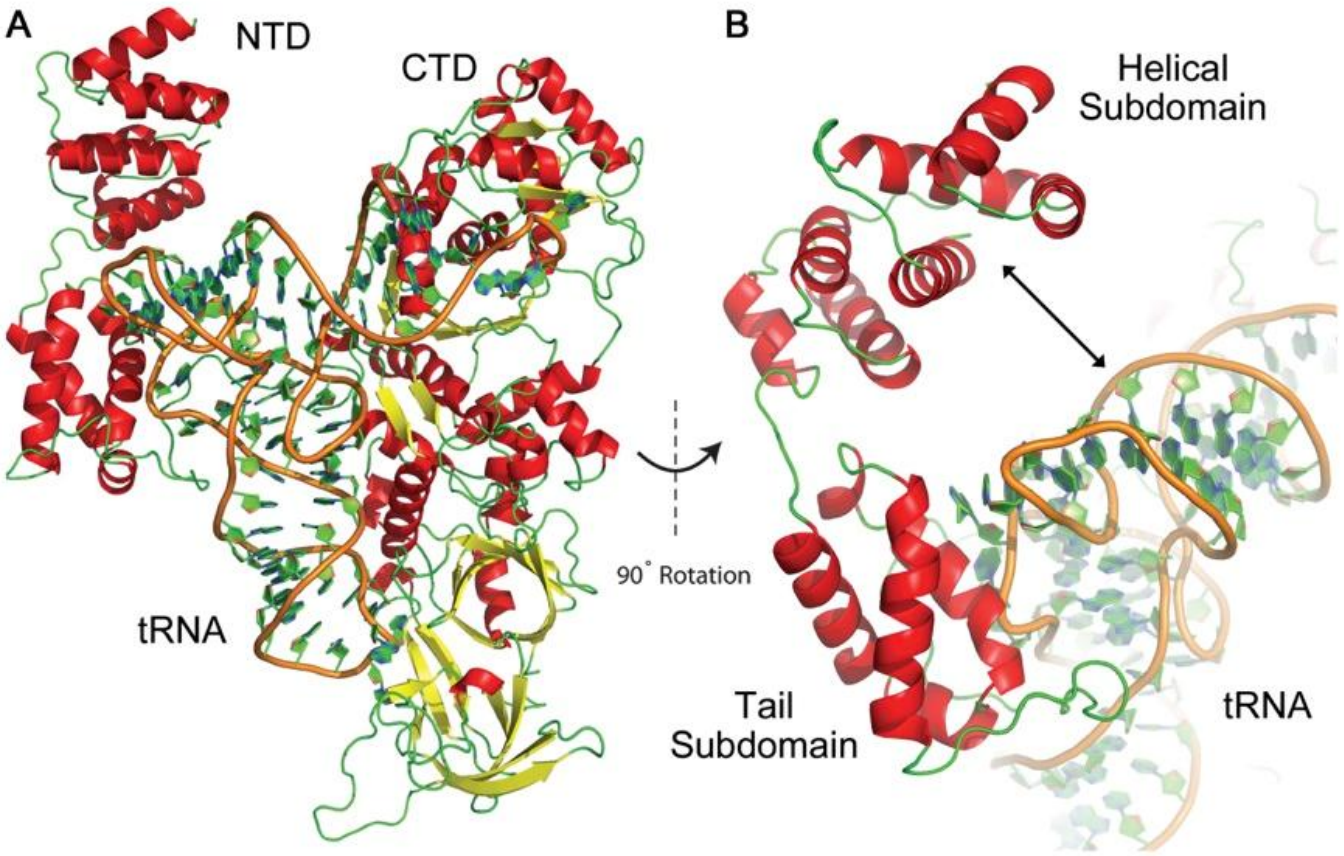


Remarkably similar to
the N-terminal domain
of Eukaryotic GlnRS

Combine the SAXS and Crystallography

Gln4 a Eukaryotic
Glutaminyl-tRNA Synthetase

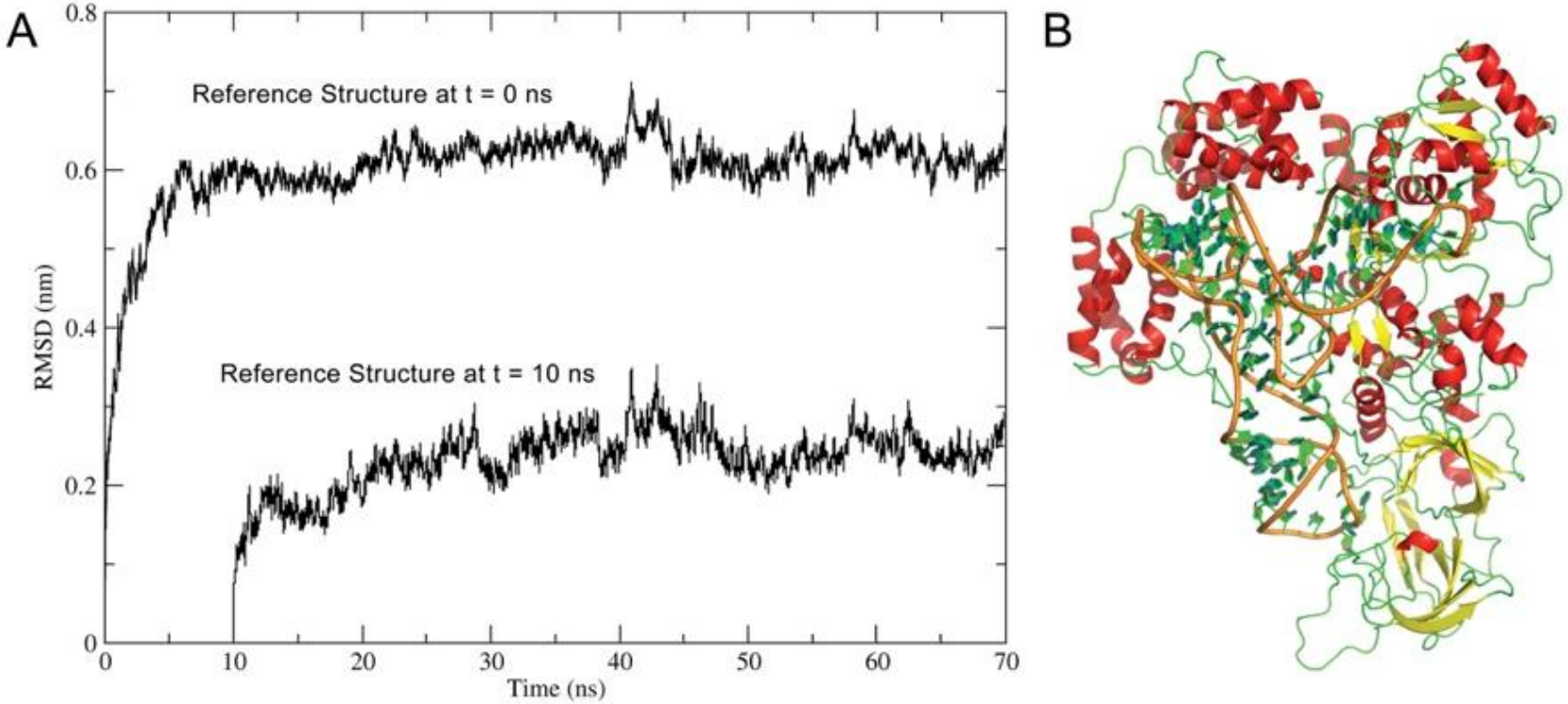


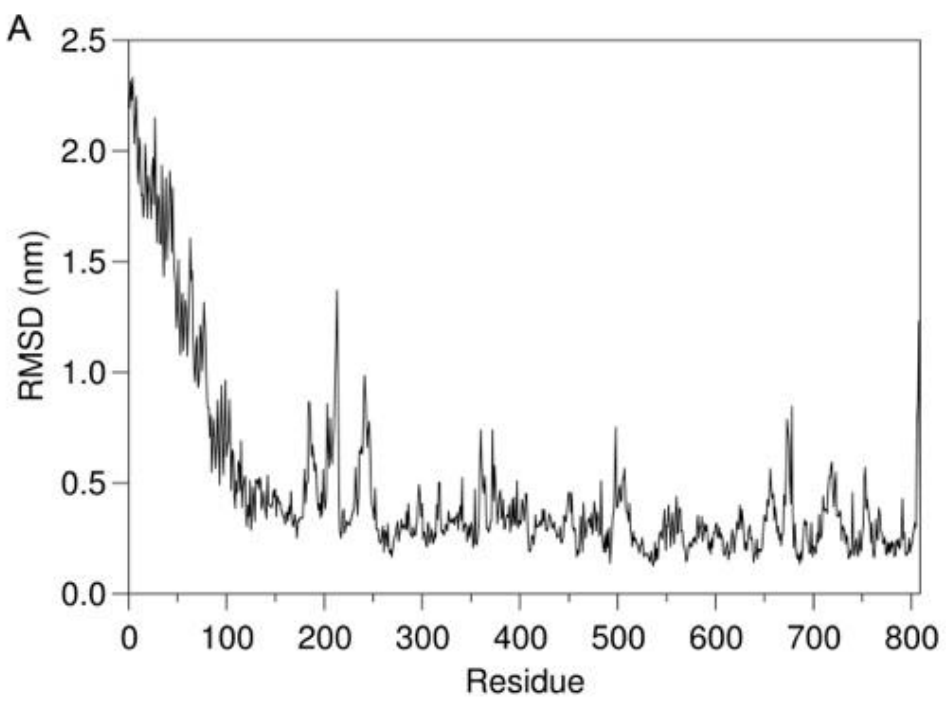


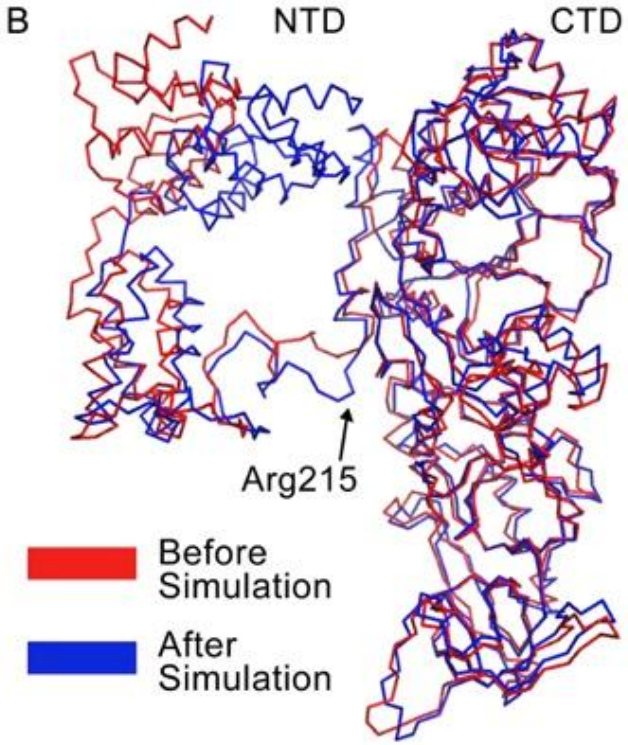
Homology Model of Full-length *Sc*GlnRS Bound to $tRNA^{gln}$. A. Full-length *Sc*GlnRS shown bound to $tRNA^{gln}$. B. Enlarged and rotated model showing gap between NTD helical subdomain and tRNA molecule.

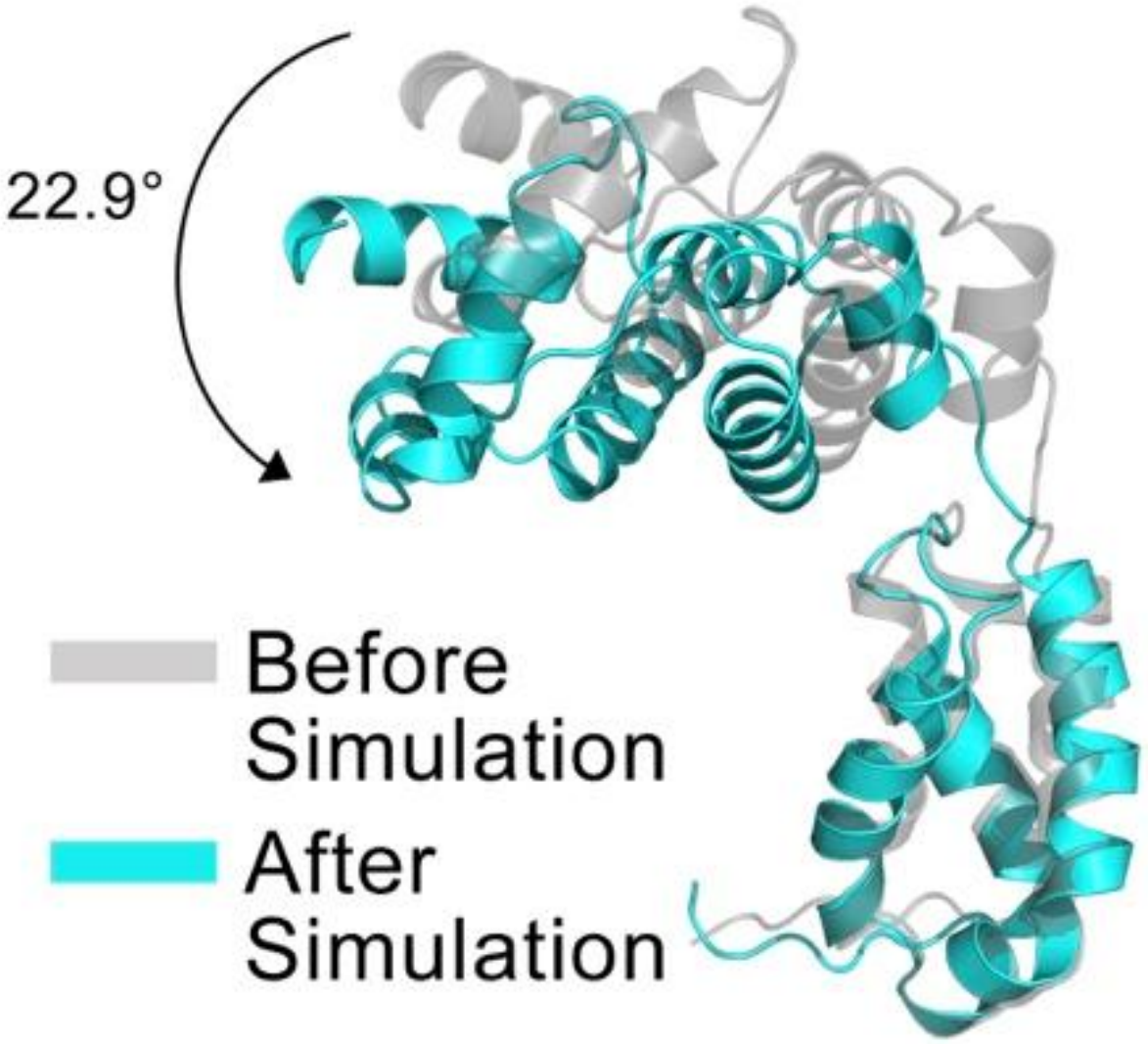
Molecular Dynamics Simulations

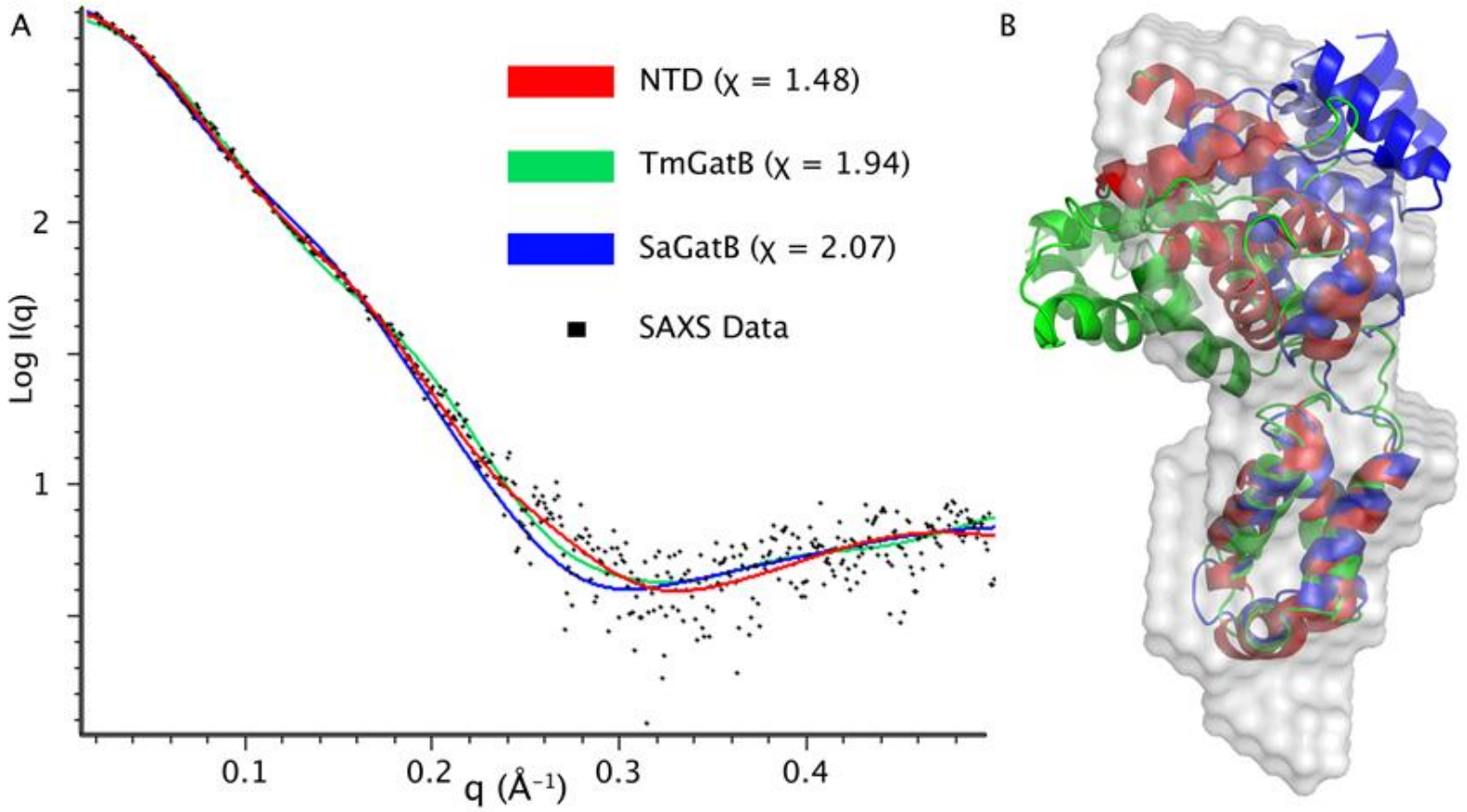
- Performed in GROMACS with the AMBER99SB force field.
- The initial model was solvated using a cubic SPC/E water model and neutralized with ions prior to minimization via steepest descents.
- Distance restraints were added to keep the zinc ion in place.
- The model was then equilibrated under an isothermal-isochoric ensemble for 100 picoseconds at 300K followed by equilibration under an isothermal-isobaric ensemble for 100 picoseconds.
- Simulations were then performed at the Center for Computational Resources on 512 processors. Total simulation time was 70 ns.


A
B
Reference Structure at t = 0 ns
Reference Structure at t = 10 ns
RMSD (nm)
Time (ns)
0
0.2
0.4
0.6
0.8
0
10
20
30
40
50
60
70


A
RMSD (nm)
2.5
2.0
1.5
1.0
0.5
0.0
0
100
200
300
400
500
600
700
800
Residue


B
NTD
CTD
Arg215
Before Simulation
After Simulation


22.9°
Before Simulation
After Simulation

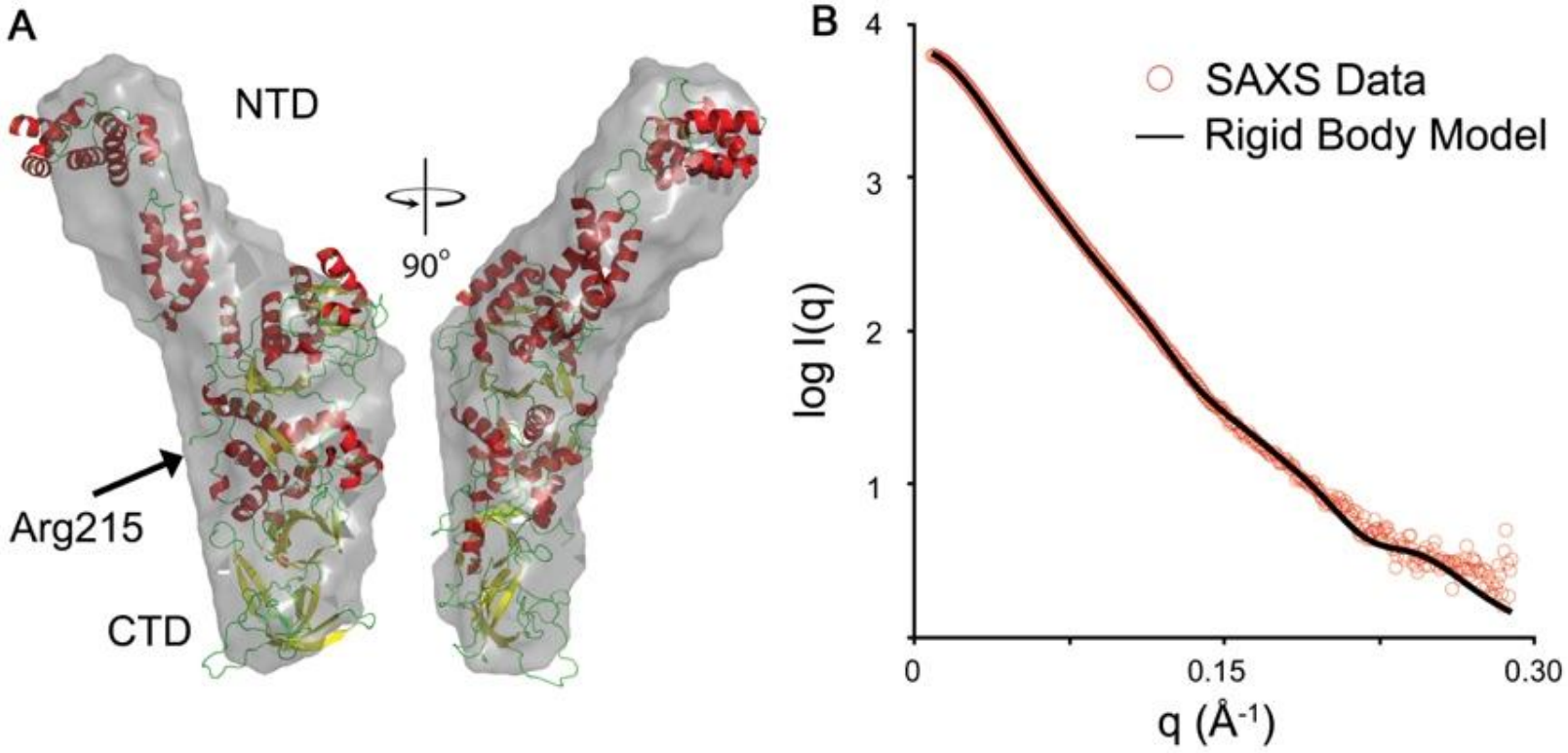


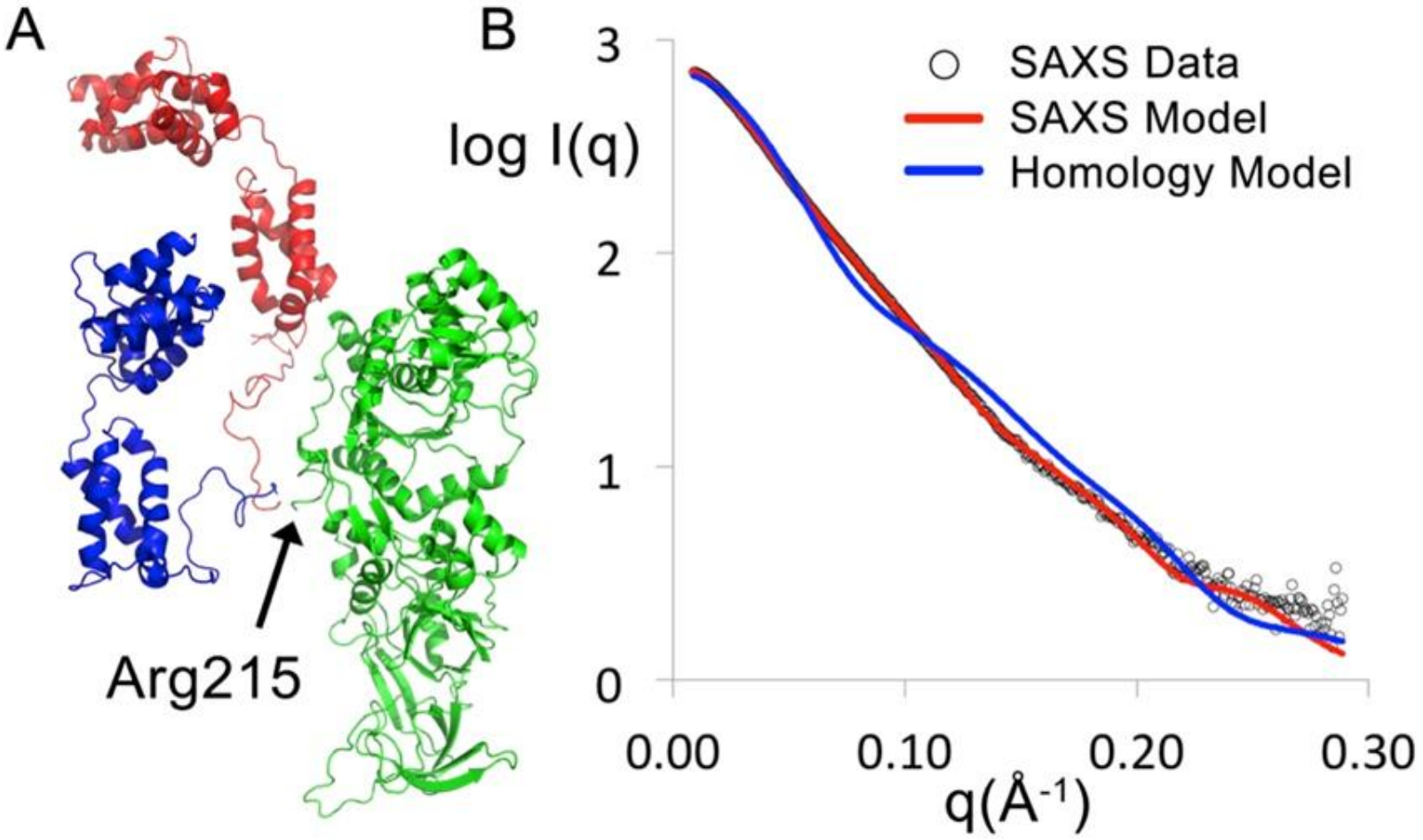
SAXS data shows that the NTD crystal structure is similar to that found in solution. A. Simulated scattering profiles calculated by CRYSOL for the Gln4 NTD (red), TmGatB (green), and SaGatB (blue) are shown overlaid on top of experimental SAXS data from the Gln4 NTD in solution. Goodness of fit values (χ) are given in parentheses. B. The ab initio envelope reconstructed from the experimental scattering profile of the Gln4 NTD is shown superimposed onto the crystal structures of the Gln4 NTD (red), TmGatB (green), and SaGatB (blue).

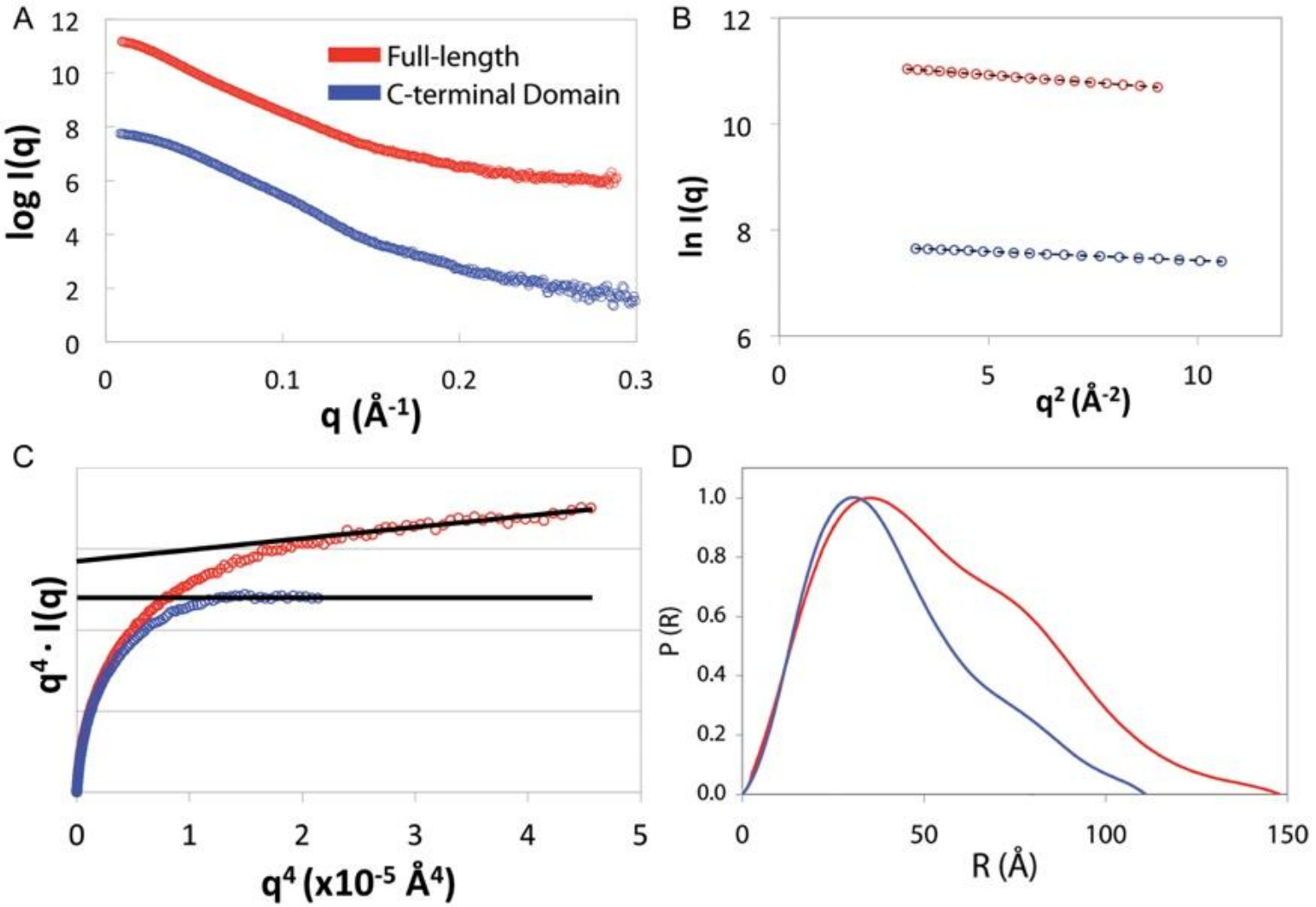
Homology model is not in agreement with solution envelope

Homology versus solution envelope

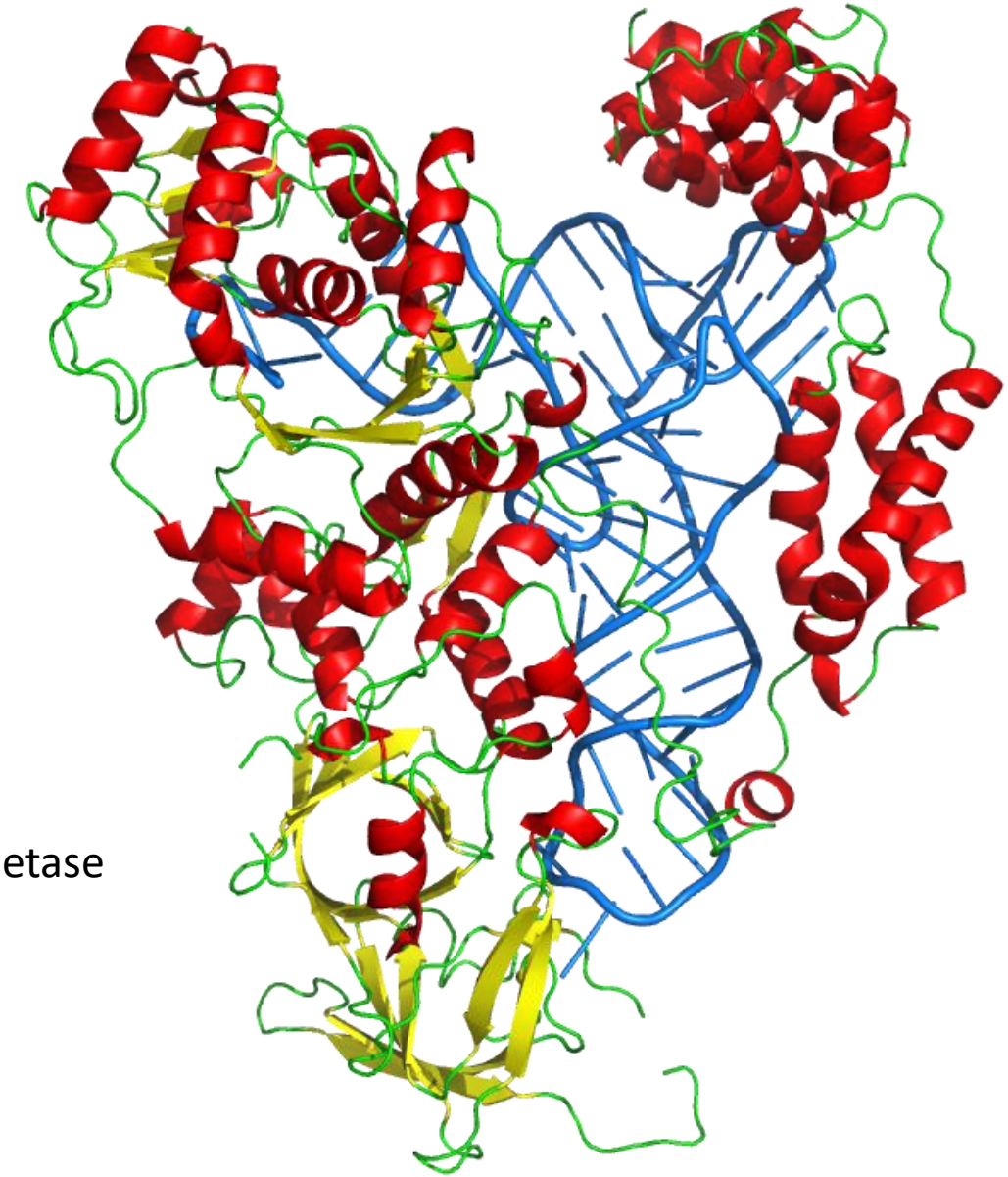
- The full-length ScGlnRS bound to $tRNA^{Gln}$ shows a significant change in the NTD position when compared to the $tRNA^{Gln}$-free, SAXS-derived conformation .
- The model shows a ~160° rotation and a ~40 Å translation of the NTD with respect to the solution conformation.
- Fitting the simulated scattering of the protein portion of the protein-tRNA complex to the experimental SAXS data resulted in a poor fit, yielding a $\chi 2 = 12.25$ compared to 1.82 for the rigid body model . The limited flexibility of the NTD, coupled with the poor fit of the simulated scattering of the protein portion of the model bound to $tRNA^{Gln}$, suggests that without tRNA bound, this conformation does not exist in solution.
- Analysis with OLIGOMER showed that only the rigid body model exists in solution, while the homology model does not.
- Taken together, these observations suggest that **CTD binding of $tRNA^{Gln}$ induces substantial conformational reorientation of the NTD required for interactions with $tRNA^{Gln}$.**


A
NTD
90°
Arg215
CTD
B
SAXS Data
Rigid Body Model
log I(q)
4
3
2
1
0
0.15
0.30
q (Å-1)


A
Arg215
B
log I(q)
3
2
1
0
SAXS Data
SAXS Model
Homology Model
0.00
0.10
0.20
0.30
q(Å-1)


A
log I(q)
q (Å-1)
Full-length
C-terminal Domain
B
ln I(q)
q2 (Å-2)
C
q4 · I(q)
q4 (x10-5 Å4)
D
P (R)
R (Å)

A combination of molecular biology, SAXS, crystallography and molecular dynamics



Gln4 a Eukaryotic
Glutaminyl-tRNA Synthetase

Summary …
what can SAXS do for you?

Model the Question

A SAXS profile can be calculated from any model

Going from a SAXS profile to a three dimensional envelope is an inherently underdetermined problem. However the reverse is not, it is completely possible (and easily done) to determine a theoretical SAXS curve from a model

The first question (if you are not looking for simple characterization or envelope information) should be can a SAXS experiment distinguish between hypotheses? Calculate model scattering profiles and determine if potential models produce noticeable difference in the scattering curve.

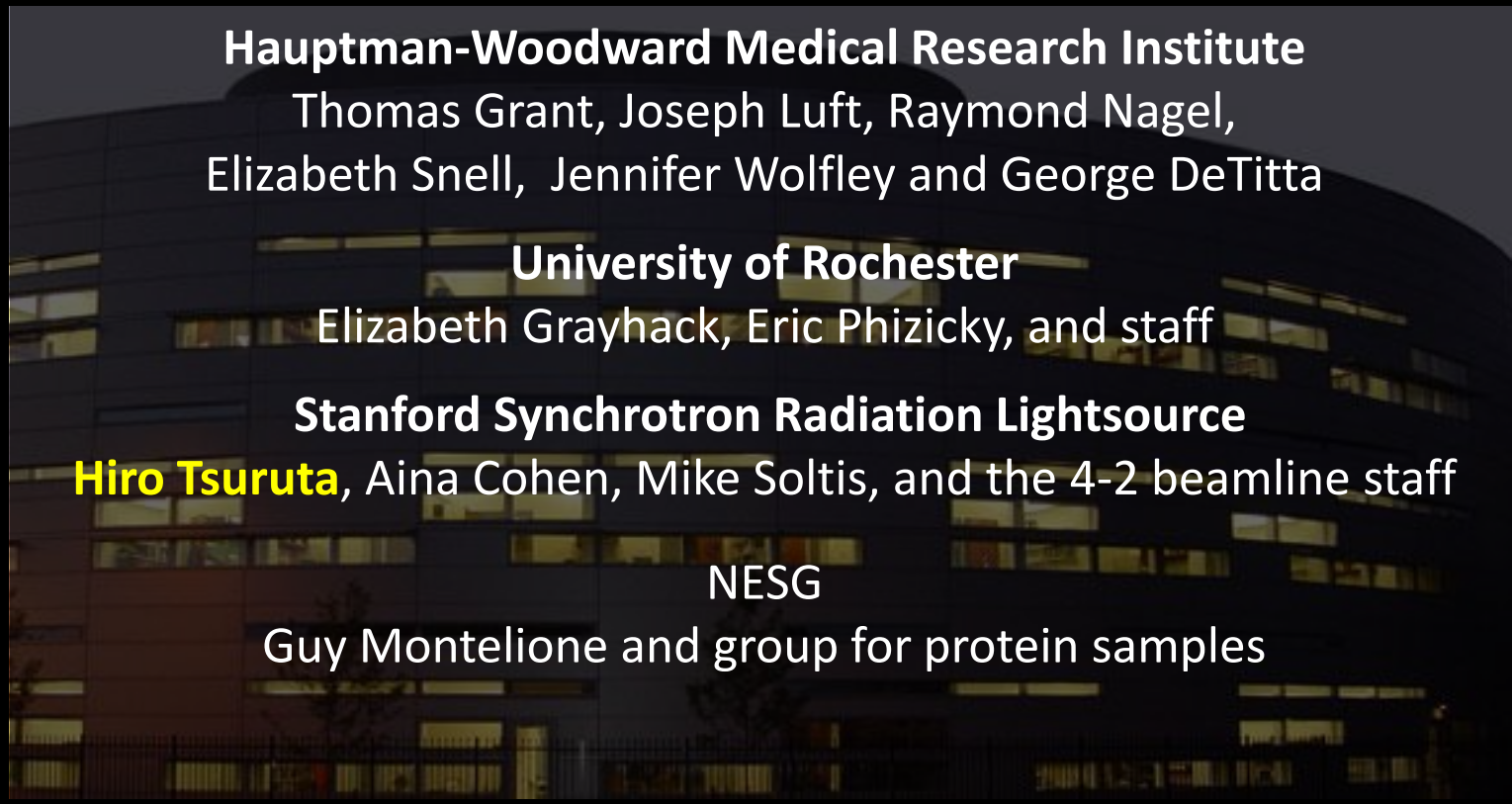
What question do you want answered?

Defining the question is fundamental to reliable conclusions
Ask yes or no questions and decide if SAXS can provide an answer
Model the question – could you see the result in the data?

Resolution of the question determines resolution and quality of the data that is needed, which can effect experimental setup

Sample-detector distance - size of particle versus resolution, oligomers?
Complexes - molecular weight difference, what resolution?
Effect of solution conditions - buffer preparation? Dialysis? Number of concentrations? Serial dilution?
Flexibility - resolution needed for accurate assessment?
Signal to noise - Concentration? Exposure time?

Acknowledgements



Hauptman-Woodward Medical Research Institute

Thomas Grant, Joseph Luft, Raymond Nagel, Elizabeth Snell, Jennifer Wolfley and George DeTitta

University of Rochester

Elizabeth Grayhack, Eric Phizicky, and staff

Stanford Synchrotron Radiation Lightsource

Hiro Tsuruta, Aina Cohen, Mike Soltis, and the 4-2 beamline staff

NESG

Guy Montelione and group for protein samples

Support and Funding

NIH, NSF and DoD support to Edward Snell

Thank you and questions?

esnell@hwi.buffalo.edu