The application of high-throughput technologies to fundamental crystallization research

Edward H. Snell, Thomas Grant, Joseph R. Luft,
and the Center for High-Throughput Structural Biology.

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

An introduction to the 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:

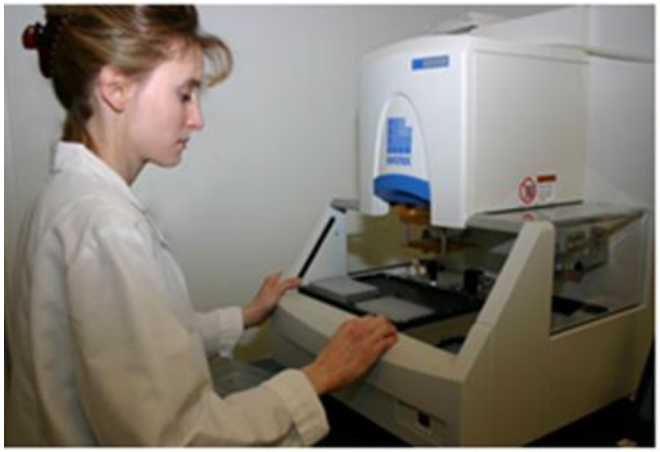
1. buffered (4<pH< 10), highly concentrated salts (35 salts total, sampling 18 different cations and 20 anions) – 229 conditions.
2. PEG/salt/buffer solutions (eight buffers (4<pH< 10), six molecular weight PEGs at three concentrations, and 35 salts at fixed 200 mM concentration) – 721 conditions.

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

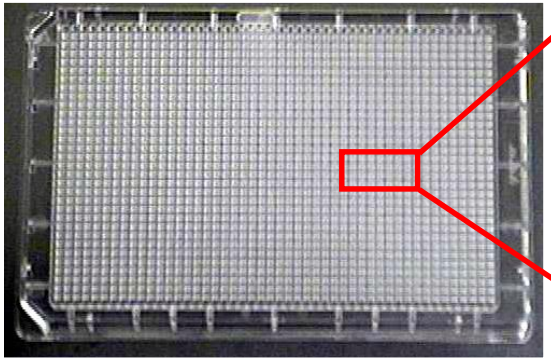
The crystallization method used is micro-batch under oil with 200 nl of protein solution being added to 200 nl of precipitant cocktail in each well of a 1536 well plate.

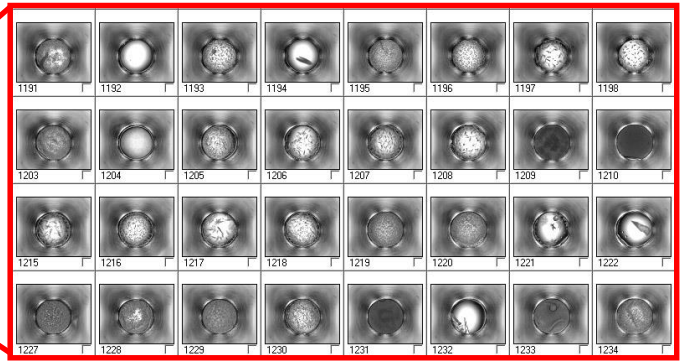
Wells are imaged before filling, immediately after filling then weekly for six weeks duration with images available immediately on a secure ftp server.

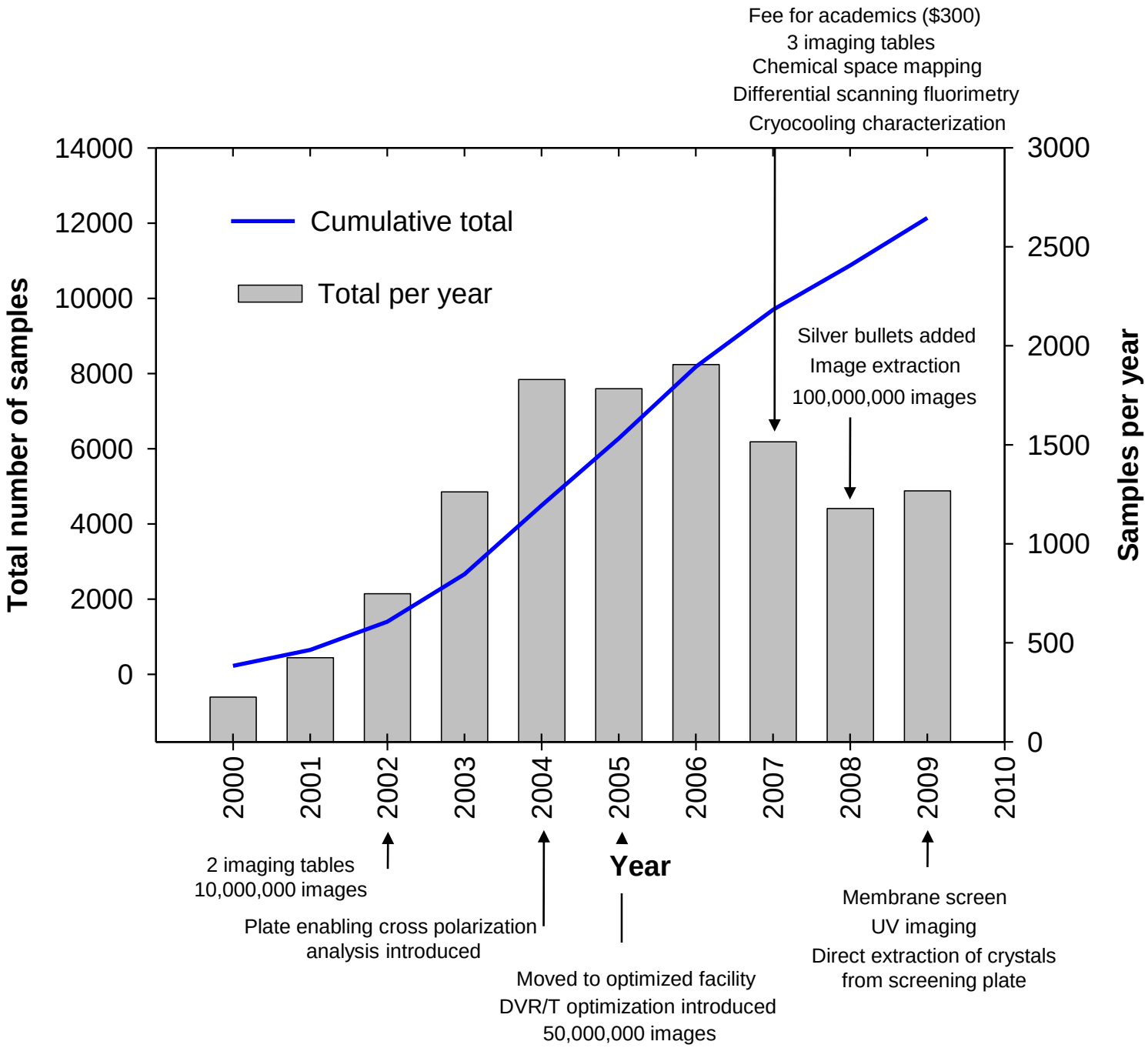
The HTSlab has investigated the crystallization properties of over 12,500 individual proteins archiving over 115,000,000 images of crystallization experiments.

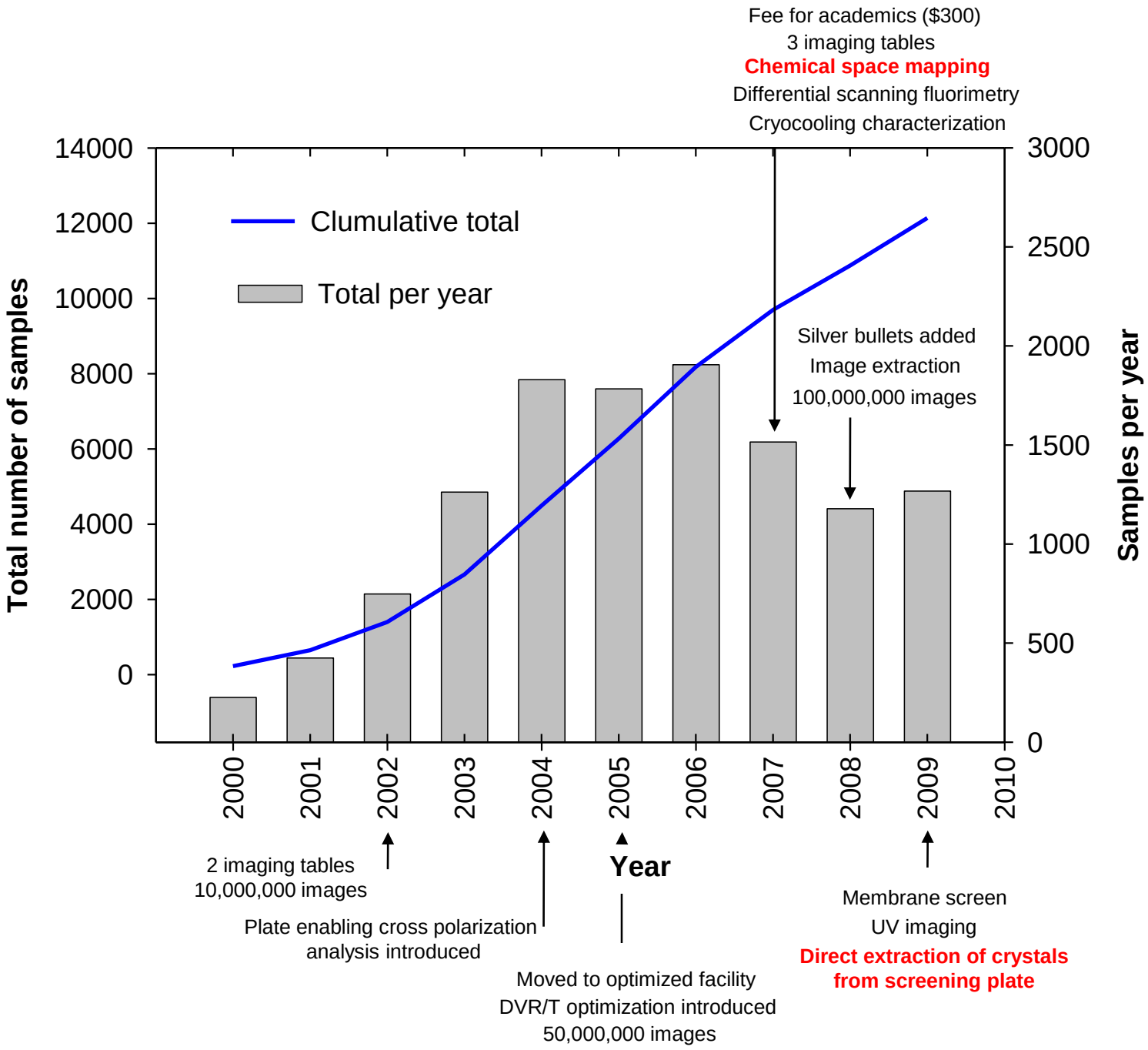


The staff, instrumentation and crystallization plate used






Fee for academics ($300)
3 imaging tables
Chemical space mapping
Differential scanning fluorimetry
Cryocooling characterization
Cumulative total
Total per year
Silver bullets added
Image extraction
100,000,000 images
Total number of samples
Samples per year
14000
12000
10000
8000
6000
4000
2000
0
3000
2500
2000
1500
1000
500
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Year
2 imaging tables
10,000,000 images
Plate enabling cross polarization
analysis introduced
Moved to optimized facility
DVR/T optimization introduced
50,000,000 images
Membrane screen
UV imaging
Direct extraction of crystals
from screening plate


Fee for academics ($300)
3 imaging tables
Chemical space mapping
Differential scanning fluorimetry
Cryocooling characterization
Clumulative total
Total per year
Silver bullets added
Image extraction
100,000,000 images
Total number of samples
Samples per year
Year
2 imaging tables
10,000,000 images
Plate enabling cross polarization
analysis introduced
Moved to optimized facility
DVR/T optimization introduced
50,000,000 images
Membrane screen
UV imaging
Direct extraction of crystals
from screening plate

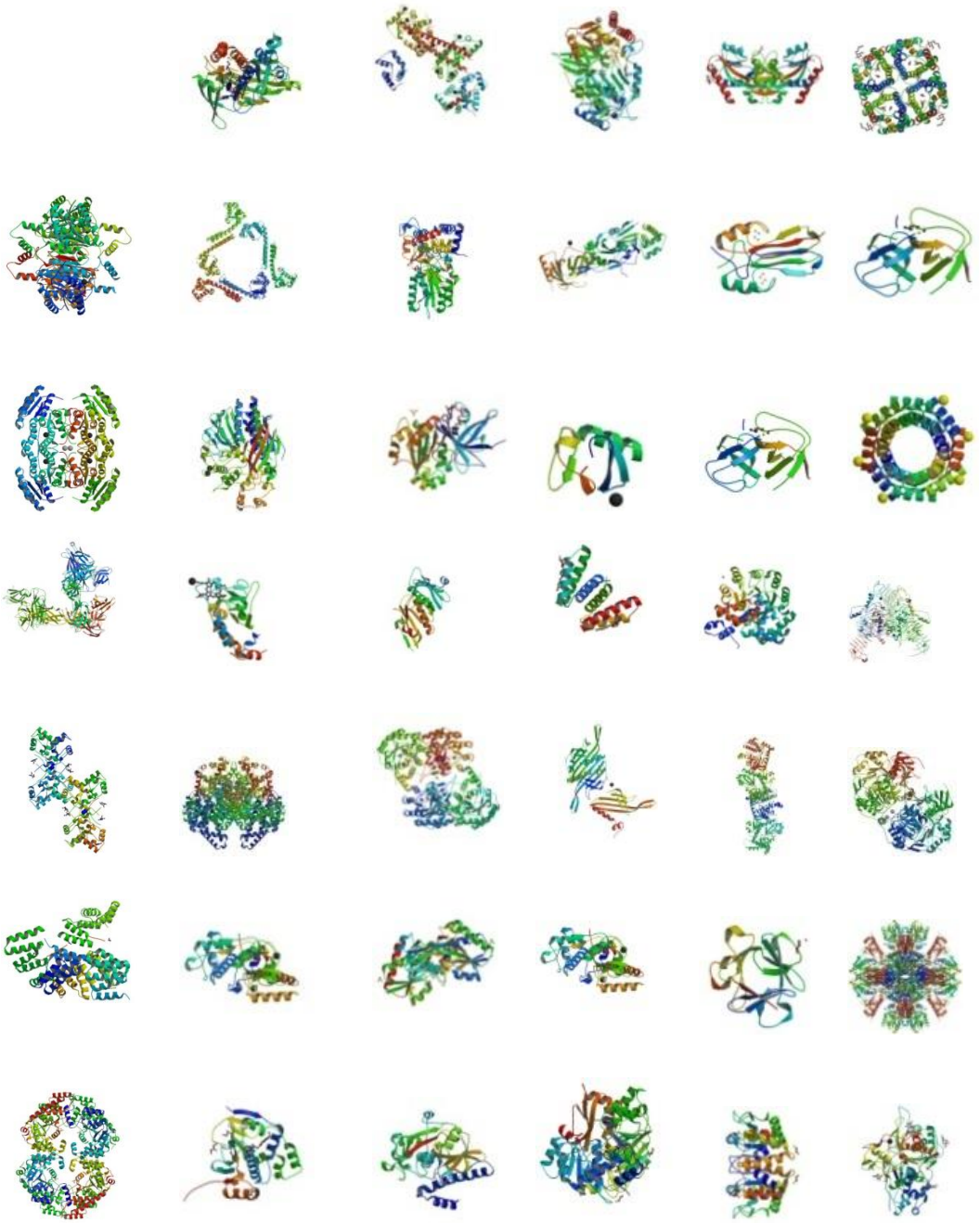
Born in Buffalo

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.



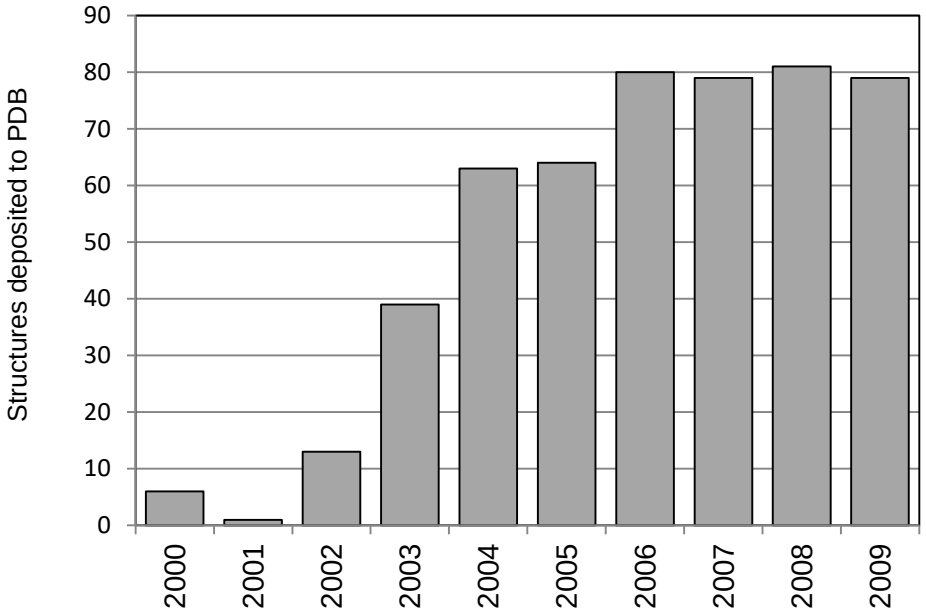
Where success is tracked.

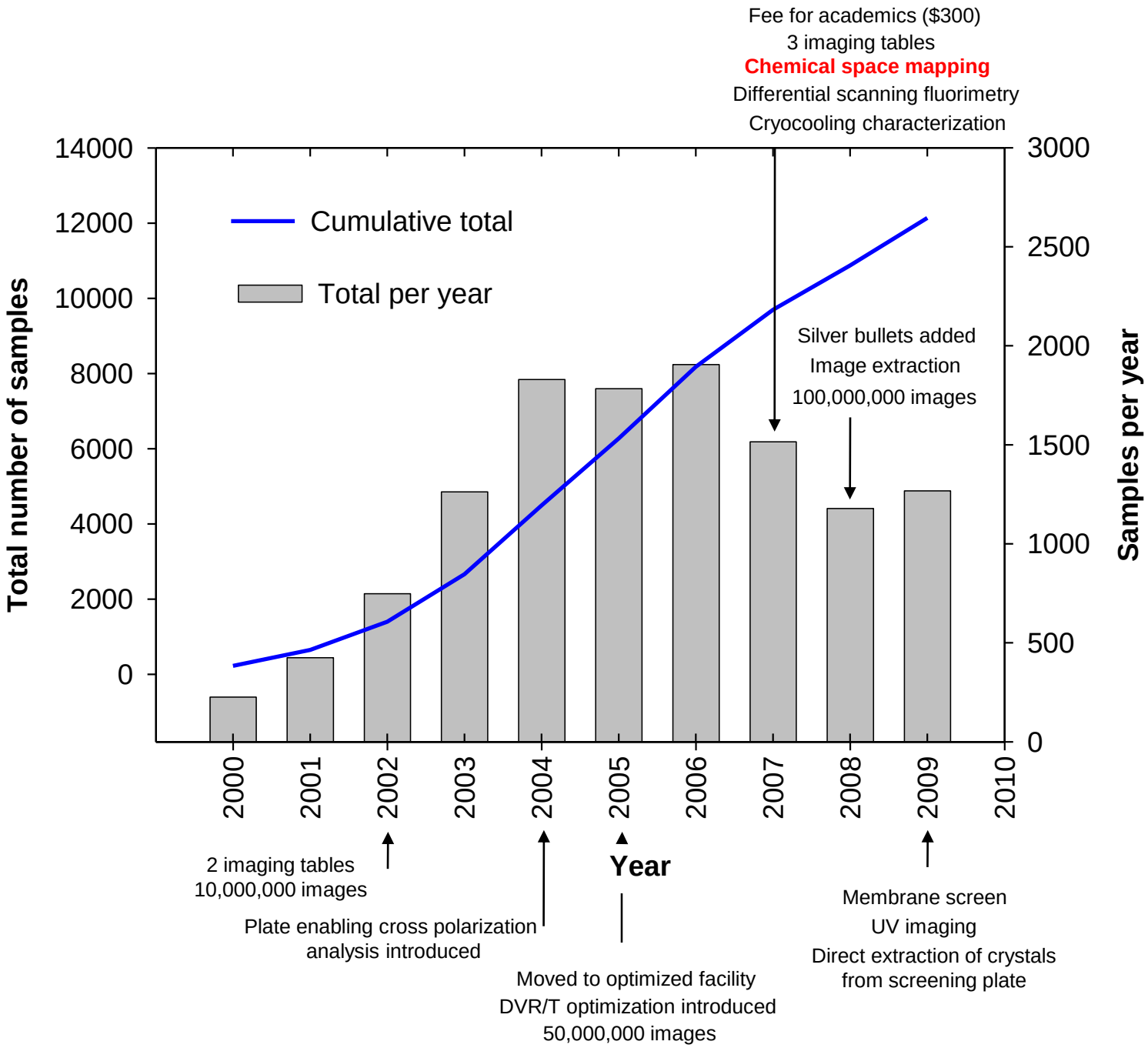
For our Protein Structure Initiative partners both success and failure is tracked. In the case of NESG our initial screening hits enable on average 80 structures per year to be deposited to the PDB.

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

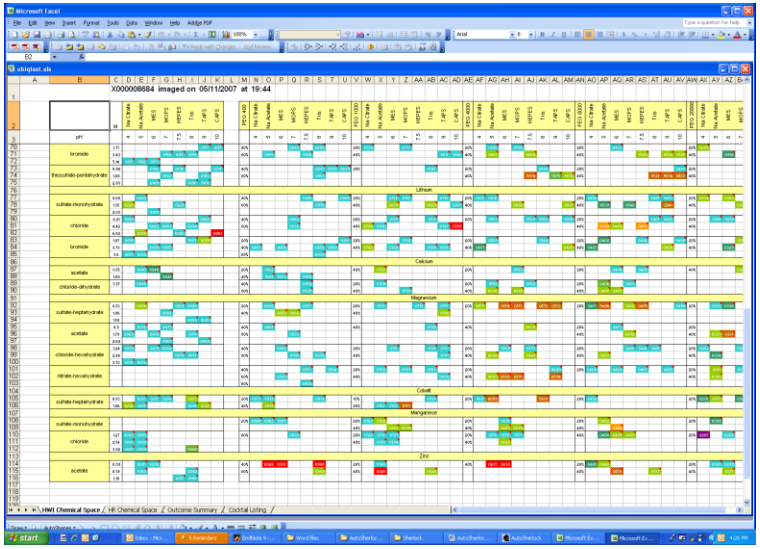
Our success rate from protein in the door to a crystallization hit leading to a PDB deposition is **22%**.

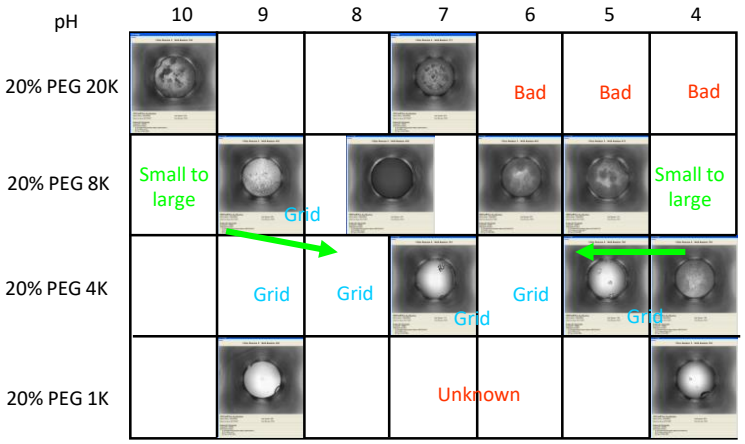
The NESG samples represent a special case in that they are well characterized beforehand – size exclusion chromatography, mass spec analysis and dynamic light scattering studies.




Fee for academics ($300)
3 imaging tables
Chemical space mapping
Differential scanning fluorimetry
Cryocooling characterization
Cumulative total
Total per year
Silver bullets added
Image extraction
100,000,000 images
Total number of samples
Samples per year
Year
14000
12000
10000
8000
6000
4000
2000
0
3000
2500
2000
1500
1000
500
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2 imaging tables
10,000,000 images
Plate enabling cross polarization
analysis introduced
Moved to optimized facility
DVR/T optimization introduced
50,000,000 images
Membrane screen
UV imaging
Direct extraction of crystals
from screening plate

Chemical space mapping





research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The application and use of chemical space mapping to interpret crystallization screening results

Edward H. Snell,[a,b]* Ray M. Nagel,[a] Ann Wojtaszcyk,[a] Hugh O'Neill,[c] Jennifer L. Wolfley[a] and Joseph R. Luft[a,b]

[a]Hauptman–Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, NY 14203, USA, [b]Department of Structural Biology, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA, and [c]Center for Structural Molecular Biology, Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

Correspondence e-mail: esnell@hwi.buffalo.edu

Received 17 June 2008
Accepted 7 October 2008

Macromolecular crystallization screening is an empirical process. It often begins by setting up experiments with a number of chemically diverse cocktails designed to sample chemical space known to promote crystallization. Where a potential crystal is seen a refined screen is set up, optimizing around that condition. By using an incomplete factorial sampling of chemical space to formulate the cocktails and presenting the results graphically, it is possible to readily identify trends relevant to crystallization, coarsely sample the phase diagram and help guide the optimization process. In this paper, chemical space mapping is applied to both single macromolecules and to a diverse set of macromolecules in order to illustrate how visual information is more readily understood and assimilated than the same information presented textually.

computer programs

Journal of
Applied Crystallography
ISSN 0021-8898

Received 17 June 2008
Accepted 9 September 2008

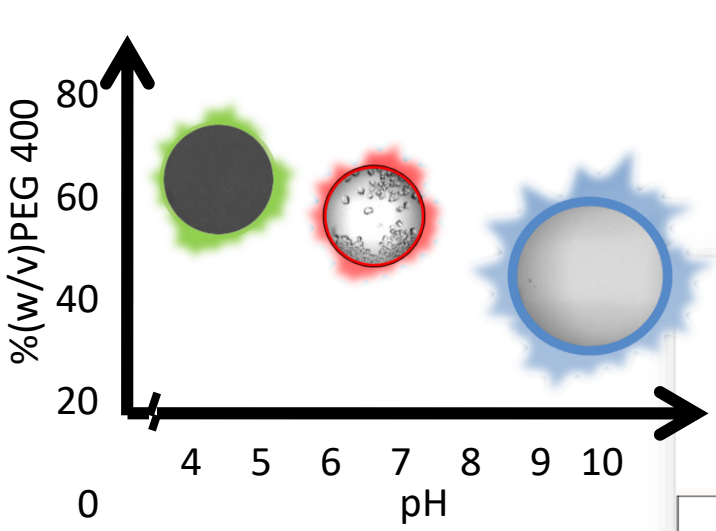
AutoSherlock: a program for effective crystallization data analysis

Raymond M. Nagel,[a] Joseph R. Luft[a,b] and Edward H. Snell[a,b]*

[a]Hauptman–Woodward Medical Research Institute, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA, and [b]Department of Structural Biology, SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203, USA. Correspondence e-mail: esnell@hwi.buffalo.edu

A program, *AutoSherlock*, has been developed to present crystallization screening results in terms of chemical space. This facilitates identification of lead conditions, rational interpretation of results and directions for the optimization of crystallization conditions.

Chemical space mapping (analysis)



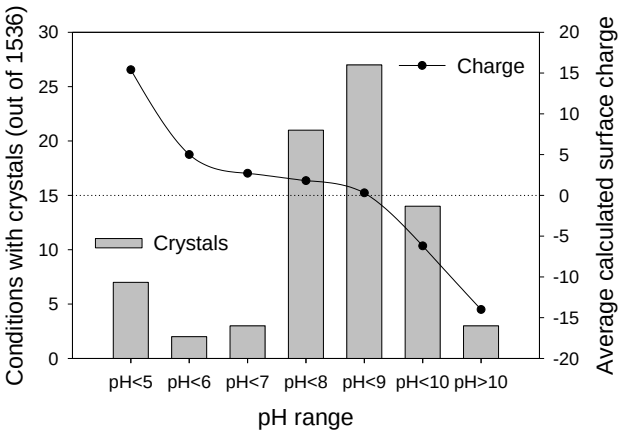
Un-sampled
Crystals
Clear
Precipitate

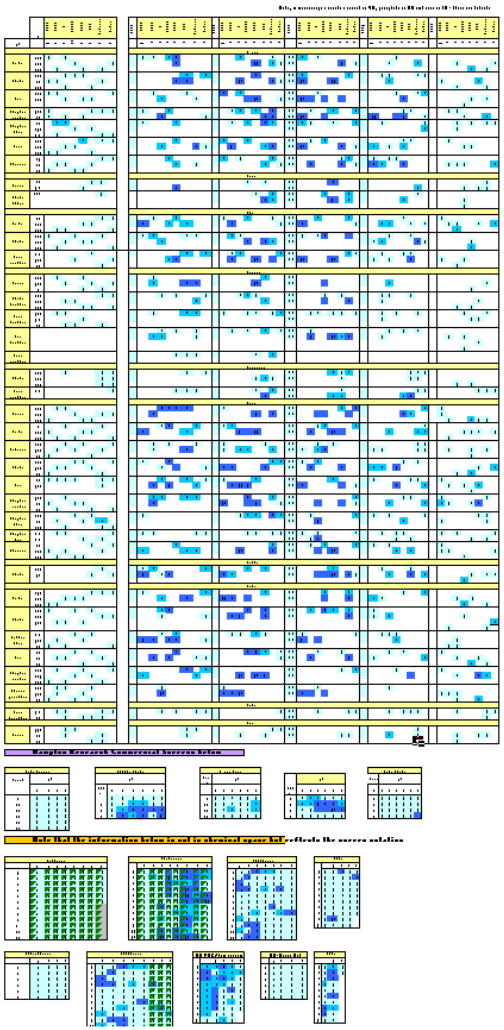
		PEG 400	Na Citrate	Na Acetate	MES	Bis-Tris Propane	HEPES	Tris	TAPS	CAPS
			Lithium							
pH			4.2	5	6	7	7.5	8	9	10
bromide		20%								●
		40%	●		●			●	●	
		80%						●		

Chemical space mapping

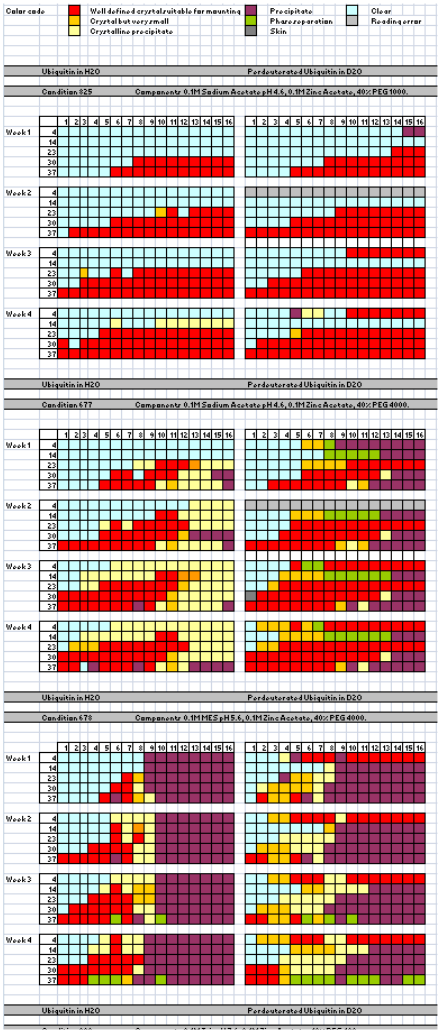
Analysis as a function of the entire cocktail screen and multiple proteins

Or a single protein and a variation on biochemical conditions e.g. pH





Or as a variable of time



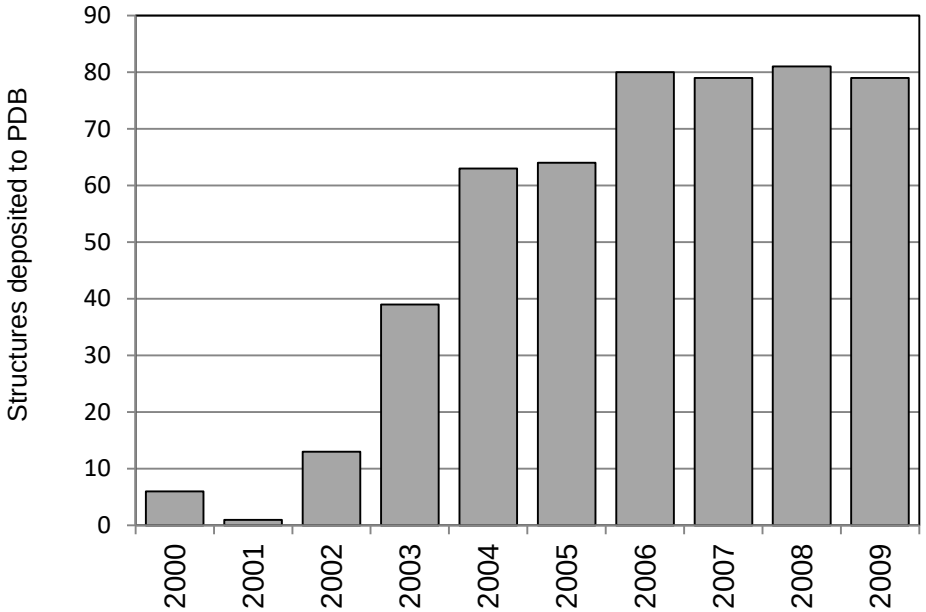
Where success is tracked.

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.

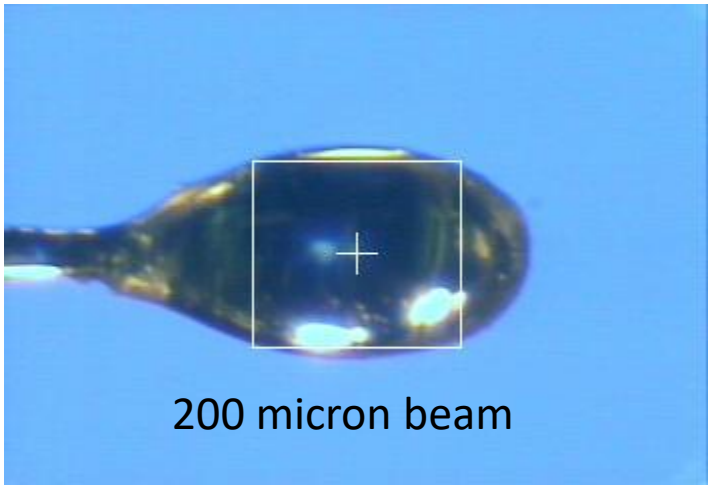


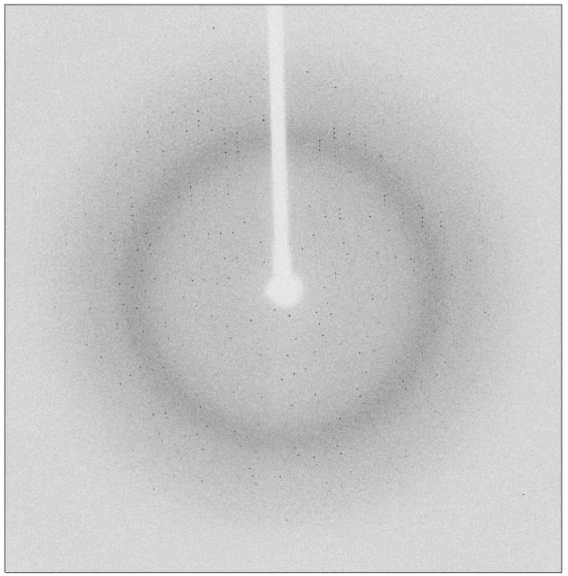
We see hits in ~50% of cases, only 22% go to the PDB – Why?

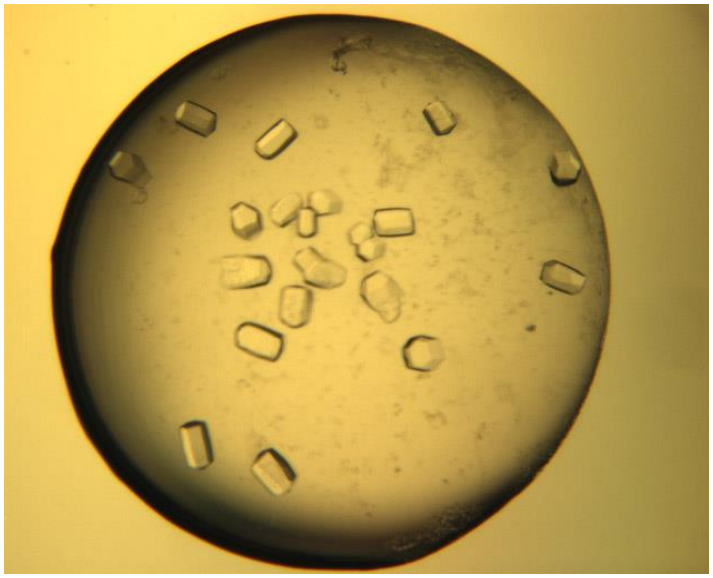
Complexes are less successful – can we overcome this?

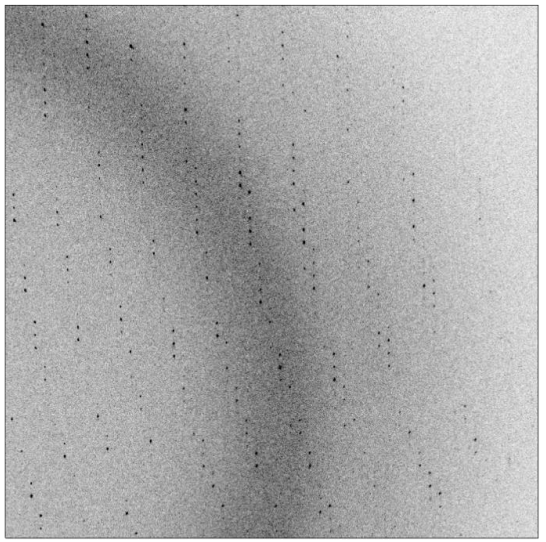
A case study

- A tRNA synthetase from a eukaryotic system







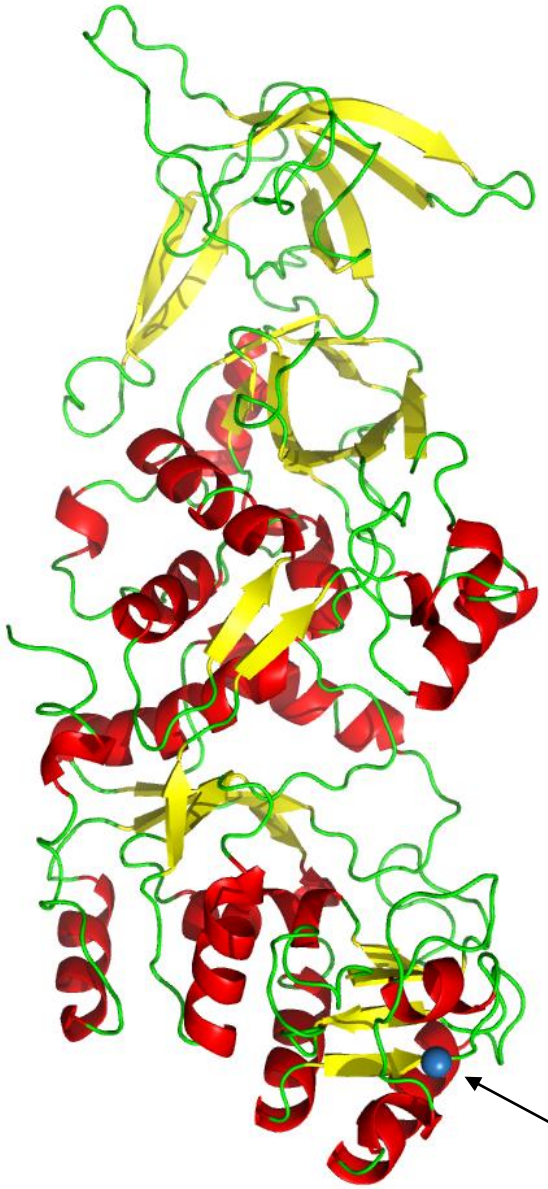


80% PEG 400 in the crystallization cocktail

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˚

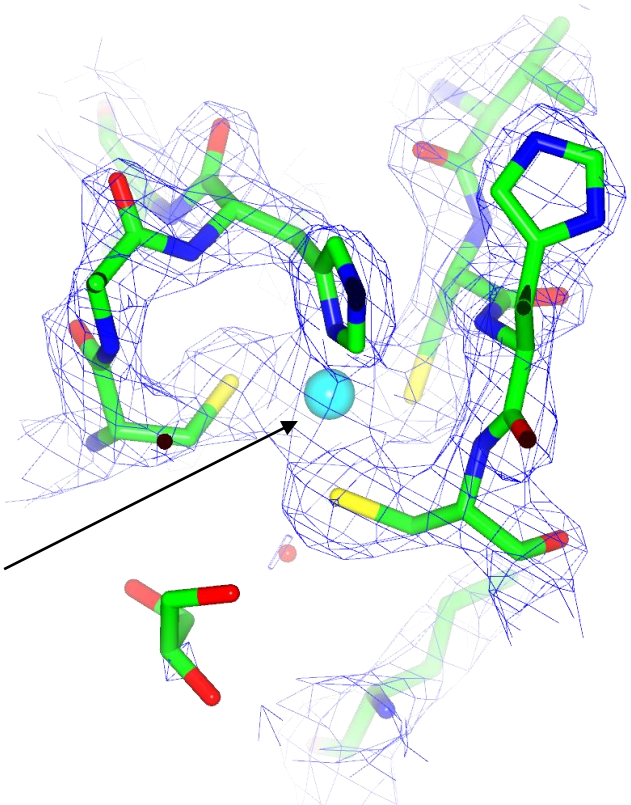
	Overall	Inner Shell	Outer Shell
Low resolution limit (Å)	40.00	40.00	2.64
High resolution limit (Å)	2.5	7.91	2.5
R_{merge}	0.104	0.036	0.743
R_{pim}	0.032	0.011	0.273
	3.2%	1.1%	27.3%
Total number of observations	508484	17694	51511
Total number unique	44752	1523	6332
Mean((I)/sd(I))	24.6	86.6	2.2
Completeness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

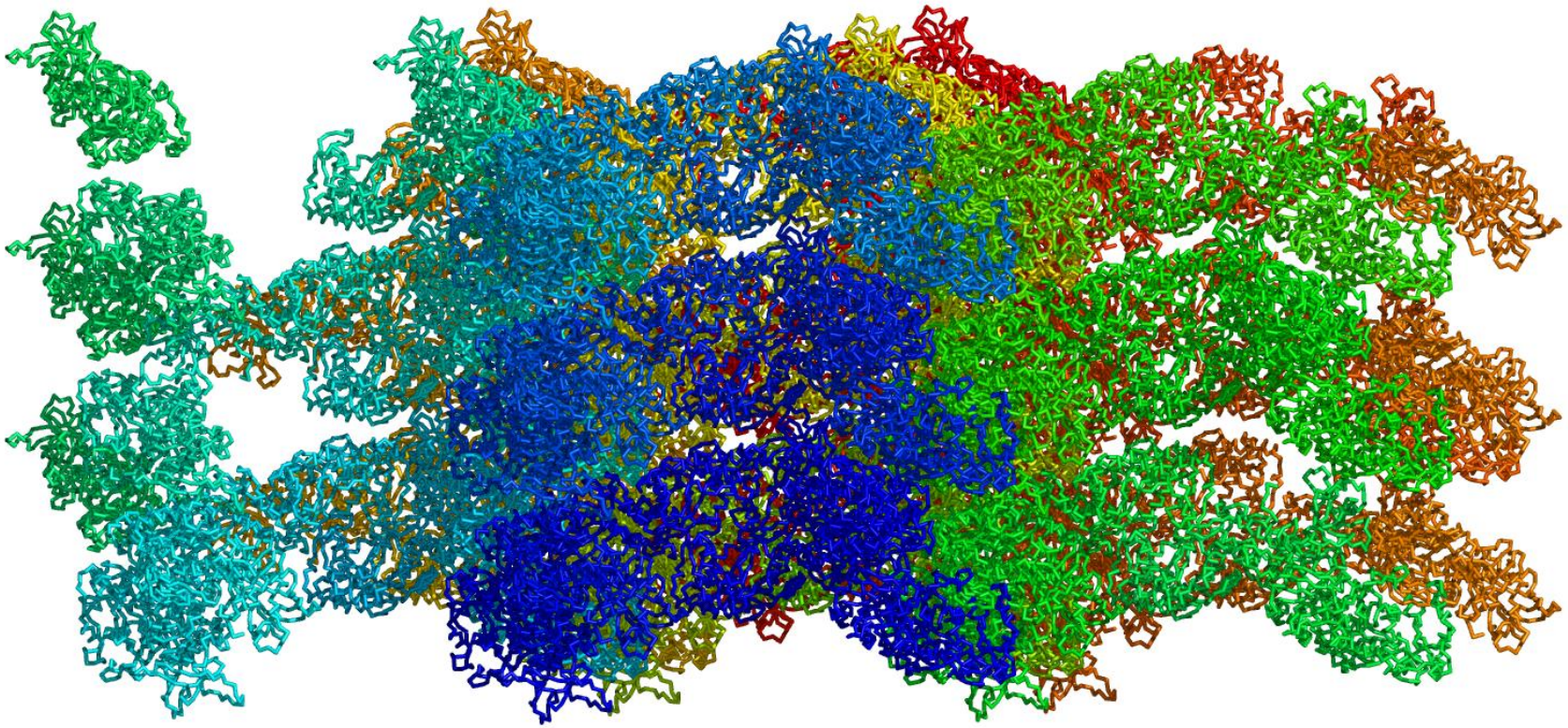


Zinc

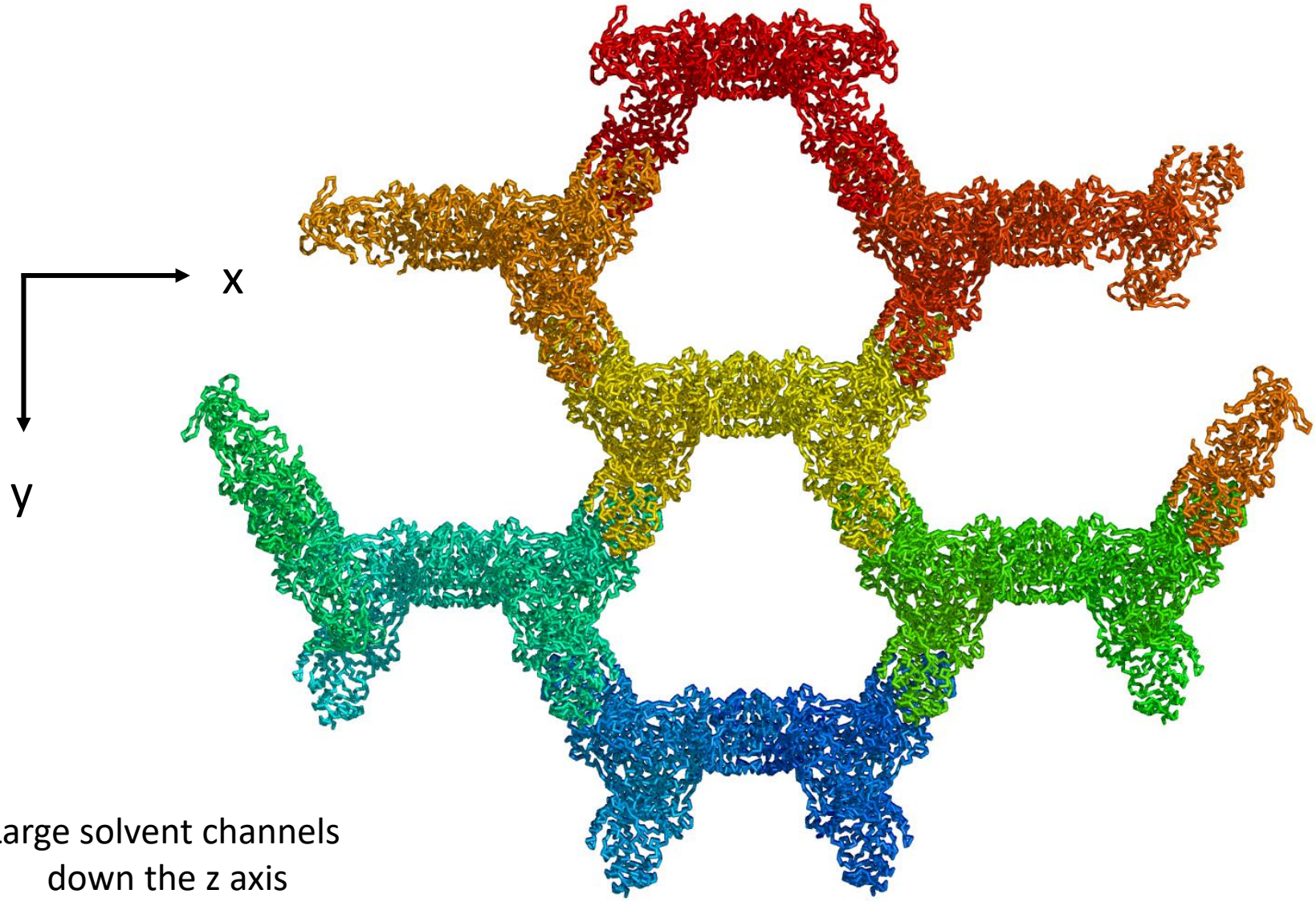
Structure solved (with help of the zinc signal) and refined with Phenix.

Overall R and R_{free} are 14.2 and 19.8% respectively.

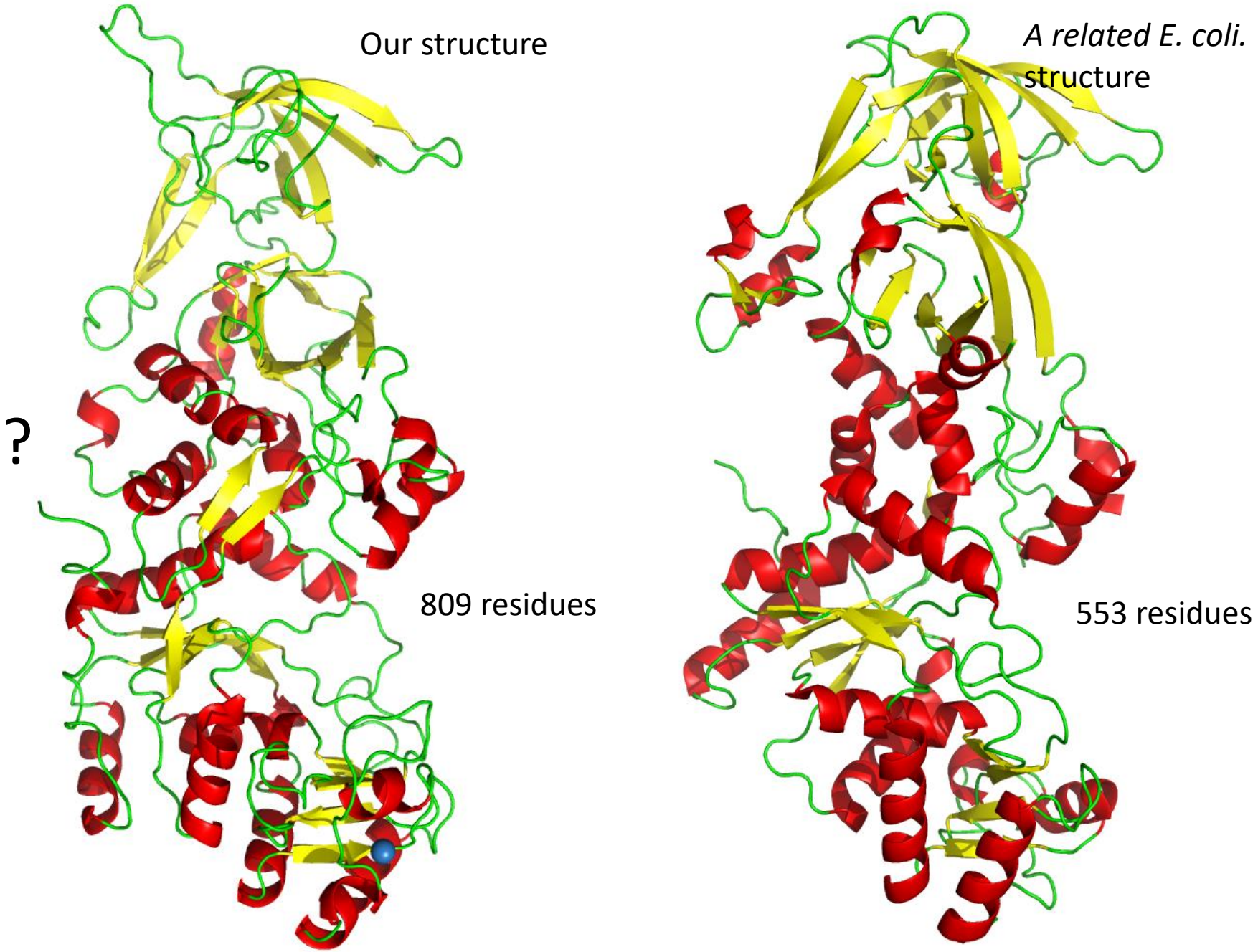


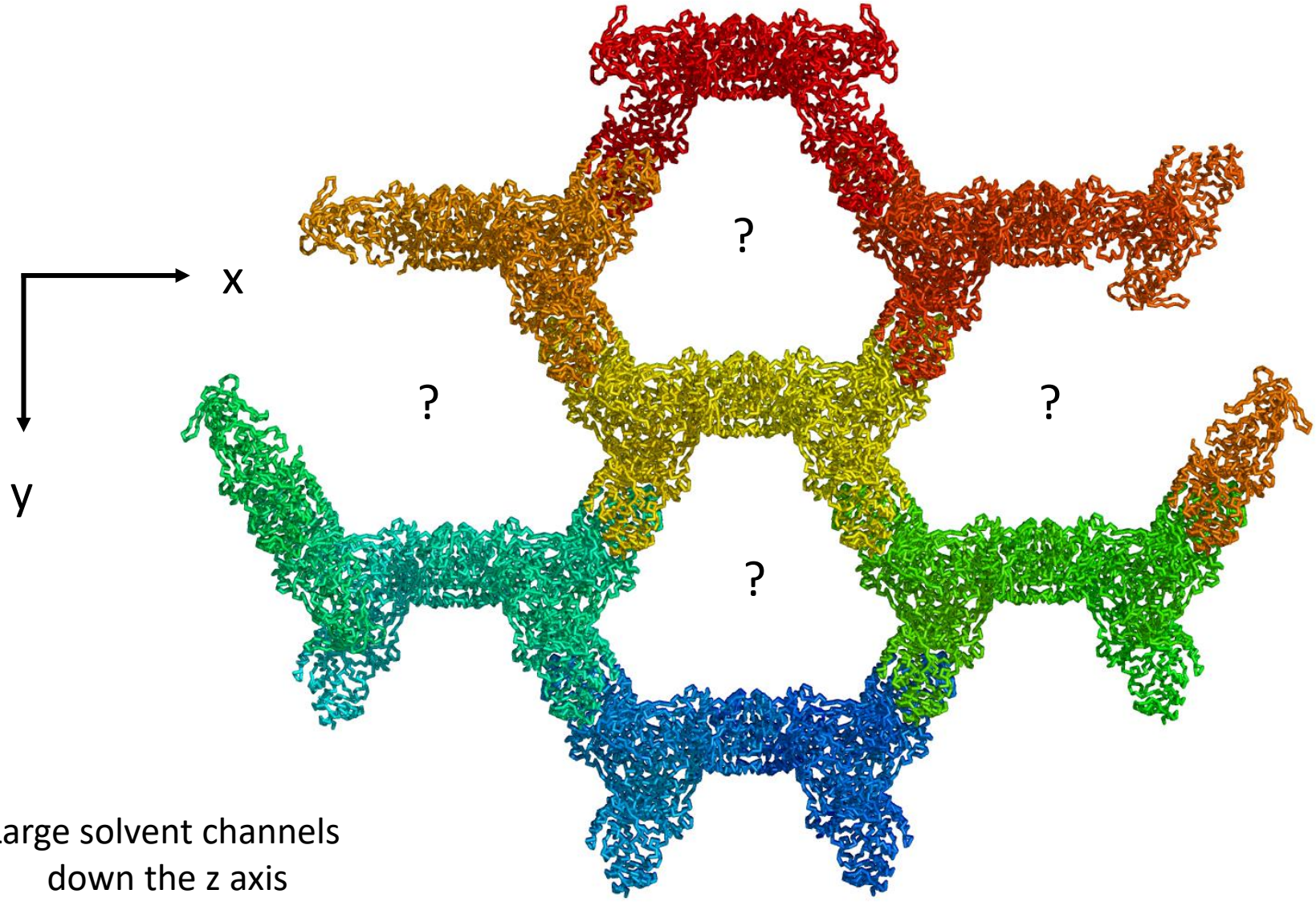


Tight packing in z and y



Large solvent channels
down the z axis


Our structure
?
809 residues
A related E. coli. structure
553 residues



Large solvent channels
down the z axis

Missing residues

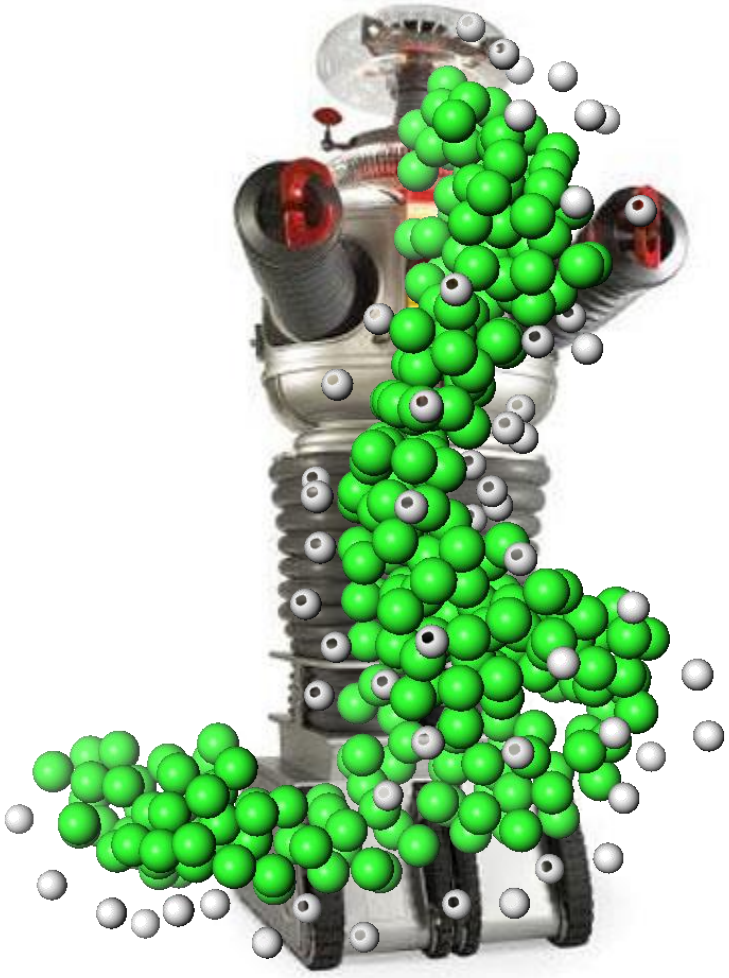
- There were 216 missing residues from the structure, 95% of the N-terminal domain.
- Where they in the mix to start with?.
- SDS PAGE gel on the remaining crystals indicated that the full length protein was present.
- For a more concrete answer the protein was re-expressed with a His tag attached to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

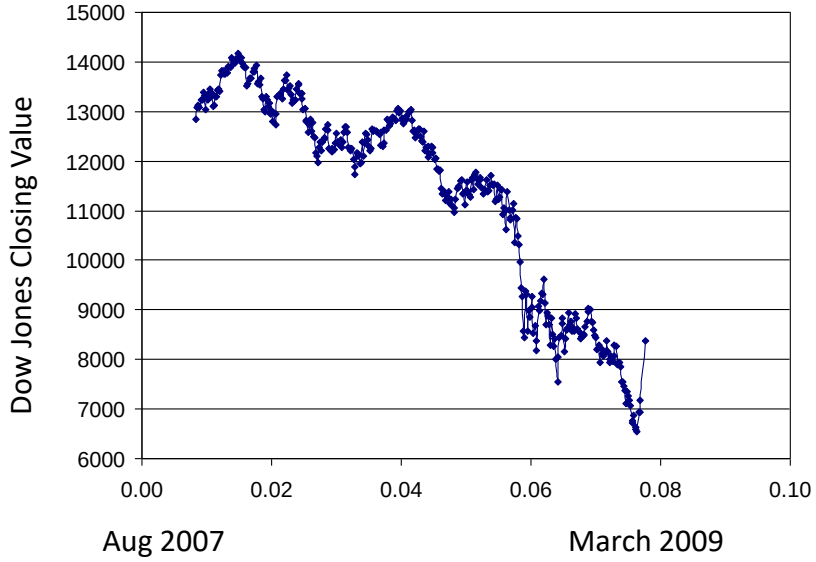
SAXS is complementary to crystallography

- The SAXS solution can be validated when other information is present (**we have crystallographic information**).
- In a non-symmetrical case (most examples) the X-ray derived structure can be fitted to the envelope (**our system is non-symmetrical, at least the known part**).
- The scattering curve from SAXS is derived from a summation of all the particles in solution. It is radially averaged over all these particles (losing 3D information) but samples all positions of the particle and all conformations (sampling dynamic information, i.e. **we can visualize the part missing in the crystal structure**).

- SAXS is sensitive to information that crystallography does not see.
- SAXS is sensitive to dynamics.
- SAXS is a low-resolution technique - Crystallography is sensitive to information SAXS cannot see.
- **Crystallography and SAXS are complementary.**

Can we use X-ray solution scattering?



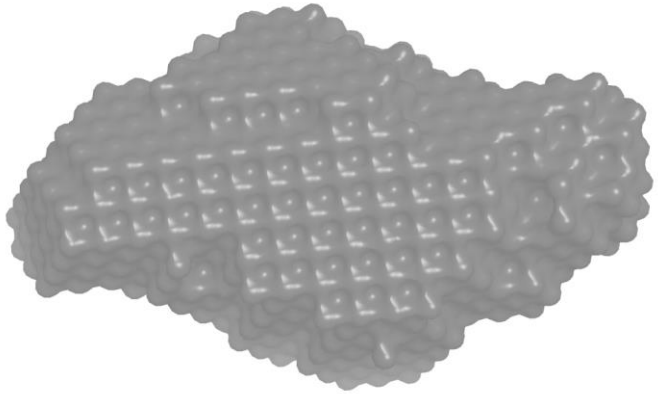


The scattering data from SAXS provides a 1D Fourier transform of the envelope of the particle.

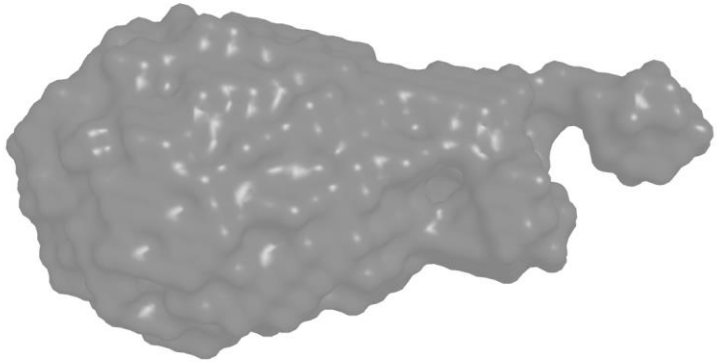
It's possible to fit multiple envelopes to the data.

You will always get an envelope despite the data!

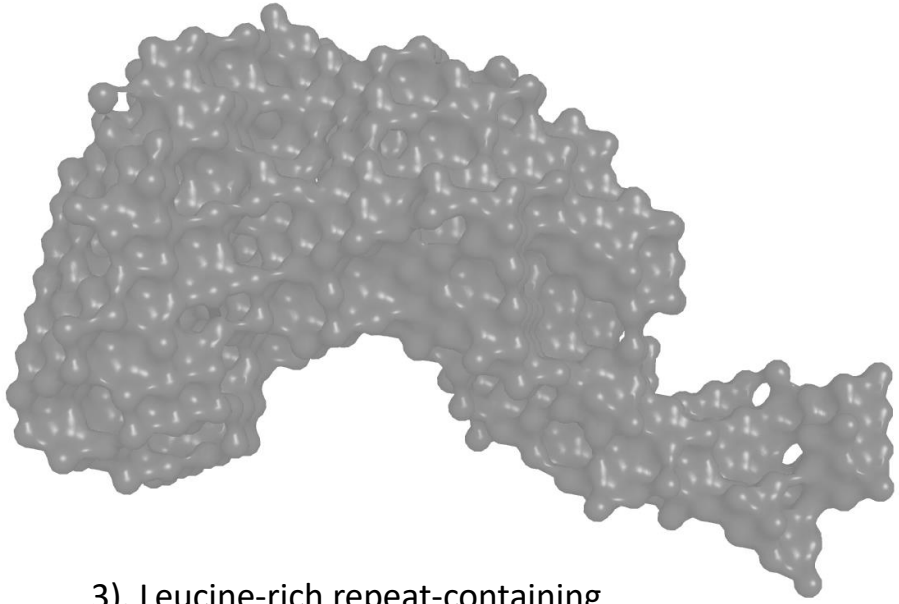
Ab intio envelopes from samples generously supplied by NESG



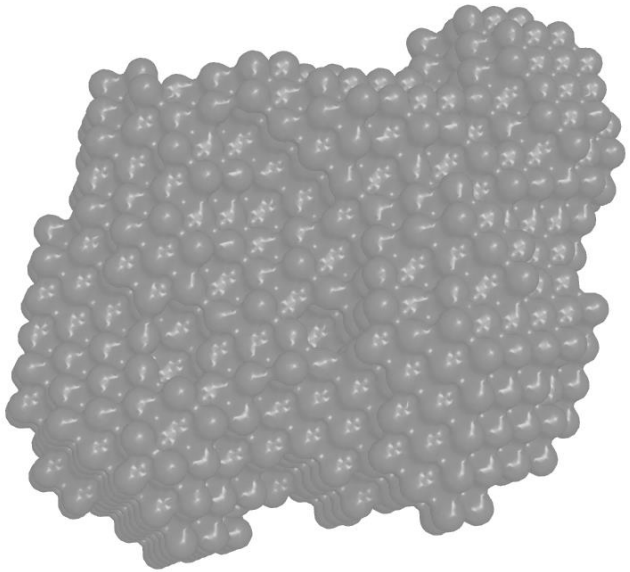
1). alr0221 protein from Nostoc (18.6 kDa)



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

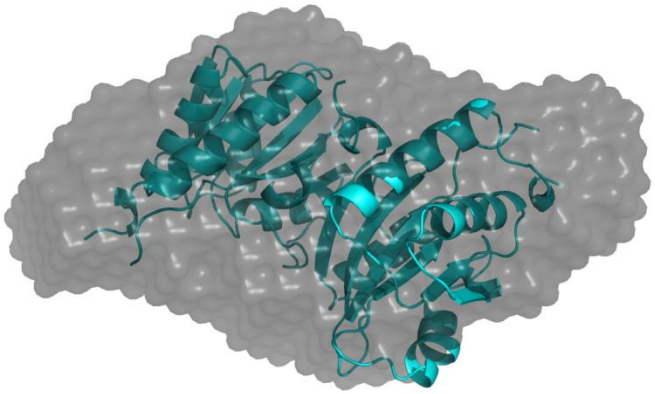


3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

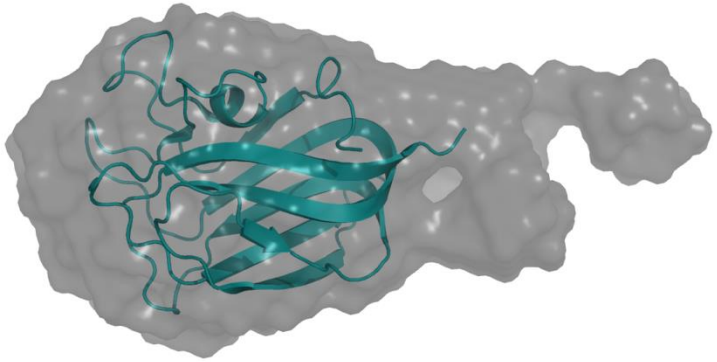


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

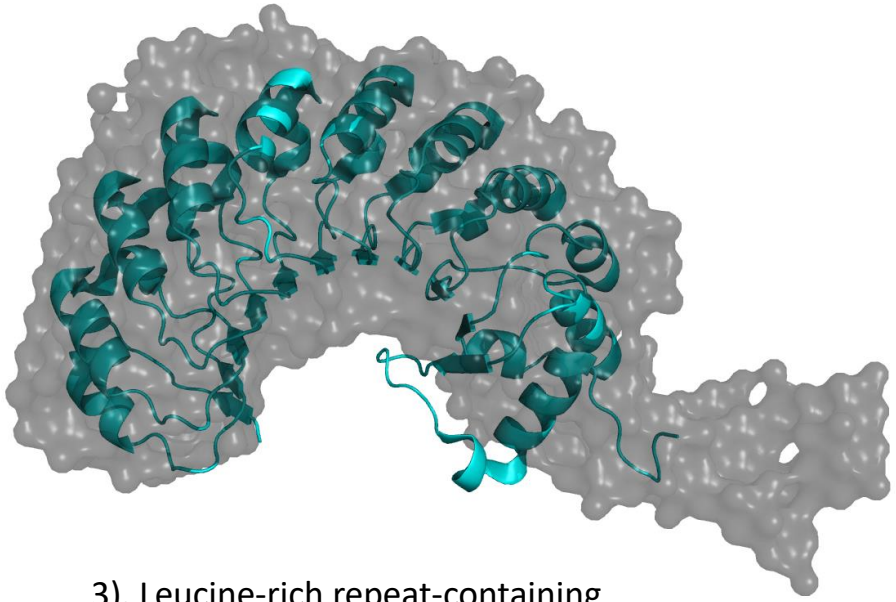
Overlaid with subsequent X-ray structures



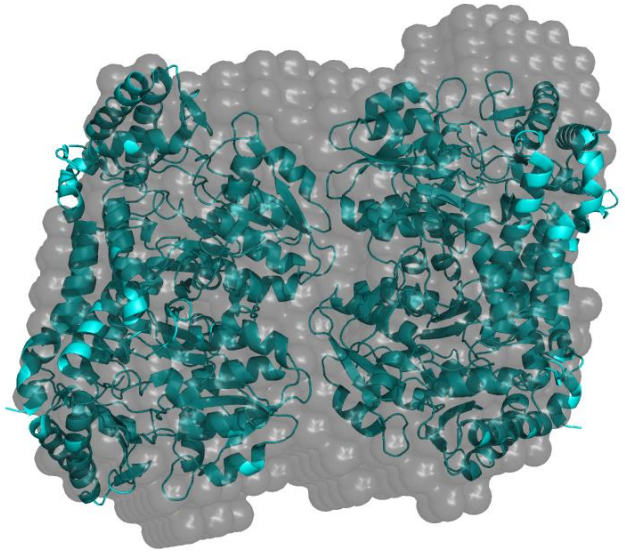
1). alr0221 protein from Nostoc (18.6 kDa)



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

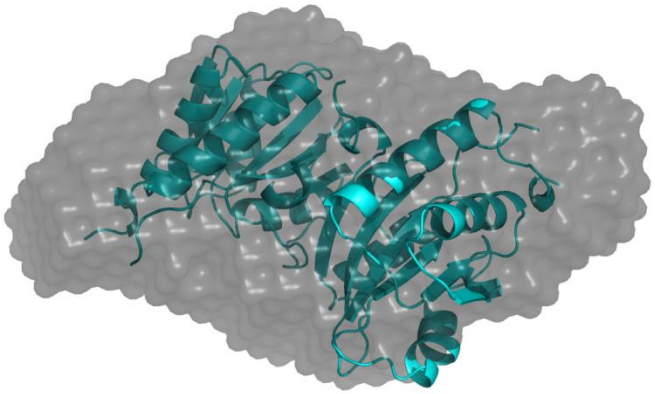


3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

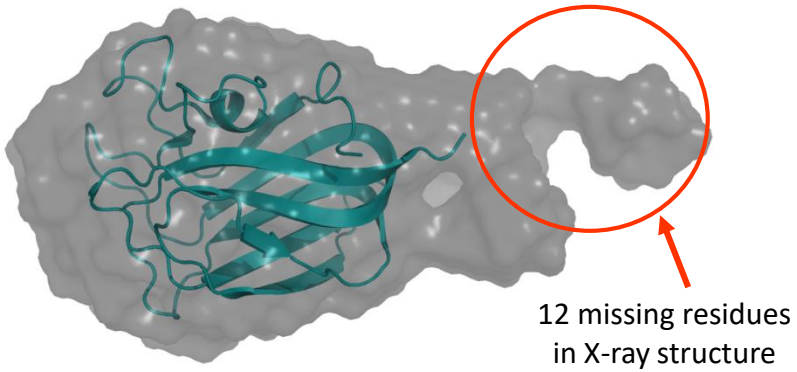


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

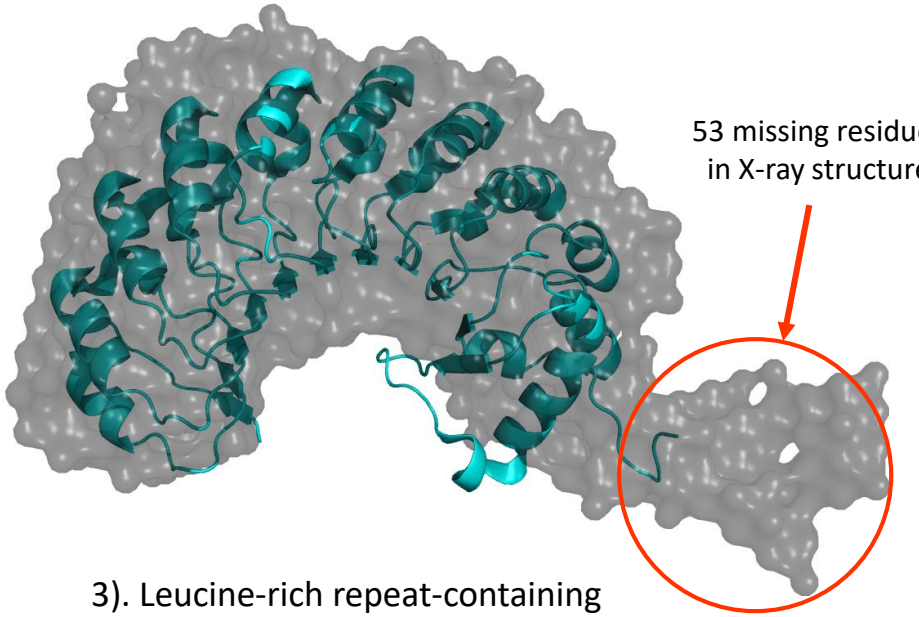
And data on what was missing … (in our case SAXS is reliable)



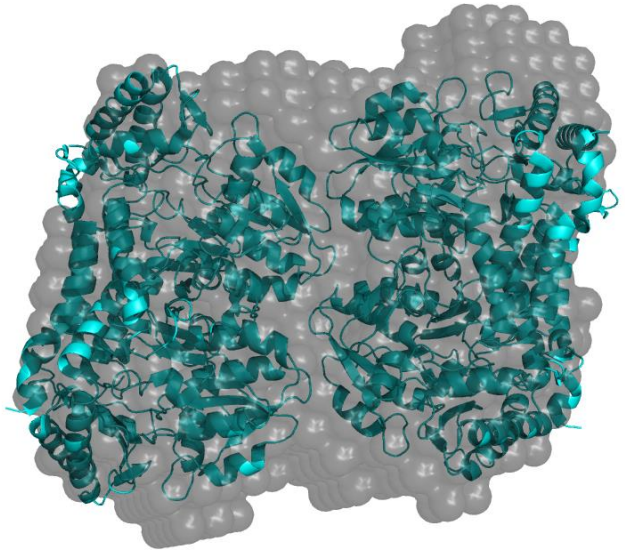
1). alr0221 protein from Nostoc (18.6 kDa)



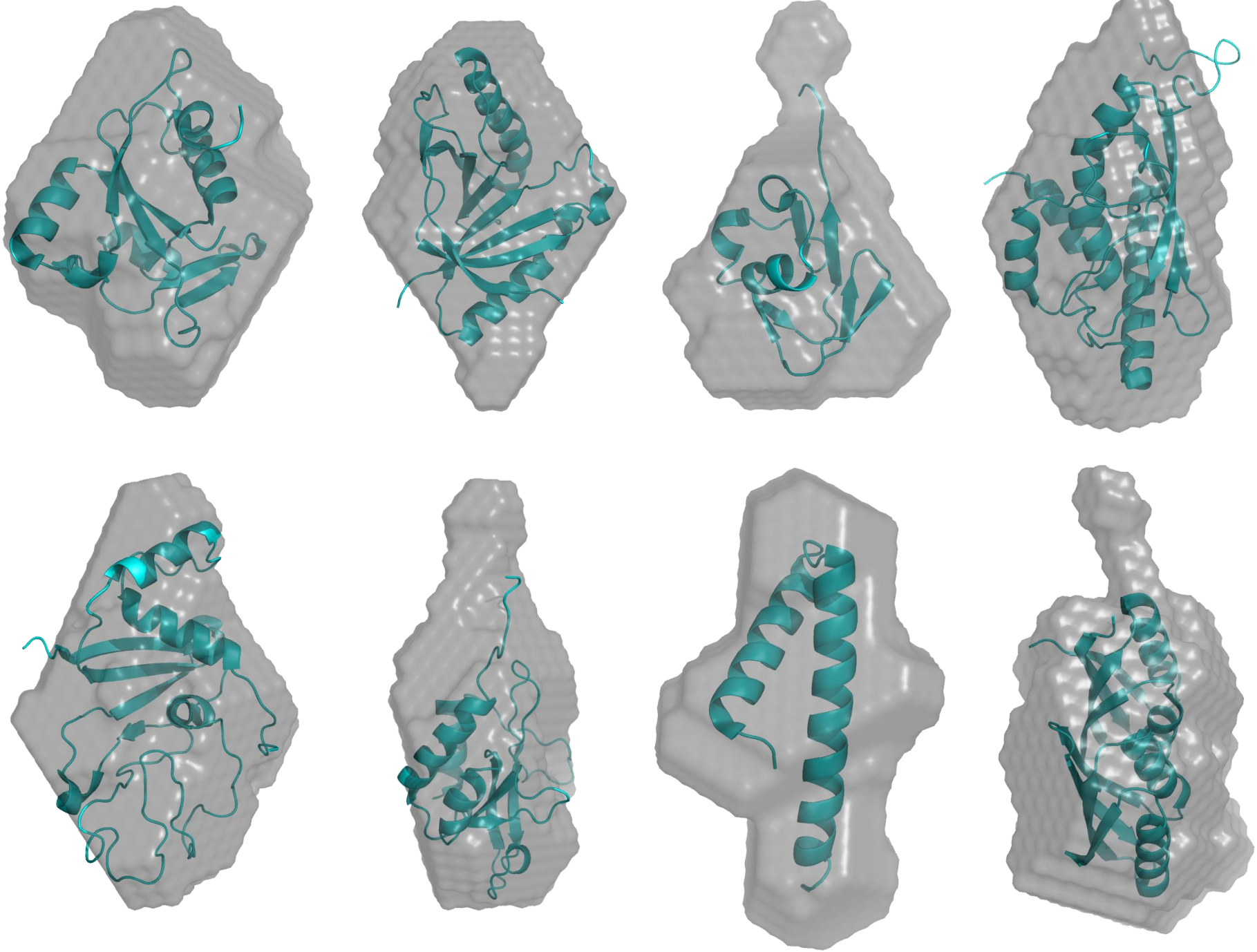
2). C-terminal domain of a chitobiase (17.9 kDa)



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

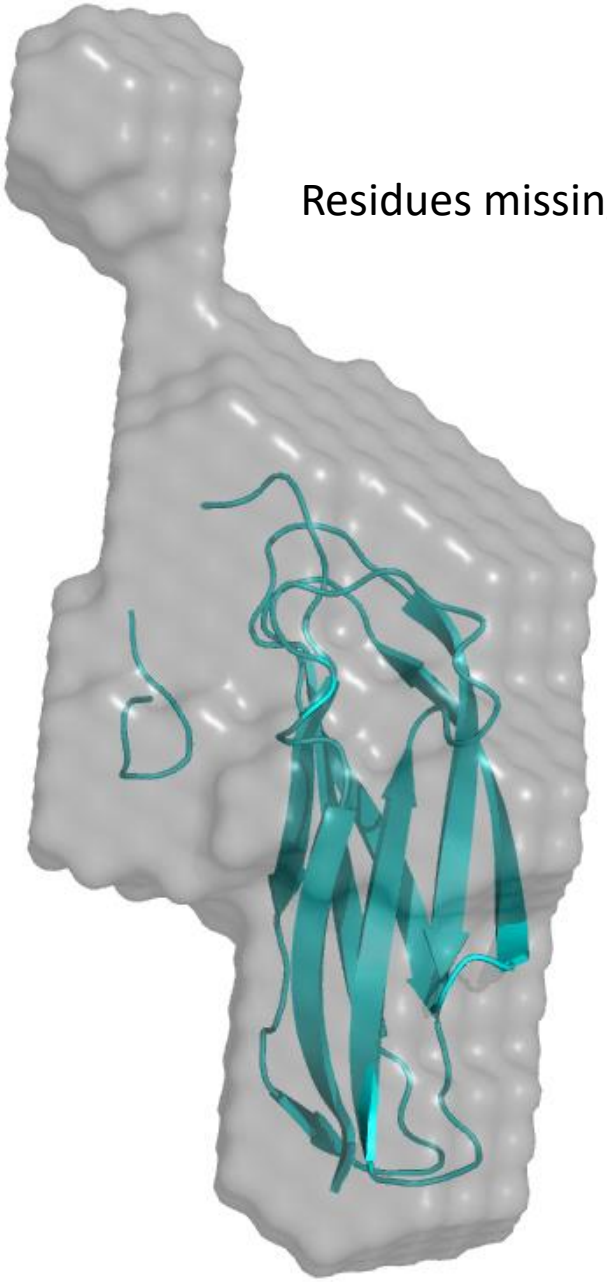


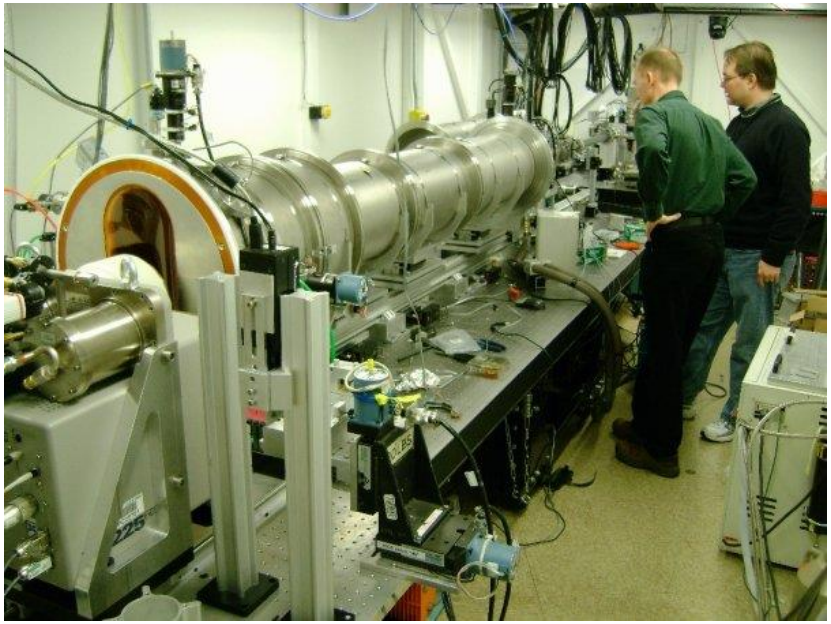
4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

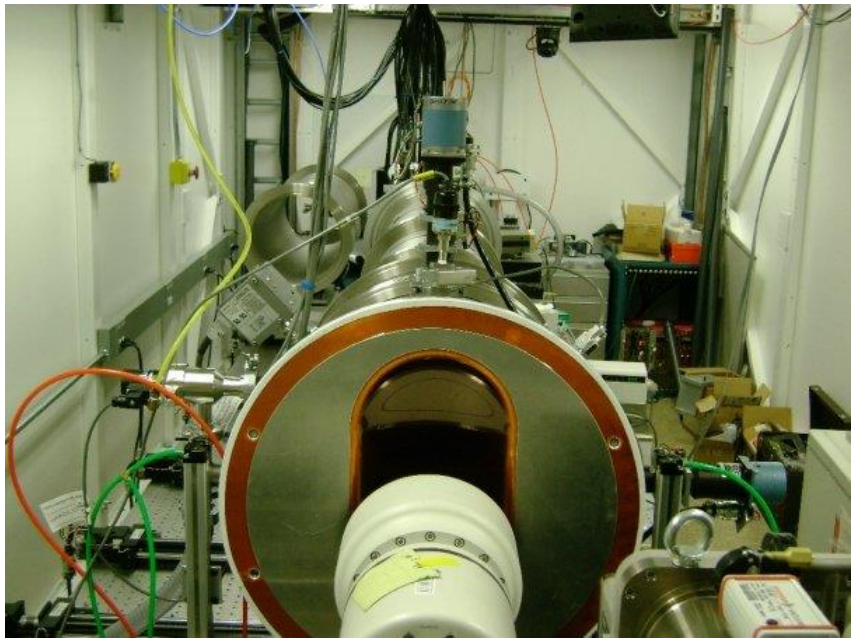


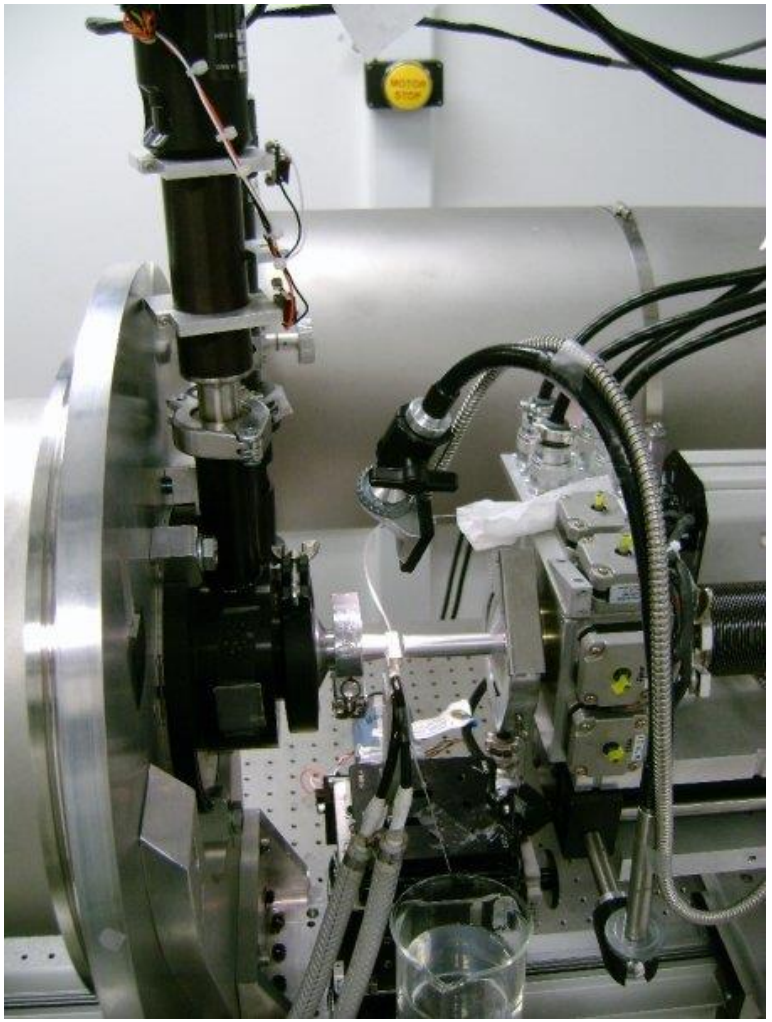
Residues missing in both X-ray and SAXS

X-ray and SAXS

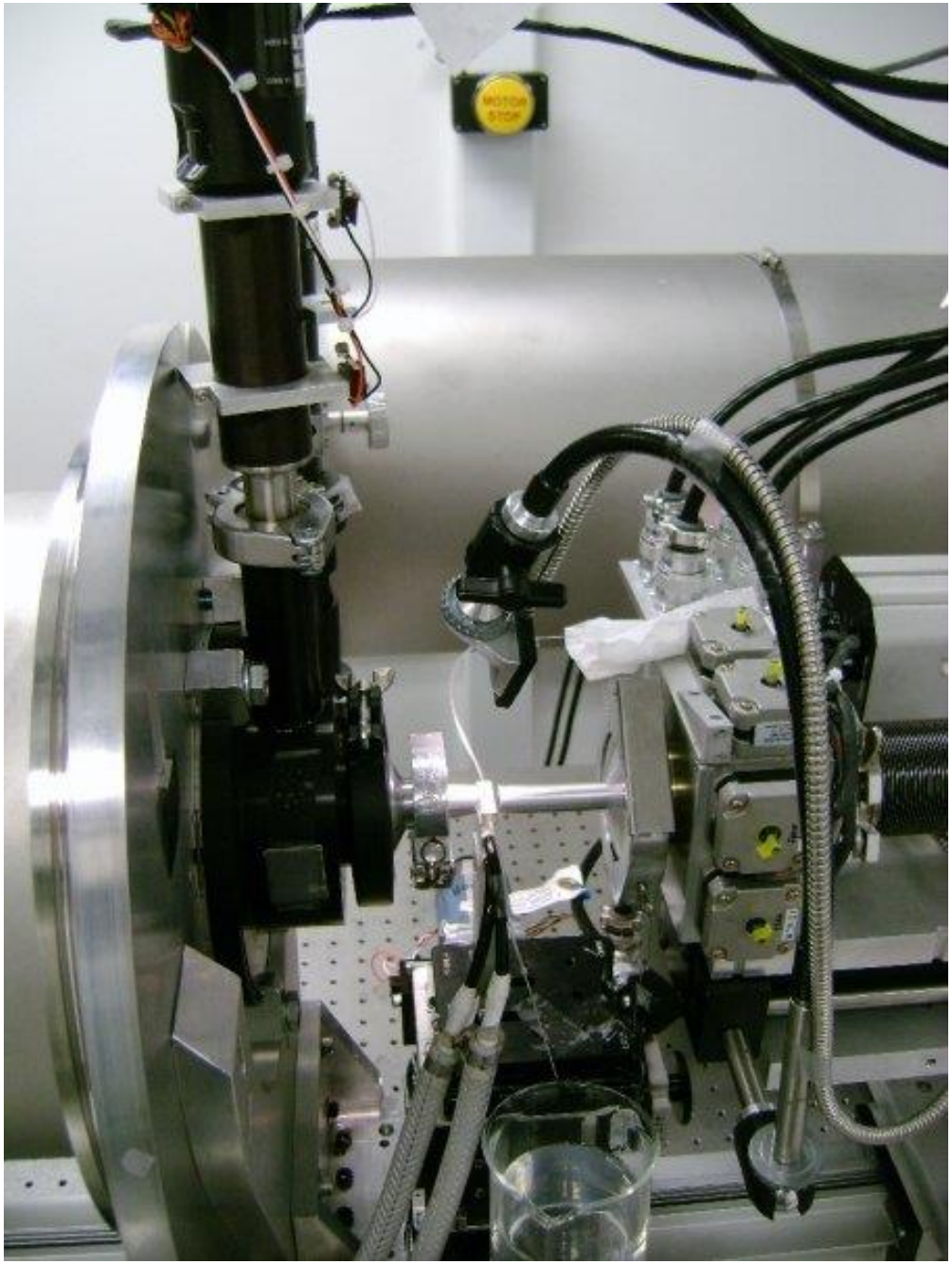








SSRL Beamline 4-2 (Being developed with a high-throughput capability)



Initial protocol

5 concentrations

Start with buffer then lowest concentration first

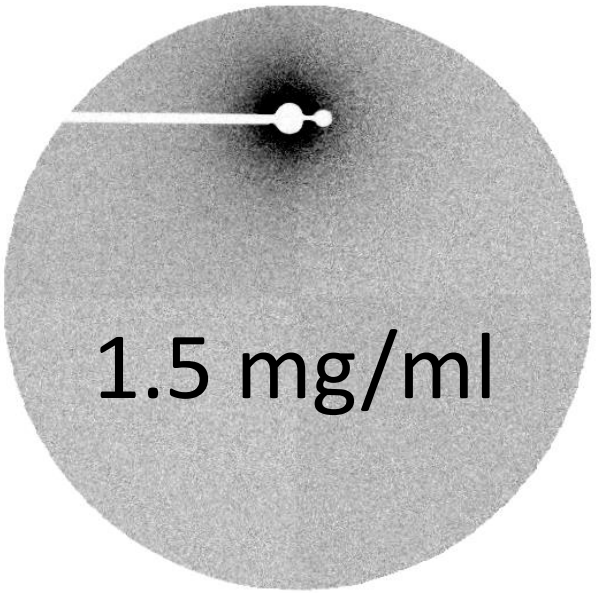
Using the lowest first means that residual protein in the capillary would not alter the assumed concentration greatly.

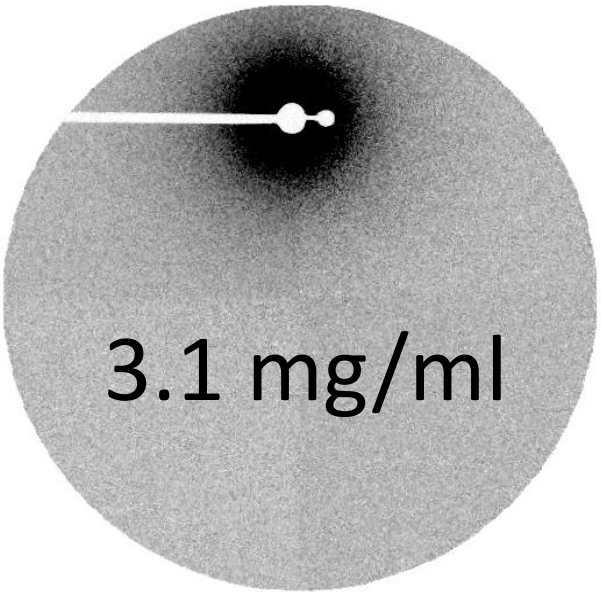
Up to 12 exposures, 1.5s each.

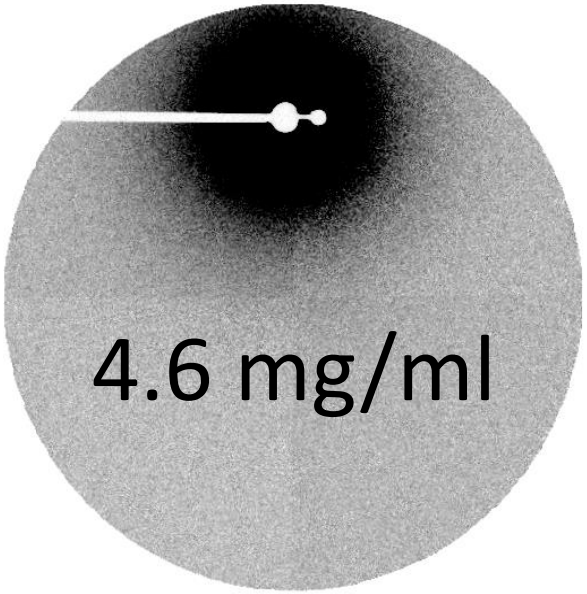
Load next concentrations and repeat.

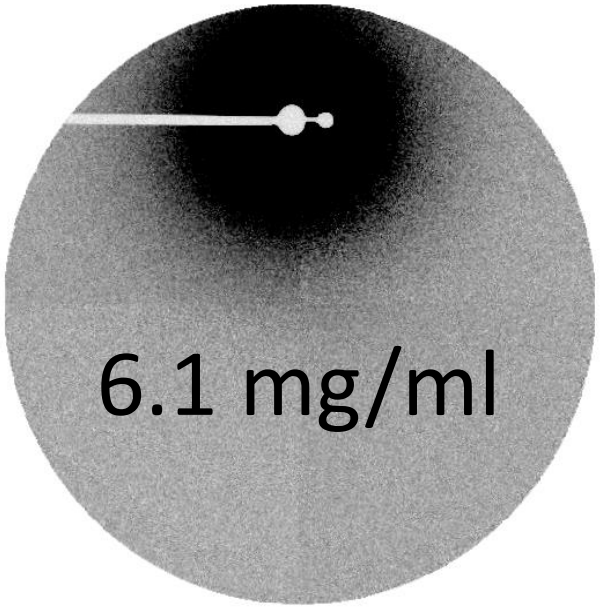
Repeat the buffer.

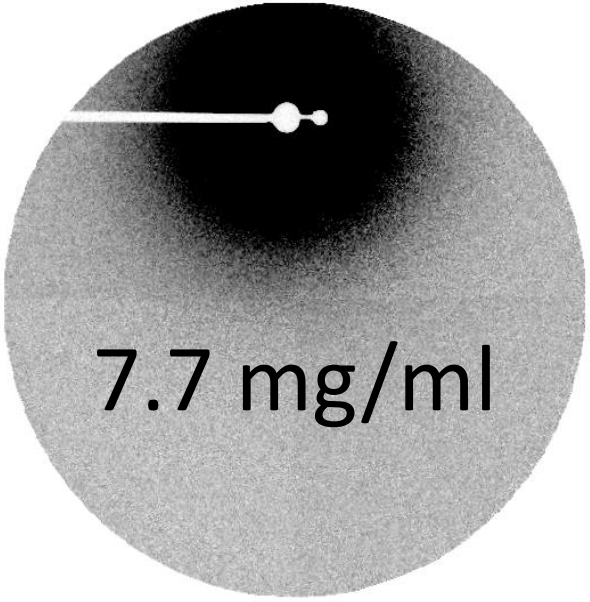
Clean the capillary with bleach followed by water.

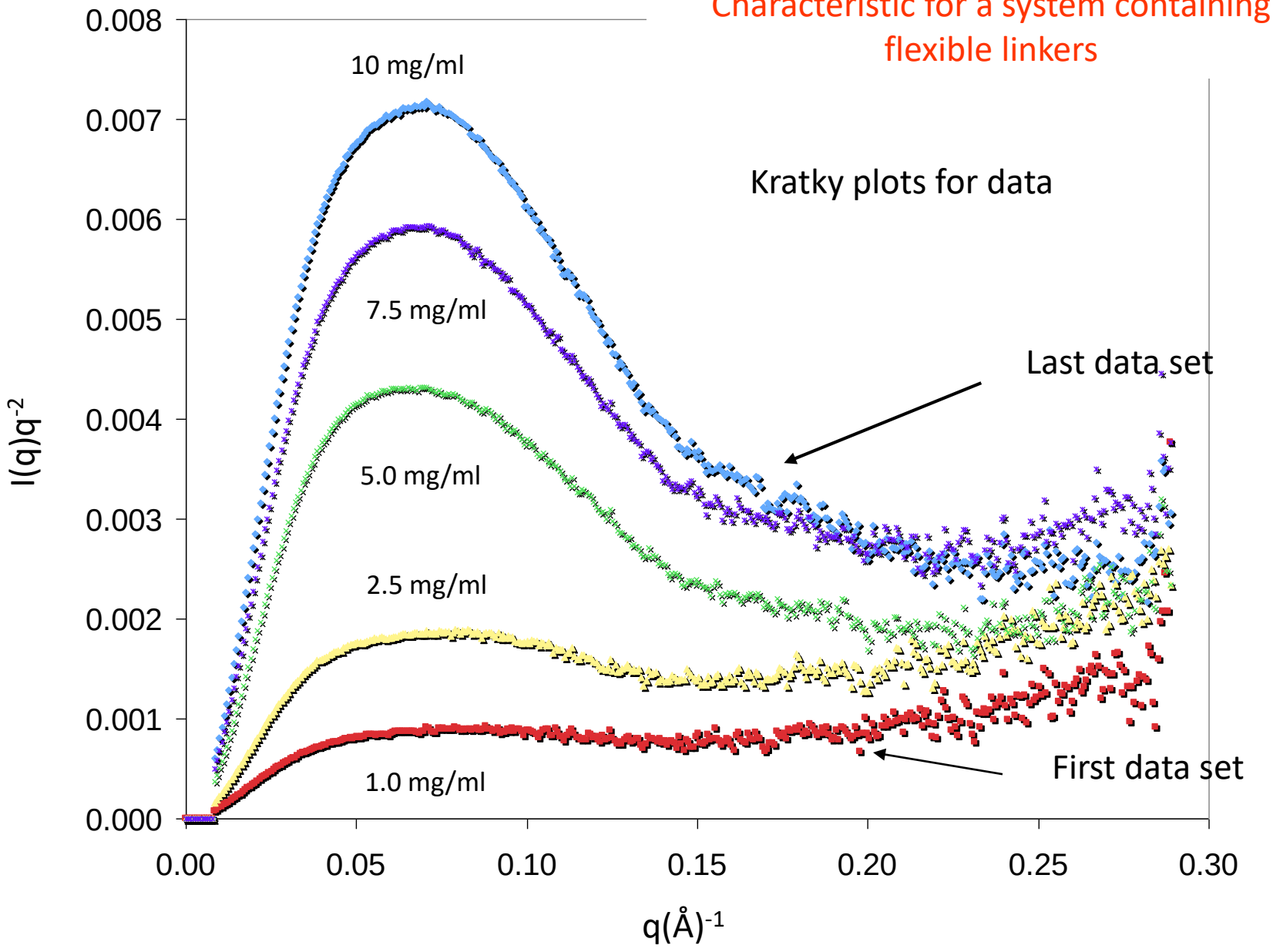

1.5 mg/ml

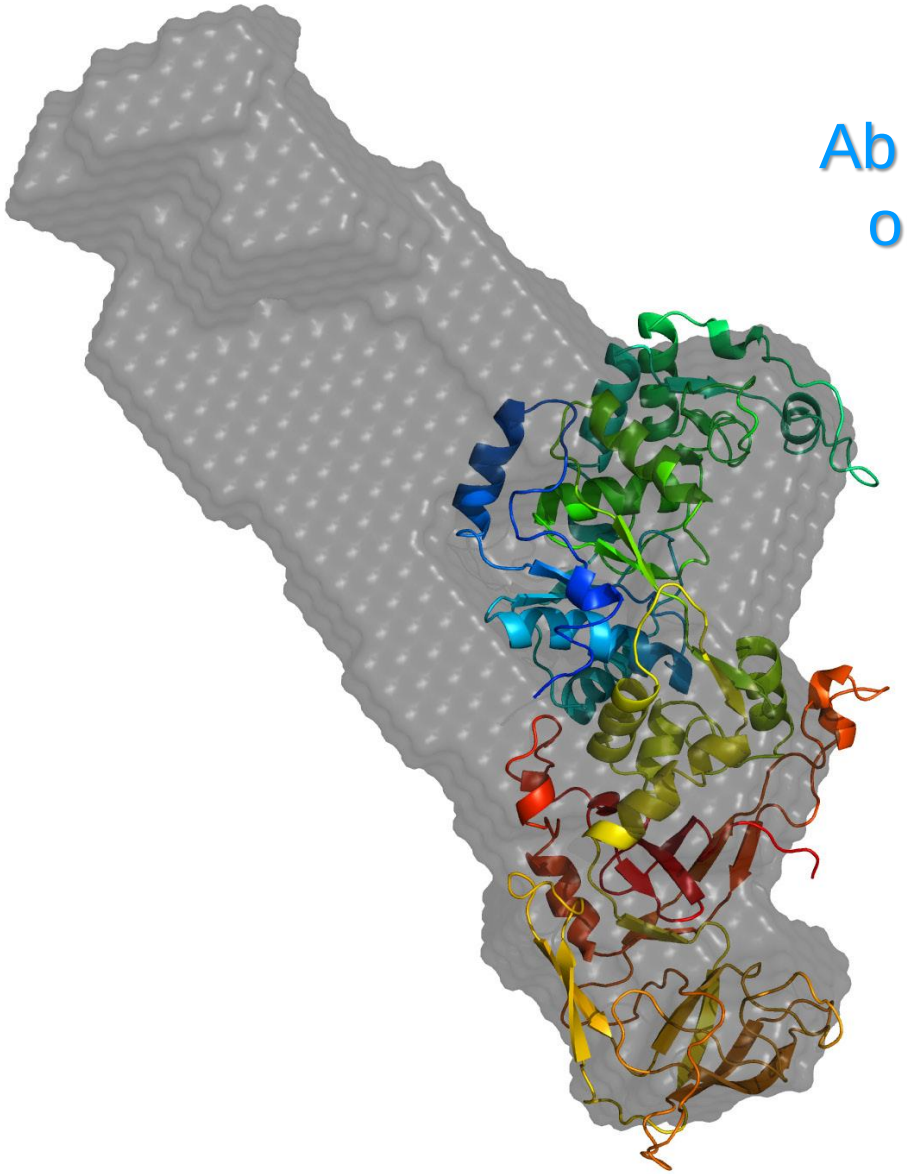

3.1 mg/ml


4.6 mg/ml

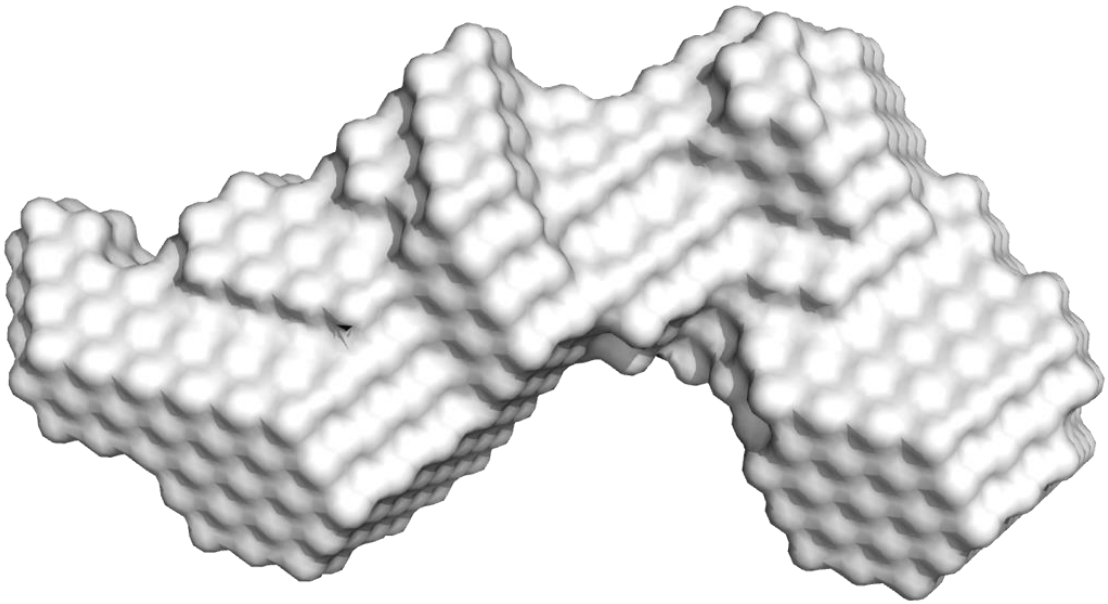

6.1 mg/ml

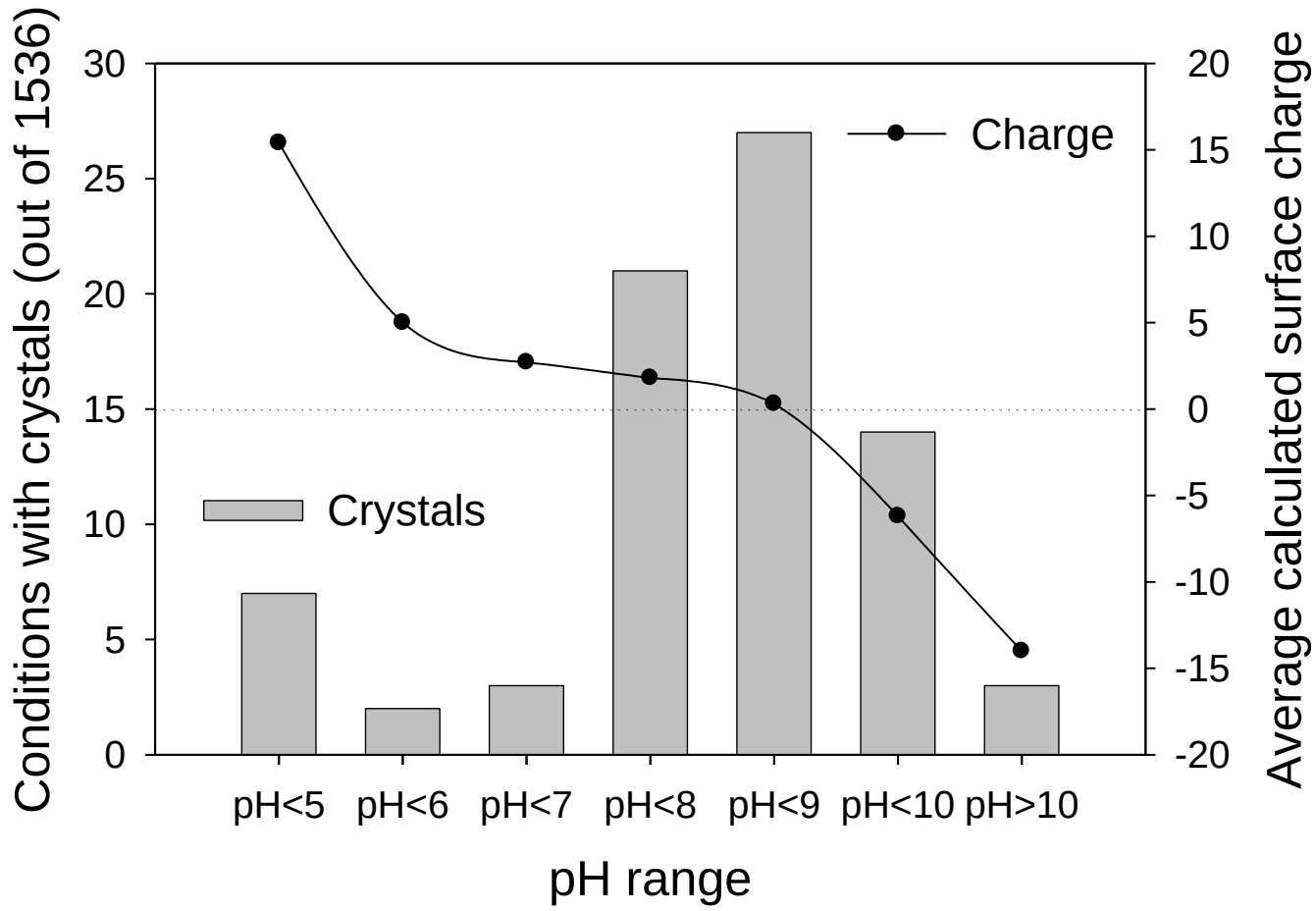

7.7 mg/ml

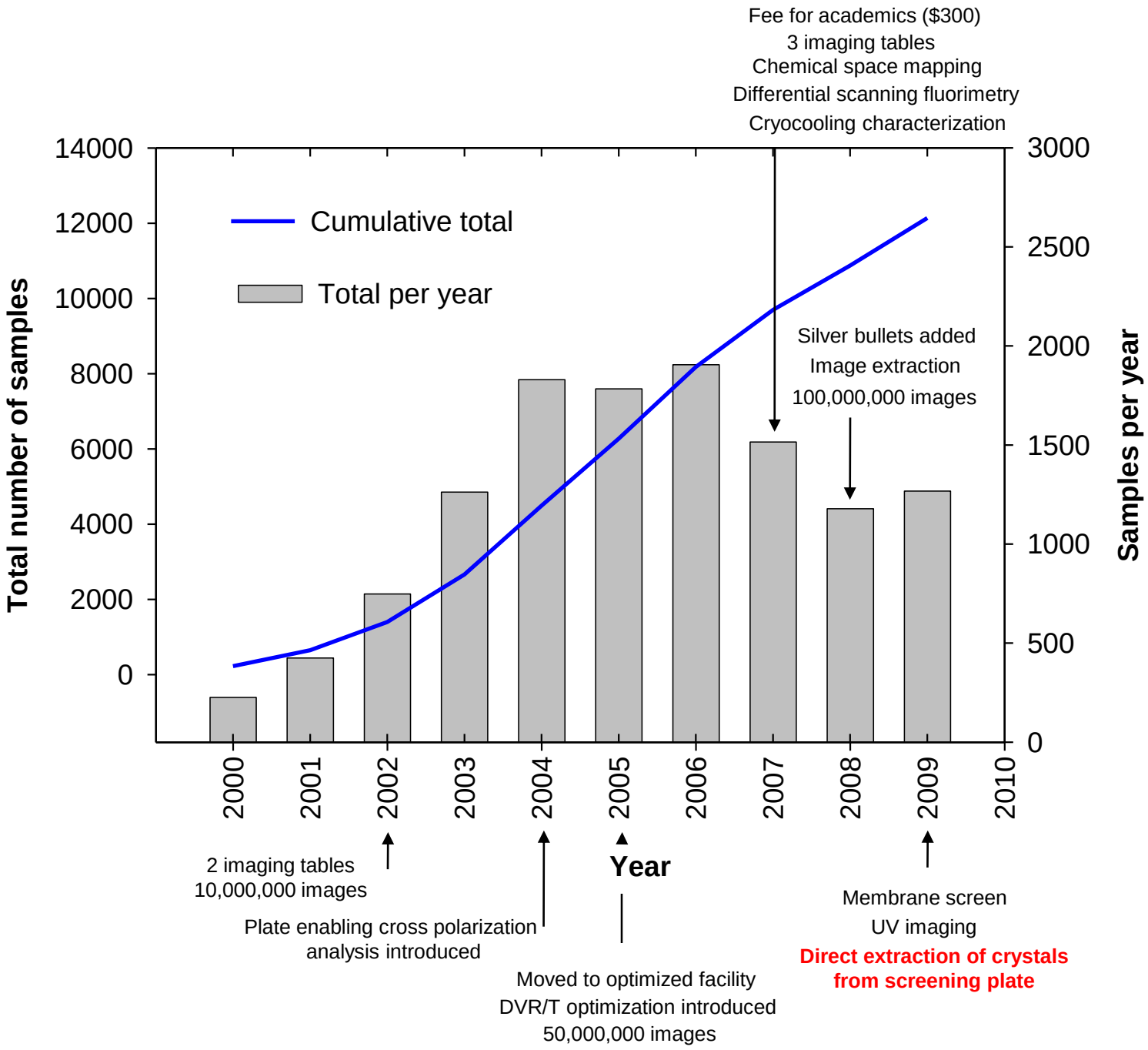

Characteristic for a system containing flexible linkers
Kratky plots for data
10 mg/ml
7.5 mg/ml
5.0 mg/ml
2.5 mg/ml
1.0 mg/ml
Last data set
First data set
I(q)q-2
q(Å)-1
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.00
0.05
0.10
0.15
0.20
0.25
0.30



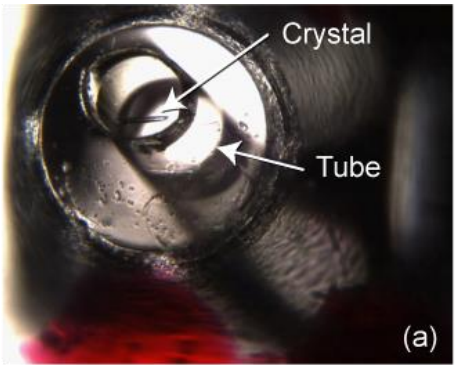
Ab initio structure overlaid on the crystallographic structure

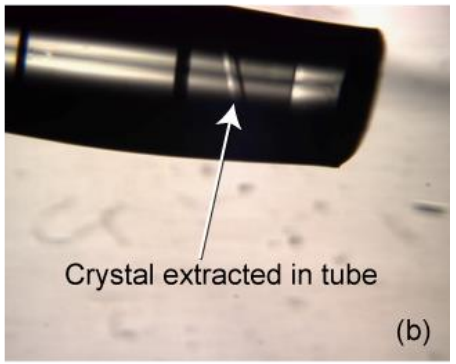


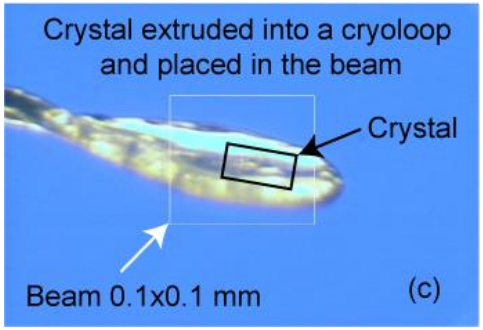

Conditions with crystals (out of 1536)
Average calculated surface charge
Charge
Crystals
30
25
20
15
10
5
0
20
15
10
5
0
-5
-10
-15
-20
pH<5
pH<6
pH<7
pH<8
pH<9
pH<10
pH>10
pH range

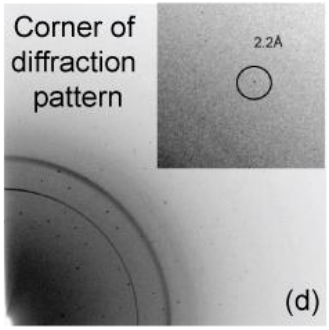

Fee for academics ($300)
3 imaging tables
Chemical space mapping
Differential scanning fluorimetry
Cryocooling characterization
Cumulative total
Total per year
Silver bullets added
Image extraction
100,000,000 images
Total number of samples
Samples per year
14000
12000
10000
8000
6000
4000
2000
0
3000
2500
2000
1500
1000
500
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Year
2 imaging tables
10,000,000 images
Plate enabling cross polarization
analysis introduced
Moved to optimized facility
DVR/T optimization introduced
50,000,000 images
Membrane screen
UV imaging
Direct extraction of crystals
from screening plate

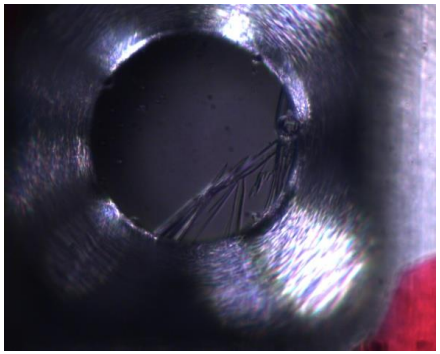
Does it diffract? Screening before the synchrotron

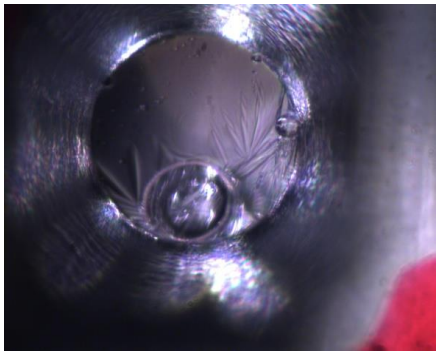


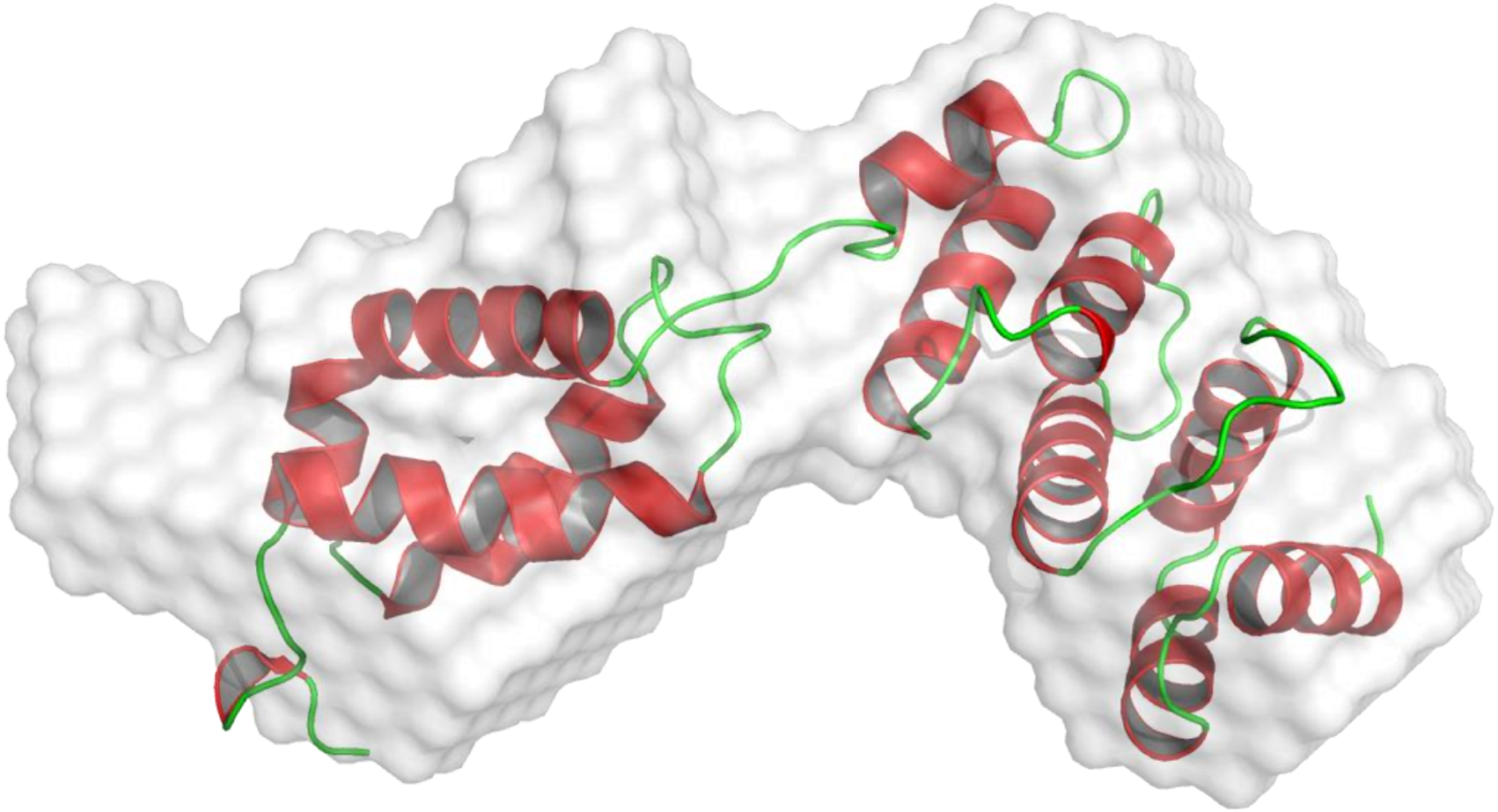




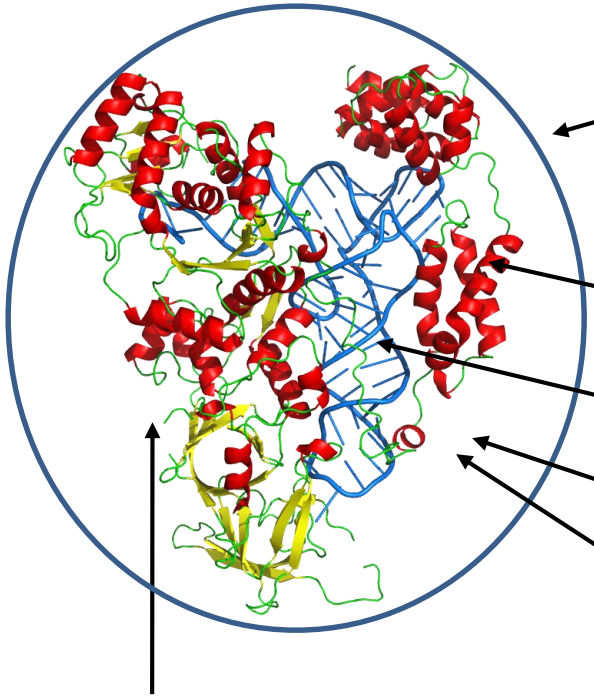








Case study applied to complexes



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

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

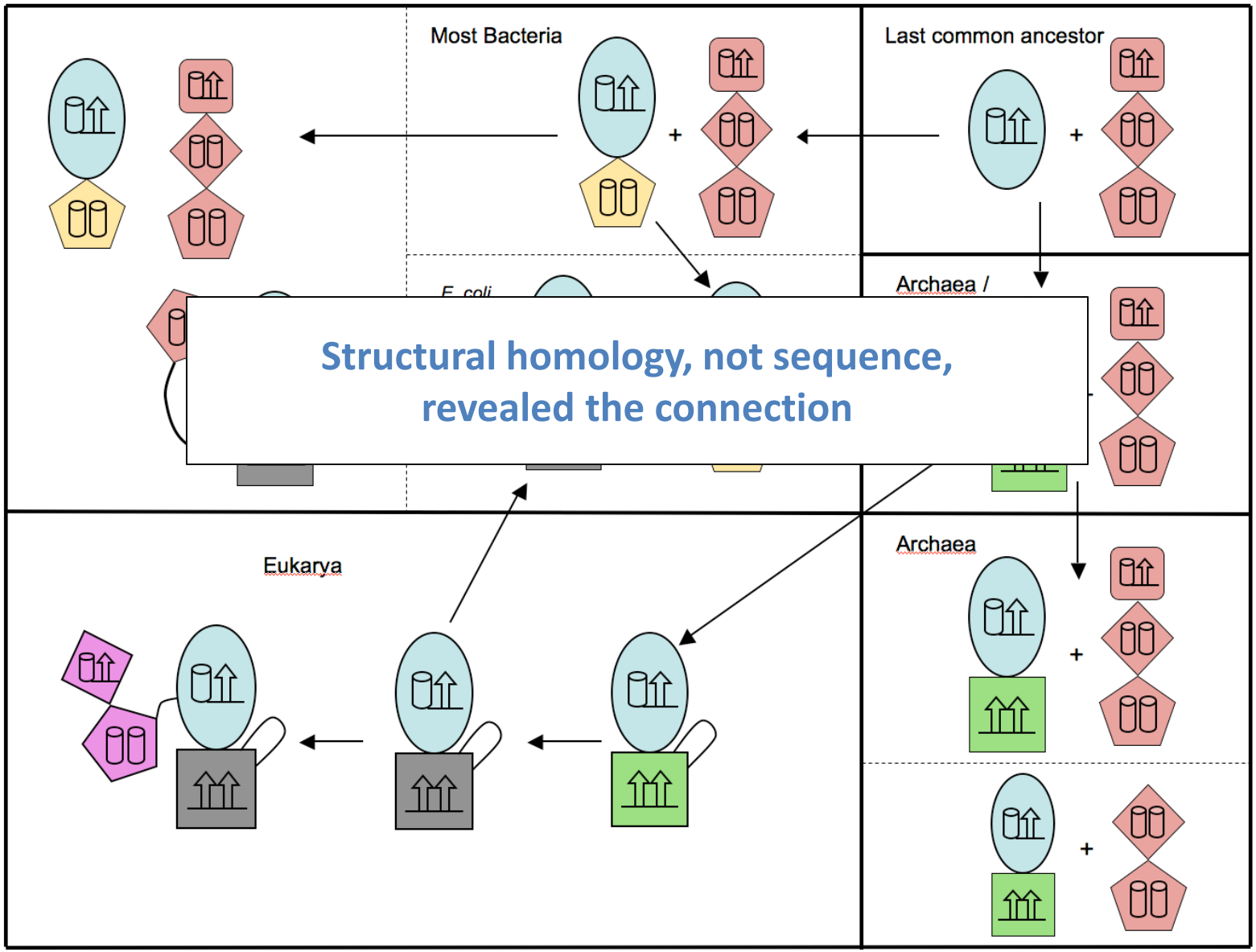
SAXS aided by sequence analysis identified a flexible region

Homology modeling gave the flexible region

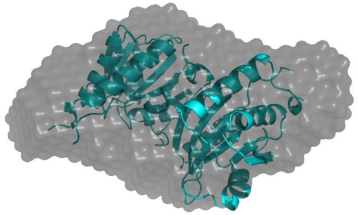
Mutation studies are underway to confirm the structure

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

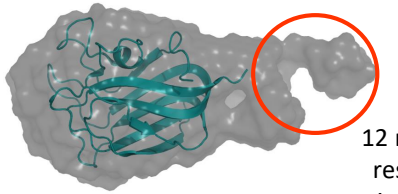
Why is this important?



SAXS Envelopes



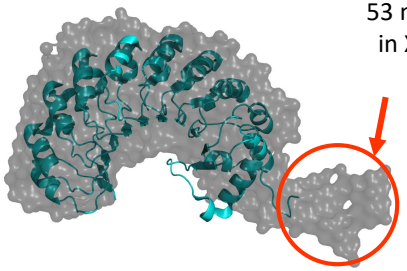
alr0221 protein from Nostoc (18.6 kDa)



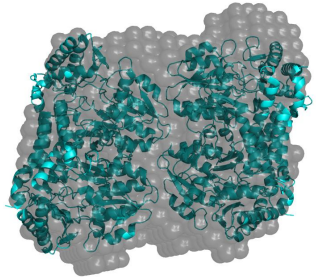
12 missing residues in X-ray structure

C-terminal domain of a chitobiase (17.9 kDa)

53 missing residues in X-ray structure



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



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

SAXS allows us to see ‘disorder’ and characterize our samples

SAXS"ess" with complexes

- Complex problems have real and imaginary components.
- The case study here is not a true complex example but demonstrates our current approach in the most complete manner.
- We remove the problem of crystallizing the complex by focusing on the individual components and piecing the complex structure together by other means.
- We attempt to crystallize the components and perform SAXS simultaneously to:
  1. Characterize the sample in terms of globularity, flexible domains, native unfoldedness (i.e. is it worth proceeding).
  2. Determine the influence of various factors on crystallization conditions.
  3. Produce a molecular envelope to predict regions suitable for truncation.
- We go back to crystallization on truncated components and perform SAXS simultaneously on the whole complex.
- We fit the high-resolution crystalllographic data into the low resolution envelope of the structure.
- For uncharacterized regions we use homology and molecular dynamics approaches.
- We test our model by mutation and biochemical analysis.

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 beamline staff

Support and Funding

NIH and DoD support to Edward Snell, NIH PSI support to George DeTitta

Questions?

esnell@hwi.buffalo.edu

Post-doc position available