# Developing tools to transition high-throughput crystallization to high-output crystallography

A case study with eukaryotic Glutaminyl-tRNA synthetase

Edward H. Snell and many others

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:

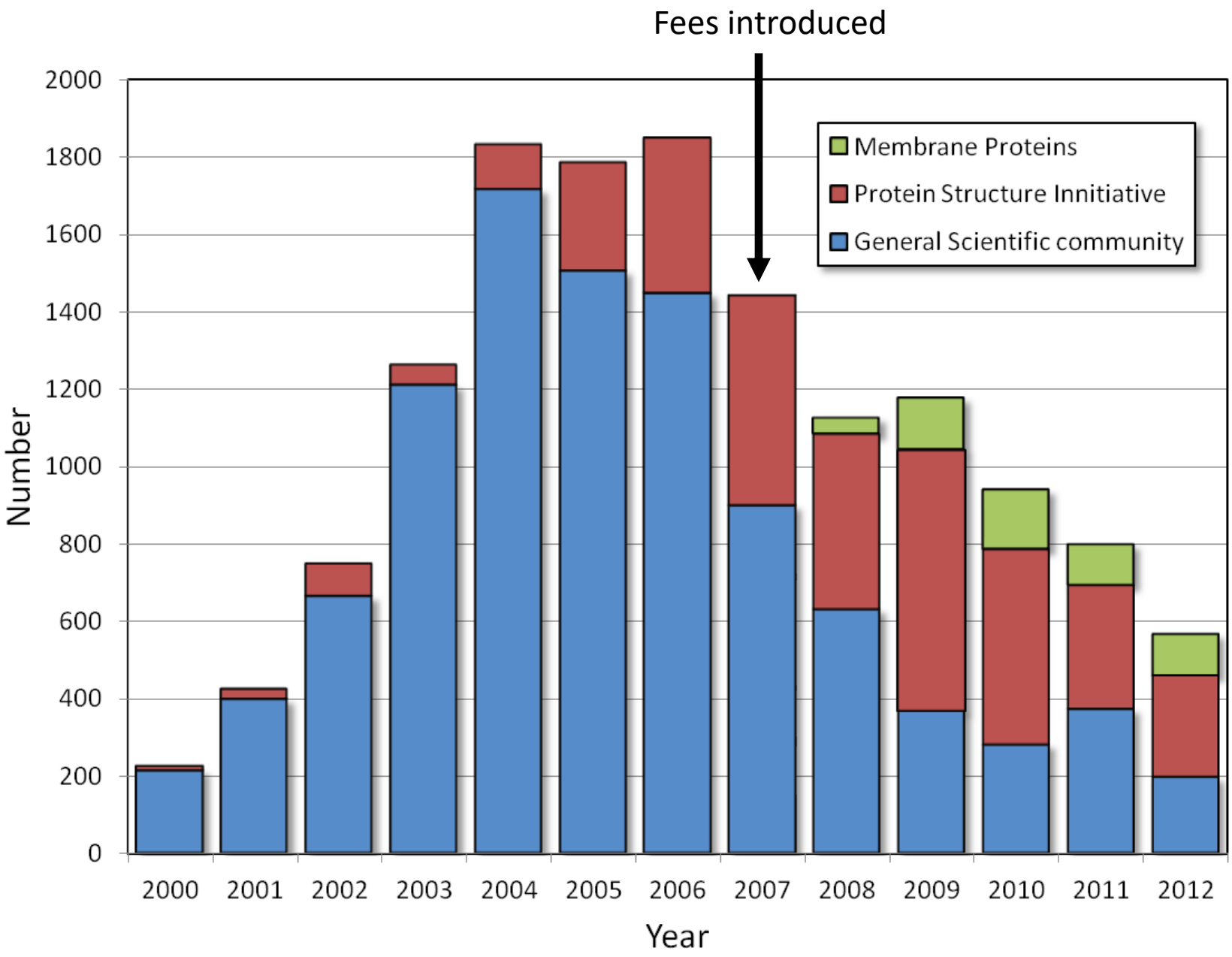
1. buffered (4<pH< 10), highly concentrated salts (35 salts total, sampling 18 different cations and 20 anions) – 229 conditions.
2. PEG/salt/buffer solutions (eight buffers (4<pH< 10), six molecular weight PEGs at three concentrations, and 35 salts at fixed 200 mM concentration) – 721 conditions.

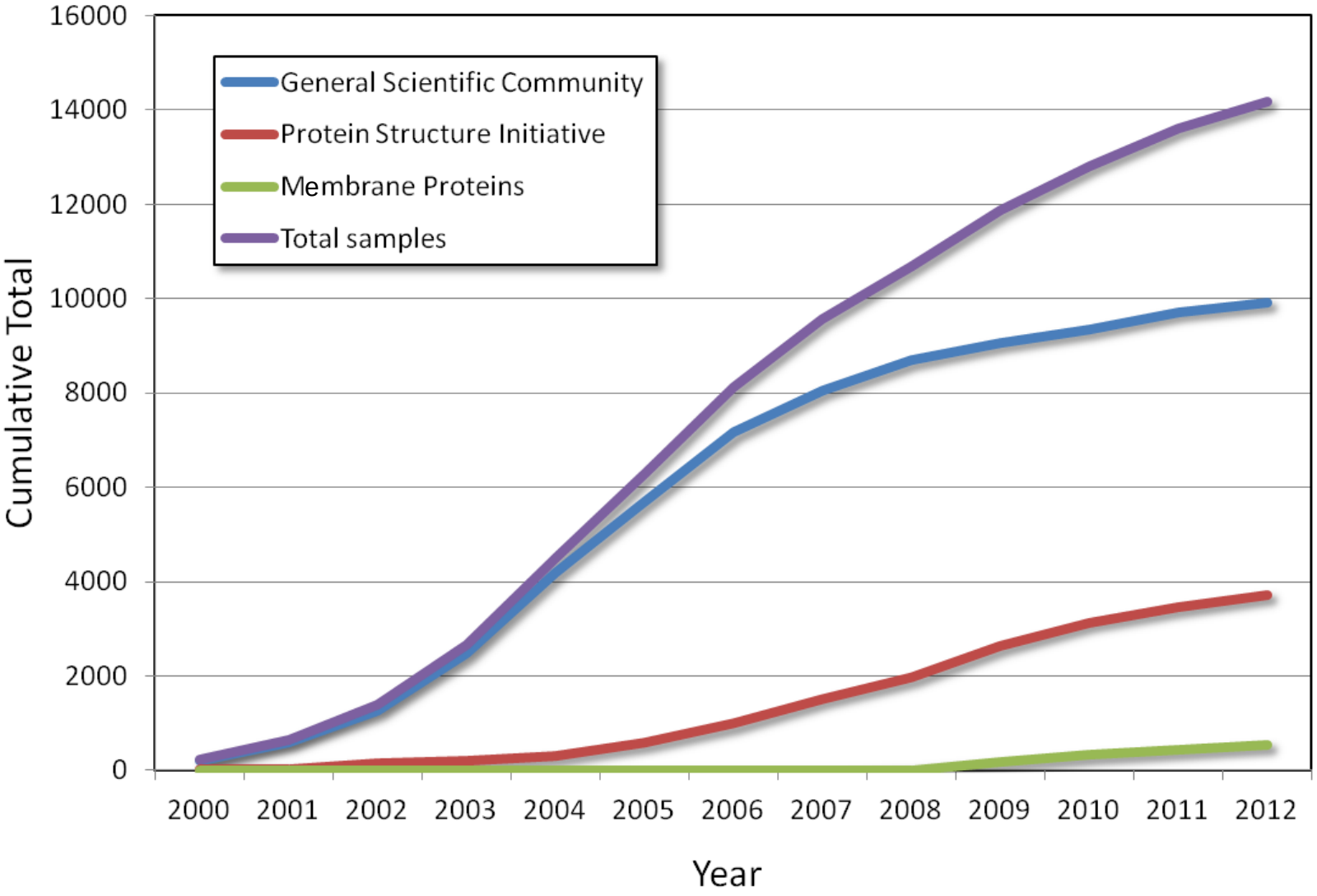
Added to this is a screen of some 586 conditions encompassing screens commercially available from Hampton Research.

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 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


16000
14000
12000
10000
8000
6000
4000
2000
0
Cumulative Total
General Scientific Community
Protein Structure Initiative
Membrane Proteins
Total samples
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Year

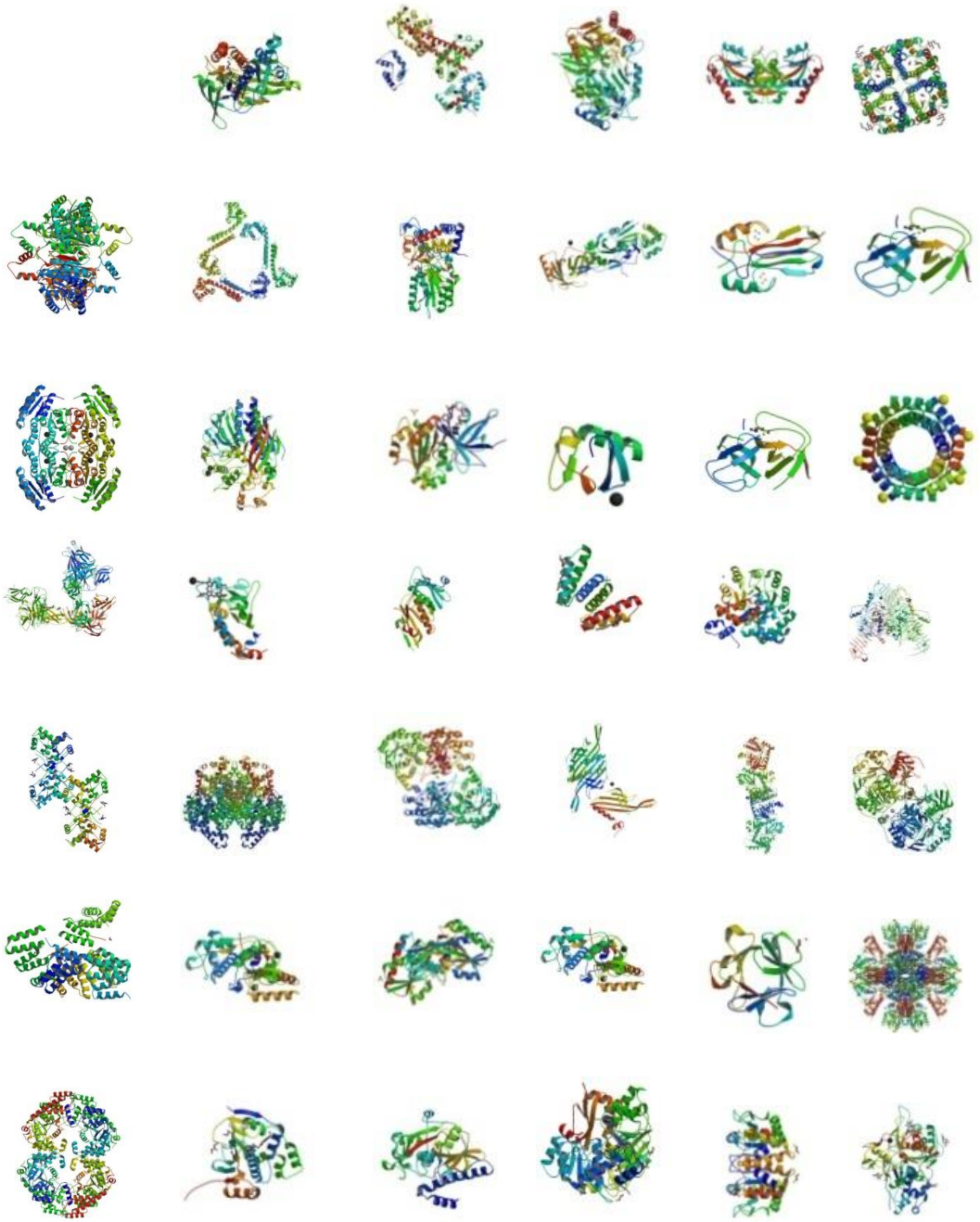
# Born in Buffalo

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.



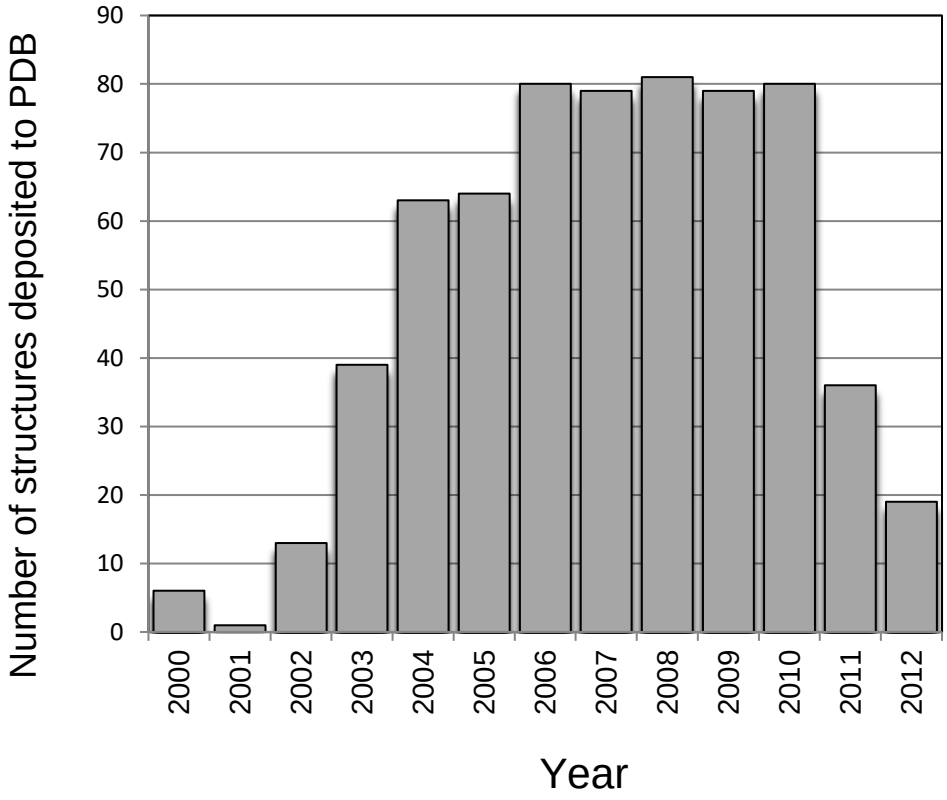
# Where success is tracked.

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

We are now working with more difficult proteins:

Complexes and disordered systems and membrane proteins

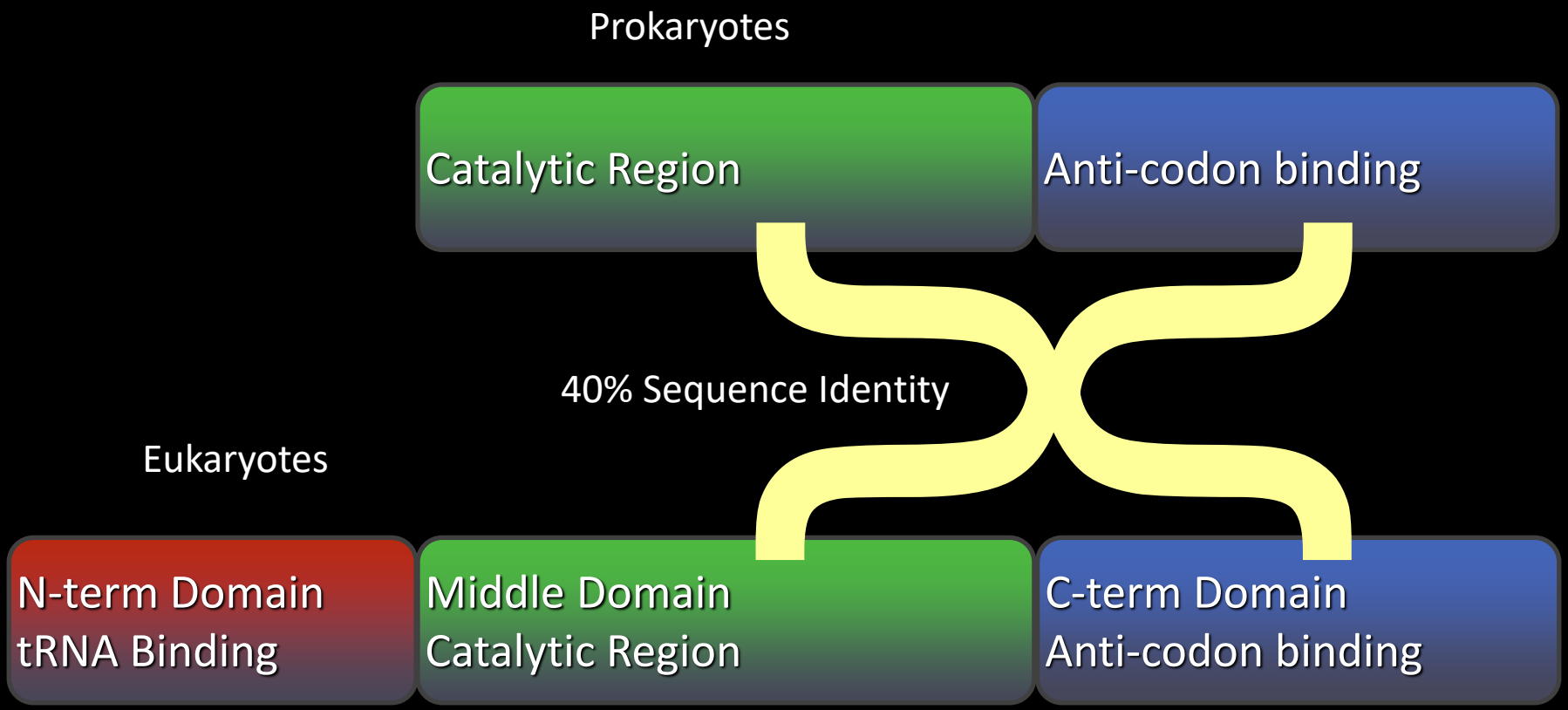
We are now working with more difficult proteins:

Complexes and disordered systems and membrane proteins

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

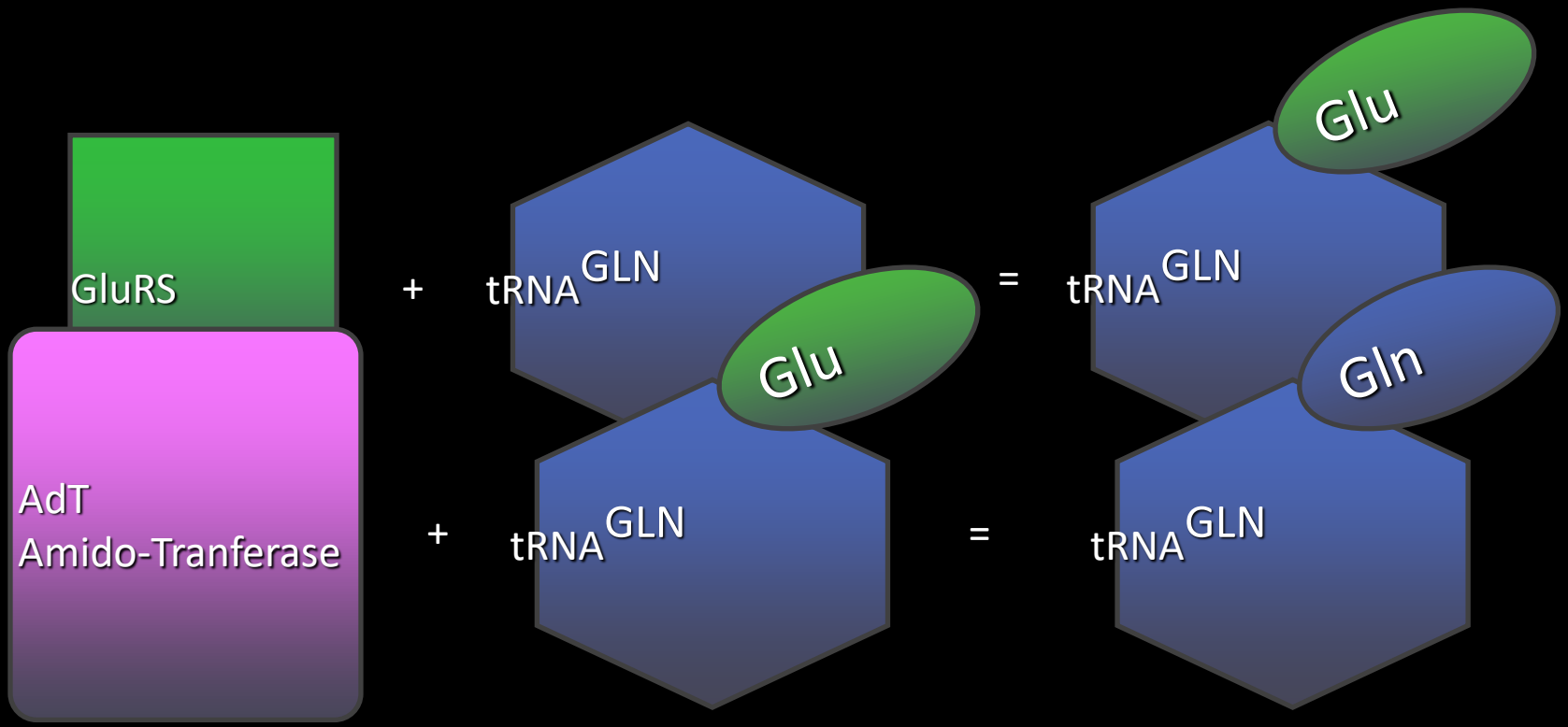
- Eukaryotic tRNA synthetases often carry appended domains not present in prokaryotic homologs
- These domains are known to bind RNA non-specifically
- Little is known about their function or structure

# Glutamine tRNA Synthetase



# Two routes of gln-tRNA$^{GLN}$ Formation

Indirect Route:  Archaea and Most Bacteria



# Two routes of gln-tRNA$^{GLN}$ Formation

Direct Route:  Eukaryotes and few bacteria

# 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

# Crystallization/Data collection

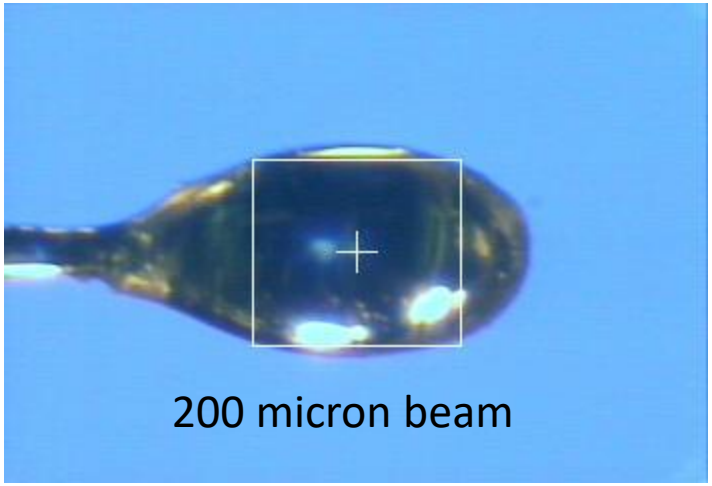
- Gln4 Screened against 1536 different biochemical conditions, ~1000 forming an incomplete factorial of chemical space and ~500 representing commercially available screens.
- Crystal leads seen, several were chosen based on ease of cryoprotection of the native hit.
- Crystals were optimized with a Drop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

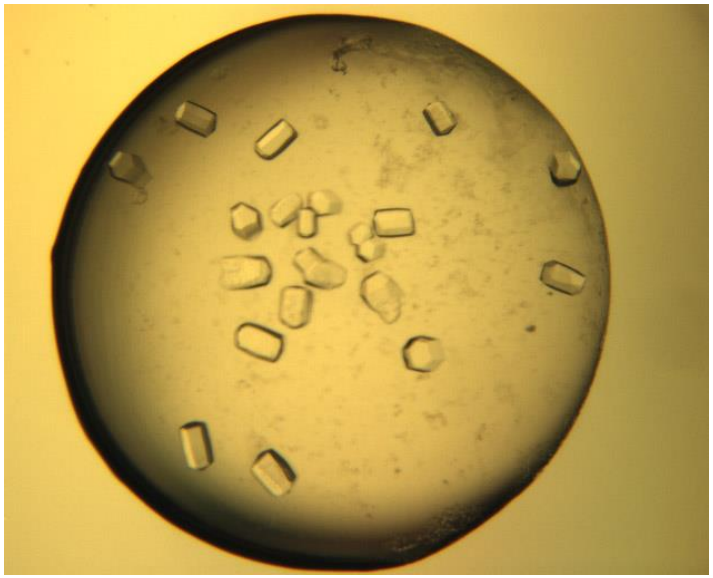
- Only 2 structures for related glutaminyl tRNA synthetases are available (~40% sequence homology), we had 228 extra residues (almost 40% more residues) therefore we expected problems in molecular replacement and didn't have a SeMet example.

- 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.

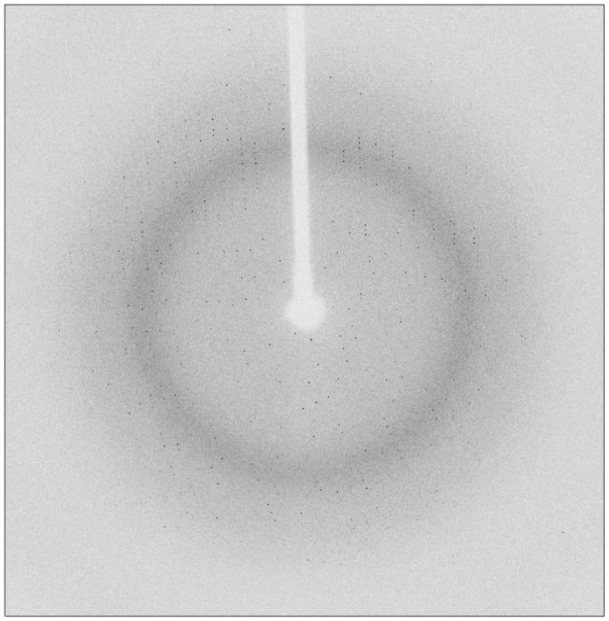
- We collected data remotely with an excitation scan to determine if Zinc was present.

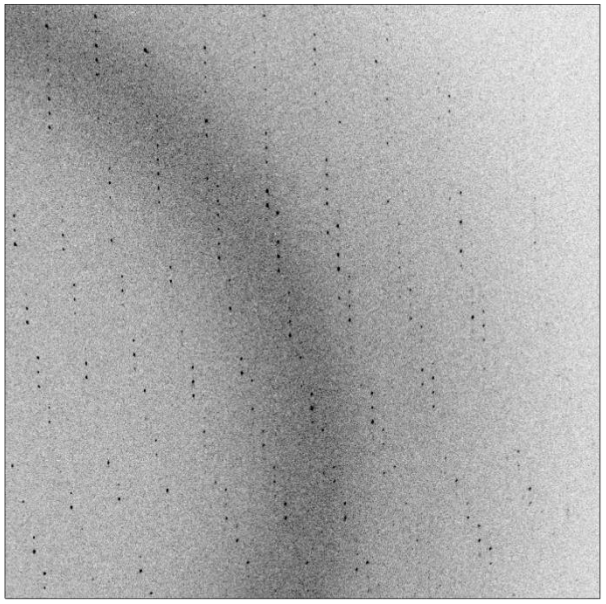
- It was!



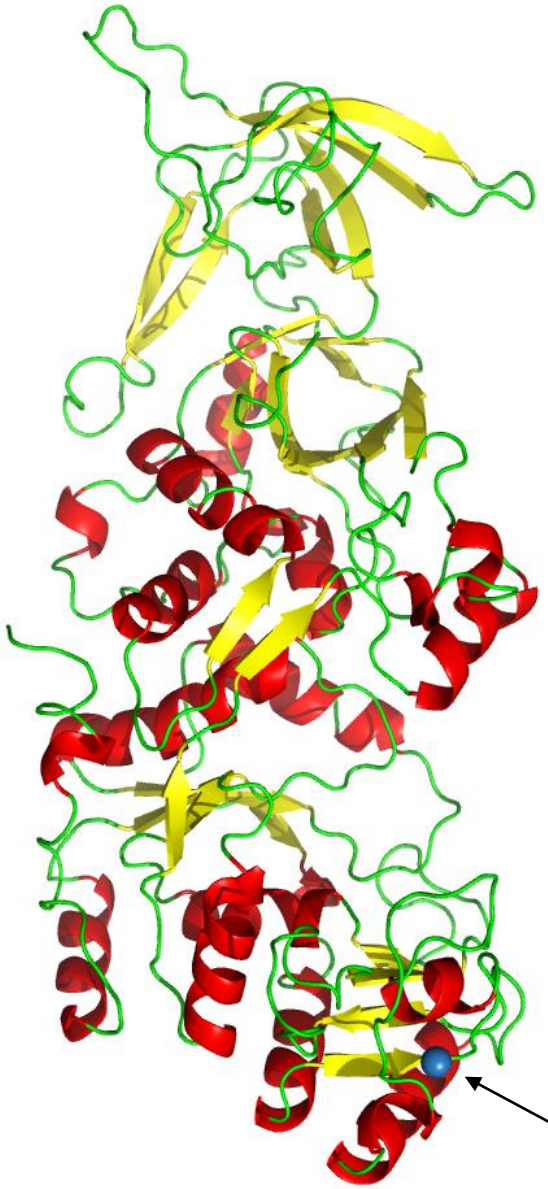



80% PEG 400 in the
crystallization cocktail

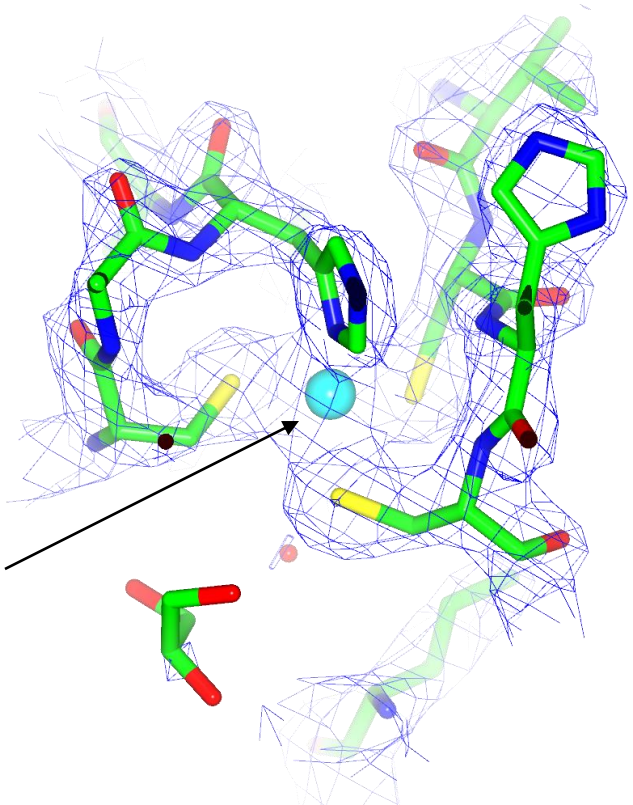


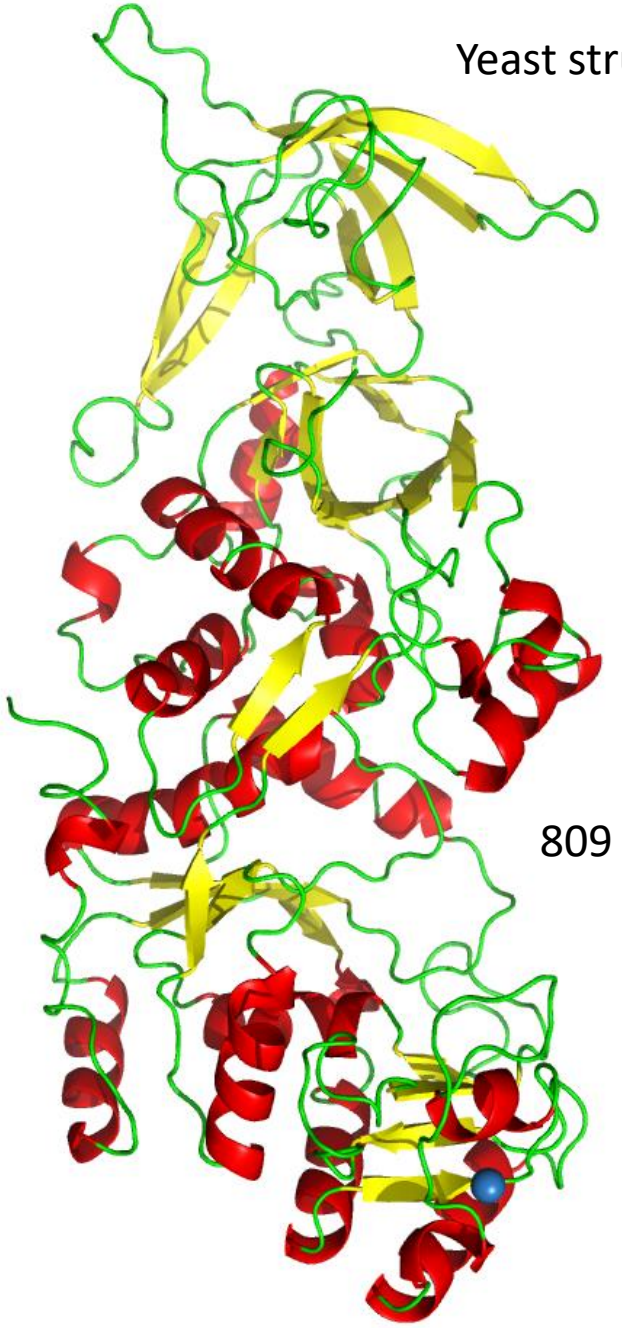


Structure solved (with help of the zinc signal) and refined with Phenix.



Zinc

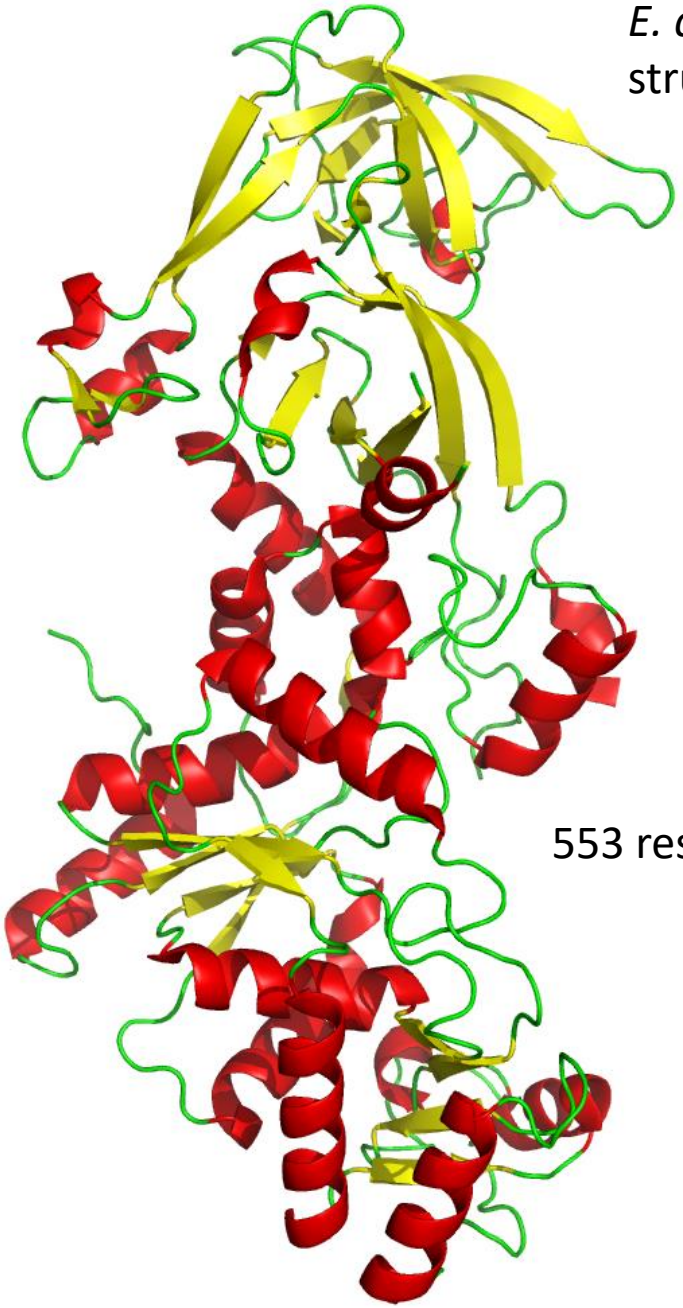




Yeast structure

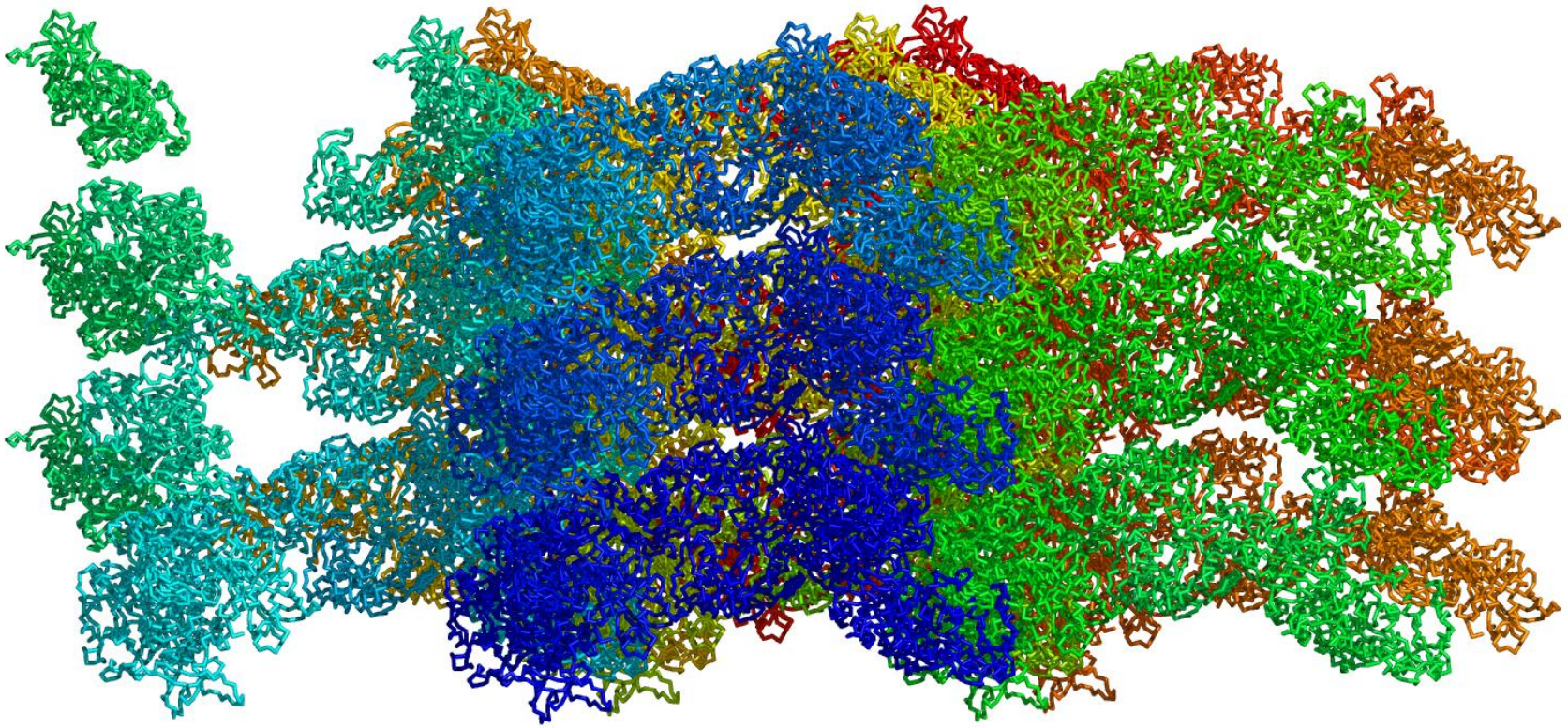
?

809 residues

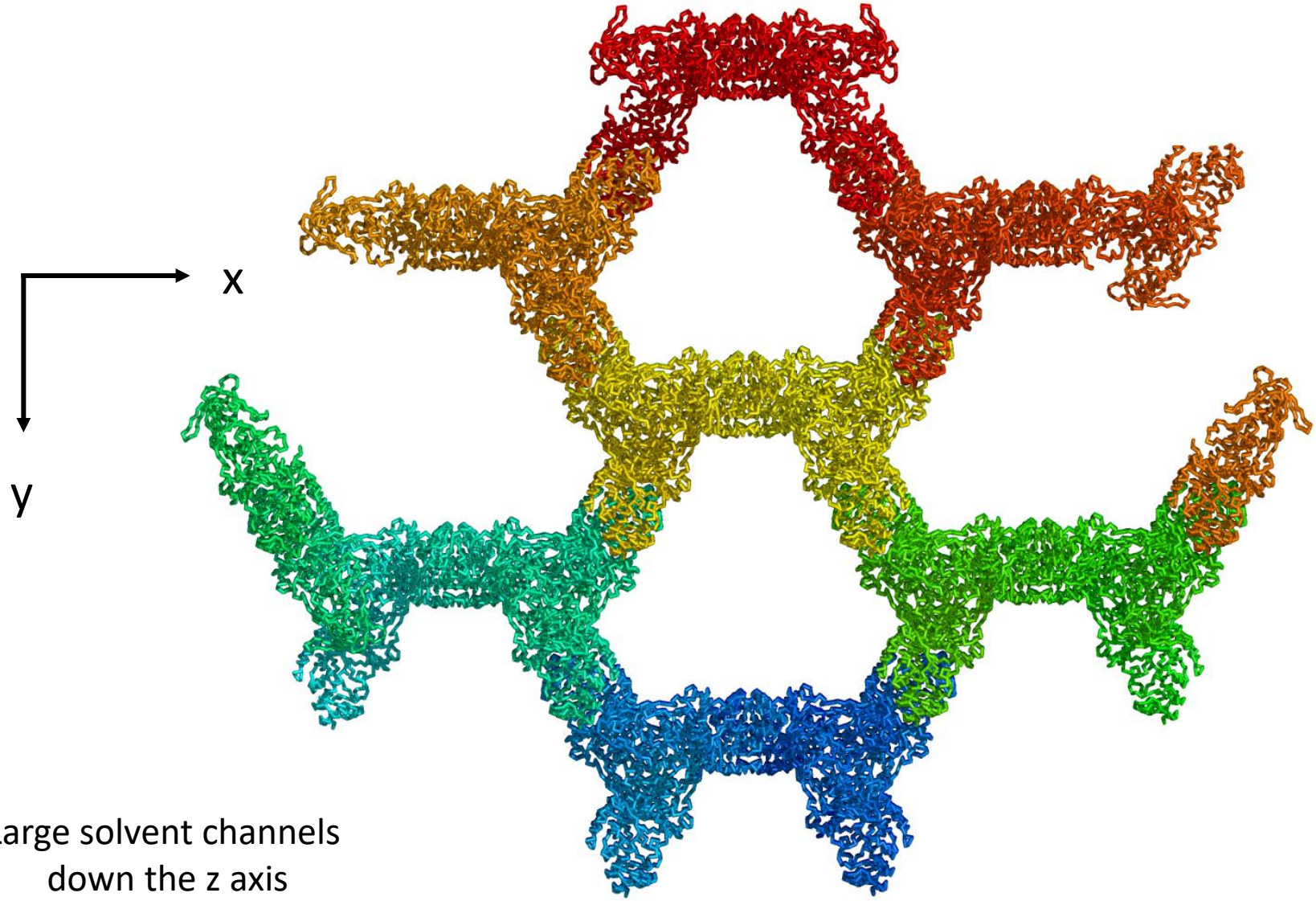


*E. coli.* structure

553 residues

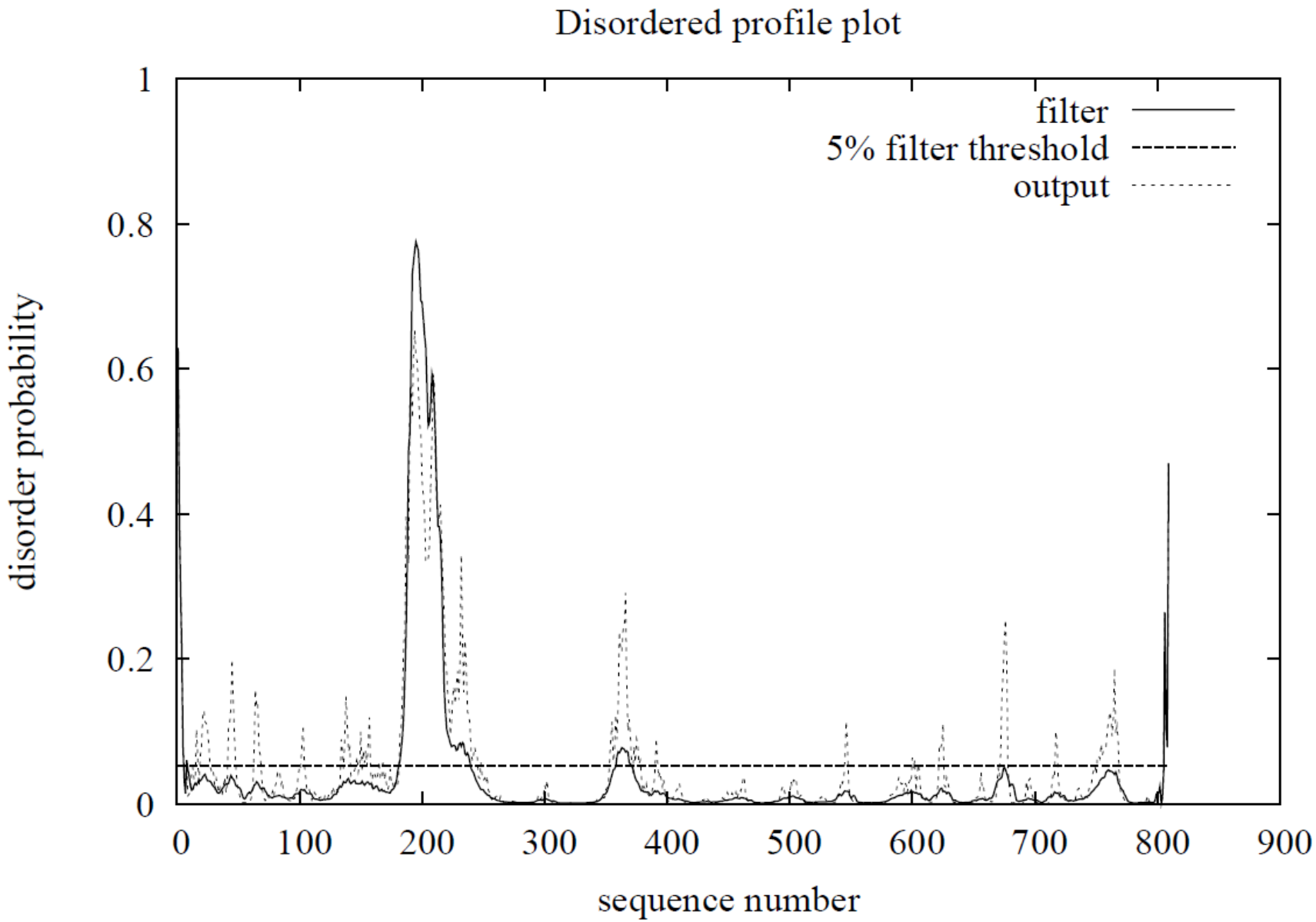



Tight packing in z and y



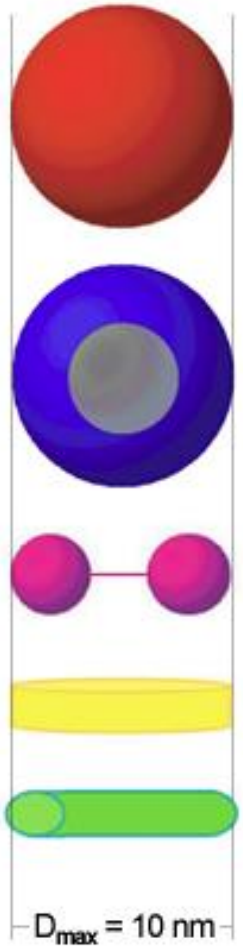

Large solvent channels
down the z axis

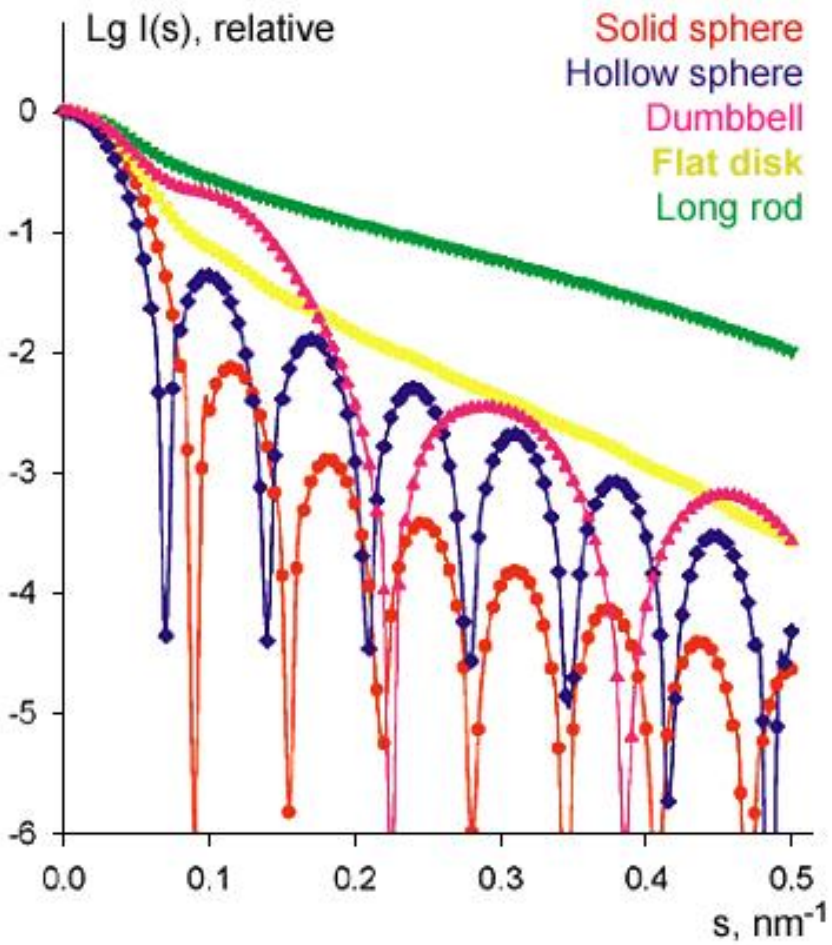
# Incorporate Bioinformatics


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

# Incorporate Other Techniques

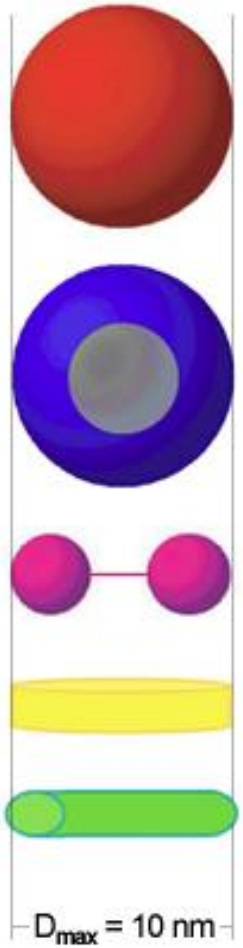
# Data

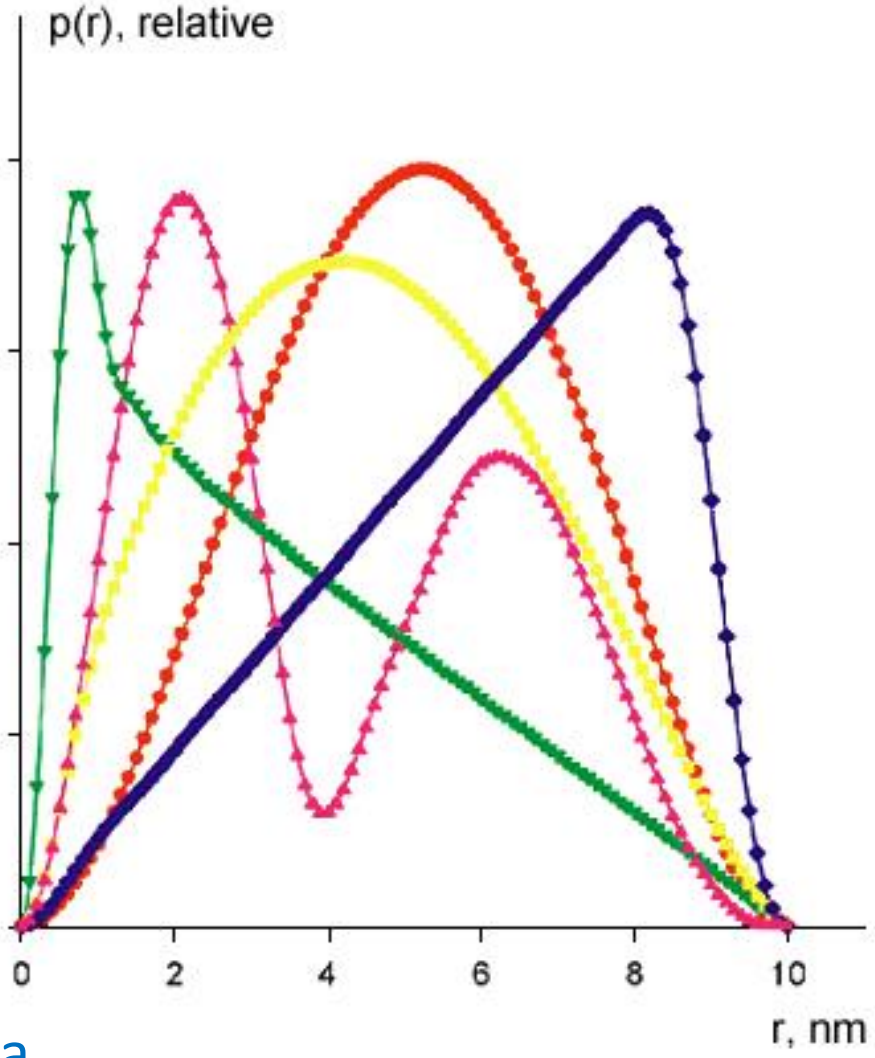



From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)

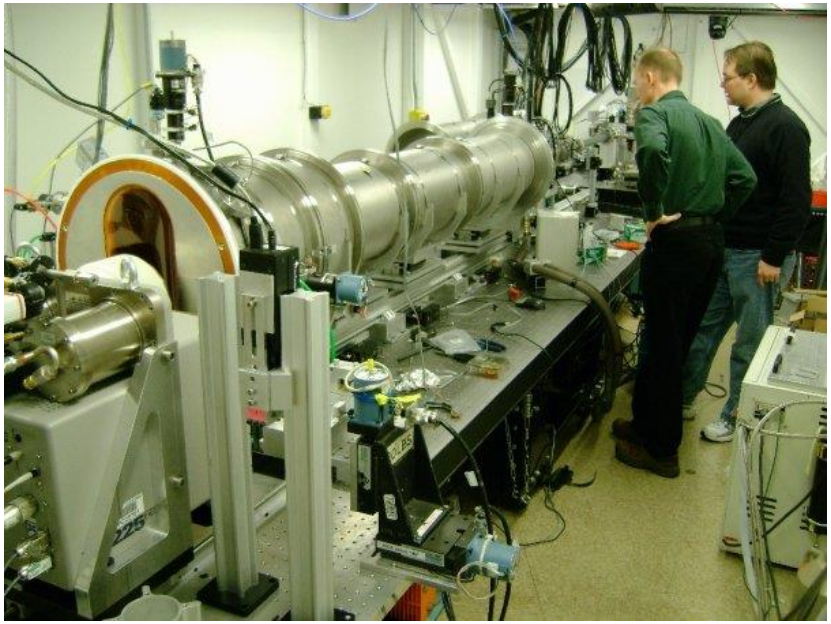
# Pair distribution function

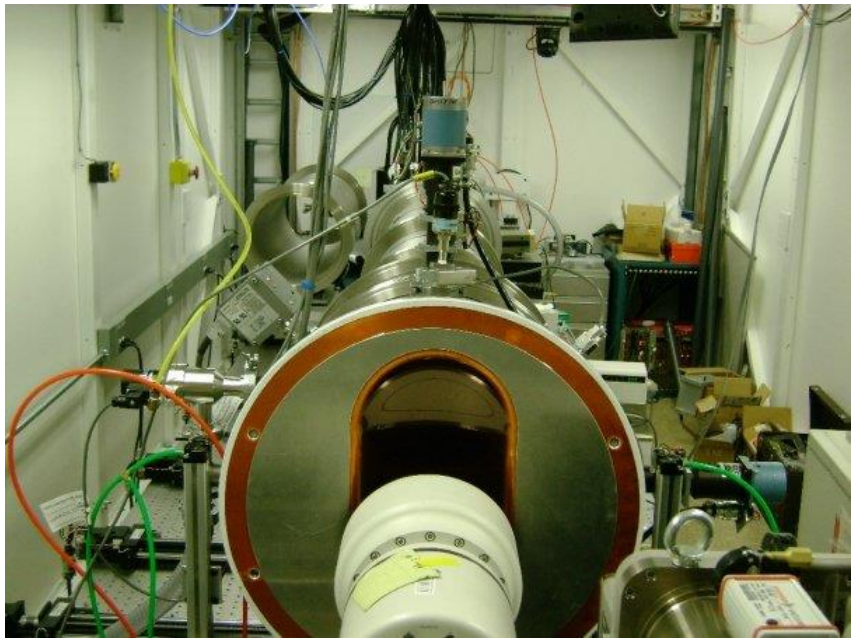


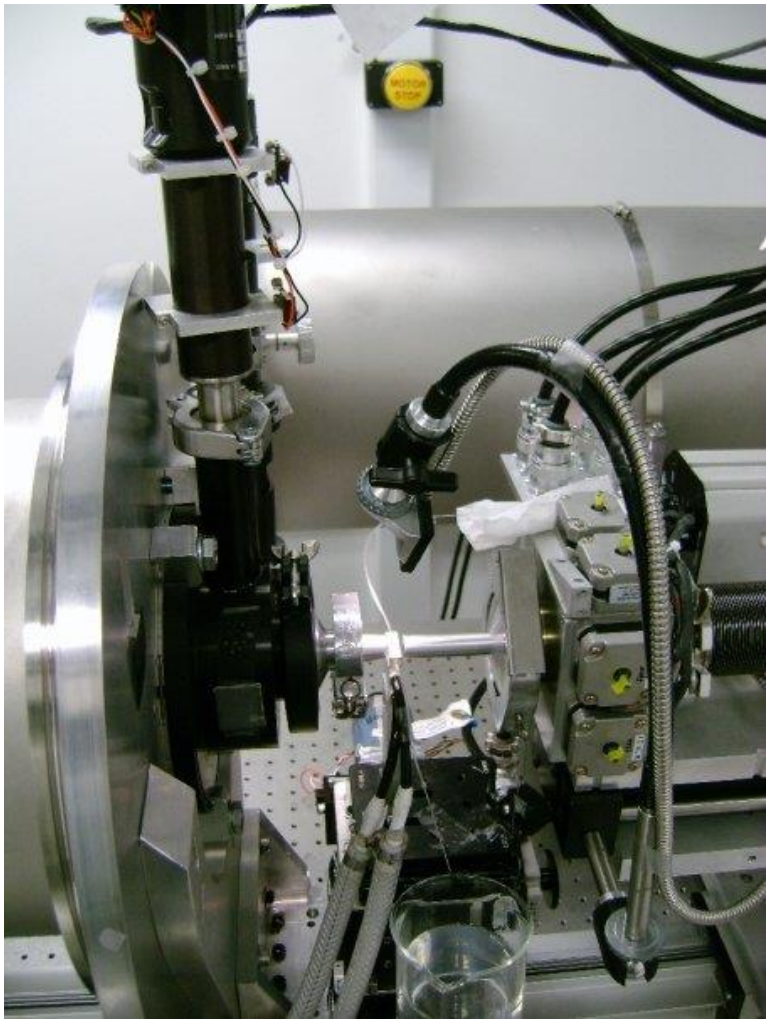




## Fourier transform of data.

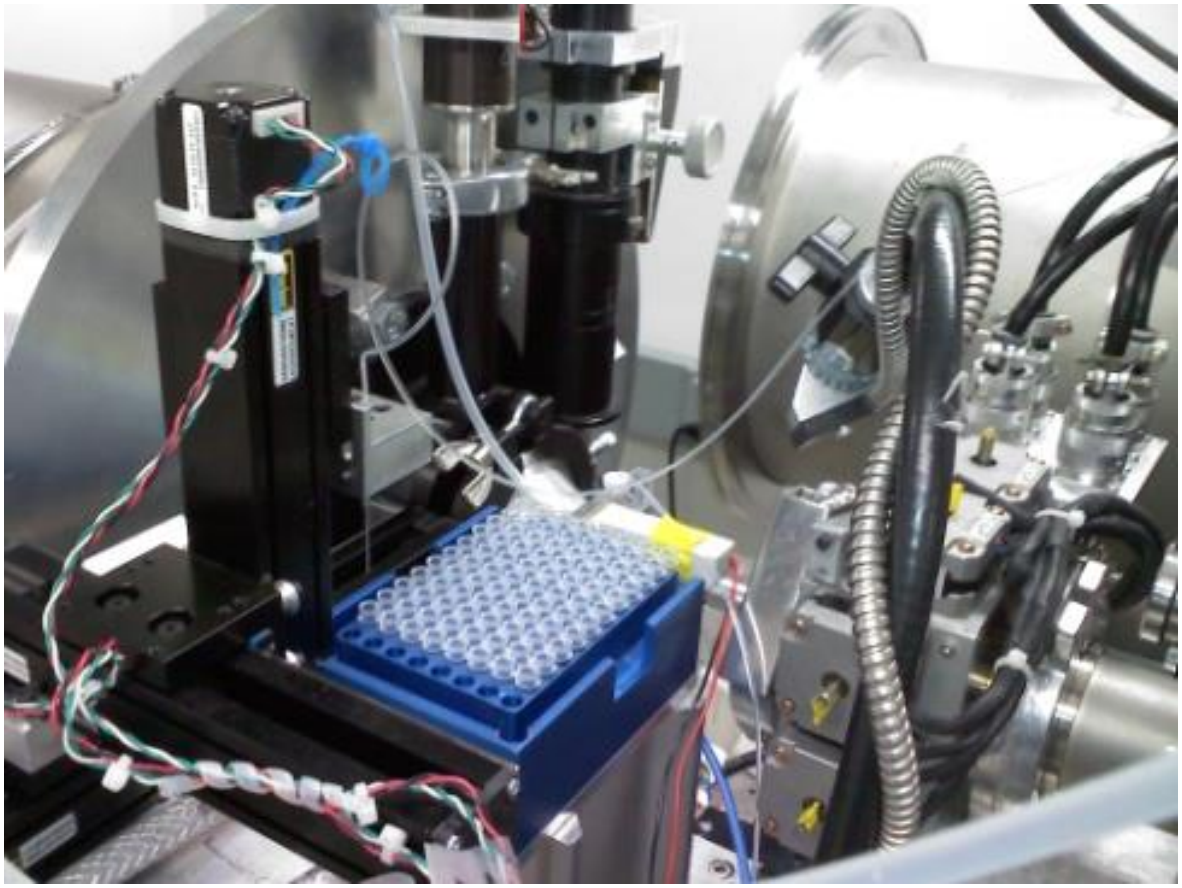
From: Small-angle scattering studies of biological macromolecules in solution, Svergun and Koch, Rep. Prog. Phys., 1735-1782 (2003)







**Beamline 4-2 SSRL**



# High throughput protocol

Up to 12 different PCR strips.

3-7 different concentrations per sample.

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

Start with buffer then lowest concentration first. End with buffer

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

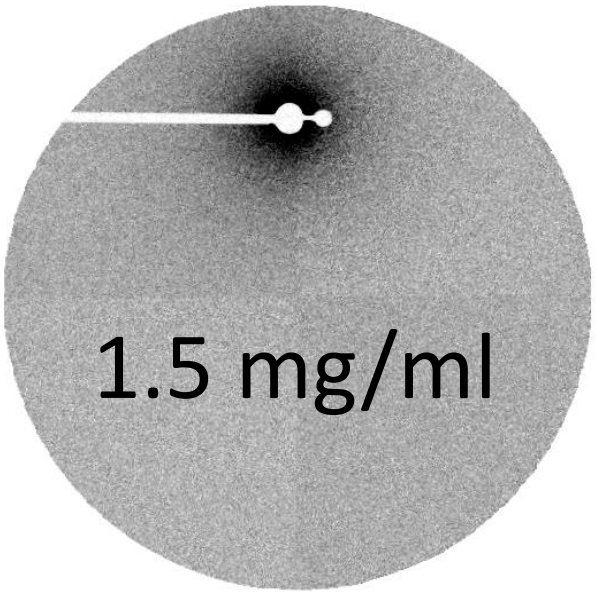
Oscillate sample to minimize radiation damage

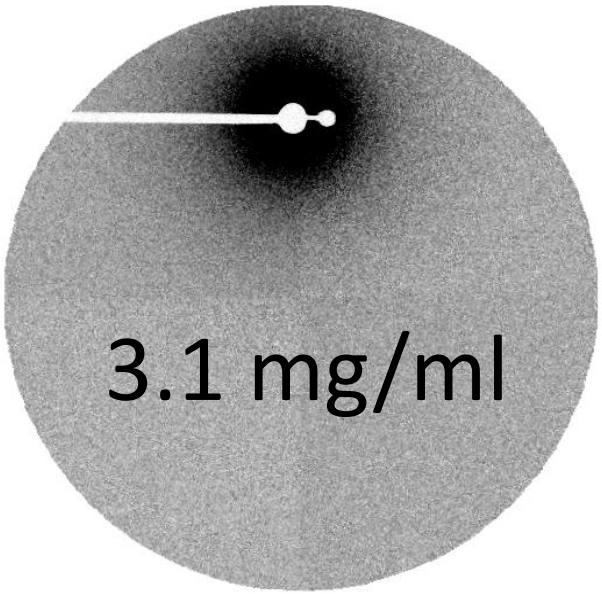
Repeat the buffer.

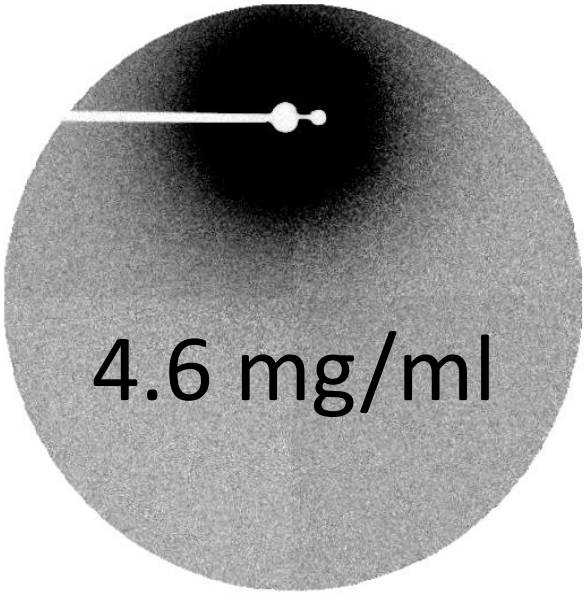
Load next sample

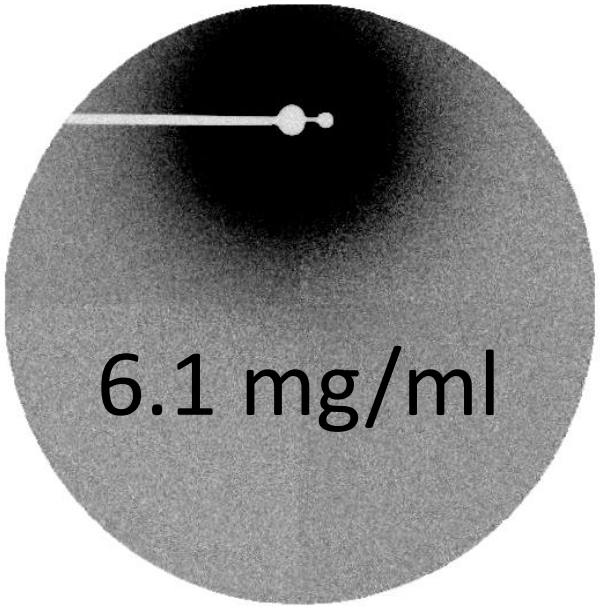
Time per concentration series – approximately 10 to 15 minutes. In high-throughput mode 24 samples in 3 to 4 hours.

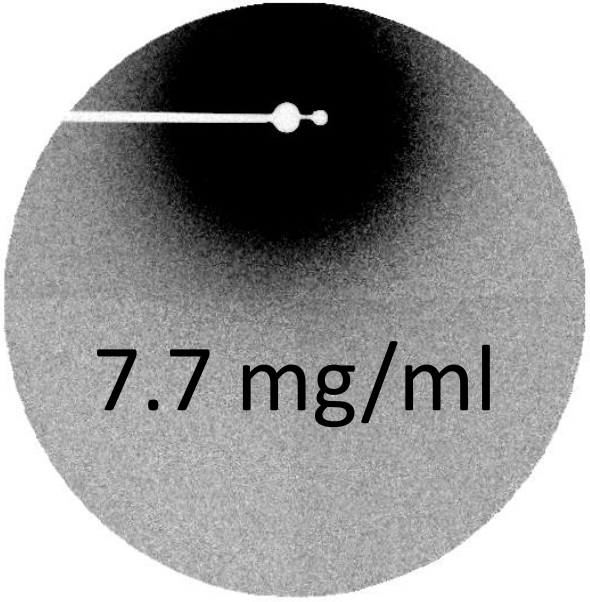
Enables two important things – eat and sleep!


1.5 mg/ml


3.1 mg/ml


4.6 mg/ml

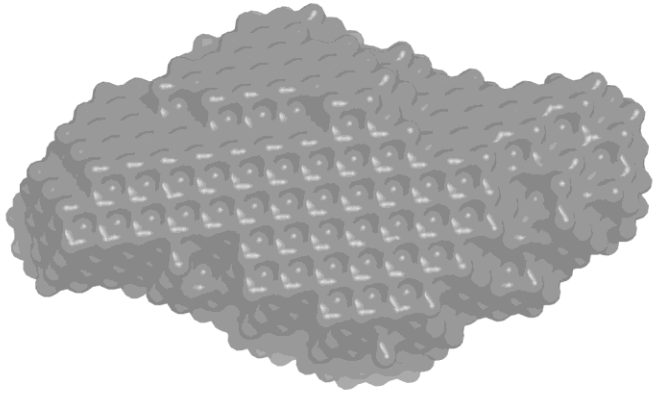

6.1 mg/ml


7.7 mg/ml

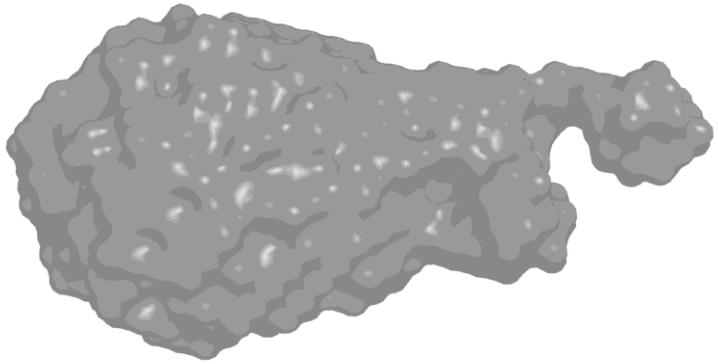
# Developments in the last decade that have revolutionized SAXS

- 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!

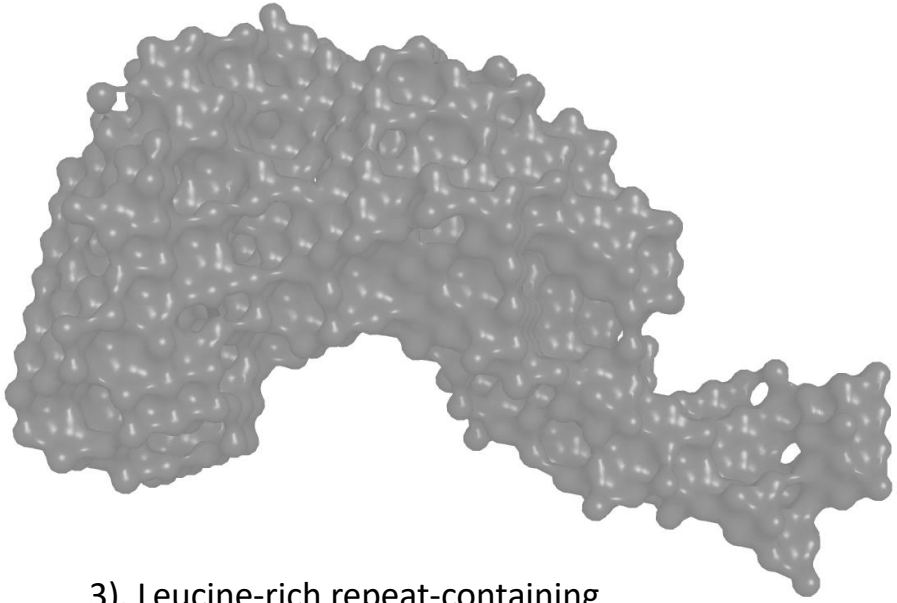
# *Ab intio* envelopes



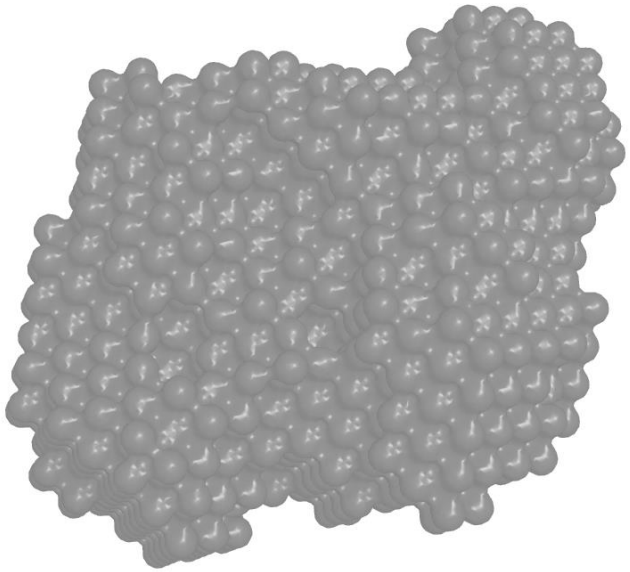
1). alr0221 protein from Nostoc (18.6 kDa)



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

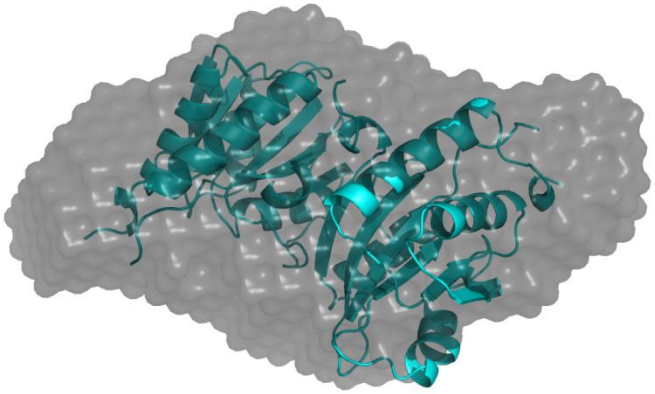


3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

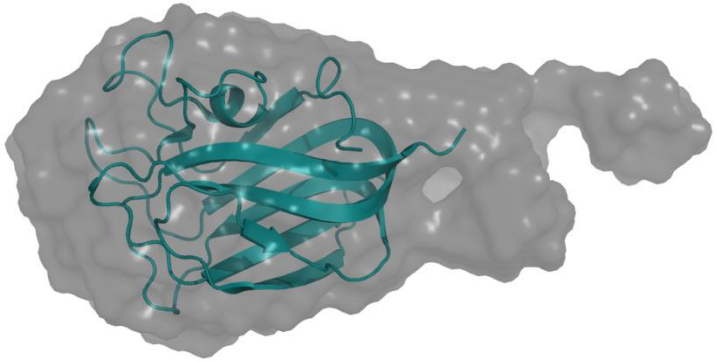


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

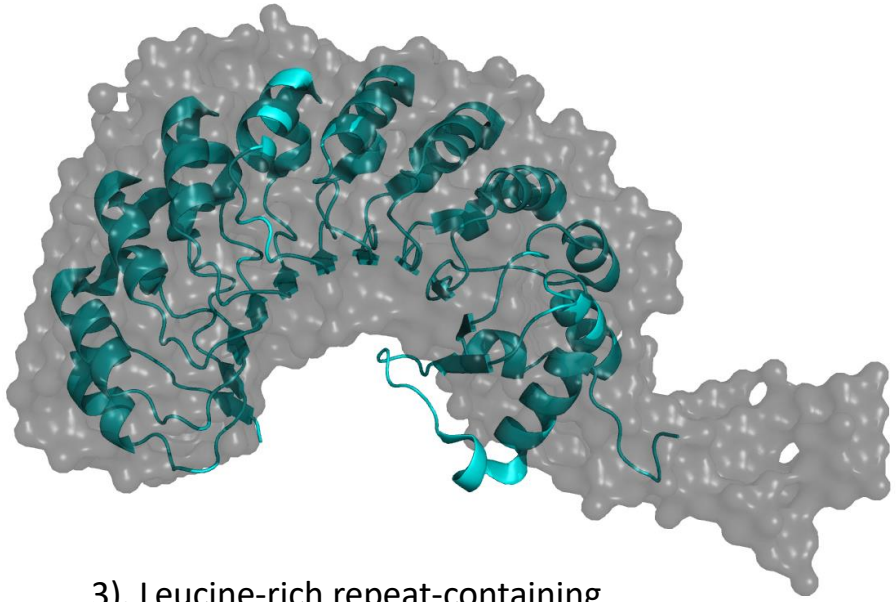
# Overlaid with subsequent X-ray structures



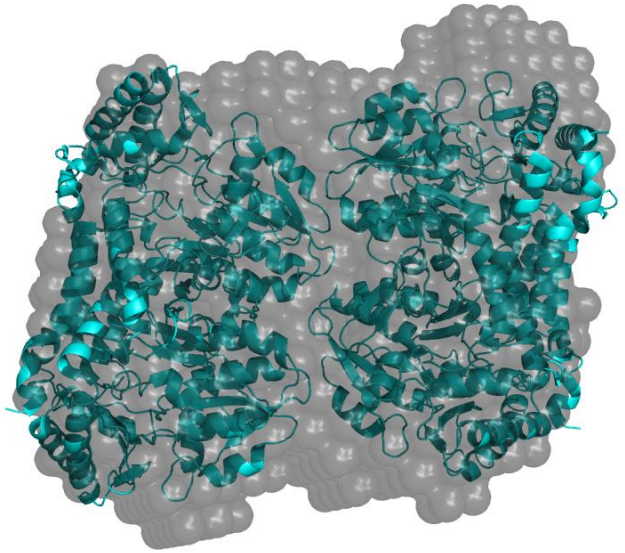
1). alr0221 protein from Nostoc (18.6 kDa)



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

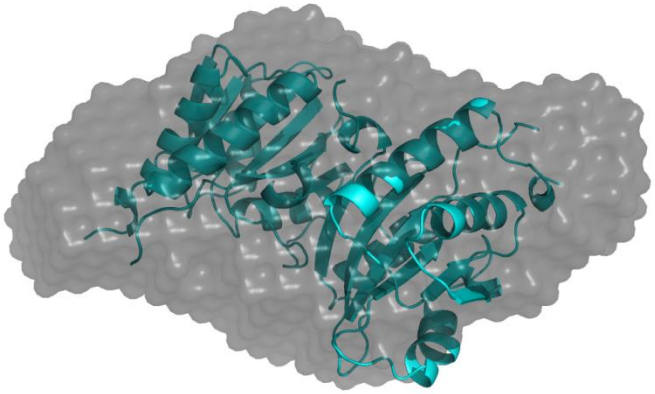


3). Leucine-rich repeat-containing protein LegL7 (39 kDa)

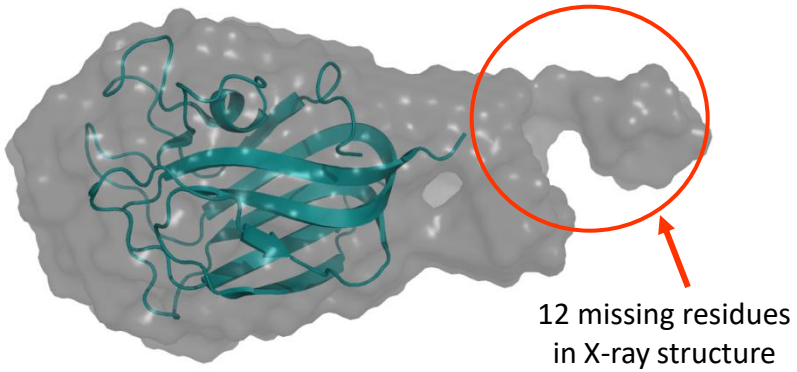


4). *E. Coli.* Cystine desulfurase activator complex (170 kDa)

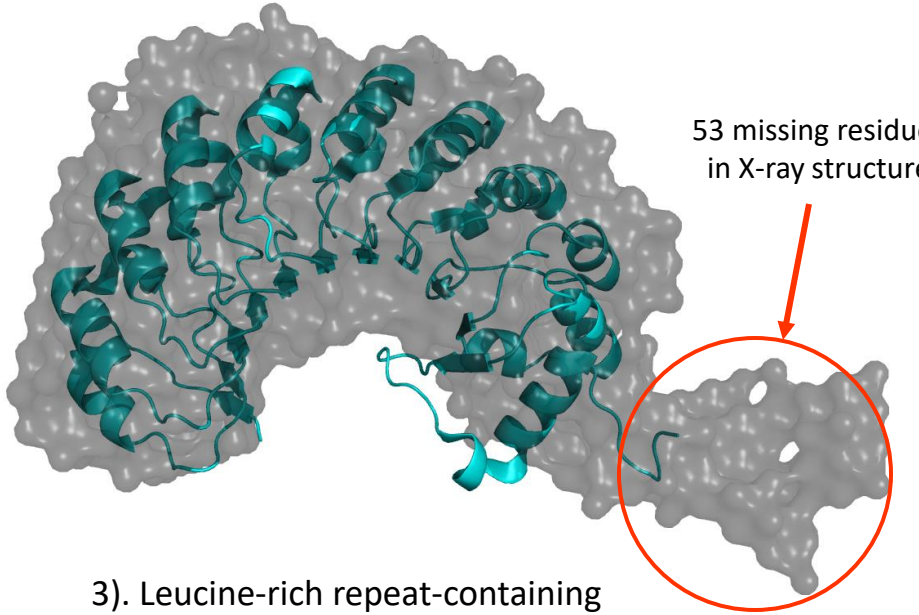
# And data on what was missing …



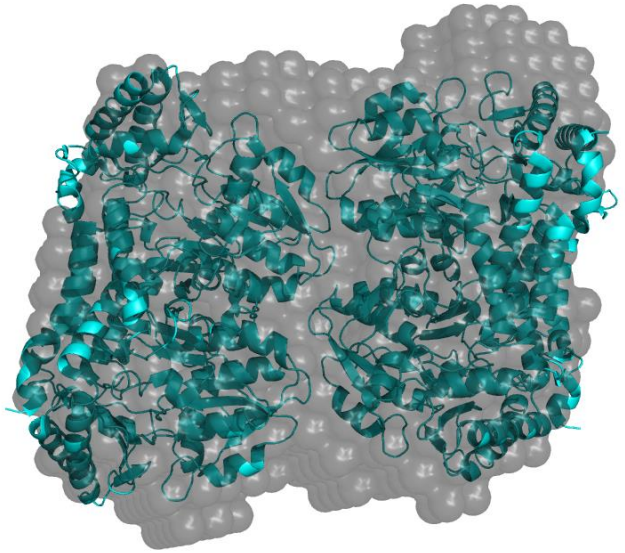
1). alr0221 protein from Nostoc (18.6 kDa)



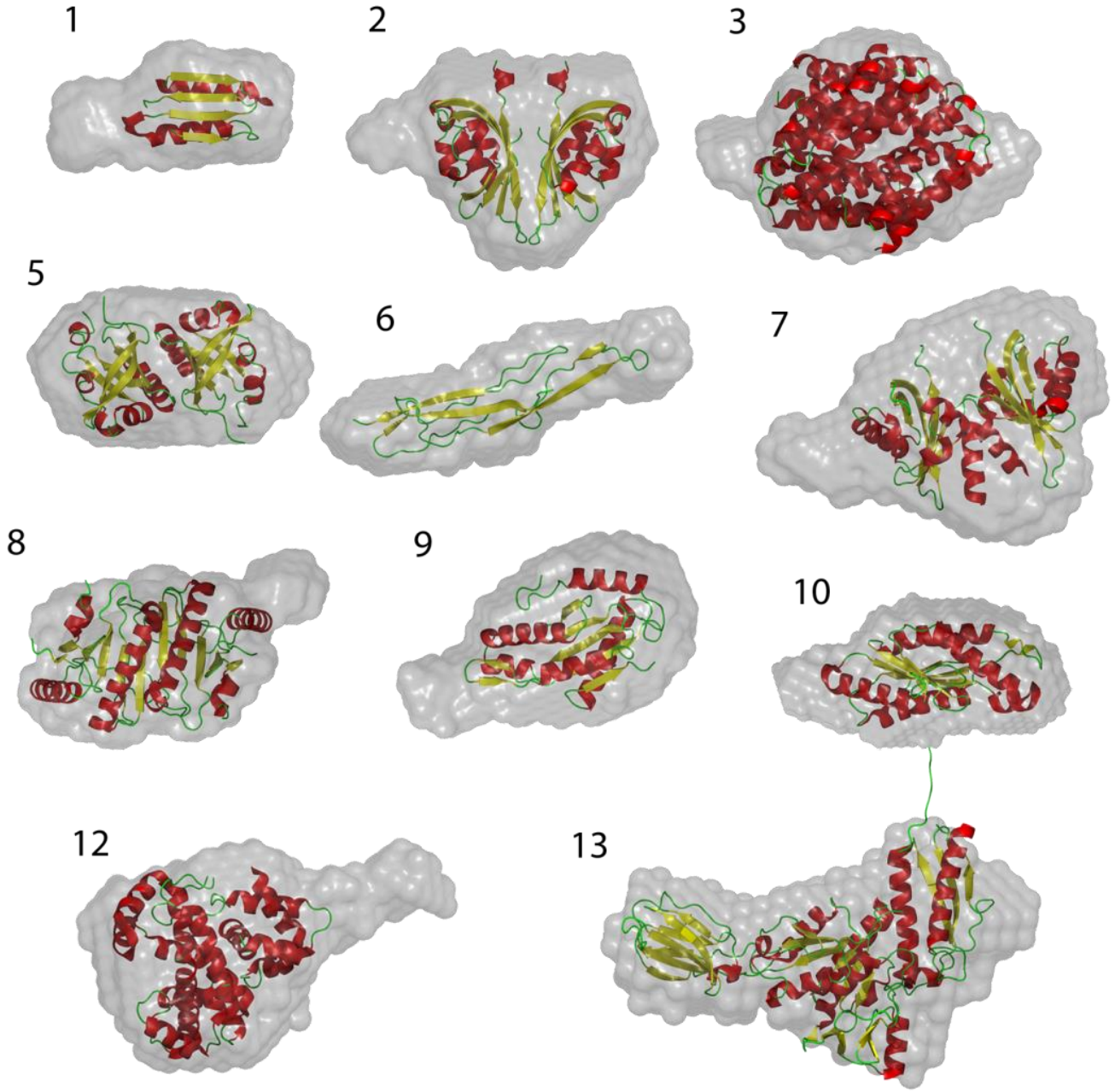

2). C-terminal domain of a chitobiase (17.9 kDa)



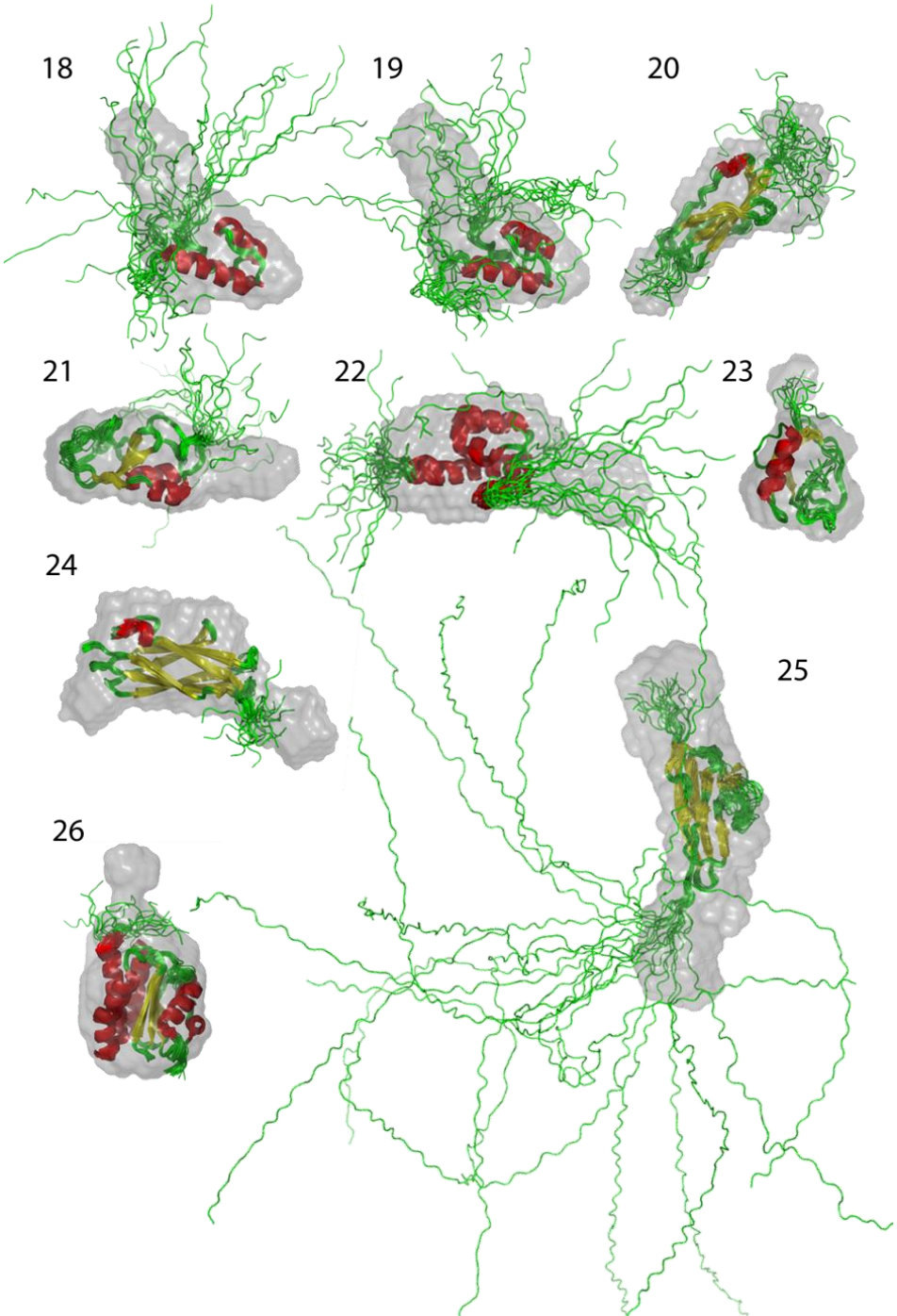

3). Leucine-rich repeat-containing protein LegL7 (39 kDa)



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



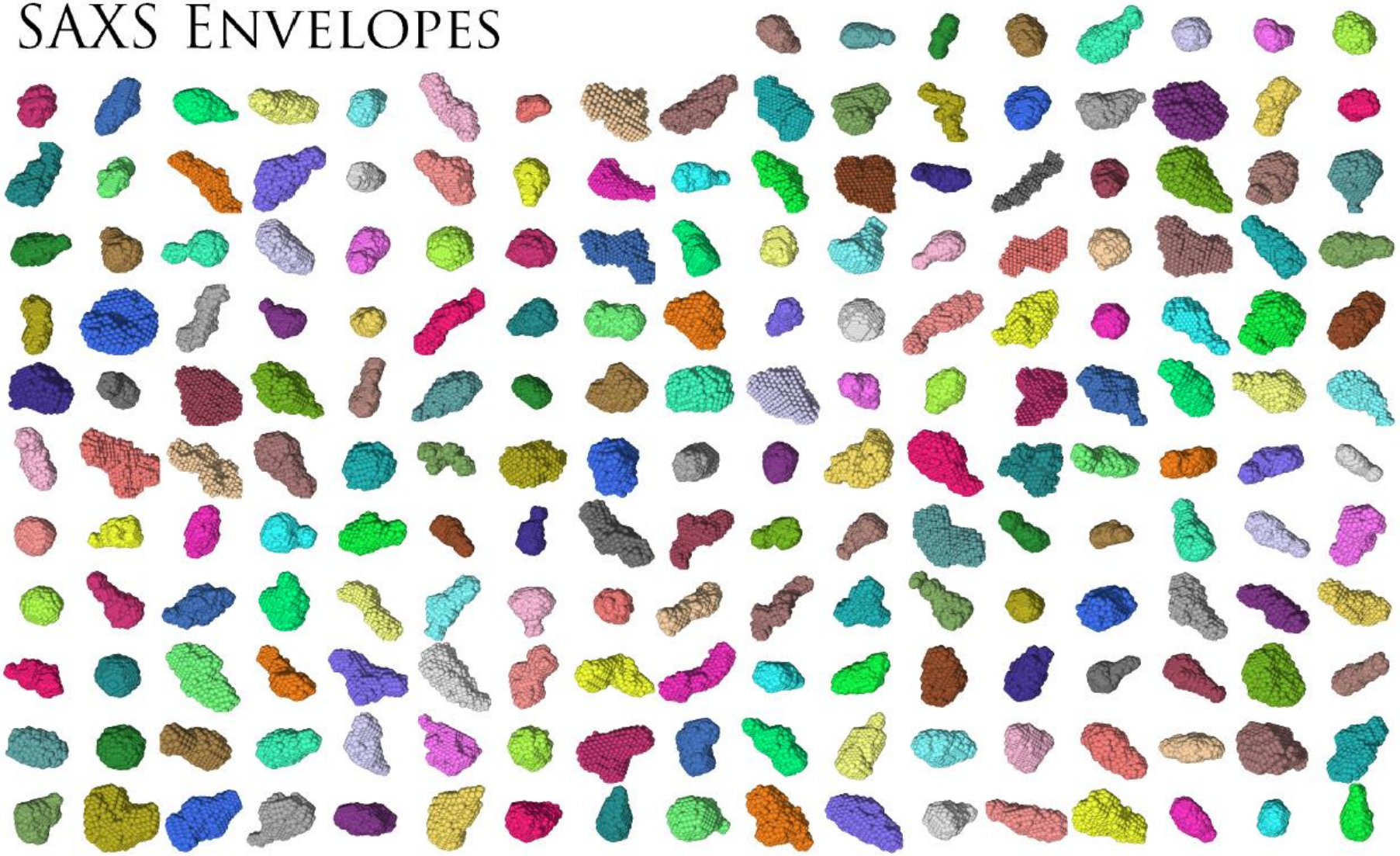

Comparing X-ray structures




## Comparing NMR structures

Data set of SAXS profiles from NESG crystallographic and NMR structures now being used by four groups (three on West Coast, one on the East,)to develop, test and validate new SAXS data processing and reconstruction developments.

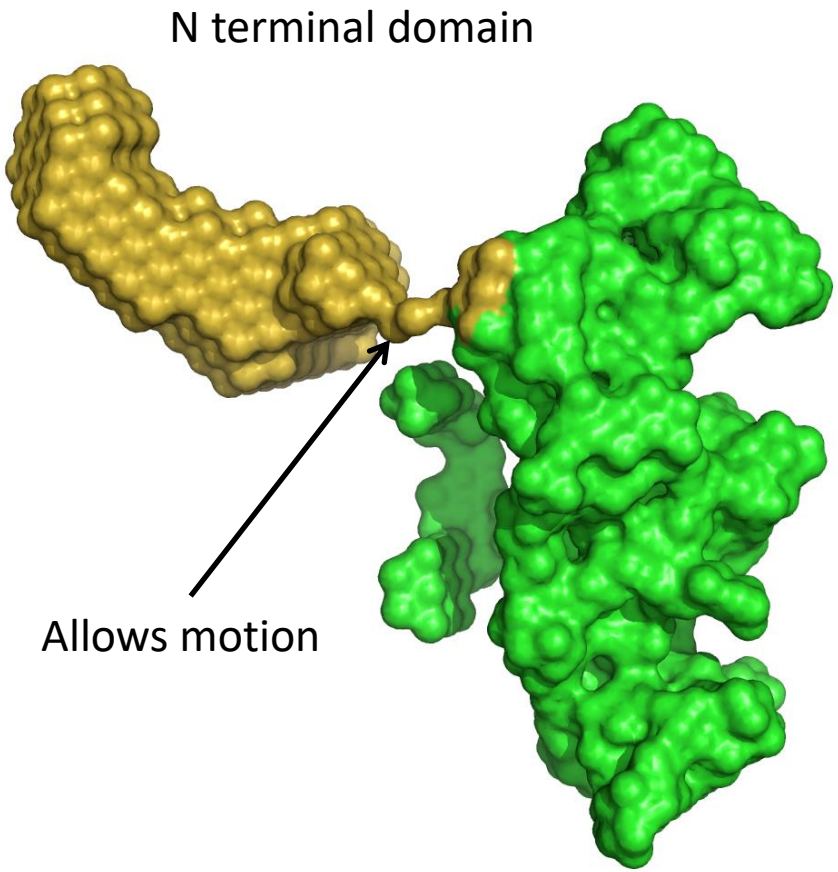
# SAXS Envelopes



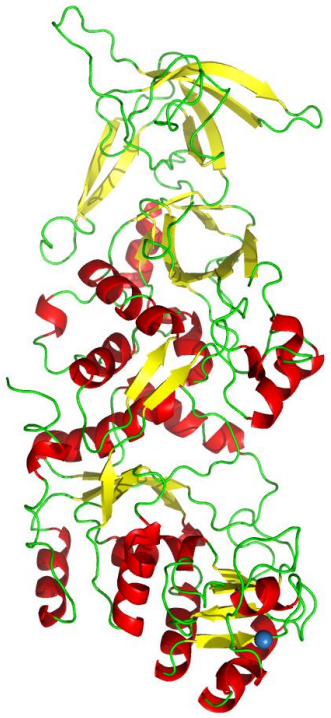
796 PSI samples, >50 other in-lab and collaborative studies

# Back to our sample of interest

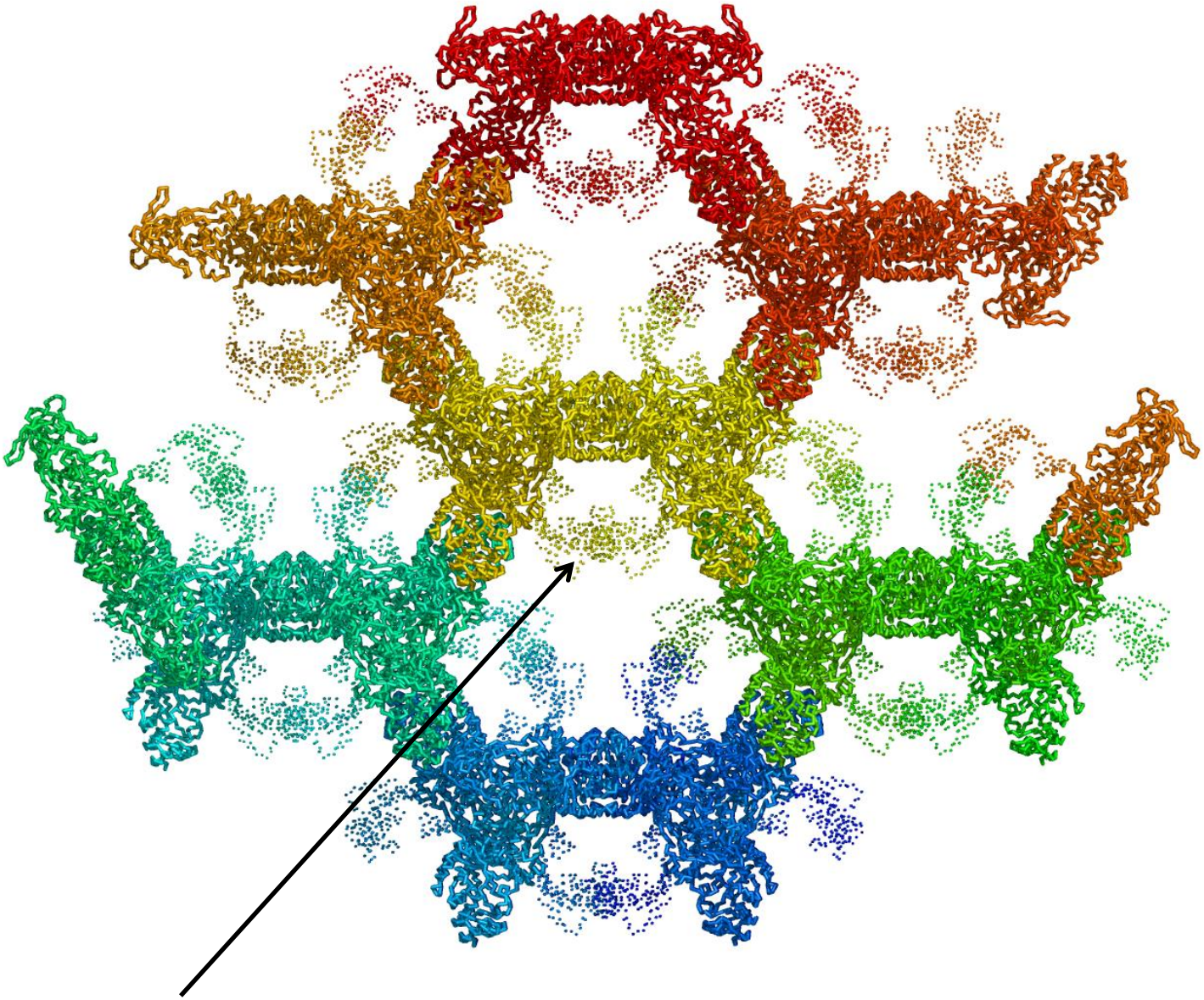
# Envelope reconstruction using the crystallographic structure




C terminal domain

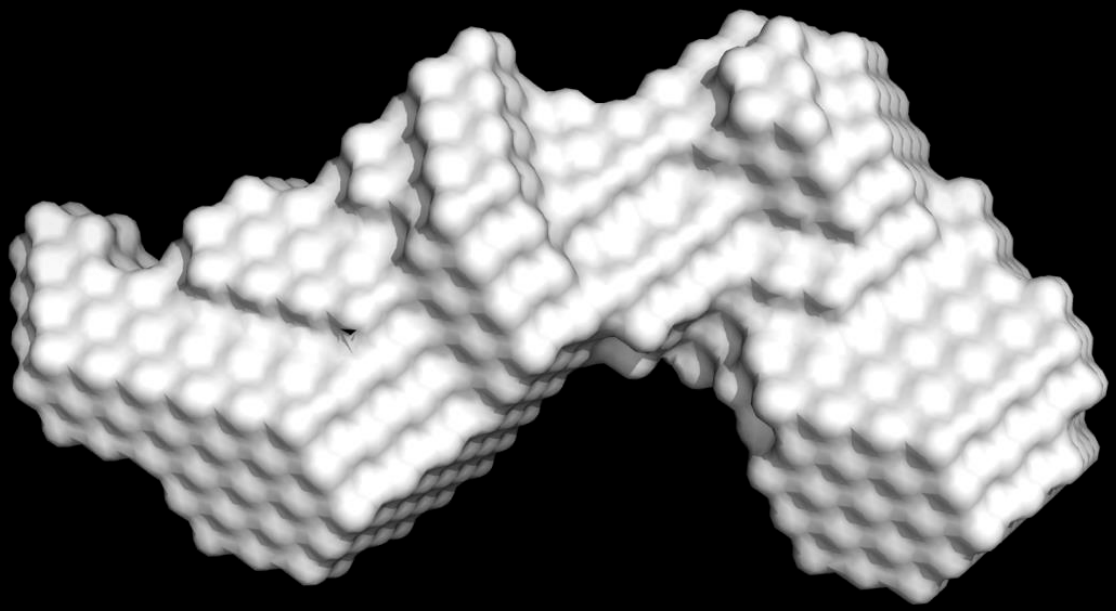


The crystal structure (which shows only the C-domain)



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

# Envelope reconstruction of the N-terminal domain

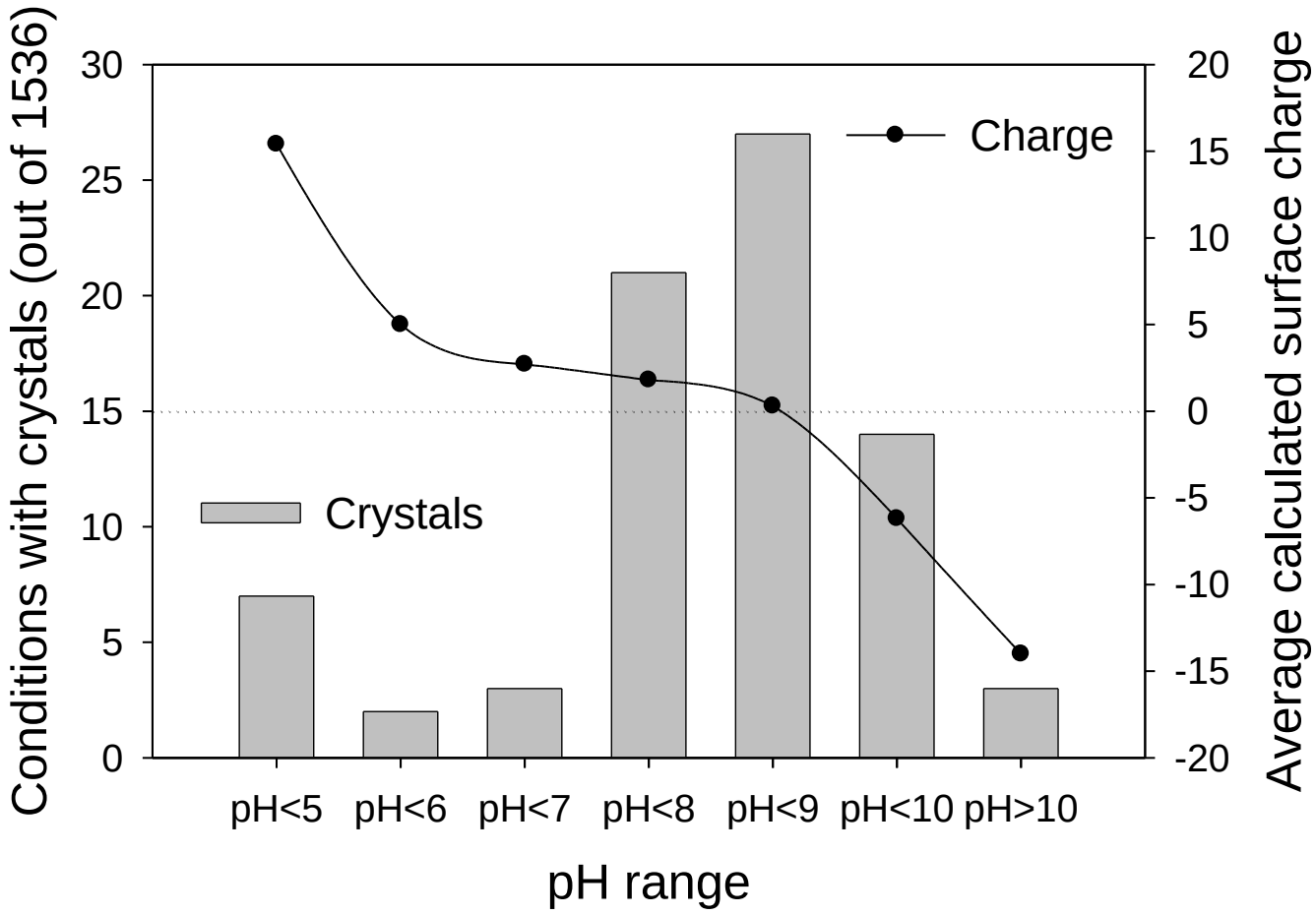


Express N-terminal domain, C-terminal domain, tRNA, SAXS studies on all

Check the crystallography again

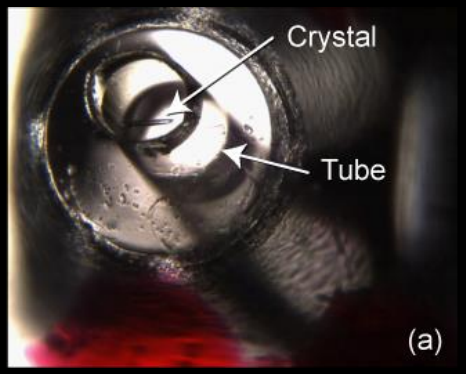
Factorial sampling of chemical space (ask the protein where it is happy)

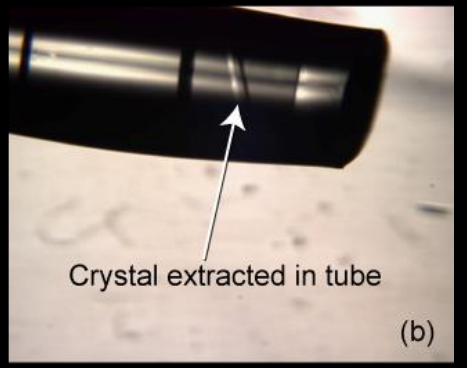
# Crystallization trials of the N-terminal domain

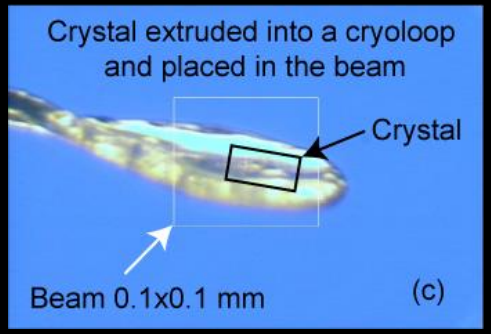


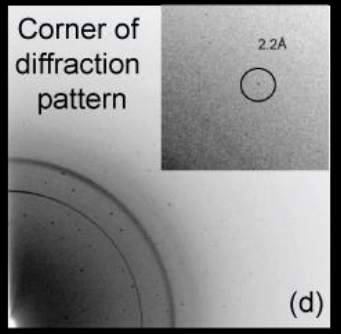
# Why bother with optimization?

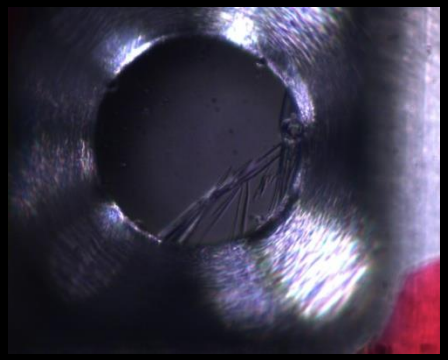
# Does it diffract? Screening before the synchrotron

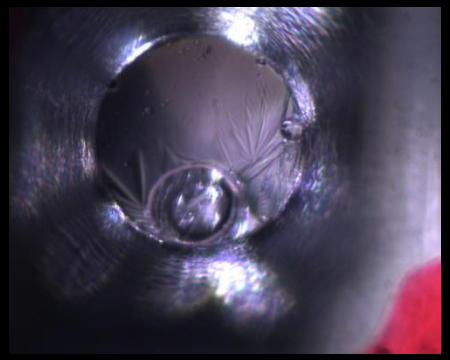


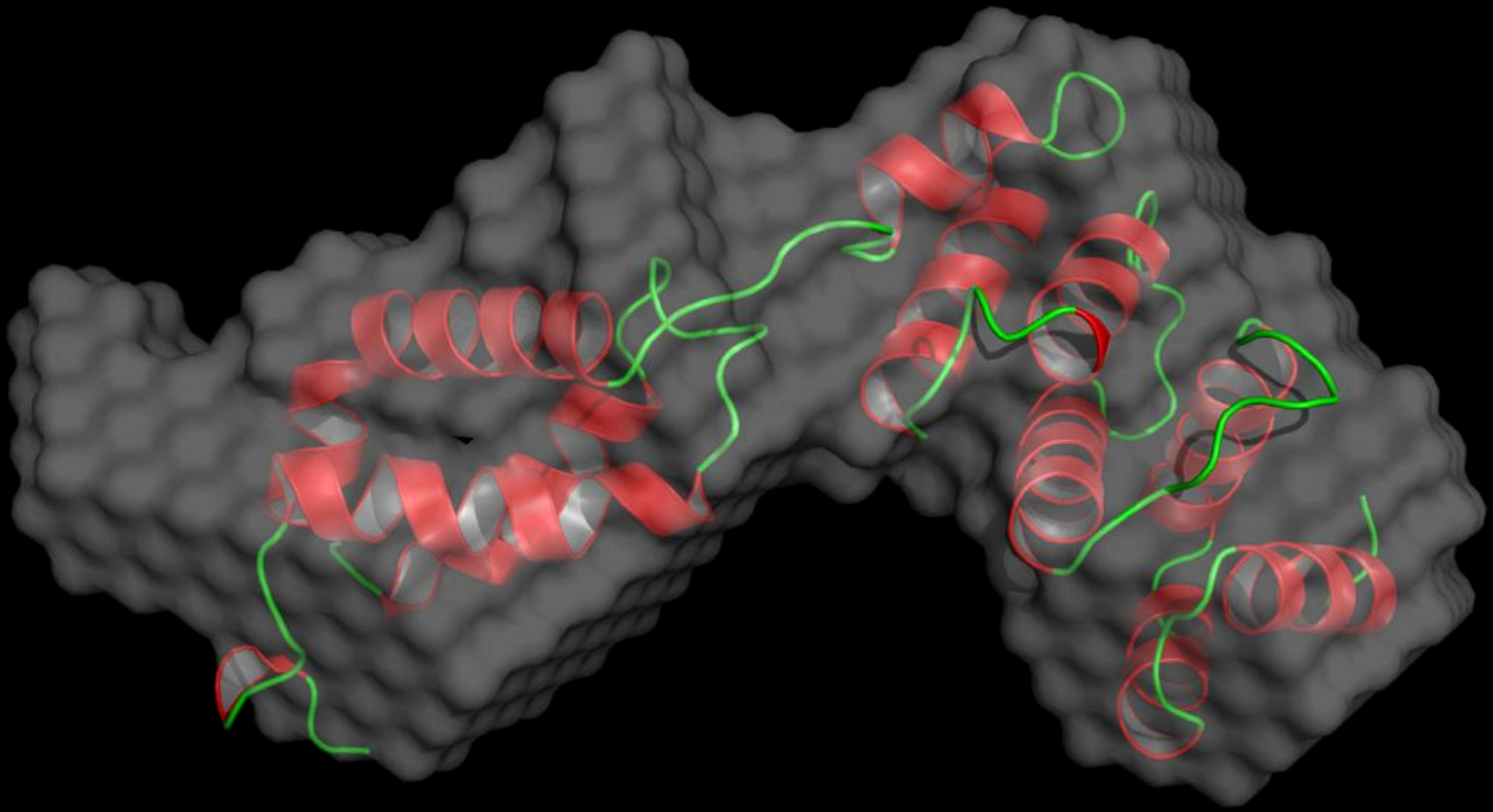



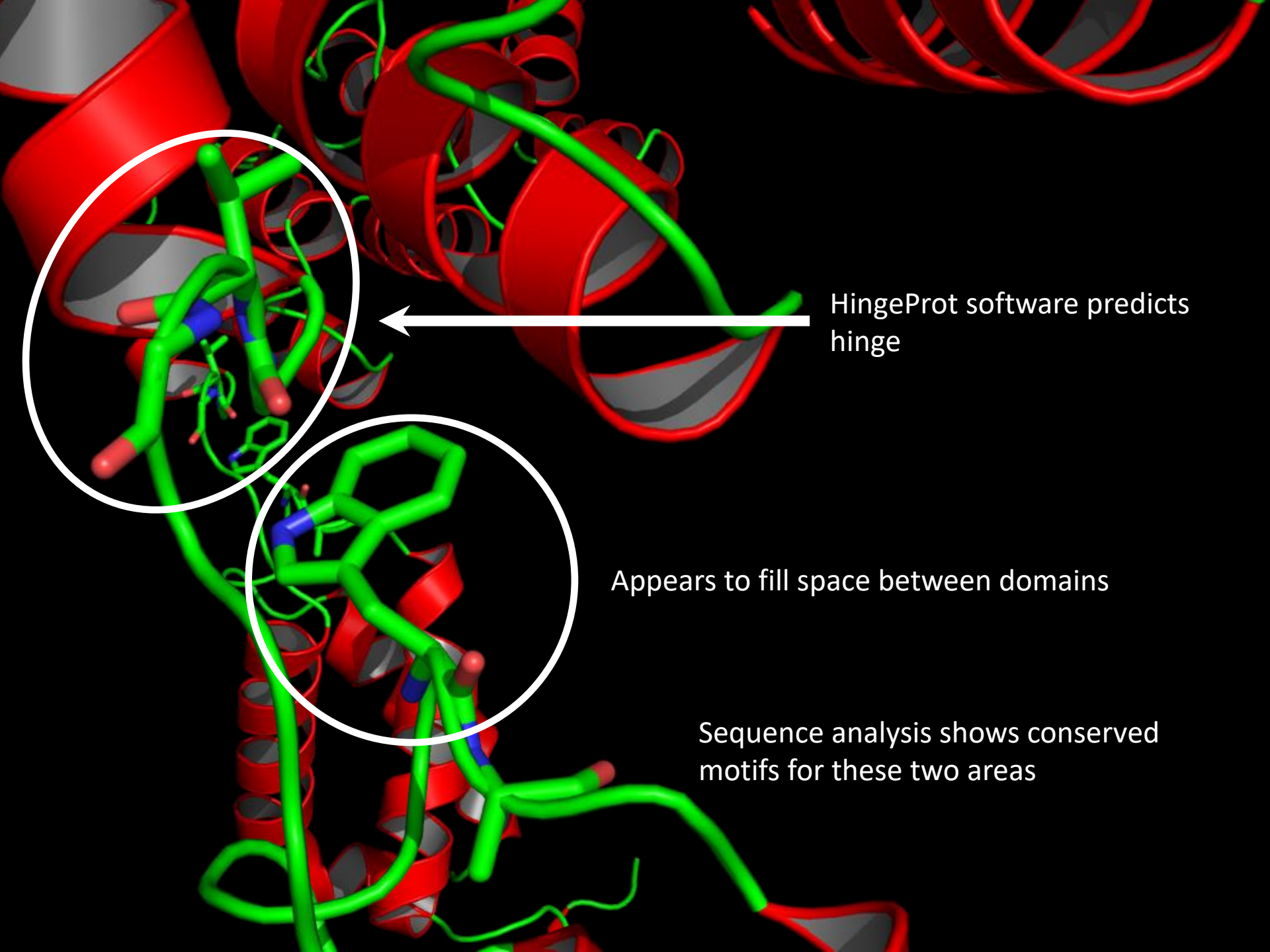







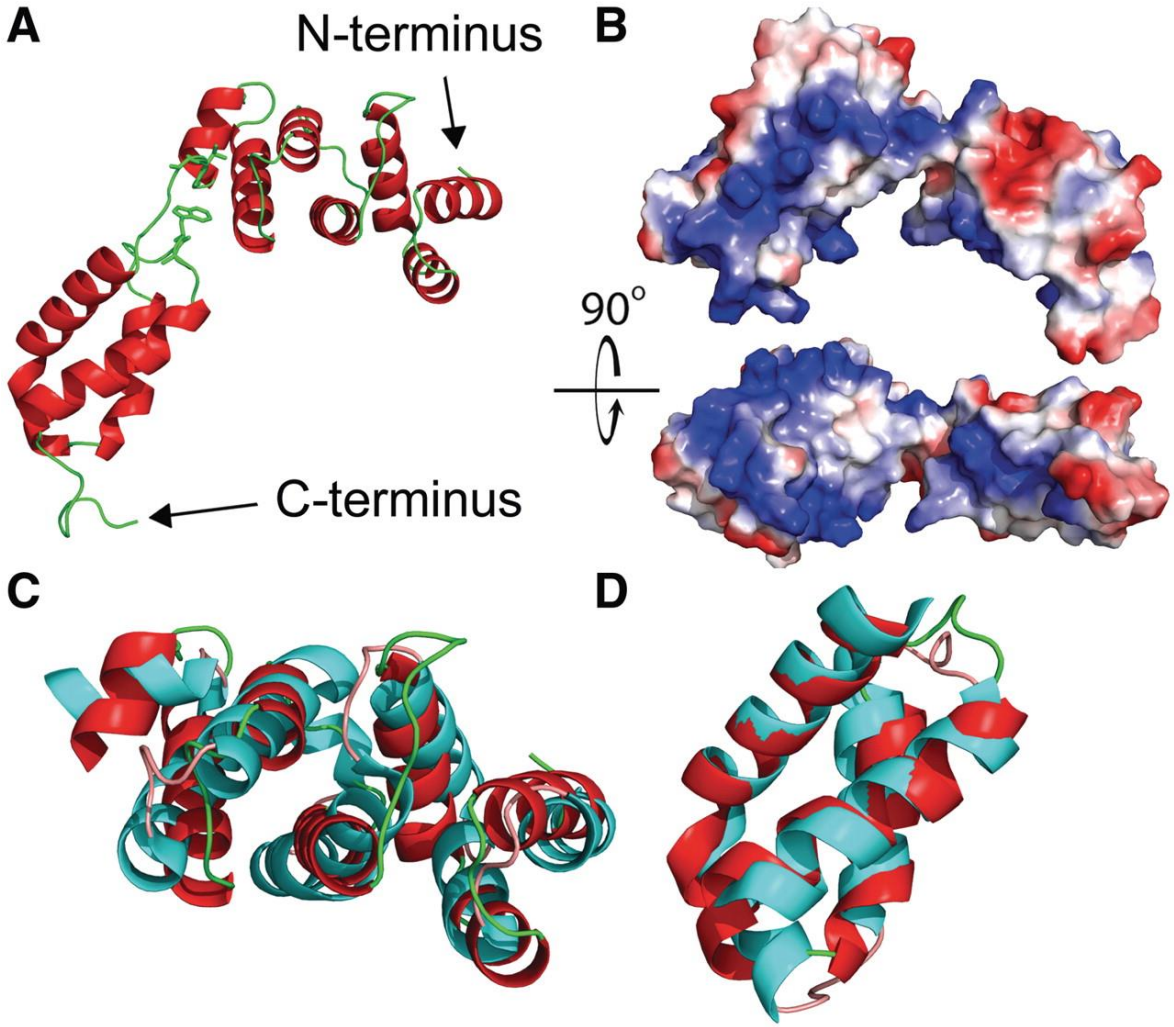

HingeProt software predicts hinge
Appears to fill space between domains
Sequence analysis shows conserved motifs for these two areas

# 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).**

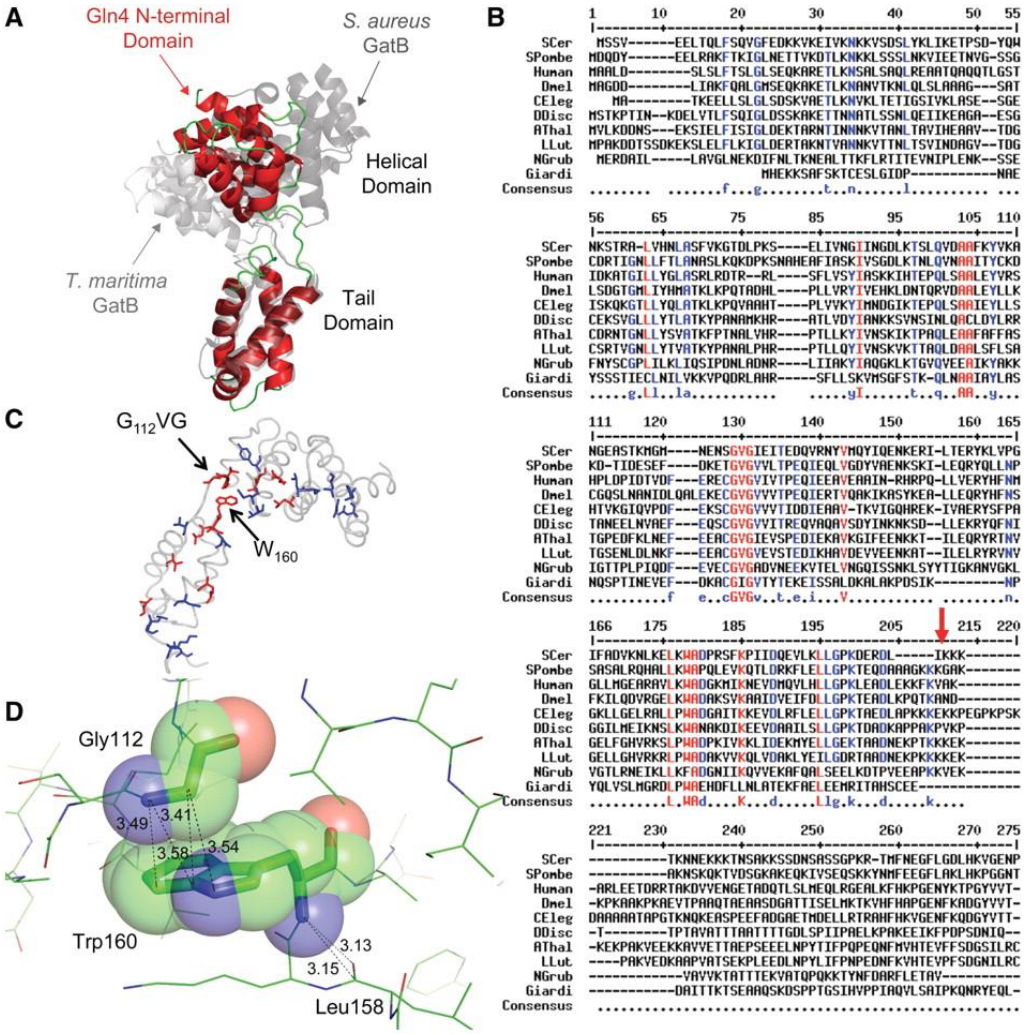



**Grant T D et al. Nucl. Acids Res. 2011;nar.gkr1223**



Nucleic Acids Research

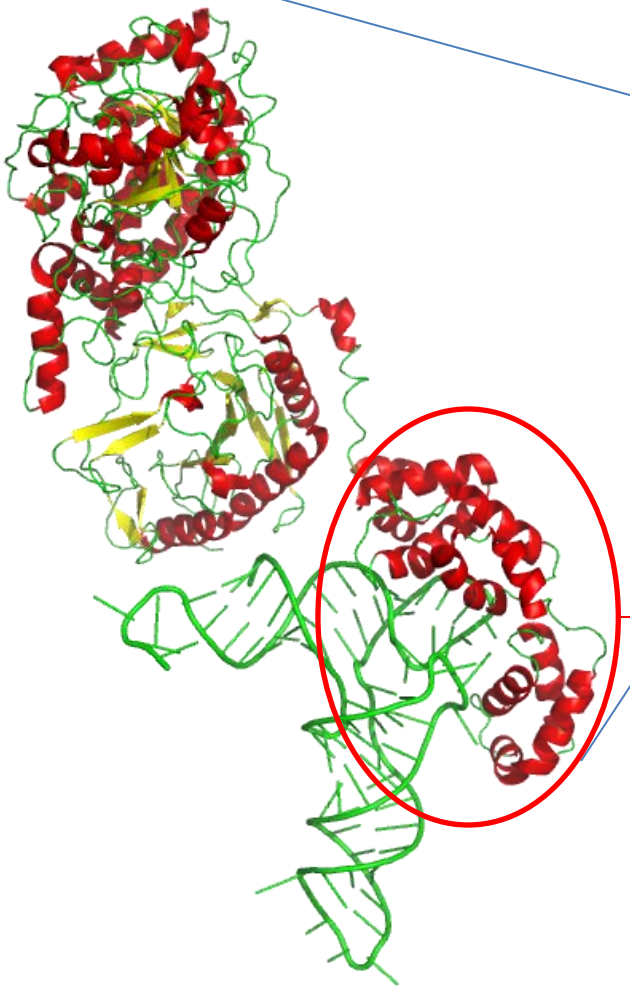
# The linker between the two domains in Gln4(1–187) likely behaves as a hinge, is highly conserved and is important for tRNA binding.

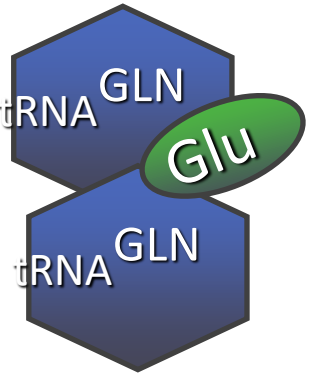


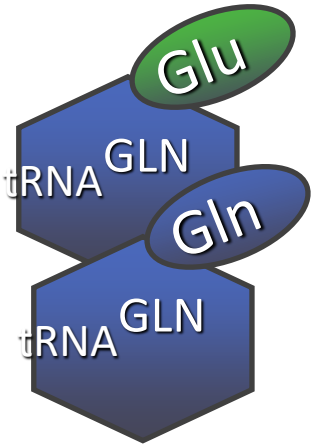

Grant T D et al. Nucl. Acids Res. 2011;nar.gkr1223



Nucleic Acids Research



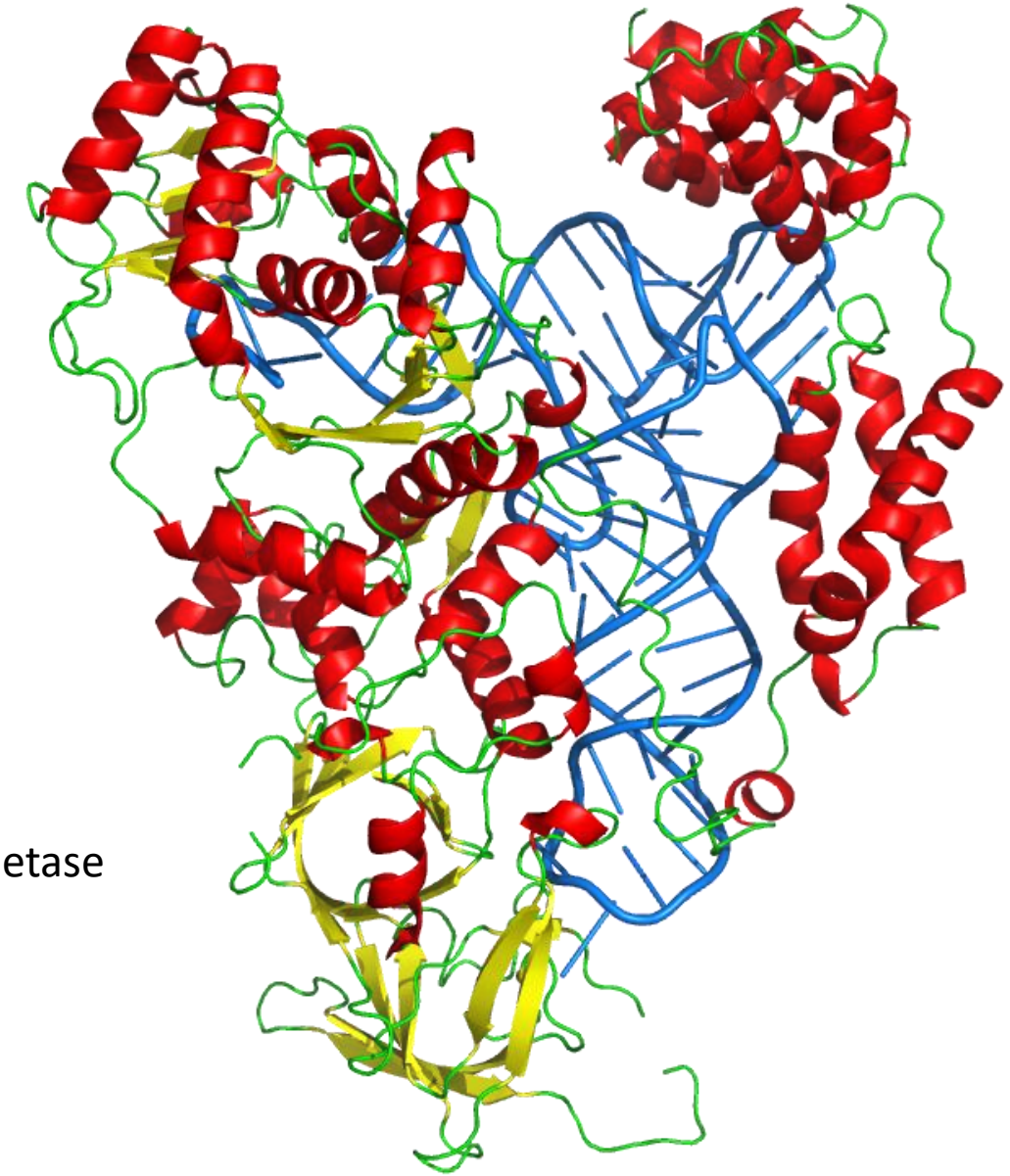




Remarkably similar to the N-terminal domain of Eukaryotic GlnRS

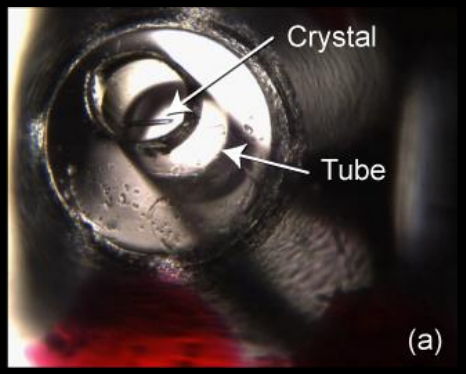
# Combine the SAXS and Crystallography

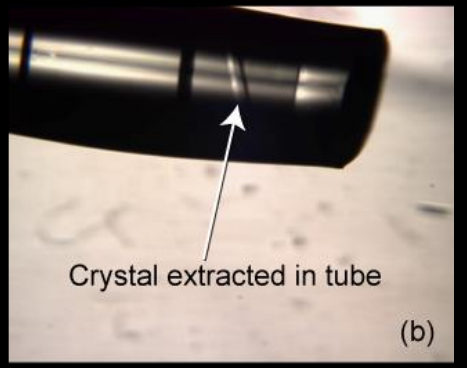
Gln4 a Eukaryotic
Glutaminyl-tRNA Synthetase

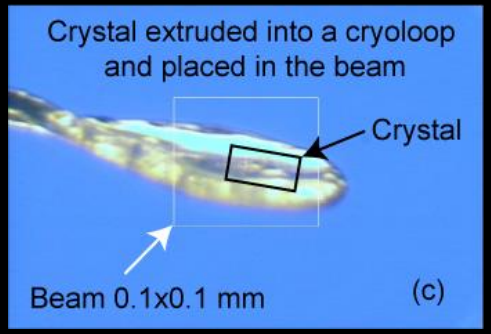


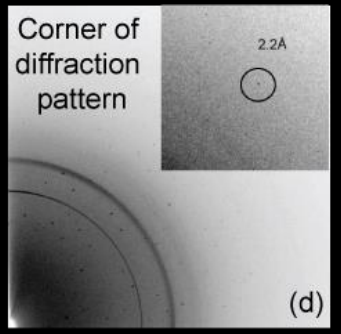
Where are we heading now?

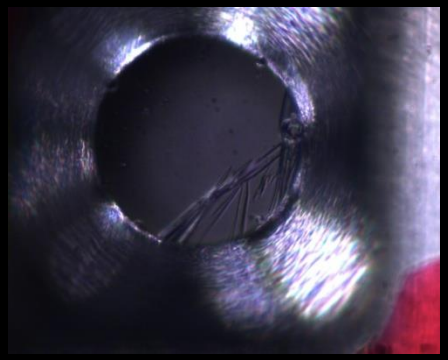
# Does it diffract? Screening before the synchrotron

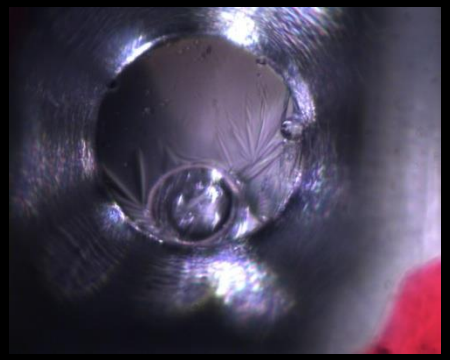


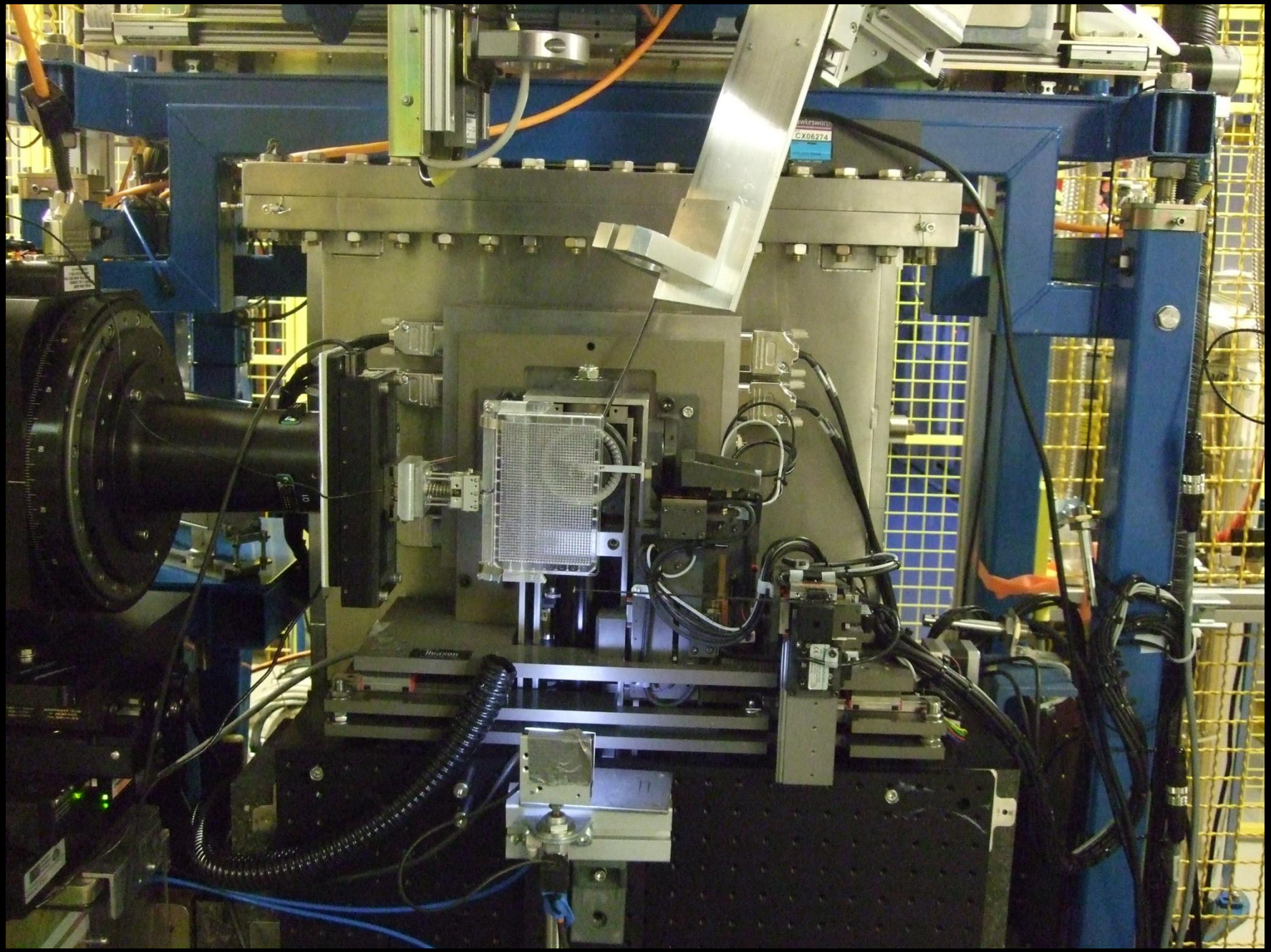



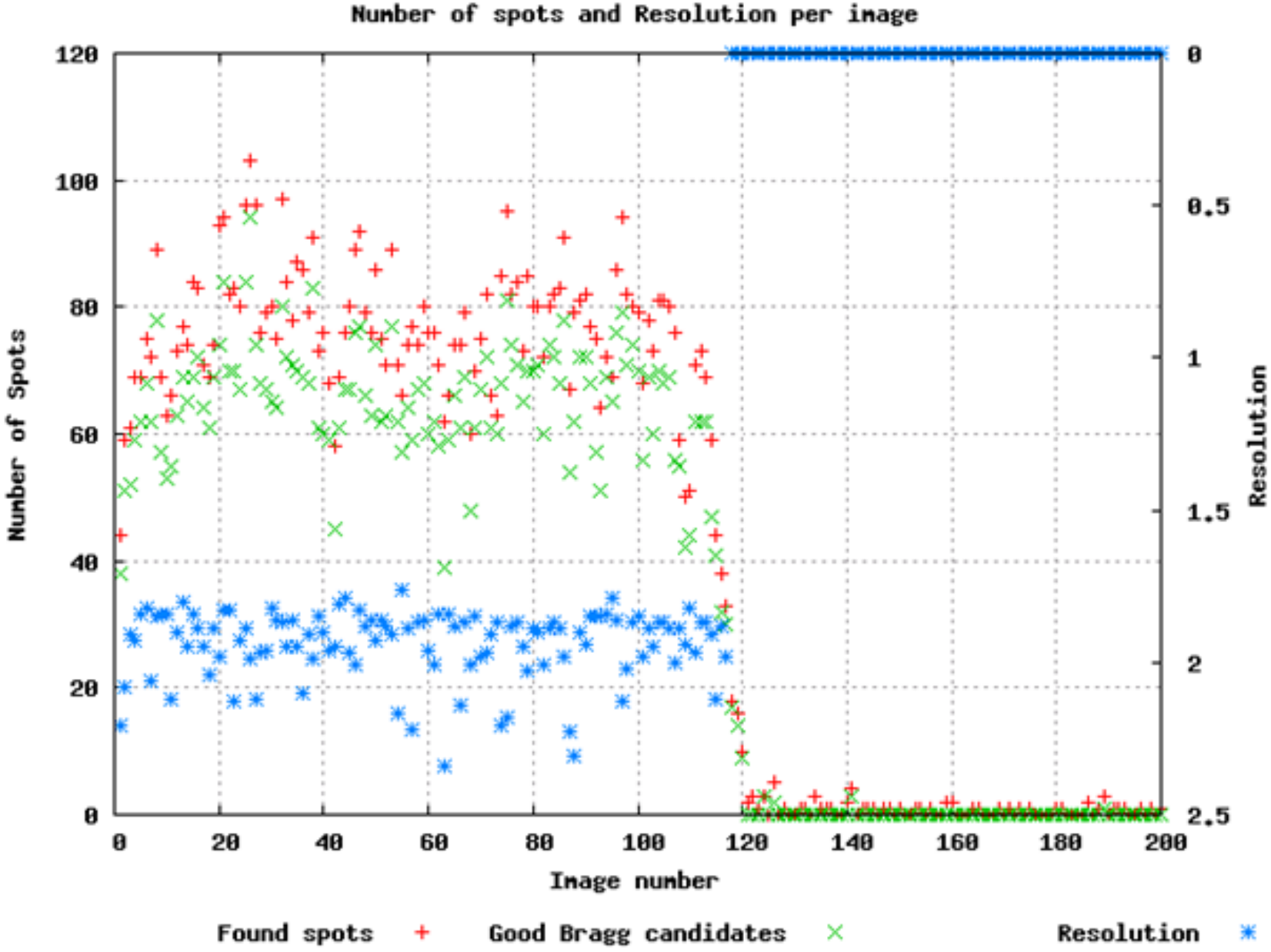






CX06274


55
56.0 µm


Number of spots and Resolution per image
Number of Spots
Resolution
Image number
0
20
40
60
80
100
120
140
160
180
200
0
0.5
1
1.5
2
2.5
Found spots
Good Bragg candidates
Resolution

# Developments

- http://xtuition.ccr.buffalo.edu/devel/ipad.php

- – 4,000 proteins, with crystallization results for 1,536 different conditions and images weekly over 6 weeks.
- Finger selection of interesting crystals with ‘fingerprint’ equaling synchrotron beam profile.
- Generation of file that can be loaded into beamline robot.
- Rapid in situ diffraction from plates (~40 degrees of data). Diffraction from individual crystals (standard beam footprint 20 micron, small 5 micron).

# 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

**Diamond Light Source**

Robin Owen

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