Wah Chiu

Stanford University

wahc@stanford.edu

Challenges and Opportunities

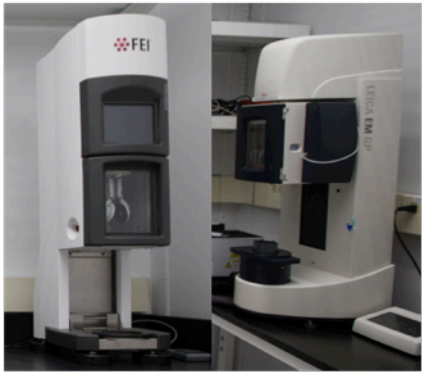
Questions from Clint and Bridget

- Validation methods have become much better established over the last couple of years.
- What are the methods that are being used?
- In what resolution realms are they useful?
- Do we need more tools?
- How do we avoid mistakes?
- Is validation at very high resolution easier than at intermediate resolutions?
- What about highly heterogeneous datasets?

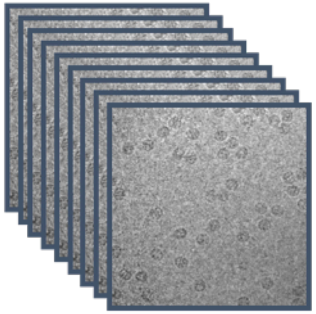
Cryo-EM Workflow

Biochemical Preparation

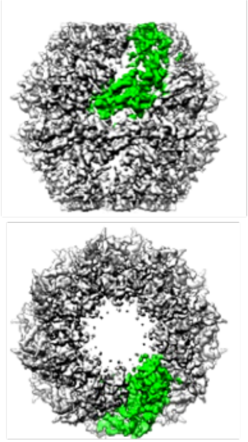
Cryo-EM Sample Preparation



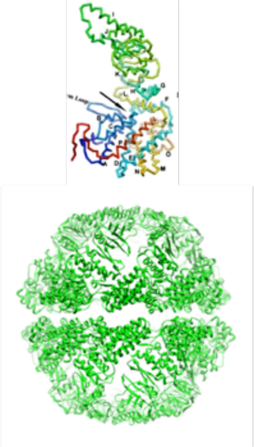
High Resolution Automated Data Collection



Data Archiving & Processing



3D Reconstruction



Model Building & Validation

WORLDWIDE PDB PROTEIN DATA BANK

Structure Deposition

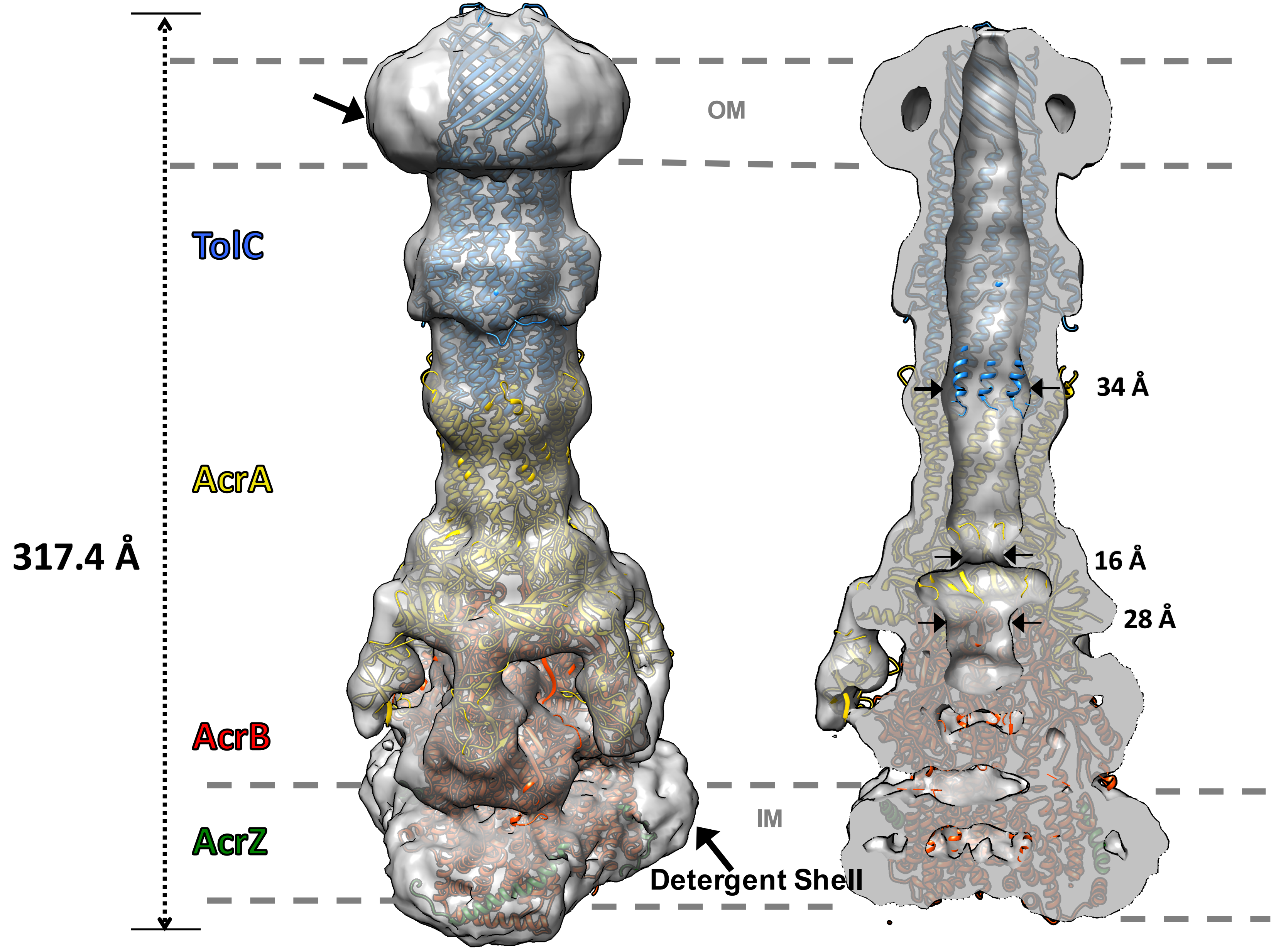
1. Challenges in Samples

- Flexibility
- Heterogeneity
- Transient

1. Challenges in Samples

- Flexibility
- Heterogeneity
- Transient

16 Å resolution cryoEM Map of AcrABZ-TolC



D Du, Z Wang, B Luisi et al Nature 2014

Improve Purification of AcrABZ-TolC

co-expression in cell culture

Cellular membrane

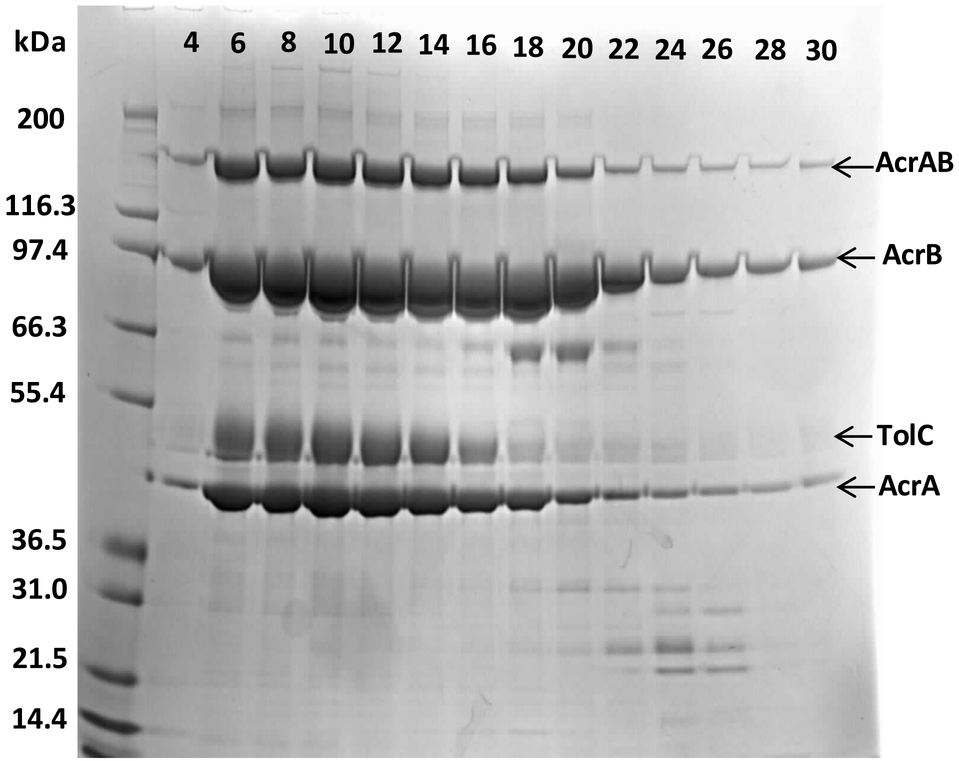
Solubilization with 1.5% DDM

Nickel affinity chromatography

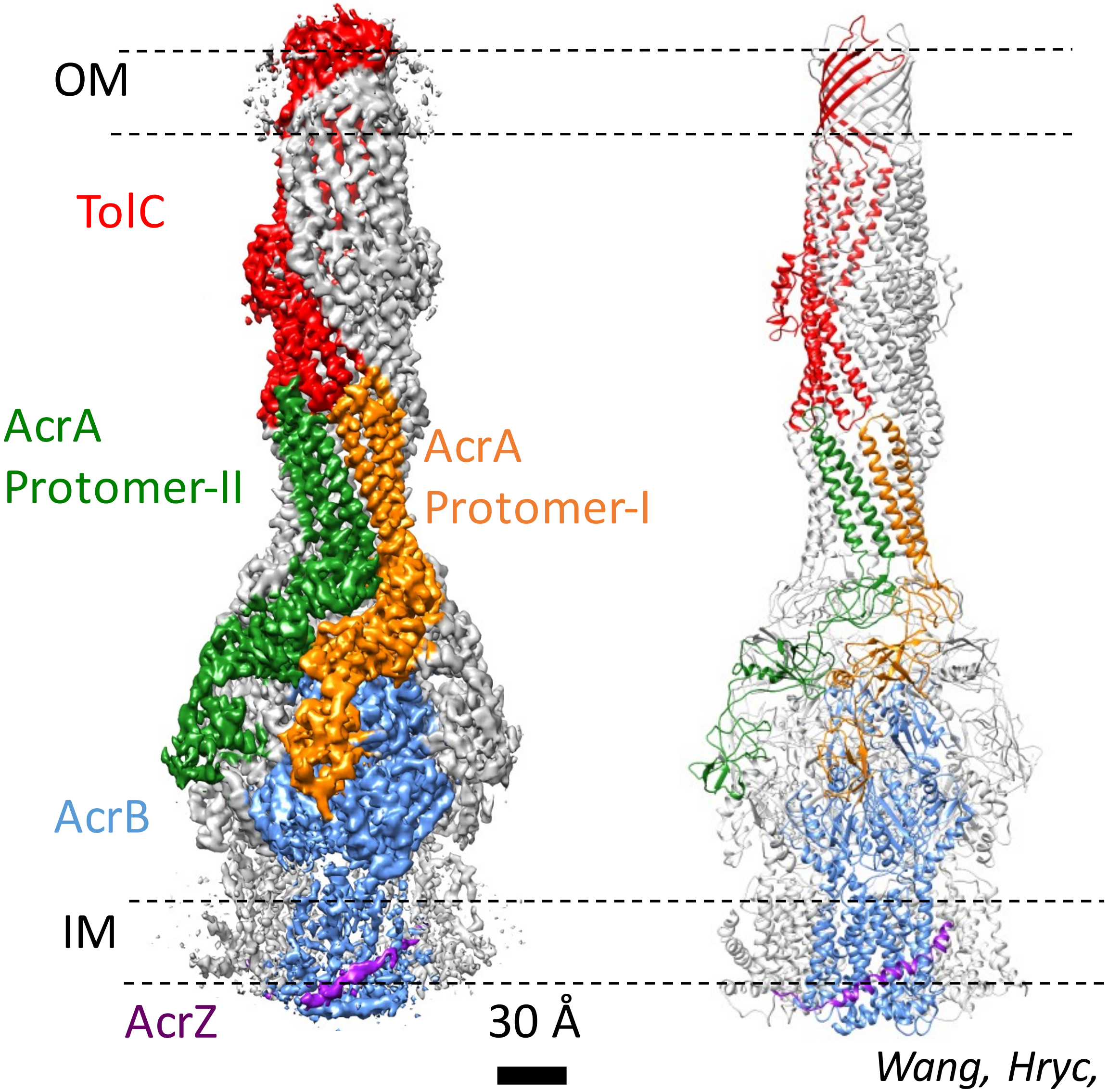
Amphipol A8-35 change

We add inhibitor MBX3132 to stabilize pump

SDS-PAGE



3.6 Å Cryo-EM Structure of the AcrABZ-TolC Pump with Inhibitor



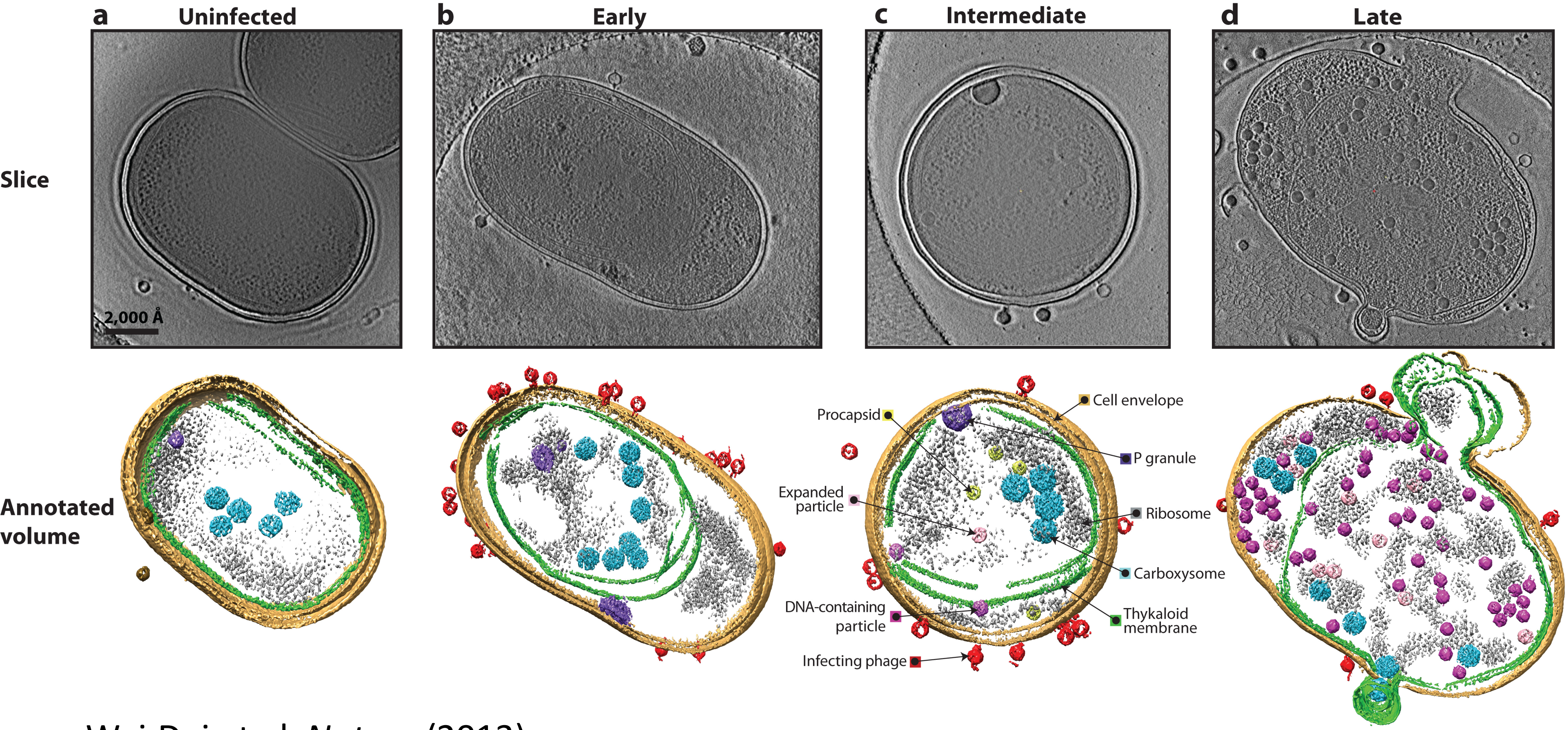
TolC :AcrA: AcrB = 3:6:3

Wang, Hryc, Du, Luisi et al Elife 2017

1. Challenges in Samples

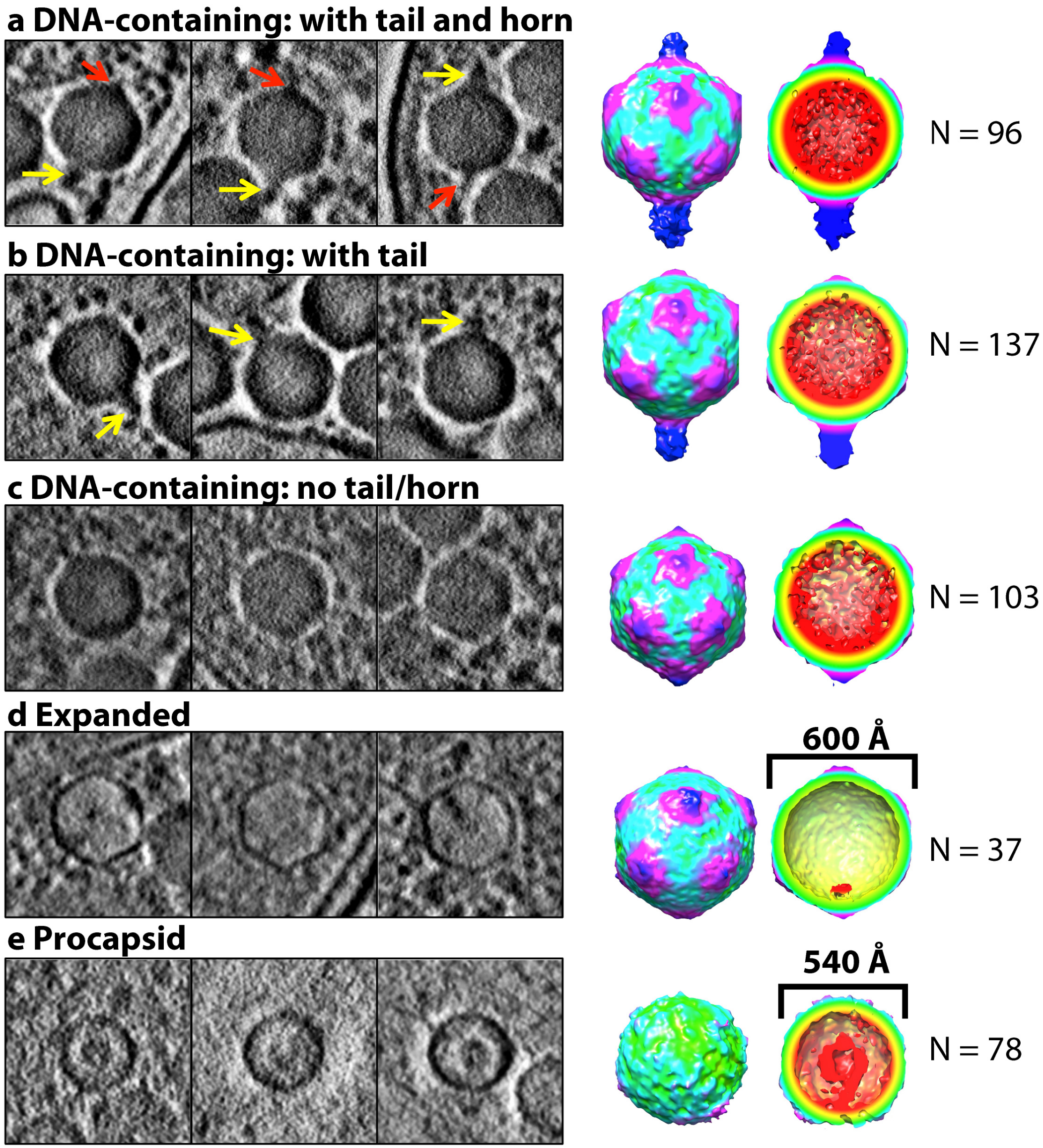
- Flexibility
- Heterogeneity
- Transient

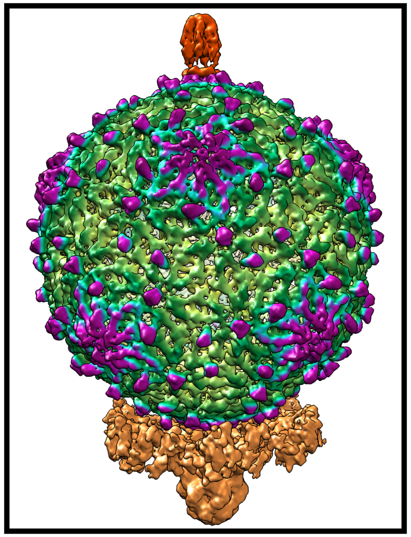
Zernike Phase Contrast Cryo-ET of Syn5 Infected Cell at Various Stages of Infection



Wei Dai et al, *Nature* (2013)

Five Types of Syn5 Particles Classified from Subtomograms





W Dai et al, *Nature* (2013)

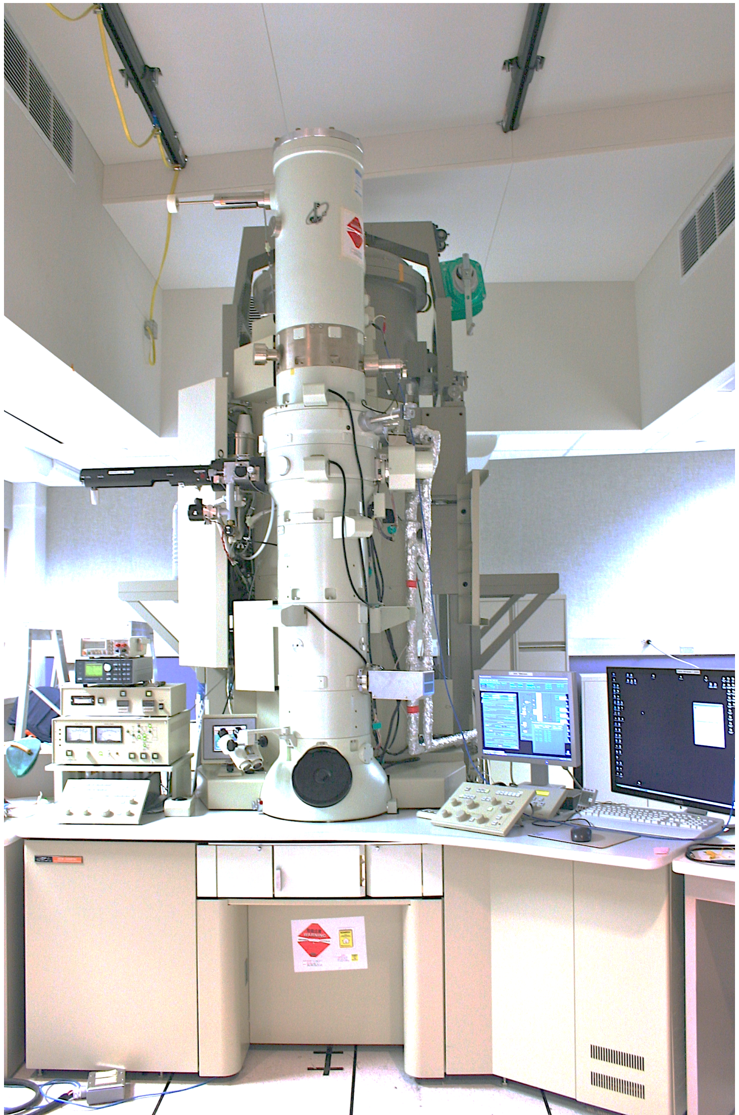
2. Unsolved issues in Cryo-Specimen Preparation

- Preferred orientation
- Falling apart in cryo-preservation
- Capture reaction intermediates
- Precious samples

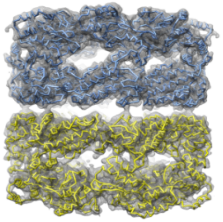
New Product Development Opportunities

- New types of grids: cheaper and faster deliveries
- New freezing devices and protocols
- Freezing device to handle potentially biohazardous (BSL2) materials
- Bioreactor chamber to allow performing biochemical reactions
- Single device for freezing and grid loading to a cryo-holder in a normal room humidity

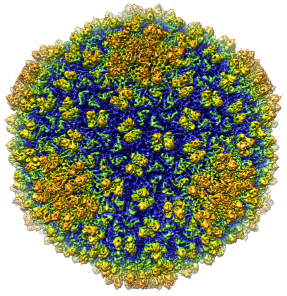
3. Electron Microscope Manufacturers



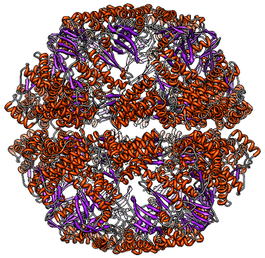
Backbone Traceable Cryo-EM Structures Using Films or CCD in a JEM3200FSC



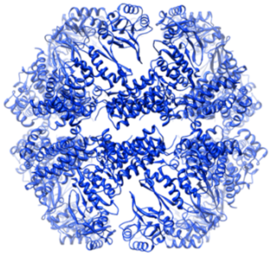
GroEL 2008



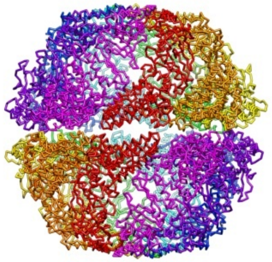
ε15 2008



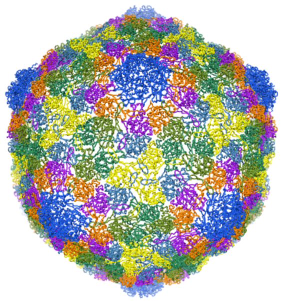
Mm-Cpn
2010



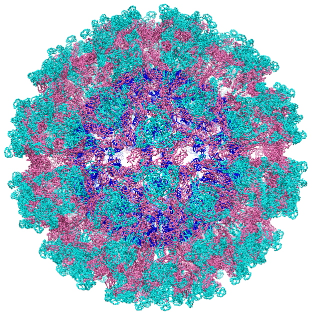
Mutant Mm-Cpn
2010



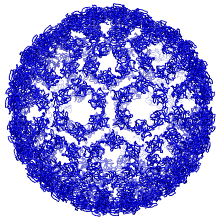
TRiC 2010



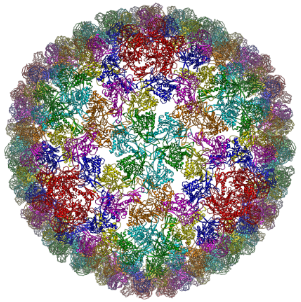
PSSP-7
2010



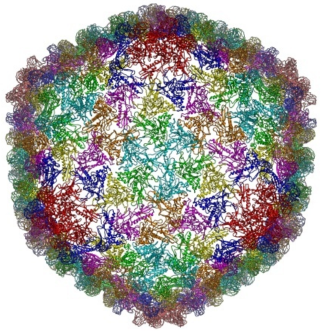
VEEV Shell
2011



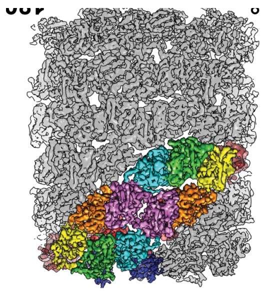
VEEV capsid
2011



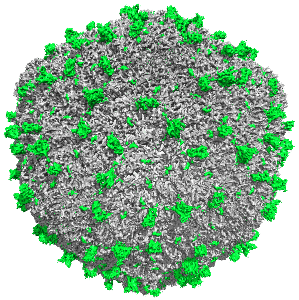
P22 procapsid
2011



P22 Phage
2011

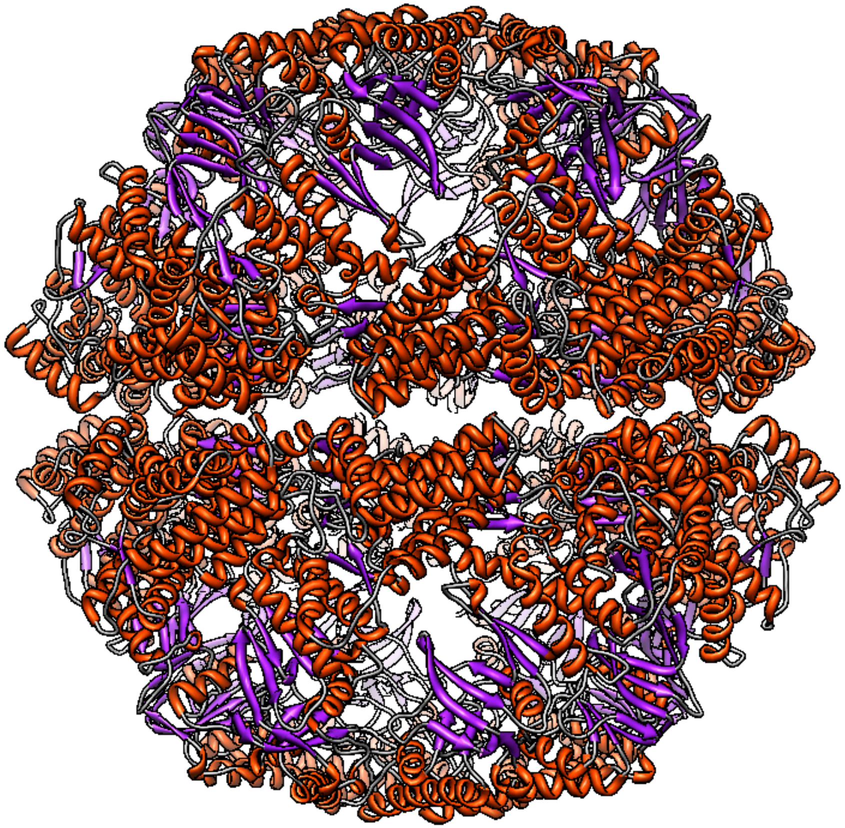


Hemocyanin
2013

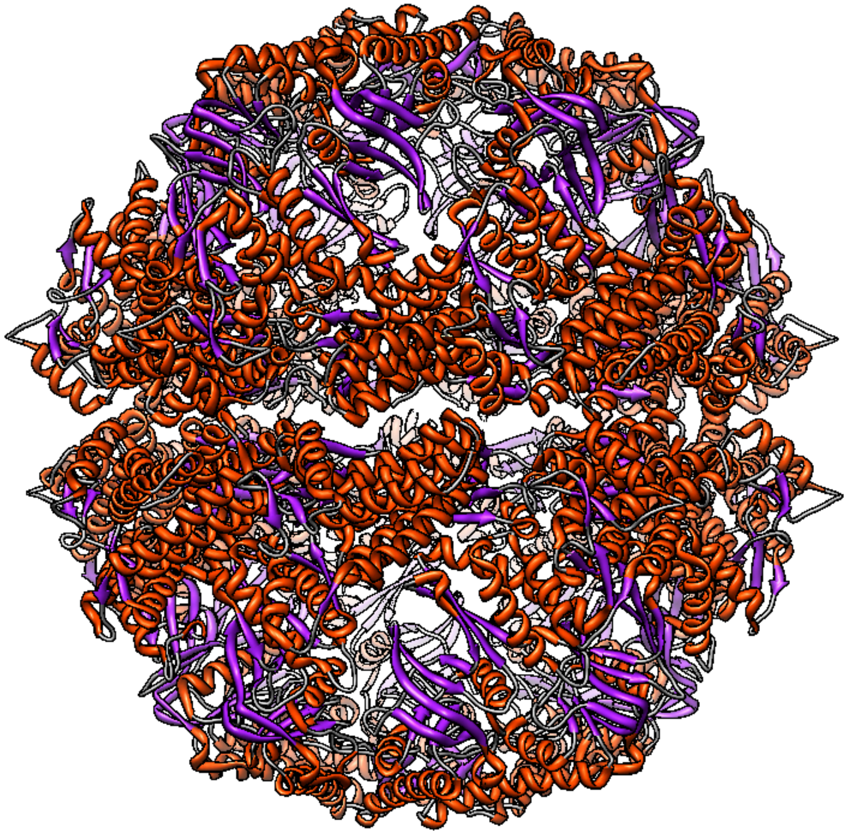


Syn5
2014

Cryo-EM vs Subsequent X-Ray Model of Mm-Cpn Chaperonin



Cryo-EM Model
Zhang et al *Nature* (**2010**) (PDB:3LOS)
DiMaio et al *Protein Science* (2013) (PDB:3J3X)



X-Ray Structure (PDB:3RUQ)
Periera et al *EMBO J* (**2012**)

3 - 4.5 Å Cryo-EM Structures Using Direct Detectors (DE12, DE20, K2) in JEM3200FSC

- 8 viruses
- 6 membrane proteins
- 14 protein complexes

Wish List on New Instrumentations

- Higher voltage
- Cold field emission gun
- Better phase plate
- Aberration free lens
- Energy filter
- Isotropic magnification
- Improved DQE and throughputs in cameras
- Less purchase and maintenance costs

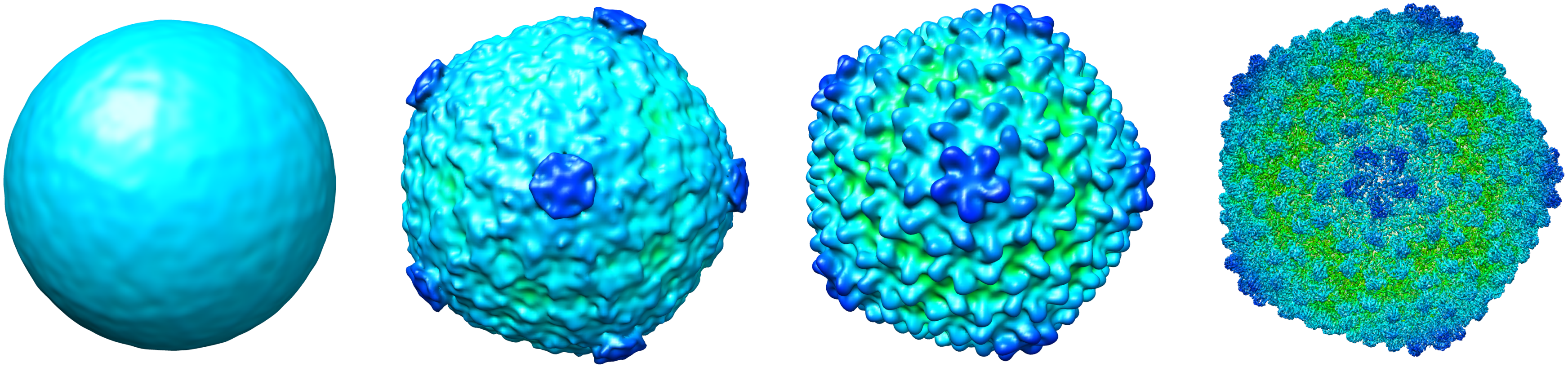
4. Opportunities and Challenges in Image Processing

- Data management
    - EMEN2 S Ludtke *Microsc Microanal* 2013
    - NRAMM Appion Clint Potter and B Carragher
    - LIM Alexis Rohou
    - CryoEM-Logbook (W Chiu, in development)
- Map validation

How to evaluate a cryo-EM map ?

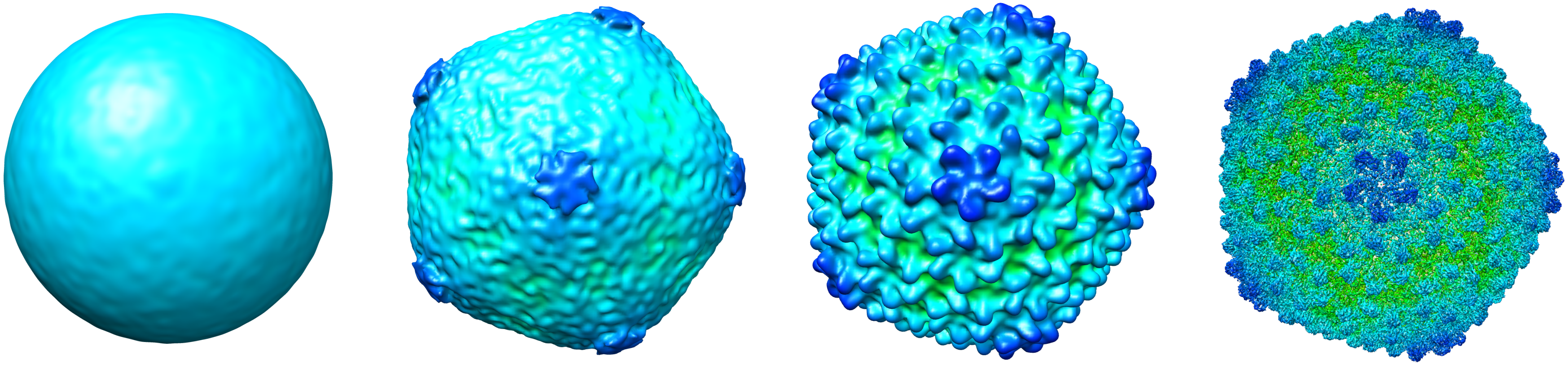
- Resolution
- Phase randomization of raw images at lower resolution
- Structure features
- ResMap
- Tilt Pair
- Use different data sets
- Look for structure signatures from EMDB

2 Independent Maps from 2 Data Subsets and 2 Initial Models



Subset 1, N=11,000

Map 1



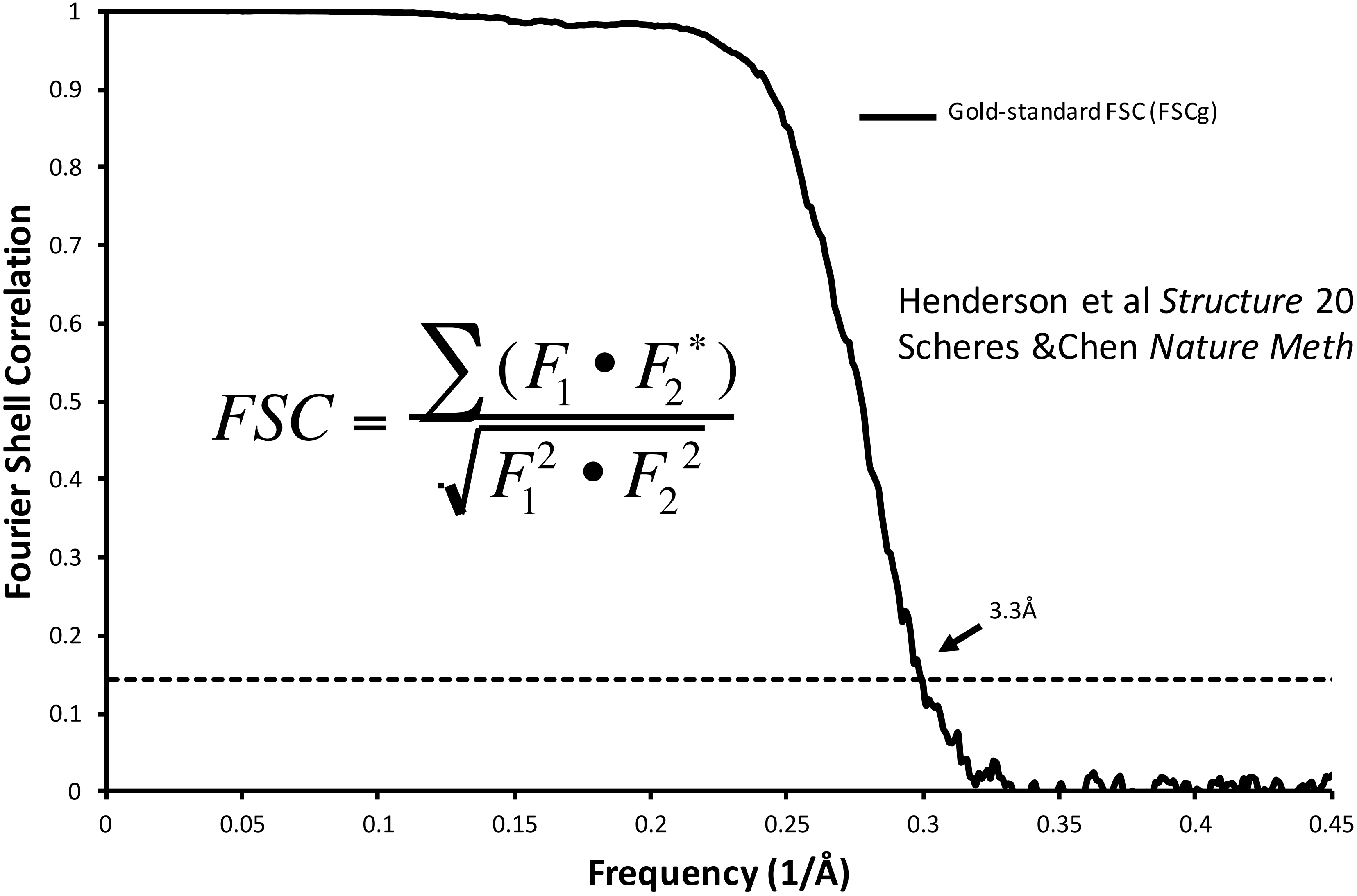
Subset 2, N=11,000

Map 2

Hryc, Chen et al ***PNAS*** 2017

EMAN2, MPSA, JSPR

Estimating the Map Resolution



$$FSC = \frac{\sum (F_1 \bullet F_2^{\ *})}{\sqrt{F_1^2 \bullet F_2^{\ 2}}}$$

Henderson et al *Structure* 2012
Scheres &Chen *Nature Meth* 2012

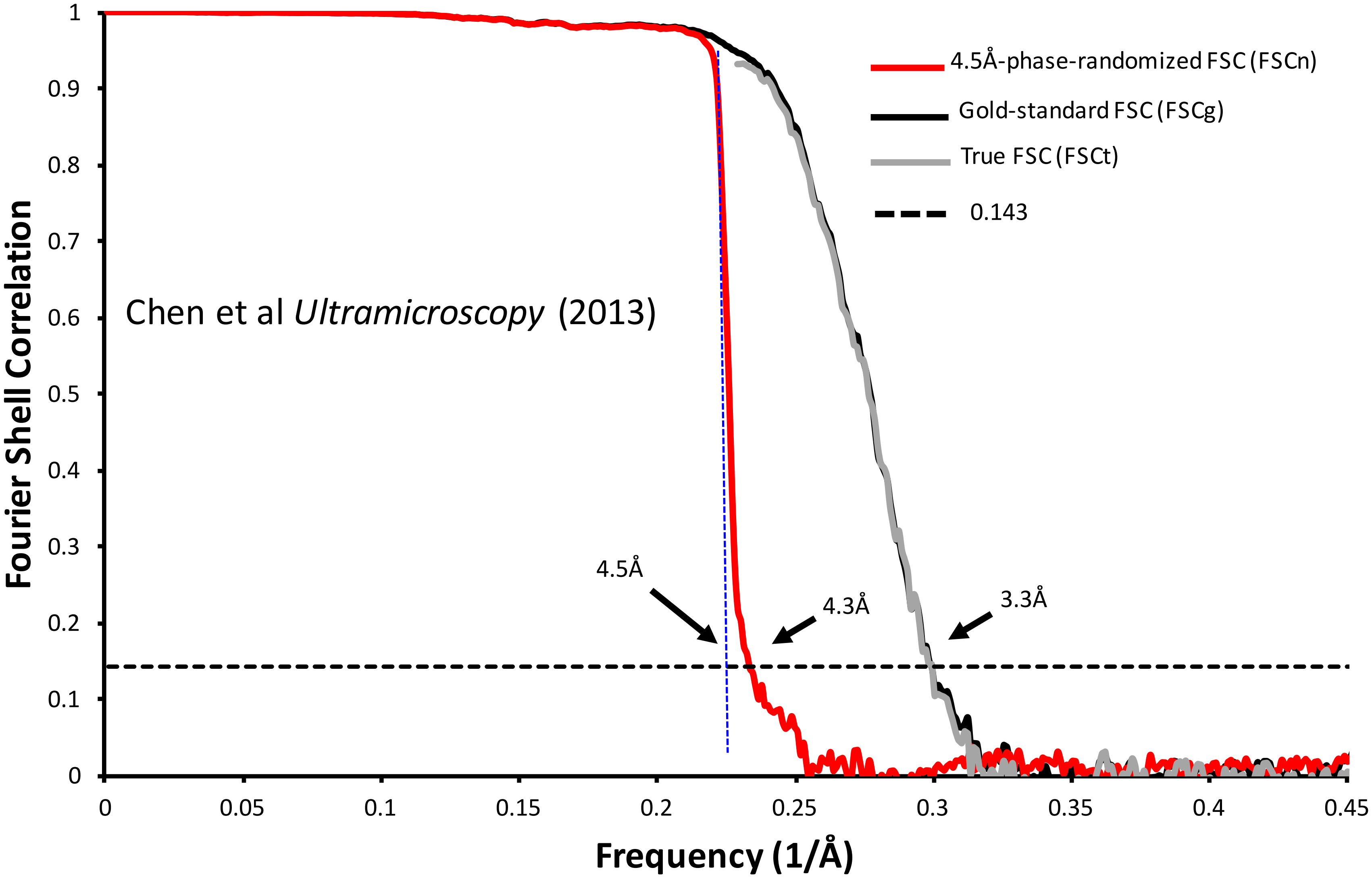
Hryc, Chen et al ***PNAS*** 2017

Validation of CryoEM Map

Assure No Over-Refinement: Randomize the phases of all the raw particle images for frequencies beyond 75% of the targeted resolution

-

Estimate the Map Resolution with 2 Maps



Hryc, Chen et al ***PNAS*** 2017

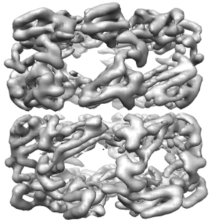
How to evaluate a cryo-EM map ?

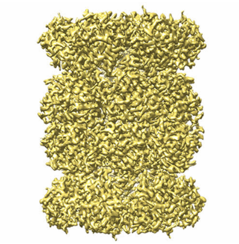
How to evaluate a cryoEM map ?

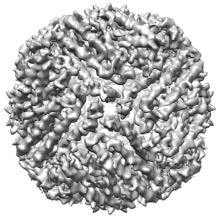
○ Resolution
○ Phase randomization of raw images at lower resolution
○ Structure features
○ ResMap
○ Tilt Pair
○ Use different software

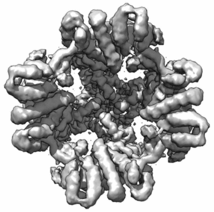
Benchmark Targets
http://challenges.emdatabank.org/

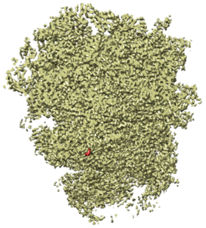
Map Challenge: Raw Images @ EMPIAR

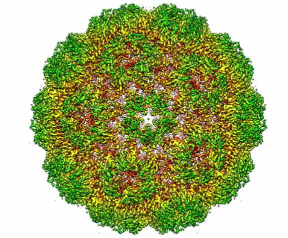

GroEL

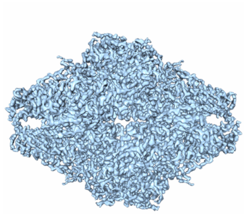

T20S
Proteasome


Apo-
Ferritin

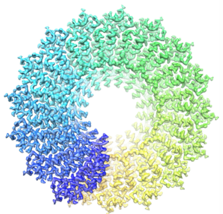

TrpV1
channel

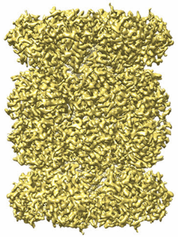

80S
Ribosome

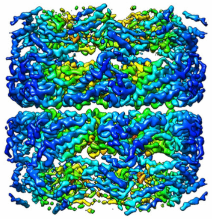

Brome
Mosaic Virus

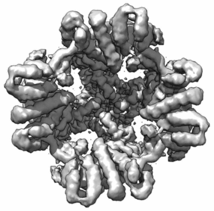

ß-galacto-
sidase

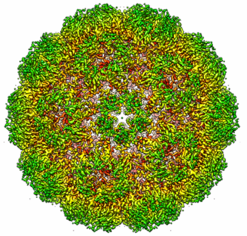
Model Challenge: Maps @ EMDB

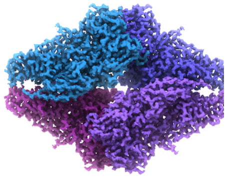

Tobacco
Mosaic Virus

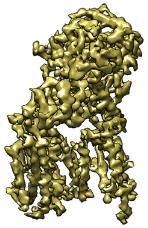

T20S
Proteasome

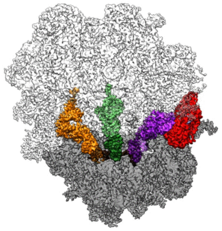

GroEL


TrpV1
channel

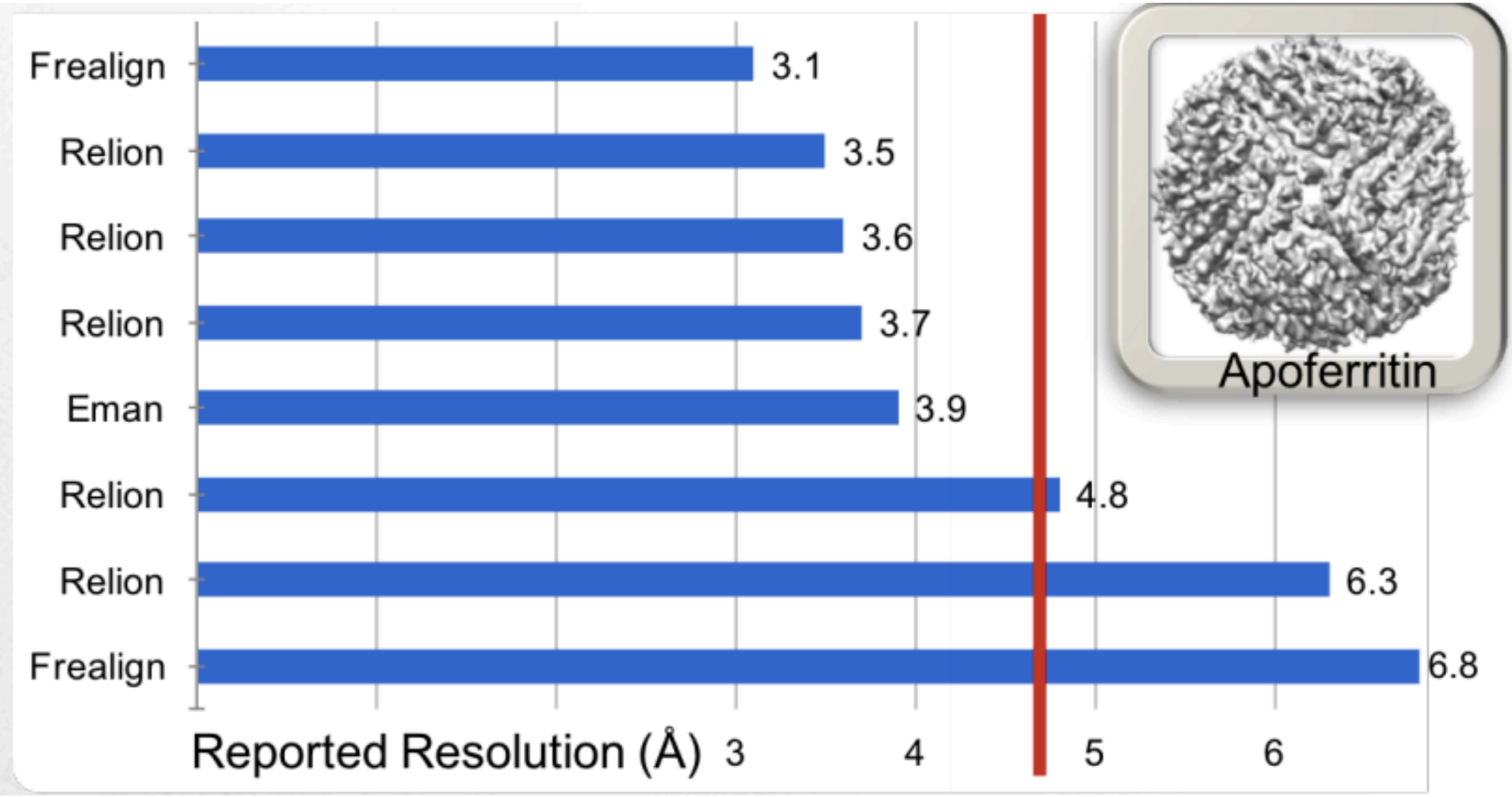

Brome
Mosaic Virus


ß-galacto-
sidase


γ-Secretase


70S
Ribosome

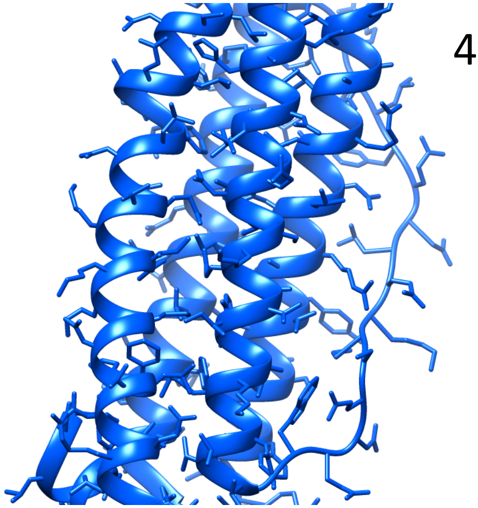
Map Challenge: Apoferritin



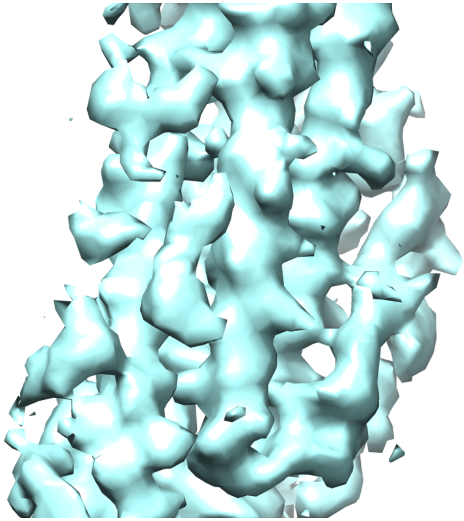
C. Lawson

Map Challenge: Apoferritin

Russo
& Passmore
(2014)



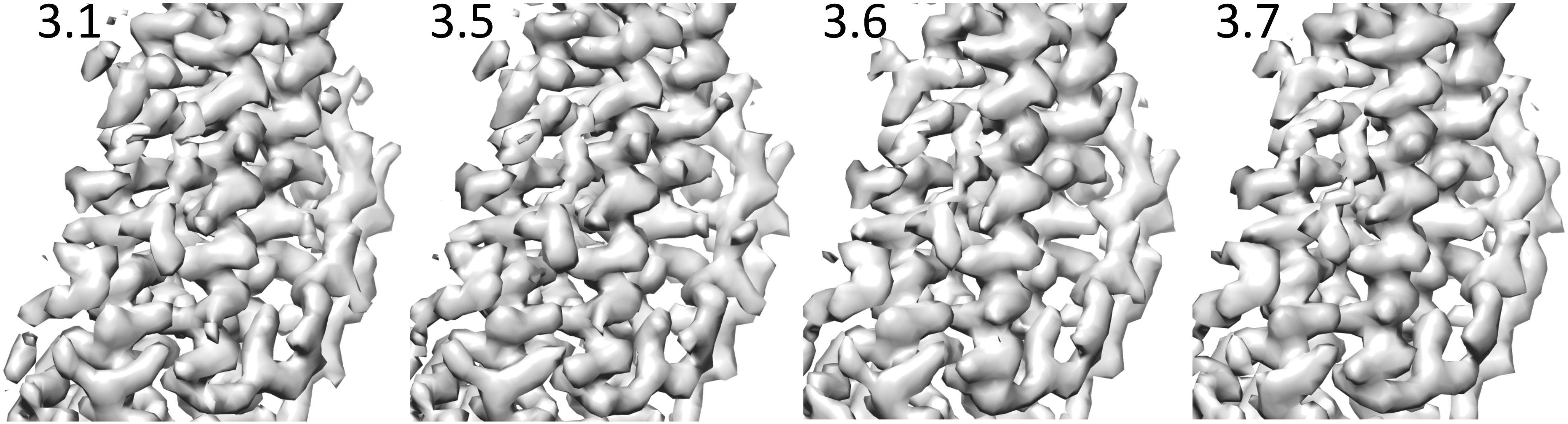
4v1w



EMD-2788

Map Challenge maps with reported resolution (Å)

C Lawson

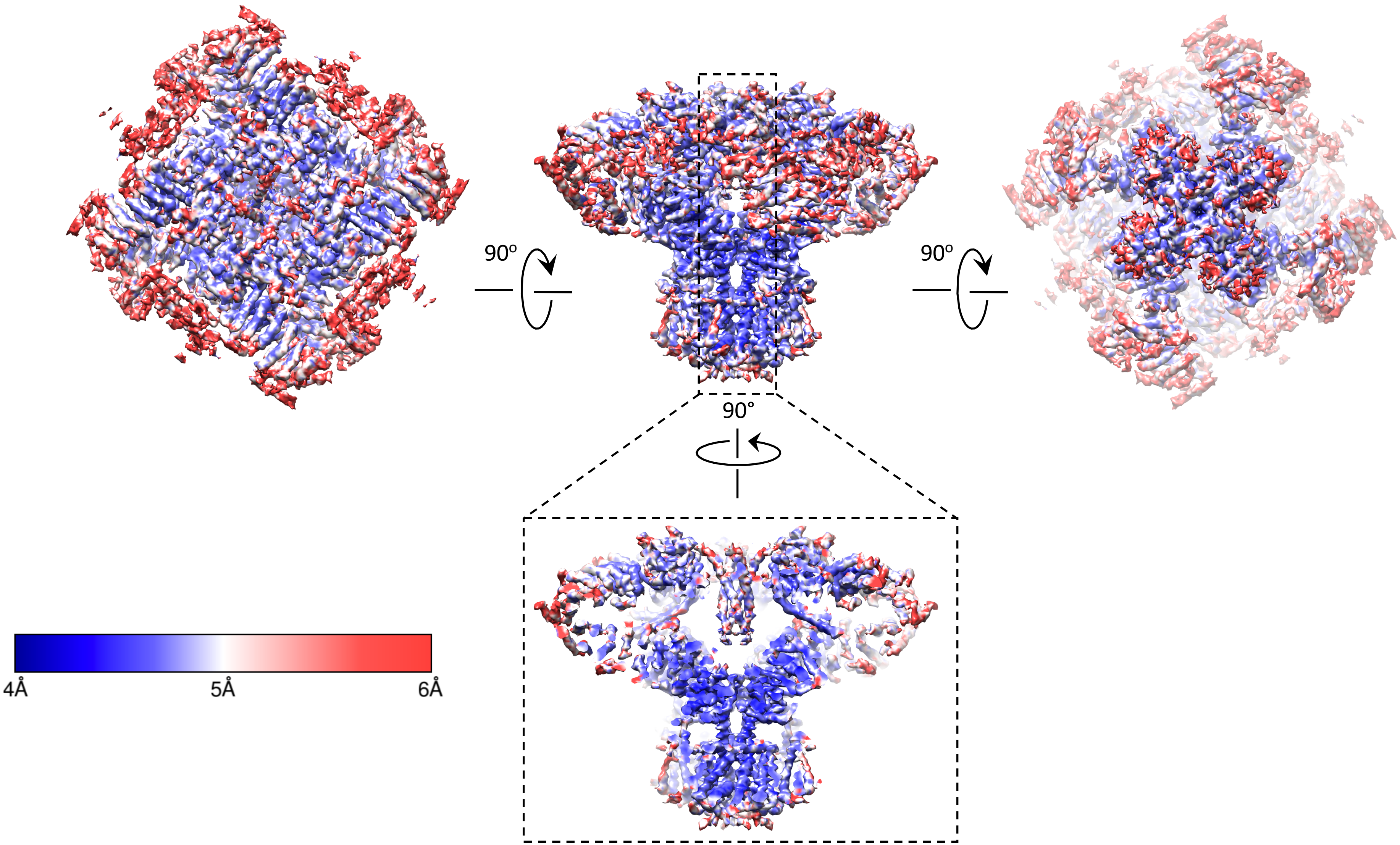


How to evaluate a cryo-EM map ?

How to evaluate a cryoEM map ?

- Resolution
- Phase randomization of raw images at lower resolution
- Structure features
- ResMap
- Tilt Pair
- Use different software

IP$_3$R1 CryoEM Map Resolution Variation



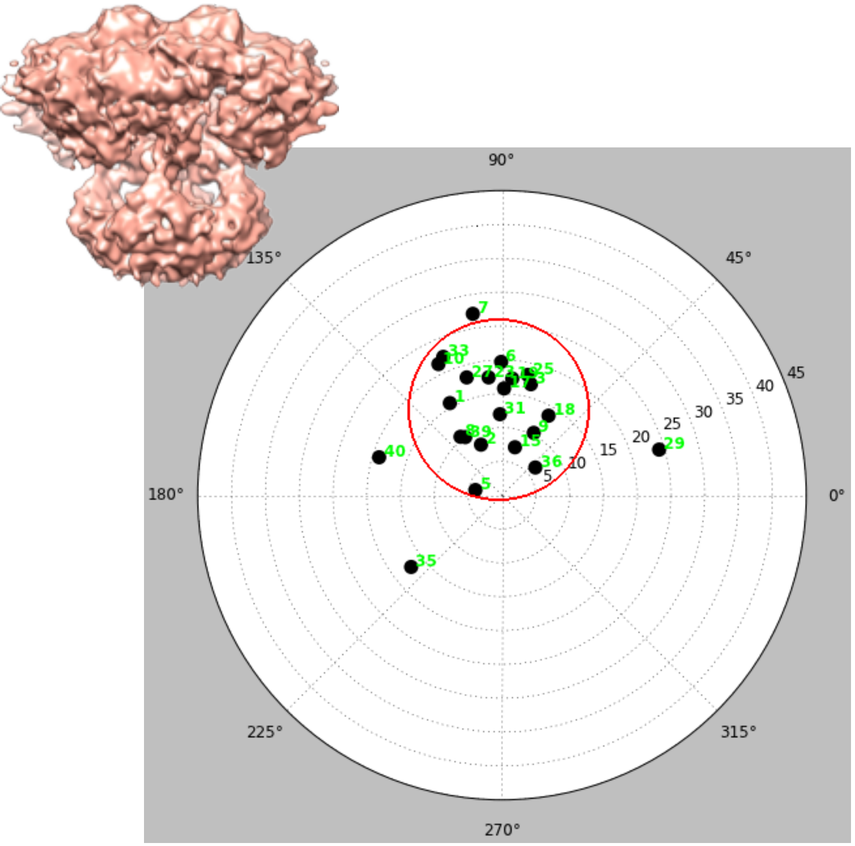
G Fan et al, *Nature* 2015

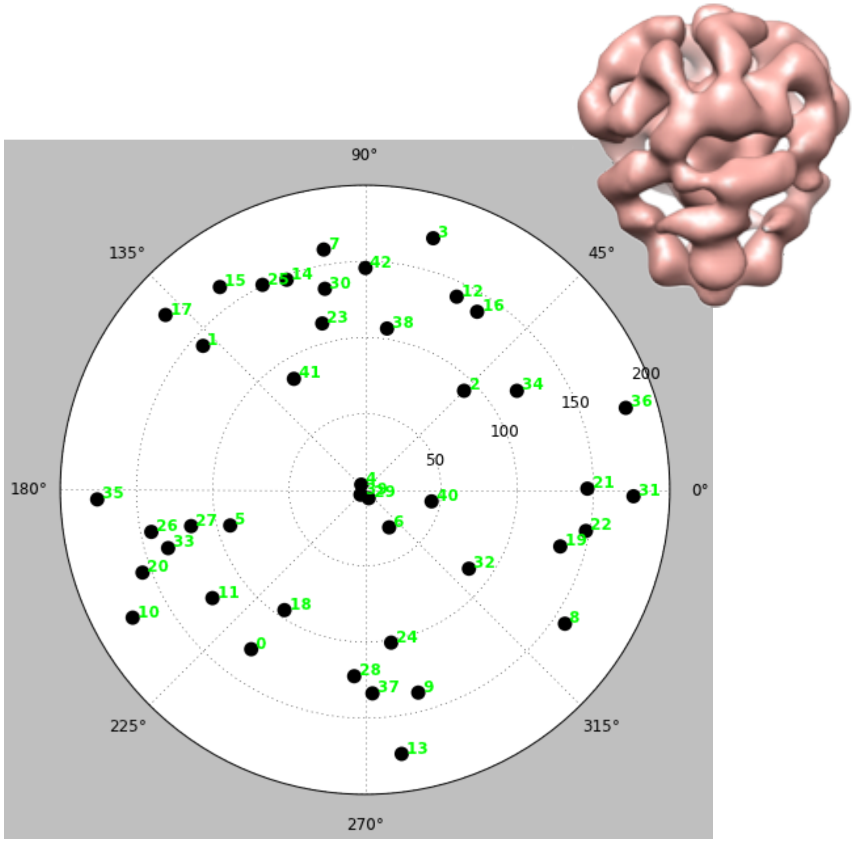
How to evaluate a cryo-EM map ?

How to evaluate a cryoEM map ?

- Resolution
- Phase randomization of raw images at lower resolution
- Structure features
- ResMap
- Tilt Pair
- Use different software

Tilt Pair Validation





Murray, Serysheva et al *Structure* 2013

➡ P. Rosenthal & R. Henderson *JMB (2003)*

5. Opportunities & Challenges in Modeling

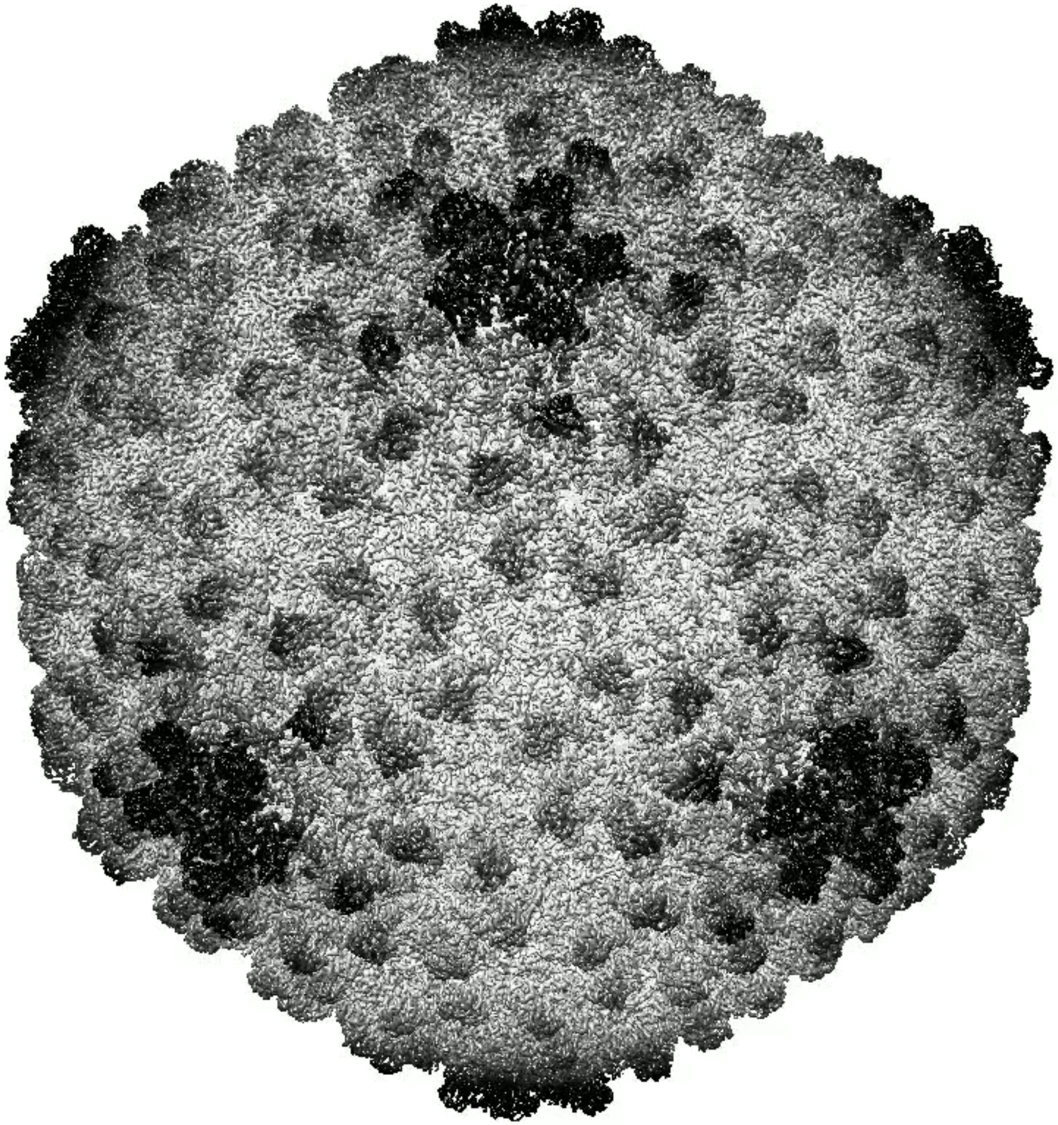
- What are the best practices for simple and complicated structures?
- How to validate models?
- Does it make any difference what software to use ?

DiMaio and Chiu *Methods in Enzymology* 2016

5. Opportunities & Challenges in Modeling

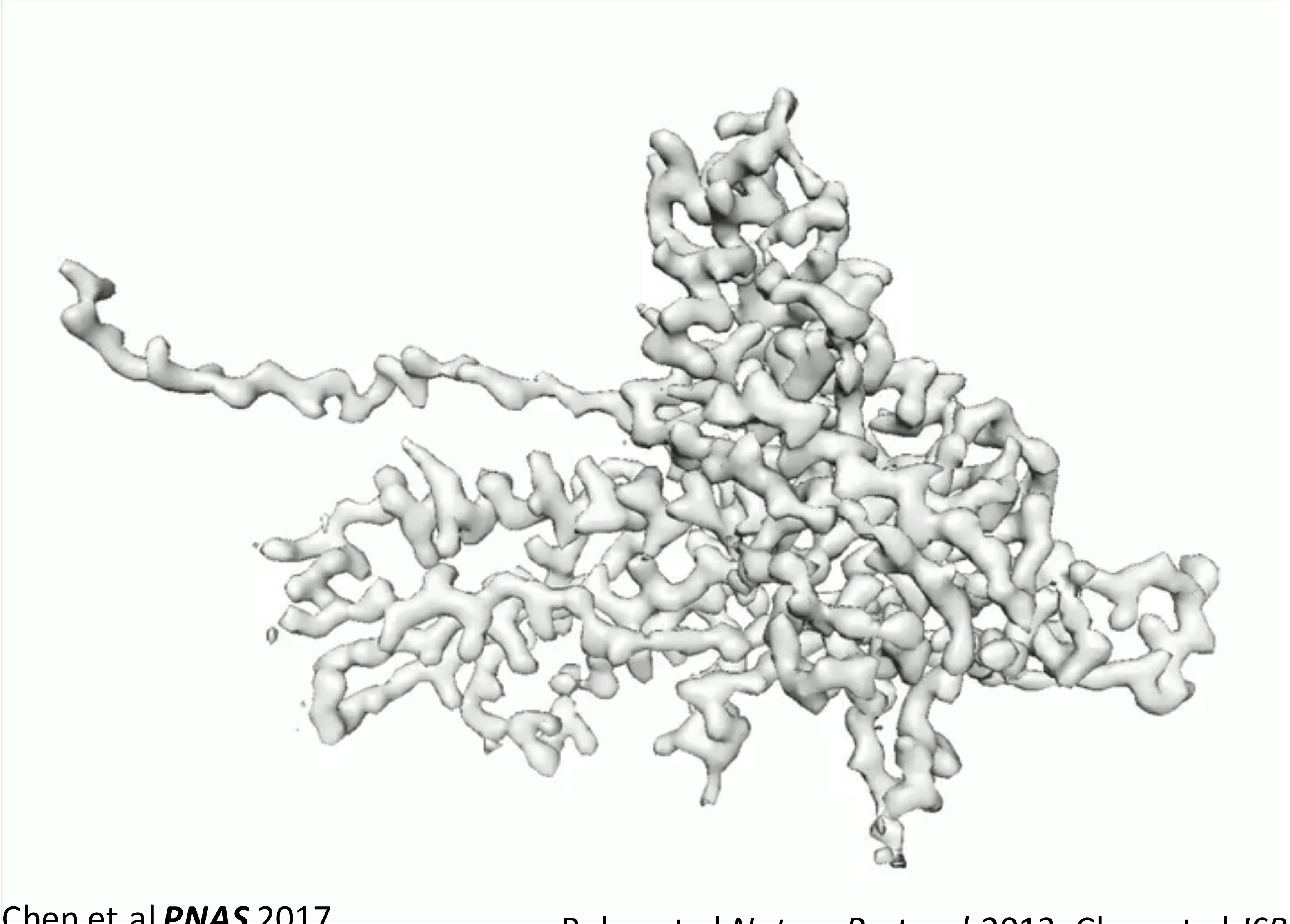
- What are the best practices for simple and complicated structures?
- How to validate models?
- Does it make any difference what software to use ?

3.3Å CryoEM Map of P22 Bacteriophage



Hryc, Chen et al *PNAS* 2017

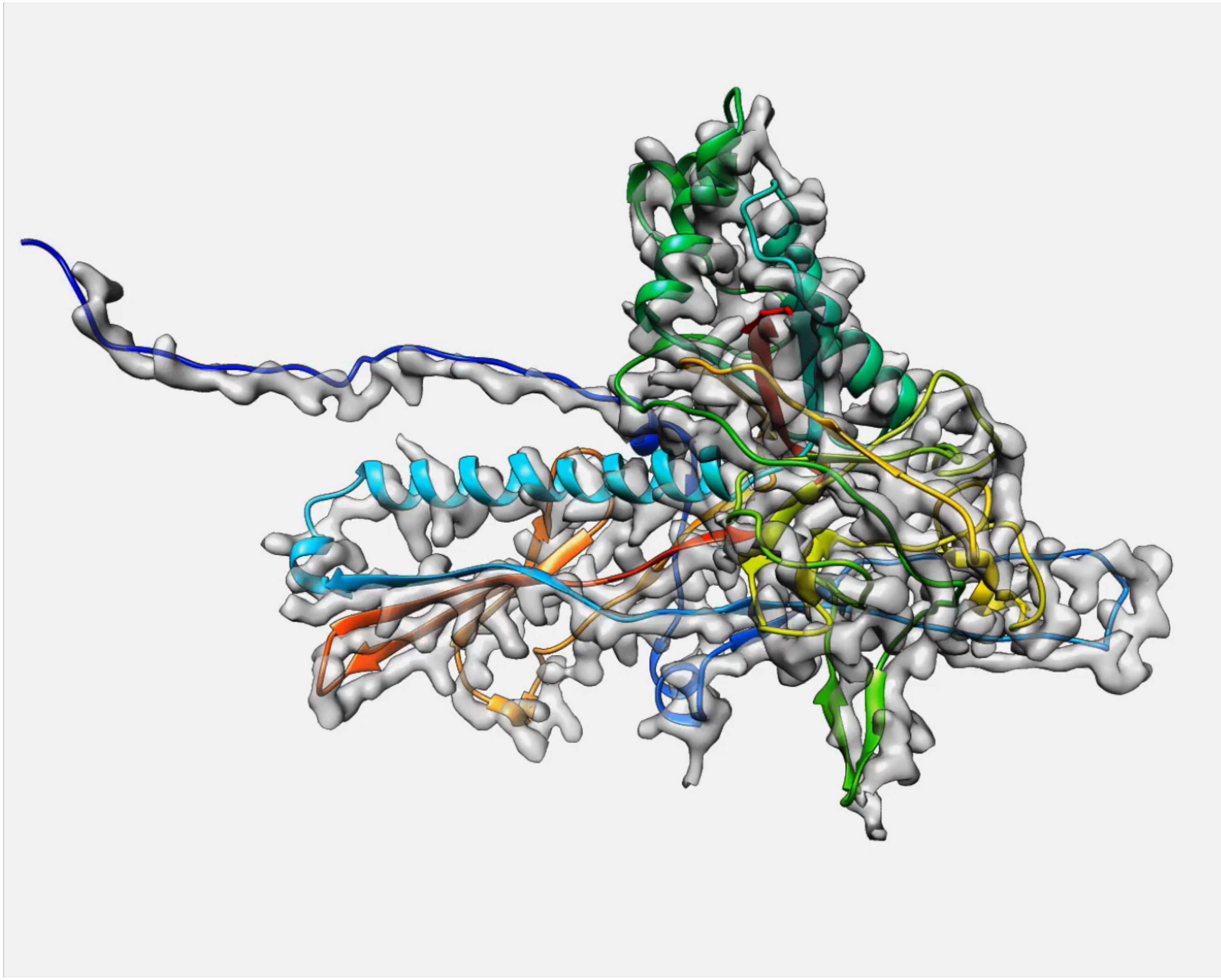
De Novo Model of Single Protein Subunit of P22 Bacteriophage



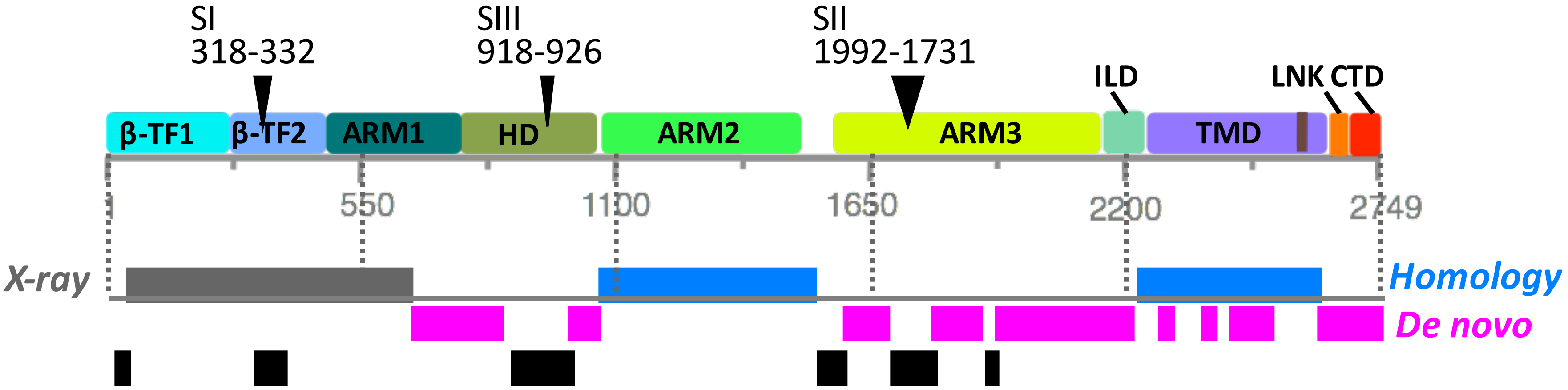
Hryc, Chen et al ***PNAS*** 2017

Baker et al *Nature Protocol* 2013; Chen et al *JSB* 2016

Full Atom Modeling
(Model Optimization/Refinement)

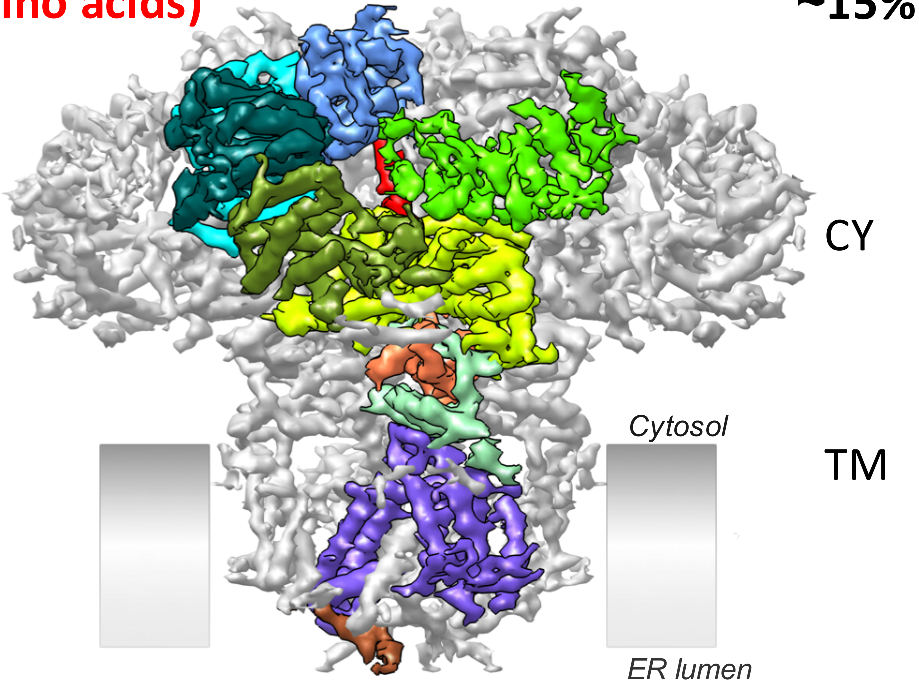


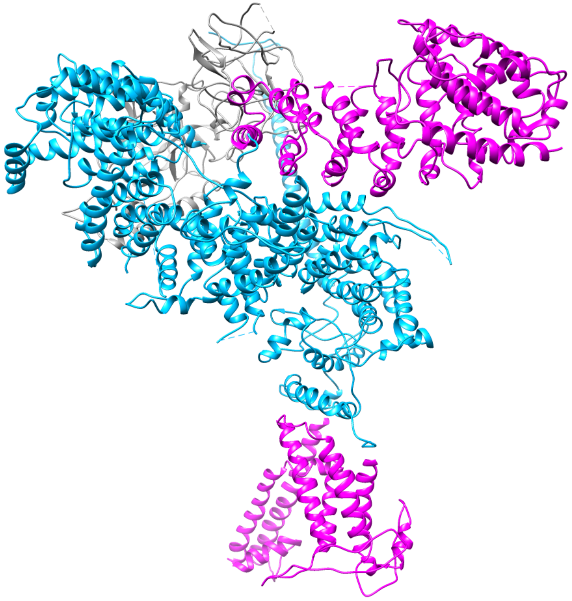
Building a Model for IP$_3$R1 Protein Subunit



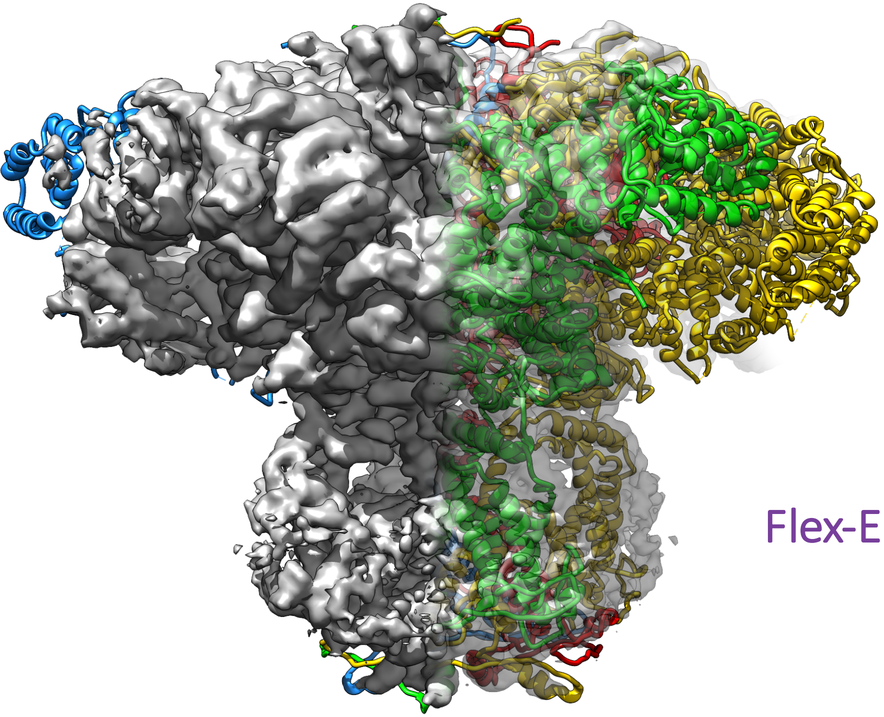
~85% of the backbone topology (2,327 of 2,749 amino acids)

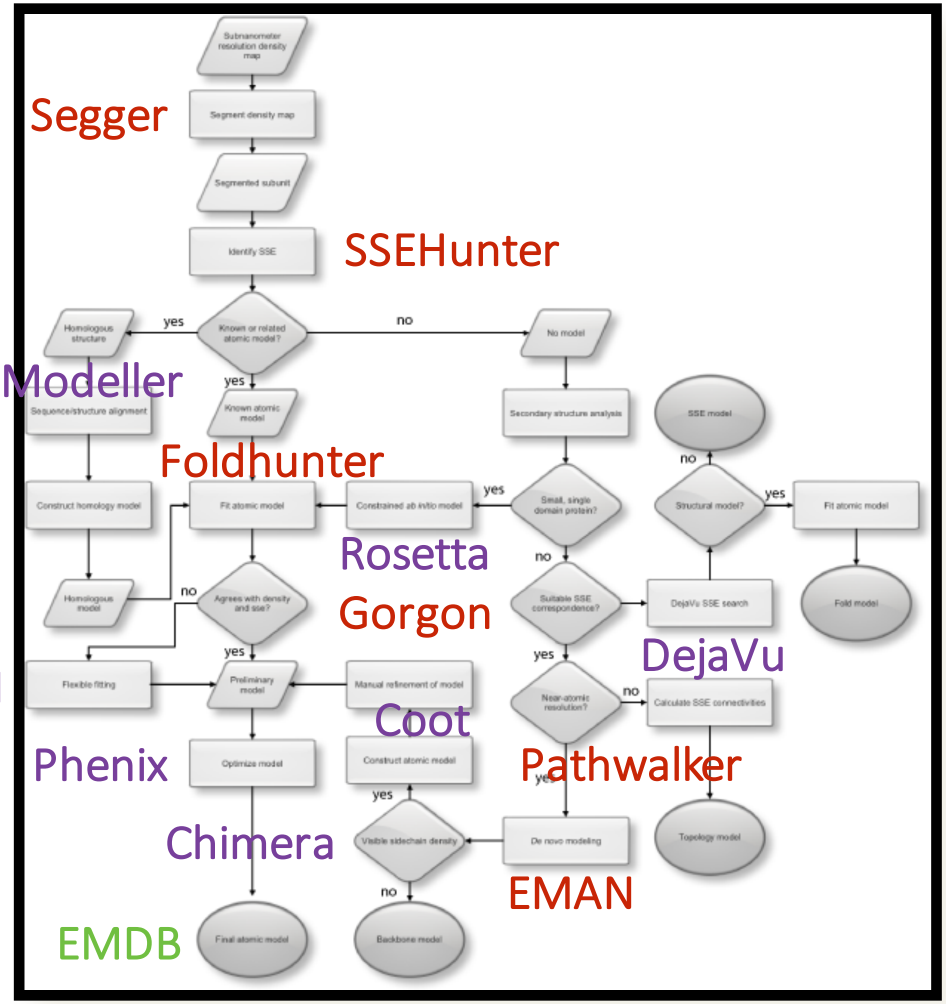
Un-modeled: ~15% (422 residues)





Modeling Pathway





Segger

SSEHunter

Modeller

Foldhunter

Rosetta

Gorgon

DejaVu

Flex-EM

Coot

Phenix

Pathwalker

Chimera

EMAN

EMDB

Matthew Baker

5. Opportunities & Challenges in Modeling

- What are the best practices for simple and complicated structures?
- How to validate models?
- Does it make any difference what software to use ?

GroEL Cryo-EM Map Derived Model Statistics

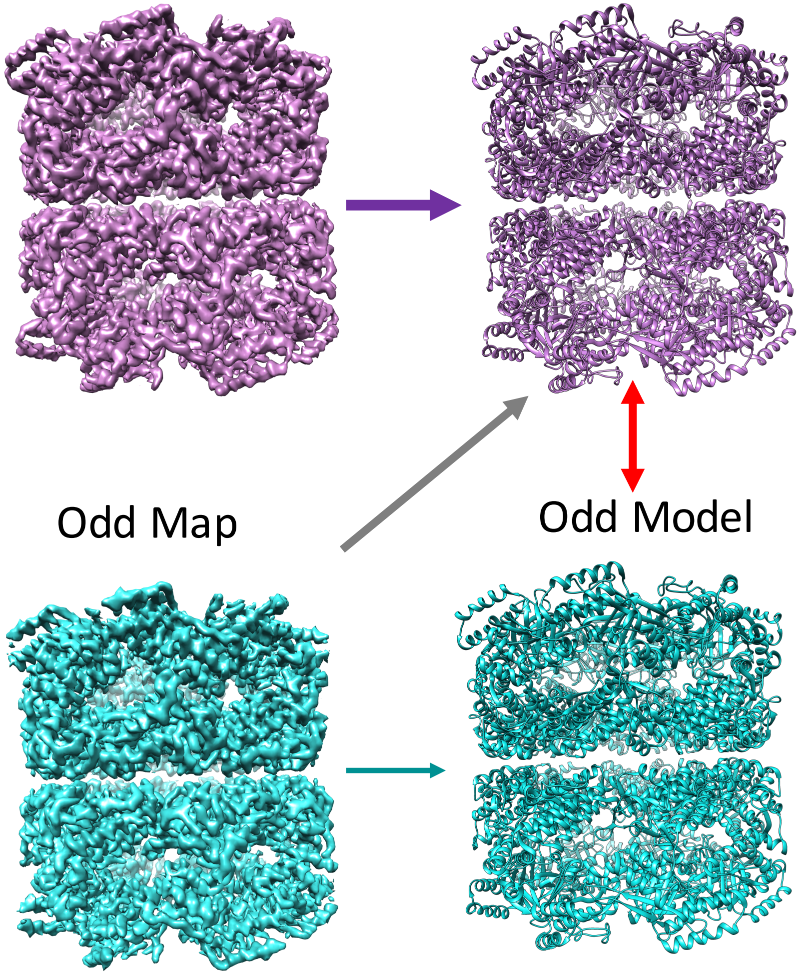
All-Atom Contacts	Clashscore, all atoms:	12.27	97th percentile* (N=37, 3Å - 9999Å)

Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

Protein Geometry			
Poor rotamers	0	0.00%	Goal: <0.3%
Favored rotamers	5600	99.01%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	7168	98.08%	Goal: >98%
MolProbity score^	1.6		100th percentile* (N=342, 3.25Å - 3.95Å)
Cβ deviations >0.25Å	0	0.00%	Goal: 0
Bad bonds:	0 / 54306	0.00%	Goal: 0%
Bad angles:	0 / 73346	0.00%	Goal: <0.1%
Peptide Omegas			
Cis Prolines:	0 / 196	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria			
Ca BLAM outliers	56	0.77%	Goal: <1.0%
Cα Geometry outliers	28	0.38%	Goal: <0.5%
Map / Model Fit			
Cross-correlation	0.926	**EmRinger Score**	1.438

Roh et al ***PNAS*** 2017

Even / Odd Maps and Models

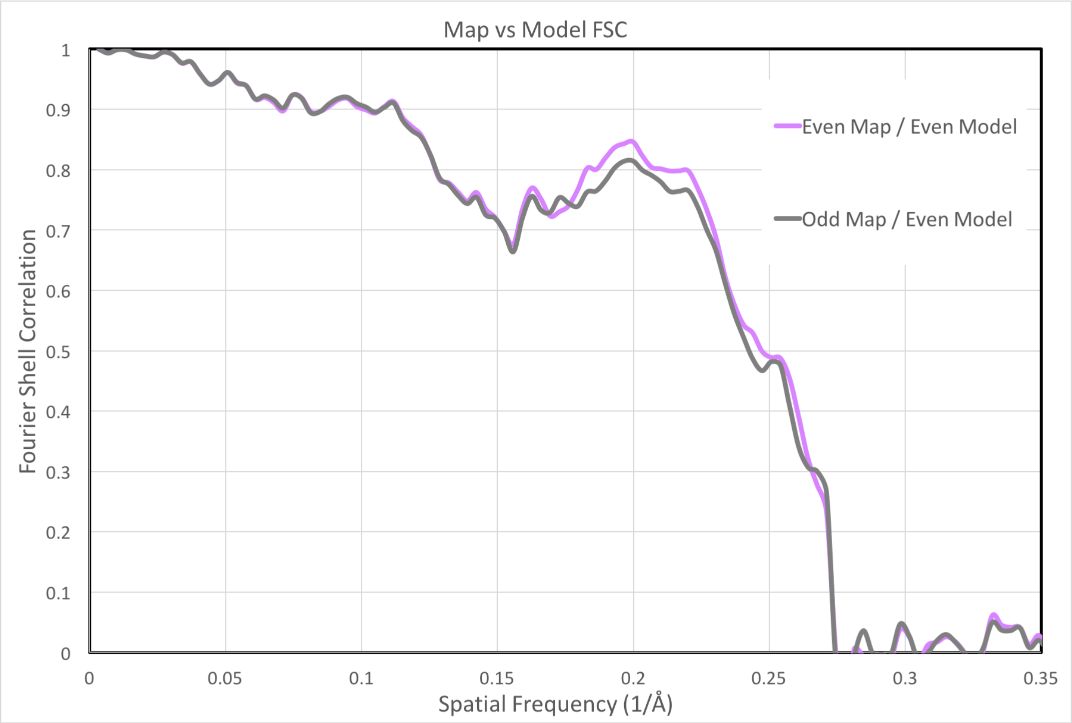


Even Map

Even Model

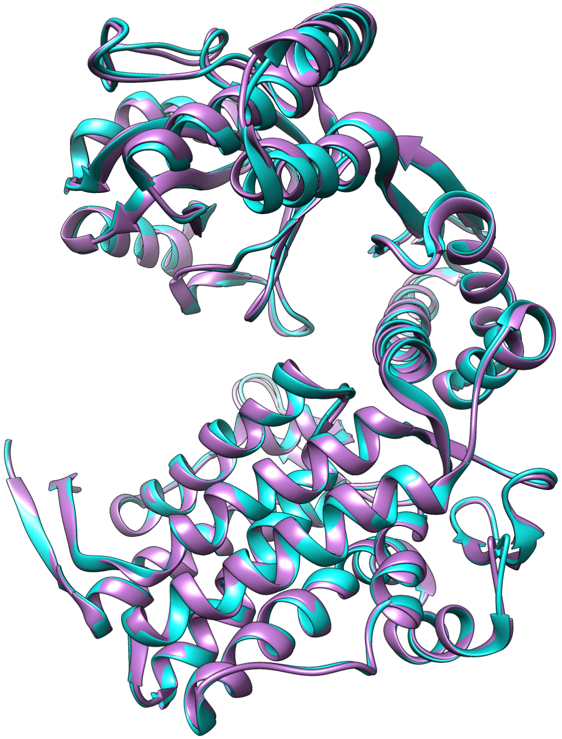
Odd Map

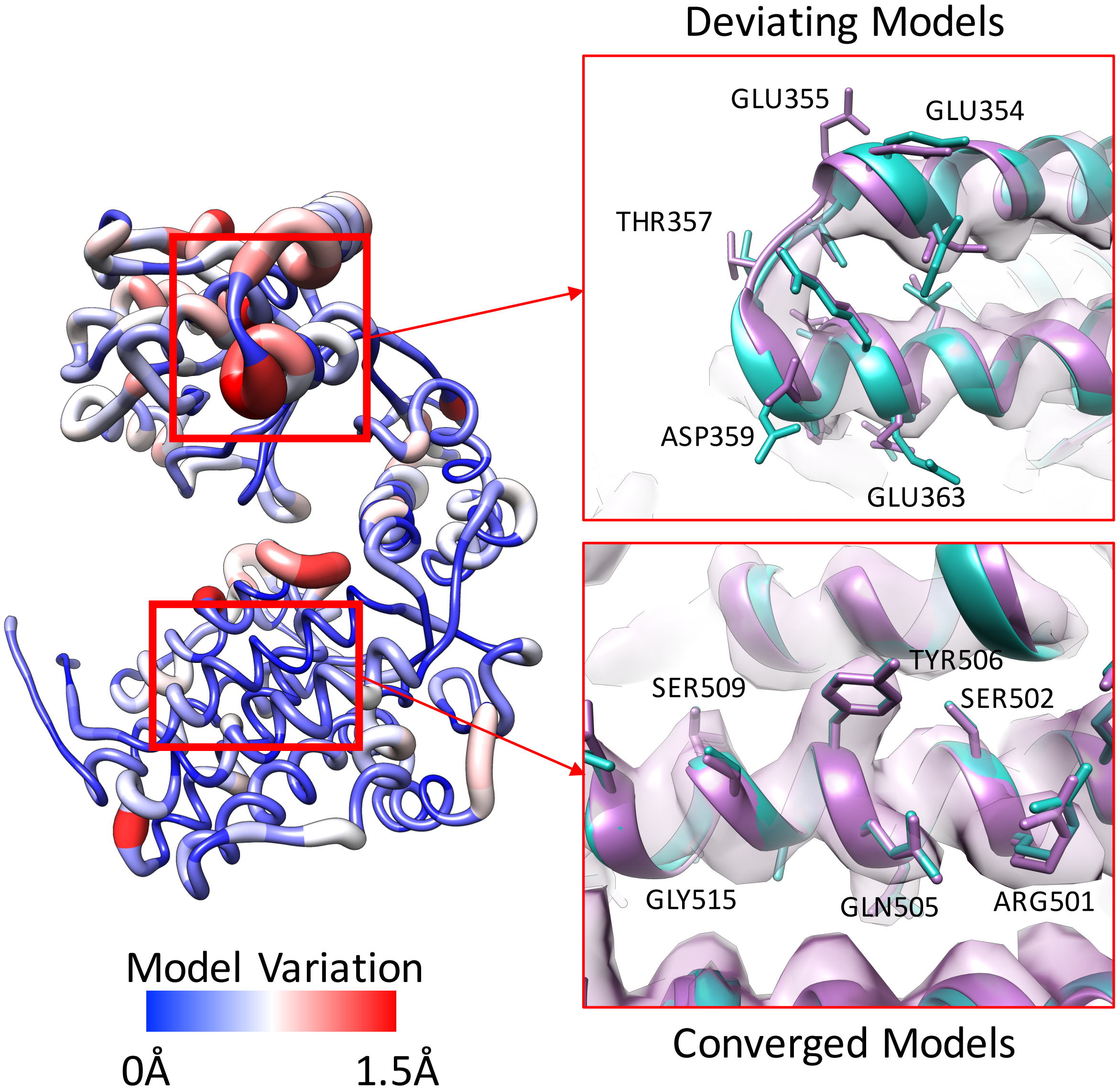
Odd Model



Roh et al ***PNAS*** 2017

Even / Odd Optimized Models



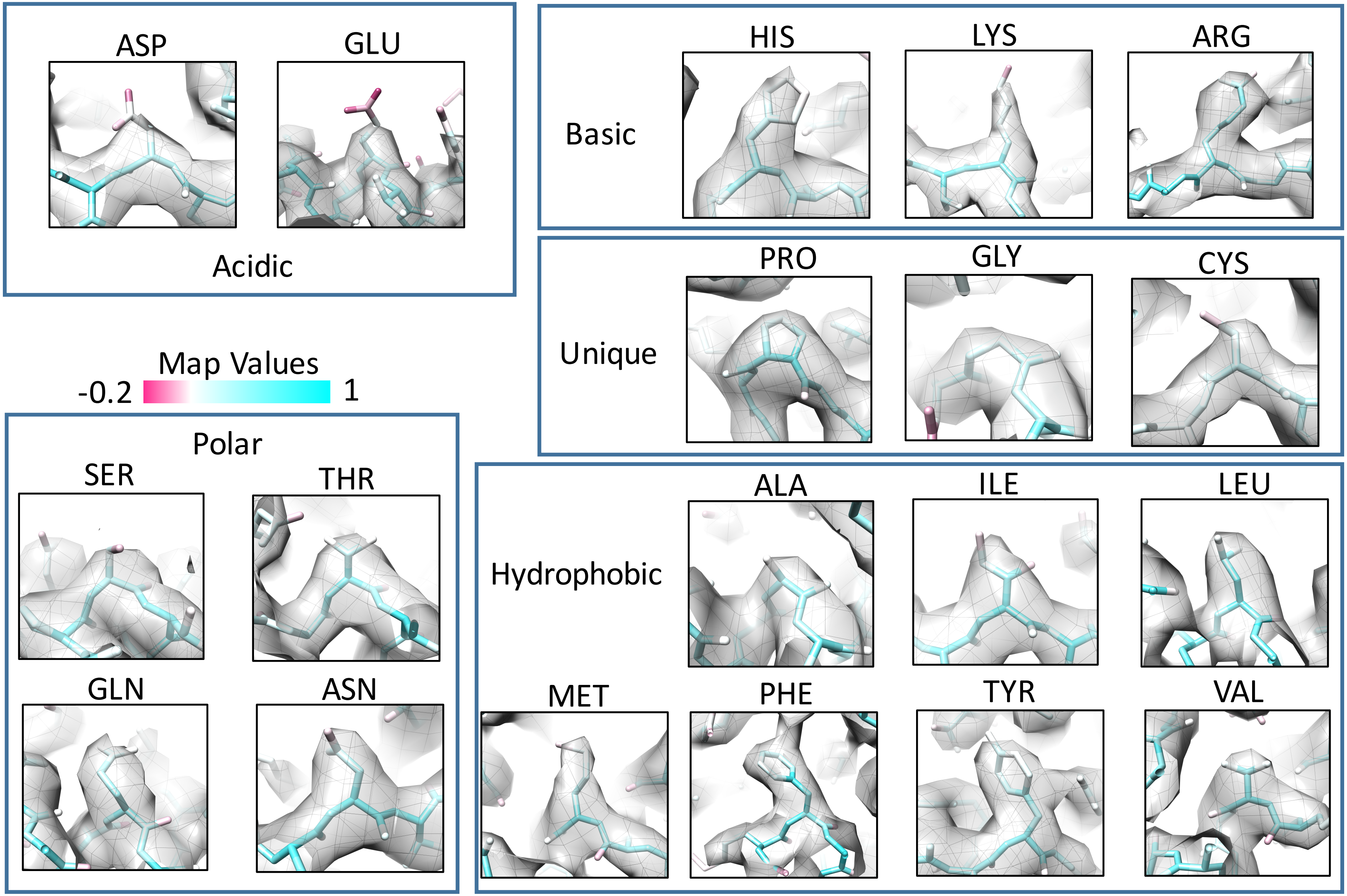


Roh et al ***PNAS*** 2017

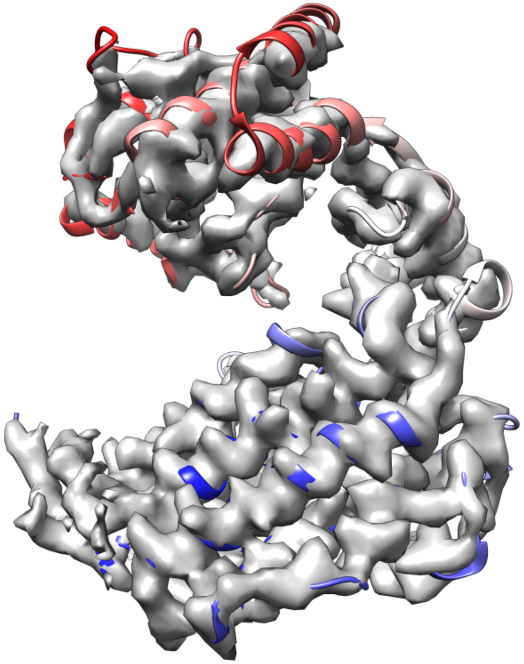
6. Structure Annotation and Archival

- Map and model have to match
- Map and model are deposited to PDB and EMDB
- Should raw data be deposited ?

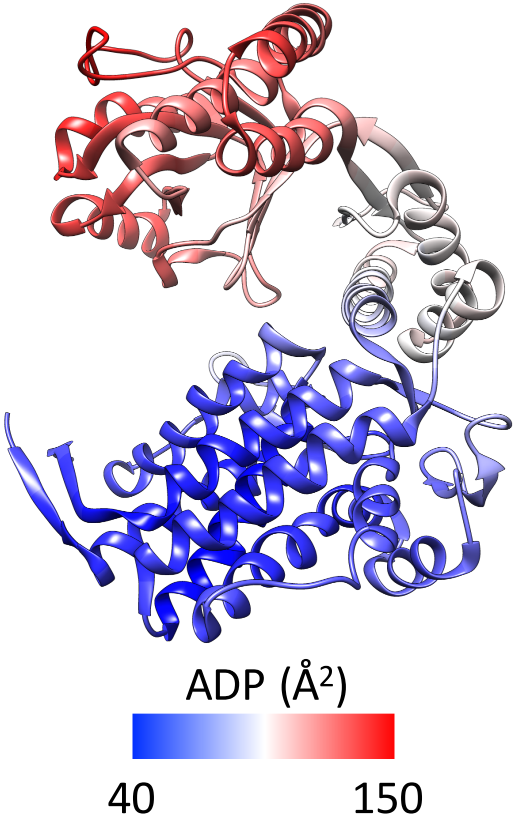
Gallery of Amino Acids of 3.5 Å GroEL Map

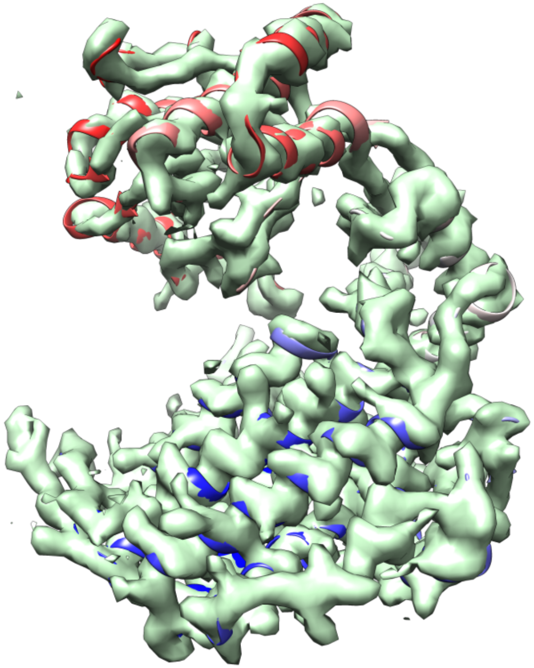


Annotate Model with Atomic Displacement Parameter (ADP)



Experimental Density
+ Model

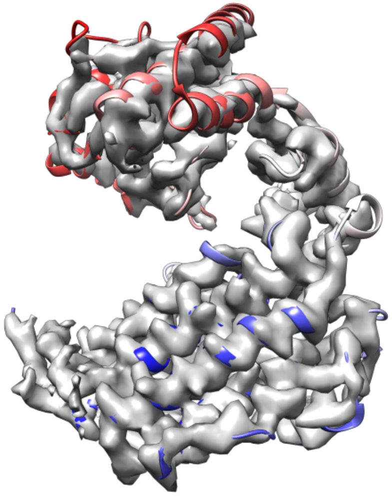




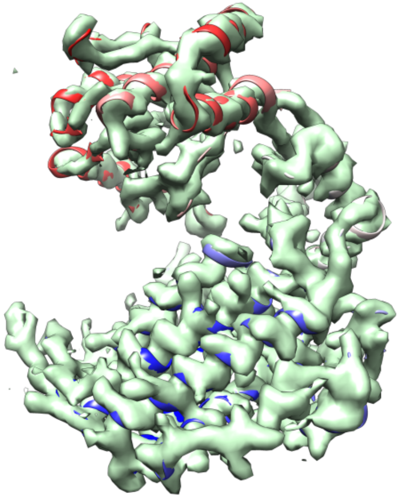
Weighted Calc. Map

Roh et al ***PNAS*** 2017

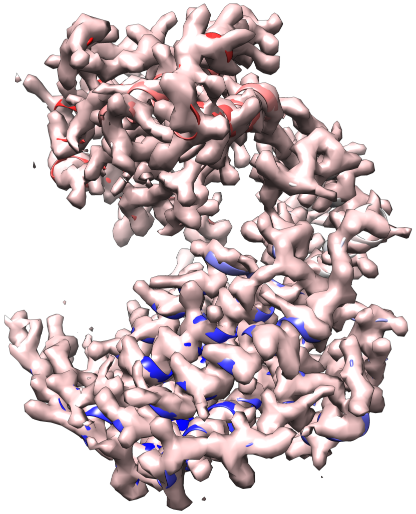
Properly Weighted Calculated Maps



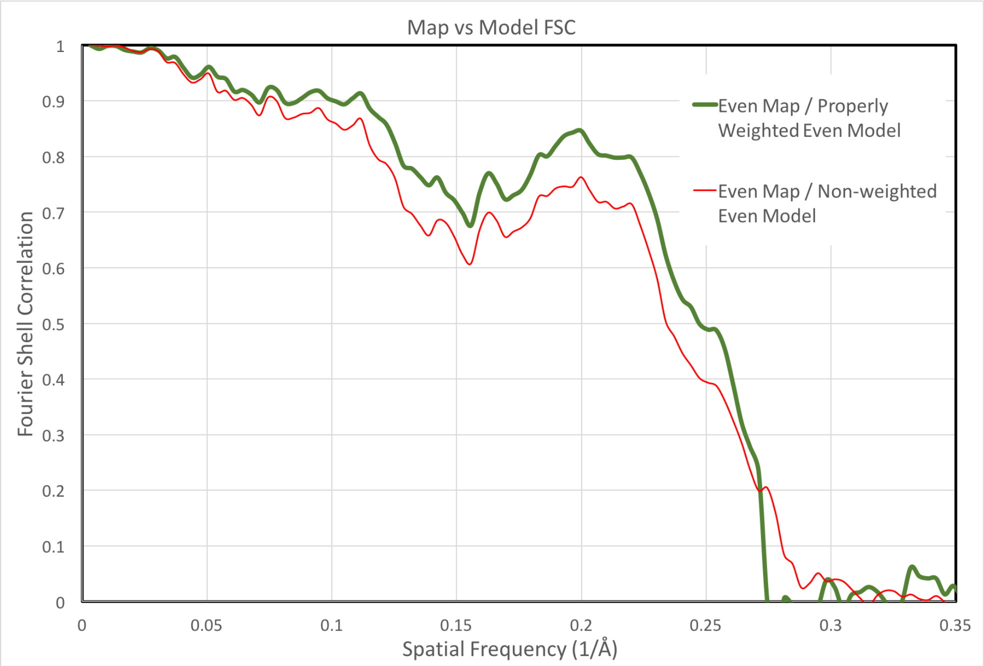
Experimental Density
+ Model



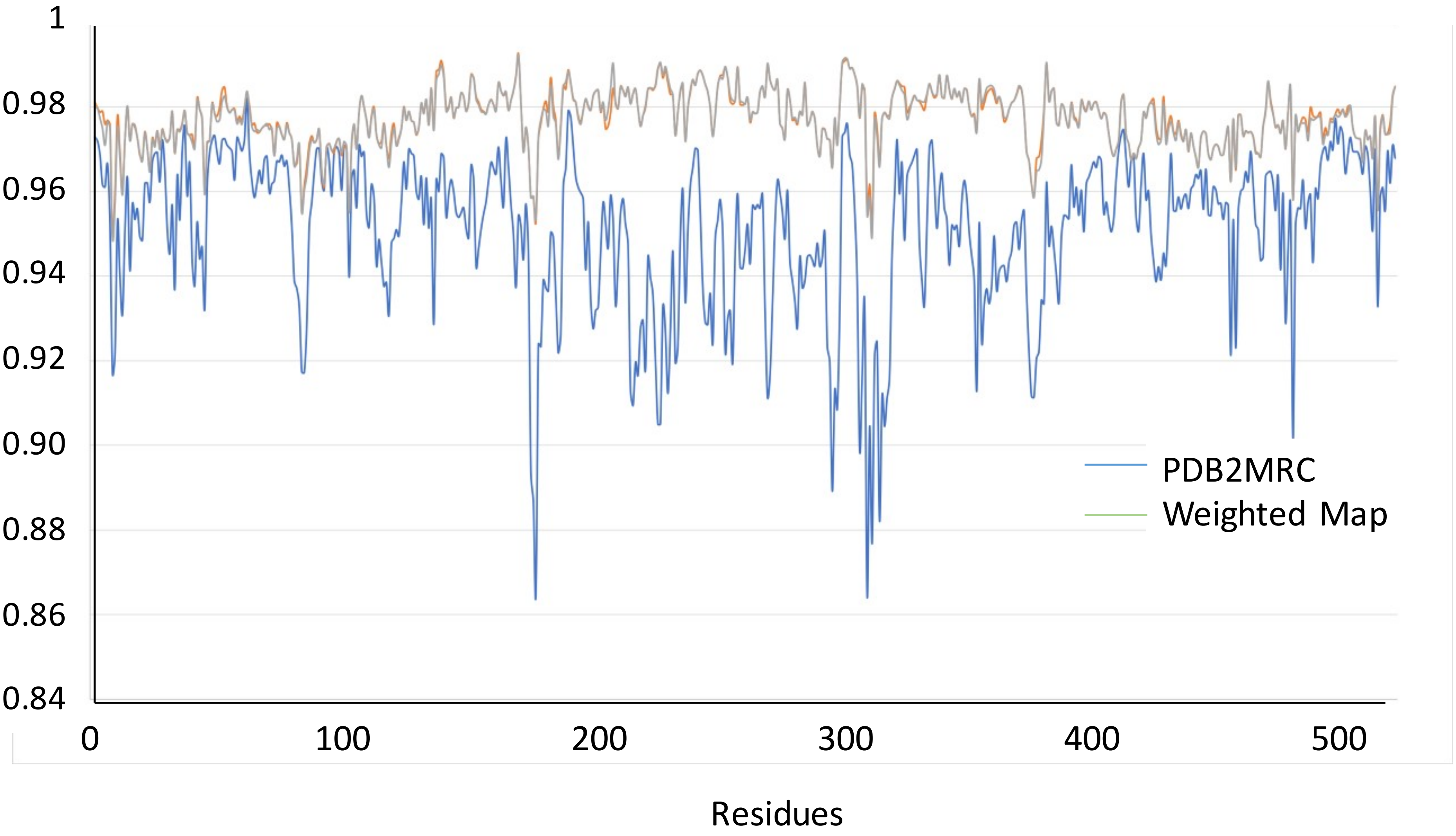
Weighted Calc. Map



Non-weighted Calc. Map



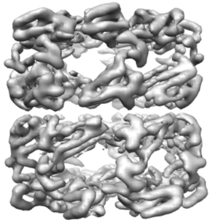
Map vs Model: Cross Correlation Per Residue

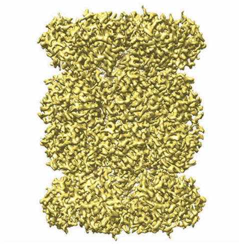


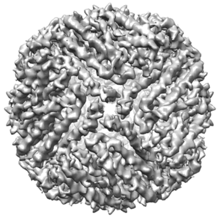
Roh et al ***PNAS*** 2017

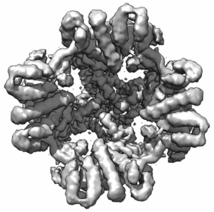
Benchmark Targets
http://challenges.emdatabank.org/

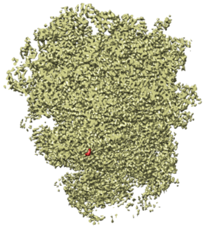
Map Challenge: Raw Images @ EMPIAR

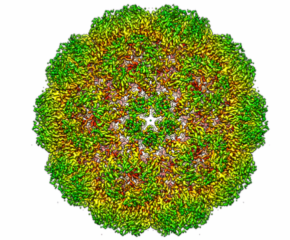

GroEL

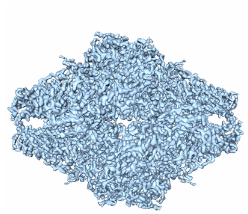

T20S Proteasome


Apo-Ferritin

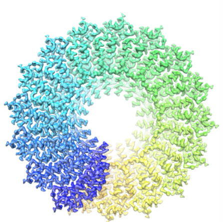

TrpV1 channel

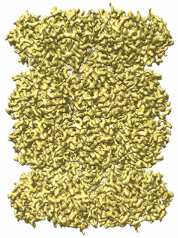

80S Ribosome

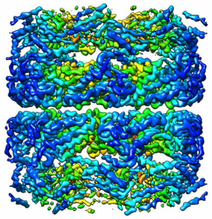

Brome Mosaic Virus

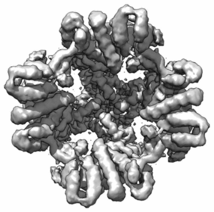

ß-galactosidase

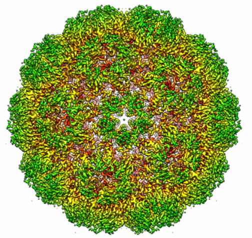
Model Challenge: Maps @ EMDB

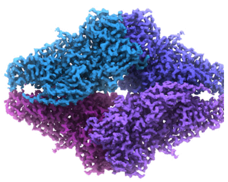

Tobacco Mosaic Virus

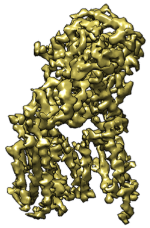

T20S Proteasome

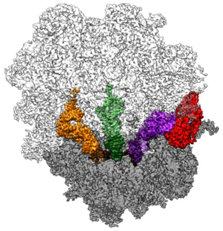

GroEL


TrpV1 channel


Brome Mosaic Virus


ß-galactosidase


γ-Secretase


70S Ribosome

Useful Reading

Glaeser, RM, Downing, KH, DeRosier, D, Chiu, W, & Frank, J (2007) *Electron crystallography of biological macromolecules* (Oxford University Press, Oxford)