EMAN Workshop, 2008

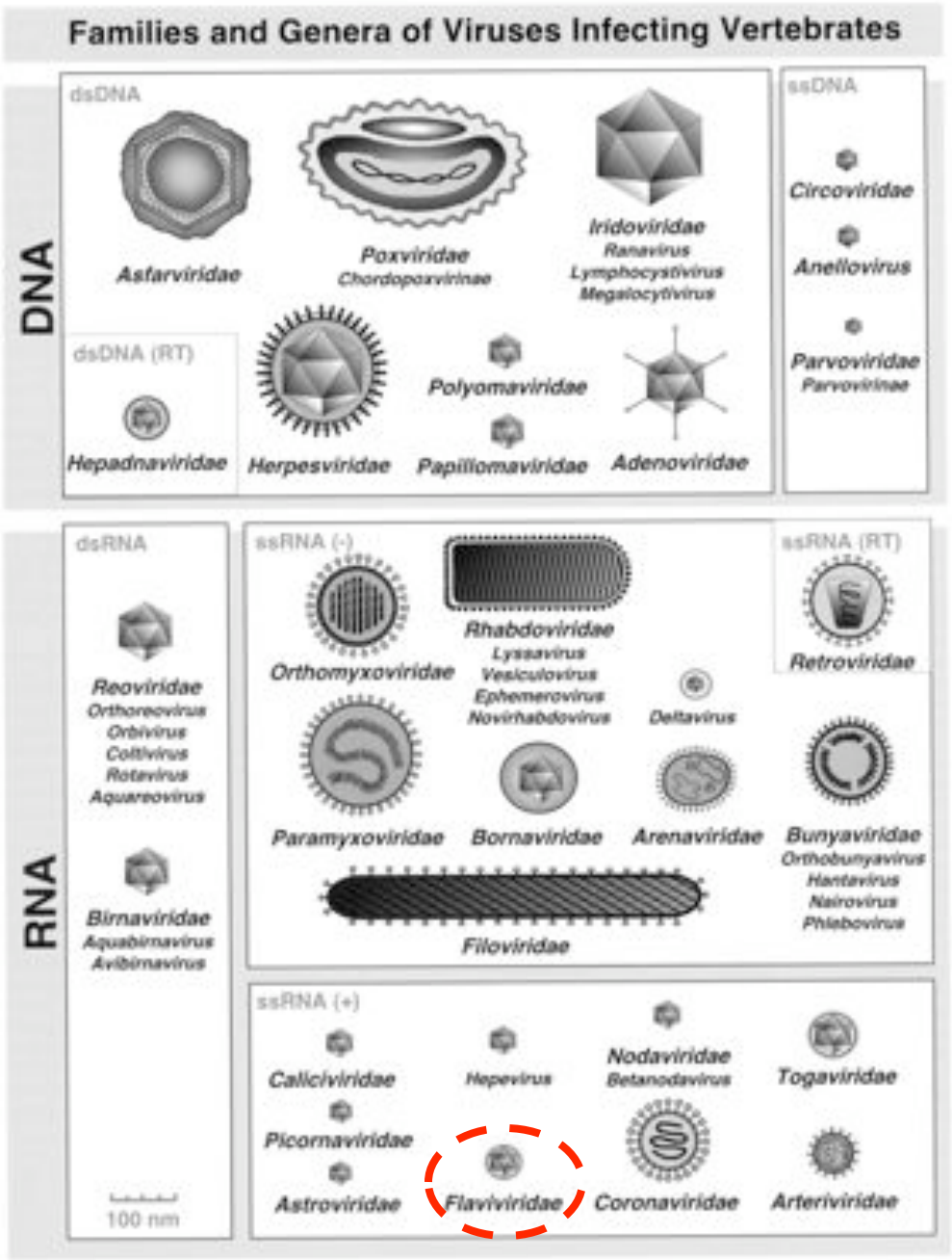
Near-atomic resolution 3-D reconstruction of viruses

Wen Jiang

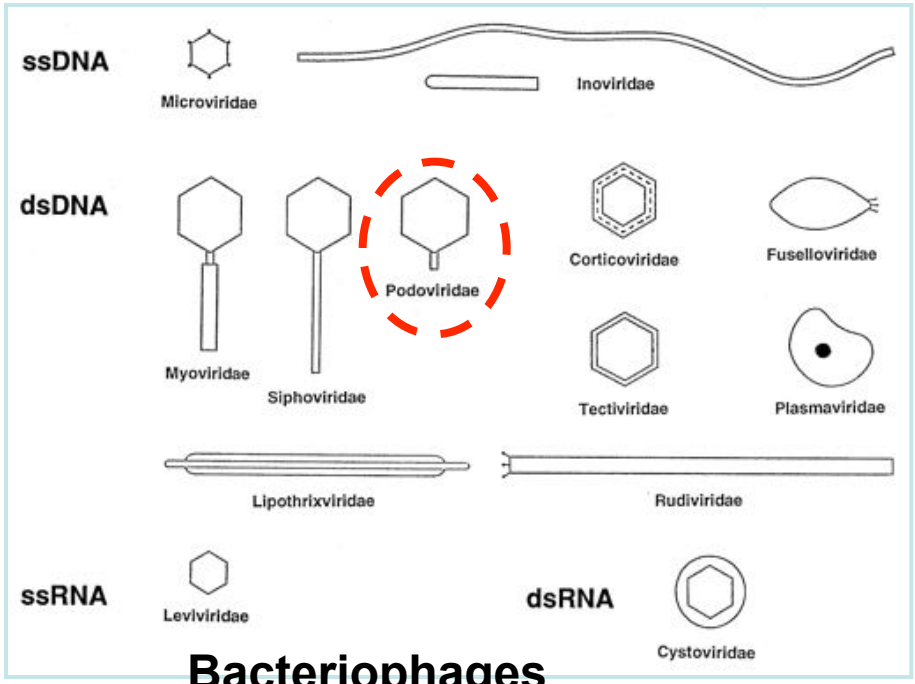
Markey Center for Structural Biology

PURDUE UNIVERSITY

Viruses



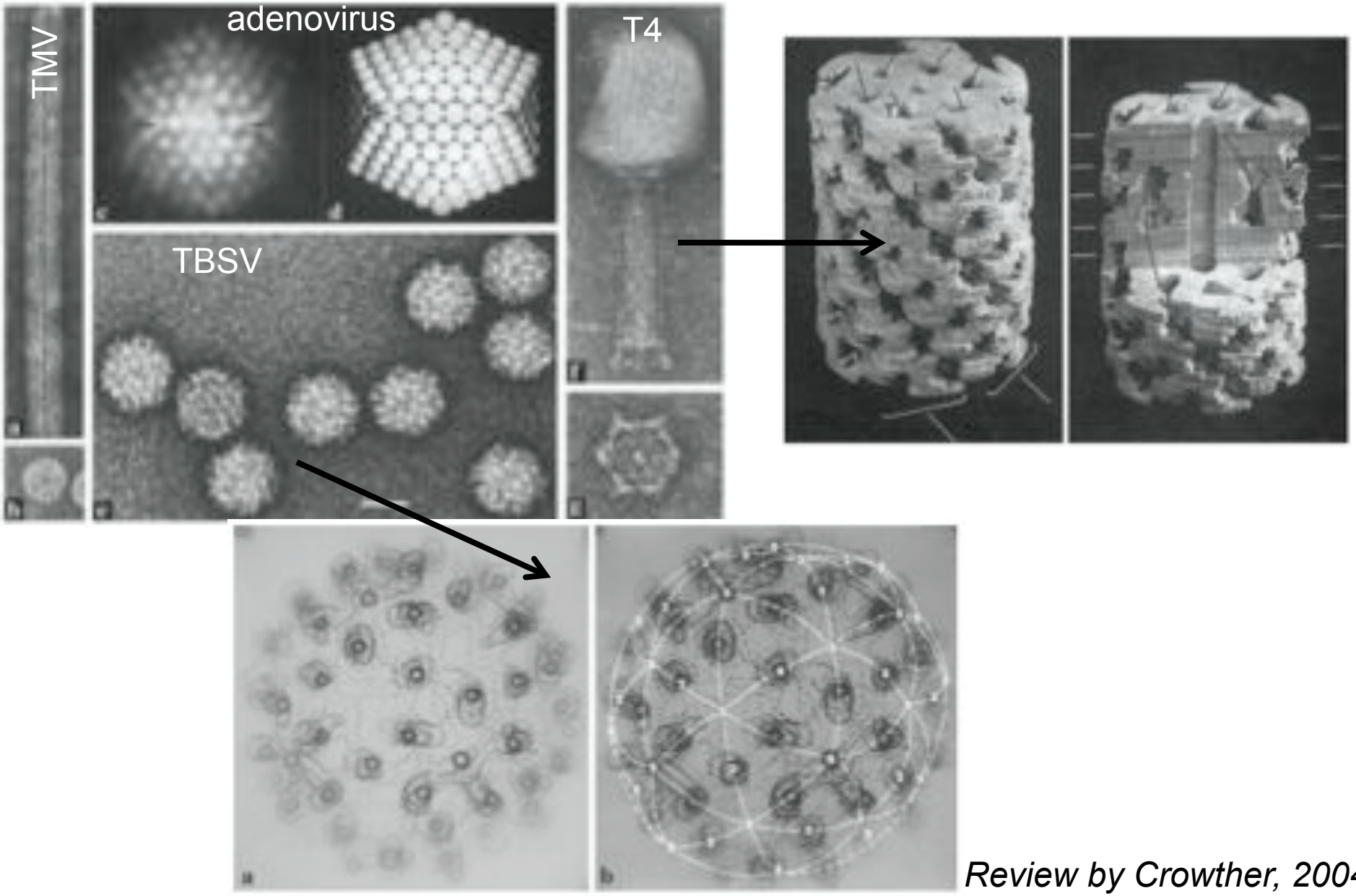
Fields Virology, 2006



Bacteriophages

- "Parasites" that infect cells
- Bacterial viruses are also called phages
- 10^2-10^3 Å in diameter
- Composition: protein shell + DNA/RNA genome. Some have lipid envelope
- Many viruses have icosahedral shell

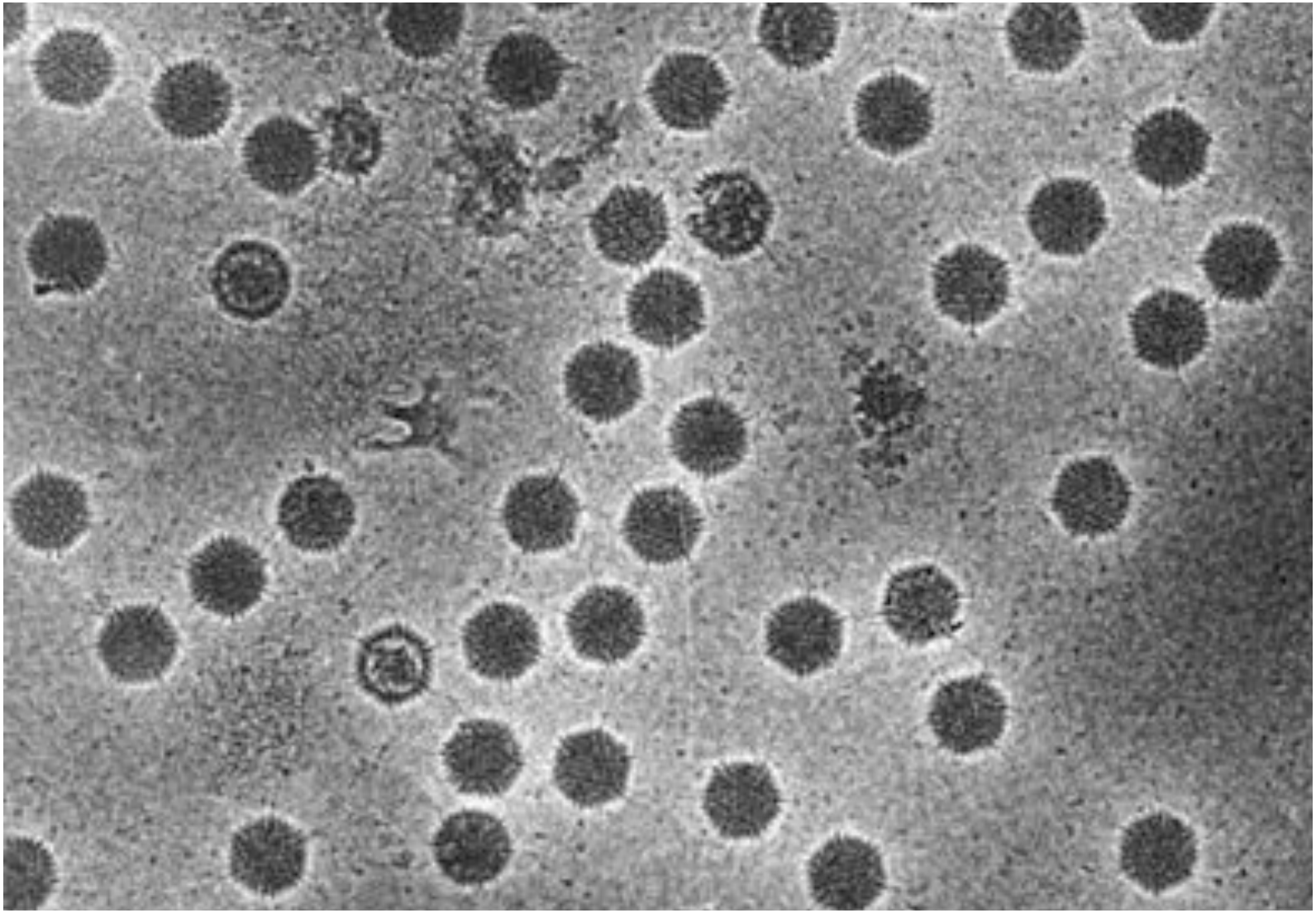
Viruses as Early 3-D EM Targets (1960s)



Review by Crowther, 2004

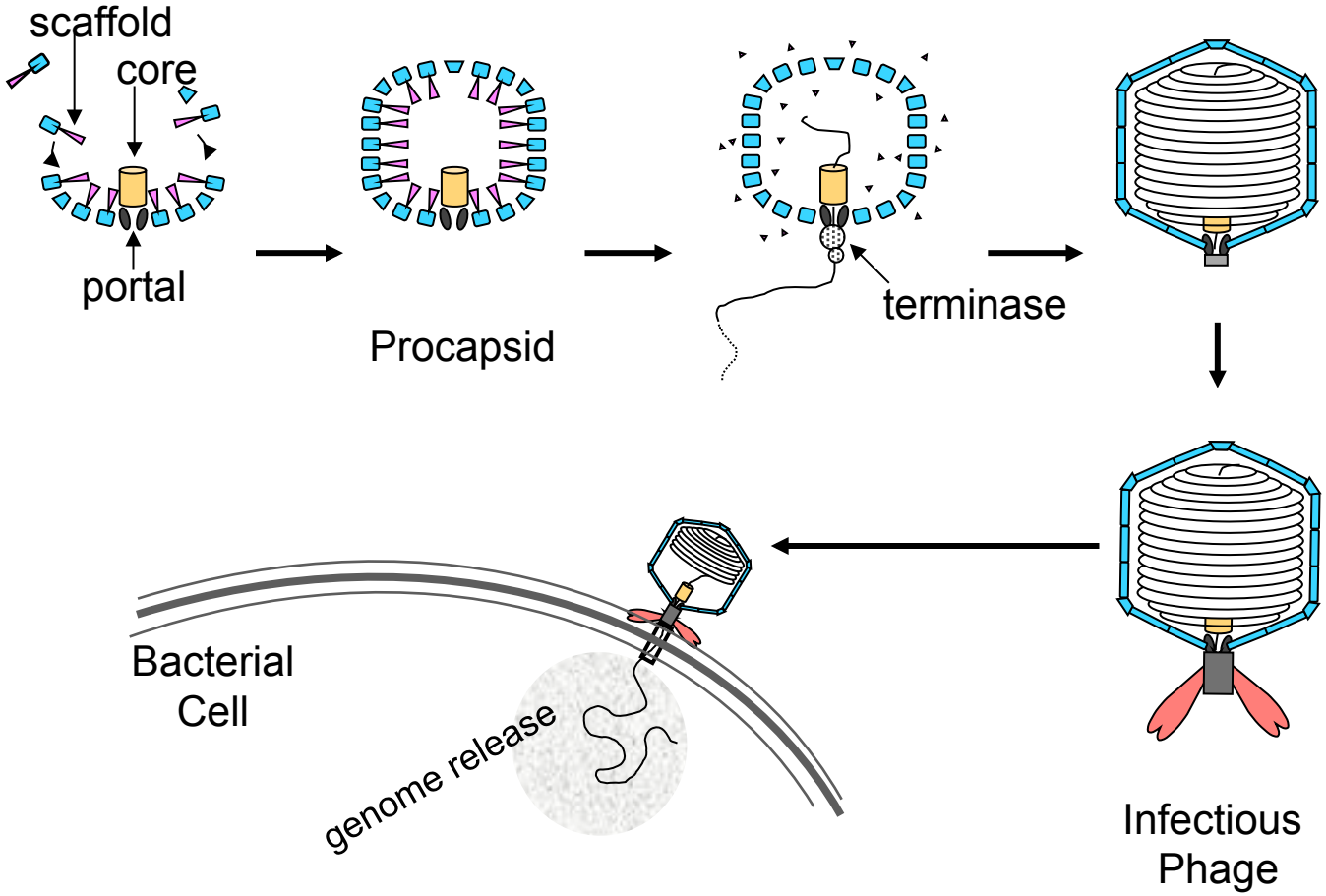
Cryo-EM: Frozen Hydrated Sample

Ice Embedded Adenovirus Virions



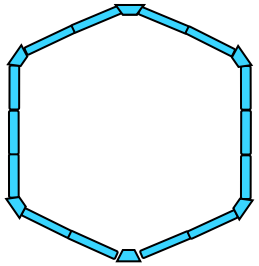
Adrian, M. *et al.* (1984). Cryo-electron microscopy of viruses. *Nature* **308**, 32.

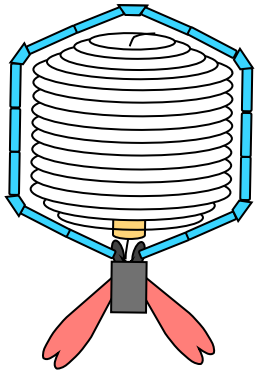
Tailed dsDNA Bacteriophages



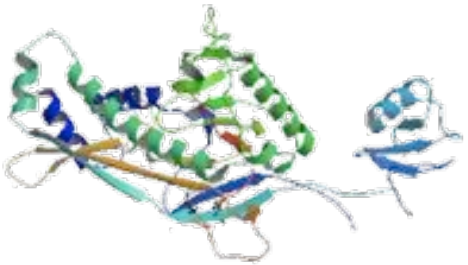
- **~10^{31} tailed dsDNA phage particles ➔ most abundant life form on Earth and enormous genetic diversity**
- **Some of the most complex macromolecular machines known**
- **Host specificity is determined by the tail fibers**

Tailed dsDNA Phage: Atomic Models

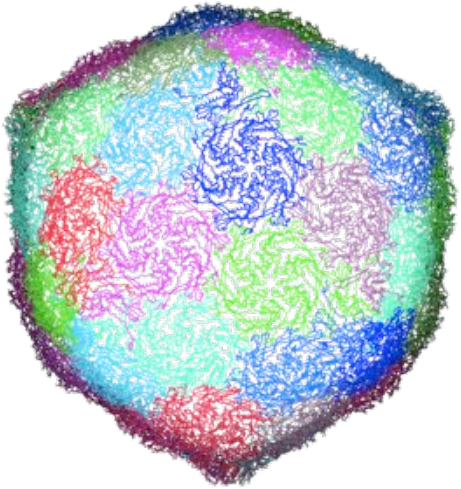




T4:
vertex protein
(2005)



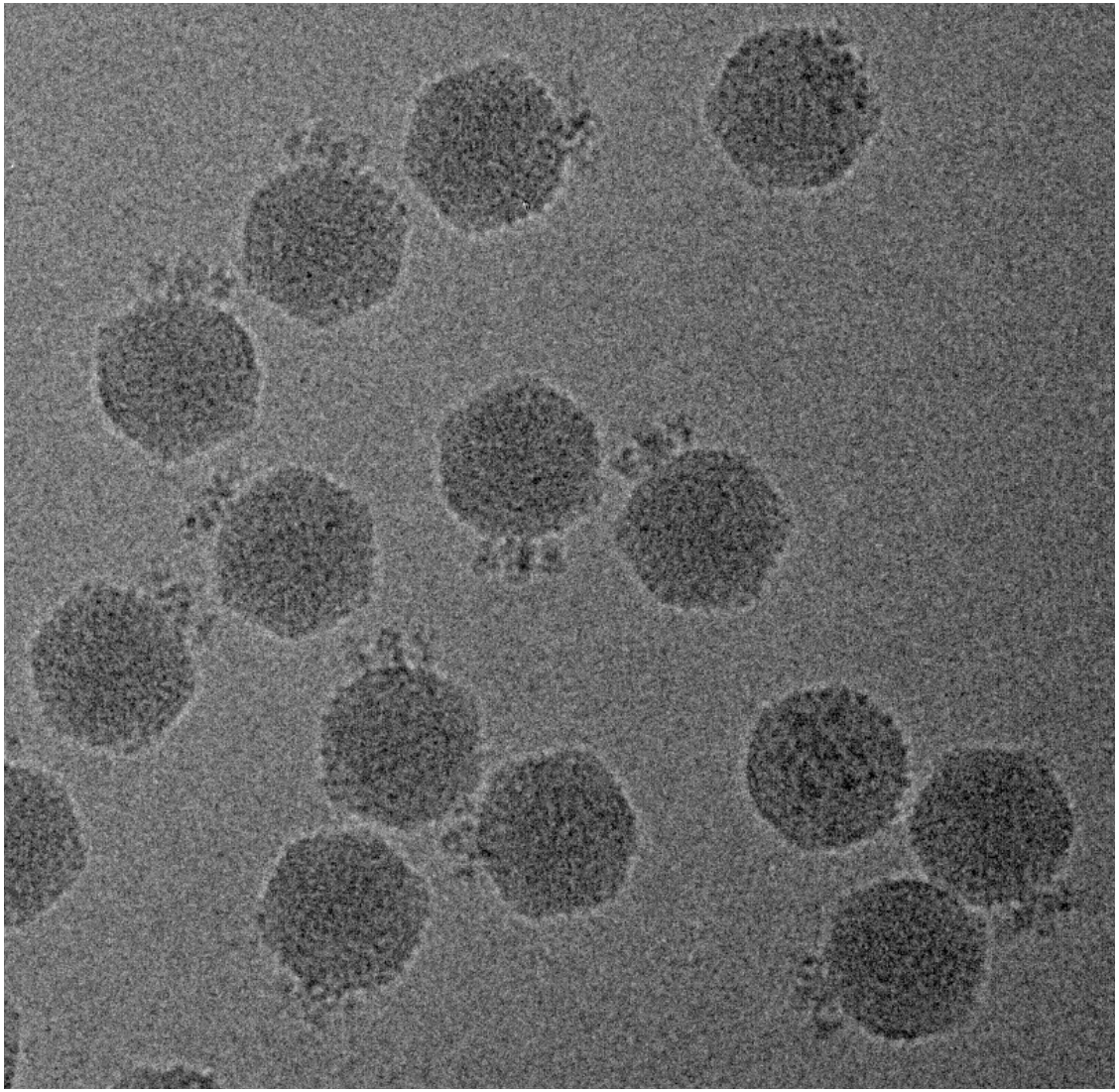
HK97:
recombinant shell
(2000)



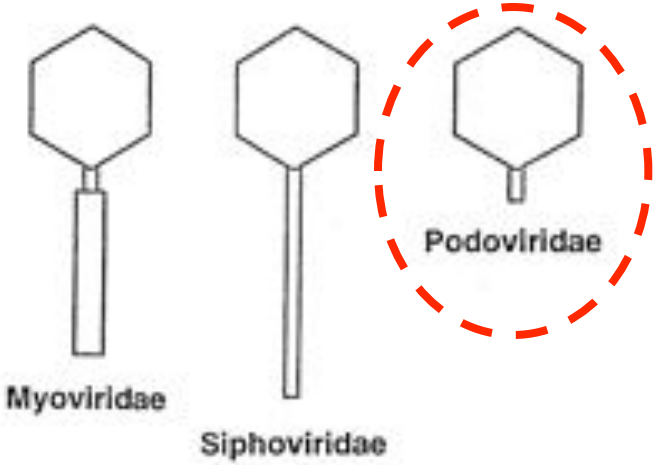
infectious phage

?

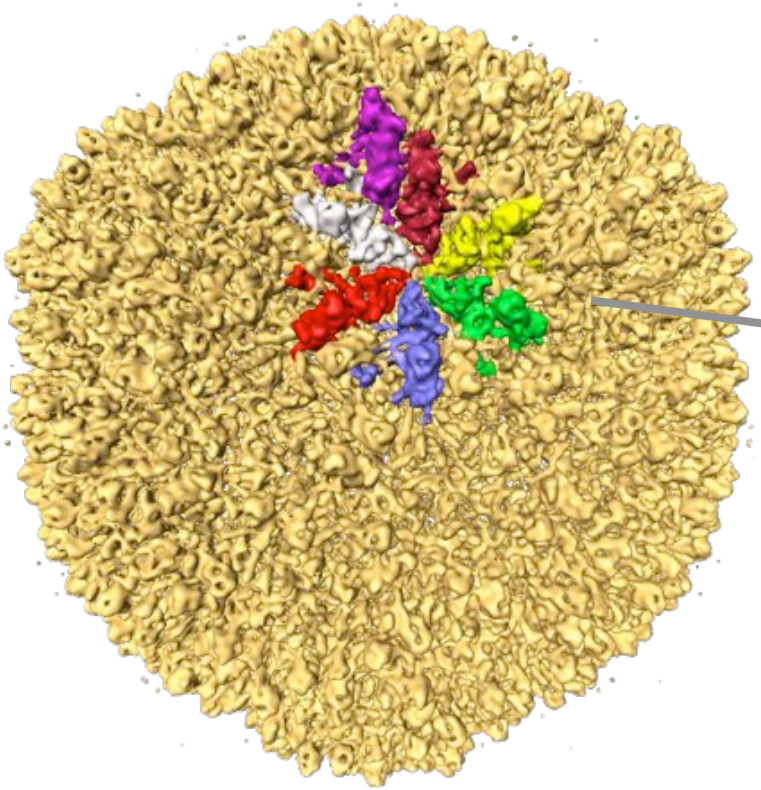
Epsilon15 Phage



- short tail dsDNA phage (*Podoviridae*)
- infects *Salmonella anatum*
- ~40kb genome
- ~700Å in diameter
- ~20 MDa in mass

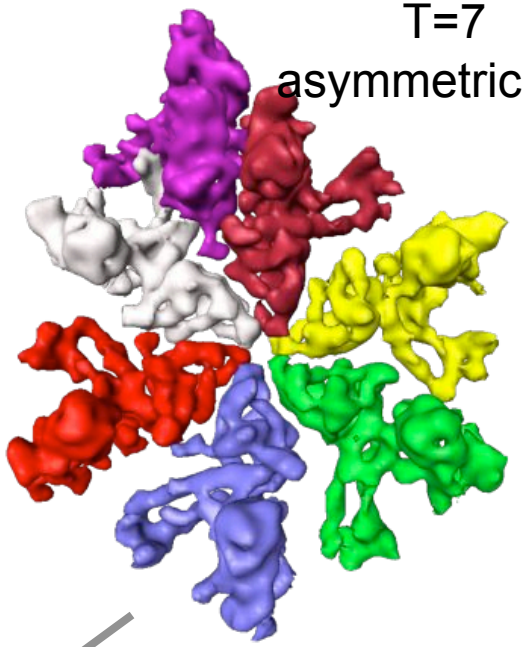


Epsilon15: Icosahedral Shell at 9.5 Å

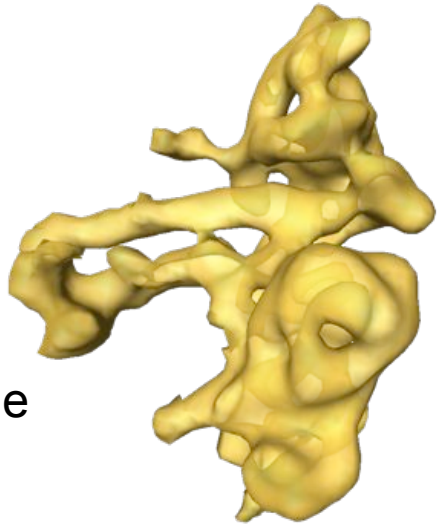


segmentation

T=7
asymmetric unit



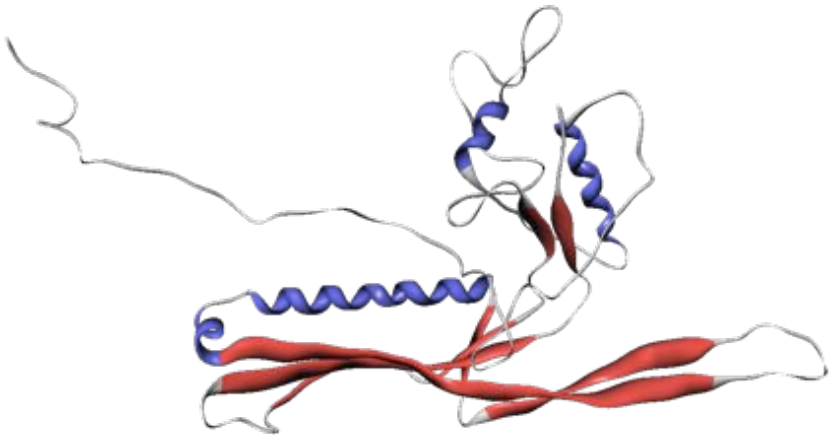
align & average



hexon average

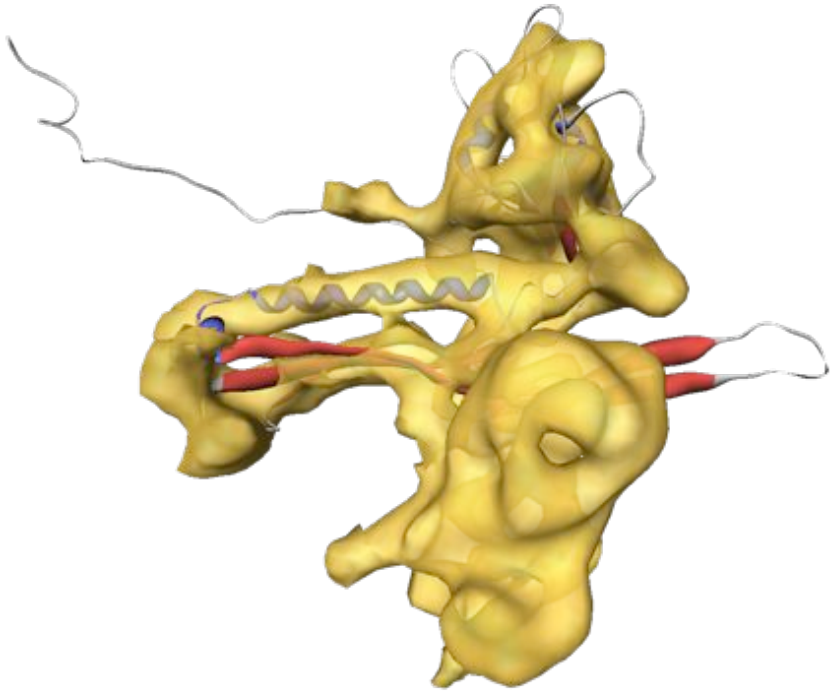
Jiang et al. Nature 2006

Shell Proteins of Epsilon15 Phage and HK97 Phage Have Common Fold



HK97

(Wikoff et al. Science 2000)



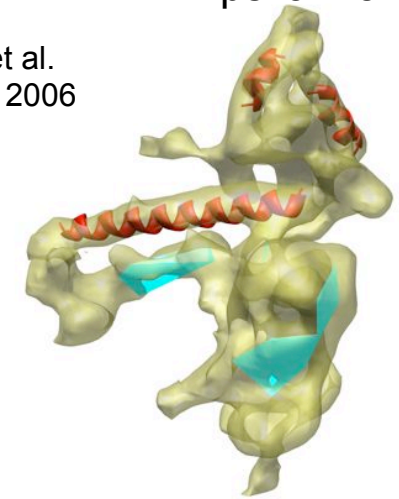
Epsilon15

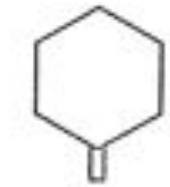
(Jiang et al. Nature 2006)

Common Shell Protein Fold for Tailed dsDNA Phages and Herpesvirus

Epsilon15

Jiang et al.
Nature 2006

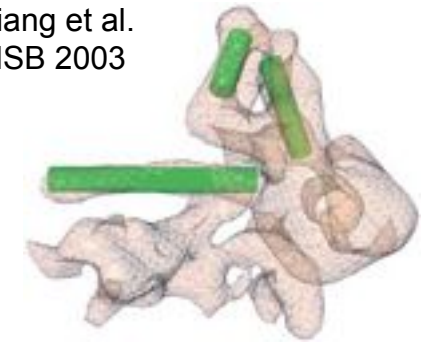




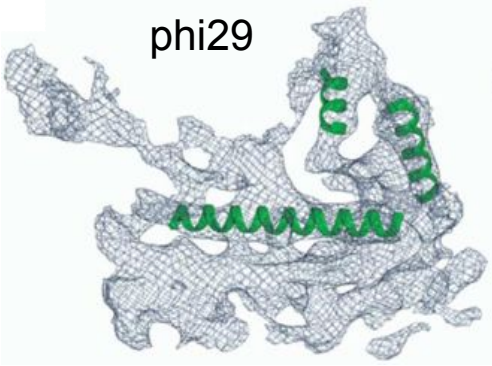
Podoviridae

P22

Jiang et al.
NSB 2003

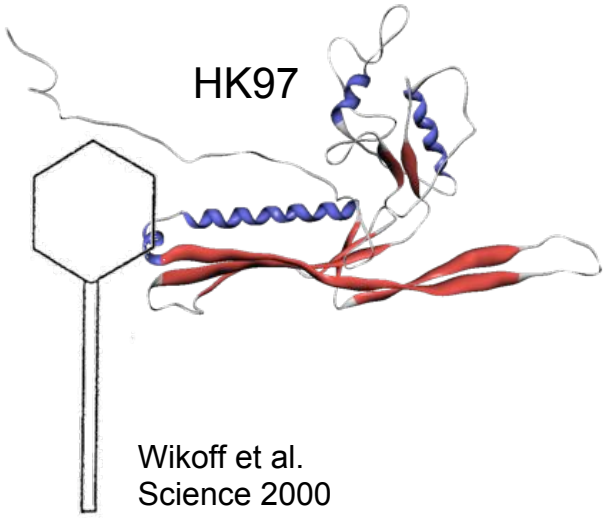


phi29



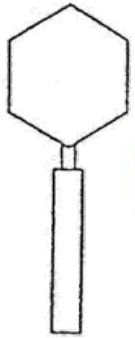
Morais et al. Mol. Cell
2005

HK97



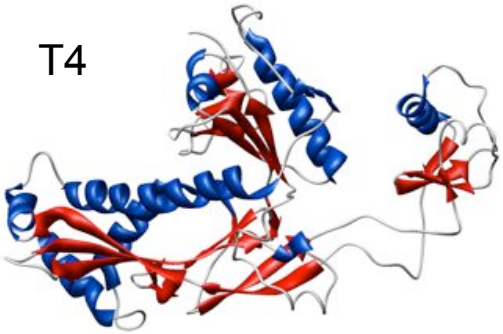
Wikoff et al.
Science 2000

Siphoviridae



Myoviridae

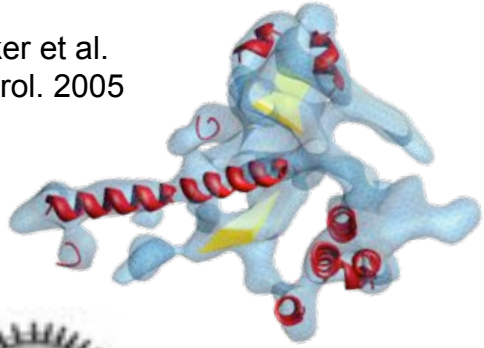
T4



Fokine et al.
PNAS 2005

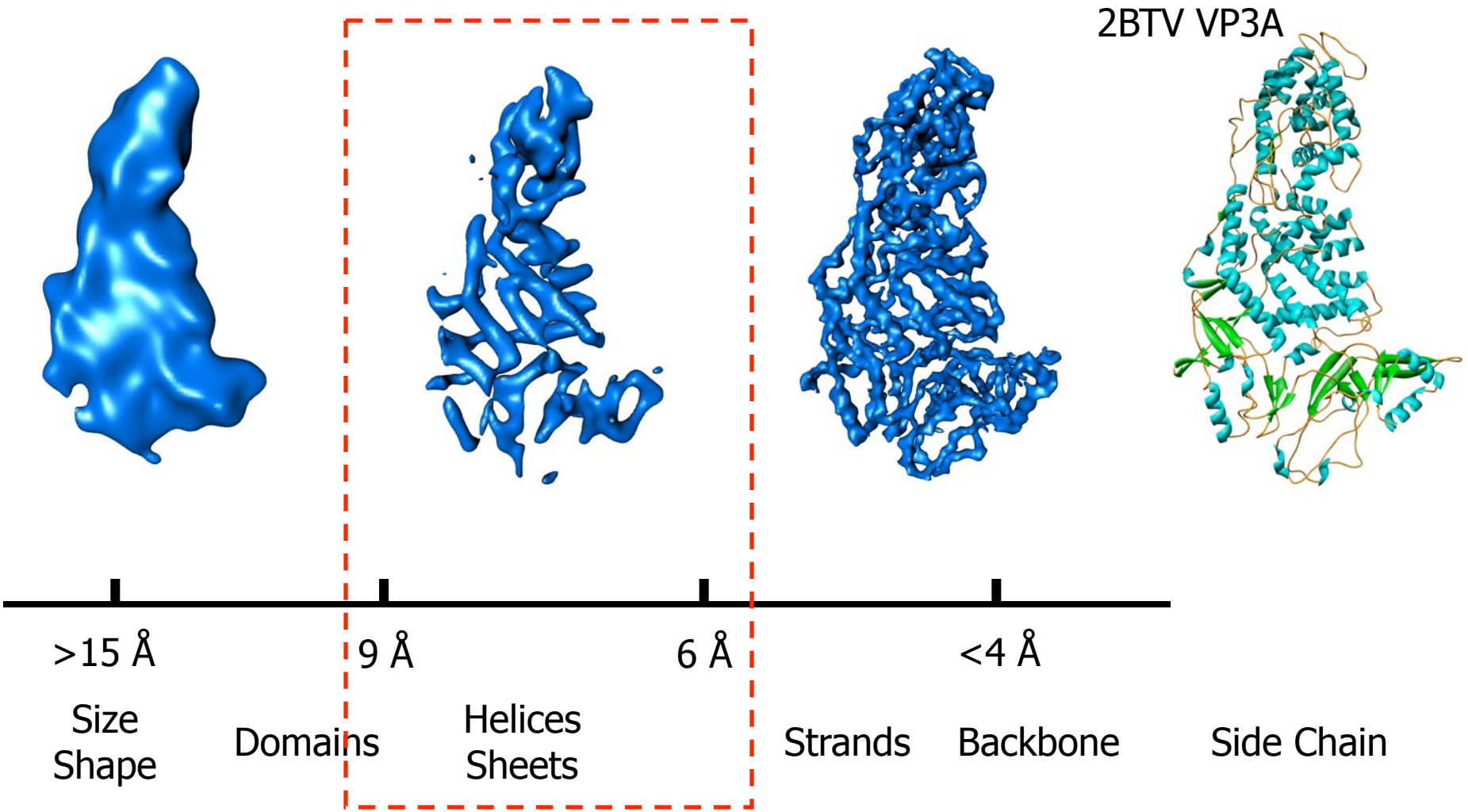
Herpes

Baker et al.
J.Virol. 2005

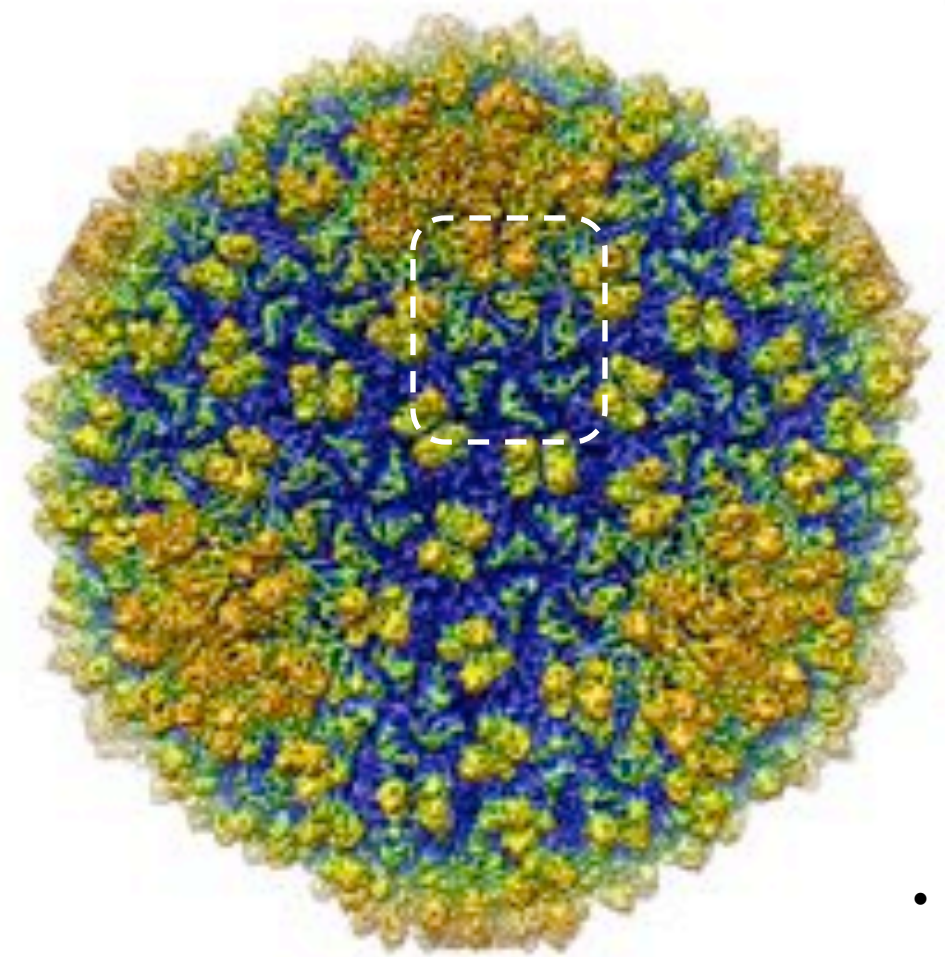


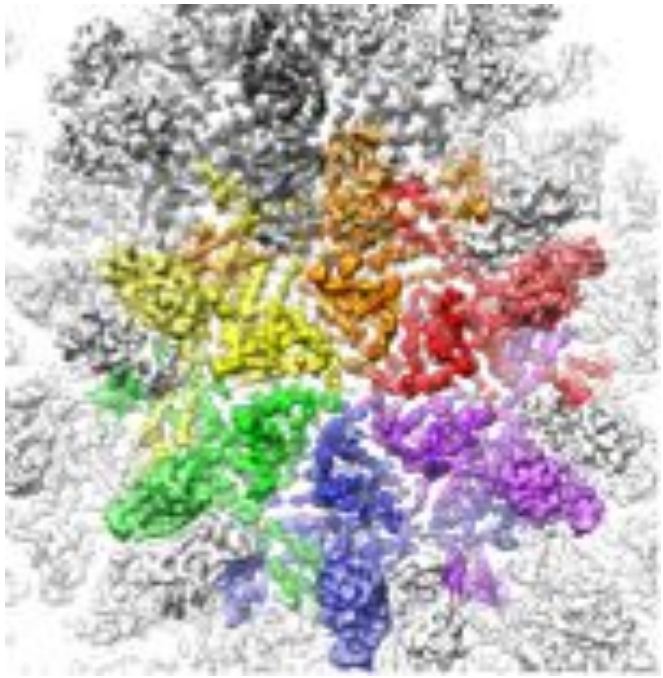
Herpesviridae

Towards Near-Atomic Resolution



4.5Å Icosahedral Structure of Epsilon15



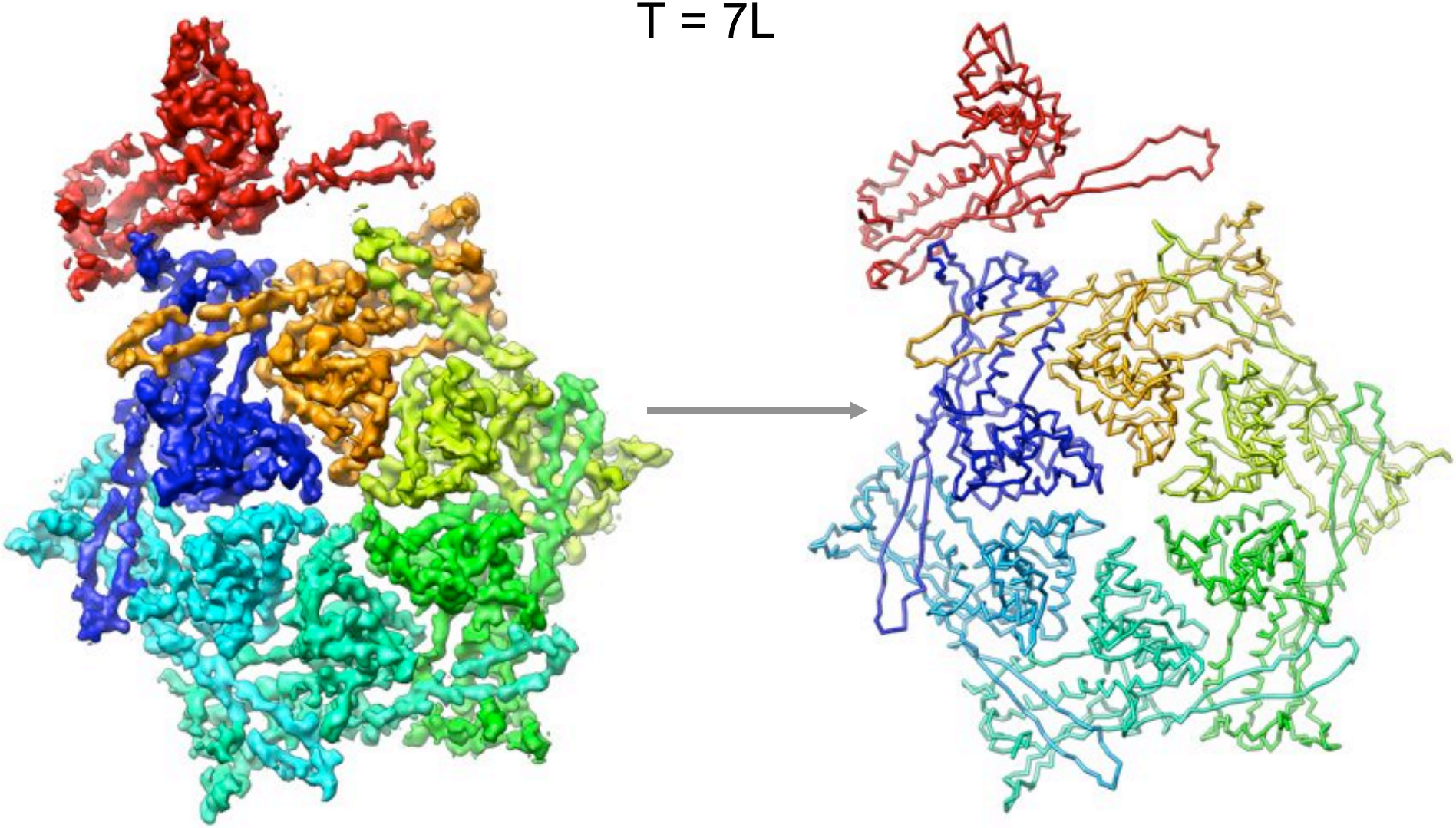


- Liquid Helium microscope
- Large dataset (~3000 micrograph)
- Image processing (EMAN)
- Large scale computation (million hours)

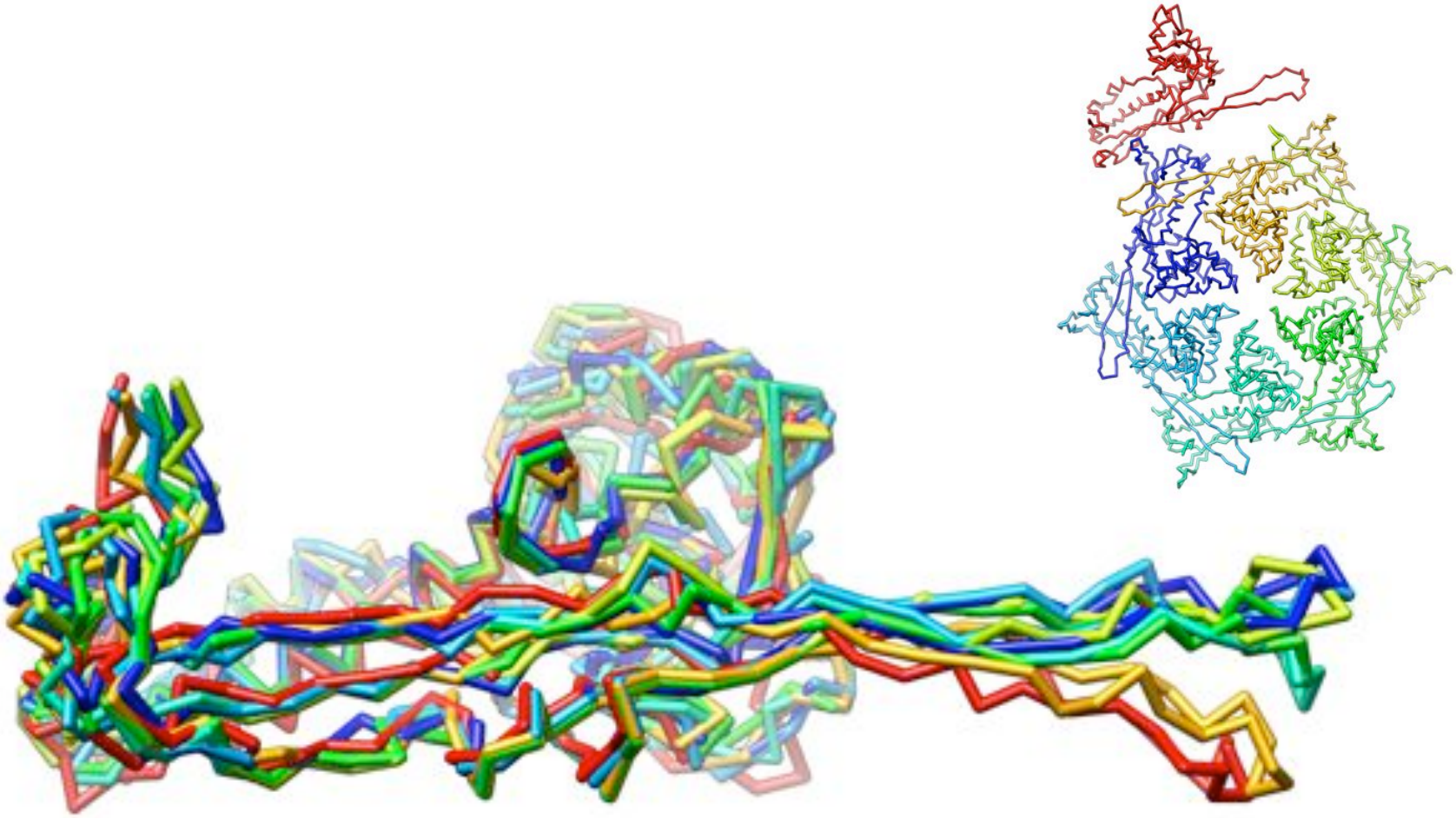
Jiang et al. Nature 2008

Cα Model of Epsilon15: Asymmetric Unit

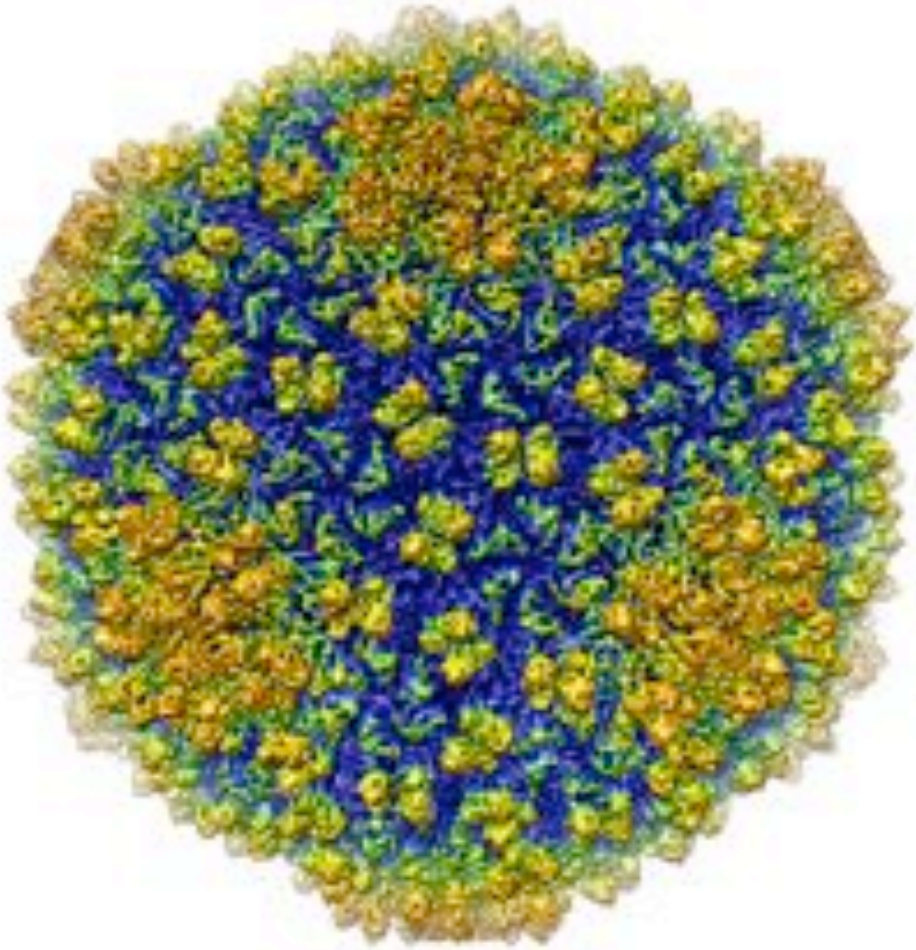
T = 7L



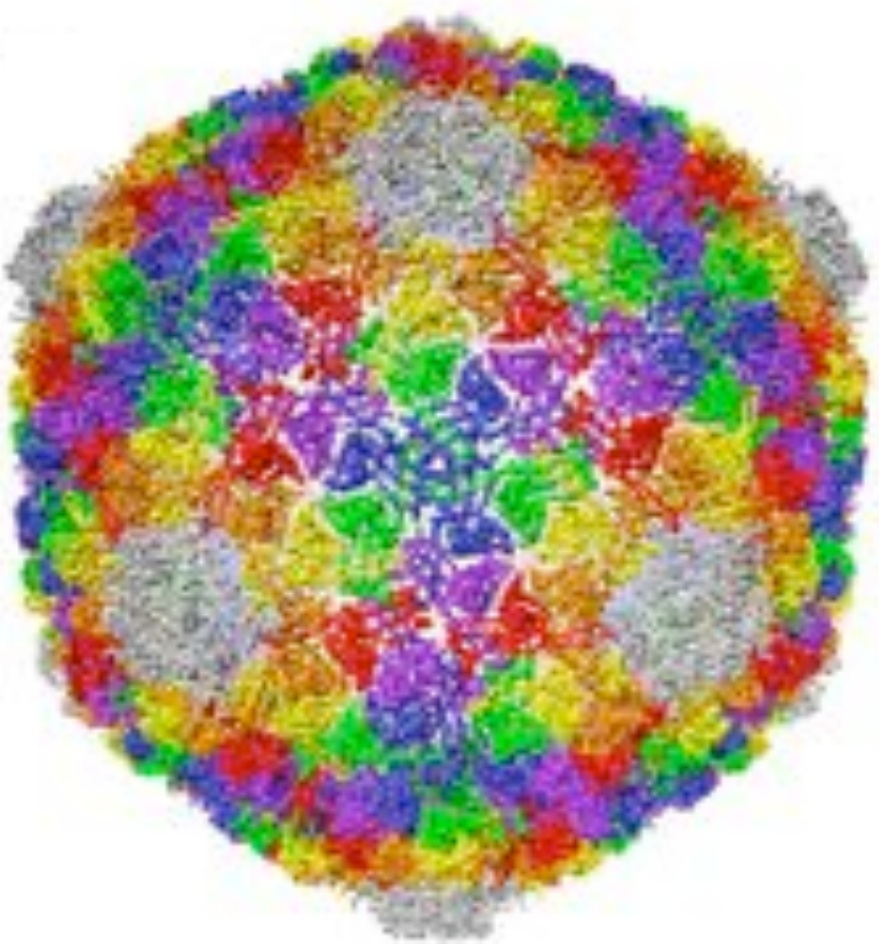
GP7: Conformation Variance



Capsid Model of Epsilon15

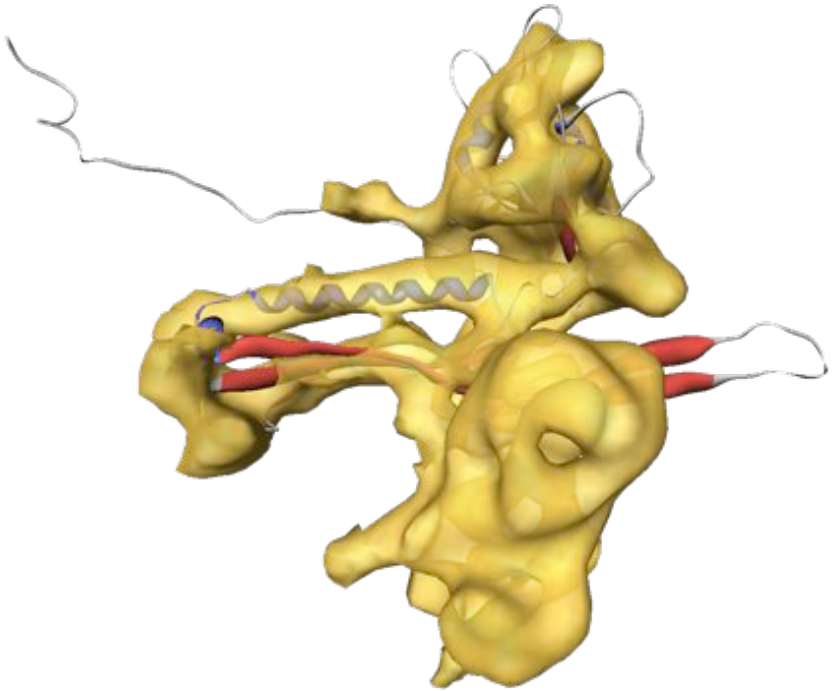


Density Map



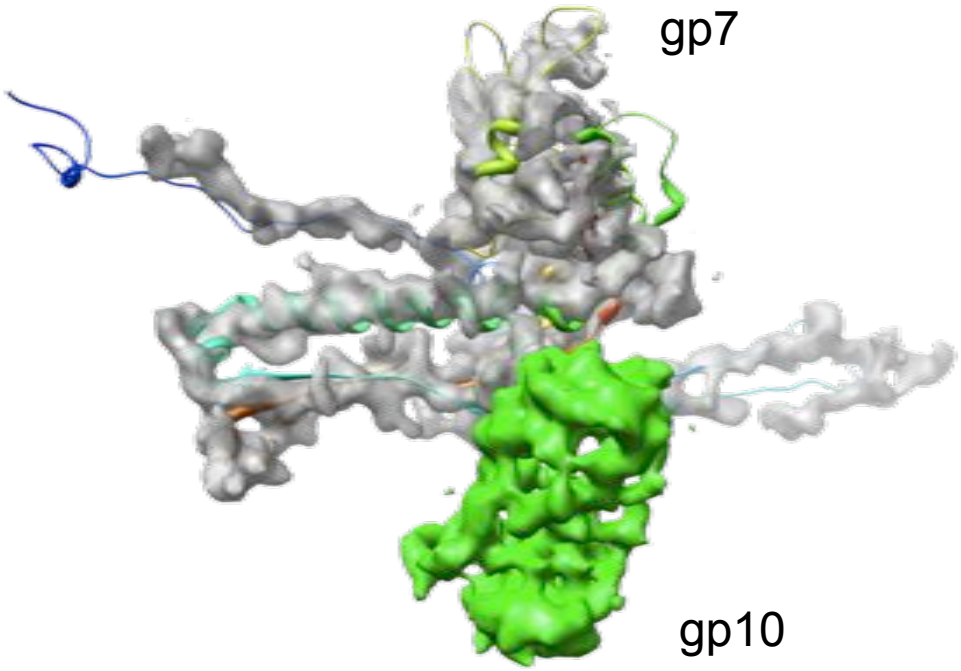
gp7 Model

9.5Å



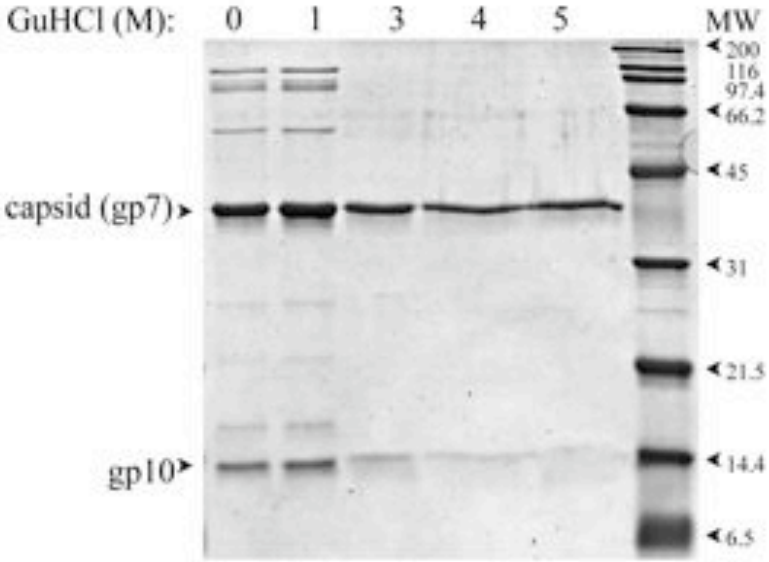
2006
(Jiang et al. Nature)

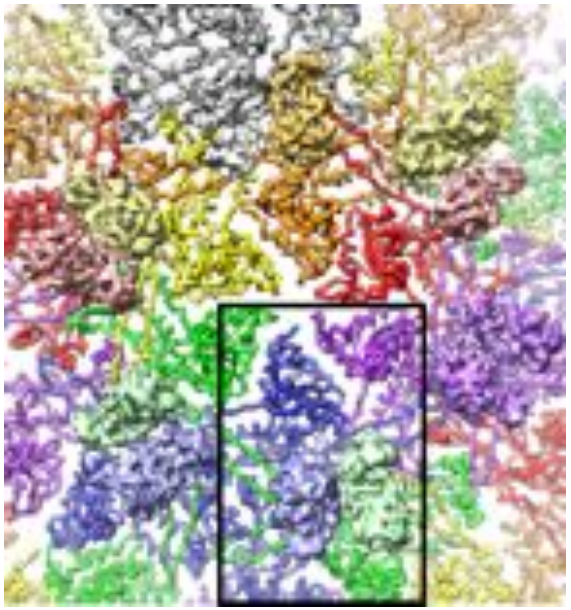
4.5Å

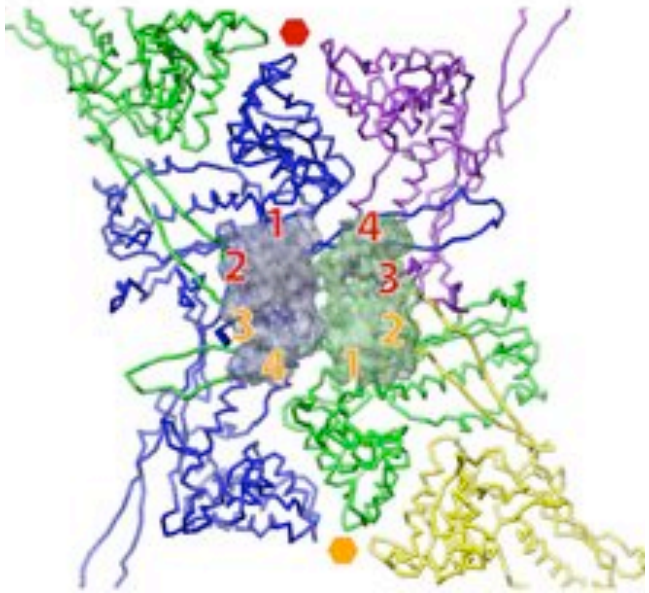


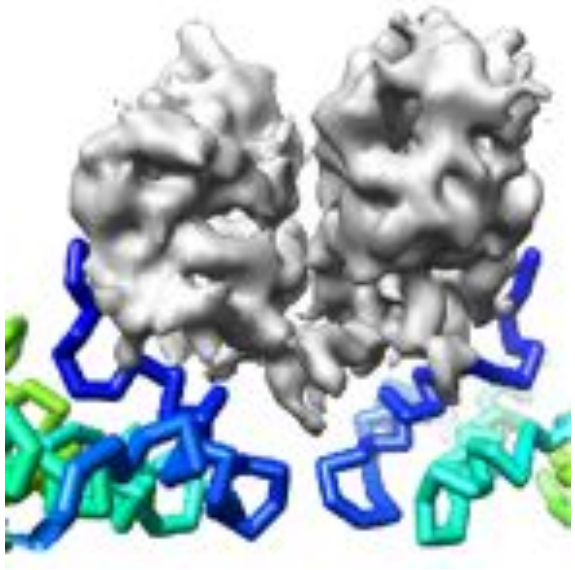
2008
(Jiang et al. Nature)

gp10: a molecular staple





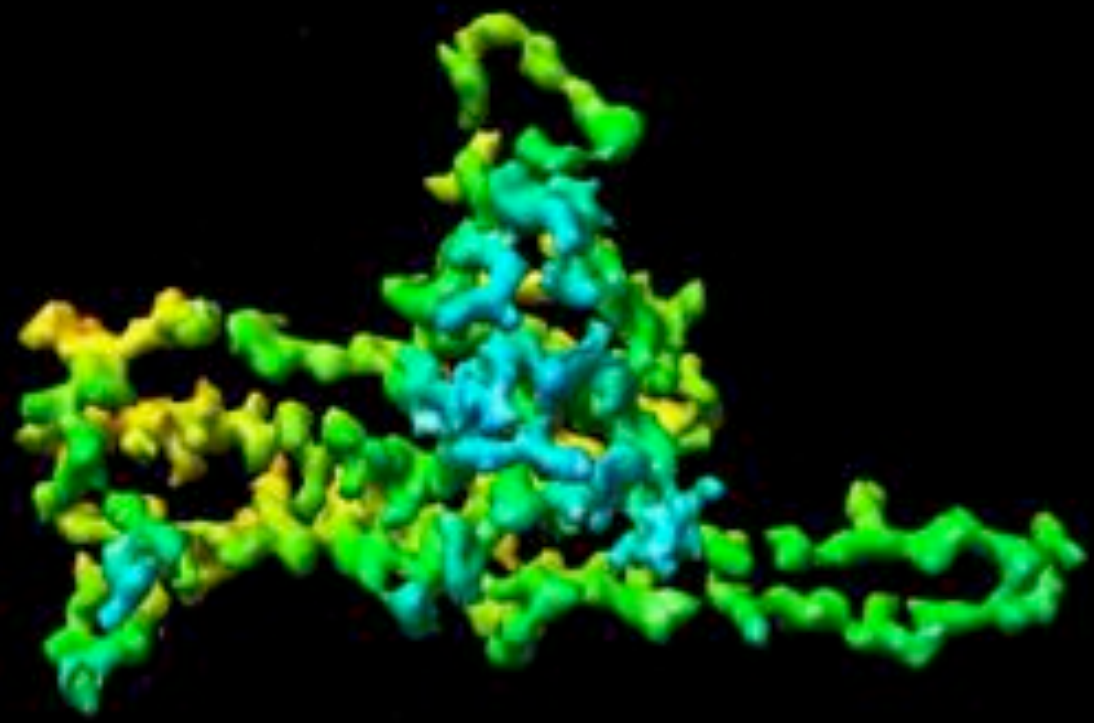




gp10

gp7

3.7 Å map: gp7



Near-atomic Resolution 3-D Reconstruction

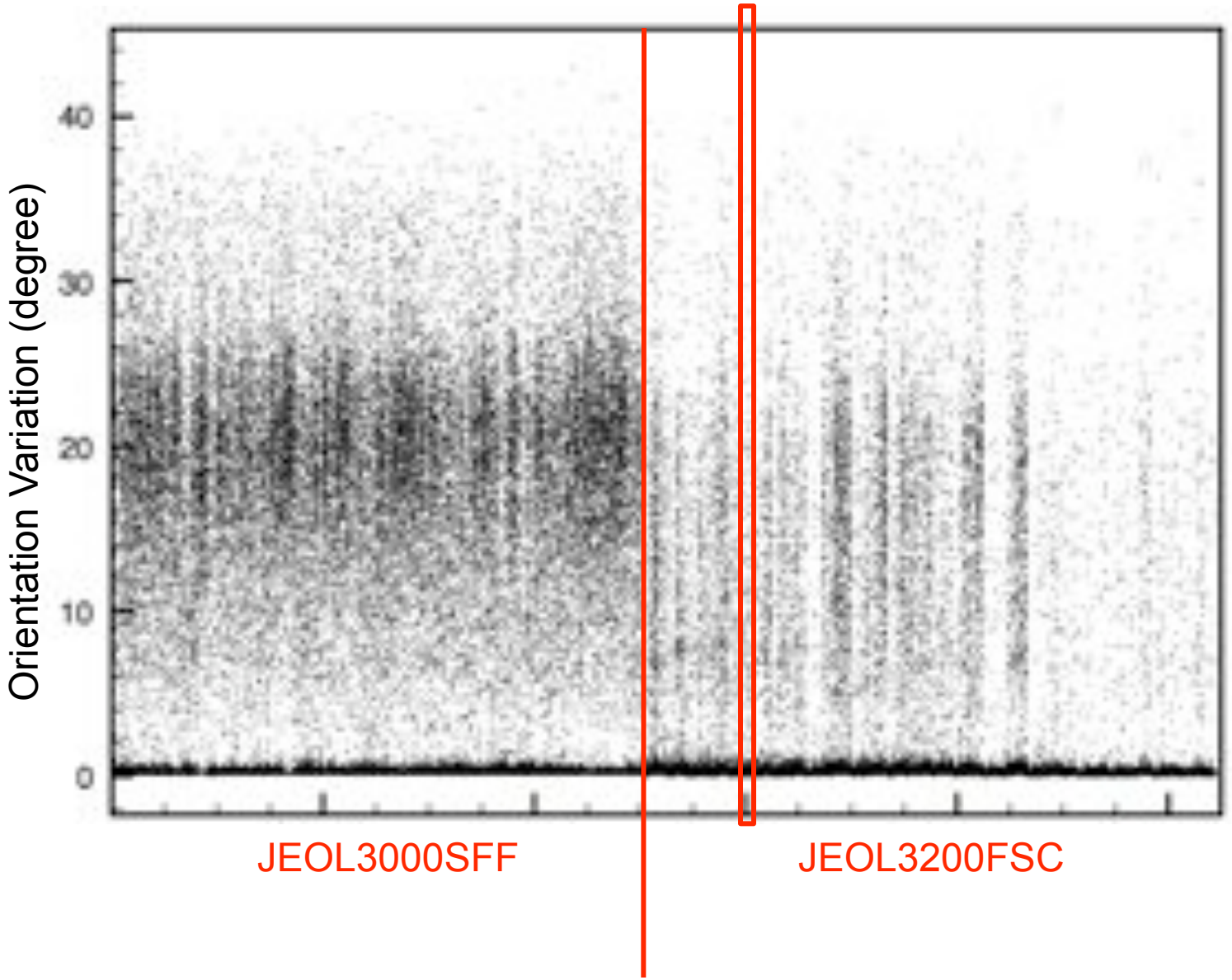
Tasks	Requirements/Methods
Sample	Stable, homogeneous
Imaging	Good microscope (LHe or LN_2 ?)
Image processing	Accurate alignment (orientation, center, defocus)
Computing	Cluster, Condor, Grid

High Resolution Images

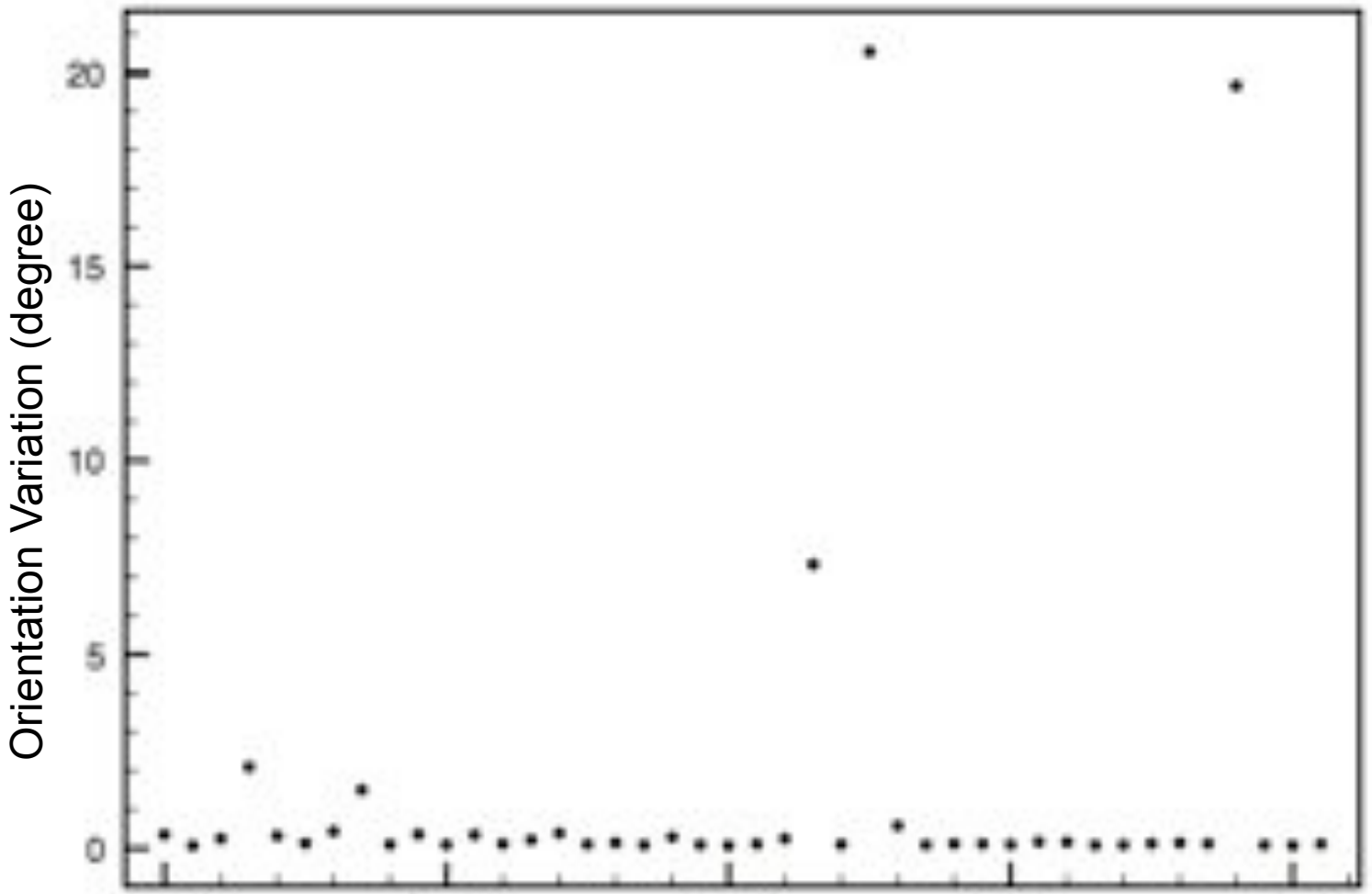
JEOL3000SFF: 300 kV, Helium
~3000 micrographs, 60k mag
1 exposure, on film, Nikon scanner,

JEOL3200FSC: 300 kV, N_2, Energy filter
~1500 micrographs, 50k mag
1 exposure, on film, Nikon scanner

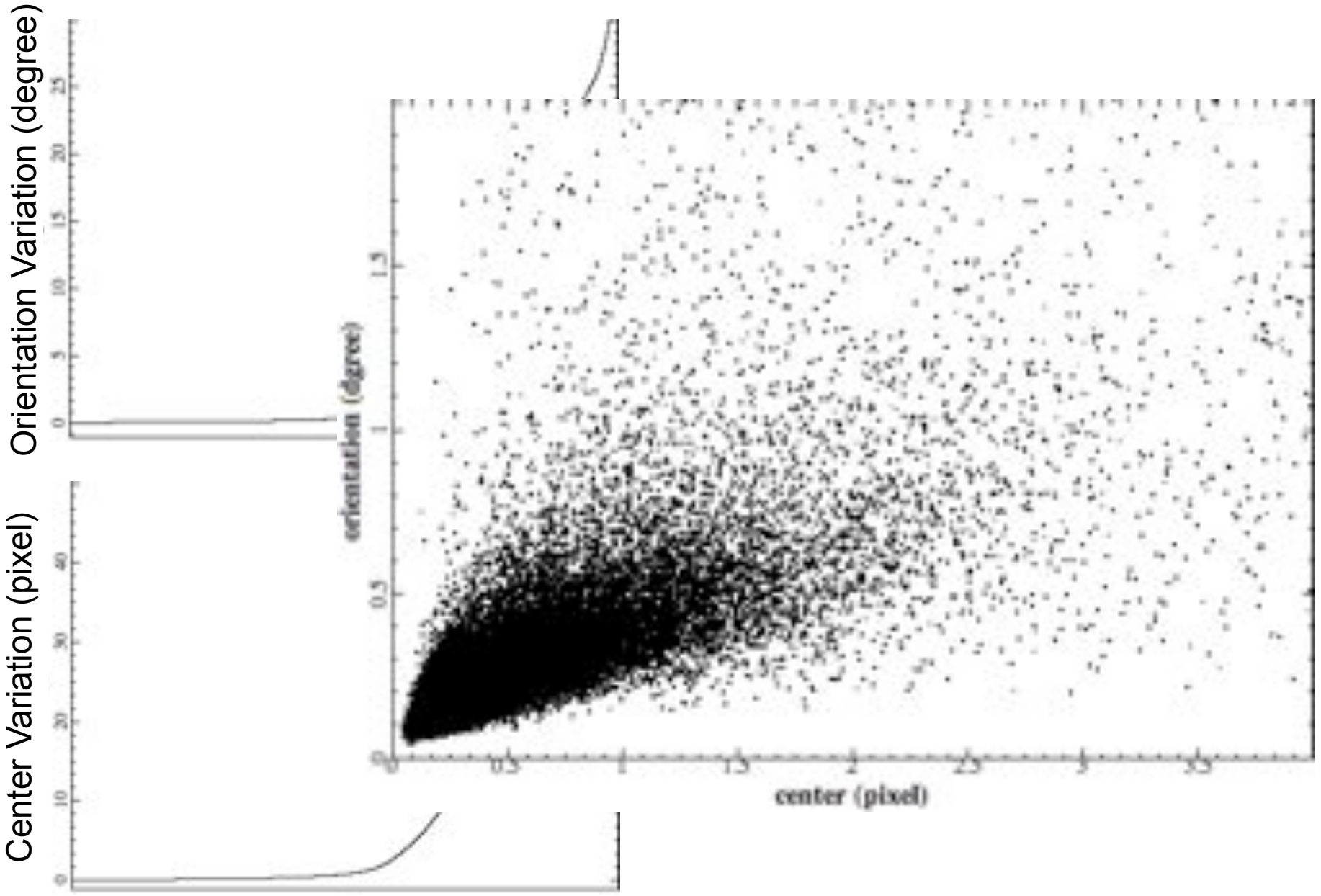
2-D Alignment



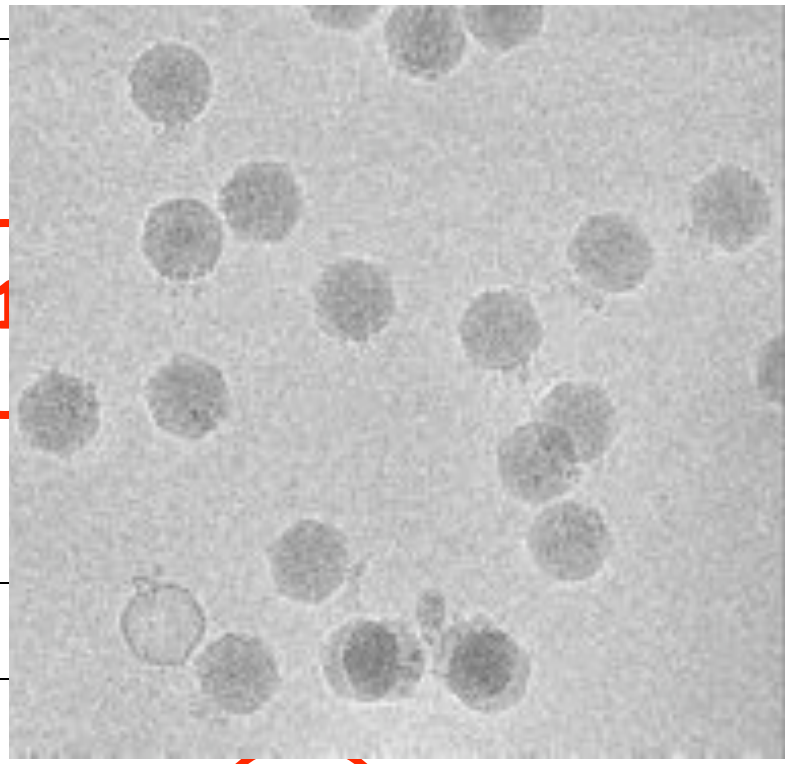
2-D Alignment

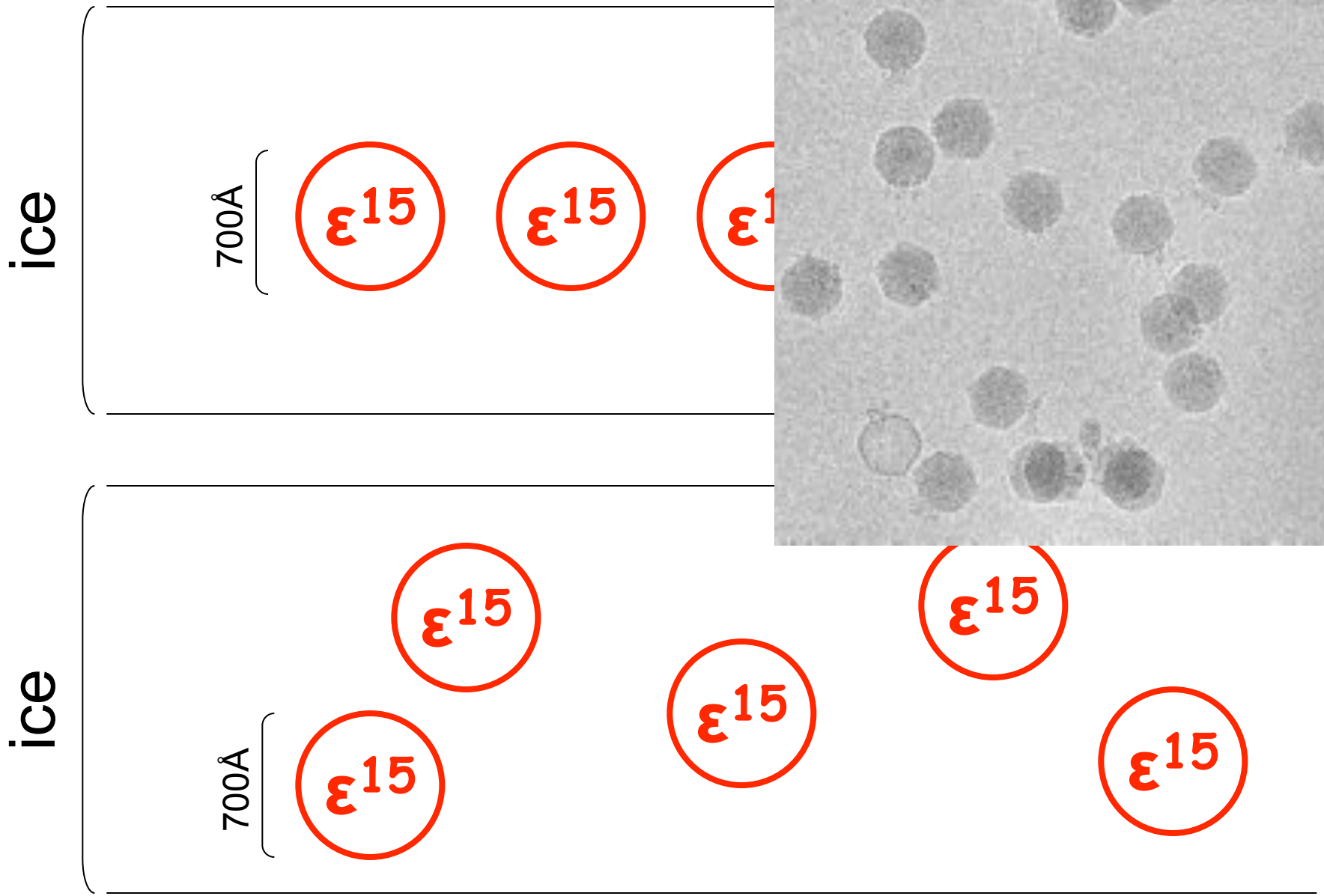


2-D Alignment

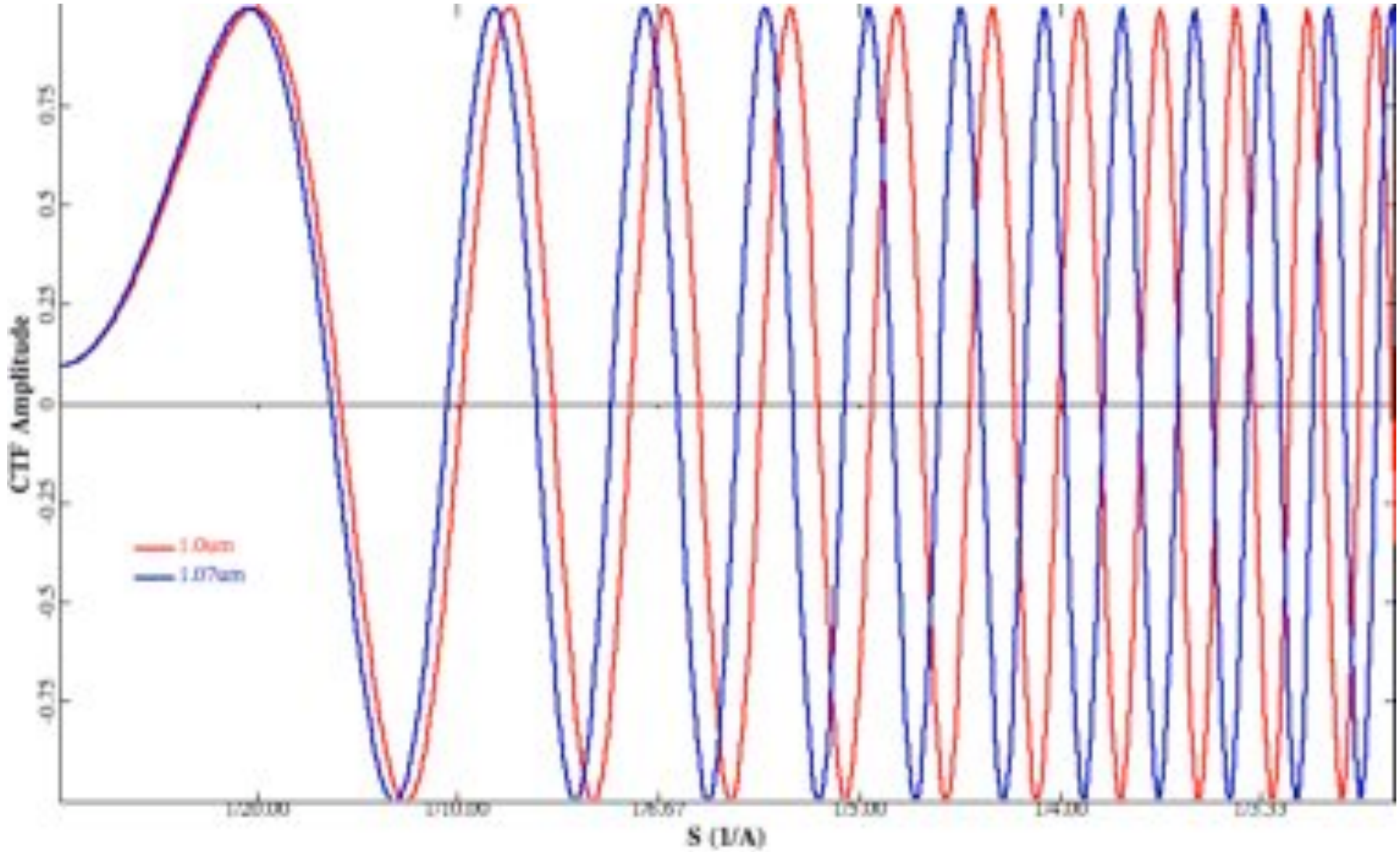


Focus Variation

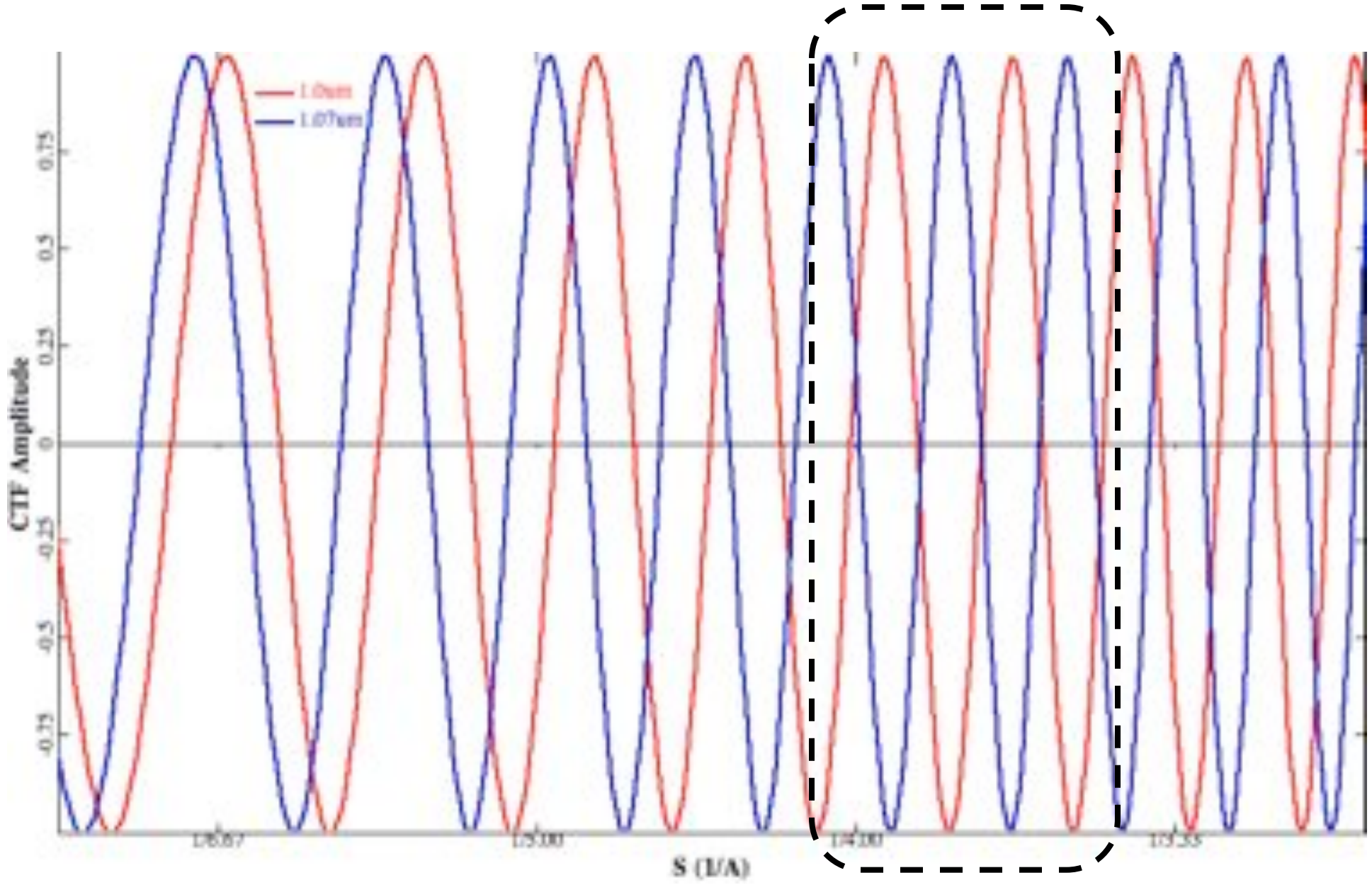




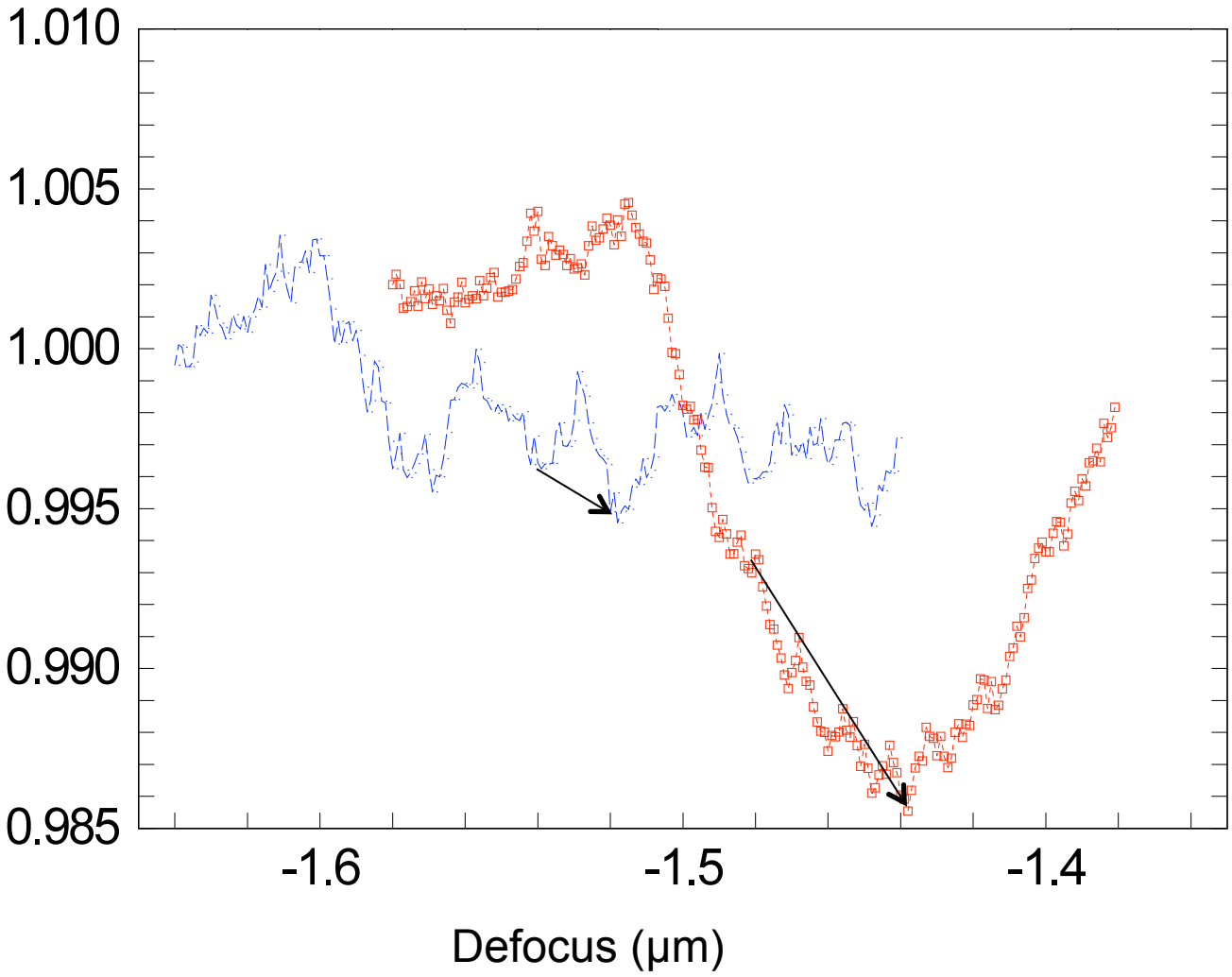
Focus Variation



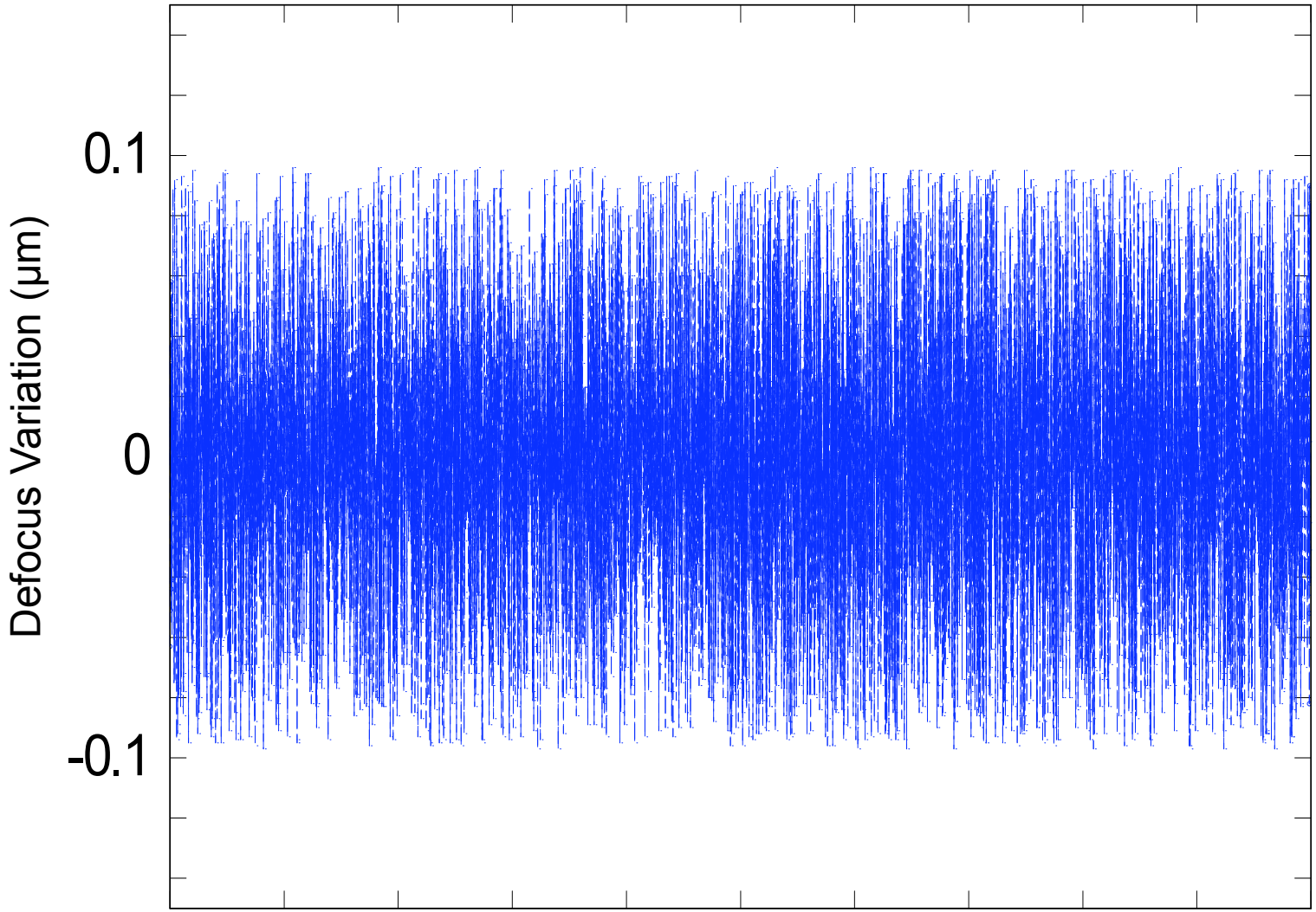
Focus Variation



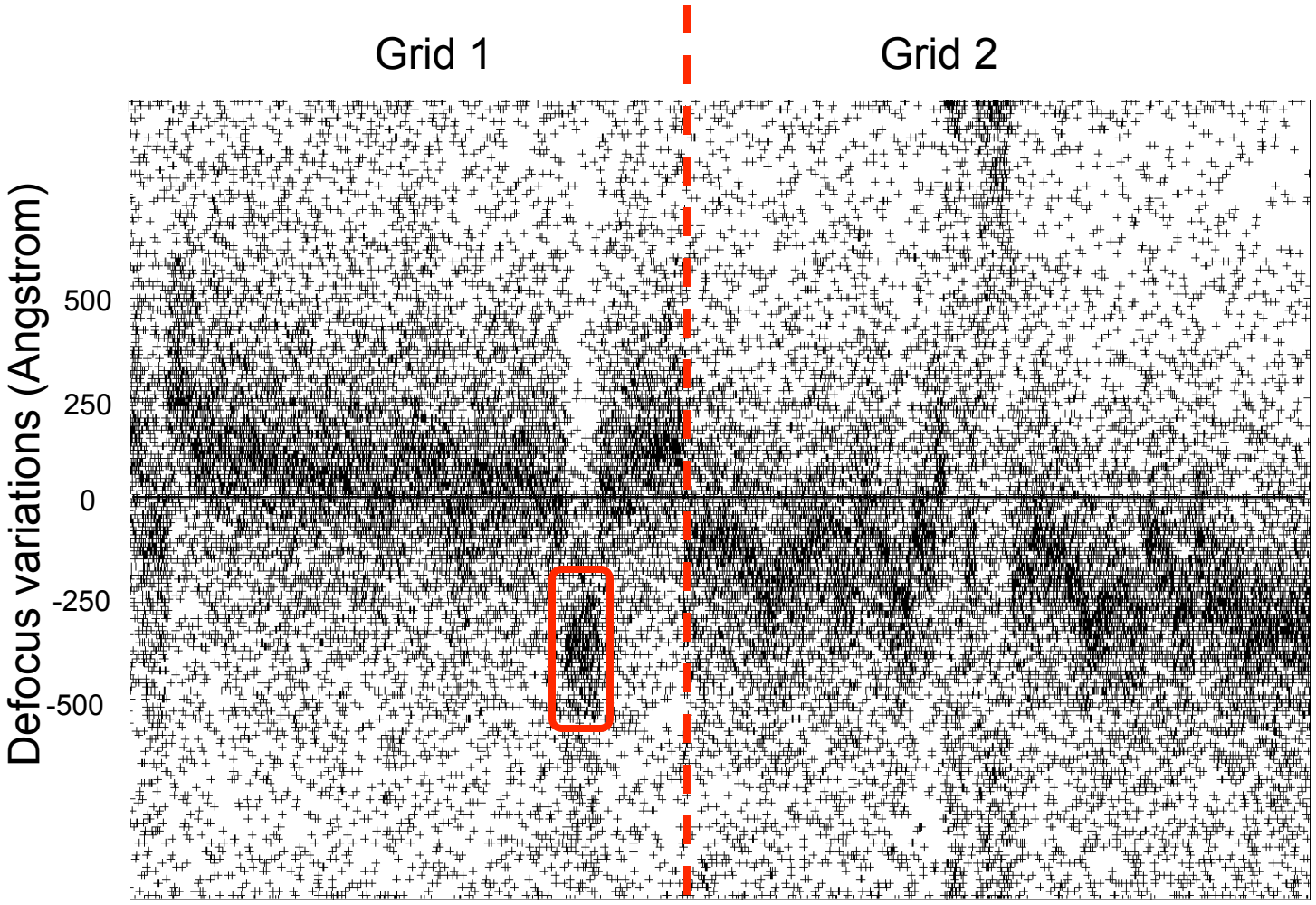
Focus Variation



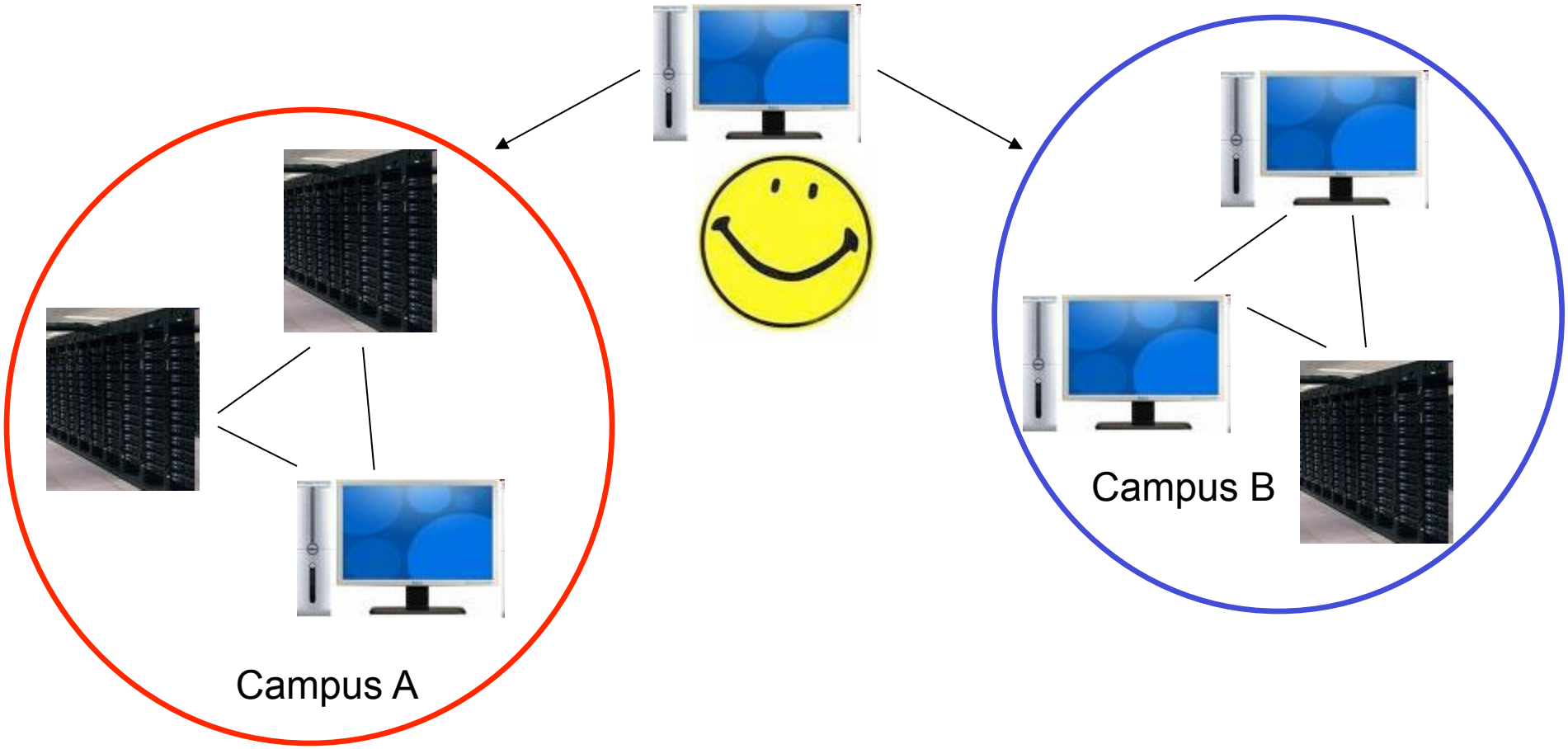
Focus Variation



Detecting CTF Fitting Errors

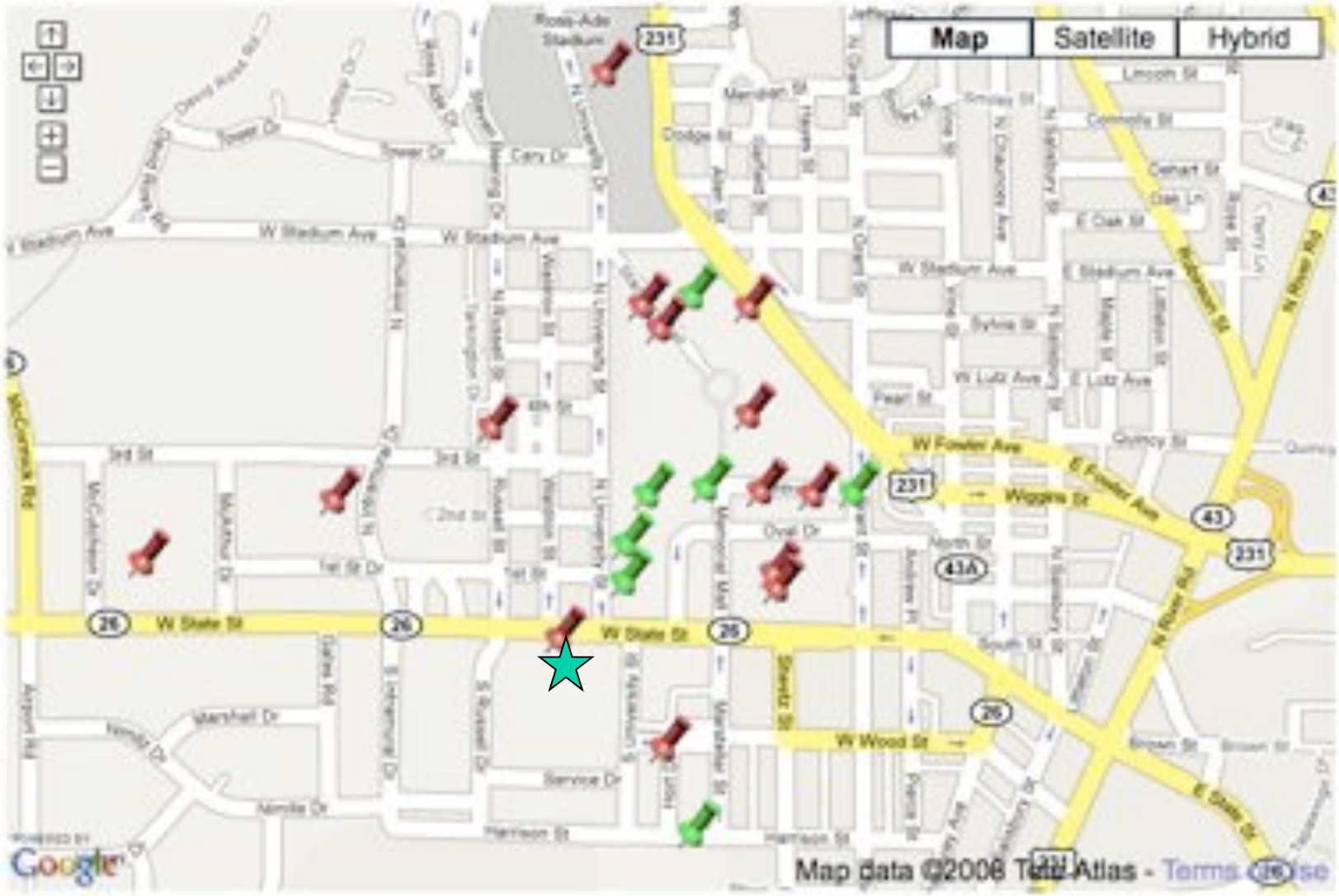


Large Scale Distributed Computing:
Condor



Developed by U Wisconsin - Madison

Large Scale Computing: Condor



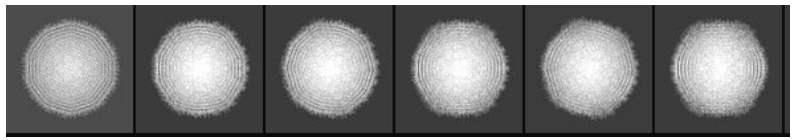
Greater than 50% available

25% - 50% available

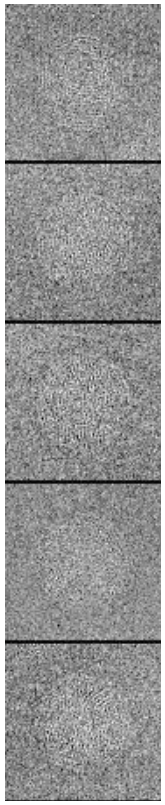
Less than 25% available

Parallelizing 2-D Alignment

reference projections of current 3-D model (n)



Cryo-EM particle images (m)



	0	1	2						n
0	M_{00}	M_{01}	M_{02}						M_{0n}
1	M_{10}	M_{11}	M_{12}						
2	M_{20}	M_{21}	M_{22}						
					M_{ij}				
m	M_{m0}								M_{mn}

- All to all comparison
- Comparisons are independent
- $m=10^2$-10^5
- $n=10^2$-10^3
- Image size: 100-1000^2
- 3 variables (2 center, 1 rotation) for comparison pair
- Block parallelization

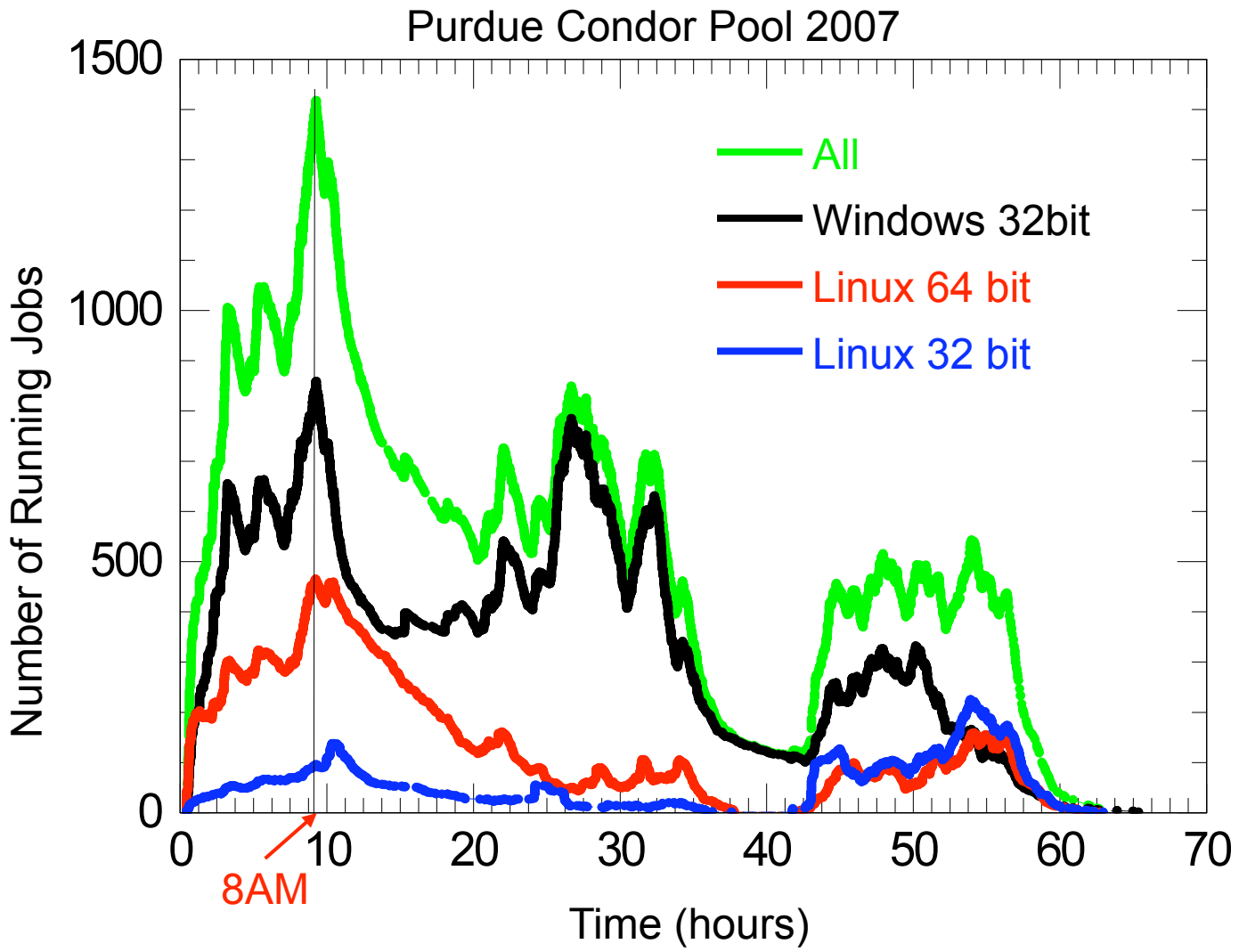
EMAN on Condor

- Implemented EMAN on Condor
- Works well with Condor on Linux (32bit and 64bit X86, PPC64) and Windows
- Simultaneously run on all Purdue Condor pools (BoilerGrid)
  - Purdue main campus
  - Purdue Calumet
  - Indiana state university
  - Notre Dame University
- Hundreds to thousands CPUs can be used

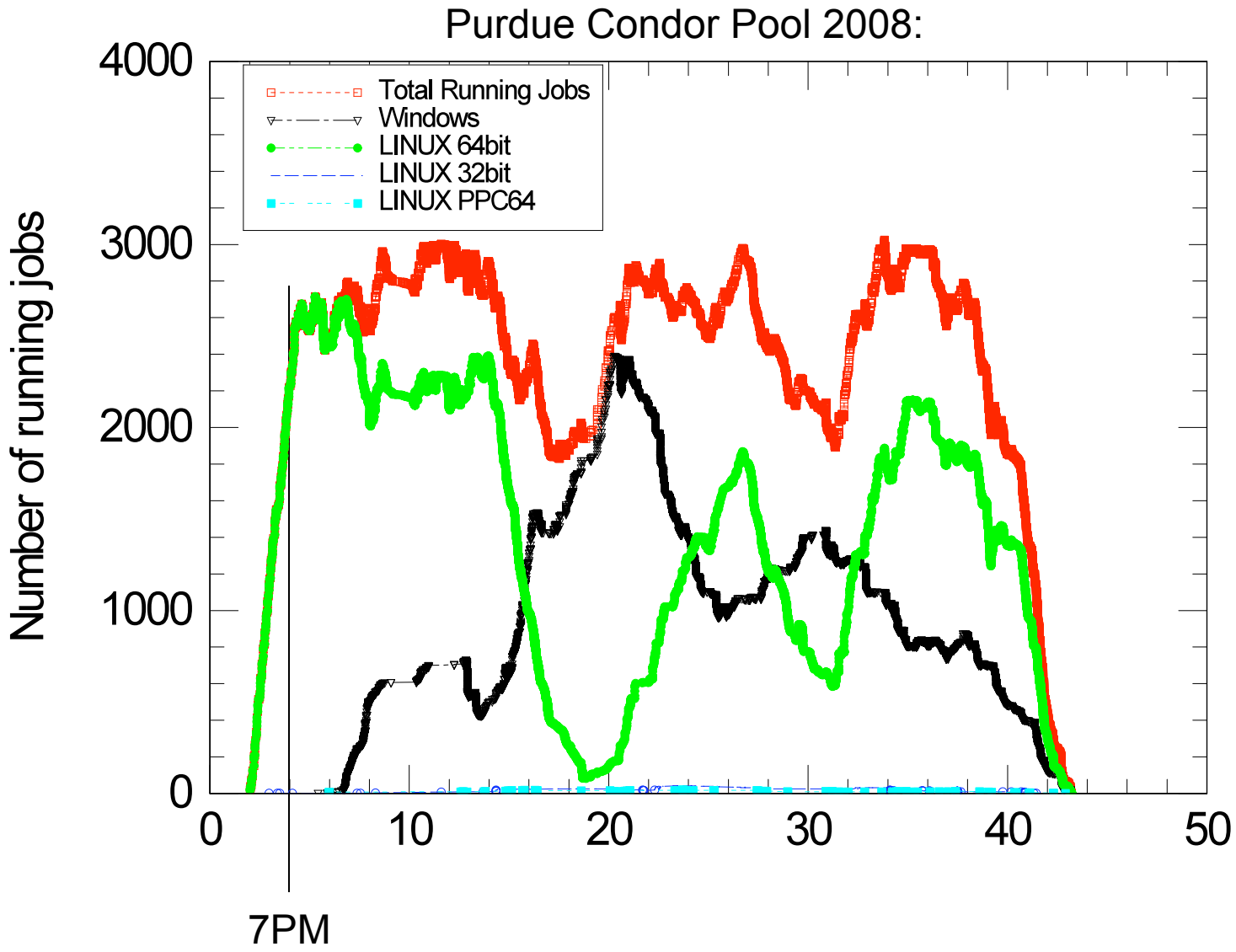
Remaining Issues:

- Data IO/image file transfer bottleneck ?
- Distributed checkpointing ?

Large Scale Computing: Condor 100 → 1000 CPUs

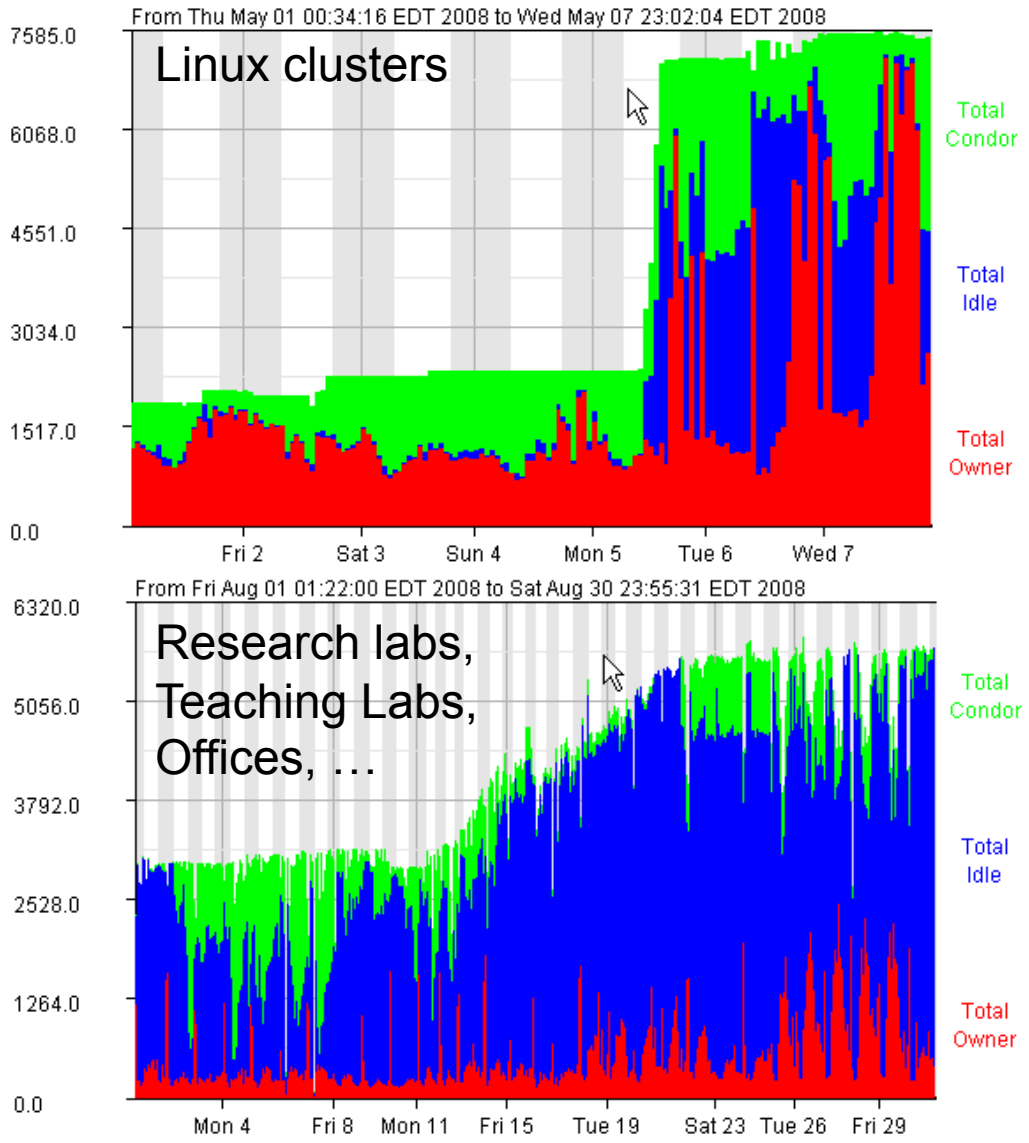


Condor Is Growing

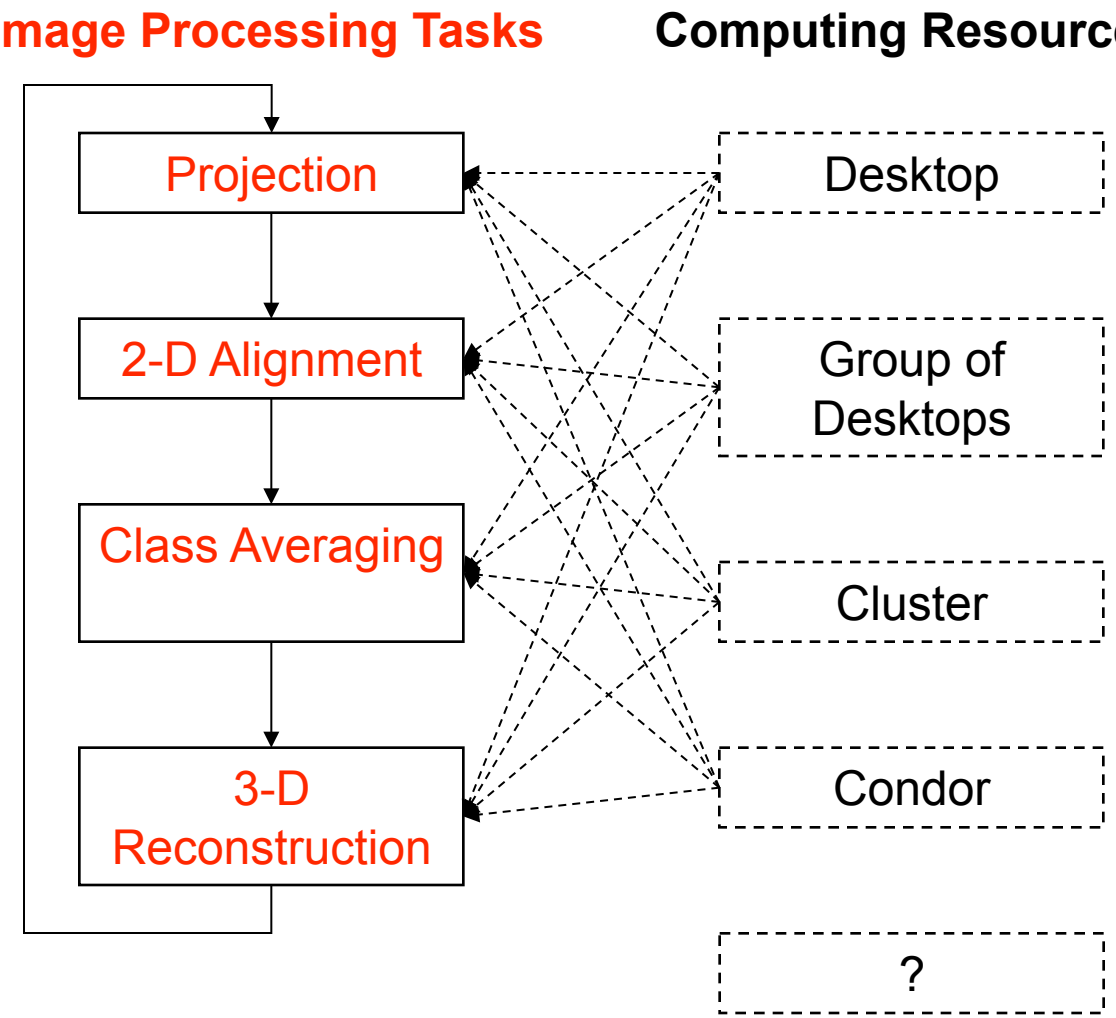


Condor Is Growing

	Total	Owner	Claime
IA64/LINUX	4	0	0
INTEL/LINUX	818	346	218
INTEL/WINNT51	**5142**	**2887**	**1**
INTEL/WINNT52	1	0	0
INTEL/WINNT60	**2373**	**250**	**0**
PPC/OSX	1	0	0
PPC64/LINUX	8	0	0
SUN4u/SOLARIS28	1	0	0
SUN4u/SOLARIS5.10	15	5	0
X86_64/LINUX	**10739**	**7738**	**2857**
Total	**19102**	**11226**	**3076**



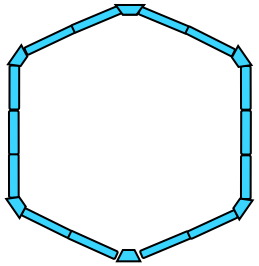
Computing Resources for Image Processing

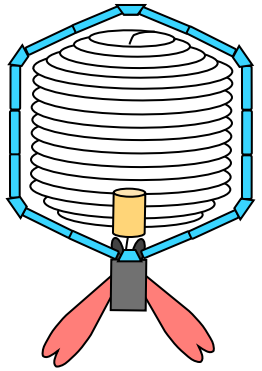


Our dream:

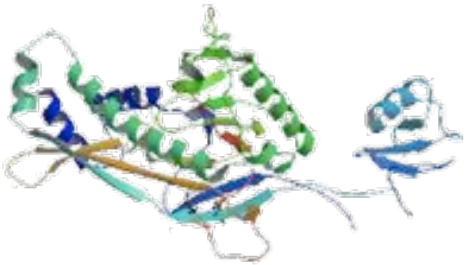
Super-computer scale resource for complete image processing from individual's desktop without worrying about complex computer issues for a biologist user

Tailed dsDNA Phage: Atomic Models

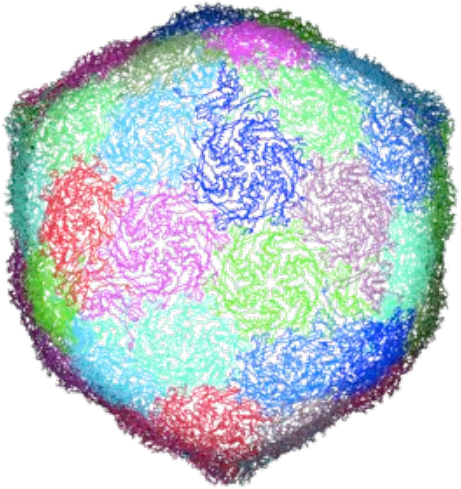




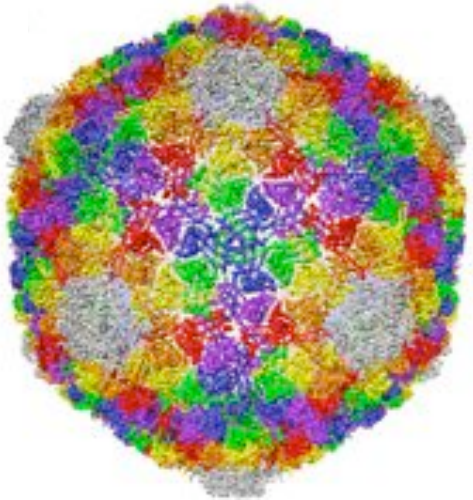
T4:
vertex protein
(2005)



HK97:
recombinant shell
(2000)

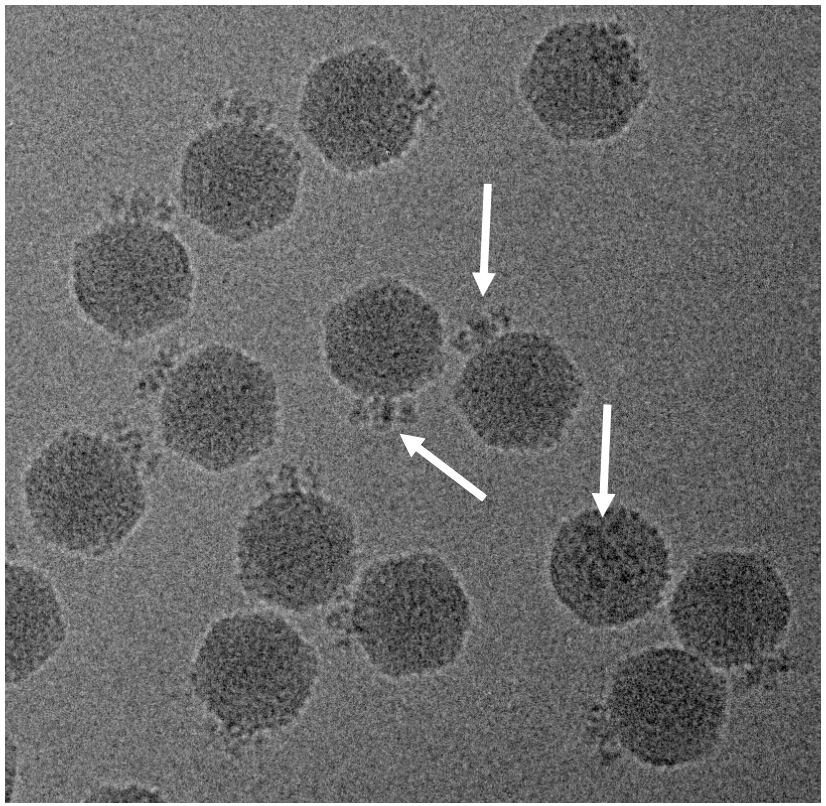


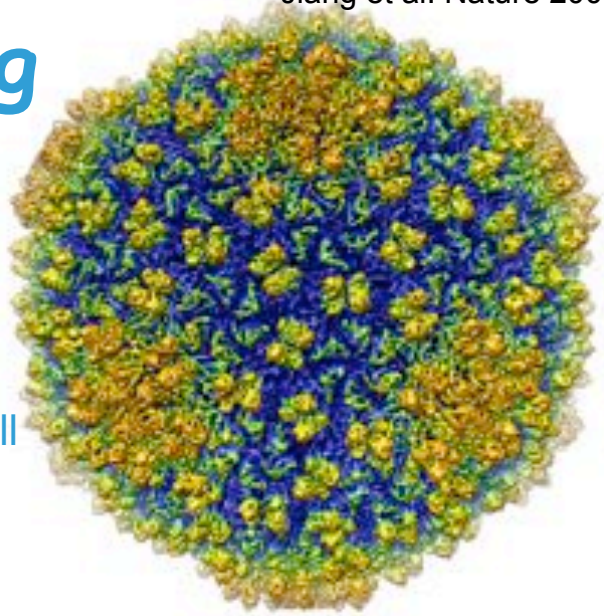
Infectious Phage
(2008)



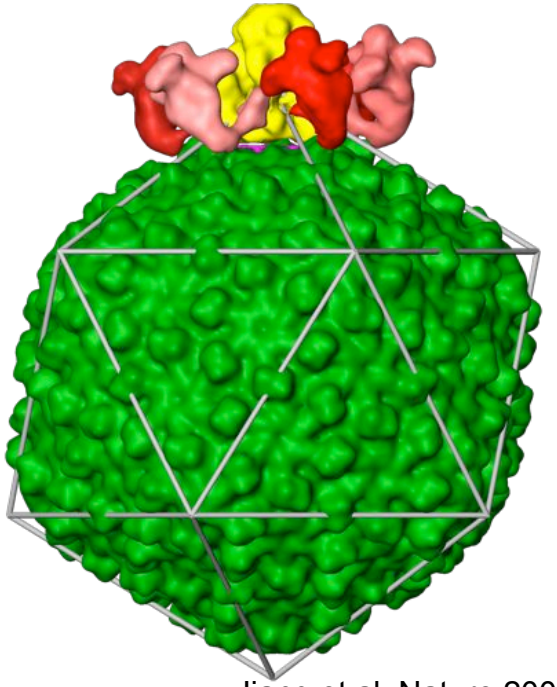
Beyond Icosahedral Averaging

Jiang et al. Nature 2008





Icosahedral shell (4.5 Å)



Complete structure (20 Å)

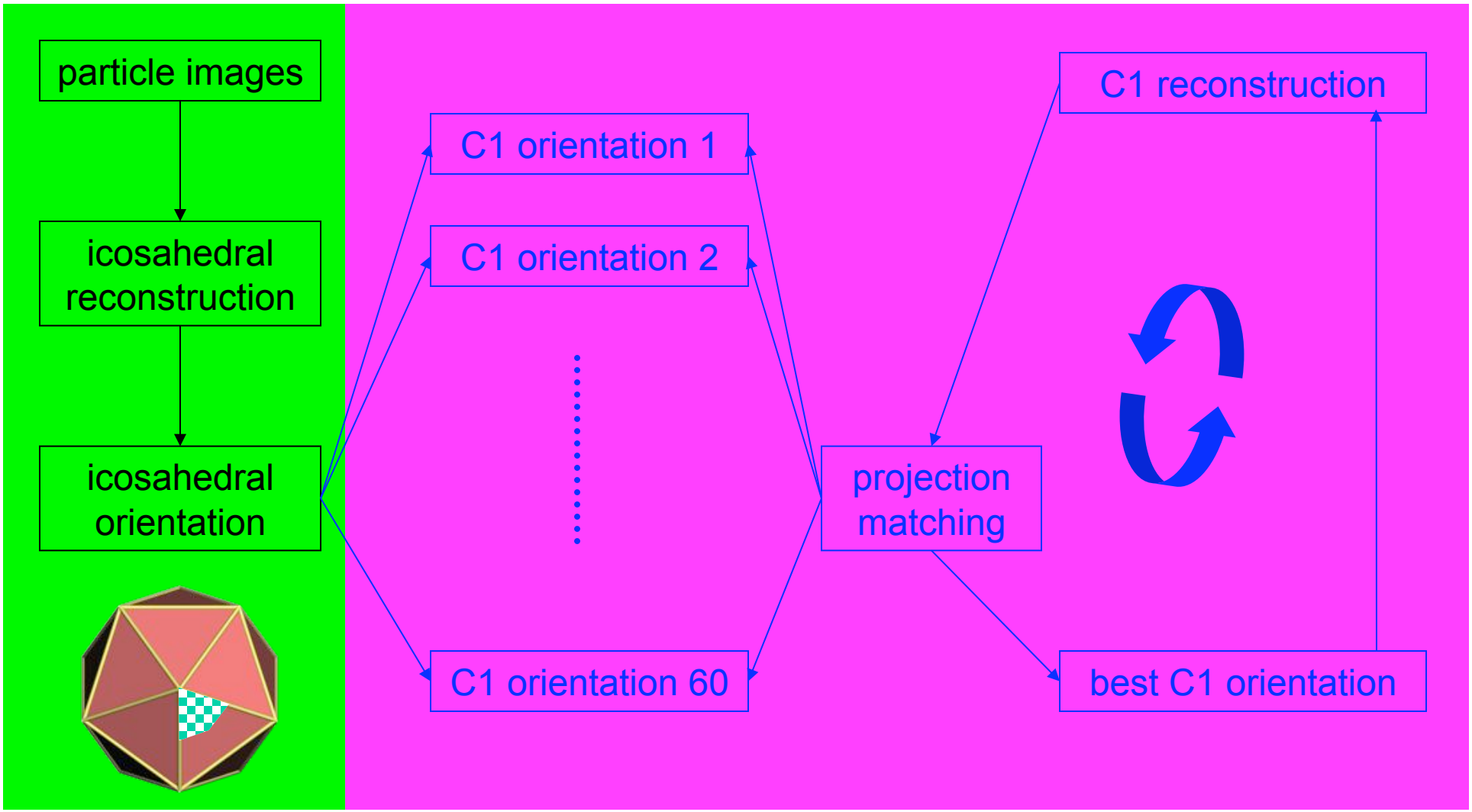
Jiang et al. Nature 2006

Can we solve the structure of non-icosahedral components: portal, tail, genome ?

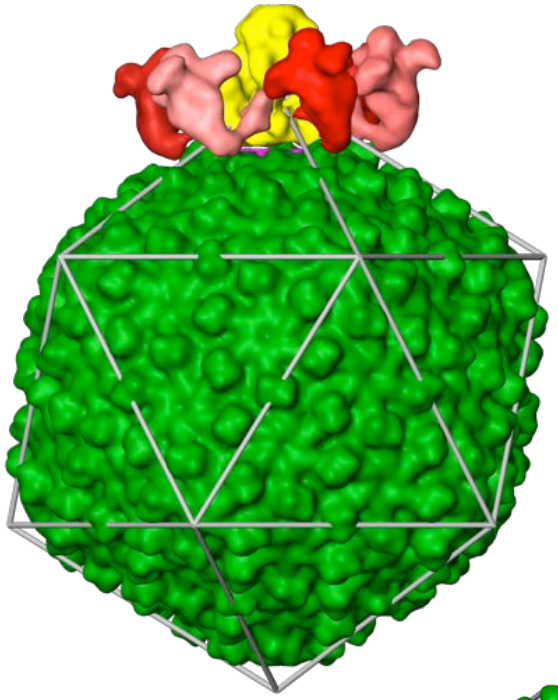
Icosahedral Reconstruction ➔ Non-Icosahedral Reconstruction

icosahedral

asymmetric reconstruction

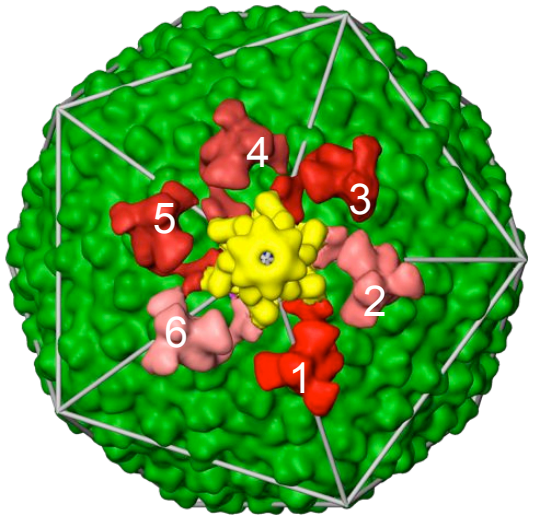


Asymmetric Reconstruction

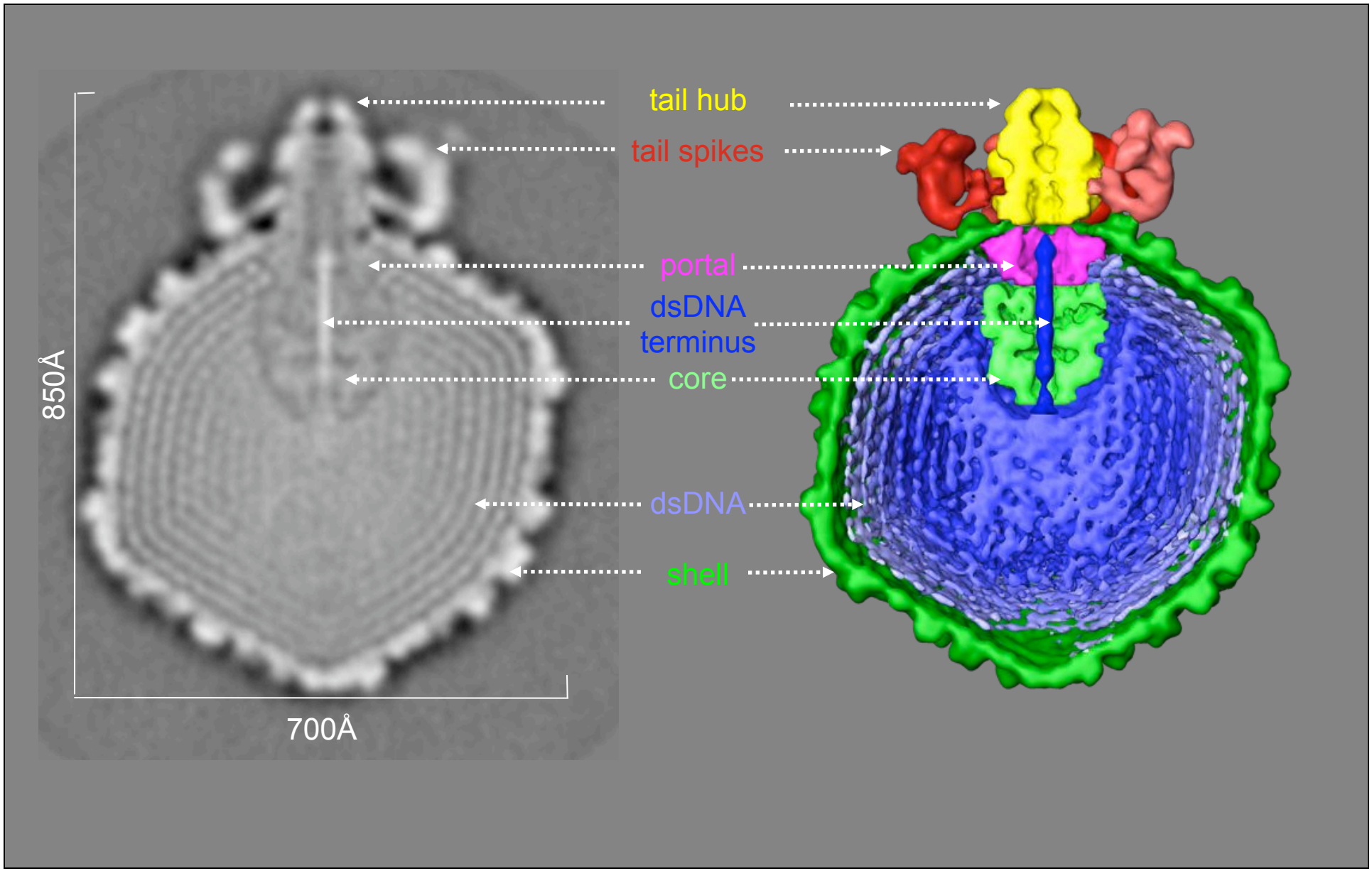


side view

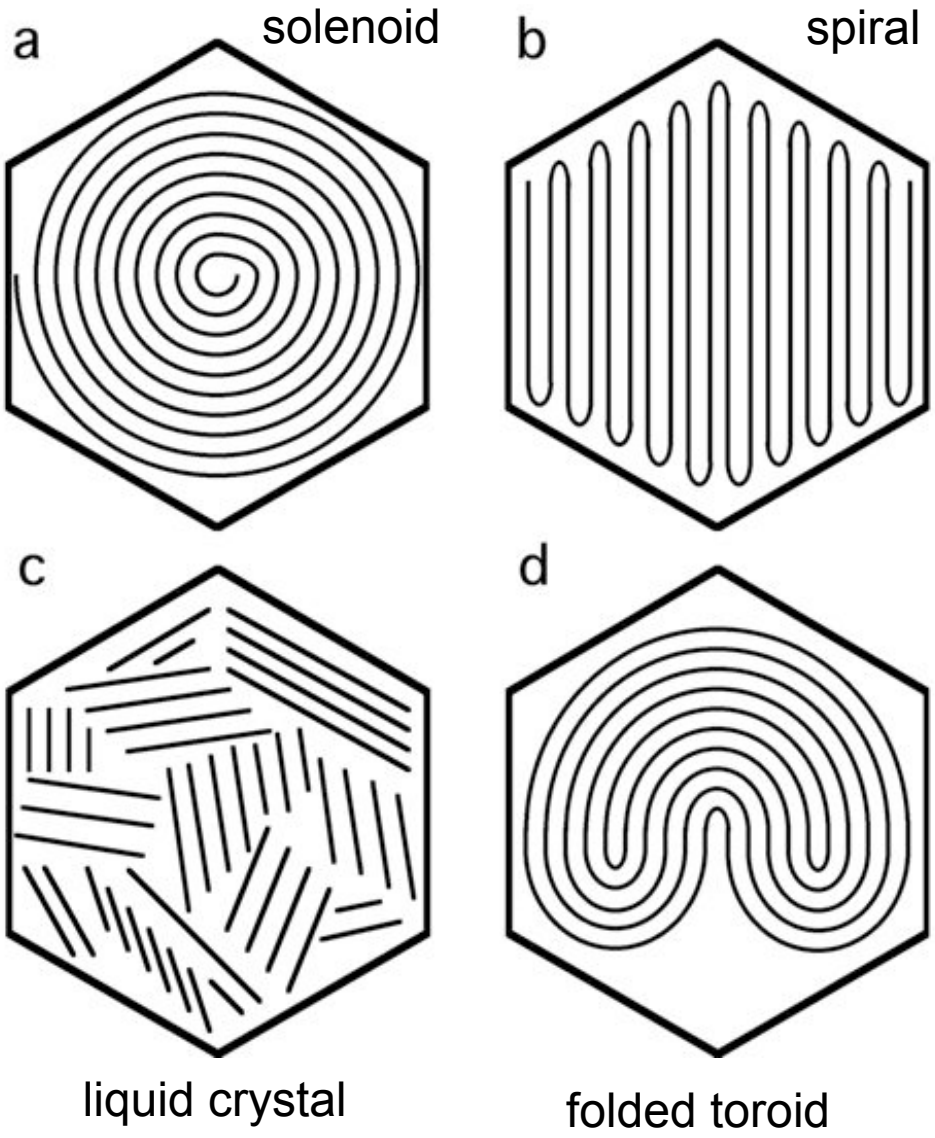
top view

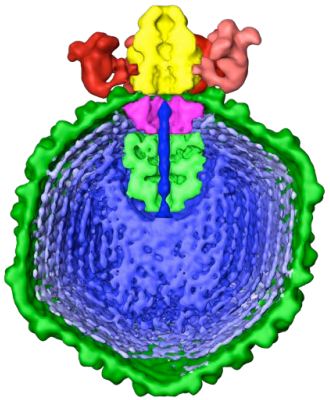


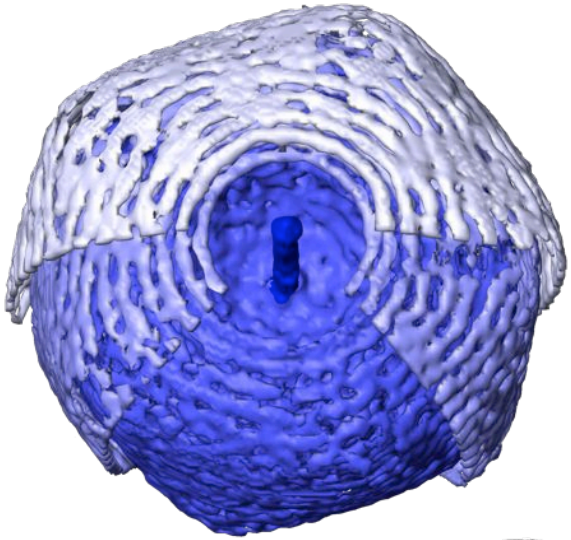
"Virus Anatomy"

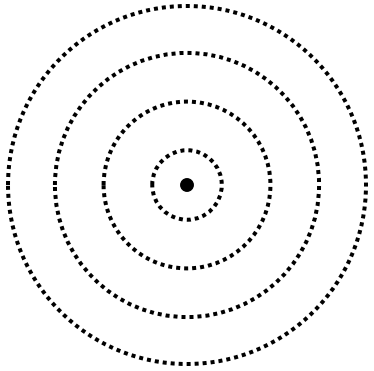


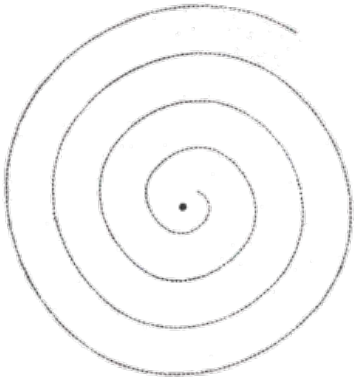
dsDNA Packing: Coaxial Spooling



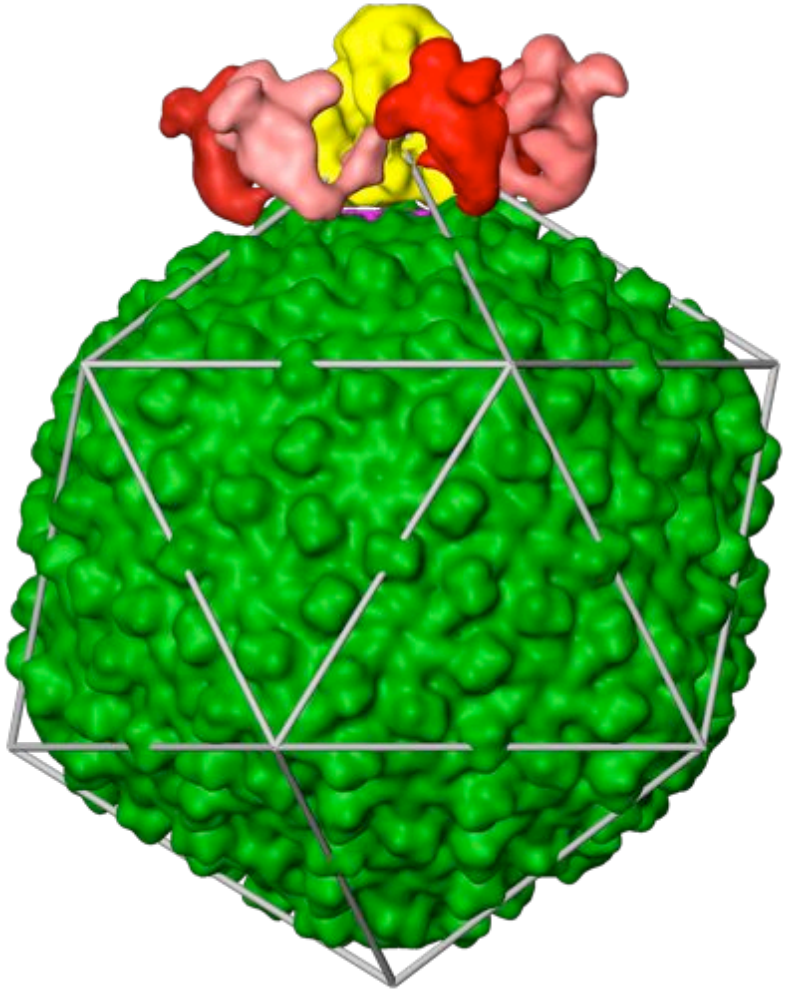




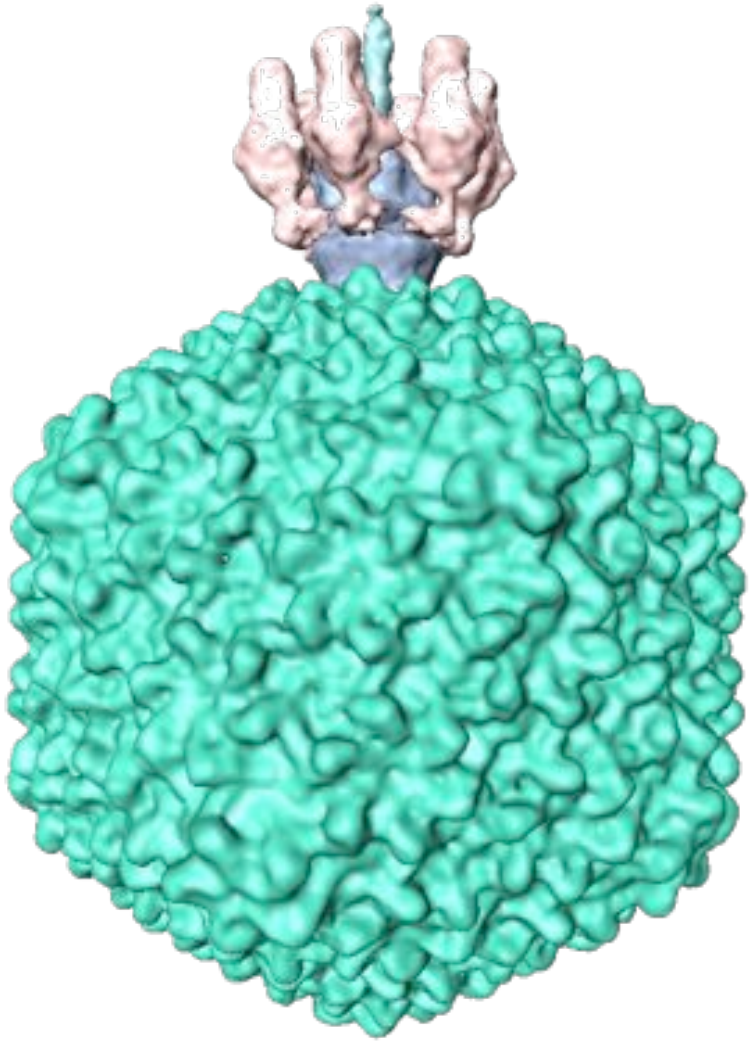




Epsilon15 vs. P22



Epsilon15



P22

Acknowledgments

Purdue

Weimin Wu
Ping-An Fang
Min Su
Teepanis Chachiyo
Rosen Center of Advanced Computing

Baylor College of Medicine

Wah Chiu
Matthew Baker
Jonita Jakana
Qinfen Zhang
Steve Ludtke
Juan Chang

MIT

Jonathan King
Peter Weigele

Funding

Purdue, NIH/NIAID