Toward Atomic Resolution CryoEM with Bioinformatics

Hong Zhou, Ph.D.
University of Texas - Medical School at Houston
Pathology & Laboratory Medicine

4/20/2006, Houston, TX

Outlines

- Introduction of our model system dsRNA virus and cytoplasmic polyhedrosis virus (CPV)
- High resolution imaging of CPV
- Data processing, software development
- CPV structures at 5 Å resolutions
- Atomic modeling by integrative bioinformatics approach

Push the limit: dsRNA viruses as models

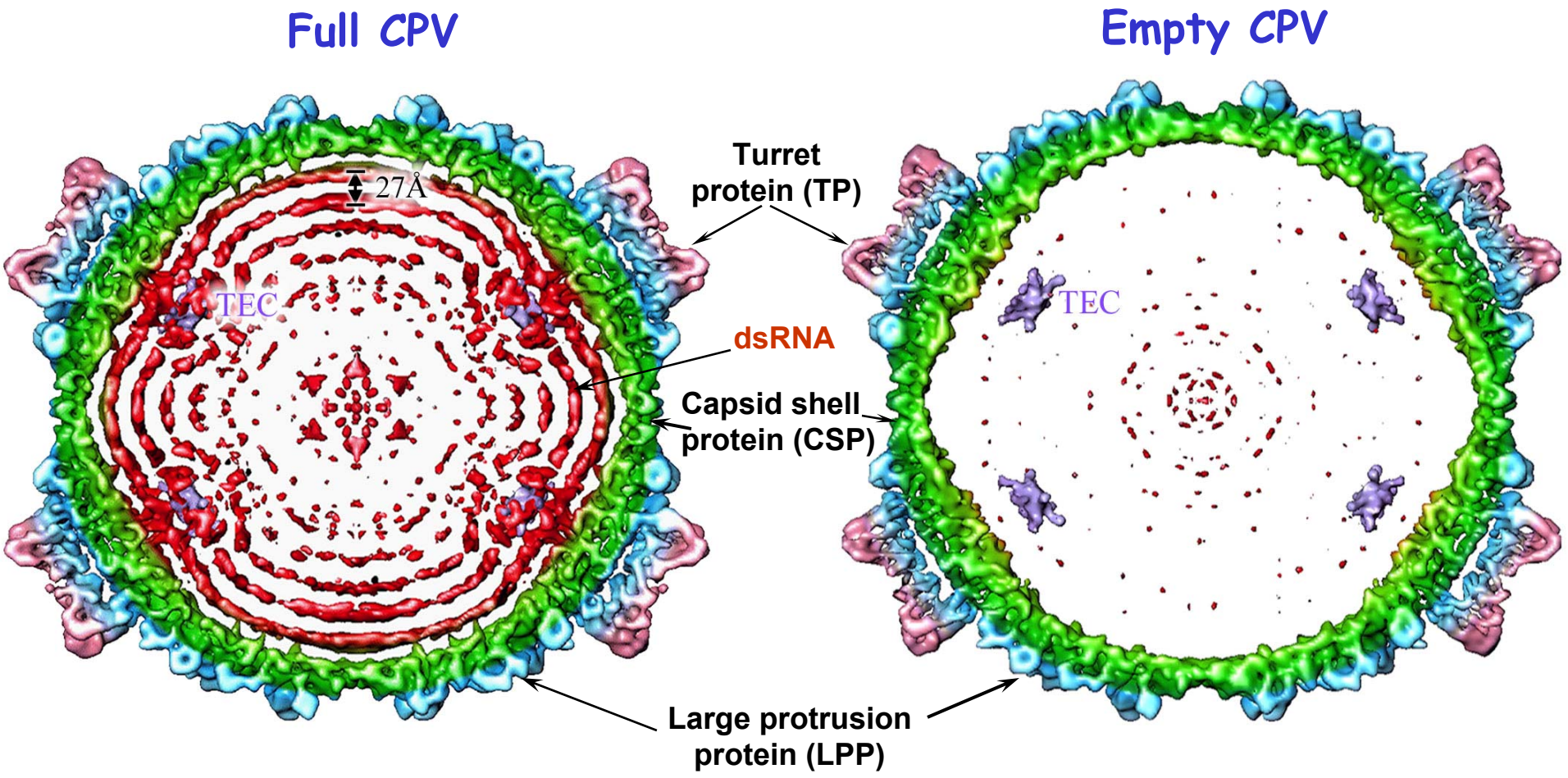
Genus	Host ranges	3D structure	# RNA segments	# Structural Proteins	Capsid size (Å)
Orthoreovirus	Mammals	X-ray & cryoEM	10	8	850
Rotavirus	Mammals/birds	cryoEM	11	6	**800/1050**
Orbivirus (BTV)	Mammals/insect vectors	X-ray & cryoEM	10	7	800
Phytoreovirus (rice dwarf virus)	Plants/insect vectors	cryoEM & X-ray	12	12	780
Cypovirus (CPV)	**Insects**	**cryoEM**	**10**	**5**	**590/800**
Fijivirus	Plant/insect vectors	no	10	11	700
Coltivirus	Mammals,invertebrate	no	12	12	800
Aquareovirus	Bony fish, crustaceans	cryoEM	11	7	800
Seadornavirus	Mammals	no	12	12	800
Mycoreovirus	fungus	no	11	10(?)	800(?)
Oryzavirus	Plant,invertebrate	no	10	10	700

Capable of endogenous mRNA transcription, capping & efficient release

Facts of CPV

- Single-shelled capsid, yet very STABLE
- Fully capable of endogenous transcription, mRNA capping and release within intact virus
- Used as a bio-control agent, an environment-friendly pesticide

Structural Organization of the CPV (13 Å)



TEC: Transcriptional Enzyme Complex

CryoEM imaging of CPV

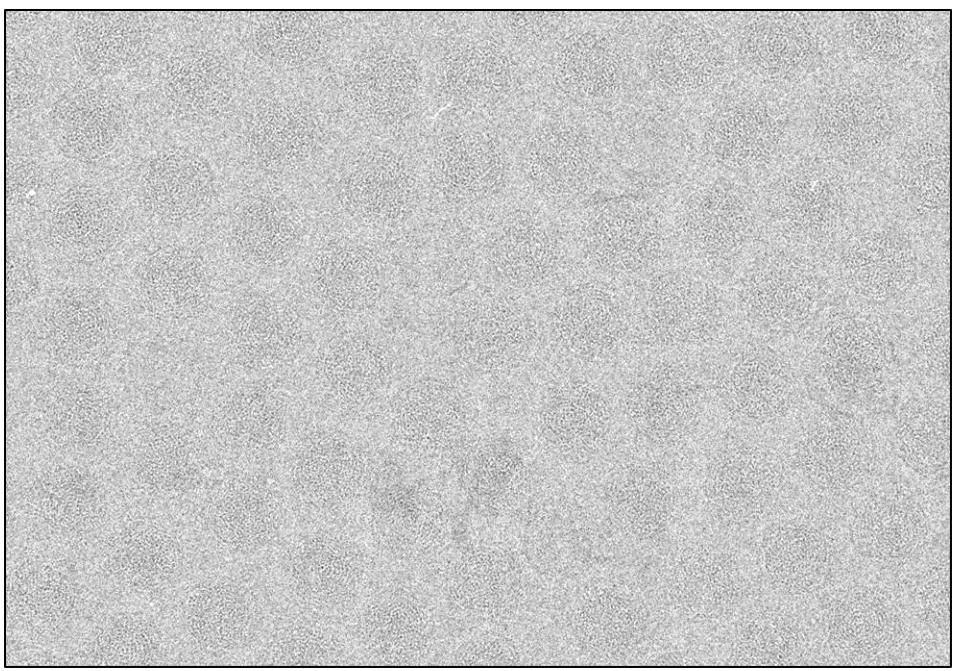
- Liquid helium-cooled specimen (4 K) (JEOL)
- Gold aperture
- Fully "Baked" *Quantifoil* holey grids
- 300kV, field emission gun (FEG)
- Kodak SO163 films at 60,000 x
- Focal pair, combined dose about 50 $e^-/Å^2$
- First micrograph < 1 µm defocus to improve accuracy of CTF correction
- Second micrograph, crucial, but only used in initial processing stage

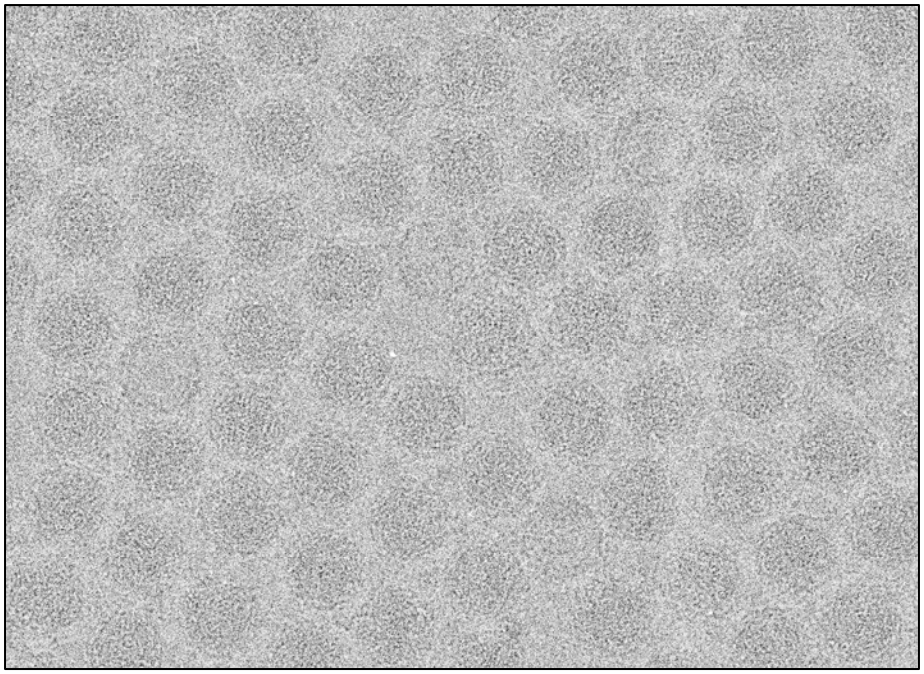
JEOL FEG 300 kV, liquid helium-cooled
National Center for Macromolecular Imaging, Baylor College of Medicine

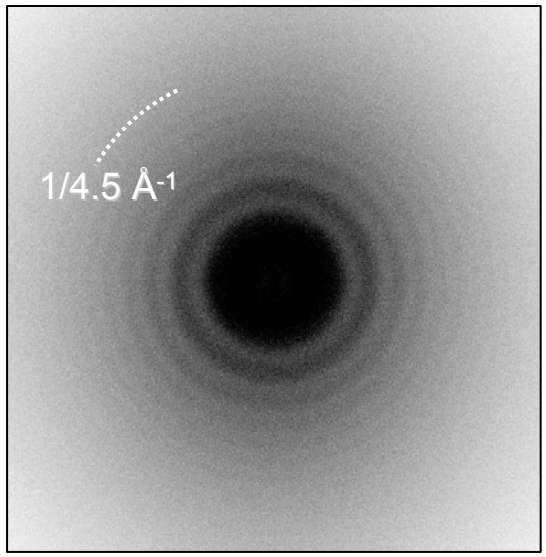
400kV JEOL4000

300kV FEI Polara G^2 F30 liquid helium (UT)

Image evaluation: incoherent averaging of Fourier transforms

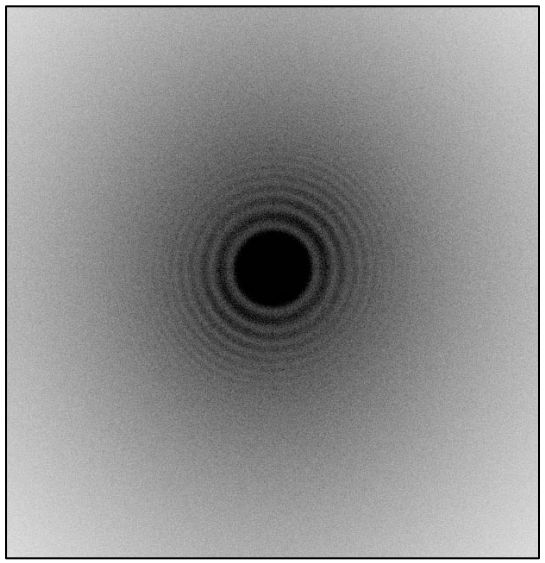






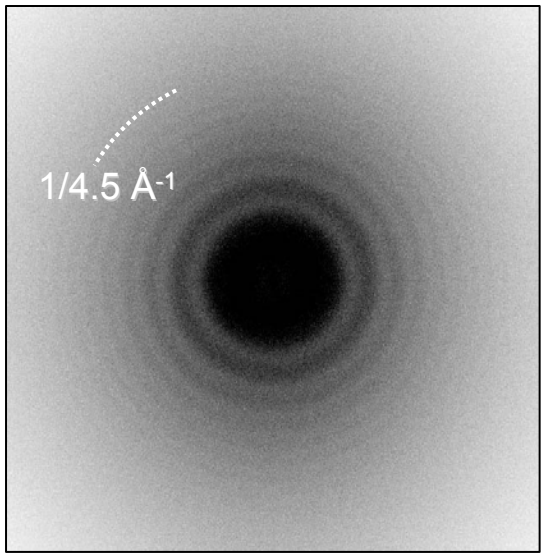
First micrograph

Incoherent average of particle Fourier transforms

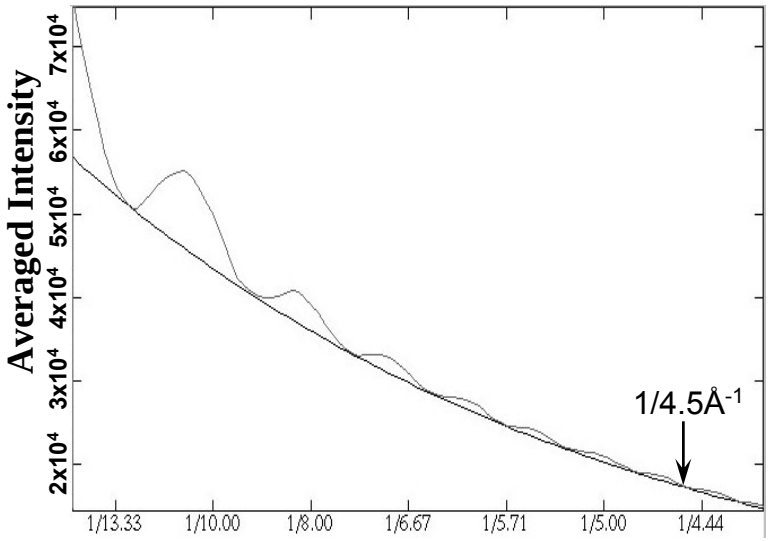


Second micrograph

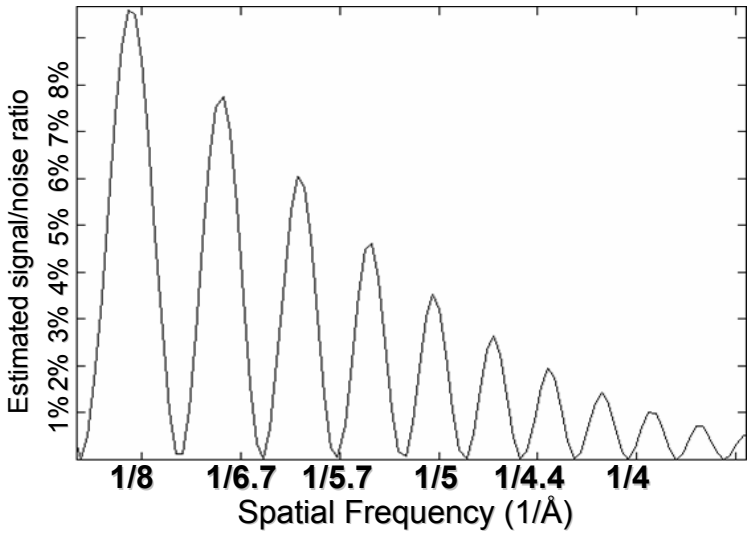
Defocus and "B" factor: exponential decay of data

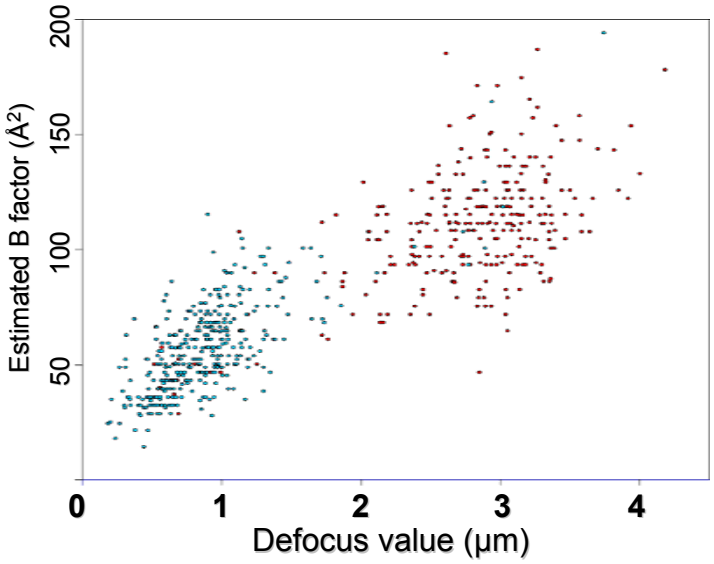


Incoherent average of FTs



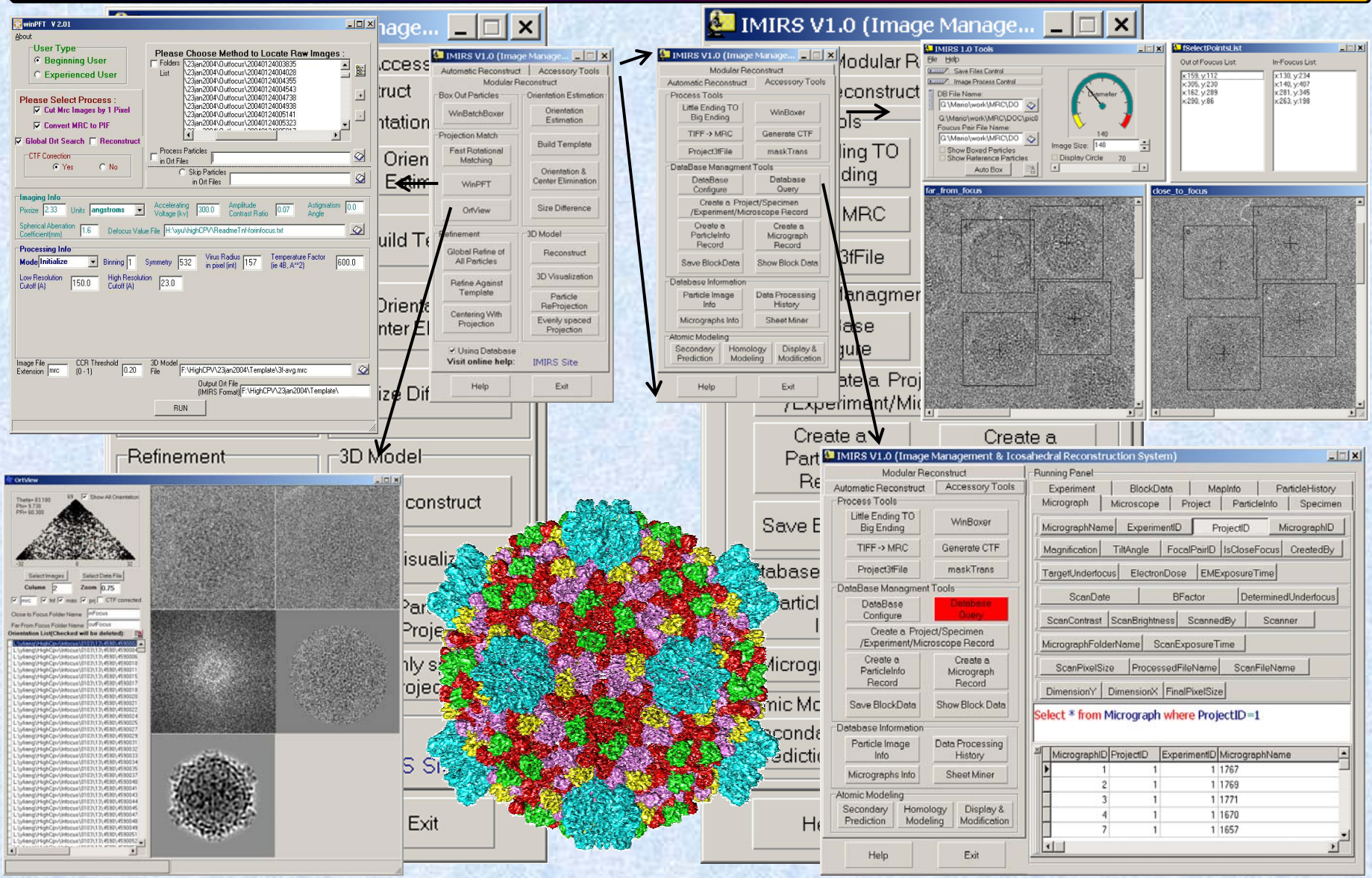
- Tens or even hundreds of thousands of particle images might be needed toward 4-5 Å resolution





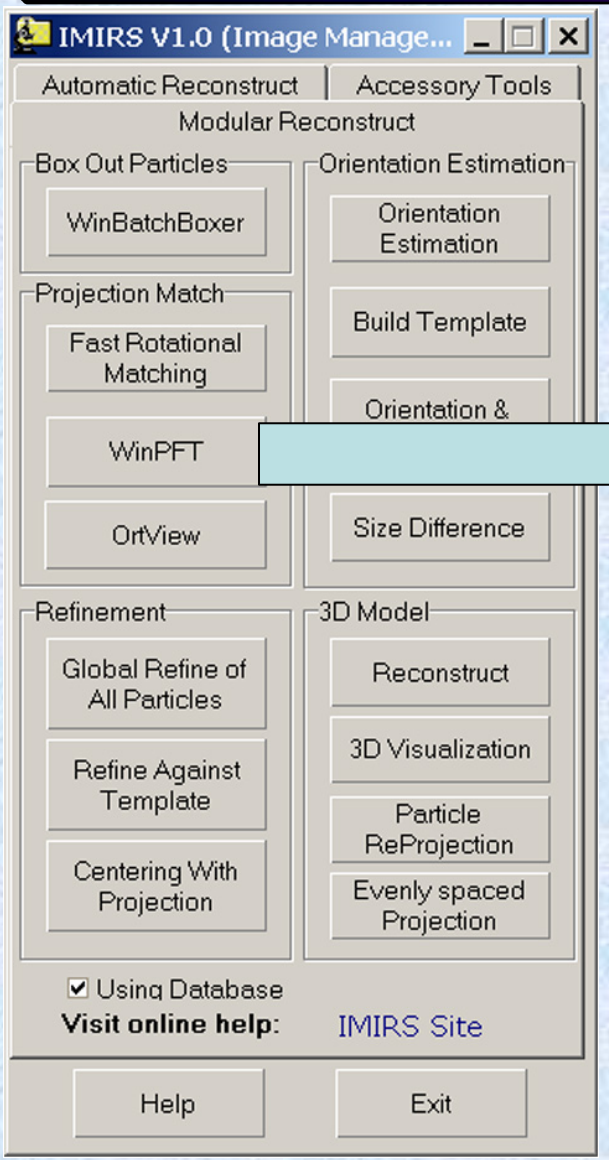
IMIRS: an integrative and modular approach

- Integrated Management & Icos. Reconst. System



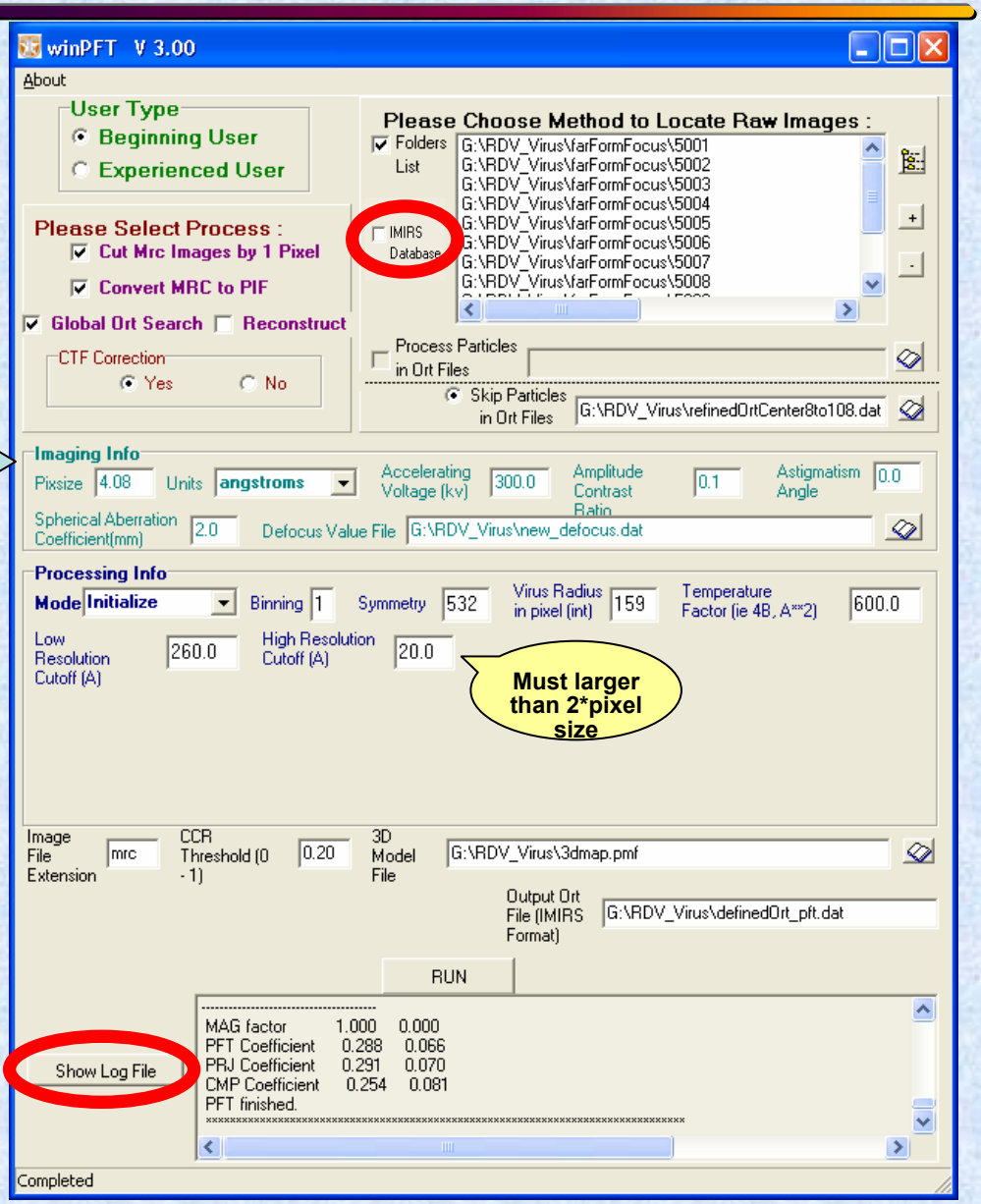
http://hub.med.uth.tmc.edu/~hong/IMIRS

Engineering efforts of IMIRS
- Partially compatible with Pam's rules



PFT by Baker *et al.*

Demo

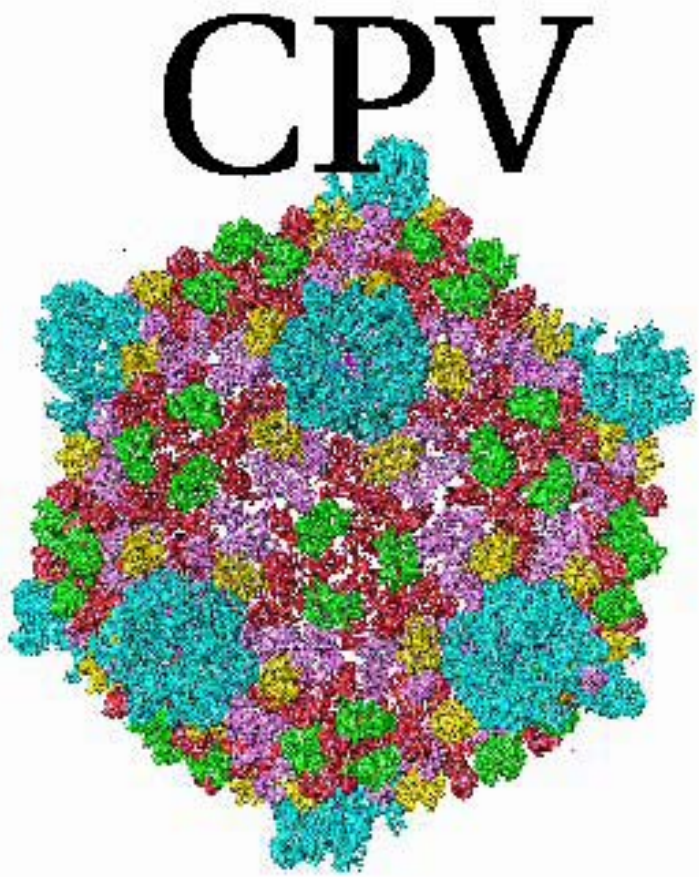


Summary of data processing statistics

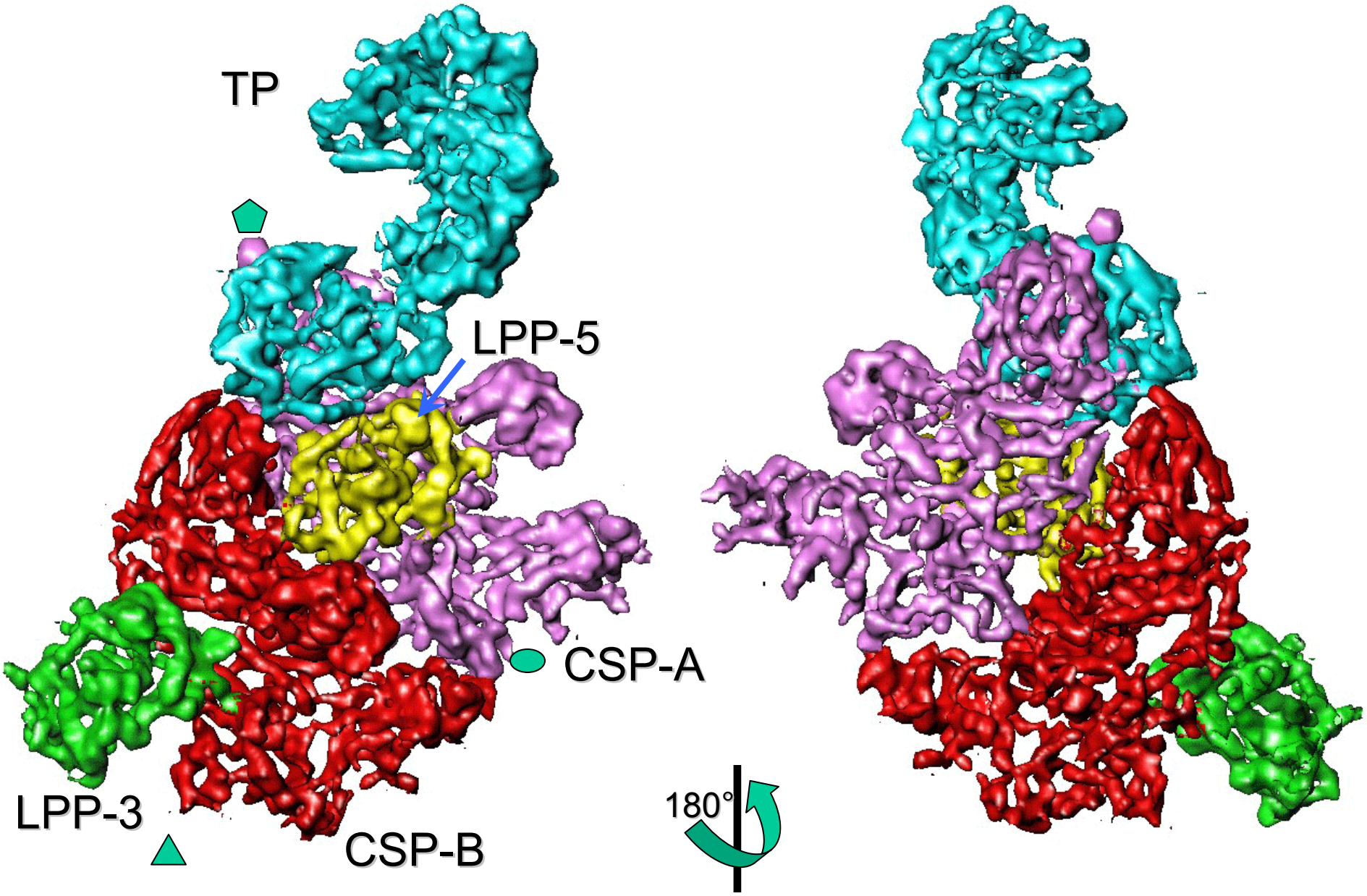
- Number of focal pairs scanned: >1,000 pairs
- Number of focal pairs refined: 646 pairs
- 1.16Å/pixel, 800x800 particle
- Number of particles processed: 135,000

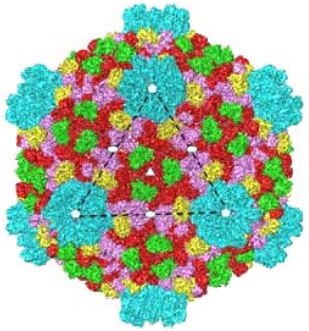
- Defocus ranges:
  1.9-3.7 μm and 0.2-1.7 μm for 1st & 2nd micrographs
- B factor: 100-210 and 40-140 $Å^2$ respectively
- Final reconstruction
  25,705 particle images used, all close-to-focus
  refined to 1/3.5 $Å^{-1}$
  effective resolution 5.2 Å
- Total averaging is about 1.5 million (25,705 x 60)

CPV Capsid Shell at 5.5 Å

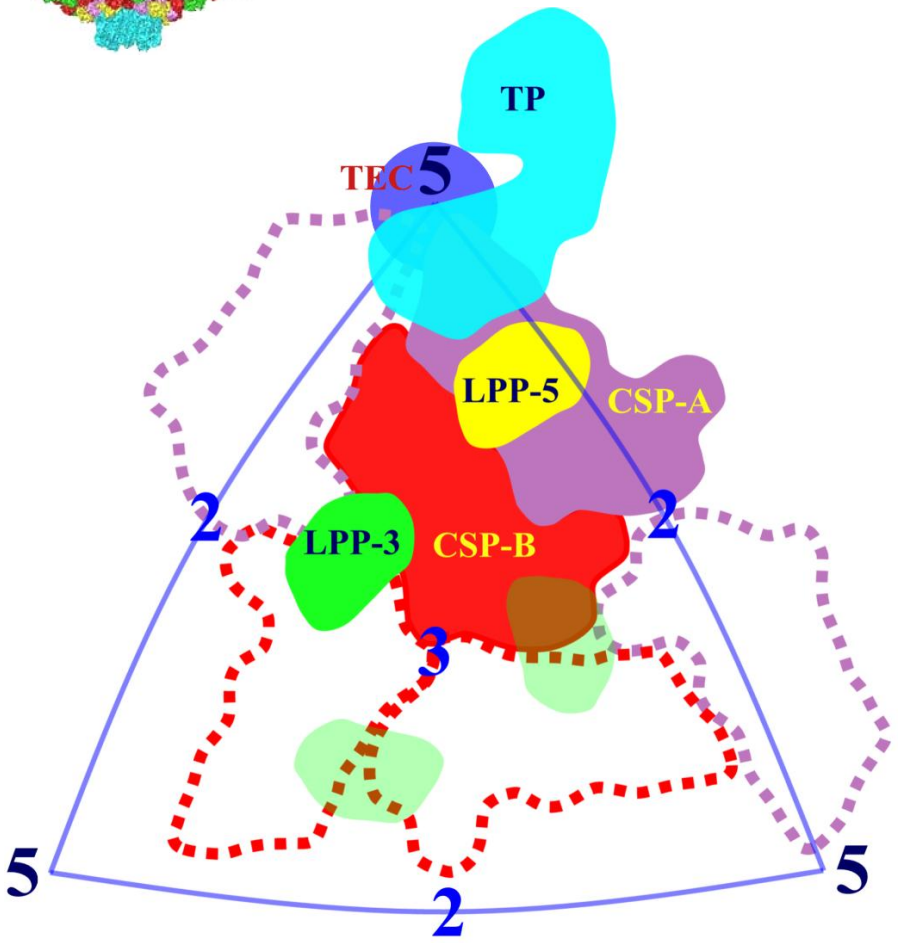


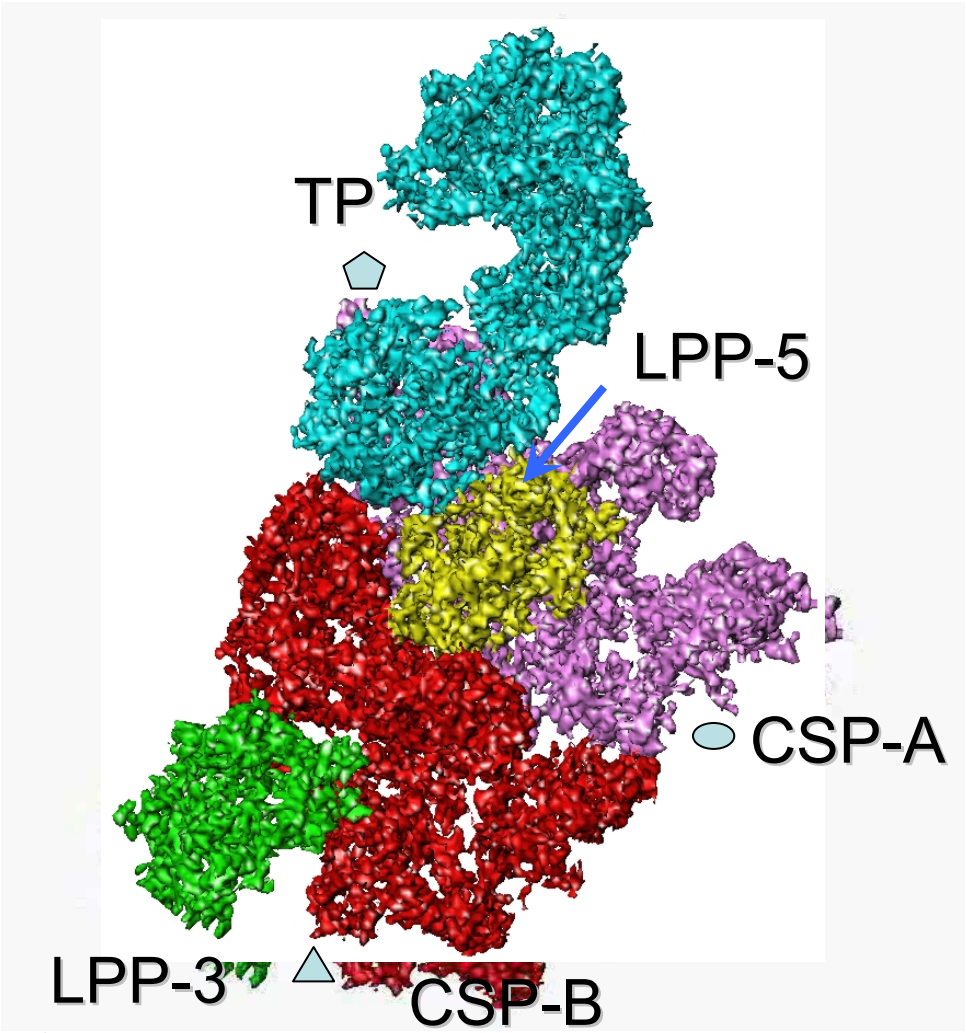
Asymmetric Unit: Molecular Interactions





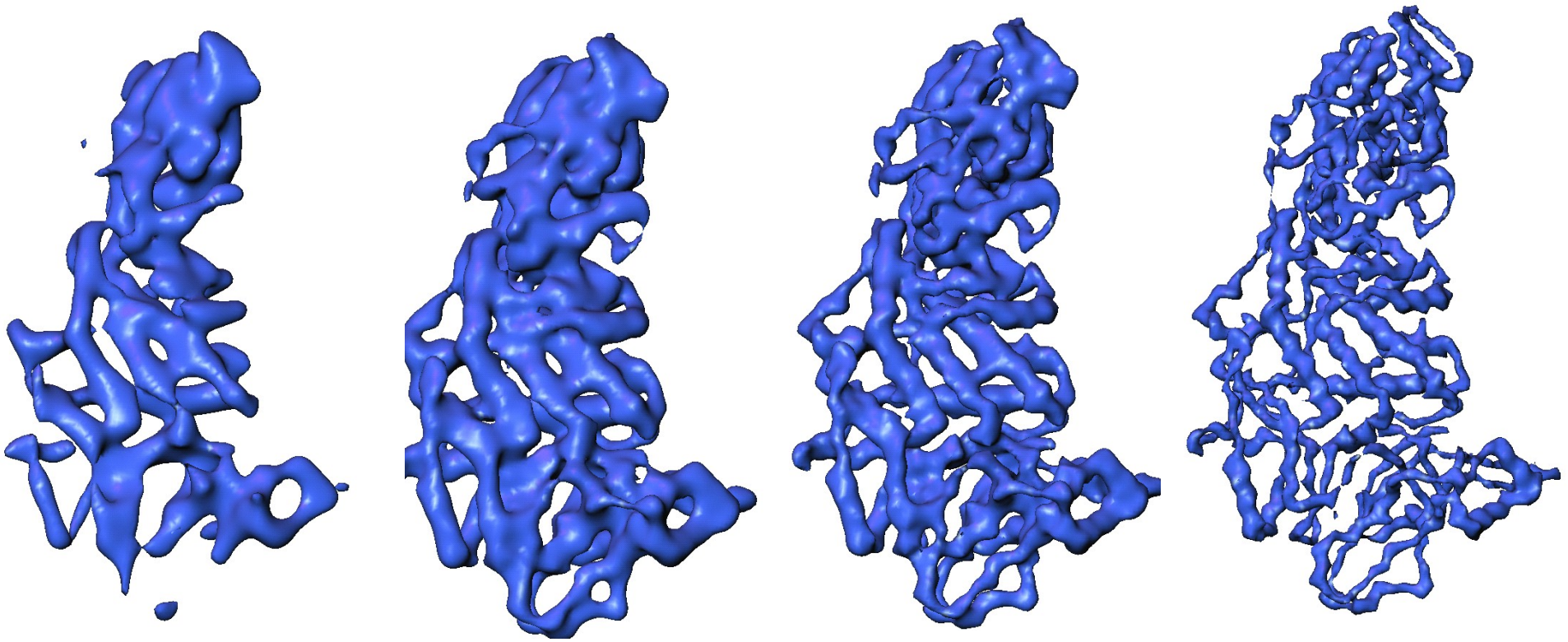
Asymmetric Unit of CPV





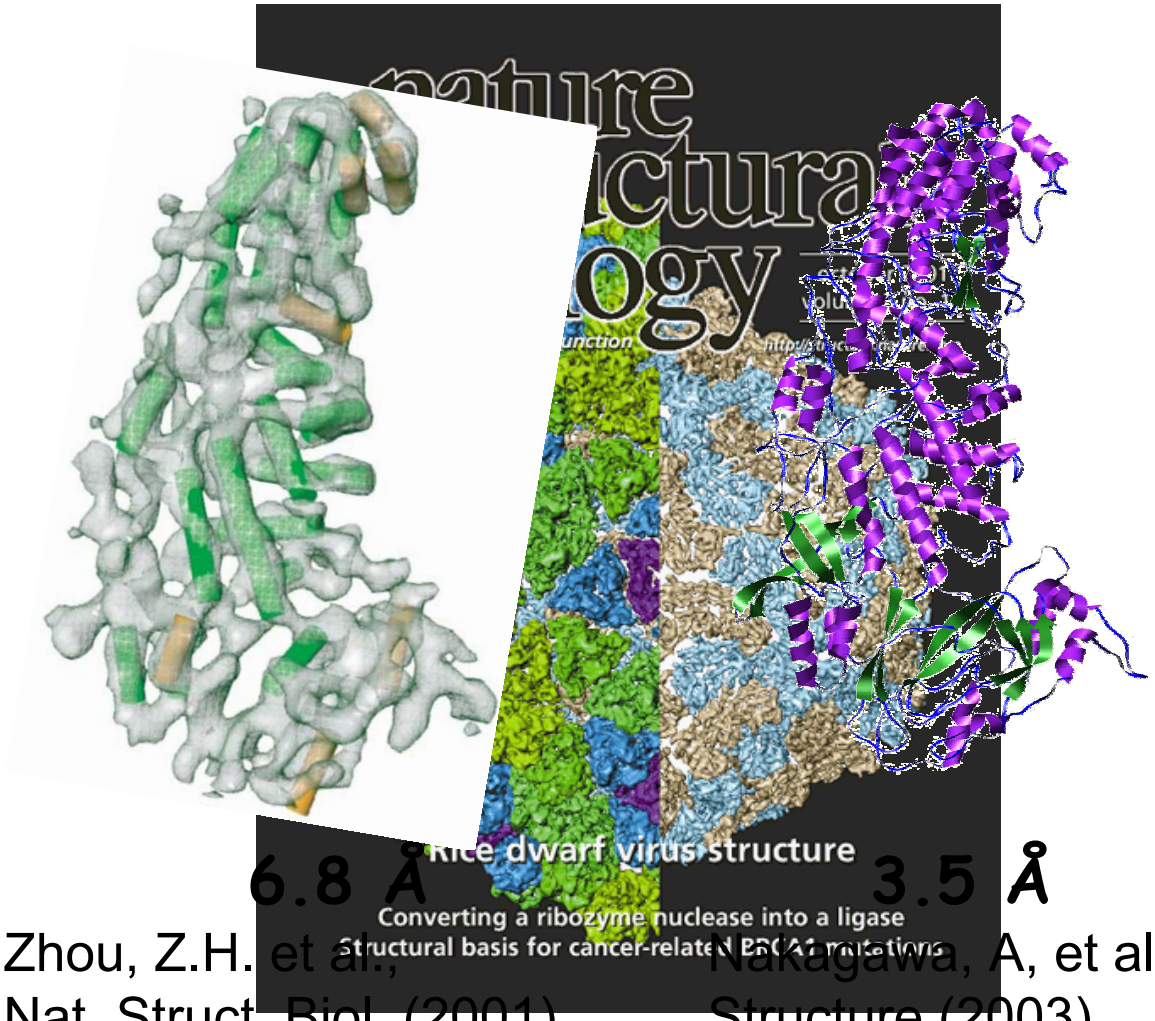
- Intensive molecular interactions
- Molecular clamps

BTV CSP density map at different resolutions



8 Å 6.5 Å 5.5 Å 4.5 Å

Simulation: What Can We See at 5-Å? (RDV: Rice Dwarf Virus)



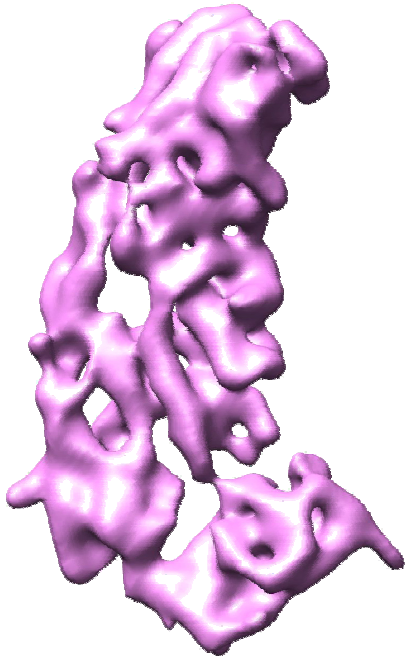
6.8 Å

Zhou, Z.H. et al.,
Nat. Struct. Biol. (2001)

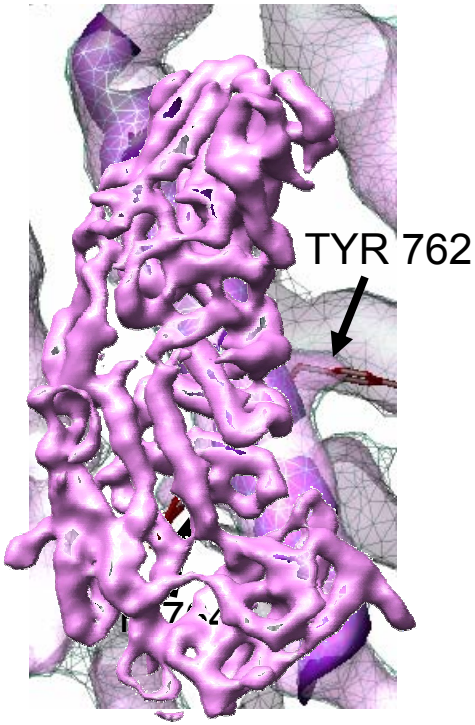
3.5 Å

Nakagawa, A, et al.,
Structure (2003)

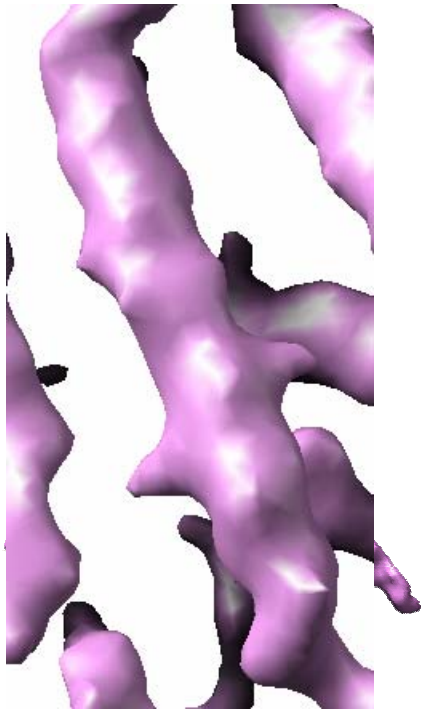
Bulky Side Chains Can Be Resolved at 5Å



8 Å



6 Å



5 Å

Bulky side chains of TYR, TRP, PHE, etc can be resolved at 5 Å

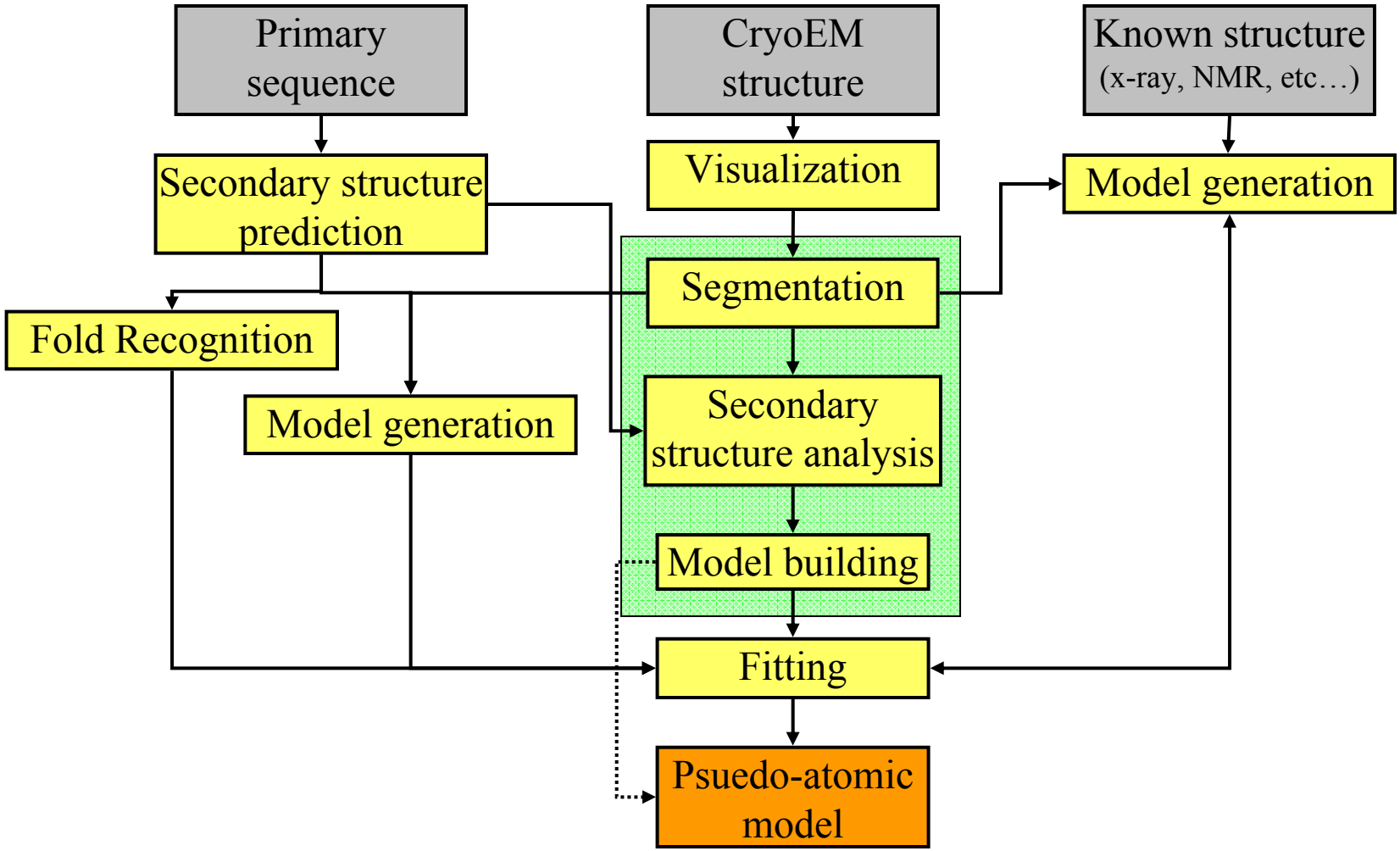
Approaches to Atomic Modeling of CryoEM Structures

***Motivation*: bottom-up approach (O, MAID, X-Build etc) *NOT* applicable to near-atomic resolution cryoEM maps**

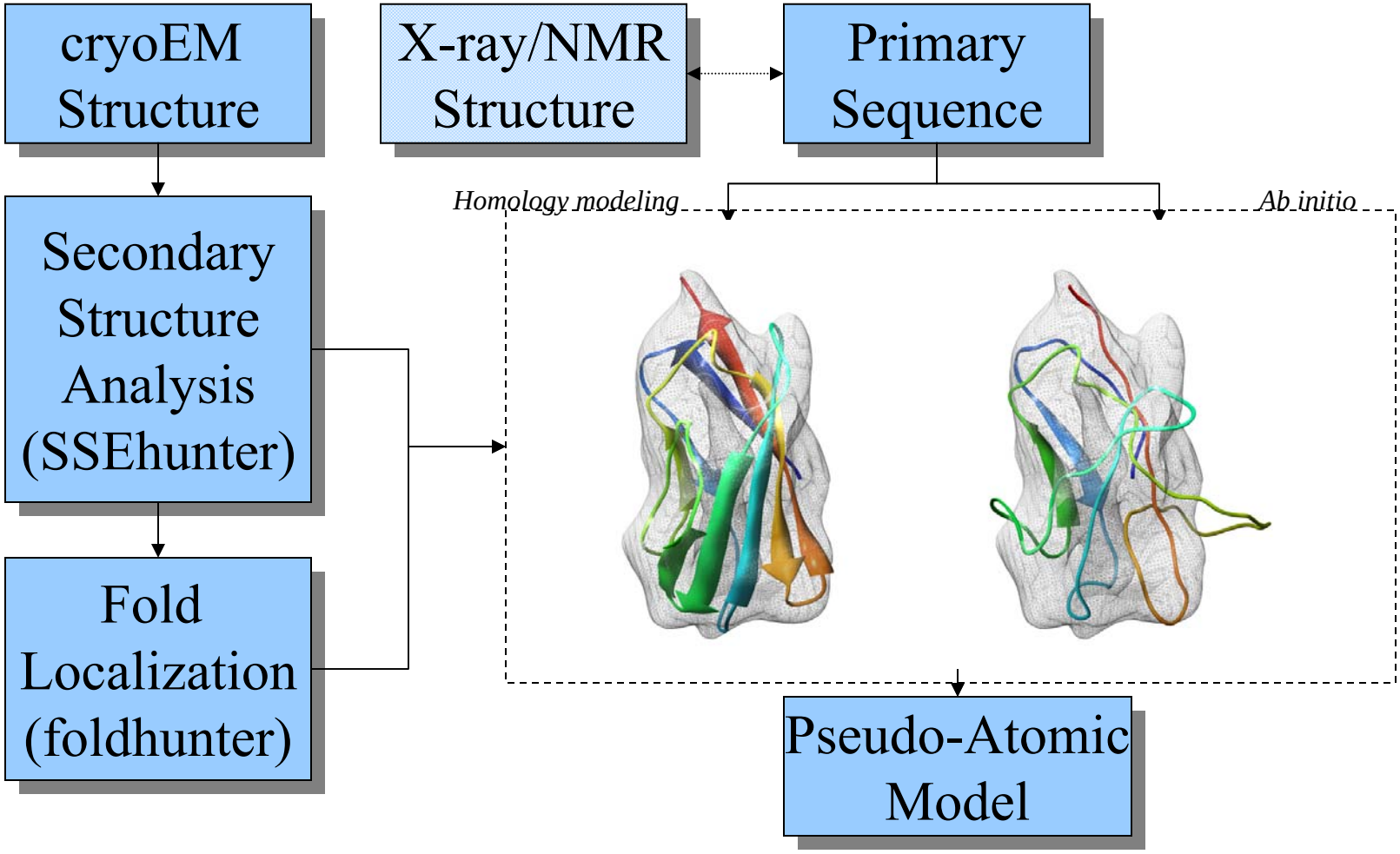
Our approach: top-down and integrate all available knowledge

1. Structural Analysis
   - Identification of SSE w/AIRS (M. Baker)
2. Sequence Analysis
   - Homologue identification
   - Template identification
   - Secondary structure prediction
3. Model Building
   1. SSE assignment
      - Consensus sse assignment
   2. Homology modeling
      - Accurate template models
   3. *ab initio* modeling
      - Domain size limitations

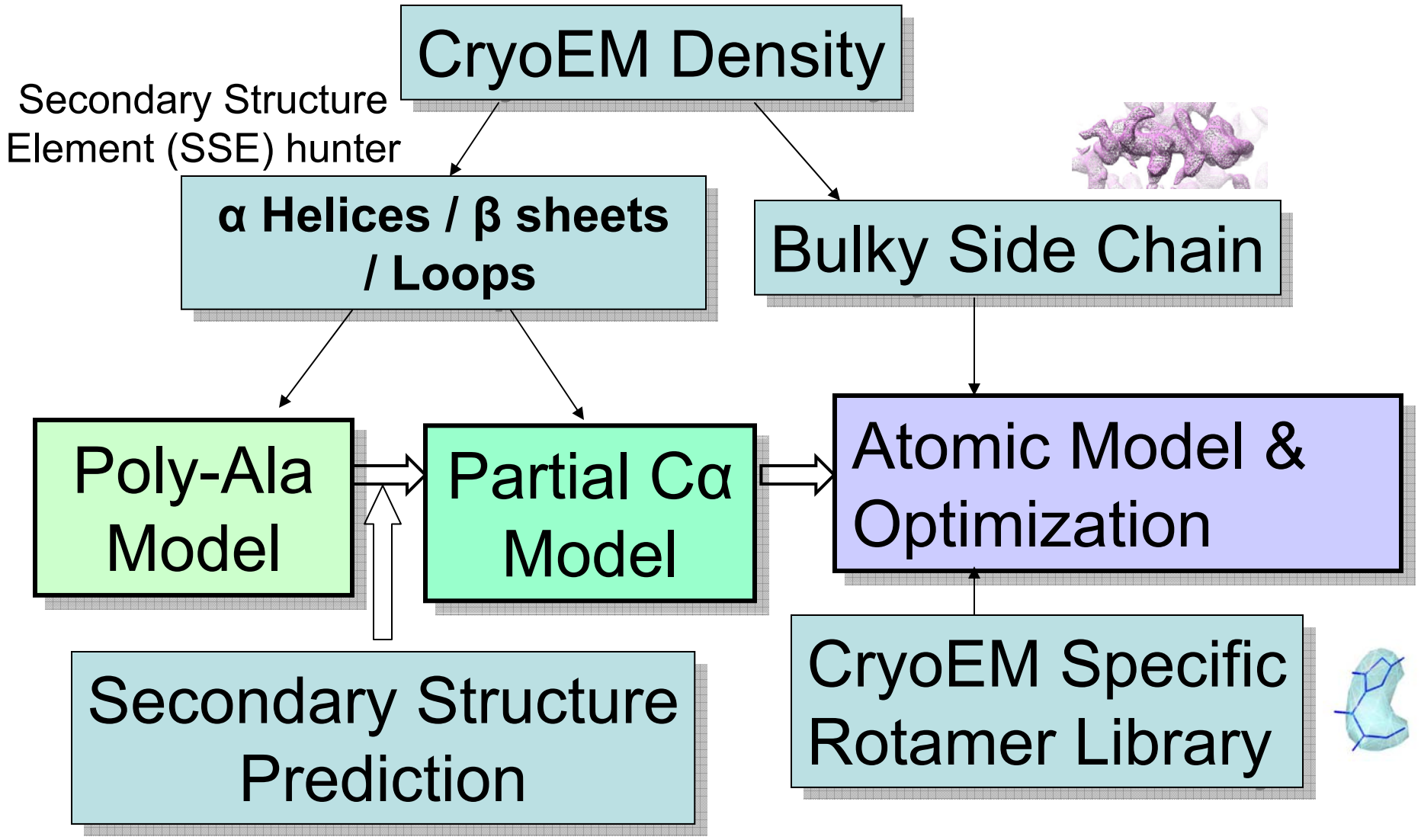
Our Generic Modeling Building Tools



Constrained Modeling



CPV Model Building Protocol



Summary

- Reconstructing CPV to 5.2 Å (including data to 3.5 Å)
- Secondary structure elements and **bulky amino-acid side chains** are resolved
- Use of an integrative modeling method to build Cα models
- Partial atomic model for regions with bulky side chains

Biology: helix as a regulating switch

Acknowledgements

University of Texas Medical School at Houston

Yuyao (Mario) Liang,

Xue-Kui Yu,

Hua Tsen

Ivo Atanasov

Baylor College of Medicine

Wah Chiu,

Joanita Jakana,

Matthew Baker

Zhongshan University, China

Jing -Qiang Zhang

Qinfen Zhang