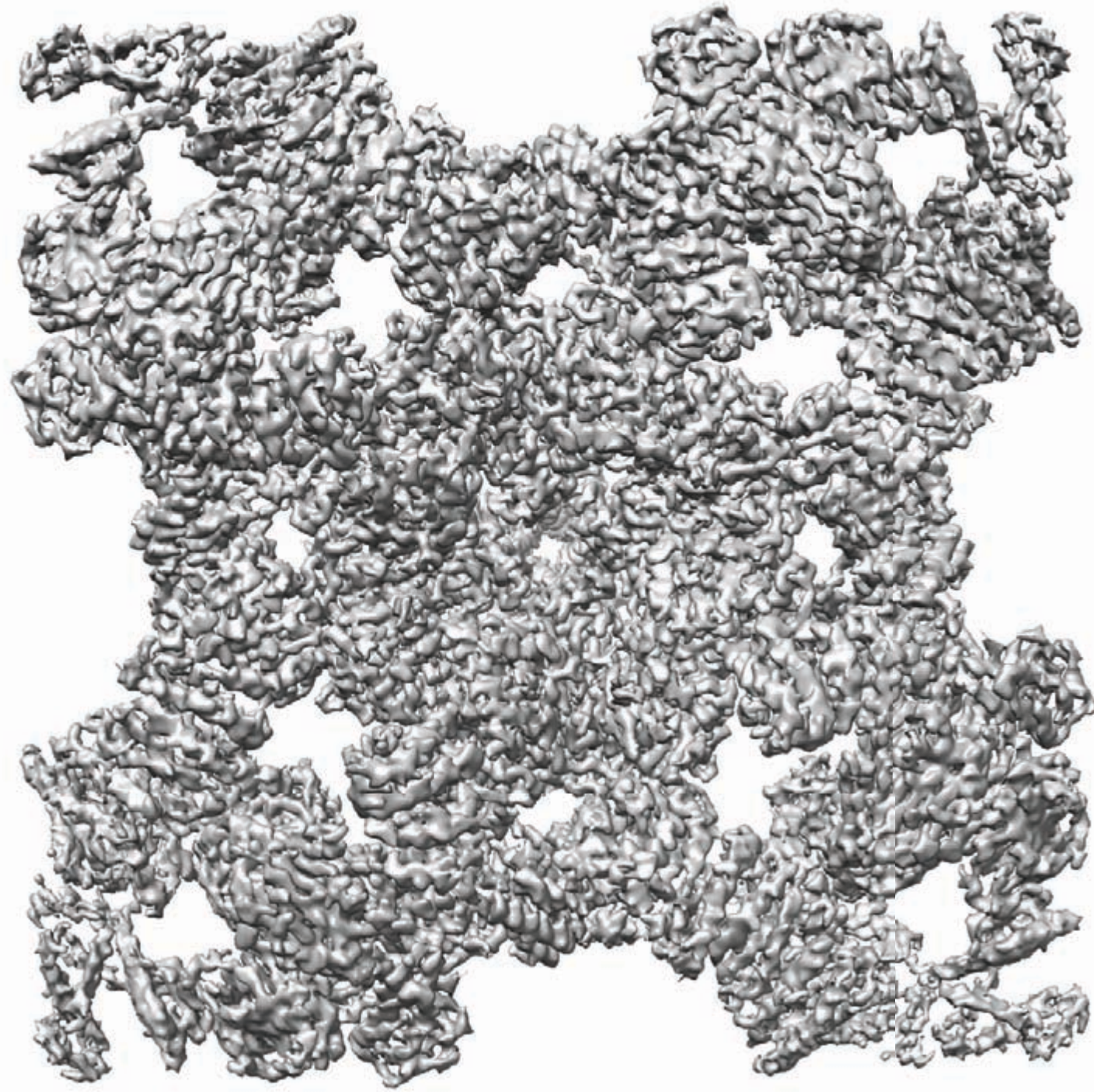


UTMB CRYO-EM WORKSHOP

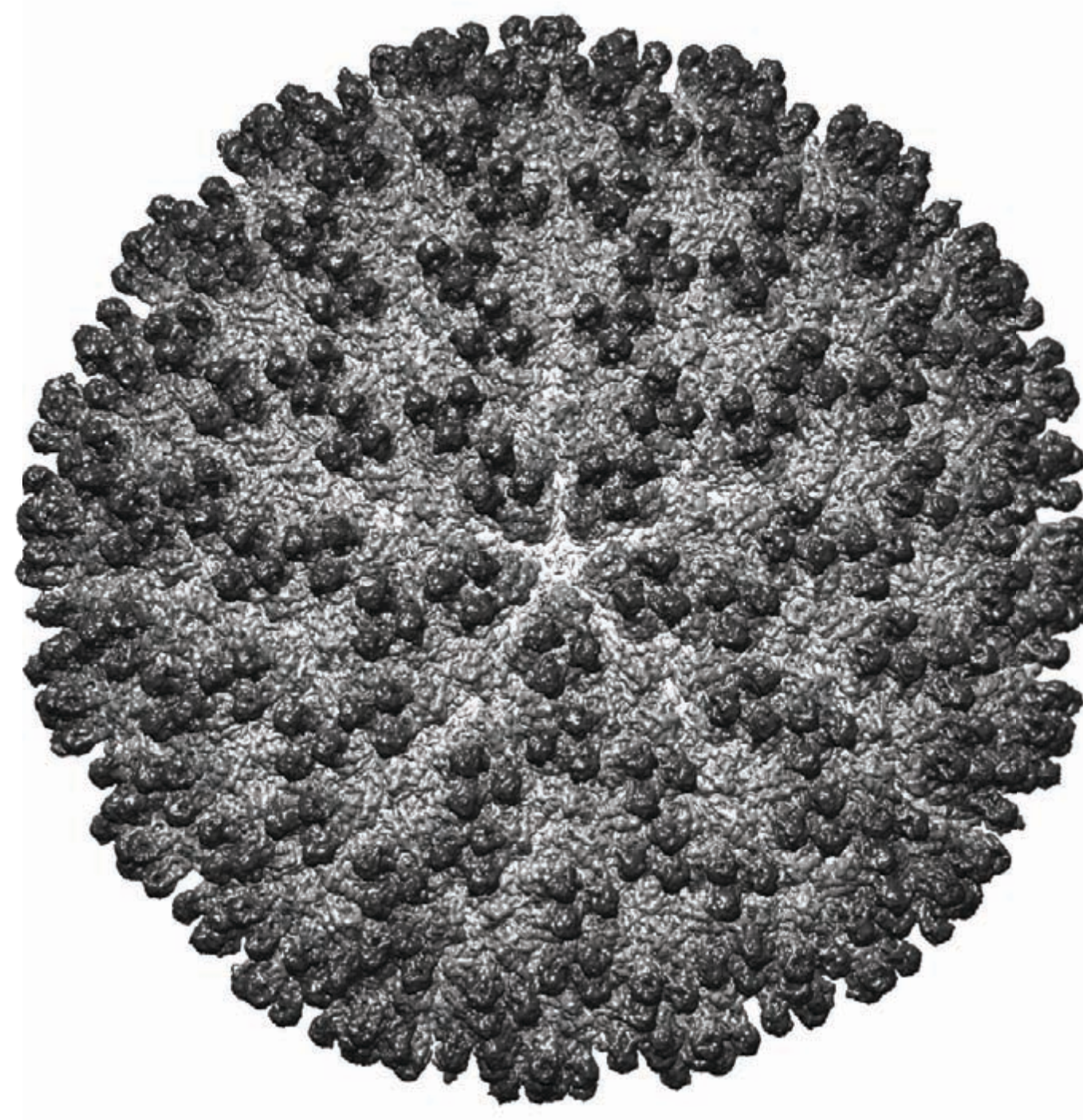
PRACTICAL SYMMETRY

Matthew Baker, Ph.D
Baylor College of Medicine

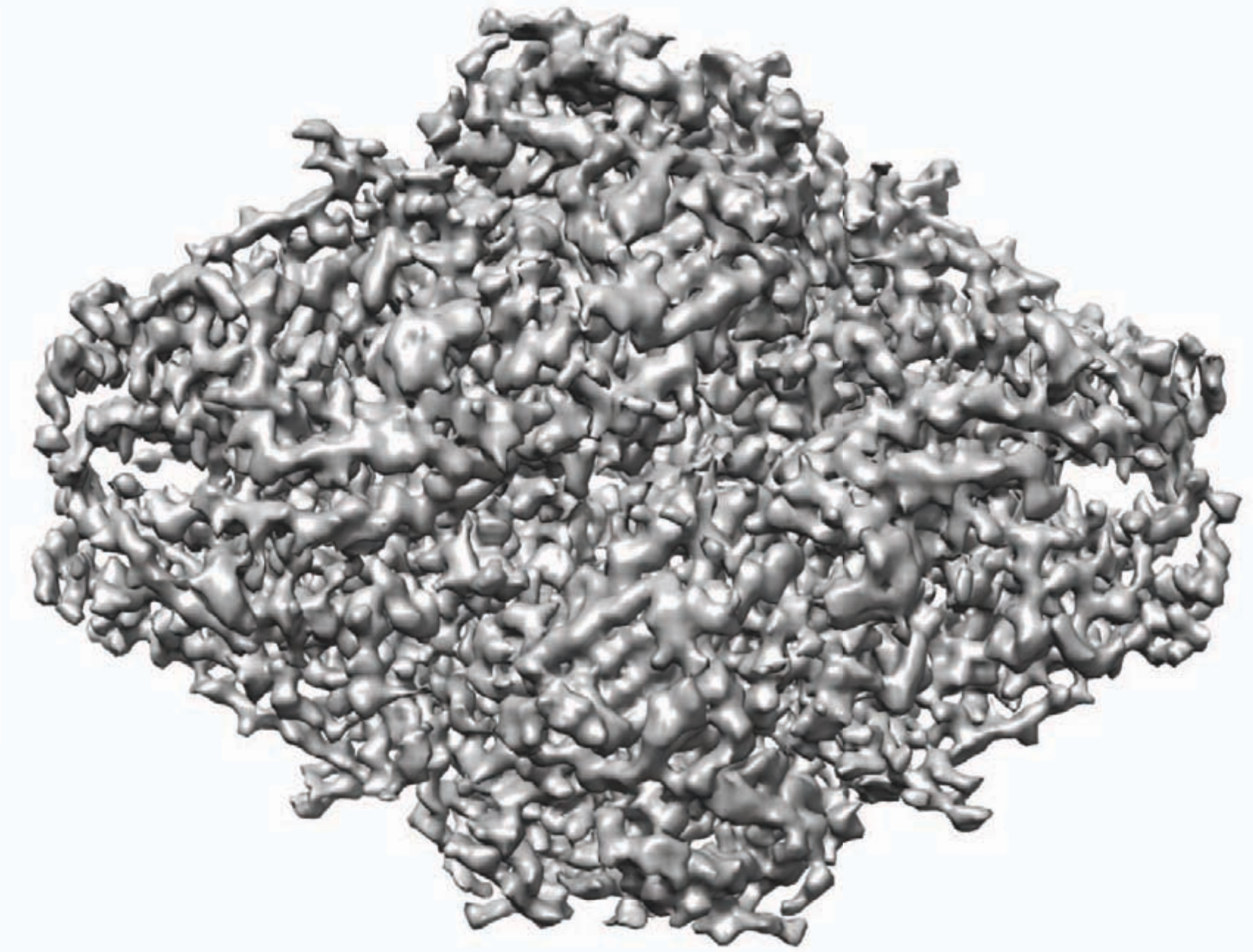
SYMMETRY IN CRYO-EM



IP3R1: C4

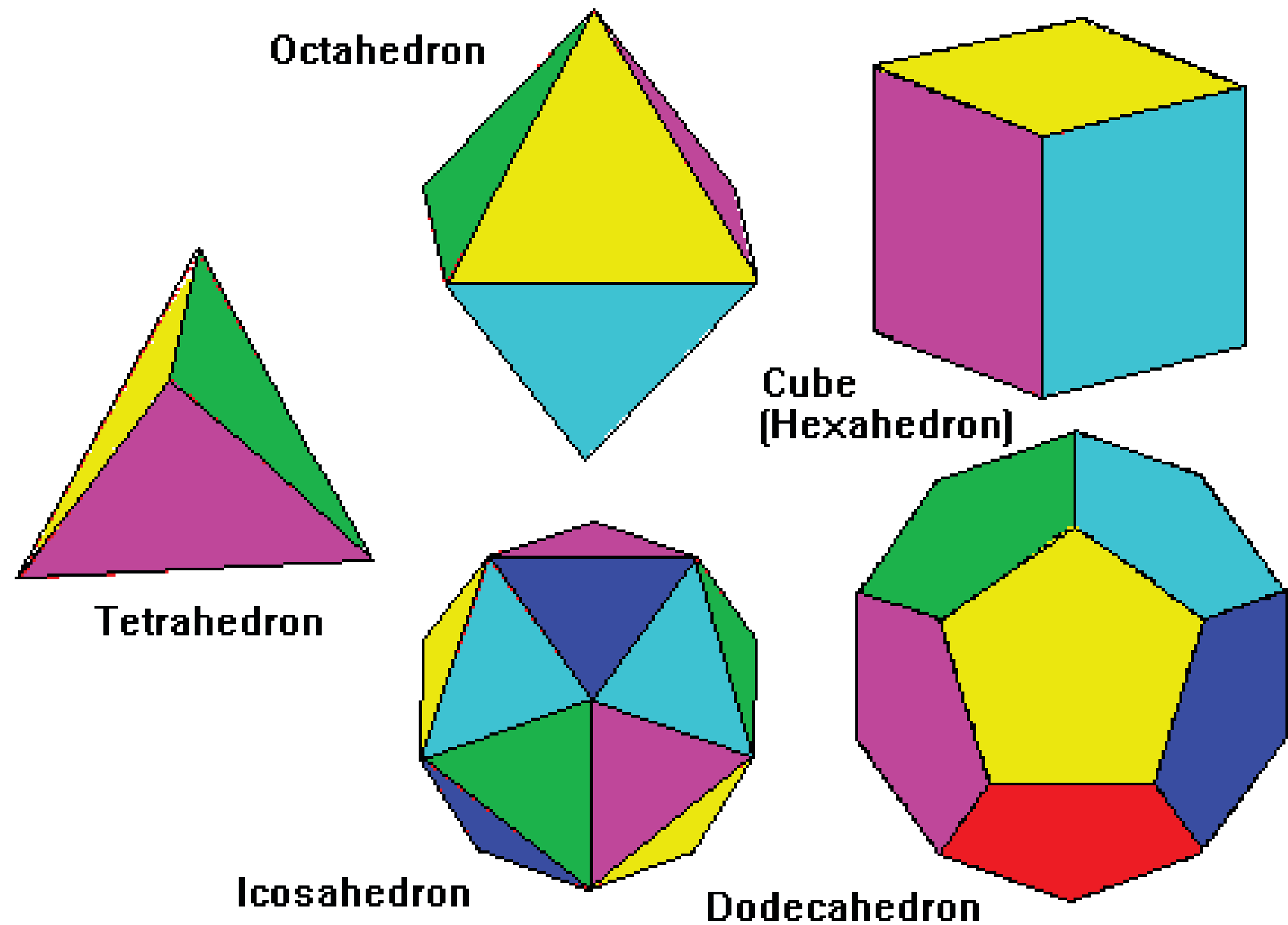


MCRV: Icos



Beta-galactosidase: D2

TYPES OF SYMMETRY IN THE FIVE PLATONIC SOLIDS

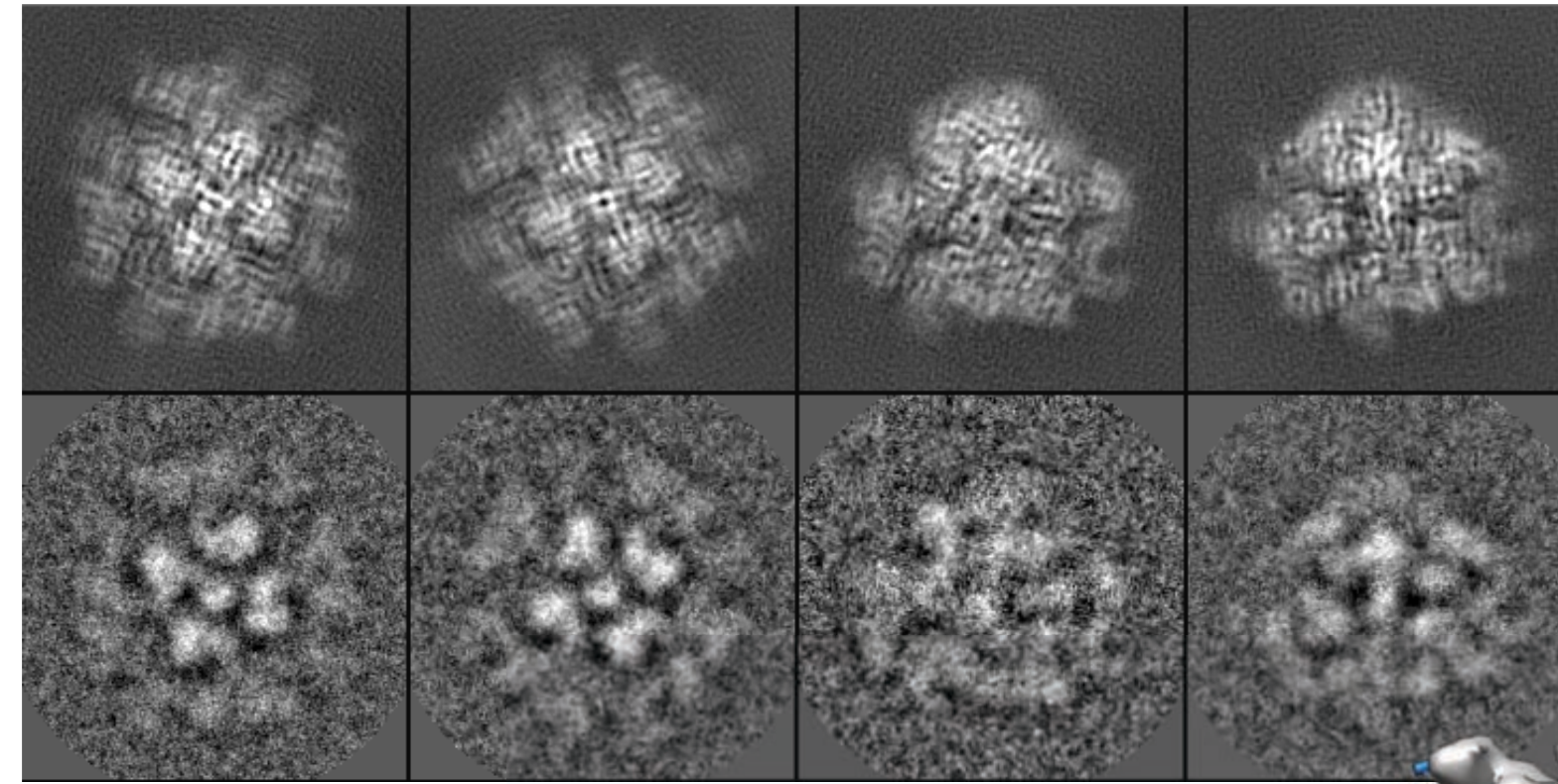


Solid	Faces	Edges in a face	Symmetries
Tetrahedron	4	3	24
Cube	6	4	48
Octahedron	8	3	48
Dodechedron	12	5	120
Icosahedron	20	3	120

SYMMETRY IN CRYO-EM

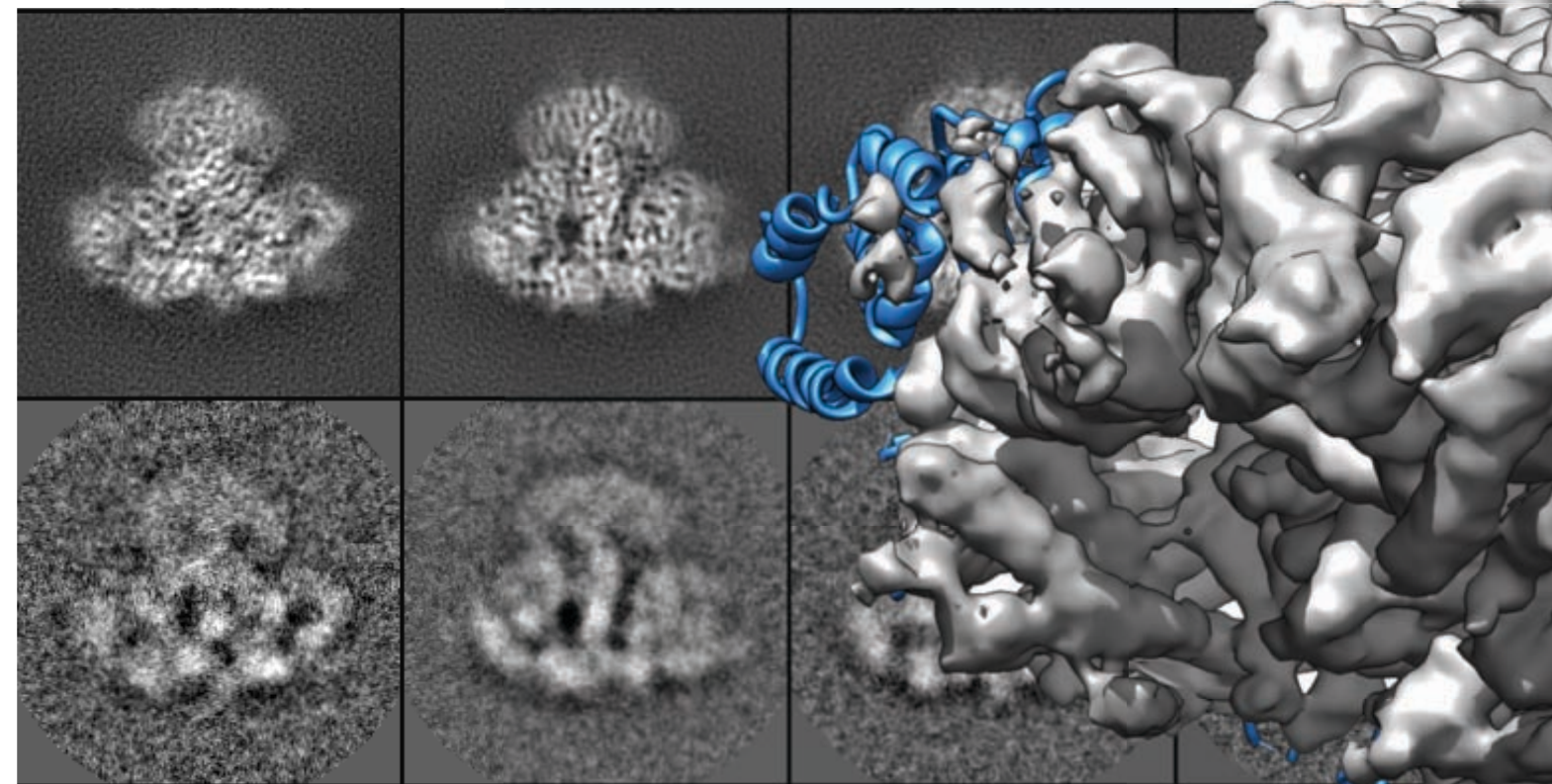
- ▶ Particle Images
- ▶ Reconstruction
- ▶ Modeling
- ▶ Refinement

reprojections

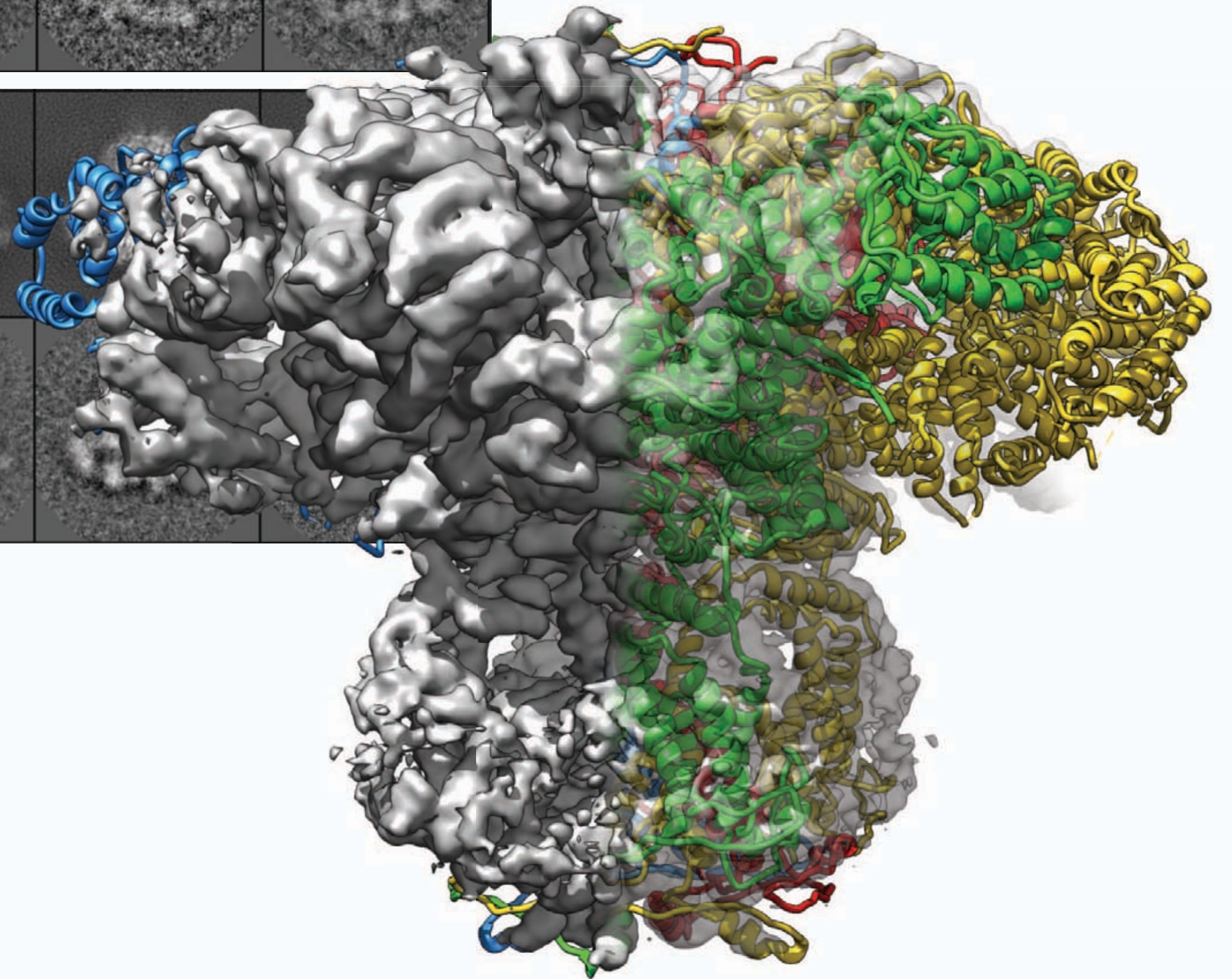


class averages

reprojections

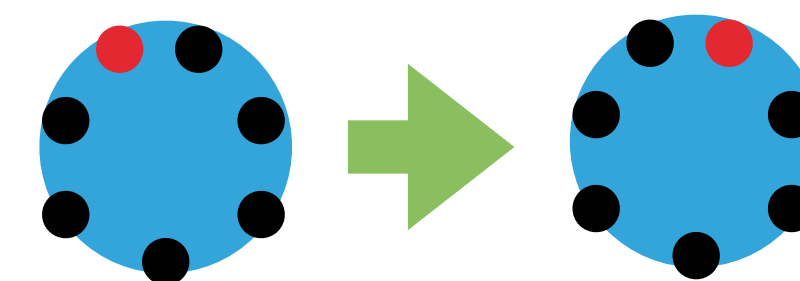
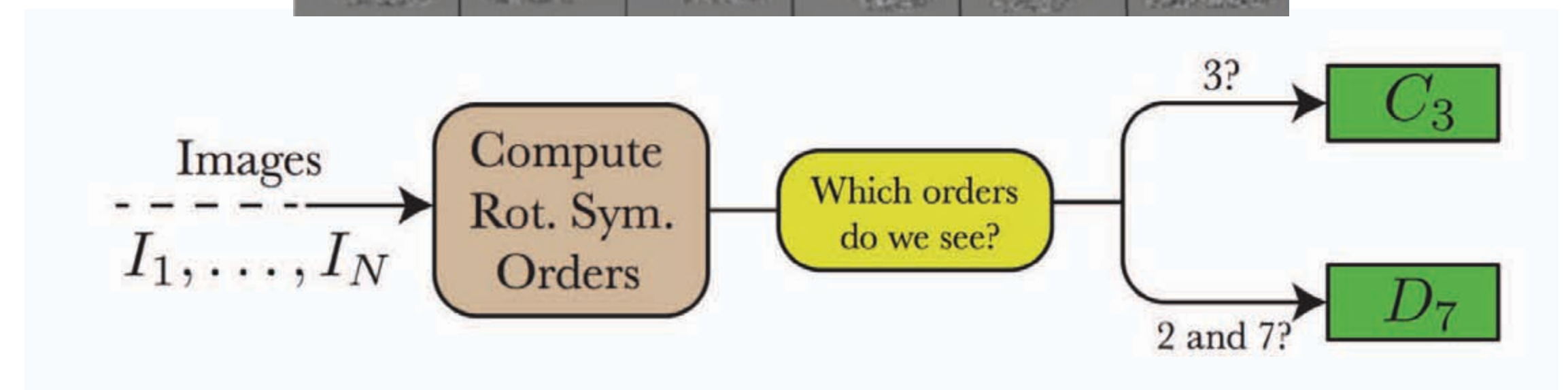
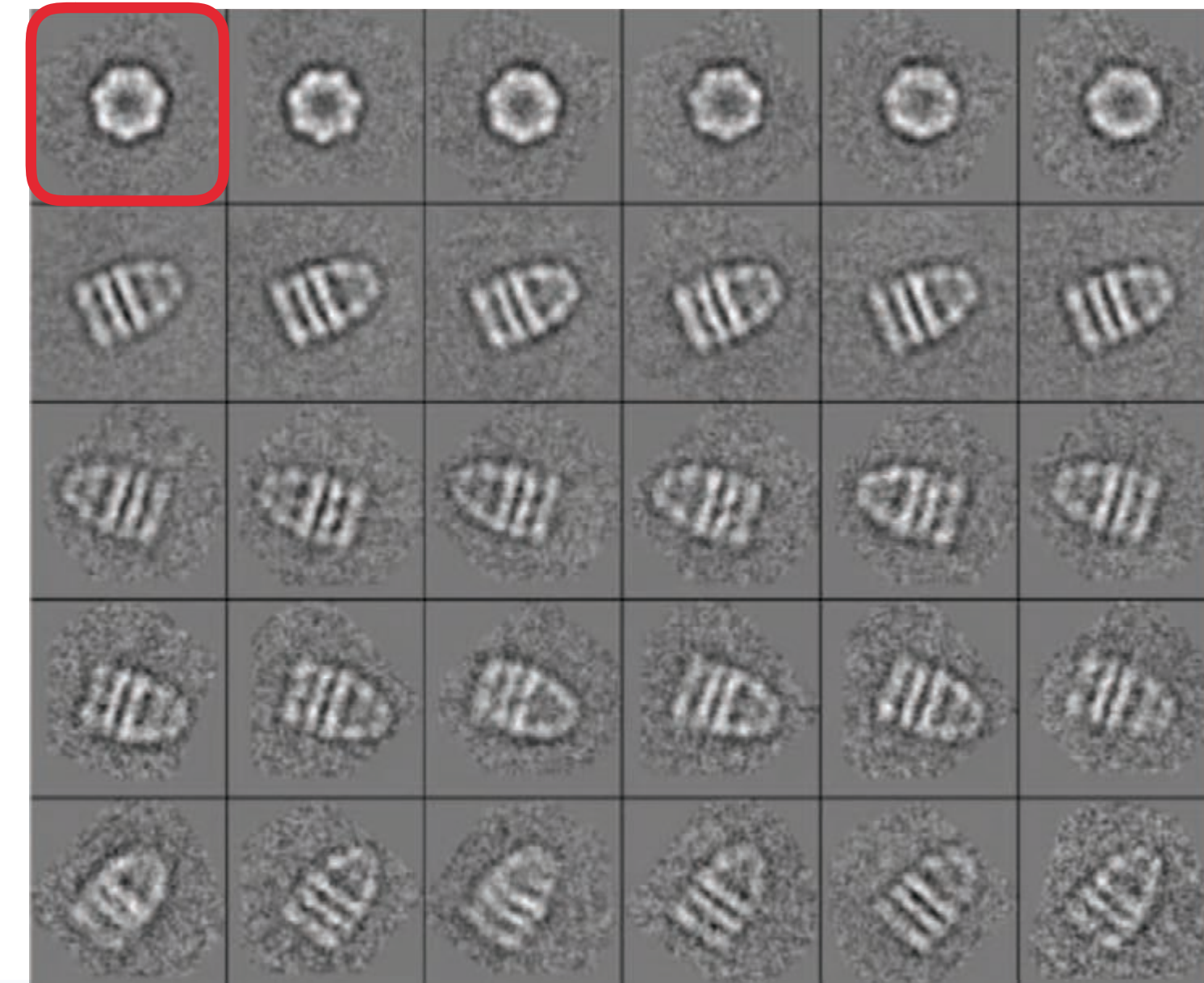


class averages



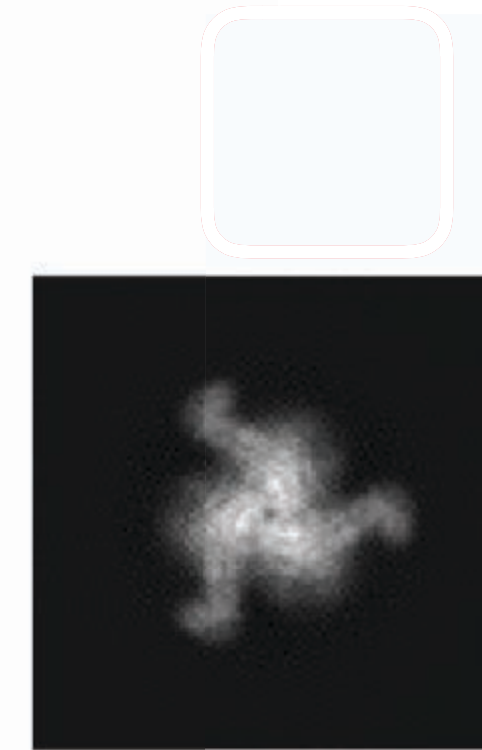
DETECTING SYMMETRY IN 2D IMAGES

- ▶ Symmetric views of particles/class averages can be extracted and examined for symmetry
- ▶ 2D images can be rotated at given intervals and correlated (rotational auto-correlation) to determine symmetry

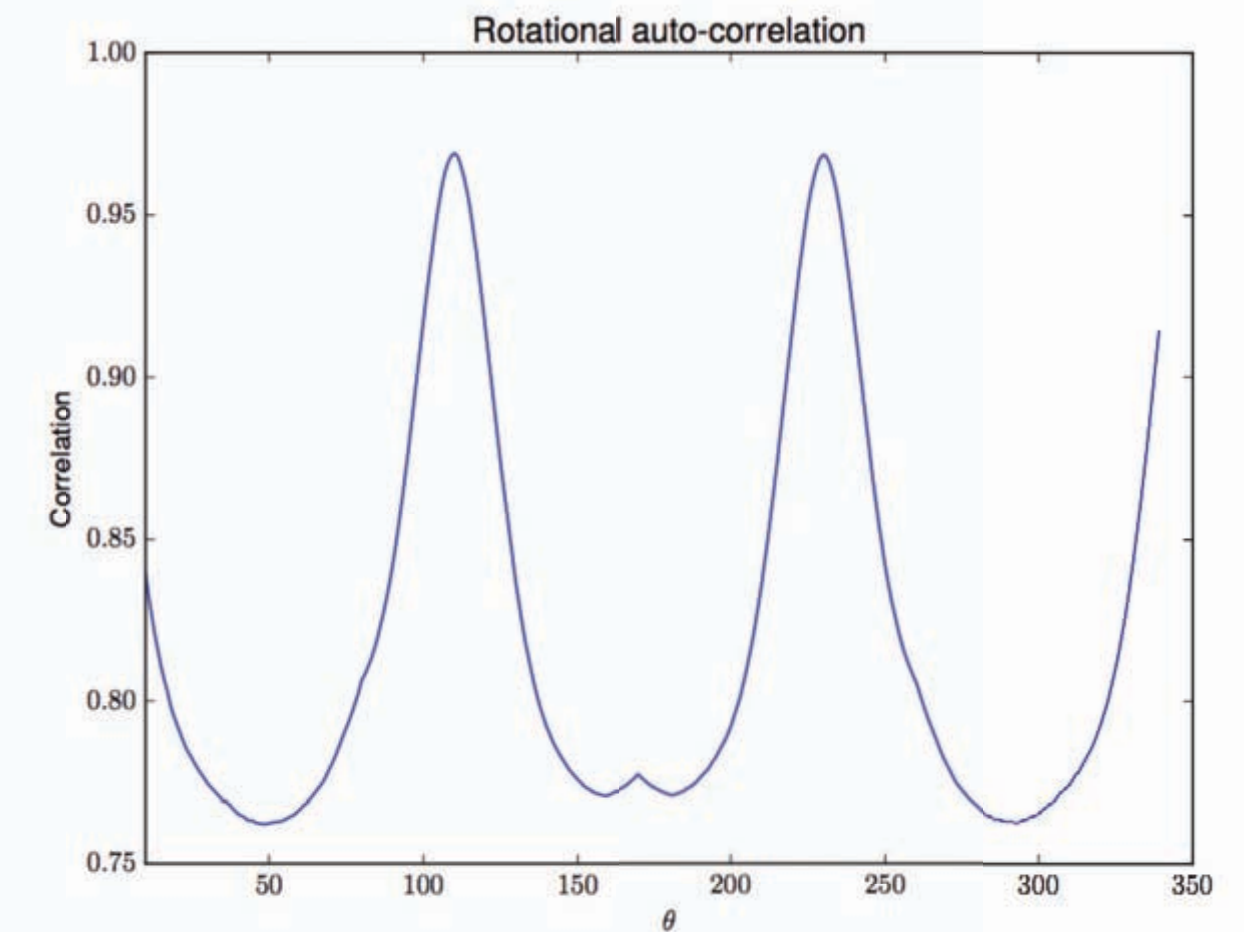


DETECTING SYMMETRY IN 2D IMAGES

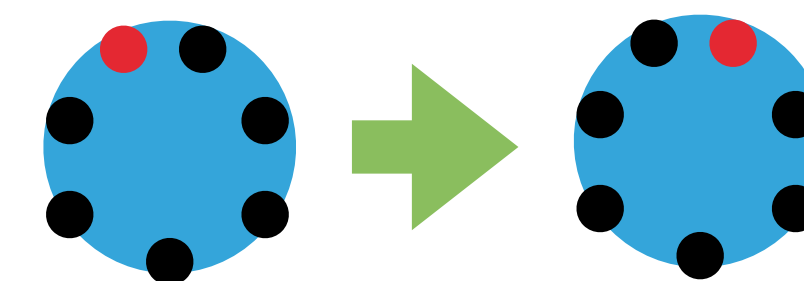
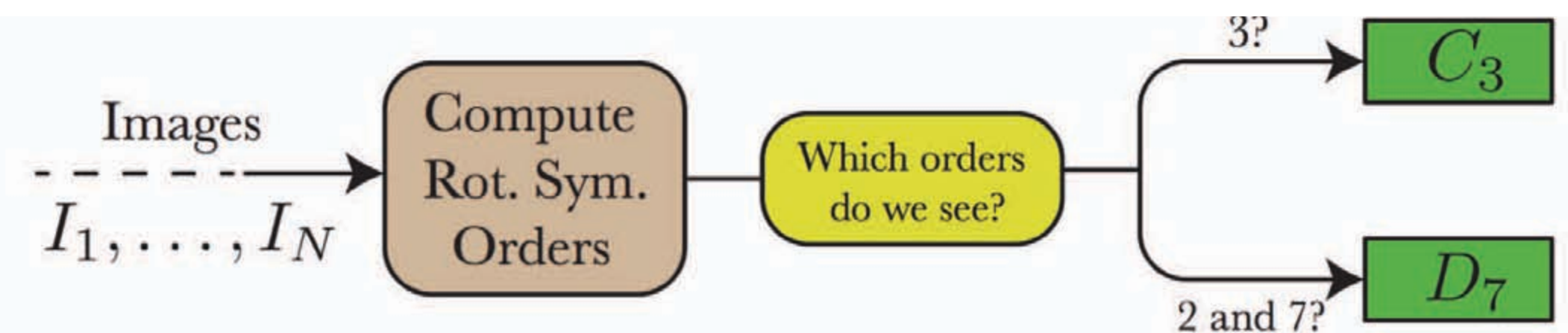
- ▶ Symmetric views of particles/class averages can be extracted and examined for symmetry
- ▶ 2D images can be rotated at given intervals and correlated (rotational auto-correlation) to determine symmetry



(a) Clean Env top view



(b) Rotational auto-correlation



DETECTING SYMMETRY IN A 3D RECONSTRUCTION

▶ EMAN

- ▶ rotate and calculate correlation with e2proc3d.py

▶ Chimera

- ▶ from the command line "measure symmetry #0"

▶ Phenix

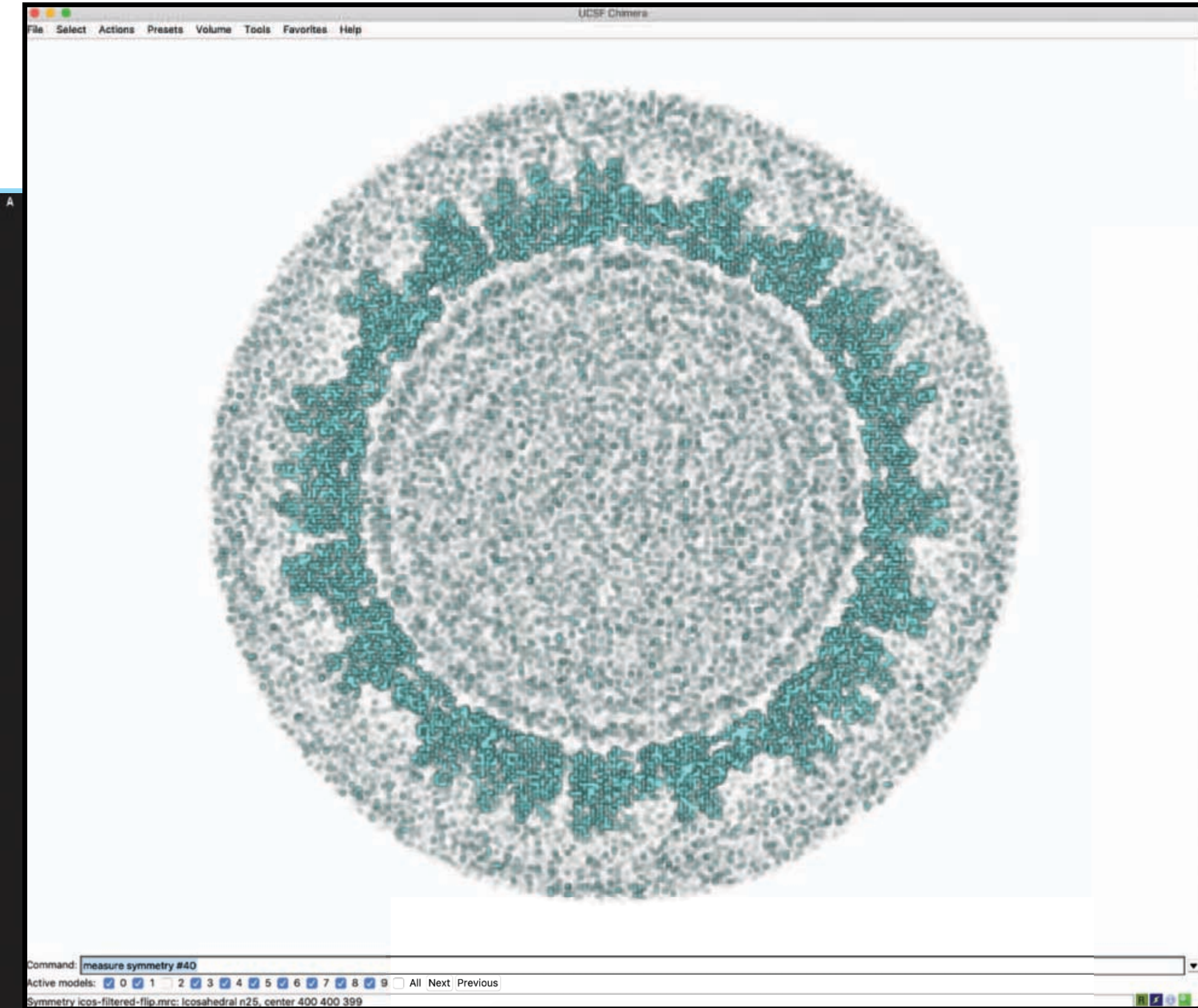
- ▶ phenix.map_symmetry

```
58.35 A 33.34 A 4.58 A 8.98 A 5.19 A 7.29 A 3.70 A
Optimizing helical parameters:
Rot Trans Score CC
Trying fraction of rot/trans
Trying 15x multiples of rot/trans
-2024.99 875.23 0.51 0.51 (Scoring for multiple)
-135.00 58.35 0.51 0.51 (Final)
Offsetting NCS center by (116.70, 116.70, 116.70) A
Total of 46 NCS types to examine...
Rescoring top 5 results
Ranking of NCS types:
SCORE CC OPERATORS SYMMETRY
2.00 1.00 4 C4
1.41 1.00 2 C2
1.31 0.70 6 C6
1.30 0.51 11 C11
0.97 0.53 12 C12
0.95 0.51 9 C9
0.93 0.62 3 C3
0.88 0.34 13 C13
0.81 0.59 8 C8
0.80 0.48 14 C14
0.79 0.52 10 C10
0.79 0.39 28 D14 (a)
0.79 0.39 28 D14 (b)
0.73 0.54 24 D12 (b)
0.73 0.54 24 D12 (a)
0.67 0.40 20 D10 (a)
0.67 0.40 20 D10 (b)
0.60 0.52 16 D8 (b)
0.60 0.52 16 D8 (a)
0.57 0.44 7 C7
0.56 0.46 5 C5
0.56 0.36 14 D7 (b)
0.56 0.36 14 D7 (a)
0.52 0.52 12 D6 (a)
0.52 0.52 12 D6 (b)
0.47 0.36 10 D5 (b)
0.47 0.36 10 D5 (a)
0.43 0.41 18 D9 (b)
0.43 0.41 18 D9 (a)
0.42 0.51 8 D4 (a)
0.42 0.51 8 D4 (b)
0.36 0.48 6 D3 (b)
0.36 0.48 6 D3 (a)
0.32 0.43 26 D13 (a)
0.32 0.43 26 D13 (b)
0.30 0.31 60 I (d)
0.30 0.43 4 D2 (b)
0.30 0.43 4 D2 (a)
0.29 0.32 60 I (c)
0.28 0.33 60 I (a)
0.23 0.39 22 D11 (b)
0.23 0.39 22 D11 (a)
0.10 0.51 5 Type: Helical -135.00 deg 58.35 Z-trans
-1.19 0.29 60 I (b)
-1.69 0.29 60 I (e)
-1.69 0.23 60 I (f)

Best NCS type is:
SCORE CC OPERATORS SYMMETRY
2.00 1.00 4 C4

Final symmetry obtained:
Correlation of symmetry-related regions: 1.00 Copies: 4
Wrote operators in .ncs_spec format to symmetry_from_map.ncs_spec
Finished with map_symmetry

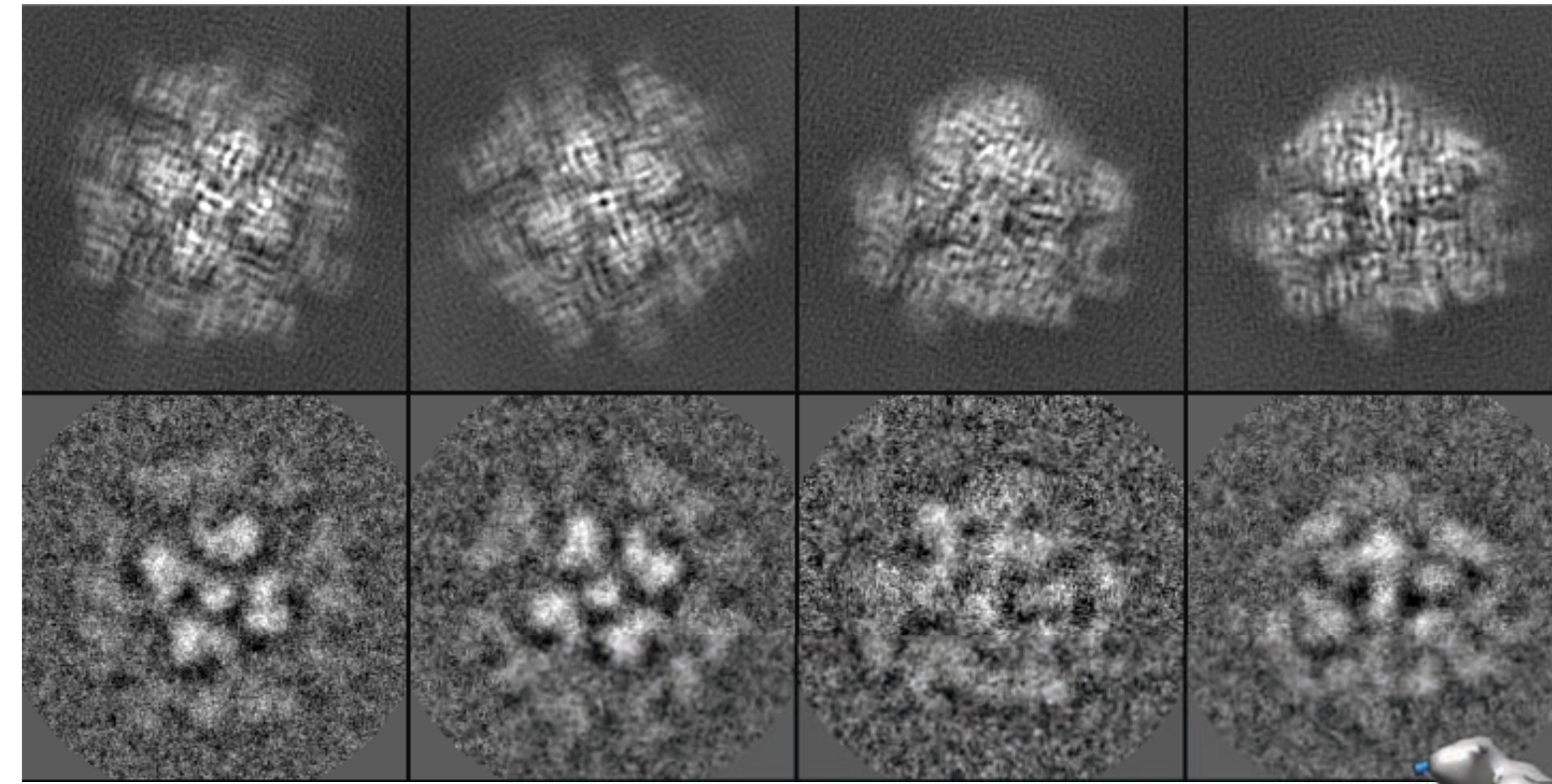
Job complete
usr+sys time: 180.58 seconds
wall clock time: 3 minutes 2.86 seconds (182.86 seconds total)
```



SYMMETRY IN CRYO-EM

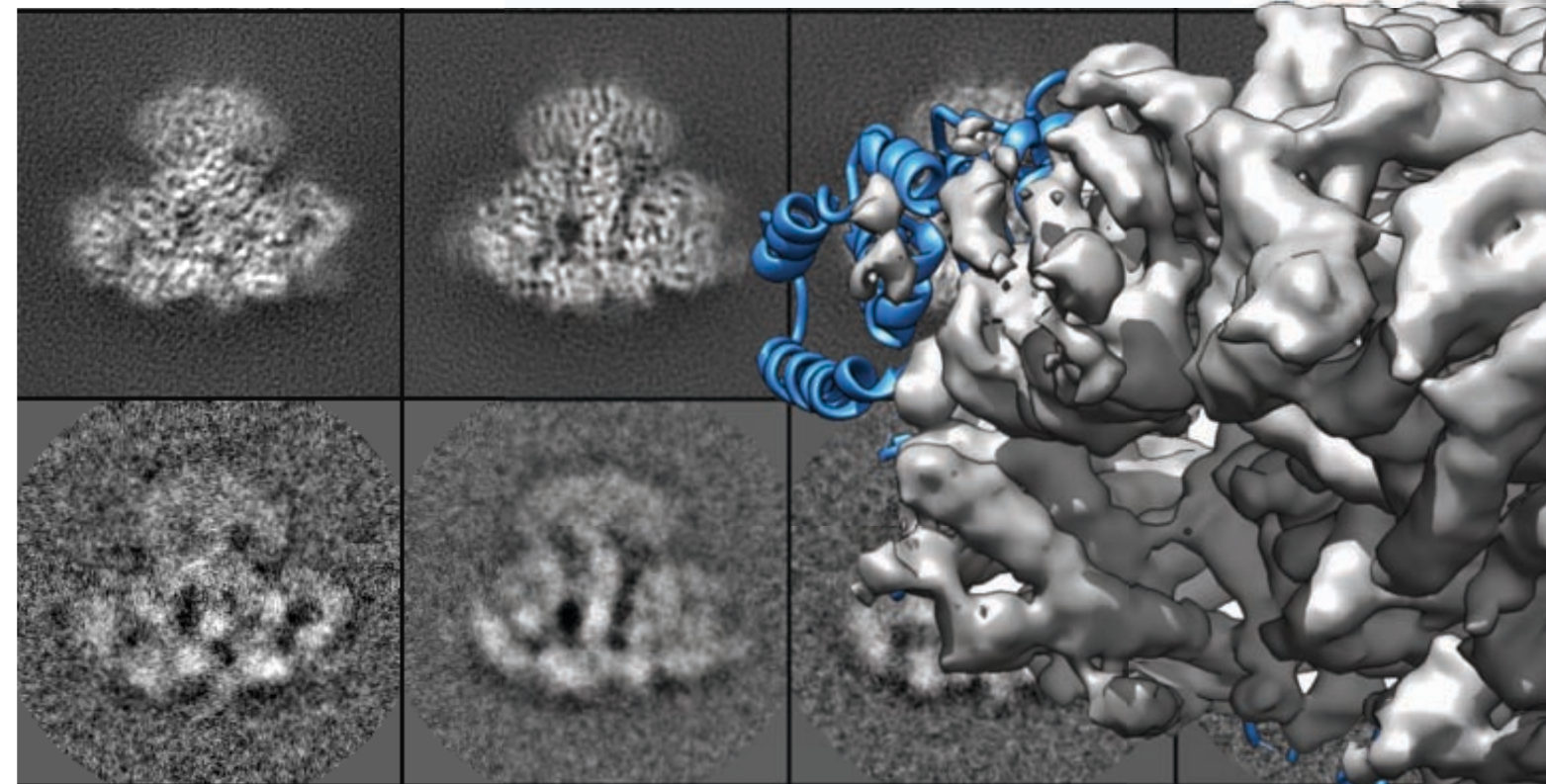
- ▶ Particle Images
- ▶ Reconstruction
- ▶ Modeling
- ▶ Refinement

reprojections

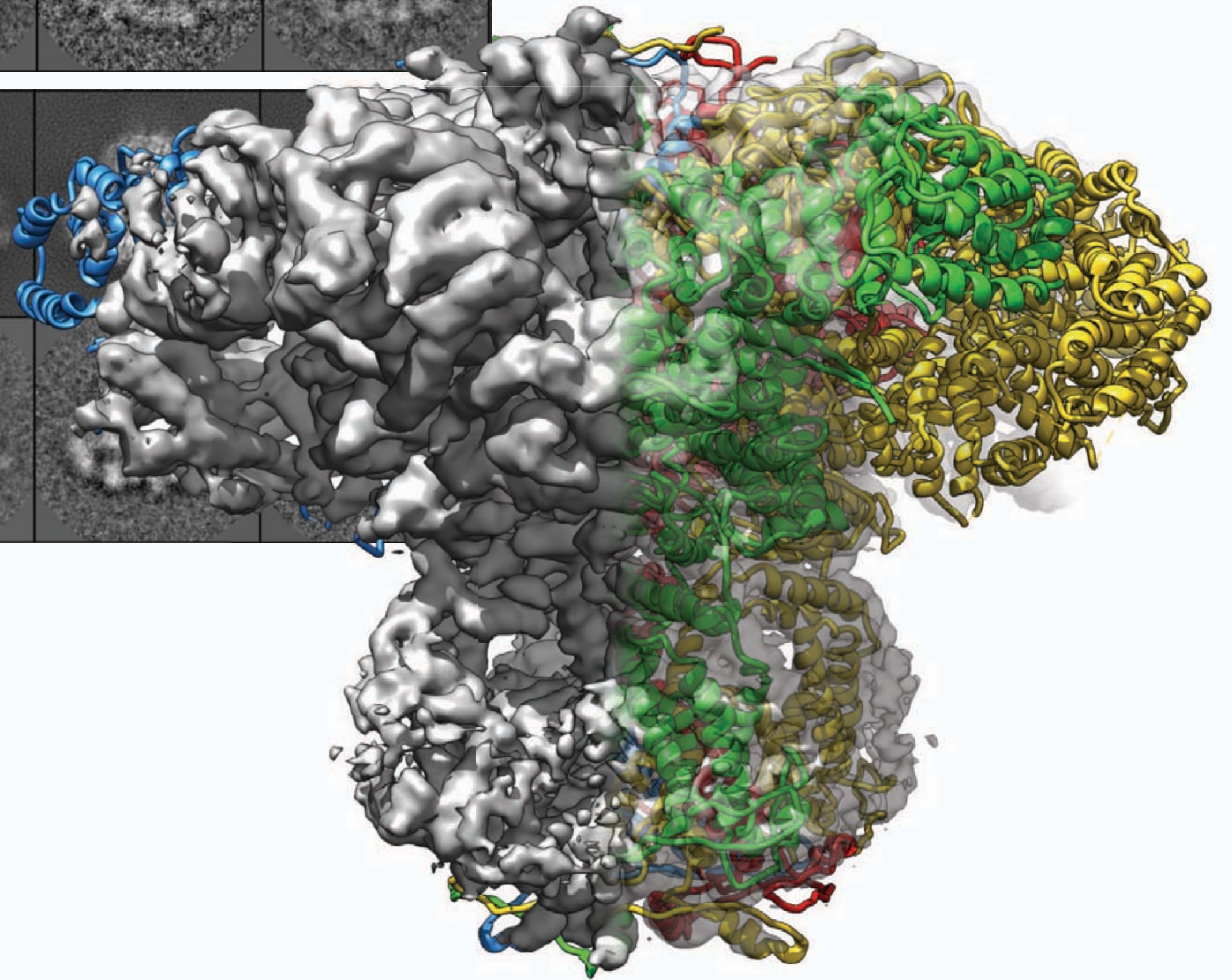


class averages

reprojections

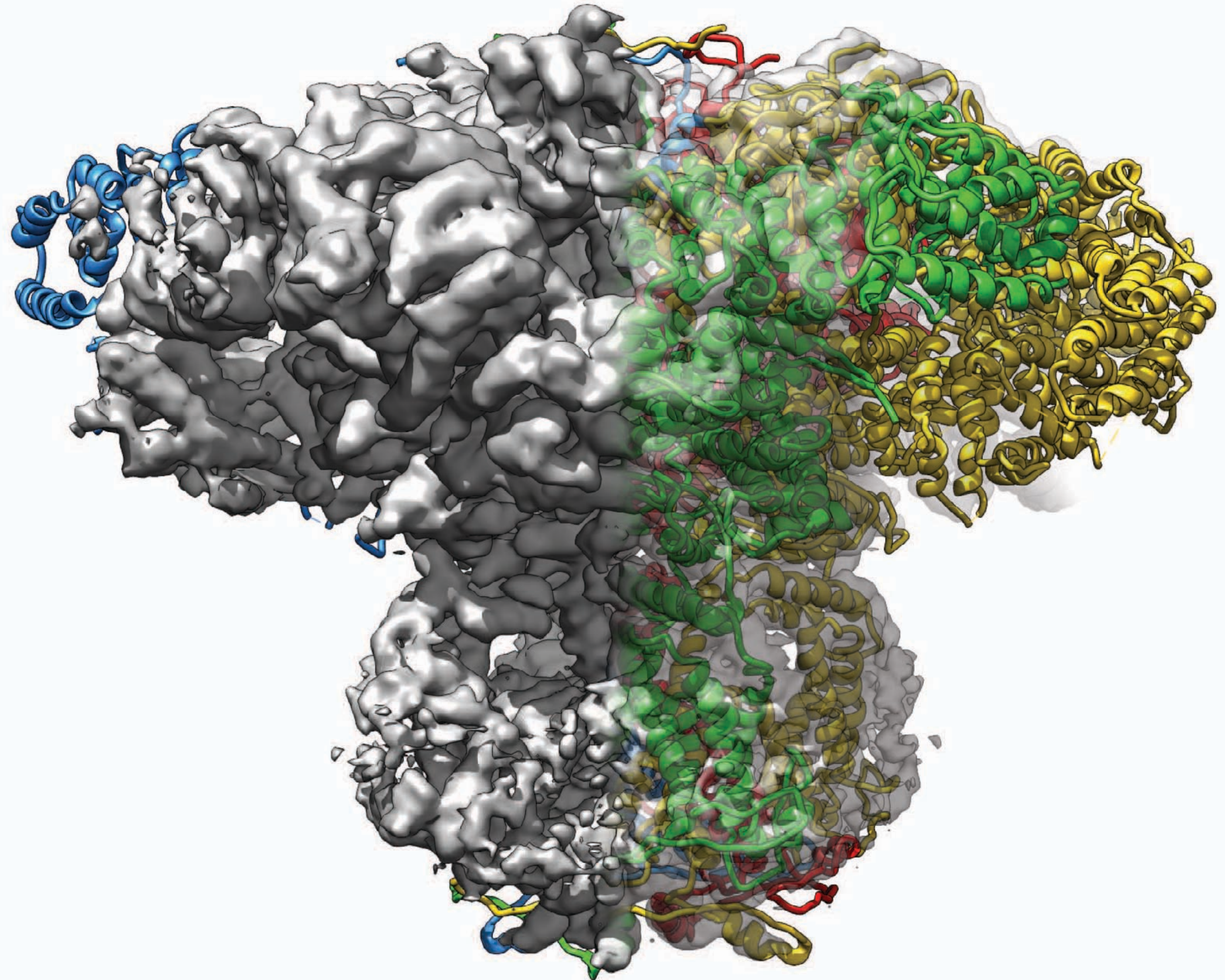


class averages



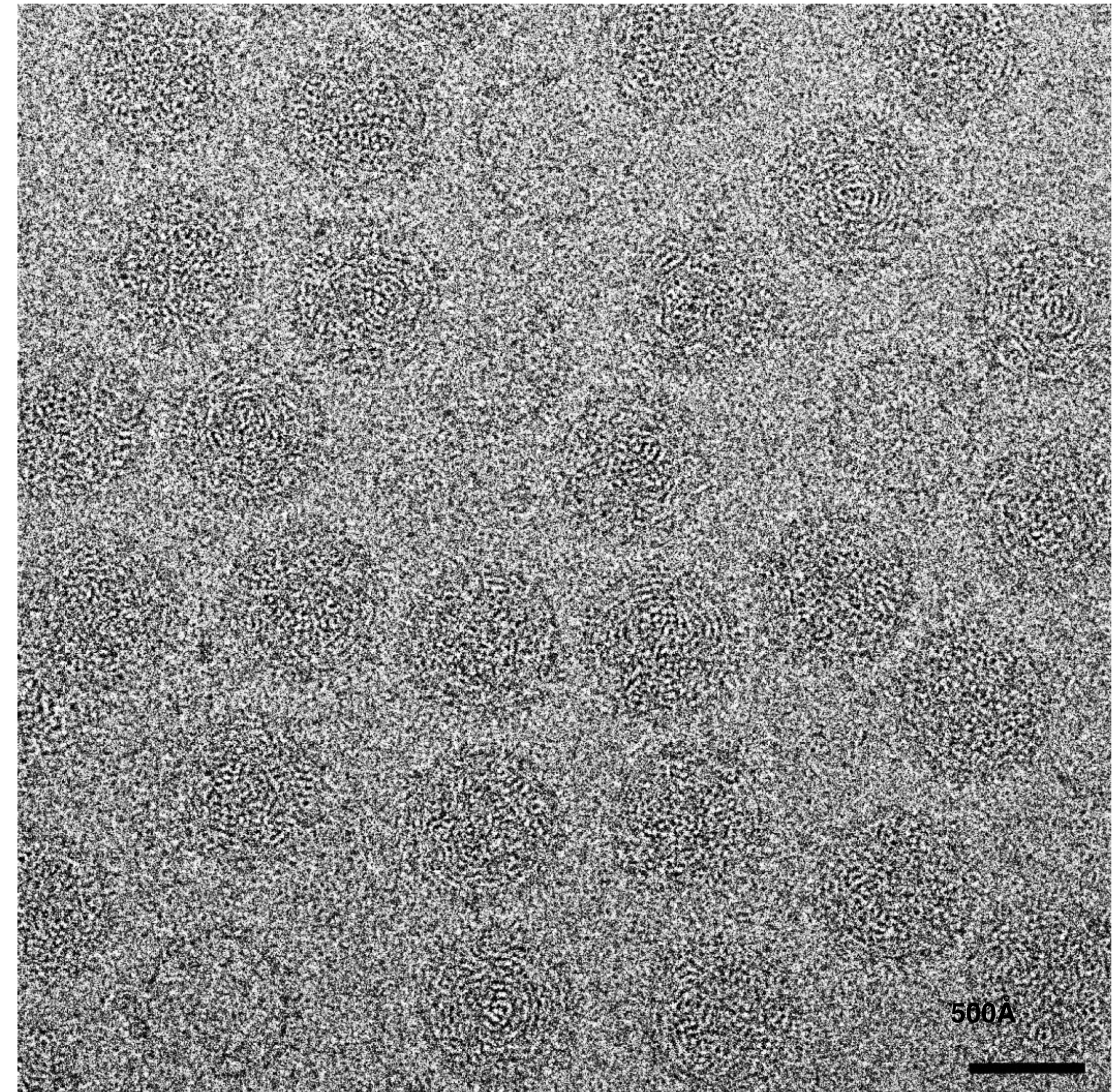
SYMMETRY IN CRYO-EM

- ▶ Particle Images
- ▶ Reconstruction
- ▶ Modeling
- ▶ Refinement



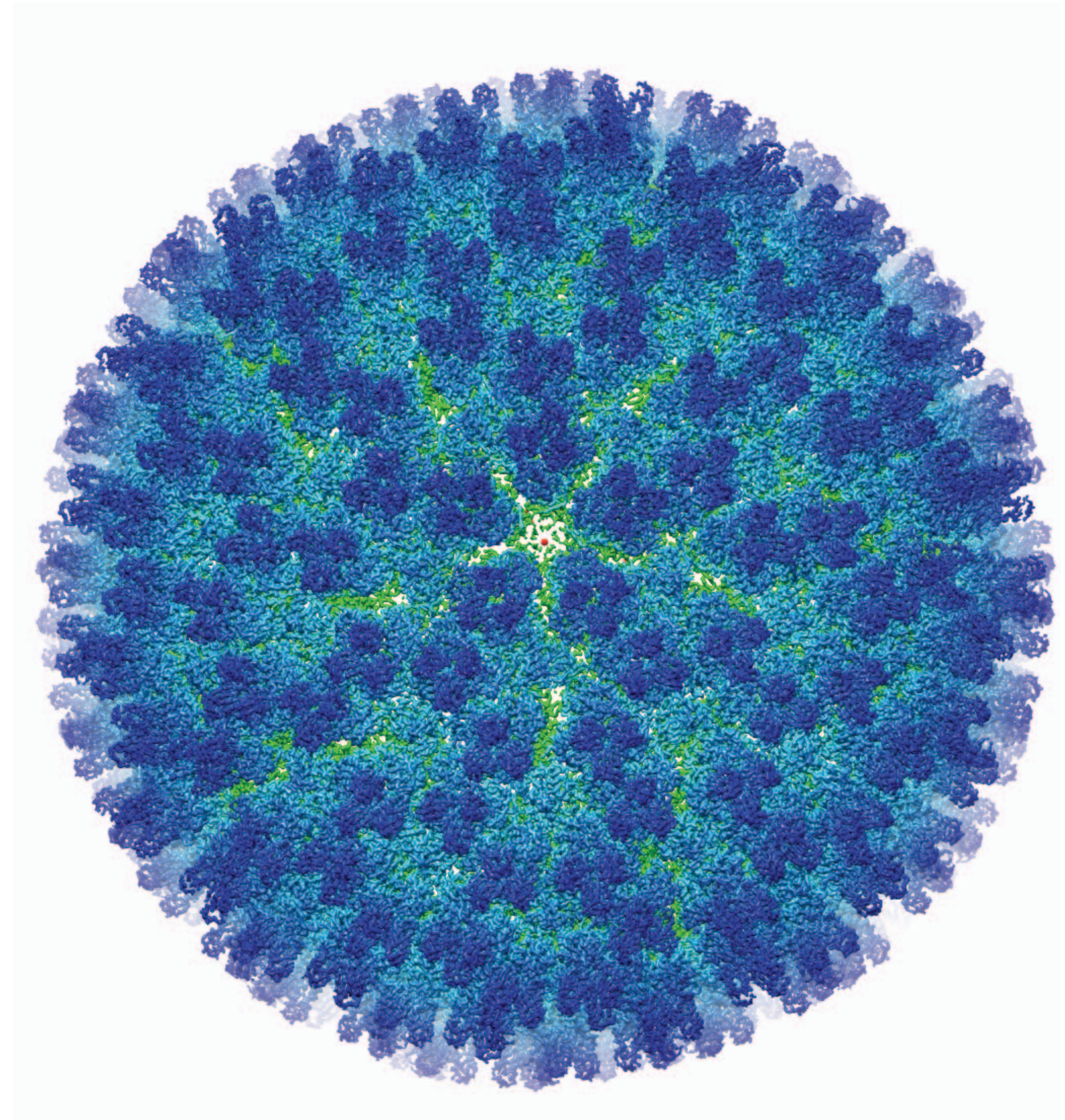
CASE STUDY: MUD CRAB REOVIRUS

- ▶ Infects *Scylla serrata*
- ▶ Sedoreovirinae, genus Crabreovirus
- ▶ 12 segmented dsRNA genome
- ▶ 70nm in diameter
- ▶ Double layer capsid
 - ▶ No capping turret; internal capping enzyme
 - ▶ 3 capsid proteins: VP3, VP11 and VP12
 - ▶ T=13I outer layer
 - ▶ Pseudo T=2 inner layer
 - ▶ RdRP located on inside of inner layer



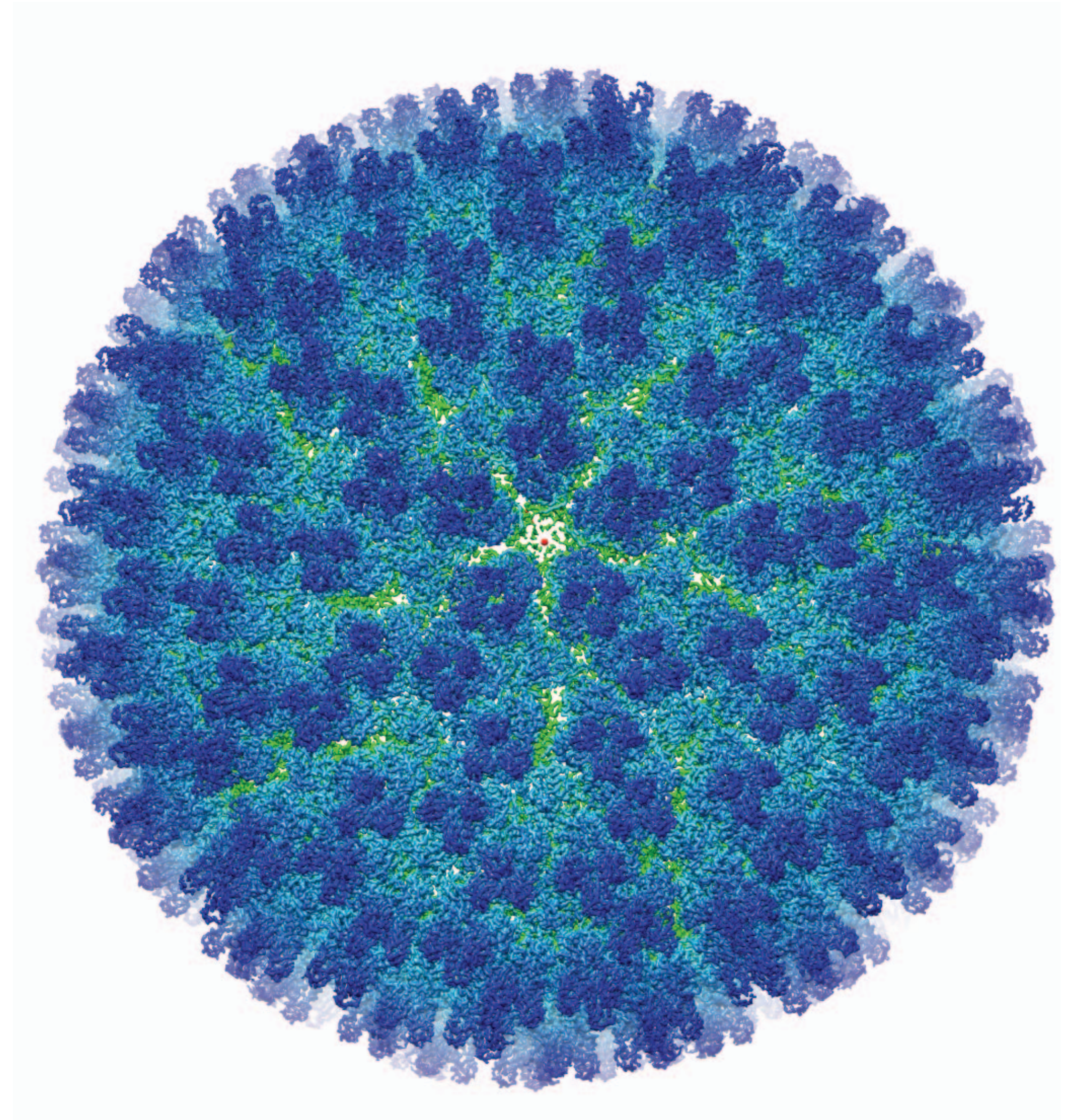
CASE STUDY: MUD CRAB REOVIRUS

- ▶ Infects *Scylla serrata*
- ▶ Sedoreovirinae, genus Crabreovirus
- ▶ 12 segmented dsRNA genome
- ▶ 70nm in diameter
- ▶ Double layer capsid
 - ▶ No capping turret; internal capping enzyme
 - ▶ 3 capsid proteins: VP3, VP11 and VP12
 - ▶ T=13I outer layer
 - ▶ Pseudo T=2 inner layer
 - ▶ RdRP located on inside of inner layer



ICOSAHEDRAL RECONSTRUCTION

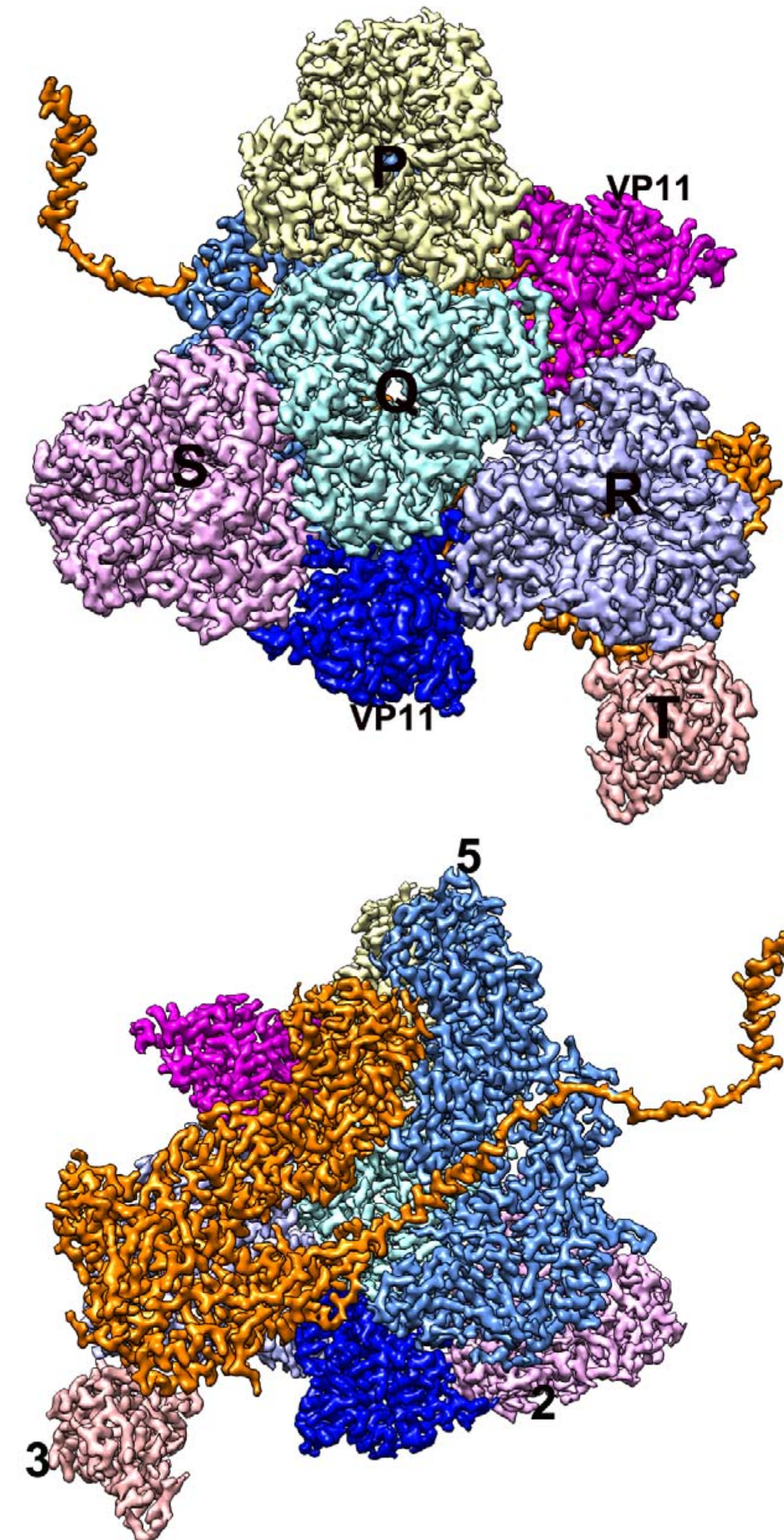
- ▶ MCRV purified from infected *Scyalla serrata*
- ▶ Imaged on FEI Titan Krios with FEI Falcon II/ Falcon III at 75K (1.09 Å/pix)
- ▶ Images processed with Motioncorr V2, EMAN2 and JSPR
- ▶ Icosahedral reconstruction
 - ▶ Inactive: 58095 particles (3.1Å resolution)
 - ▶ Transcribing: 9937 particles (3.4Å resolution)



ICOSAHEDRAL RECONSTRUCTION

- ▶ MCRV purified from infected *Scylla serrata*
- ▶ Imaged on FEI Titan Krios with FEI Falcon II/ Falcon III at 75K (1.09 Å/pix)
- ▶ Images processed with Motioncorr V2, EMAN2 and JSPR
- ▶ Icosahedral reconstruction
 - ▶ Inactive: 58095 particles (3.1Å resolution)
 - ▶ Transcribing: 9937 particles (3.4Å resolution)

- ● VP3 - inner capsid protein (854aa)
- ● VP12 - outer capsid protein (274aa)
- ● VP11 - hexonal outer capsid protein (203aa)



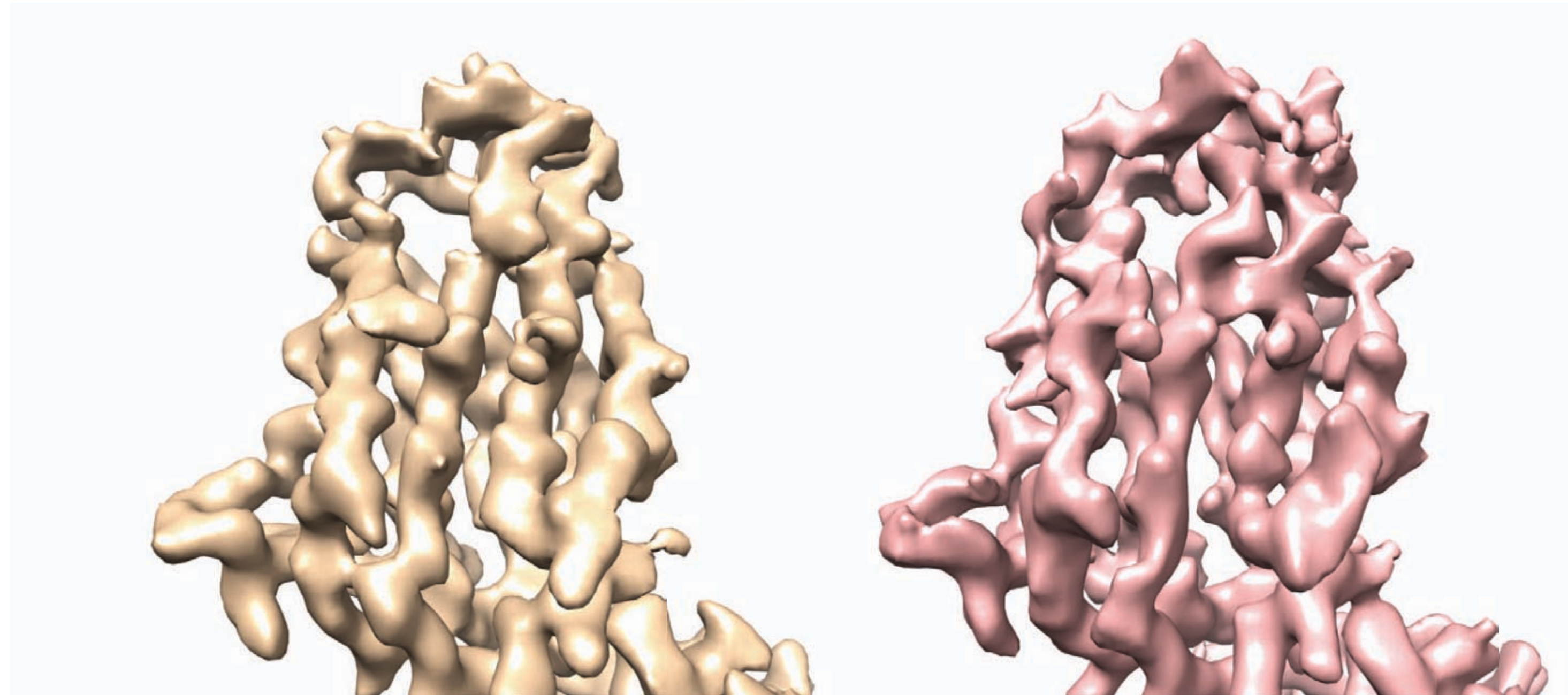
VP11: T=13L OUTER SHELL CAPSID PROTEIN

- ▶ 13 quasi-equivalent copies of VP11 per asymmetric unit
- ▶ Overall density appears similar
- ▶ Resolvability varies in each of the subunits



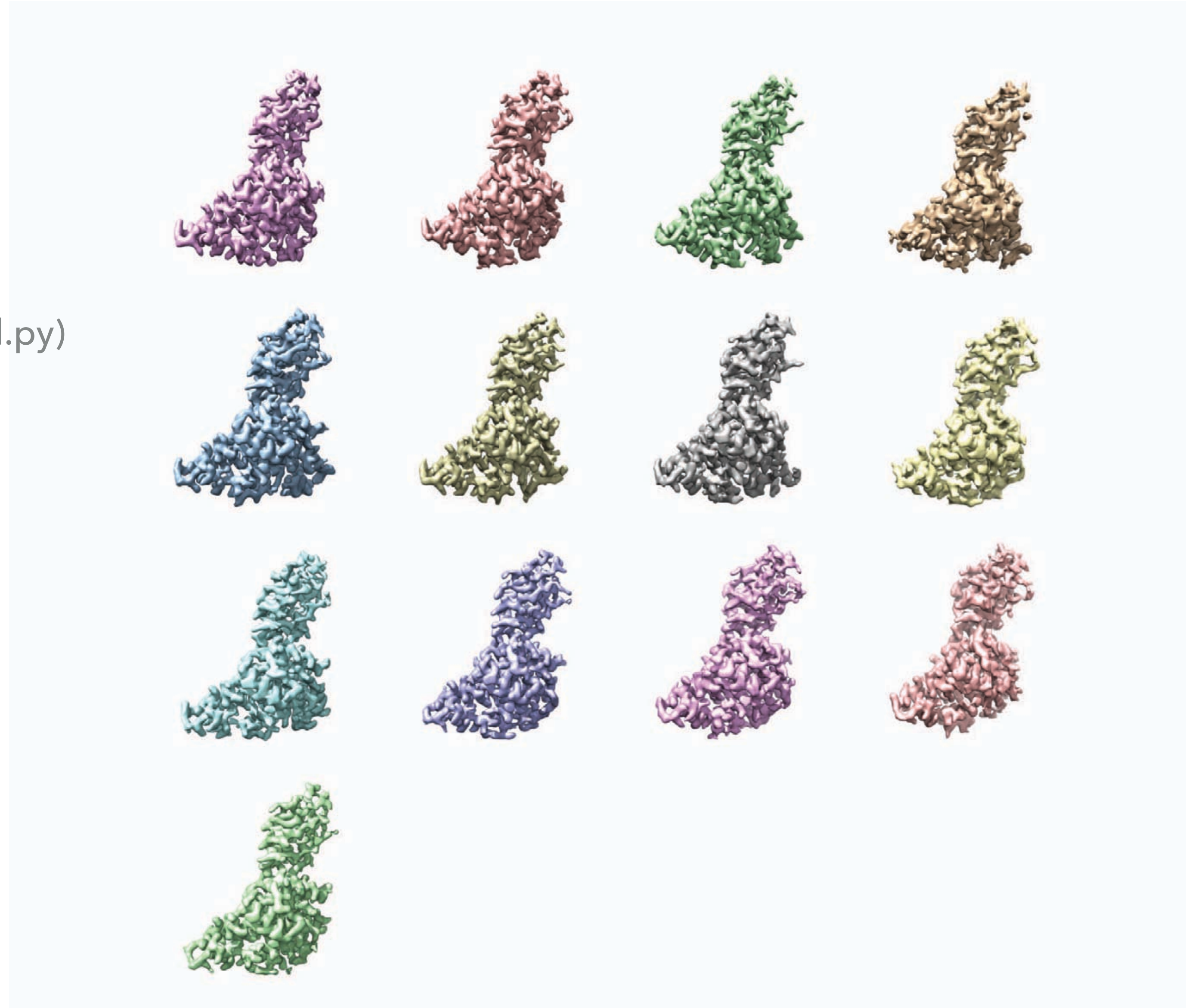
VP11: T=13L OUTER SHELL CAPSID PROTEIN

- ▶ 13 quasi-equivalent copies of VP11 per asymmetric unit
- ▶ Overall density appears similar
- ▶ Resolvability varies in each of the subunits



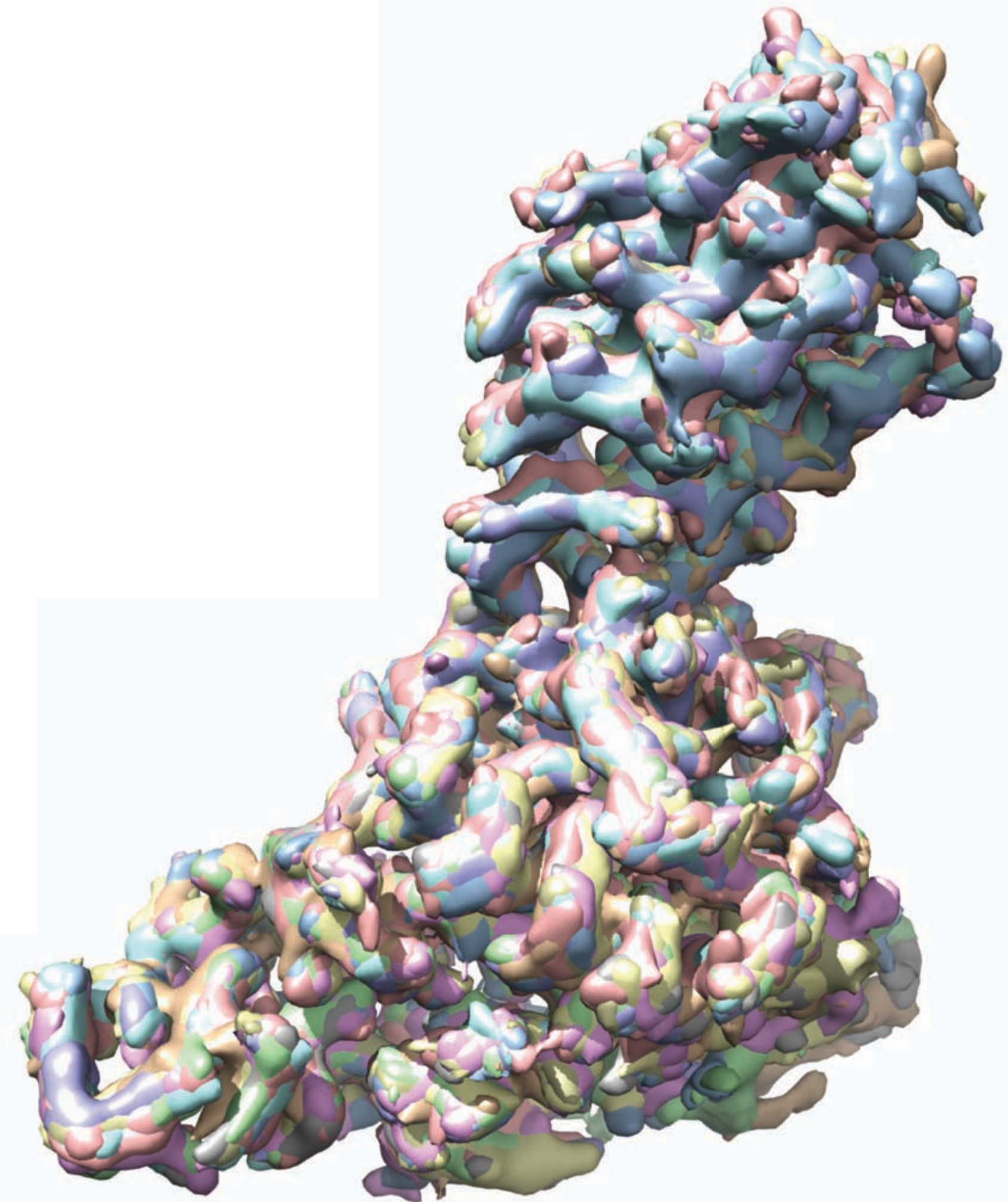
VP11 AVERAGE

- ▶ Building a model for VP11
 - ▶ Rough segmentation of single subunits (Chimera: "volume eraser" or "Segger", Eman2: e2segment3d.py)
 - ▶ Pad all subunits to the same size box (Chimera: vop resample, Eman2: e2proc3d.py)
 - ▶ Align subunits (Chimera: "fit in map", Eman2: e2foldhunter.py,)
 - ▶ Sum aligned subunits (Chimera: vop add, Eman2: e2proc3d.py)
 - ▶ Build a model for the average subunit (Eman2: e2pathwalker.py)
 - ▶ Refine model (Phenix: phenix.real_space_refine)



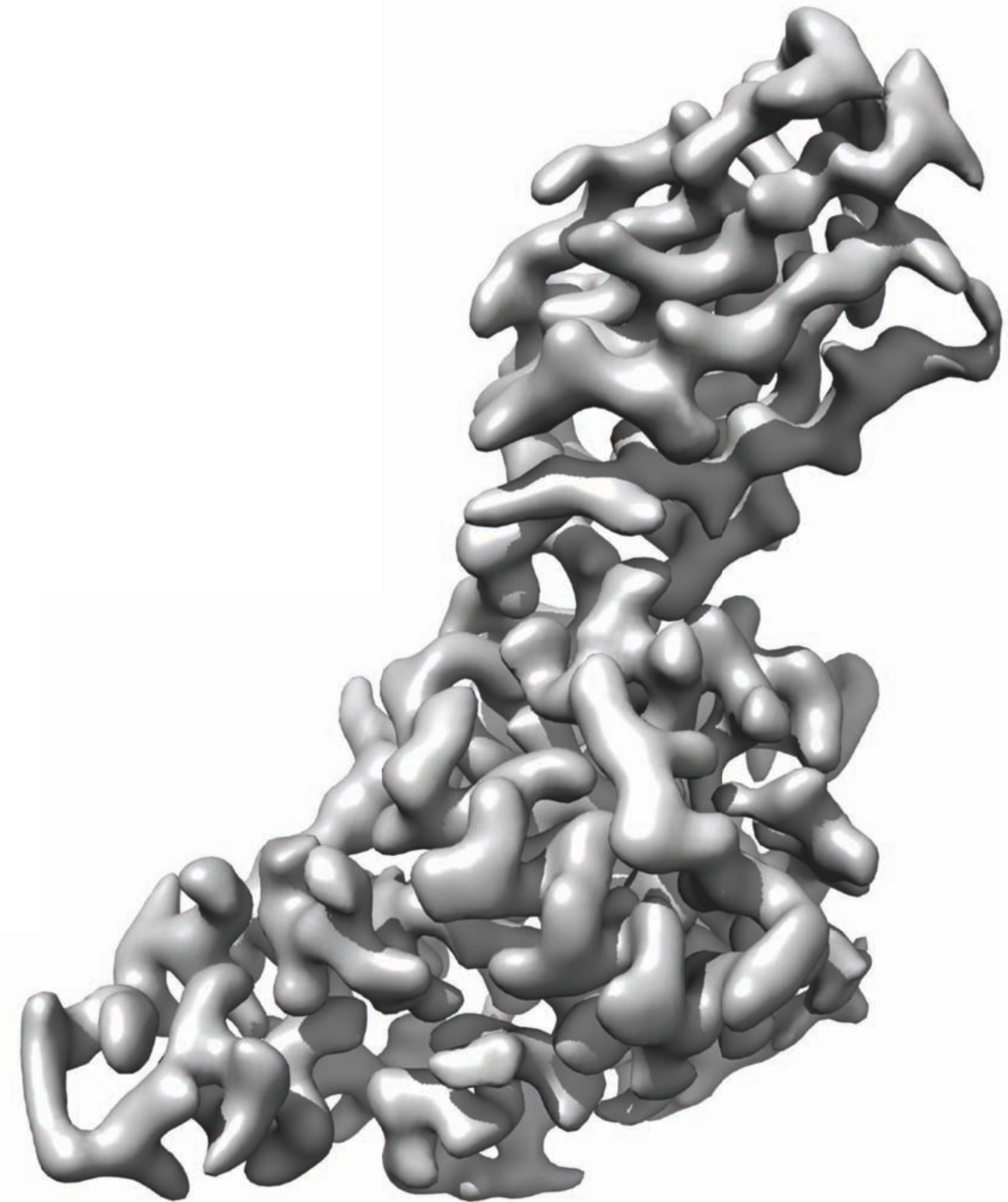
VP11 AVERAGE

- ▶ Building a model for VP11
 - ▶ Rough segmentation of single subunits (Chimera: "volume eraser" or "Segger", Eman2: e2segment3d.py)
 - ▶ Pad all subunits to the same size box (Chimera: vop resample, Eman2: e2proc3d.py)
 - ▶ Align subunits (Chimera: "fit in map", Eman2: e2foldhunter.py,)
 - ▶ Sum aligned subunits (Chimera: vop add, Eman2: e2proc3d.py)
 - ▶ Build a model for the average subunit (Eman2: e2pathwalker.py)
 - ▶ Refine model (Phenix: phenix.real_space_refine)



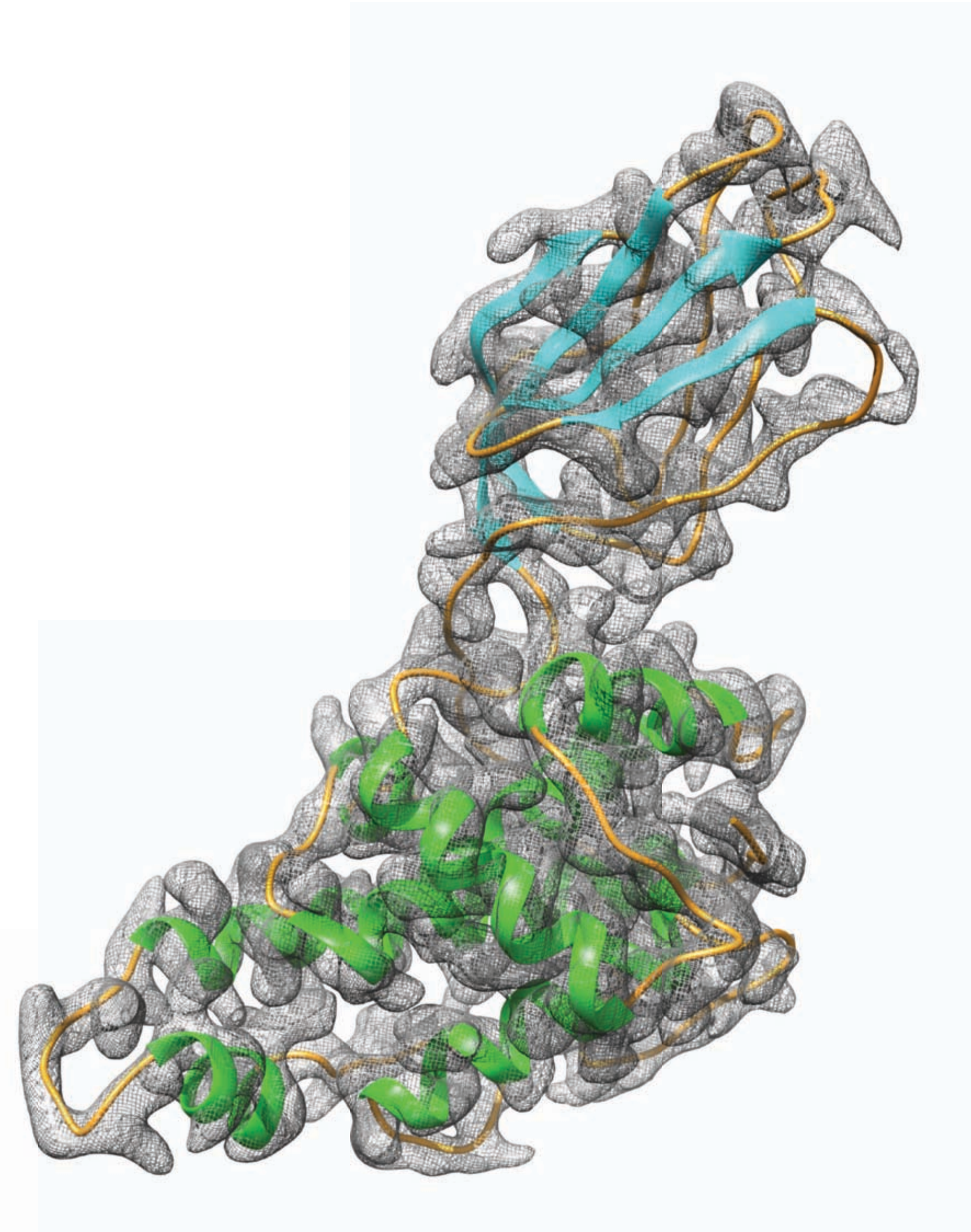
VP11 AVERAGE

- ▶ Building a model for VP11
 - ▶ Rough segmentation of single subunits (Chimera: "volume eraser" or "Segger", Eman2: e2segment3d.py)
 - ▶ Pad all subunits to the same size box (Chimera: vop resample, Eman2: e2proc3d.py)
 - ▶ Align subunits (Chimera: "fit in map", Eman2: e2foldhunter.py,)
 - ▶ Sum aligned subunits (Chimera: vop add, Eman2: e2proc3d.py)
 - ▶ Build a model for the average subunit (Eman2: e2pathwalker.py)
 - ▶ Refine model (Phenix: phenix.real_space_refine)



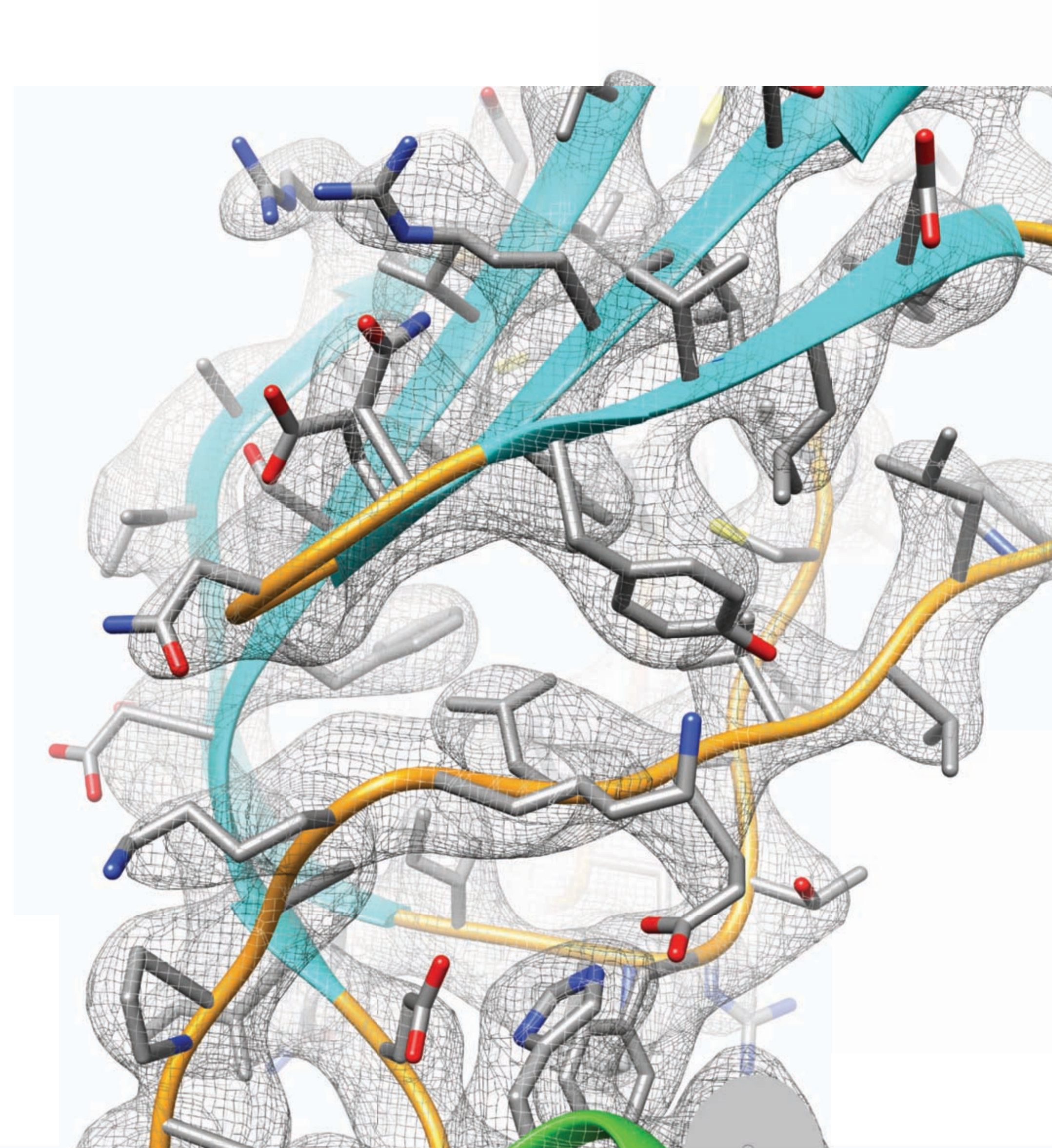
VP11 AVERAGE

- ▶ Building a model for VP11
 - ▶ Rough segmentation of single subunits (Chimera: "volume eraser" or "Segger", Eman2: e2segment3d.py)
 - ▶ Pad all subunits to the same size box (Chimera: vop resample, Eman2: e2proc3d.py)
 - ▶ Align subunits (Chimera: "fit in map", Eman2: e2foldhunter.py,)
 - ▶ Sum aligned subunits (Chimera: vop add, Eman2: e2proc3d.py)
 - ▶ Build a model for the average subunit (Eman2: e2pathwalker.py)
 - ▶ Refine model (Phenix: phenix.real_space_refine)



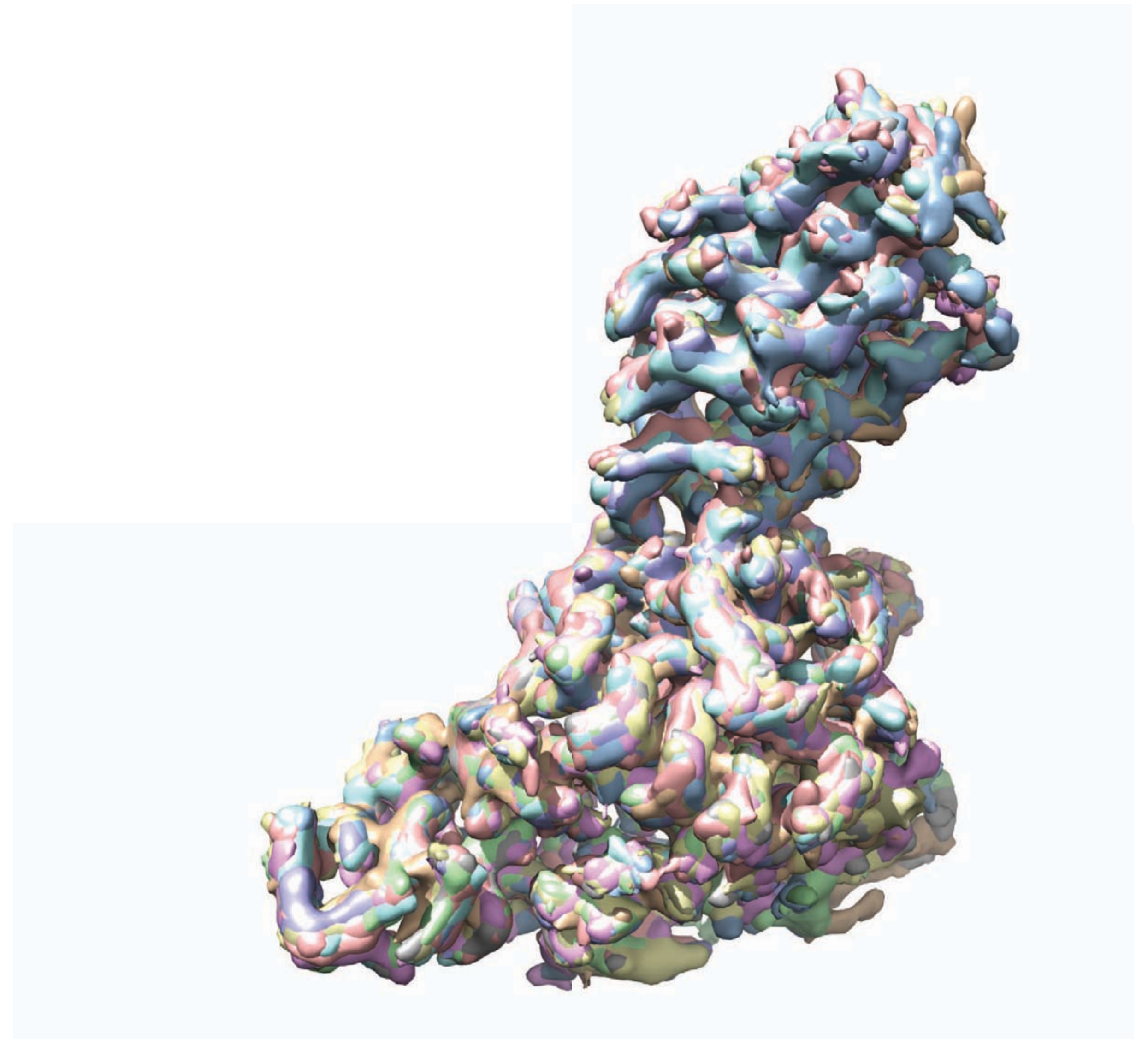
VP11 AVERAGE

- ▶ Building a model for VP11
 - ▶ Rough segmentation of single subunits (Chimera: "volume eraser" or "Segger", Eman2: e2segment3d.py)
 - ▶ Pad all subunits to the same size box (Chimera: vop resample, Eman2: e2proc3d.py)
 - ▶ Align subunits (Chimera: "fit in map", Eman2: e2foldhunter.py,)
 - ▶ Sum aligned subunits (Chimera: vop add, Eman2: e2proc3d.py)
 - ▶ Build a model for the average subunit (Eman2: e2pathwalker.py)
 - ▶ Refine model (Phenix: phenix.real_space_refine)



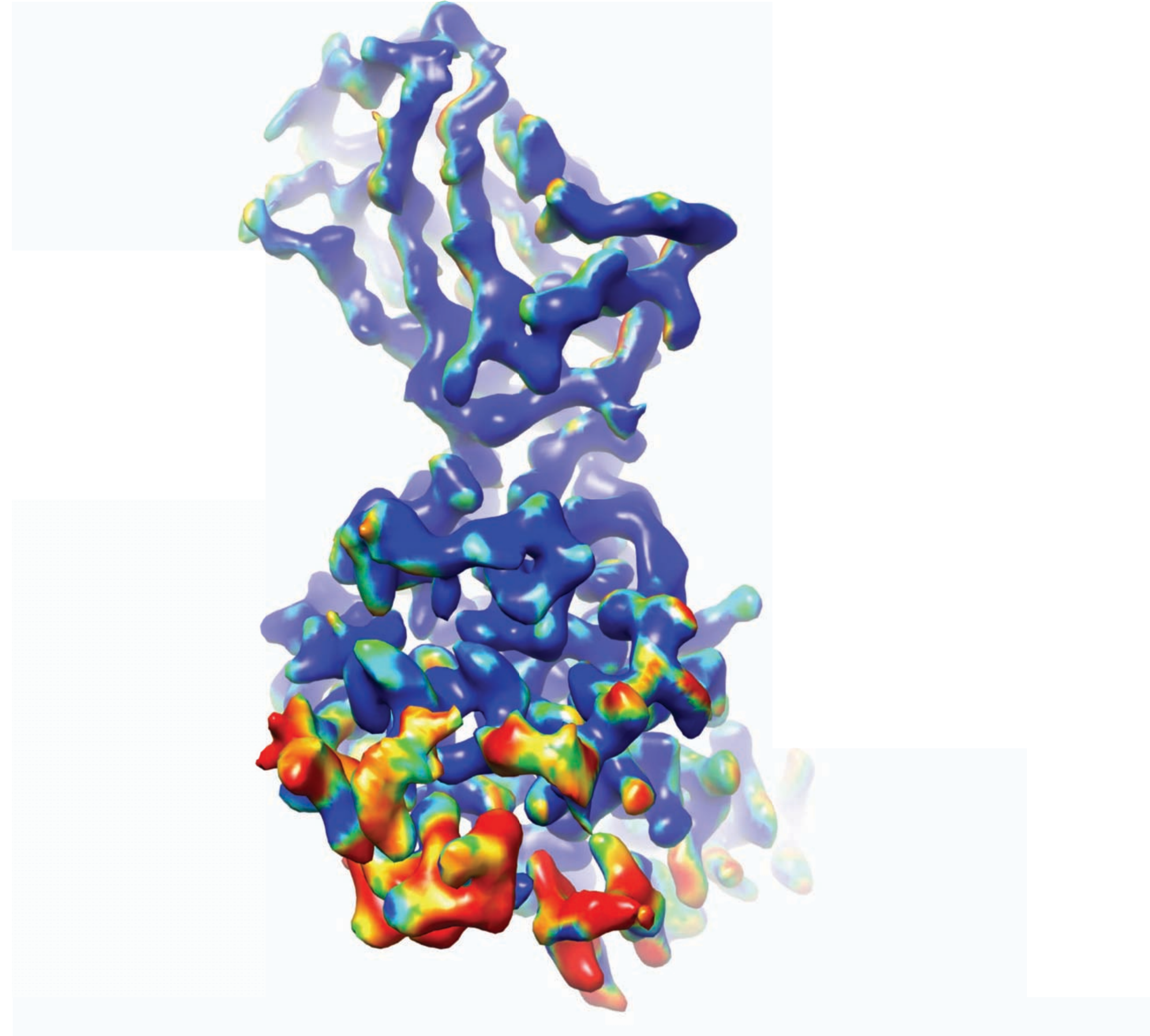
QUASI-EQUIVALENCE?

- ▶ Aligned maps are similar but have “real” differences at subunit interfaces
- ▶ Difference map shows a region corresponding to a loop is re-arranged in at least 2 VP11 subunits when compared to average (UCSF Chimera “Surface color”)



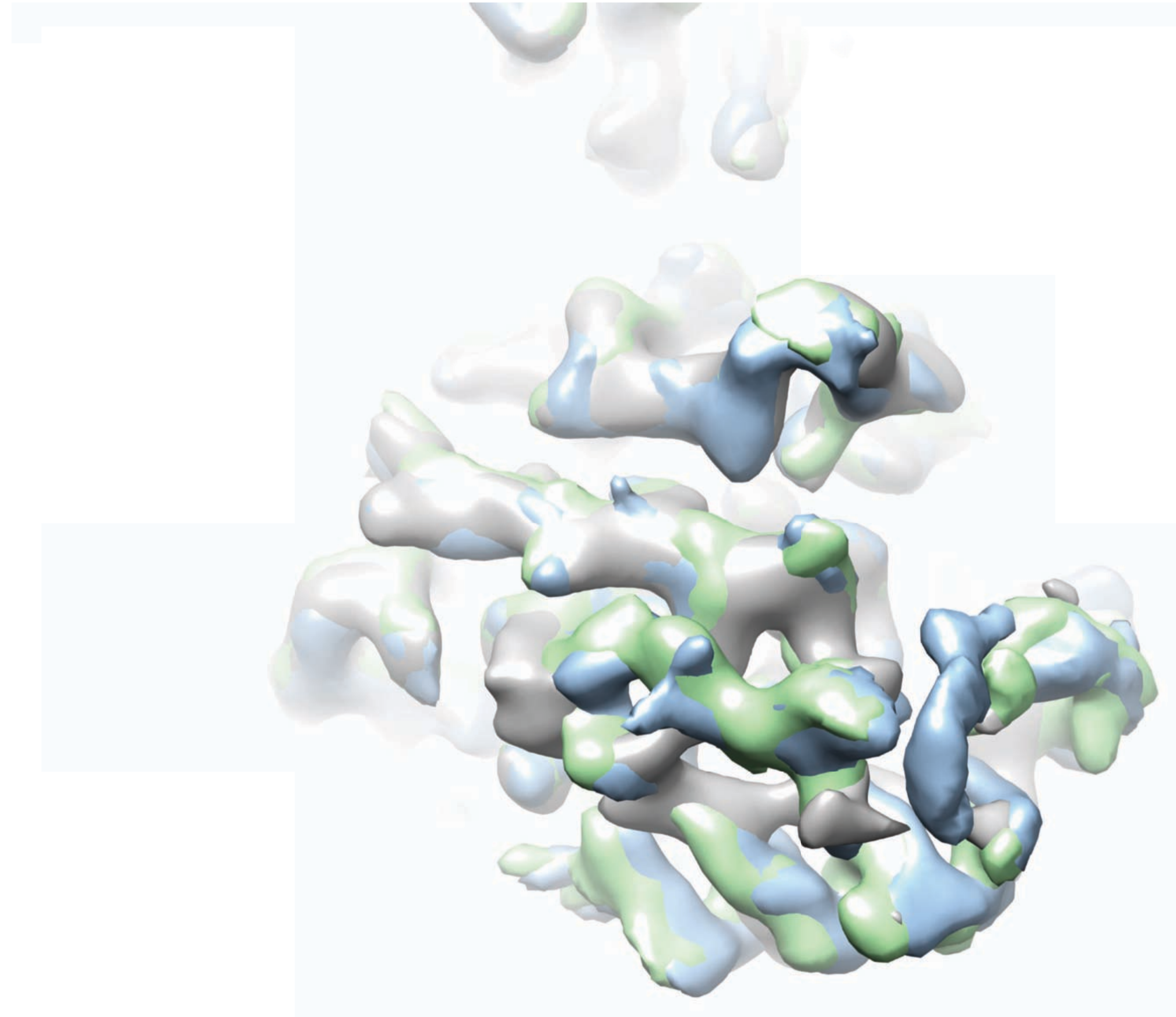
QUASI-EQUIVALENCE?

- ▶ Aligned maps are similar but have “real” differences at subunit interfaces
- ▶ Difference map shows a region corresponding to a loop is re-arranged in at least 2 VP11 subunits when compared to average (UCSF Chimera “Surface color)



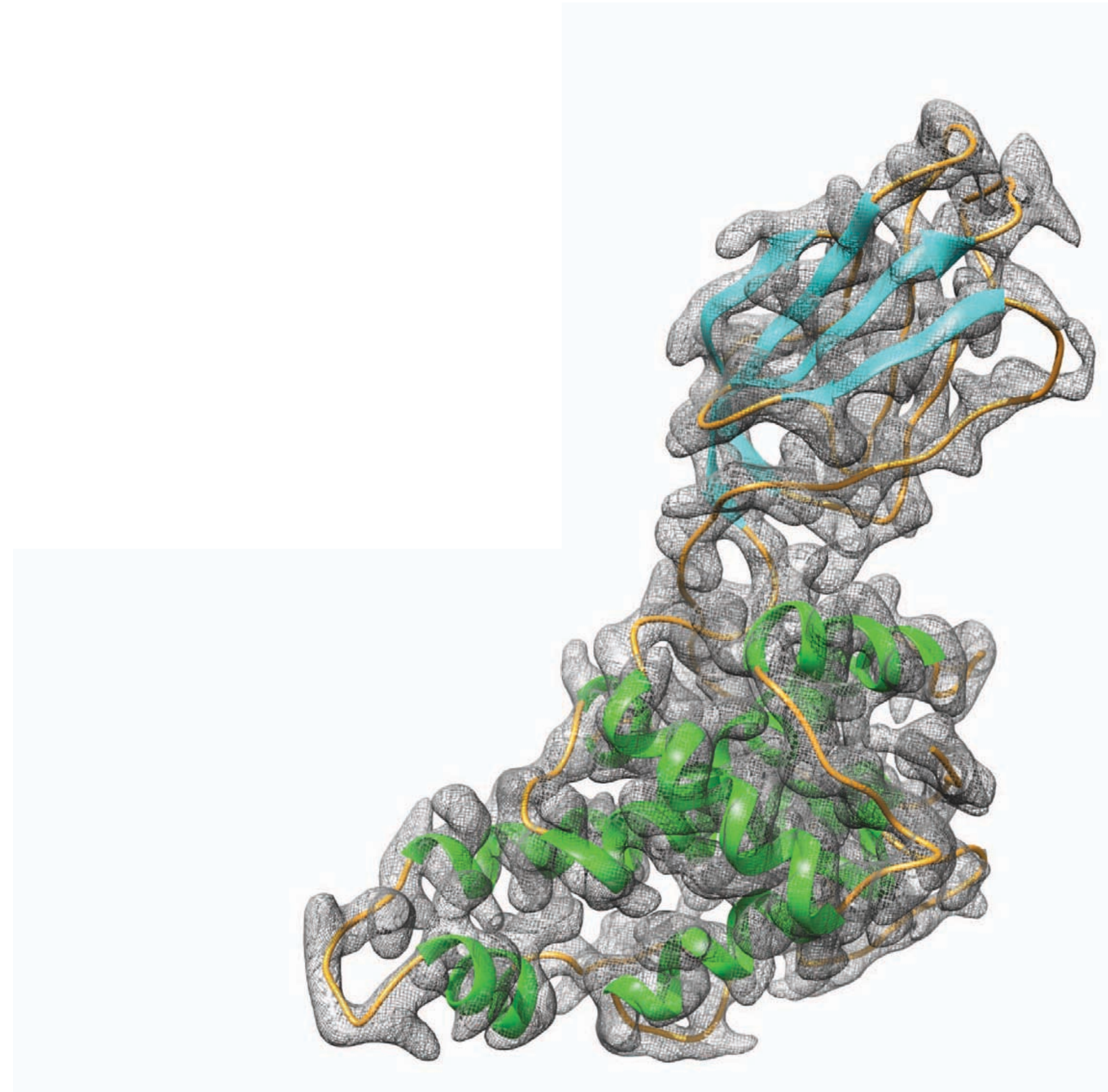
QUASI-EQUIVALENCE?

- ▶ Aligned maps are similar but have “real” differences at subunit interfaces
- ▶ Difference map shows a region corresponding to a loop is re-arranged in at least 2 VP11 subunits when compared to average (UCSF Chimera “Surface color)



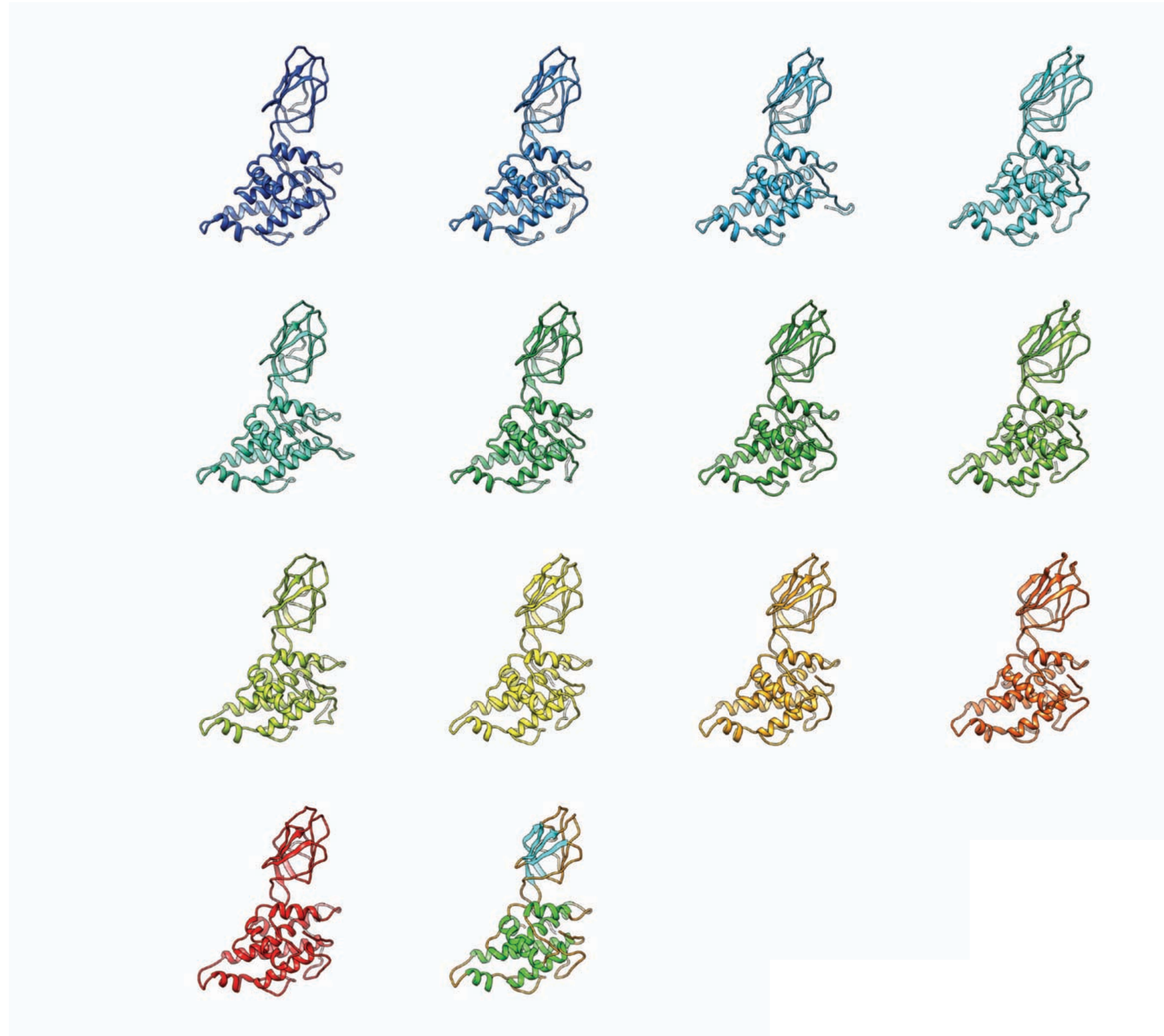
13 MODELS FOR VP11

- ▶ Starting with averaged VP11 model
 - ▶ Fit an averaged VP11 subunit into the density (Chimera "fit in map")
 - ▶ Refine model with Phenix (phenix.real_space_refine run=minimization_global+simulated_annealing+adp)
- ▶ RMSD between average and individual subunit models (Chimera "Matchmaker")
 - ▶ ~1Å in 11 of 13 subunits
 - ▶ >2.5Å RMSD in 2 subunits



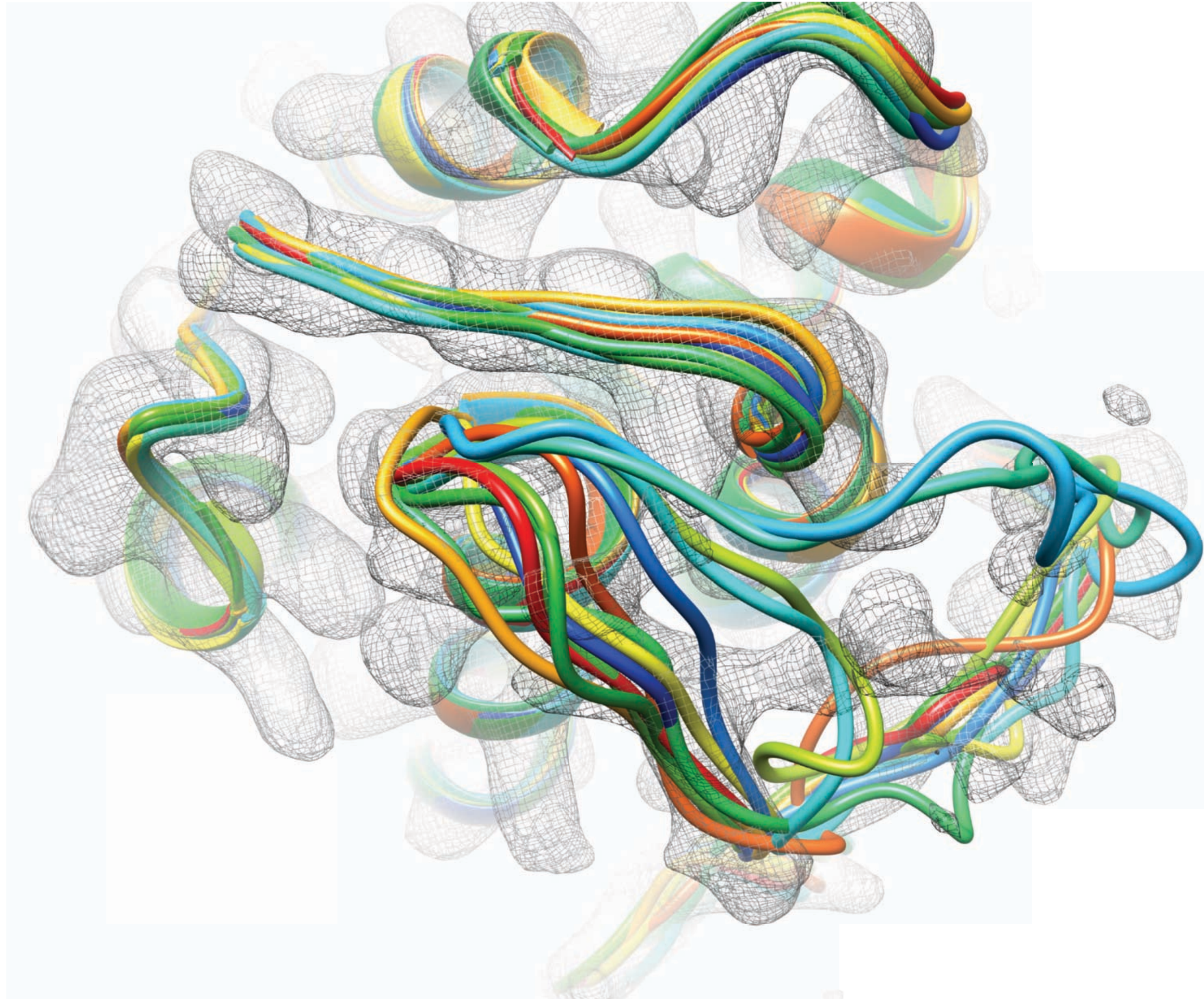
13 MODELS FOR VP11

- ▶ Starting with averaged VP11 model
 - ▶ Fit an averaged VP11 subunit into the density (Chimera "fit in map")
 - ▶ Refine model with Phenix (phenix.real_space_refine run=minimization_global+simulated_annealing+adp)
- ▶ RMSD between average and individual subunit models (Chimera "Matchmaker")
 - ▶ ~1Å in 11 of 13 subunits
 - ▶ >2.5Å RMSD in 2 subunits

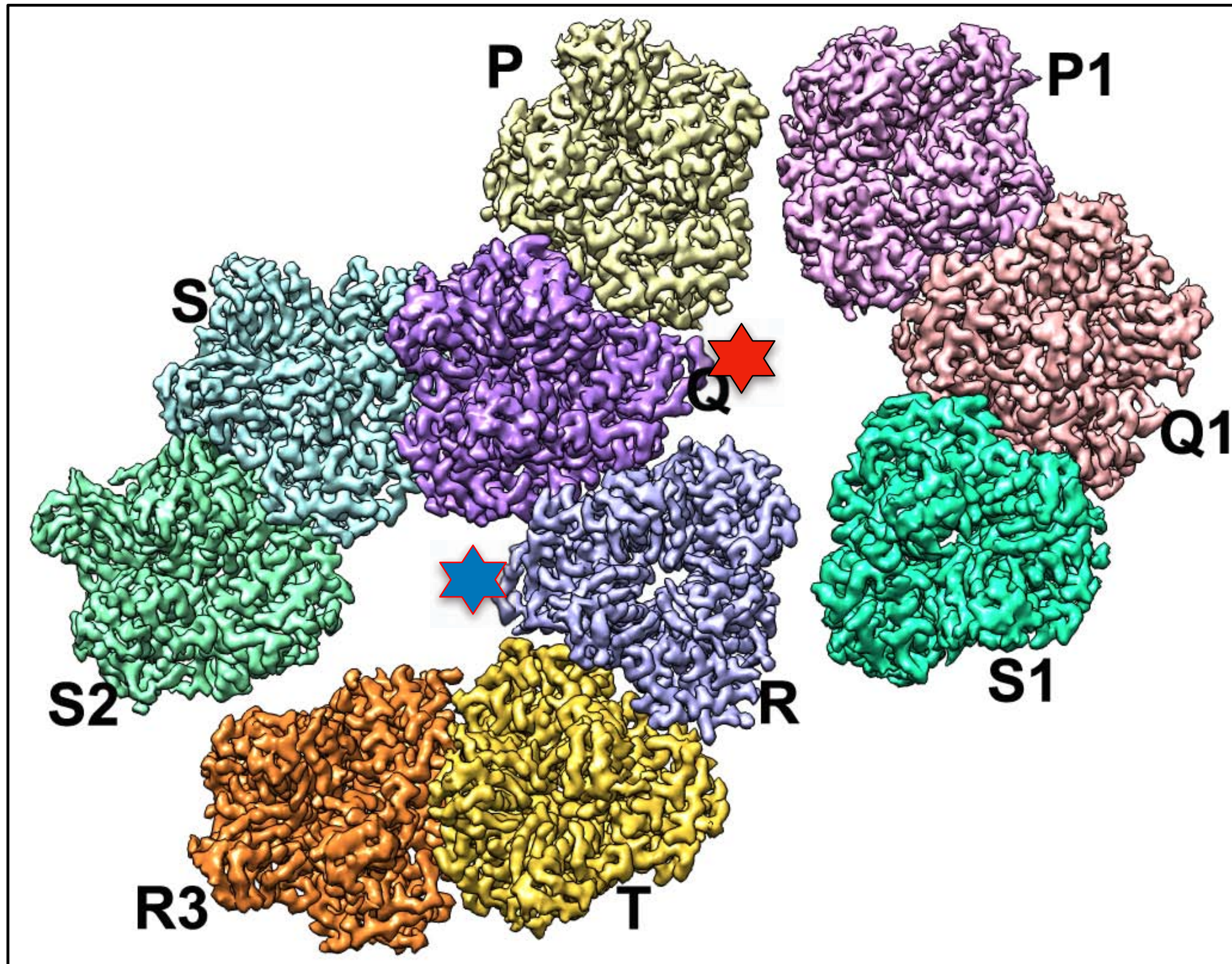


13 MODELS FOR VP11

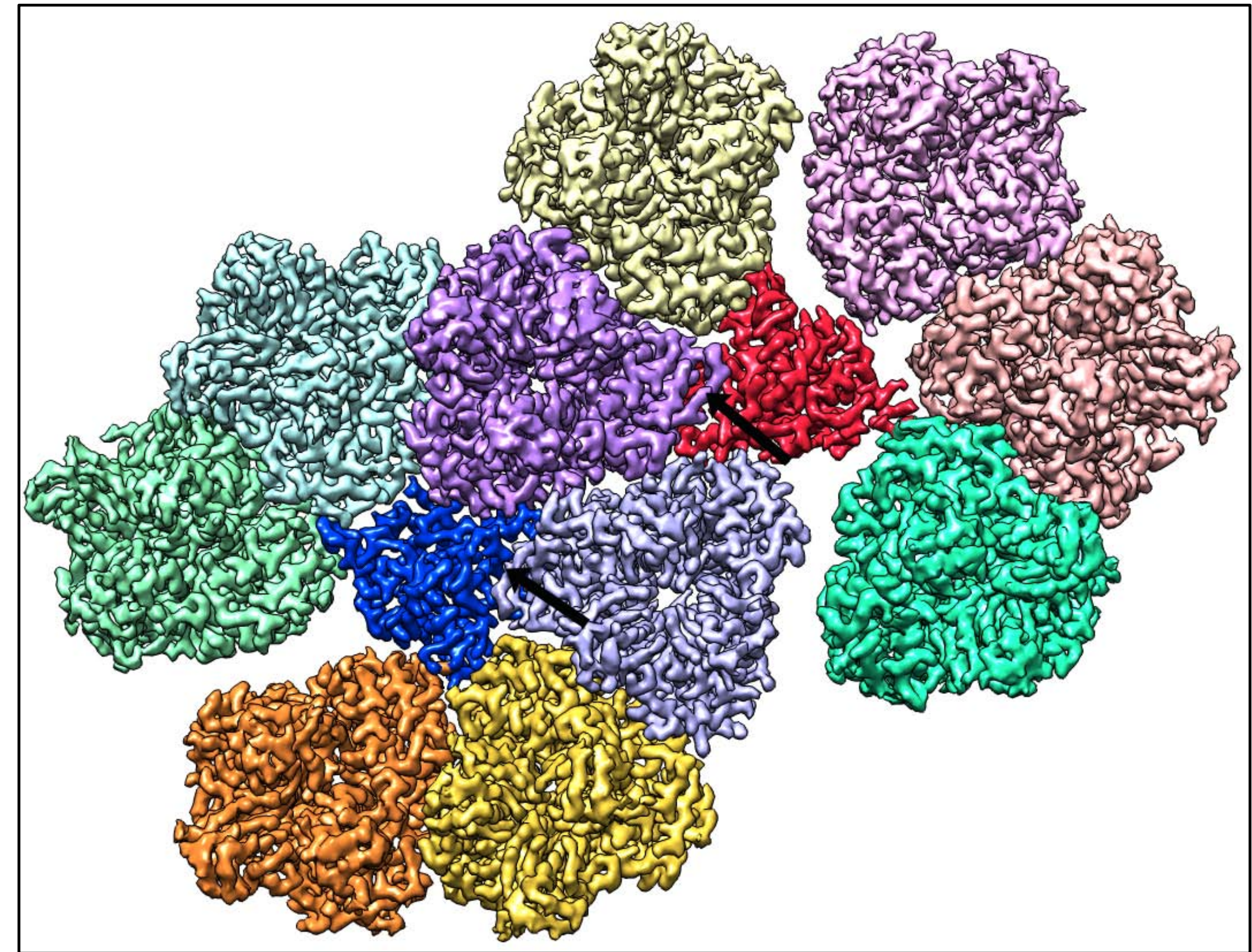
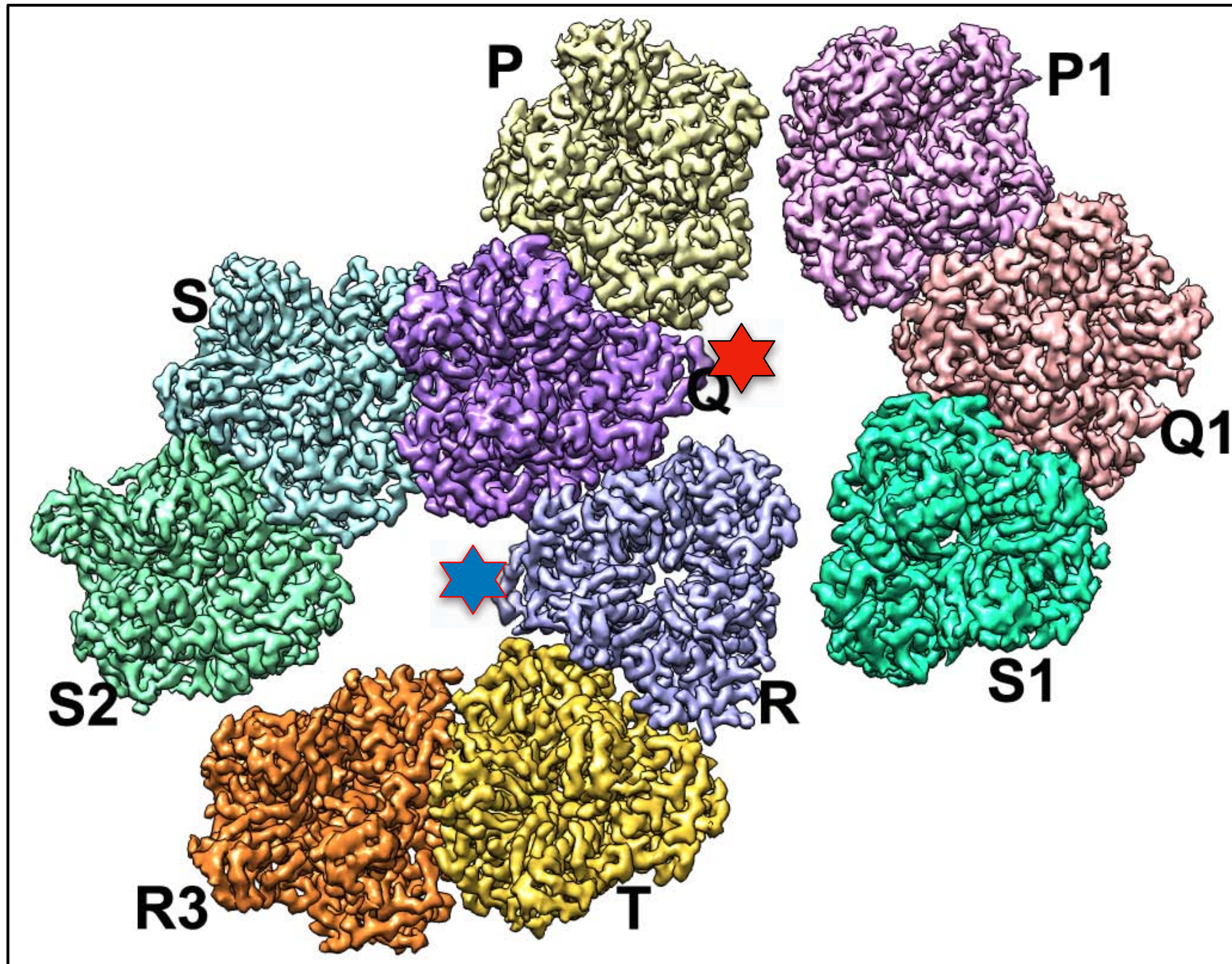
- ▶ Starting with averaged VP11 model
 - ▶ Fit an averaged VP11 subunit into the density (Chimera "fit in map")
 - ▶ Refine model with Phenix (`phenix.real_space_refine run=minimization_global+simulated_annealing+adp`)
- ▶ RMSD between average and individual subunit models (Chimera "Matchmaker")
 - ▶ $\sim 1\text{\AA}$ in 11 of 13 subunits
 - ▶ $>2.5\text{\AA}$ RMSD in 2 subunits



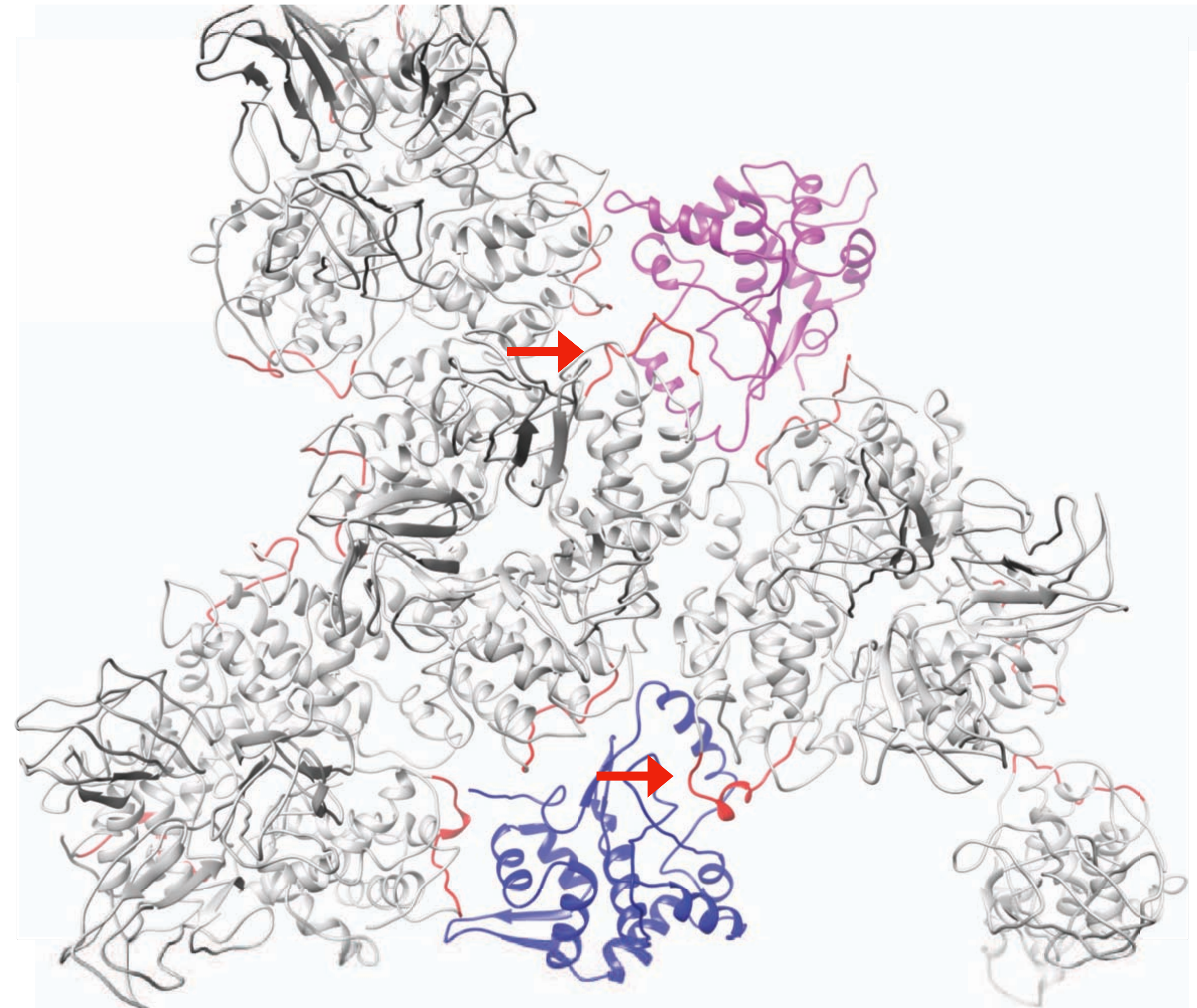
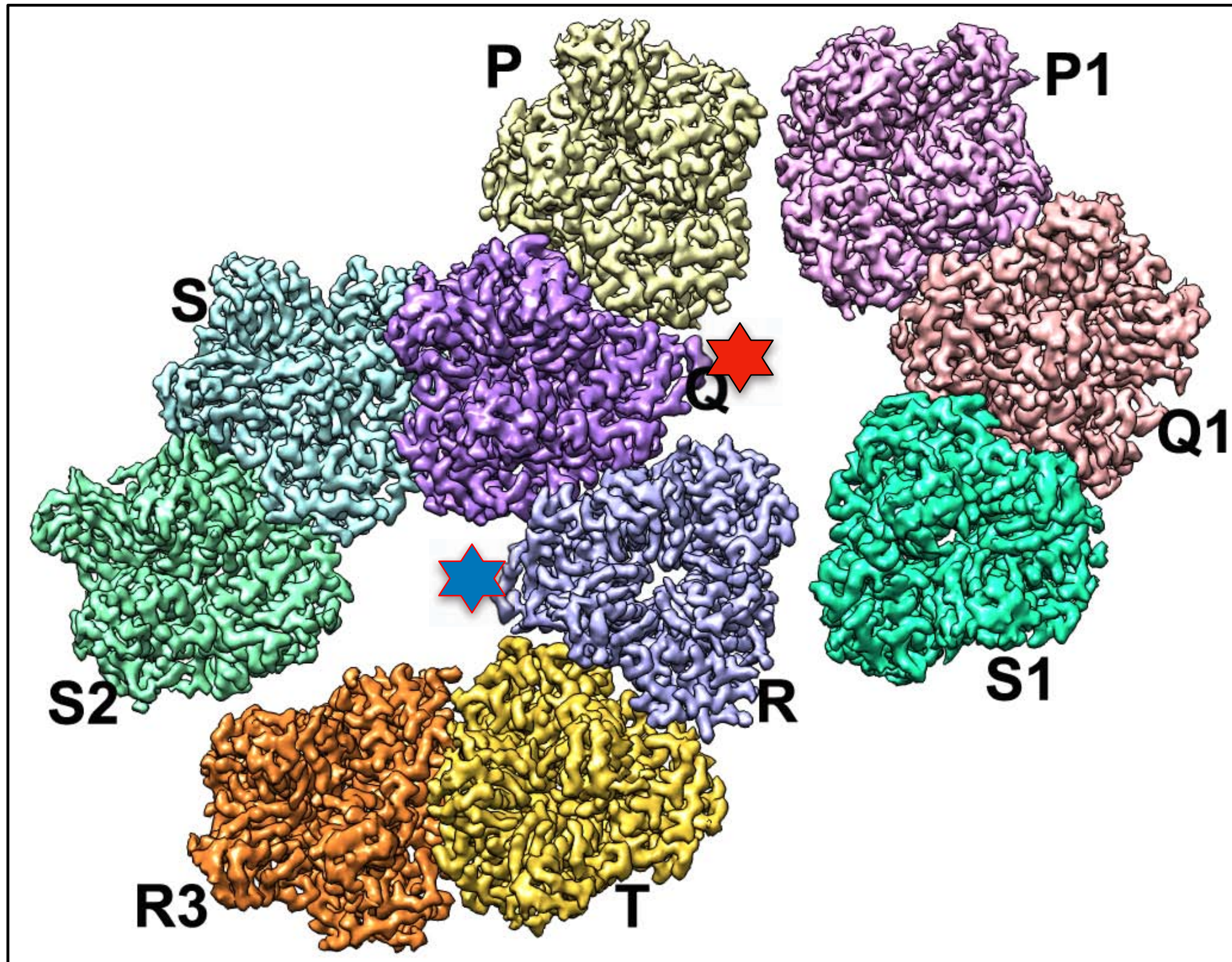
QUASI-EQUIVALENCE?



QUASI-EQUIVALENCE?

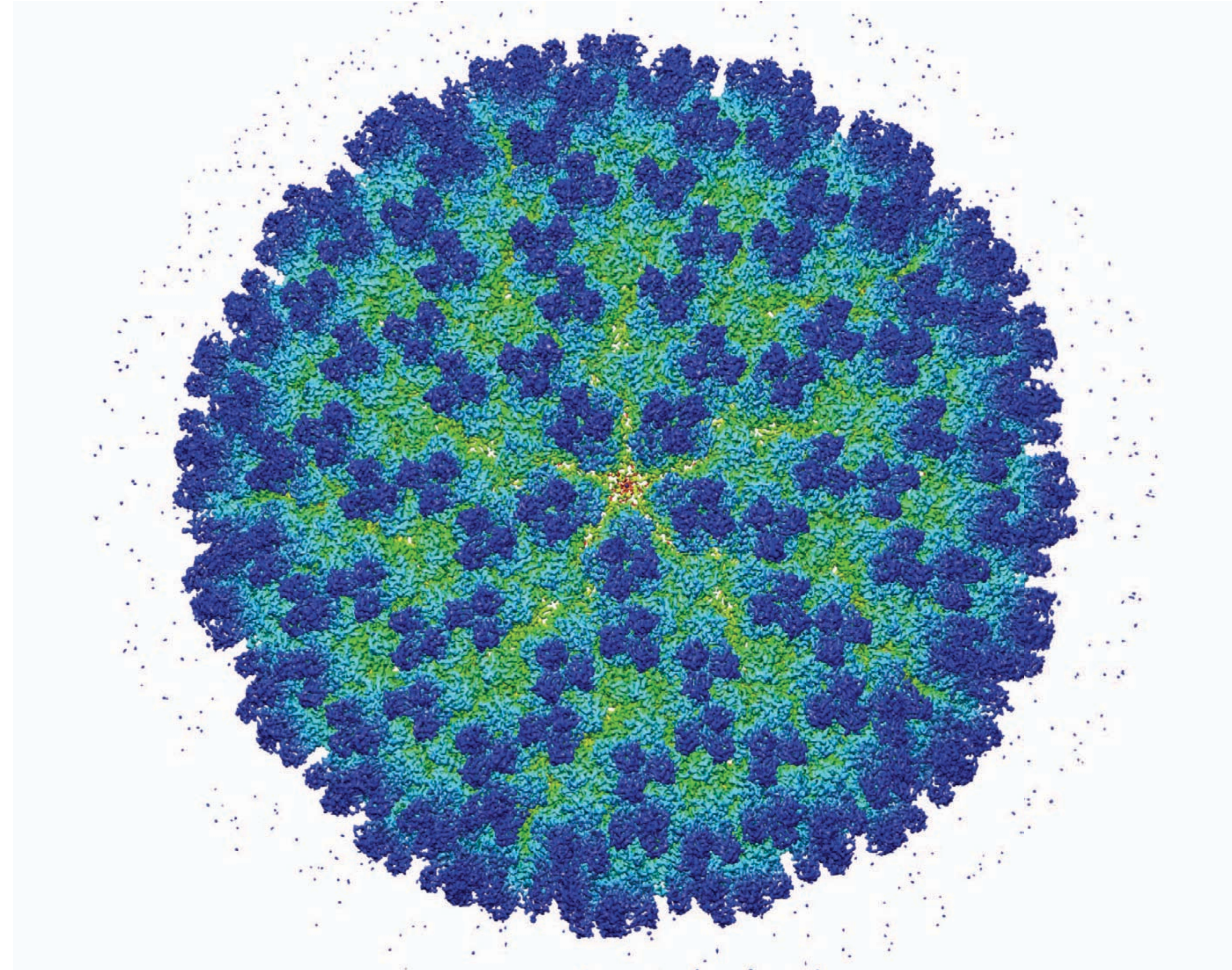


QUASI-EQUIVALENCE?



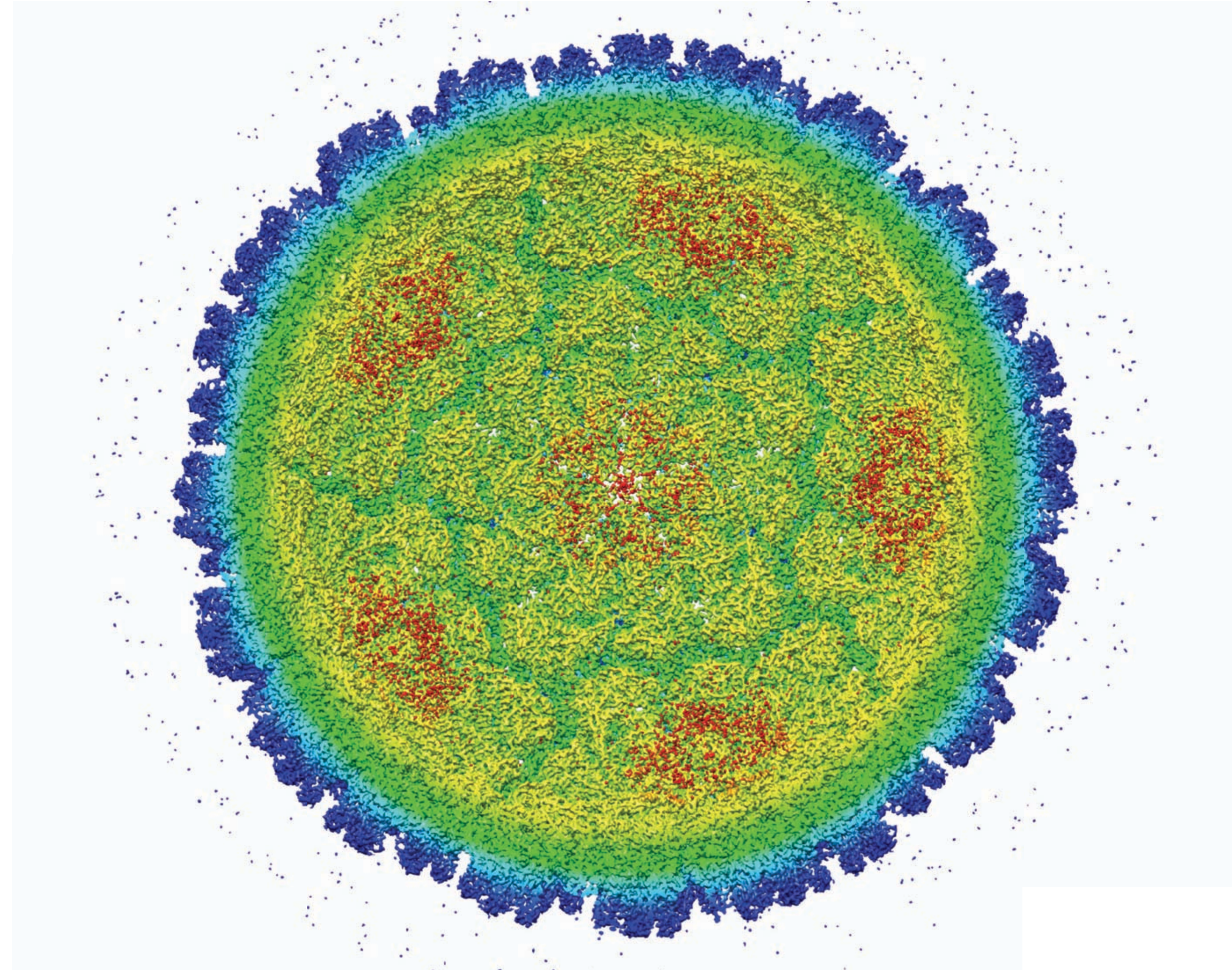
ICOSAHEDRAL SYMMETRY?

- ▶ Icosahedral symmetry used during reconstruction
- ▶ VP11/VP12 and VP3 layers appear to follow icosahedral symmetry
- ▶ RdRP should be located on the interior of the capsid near the 5-fold axis and surrounded by segmented dsRNA genome



ICOSAHEDRAL SYMMETRY?

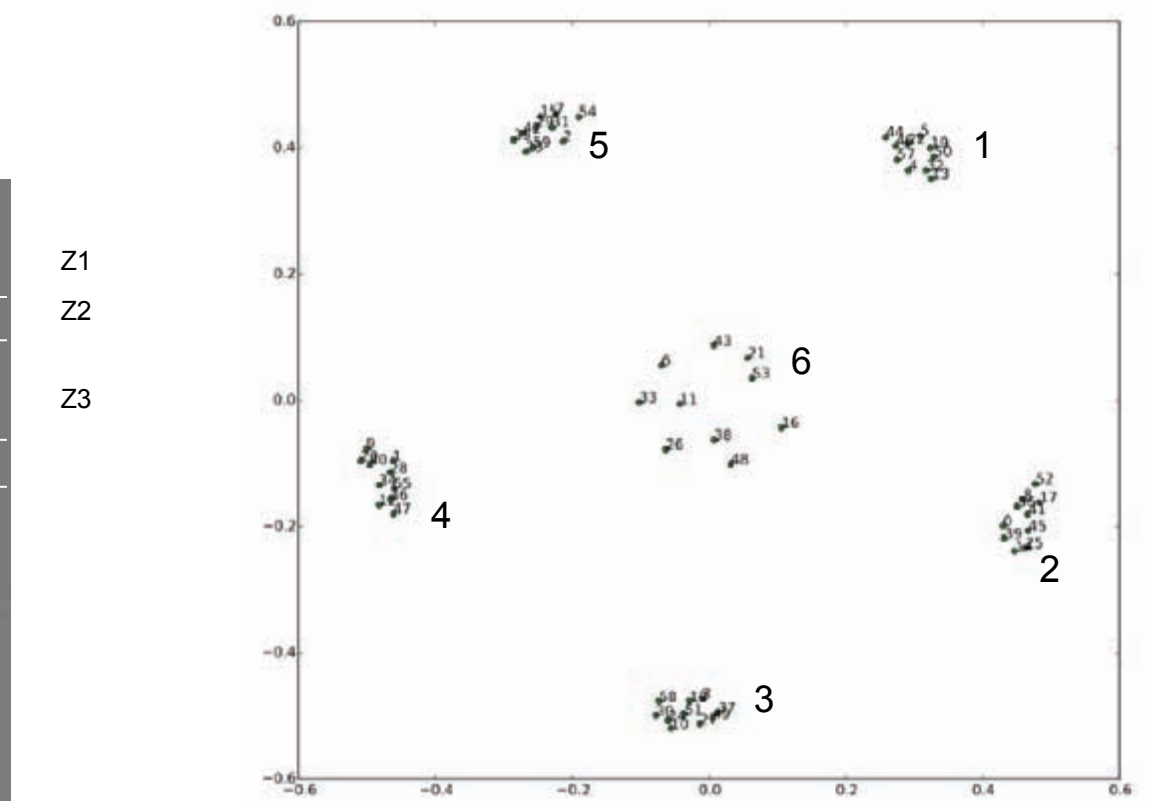
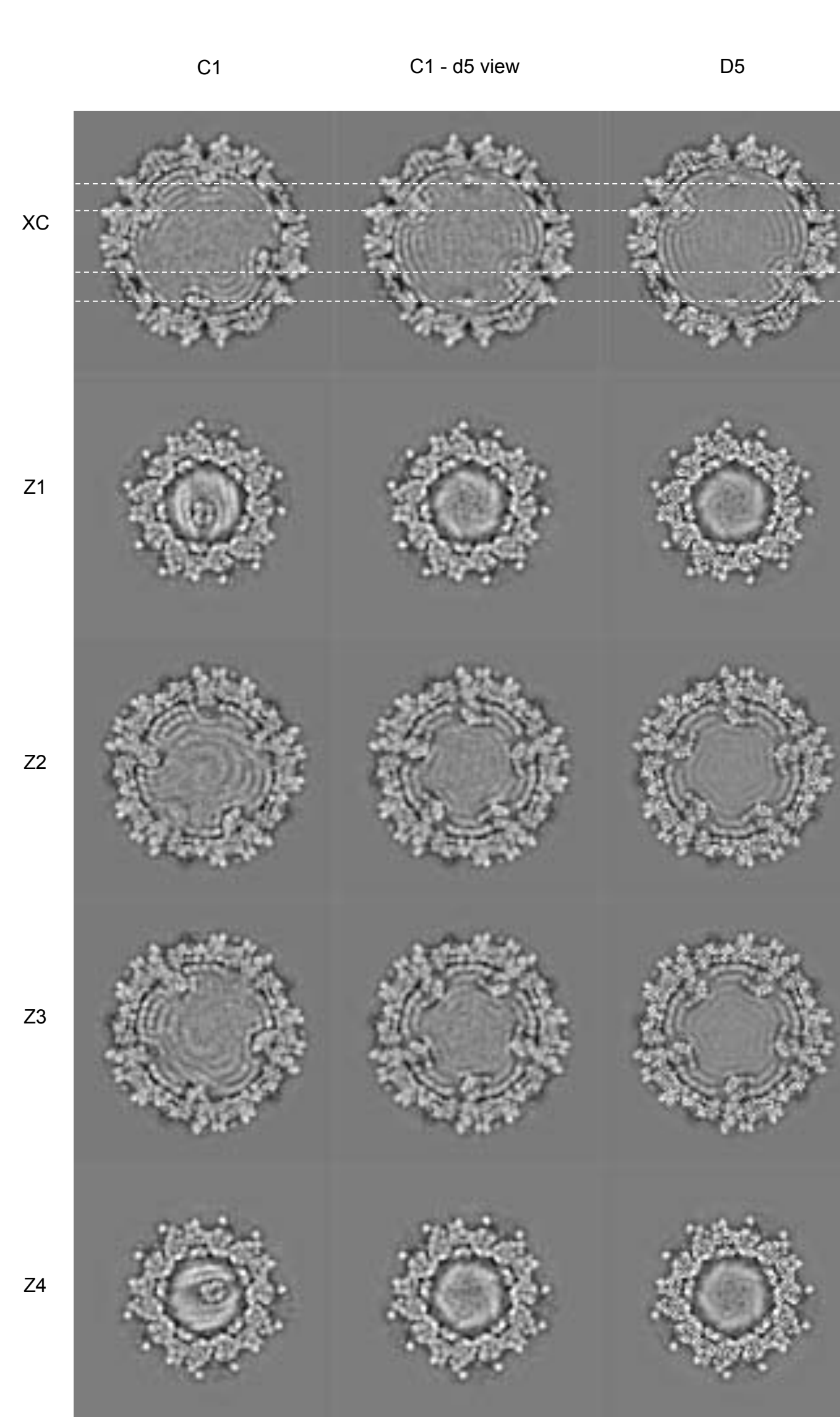
- ▶ Icosahedral symmetry used during reconstruction
- ▶ VP11/VP12 and VP3 layers appear to follow icosahedral symmetry
- ▶ RdRP should be located on the interior of the capsid near the 5-fold axis and surrounded by segmented dsRNA genome



DETERMINING SYMMETRY

▶ Asymmetric reconstruction

- ▶ SAR: SAR (without imposing any symmetry): iterative search for best one among the 60 views related by icosahedral symmetry (i.e., icos → c1 symmetry relax) that matches the model projection to the particle image
- ▶ FAR: extended SAR whereby limited features (i.e. ROI) are used for alignment in both 3D reference map and 2D images. Focusing on an ROI helps exclude the contribution of other regions to the alignment
- ▶ Symmetry relaxation and symmetry search done with JSPR and EMAN2

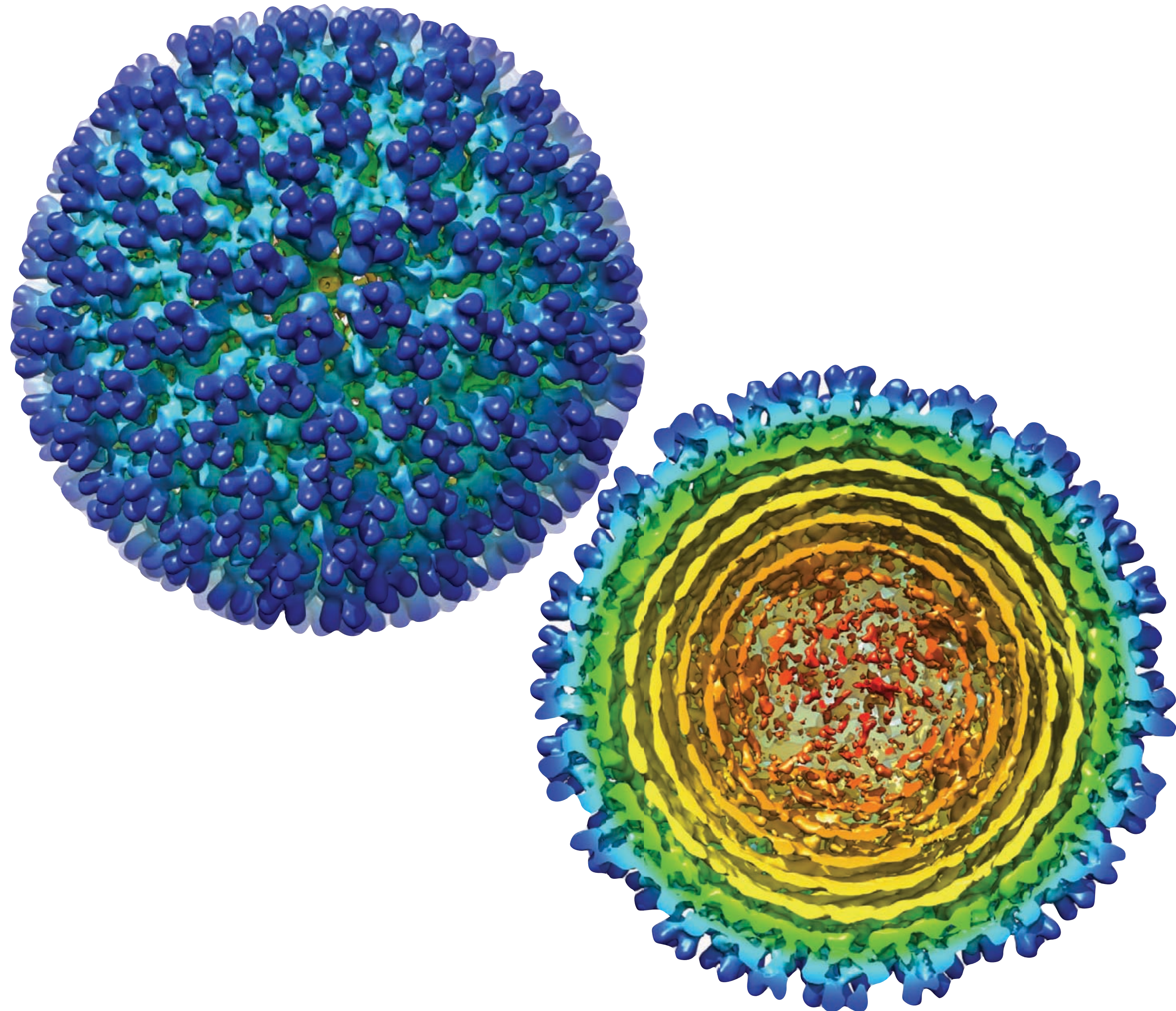


Cluster	Sym (0-59, Icosahedral)	Counts of Angles Between Sym Pairs		
		180°	144°	72°
1	4, 5, 13, 19, 22, 32, 44, 46, 50, 57	25	10	10
2	0, 8, 14, 17, 25, 39, 41, 45, 52, 56	25	10	10
3	3, 10, 18, 24, 27, 30, 37, 49, 51, 58	25	10	10
4	1, 9, 12, 20, 28, 34, 36, 40, 47, 55	25	10	10
5	2, 7, 15, 23, 29, 31, 35, 42, 54, 59	25	10	10
6	6, 11, 16, 21, 26, 33, 38, 43, 48, 53	25	10	10

Sym	# Sym	Counts of Angles Between Sym Pairs					
		180°	144°	120°	108°	90°	72°
icos	60	450	360	600			360
tet	12	18		48			
d5	10	25	10				10
c10	10	5	10		10		10

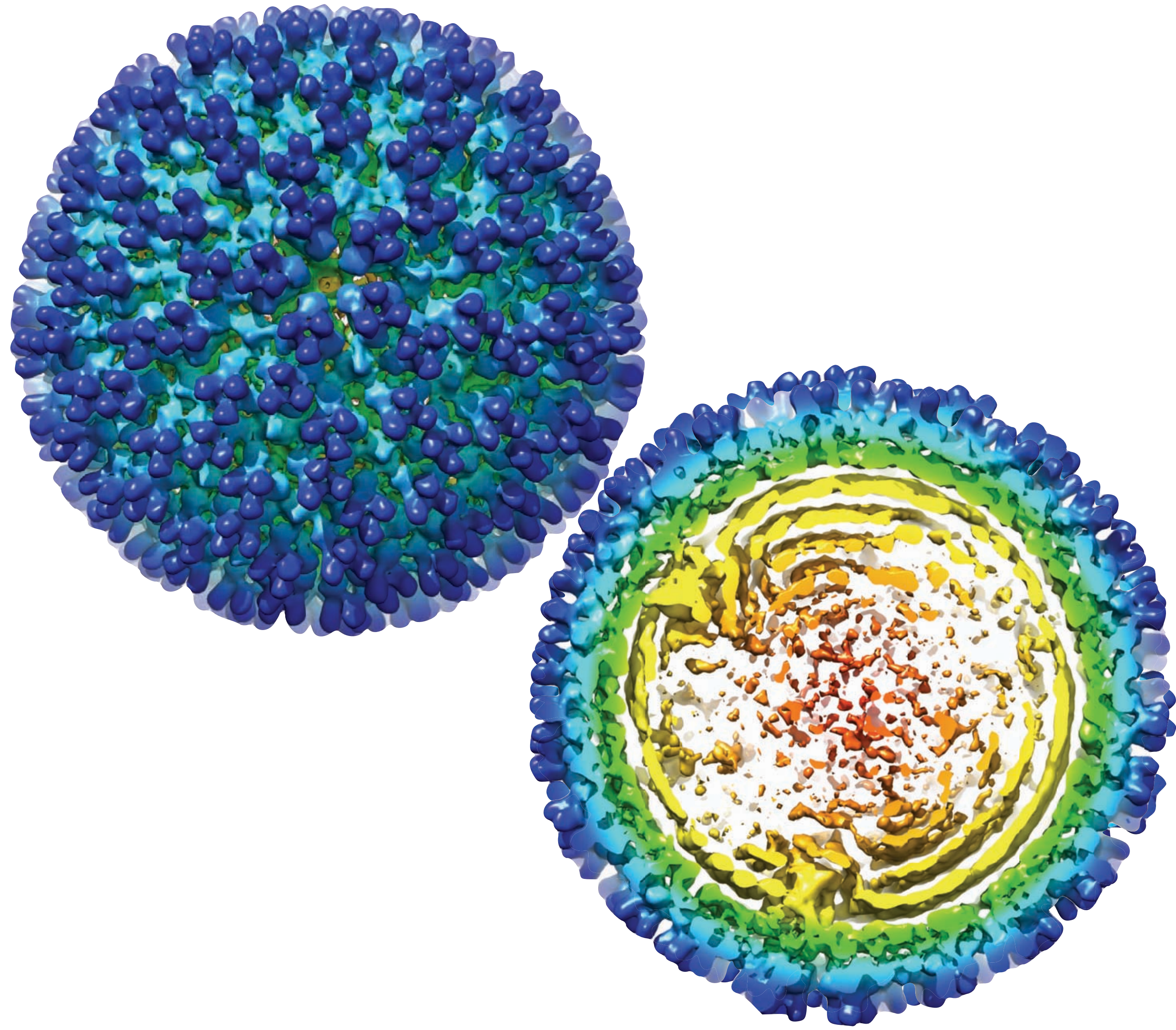
MCRV WITH D5 SYMMETRY

- ▶ Same particle data as icosahedral reconstruction
- ▶ C1 reconstruction: VP3, VP11 and VP12 densities still icosahedral
 - ▶ RdRP density observed offset from 5-fold axis
 - ▶ dsRNA genome clearly visible
 - ▶ ~10Å resolution
- ▶ D5 focused asymmetric reconstruction:
 - ▶ Resolved RdRP and genome
 - ▶ 3.4Å, 3.7Å resolution



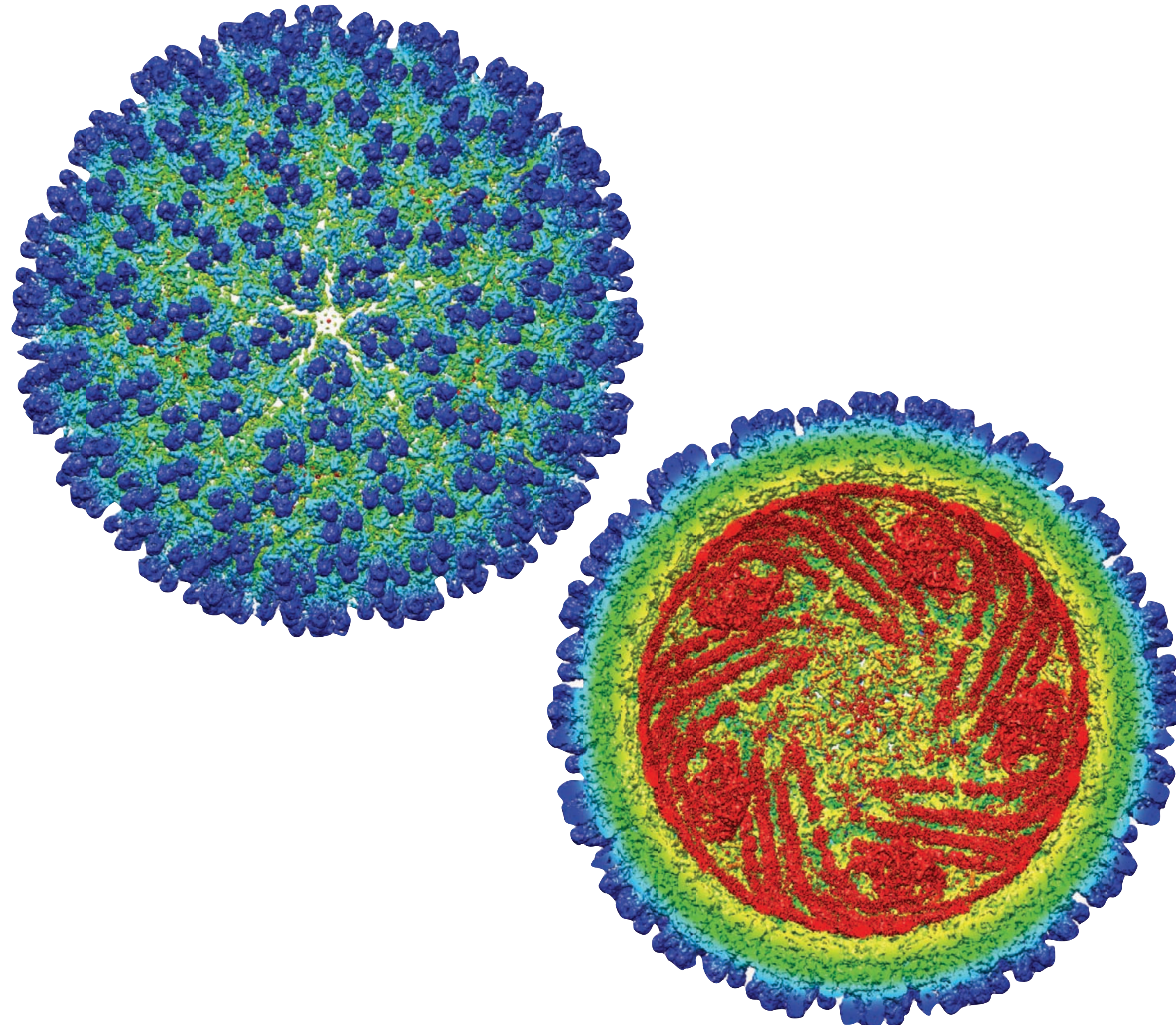
MCRV WITH D5 SYMMETRY

- ▶ Same particle data as icosahedral reconstruction
- ▶ C1 reconstruction: VP3, VP11 and VP12 densities still icosahedral
 - ▶ RdRP density observed offset from 5-fold axis
 - ▶ dsRNA genome clearly visible
 - ▶ ~10Å resolution
- ▶ D5 focused asymmetric reconstruction:
 - ▶ Resolved RdRP and genome
 - ▶ 3.4Å, 3.7Å resolution



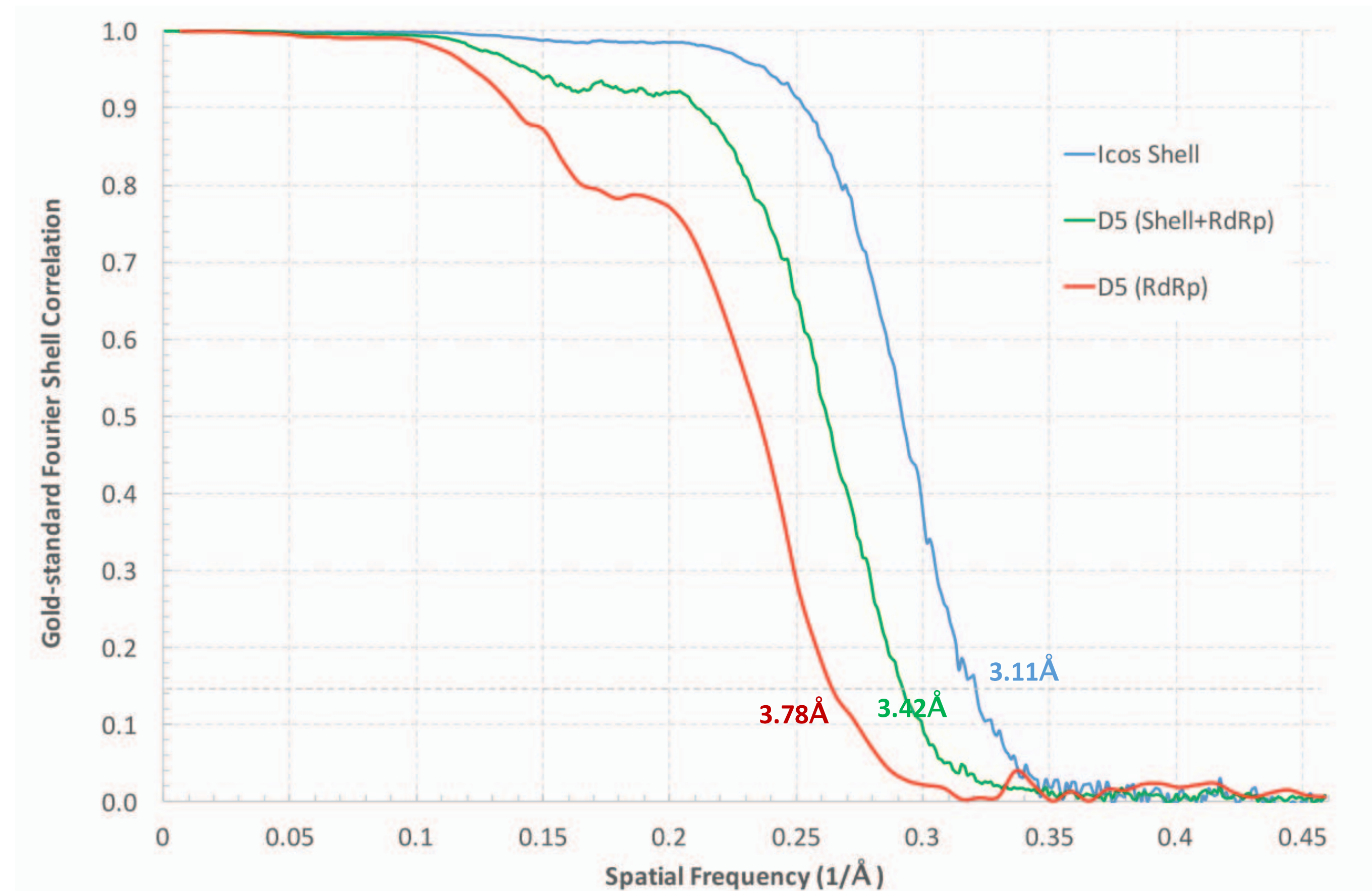
MCRV WITH D5 SYMMETRY

- ▶ Same particle data as icosahedral reconstruction
- ▶ C1 reconstruction: VP3, VP11 and VP12 densities still icosahedral
 - ▶ RdRP density observed offset from 5-fold axis
 - ▶ dsRNA genome clearly visible
 - ▶ ~10Å resolution
- ▶ D5 focused asymmetric reconstruction:
 - ▶ Resolved RdRP and genome
 - ▶ 3.4Å, 3.7Å resolution



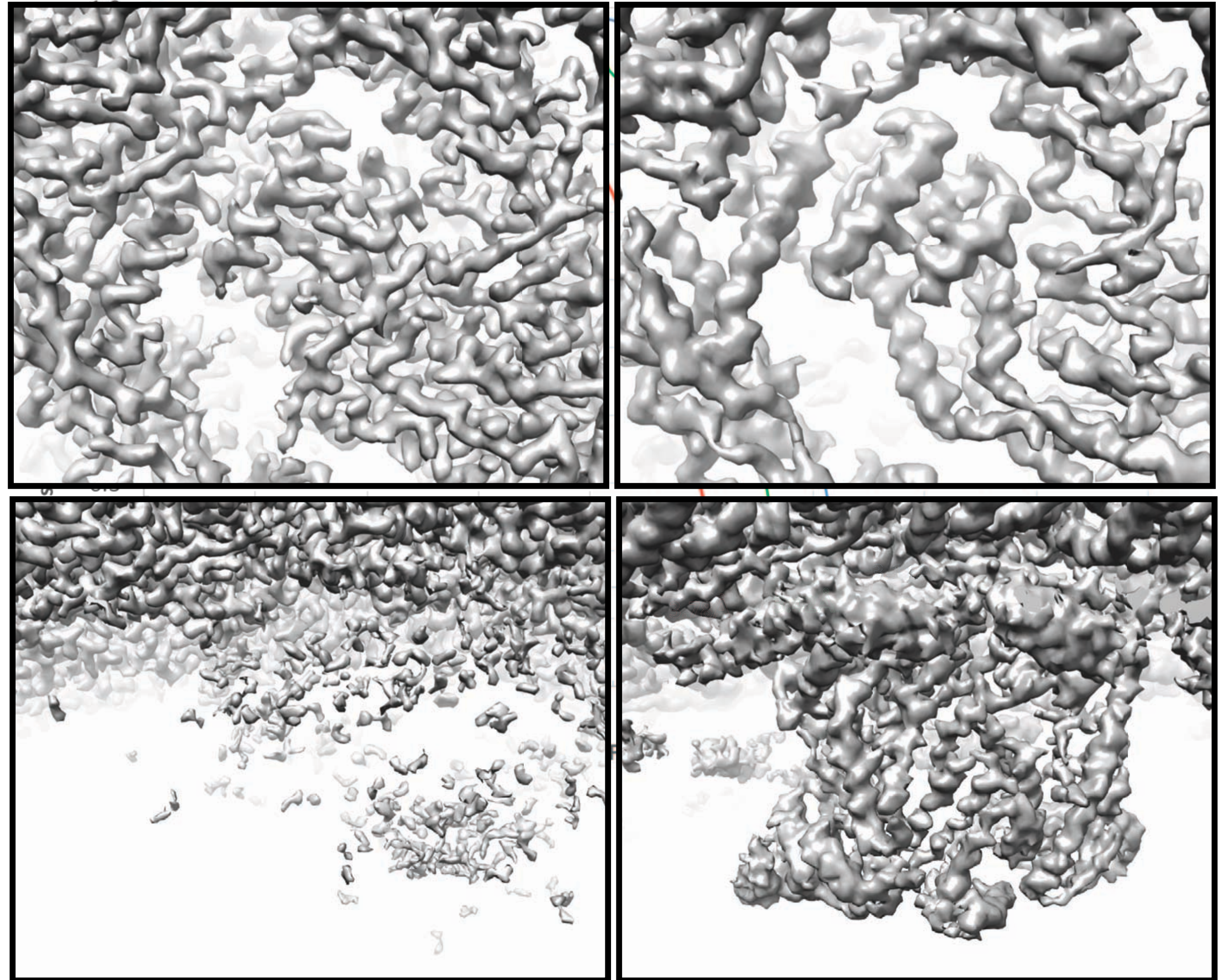
RESOLUTION VS RESOLVABILITY

- ▶ C1/D5 reconstructions lose $\sim 0.5\text{\AA}$ resolution as compared to icosahedral reconstruction
- ▶ C1/D5 reconstructions still have sufficient detail
 - ▶ Model relaxation: fit models from icosahedral reconstructions and refine using Phenix (phenix.real_space_refine run=minimization_global+simulated_annealing+adp)
 - ▶ De novo model building (Pathwalking + Phenix + Coot) for RdRP model building
 - ▶ RNA observed interacting with RdRP



RESOLUTION VS RESOLVABILITY

- ▶ C1/D5 reconstructions lose $\sim 0.5\text{\AA}$ resolution as compared to icosahedral reconstruction
- ▶ C1/D5 reconstructions still have sufficient detail
 - ▶ Model relaxation: fit models from icosahedral reconstructions and refine using Phenix (phenix.real_space_refine run=minimization_global+simulated_annealing+adp)
 - ▶ De novo model building (Pathwalking + Phenix + Coot) for RdRP model building
 - ▶ RNA observed interacting with RdRP

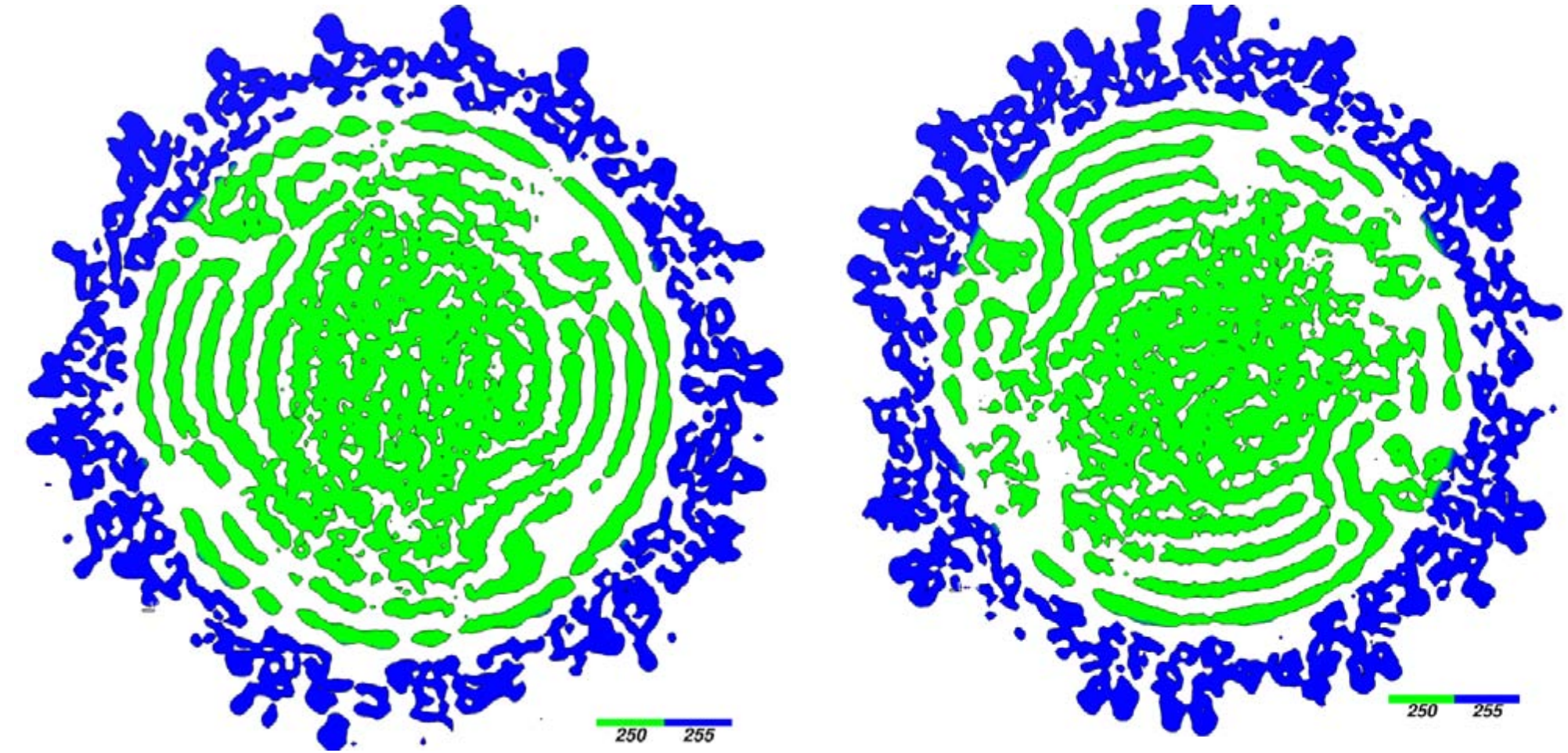


icos

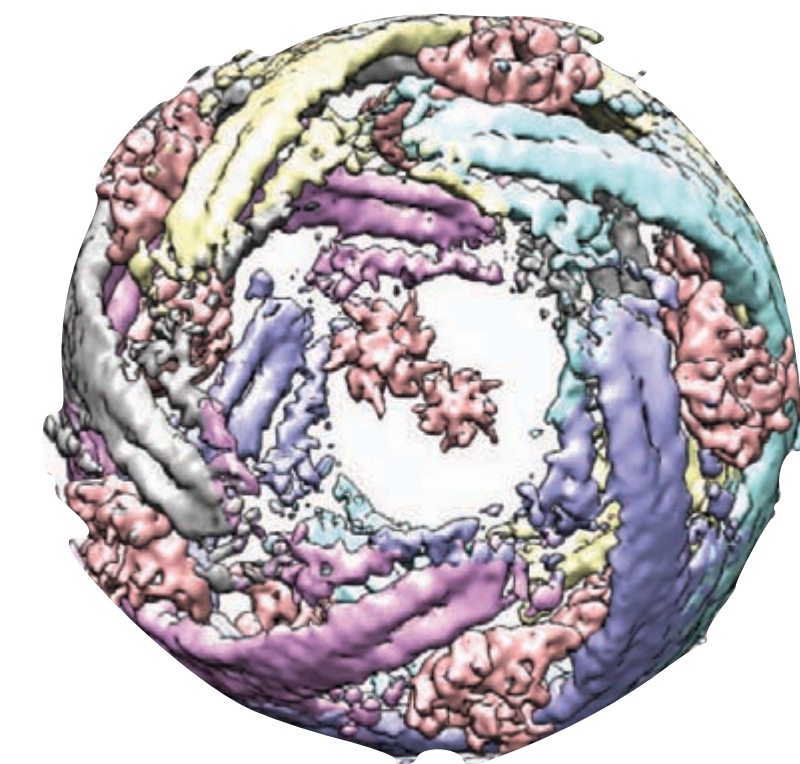
D5

DSRNA GENOME

- ▶ 7 layers of segmented dsRNA genome visible
- ▶ dsRNA interacts and may help to position RdRP
- ▶ ssRNA seen entering transcribing MCRV
- ▶ Nucleotides seen inside RdRP near priming loop



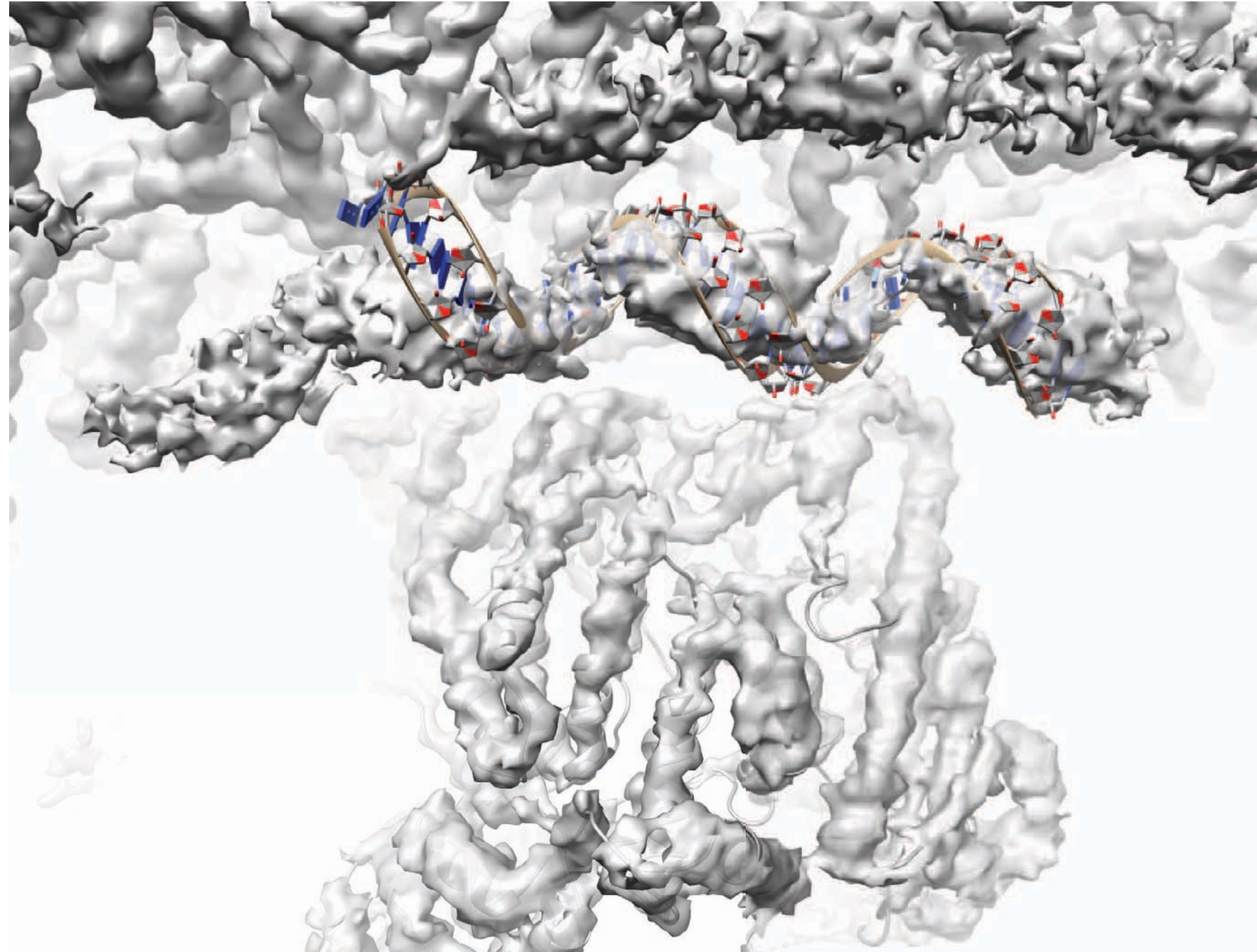
1st layer



2nd layer

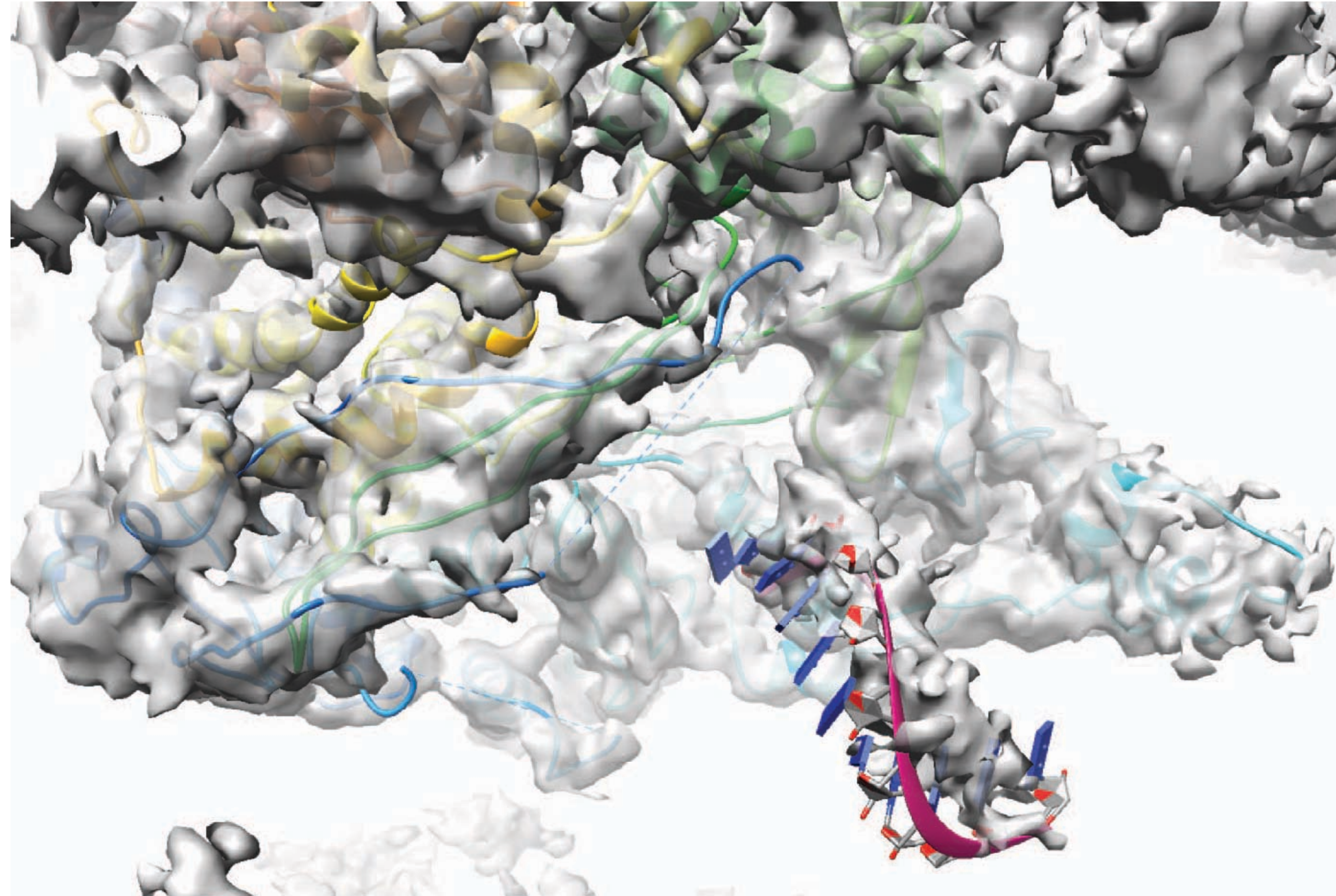
DSRNA GENOME

- ▶ 7 layers of segmented dsRNA genome visible
- ▶ dsRNA interacts and may help to position RdRP
- ▶ ssRNA seen entering transcribing MCRV
- ▶ Nucleotides seen inside RdRP near priming loop



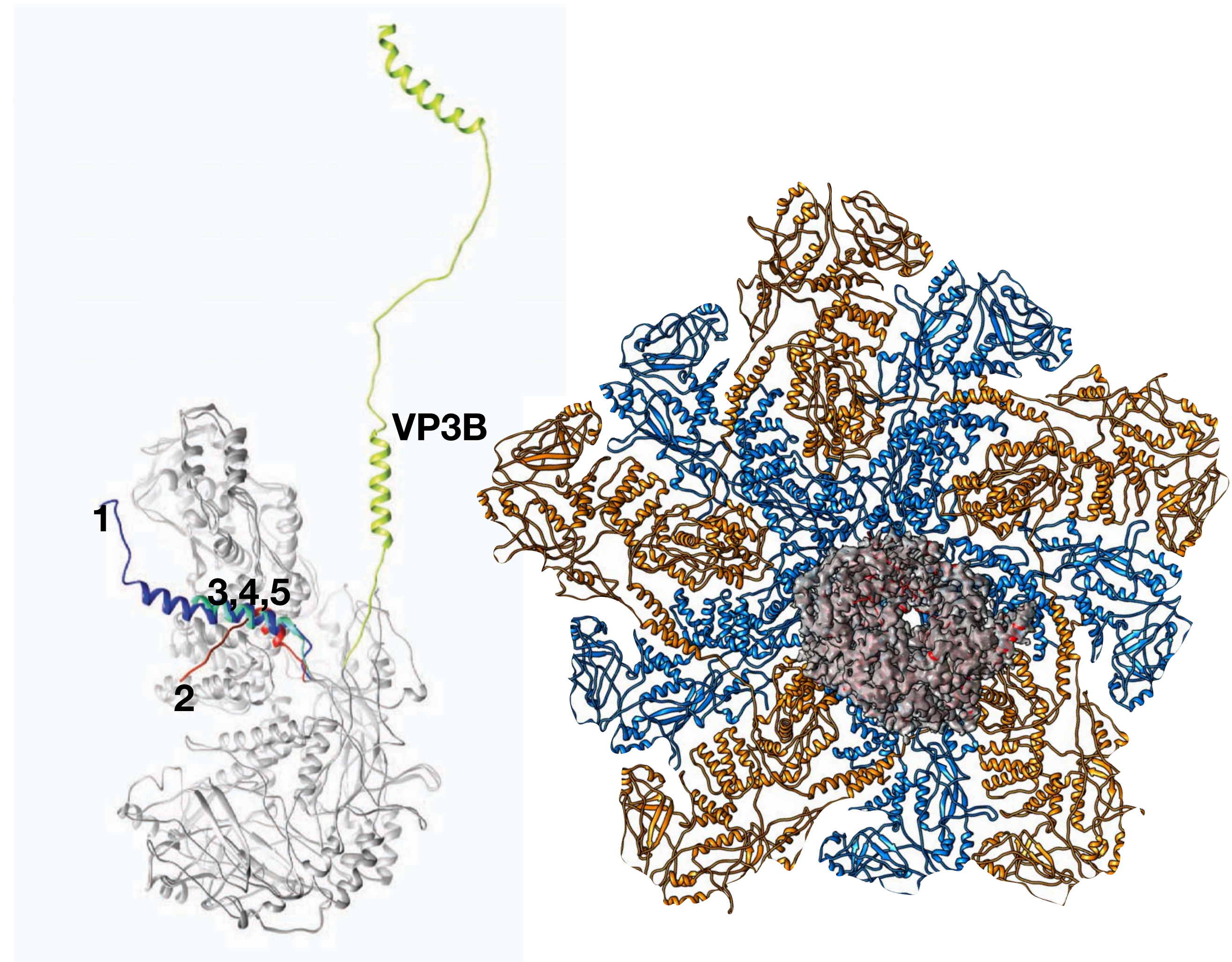
DSRNA GENOME

- ▶ 7 layers of segmented dsRNA genome visible
- ▶ dsRNA interacts and may help to position RdRP
- ▶ ssRNA seen entering transcribing MCRV
- ▶ Nucleotides seen inside RdRP near priming loop



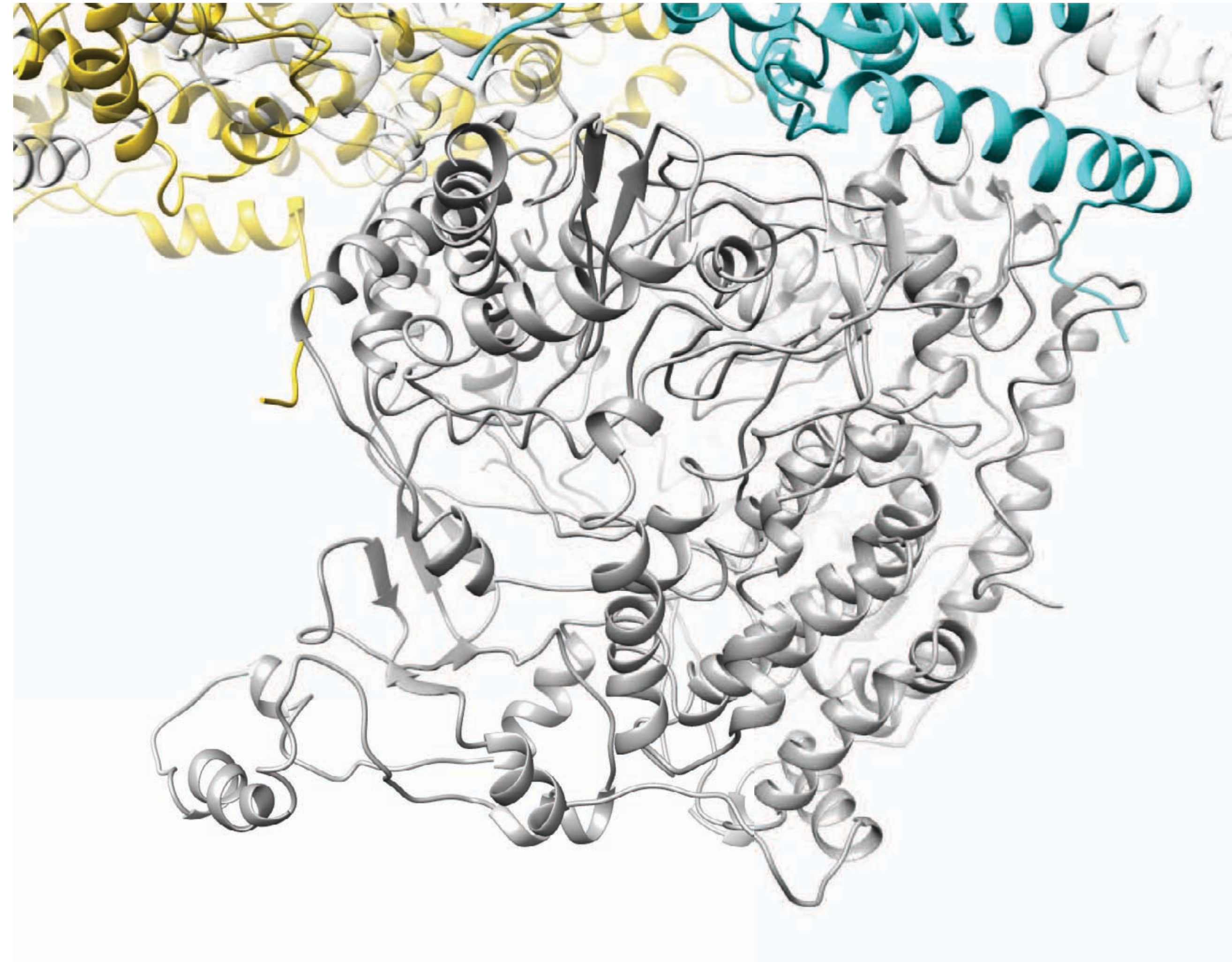
VP3 INTERACTIONS

- ▶ N-terminus of VP3 adopts 5 unique conformations and interacts with neighboring VP3s and RdRP



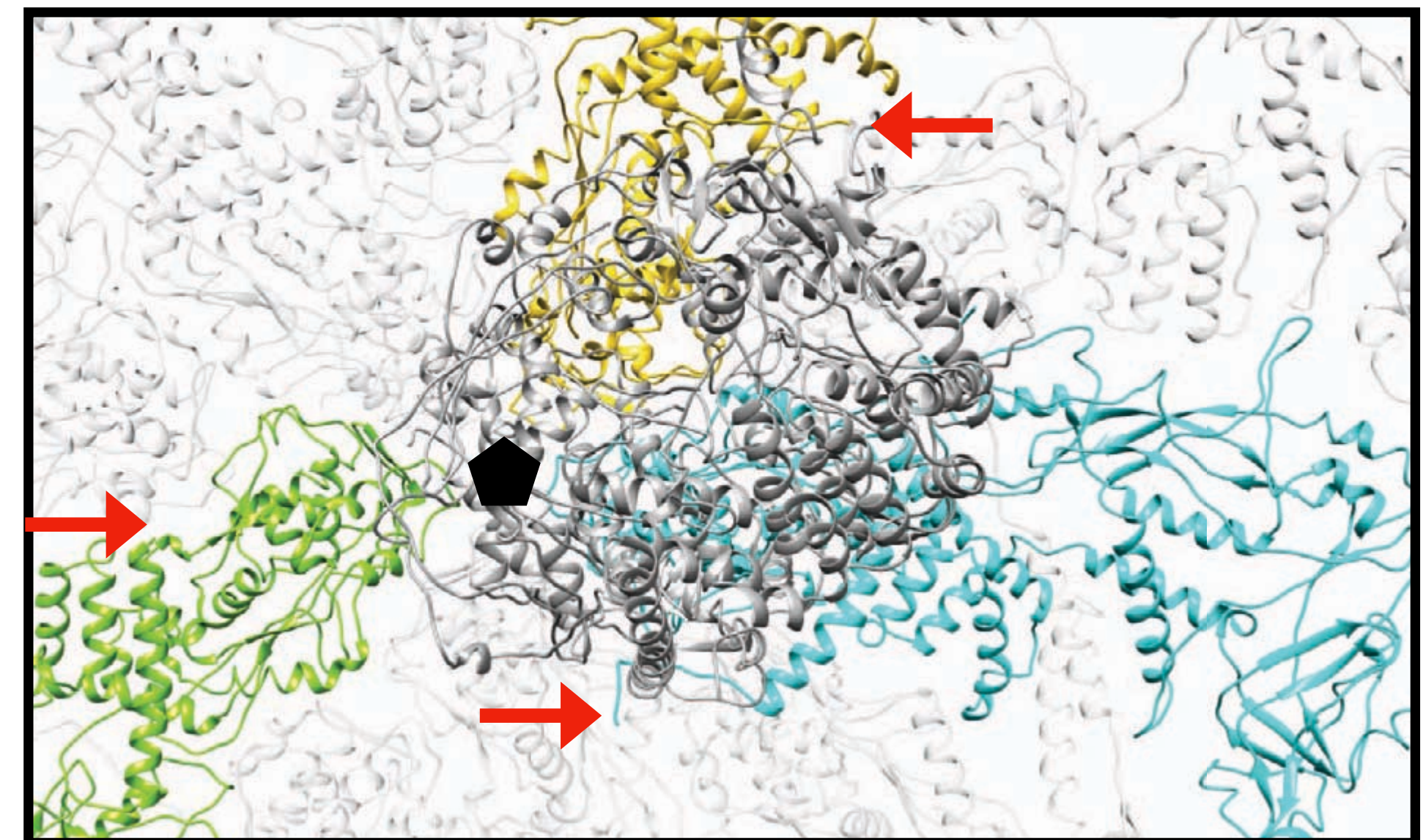
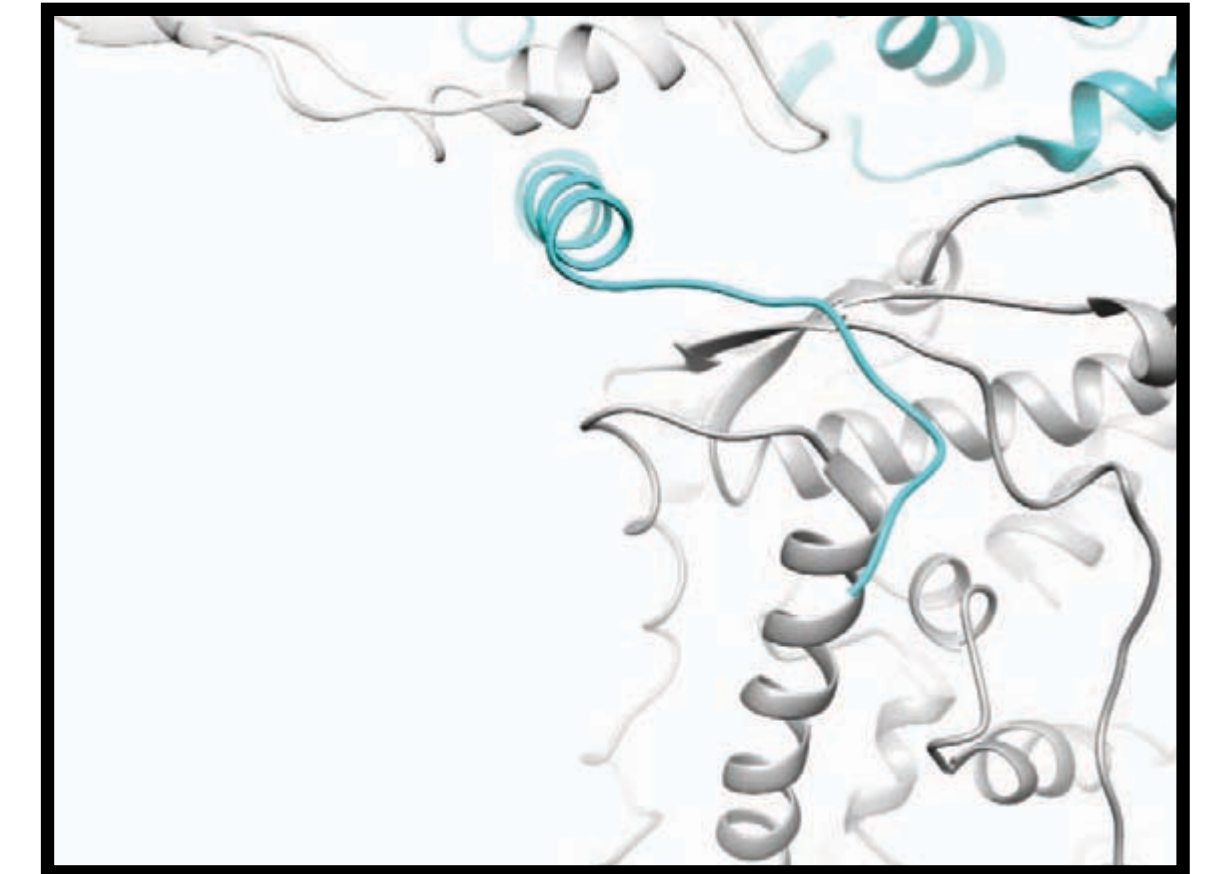
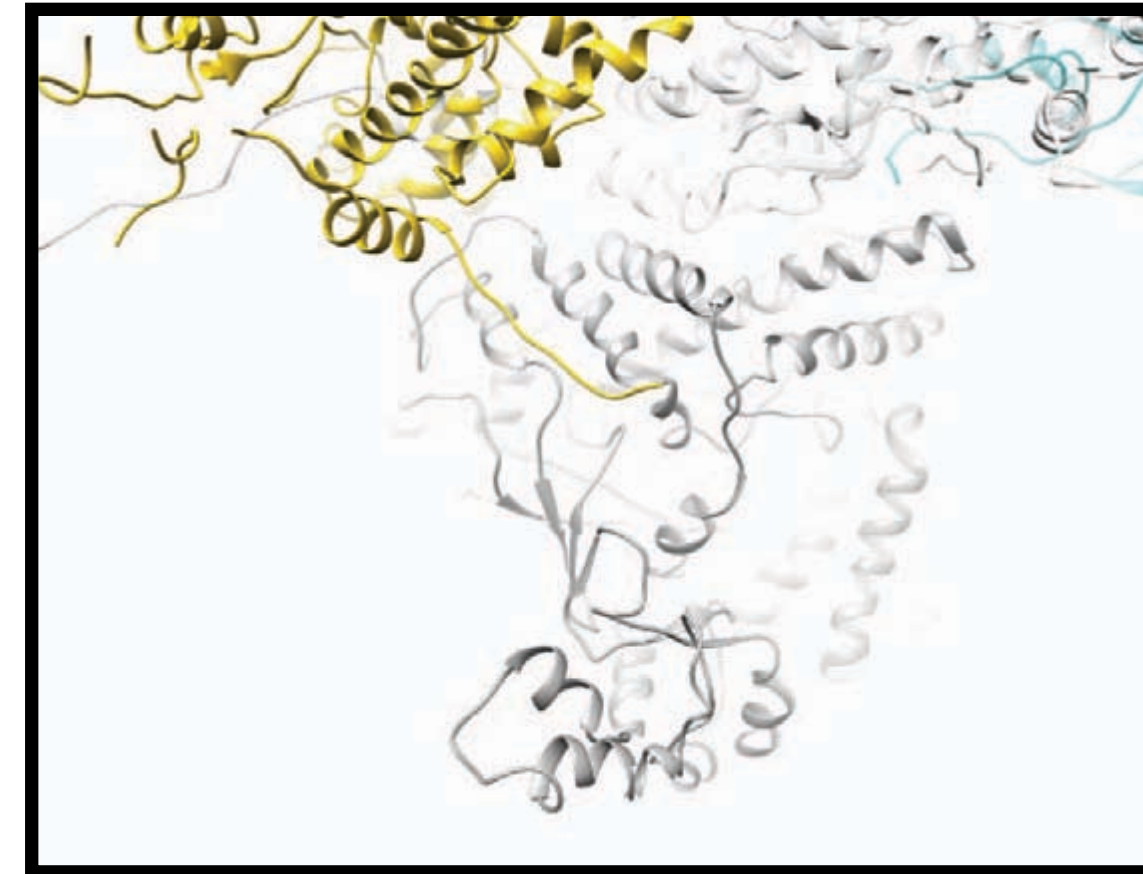
VP3 INTERACTIONS

- ▶ N-terminus of VP3 adopts 5 unique conformations and interacts with neighboring VP3s and RdRP
- ▶ VP3A residues 40-48 form an inter-molecular anti-parallel beta sheet with residues 1220-1226 and 1390-1395 of the RdRP



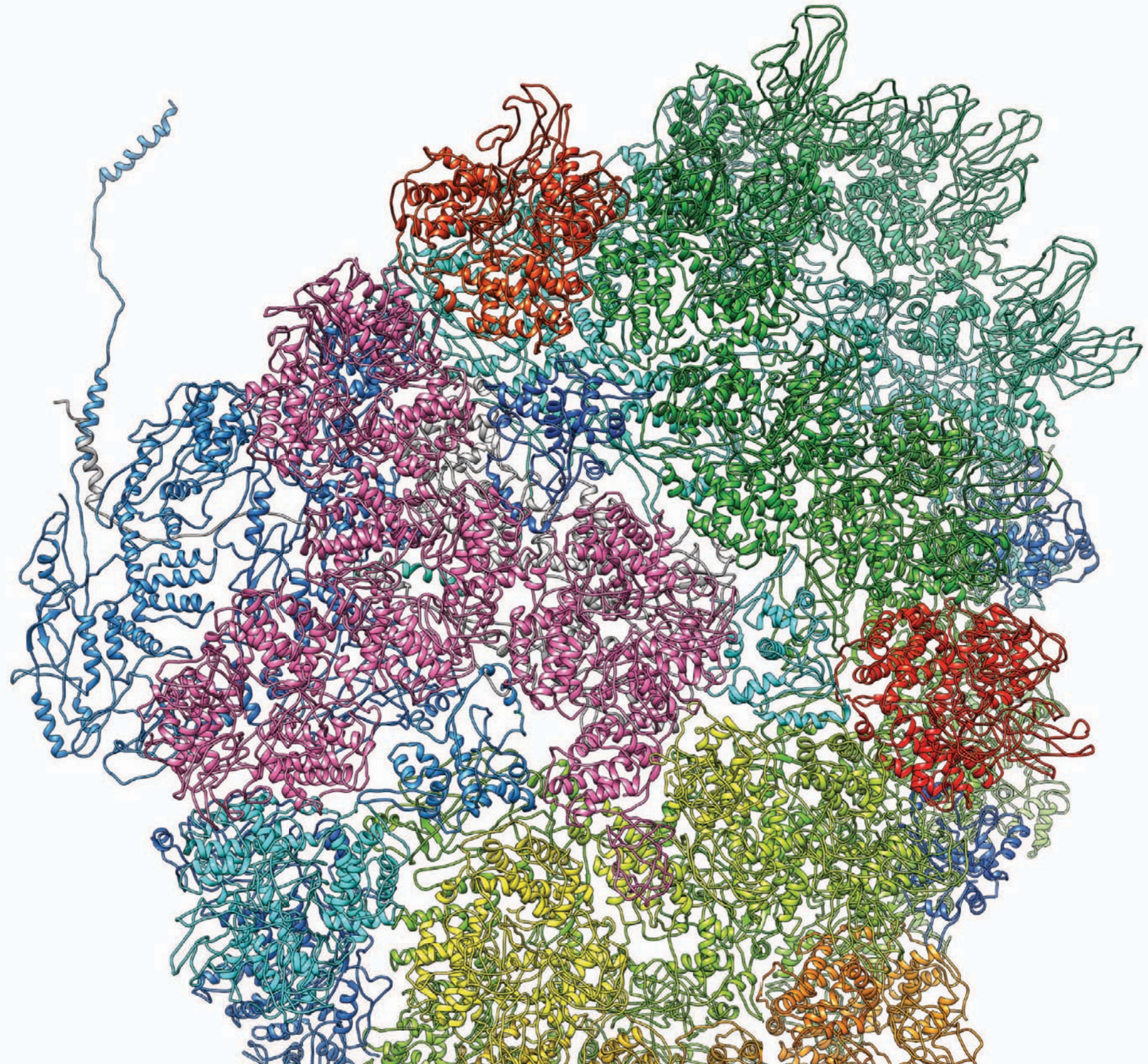
VP3 INTERACTIONS

- ▶ N-terminus of VP3 adopts 5 unique conformations and interacts with neighboring VP3s and RdRP
- ▶ VP3A residues 40-48 form an inter-molecular anti-parallel beta sheet with residues 1220-1226 and 1390-1395 of the RdRP



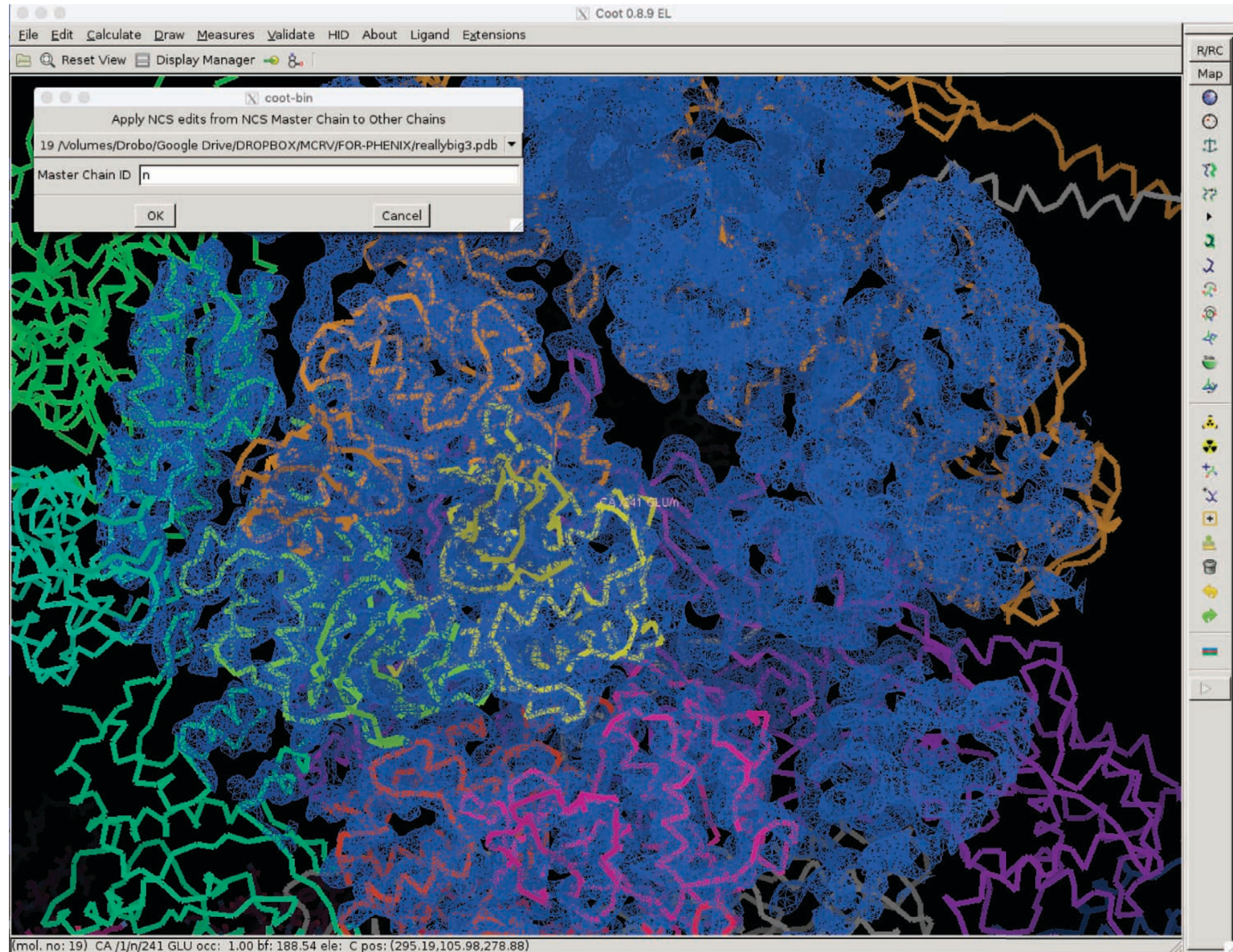
REFINING THE MODELS

- ▶ Build and refine individual subunits
- ▶ Create asymmetric unit structure
- ▶ Create multi-asymmetric unit structure
 - ▶ Refine with NCS symmetry in Phenix
 - ▶ Create "Ghosts" in Coot
 - ▶ Refine and apply NCS symmetry in Coot

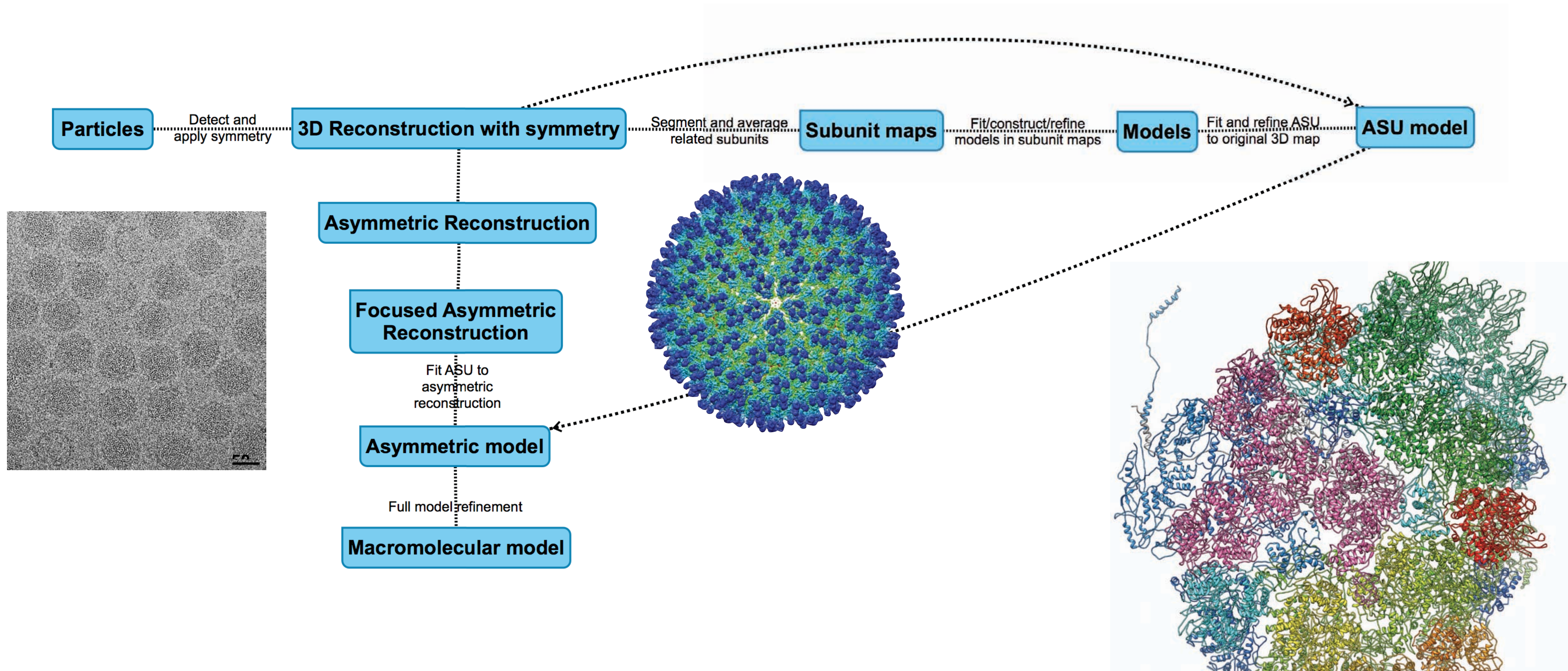


REFINING THE MODELS

- ▶ Build and refine individual subunits
- ▶ Create asymmetric unit structure
- ▶ Create multi-asymmetric unit structure
 - ▶ Refine with NCS symmetry in Phenix
 - ▶ Create "Ghosts" in Coot
 - ▶ Refine and apply NCS symmetry in Coot

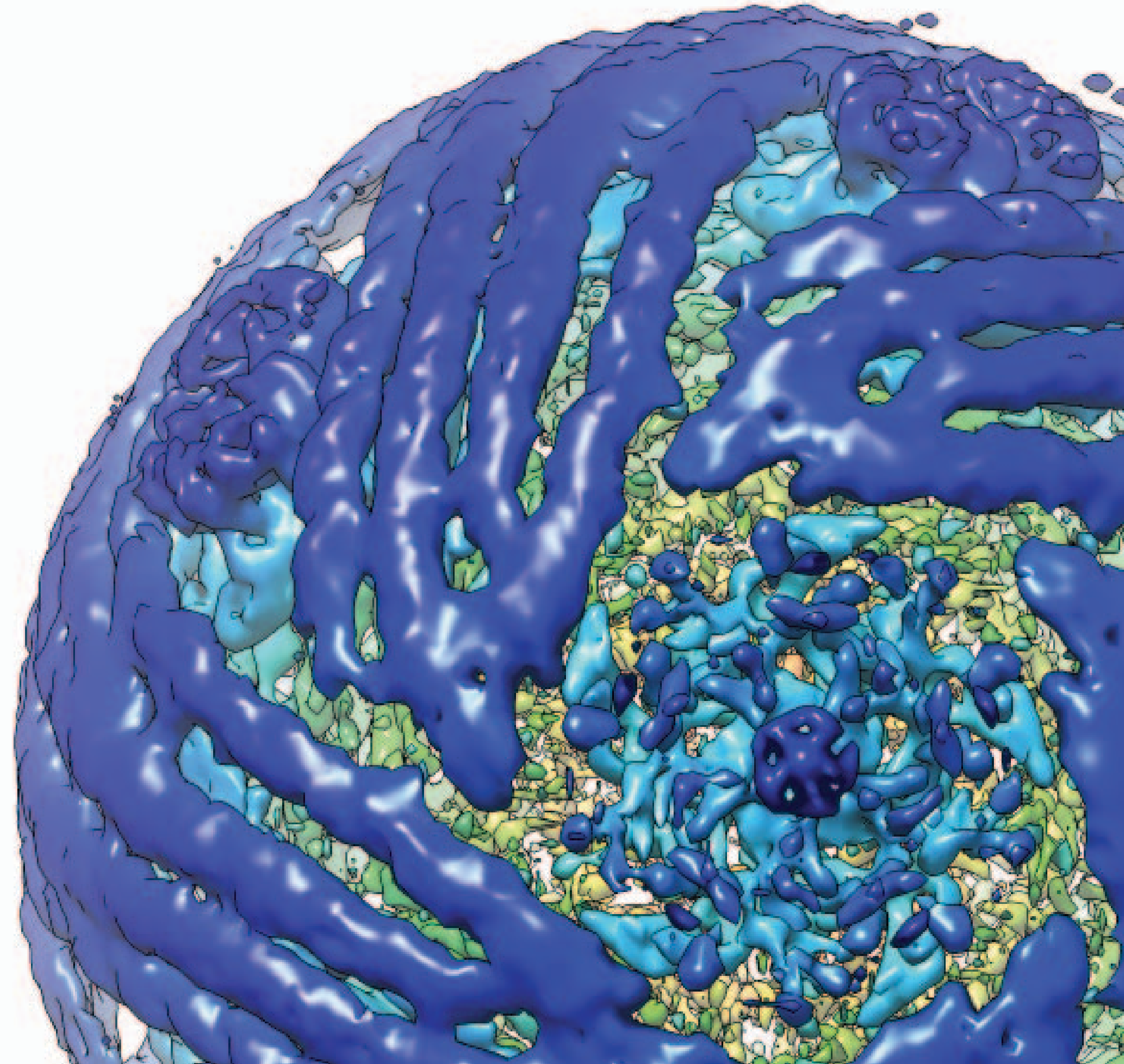


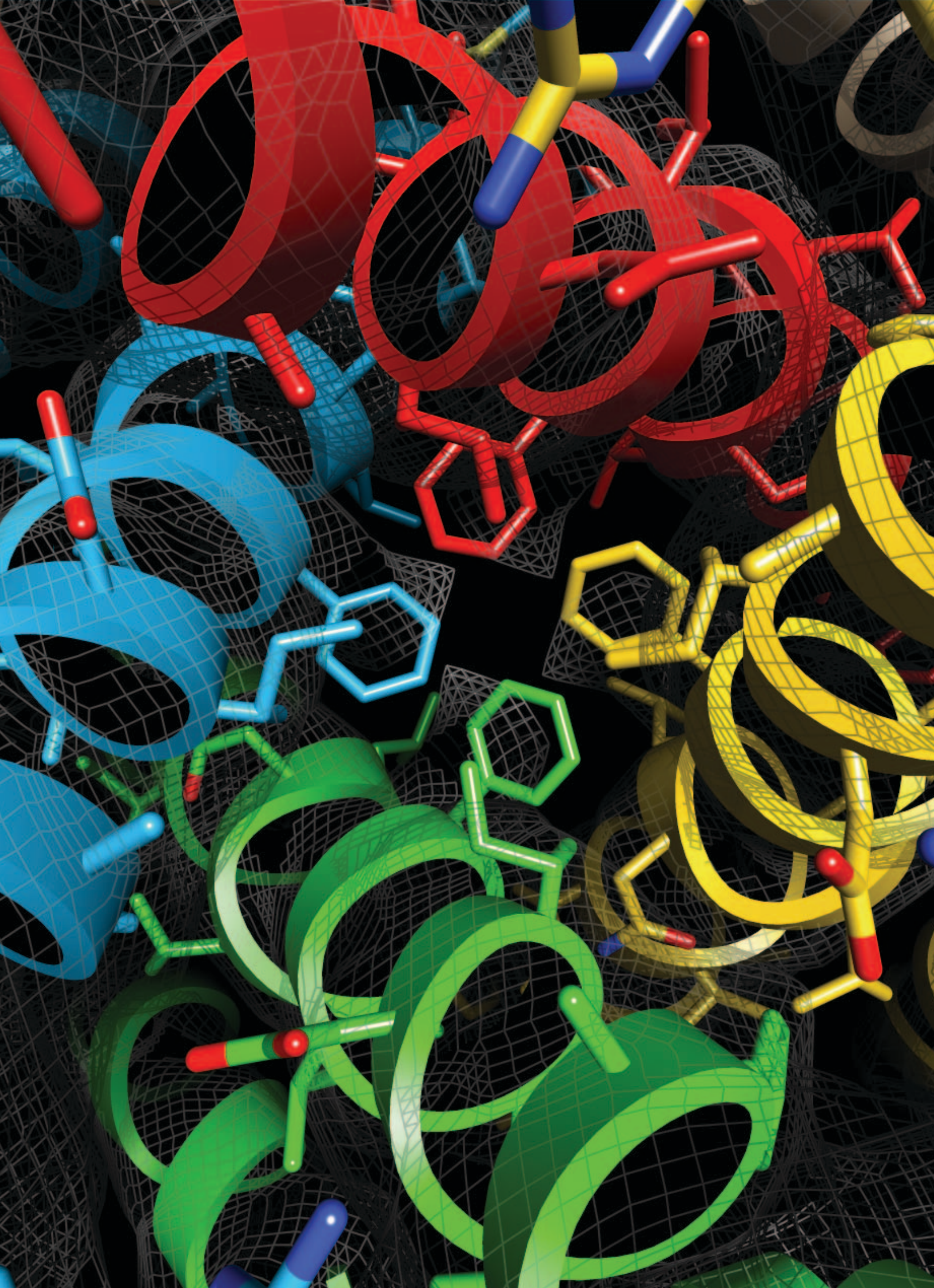
SYMMETRY IN CRYO-EM



CONCLUSIONS

- ▶ Symmetry can be detected from particle level to model level in cryoEM
- ▶ Workflow has to compensate for symmetry at various stages
- ▶ Symmetry can help reduce the number of particles to get to desired resolution





MCRV

- ▶ Wen Jiang (Purdue)
- ▶ Jason Kaebler (Rutgers)
- ▶ Qinfen Zhang (Zhongshen)

IP3R1

- ▶ Irina Serysheva (UTHealth)
- ▶ Mariah Baker (UTHealth)
- ▶ Guizhen Fan (UTHealth)

Modeling

- ▶ Muyuan Chen (BCM)
- ▶ Corey Hryc (BCM)

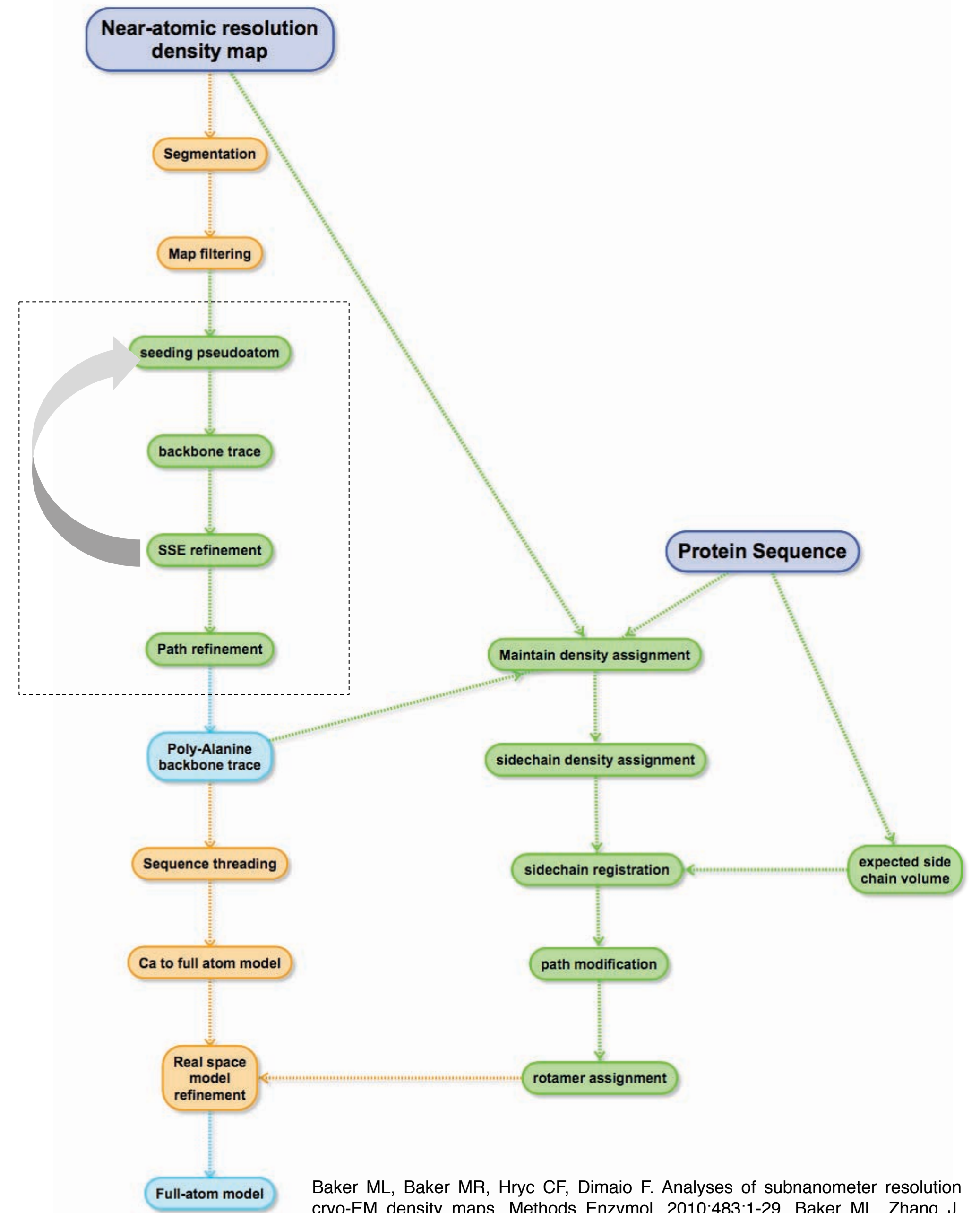
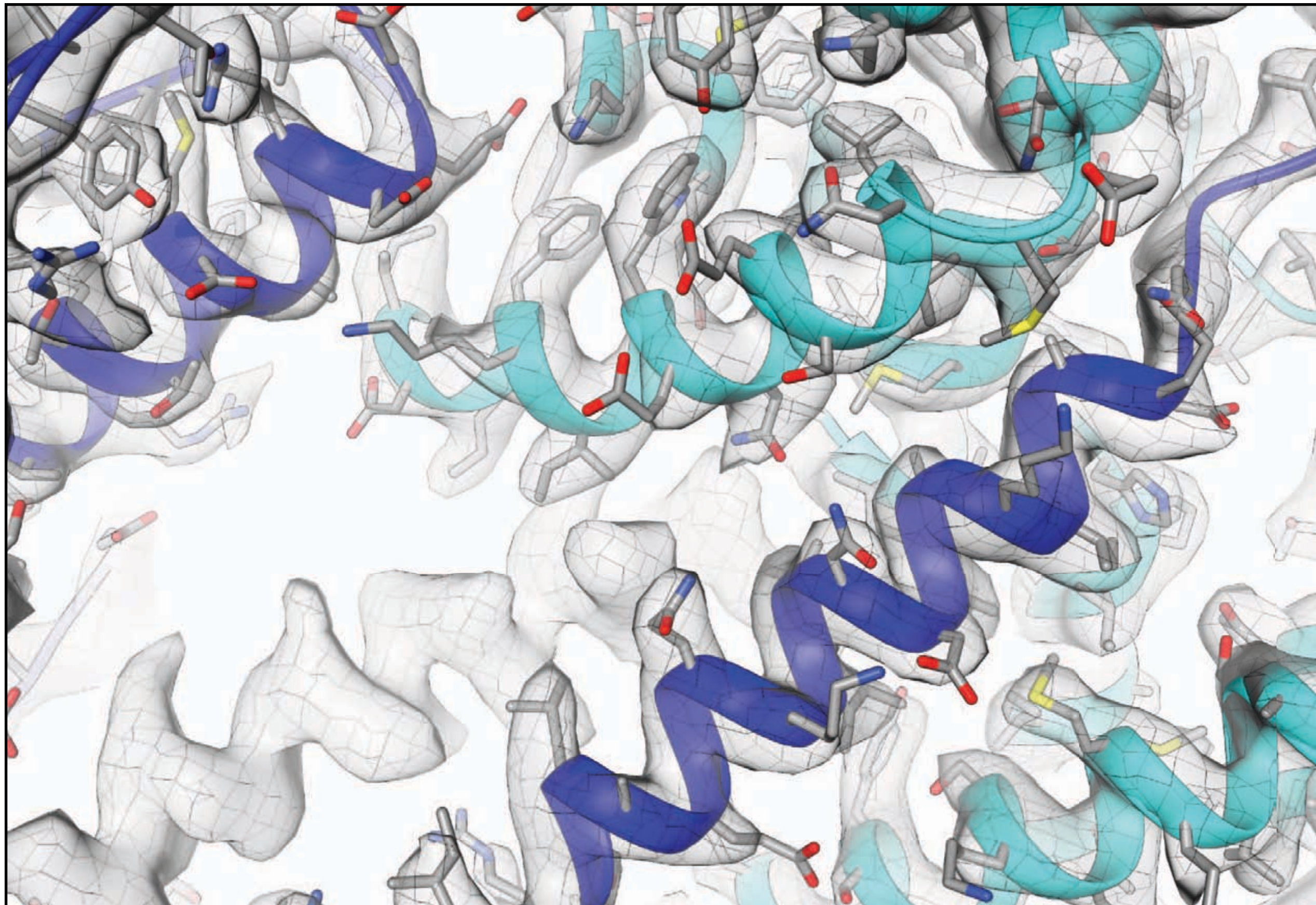
- ▶ Zhao Wang (BCM)
- ▶ Steve Ludtke (BCM)

ACKNOWLEDGEMENTS



DISCUSSION

Modeling in Cryo-EM

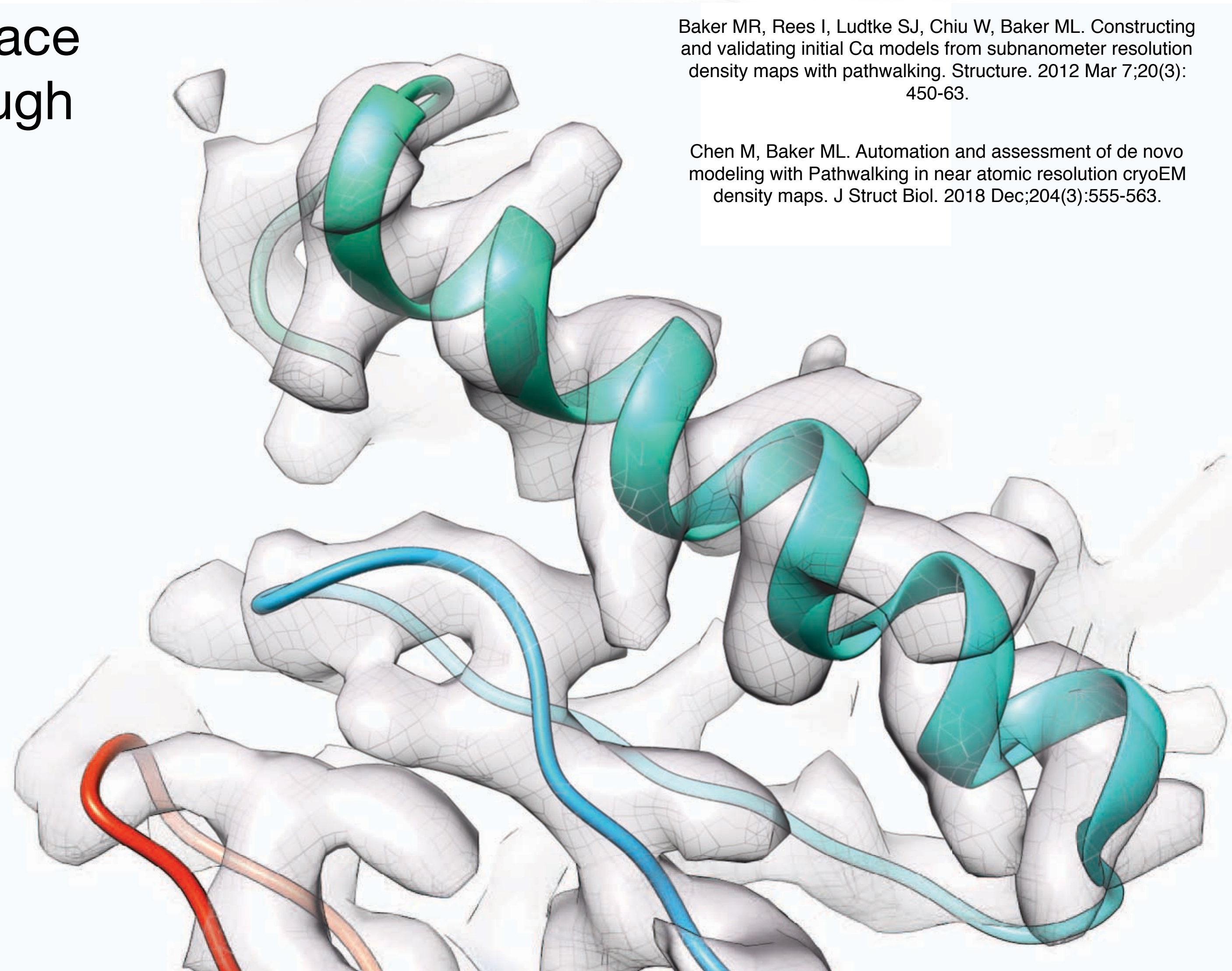


Baker ML, Baker MR, Hryc CF, Dimaio F. Analyses of subnanometer resolution cryo-EM density maps. *Methods Enzymol.* 2010;483:1-29. Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat Protoc.* 2010 Sep;5(10):1697-708.

Pathwalking

Finds a path or sets of paths that trace the complete path of a protein through a density map at near-atomic resolutions

- ➡ No structural template required
- ➡ No SSEs required
- ➡ No explicit sequence information required
- ➡ Optimized against biophysical constraints
- ➡ Automated



The Traveling Salesman Problem

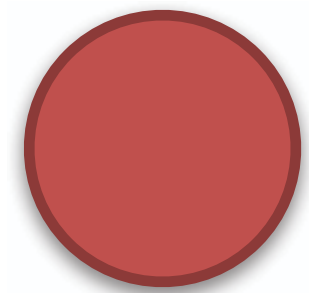


- TSP solvers calculate optimal route between nodes by minimizing distance travelled, where each city can only be visited once

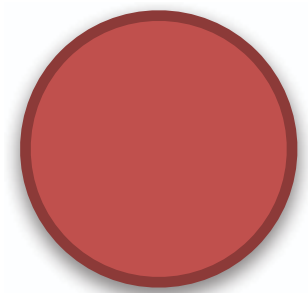
Protein Folds and a TSP Solver



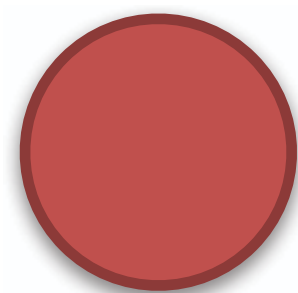
Seattle



San Francisco



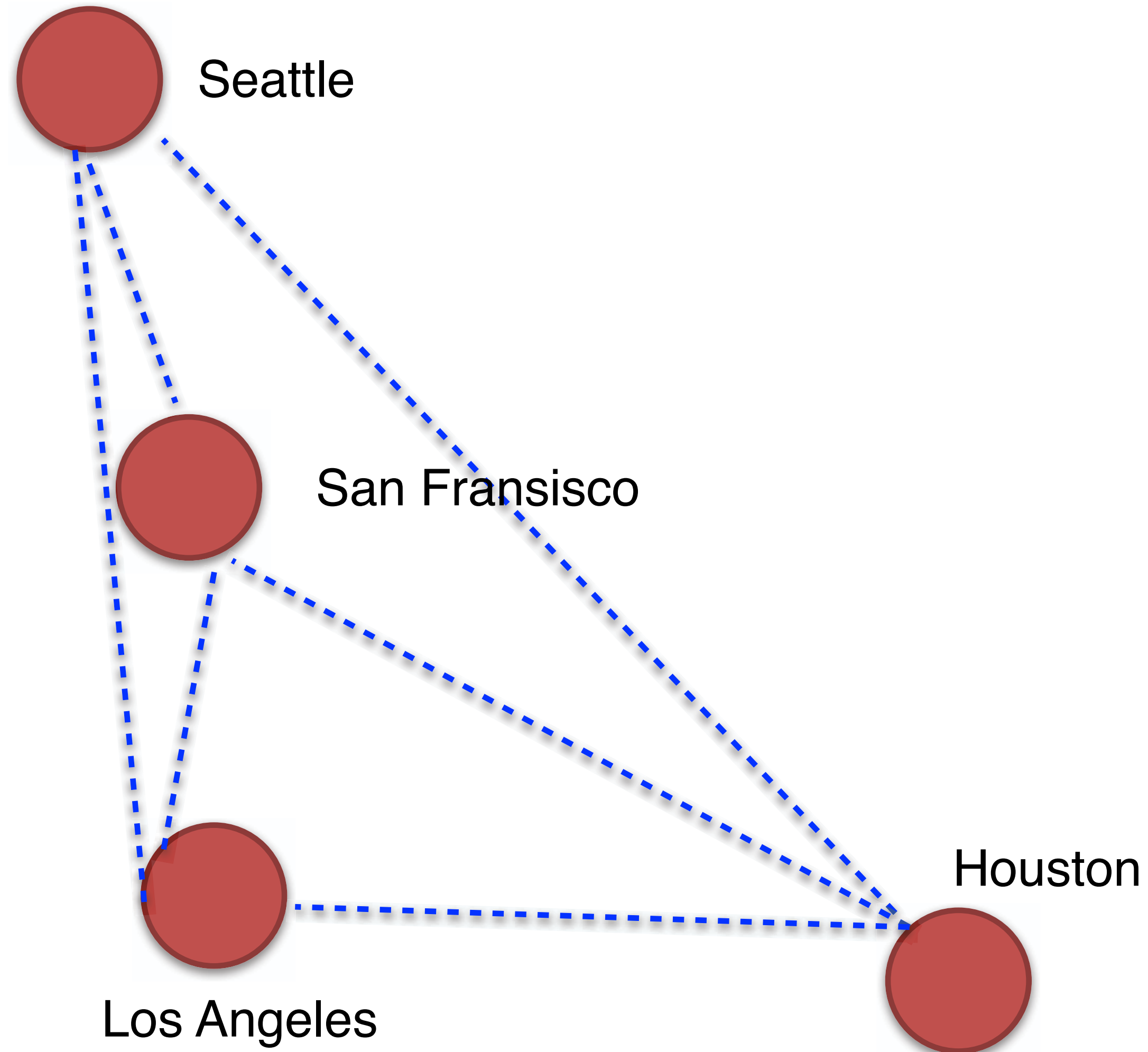
Los Angeles



Houston

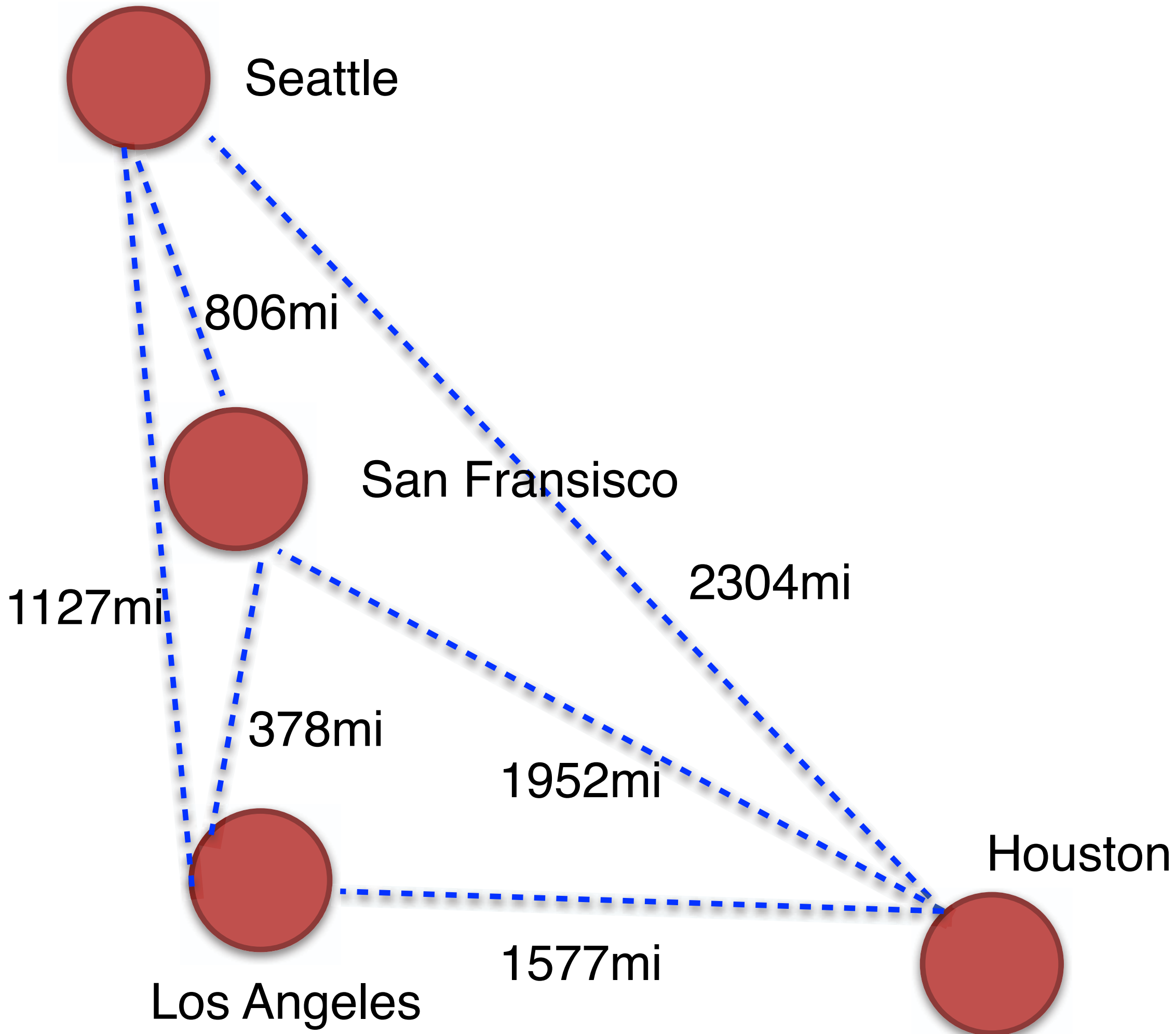
Conventional Application of a TSP Solver

Protein Folds and a TSP Solver



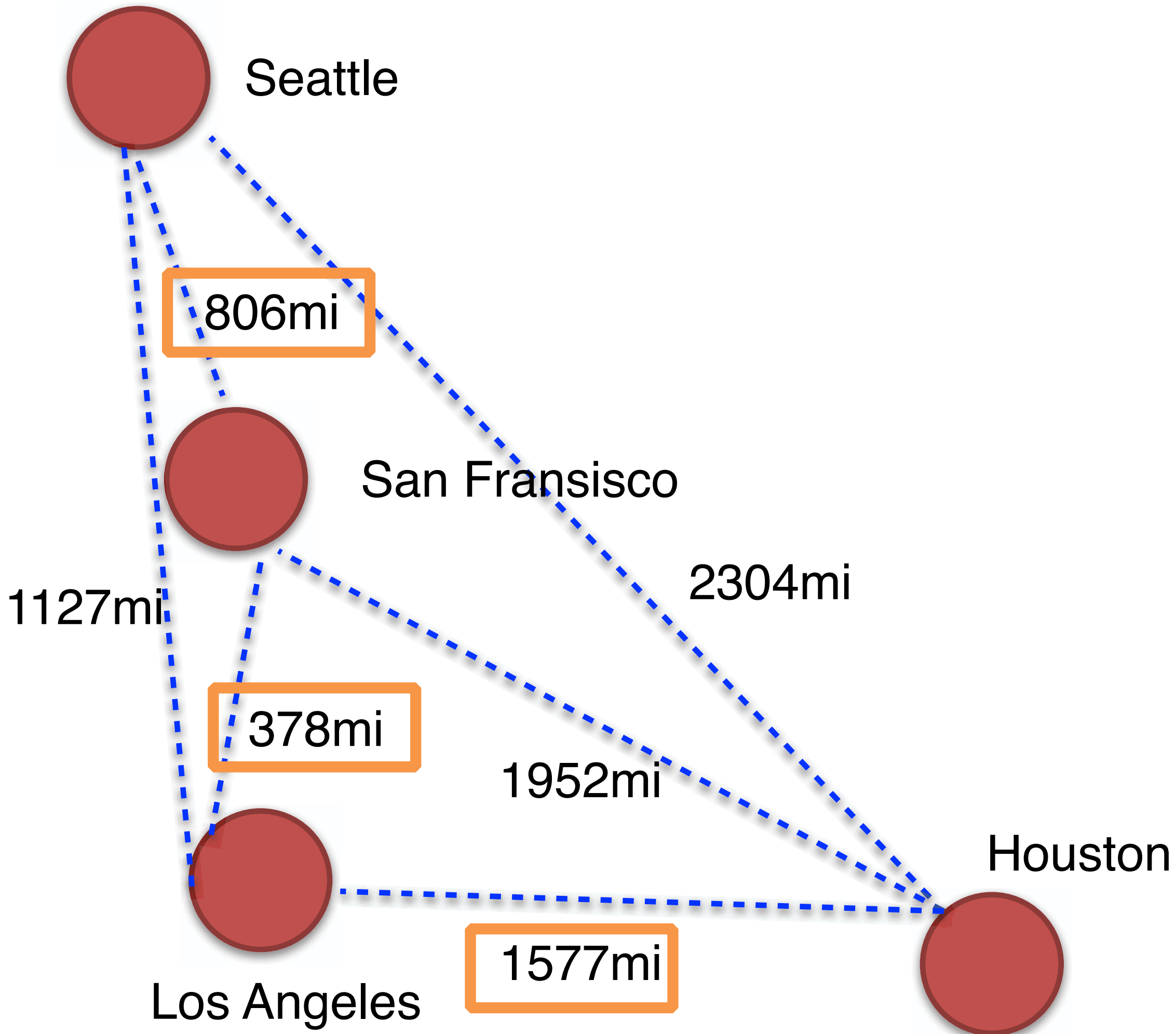
Conventional Application of a TSP Solver

Protein Folds and a TSP Solver



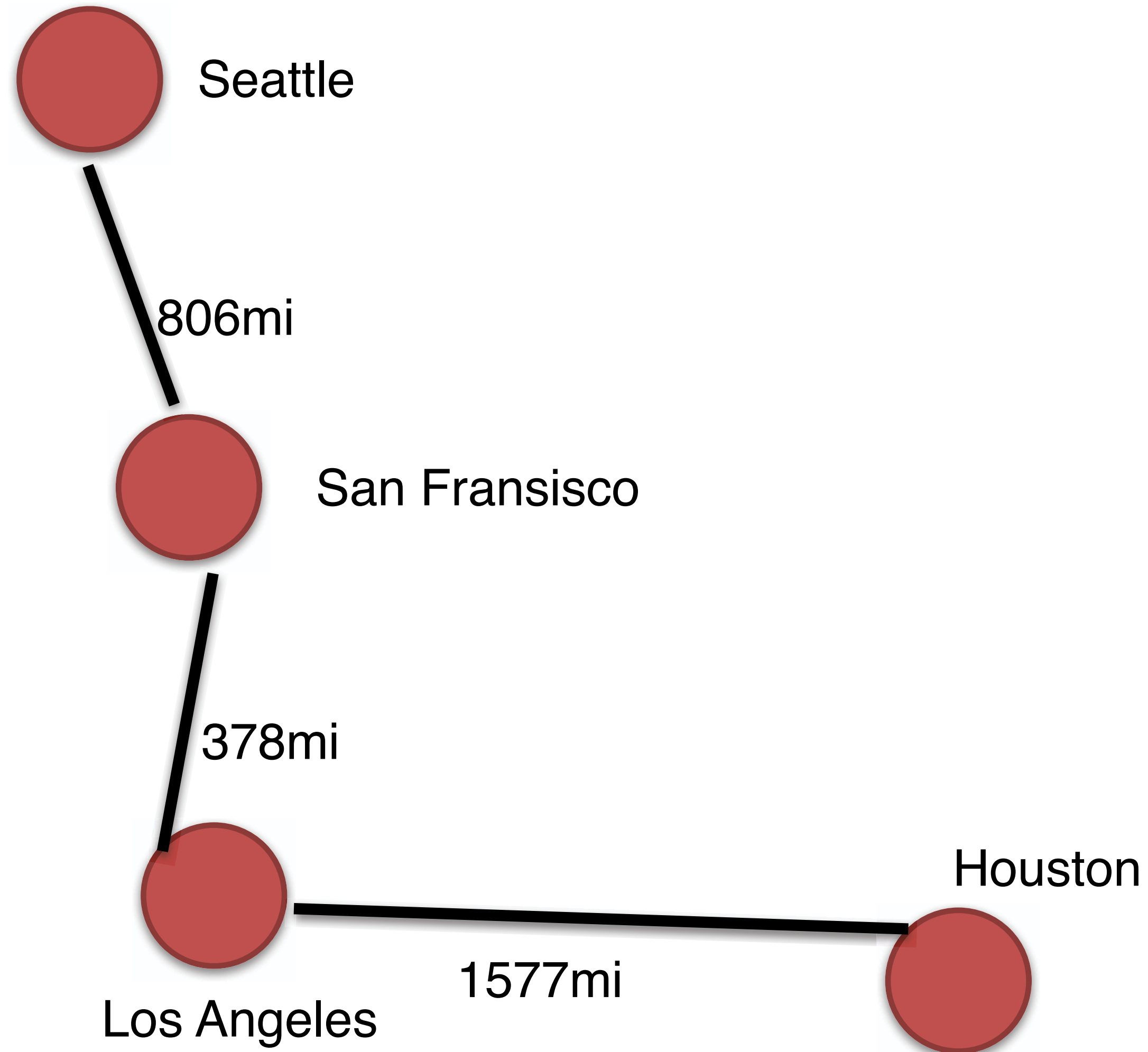
Conventional Application of a TSP Solver

Protein Folds and a TSP Solver



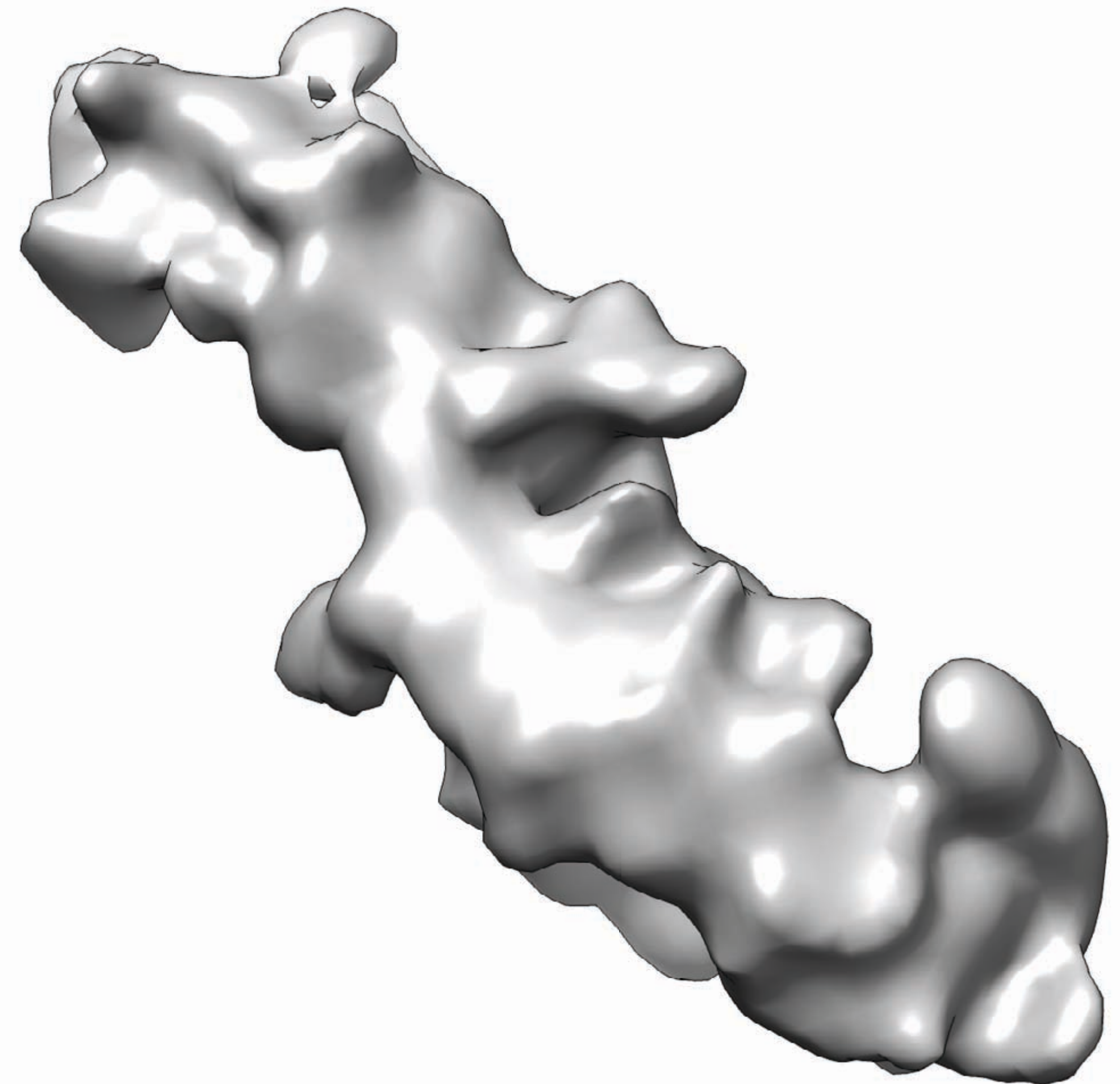
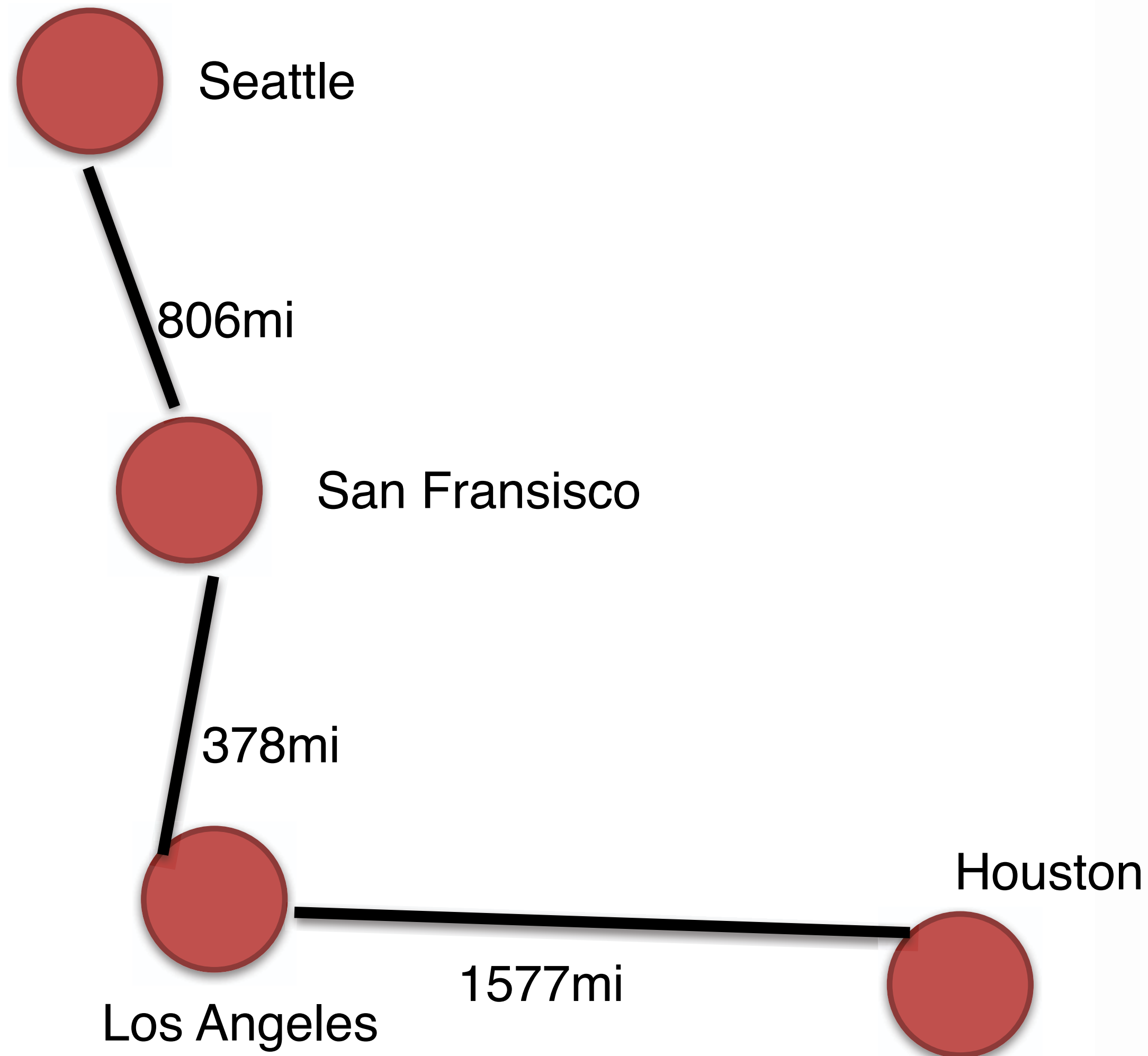
Conventional Application of a TSP Solver

Protein Folds and a TSP Solver



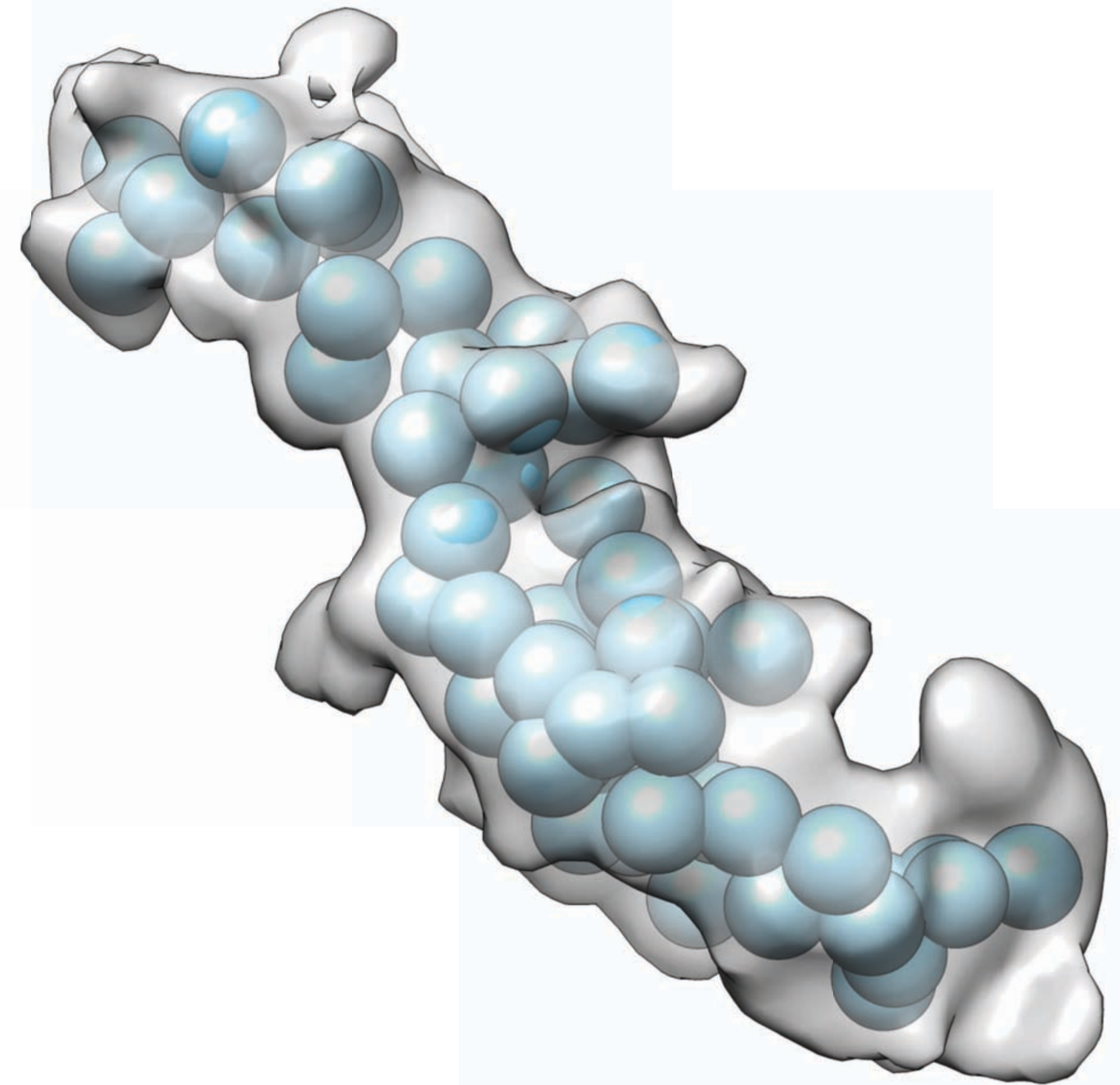
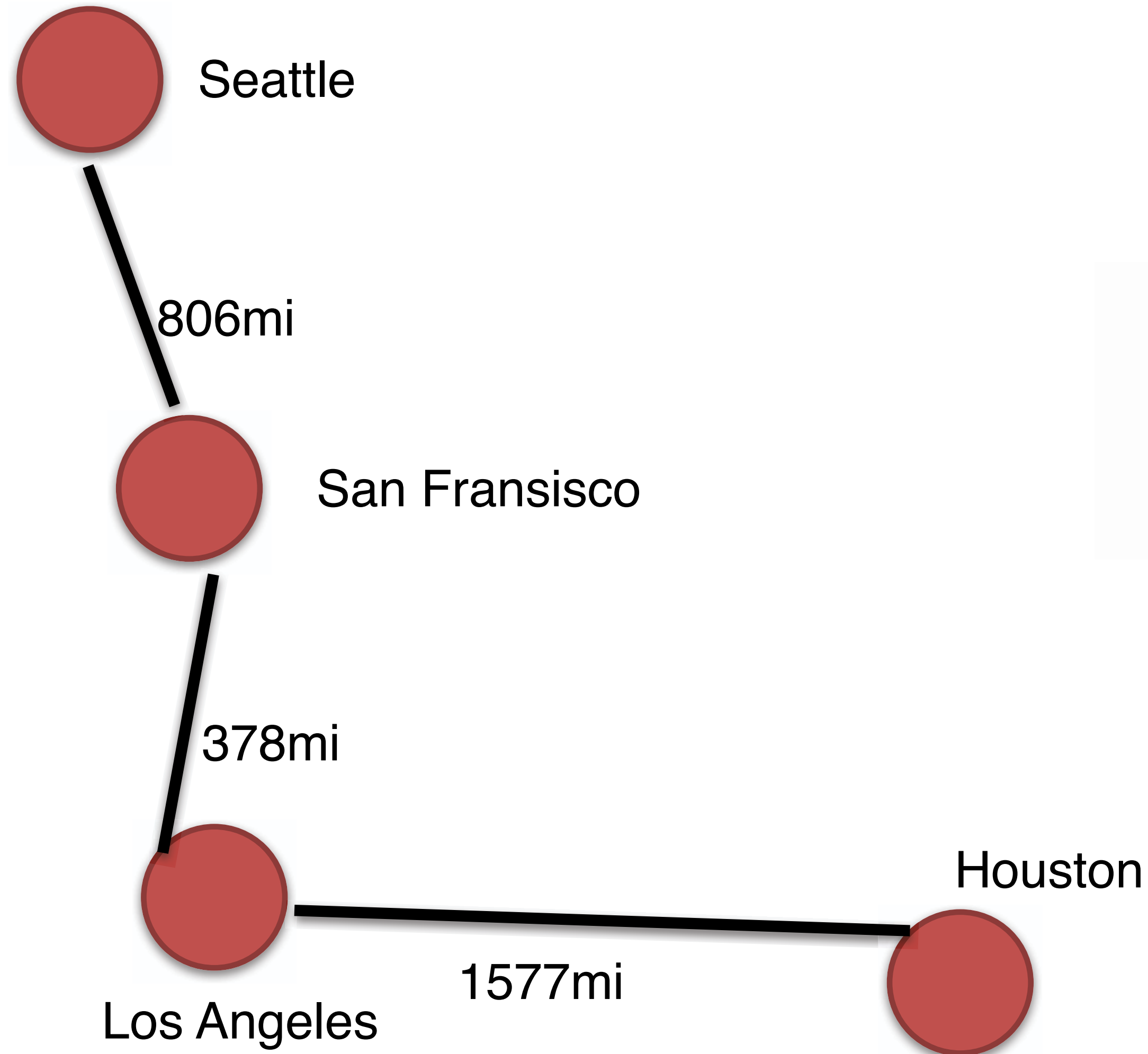
Conventional Application of a TSP Solver

Protein Folds and a TSP Solver



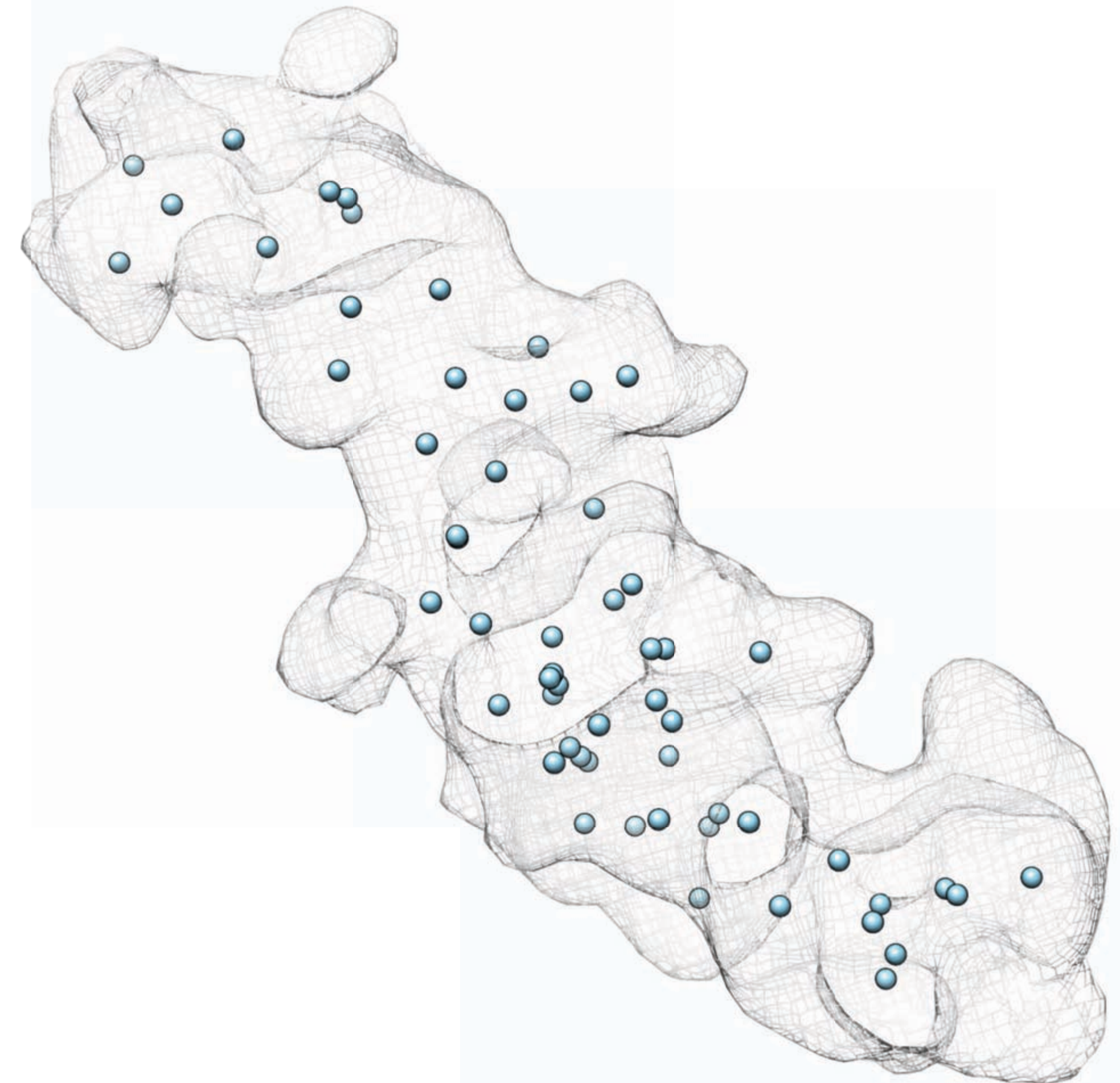
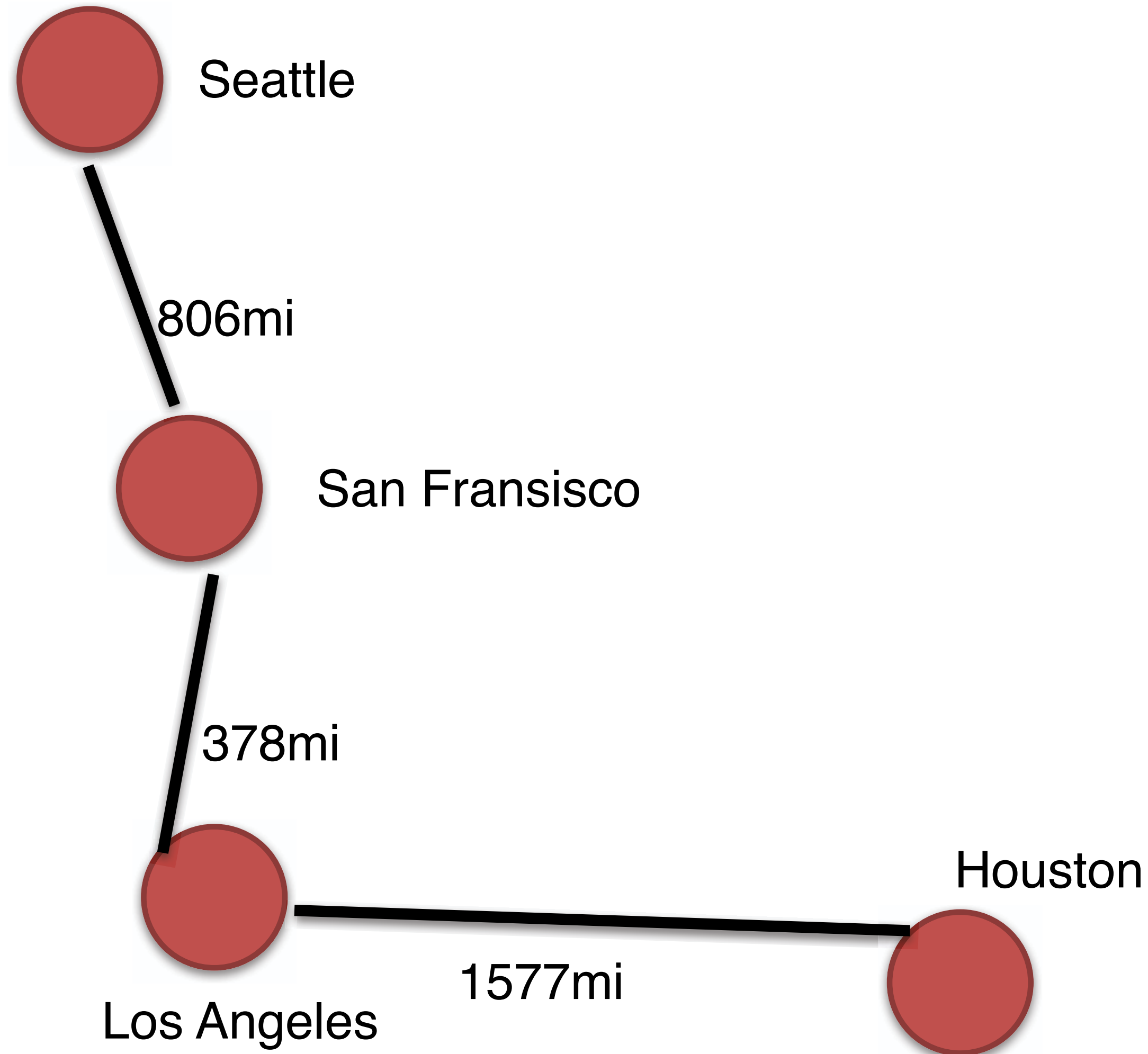
Application of a TSP Solver to CryoEM

Protein Folds and a TSP Solver



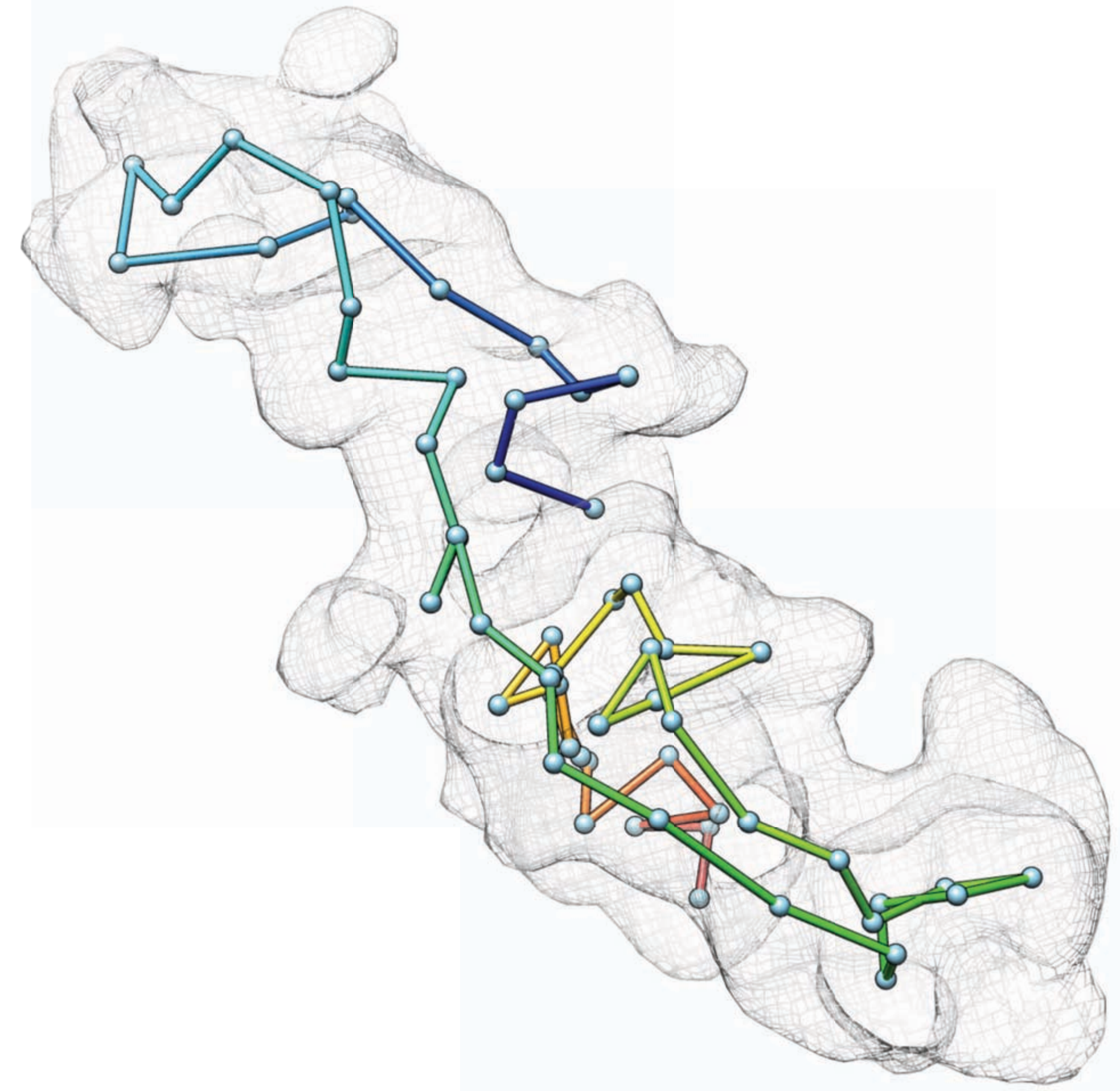
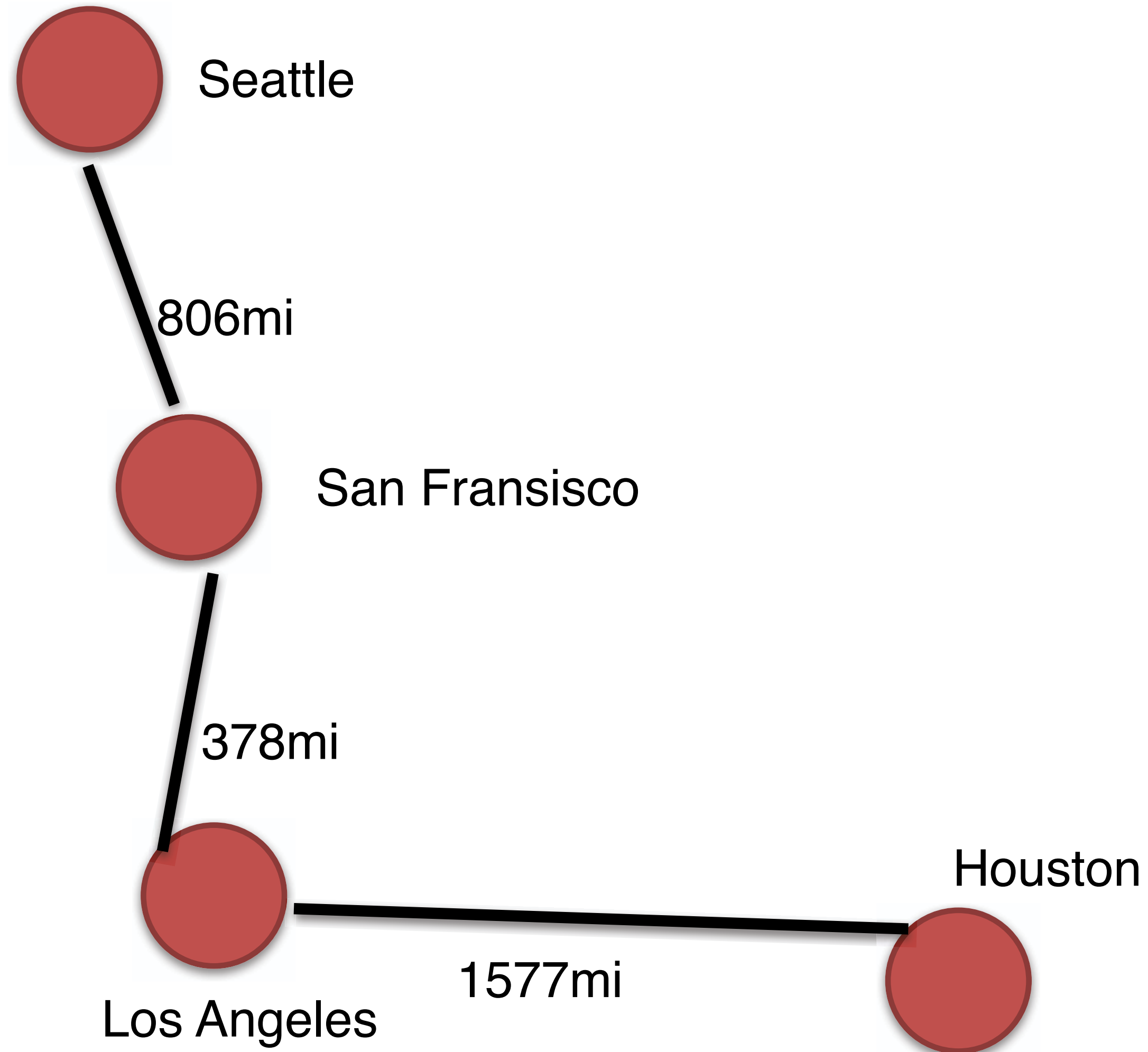
Application of a TSP Solver to CryoEM

Protein Folds and a TSP Solver



Application of a TSP Solver to CryoEM

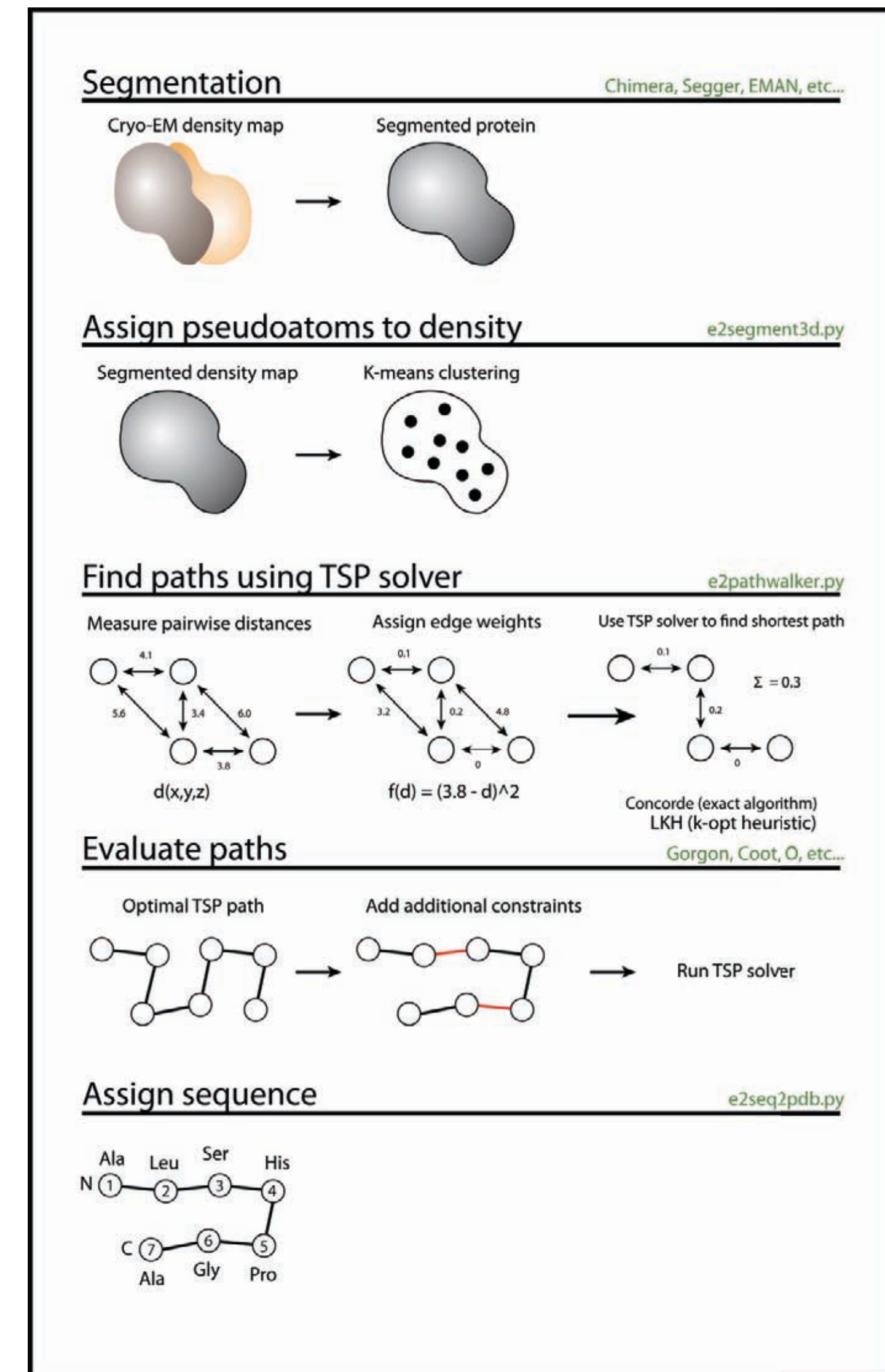
Protein Folds and a TSP Solver



Application of a TSP Solver to CryoEM

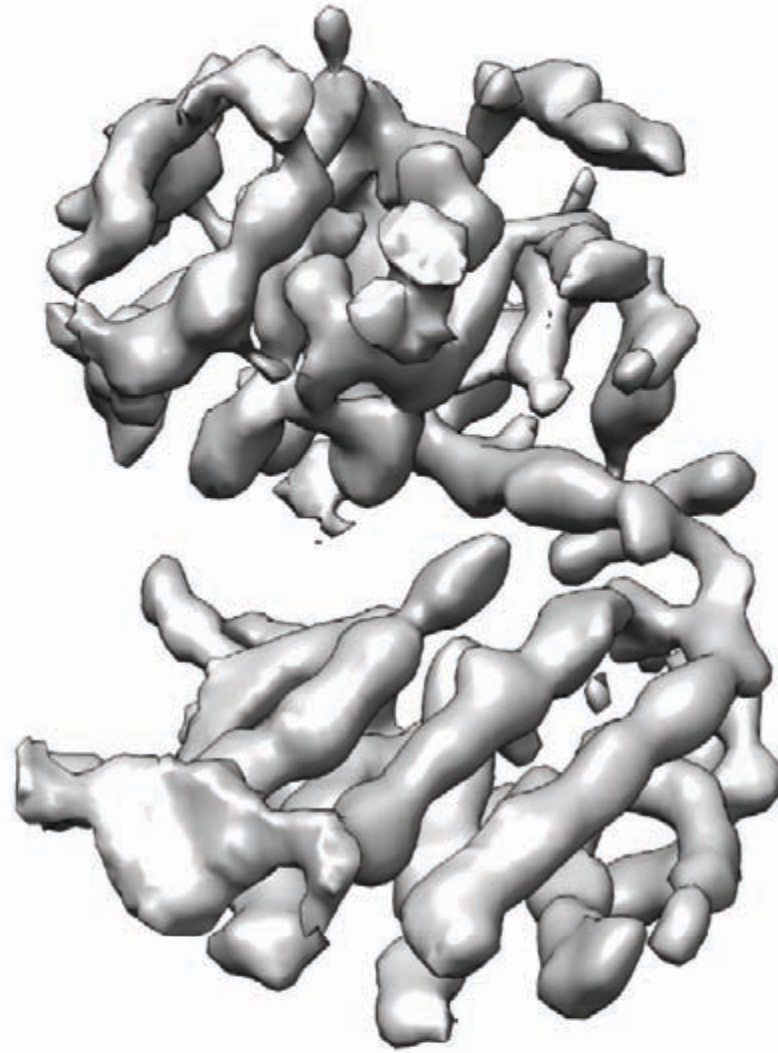
de novo Modeling as a 3D TSP Problem

- ➔ “Nodes” are pseudoatoms (C-alpha atoms representative of aa position in a backbone trace)
- ➔ Rather than calculating a distance between all atoms, TSP distances are expressed as an error function
 - ➔ Error distances are expressed as a deviation from the ideal C-alpha/C-alpha distance of 3.8Å
- ➔ TSP solver searches for minimal error path, which results in paths with near ideal protein geometry

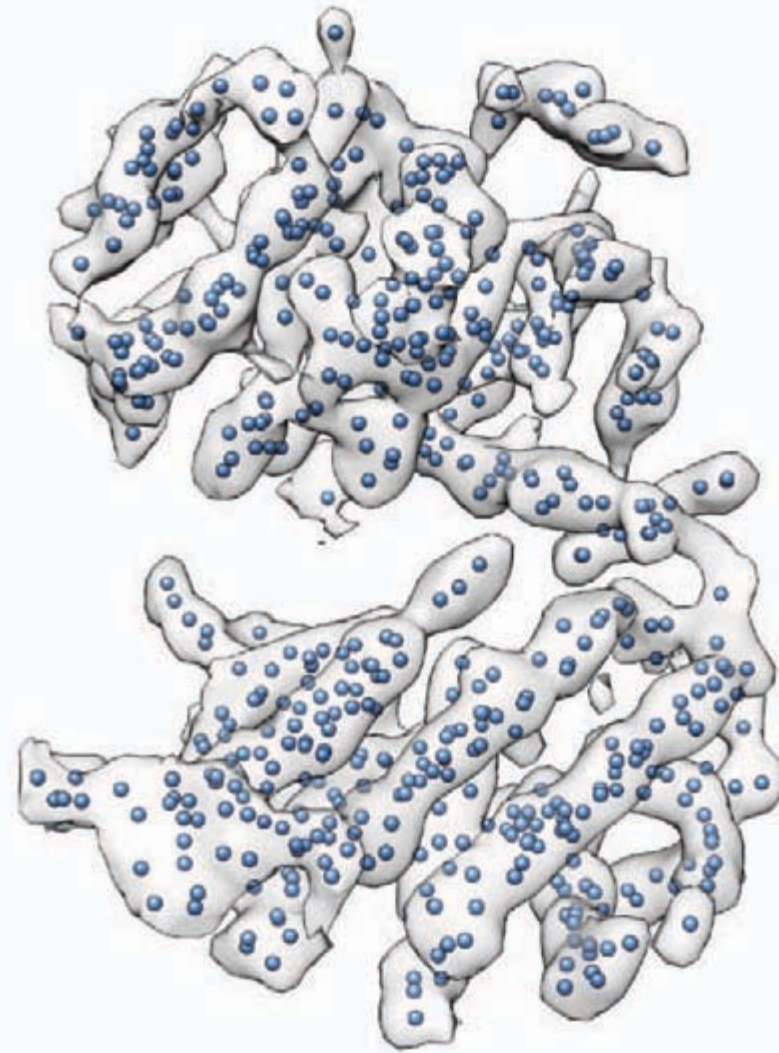


Pathwalking Protocol

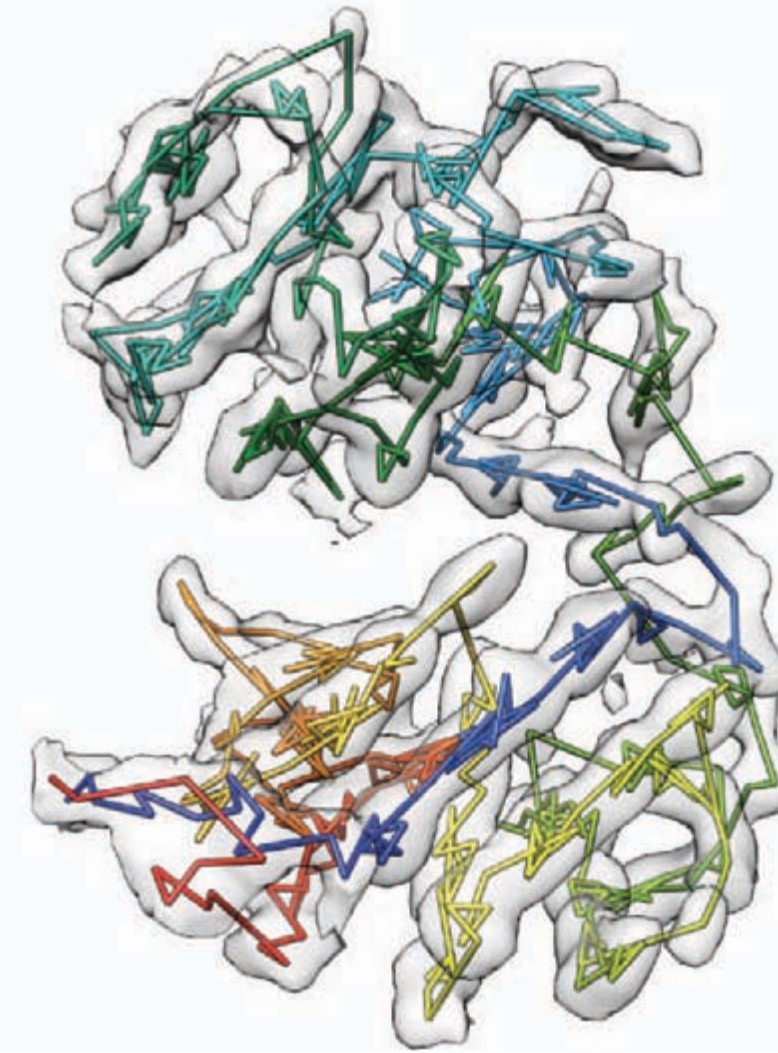
A



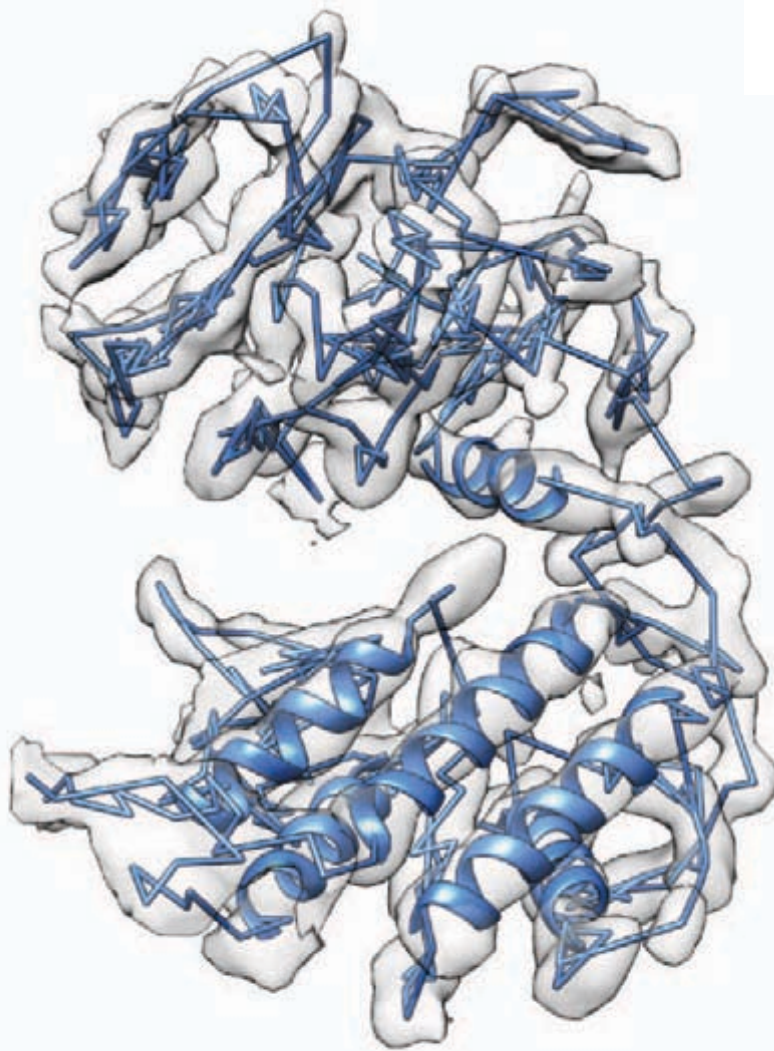
B



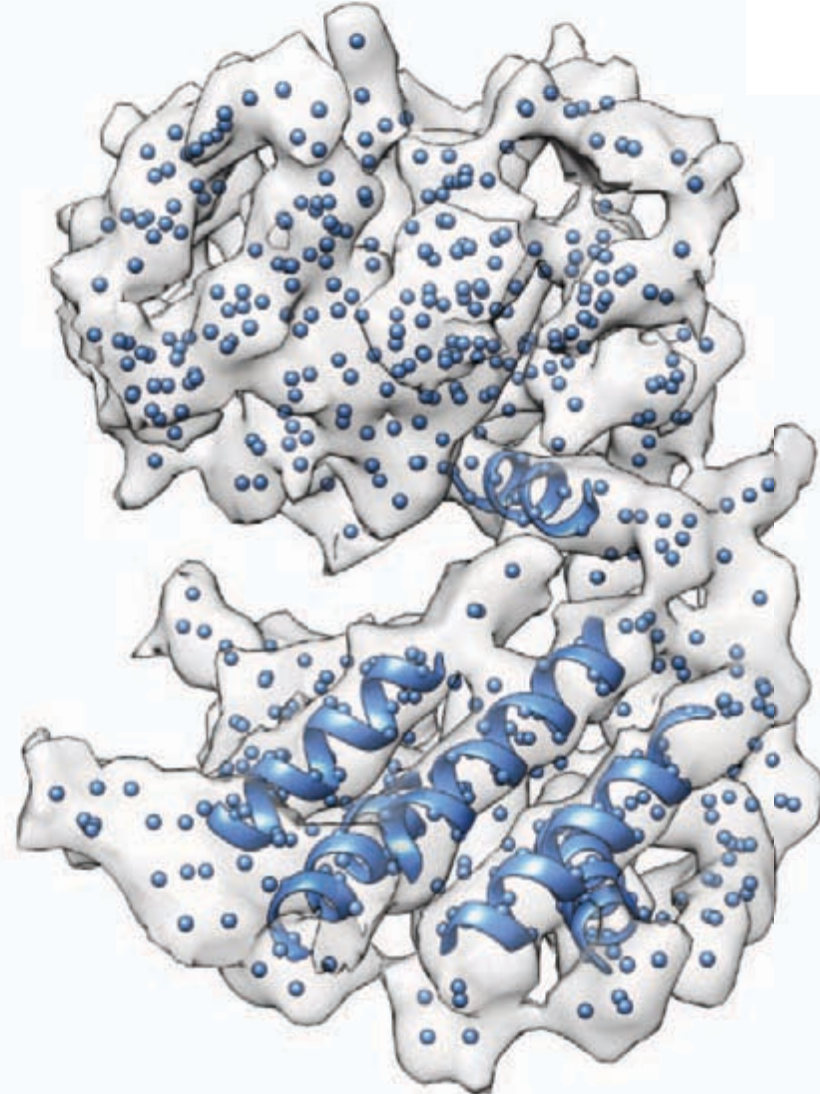
C



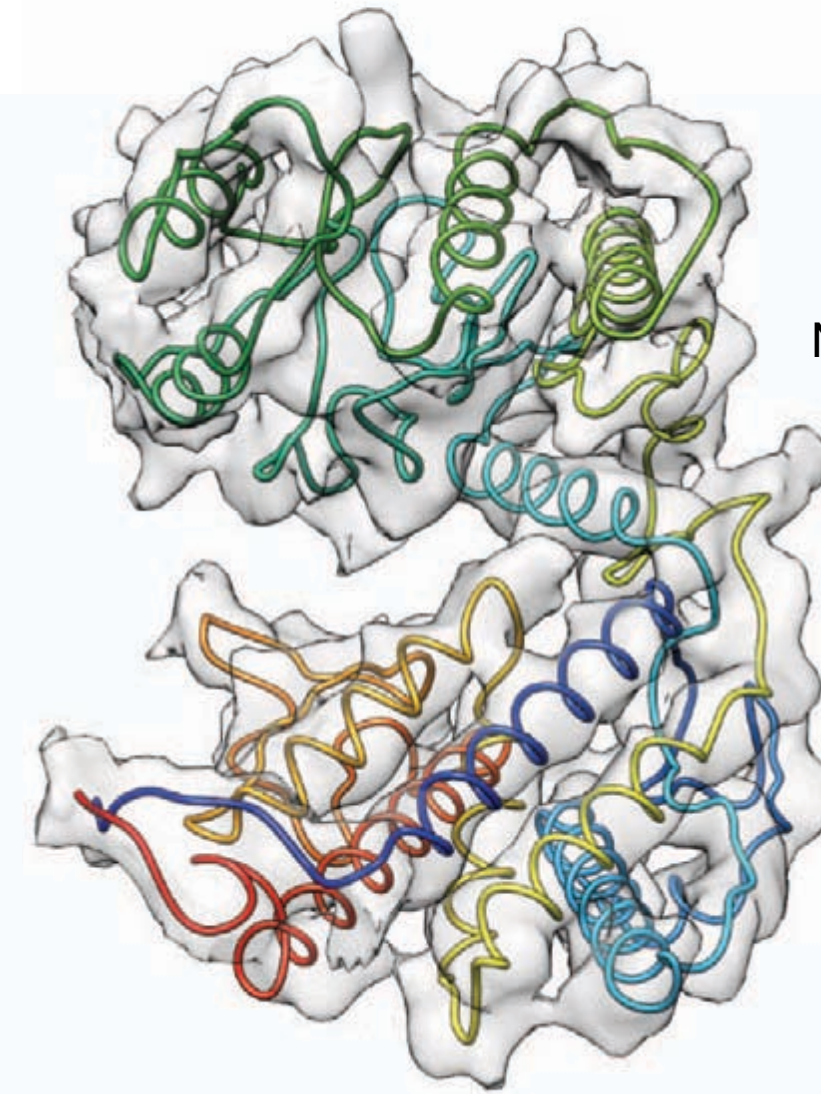
D



E



F



Preprocess density map

(A)

Seed pseudoatoms

(B)

Backbone tracing

(C)

Find and correct

SSEs (D)

Fix SSEs in map

(E)

No new SSE found

Final modification

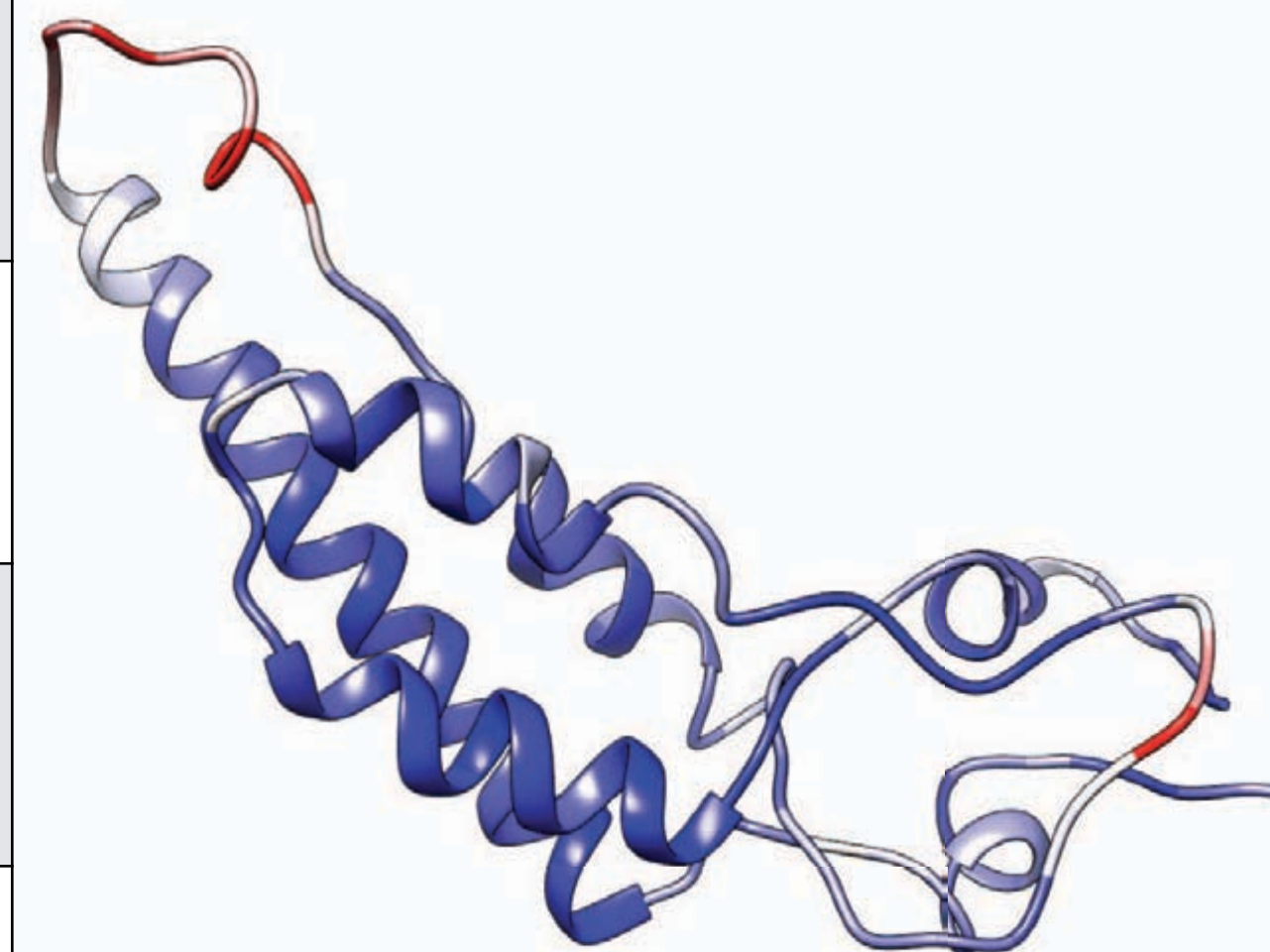
(F)

Finish

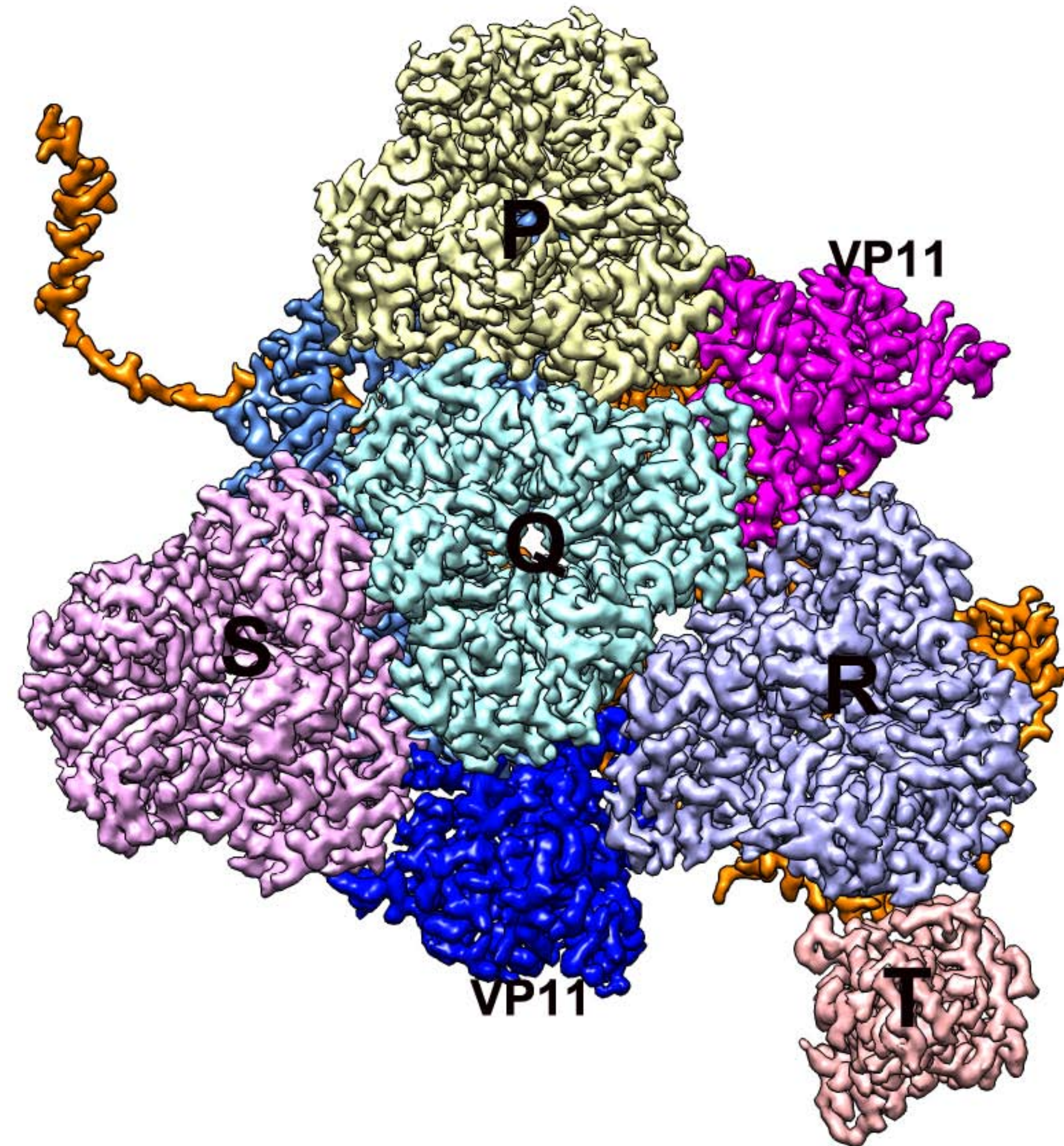
New SSE found

2015/2016 CryoEM Model Challenge

Target	Map Res (Å)	RMSD (Å)	CC	Clash score	Rama Outliers (%)	CA score	Sequence match (%)	length
Tobacco Mosaic Virus	3.3	1.558	0.7373	7.54	2.13	201.9	98.7	153
TRPV1	3.3	11.040	0.8792*	-	-	440.2	15.4	488
Brome Mosaic Virus	3.8	0.855 (A) 15.169 (B) 16.561 (C)	0.6544 0.5459 0.5416	6.04	0.640	221.7	60.8	149 164 164
b-galactosidase	3.2	0.780	0.6999	10.59	0.2	1151.8	98	1018
g-secretase	3.4	0.831 (C) 0.758 (D)	0.7994 0.7825	6.03	3.03	491.5	99.1	243 99
20S proteasome	3.3	9.061 (S) 8.843 (Z)	0.7602 0.7255	15.19 18.95	6.36 4.48	118.9 119.0	88.8 91.6	222 203



MCRV Asymmetric Unit

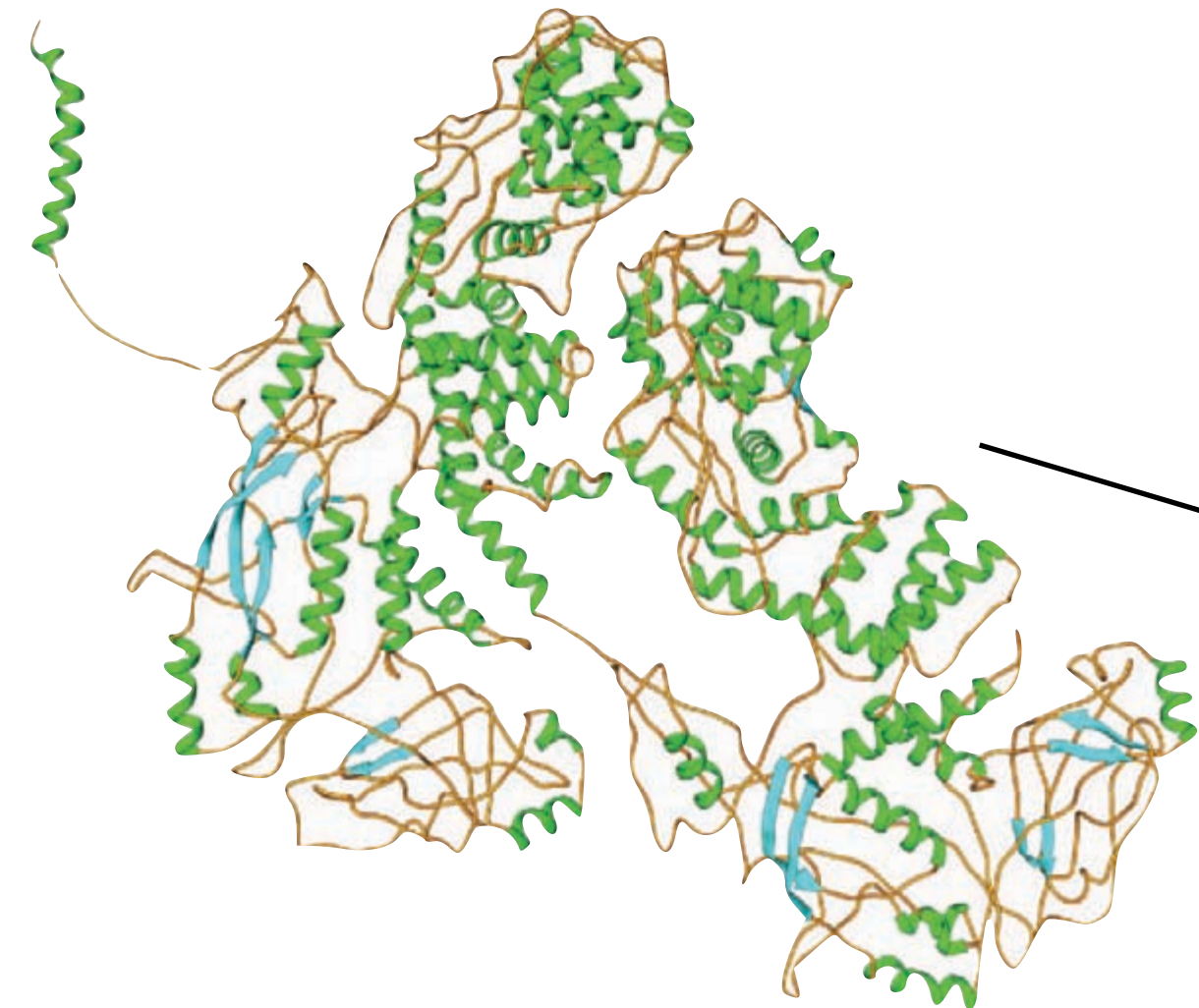
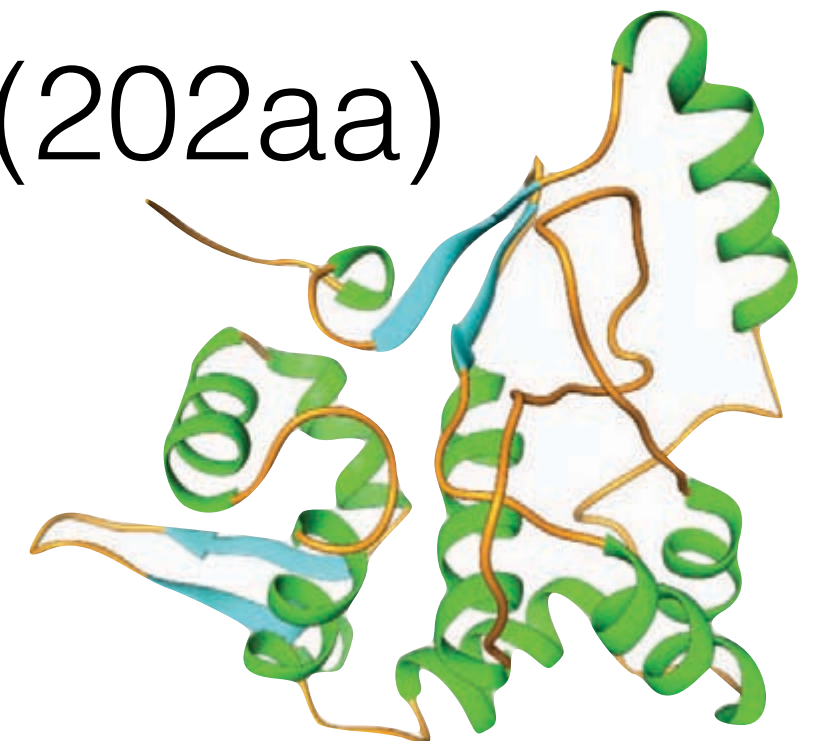


MCRV Asymmetric Unit

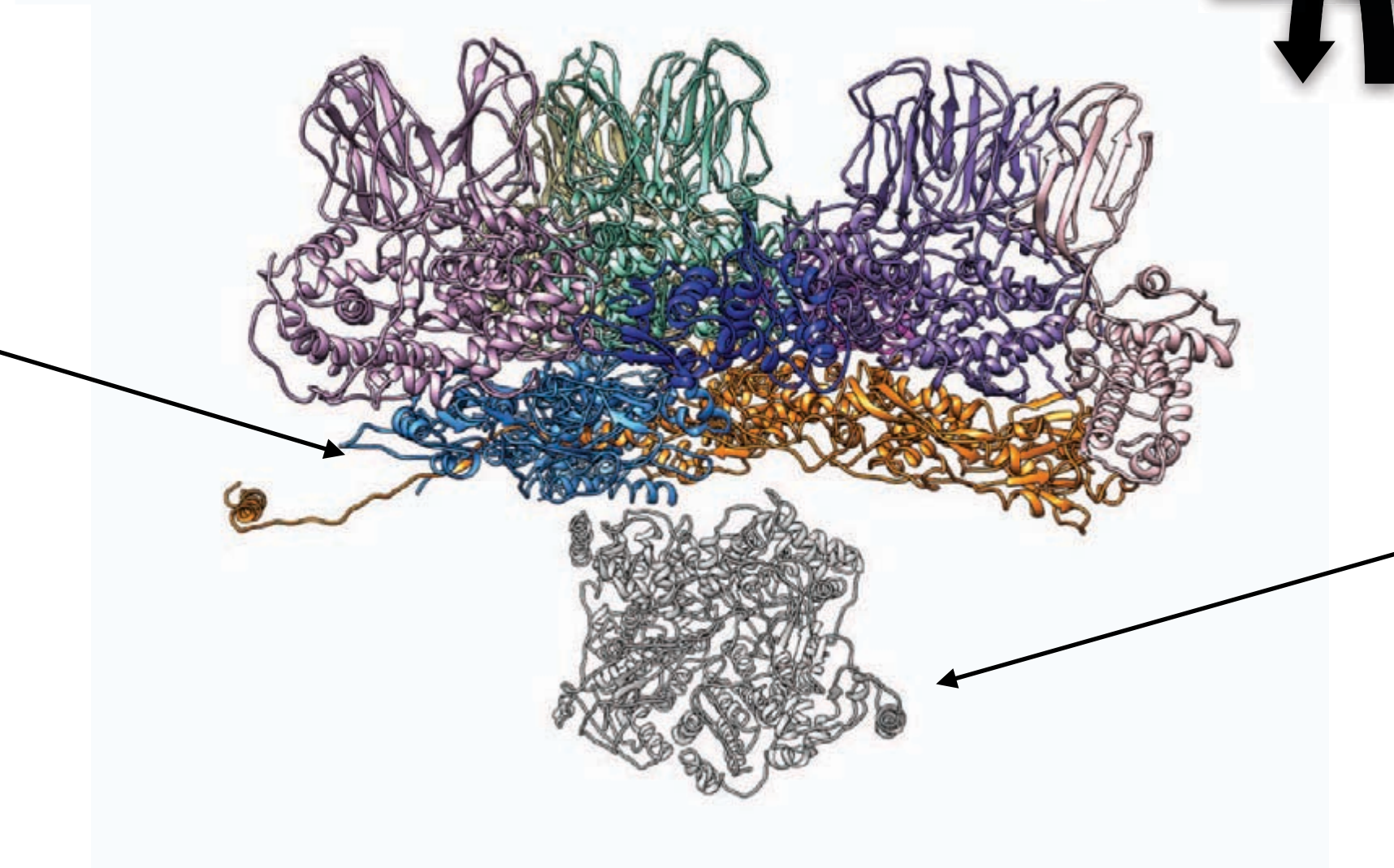
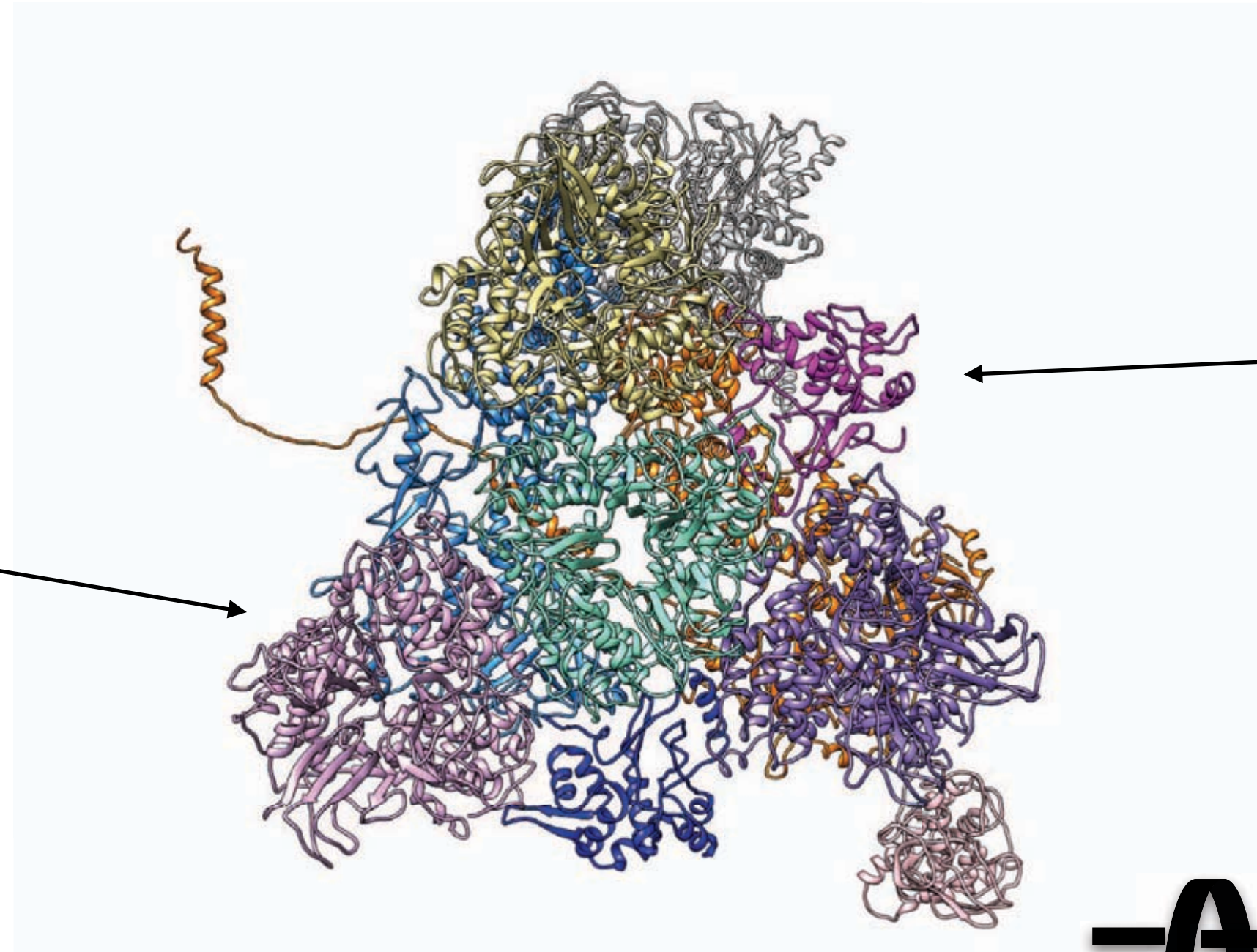
VP12 (274aa)



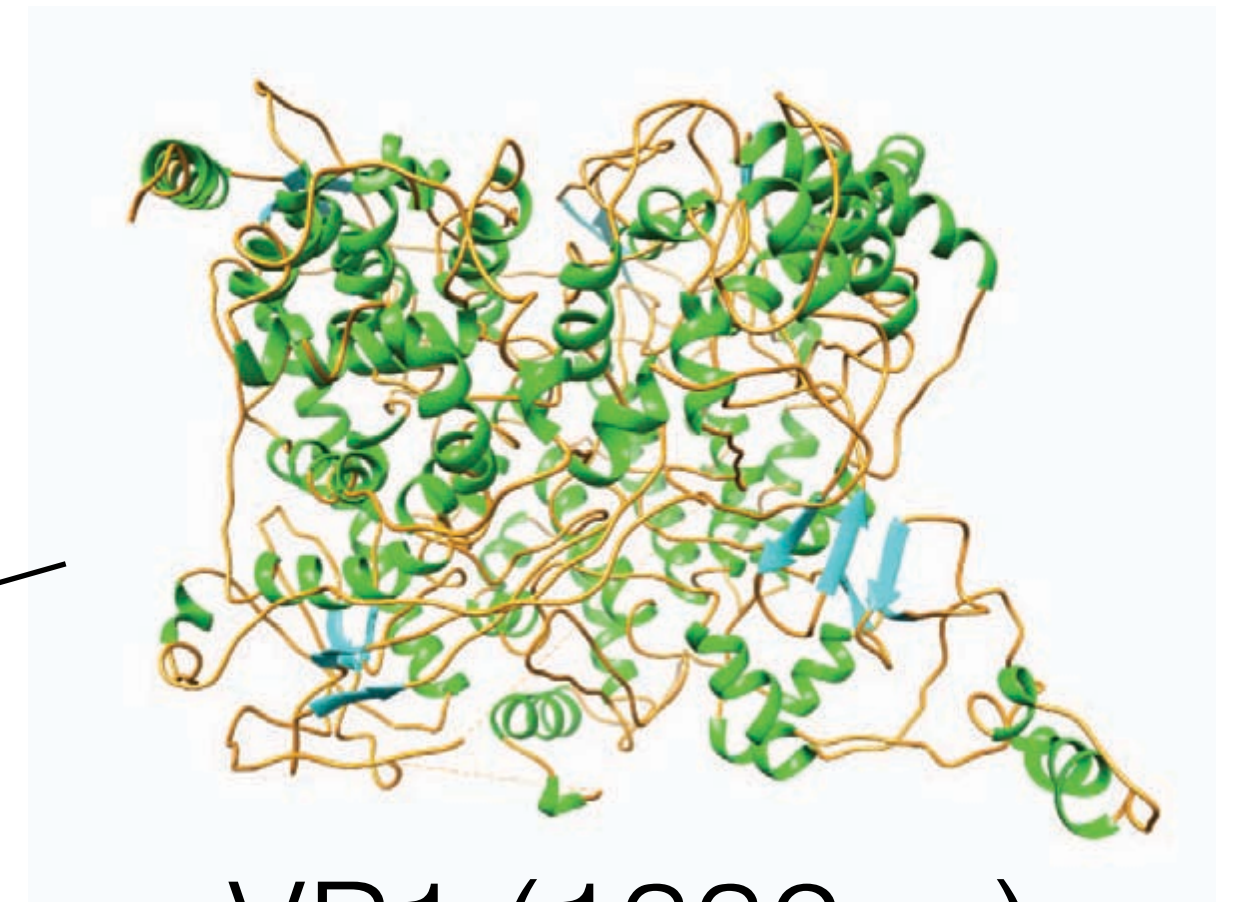
VP11 (202aa)



VP3 (803 and 854aa)

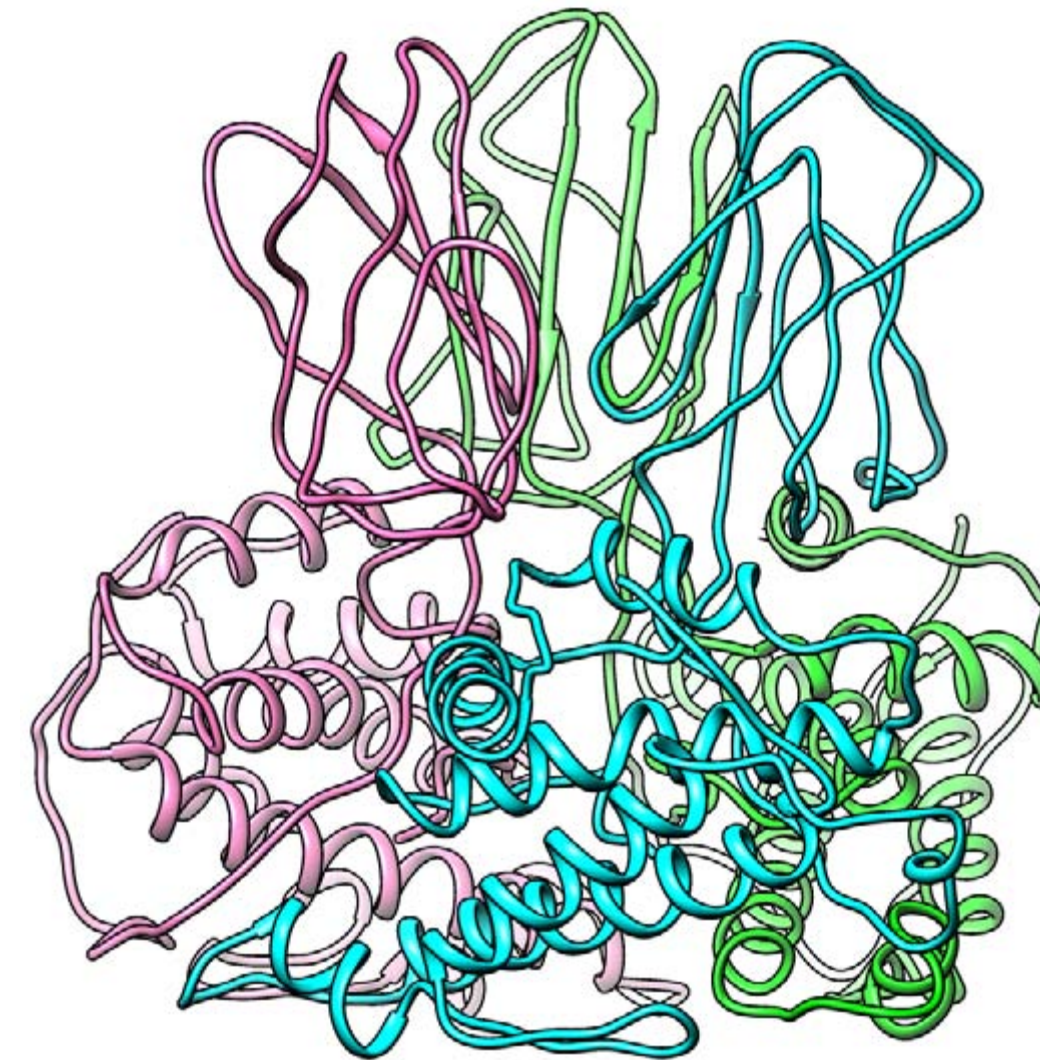
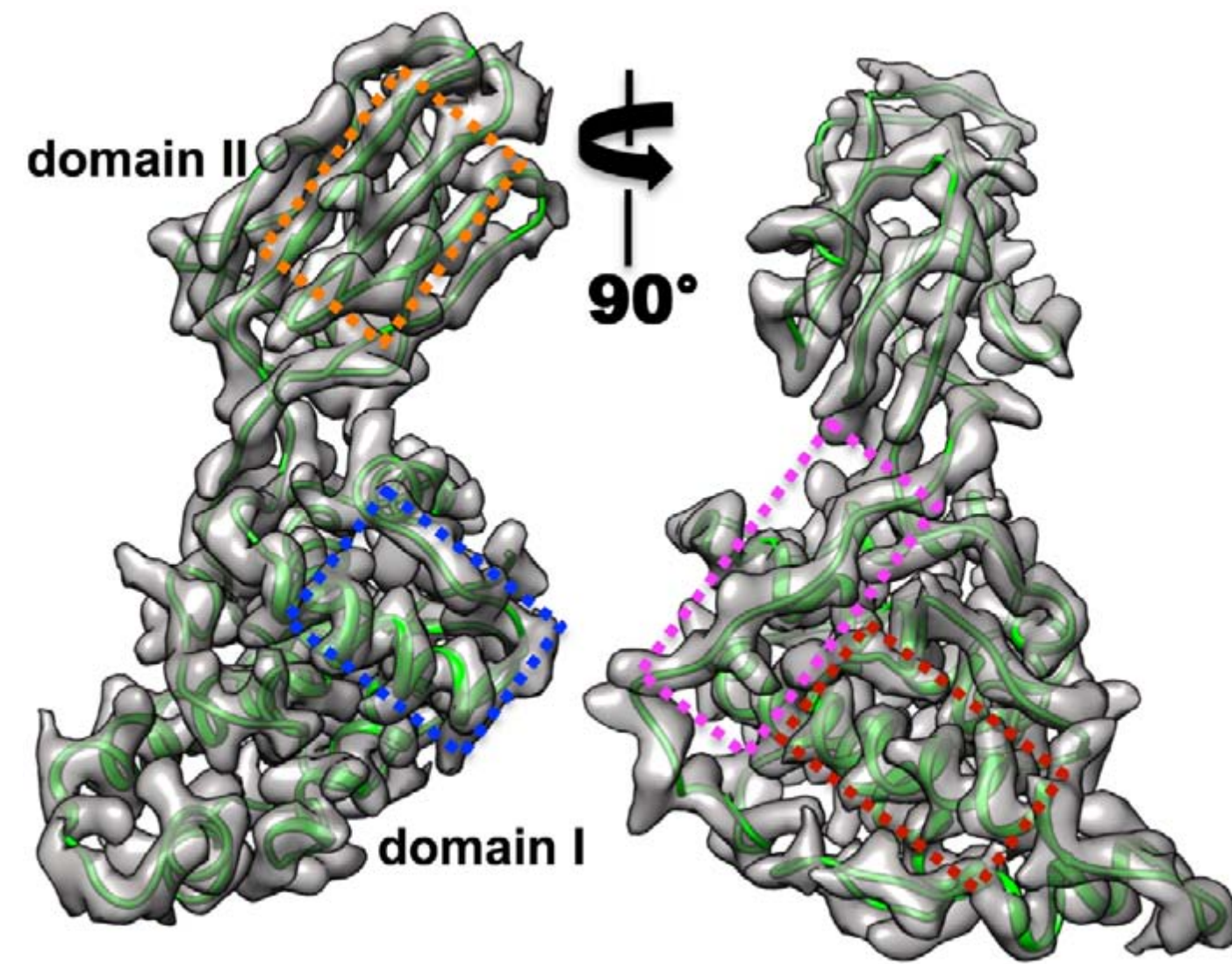
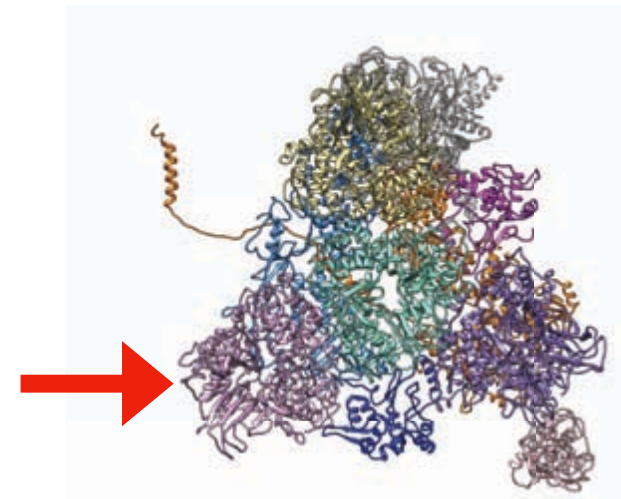


90°



VP1 (1332aa)

VP12

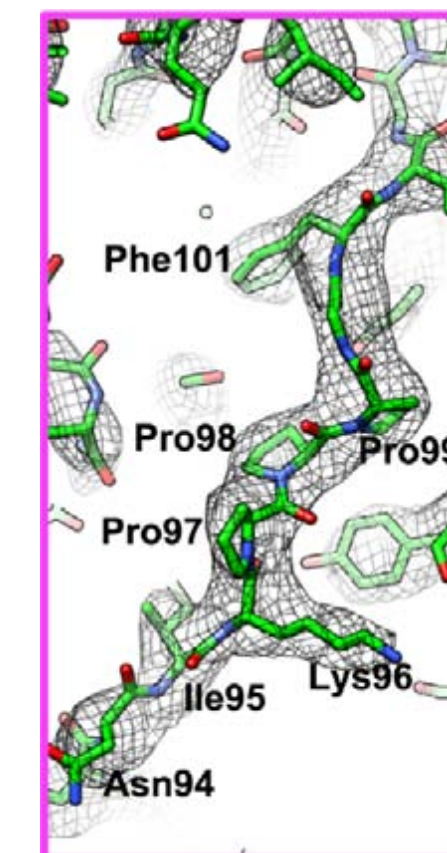
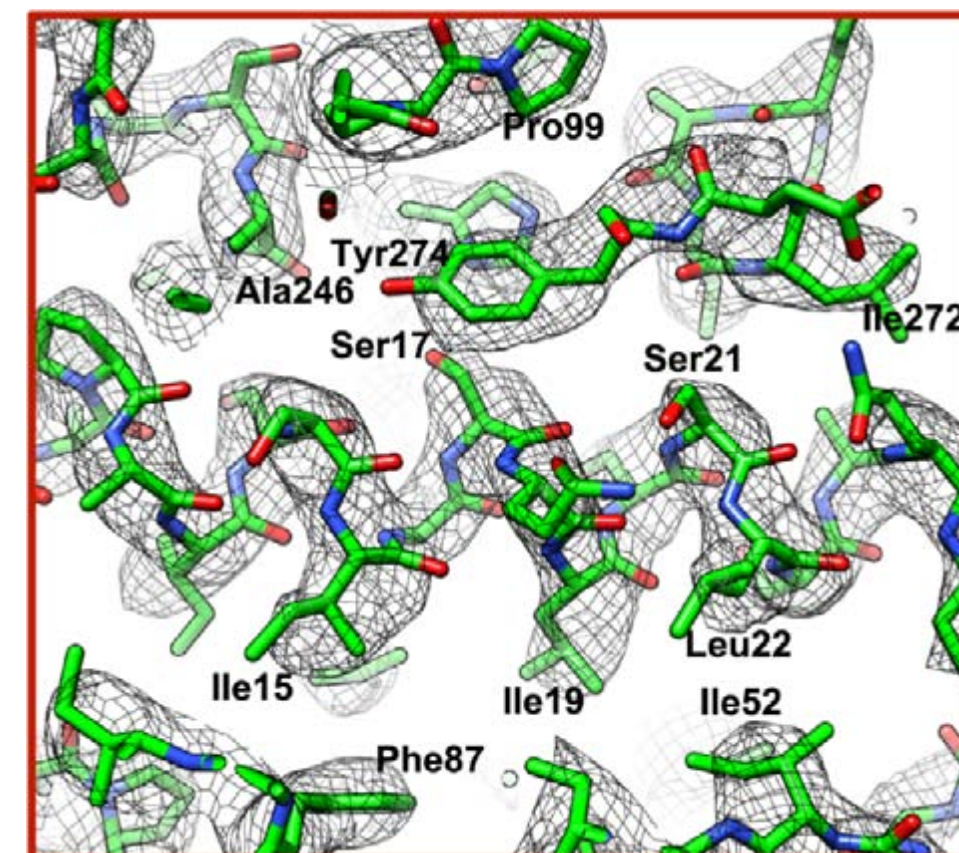
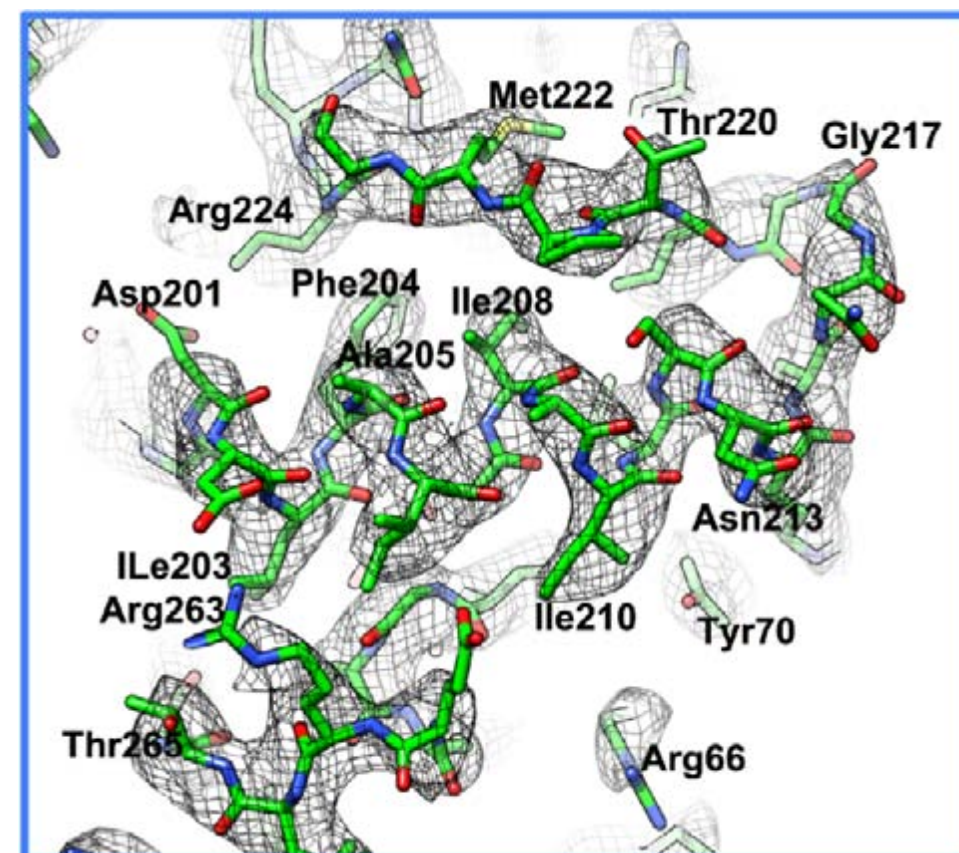
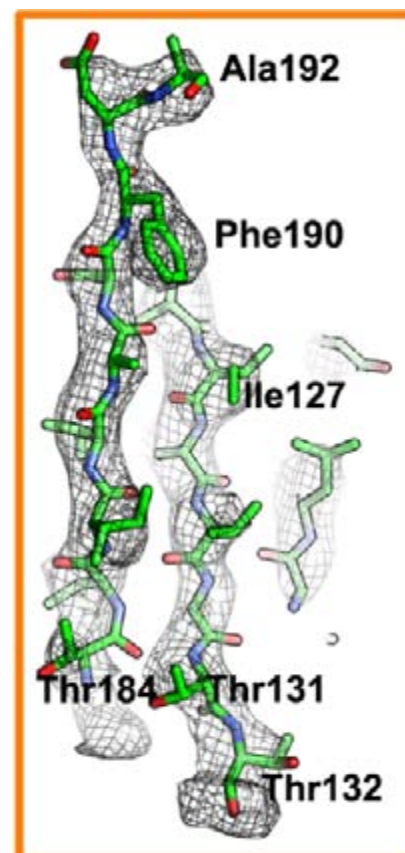


- Primary outer capsid protein
- 4 1/3 unique trimers per ASU
- Structurally homologous to other reovirus outer capsid proteins

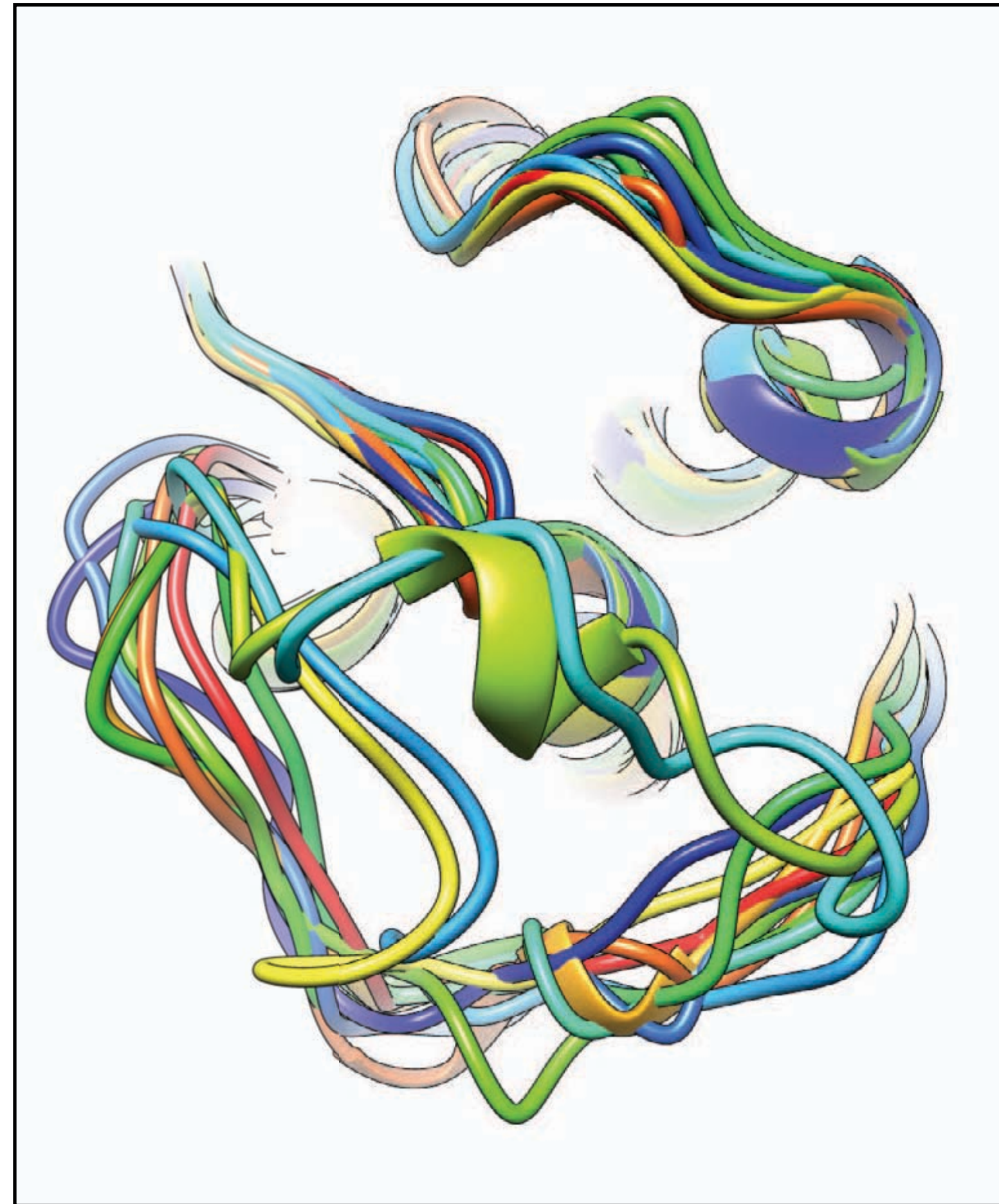
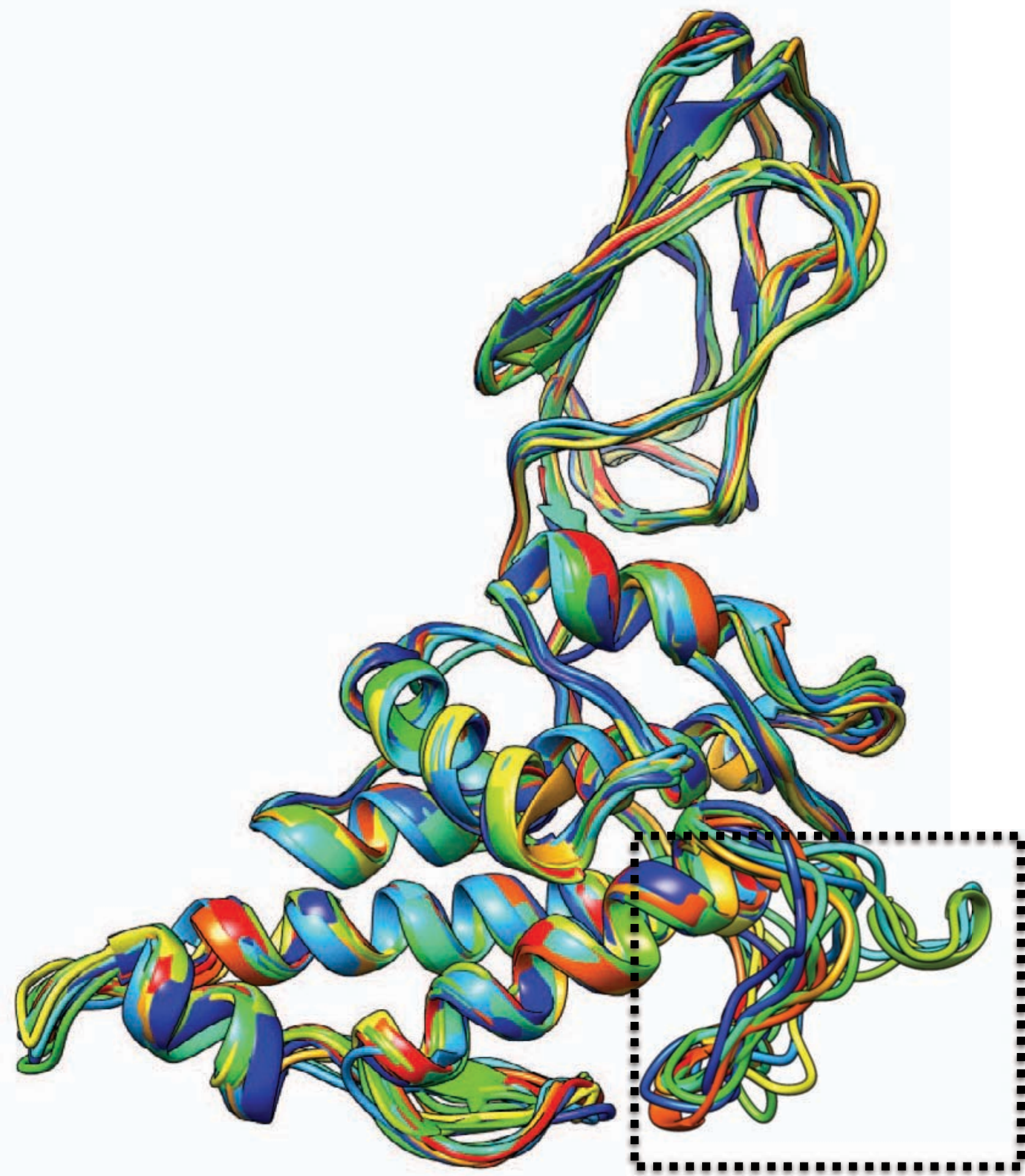
➡ Upper domain beta sandwich

➡ PPPGY motif between domains

➡ Lower helical domain resist on inner capsid protein VP3

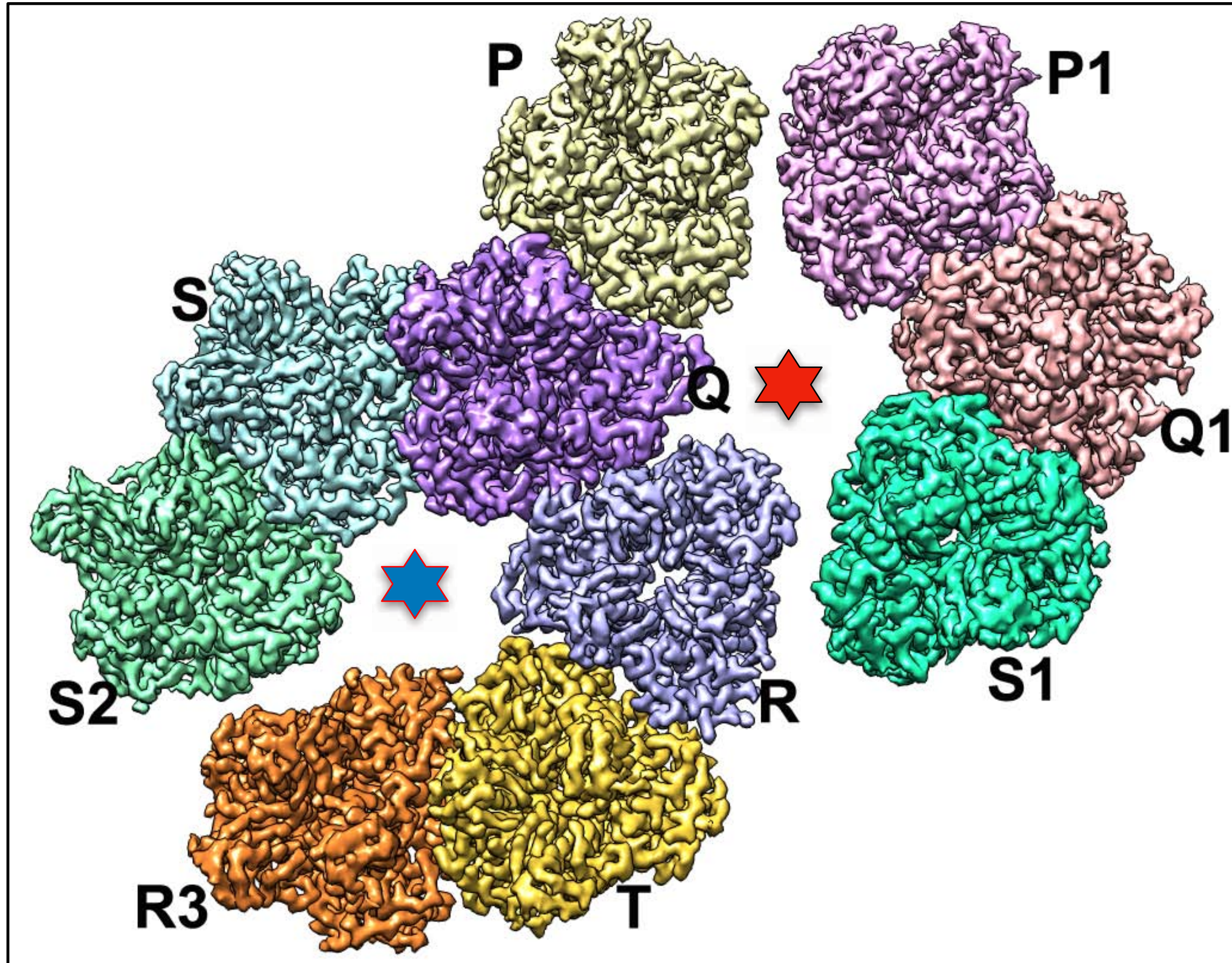


Variability in VP12

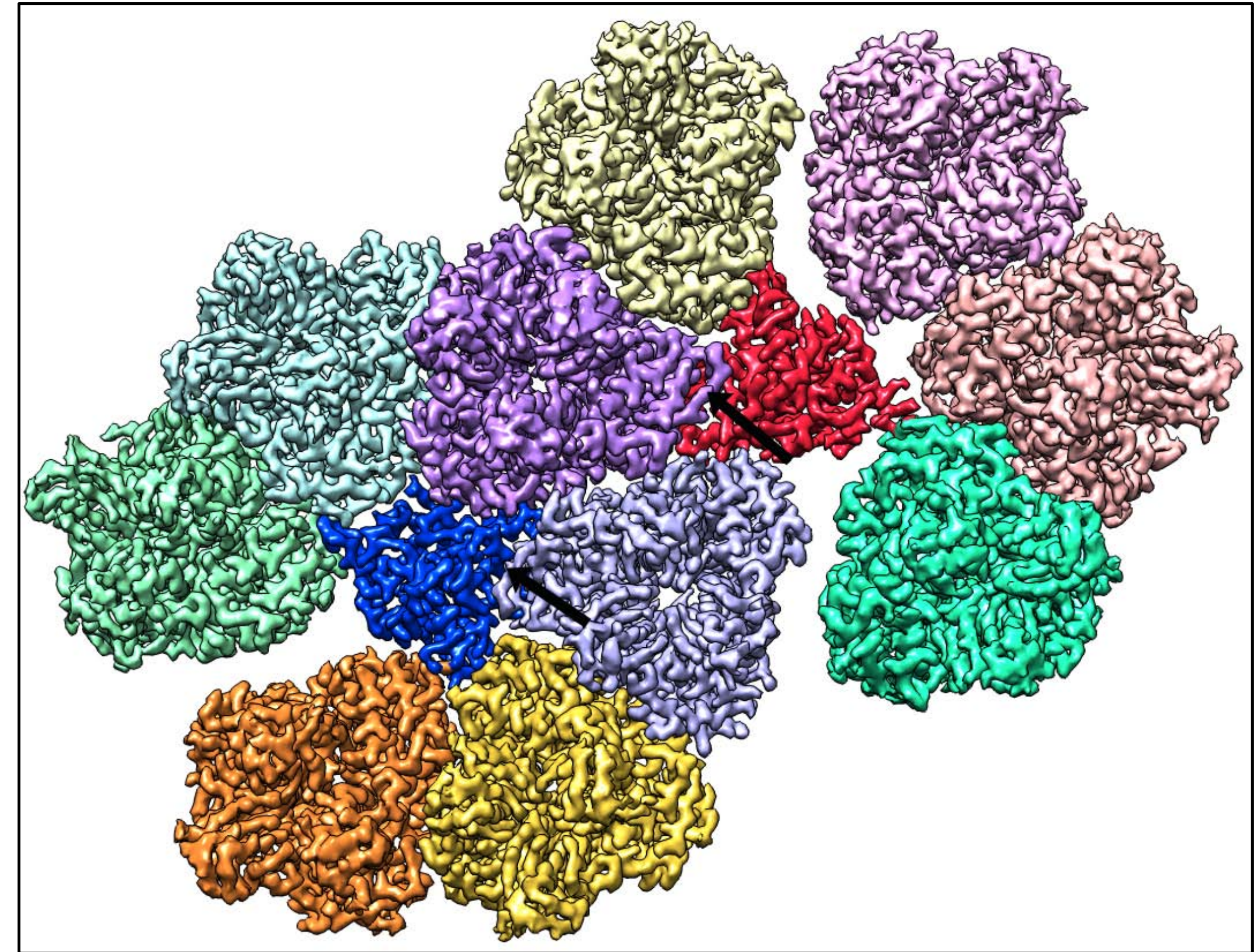
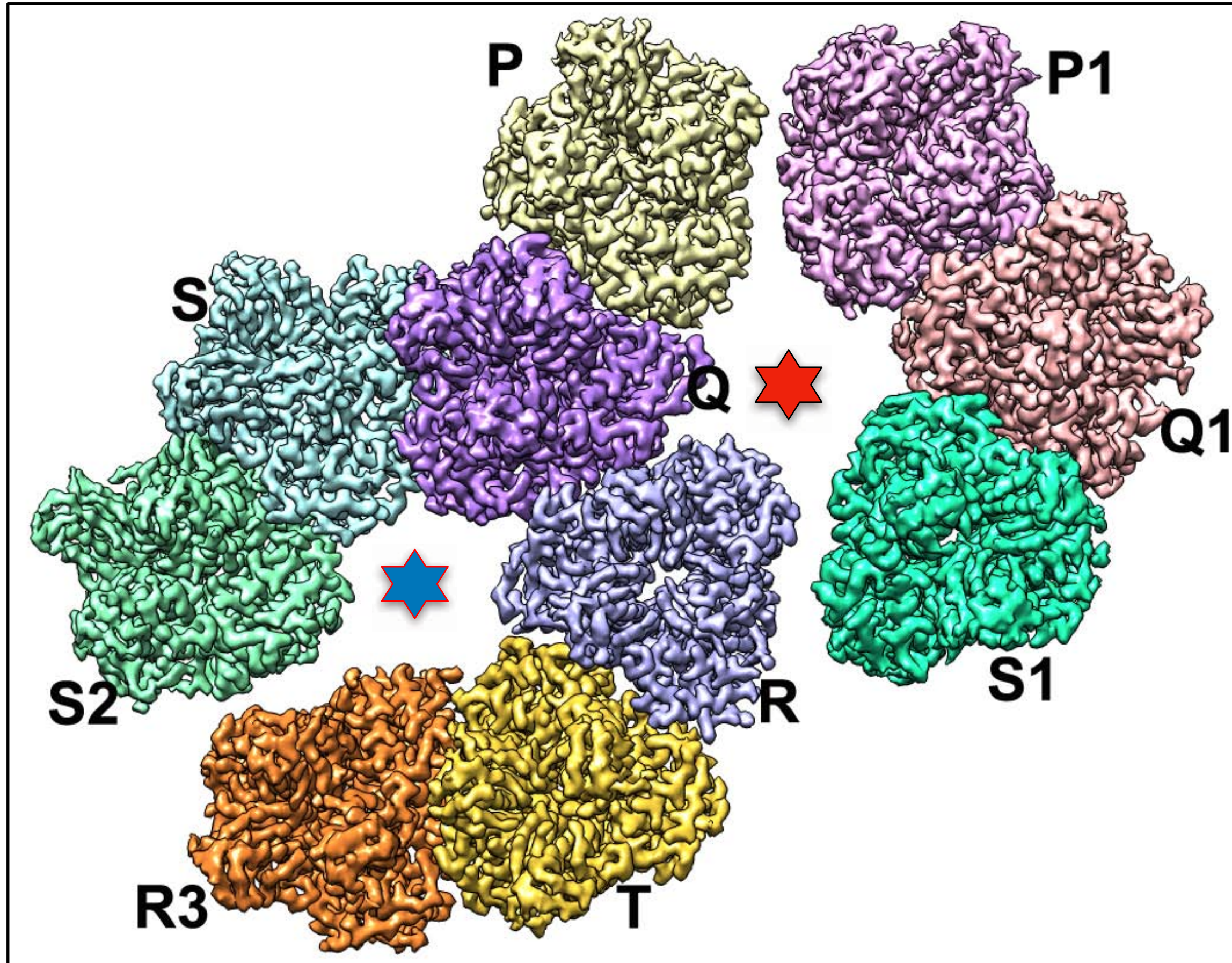


- RMSD across all 13 VP12s is $\sim 1\text{\AA}$, except in 2 subunits where RMSD is $\sim 2.9\text{\AA}$
- Residues 81-94 form a loop and have $>15\text{\AA}$ atom-atom deviations
 - ➔ Largely polar
 - ➔ can assume α -helical or loop conformations in different contexts
 - ➔ Loops faces toward local 6-fold
 - ➔ Only MCRV in reoviruses has this motif

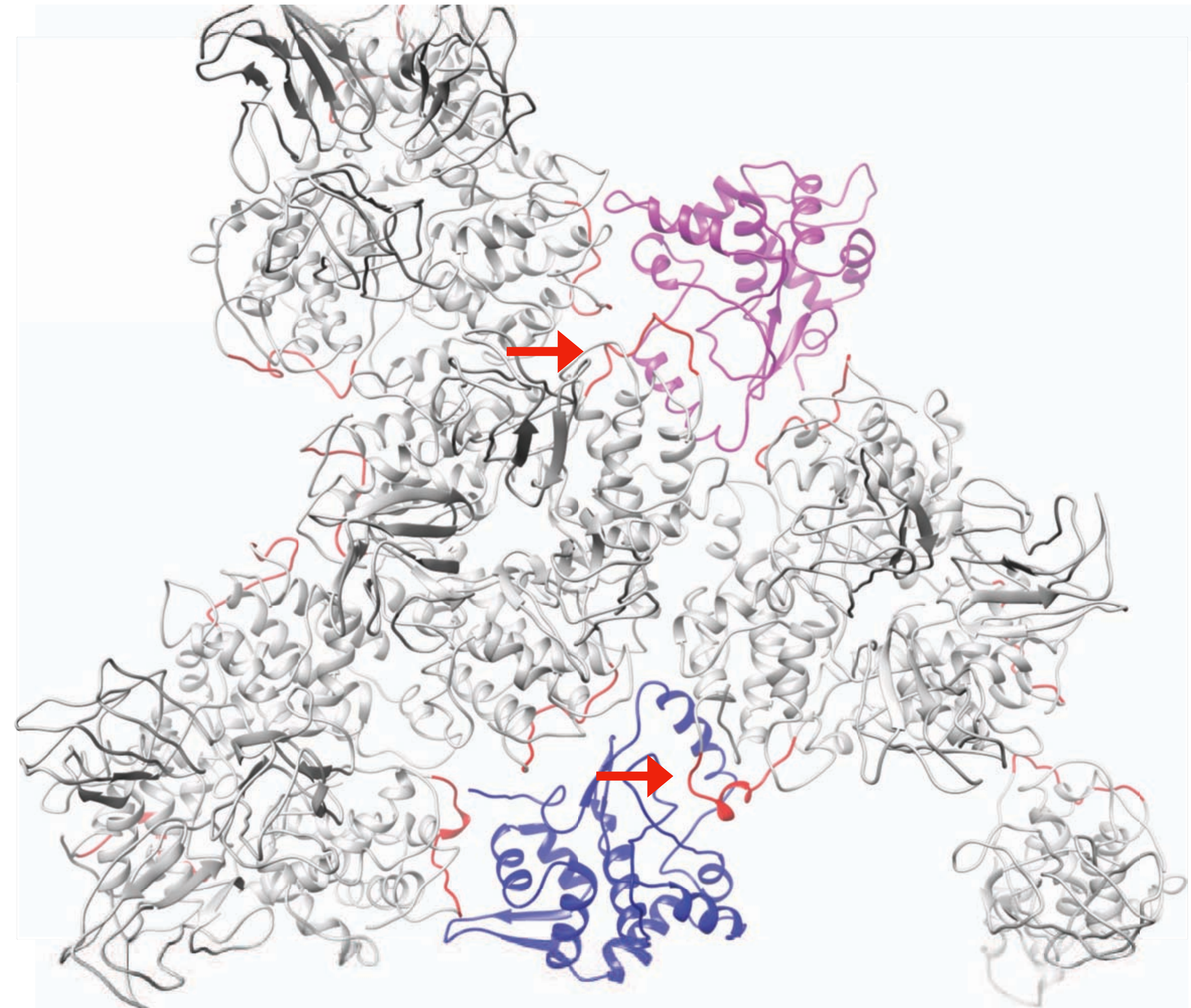
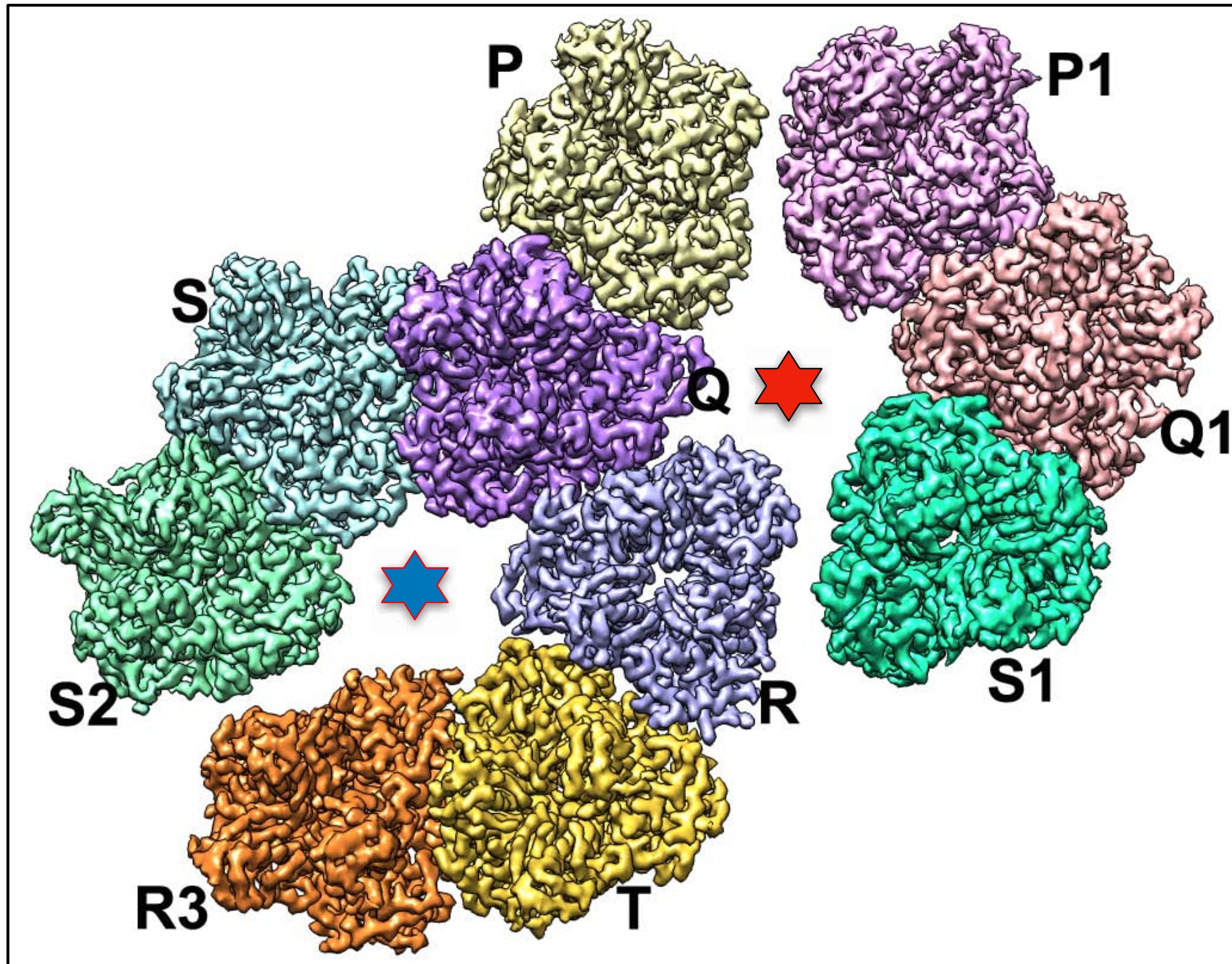
VP11-VP12 Interactions



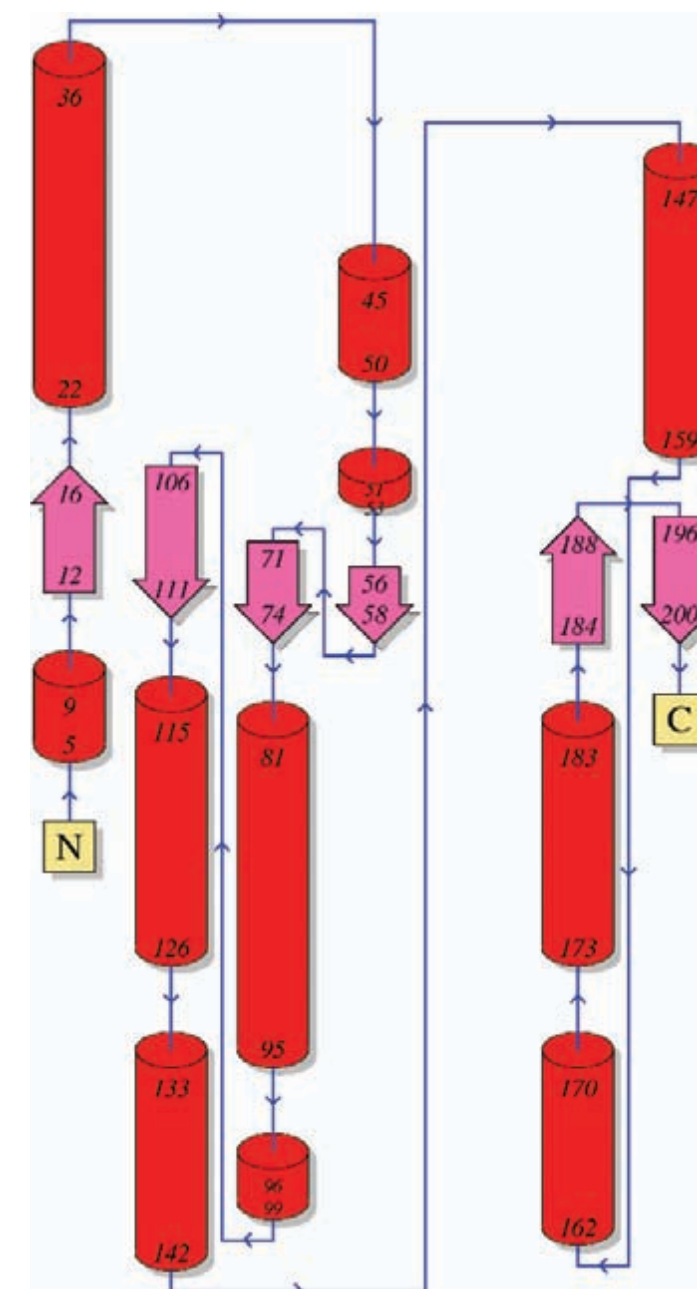
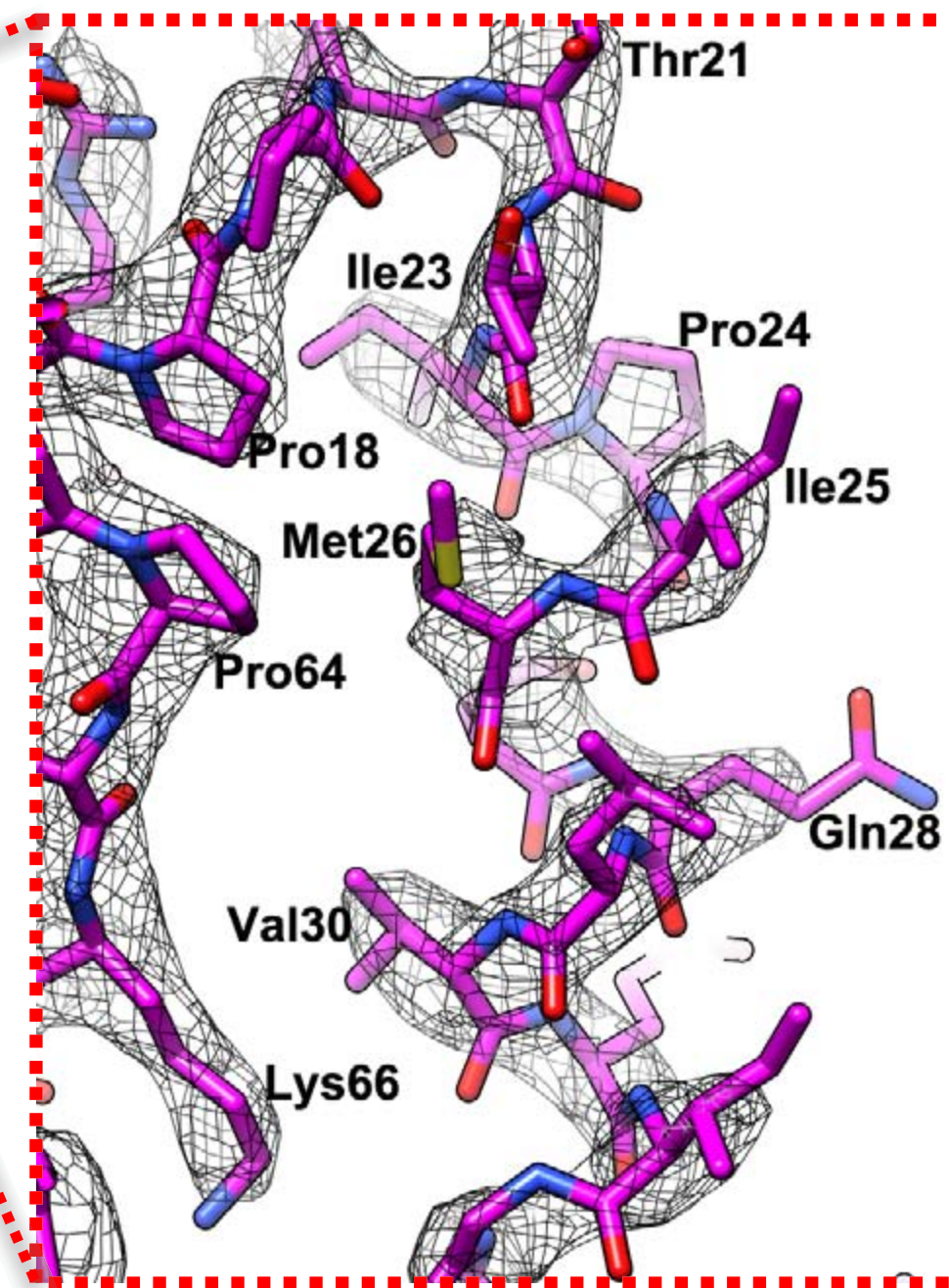
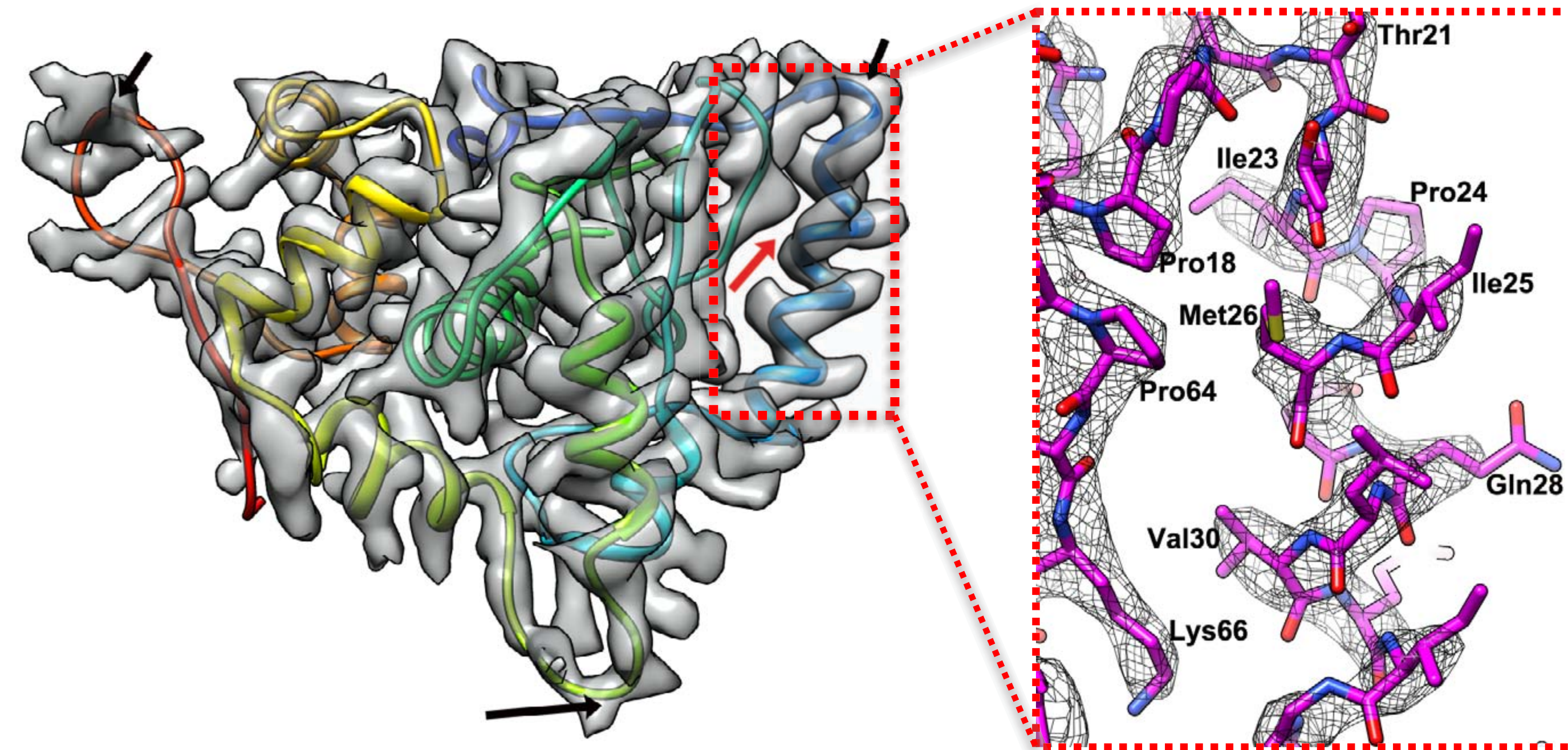
VP11-VP12 Interactions



VP11-VP12 Interactions



VP11 - A novel capsid protein

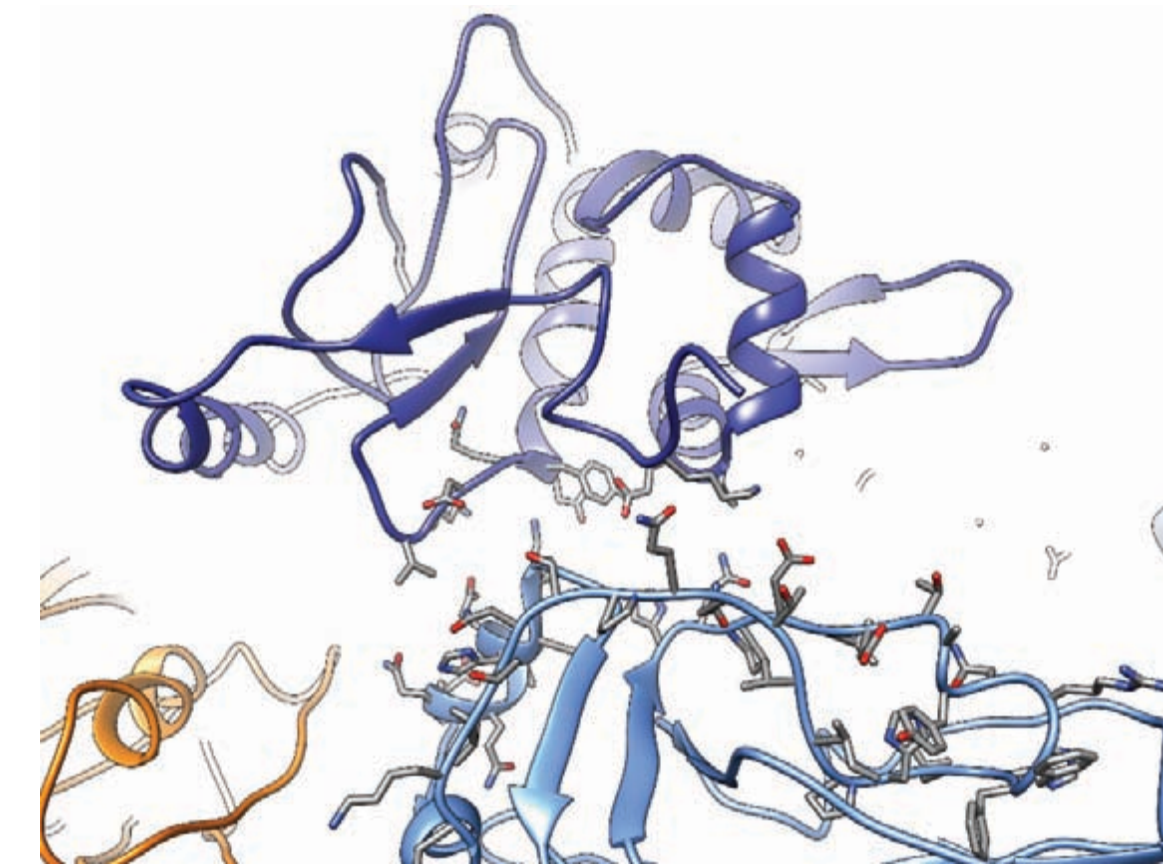
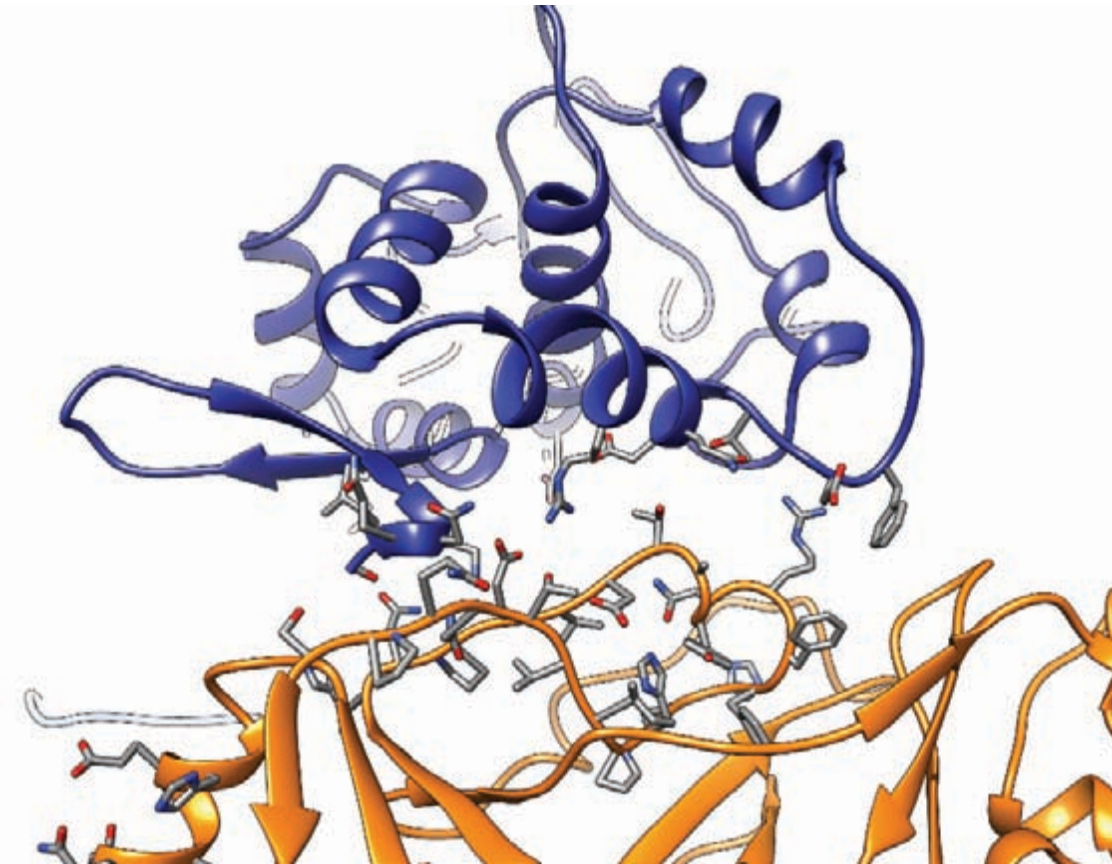
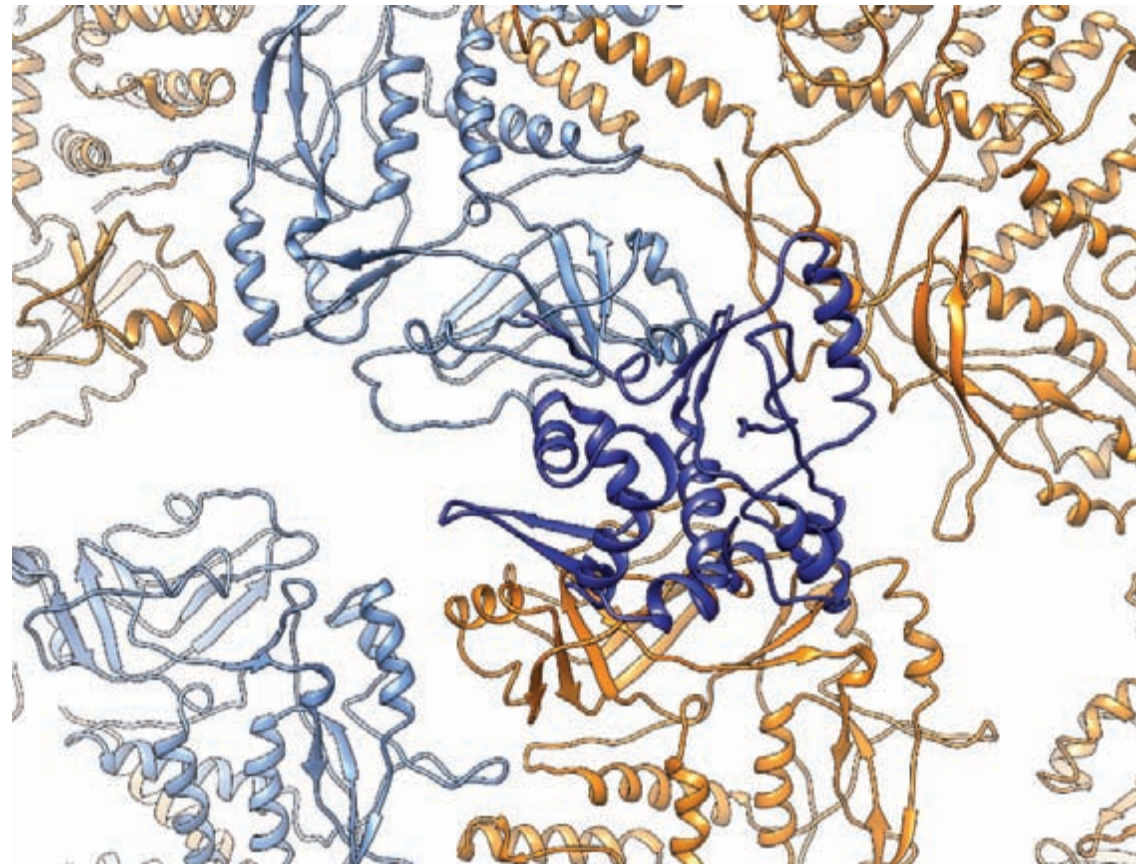


- No sequence or structural homologues
- Occupies same location as clamp proteins in other reoviruses (CPV, Fako, orthoreovirus and aquareovirus)
 - ➔ similar interactions with sites i and ii on inner capsid proteins
 - ➔ conserved hydrophobic helix (L82-Y98) with 5 aromatics

VP11 - Stabilizing the capsid

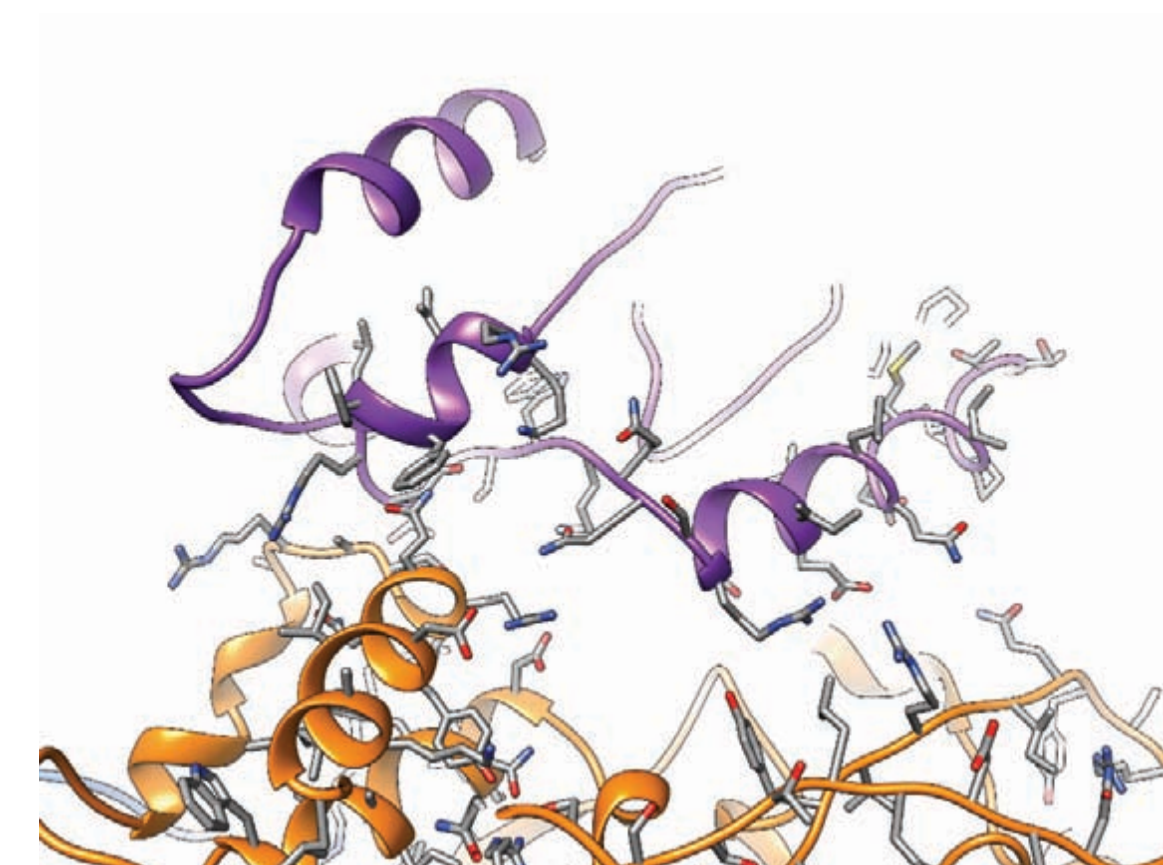
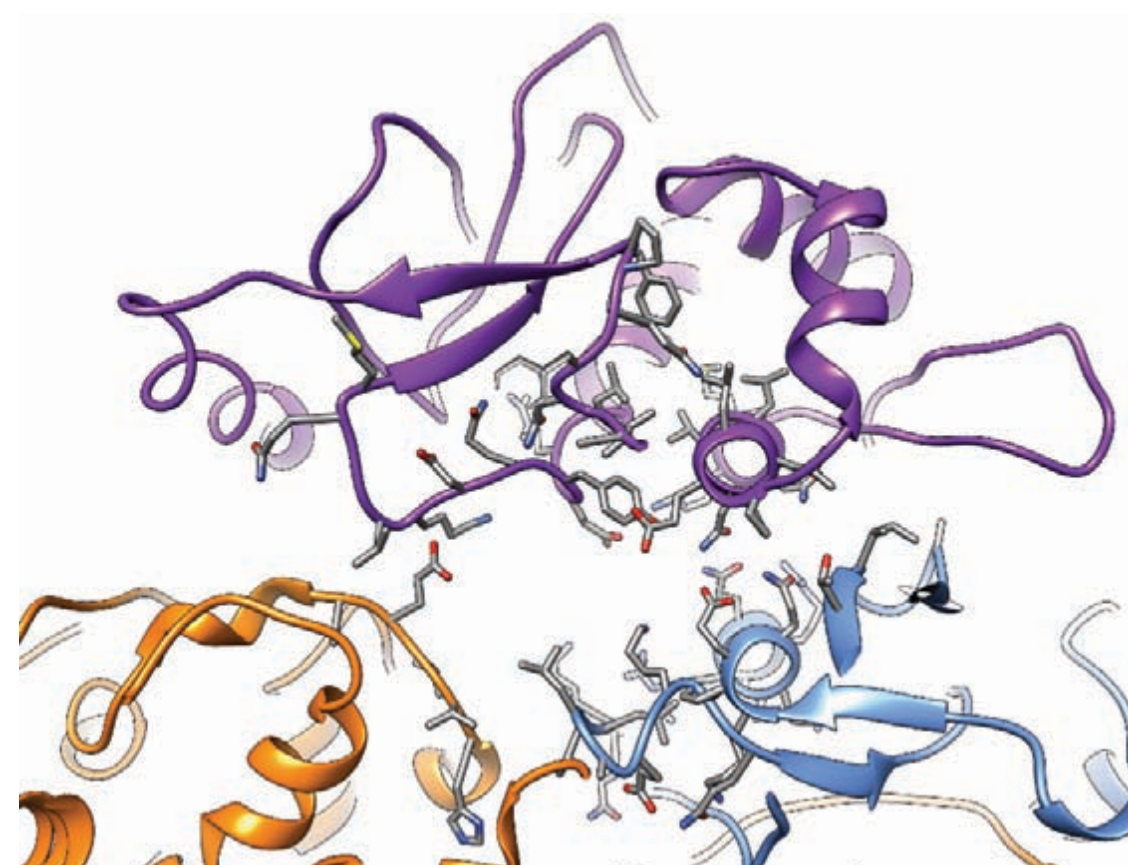
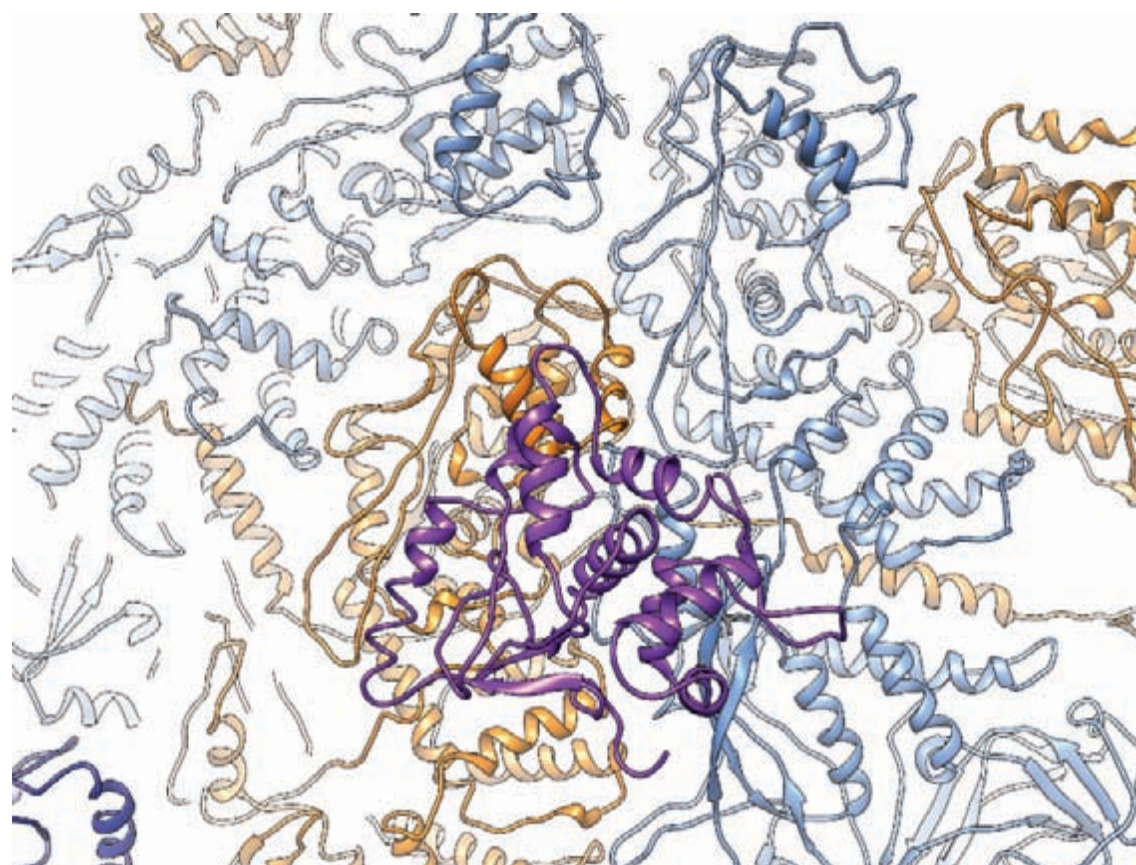


site ii



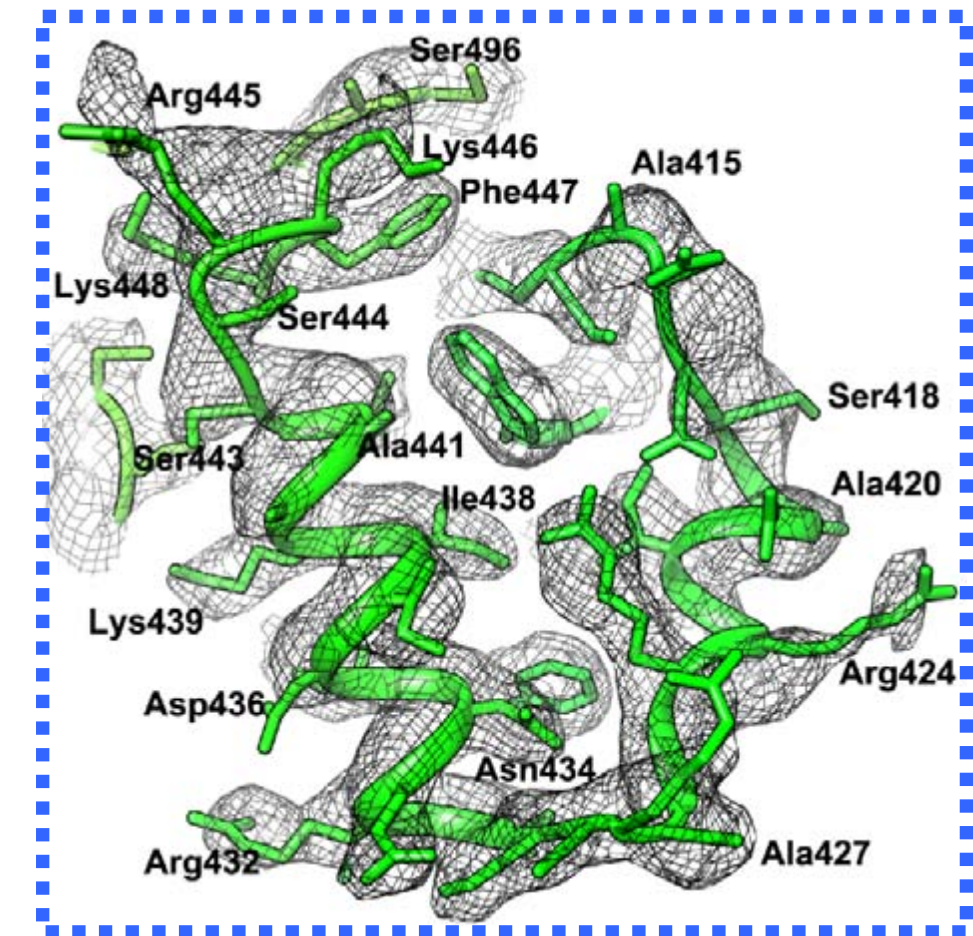
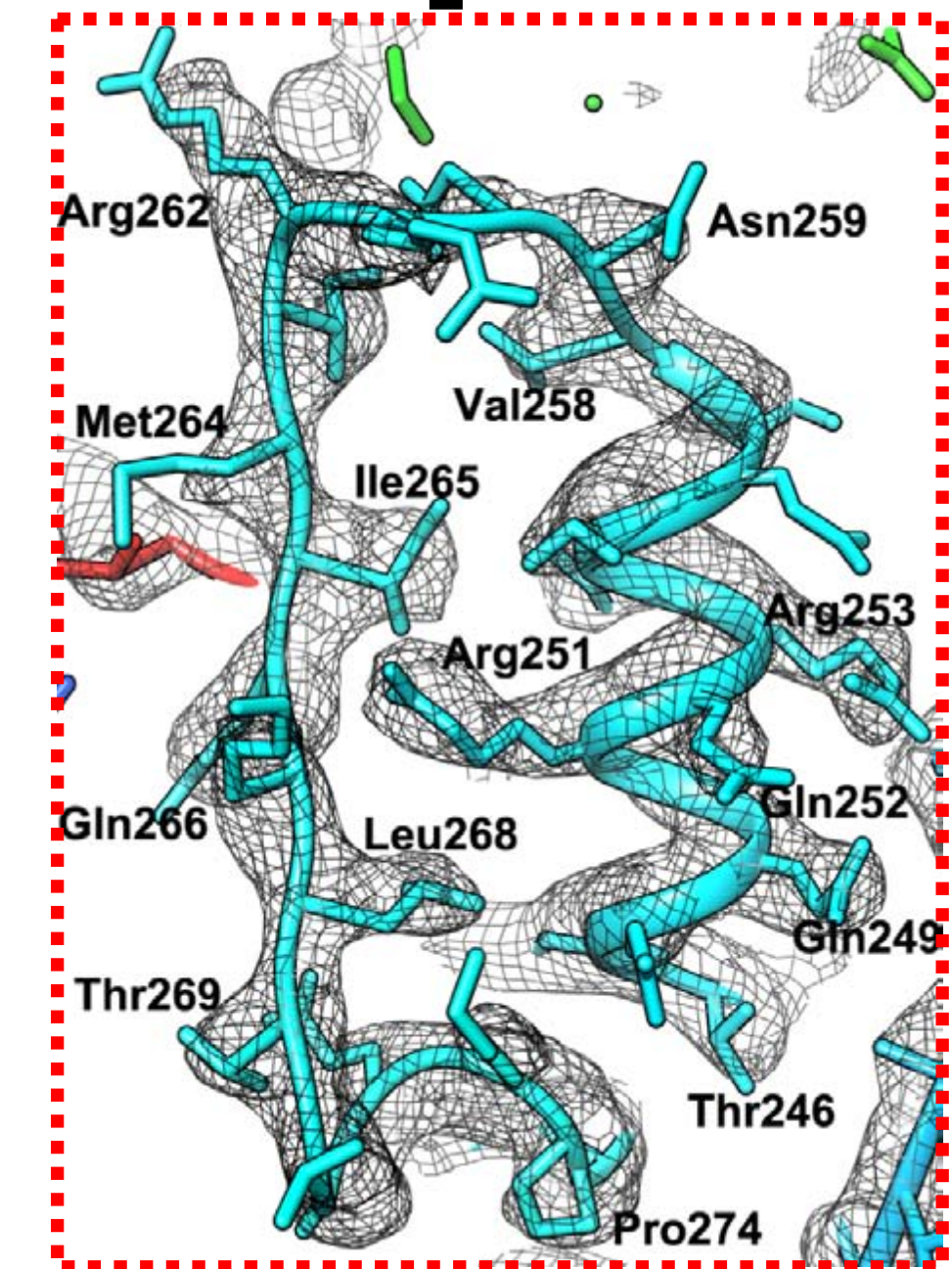
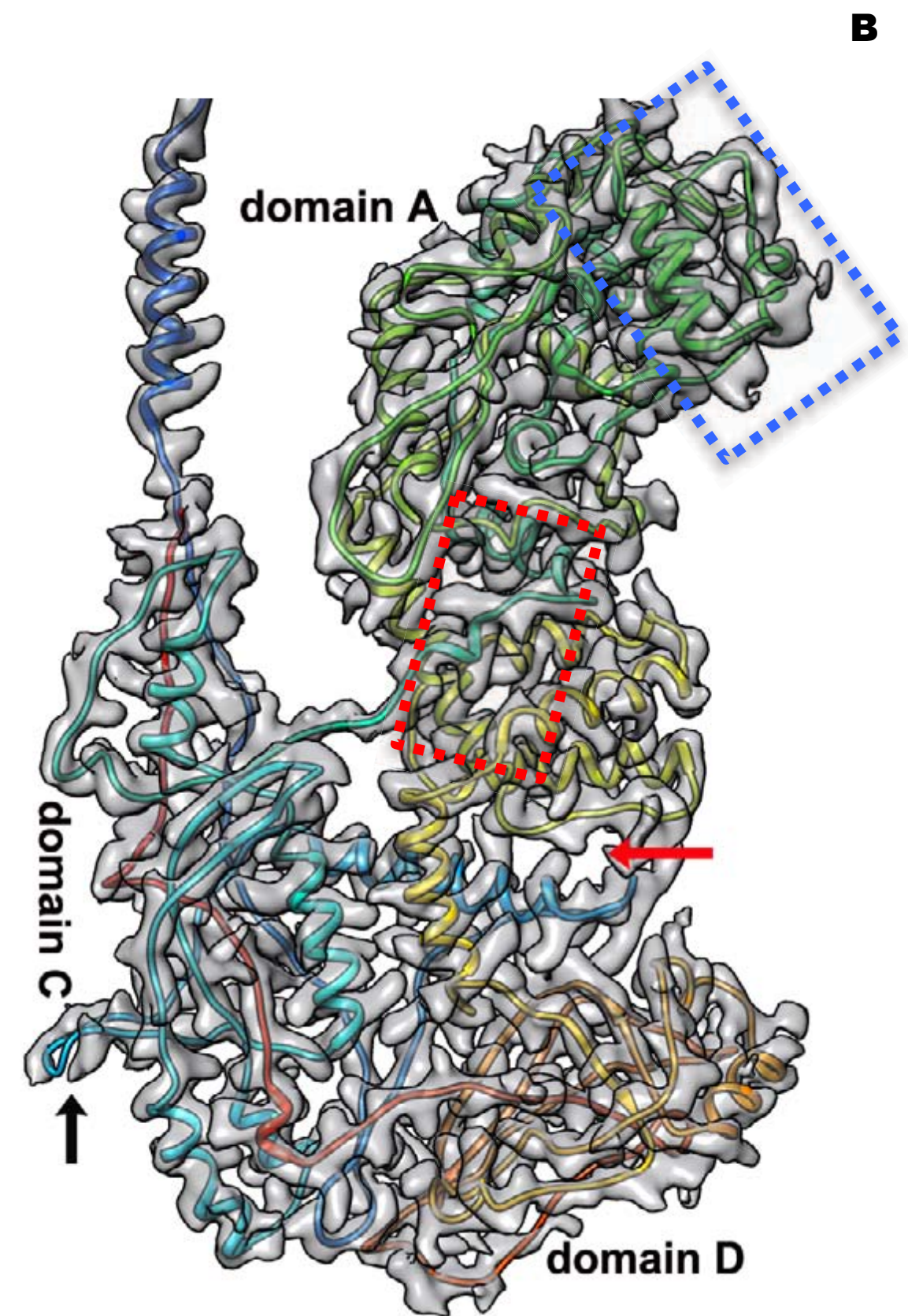
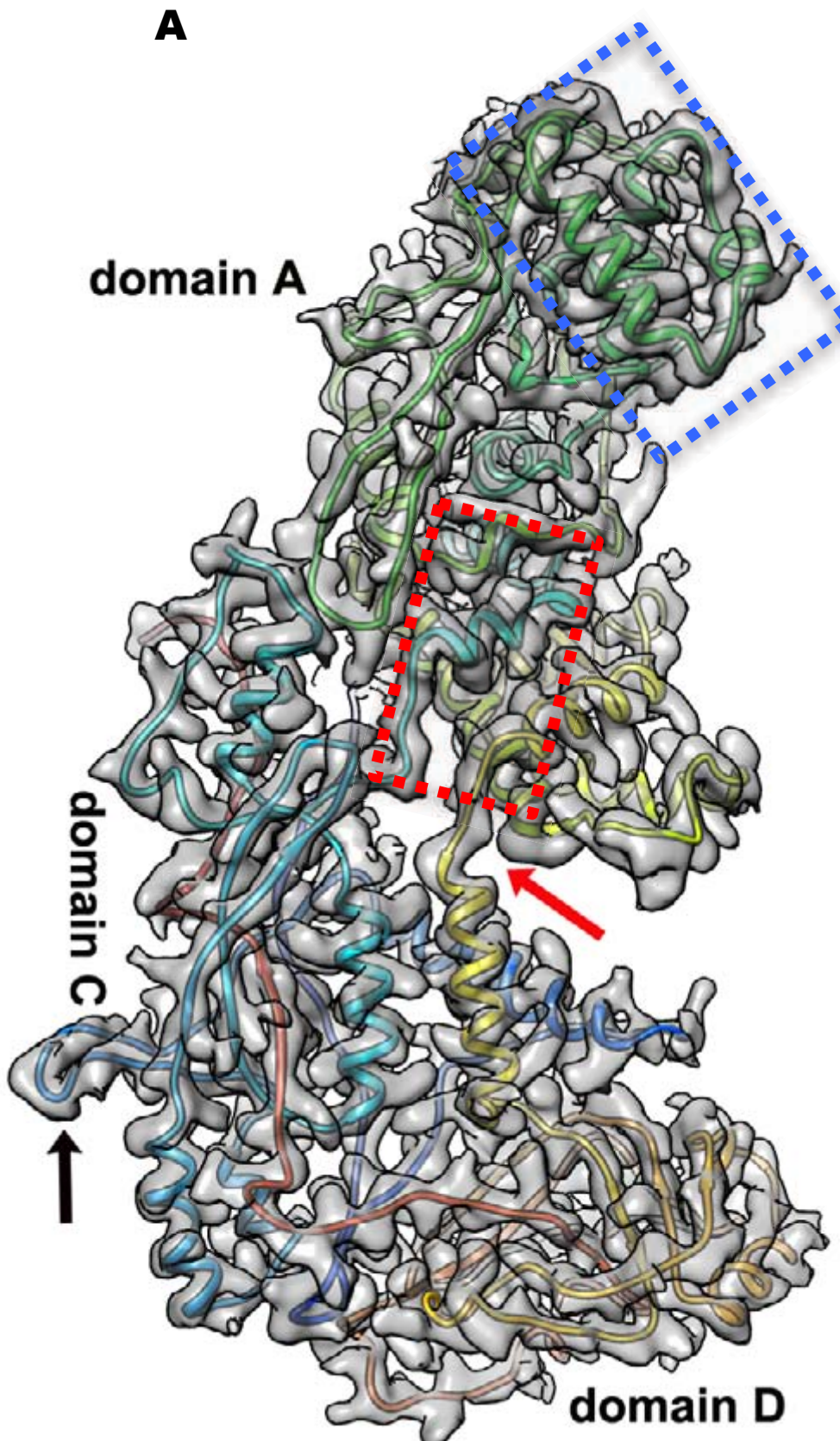
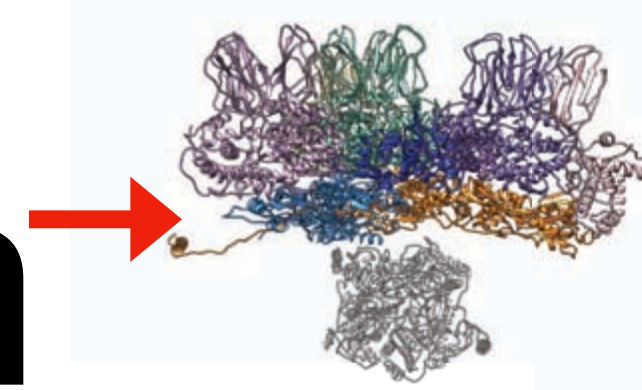
VP11 interacts with 2 VP3 s across the dimerization domain (residues 696-702 and 734-744)

site i



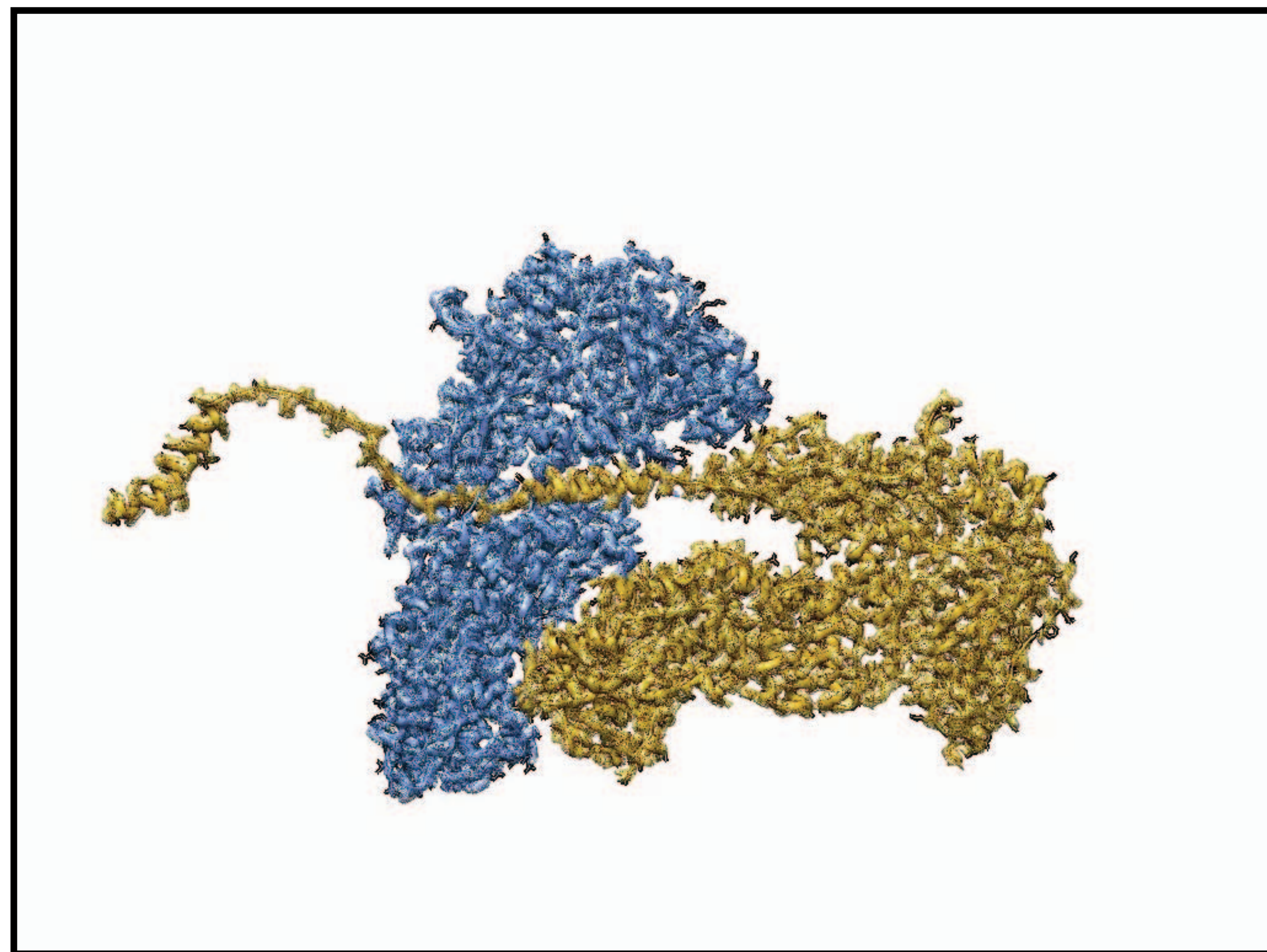
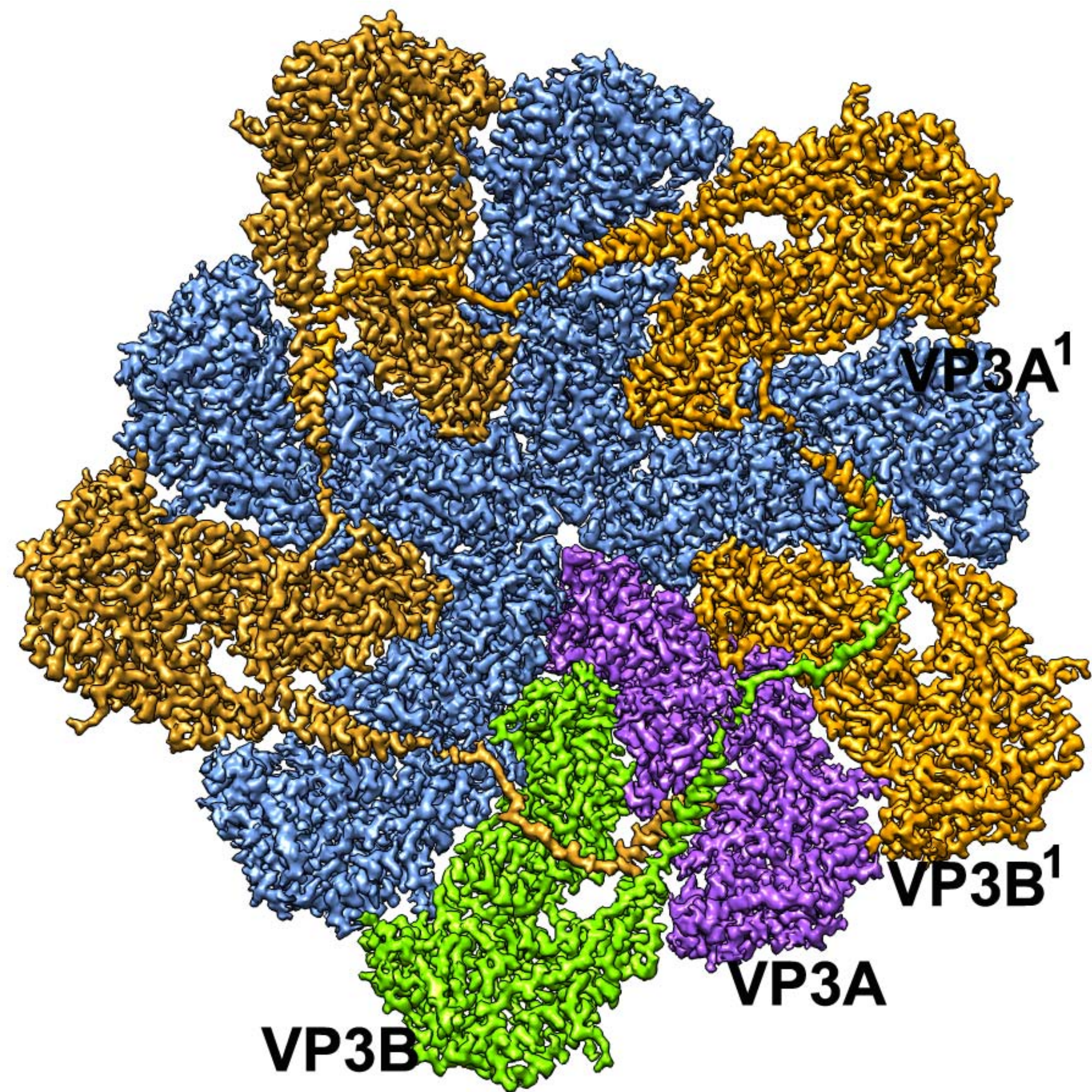
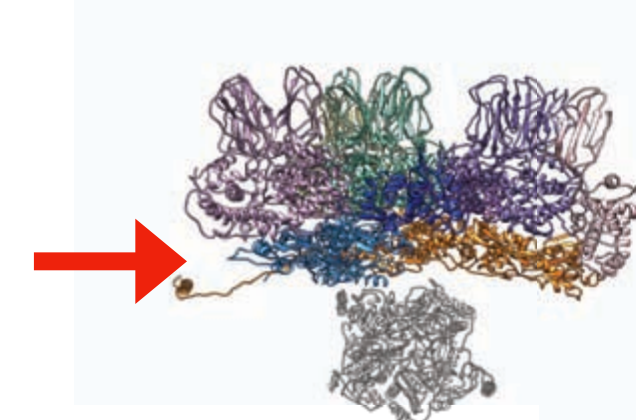
VP11 interacts with 1 apical domain (419-441) and 1 carapace (200-220, 247-256) domain from adjacent ASUs

VP3- Inner shell capsid protein

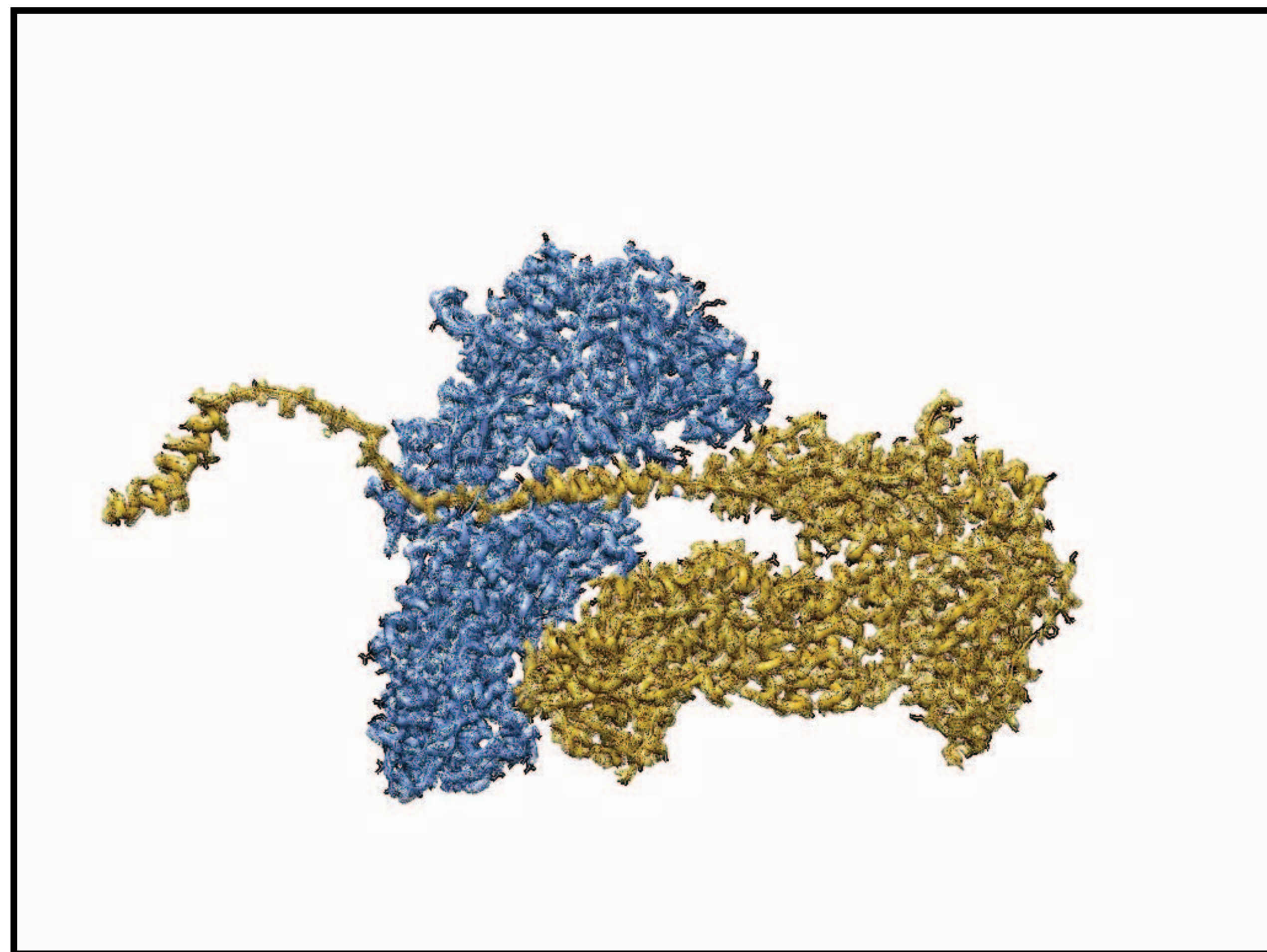
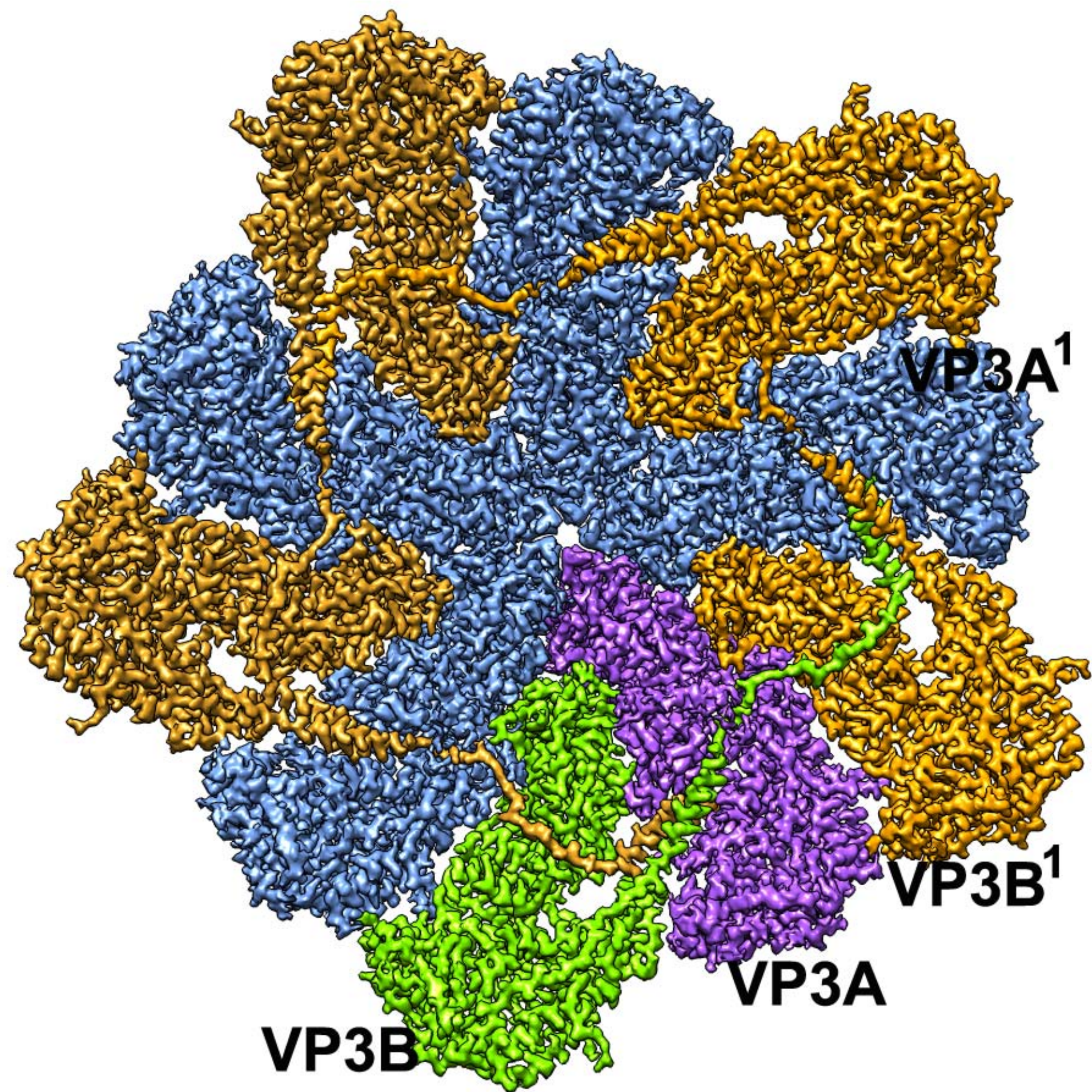
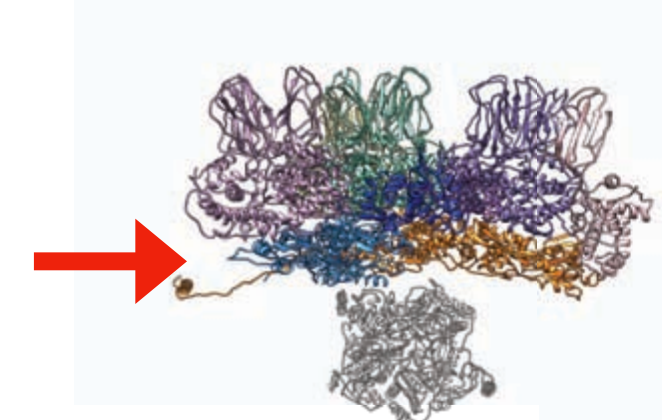


- Pseudo-dimer of VP3A and VP3B
- ~6.9Å RMSD between VP3A and VP3B
- Largest differences in the central and dimerization domain
- VP3B N-arm (1-85) contains Helix-loop-helix-loop motif that extends across 2 adjacent VP3 molecules
- VP3 N-arm is largely disordered and adopts 3 unique conformations
- VP3AB are nearly identical in active and quiescent states

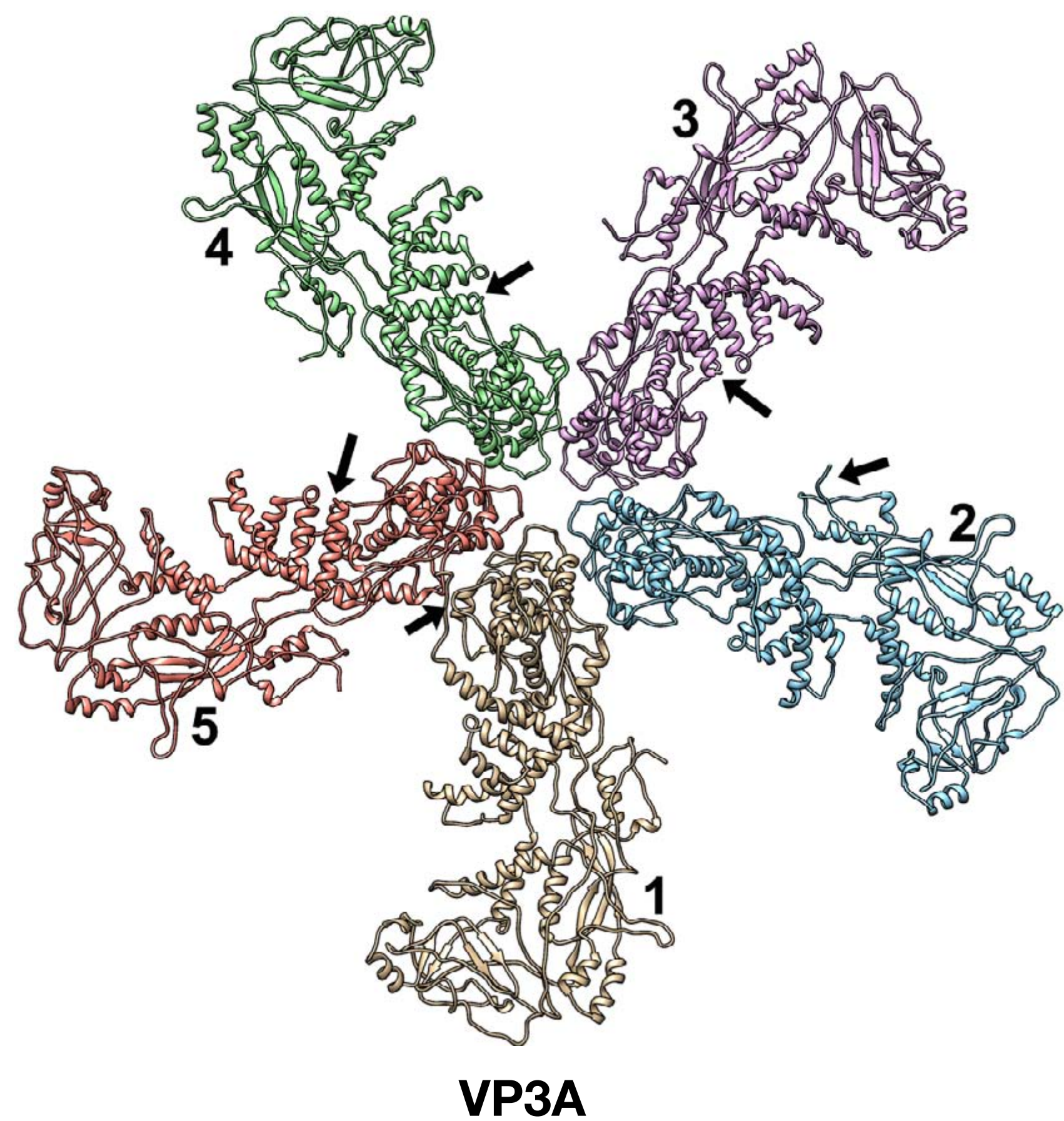
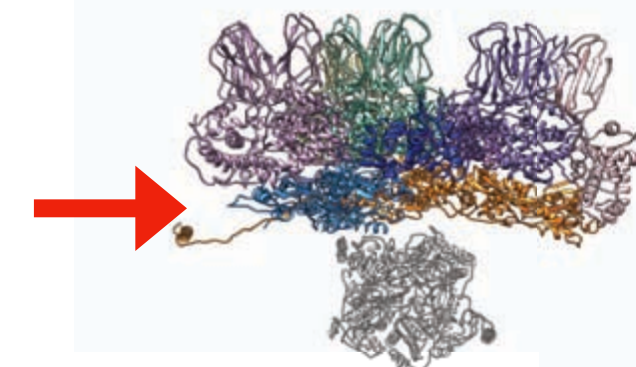
VP3B N-arm



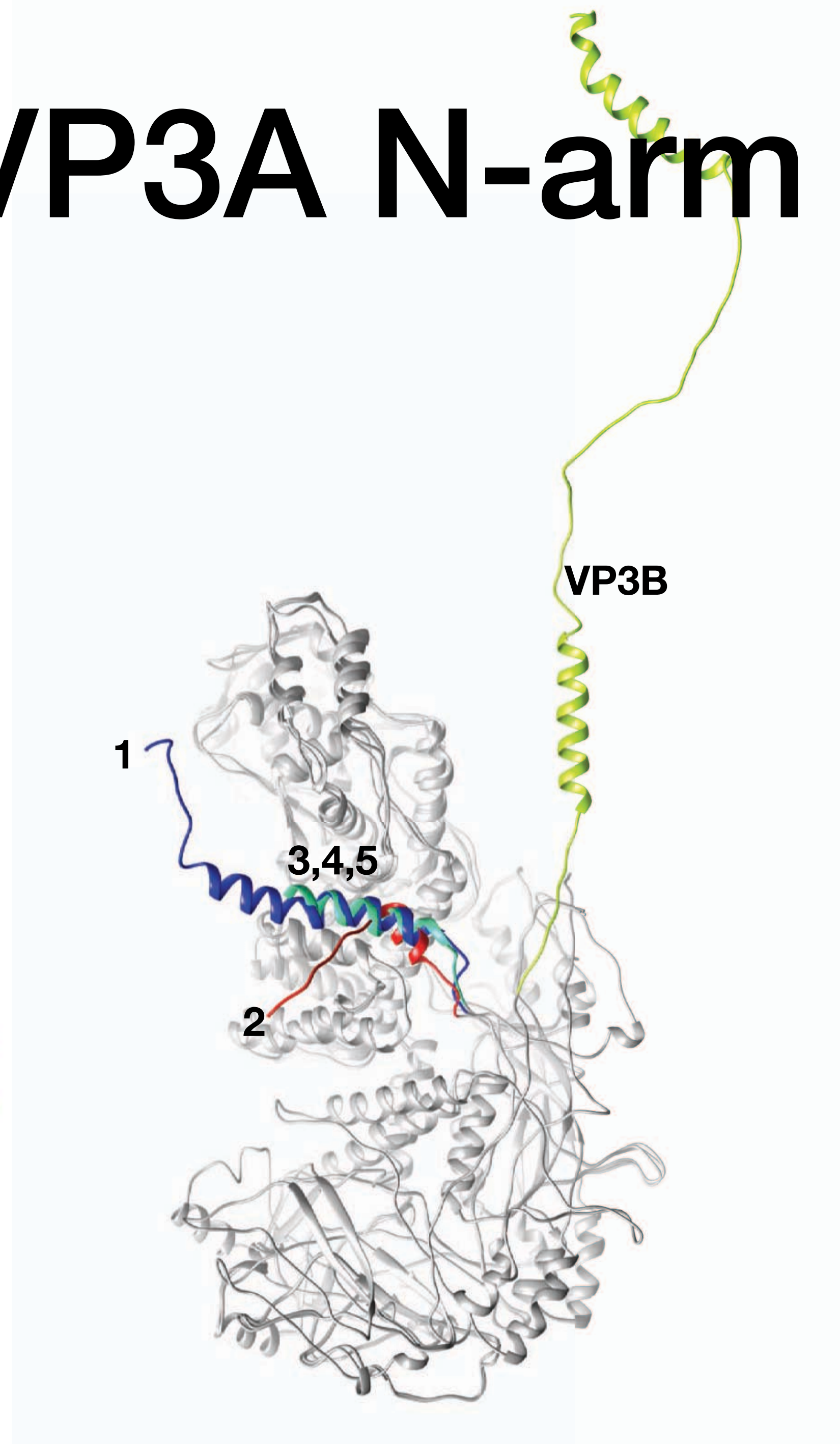
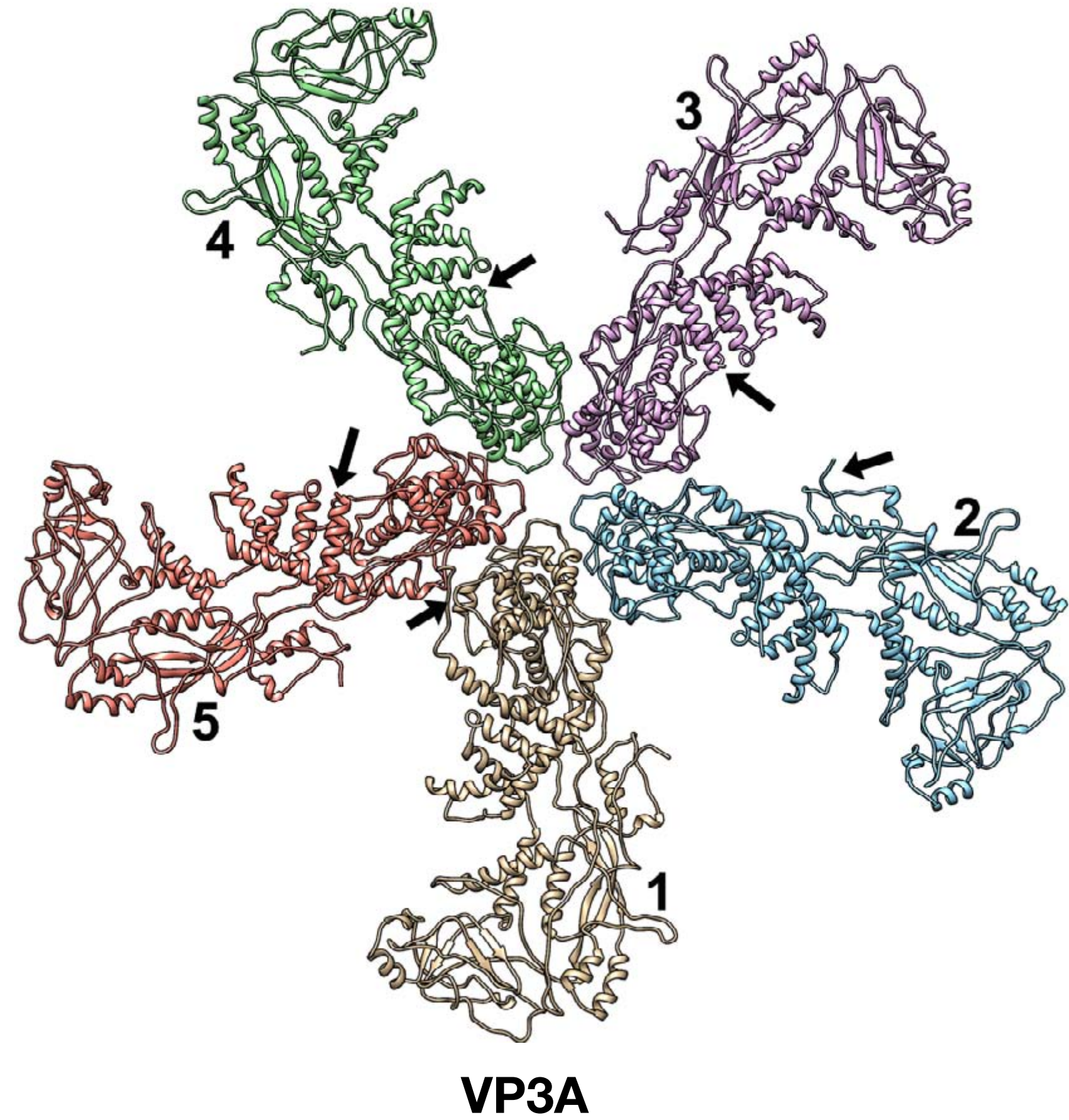
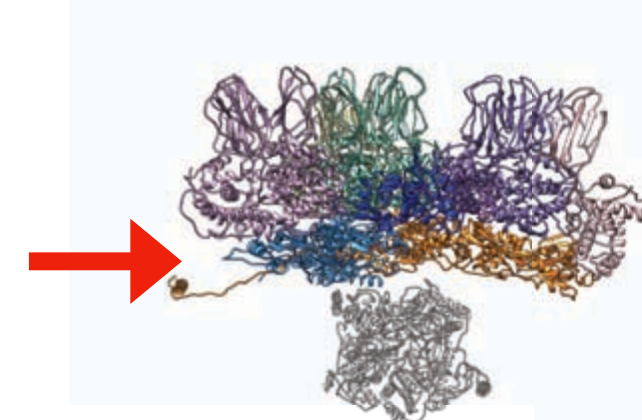
VP3B N-arm



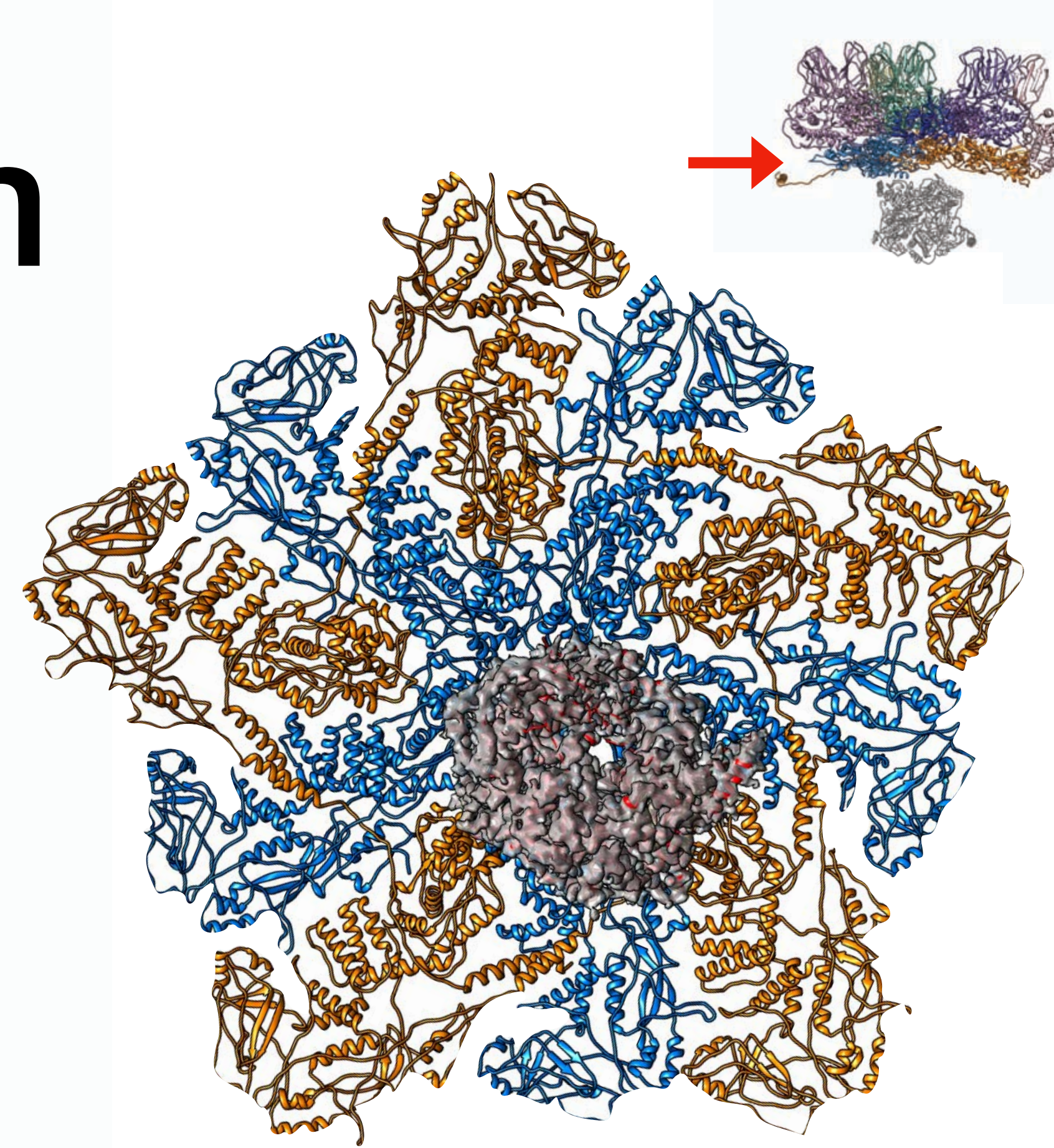
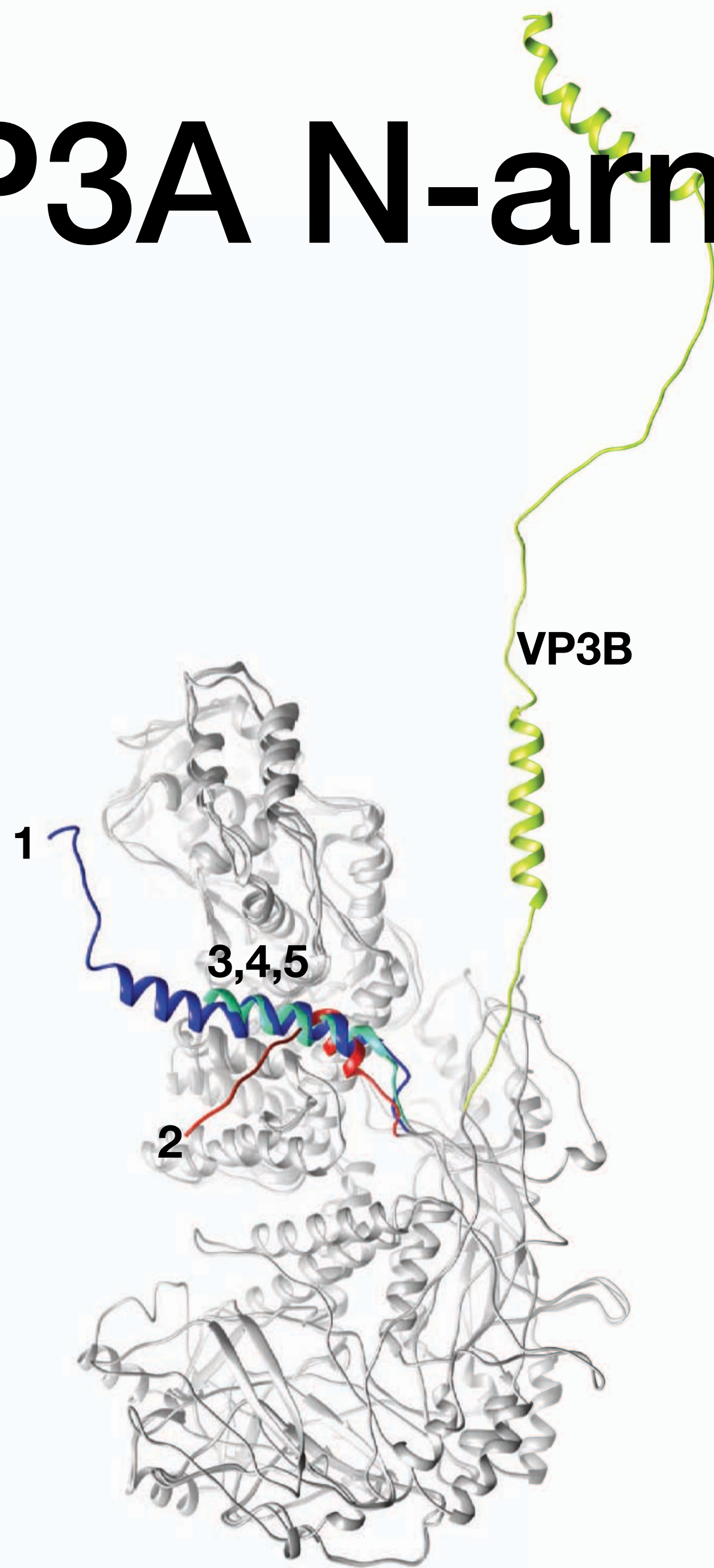
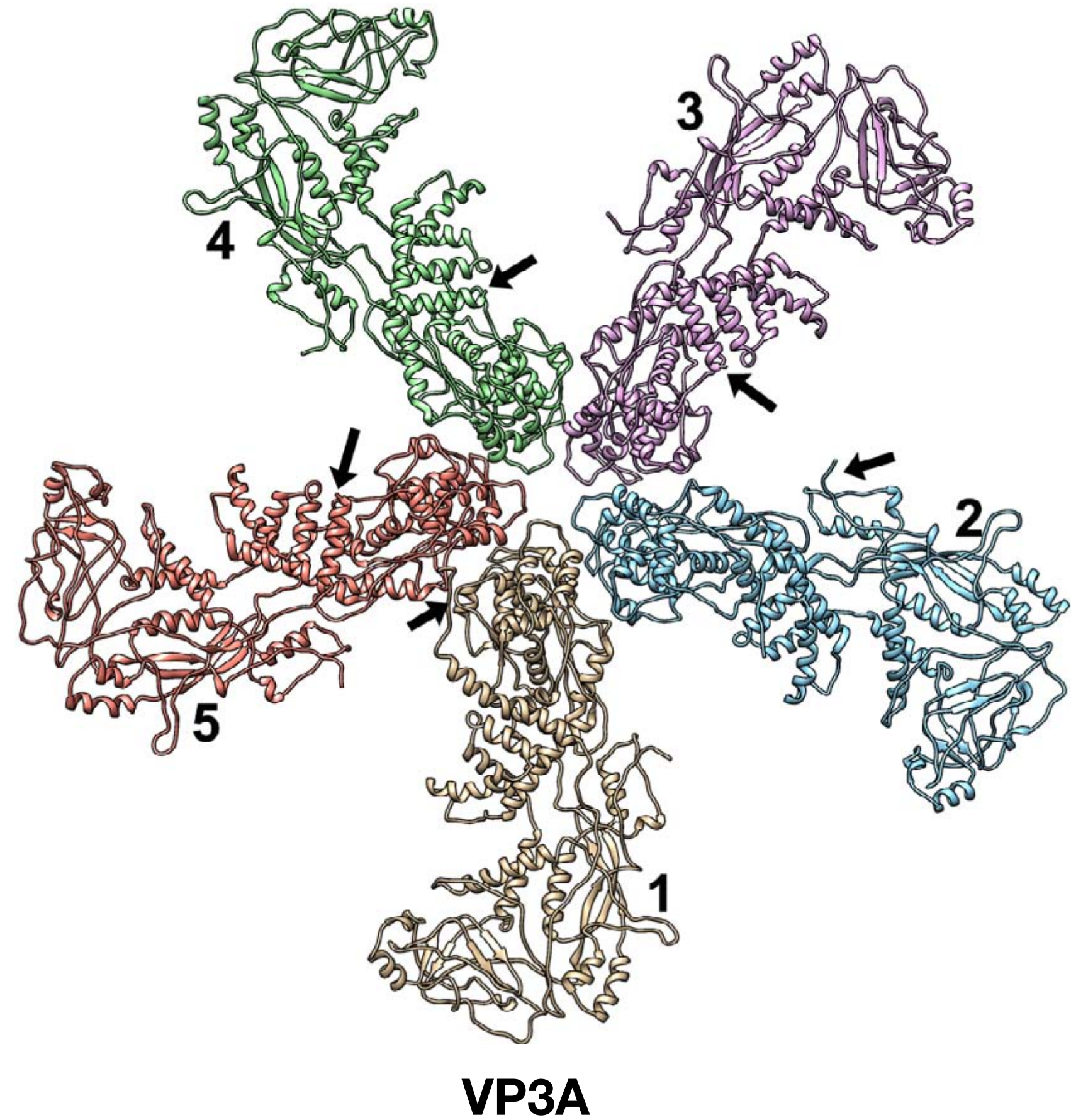
VP3A N-arm



VP3A N-arm

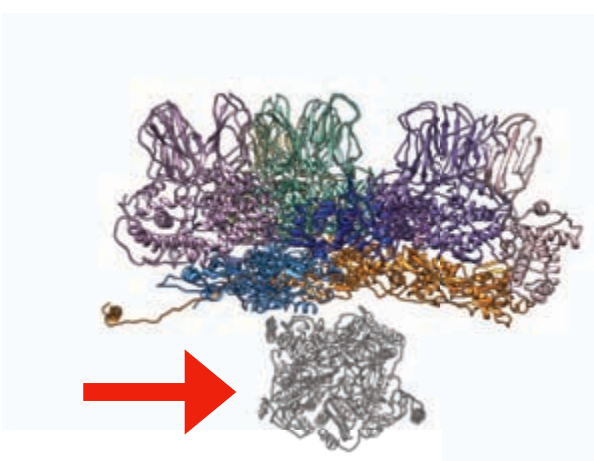


VP3A N-arm

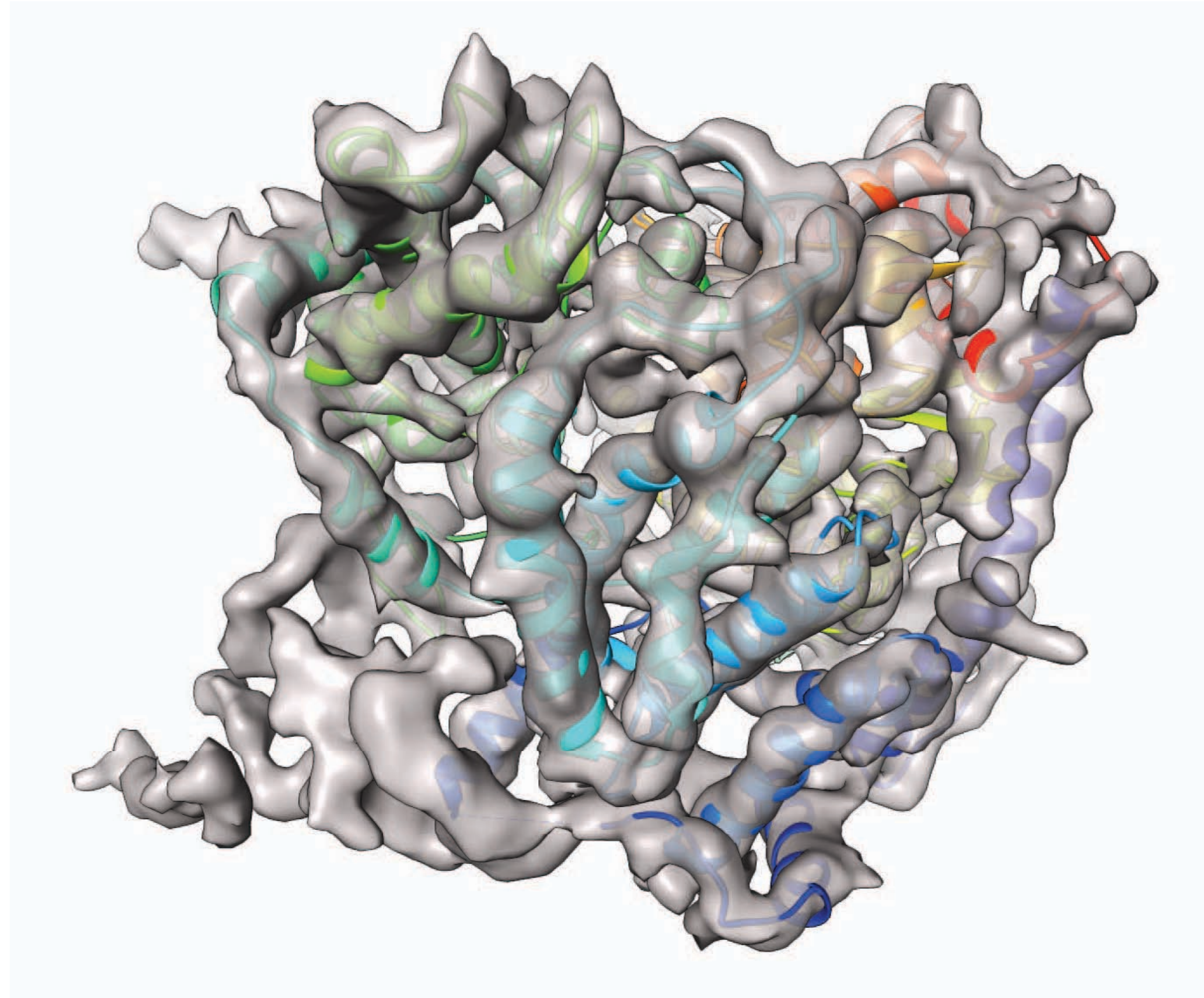


- 2 VP3A N-arm conformations extend to interact with RdRP

RNA dependent RNA Polymerase



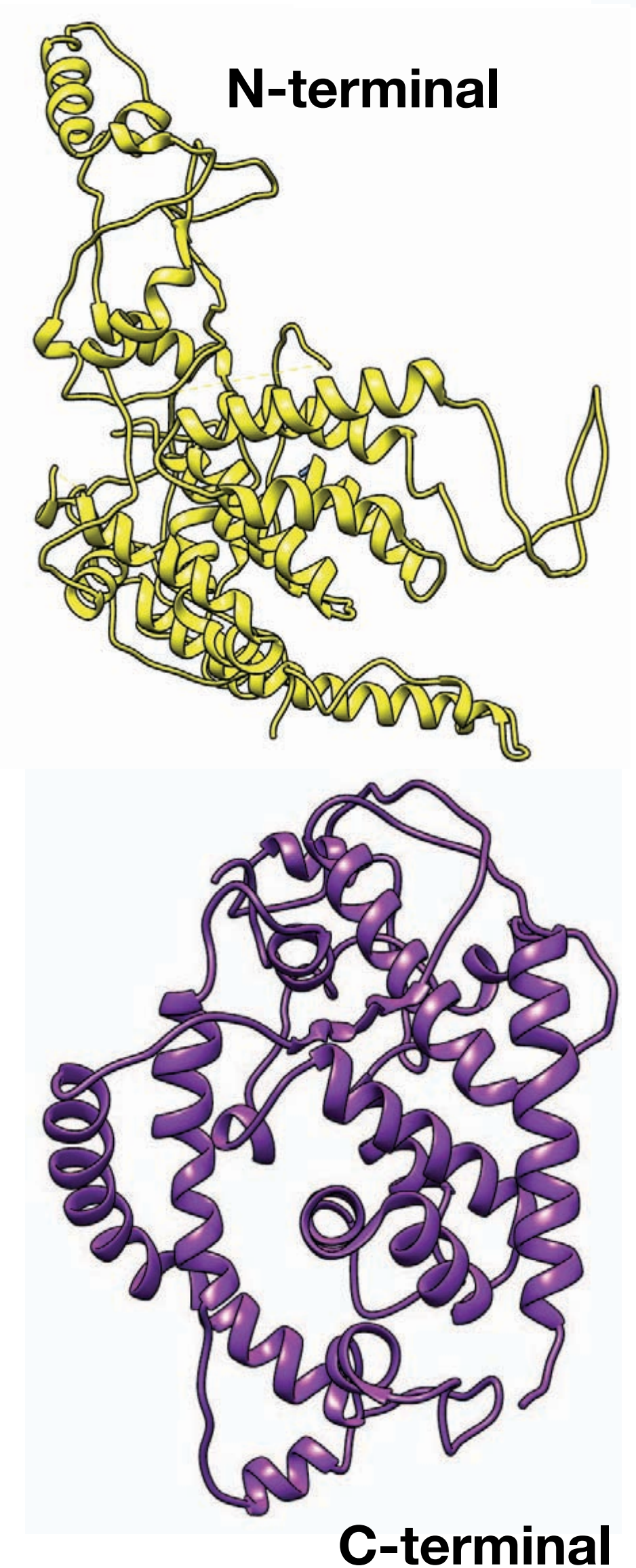
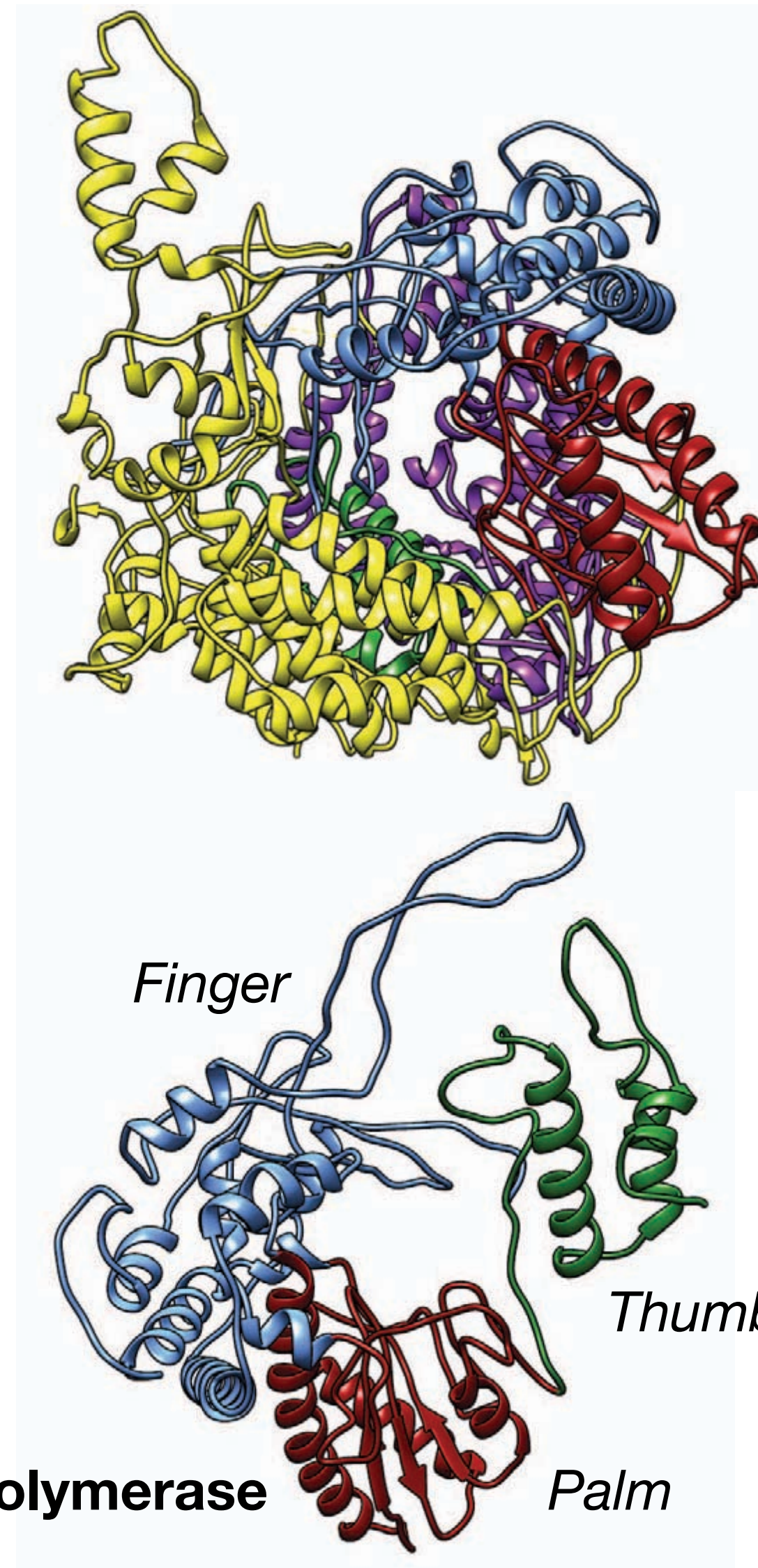
- 1.62Å RMSD between actively transcribing and quiescent MCRV RdRPs
- Priming loop (776-796) differs by ~3.5Å between transcribing and quiescent MCRV
- Cage like structure with 3 domains
 - ➡ N-terminal (1-617)
 - ➡ Central polymerase (618-1090)
 - ➡ C-terminal bracelet (1091-1422)
 - ➡ Core catalytic residues: D819, D824, D947, D948



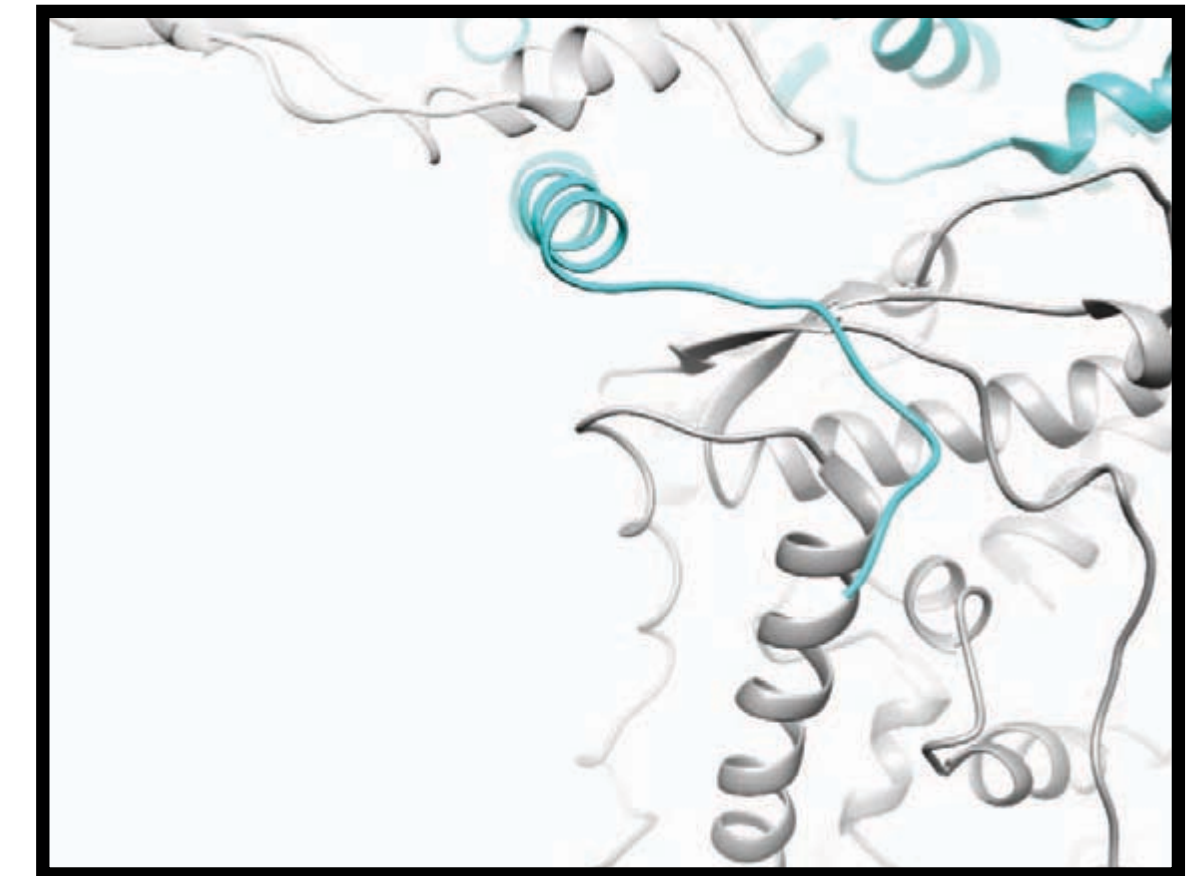
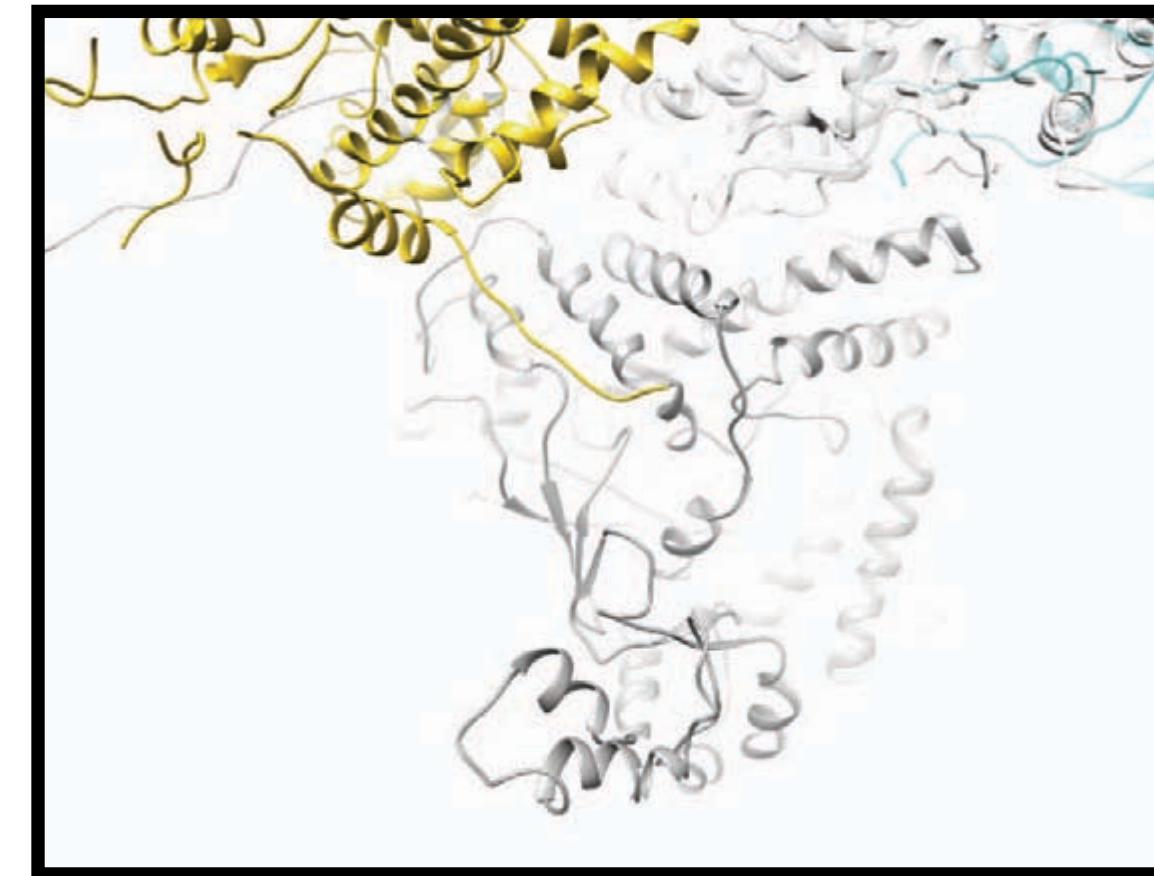
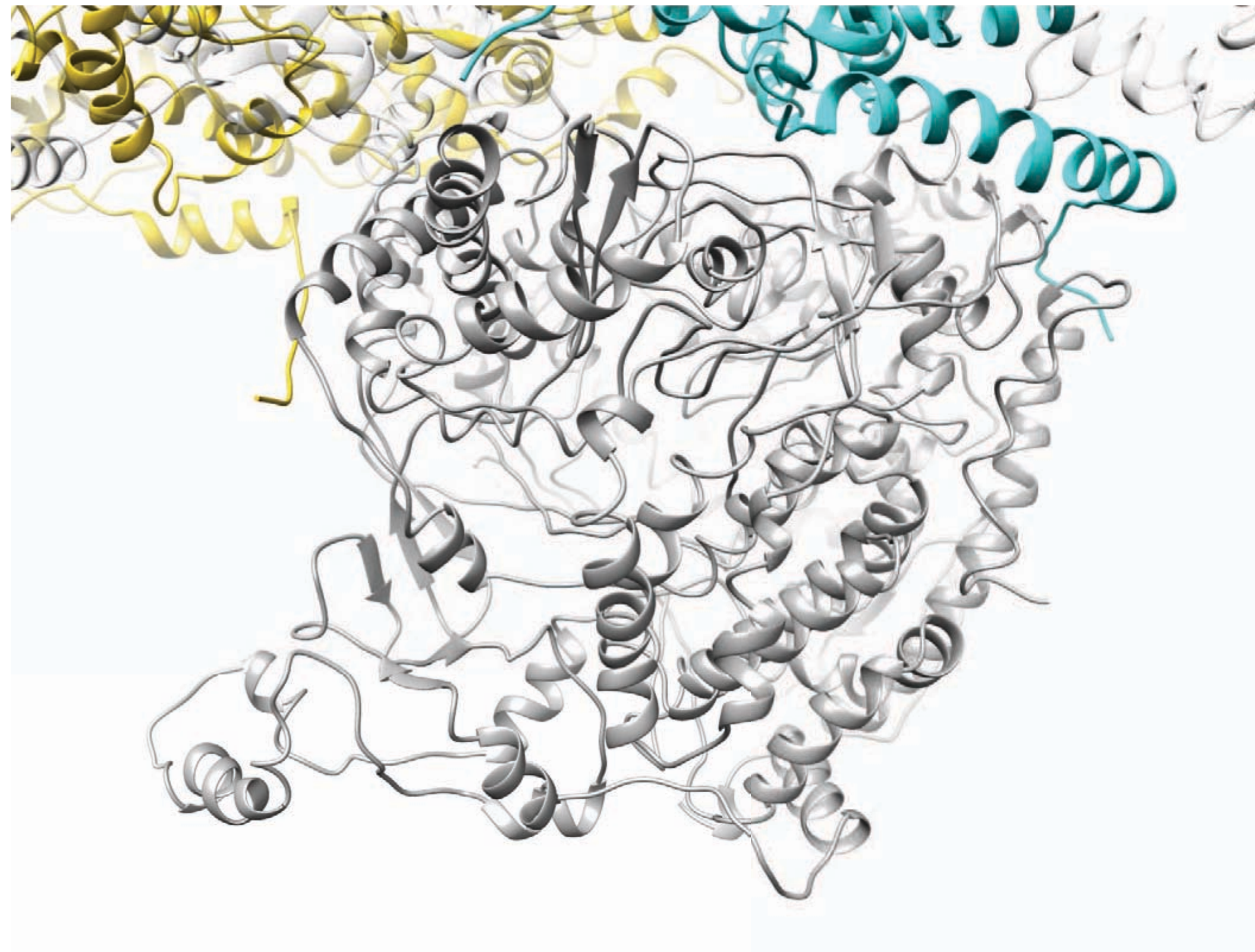
RNA dependent RNA Polymerase



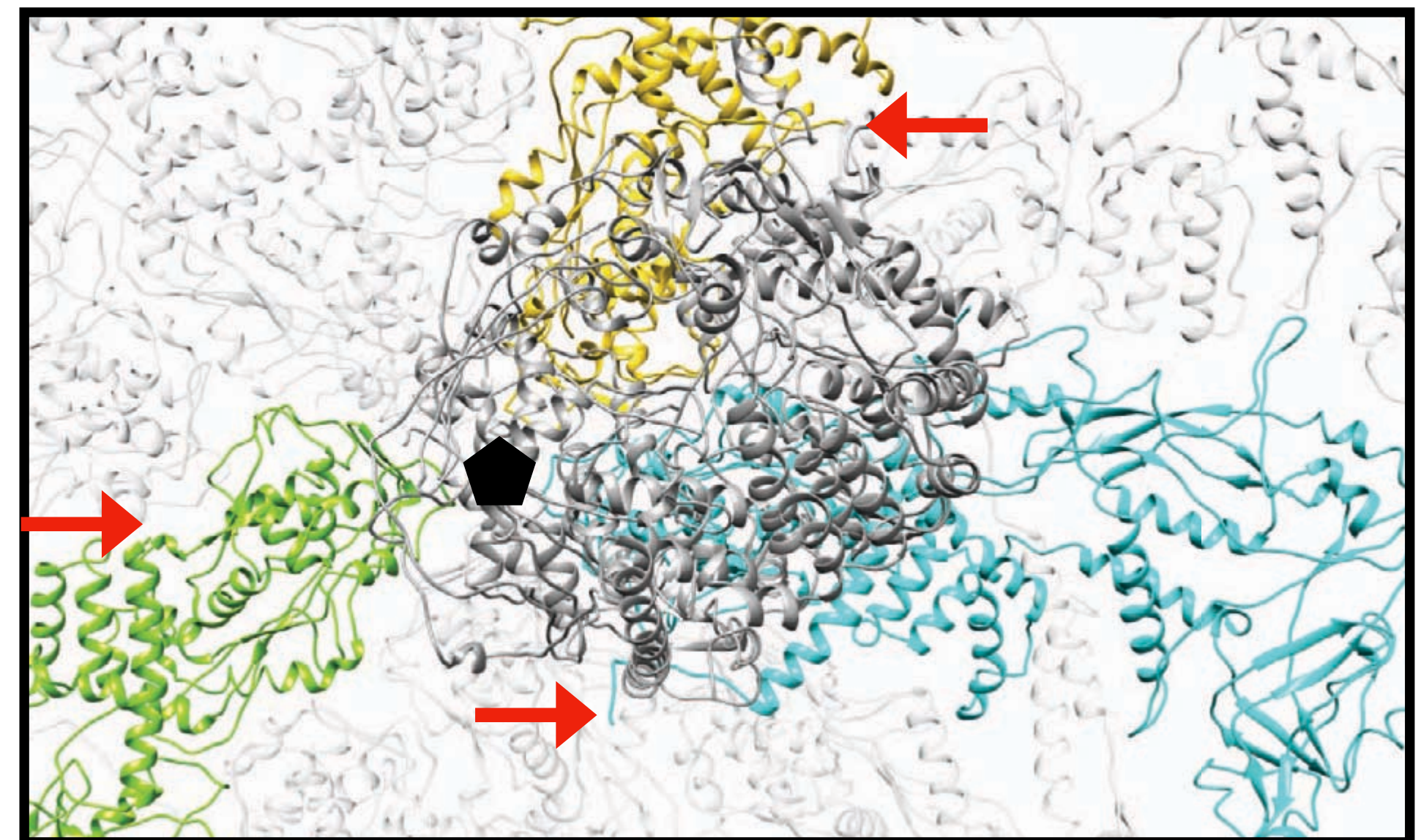
- 1.62Å RMSD between actively transcribing and quiescent MCRV RdRPs
- Priming loop (776-796) differs by ~3.5Å between transcribing and quiescent MCRV
- Cage like structure with 3 domains
 - ➔ N-terminal (1-617)
 - ➔ Central polymerase (618-1090)
 - ➔ C-terminal bracelet (1091-1422)
 - ➔ Core catalytic residues: D819, D824, D947, D948



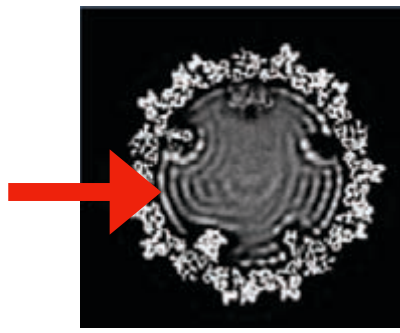
RdRP and the VP3A N-Arms



VP3A residues 40-48 form an inter-molecular anti-parallel beta sheet with residues 1220-1226 and 1390-1395 of the RdRP



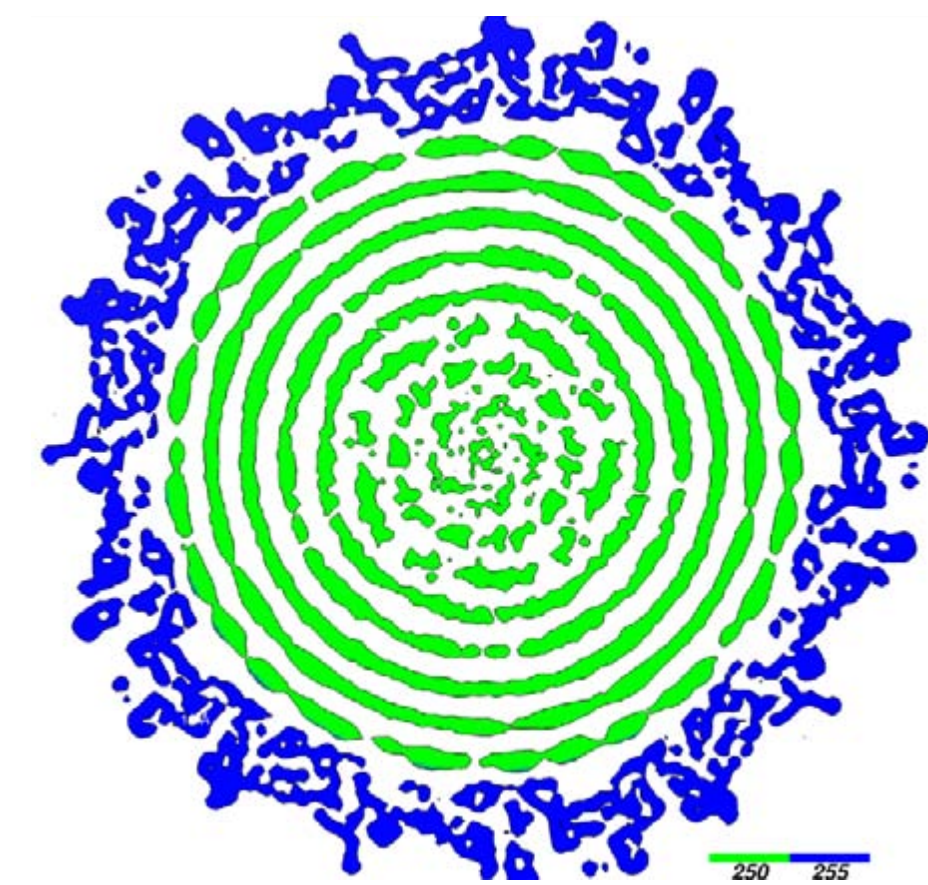
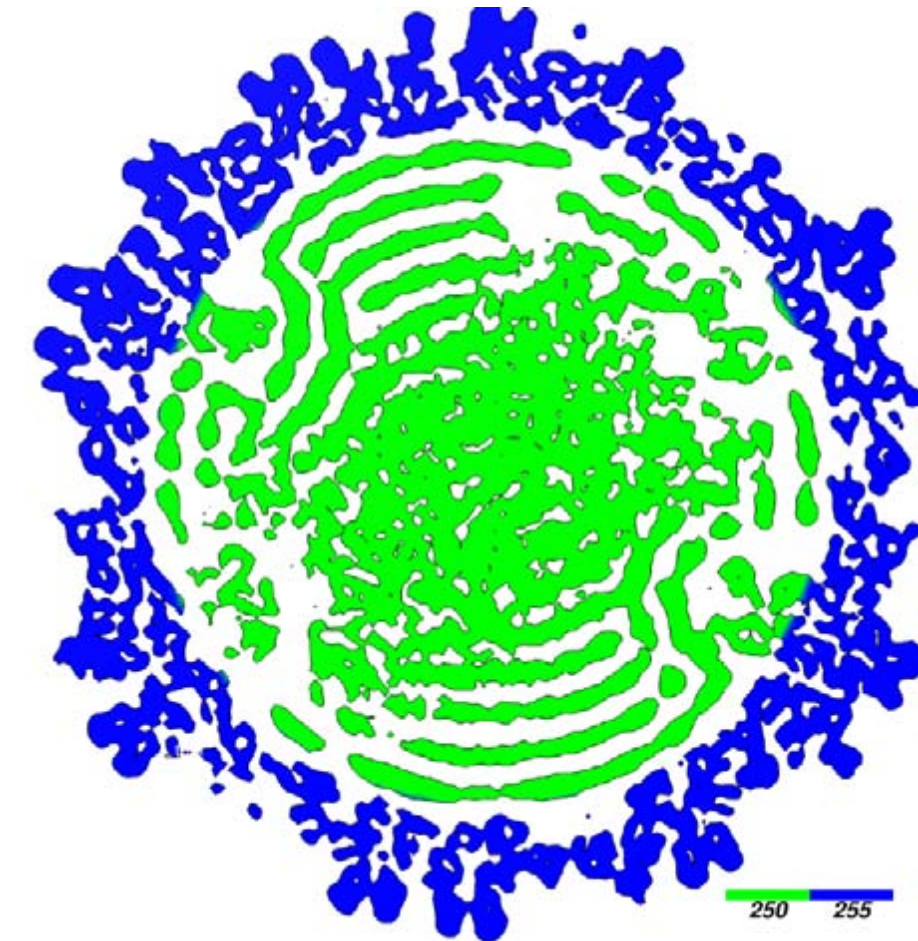
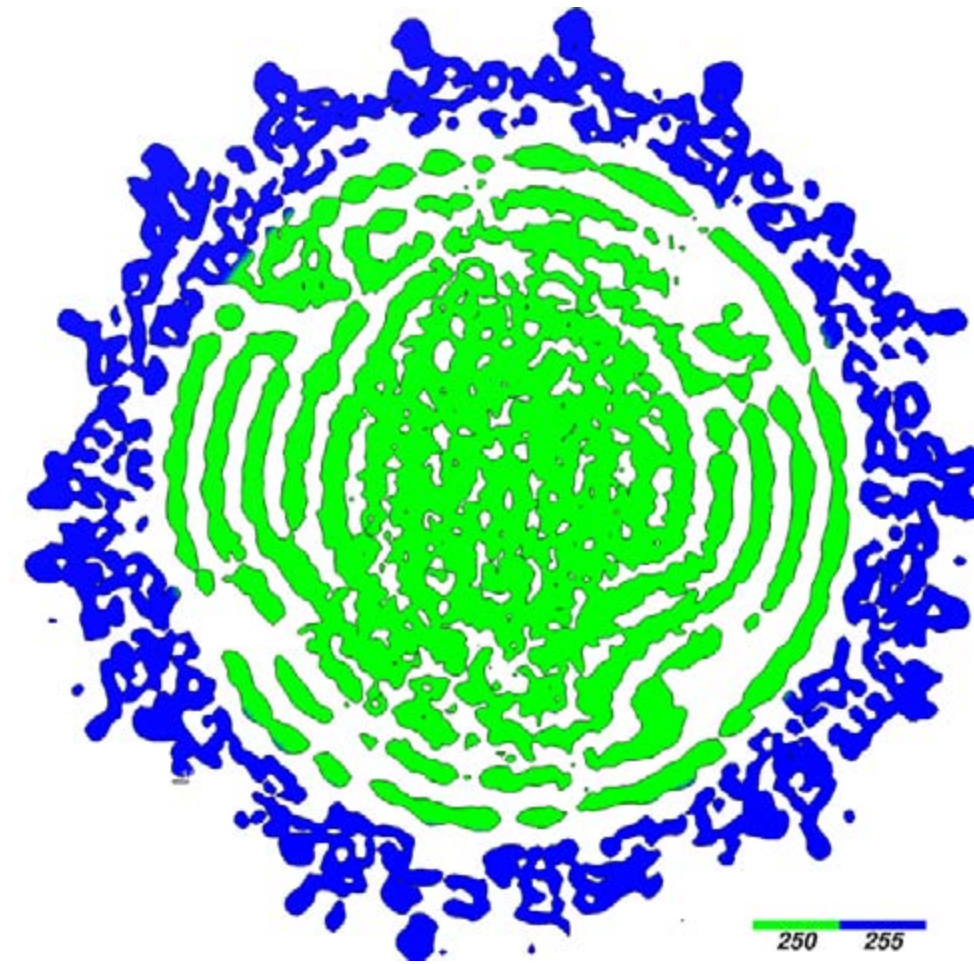
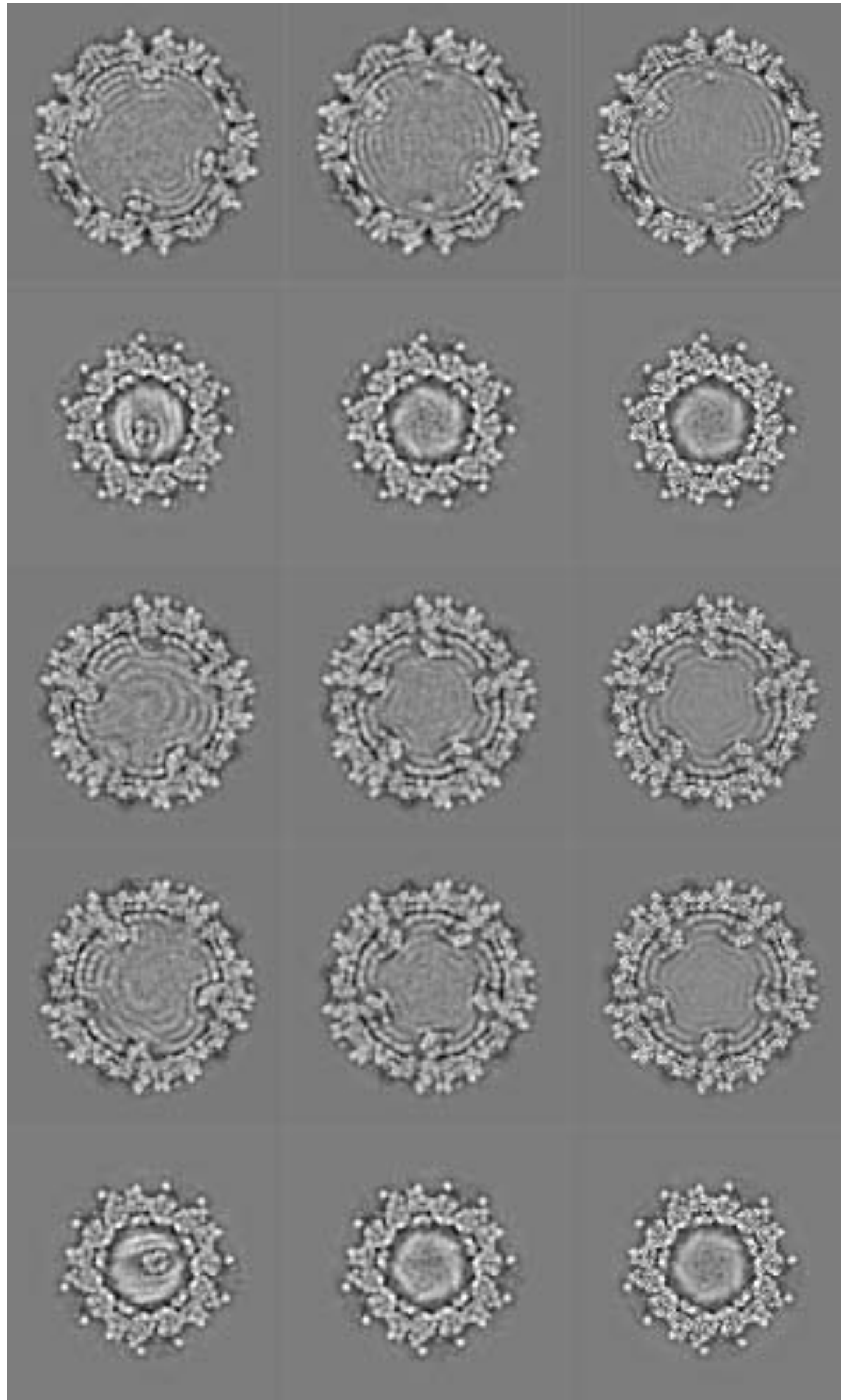
RNA in MCRV



C1

C1 - d5 view

D5



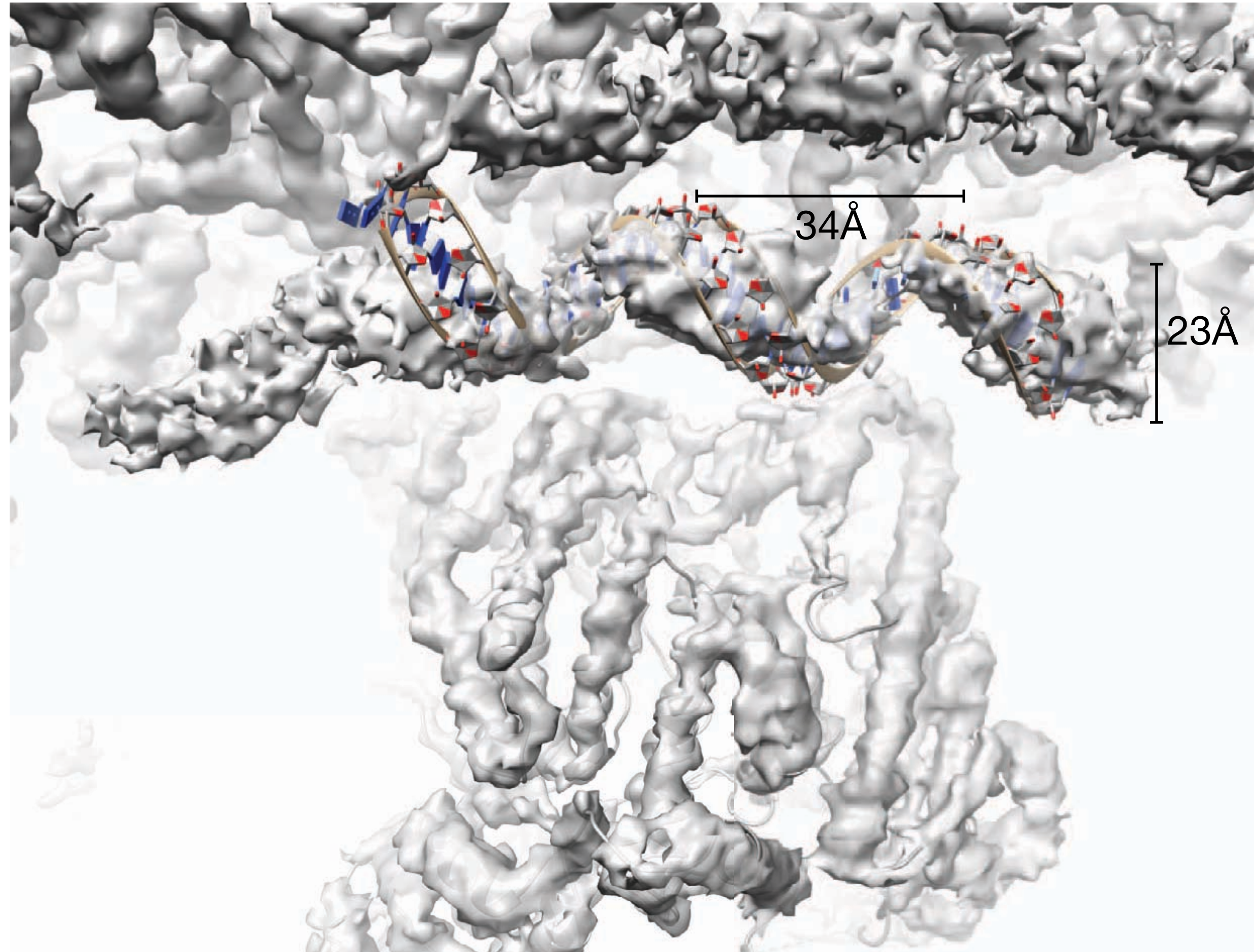
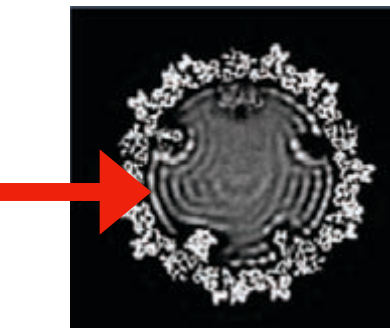
1st layer



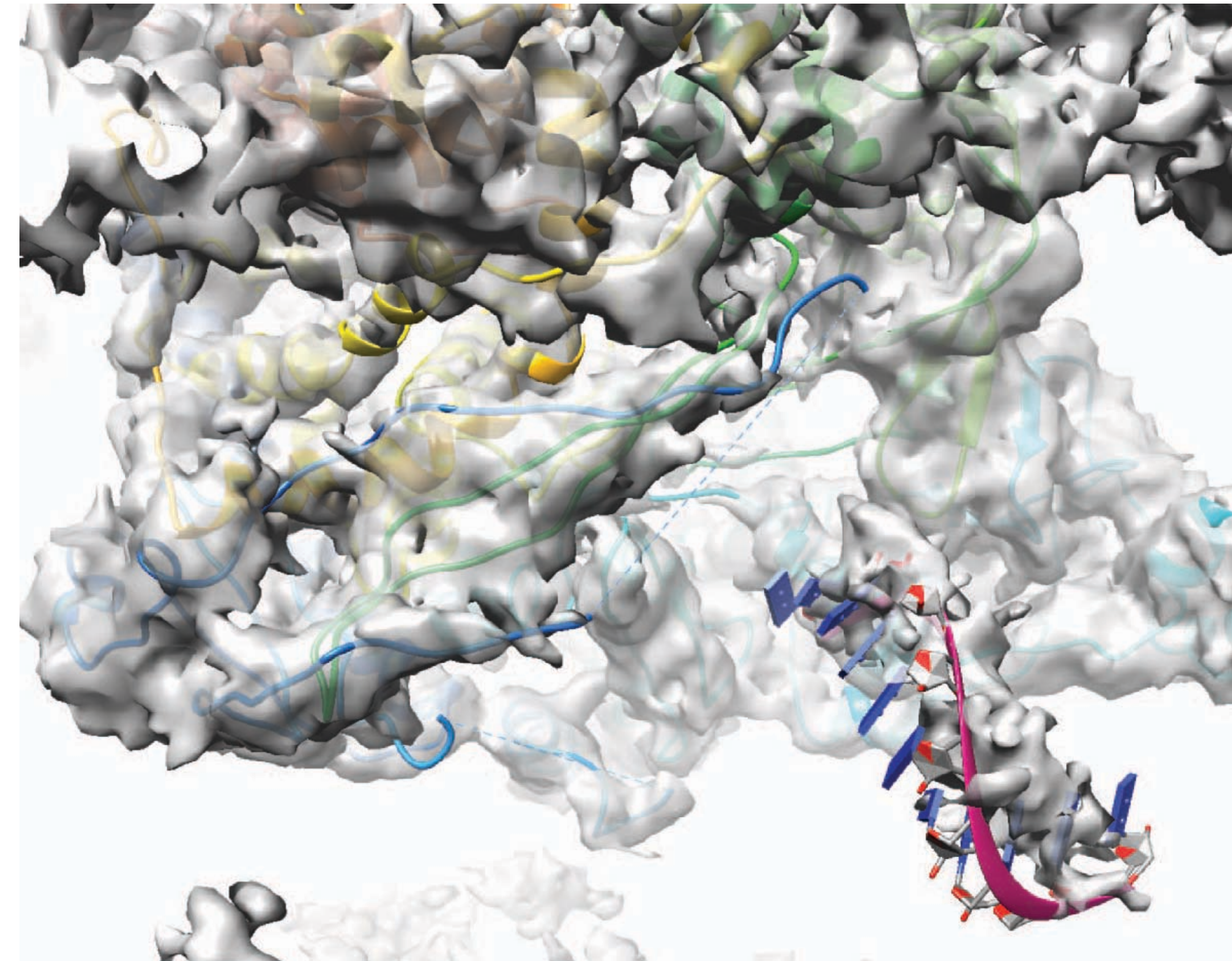
2nd layer

- In C1 reconstructions, clear d5 symmetry is present
- Only 10 of the 12 5-fold vertices have RdRP
- 7+ “rings” of dsRNA; 32Å spacing between layers

Genomic RNA

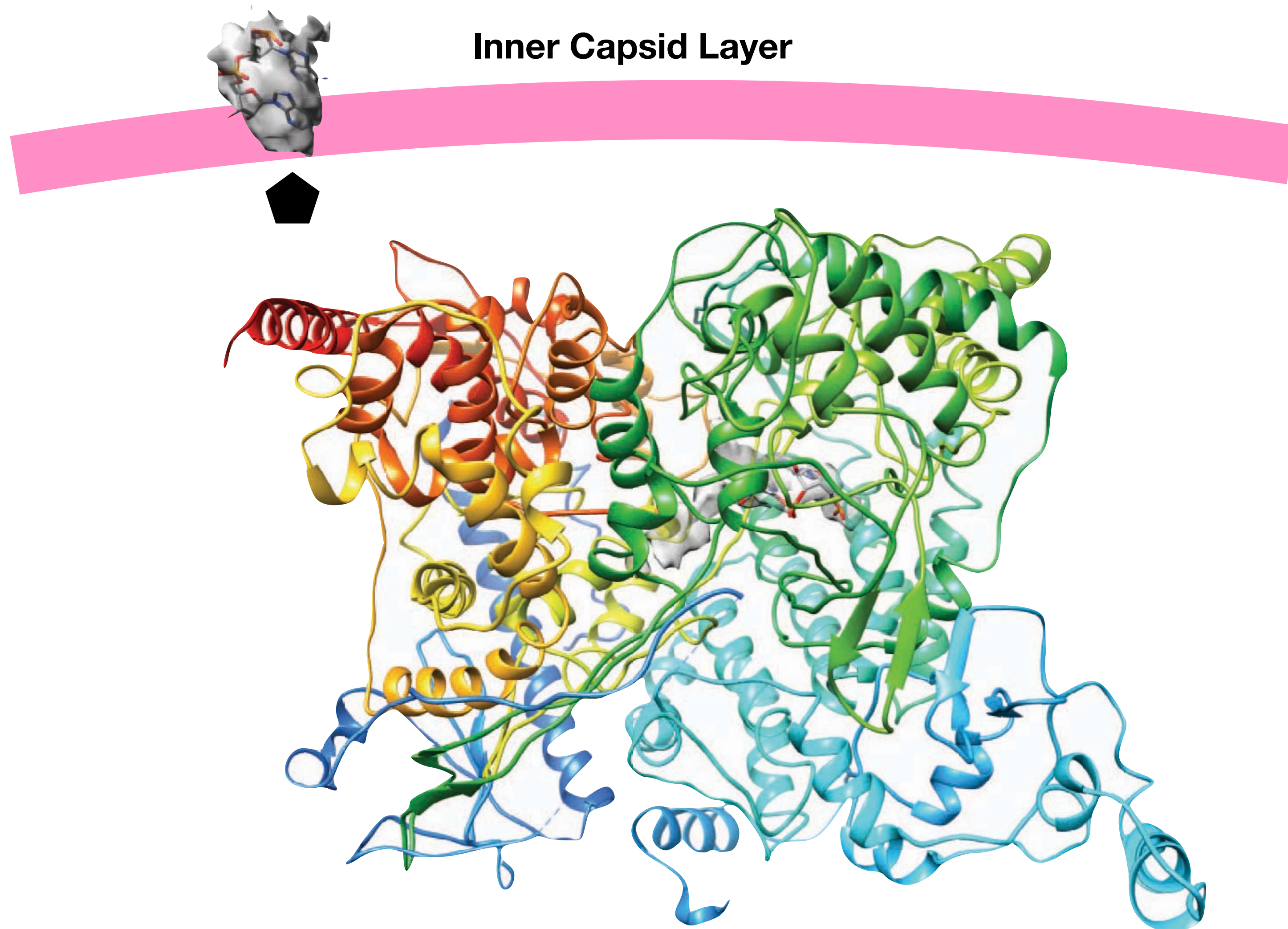
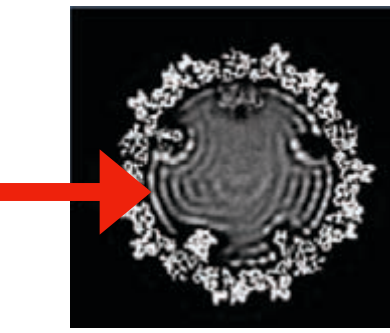


dsRNA



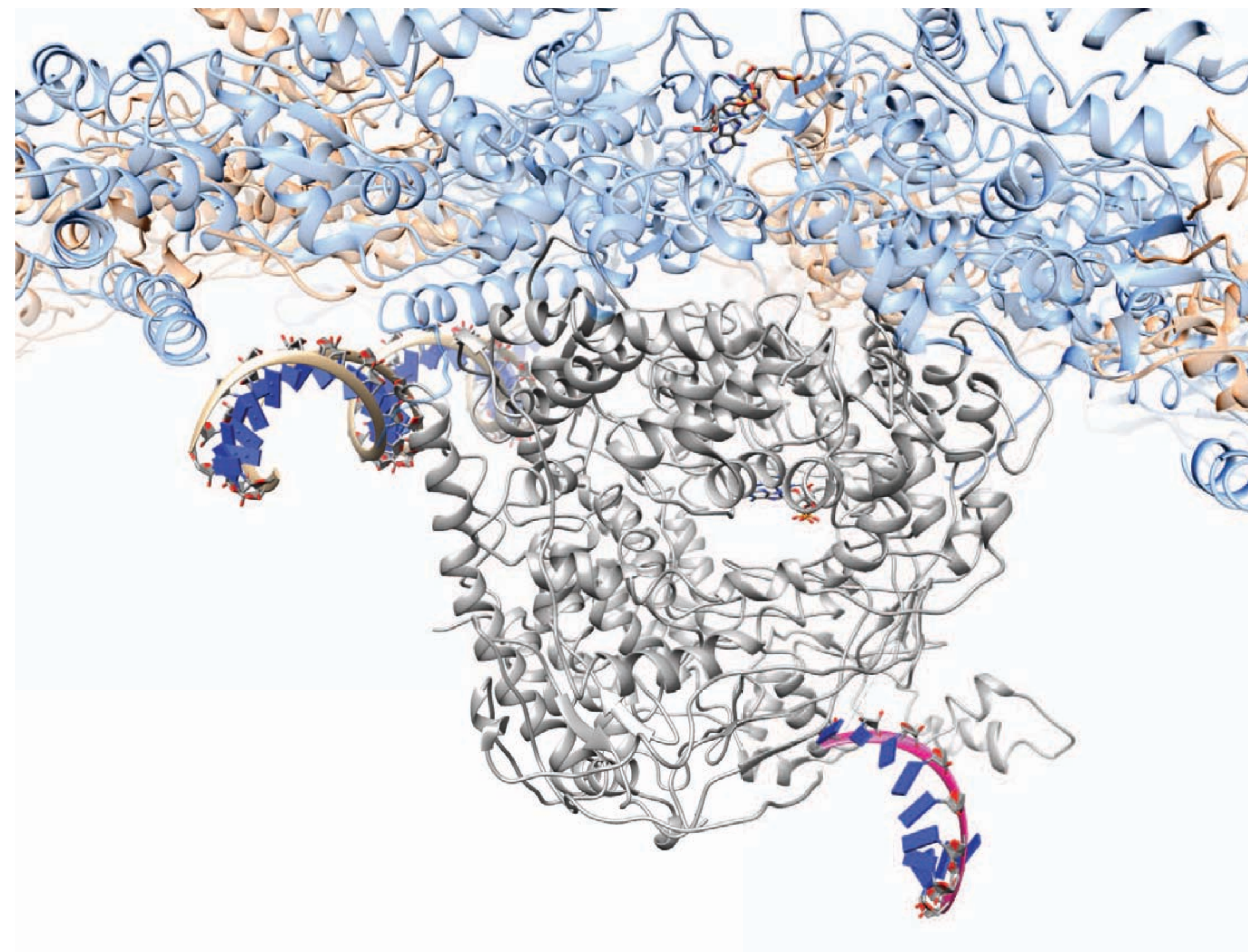
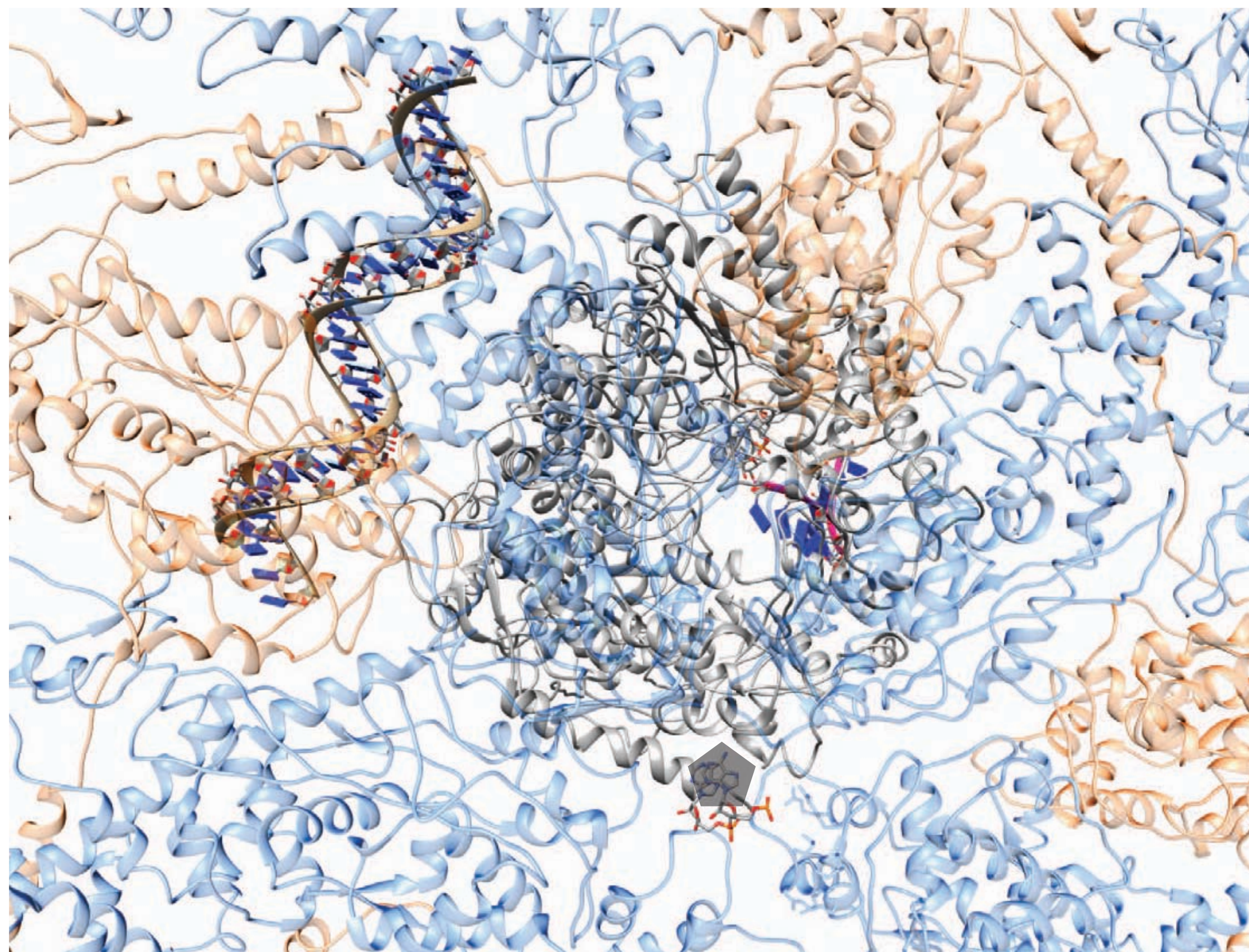
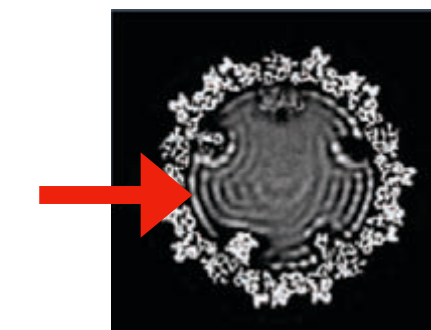
ssRNA

Non-genomic RNA

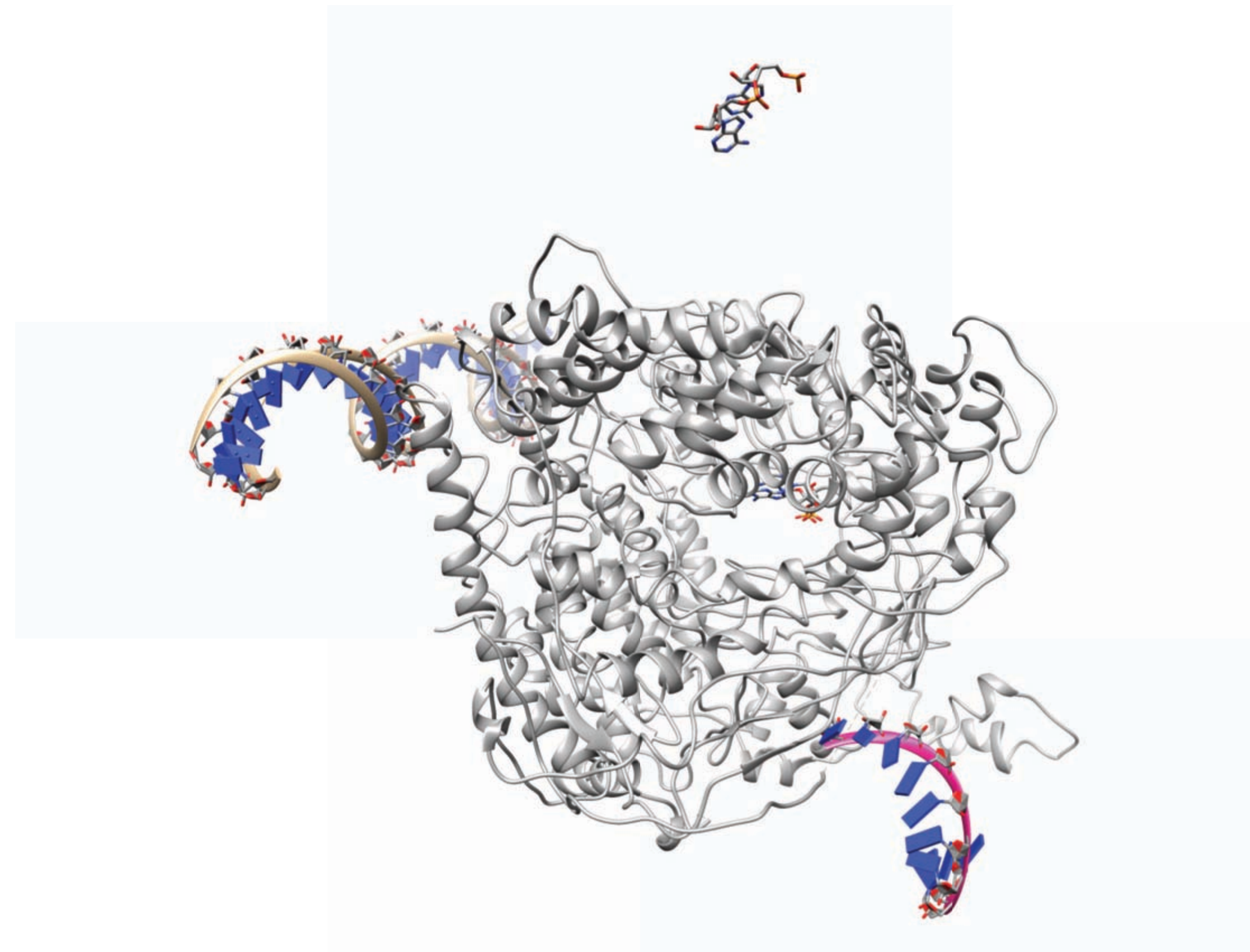
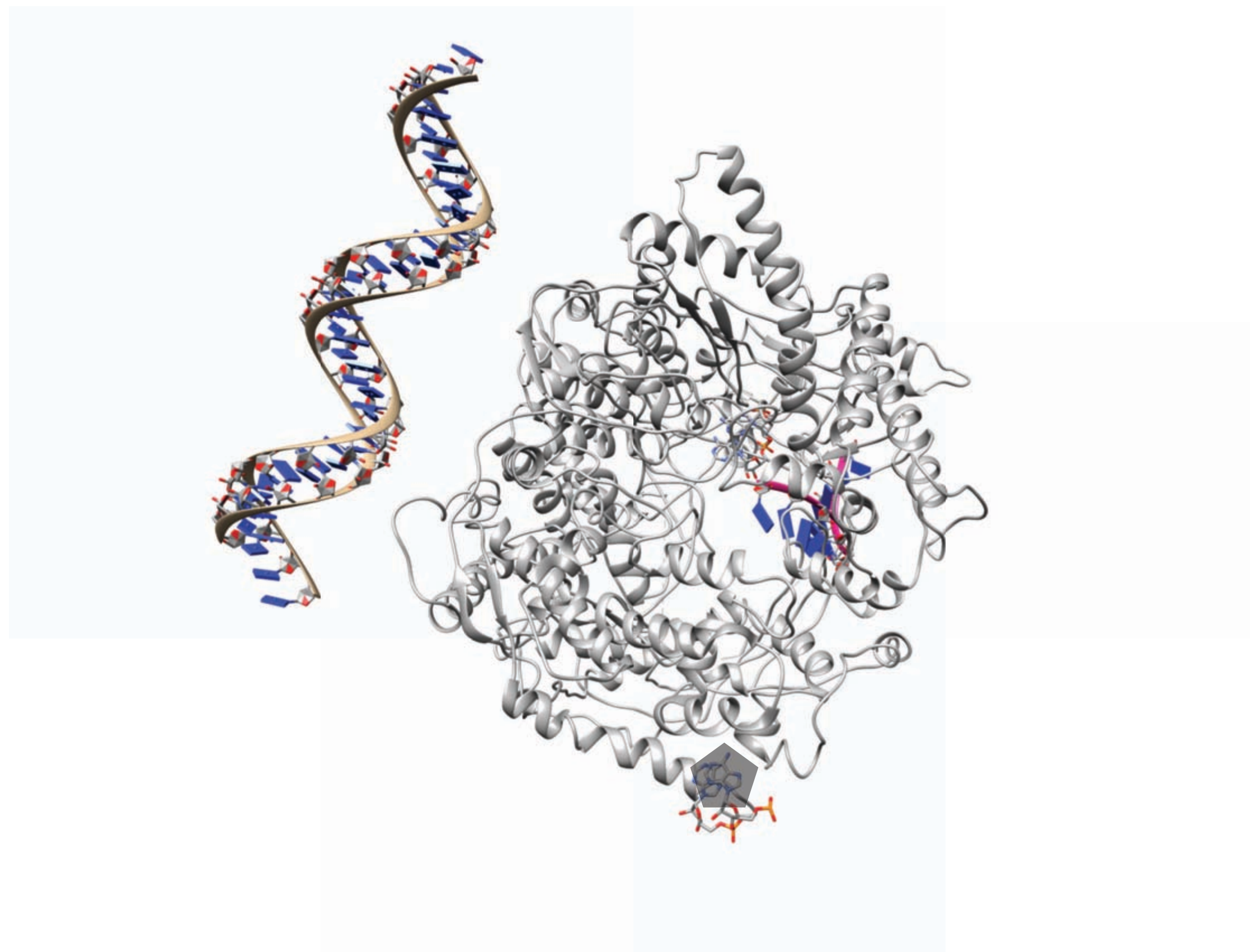
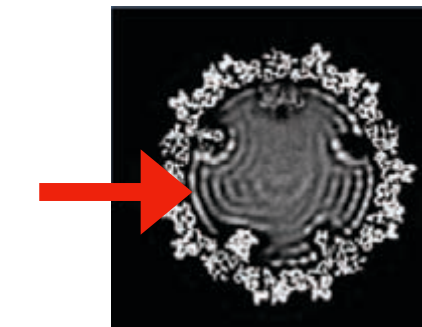


- Two major density differences between active and quiescent MCRV
- Putative nucleotide in catalytic core and priming loop of the central polymerase domain
- Density corresponding to RNA at 5-fold VP3A channel can be seen only in the actively transcribing MCRV.

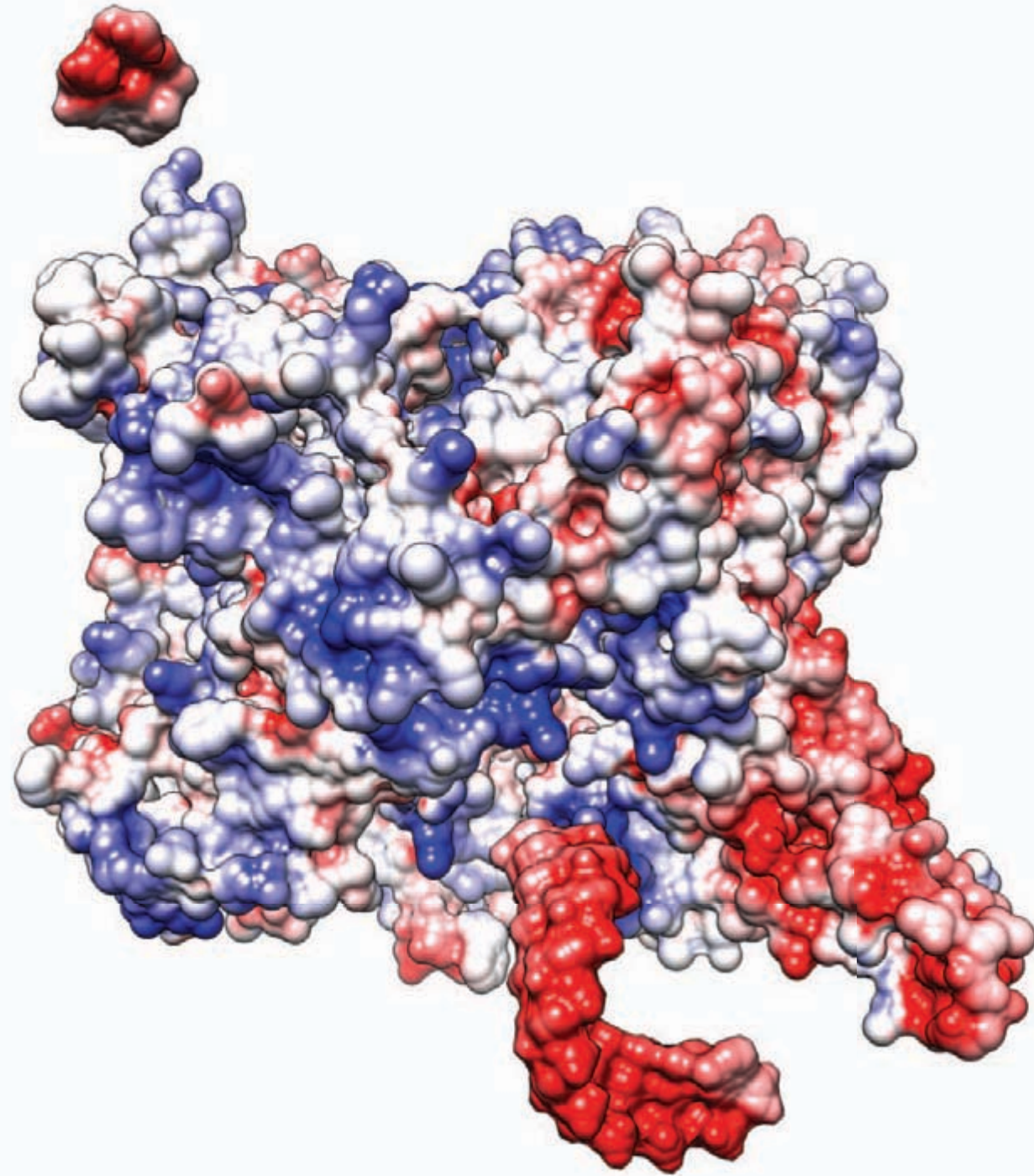
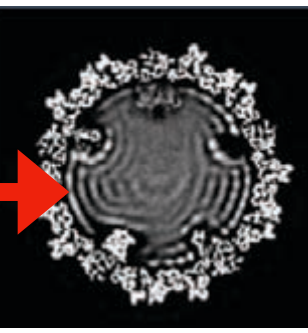
RdRP and RNA



RdRP and RNA

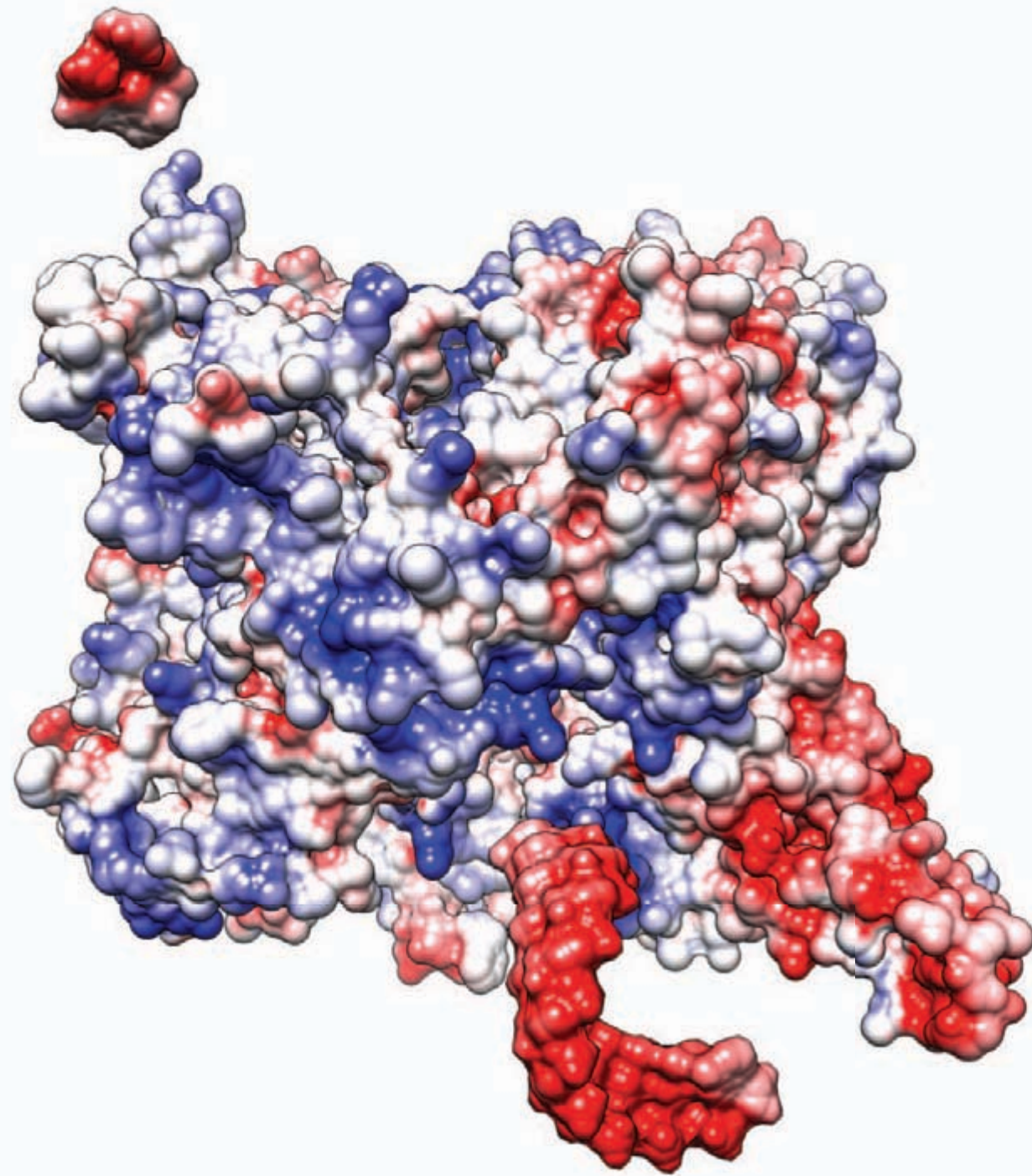
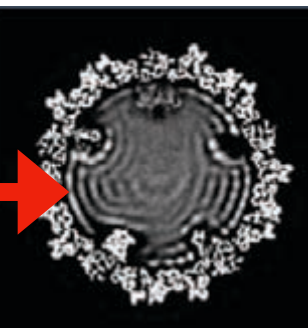


RNA interactions in transcribing MCRV

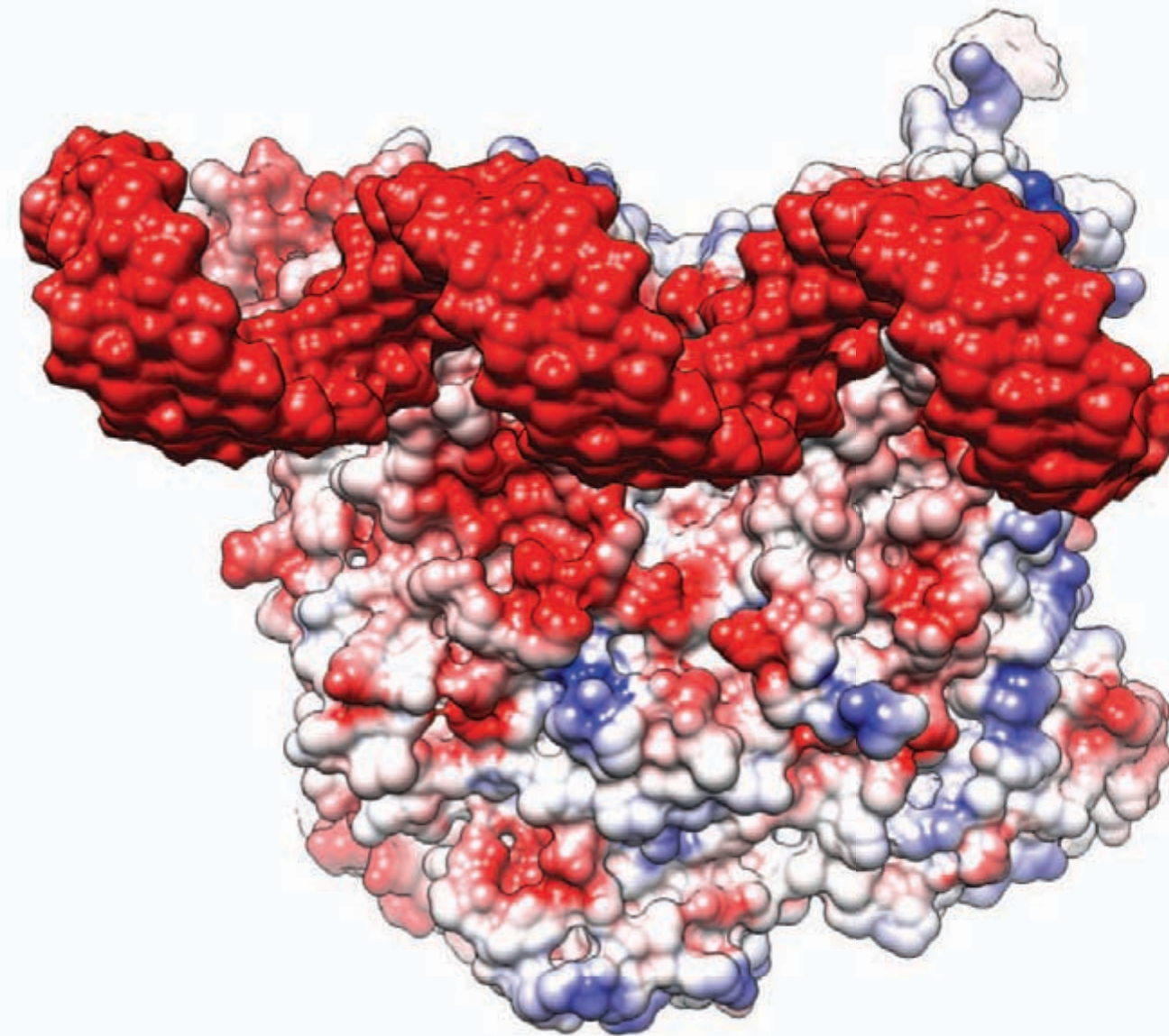


ssRNA (~24 nucleotides) enters the RdRP at an entry site formed by the positively charged N-terminal and negatively charged finger domain.

RNA interactions in transcribing MCRV

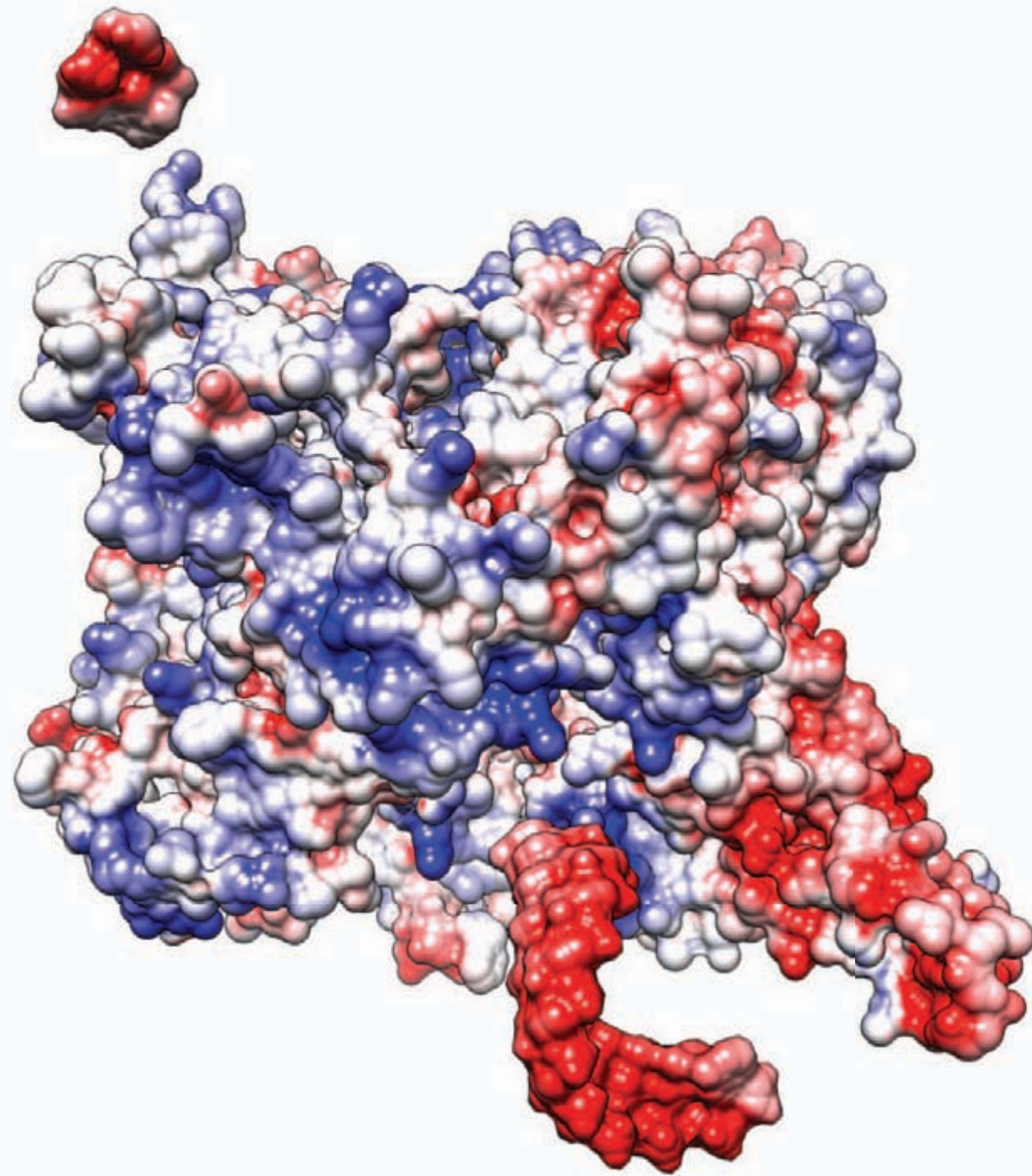
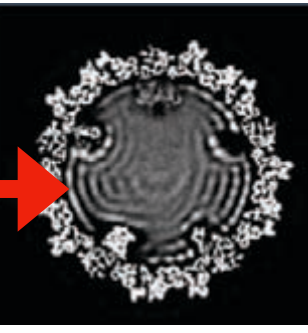


ssRNA (~24 nucleotides) enters the RdRP at an entry site formed by the positively charged N-terminal and negatively charged finger domain.

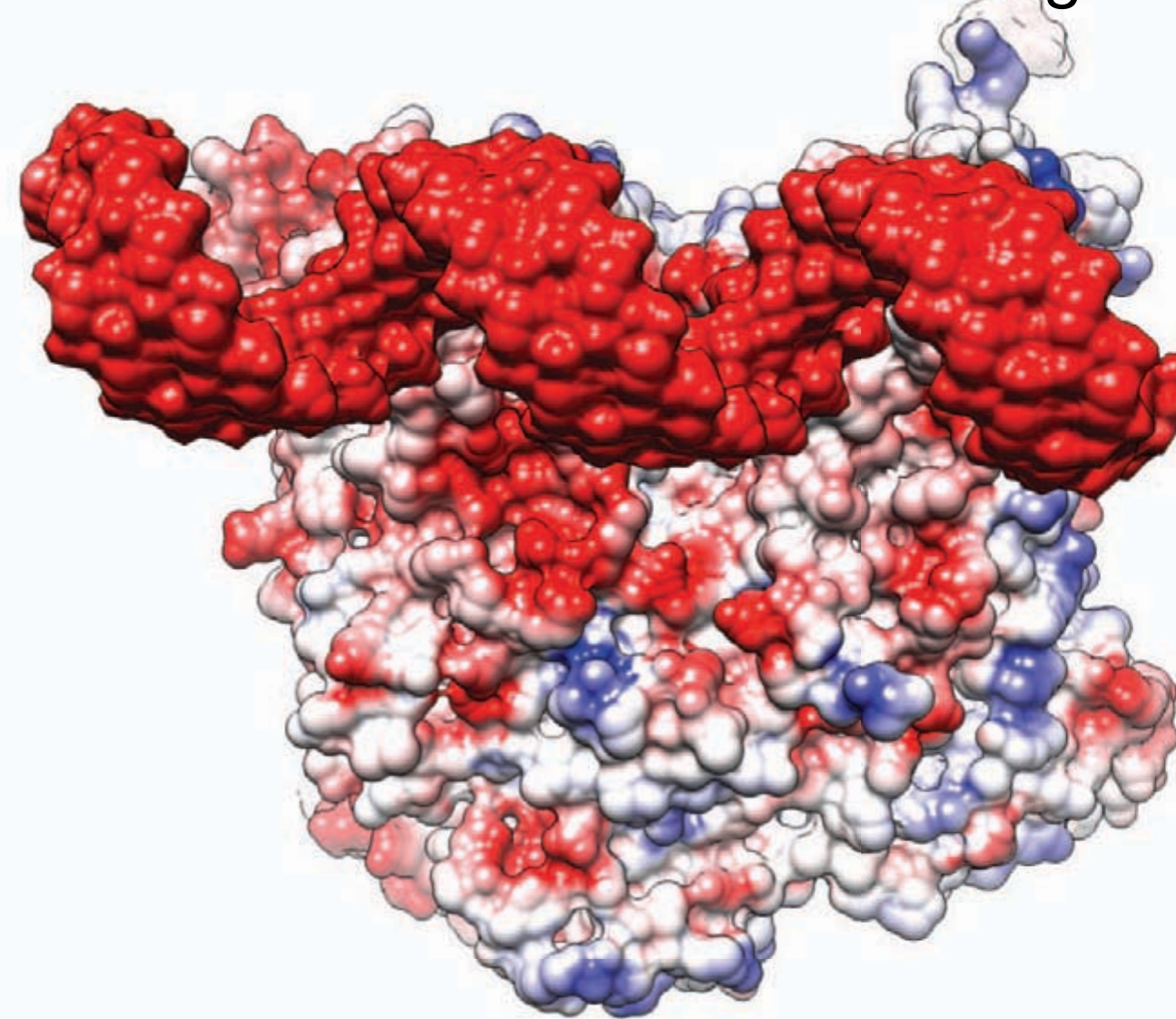


dsRNA runs along the opposite side of the RdRP, which is largely negatively charged

RNA interactions in transcribing MCRV

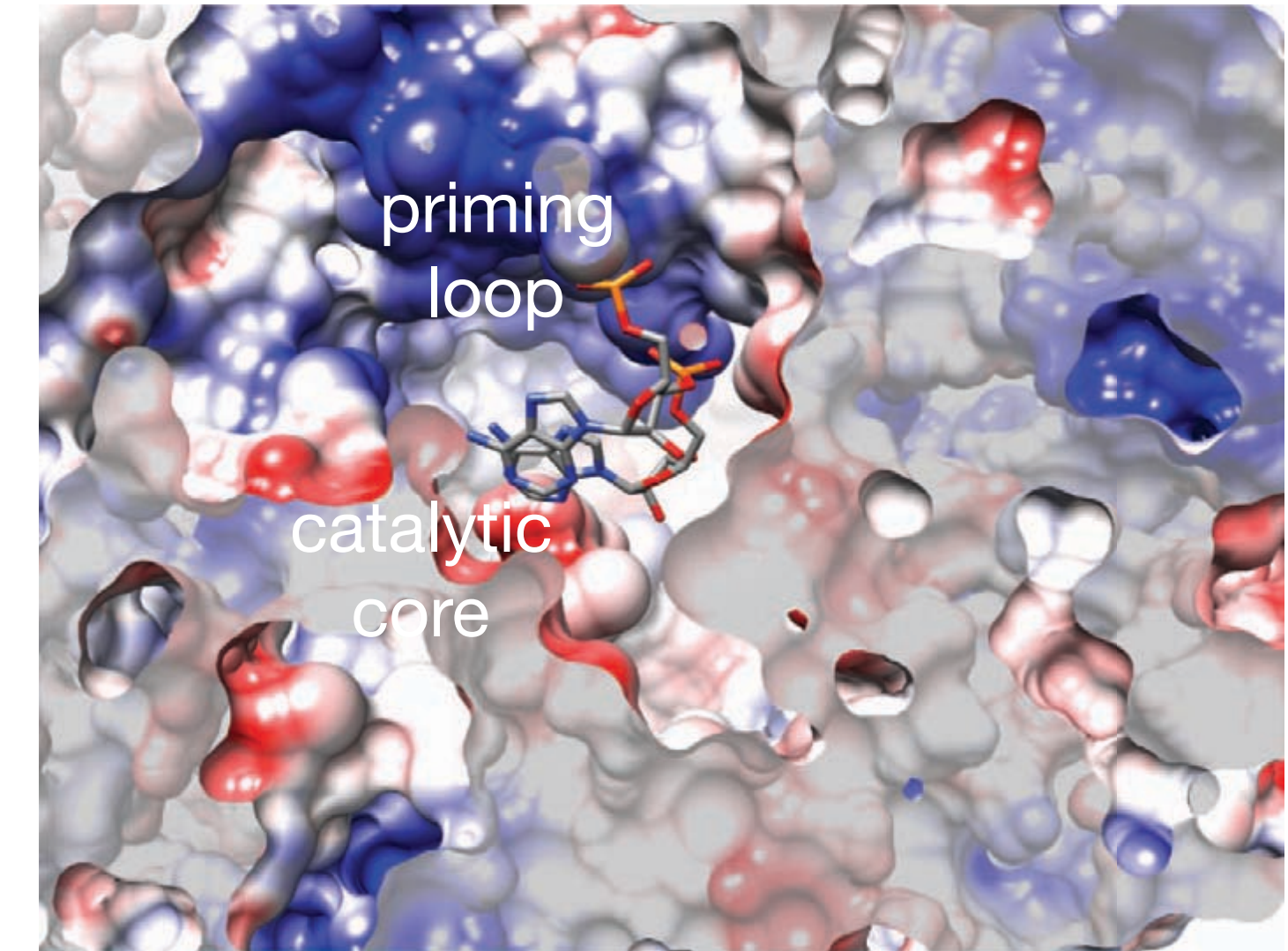


ssRNA (~24 nucleotides) enters the RdRP at an entry site formed by the positively charged N-terminal and negatively charged finger domain.

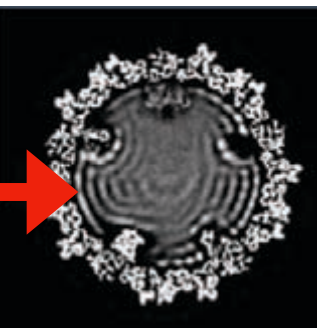


dsRNA runs along the opposite side of the RdRP, which is largely negatively charged

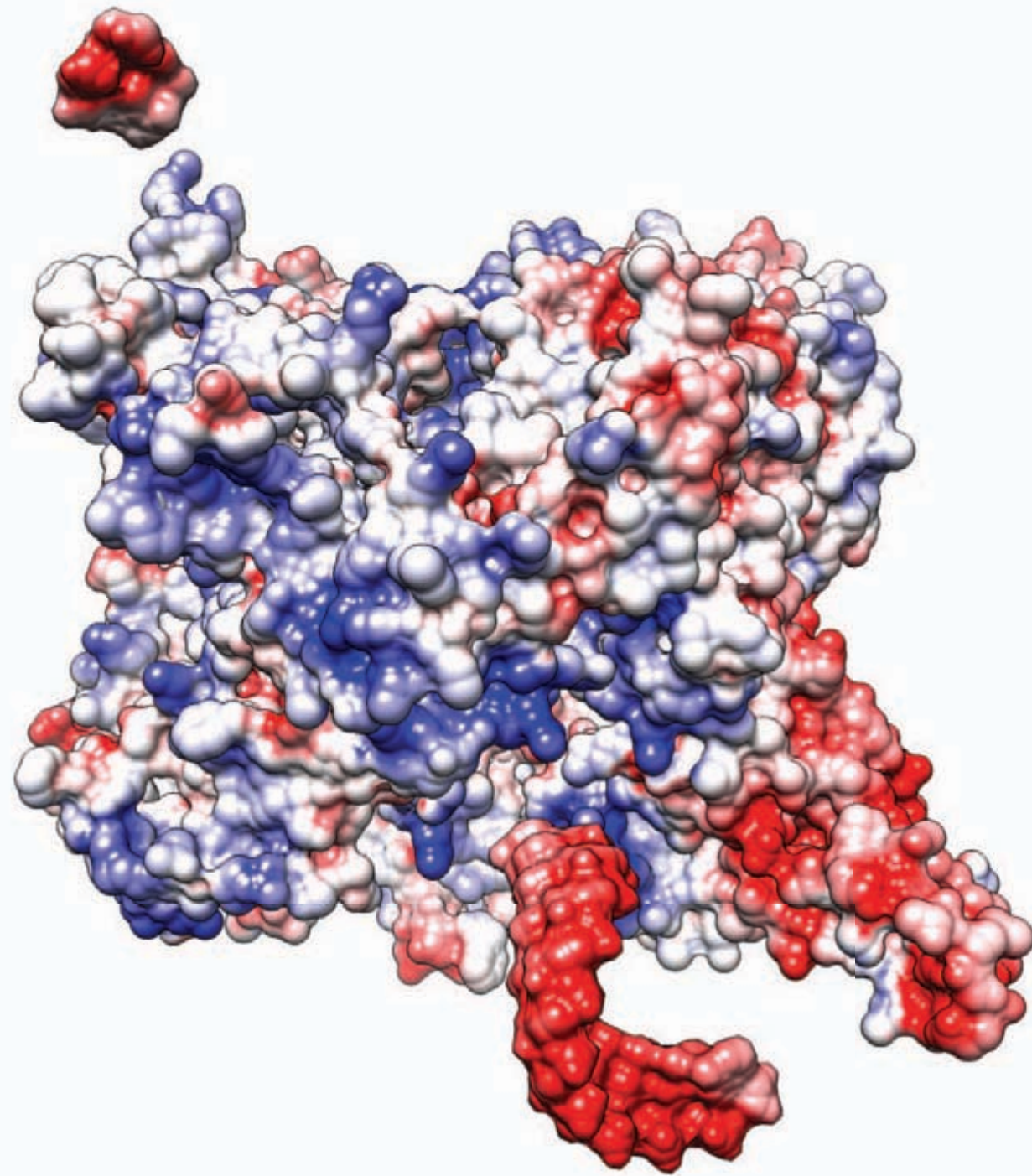
Nucleotide in the catalytic core interacts with the negatively charged catalytic core and the positively charged priming loop



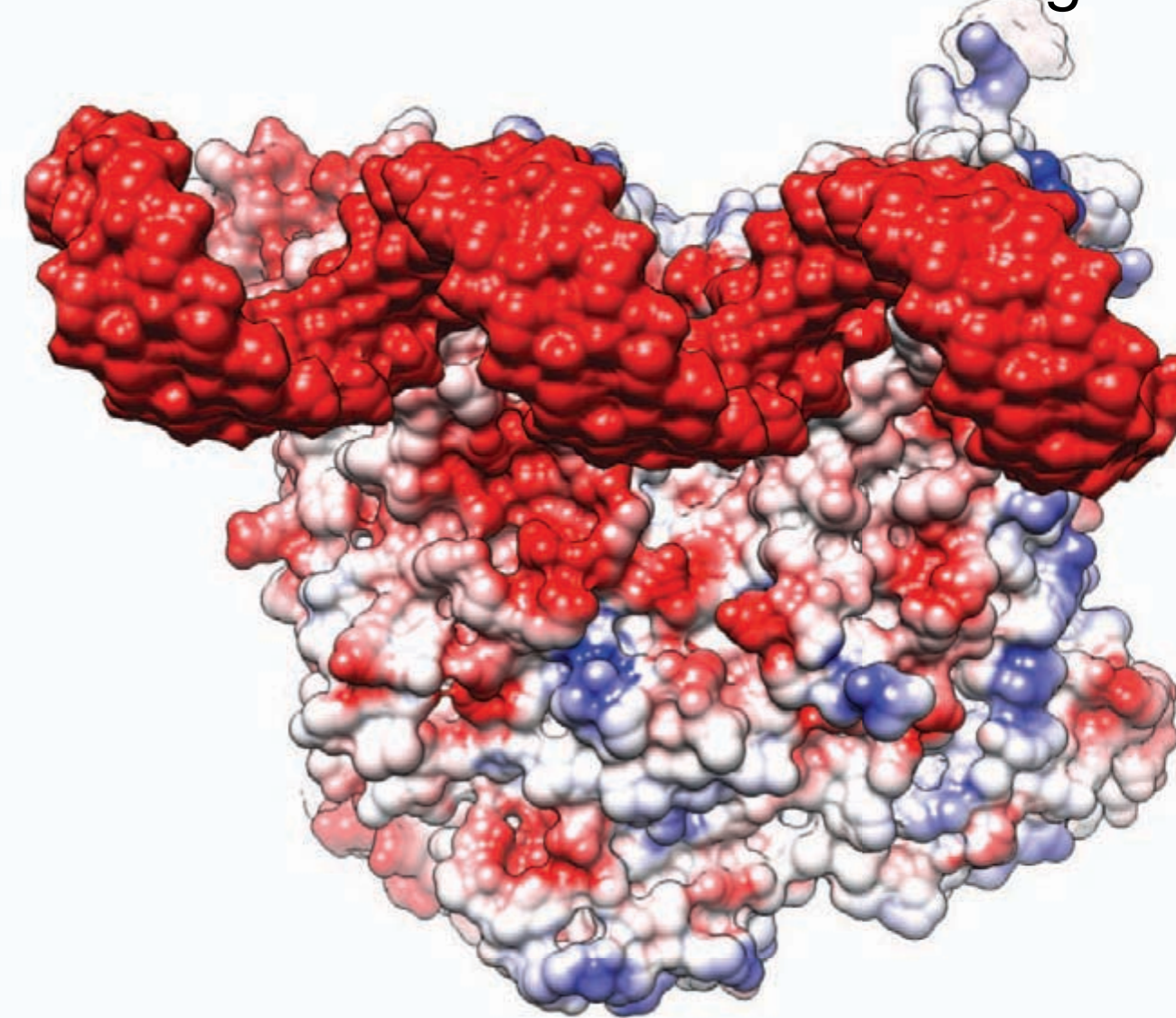
RNA interactions in transcribing MCRV



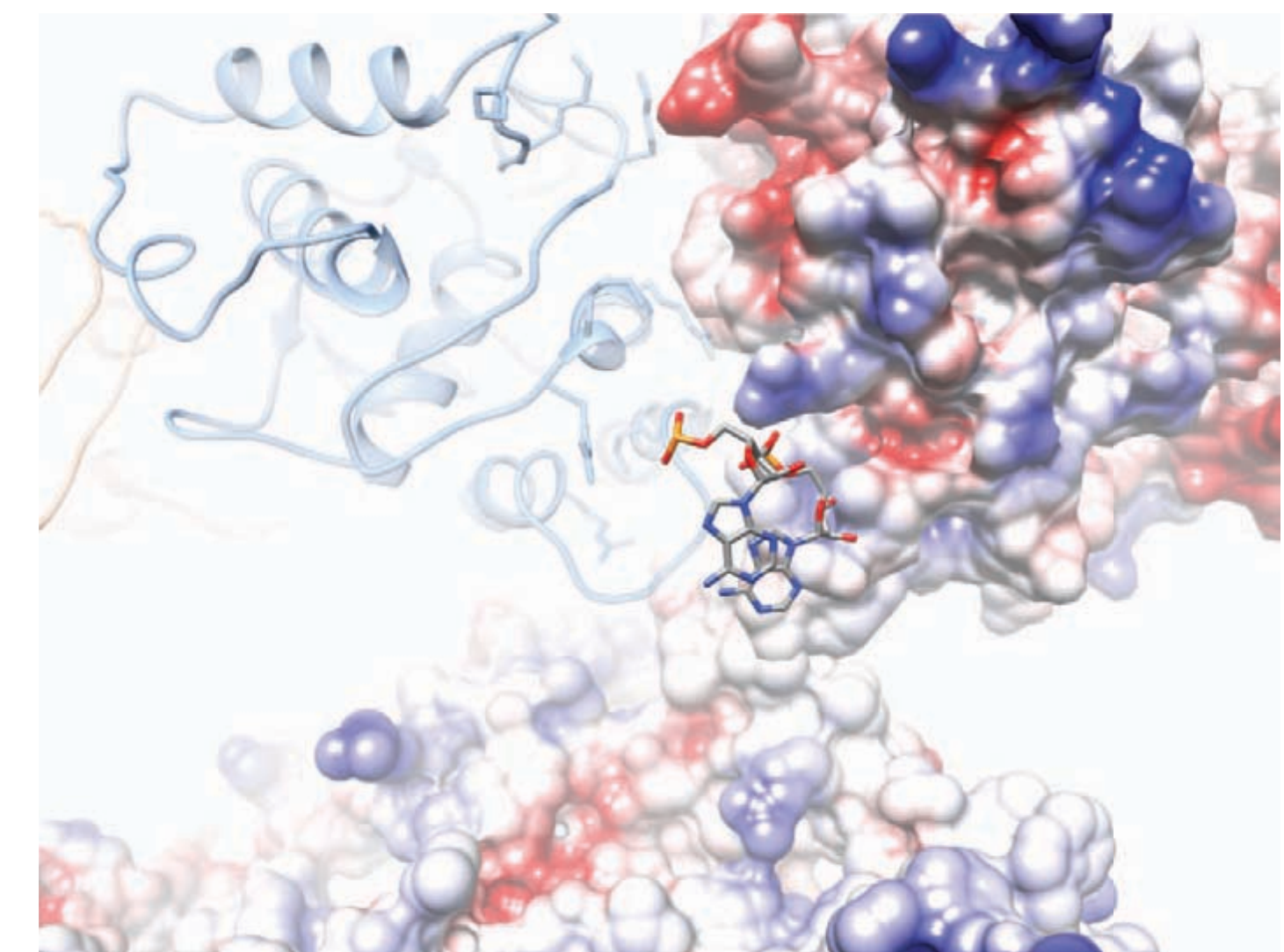
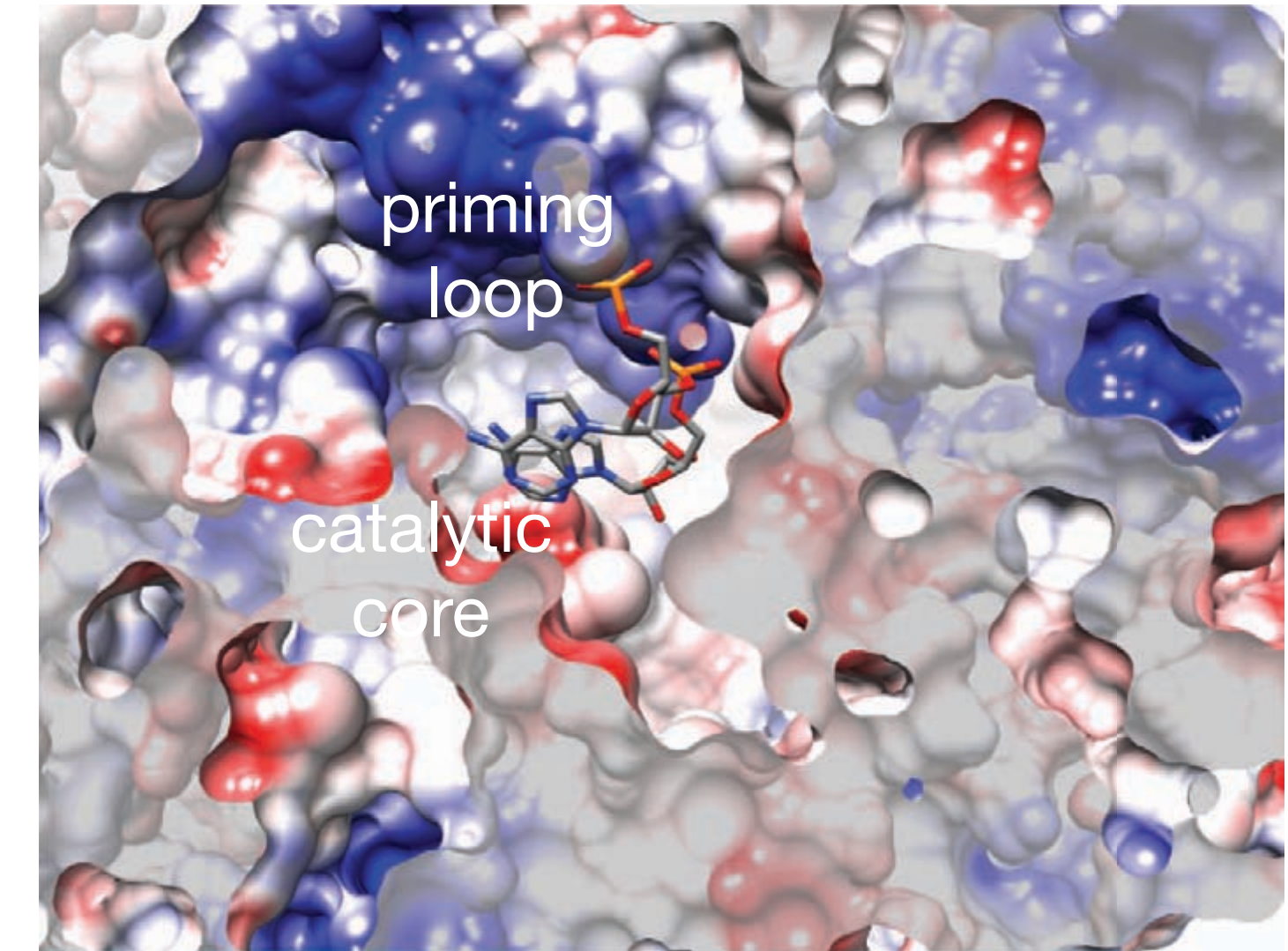
Nucleotide in the catalytic core interacts with the negatively charged catalytic core and the positively charged priming loop



ssRNA (~24 nucleotides) enters the RdRP at an entry site formed by the positively charged N-terminal and negatively charged finger domain.



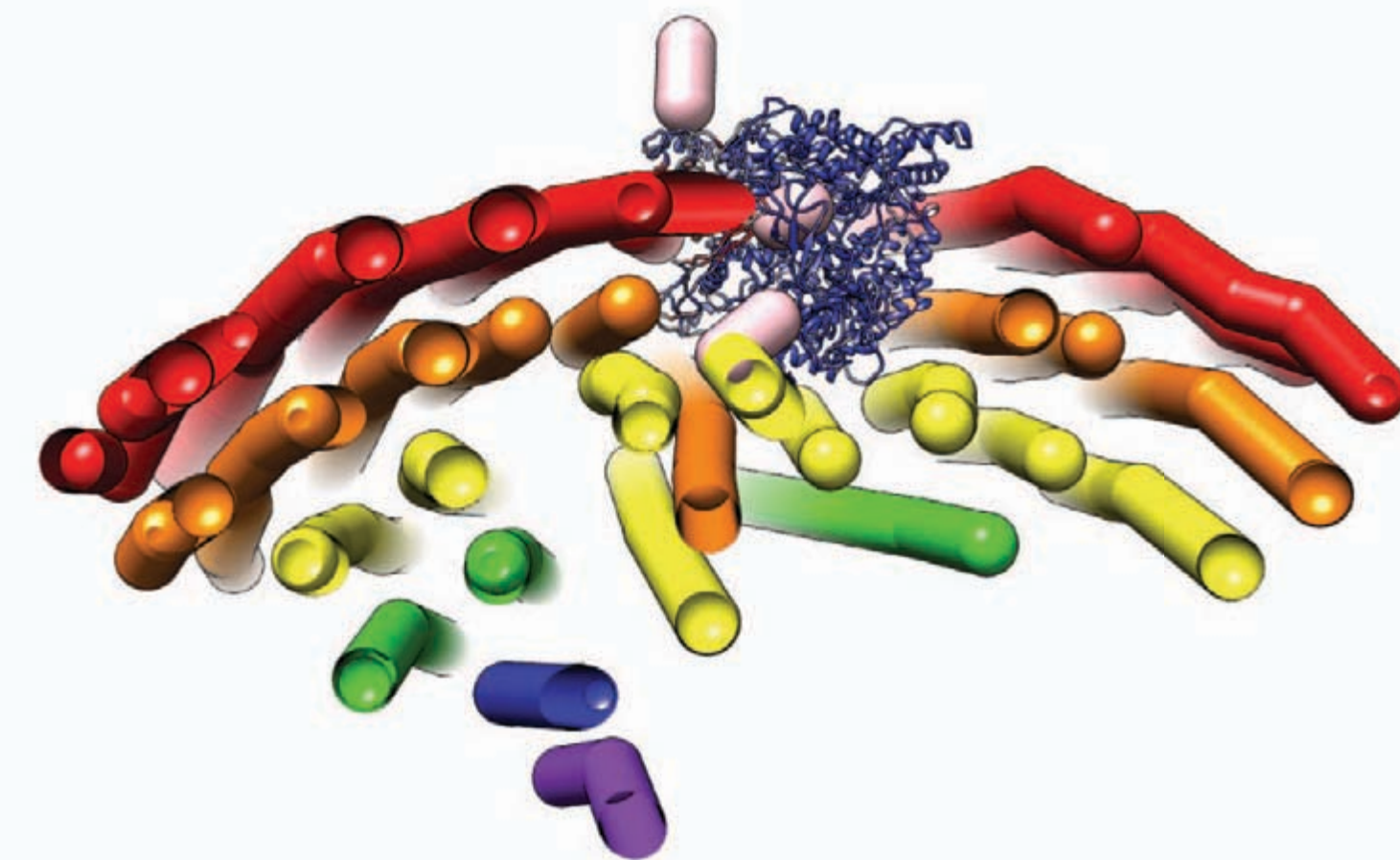
dsRNA runs along the opposite side of the RdRP, which is largely negatively charged



Residues 316-346 and 496-506 of VP3A form a positively charged constriction point and interact with RNA at the inner capsid shell

Conclusions

- VP11 has a novel fold but shares similar location clamp proteins in Spinareovirinae. This suggests the most recent common ancestor of all Reoviridae possessed a clamp (VP11 is a relic) and contraindicates the origin of Spinareovirinae from Sedoreovirinae.
- MCRV genome is organized in a spiral arrangement around the RdRP in layers and not in a coned-coil model as previously suggested.
- The VP3 N-arm is multifunctional and adopts 4 different conformations, not only connecting adjacent VP3 subunits, but also anchoring the RdRP. N-terminal fragments of VP3 have been shown to be involved in RNA organization and movement around the RdRP.
- Unlike other reoviruses the number of RdRPs is NOT consistent with the number of genome segments. VP3 and RNA may be involved in the organization of the RdRP.



Acknowledgments

MCRV

- Wen Jiang (Purdue)
- Jason Kaebler (Rutgers)
- Qinfen Zhang (Zhongshen)

Pathwalker

- Mariah Baker (BCM,UTH)
- Ian Rees (BCM)
- Steve Ludtke (BCM)
- *Muyuan Chen (BCM)*

