

SIMONS
ELECTRON
MICROSCOPY
CENTER

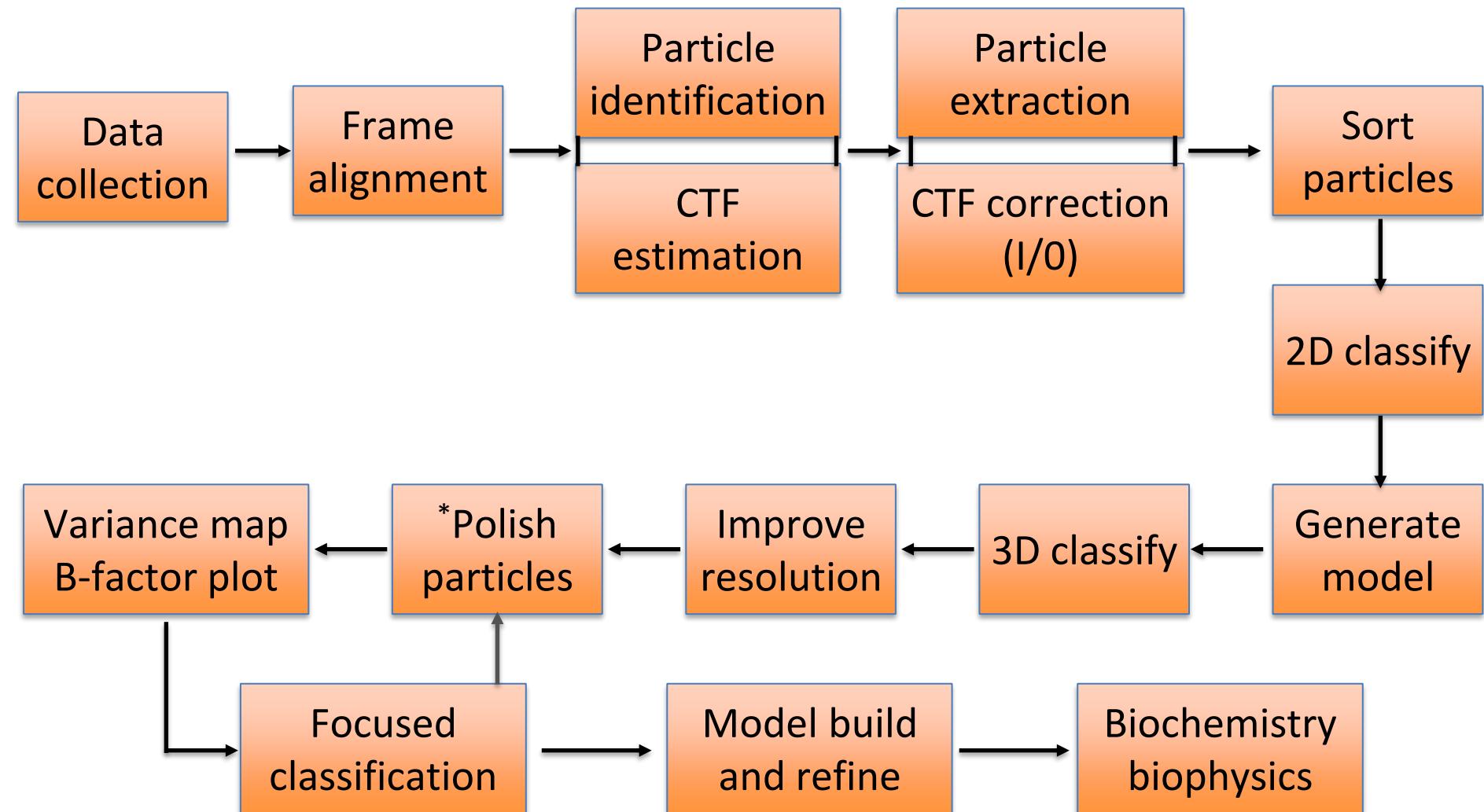
NYSBC

Winter-Spring 2024 EM Course

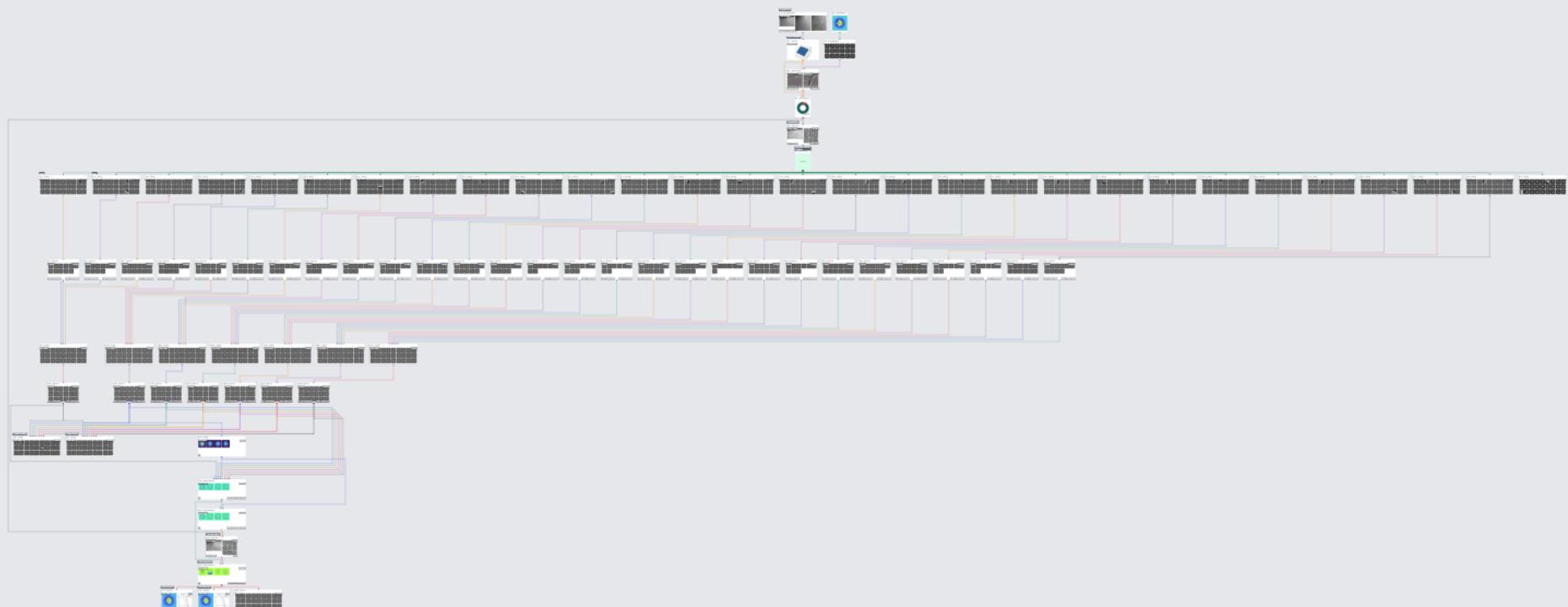
Single-particle analysis (part II)

Reza Khayat & Amedee des Georges

SPA Image analysis

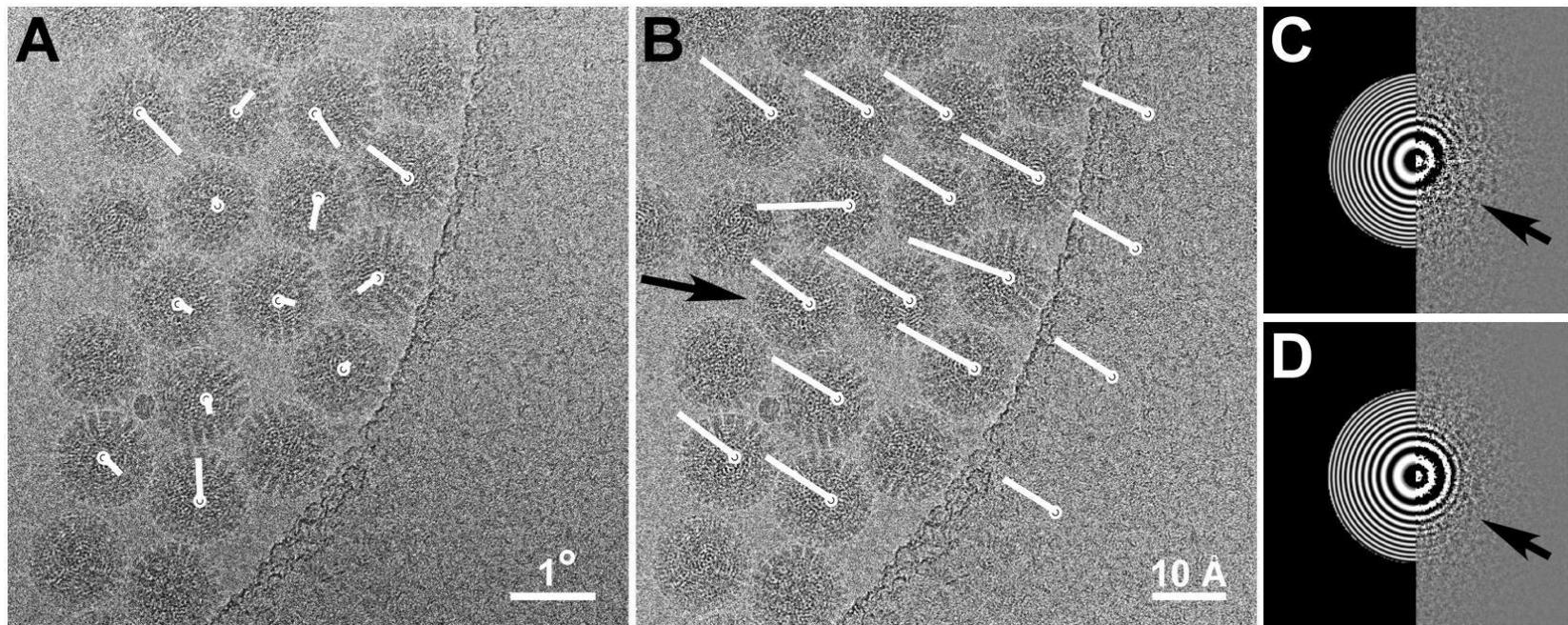


Workflow in cryoSPARC



Movie frame alignment

- UCSF MotionCor2
- Unblurr
- Warp
- cryoSPARC



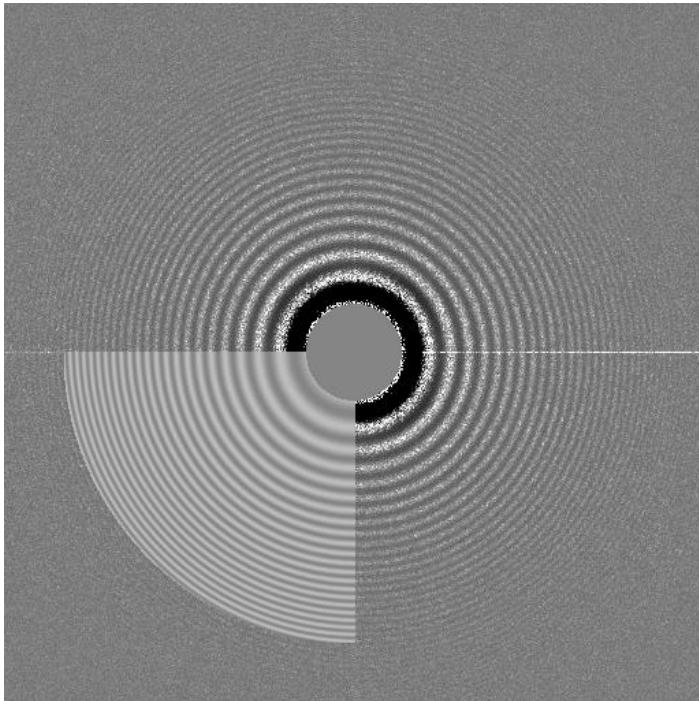
A/C: Unaligned micrographs and power spectrum

B/D: Aligned micrographs and power spectrum

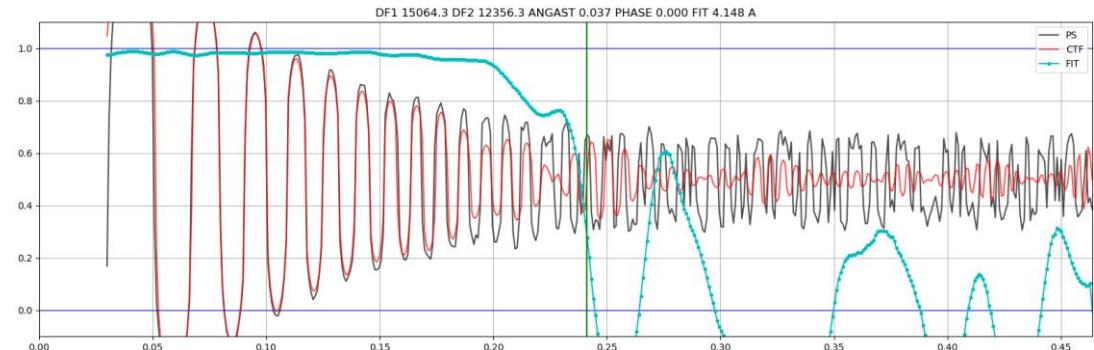
Campbell et al., 2012

CTF Estimation

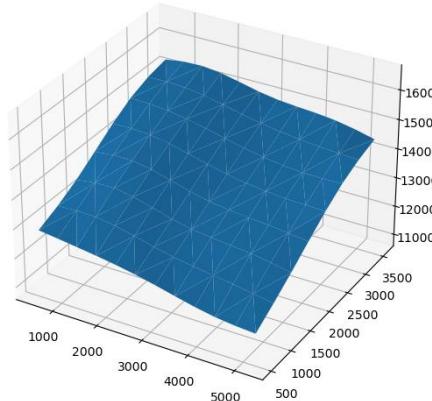
- ACE
- CTFFIND4
- Sparx/EMAN
- GCTF
- Warp
- CryoSPARC (patch method)



CTFFIND4



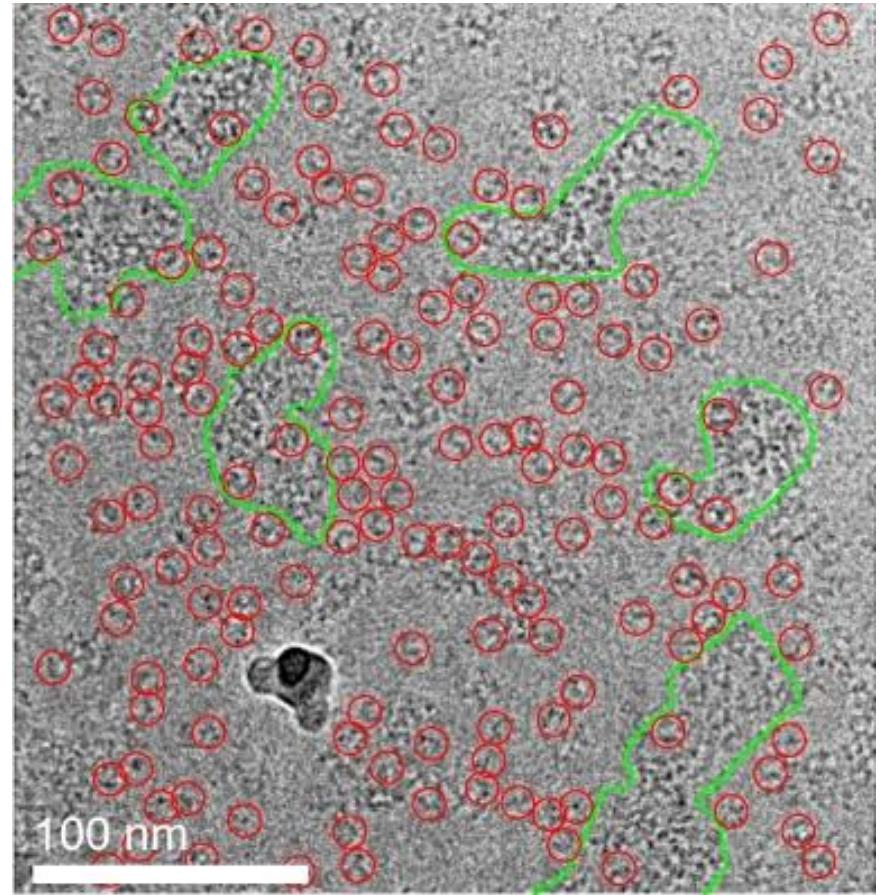
My stuff (CryoSPARC)



Particle Picking

Manual – Blob – Template – Neural network

- FindEM
- DoG Picker
- Gautomatch
- Topaz
- Warp
- DeepPicker
- DeepEM
- crYOLO
- PIXER
- DRPnet
- DeepCryoPicker
- AutoCryoPicker
- *Start with provided model, get 2D classes, and retrain*



Bepler et al., 2019

Curation (cryoSPARC)

CryoSPARC P105 Harris's Project W93 Valentin's 10853 Refmotion J2633 Manually Curate Exposures 22:27:15

Event Log Interactive Inputs and Parameters Outputs Metadata

Total: 12,299 Accepted: 7,165 Rejected: 5,134 Thresholds: 1 Done

Exposure Plot

Type Scatter Average defocus (Å) X-axis CTF fit resolution (Å) Y-axis Average defocus (Å)

Micrographs

Index	Avg Inten.	DF Avg	Astig	Phase	CTF Fit	CTF CC	DF Range	Tilt Angle	Rel Ice Thick.
5	10900.414	04.40	0	0	3.104	0	251	2.7	0.998
6	1.06	17684.855	47.93	0	3.077	0	282	2.8	0.997
7	0.18	16365.057	68.24	0	3.269	0	291	2.9	0.997
8	0.83	14951.990	80.97	0	3.015	0	224	3.1	0.998
9	0.96	13542.951	89.22	0	3.078	0	302	3.6	0.999
10	-0.13	12583.302	126.12	0	2.886	0	287	3.5	0.998
11	0.75	14221.251	105.79	0	3.113	0	282	3.6	0.996
12	0.57	15795.734	87.17	0	3.176	0	242	3.8	0.997
13	0.06	17363.037	67.37	0	3.286	0	216	3.2	0.997
14	0.83	18951.289	72.05	0	3.120	0	230	3.2	0.996
15	1.02	18145.645	84.93	0	3.259	0	266	3.2	0.999
16	0.73	16846.275	62.54	0	3.246	0	262	3	0.998
17	0.56	15491.193	101.52	0	3.316	0	281	3	0.997
18	0.95	14124.570	153.63	0	3.110	0	228	2.9	0.998
19	-0.61	12917.520	191.85	0	3.161	0	259	2.8	1
20	-0.44	13354.119	258.2	0	3.115	0	302	3.1	1.001
21	0.21	14675.084	220.33	0	3.162	0	314	3.3	0.998

Thresholds Micrograph Diagnostics Current: 10 Reject

Average defocus (Å) Min: 9191.816 → Max: 14956.238

Minimum Exposure 70 (1298.322) Maximum Exposure 4041 (23272.672) Included 7,166 Excluded 5,133

CTF fit resolution (Å) Minimum Exposure 557 (2.351) Maximum Exposure 9463 (6.461)

Average Intensity

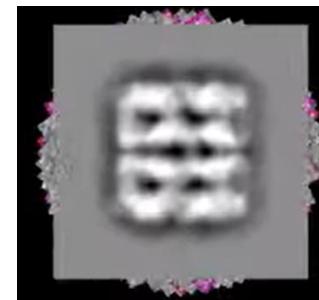
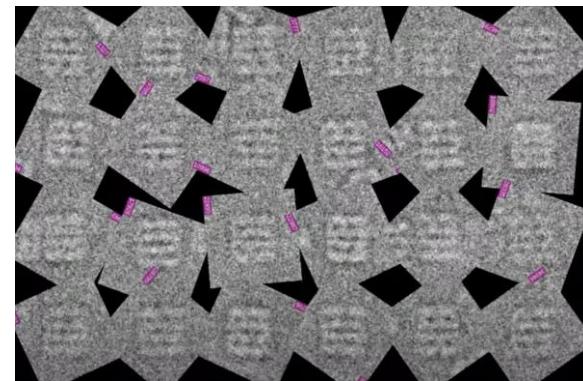
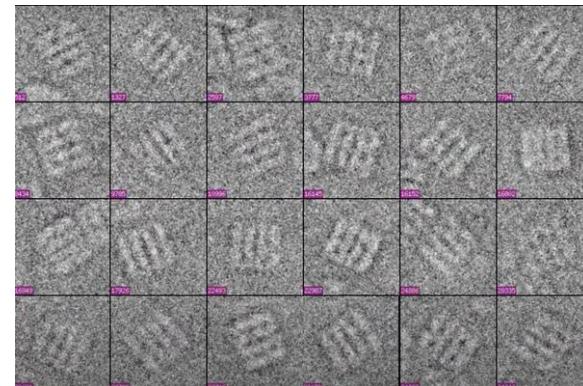
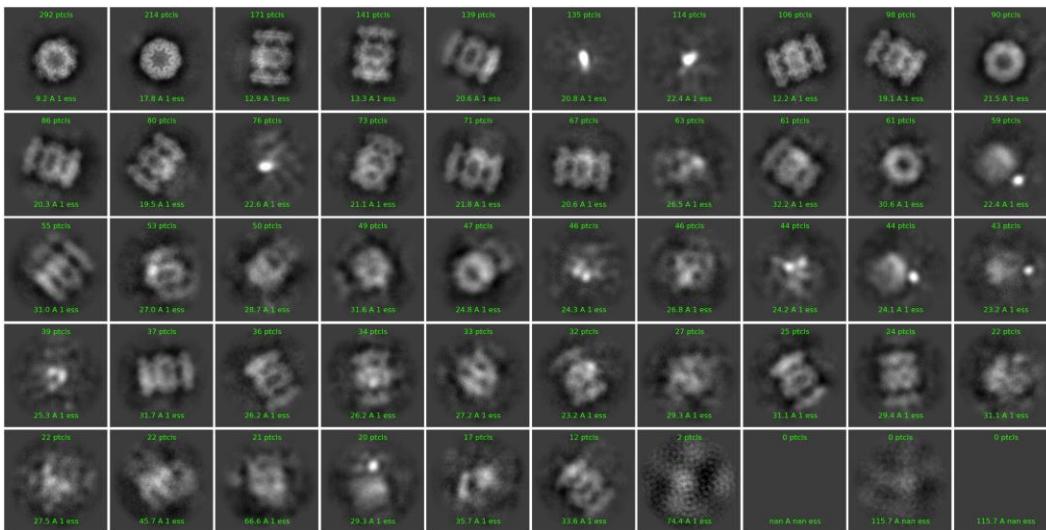
- Defocus range
- CTF fit resolution
- Number of particles
- Tilt angle
- Astigmatism
- Ice thickness

Particle Extraction

- Signal is delocalized according to energy of electron and defocus value of image:
 - Fred Sigworth 2022 lecture: $r = \delta * \lambda * f$
 - r is the radius that surrounds the particle. It describes how far the signal is delocalized.
 - $\delta = \text{defocus in } \text{\AA}$
 - $\lambda = 0.02 \text{ \AA}$ (wavelength of e- at 300kV)
 - Frequency = desired resolution (e.g. 0.33\AA^{-1} for 3 \AA)
- Even number with low prime factors (2, 3, 5, and 7)
 - I like 32, 64, 128, 256, 320, 384
- You may want to downsize (fourier bin) the particles to expedite initial data processing, and save on drive space.

2D Classification

- Spider
- Sparx/EMAN2
- ISAC
- Relion
- Simple
- cryoSPARC
- Warp
- *Remove “bad” particles*



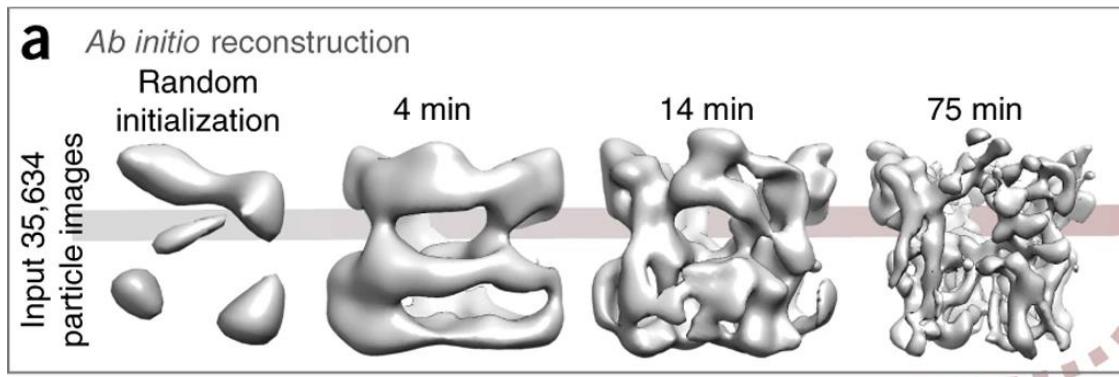
Lander 2009

2D Classification

- Anticipate 100 to 400 particles per class
- Don't ask for too many
- I split my particle stack into stacks of 100K particles and process each separately to get clean-vs-dirty particles
 - Radius of search
 - Relion
 - Tau fudge
 - CTF
 - cryoSPARC
 - Turn off Force Max over poses/shifts
 - Initial classification uncertainty factory (2 and above)
 - Number of iteration to anneal sigma as high as 25
 - Set online-EM iterations to 40
 - Set Batchsize per class to 400
 - Change Re-center mask threshold (possibly as high as 0.75) for centering particles and smearing neighbors
 - set White noise model to off

Initial Model

- Random conical tilt
- Orthogonal conical tilt
- Common-lines
- Tomography with STA
- Random initial parameters, optimize with stochastic gradient descent (SIMPLE, cryoSPARC, and Relion).



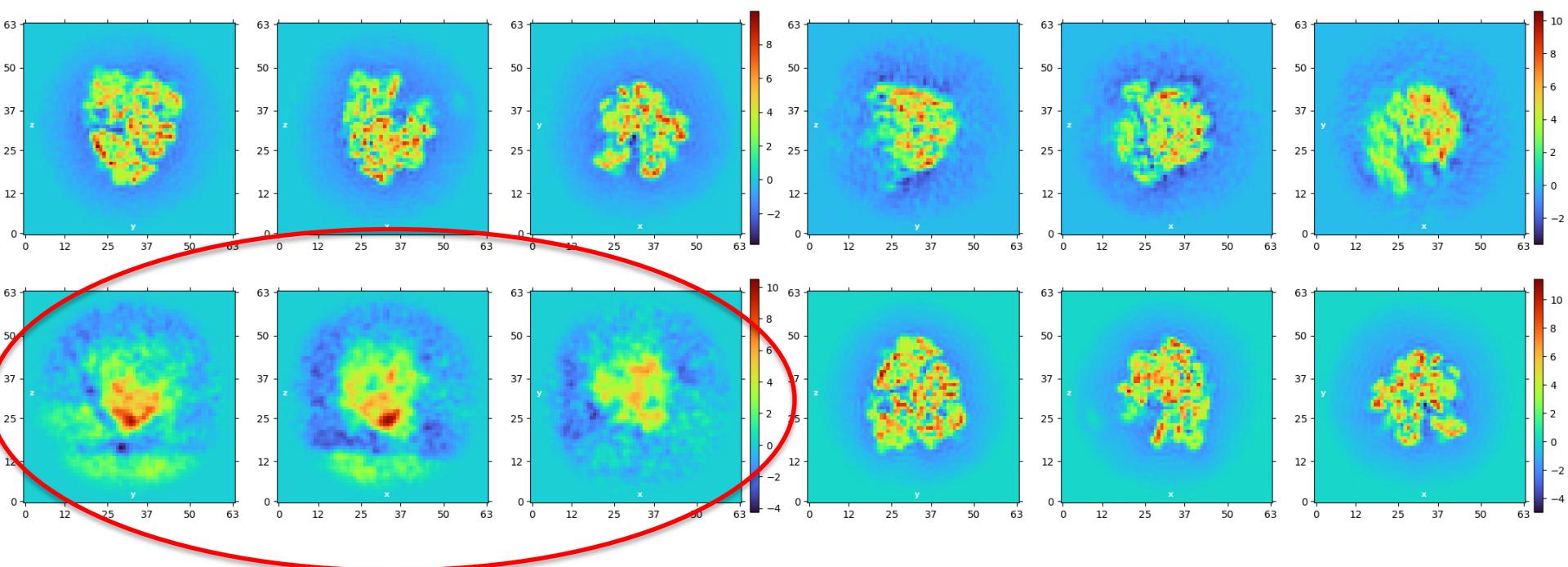
Punjani et al., 2017

Initial Model

- Generate multiple initial models if uncertain in model
 - Look for continuity in density
 - Look for sausages to indicate α -helices
 - Are projections comparable to class averages?
- Ask for multiple models to be generated
- Starting frequency should have more information than `particle_size / 5`
- Use C_1 symmetry

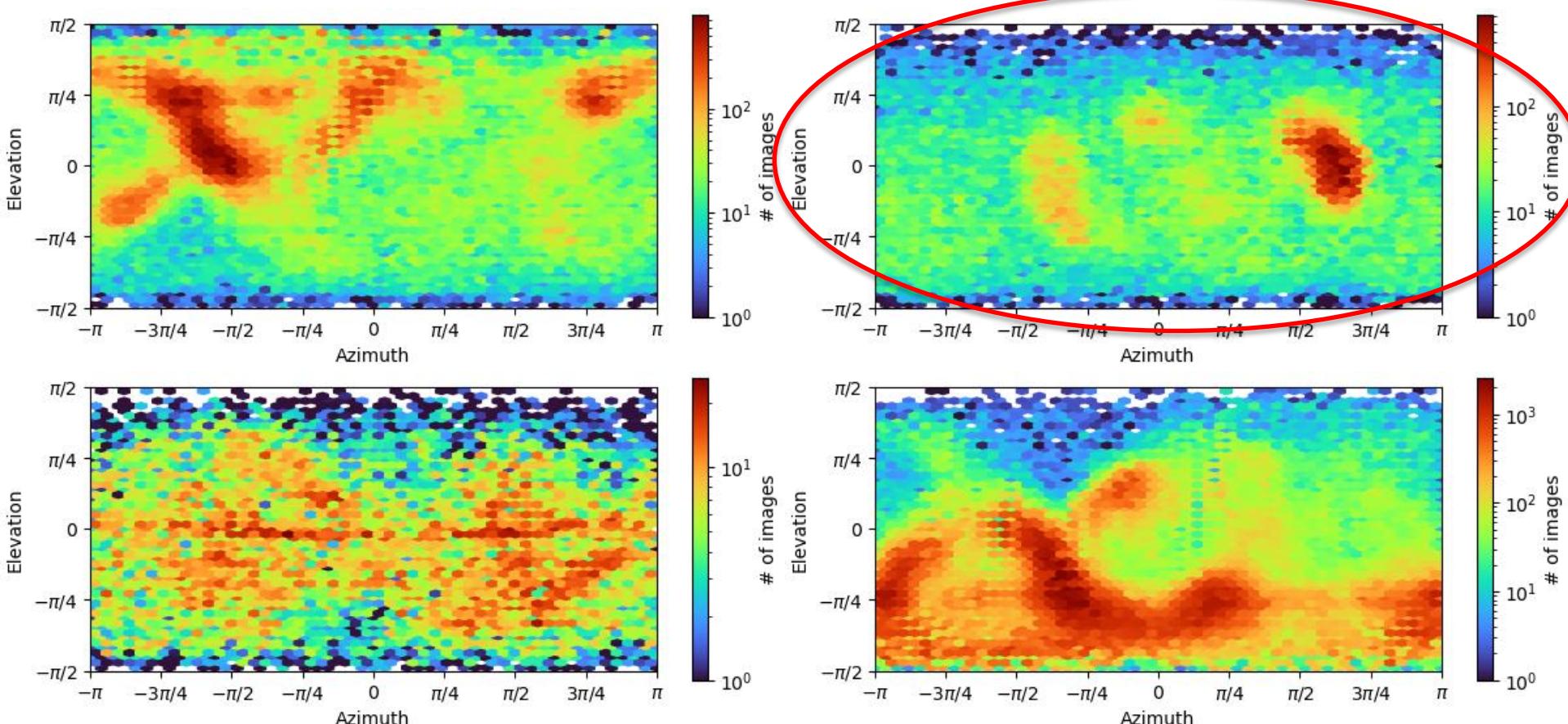
3D Classification

- Can be used to clean data further
 - Discard “bad” particles



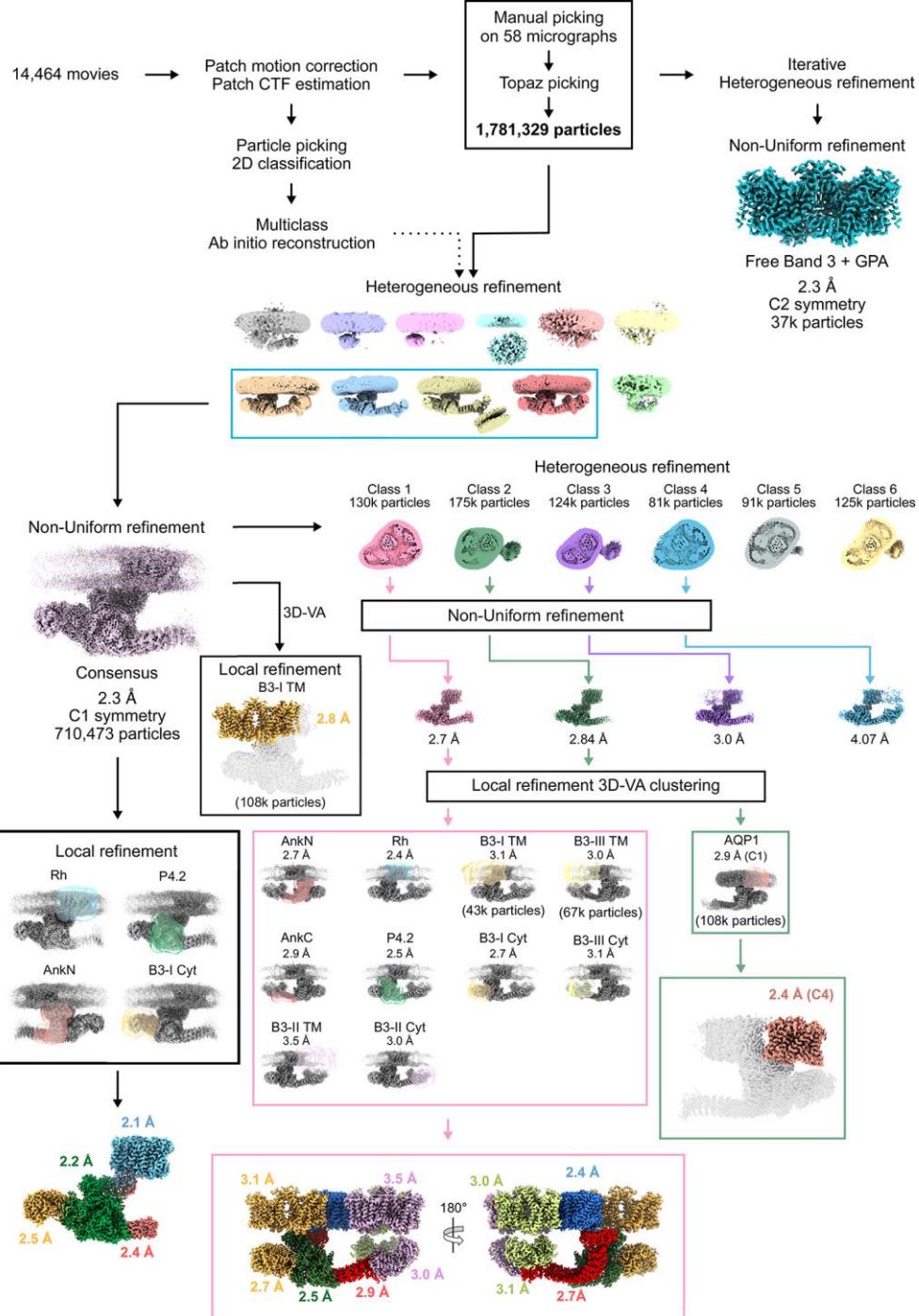
3D Classification

- Can be used to clean data further
 - Discard “bad” particles
 - Discard some preferred orientations



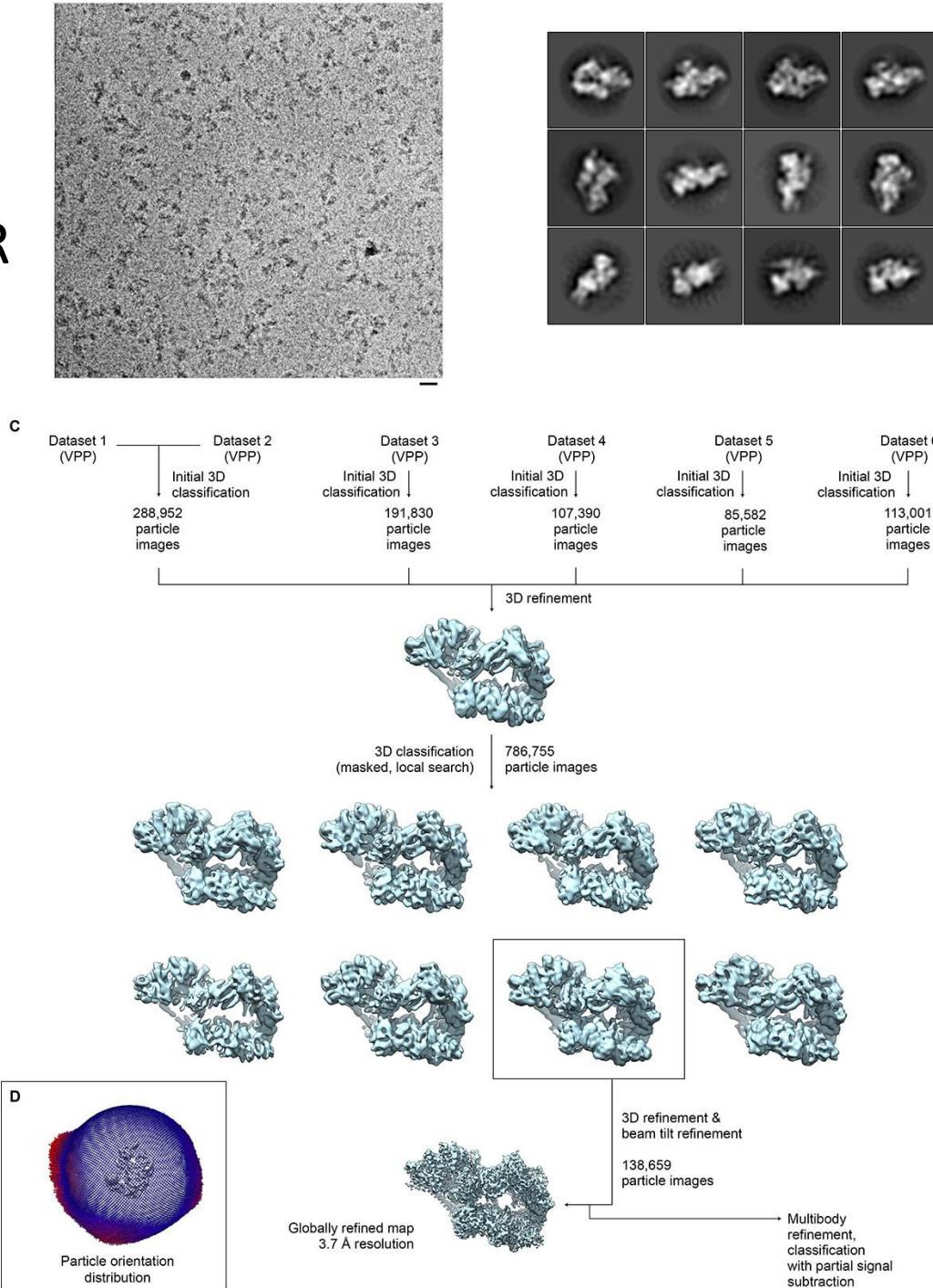
3D Classification

- Configurational and conformational differences
- Multiple ab initio models generate
- 3D classification use to identify configurational differences.
- Local refinement used to identify conformational differences



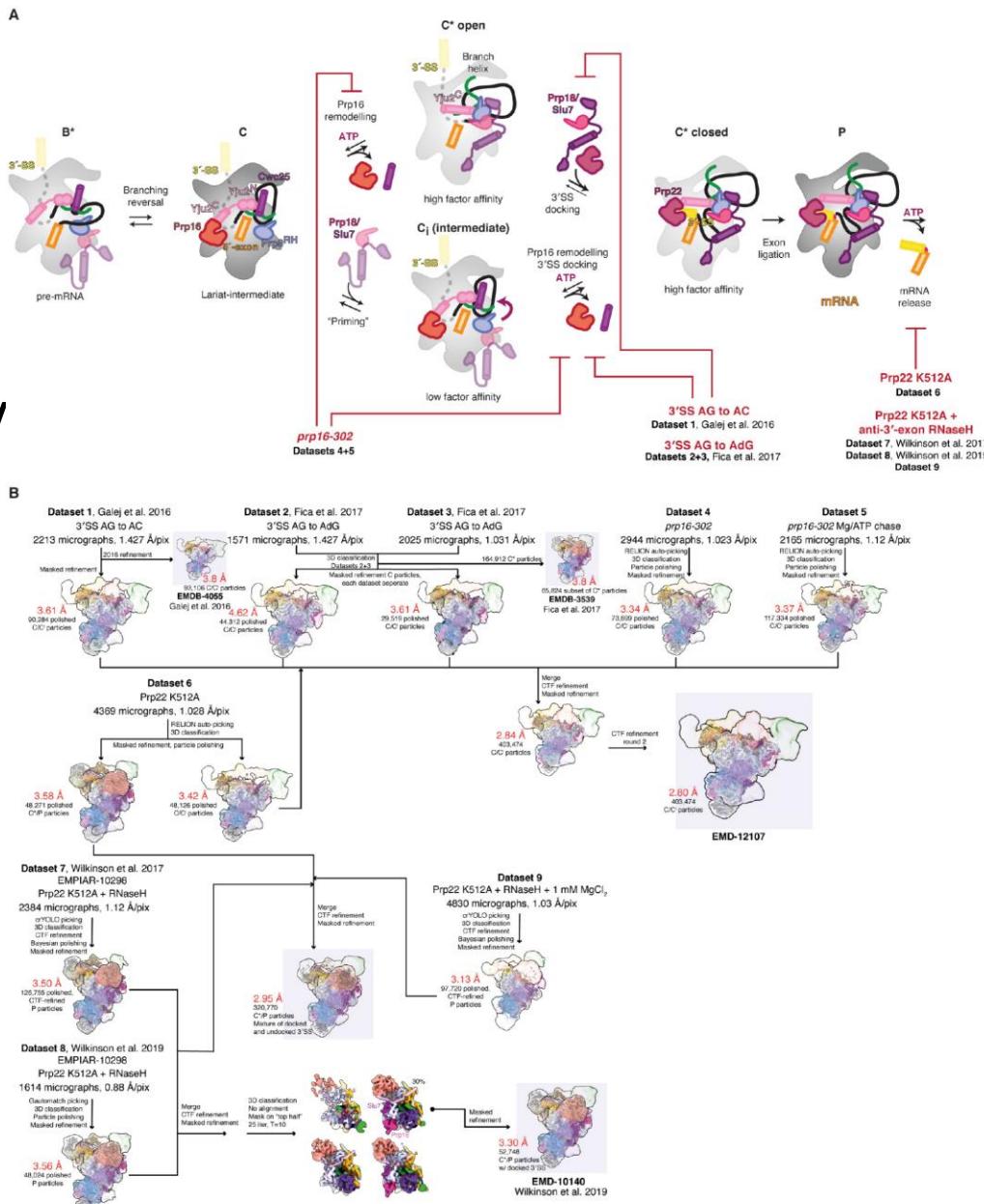
3D Classification

- TFIIH, transcription initiation by RNAPII and NER
- Enrich rare views that 2D classification would discard
- Do 2D classification (B) for sanity check
- Extract particles and perform 3D classification with high tau2_fudge value to enrich for rare views (D).
 - Value empirically determined. Try 1, 5, 10, 20, 50, 100, 200, 500, 1000



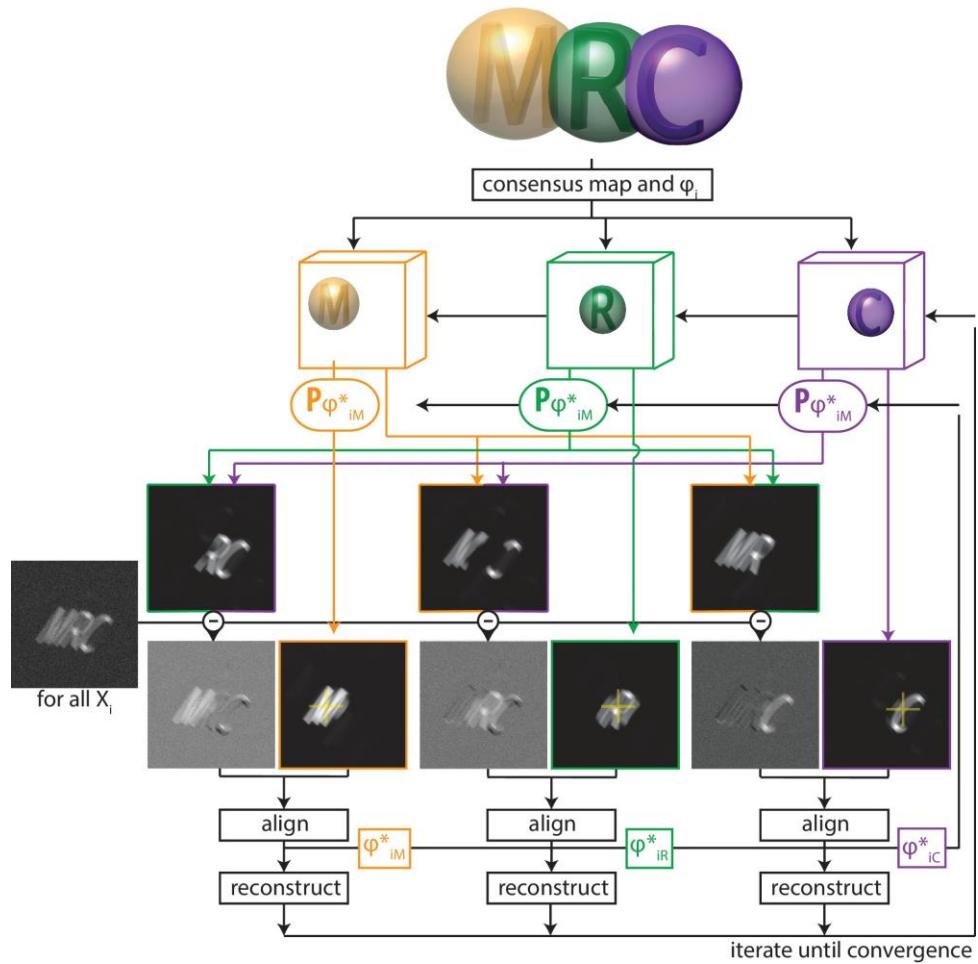
3D Classification

- Spliceosome
- Initial reconstruction is at 2.8 Å; however, lots of domains/proteins at periphery have poor density
- Signal subtraction coupled with focused classification and empirically determined tau2_fudge values (Relion) improve their resolutions
 - Try 1, 5, 10, 20, 50, 100, 200, 500, 1000 in parallel



Signal subtraction

- Improve resolution of desired region(s)
- Generate consensus map
- Mask region of interest
- Use consensus map to subtract everything outside/inside of mask from each particle
- Refine map of remaining signal
- Subtraction does not always work completely. May need to iterate through this process.



Nakane et al., 2018