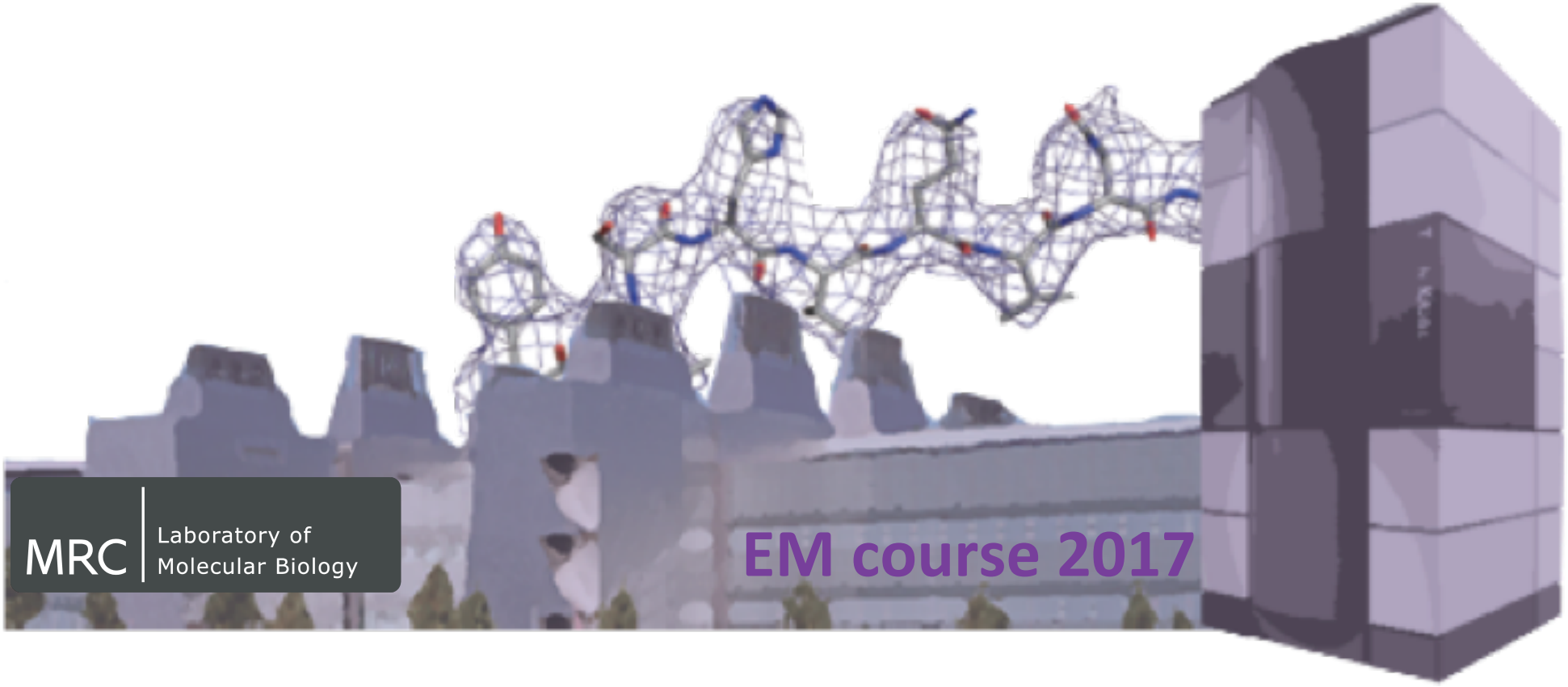
7. Data Processing Strategy

Rafael Fernandez-Leiro

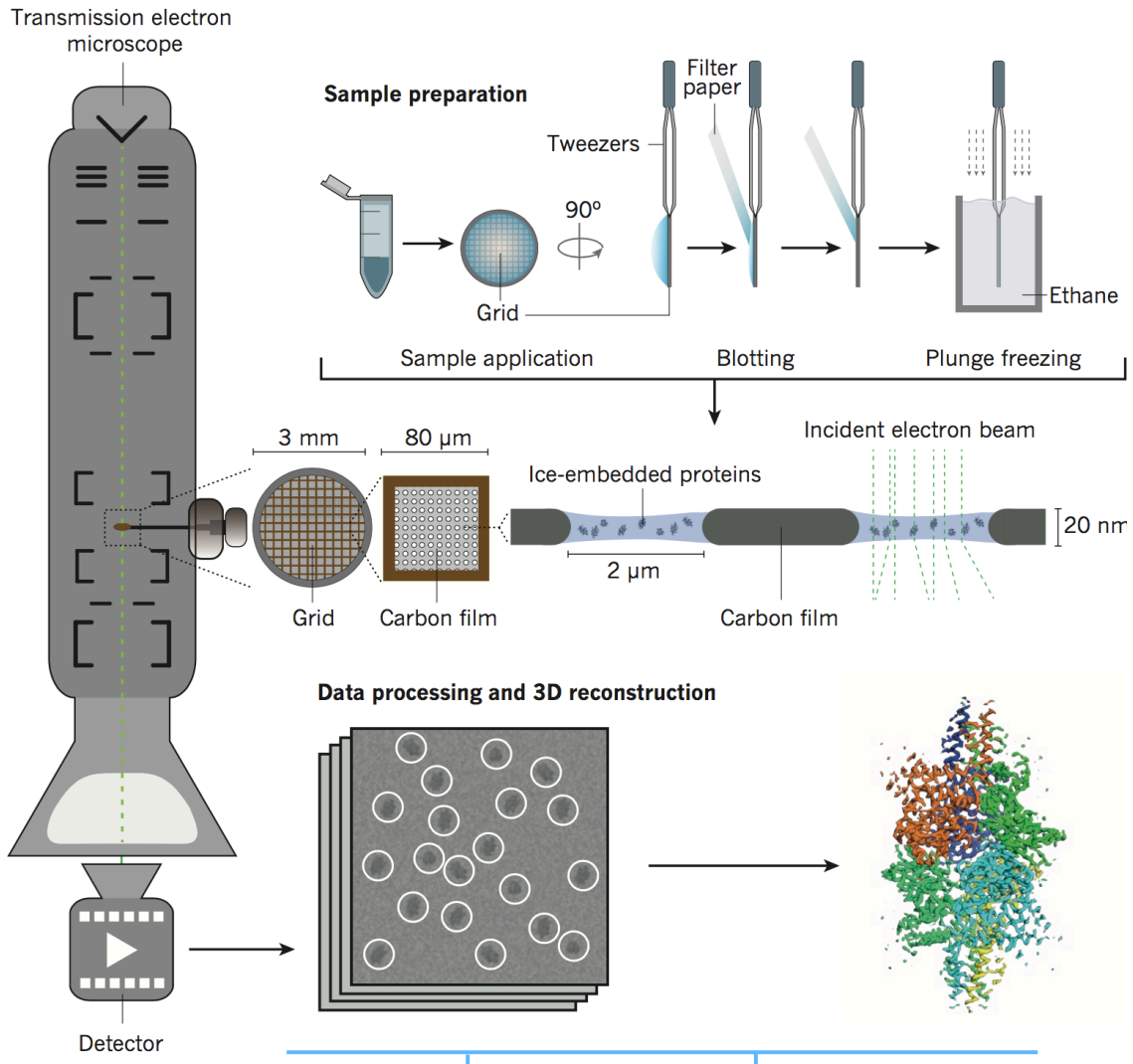


Cryo-EM - Single Particle

Richard Henderson

Chris Russo
Christos Savva

Paula da Fonseca

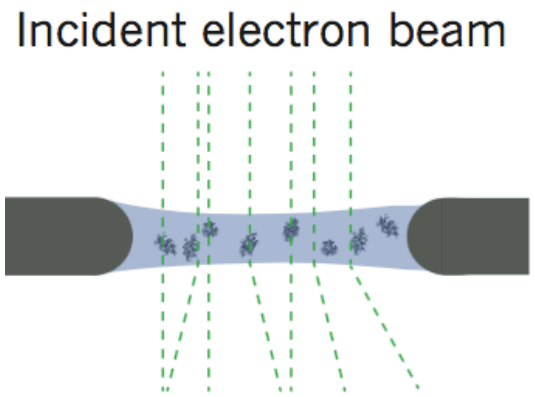


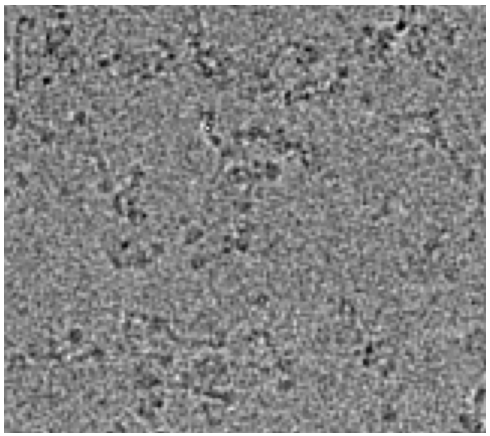
Lori Passmore

Sjors Scheres

Rafa F. Leiro

Cryo-EM - From 2D images to 3D structure(s)





projections of the original object in multiple (unknown) orientations

We collect data in 2D
(projections/integral through object)
We want to go back to 3D

↓

Microscope introduces artefacts
We don't know the orientations
Data is noisy

↓

We can perform a 3D reconstruction provided we:
- Correct for the **CTF** artefacts
- Know the **orientations** of all particles
- **Averaging** will take care of low SNR

Cryo-EM - From 2D images to 3D structure(s)

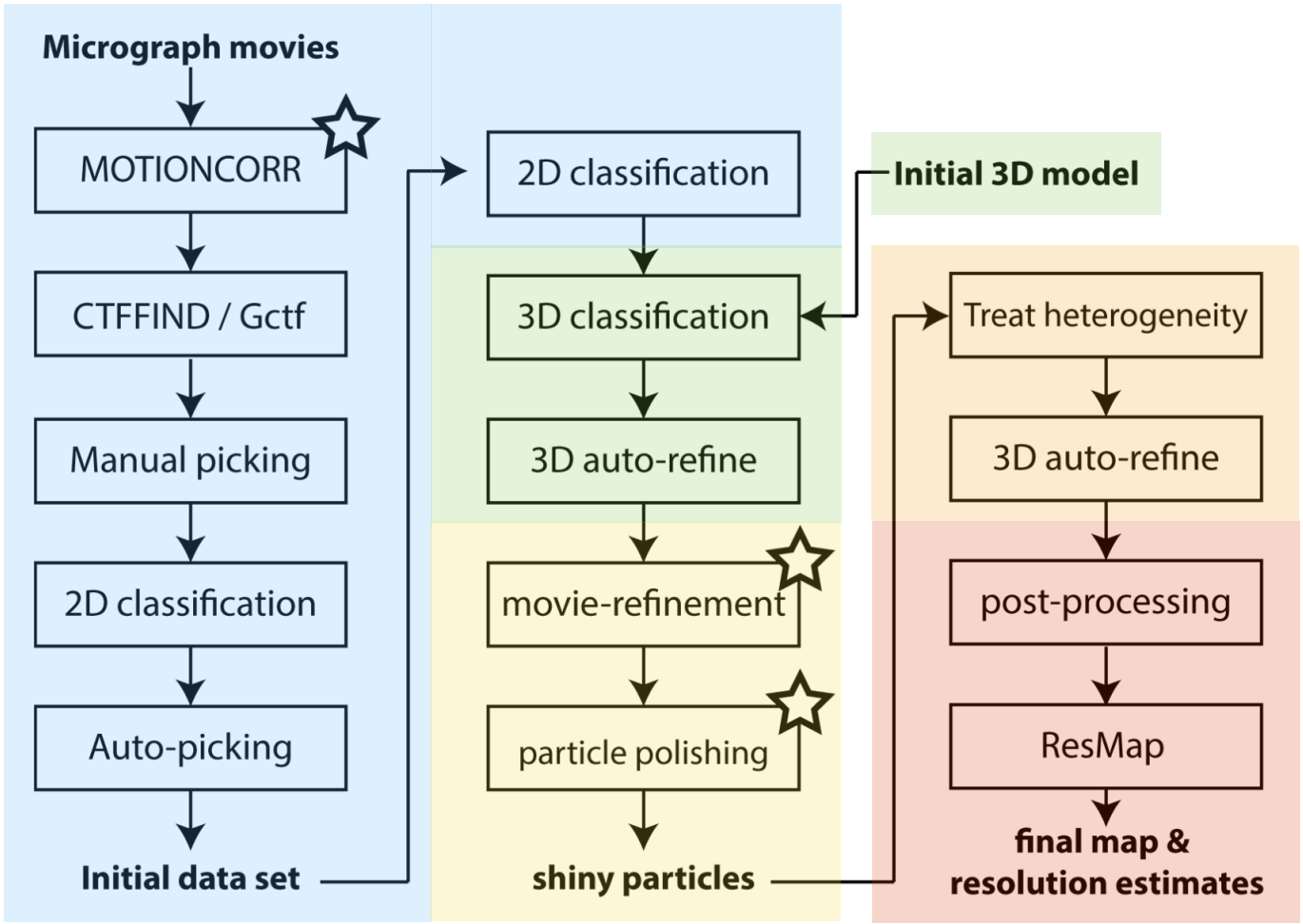
RELION (REgularised LIkelihood OptimisatioN)

Instead of assigning discrete values assign probabilities(each particle contributes to all references and in all orientations)

Learns critical parameters from the data themselves

LOTS of other software: Frealign, EMAN, SIMPLE, Xmipp, cryoSPARC, SPHIRE, and many more…

Single particle data processing strategy



*typical scheme for a well behaved dataset

Single particle data processing strategy

24h K2 data acquisition

CISTA (TEMP)

775 movies

1.86 TB

TERARAID

[…/DATA/RAW/.]

775 compressed movies

800 GB

[…/DATA/MC2/.]

775 movies + mrc

3.7 TB

[…/PROCESS/.]

RELION
symlink movies
symlink MC output

<50 GB (Initial)
<500 GB (Final)

rsync

clean jobs

rsync

BEEGFS

[…/PROCESS/.]

RELION
symlink movies
symlink MC output

<50 GB (Initial)
<500 GB (Final)

EXTERNAL HD

Backup

[…/DATA/RAW]

800 GB + 200 GB (+3.7 TB)

clean jobs

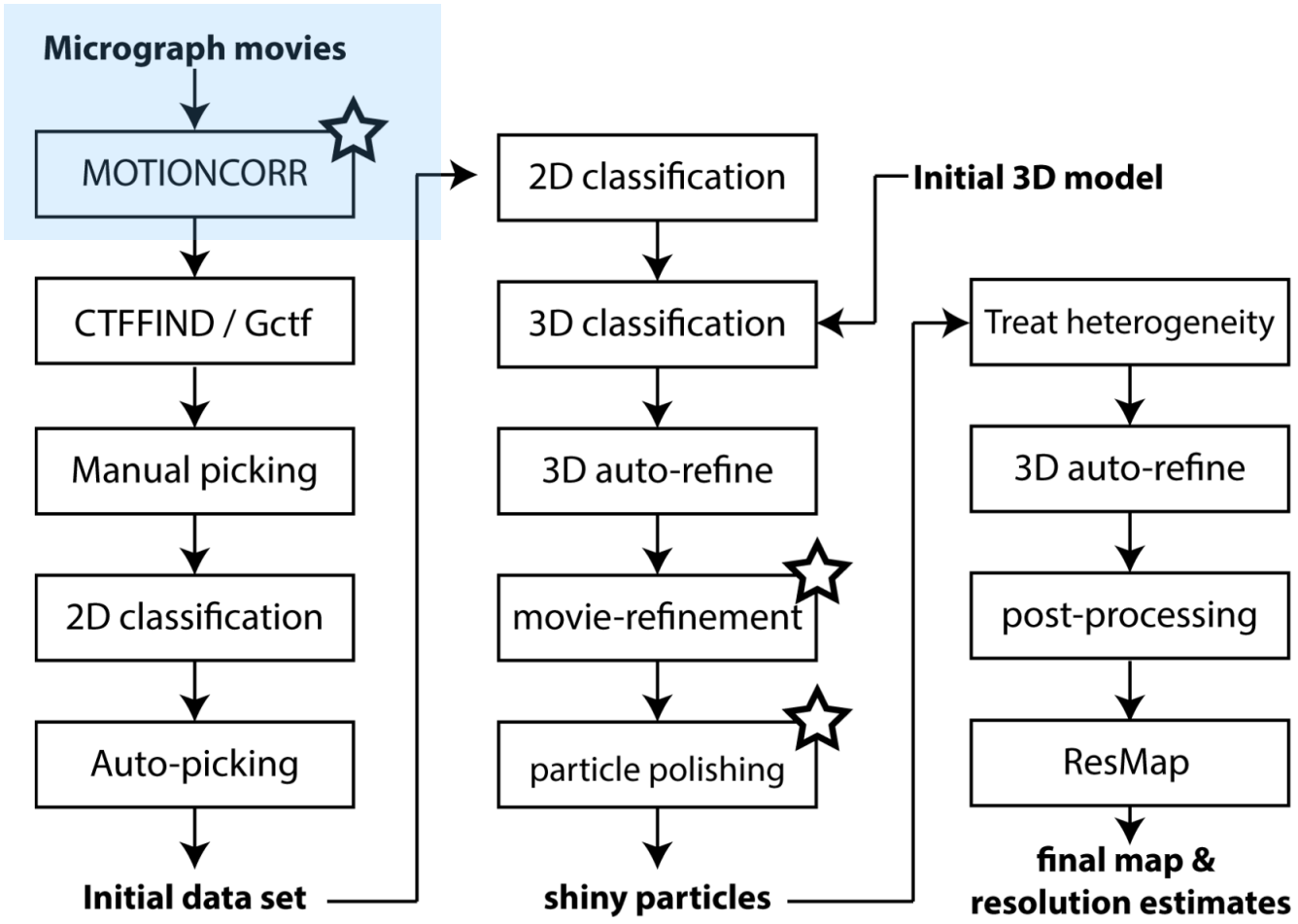
MOVIE PROCESSING WITH SOFT-LINKS TO TERARAID
[…/DATA/MC2/.]

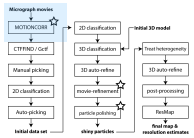
Start running “MovieRefine” on hal/hex (not submitting to queue)

After particle movie extraction you can stop the job and continue on the cluster as particle files are now on beegfs

Remember to gentle-clean MovieRefine jobs after you are done with Polishing!

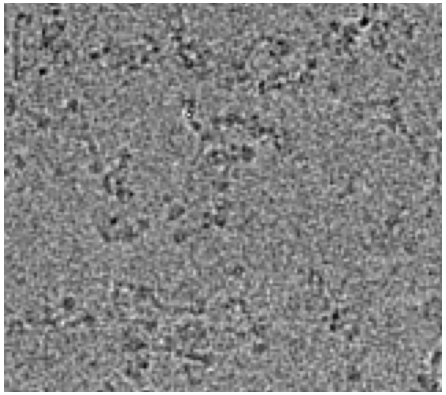
Motion Correction



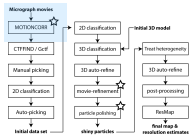


Motion Correction

Micrograph
$40e^-/Å^2$



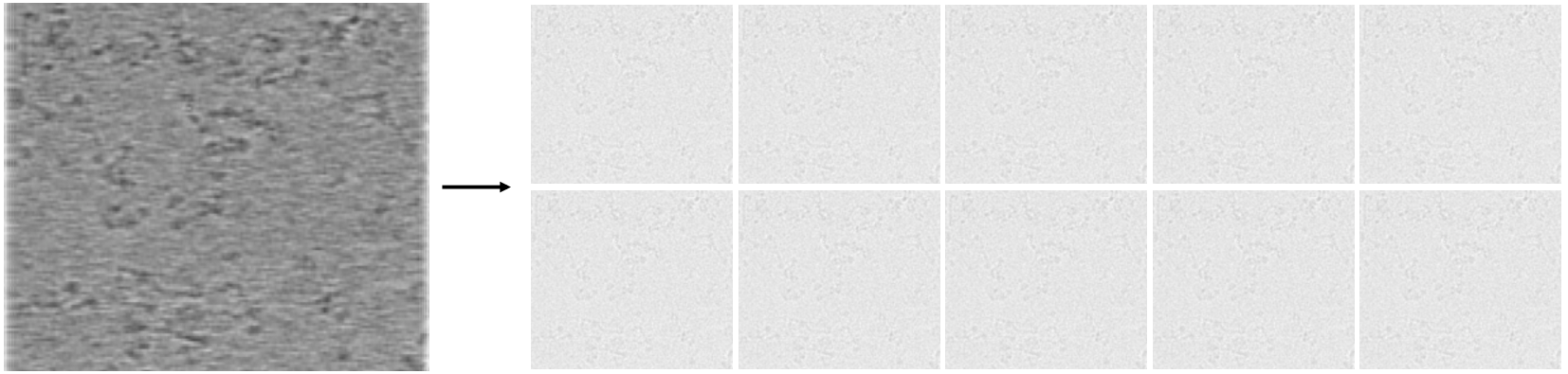
Movie or Micrograph Stack
$4e^-/Å^2$.frame

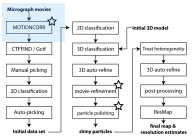


Motion Correction

Micrograph
40e$^-$/Å^2

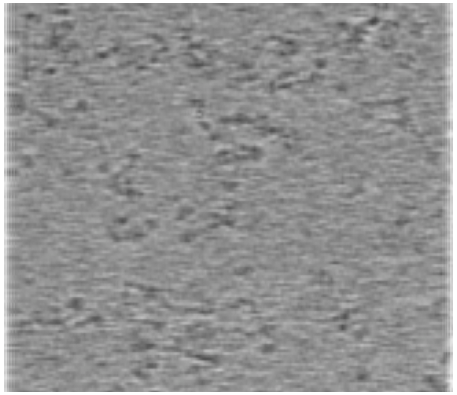
Movie or Micrograph Stack
4e$^-$/Å^2.frame





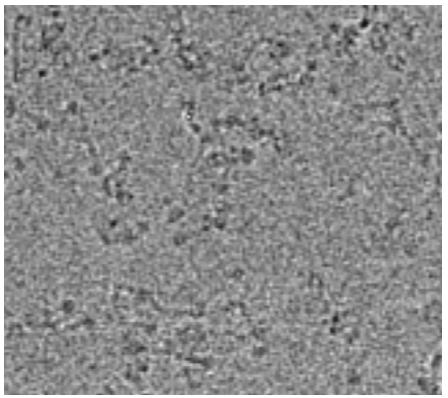
Motion Correction

Micrograph
$40e^-/Å^2$



Movie or Micrograph Stack
$4e^-/Å^2$.frame

Motion Correction



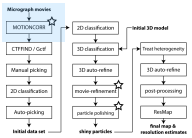
Corrected Micrograph

Brilot 2012 (JSB) **Unblur**

Zheng 2016 (Nature Methods) **MoitonCor2**

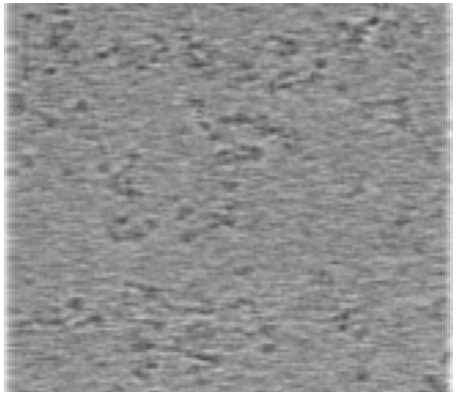
Rubinstein 2015 (JSB) **alignparts_lmbfgs & alignframes_lmbfgs**

Abrishami 2015 (JSB) **Xmipp (Optical Flow)**



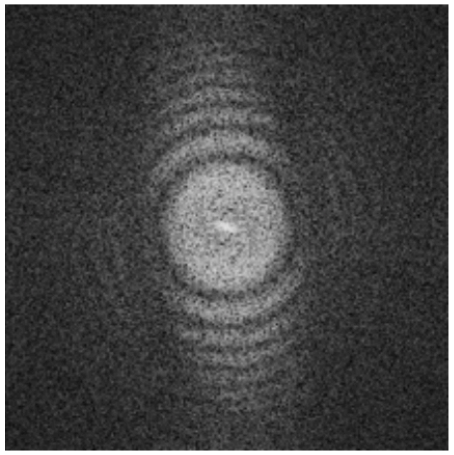
Motion Correction

Micrograph
40e$^-$/Å^2



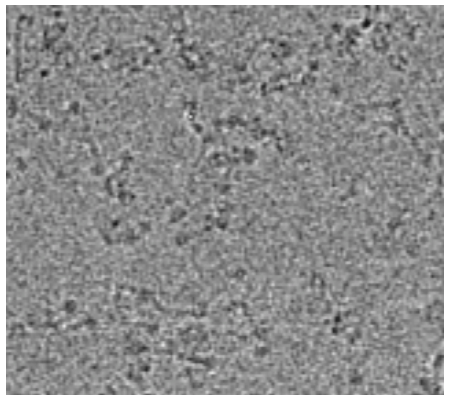
Movie or Micrograph Stack
4e$^-$/Å^2.frame

FFT



Optimise data acquisition parameters to record **GOOD DATA**

Motion Correction



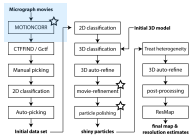
Corrected Micrograph

Brilot 2012 (JSB) **Unblur**

Zheng 2016 (Nature Methods) **MoitonCor2**

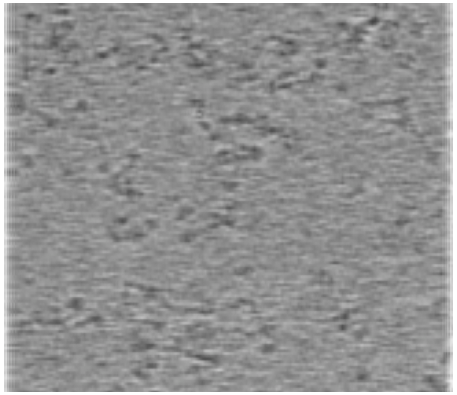
Rubinstein 2015 (JSB) **alignparts_lmbfgs & alignframes_lmbfgs**

Abrishami 2015 (JSB) **Xmipp (Optical Flow)**



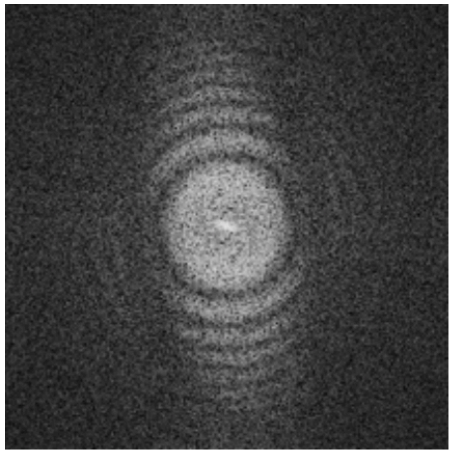
Motion Correction

Micrograph
40e$^-$/Å^2



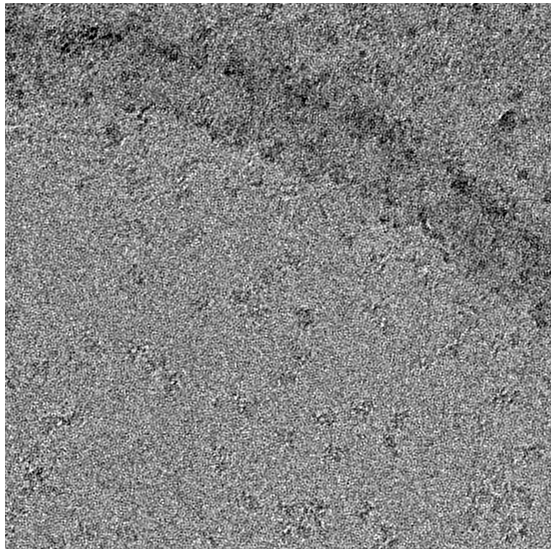
Movie or Micrograph Stack
4e$^-$/Å^2.frame

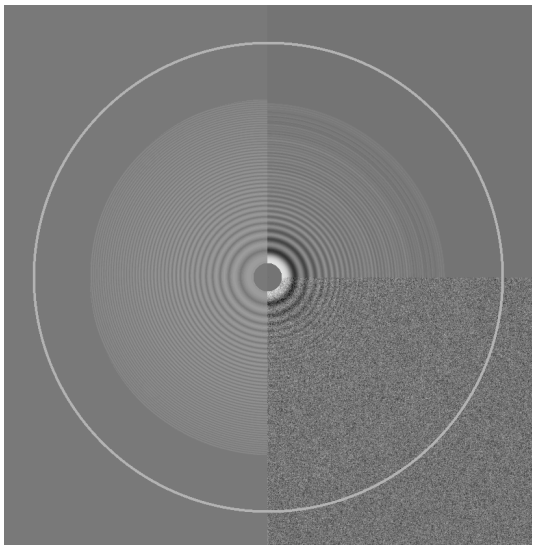
FFT

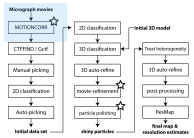


Optimise data acquisition parameters to record **GOOD DATA**

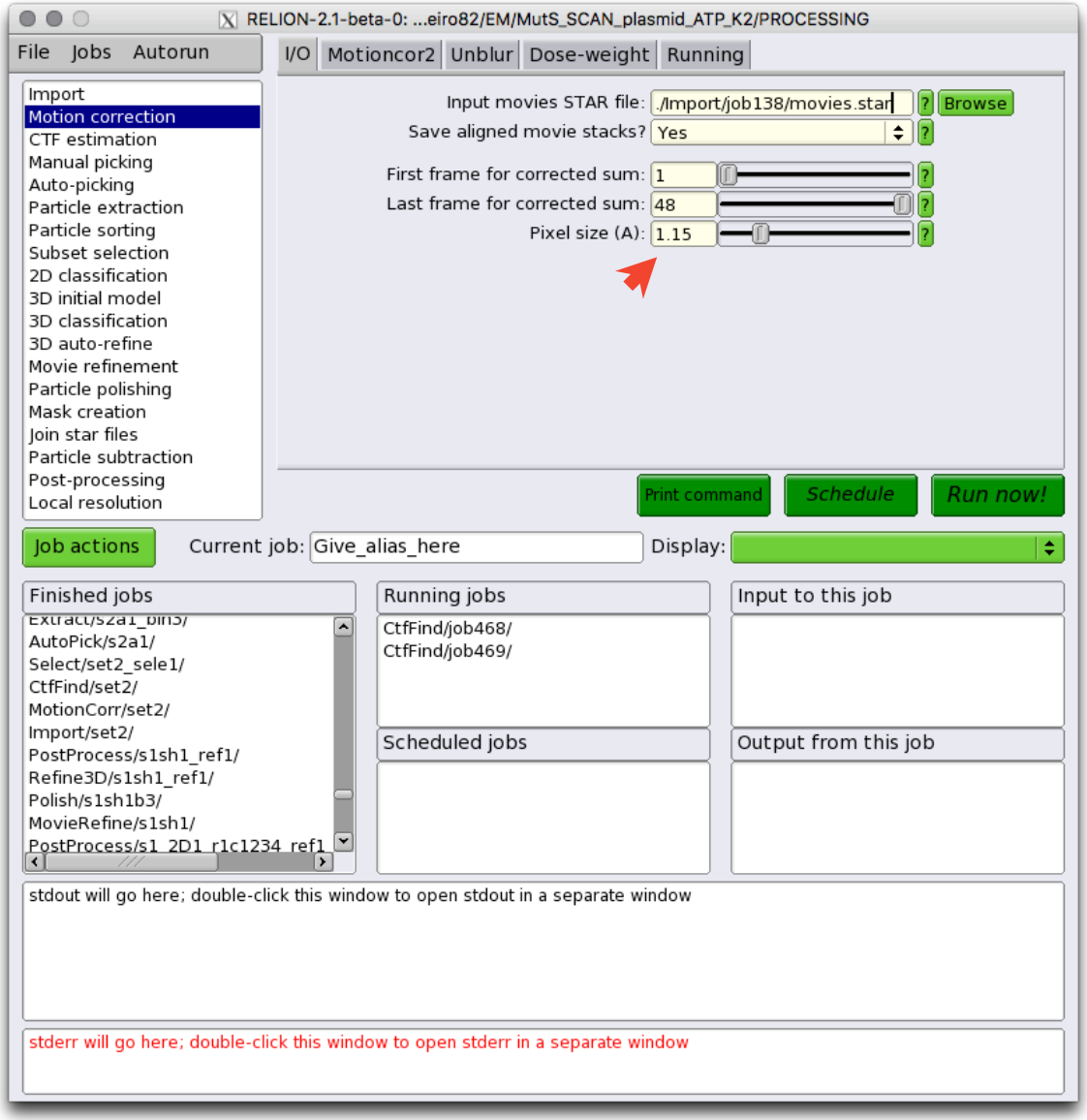
Motion Correction



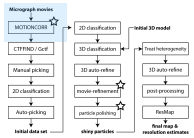




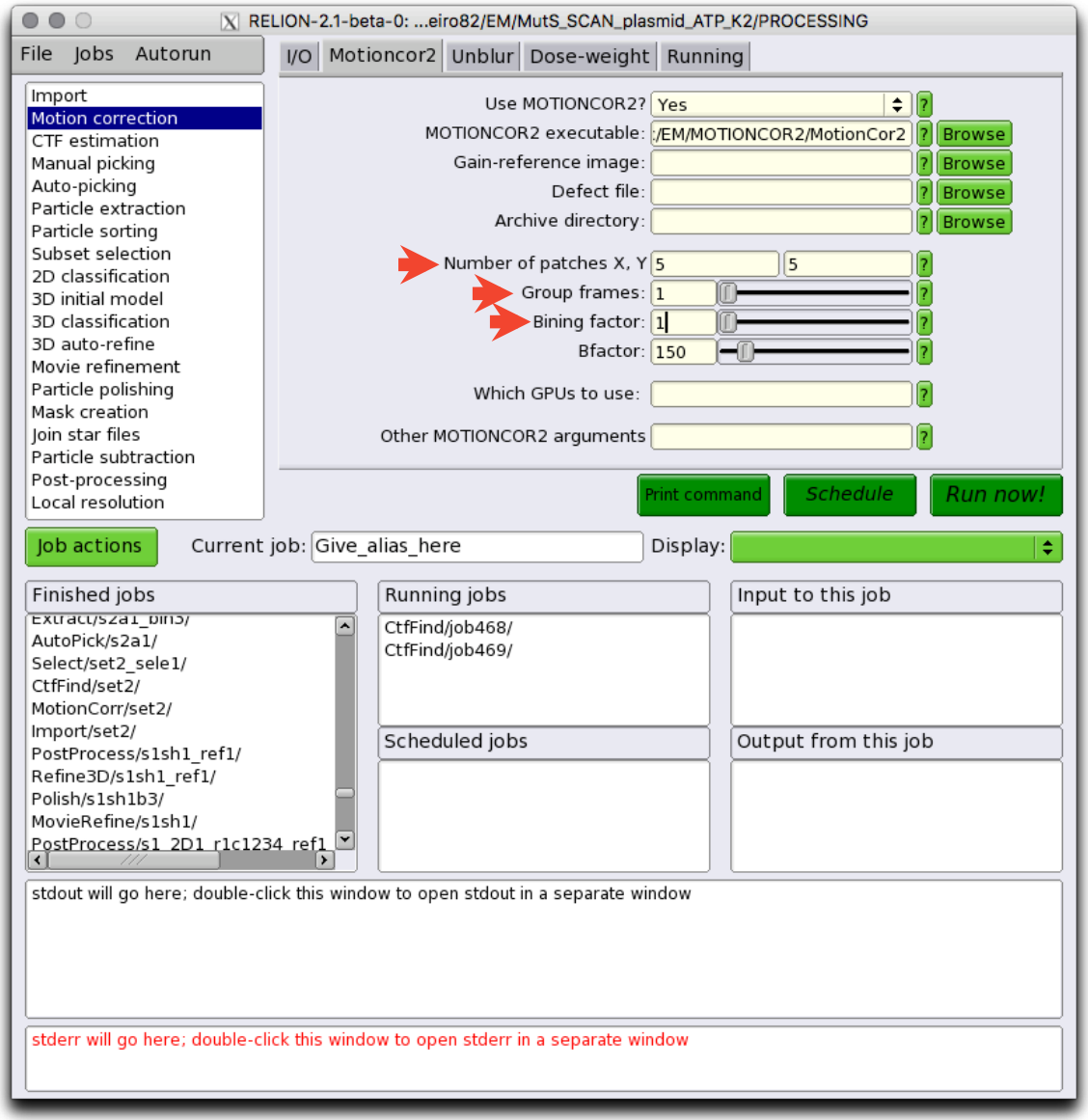
Motion Correction



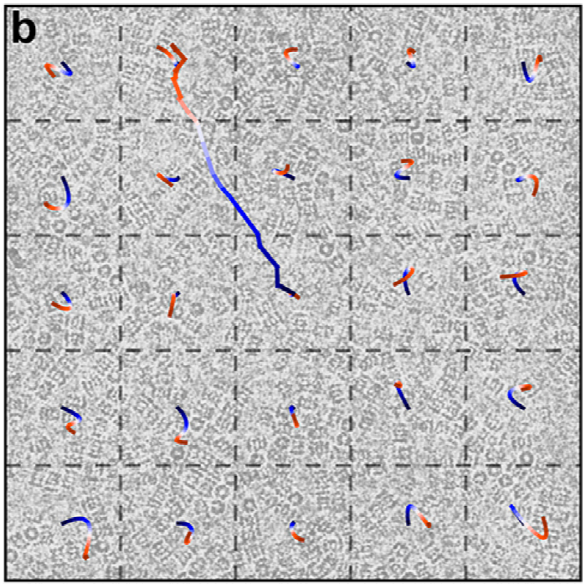
Frames
Pixel Size (—> dose weighting)



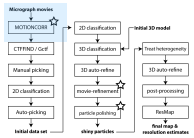
Motion Correction



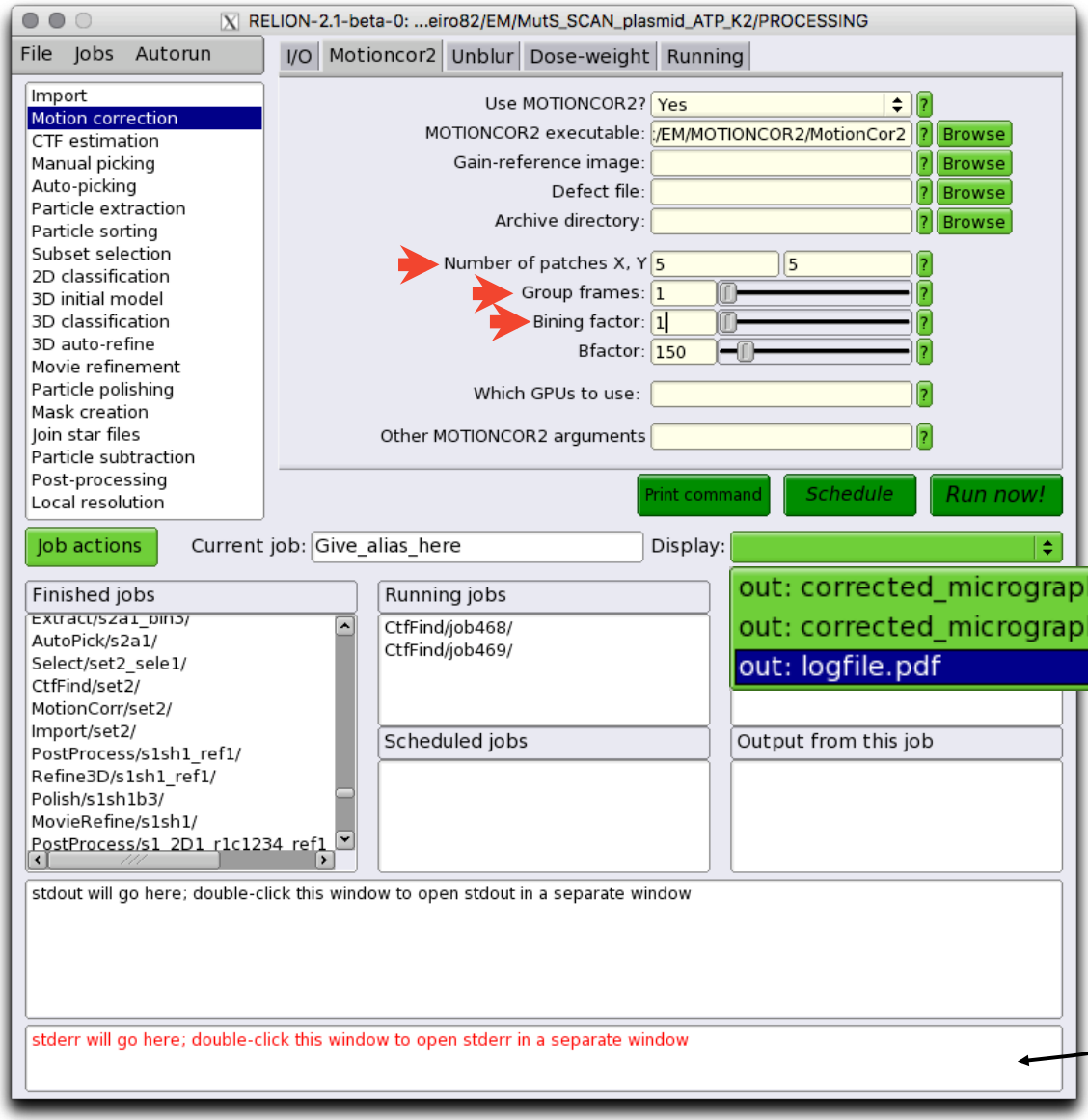
Frames
Pixel Size (—> dose weighting)
Patches
Grouping
Binning



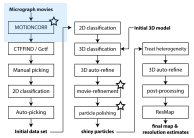
MotionCor2 patches



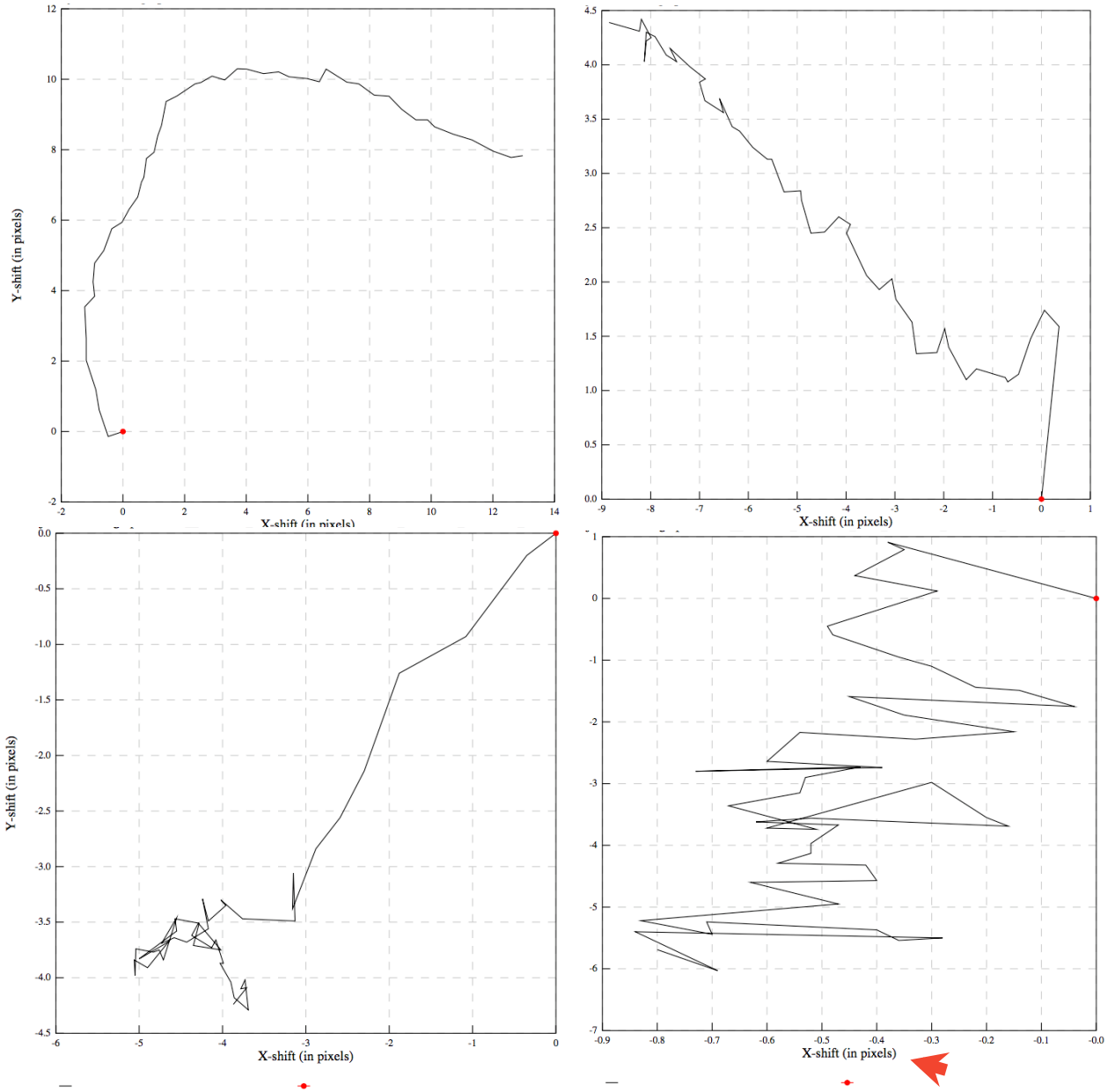
Motion Correction



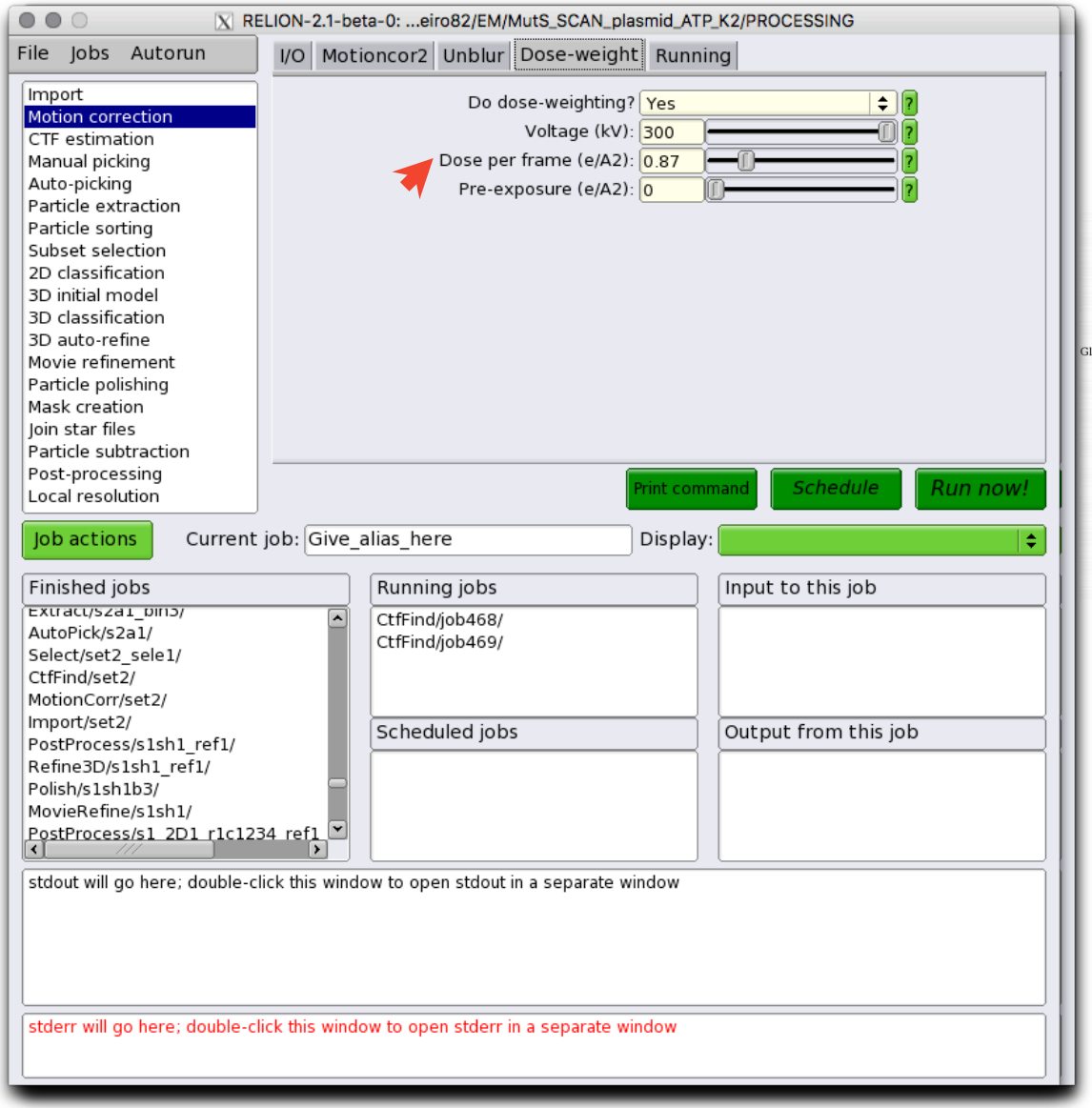
Frames
Pixel Size (—> dose weighting)
Patches
Grouping
Binning
Check your results! (errors?)



Motion Correction

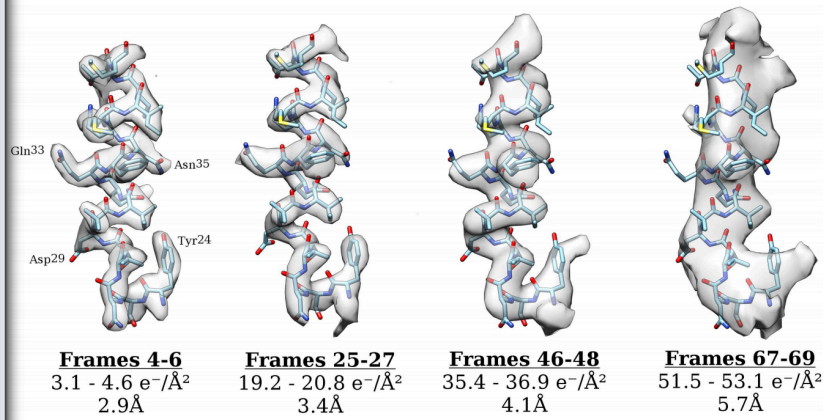


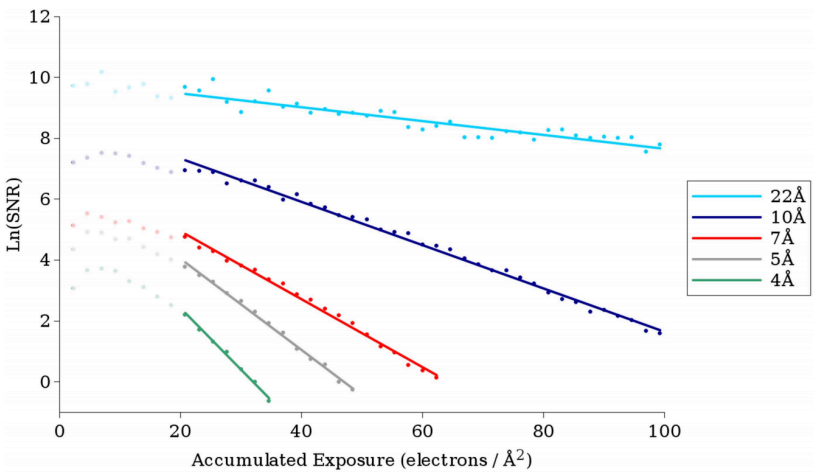
Motion Correction - Dose Weighting



Dose weighting (radiation damage)
Possibility to refine it later (Polishing)

Based on theoretical model

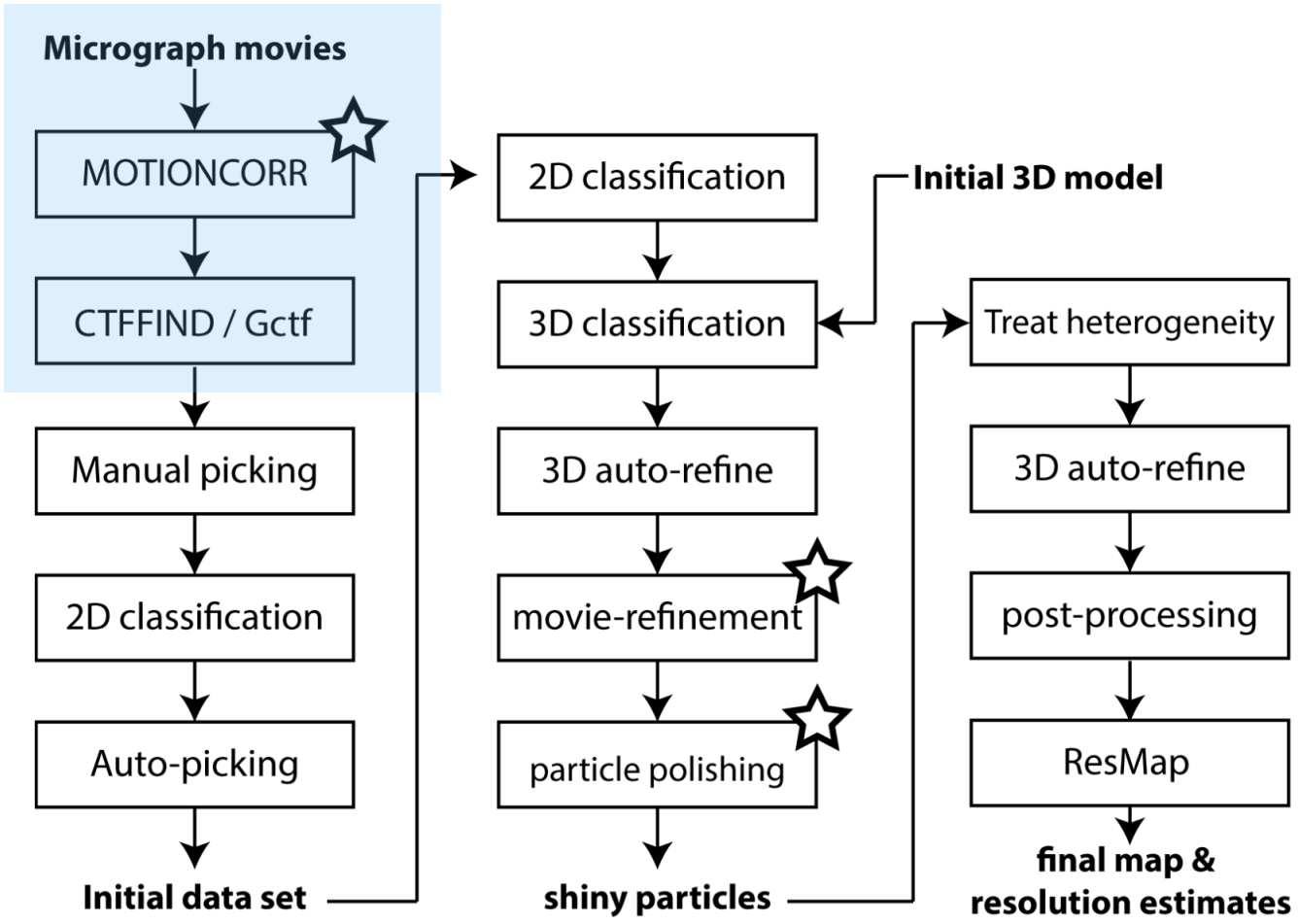


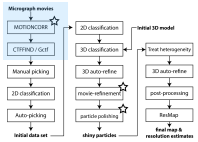


Unblur - Grant 2015 (eLife)
MotionCor2 - Zheng 2016 (Nature Methods)

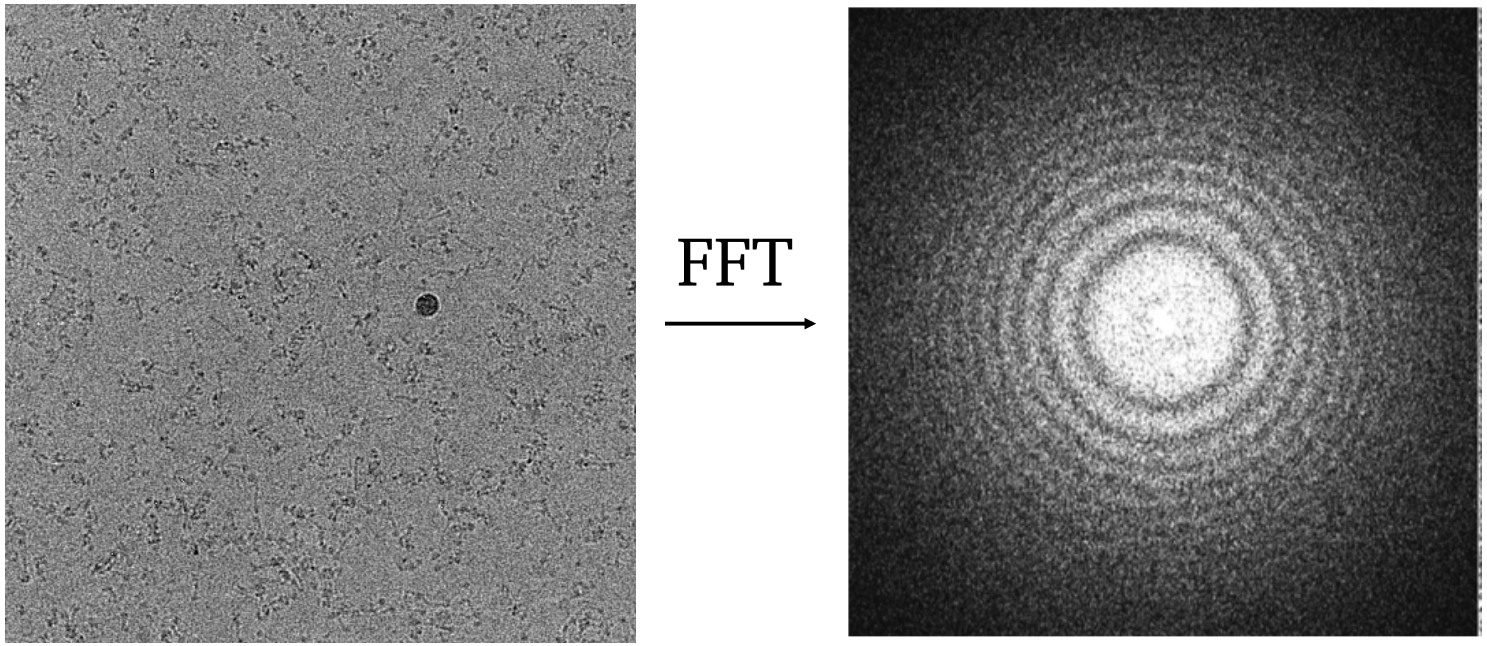
Grant 2015 (eLife)

CTF Estimation





CTF Correction

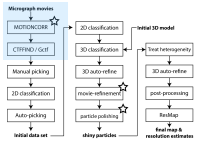


cryo-EM <— phase contrast

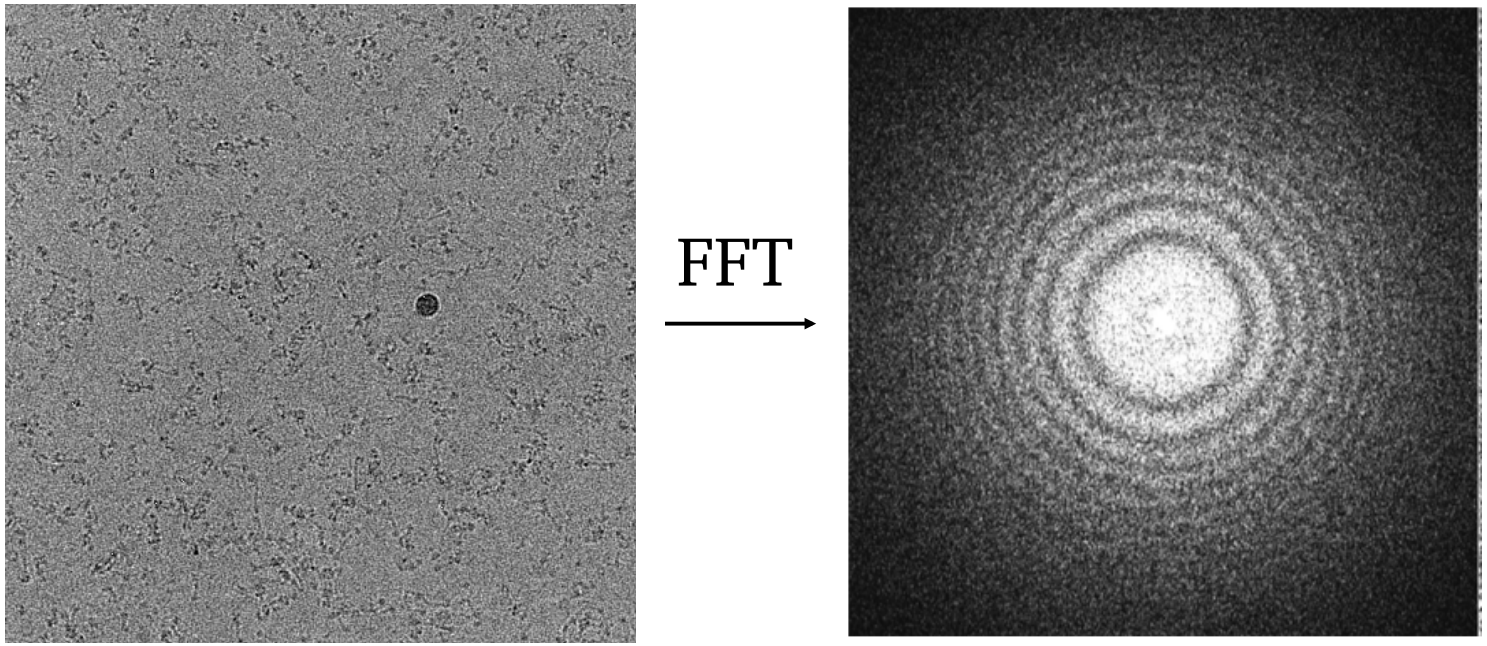
microscope aberrations + defocus increase contrast

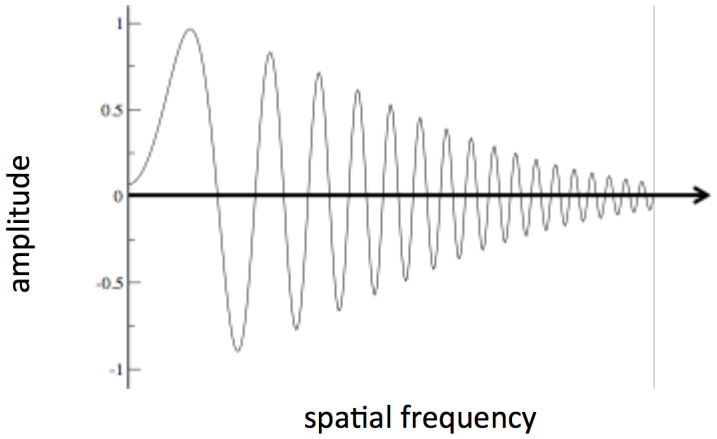
Final image is modified by these aberrations

The CTF (Contrast Transfer Function) describes these effects
(Fourier Transform of the scope's point spread function)

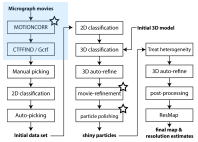


CTF Correction

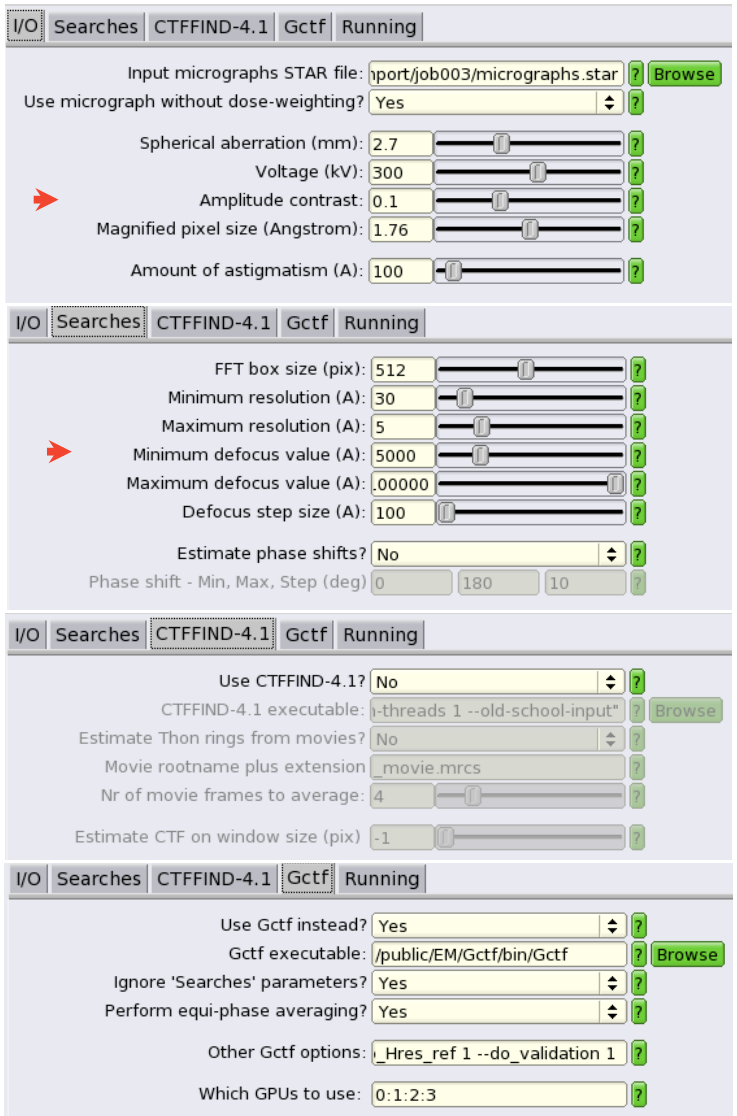




Ctffind - Niko Grigorieff (http://grigoriefflab.janelia.org/ctf) - Rohou 2015 (JSB)
Gctf - Kay Zhang (http://www.mrc-lmb.cam.ac.uk/kzhang/) - Zhang 2015 (JSB)

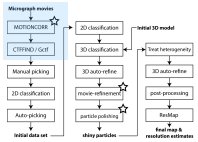


CTF Correction



Microscope parameters
Search parameters

Ctffind - Niko Grigorieff (http://grigoriefflab.janelia.org/ctf) - Rohou 2015 (JSB)
Gctf - Kay Zhang (http://www.mrc-lmb.cam.ac.uk/kzhang/) - Zhang 2015 (JSB)



CTF Correction

I/O | Searches | CTFFIND-4.1 | Gctf | Running

Input micrographs STAR file: ıport/job003/micrographs.star [Browse]
Use micrograph without dose-weighting? Yes

Spherical aberration (mm): 2.7
Voltage (kV): 300
Amplitude contrast: 0.1
Magnified pixel size (Angstrom): 1.76

Amount of astigmatism (A): 100

I/O | Searches | CTFFIND-4.1 | Gctf | Running

FFT box size (pix): 512
Minimum resolution (A): 30
Maximum resolution (A): 5
Minimum defocus value (A): 5000
Maximum defocus value (A): 50000
Defocus step size (A): 100

Estimate phase shifts? No
Phase shift - Min, Max, Step (deg) 0 180 10

I/O | Searches | CTFFIND-4.1 | Gctf | Running

Use CTFFIND-4.1? No
CTFFIND-4.1 executable: ı-threads 1 --old-school-input" [Browse]
Estimate Thon rings from movies? No
Movie rootname plus extension _movie.mrcs
Nr of movie frames to average: 4

Estimate CTF on window size (pix) -1

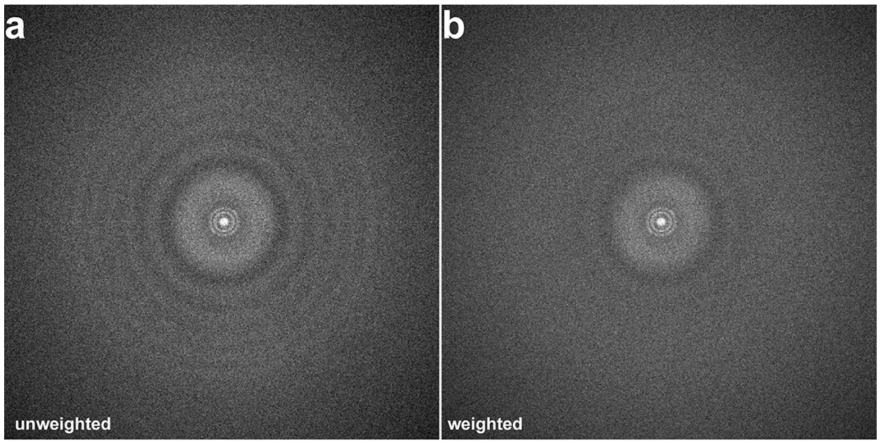
I/O | Searches | CTFFIND-4.1 | Gctf | Running

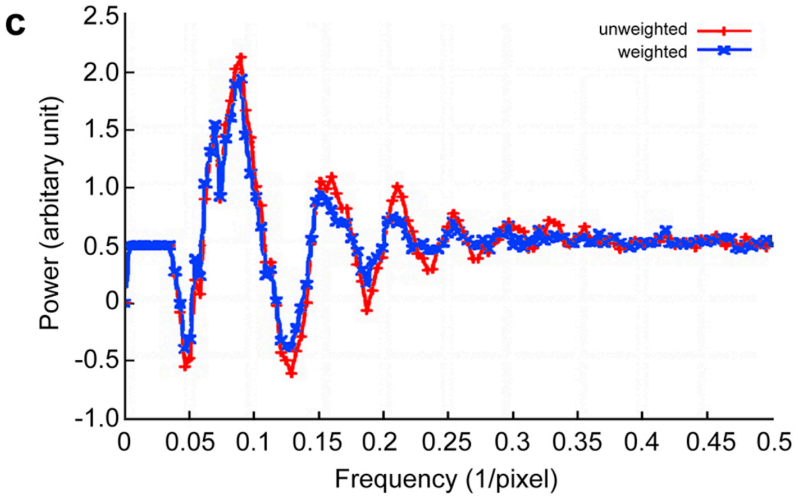
Use Gctf instead? Yes
Gctf executable: /public/EM/Gctf/bin/Gctf [Browse]
Ignore 'Searches' parameters? Yes
Perform equi-phase averaging? Yes

Other Gctf options: ı_Hres_ref 1 --do_validation 1

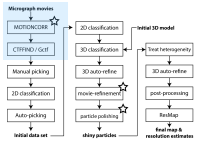
Which GPUs to use: 0:1:2:3

Microscope parameters
Search parameters





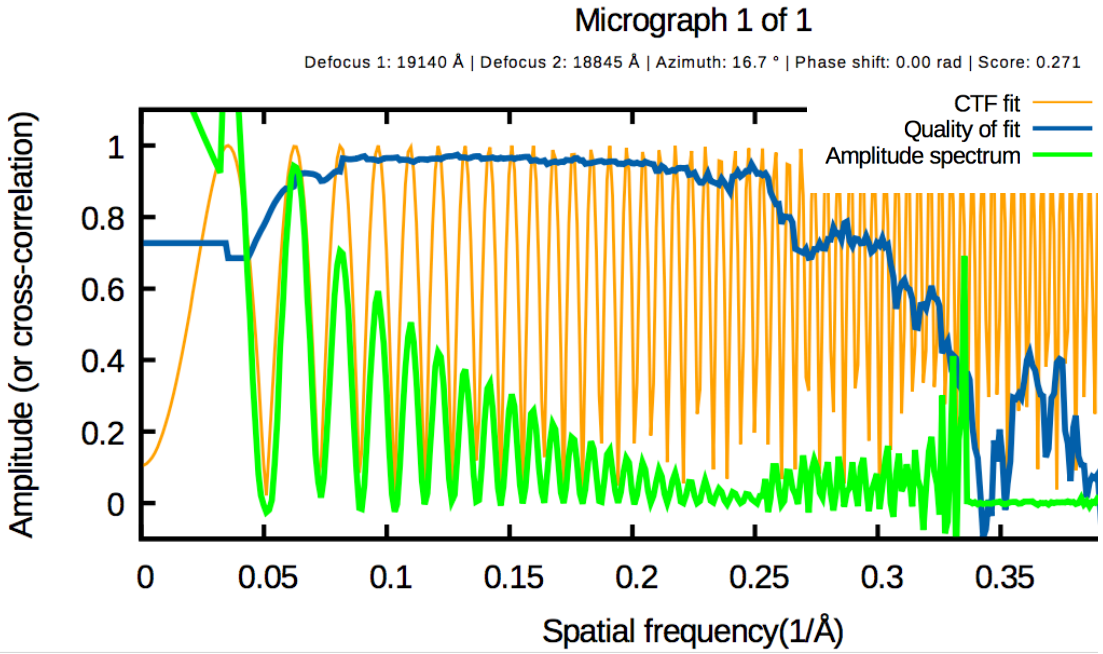
Zheng 2016 (Nature Methods)

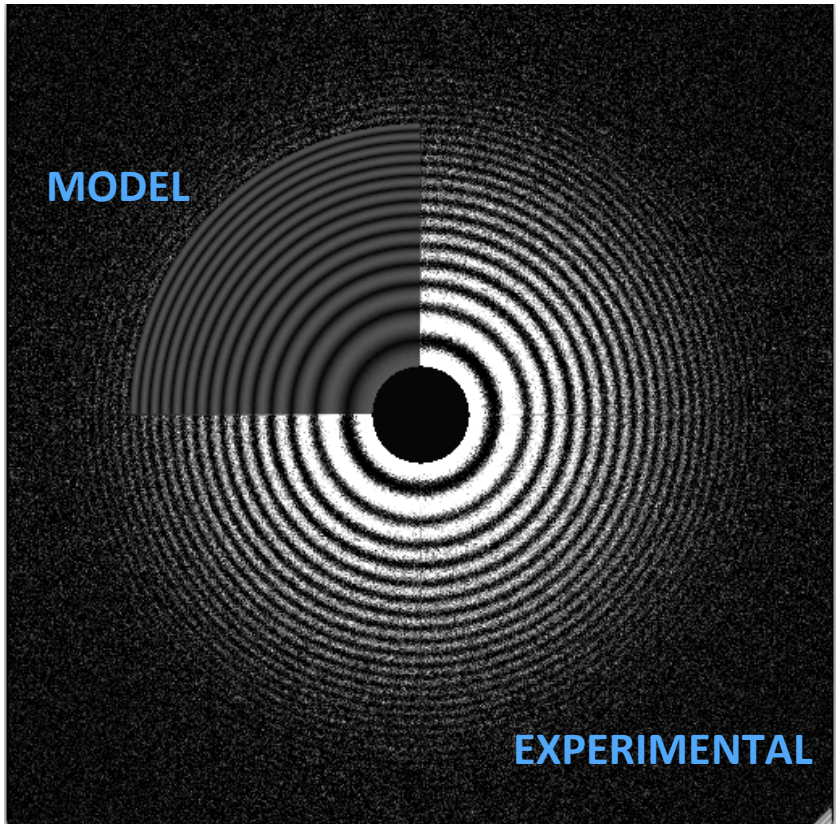


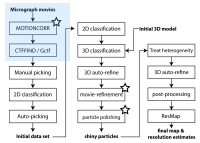
CTF Correction

Ctffind - Niko Grigorieff (http://grigoriefflab.janelia.org/ctf) - Rohou 2015 (JSB)
Gctf - Kay Zhang (http://www.mrc-lmb.cam.ac.uk/kzhang/) - Zhang 2015 (JSB)

Output from Ctffind



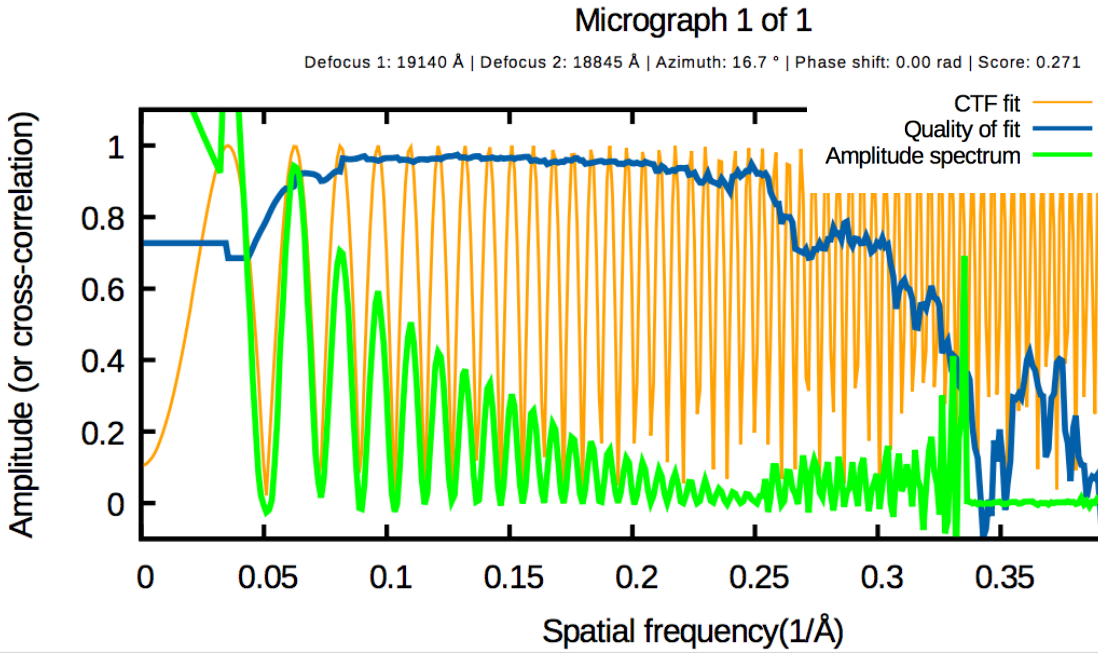


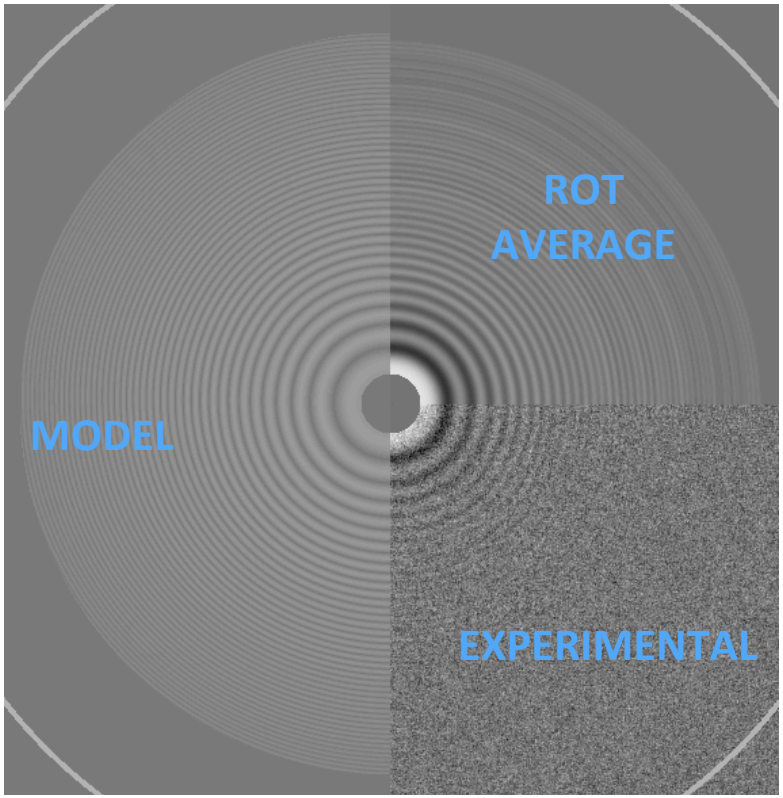


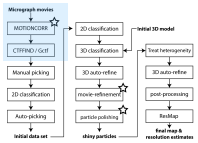
CTF Correction

Ctffind - Niko Grigorieff (http://grigoriefflab.janelia.org/ctf) - Rohou 2015 (JSB)
Gctf - Kay Zhang (http://www.mrc-lmb.cam.ac.uk/kzhang/) - Zhang 2015 (JSB)

Output from Ctffind



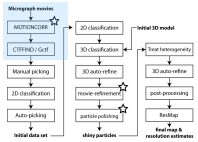




CTF Correction

```
data_

loop_
_rlnMicrographNameNoDW #1
_rlnMicrographName #2
_rlnCtfImage #3
_rlnDefocusU #4
_rlnDefocusV #5
_rlnDefocusAngle #6
_rlnVoltage #7
_rlnSphericalAberration #8
_rlnAmplitudeContrast #9
_rlnMagnification #10
_rlnDetectorPixelSize #11
_rlnCtfFigureOfMerit #12
_rlnCtfMaxResolution #13
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1616-0004_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1616-0004.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1616-0004_noDW.ctf:mrc
22994.210938 23644.460938    16.950001   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.128004     3.264000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1618-0005_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1618-0005.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1618-0005_noDW.ctf:mrc
24142.710938 24642.449219     8.200000   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.122095     3.551000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1618-0006_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1618-0006.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1618-0006_noDW.ctf:mrc
14346.419922 14696.559570    19.700001   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.123745     3.801000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1621-0007_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1621-0007.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1621-0007_noDW.ctf:mrc
25361.220703 25986.460938    18.340000   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.132673     3.040000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1622-0008_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1622-0008.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1622-0008_noDW.ctf:mrc
26172.789062 26672.929688    15.950000   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.115742     3.206000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1623-0009_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1623-0009.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1623-0009_noDW.ctf:mrc
18321.019531 18733.669922    37.980000   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.087915     3.204000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1626-0010_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1626-0010.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1626-0010_noDW.ctf:mrc
27898.339844 28468.849609     7.790000   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.104870     3.354000
MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1626-0011_noDW.mrc MotionCorr/job139/Micrographs/MutS_ATP_plasmid_20170525_1626-0011.mrc CtfFind/job156/Micrographs/MutS_ATP_plasmid_20170525_1626-0011_noDW.ctf:mrc
28418.019531 28943.410156     6.260000   300.000000     2.700000     0.070000 1.217391e+05    14.000000     0.089664     3.525000
```



CTF Correction

I/O | Searches | CTFFIND-4.1 | Gctf | Running

Input micrographs STAR file: `nport/job003/micrographs.star` Browse
Use micrograph without dose-weighting? Yes
Spherical aberration (mm): 2.7
Voltage (kV): 300
Amplitude contrast: 0.1
Magnified pixel size (Angstrom): 1.76
Amount of astigmatism (A): 100

I/O | Searches | CTFFIND-4.1 | Gctf | Running

FFT box size (pix): 512
Minimum resolution (A): 30
Maximum resolution (A): 5
Minimum defocus value (A): 5000
Maximum defocus value (A): 00000
Defocus step size (A): 100
Estimate phase shifts? No
Phase shift - Min, Max, Step (deg) 0 180 10

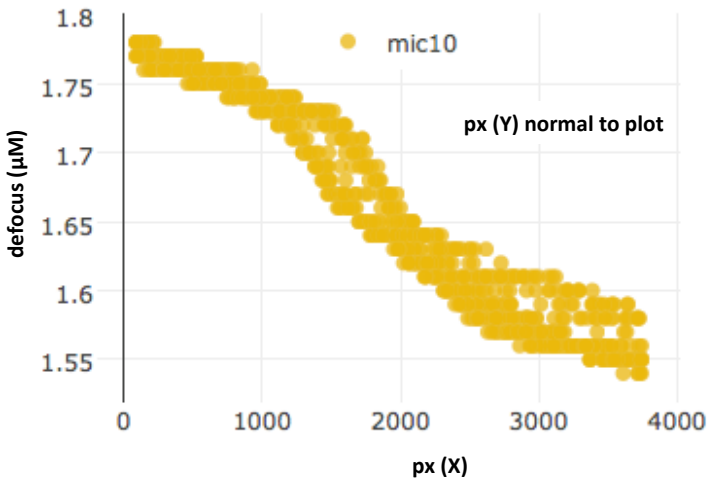
I/O | Searches | CTFFIND-4.1 | Gctf | Running

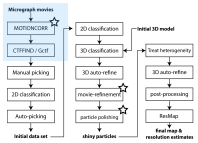
Use CTFFIND-4.1? No
CTFFIND-4.1 executable: `-threads 1 --old-school-input"` Browse
Estimate Thon rings from movies? No
Movie rootname plus extension _movie.mrcs
Nr of movie frames to average: 4
Estimate CTF on window size (pix) -1

I/O | Searches | CTFFIND-4.1 | Gctf | Running

Use Gctf instead? Yes
Gctf executable: /public/EM/Gctf/bin/Gctf Browse
Ignore 'Searches' parameters? Yes
Perform equi-phase averaging? Yes
Other Gctf options: `_Hres_ref 1 --do_validation 1`
Which GPUs to use: 0:1:2:3

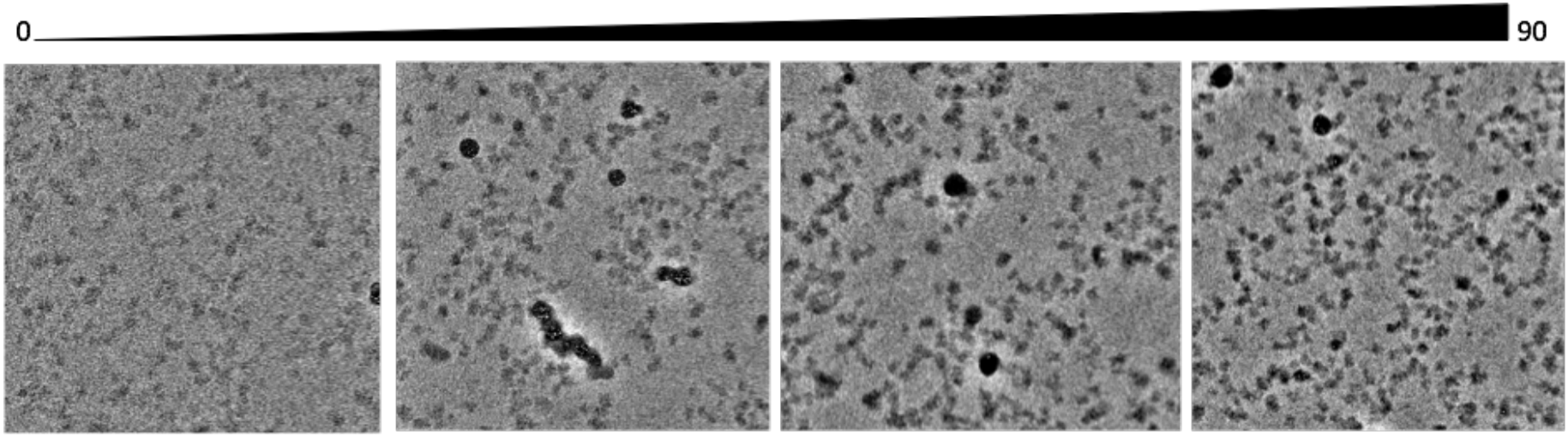
Per-Particle CTF refinement (Gctf)

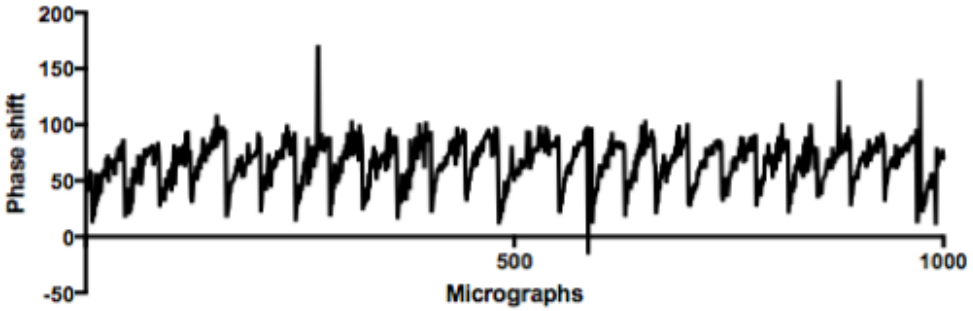


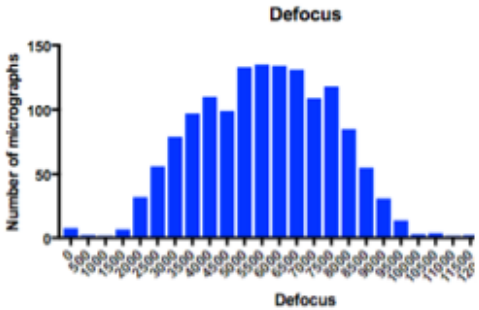


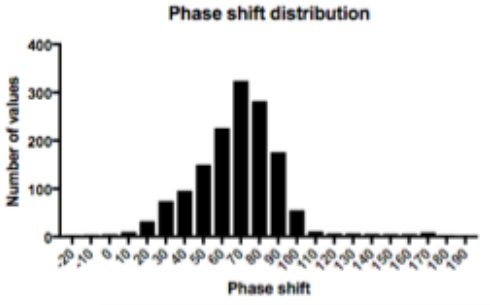
CTF Correction - Phase Plate Data

Phase Shift

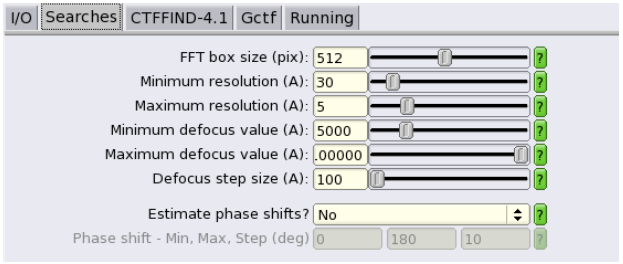








GCTF
--ac 0
--phase_shift_L 0.0
--phase_shift_H 180
--astm 3000 (tune)



Data from Javier García Nafría

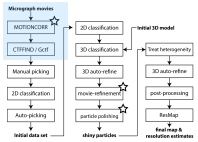
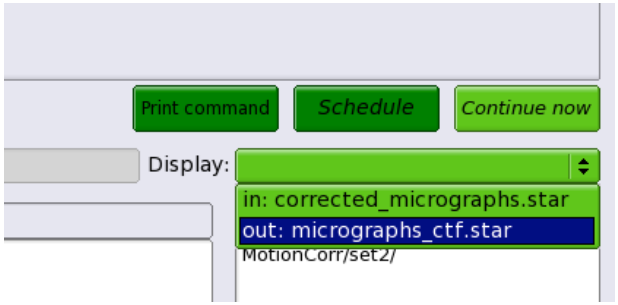
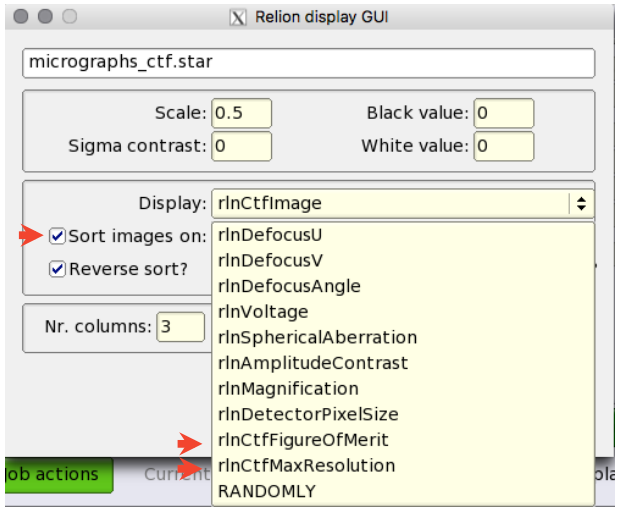
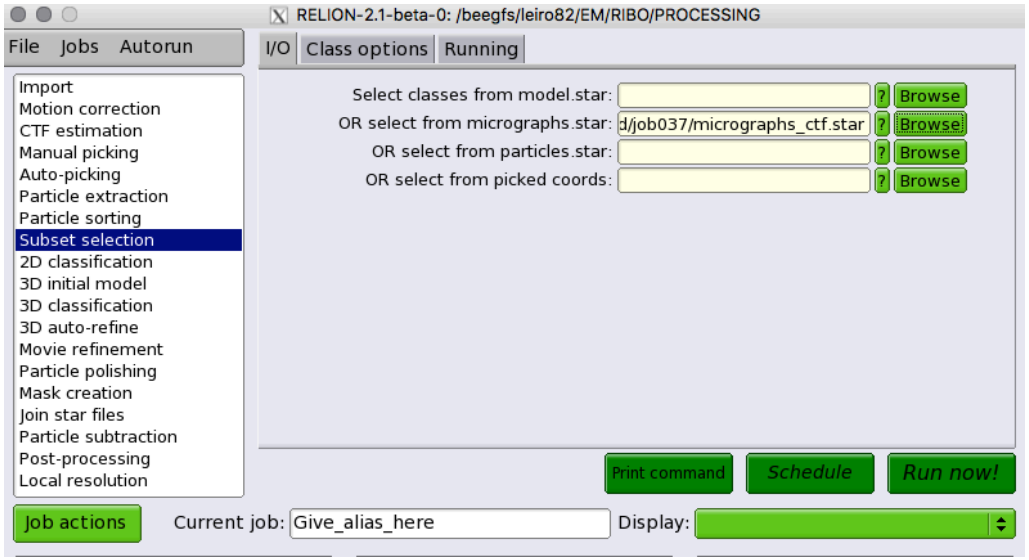


Image Selection

Let go of bad micrographs… it is hard, I know…







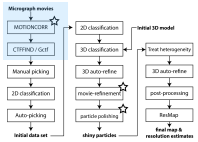
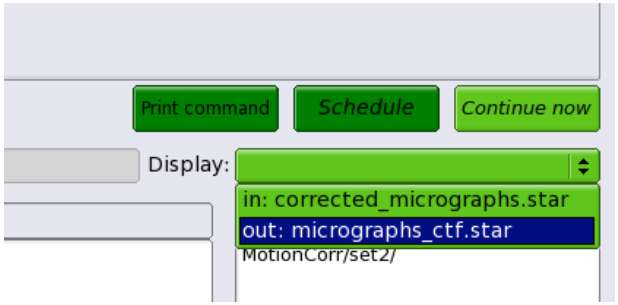
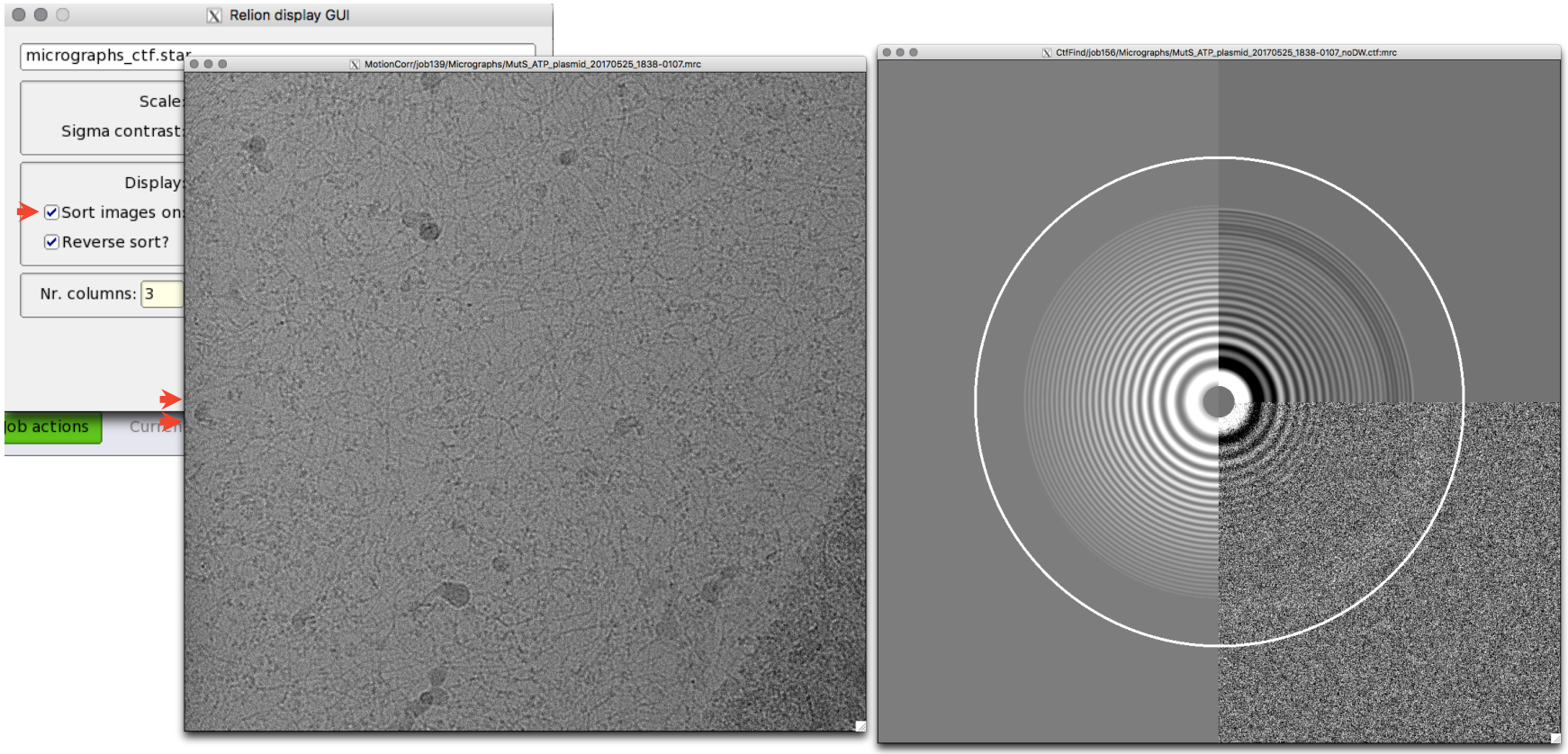


Image Selection

Let go of bad micrographs… it is hard, I know…



Ice quality
Drifty images
Check for strong astigmatism
Are there any particles??



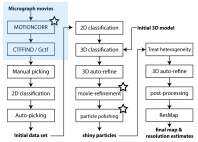
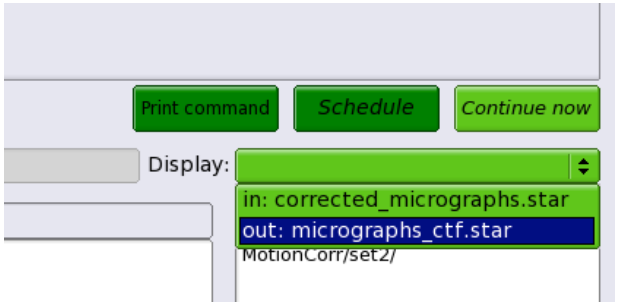
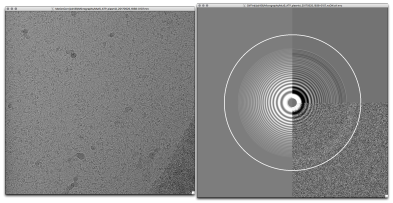
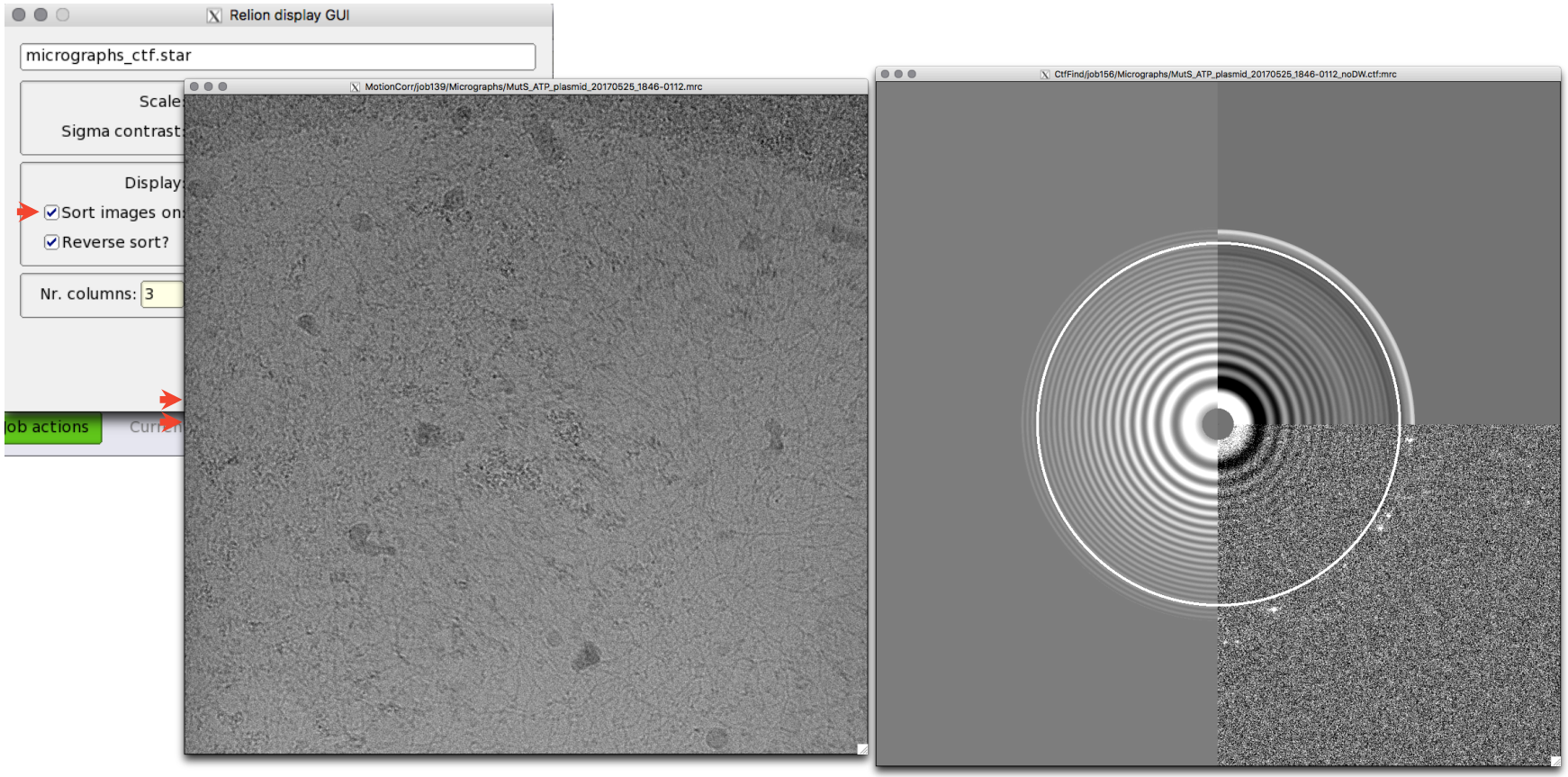


Image Selection

Let go of bad micrographs… it is hard, I know…







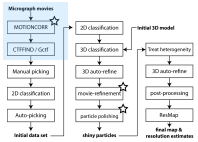
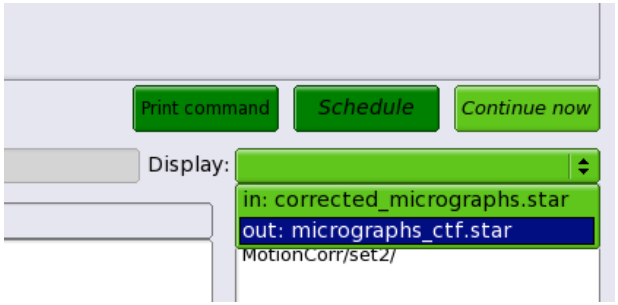
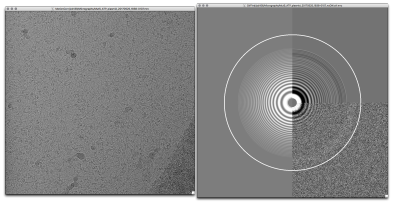
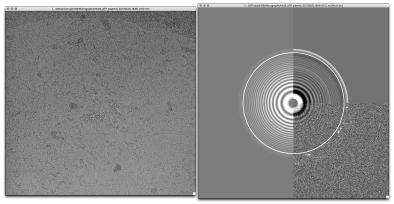


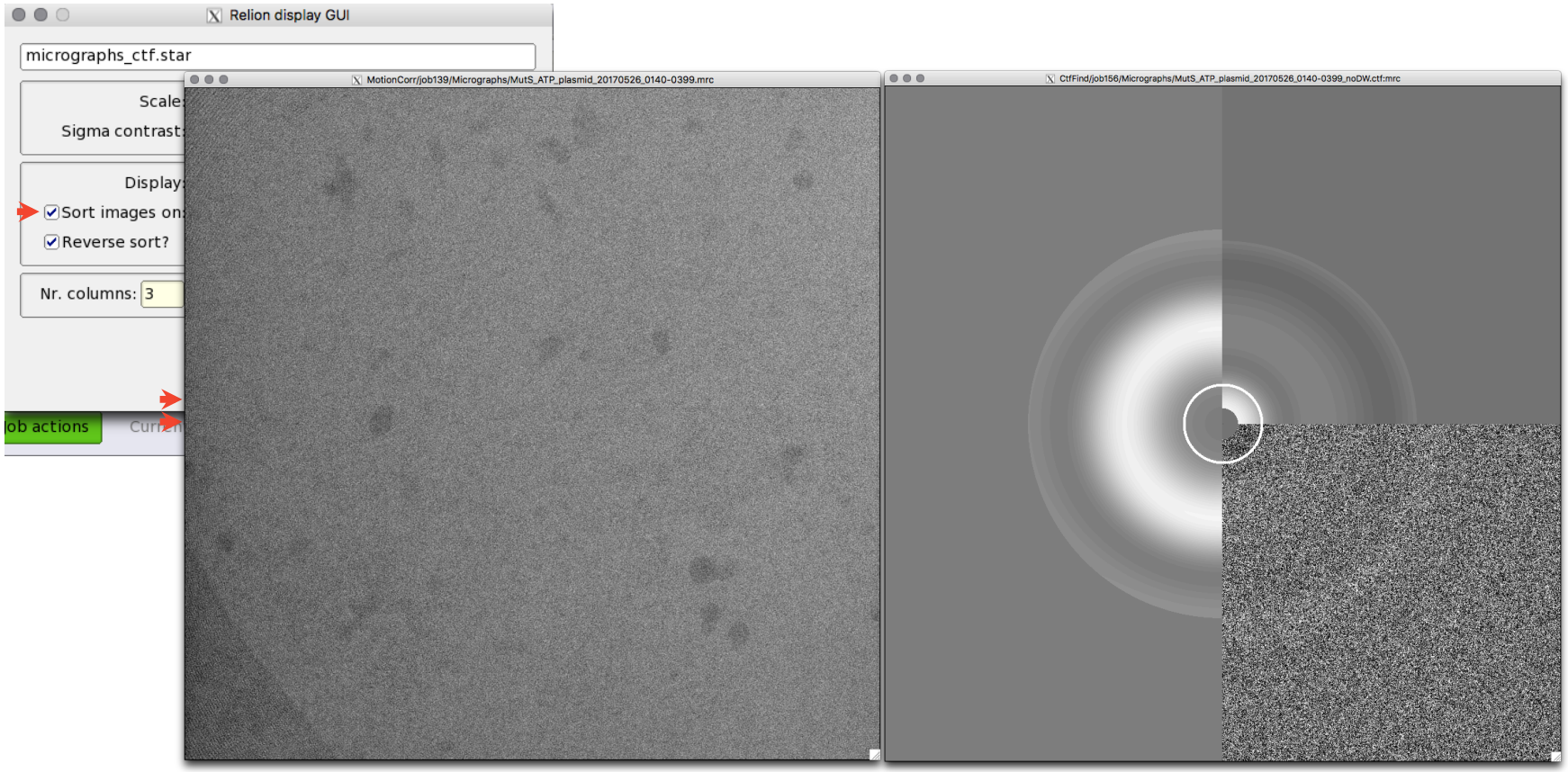
Image Selection

Let go of bad micrographs… it is hard, I know…









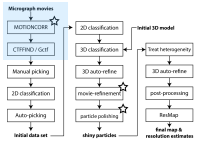
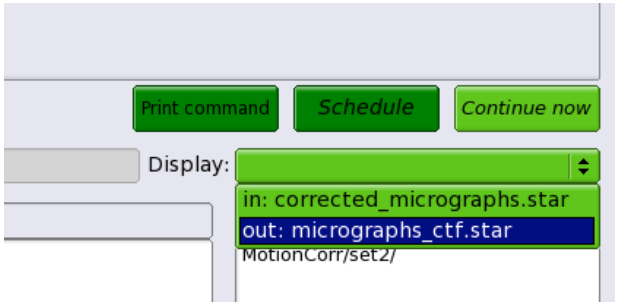
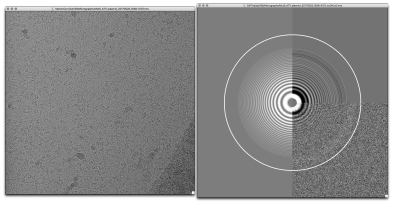
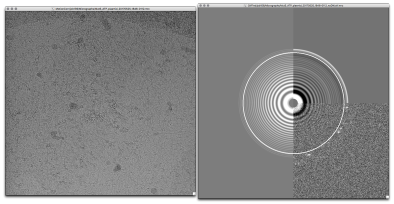


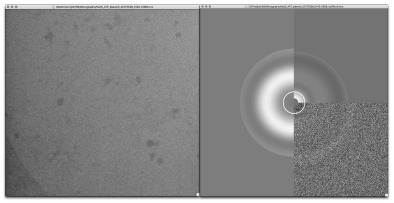
Image Selection

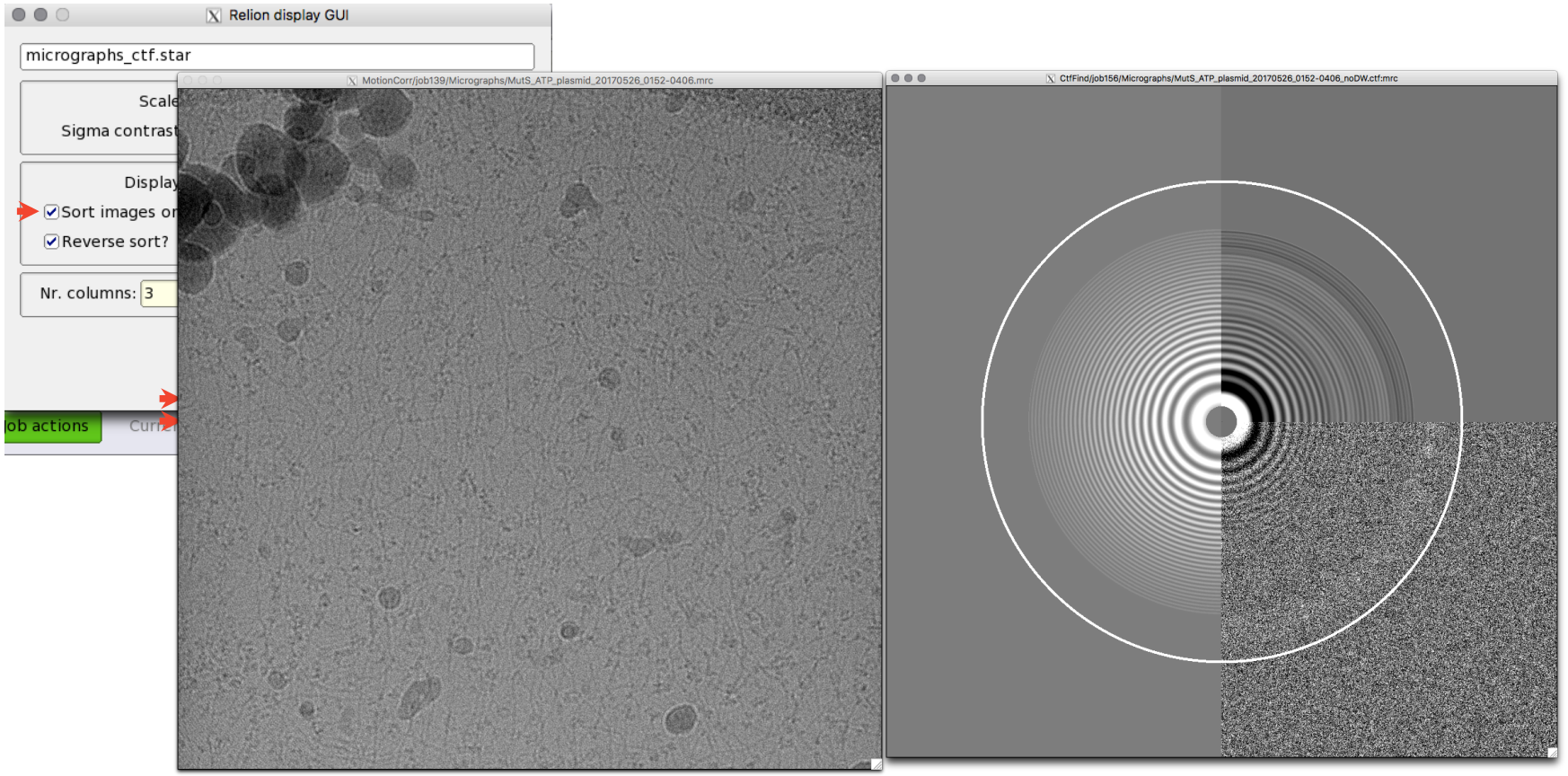
Let go of bad micrographs… it is hard, I know…











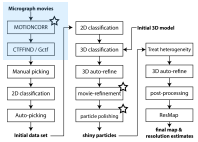
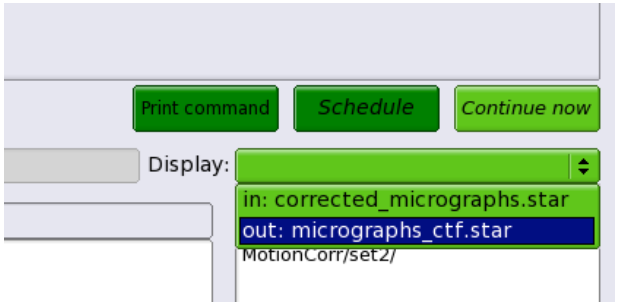
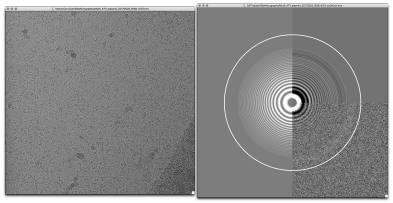
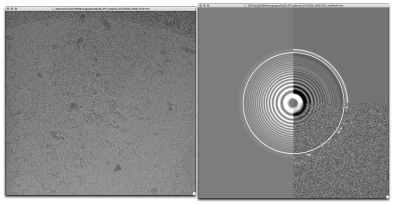


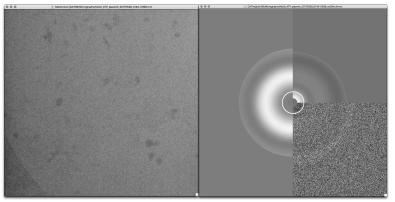
Image Selection

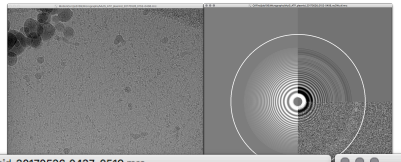
Let go of bad micrographs… it is hard, I know…

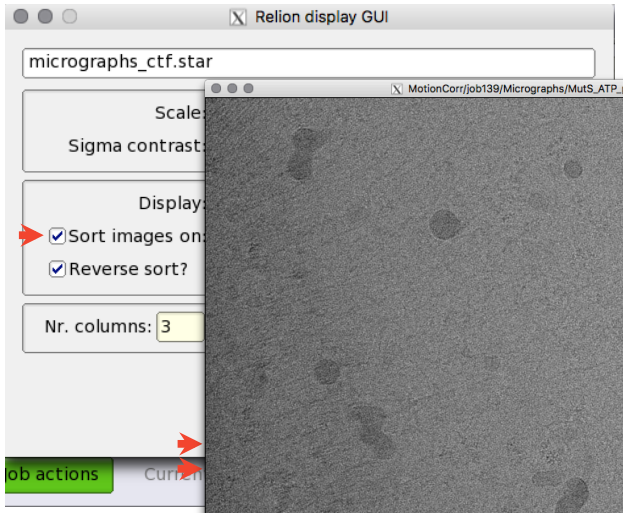


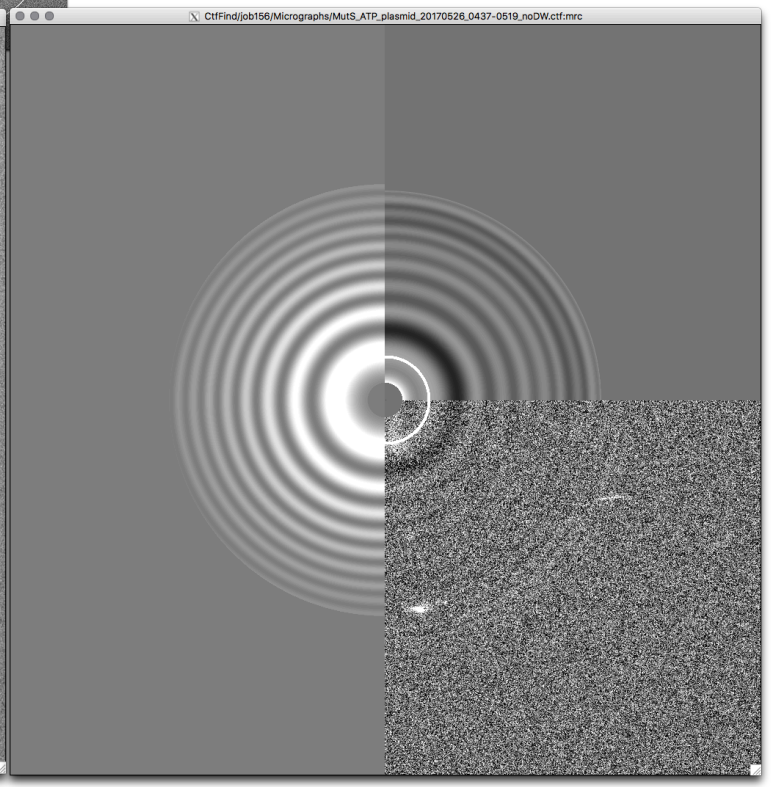












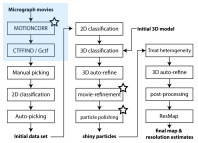
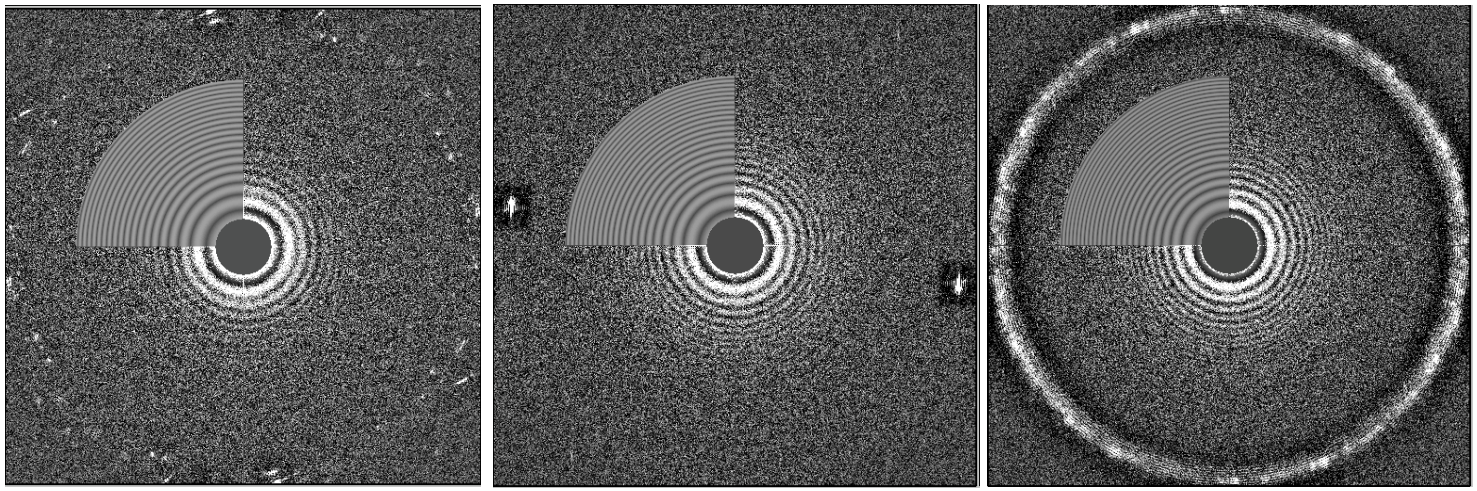


Image Selection

Let go of bad micrographs… it is hard, I know…

Ice quality
Drifty images
Check for strong astigmatism
Are there any particles??

Ice



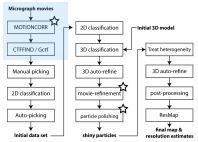
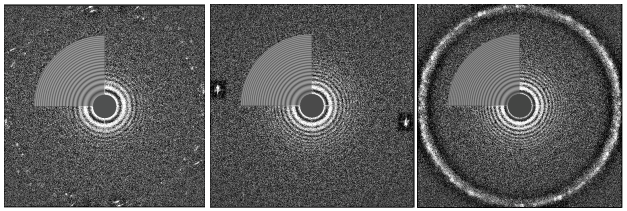


Image Selection

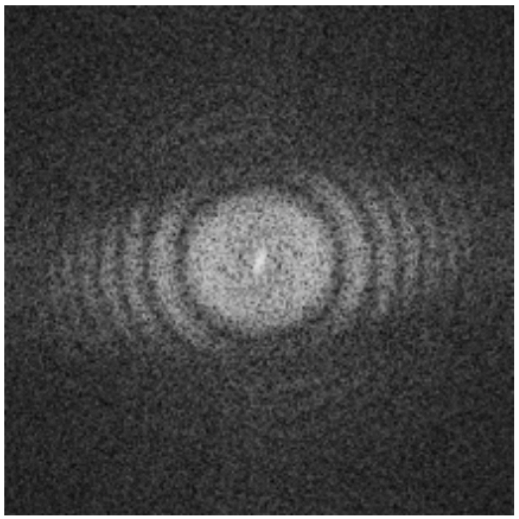
Let go of bad micrographs… it is hard, I know…

Ice quality
Drifty images
Check for strong astigmatism
Are there any particles??

Ice



Drift



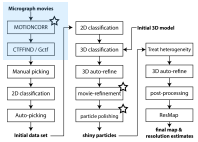
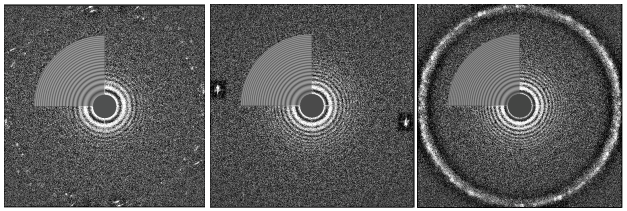


Image Selection

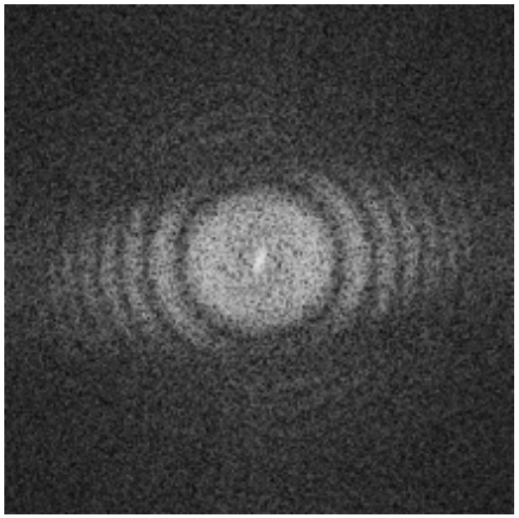
Let go of bad micrographs… it is hard, I know…

Ice quality
Drifty images
Check for strong astigmatism
Are there any particles??

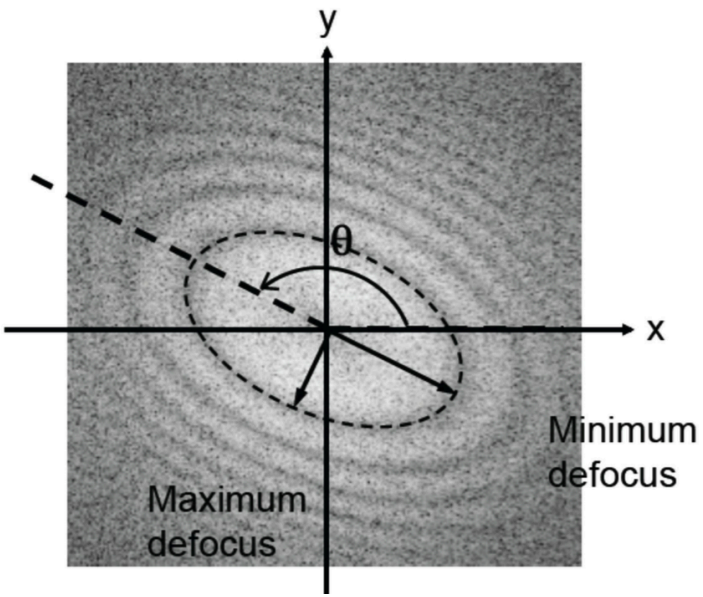
Ice



Drift



astigmatism



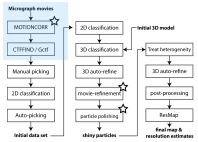


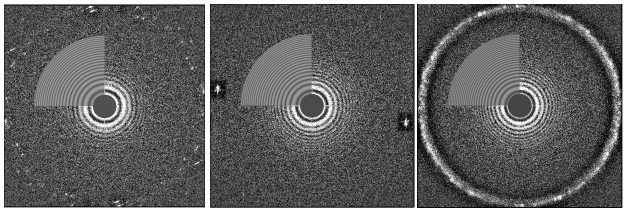
Image Selection

Let go of bad micrographs… it is hard, I know…

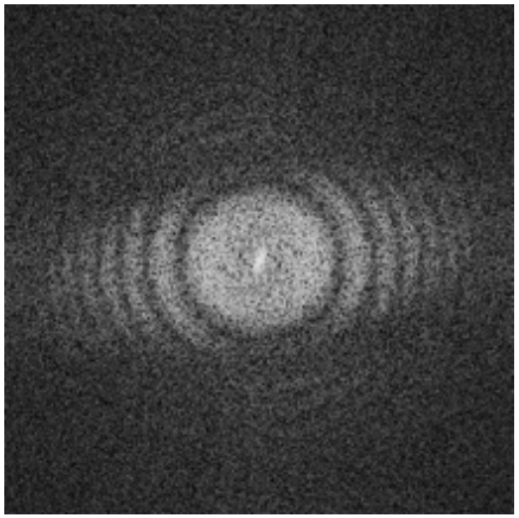
Ice quality
Drifty images
Check for strong astigmatism
Are there any particles??

Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise (Henderson 2013 PNAS)

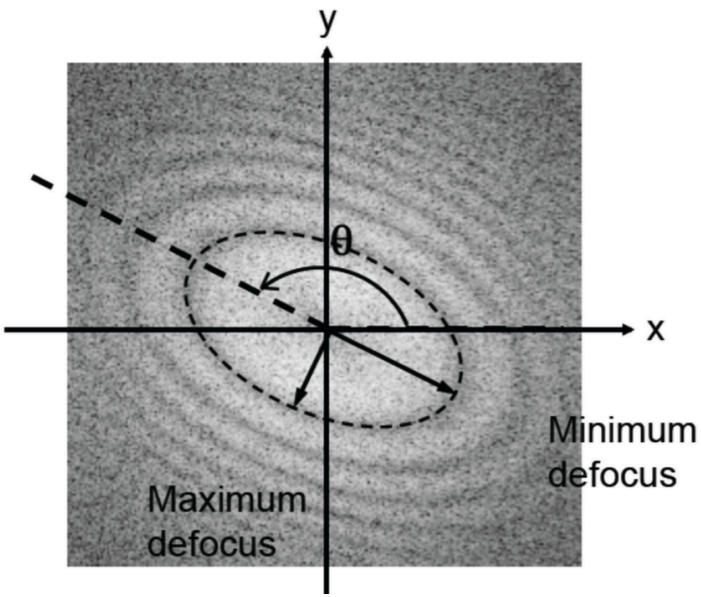
Ice



Drift

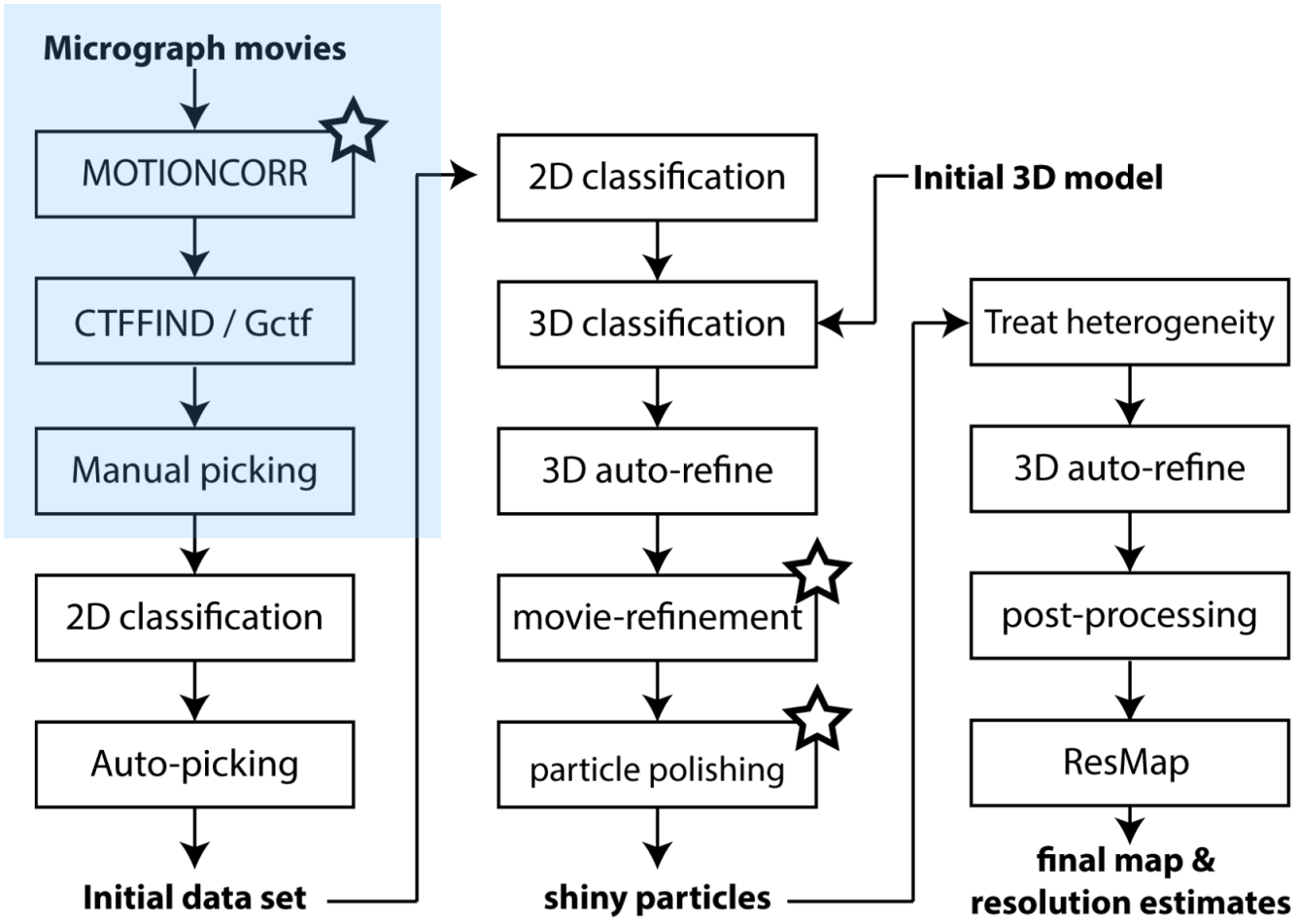


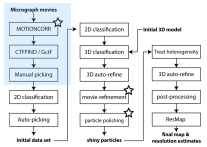
astigmatism



"Einstein from noise"

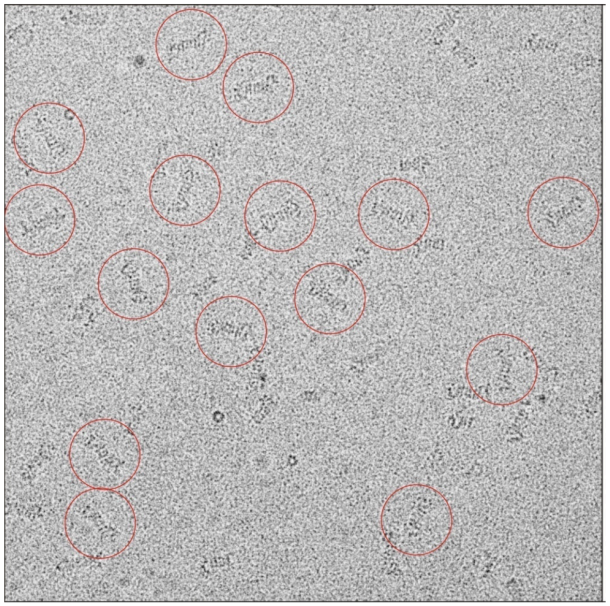
Particle Picking

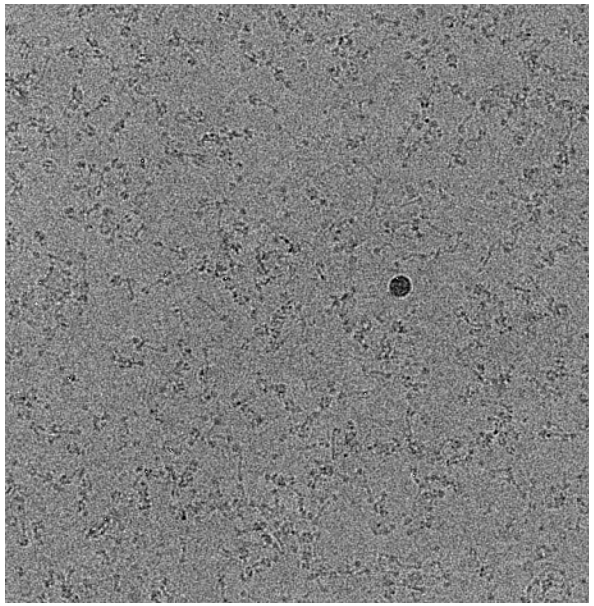


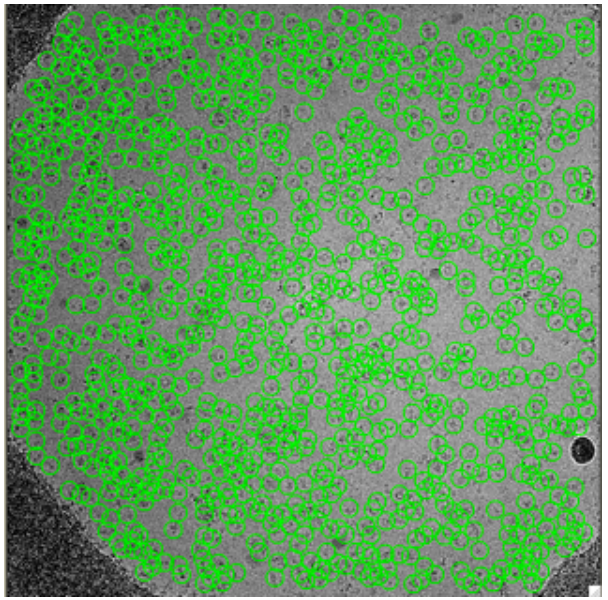


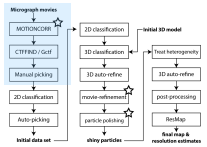
Particle Picking

Manual Picking





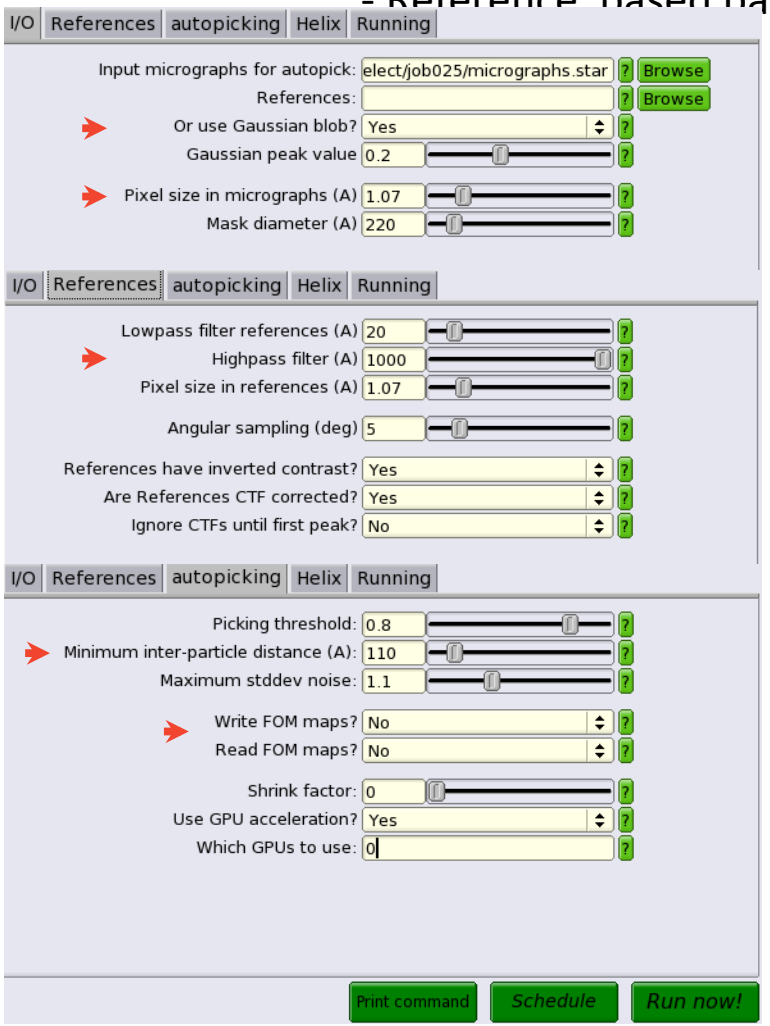


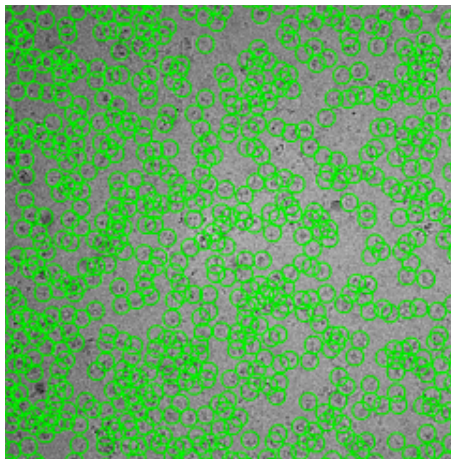


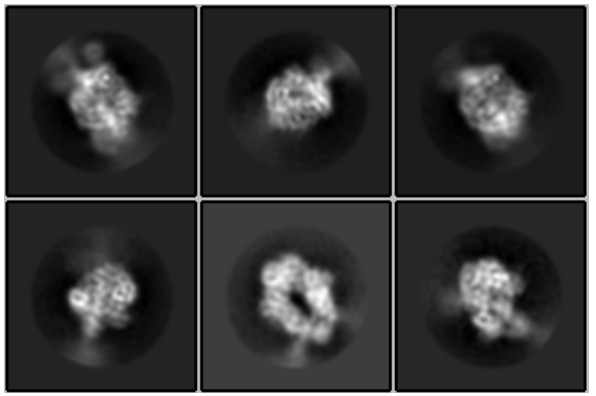
Particle Picking

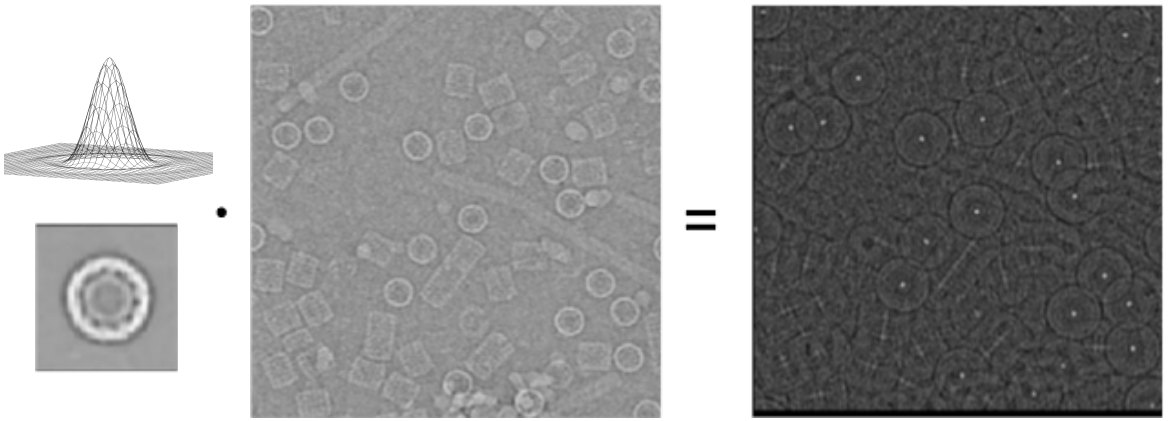
Automated Picking

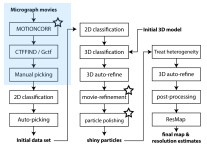
- There are many packages for automatic particle picking!
- Many support reference-free particle picking (Gautomach, RELION)
- Reference-based particle picking (RELION, EMAN, findEM, etc.)







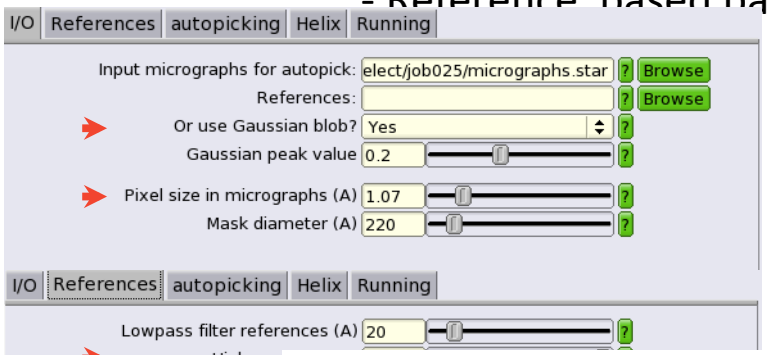


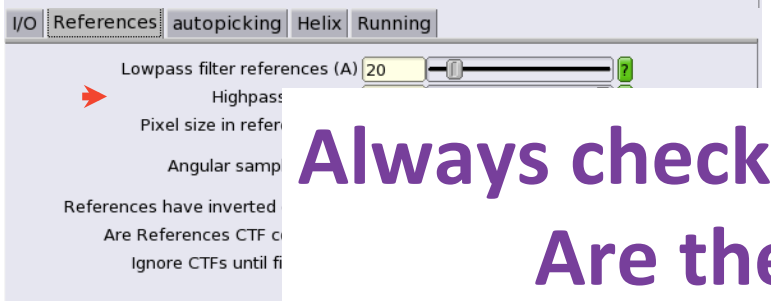


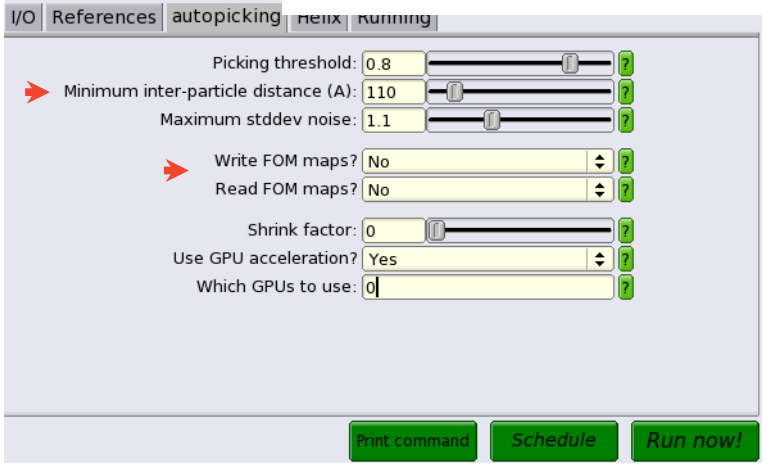
Particle Picking

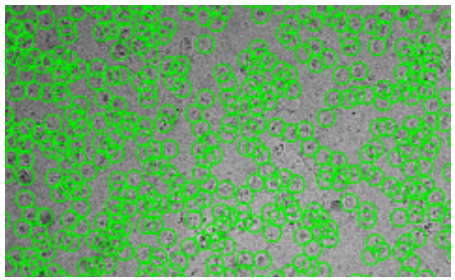
Automated Picking

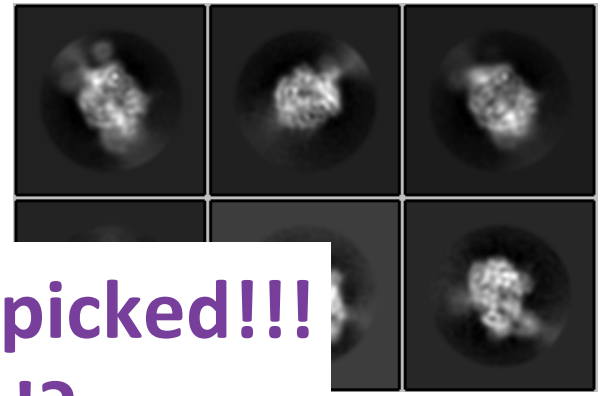
- There are many packages for automatic particle picking!
- Many support reference-free particle picking (Gautomach, RELION)
- Reference-based particle picking (RELION, EMAN, findEM, etc.)





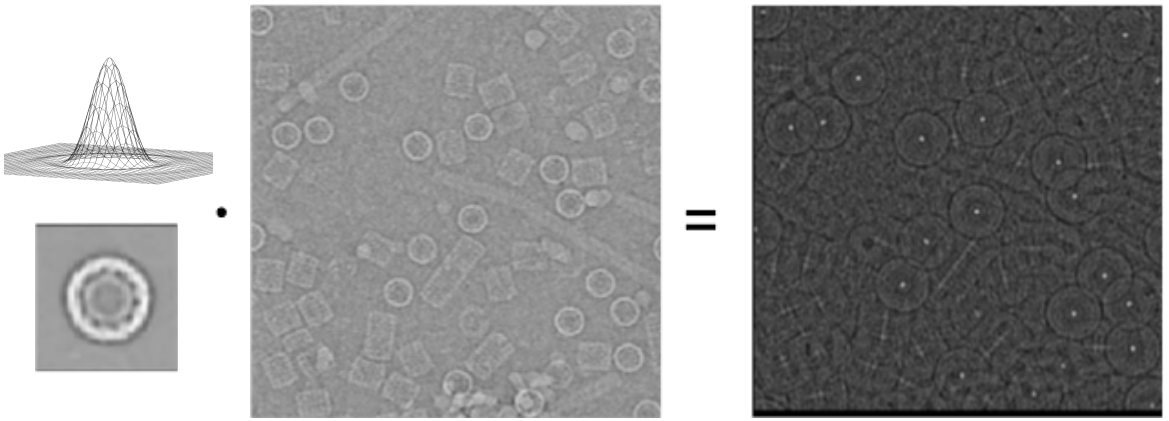


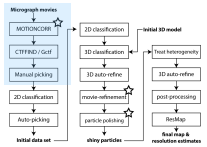




Always check what has been picked!!!

Are there any particles!?

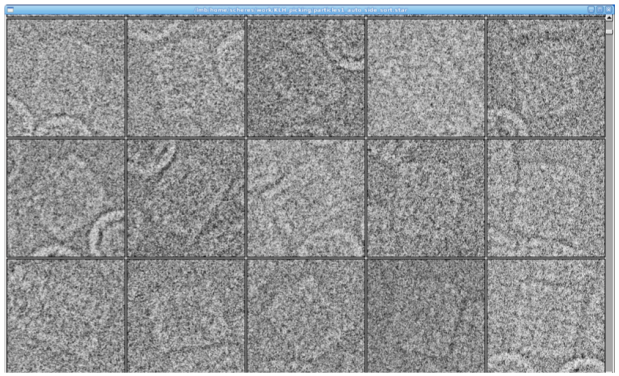




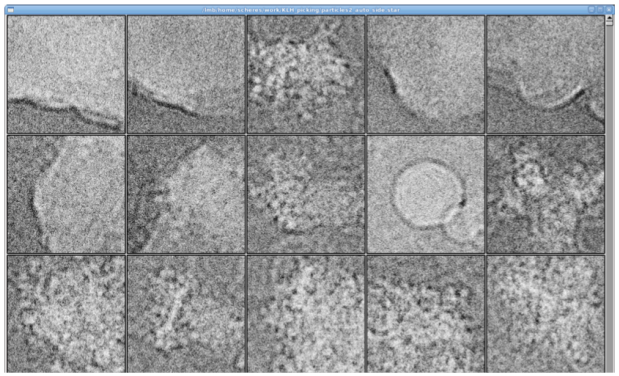
Particle Picking

Always check what has been picked!!! Are there any particles!?

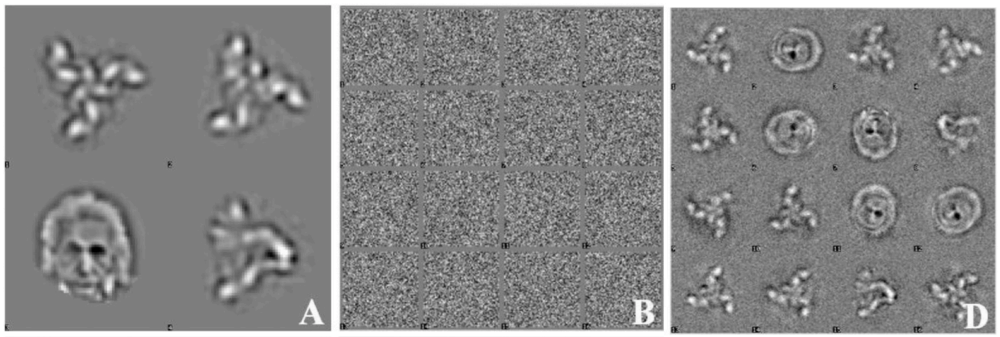
Relion uses a sorting algorithm to assist on this



low Z-scores



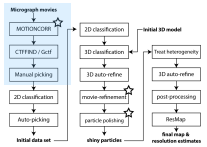
high Z-scores



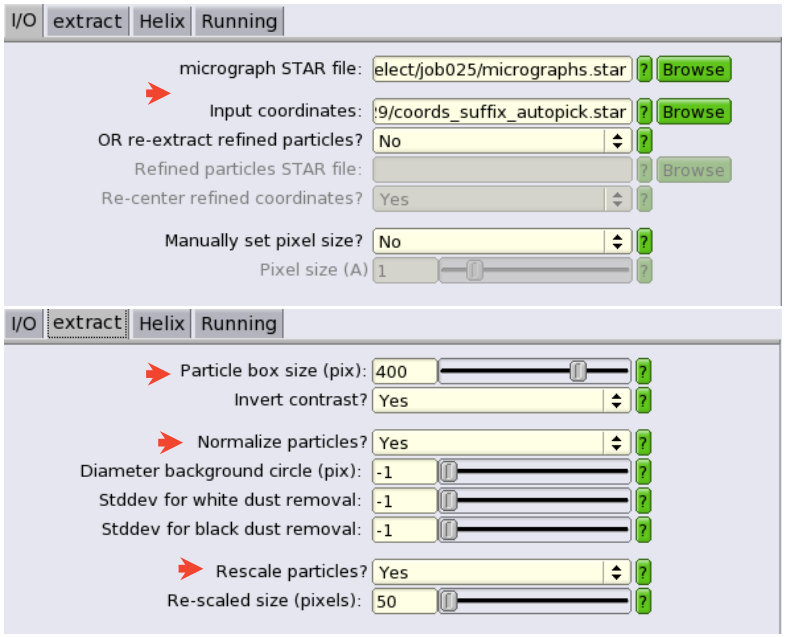
van Heel, 2013 (PNAS)

"Einstein from noise"

Henderson 2013 (PNAS)



Particle Extraction



"Rule of thumb…"

Mask = Max-particle-distance * 1.25

Box = Max-particle-distance * 1.5-2

BUT

Information is displaced with defocus!

Big boxes take longer to process & use more memory

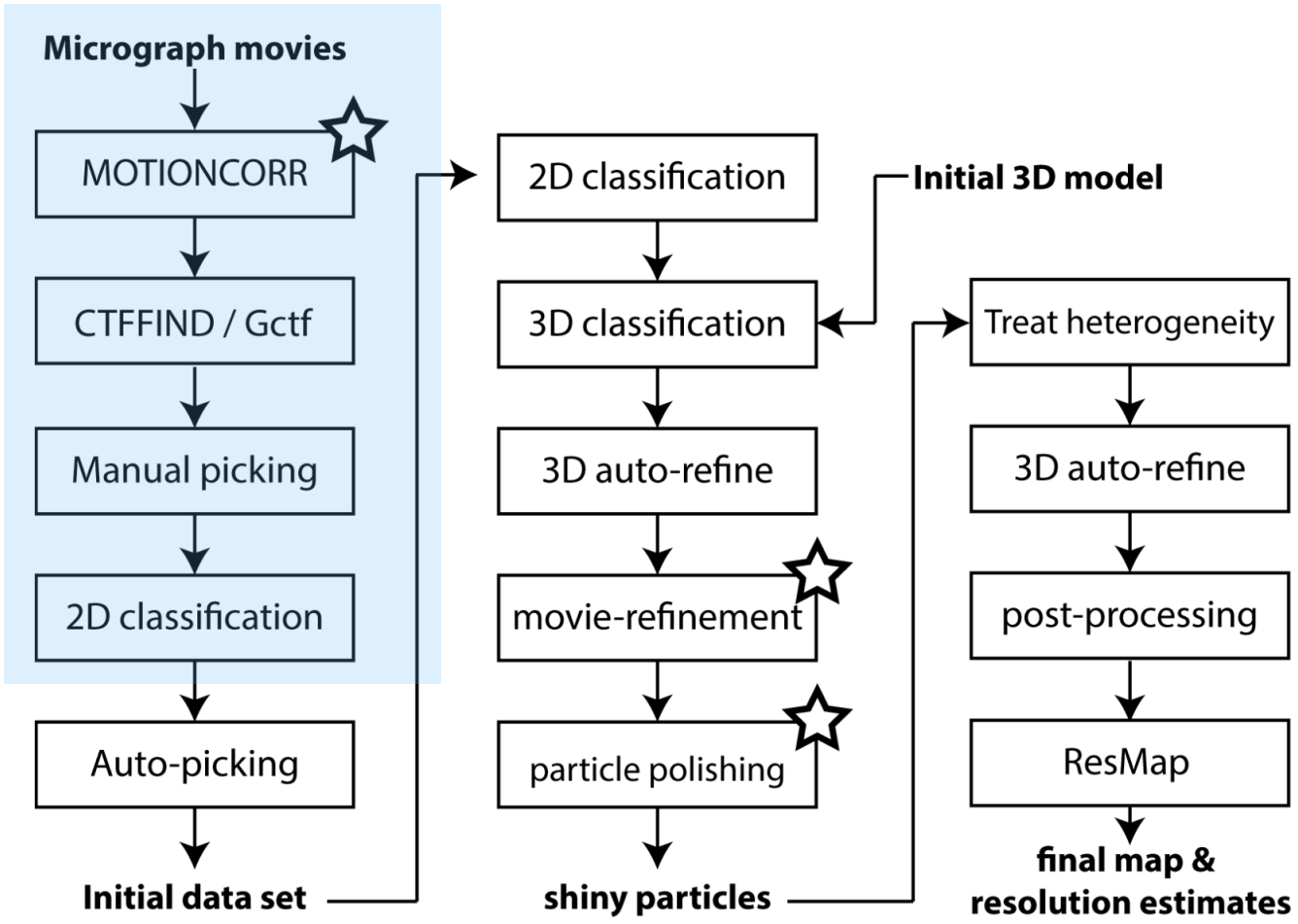
Make a reasonable box initially and play with it for your final reconstruction

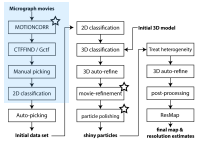
Very high defocus —> might require bigger box earlier!

Rosenthal & Henderson 2003 (JMB)

An effect of defocus (ΔF) is to displace image spacings of resolution d by a distance λΔF/d; where λ is the electron wavelength. In order to include all the information, the raw image box size was chosen to be D + 2R; where D is the particle diameter and R is the displacement expected for the most defocused image at the maximum resolution spacing expected in the reconstruction

2D Classification





2D Classification

I/O | CTF | Optimisation | Sampling | Helix | Compute | Running

Input images STAR file: Extract/job084/particles.star [?] Browse
Continue from here: p074/run_it004_optimiser.star [?] Browse

I/O | CTF | Optimisation | Sampling | Helix | Compute | Running

Do CTF-correction? Yes [?]
Have data been phase-flipped? No [?]
Ignore CTFs until first peak? No [?]

I/O | CTF | Optimisation | Sampling | Helix | Compute | Running

Number of classes: 200 [?]
Regularisation parameter T: 2 [?]
Number of iterations: 25 [?]
Use subsets for initial updates? Yes [?]
Initial subset size: 10000 [?]
Number of subset updates: 3 [?]
Mask diameter (A): 290 [?]
Mask individual particles with zeros? Yes [?]
Limit resolution E-step to (A): 6 [?]

I/O | CTF | Optimisation | Sampling | Helix | Compute | Running

Perform image alignment? Yes [?]
In-plane angular sampling: 6 [?]
Offset search range (pix): 5 [?]
Offset search step (pix): 1 [?]

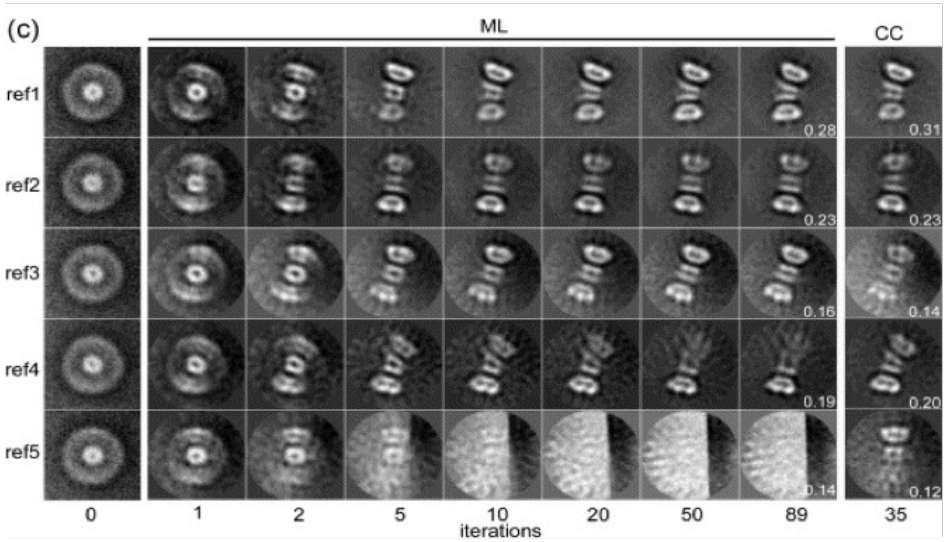
Print command | Schedule | Run now!

Use 2D classification to asses the quality of your dataset

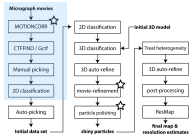
and to cleanup false positives from picking

Number of classes
CTF correction
Mask
Resolution limit

Reference-free 2D class averaging



Scheres et al (2005) J.Mol.Biol.



2D Classification

Use 2D classification to asses the quality of your dataset

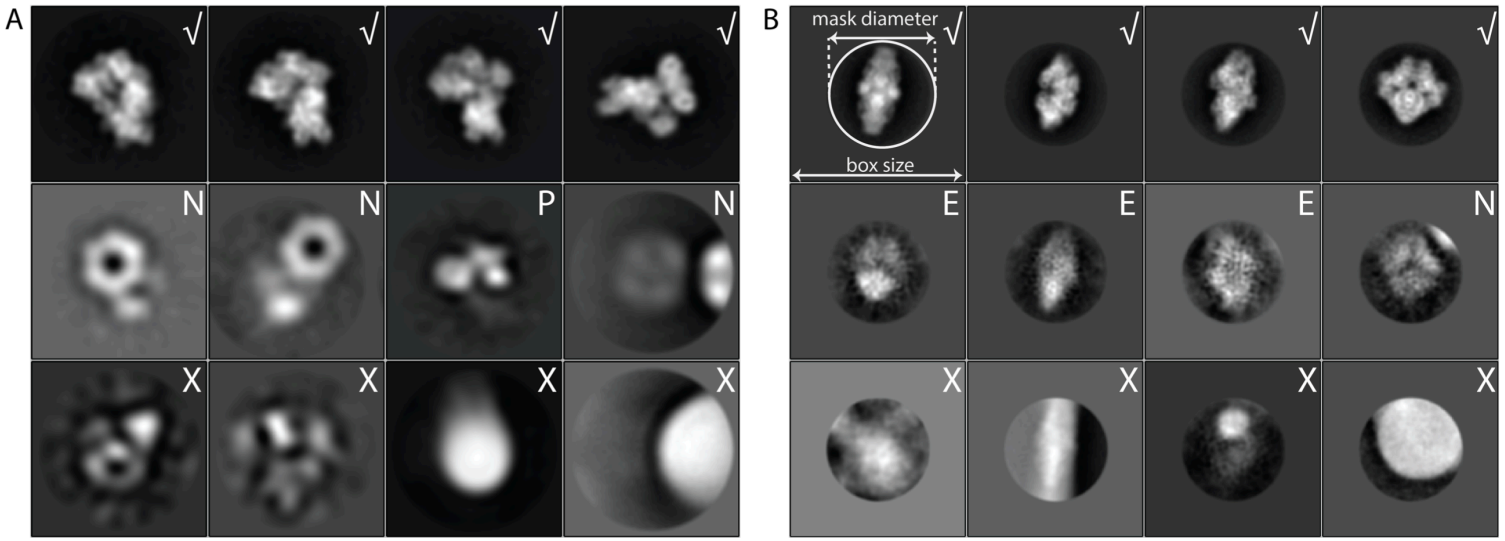
and to cleanup false positives from picking

V Correct

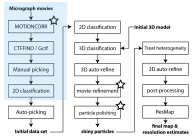
P Partial complex

N Neighbouring particles

X False positives / Contaminants

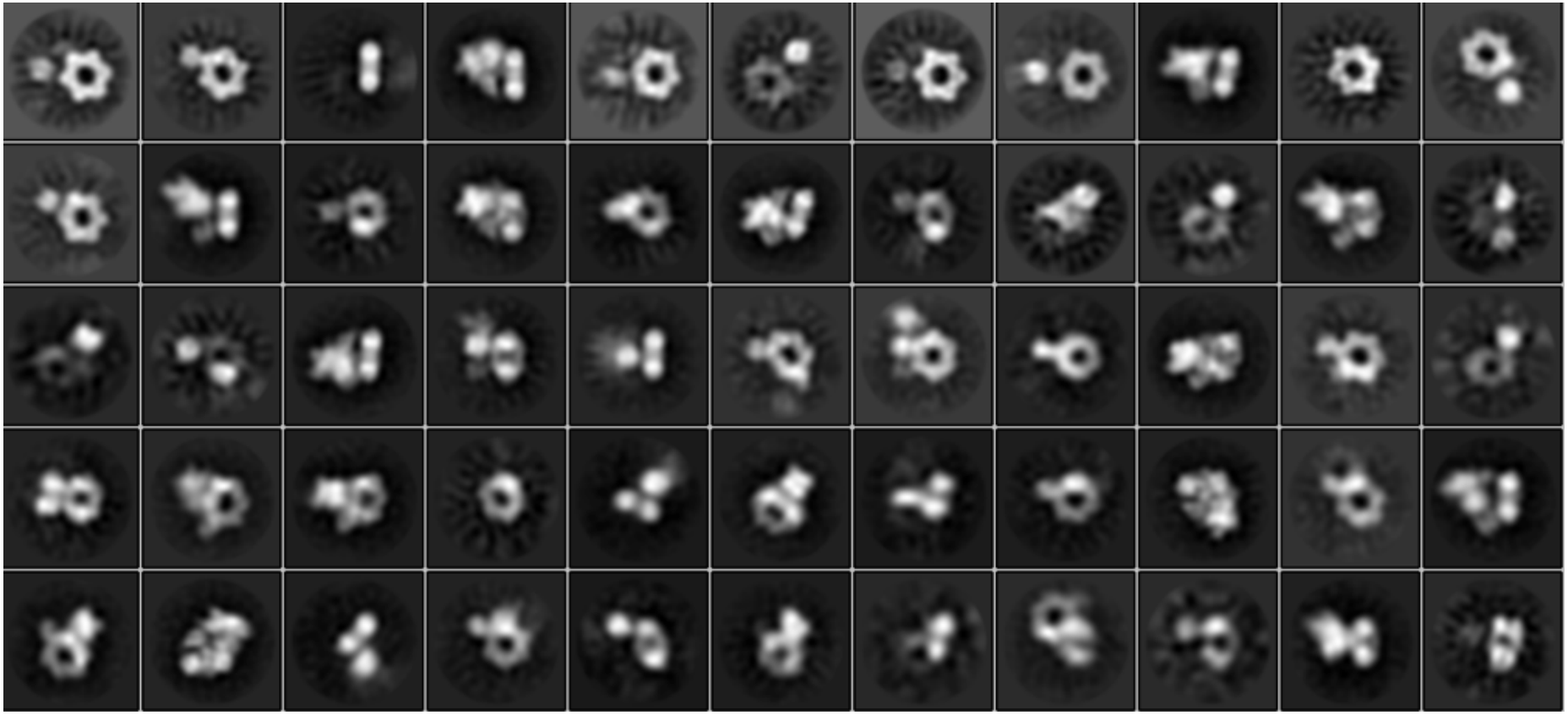


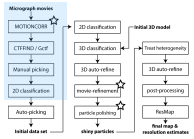
Scheres 2016 (Methods Enzymology)



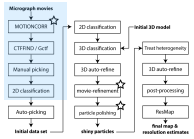
2D Classification

Use 2D classification to asses the quality of your dataset and sample

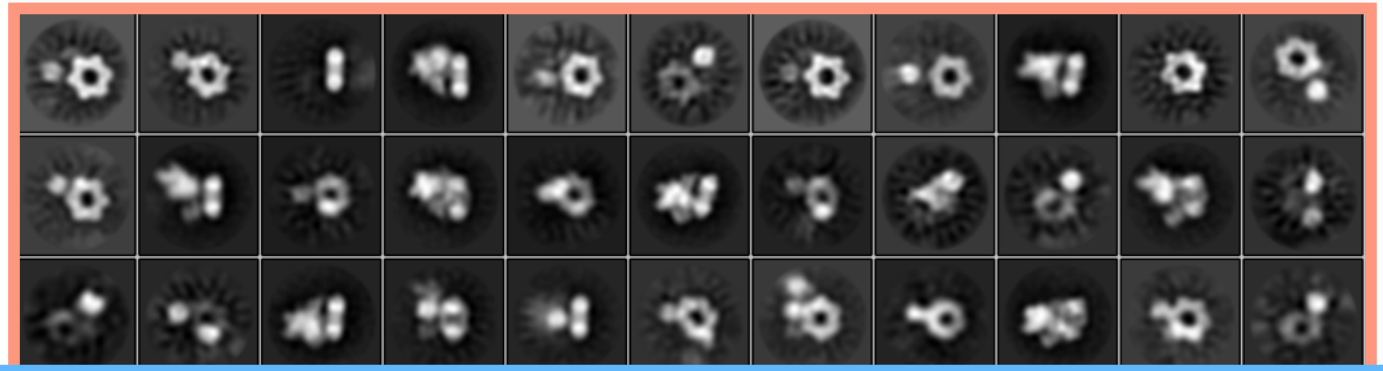


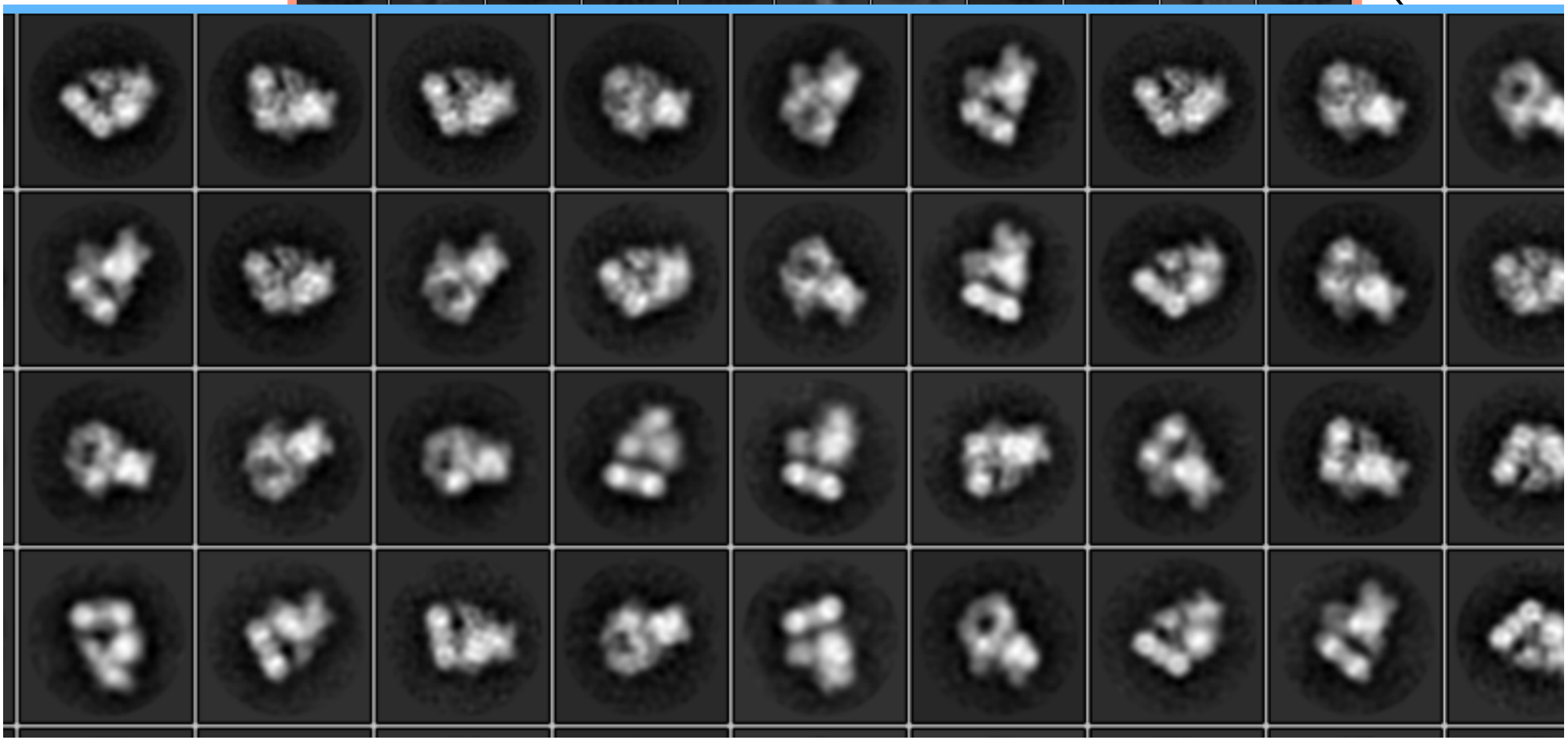


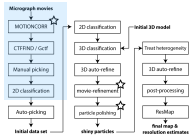
2D Classification



2D Classification

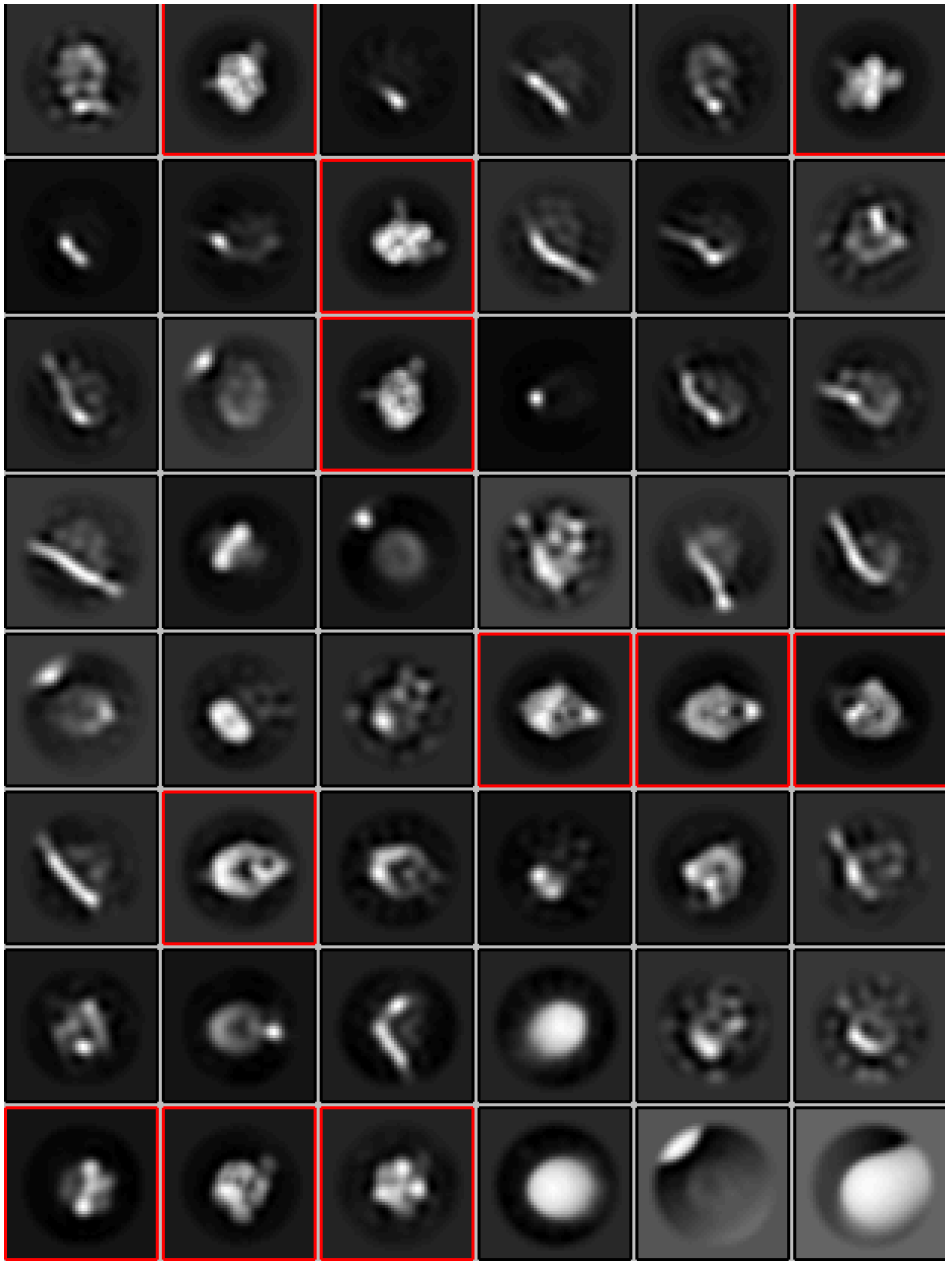


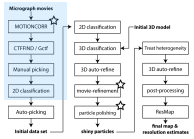




2D Classification

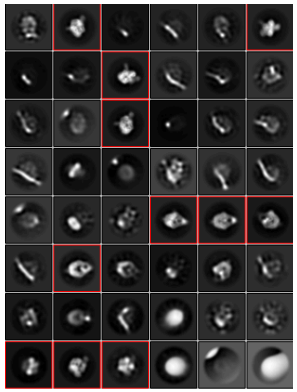
Run 1 (25Å res limit / 50px binx3)
Run 2 (25Å res limit / 50px binx3)
[…]
Run 5 (6Å res limit / 200px)

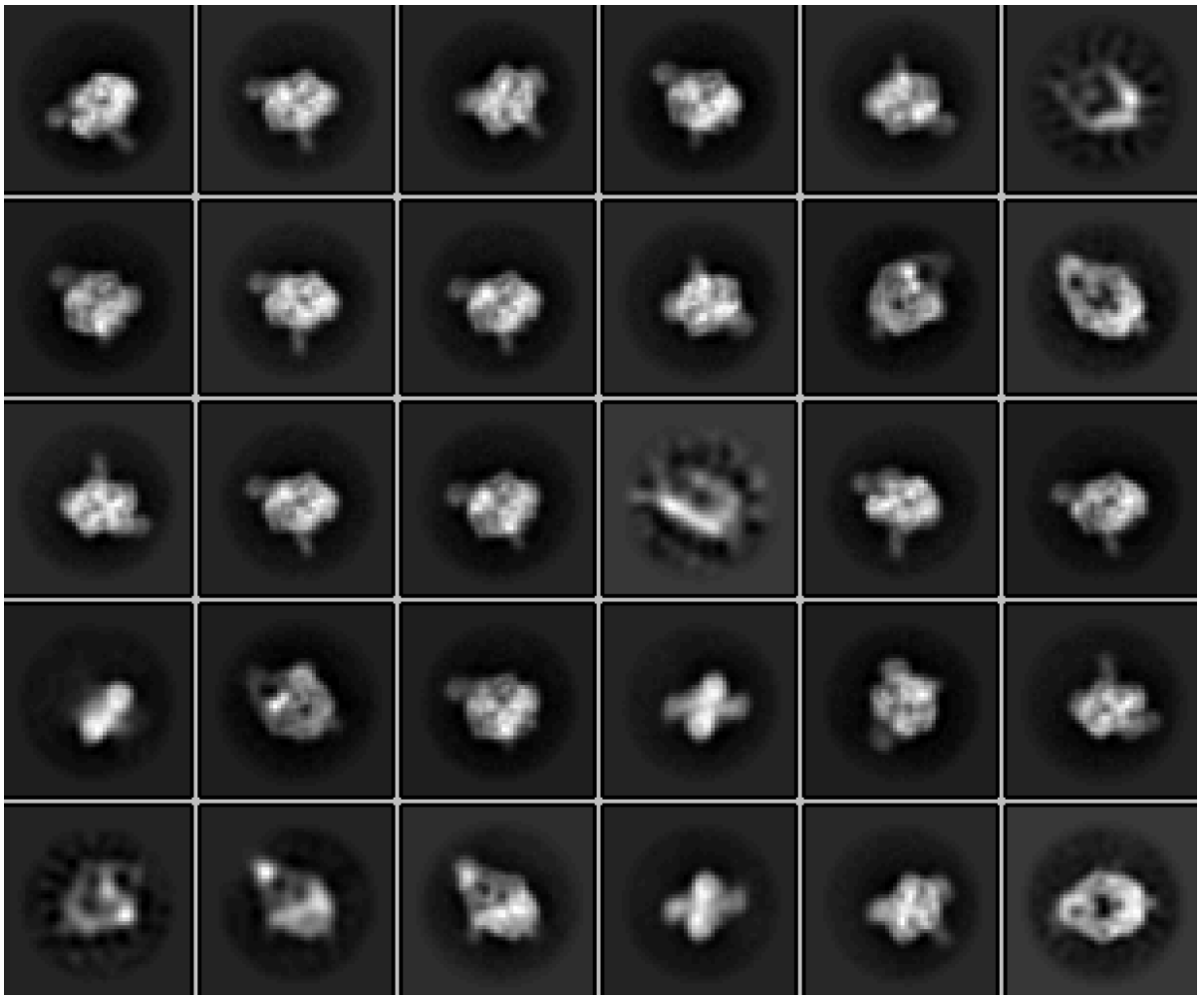


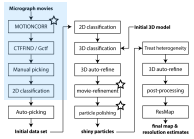


2D Classification

Run 1 (25Å res limit / 50px binx3)
Run 2 (25Å res limit / 50px binx3)
[…]
Run 5 (6Å res limit / 200px)

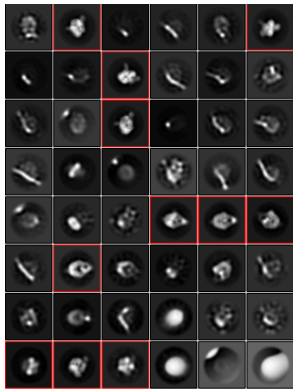


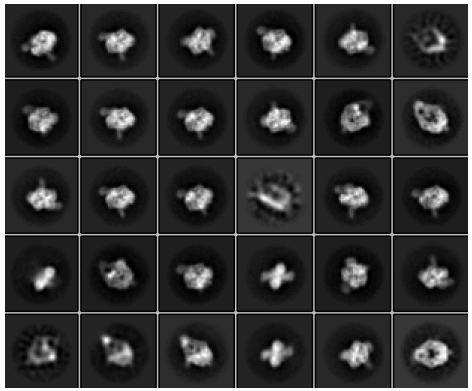


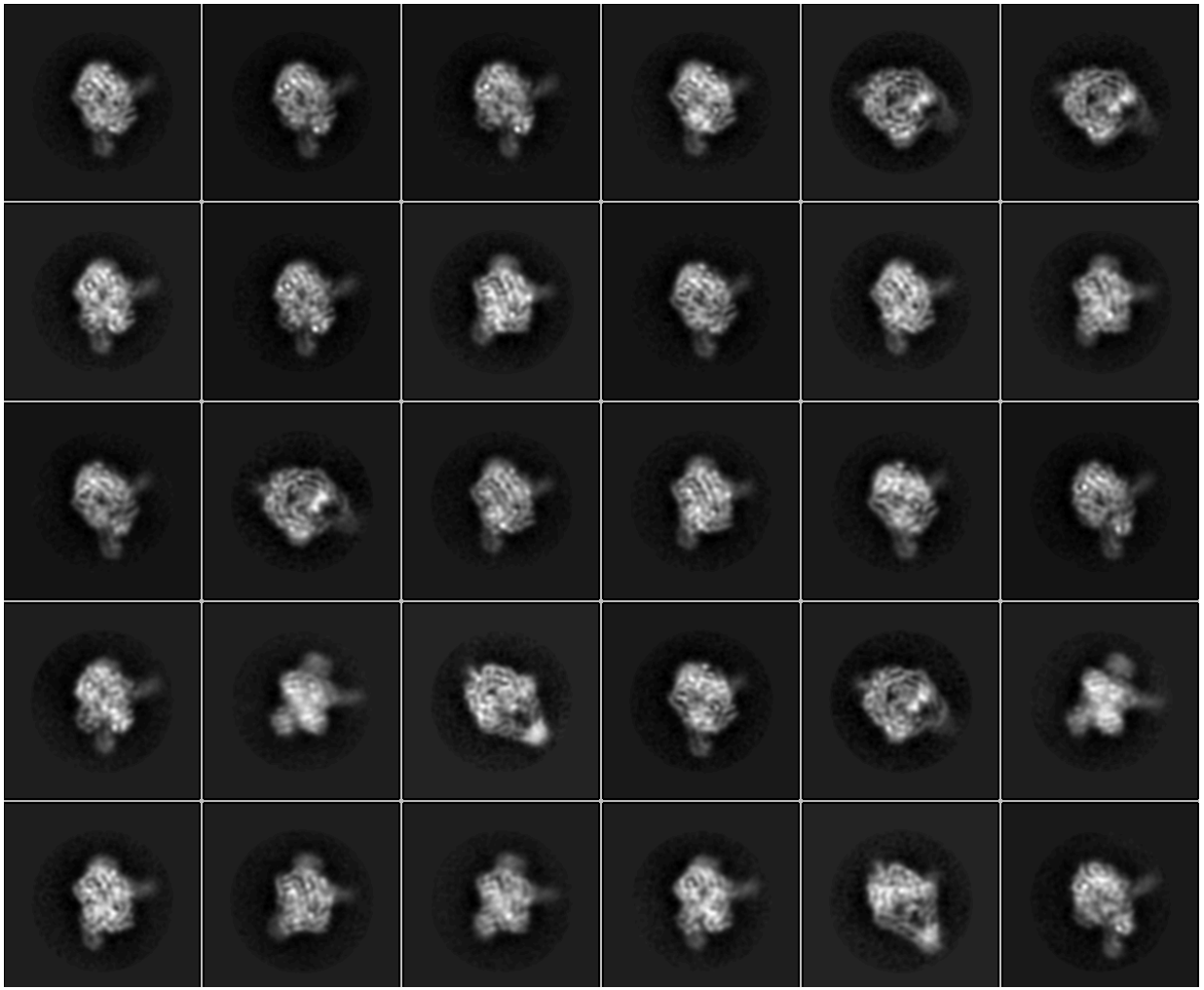


2D Classification

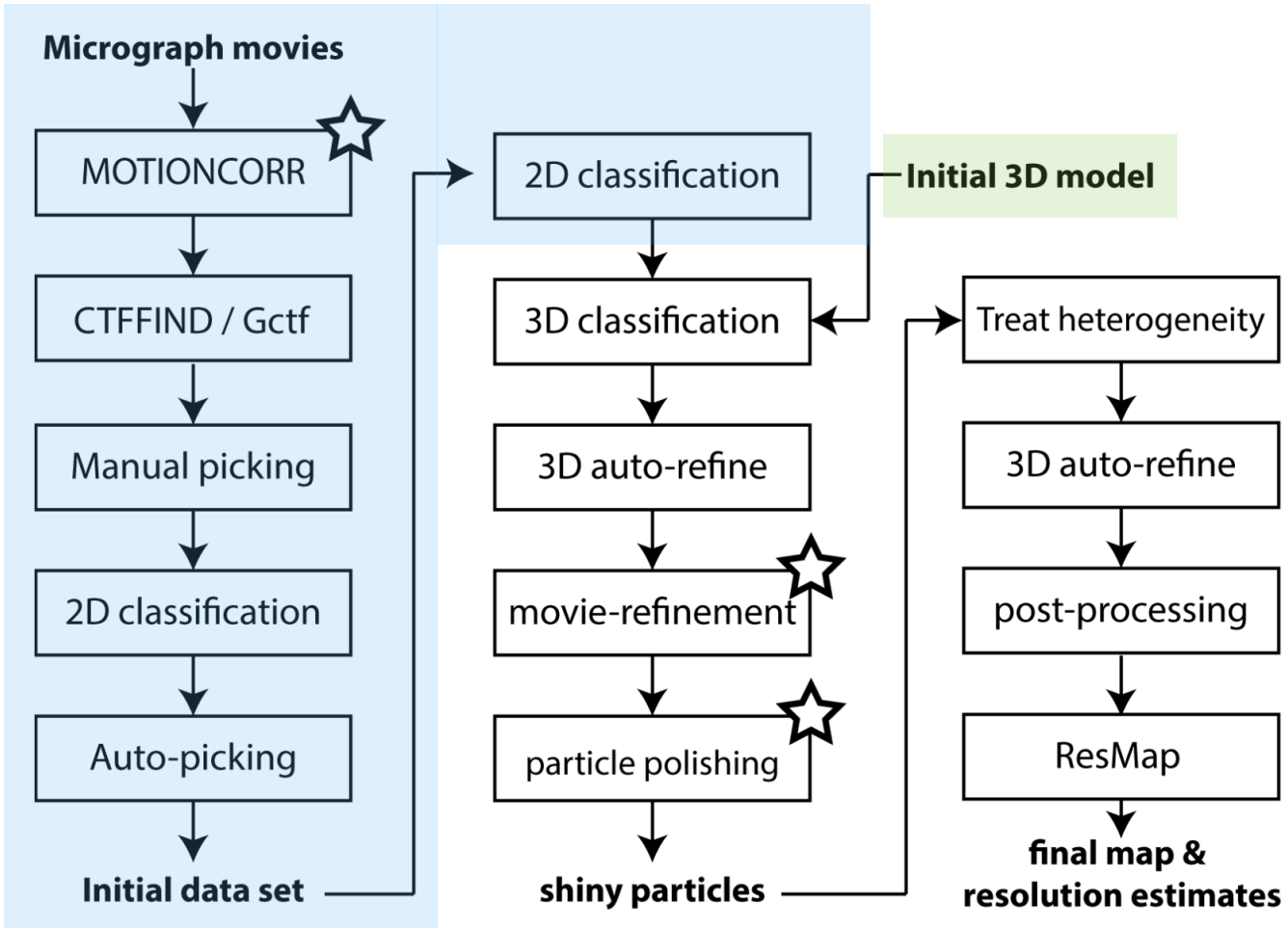
Run 1 (25Å res limit / 50px binx3)
Run 2 (25Å res limit / 50px binx3)
[…]
Run 5 (8Å res limit / 200px)

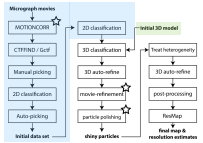






Initial Model



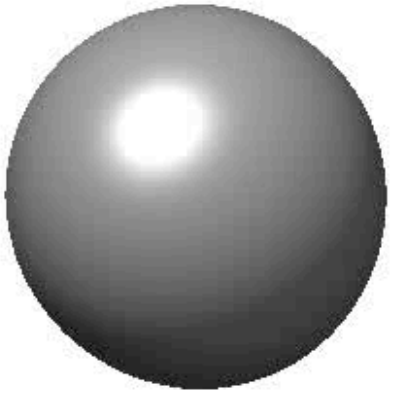


Initial Model

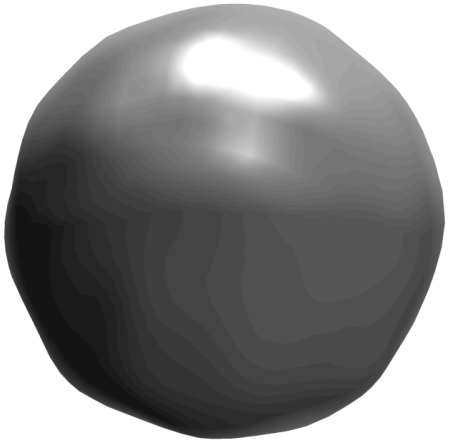
What initial reference 3d map should we use for the first 3D refinement/classification?

- PDB or EMDB structure closely related to your sample (e.g. ribosome) <— **low-pass filter!**
- Use the 2D class averages to calculate a first 3D map (e.g. ab-initio generation in EMAN)
- **Ab-initio model generation from a random subset of particles:**
  - **SIMPLE PRIME** (Stochastic Hill Climbing)
  - **CryoSPARCS** (Stochastic Gradient Descent)
  - **RELION** (Stochastic Gradient Descent)
- Subtomogram averaging
- Random conical tilt
- Common lines

RELION (SGD)

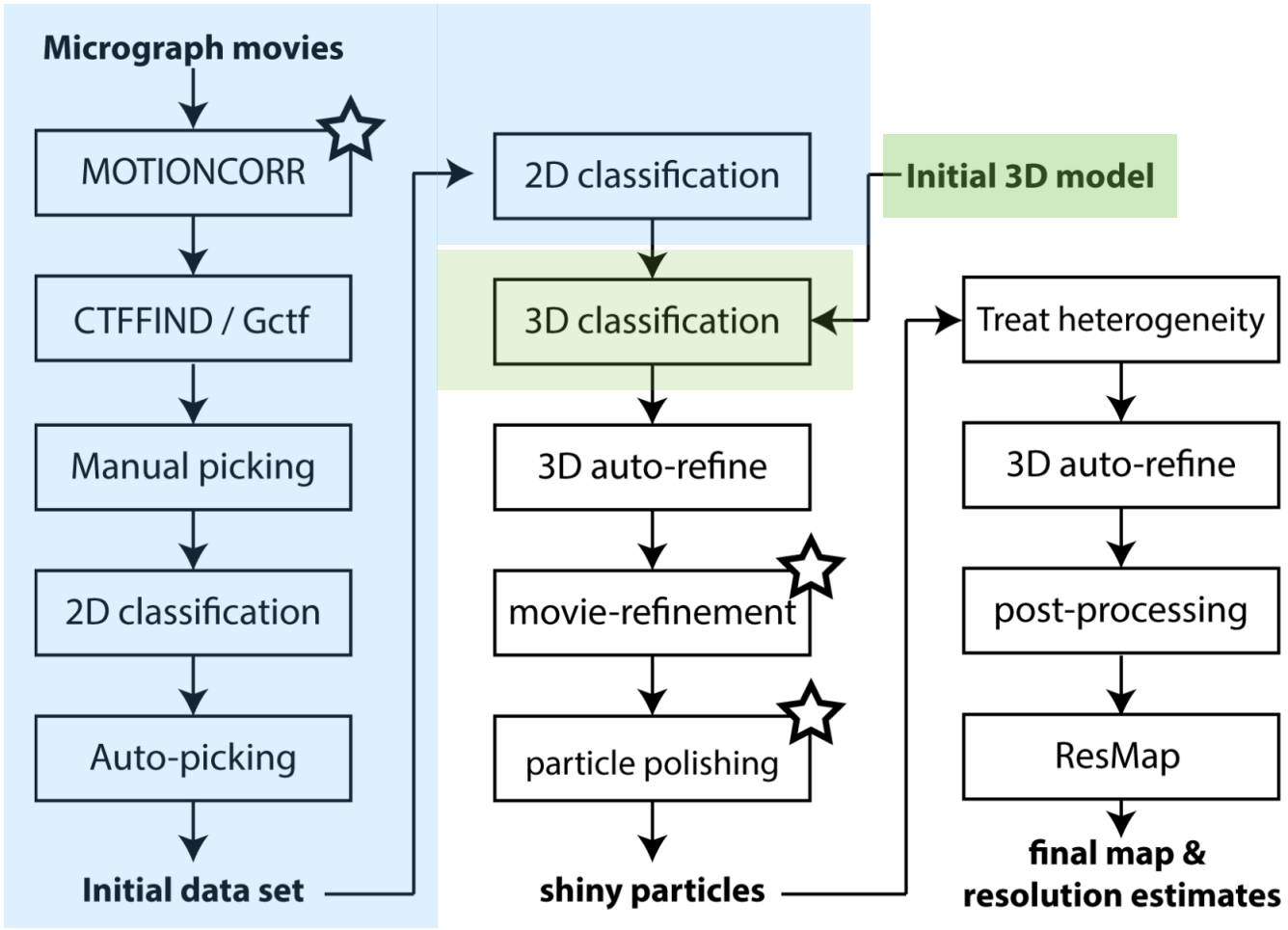


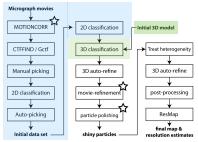
SIMPLE PRIME



SIMPLE - Elmund 2013 (Structure)
cryoSPARC - Punjani 2017 (Nature Methods)

3D Classification





3D Classification

I/O | Reference | CTF | Optimisation | Sampling | Helix | Compute | Running

Input images STAR file: Refine3D/job054/run_data.star [Browse]
Continue from here: [Browse]
Reference map: _it001_class001_rescaled.mrc [Browse]
Reference mask (optional): [Browse]

I/O | **Reference** | CTF | Optimisation | Sampling | Helix | Compute | Running

Ref. map is on absolute greyscale? Yes
Initial low-pass filter (A): 60
Symmetry: C1

I/O | Reference | **CTF** | Optimisation | Sampling | Helix | Compute | Running

Do CTF-correction? Yes
Has reference been CTF-corrected? Yes
Have data been phase-flipped? No
Ignore CTFs until first peak? No

I/O | Reference | CTF | **Optimisation** | Sampling | Helix | Compute | Running

Number of classes: 3
Regularisation parameter T: 4
Number of iterations: 25
Use subsets for initial updates? No
Initial subset size: 10000
Number of subset updates: 3
Mask diameter (A): 200
Mask individual particles with zeros? Yes
Limit resolution E-step to (A): 8

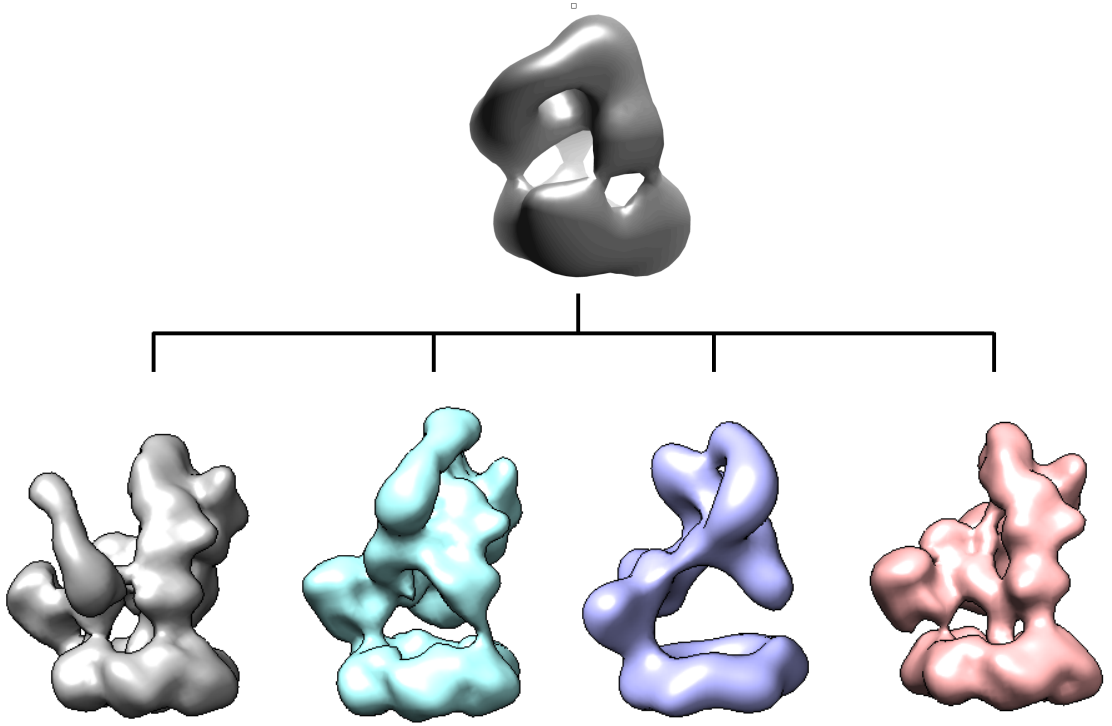
I/O | Reference | CTF | Optimisation | **Sampling** | Helix | Compute | Running

Perform image alignment? Yes
Angular sampling interval: 7.5 degrees
Offset search range (pix): 5
Offset search step (pix): 1
Perform local angular searches? No
Local angular search range: 5

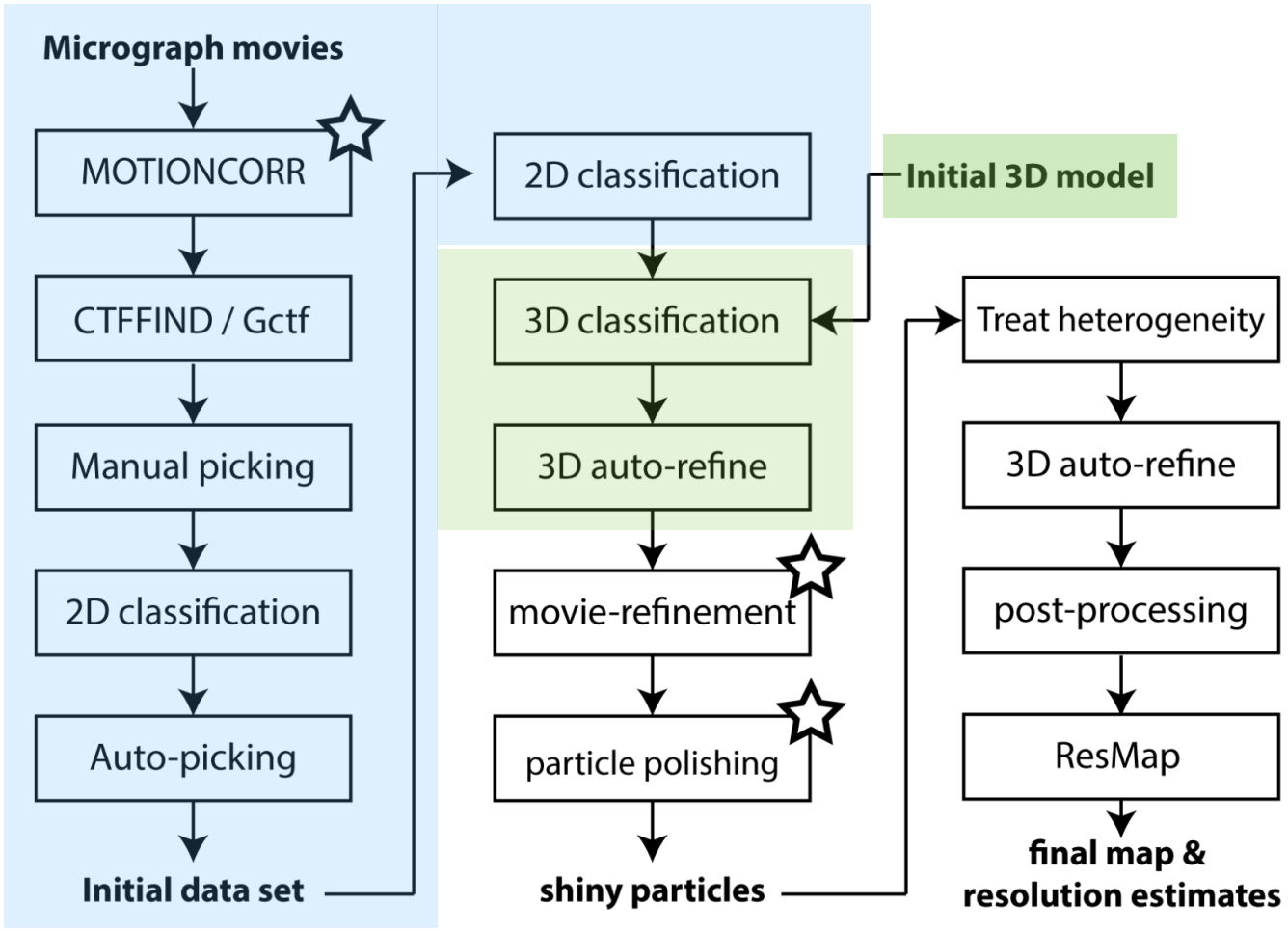
[Print command] [Schedule] [Run now!]

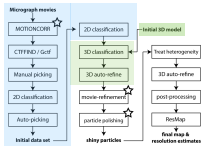
Use initial 3D classification to further cleanup your dataset

Number of classes
CTF correction
Mask
Resolution limit

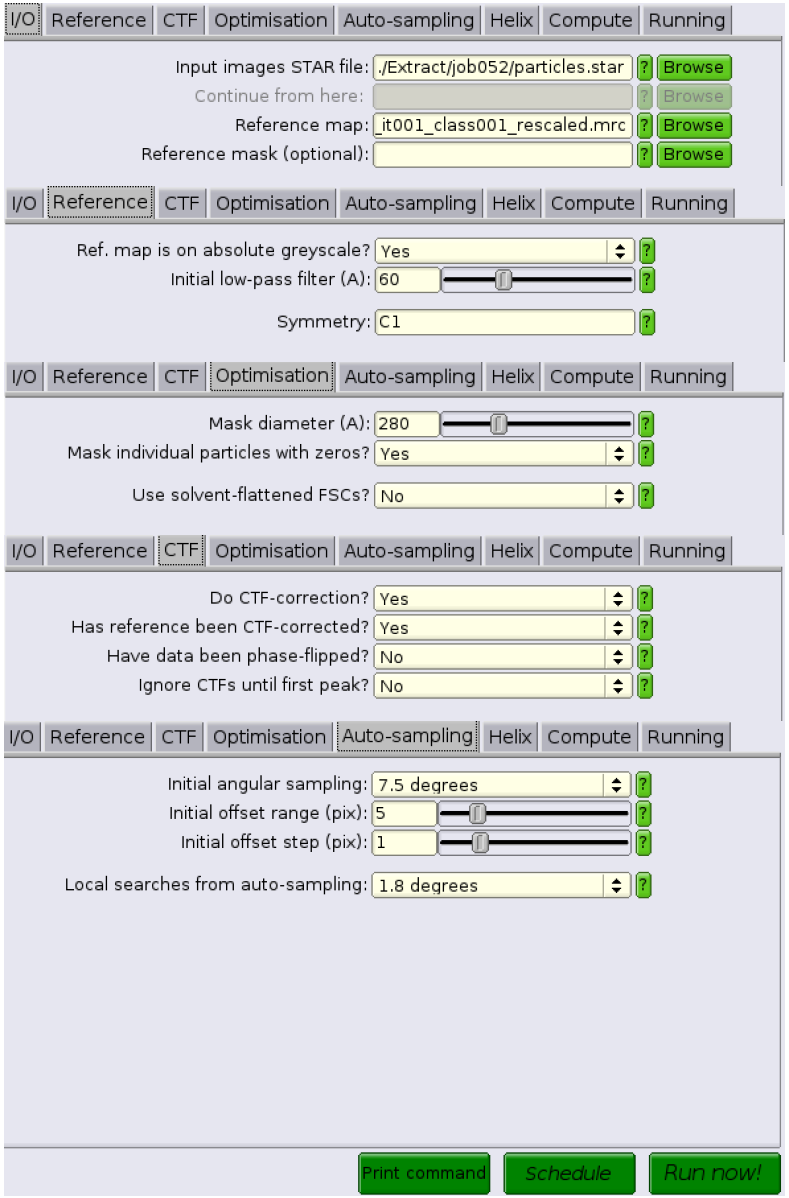


3D Refinement





3D Refinement



Gold-Standard FSC

Data is split in **two halves** at the beguinning

Refinement starts with **same initial model** for each half dataset

Both halves **refined independently** against independent models

FSC in between the two halves after every iteration helps **RELION tune parameters automatically** and **prevent overfitting**

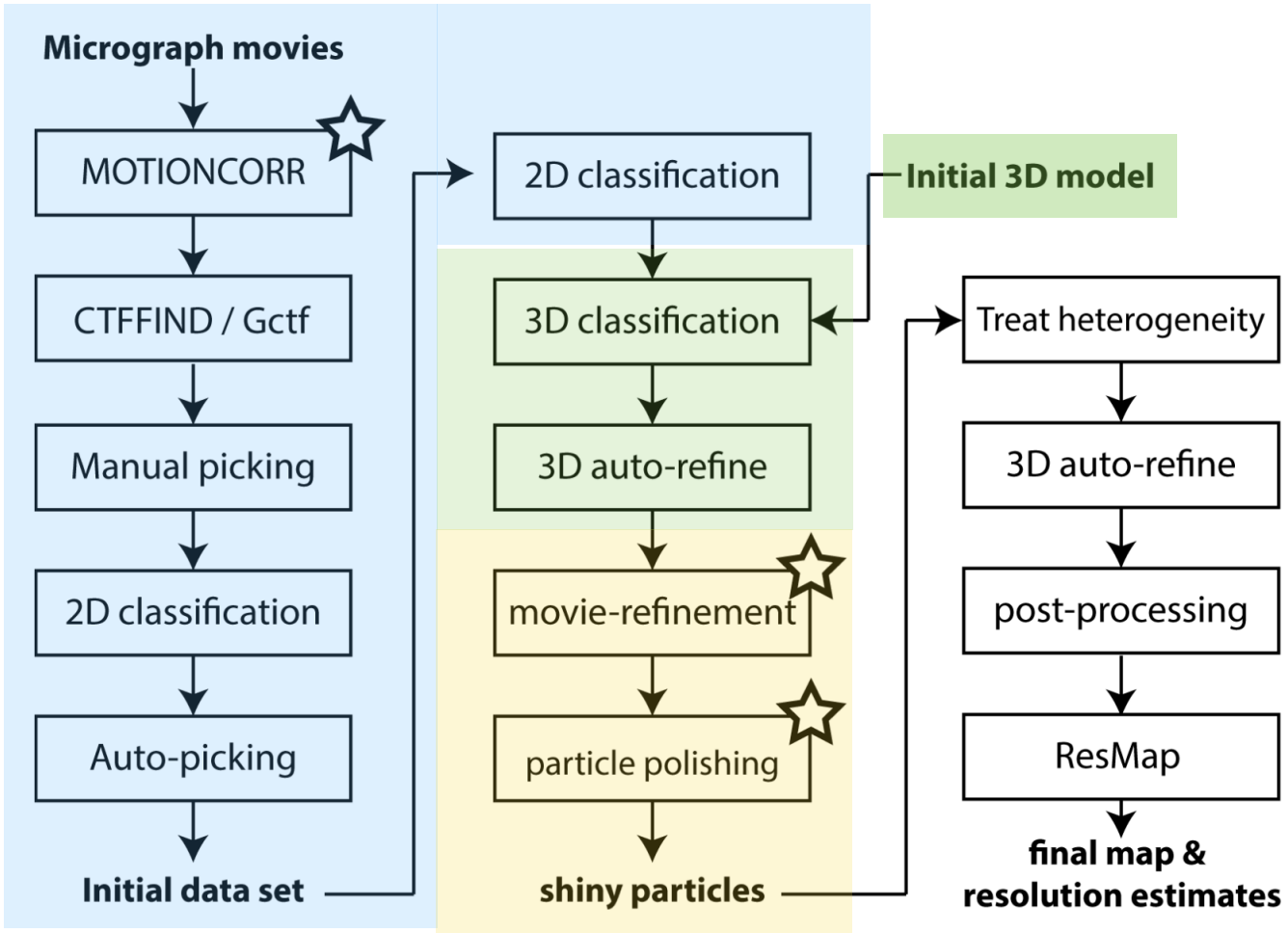
Follow progression via the run.out file

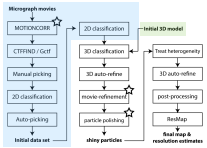
After **convergence** (no improvement in angles and resolution) two **unmasked half maps** are written and a last iteration with all data together is performed

Unfiltered-unmasked-half-maps will be used for postprocessing later on

Scheres 2012 (Nature Methods)

Movie Procesing & Particle Polishing





Movie Processing & Particle Polishing

I/O | extract | normalise | refine | Running

Input movies STAR file: cted_micrograph_movies.star ? Browse
Rootname of movies files: movie ?

Continue 3D auto-refine from: o371/run_it016_optimiser.star ? Browse

Process micrographs in batches of: 50 ?

I/O | extract | normalise | refine | Running

First movie frame to extract: 1 ?
Last movie frame to extract: 0 ?
Average every so many frames: 1 ?
Maximum number of MPI nodes: 8 ?

Particle box size (pix): 200 ?
Invert contrast? Yes ?

Rescale particles? No ?
Re-scaled size (pixels): 128 ?

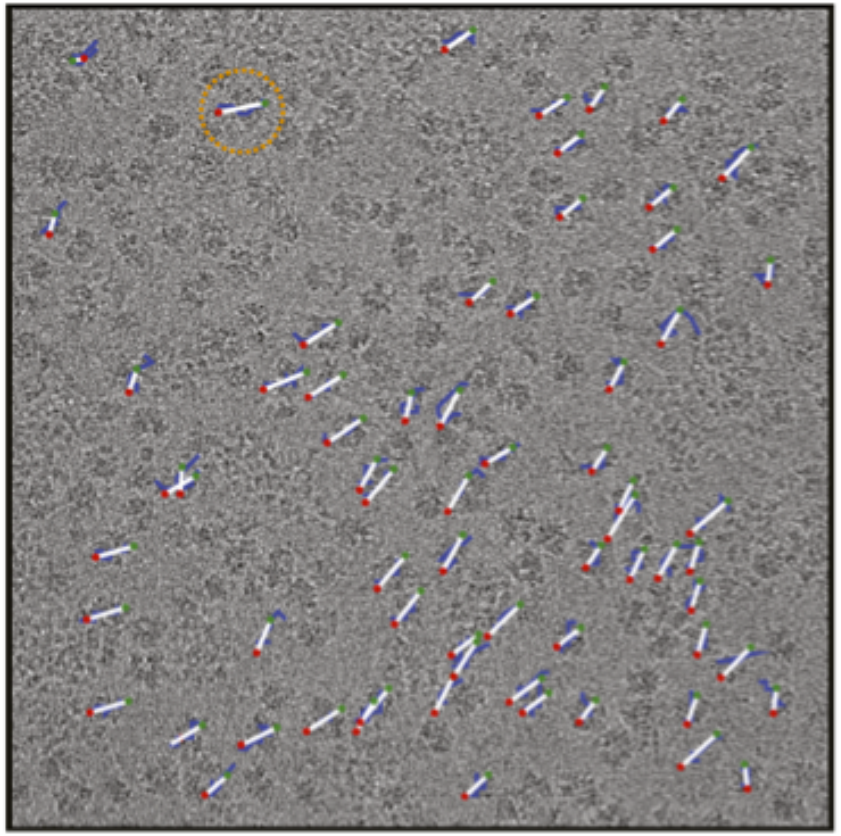
I/O | extract | normalise | refine | Running

Running average window: 7 ?
Stddev on the translations (pix): 1 ?
Also include rotational searches? No ?
Stddev on the rotations (deg): 1 ?

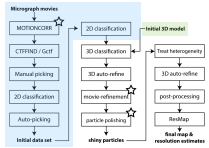
Print command | Schedule | Run now!

Movie Processing

Continuation of a refinement
Running average window



Scheres, 2014, Elife



Movie Processing & Particle Polishing

I/O | Movement | Damage | Normalise | Helix | Running

Input STAR file with aligned movies: 379/run_ravg7_off1_data.star ? Browse
Mask for the reconstructions MaskCreate/job375/mask.mrc ? Browse

I/O | Movement | Damage | Normalise | Helix | Running

Linear fit particle movements? Yes ?
Stddev on particle distance (pix) 500 ?

I/O | Movement | Damage | Normalise | Helix | Running

Perform B-factor weighting? Yes ?
Highres-limit per-frame maps (A) 6 ?
Lowres-limit B-factor estimation (A) 20 ?
Average frames B-factor estimation 1 ?

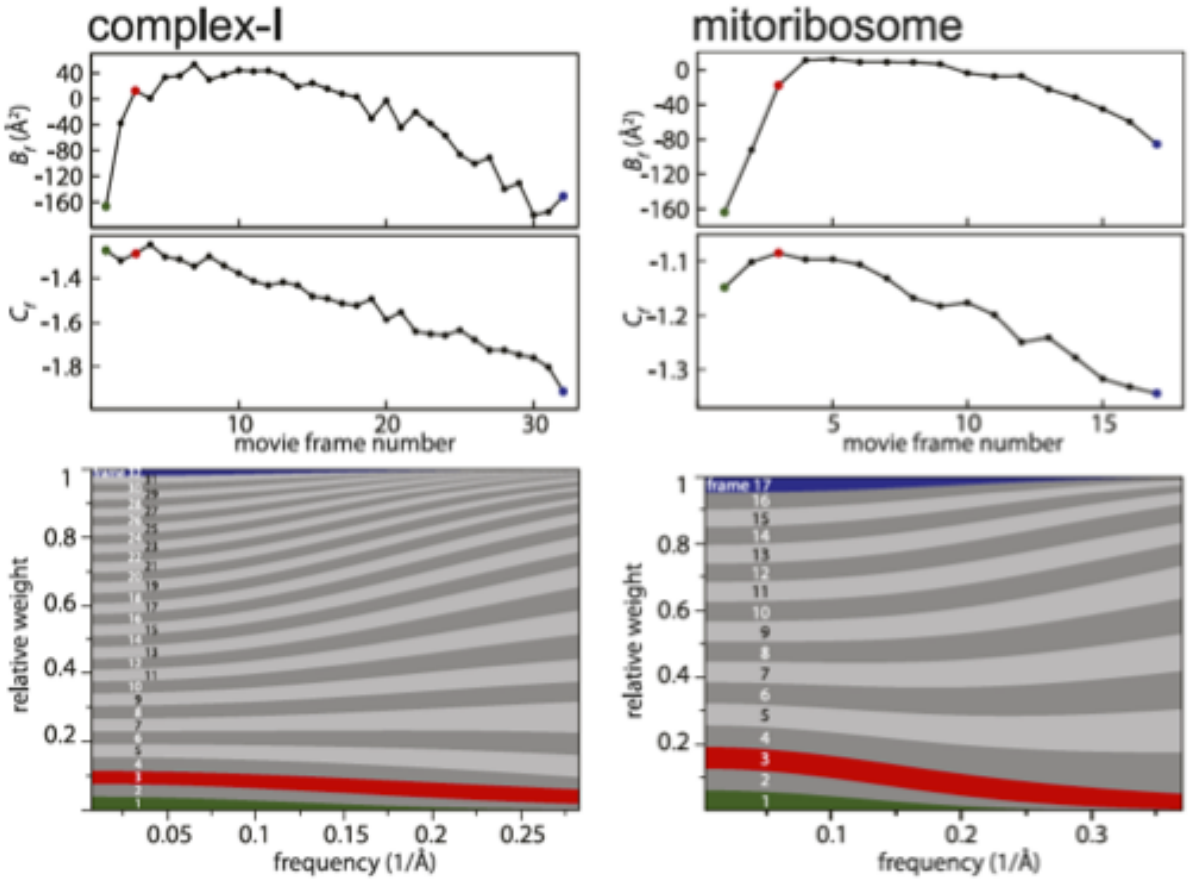
Symmetry: C1 ?

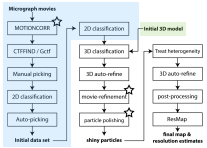
Print command | Schedule | Run now!

Particle Polishing

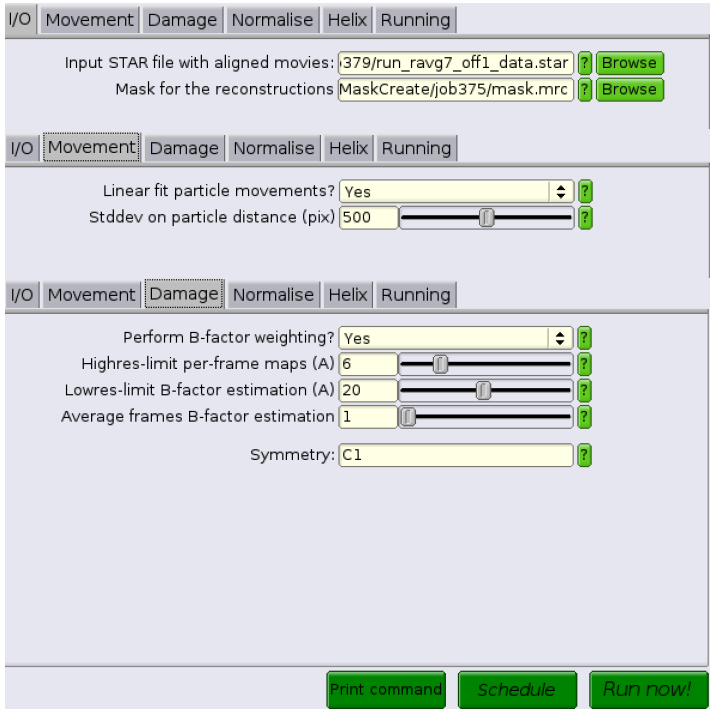
Correct Movement (fit linear path)

Account for radiation damage (Dose Weighting)





Movie Processing & Particle Polishing



Dose weighting in motioncor2?
Polishing in RELION?

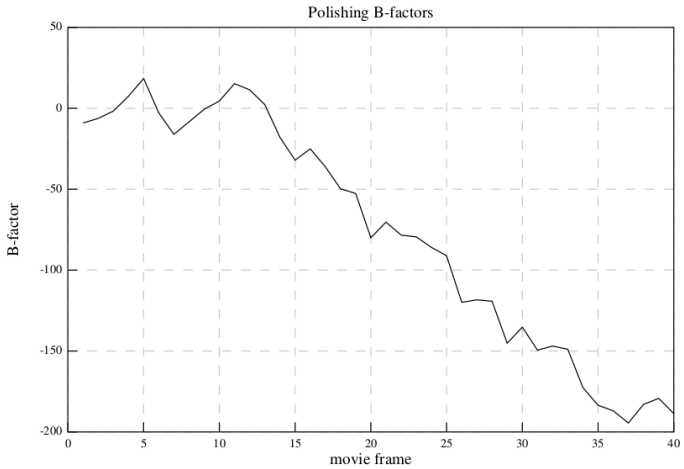
Use both

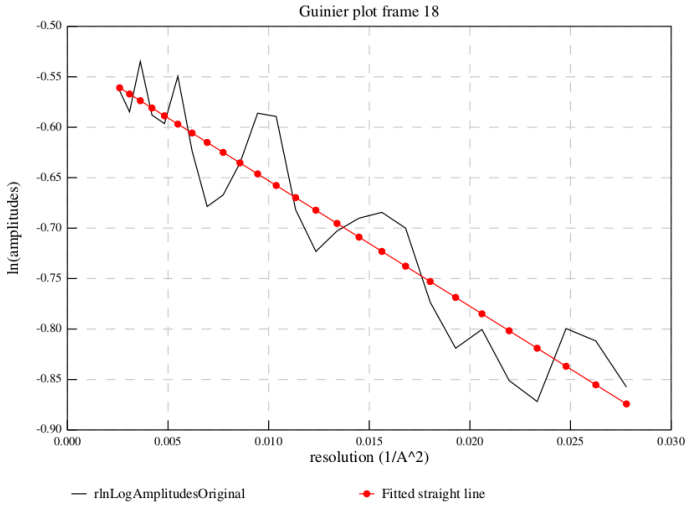
Very useful to have dose weighting from the beginning
(provided you know the dose you put on your sample)

We have seen always improvements using polishing

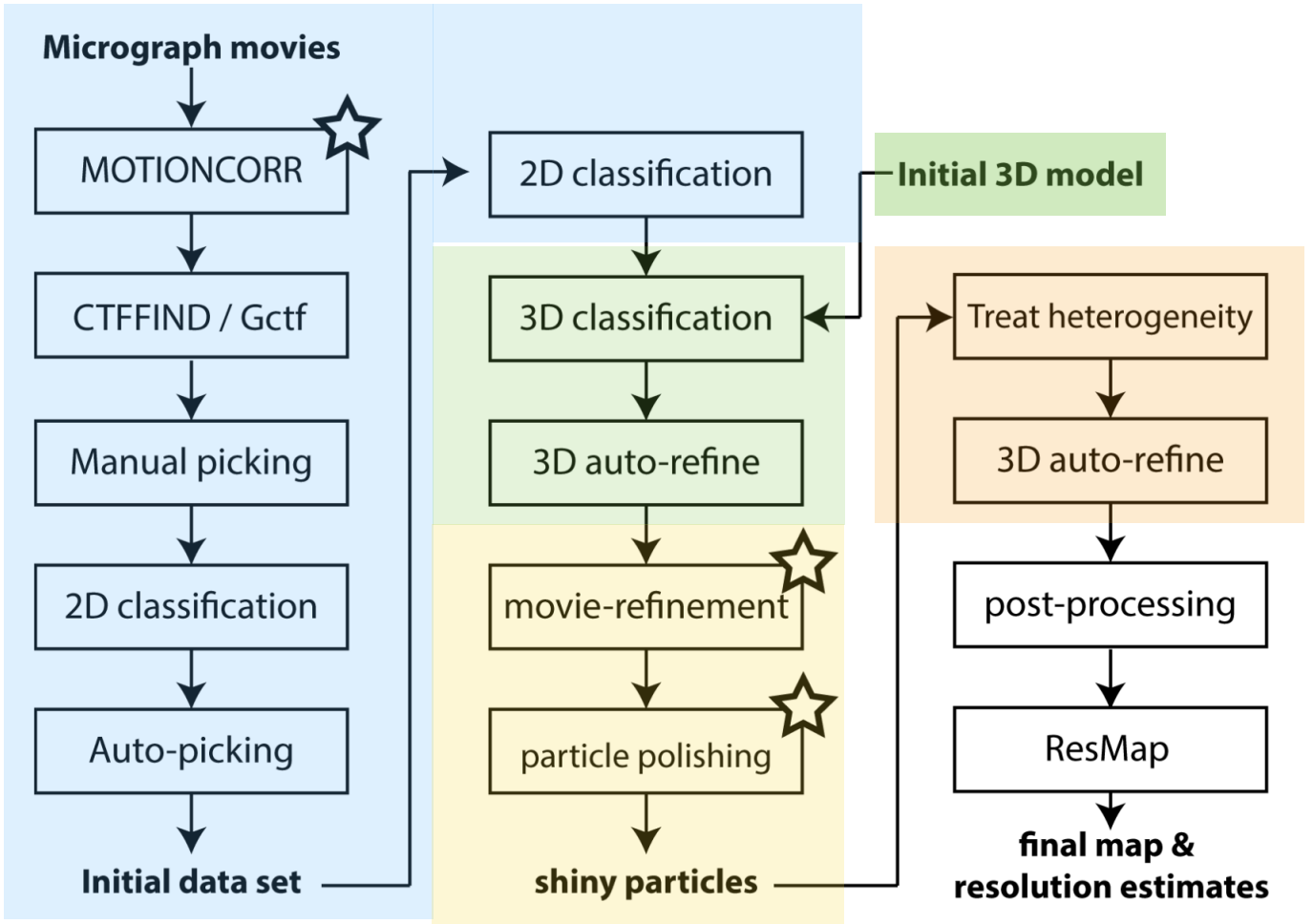
Particle Polishing

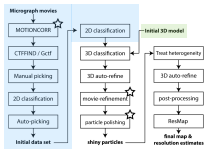
Correct Movement (fit linear path)
Account for radiation damage (Dose Weighting)





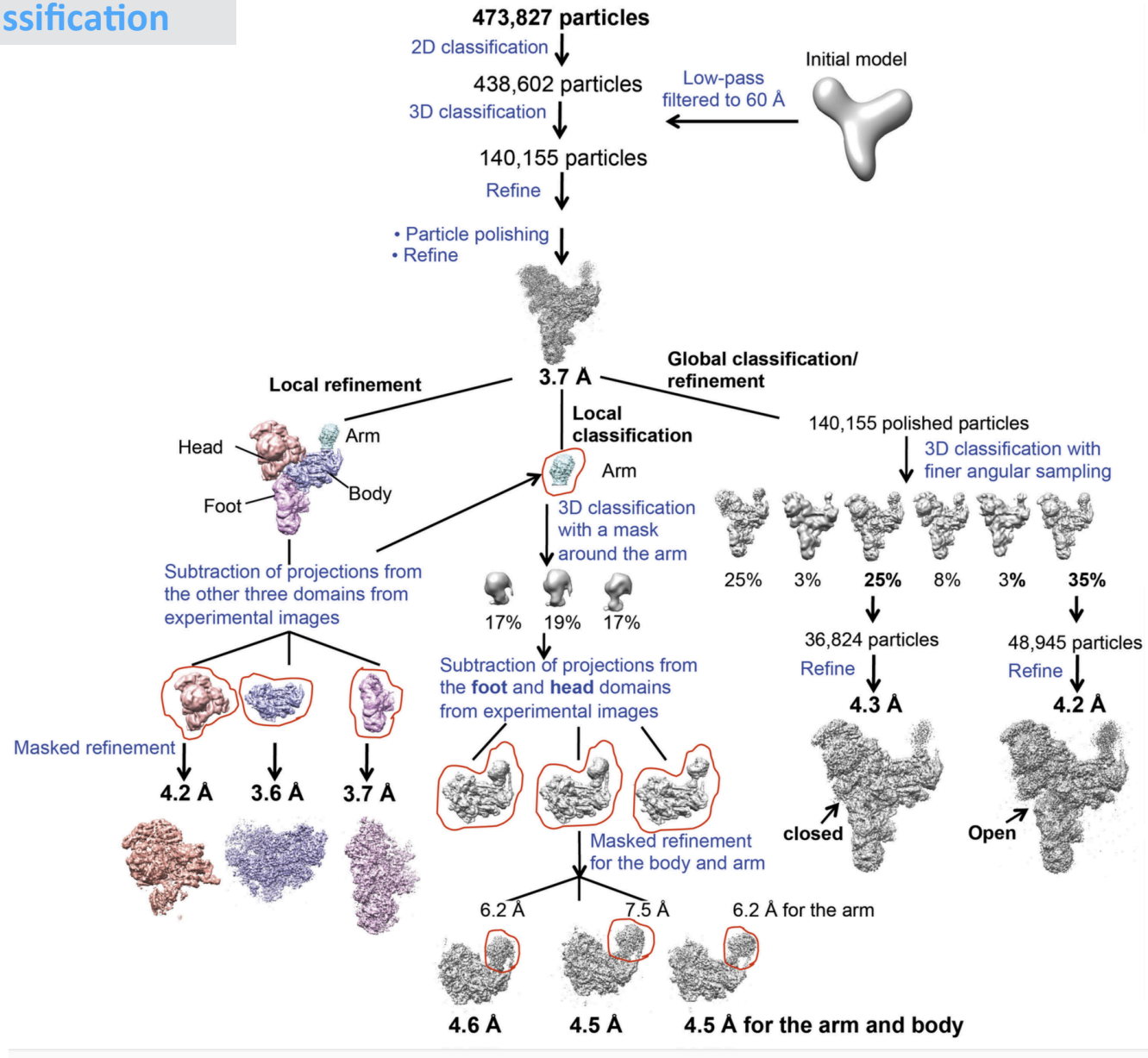
Treating Heterogeneity





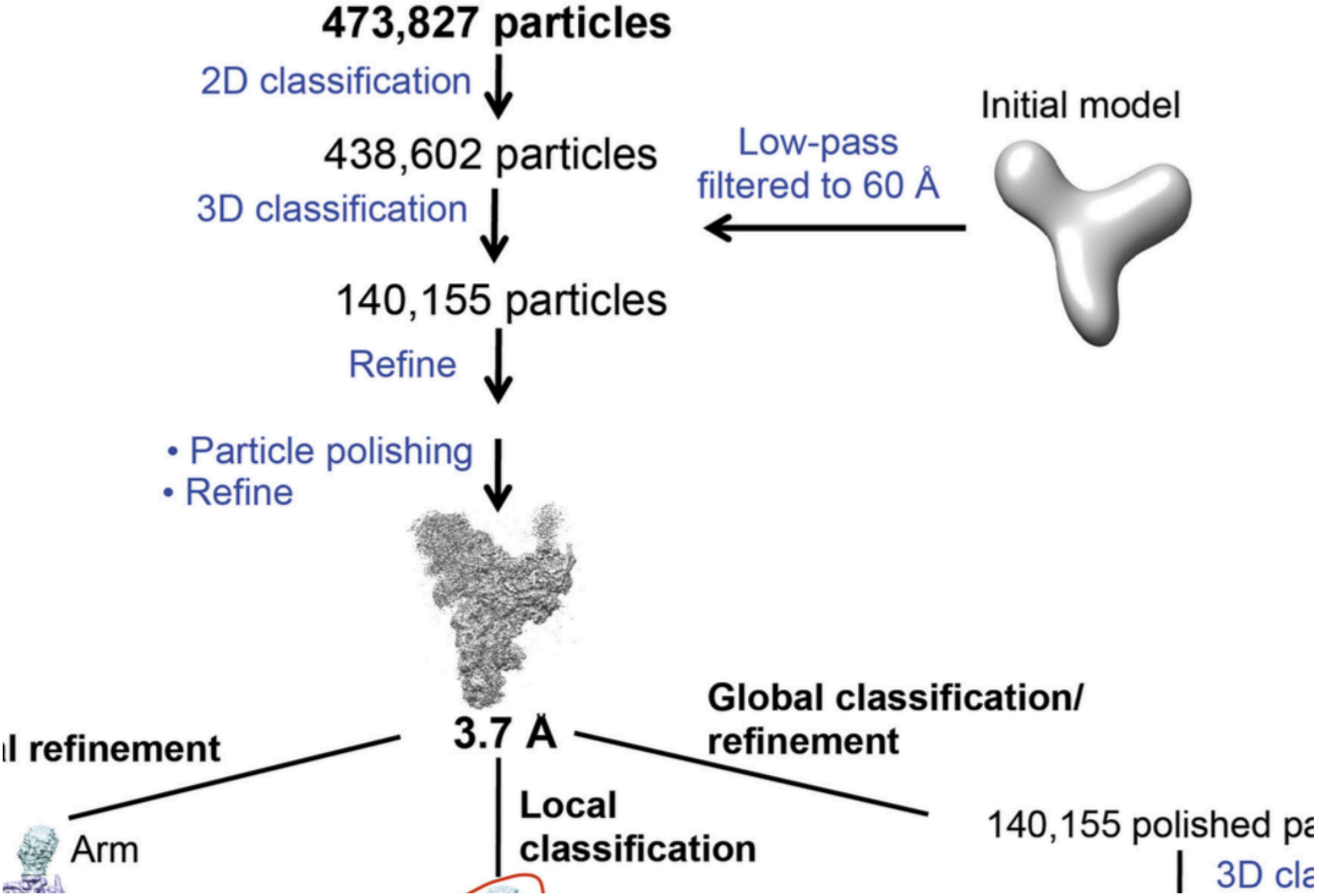
Treating Heterogeneity

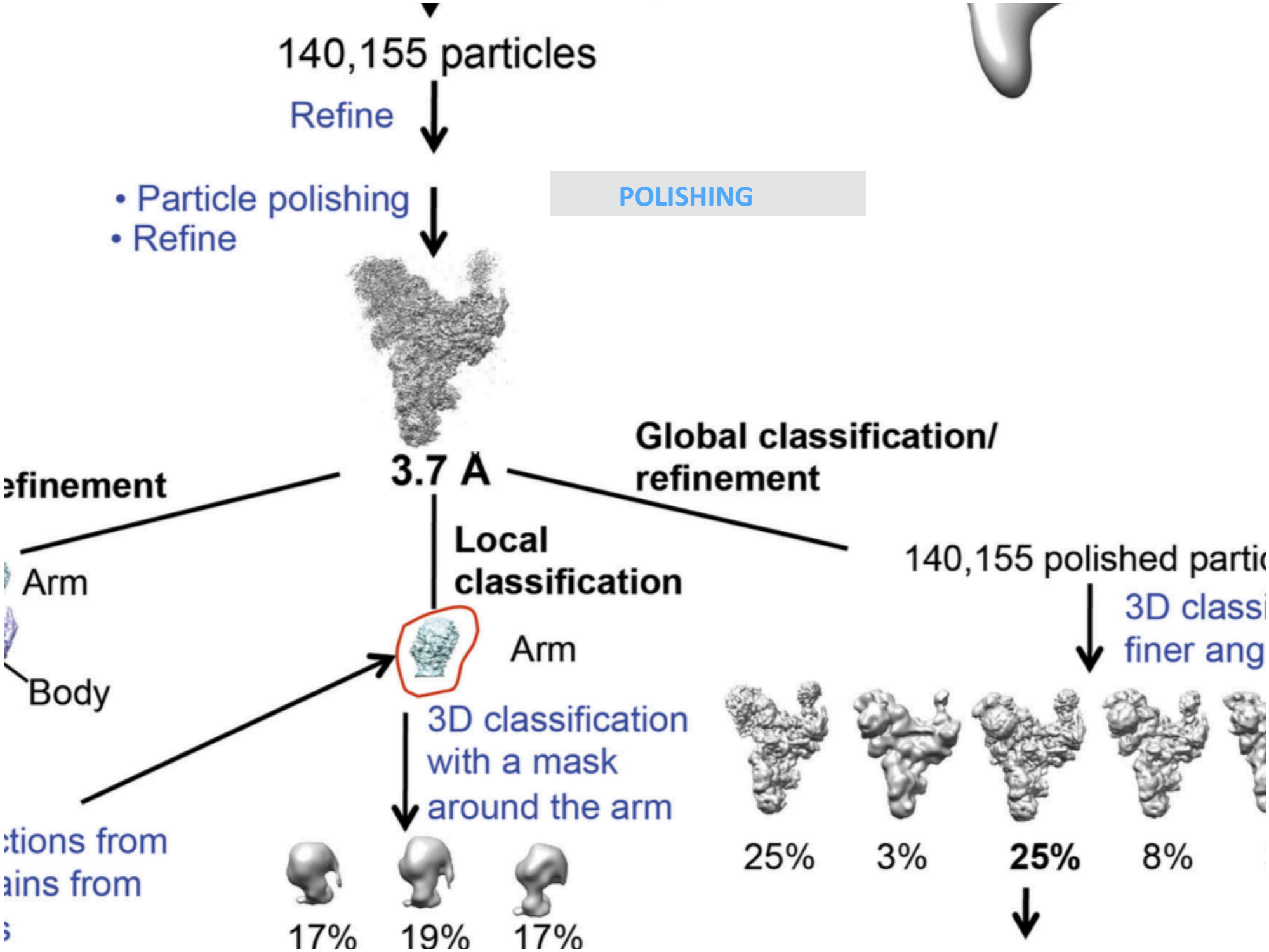
Initial Classification



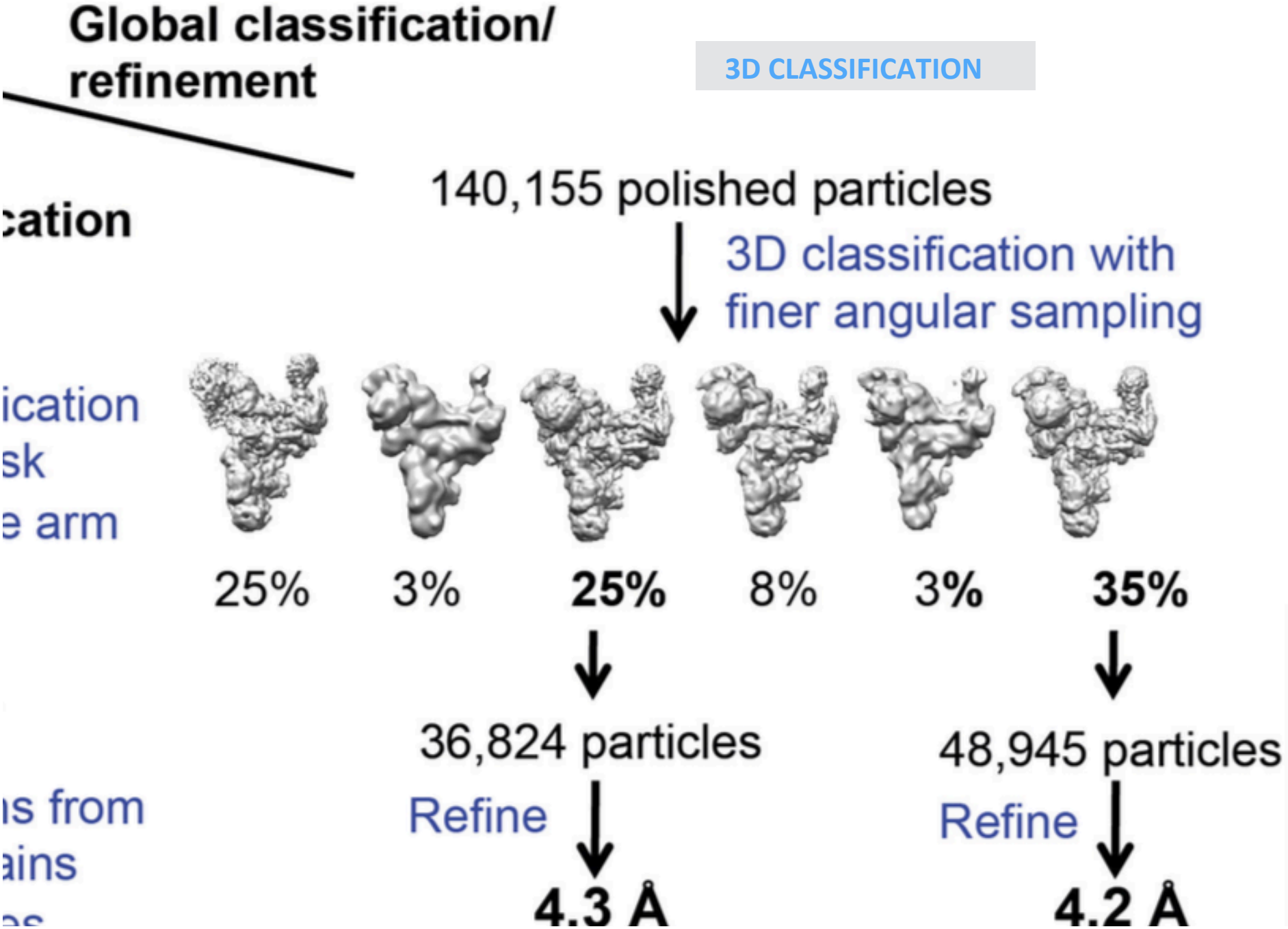
Nguyen 2016 (Nature)

Initial Classification




140,155 particles
Refine
• Particle polishing
• Refine
POLISHING
3.7 Å
efinement
Arm
Body
Local classification
Arm
3D classification with a mask around the arm
ctions from
ains from
17%
19%
17%
Global classification/ refinement
140,155 polished partic
3D classi
finer ang
25%
3%
25%
8%

3D CLASSIFICATION



25% 3% **25%** 8% 3% **35%**

36,824 particles

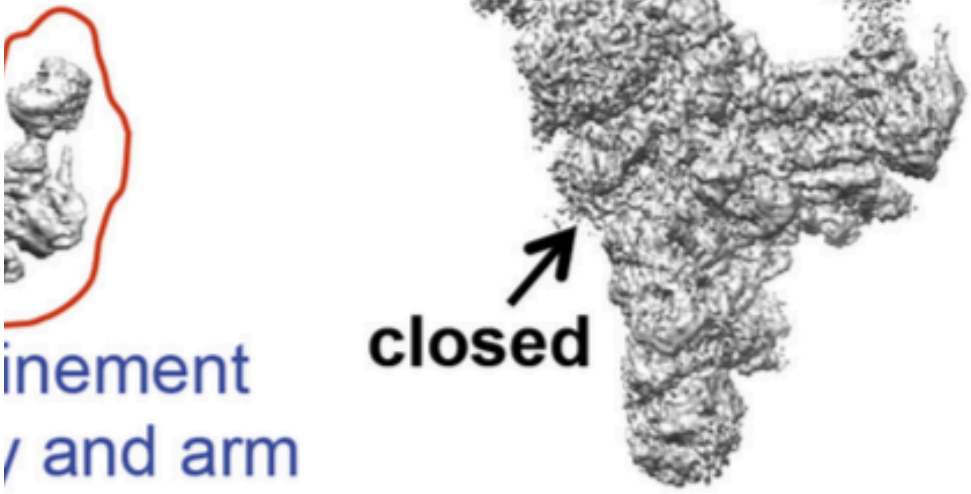
Refine

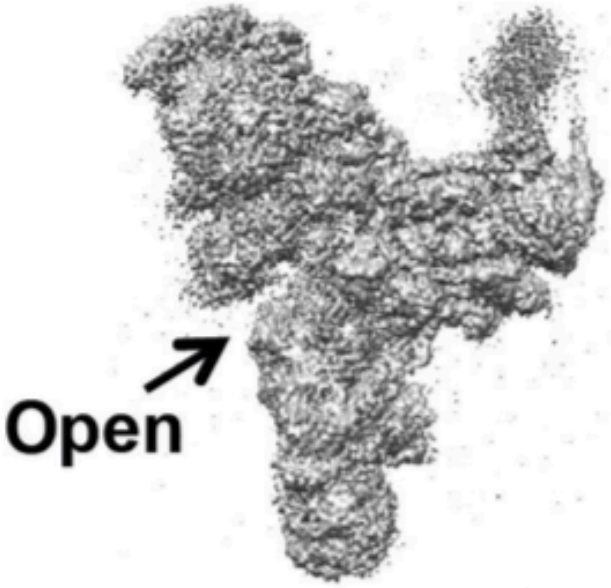
4.3 Å

48,945 particles

Refine

4.2 Å

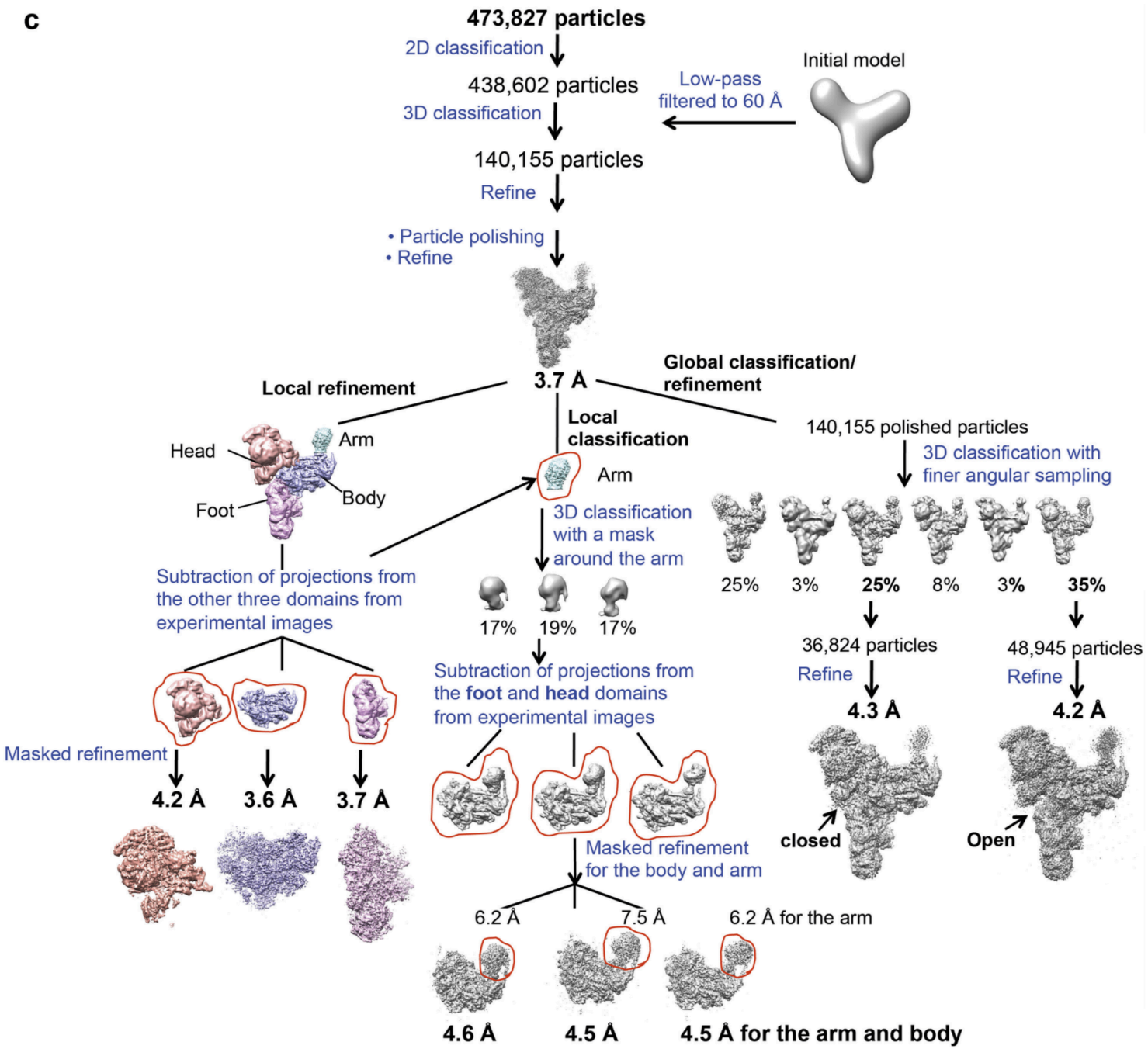


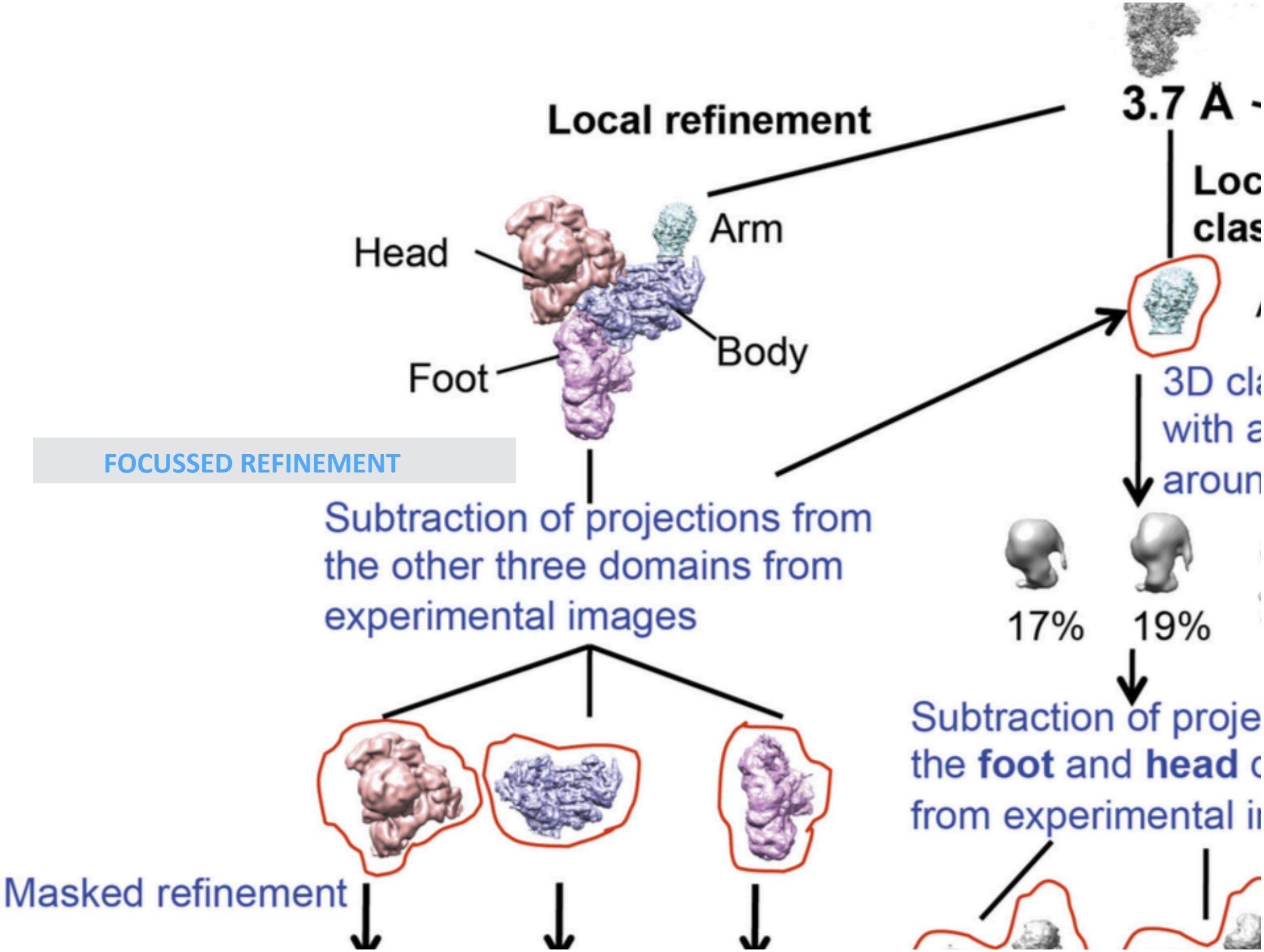


inement
/ and arm

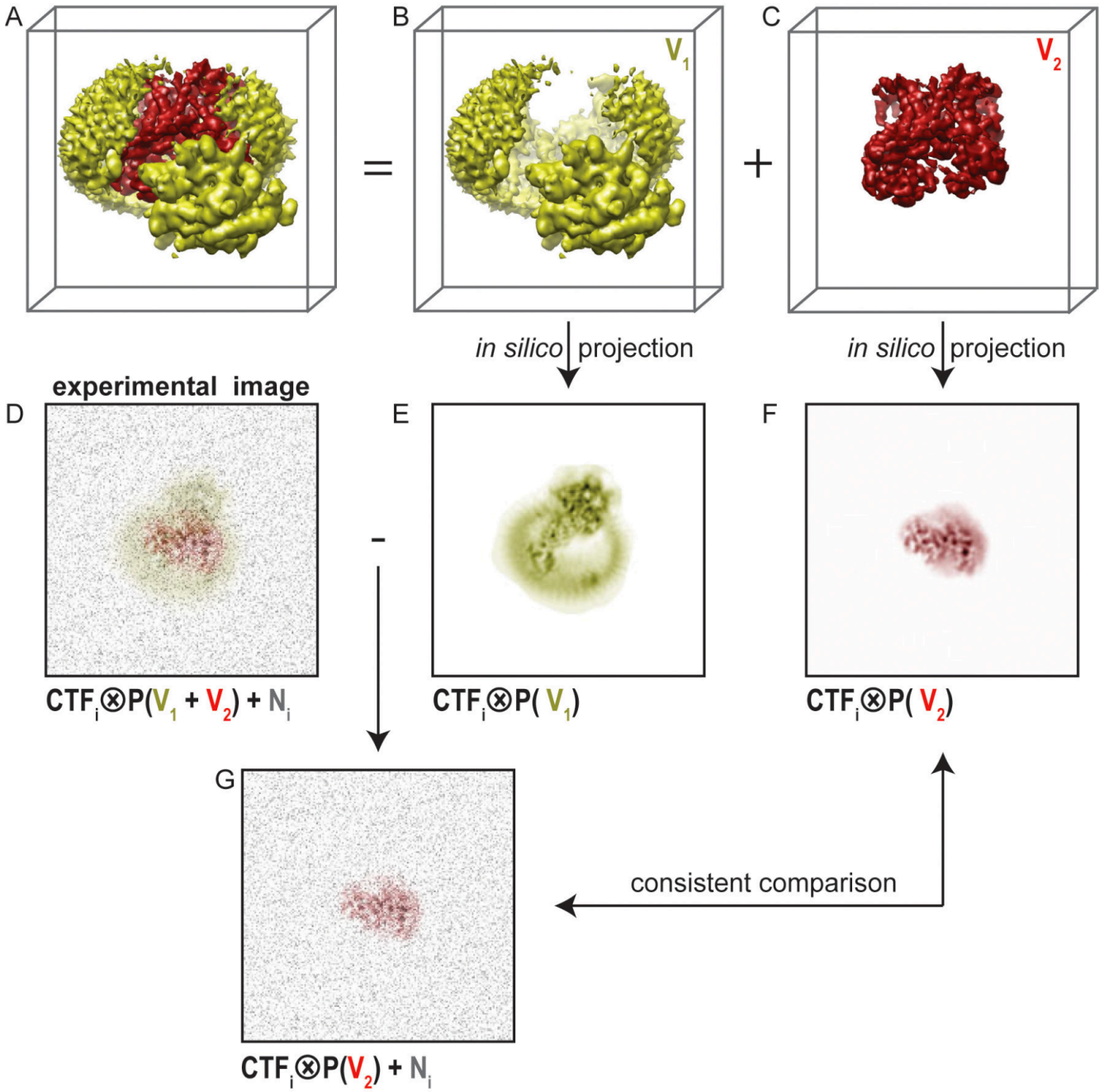
6.2 Å for the arm

3D REFINAMENT MAIN CLASSES

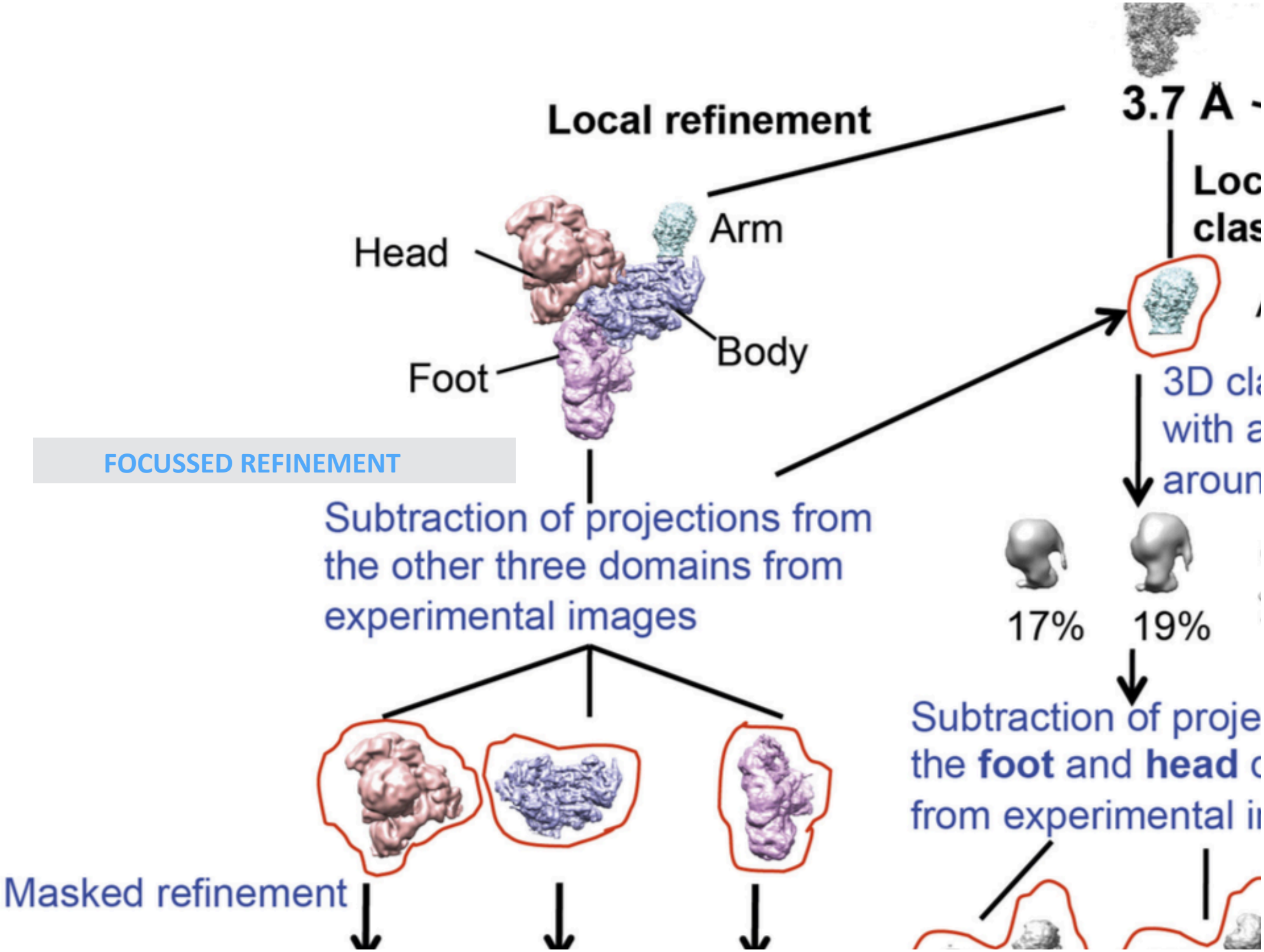

c
473,827 particles
2D classification
438,602 particles
3D classification
140,155 particles
Refine
• Particle polishing
• Refine
Initial model
Low-pass filtered to 60 Å
3.7 Å
Local refinement
Global classification/ refinement
Local classification
Head
Arm
Body
Foot
Subtraction of projections from the other three domains from experimental images
Masked refinement
4.2 Å
3.6 Å
3.7 Å
Arm
3D classification with a mask around the arm
17%
19%
17%
Subtraction of projections from the foot and head domains from experimental images
Masked refinement for the body and arm
6.2 Å
7.5 Å
6.2 Å for the arm
4.6 Å
4.5 Å
4.5 Å for the arm and body
140,155 polished particles
3D classification with finer angular sampling
25%
3%
25%
8%
3%
35%
36,824 particles
Refine
4.3 Å
closed
48,945 particles
Refine
4.2 Å
Open


Local refinement
3.7 Å
Head
Arm
Body
Foot
FOCUSSED REFINEMENT
Subtraction of projections from the other three domains from experimental images
17%
19%
Masked refinement

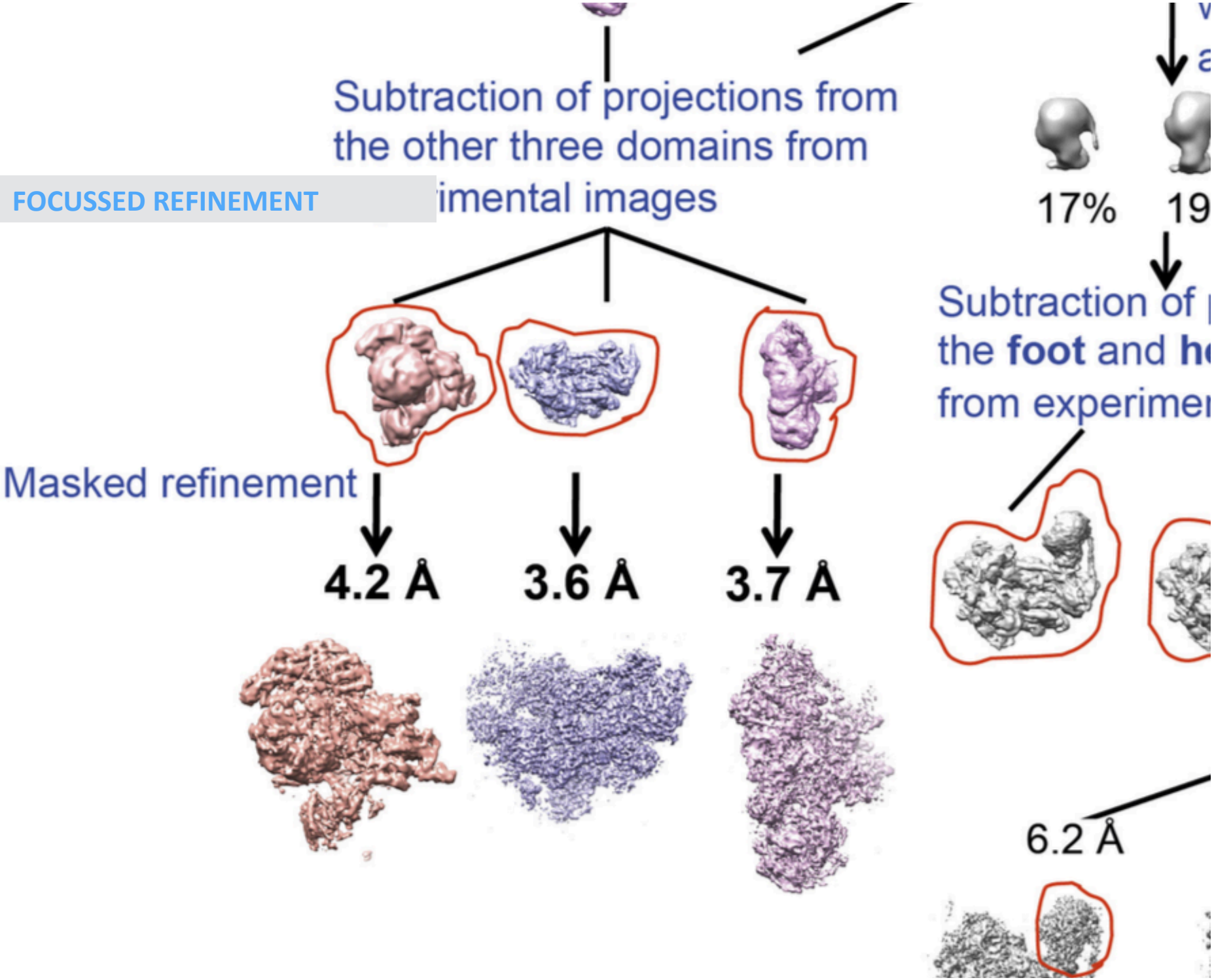
SIGNAL SUBTRACTION

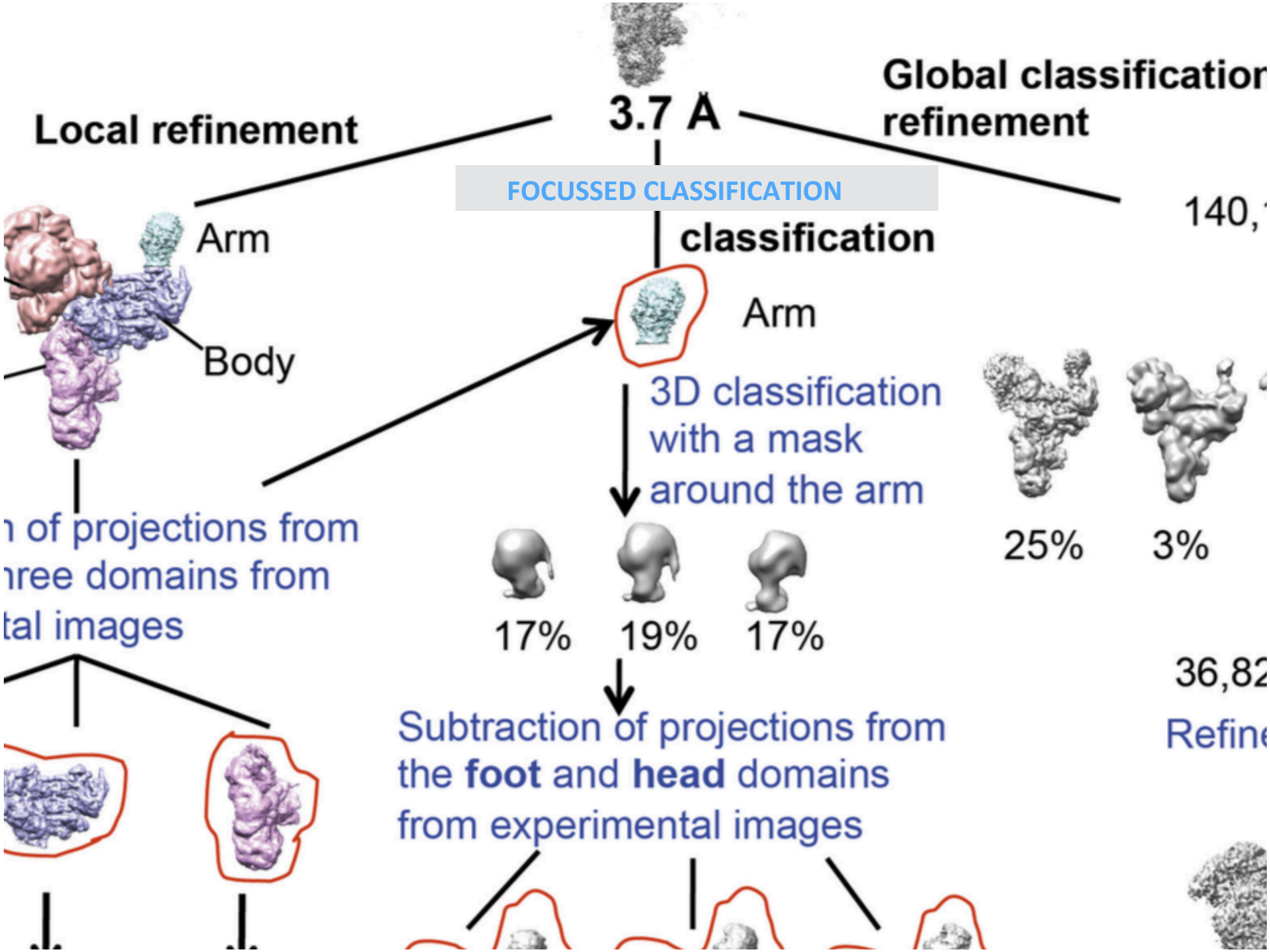


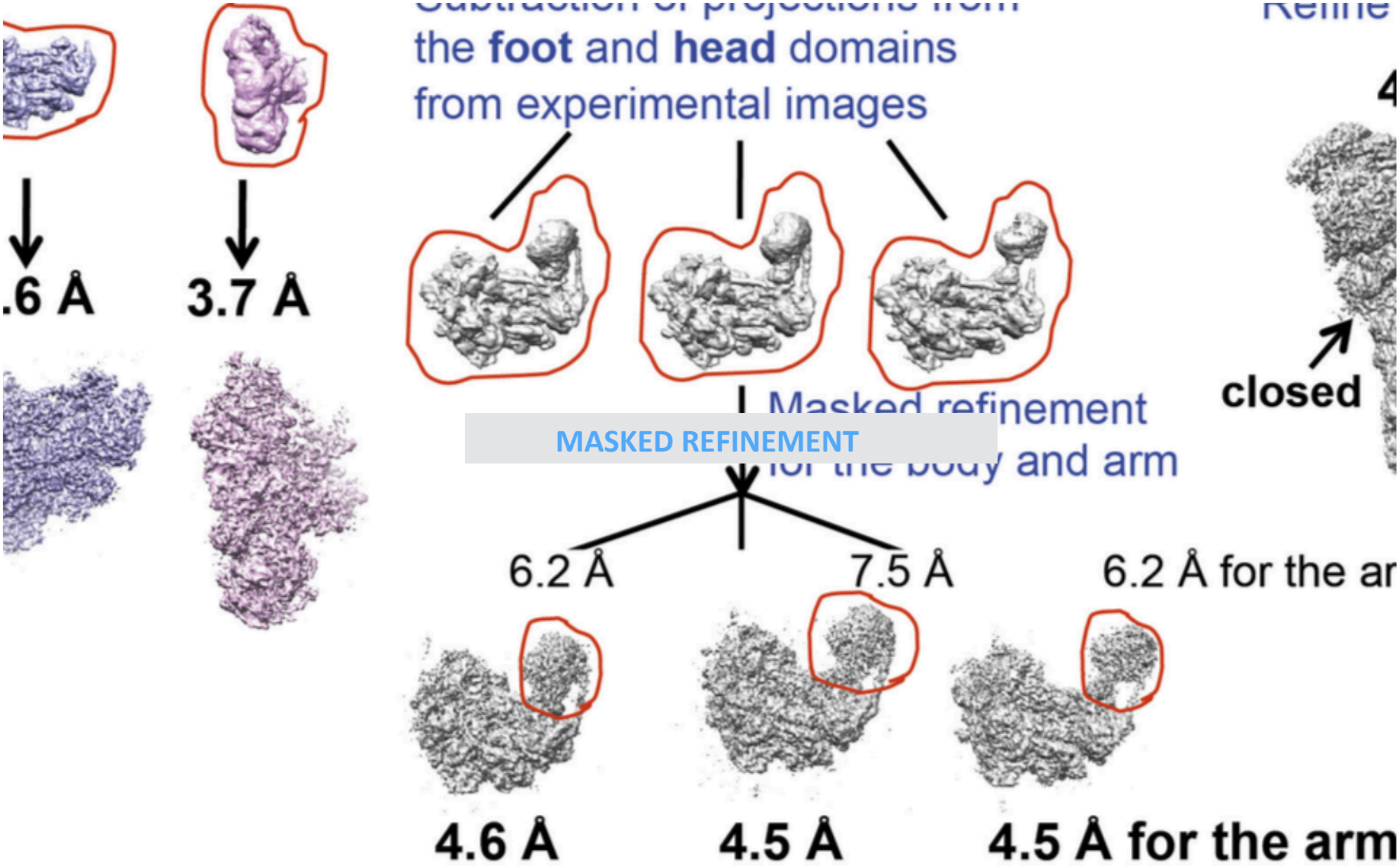
Bai 2016 (eLife)

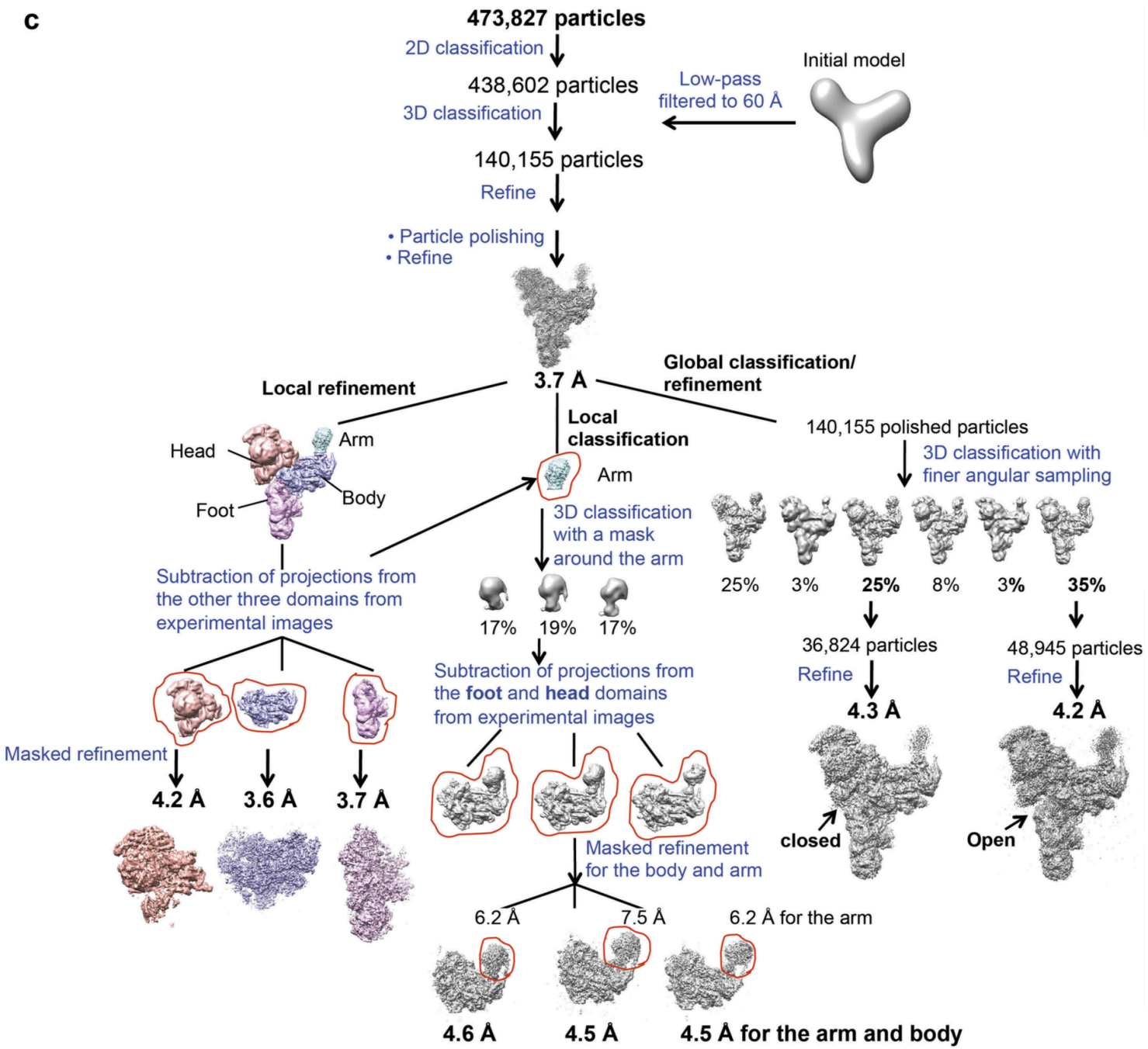

Local refinement
3.7 Å
Head
Arm
Body
Foot
FOCUSSED REFINEMENT
Subtraction of projections from the other three domains from experimental images
17%
19%
Masked refinement

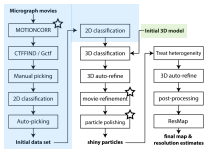
FOCUSSED REFINEMENT



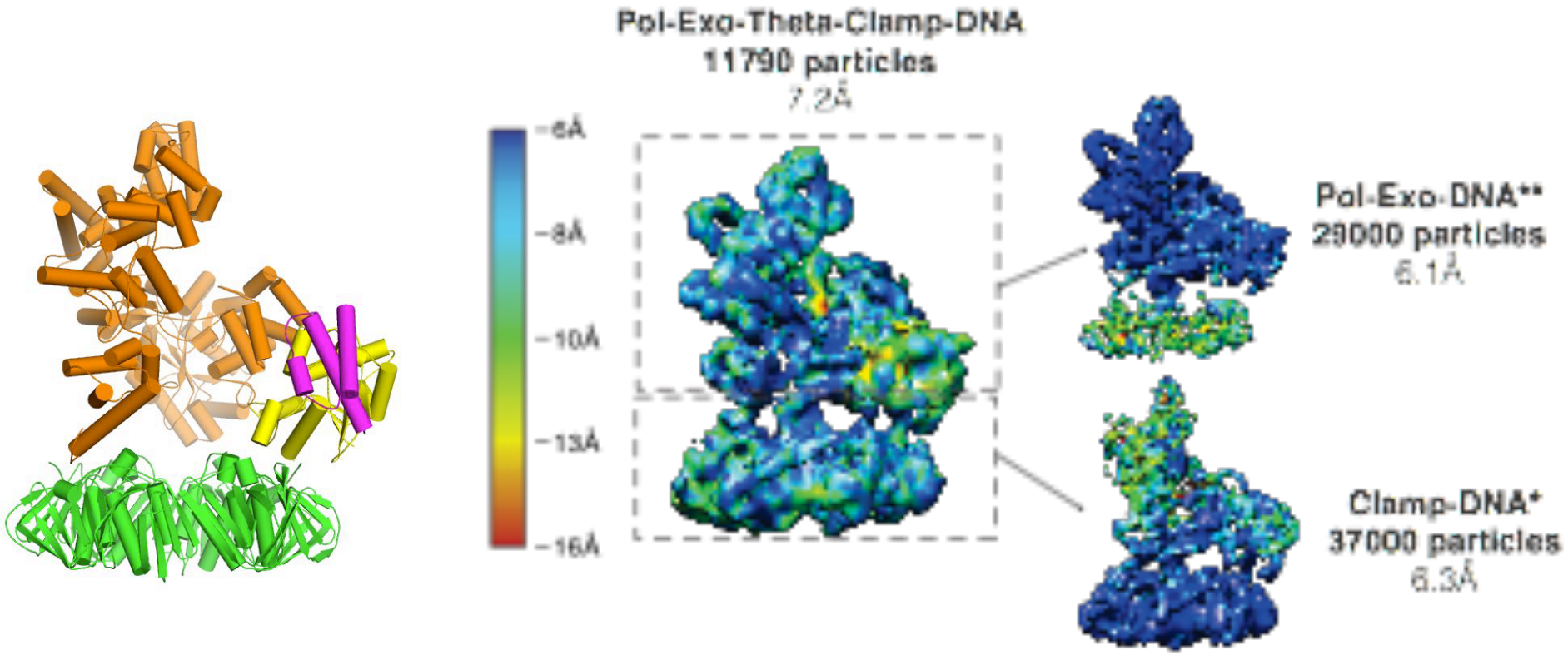

3.7 Å
Local refinement
Global classification
refinement
FOCUSSED CLASSIFICATION
Arm
Body
classification
Arm
3D classification
with a mask
around the arm
25%
3%
17%
19%
17%
Subtraction of projections from
the foot and head domains
from experimental images


.6 Å
3.7 Å
the foot and head domains
from experimental images
Refine
4
closed
Masked refinement
MASKED REFINEMENT
and arm
6.2 Å
7.5 Å
6.2 Å for the ar
4.6 Å
4.5 Å
4.5 Å for the arm


c
473,827 particles
2D classification
438,602 particles
3D classification
140,155 particles
Refine
• Particle polishing
• Refine
Initial model
Low-pass filtered to 60 Å
3.7 Å
Local refinement
Global classification/ refinement
Local classification
Head
Arm
Body
Foot
Subtraction of projections from the other three domains from experimental images
Masked refinement
4.2 Å
3.6 Å
3.7 Å
Arm
3D classification with a mask around the arm
17%
19%
17%
Subtraction of projections from the foot and head domains from experimental images
Masked refinement for the body and arm
6.2 Å
7.5 Å
6.2 Å for the arm
4.6 Å
4.5 Å
4.5 Å for the arm and body
140,155 polished particles
3D classification with finer angular sampling
25%
3%
25%
8%
3%
35%
36,824 particles
Refine
4.3 Å
closed
48,945 particles
Refine
4.2 Å
Open

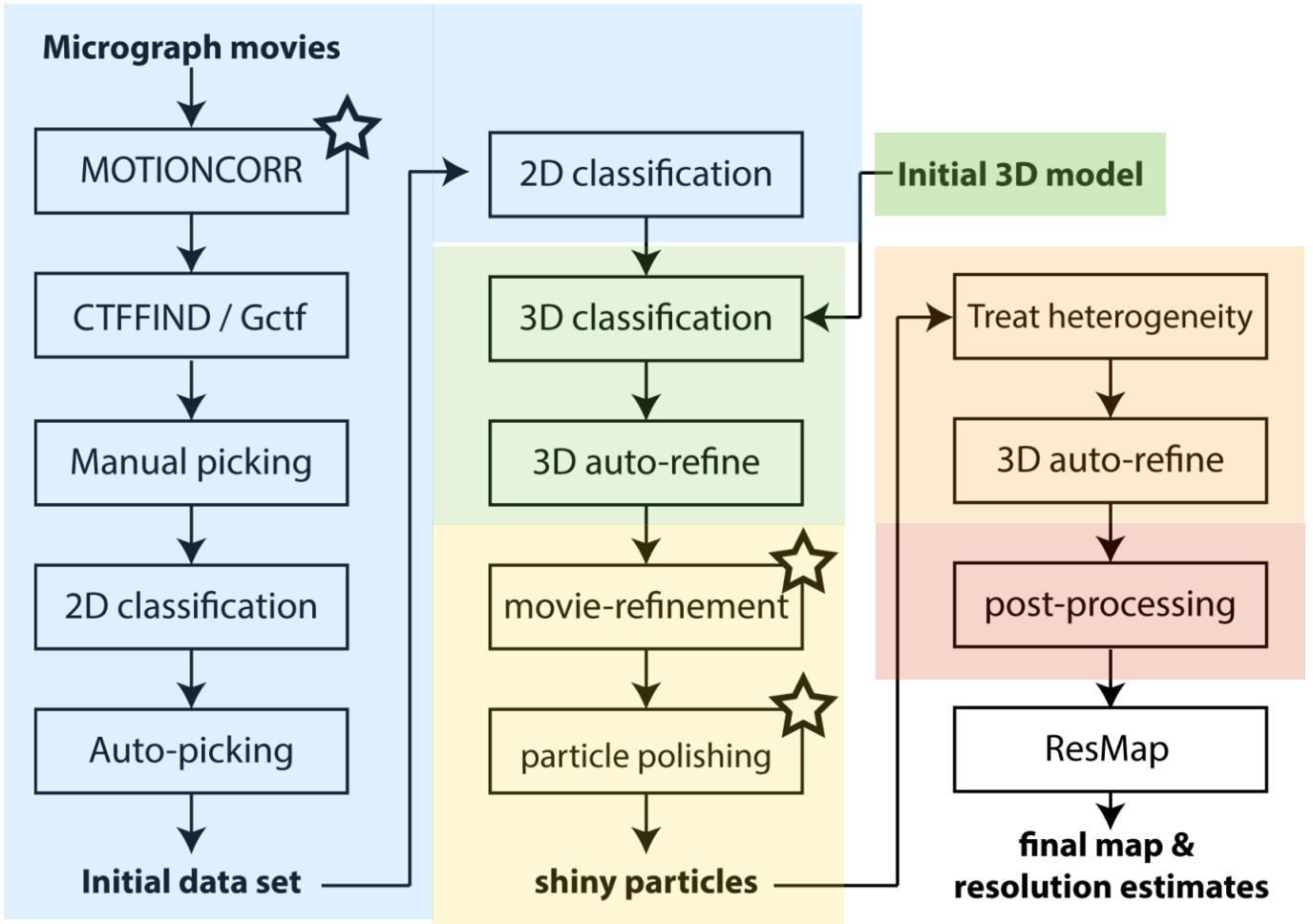


Treating Heterogeneity

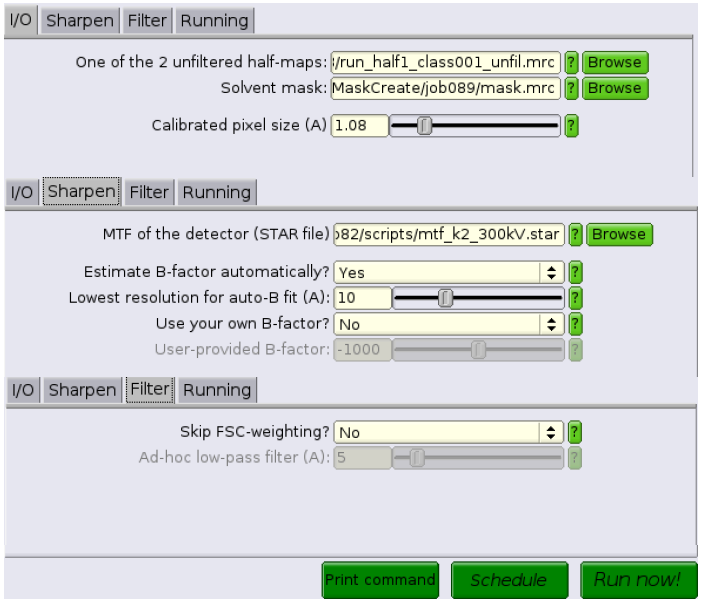


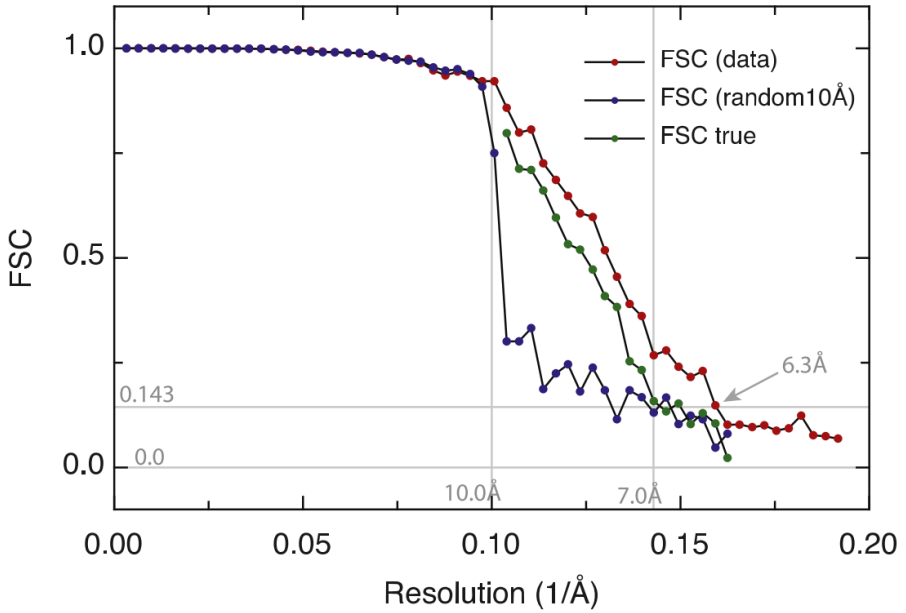
Fernández-Leiro 2016 (NSMB)

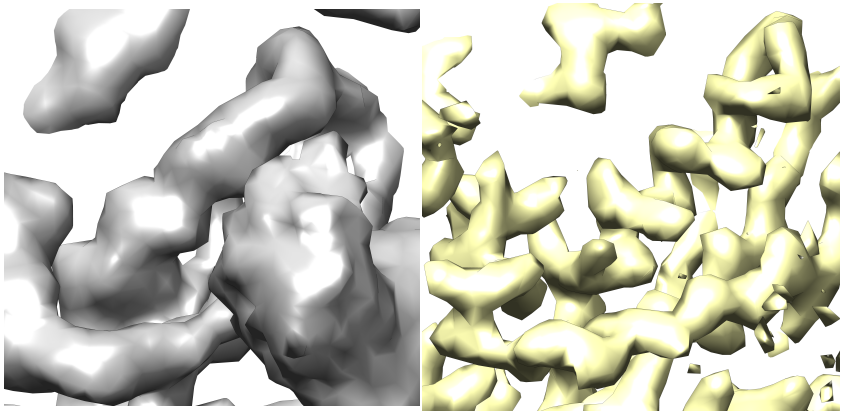
Postprocessing

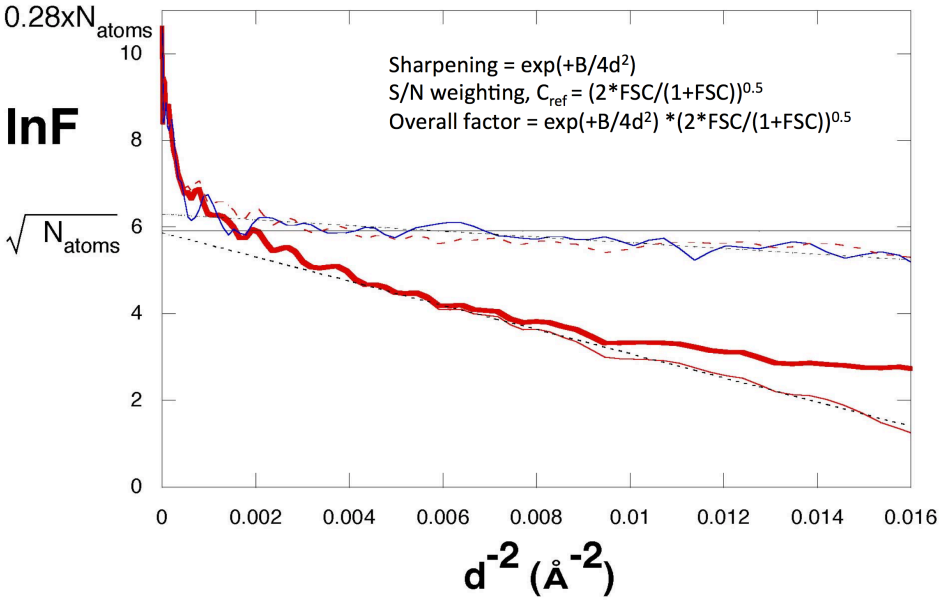


Postprocessing

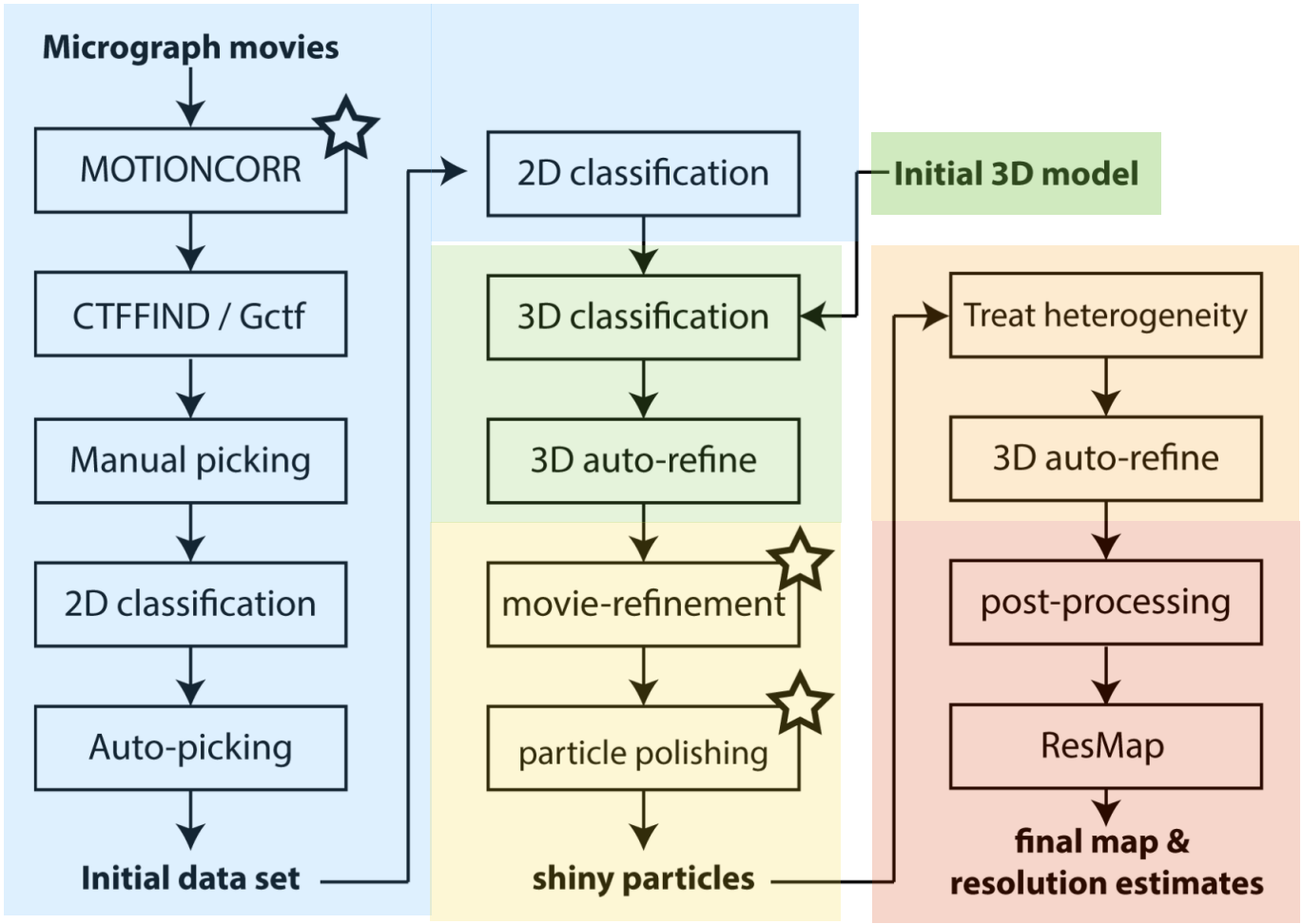


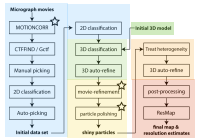






Validation


Micrograph movies
MOTIONCORR
CTFFIND / Gctf
Manual picking
2D classification
Auto-picking
Initial data set
2D classification
3D classification
3D auto-refine
movie-refinement
particle polishing
shiny particles
Initial 3D model
Treat heterogeneity
3D auto-refine
post-processing
ResMap
final map &
resolution estimates



Validation

How do we know if the structure is right?

Can you see the expected structure features for the resolution?

Alpha helices @ 9Å - Beta strands @ 4.8Å

Tilt pair validation

Check that your structure is correct by collecting pairs of micrographs with a known tilt

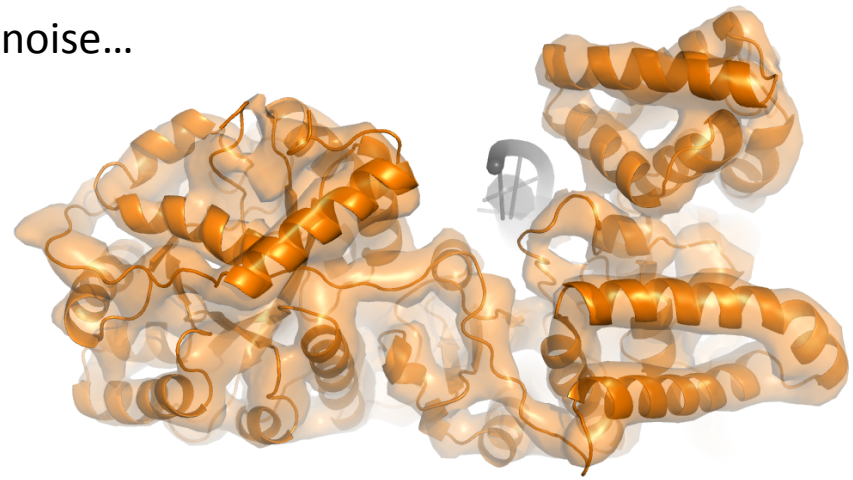
Handedness!

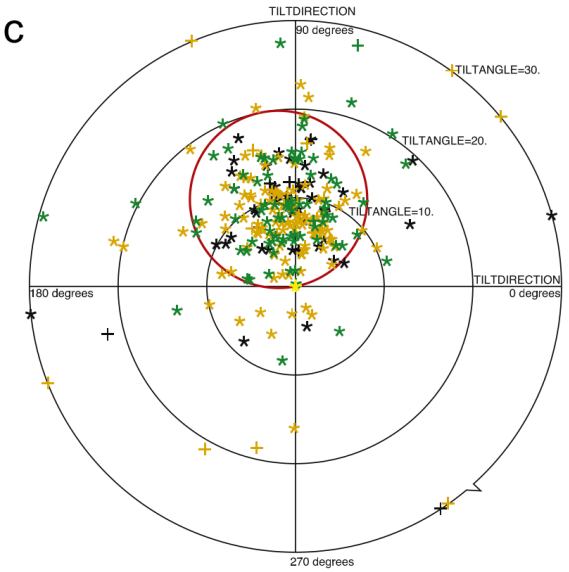
Use secondary structure features to make sure you have the right hand

Tilt pair plot will also help you find the right hand

Can you see your particles?

Avoid Einstein from noise…





Chen 2013 (Ultramicroscopy)

Single particle data processing strategy