ETH zürich

Basics of single-particle image processing

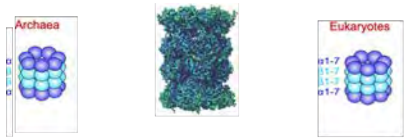
Pavel Afanasyev
Cryo-EM Knowledge Hub

16 March 2021, Midlands Regional Cryo-EM Workshop 2021

1

Agenda

- Describe the single-particle pipeline
- Explain basic ideas
- Illustration with a real-life example (20S parasite proteasome)



- barrel of rings of 2x7 identical α and 2x7 identical β subunits
- D7 symmetry
- well behaved sample on standard EM grids

- Very good test sample for EM -

- barrel of 2 copies of 7 alpha and 7 beta homologous subunits (pseudo 7-fold rings)
- C2 symmetry + pseudosymmetry
- complexes from higher eukaryotes take strongly preferential orientations on EM grids

- Not so good sample for EM -

ETH zürich 16 March 2021 Pavel Afanasyev 2

2

Transmission Cryo-Electron Microscopy
A tool used by structural biologists to study molecular nanomachines

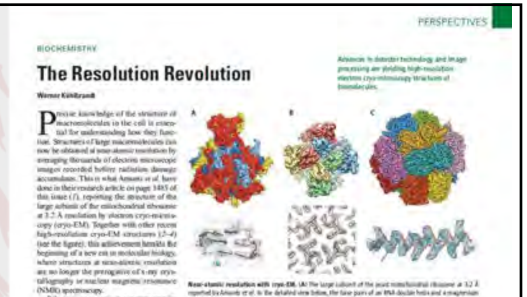
Gabriel Lander, Thesis Defense 2009

ETH zürich 16 March 2021 Pavel Afanasyev 3

3

Advances

- Direct electron detectors
- Microscopes
- Automatic faster data collection
- Better software/methodology/ algorithms
- Computational resources/ costs


PERSPECTIVES
BIOCHEMISTRY
The Resolution Revolution

ETH zürich

4

What is cryo-EM limited by?

- Radiation damage (proteins typically tolerate up to ~80 e/Å^2)
  - (very) high noise levels
- Instrumentation limits (detection of electrons)
  - usually low signal at both low and high frequencies
- Sample size, heterogeneity and behaviour on the EM grids
  - difficulties with detecting particles, their orientations and assignment to distinct conformations/components
  - preferred orientation, low concentration, aggregation etc.

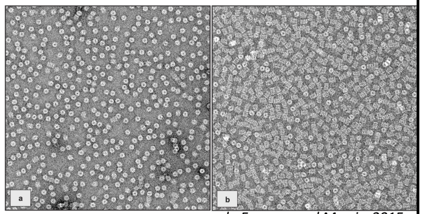
ETH zürich 16 March 2021 Pavel Afanasyev 5

5

Outline of the cryo-EM experiment

- Biochemistry optimisation (buffer components; cross-linking; detergents etc.)
- Negative staining: sample optimisation (concentration, grids, support, glow discharge settings, treatment of grids etc.)
- Cryo-EM optimisation with a screening microscope (concentration, ice, air-water interface etc.)
- High-resolution data collection and processing
- Iterative procedure!

20S proteasomes with and without pentylamine



da Fonseca and Morris, 2015

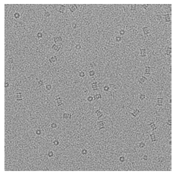
ETH zürich 16 March 2021 Pavel Afanasyev 6

6

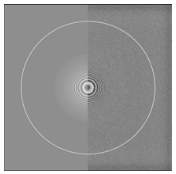
Pipeline of image processing

Micrographs pre-processing

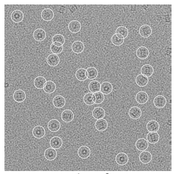
1. Movie alignments



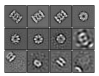
2. CTF-determination



3. Particle picking

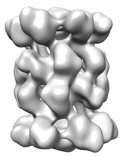


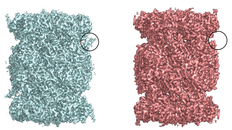
4. 2D-classification



5. Initial model generation

6. 3D-refinements and 3D-classification



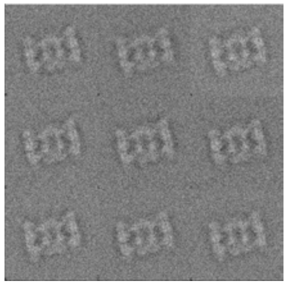


7. Particle polishing and CTF-refinement

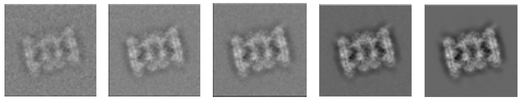
ETH zürich 16 March 2021 Pavel Afanasyev 7

7

Denoising by averaging



Sum of 3 | Sum of 10 | Sum of 20 | Sum of 100 | Sum of 1000

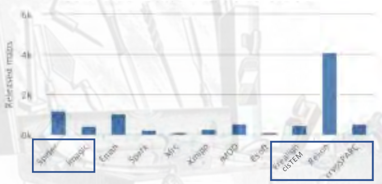


ETH zürich 16 March 2021 Pavel Afanasyev 8

8

Software

Pioneering software:
- large libraries of efficient "old-school" tools
- often require expert knowledge and take time to learn



Modern software:
- computationally faster (GPU)
- user-friendly
- rapidly developing and well-maintained

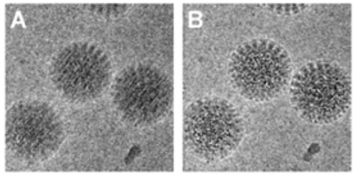

Joachim Frank

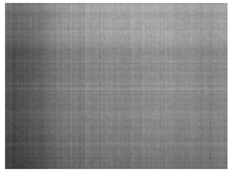

Marin van Heel

Source: EMDB-EBI, 2020

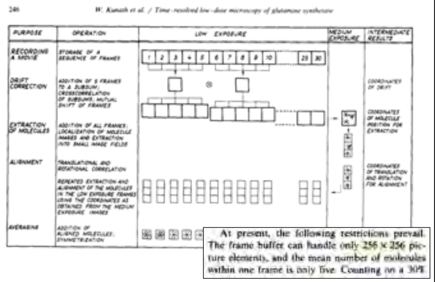
9

Movie-mode data collection





Brilot et al. 2012



Kunath et al. 1984

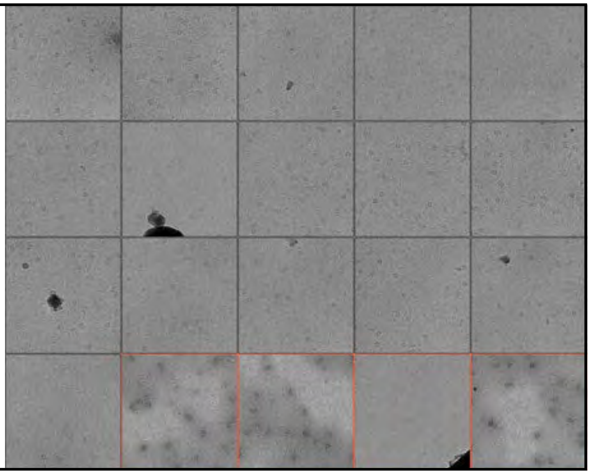
Correction for:
- Drifts during data acquisition (global motion)
- Beam-induced motion (local motion)

+ dose weighting (to account for radiation damage)

10

Data screening

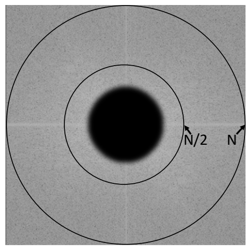
- Bin micrographs
- Screen all and select bad:
  - thick ice
  - dry/empty areas
  - contamination
  - cracked support
  - high drift/charging



11

Understanding image data in reciprocal space

Low frequency details (contrast)



Signal sampled at a rate f can be fully reconstructed if it contains only frequency components below half that sampling frequency: f/2 ("*Nyquist frequency*").

High frequency details (resolution)

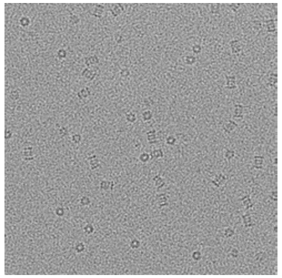
12

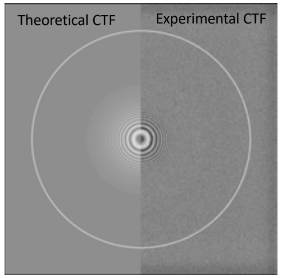
CTF-determination

CTF correction:

Phase flipping

Amplitude correction



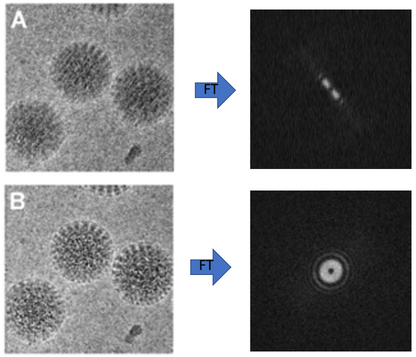


Sample: Parasite 20S proteasome
Microscope: Titan Krios
Camera: Falcon 3 (counting mode)
Pixel size: 0.81 Å
Total dose: 45 e/Å^2

The research performed at P. da Fonseca lab
Sample and inhibitors provided by M. Bogyo lab

13

Analysis of the power spectra

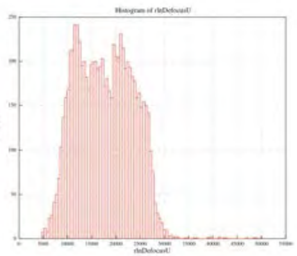


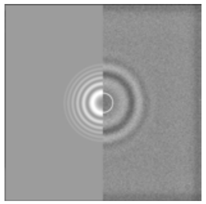
Power spectrum of an image with large drift *before* alignments

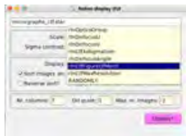
Power spectrum of an image with large drift *after* alignments

14

Selection of micrographs based on the power spectrum



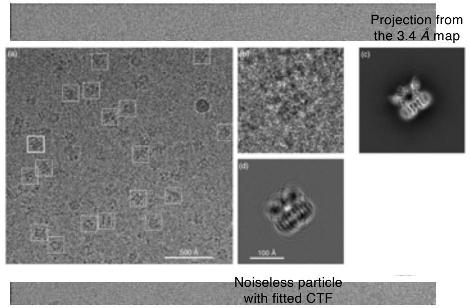




Checklist
- Histograms (defoci, astigmatism etc.)
- Quality of the fits
- Resolution
- Diffraction from crystalline ice

15

Typical case: low signal, problems with particle picking



FEI Titan Krios; Carbon-coated golden grids; detector: K3 (counting mode); 0.84 Å/pix; total dose: 45 e/Å^2

Simulations on TRPV1 protein from Sigworth, 2015: 300 keV, 1.22 Å/pix, -2.2 μm defocus

- Why don't we see particles?

Imaging:
- low defocus
- low dose
- microscope (voltage, camera etc.)

Sample:
- thick ice
- particle size
- buffer components (glycerol)
- low concentration
- aggregation on the quantifoil

16

Particle picking approaches

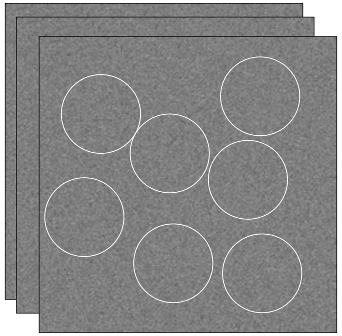
- Manual selection
- Automated reference-free
- Automated template-based (internal and external references)
- Machine-learning based approaches
  - -Topaz, CRYOLO, cryoSPARC

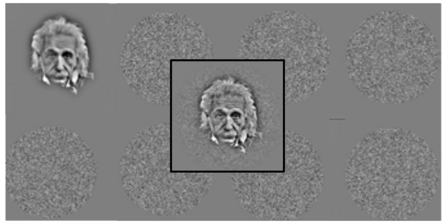
Issues:

- False-negatives (poor contrast)
- False-positives (reference-bias)

17

Reference bias: Einstein from random white noise



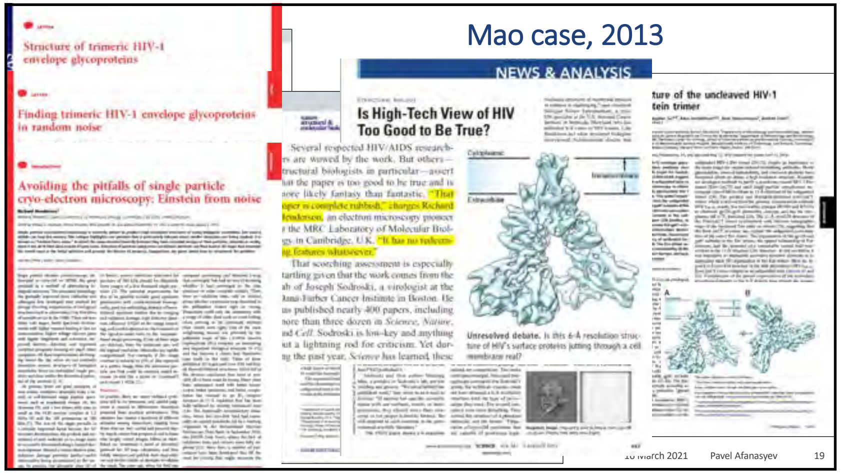


Micrographs with simulated noise, no particles or Einstein photos added

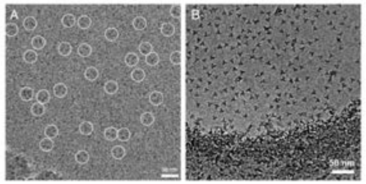
Experiment from M. Schatz

18

Mao case, 2013



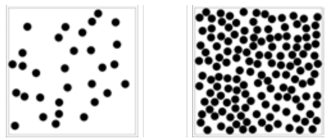
19

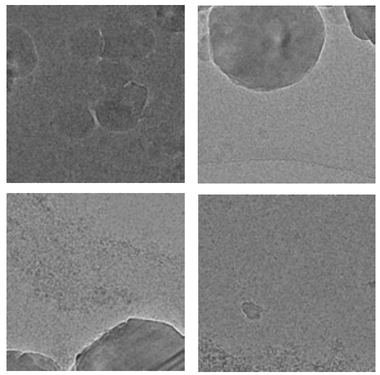


Subramaniam, 2013

Particle distribution (concentration)

Presented in the paper | Minimal to pick 670,000 particles

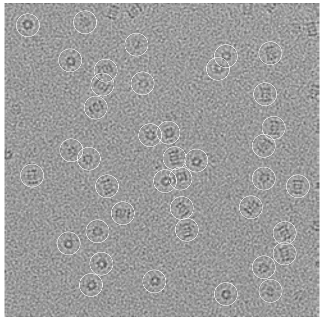




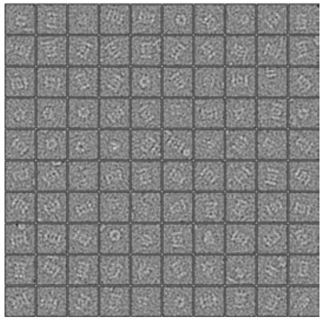
EMPIAR-10003: ~65% of micrographs are unusable (Deposited on 12/7/2015)

20

Particle extraction



Applied operations
- Cutting
- Normalization
- Binning
- Contrast inversion

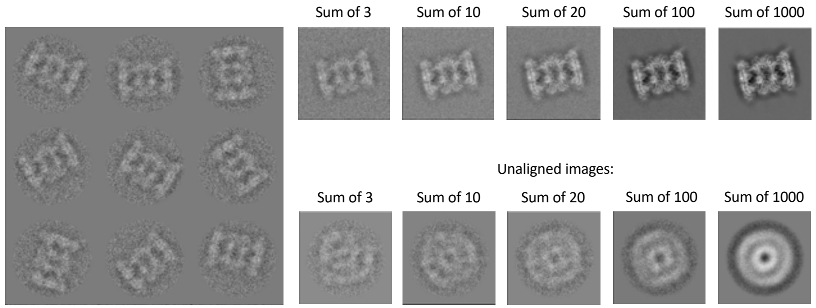


Checklist
- Centering
- Box size
- Binning
- Junk

*ETH*zürich 16 March 2021 Pavel Afanasyev 21

21

Denoising by averaging



Unknowns: X, Y: shifts + Ψ: rotational angle

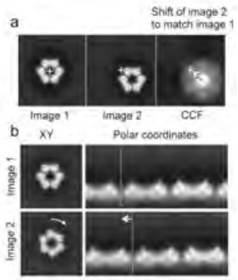
*ETH*zürich 16 March 2021 Pavel Afanasyev 22

22

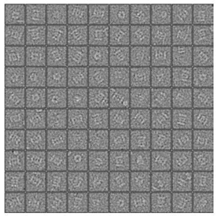
Alignments and 2D-classification

Reference-free probabilistic approach

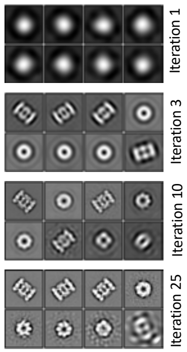
Unknowns: X, Y: shifts + Ψ: rotational angle



Orlova and Sabil, 2010



Alignments and classification



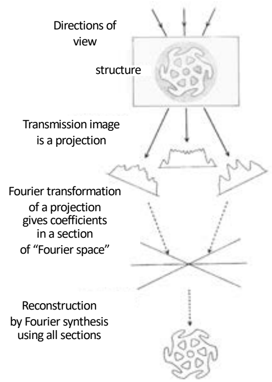
Results:
X, Y: shifts
Ψ: rotational angle
probability to belong to certain class-average

*ETH*zürich 16 March 2021 Pavel Afanasyev 23

23

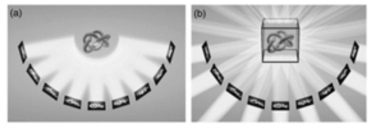
3D-reconstruction

Central section theorem:

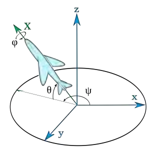


De Roisier and Klug, 1968

Back-projection method



Baumeister et al., 1999



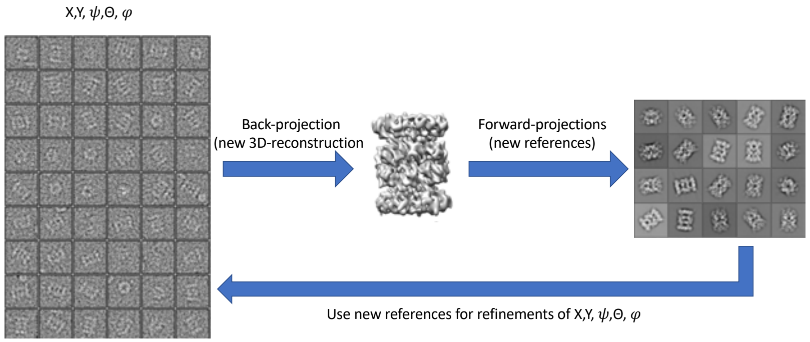
From Wikipedia

Unknowns
X, Y: shifts
ψ: rotational angle (in plane)
θ: out of plane tilt
φ: rotation around Z

*ETH*zürich 16 March 2021 Pavel Afanasyev 24

24

Projection matching algorithm



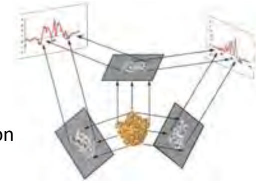
Warning: overfitting!

ETH zürich 16 March 2021 Pavel Afanasyev 25

25

Initial 3D-reconstruction

- External reference (strongly filtered)
- Stochastic gradient descent (cryoSPARC, relion)
- Angular reconstitution
  - Random startup or manually in IMAGIC or Spider based on 2D class-averages (very powerful tool in tricky cases)
- Tomography/ Random conical tilt



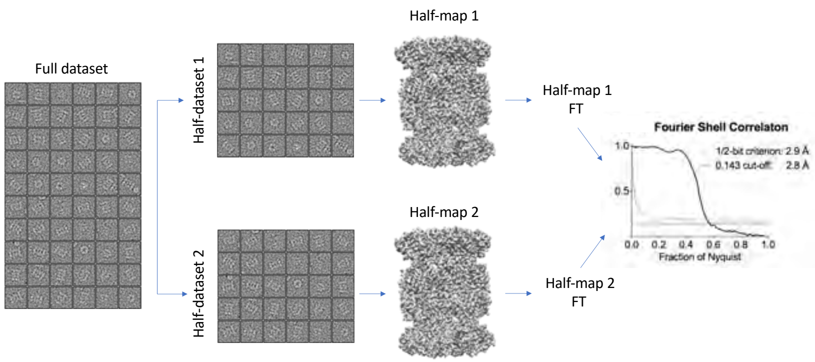
van Heel et al., 2000

3D-refinements are sensitive to the initial 3D-reconstruction!

ETH zürich 16 March 2021 Pavel Afanasyev 26

26

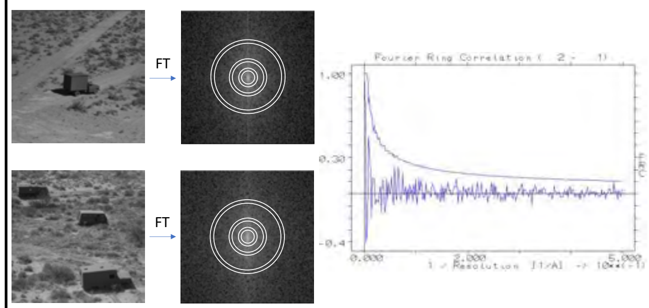
Refinements and Validation



ETH zürich 16 March 2021 Pavel Afanasyev 27

27

Fourier Shell (Ring) Correlation



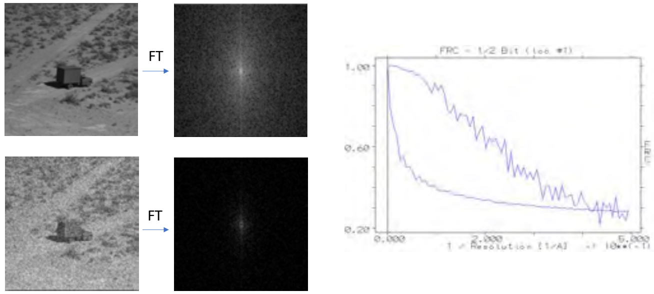
$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt[2]{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}}$$

van Heel et al., 1982
Saxton & Baumeister, 1982
Harauz & van Heel, 1986

ETH zürich 16 March 2021 Pavel Afanasyev 28

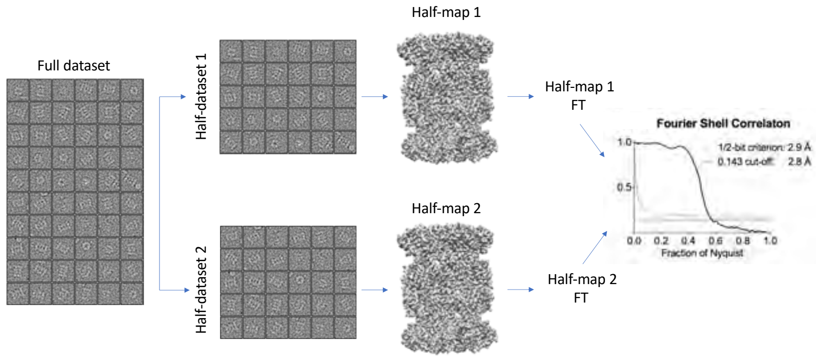
28

Fourier Shell (Ring) Correlation



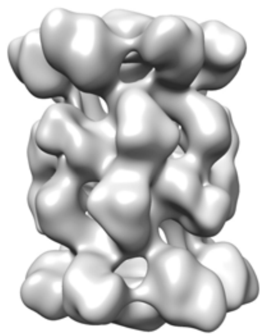
29

Refinements and Validation



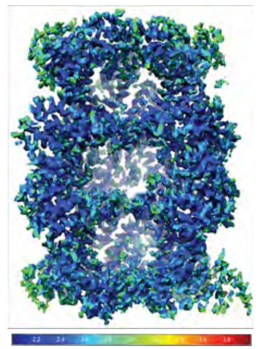
30

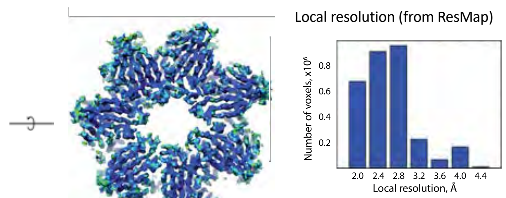
3D-refinements



31

Evaluation of the quality of the map

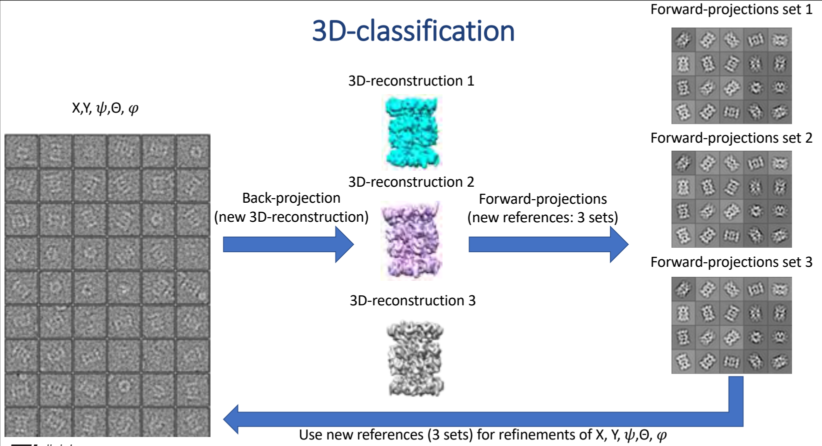




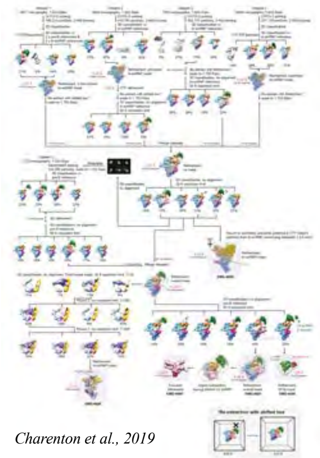
Note: local resolution estimations are often not reliable!

32

3D-classification



33



Charenton et al., 2019

3D-classification

Heterogeneity

- Compositional (can be sorted at 2D-classification)
- Conformational

Approaches:

- Using multiple 3D-reconstructions for creating projection-matching references
- Using masks and signal subtraction (focused classification)
- Multibody-refinement
- Other approaches (3D variability analysis, 3D-MSA)

34

Further steps

- CTF refinement (defocus, beamtilt values)
- Particle polishing (beam-induced motion correction)
- Post-processing (masking and filtering)
- Validation! (resolution, handedness)

35

Building a pipeline

- All the steps can (in theory) be fully done with each of these packages: cryoSPARC/Relion/cisTEM/Scipion.
- Once a reliable and efficient pipeline is tailored for the same type of data, structural analysis of different mutants or binding of substrates is becoming a routine (at high resolution)

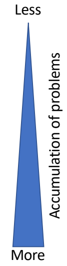
36

On-the-fly data processing

Often done during the data collection

Software: cryoSPARC, Relion, Scipion, Warp, cryoFLARE etc.

- Movie alignments
  - motioncor2, UNBLUR
- CTF-determination
  - GCTF, CTFFIND4
- Particle picking
  - Gautomatch, CRYOLO, Topaz
- 2D-classification
- Initial 3D



37

Take-home messages

- Single-particle analysis is an iterative process, requiring optimisation:
  - **purification**
  - **sample preparation**
  - data collection
  - image processing
- There are many software packages and programs
  - none of them is the best and they are often complementary
  - the image-processing pipeline should be tailored to your sample
- Validation of the results at each step is of paramount importance!

38

Thank you for your attention!

39