Lecture 9

# Interpretation and optimization of cryo-EM maps

9th December 2020
Jiri Novacek

Content

- symmetries

- map validation

- map interpretation

- model building

- map improvement

Symmetries

- regular assemblies of protein oligomers are common in nature

- oligomeric protein structures obey certain rules
    → no mirror symmetry

- understanding symmetry rules may prevent incorrect interpretation of the data

- presence of symmetry generally facilitates determination of the density map
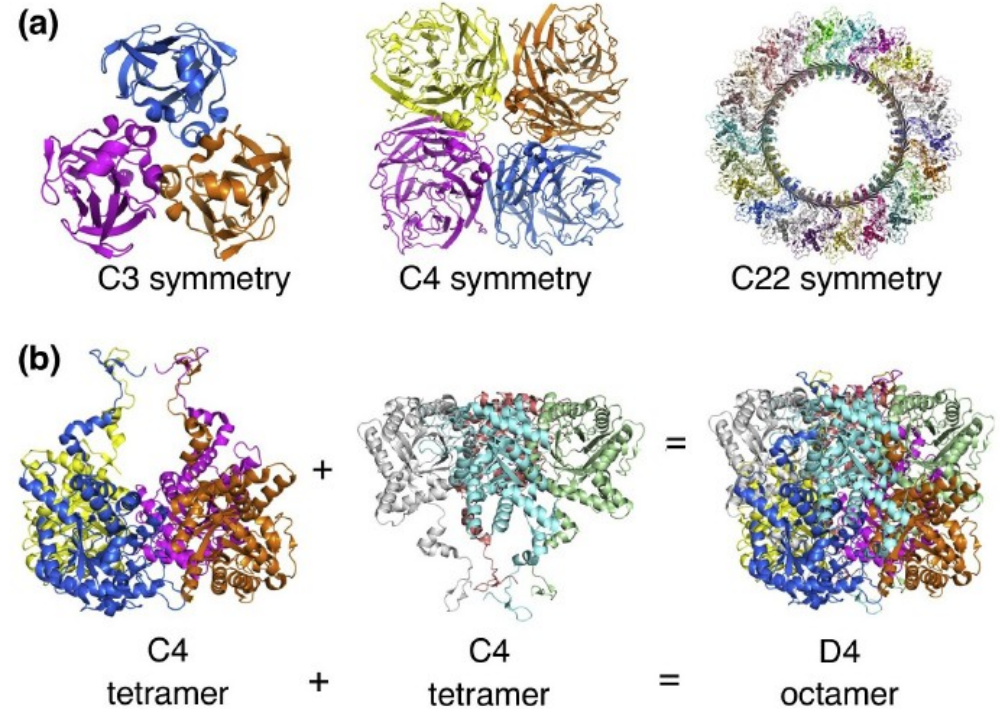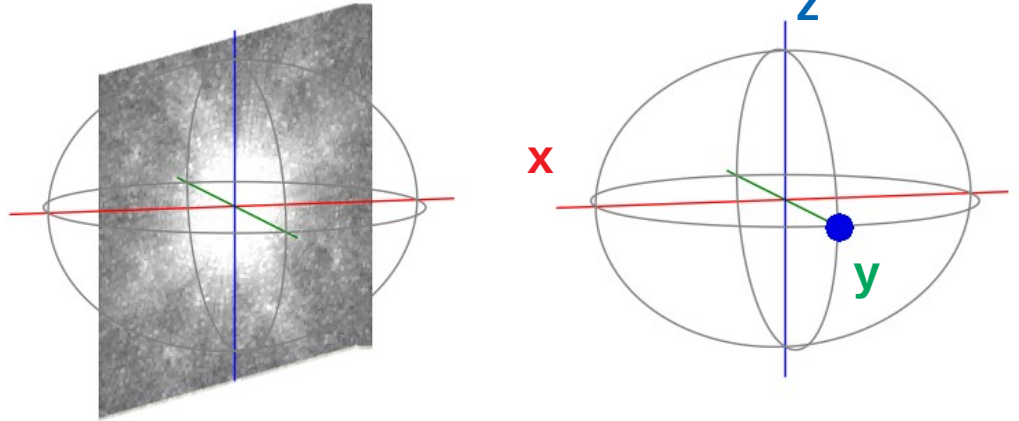

(Xu et al., Curr Opin Struct Biol 2019)

Symmetries

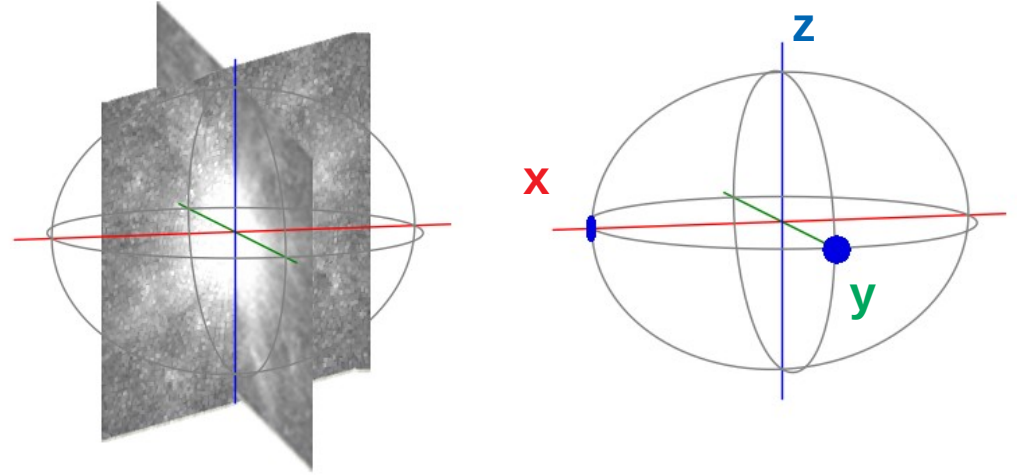## Projection Theorem, Euler angles

- A central section through the 3D Fourier transform is the Fourier transform to the projection in that direction

Symmetries

## Projection Theorem, Euler angles

- A central section through the 3D Fourier transform is the Fourier transform to the projection in that direction

- Images for all possible projection directions are required to obtain structure with homogeneous resolution in all directions

Symmetries

## Projection Theorem, Euler angles

- A central section through the 3D Fourier transform is the Fourier transform to the projection in that direction

- Images for all possible projection directions are required to obtain structure with homogeneous resolution in all directions

- Euler angles φ and θ cover ranges of (0° - 360°) and (-90° - +90°)
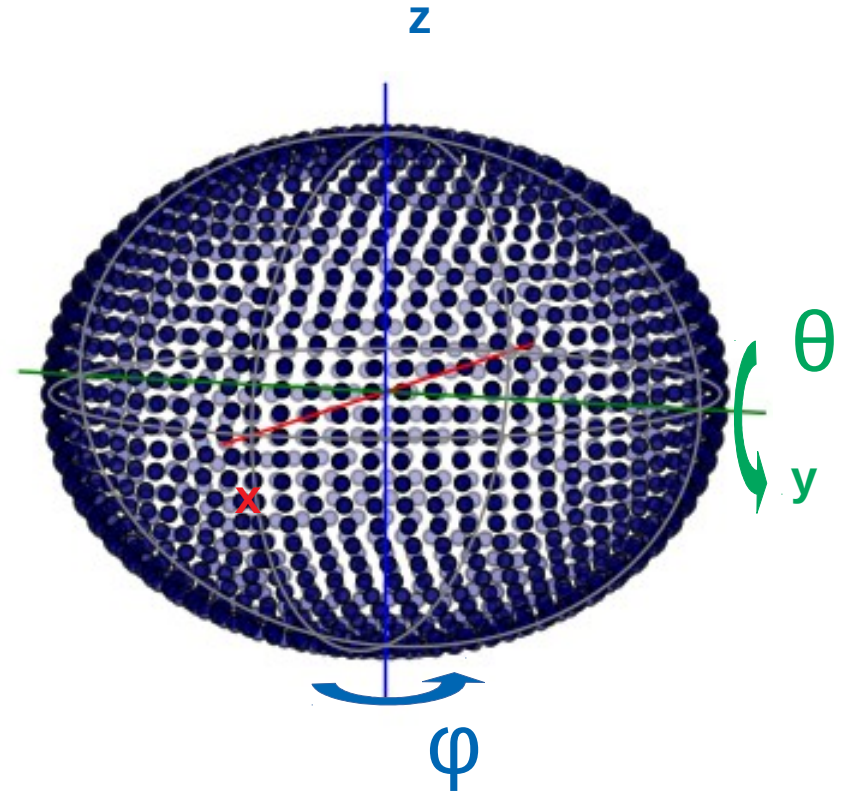
Symmetries

Rotational (cyclic) symmetries

 - one symmetry axis (usually molecules oriented with the symmetry axis alongside z)

 - Asymmetric unit – the smallest portion of the angular space to which symmetry operation can be applied in order to completely fill the angular space

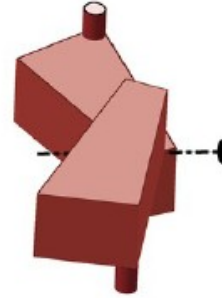 - C1 – the most trivial case, no symmetry, φ (0° - 360°), θ (-90° - +90°)
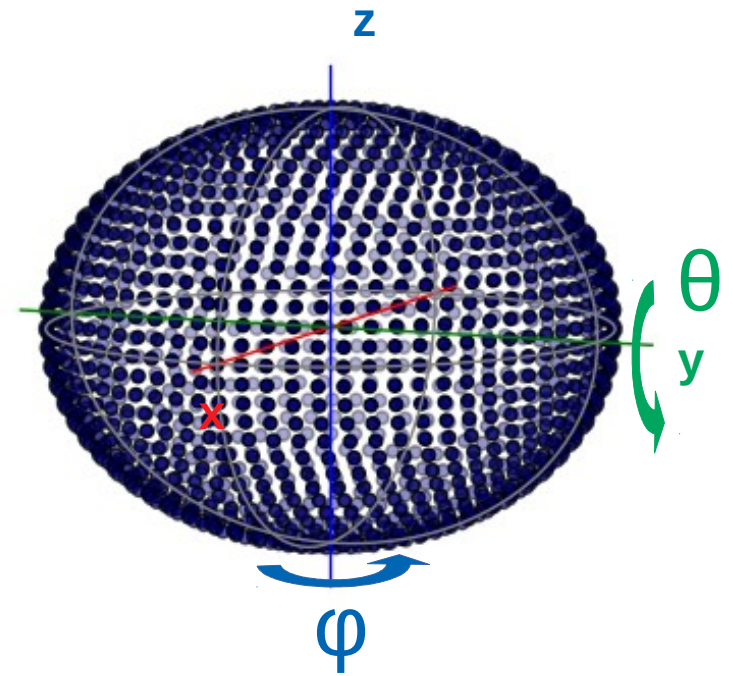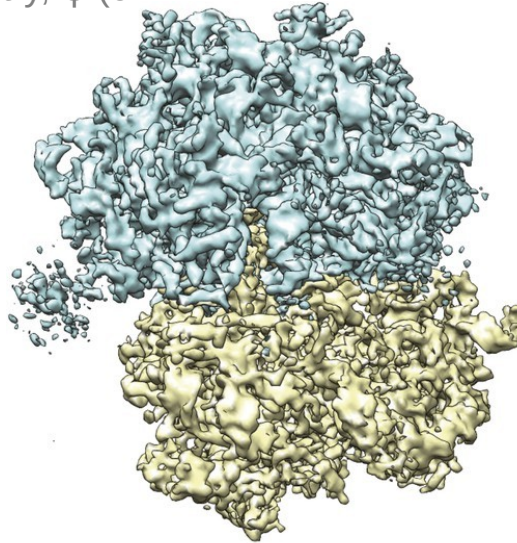
Symmetries

Rotational (cyclic) symmetries

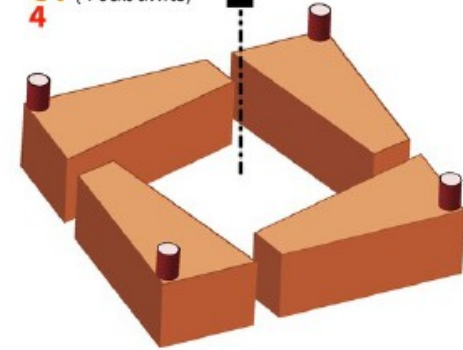- one symmetry axis (usually molecules oriented with the symmetry axis alongside z)

- Asymmetric unit – the smallest portion of the angular space to which symmetry operation can be applied in order to completely fill the angular space

- C2 – φ (0° - 180°), θ (-90° - +90°)
- C3 – φ (0° - 120°), θ (-90° - +90°)
- C4 – φ (0° - 90°), θ (-90° - +90°)
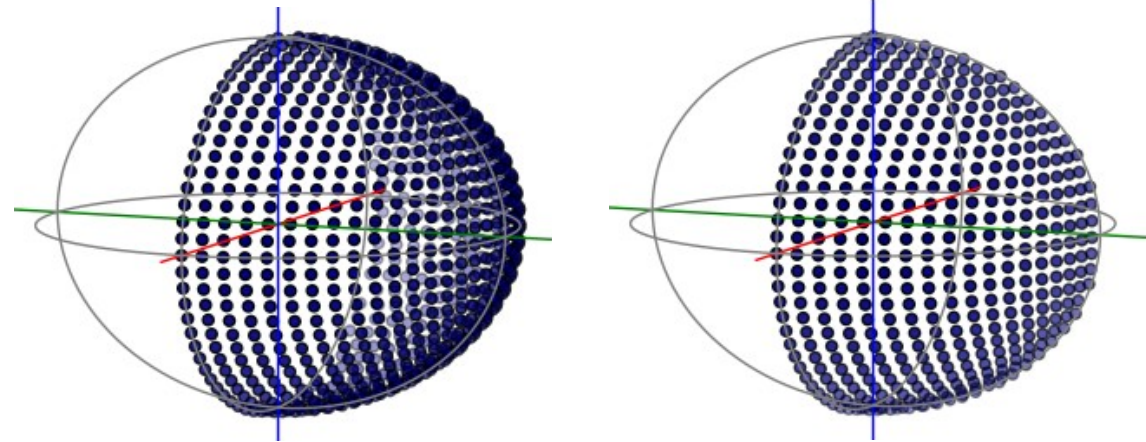- C6 – φ (0° - 60°), θ (-90° - +90°)
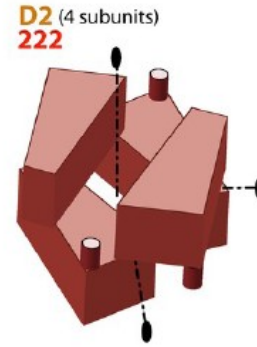

(Levy et al., PLoS computational Biology 2006)

Symmetries

Dihedral symmetries

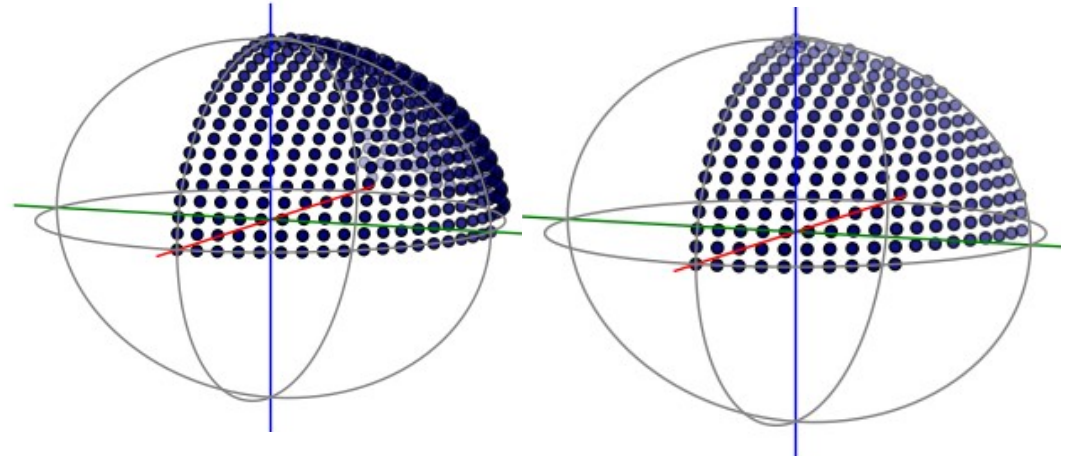- one n-fold rotational axis and two-fold axis perpendicular to it

- Asymmetric unit
  - D2 – φ (0° - 180°), θ (0° - +90°)
  - D5 – φ (0° - 72°), θ (0° - +90°)
  - D7 – φ (0° - ~51°), θ (0° - +90°)

(Levy et al., PLoS computational Biology 2006)

Symmetries

Platonic symmetries
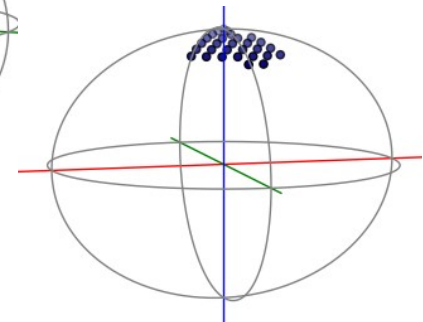
- faces, edges, and corners are related by symmetry operations

- tetrahedral – 4 3-fold axes and 3 2-fold axes
- octahedral – 3 4-fold axes of symmetry, 4 3-fold axes of symmetry, and 6 2-fold axes
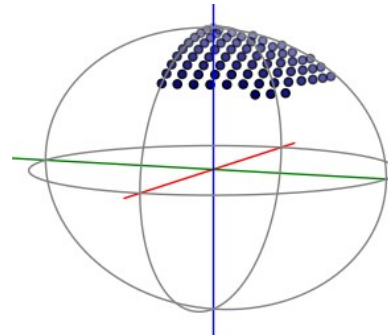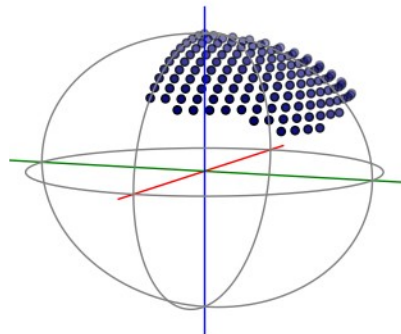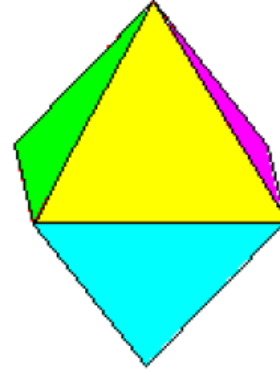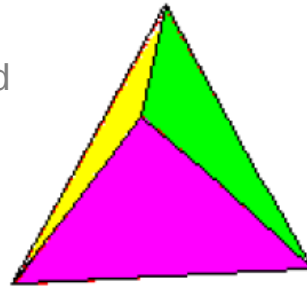- icosahedral – 6 5-fold, 10 3-fold and 15 2-fold axes

Symmetries

Platonic symmetries

- faces, edges, and corners are related by symmetry operations

- tetrahedral – 4 3-fold axes and 3 2-fold axes
- octahedral – 3 4-fold axes of symmetry, 4 3-fold axes of symmetry, and 6 2-fold axes
- icosahedral – 6 5-fold, 10 3-fold and 15 2-fold axes

EMAN2
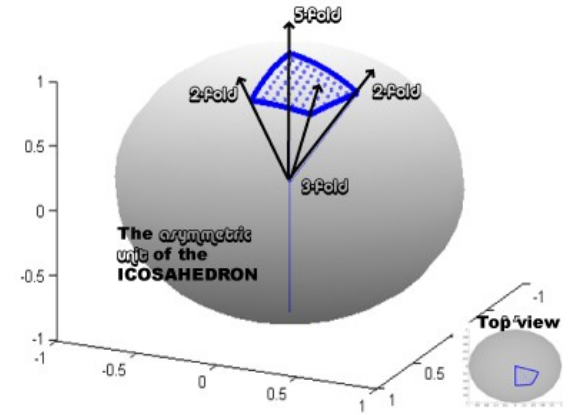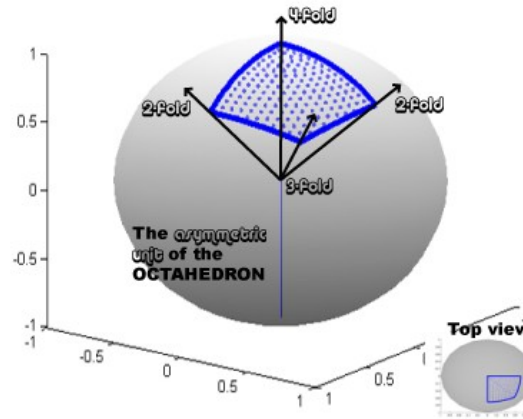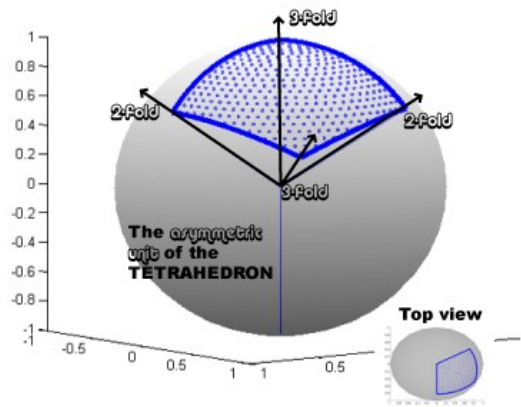
Symmetries

Helical symmetry

 - A single view contains all the necessary info for 3D reconstruction

 - 2D surface lattice rolled into 3D
 - 3D reconstruction approaches:
   - Fourier-Bessel analysis
   - Iterative Real-Space Refinement (IHRSR)
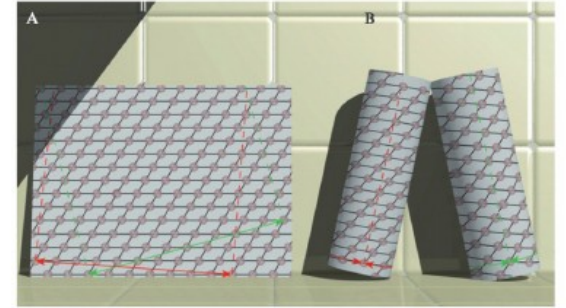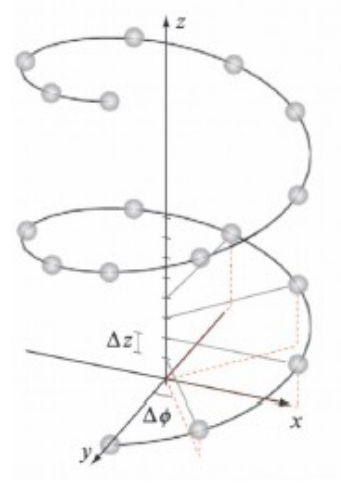
(Diaz et al. Methods in Enzymology 2010)
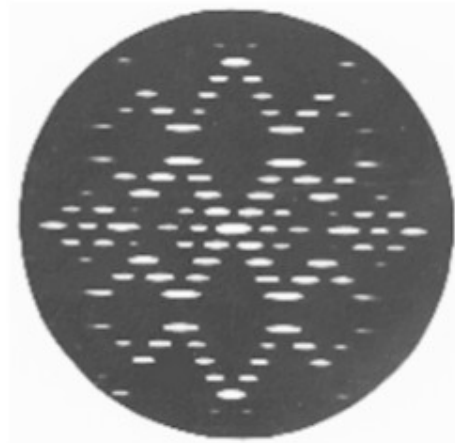
Symmetries

Helical symmetry

- A single view contains all the necessary info for 3D reconstruction

- 2D surface lattice rolled into 3D
- 3D reconstruction approaches:
  - **Fourier-Bessel analysis**
  - Iterative Real-Space Refinement (IHRSR)

- small inaccuracies in indexing lead to incorrect structure
- requires strict helical symmetry
- requires flat straight helices
- laborious

Symmetries

Helical symmetry

- A single view contains all the necessary info for 3D reconstruction

- 2D surface lattice rolled into 3D
- 3D reconstruction approaches:
    - Fourier-Bessel analysis
    - **Iterative Real-Space Refinement (IHRSR)**

- requires fairly good estimate of the cylinder diameter, rise, and twist
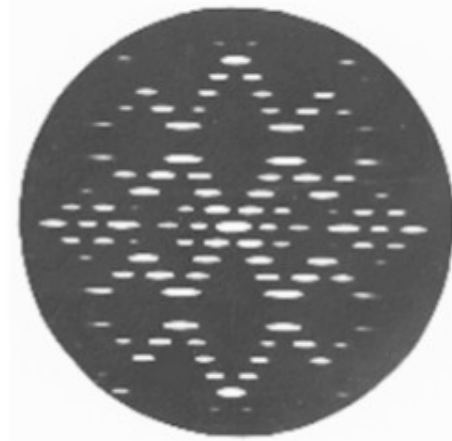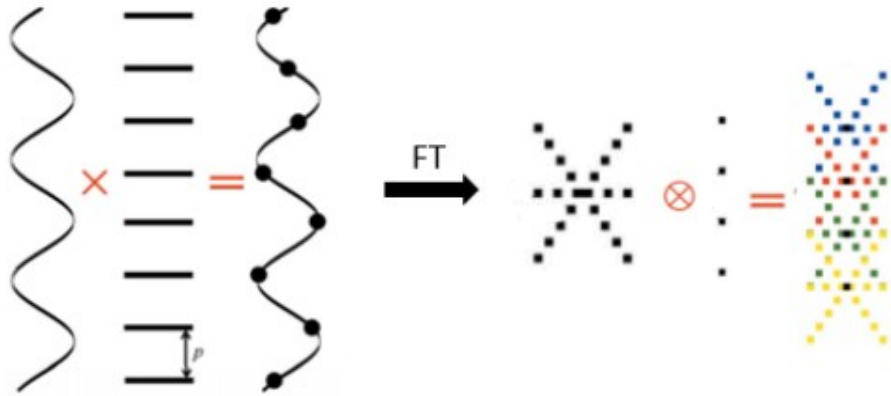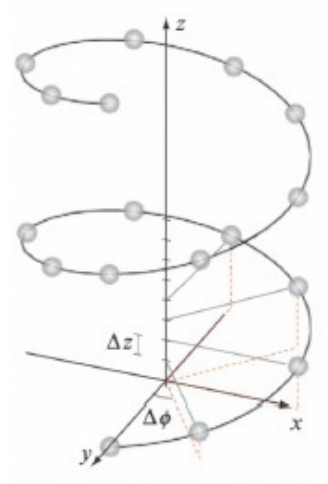- can cope with heterogeneous data
- manages to reconstruct weakly diffracting filaments (where layer lines are not visible)

Input
Set of 2D images
3D template structure
Initial symmetry guess

2D Templates
Systematically generated reference projections

Projection Matching
Orientation parameters: shifts and Euler angles

no: next iteration

Output
Symmetry parameters
Orientation parameters
Symmetrized structure

yes

Convergence
Symmetry and orientation parameters stable?

3D Reconstruction
(optional point symmetry)
Structure without helical symmetry

Symmetrization
Low-pass filtered and symmetrized helical structure

Symmetry Search
Helical symmetry parameters ($\Delta\varphi$, Δz)

(Behrmann et al., J Struct Biol 2012)
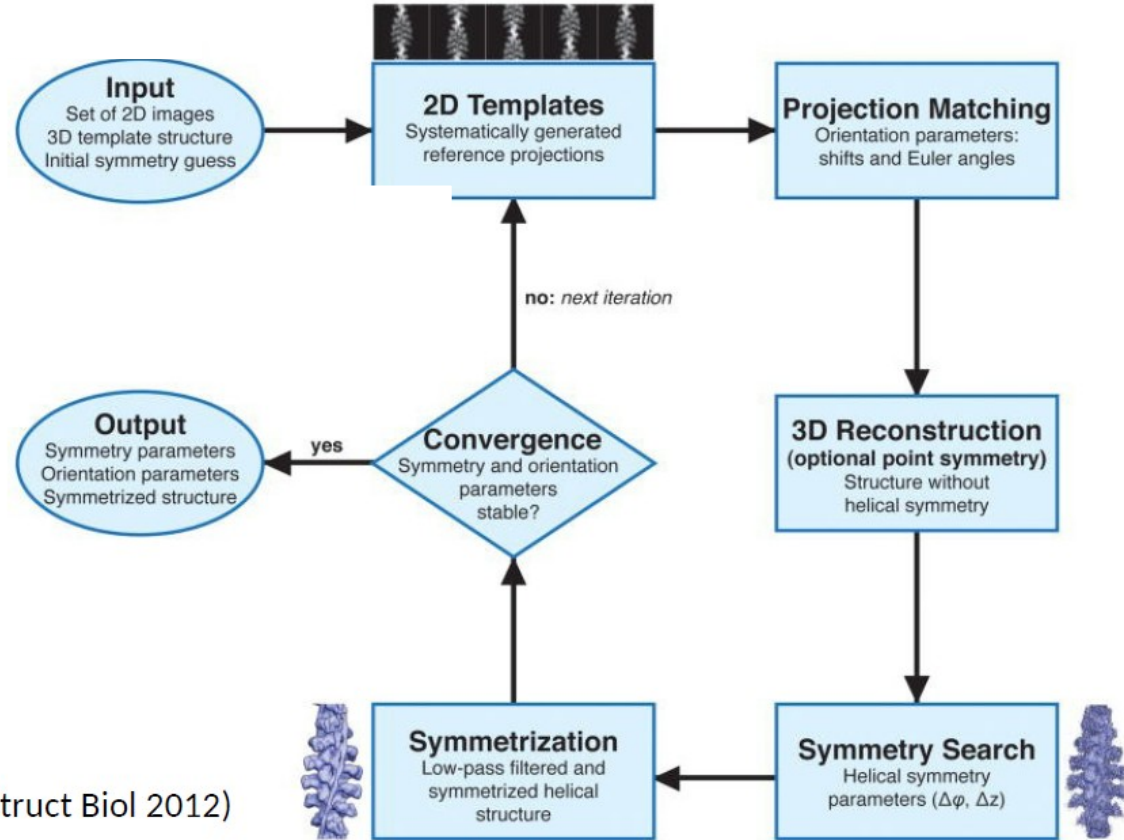
Symmetries

- Smaller asymmetric unit

- Decreased computational demands

- Improved signal to noise due to better averaging

- Lower number of particles required
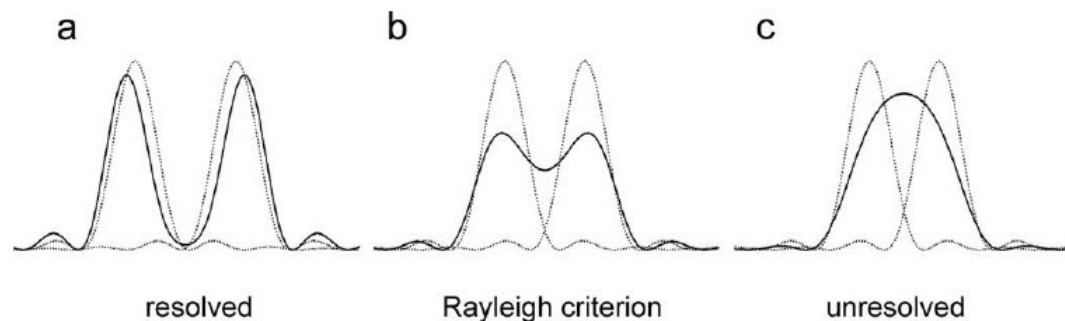
Resolution

Resolution definition by separation of features. (a) When two points are far apart, there is a deep trough of density between them. (b) Two points are regarded as just resolved when the peak of one point spread function overlaps the first minimum of the other (Rayleigh criterion), see ref 60. (c) The point spread functions of two dots close together overlap to form one maximum, so that the points are not resolved.

Map validation

- cryo-EM data – low signal to noise (VERY)

- model bias – persistence of an incorrect map or map features during refinement

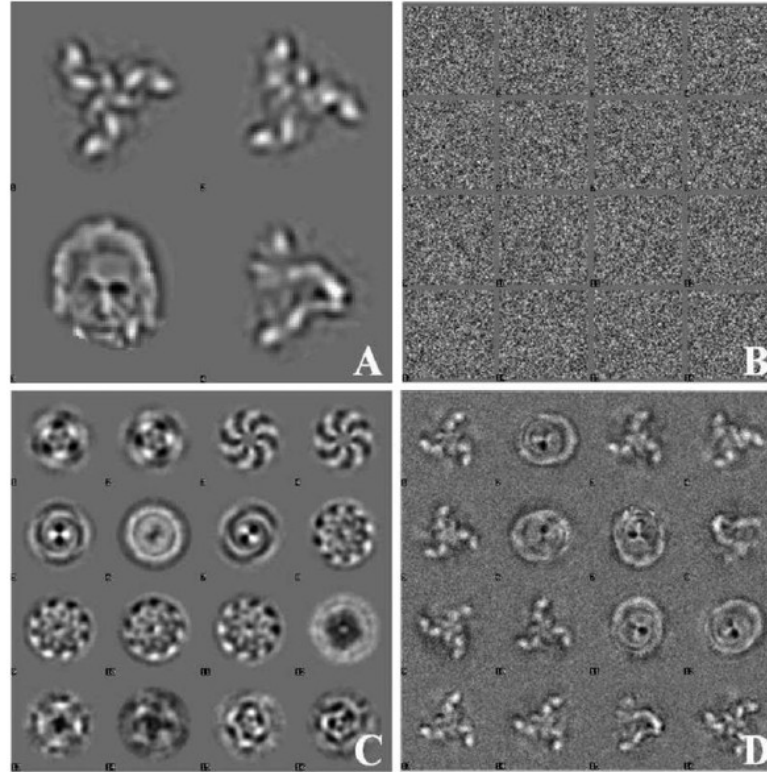(A) Four reference images (each 64 × 64 pixels) used for picking from 1,024 random noise images (of 1,024 × 1,024 pixels).
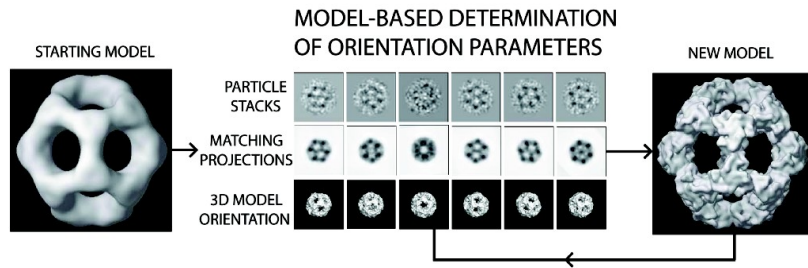
van Heel M PNAS 2013;110:E4175-E4177

STARTING MODEL

MODEL-BASED DETERMINATION OF ORIENTATION PARAMETERS

NEW MODEL

PARTICLE STACKS

MATCHING PROJECTIONS

3D MODEL ORIENTATION

PNAS

Map validation

- in order to minimize the model bias – separate the data into two halves, refine each half separately using the standard SPA protocol
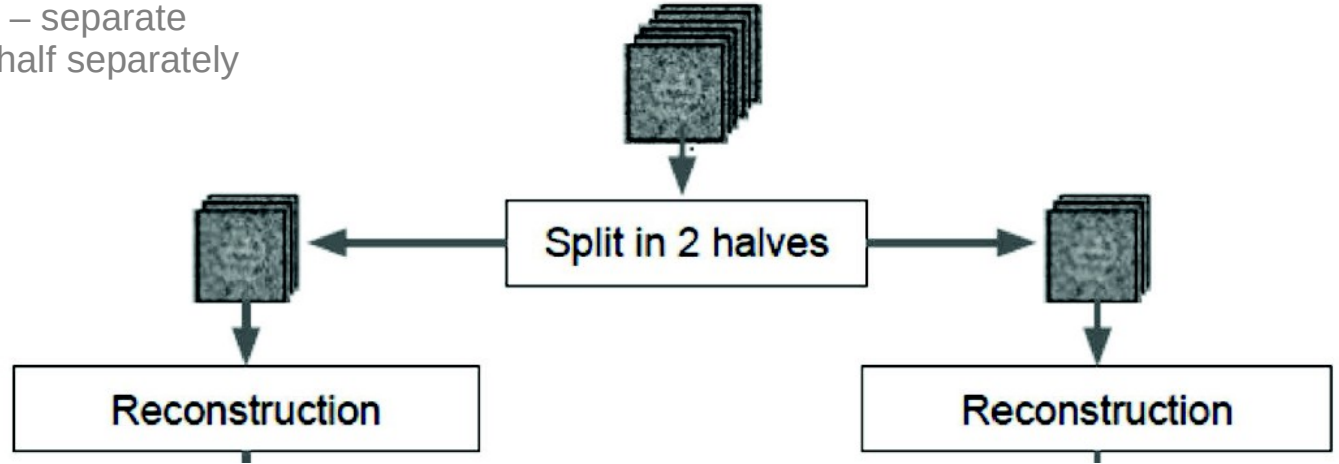
Split in 2 halves

Reconstruction

Reconstruction

Map validation

Fourier shell correlation

 - nowadays used as a metrics for map resolution estimation
 - calculate 3D Fourier transform of each map
 - calculate cross-correlation coefficients between the two 3D FTs for individual resolution shells
 - plot the CCC against the resolution for which it was calculated
 - determined CCC threshold for which resolution is reported

Map validation

Fourier shell correlation

 - nowadays used as a metrics for map resolution estimation
 - calculate 3D Fourier transform of each map
 - calculate cross-correlation coefficients between the two 3D FTs for individual resolution shells
 - plot the CCC against the resolution for which it was calculated
 - determine CCC threshold for which resolution is reported
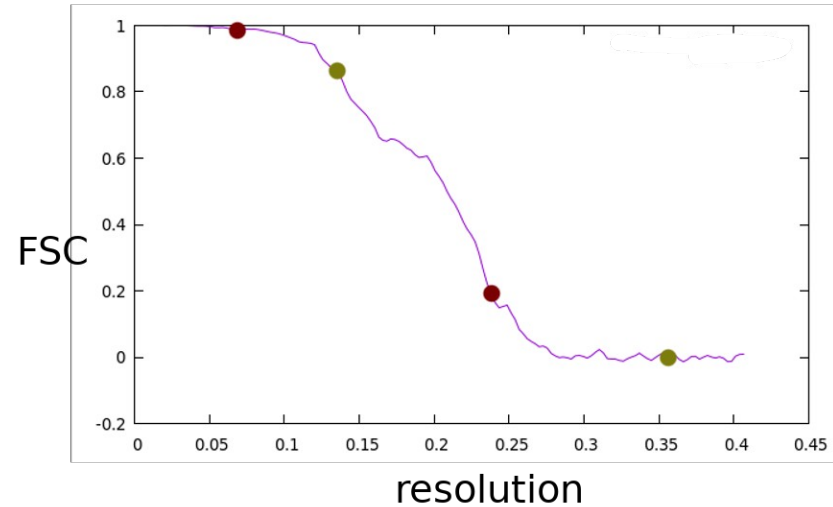
Map validation

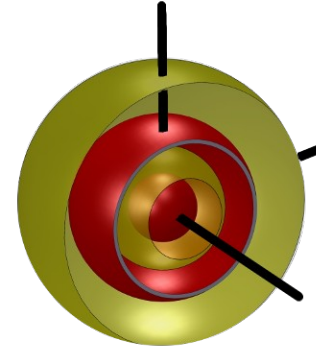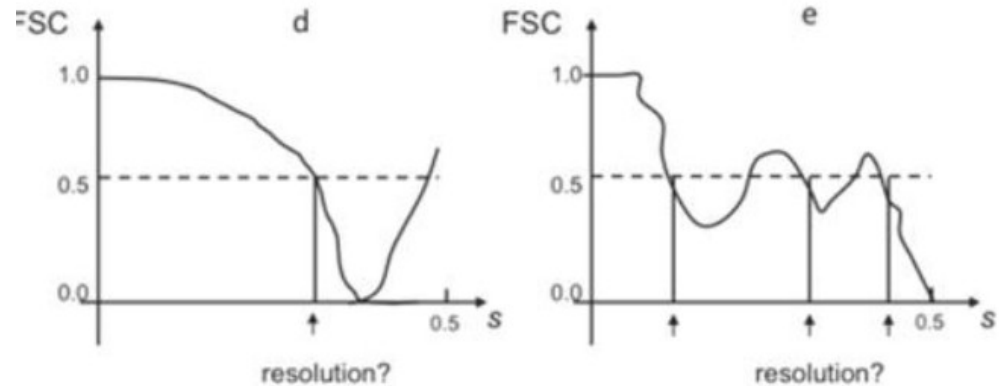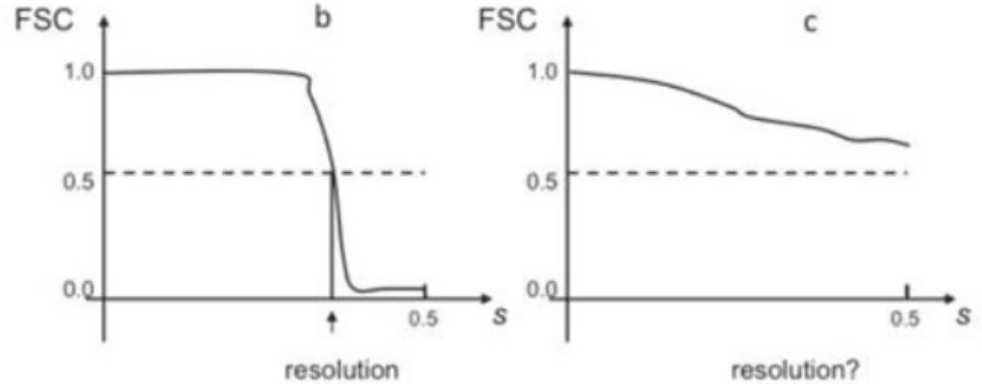## Fourier shell correlation

 - nowadays used as a metrics for map resolution estimation
 - calculate 3D Fourier transform of each map
 - calculate cross-correlation coefficients between the two 3D FTs for individual resolution shells
 - plot the CCC against the resolution for which it was calculated
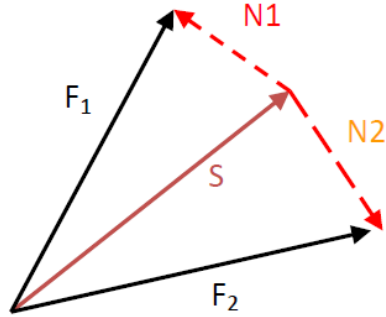 - determine CCC threshold for which resolution is reported

Penczek (2010), Meth. Enzym.

Map validation

Fourier shell correlation

 - which FSC value to report?

$$\frac{\sum (S + N1) \cdot (S + N2)^*}{\sqrt{\sum |S + N1|^2 \sum |S + N2|^2}}$$

$$FSC = \frac{\sum |S|^2}{\sum |S^2 + N^2|}$$ S~N FSC=0.5

$$FSC = \frac{\sum F_1 \cdot F_2{}^*}{\sqrt{\sum |F_1|^2 \sum |F_2|^2}}$$

Resolution Å

 - Figure of merit (C_{ref}=0.5)

$$C_{ref} = \frac{\sum F_1 \cdot F_{ref}}{\sqrt{\sum |F_1|^2 \sum |F_{ref}|^2}} = \frac{\sum |F_1||F_{ref}|\cos(\Delta\varphi)}{\sqrt{\sum |F_1|^2 \sum |F_{ref}|^2}}$$

Looks like figure-of-merit (Blow and Crick,1959)
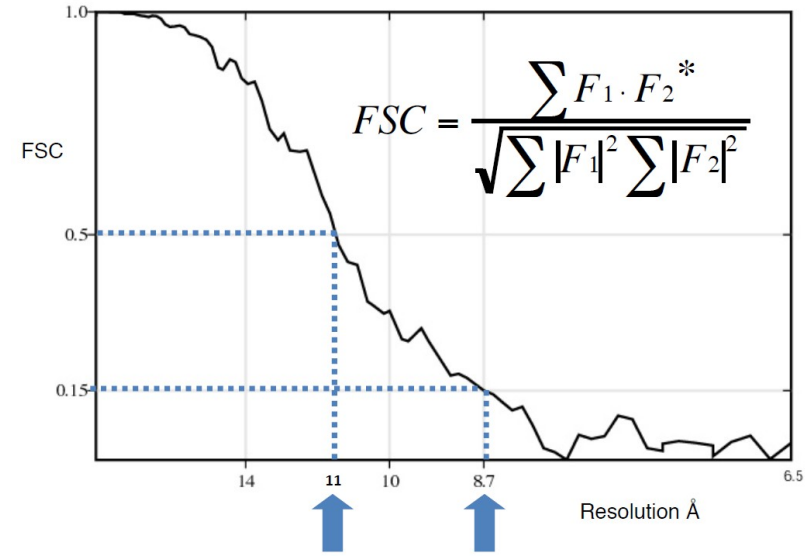
FSC	FSC_{FULL}	C_{REF}	PHASE ERROR	$S/N_{1/2}$
0.50	0.67	0.82	35°	1.00
0.33	0.50	0.71	45°	0.71
0.14	0.25	0.50	60°	0.41

Map validation

- observed features of the map should be consistent with the resolution assessment

- visibility of expected structural features
  - helices visible at 8Å
  - strands separated at 4.8Å
  - side-chains visible beyond 4Å
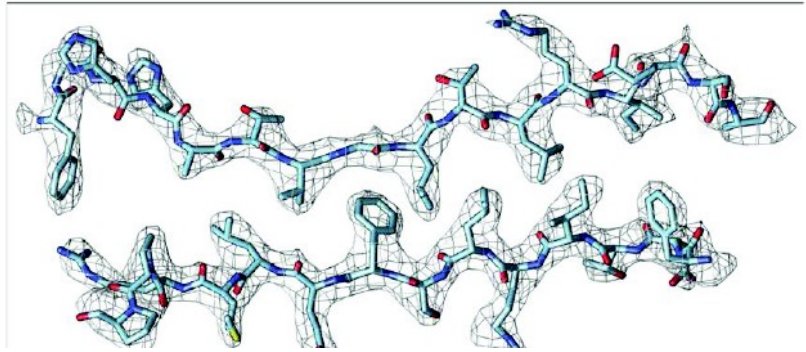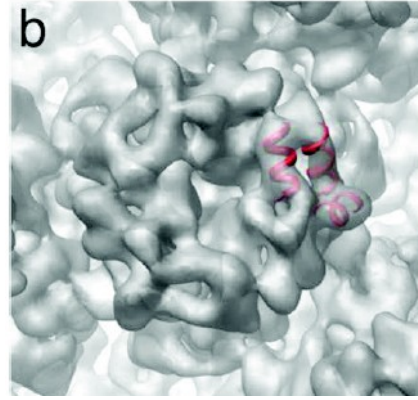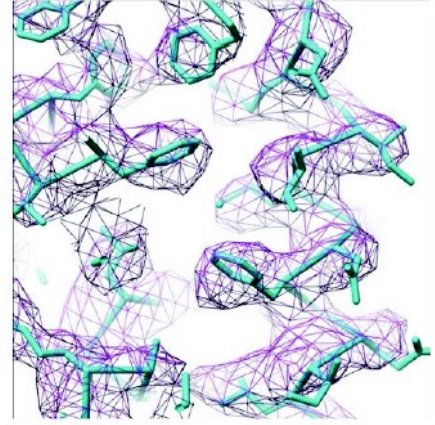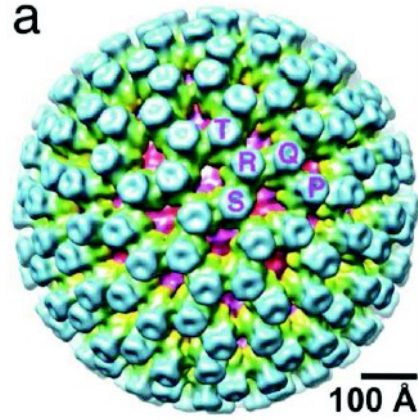
Map validation

- experimental
  - cryo-ET
  - tilt pairs (common-lines)
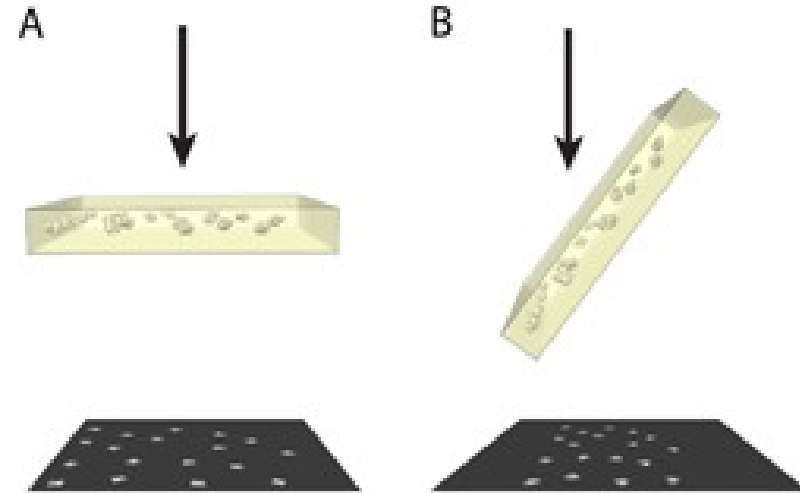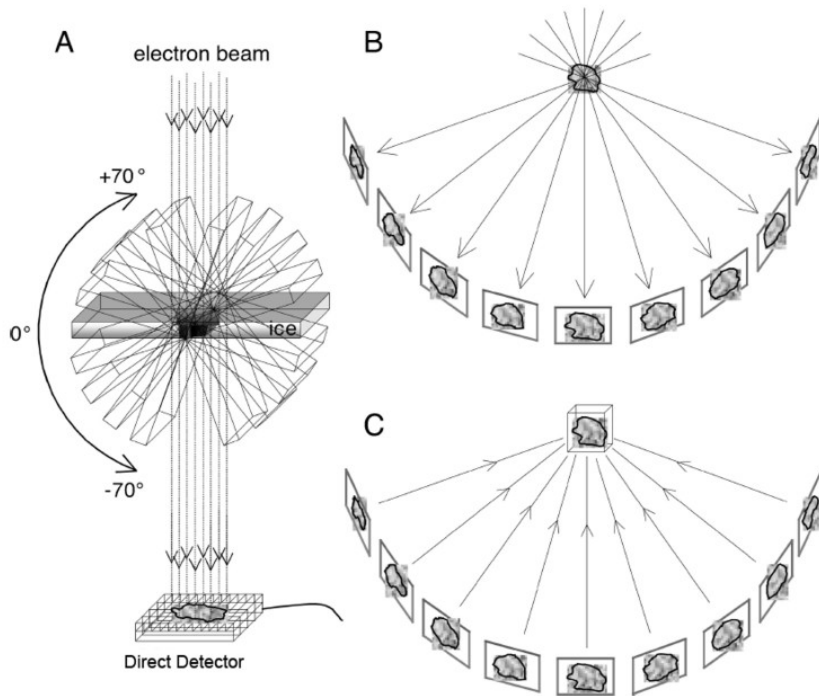
Map validation

- Steps:
    - map is correct at low resolution

    - spurious noise features are not present (noise overfitting, over-refinement)

    - FSC curve has a proper shape

    - resolution estimate corresponds to the observed structural features

    - acquisition of complementary data to confirm the model (e.g. in low resolution)

Map interpretation
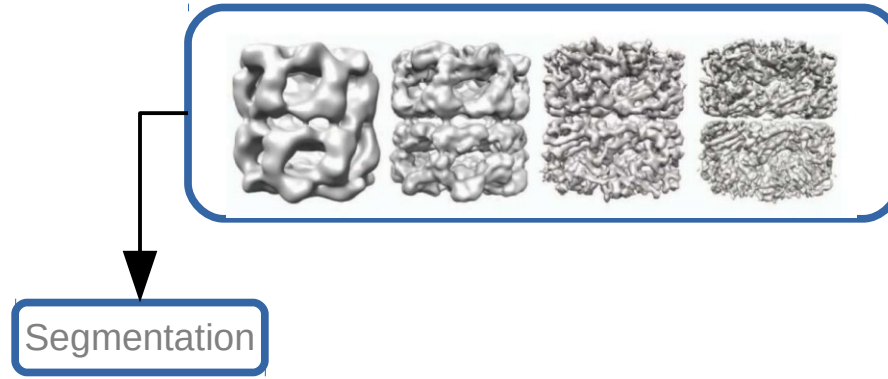
Segmentation

Map interpretation
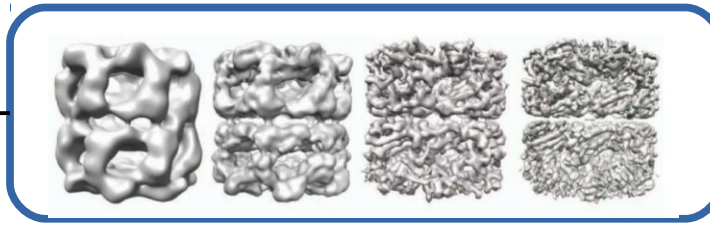
Segmentation

Visualization tools
- Chimera/ChimeraX
- Coot
- PyMol
- VMD
- Amira (Commercial)
- ...
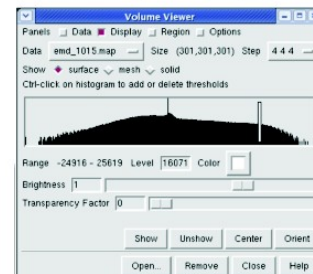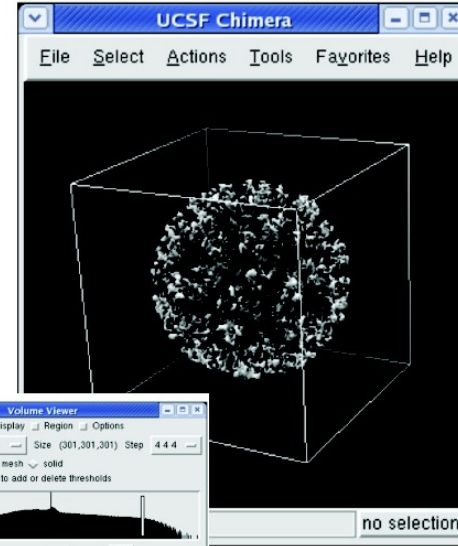
Segmentation

- identify boundaries map regions which represent different structural components
- component structures can be positions into the identified segments
- the size of the segmented components is related to the map resolution
- manual segmentation | automated segmentation | knowledge-based segmentation

Map interpretation

Segmentation

?Resolution?

25 Å   10 Å   4.5 Å

Map interpretation

Segmentation

known component
structure

?Resolution?

10-20A   4-10A   <4A

Fit known structures   Sec. str. assign.   Model building

Fold assignment
from sequence

Sec. Str. Sequence
assignment

No   template found   Yes

'Template-free'
modelling

Homology
modelling

Rigid body fit

fit different from map

Multiple conformations
EMN/NMA
Real-space methods
MD-based methods
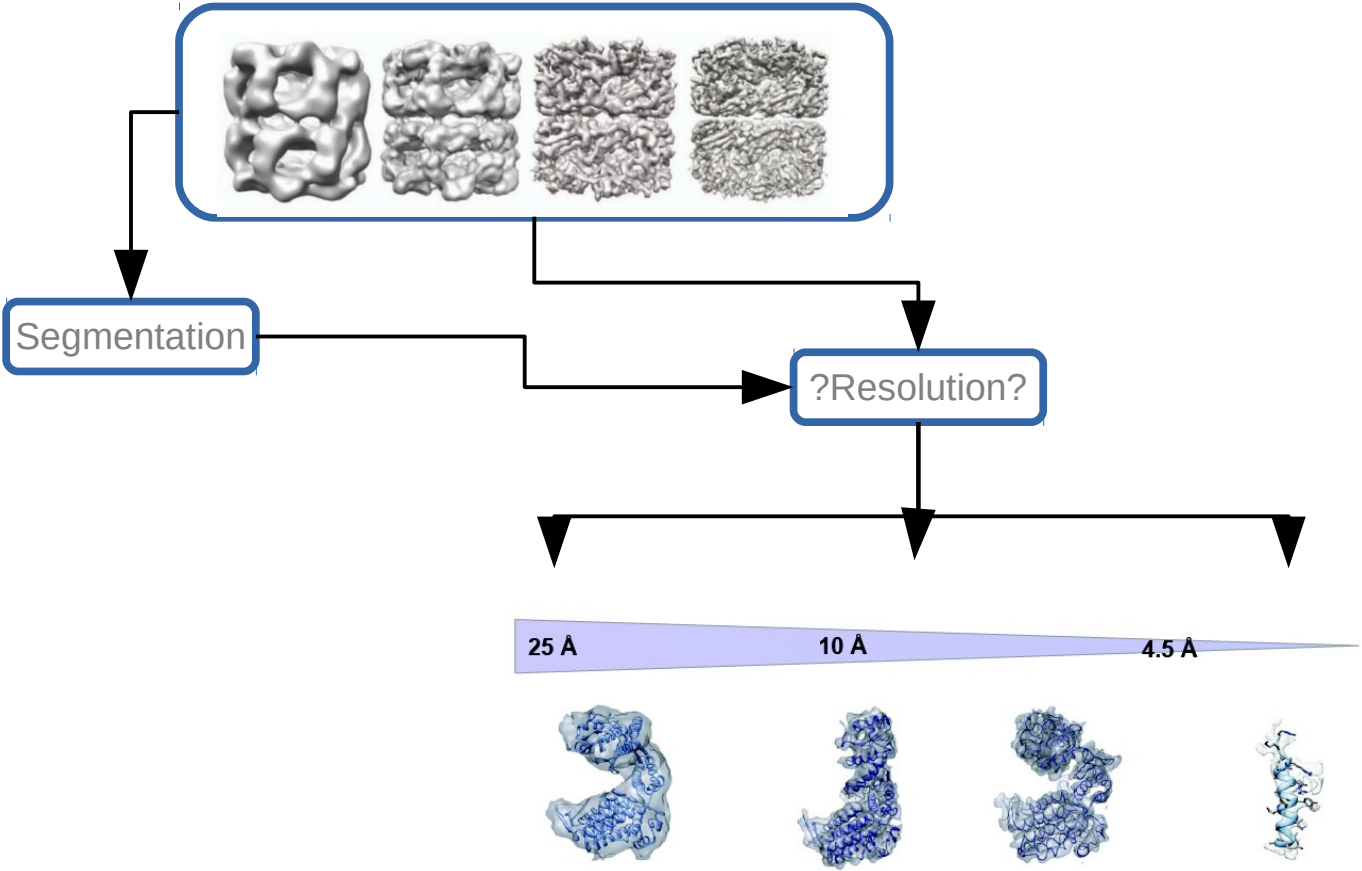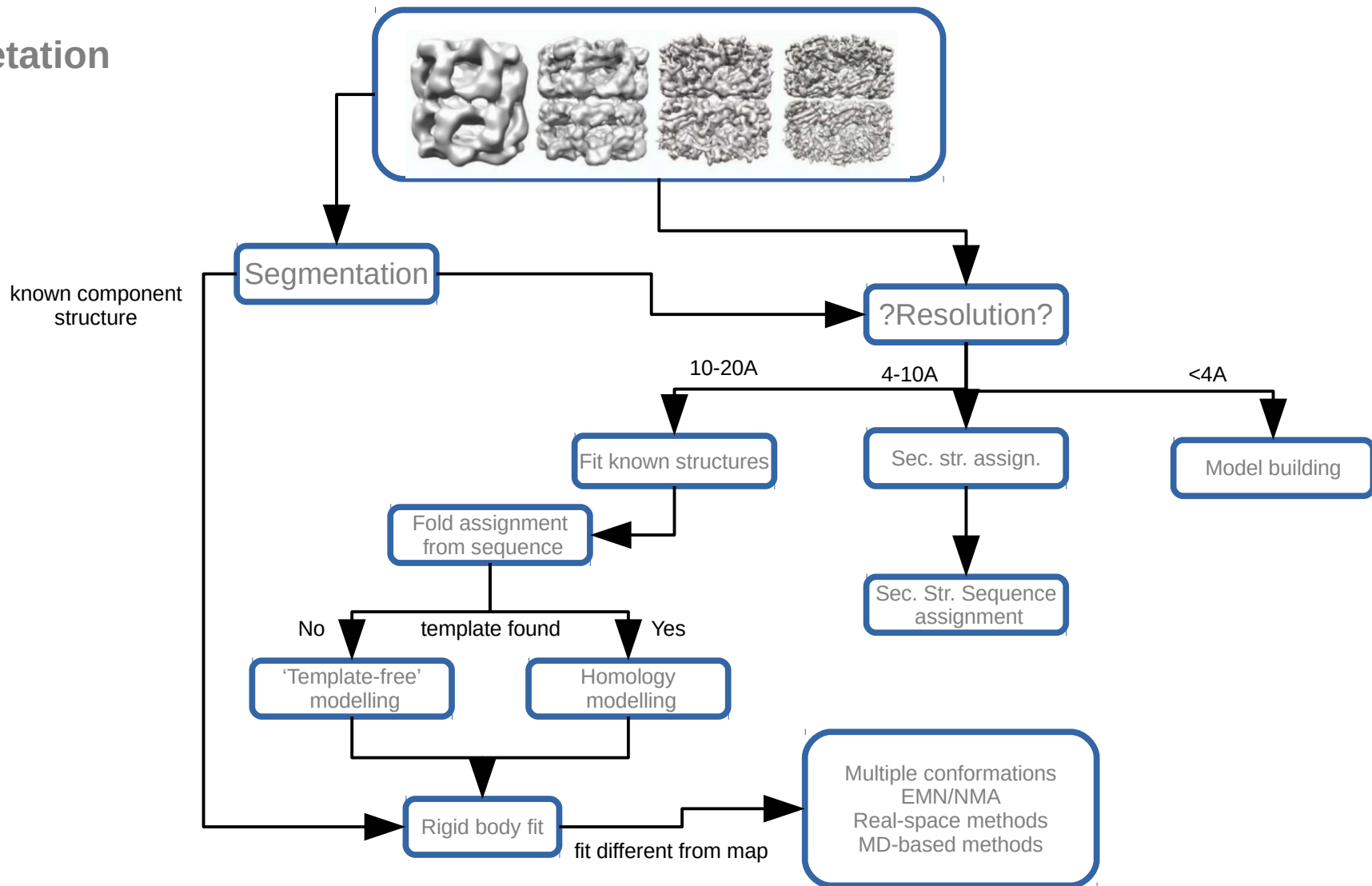
Map interpretation

Fold recognition from density

- 4.5-10A: secondary structure detection
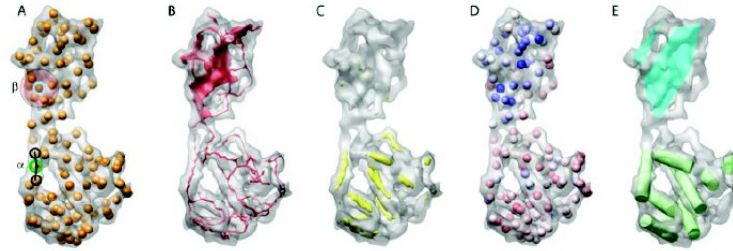
Baker et al. *Structure* 2007

- 4.5A and better: de novo CA tracing and model building
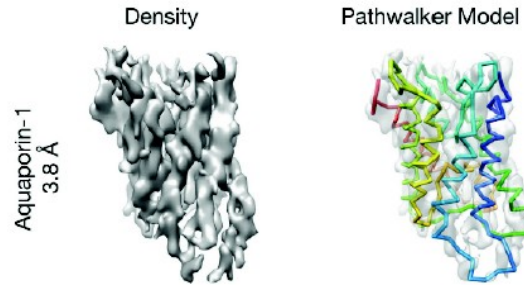
Baker et al. *Structure* 2012

- programs: SSEhunter, SSEtracer, Ematch, Pathwalker, Coot, Buccaneer, EM-fold, Rosseta, Phenix, ARP/wARP, MAINMAST

Map interpretation

Segmentation

?Resolution?

10-20A 4-10A <4A

known component
structure

Fit known structures

Sec. str. assign.

Model building

Fold assignment
from sequence

Sec. Str. Sequence
assignment

No template found Yes

'Template-free'
modelling

Homology
modelling

Rigid body fit

fit different from map

Multiple conformations
EMN/NMA
Real-space methods
MD-based methods
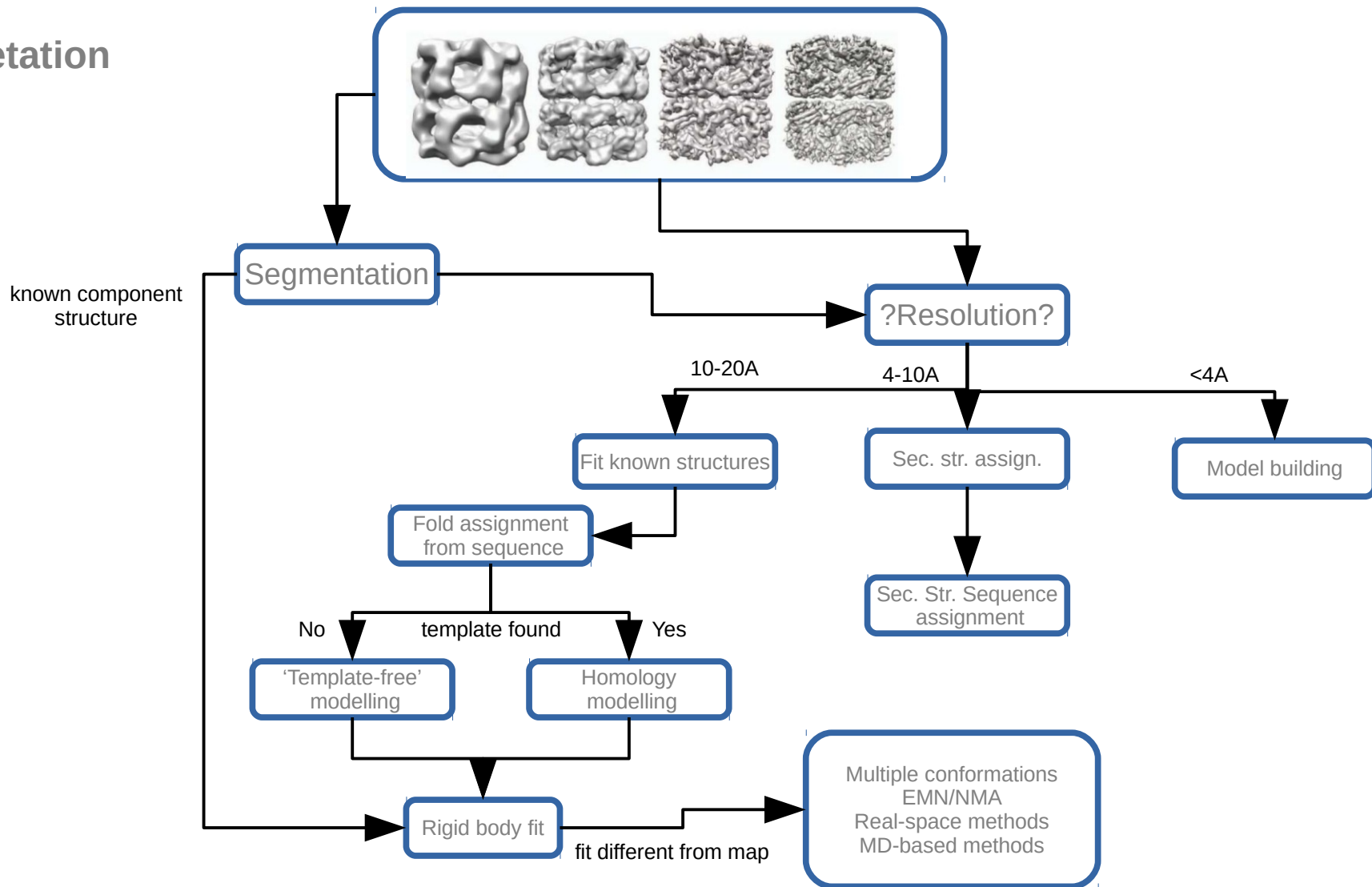
Map interpretation

Fold recognition from sequence

GFCHIKAYTRLIMVG…
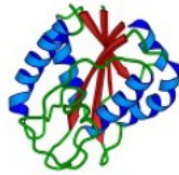
Template-free

Ab initio (*de novo*) prediction
Fragment Assembly
Evolutionary Couplings
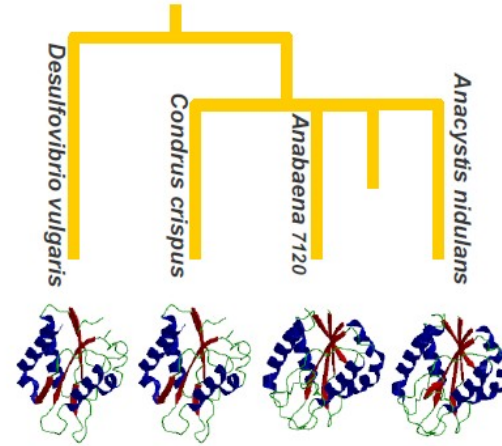
Template-based

Threading
Comparative (Homology) Modelling

- programs: MODELLER, SWISS-MODEL, Phyre2, RaptorX, I-TASSER, Rosetta, EVfold

Map interpretation

Density fitting

Rigid-Body Fitting / Assembly Fitting

Flexible Fitting

De novo

25 Å   10 Å   4.5 Å

Conformational Changes / Shape

Domain boundaries

α-helix

β-sheets

Side chains

Villa & Lasker, *Curr Opin Struct Biol*, 2014,
Cassidy et al, *Curr Opin Microbiology* 2018
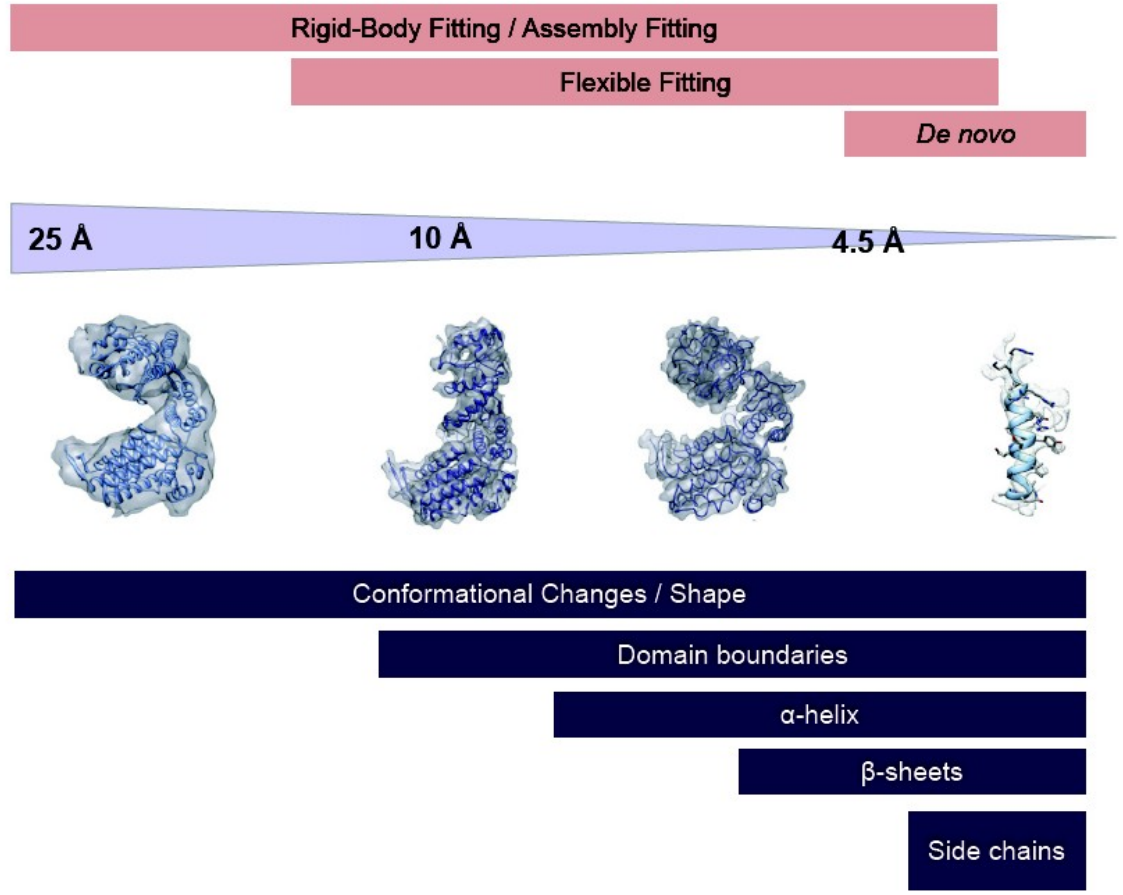
Map interpretation

Density fitting

- manual fitting

     - positioning of the atomic structure into the cryo-EM density using visualization programs

     - usually efficient (human brain efficient in pattern recognition)

     - direct feedback

     - good for initial placement of the component in to the map

     - high level of subjectivity may lead to errors

     - depends on contour level at which the map is visualized

     - conformational rearangements cannot be modelled

Map interpretation

Density fitting

- automated fitting
  - requires common representation of both the structure and the density map

  - measure of the quality of the fit

  - optimization protocol for fit improvement

Density map

Component atomic structure

Component representation and placement

Optimisation based on goodness-of-fit
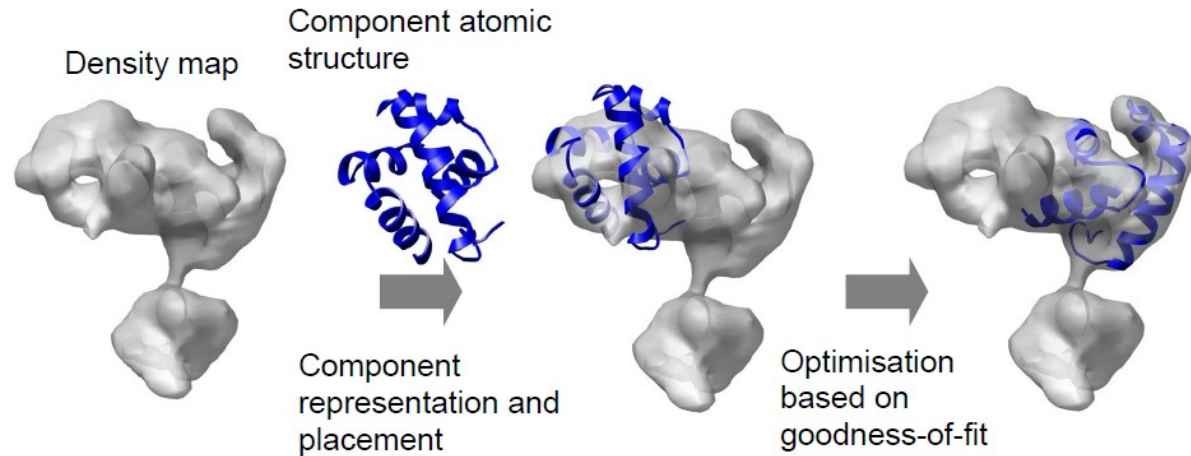
$$CCC_{\rho,lin} = \int_x (\rho_{obs}(x)\rho_{calc}(x,m))^2 d^3x$$

Map interpretation

Problems of density fitting

- limited resolution
    - many local optima with similar numerical values at low resolution

    - local resolution, noise, scaling, filtering, masking

    - blurring of the atomic structure

→ better resolution
→ improve scoring for goodness-of-fit
→ coarse-graining (change represenation)
→ fit/model validation

Correct fit   Flipped 180°

2 Å

10 Å

20 Å
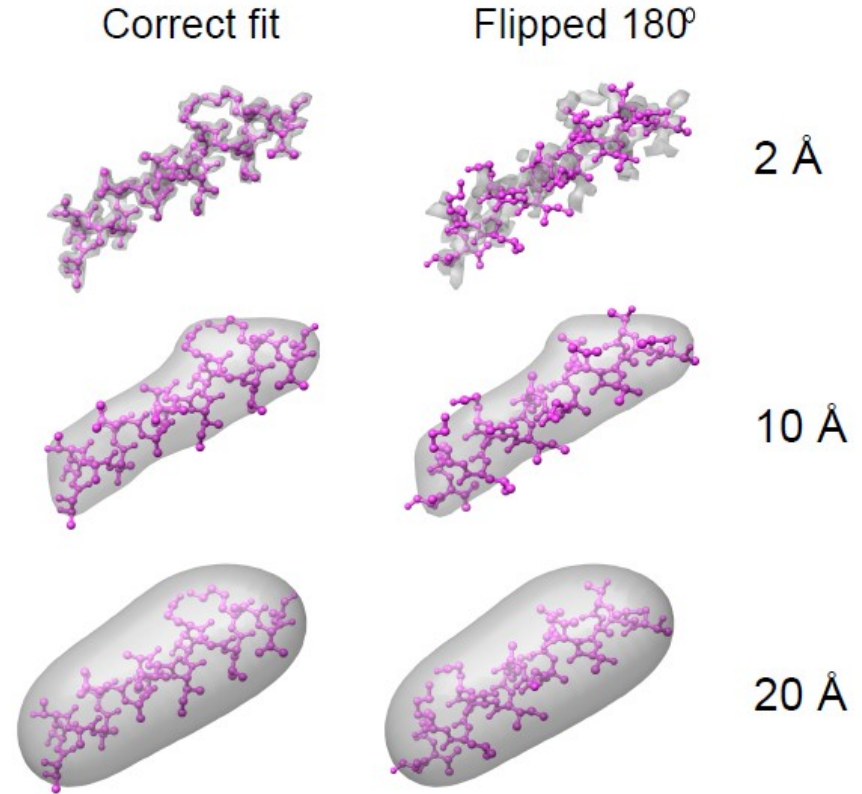
Map interpretation

Problems of density fitting

- conformational variability
  - many lconformations which are observed in density maps deviate from the conformations of the atomic models which are fitted

  - dynamics
  - crystal packing effects
  - errors in structure prediction

→ allow for the conformational changes
during model fitting process = flexible fitting
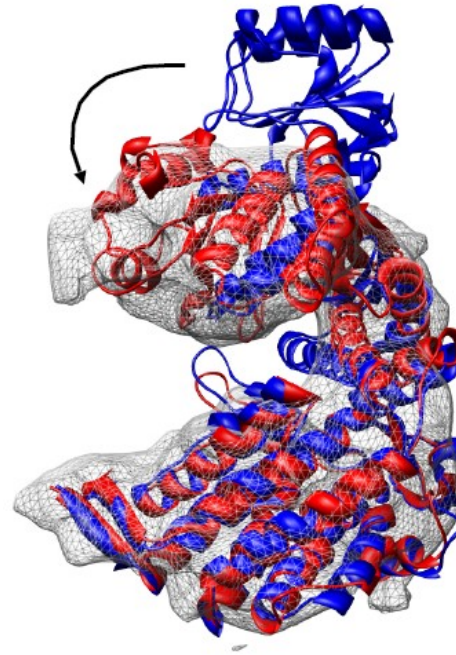
Map interpretation

Model refinement

- without any restraints a model may fit well with a high score in near-atomic to low resolution density

- such a model will, however, not have standard protein geometry: backbone torsions (Ramachandran diagram), peptide planarity, chirality (trans/cis), bond lengths and angles, side chain torsions / rotamers

- refinement methods try to maintain standard geometry while fitting the model into the density map. The geometry restraints reduce the levels of freedom.

- map density contributes as an additional penalty in the scoring function

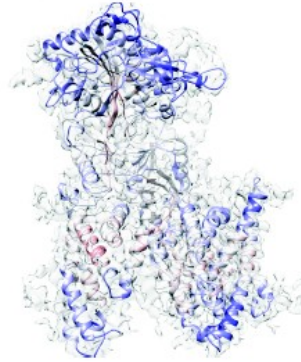Programs: MDFF, Refmac, Rosetta, Coot, Phenix, Isolde, iMODFIT

Map interpretation

Model validation

Model fit

Model geometry

peptide planarity
backbone torsions (Ramachandran)
bond lengths
bond angles
side chain rotamers

Molprobity: http://molprobity.biochem.duke.edu/
What check: http://swift.cmbi.ru.nl/gv/whatcheck/
PROCHECK: http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/

Map improvement

- in order to facilitate map interpretation, the data processing should correct for the imperfections of the imaging system to the highest possible level

- these imperfections comprise:
   - aberrations of microscope optical system (higher-order)
   - sample drift and distortions caused by interaction of the electrons with a matter

- the effect is primarily pronounced at high frequencies (resolution) → parameter optimization and additional data processing primarily concerns improving the quality of high resolution maps (<4.5A resolution)

- the effect on medium and low resolution (>8A) is limited and additional data processing usually does not result in any map improvement

Map improvement

Electron lens aberrations

- objective lens of the transmission electron microscope is really bad

$$B(\mathbf{k}) = \exp\left[i\frac{2\pi}{\lambda}W(\mathbf{k})\right]$$

2.2: Description of aberration constants to 6th order

A_0	Lateral image shift
A_1	Two-fold astigmatism
C_1	Defocus
A_2	Three-fold astigmatism
B_2	Axial coma
A_3	Four-fold astigmatism
S_3	Axial star aberration
$C_3 = C_s$	Spherical aberration
A_4	Five-fold astigmatism
D_4	Three-lobe aberration
B_4	Fourth-order axial coma
A_5	Six-fold astigmatism
S_5	Fifth-order star aberration
C_5	Fifth-order spherical aberrat
R_5	Fifth-order rosette aberratior

$$W(\mathbf{k}) = \Re\{A_0\lambda\mathbf{k}^*$$

$$+ \frac{1}{2}A_1\lambda^2\mathbf{k}^{*2} + \frac{1}{2}C_1\lambda^2\mathbf{k}^*\mathbf{k}$$

$$+ \frac{1}{3}A_2\lambda^3\mathbf{k}^{*3} + \frac{1}{3}B_2\lambda^3\mathbf{k}^{*2}\mathbf{k}$$

$$+ \frac{1}{4}A_3\lambda^4\mathbf{k}^{*4} + \frac{1}{4}S_3\lambda^4\mathbf{k}^{*3}\mathbf{k} + \frac{1}{4}C_3\lambda^4\mathbf{k}^{*2}\mathbf{k}^2$$

$$+ \frac{1}{5}A_4\lambda^5\mathbf{k}^{*5} + \frac{1}{5}D_4\lambda^5\mathbf{k}^{*4}\mathbf{k} + \frac{1}{5}B_4\lambda^5\mathbf{k}^{*3}\mathbf{k}^2$$

$$+ \frac{1}{6}A_5\lambda^6\mathbf{k}^{*6} + \frac{1}{6}S_5\lambda^6\mathbf{k}^{*4}\mathbf{k}^2 + \frac{1}{6}C_5\lambda^6\mathbf{k}^{*3}\mathbf{k}^3 + \frac{1}{6}R_5\lambda^6\mathbf{k}^{*5}\mathbf{k}$$
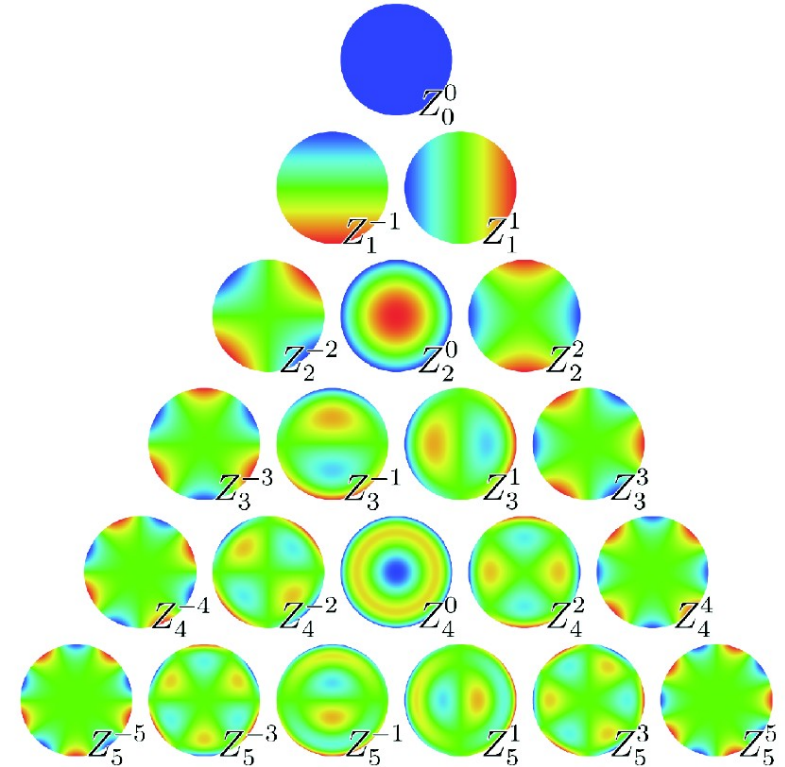
Map improvement

Zernike polynomials

- complete set of orthogonal functions

- Zernike transform analogous to Fourier transform

- can be used to visualize lens aberrations

- the aberrations can be corrected for by introducing additional lens to the microscope or by software during the image processing
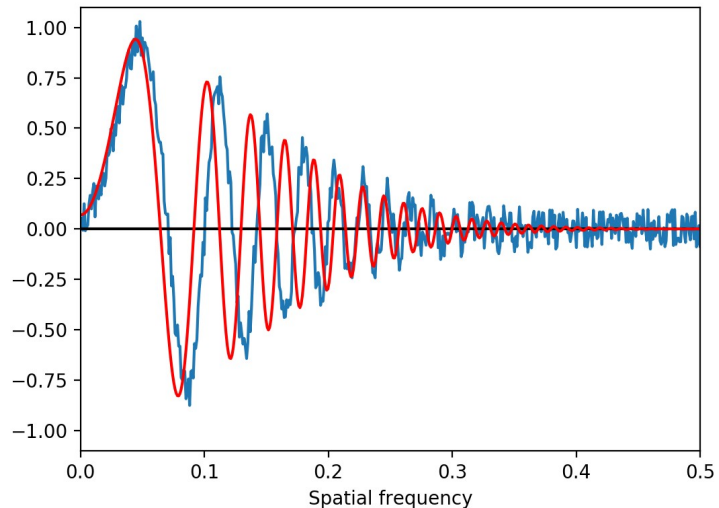
Frits Zernike,
1953 Nobel Prize in Physics
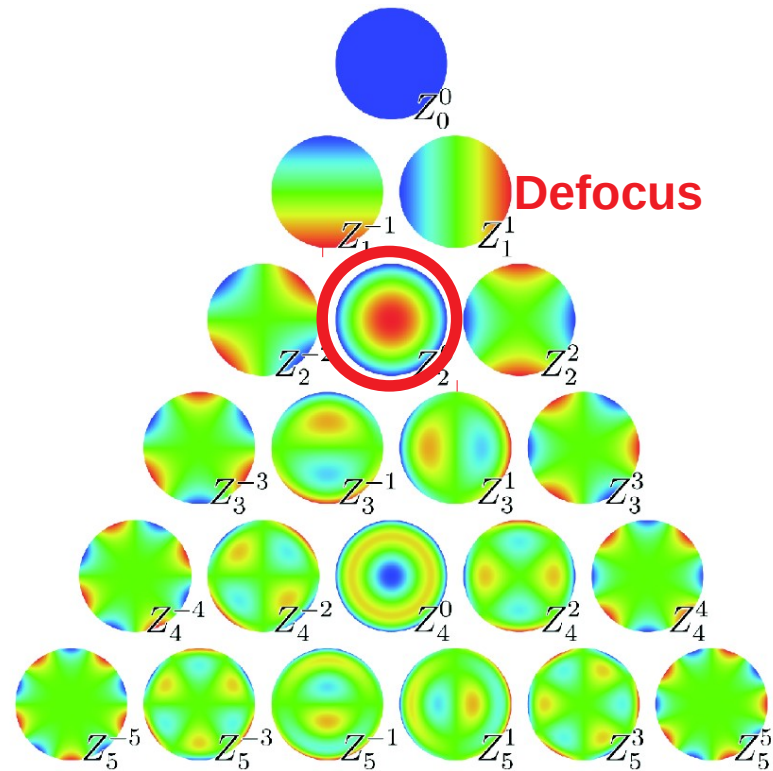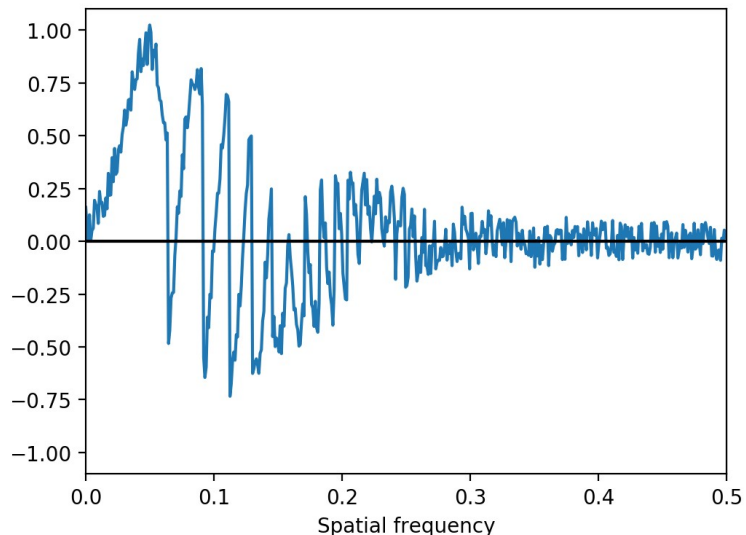inventor of phase contrast
microscopy

Map improvement

Lens aberrations

- 200nm error in defocus estimation (1.2um instead of 1.0um)

Defocus

Map improvement
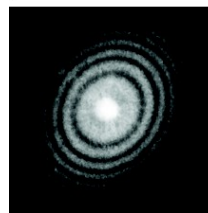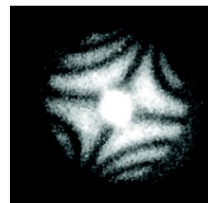
Lens aberrations

Original | Compromise

aio | aio

Horizontal Focus | Vertical Focus

aio | aio

Astigmatism

Z_0^0

Z_1^{-1} Z_1^1

Z_2^{-2} Z_2^0 Z_2^2

Z_3^{-3} Z_3^{-1} Z_3^1 Z_3^3

Z_4^{-4} Z_4^{-2} Z_4^0 Z_4^2 Z_4^4

Z_5^{-5} Z_5^{-3} Z_5^{-1} Z_5^1 Z_5^3 Z_5^5
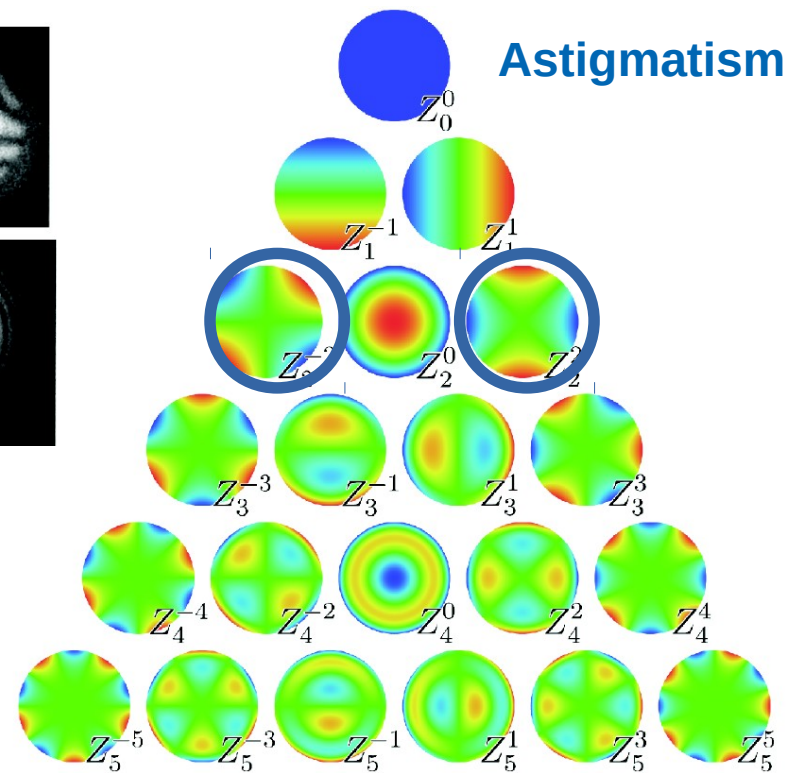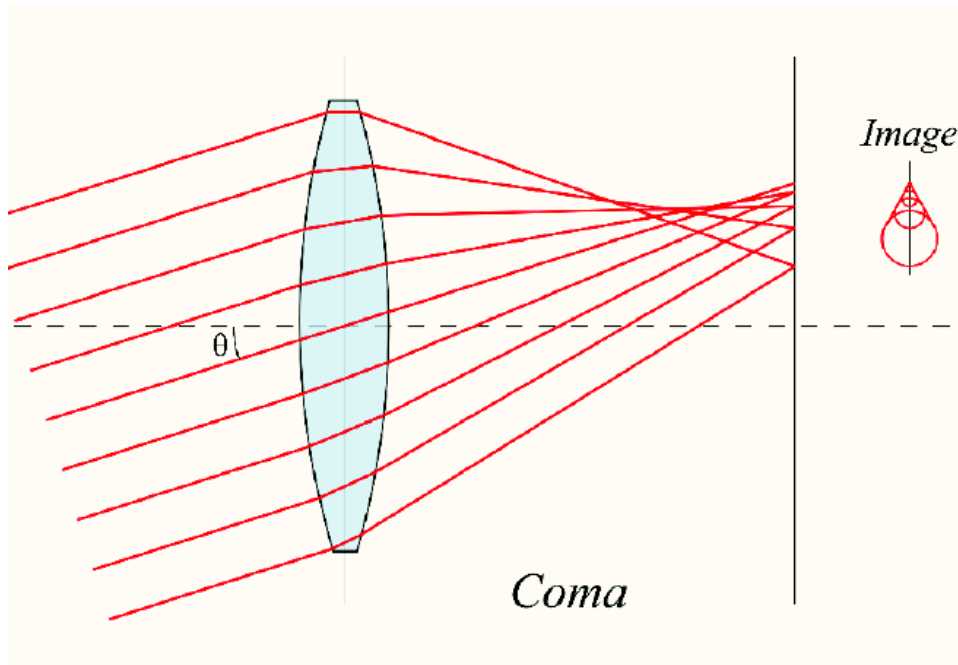
Map improvement

Lens aberrations

- certain level of underfocus is necessary during cryo-EM data collection
    → corrected during CTF correction
- astigmatism can be eliminated to high extent by proper microscope alignment

- only aberrations which are relevant for the quality of medium and low resolution maps

- correct estimation of CTF parameters (defocus,astigmatism)
    → quality control – goodness of fit
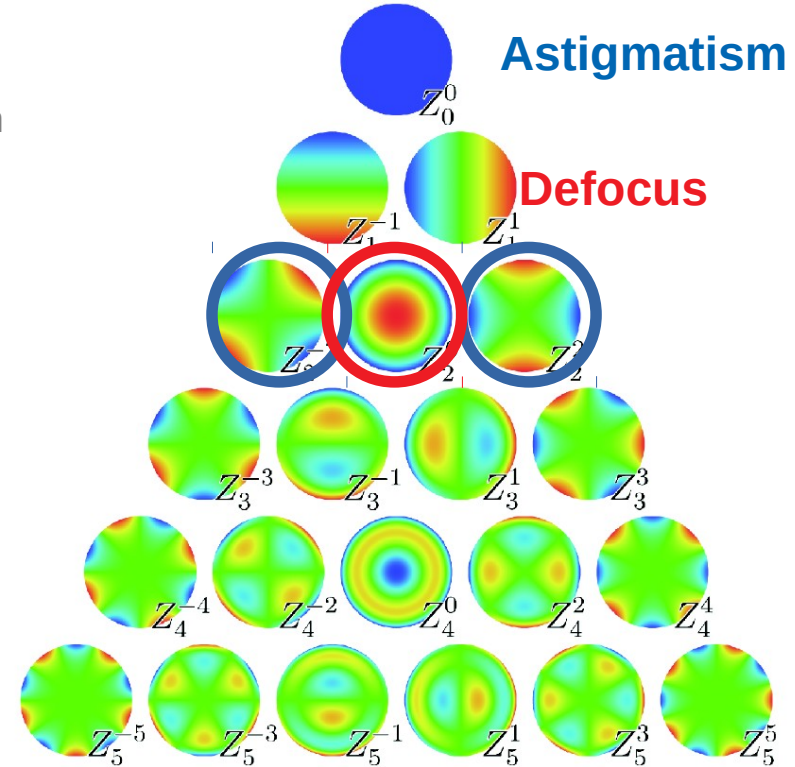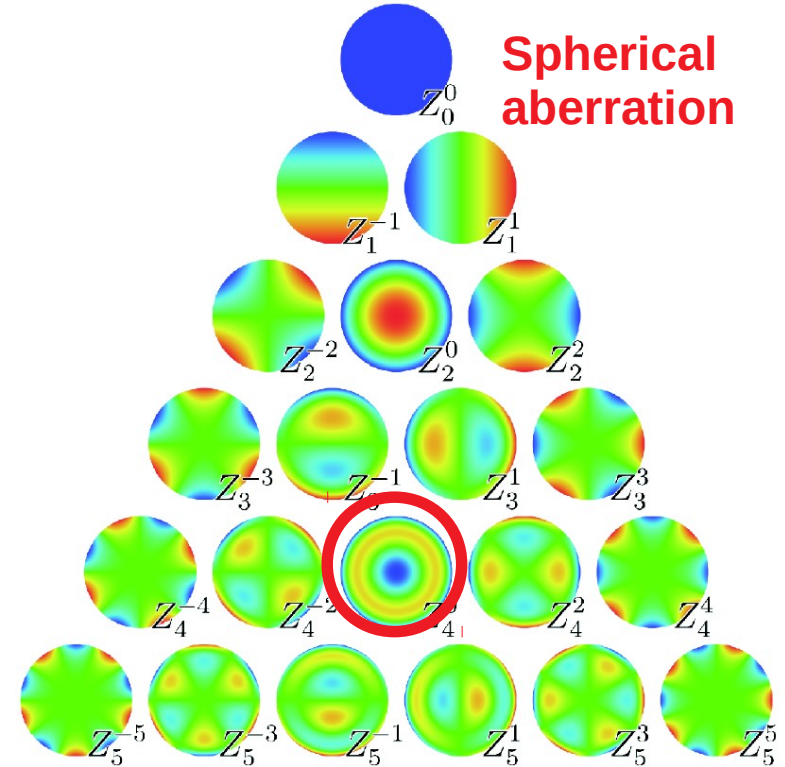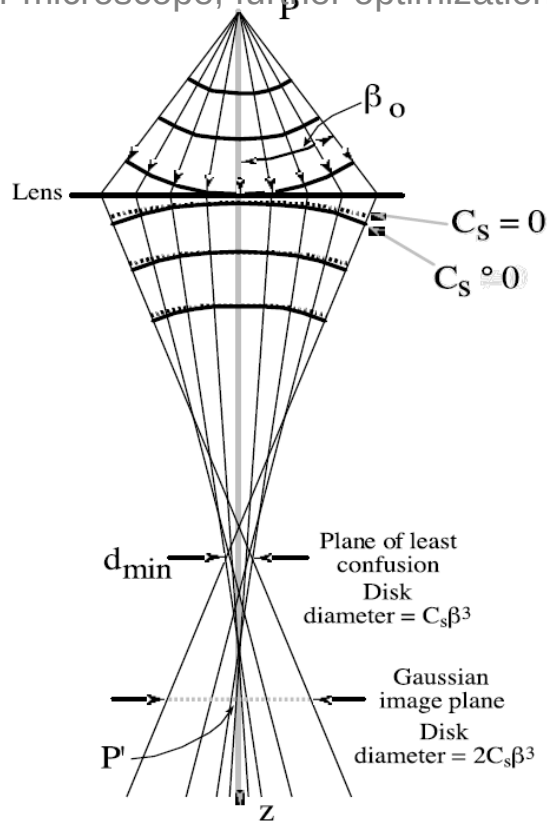
Map improvement

Lens aberrations

- dependence with third power of frequency
- can be primarily removed by proper microscope alignment and further during data analysis in software
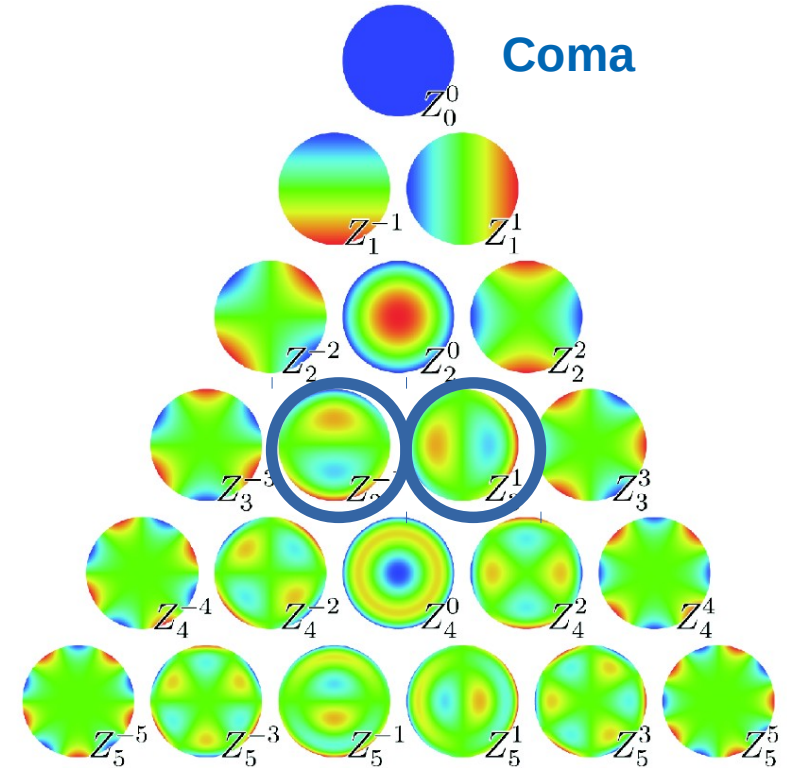
Map improvement
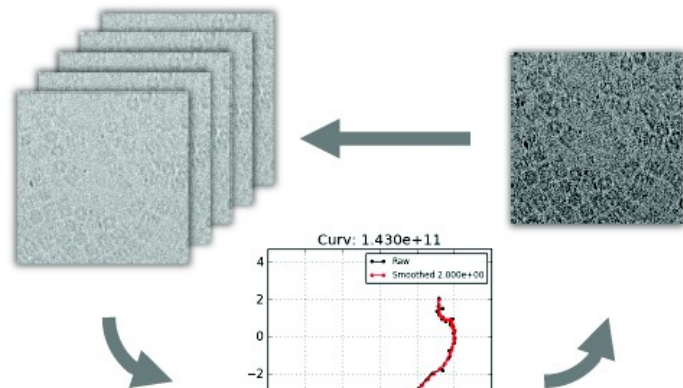
Lens aberrations

- dependence on fourth power of the frequency
- lens is stronger off axis, plane of least confusion
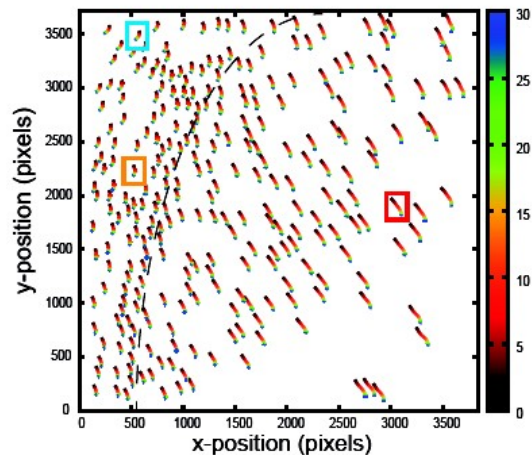- considered constant for microscope, further optimization in software possible

Spherical aberration

Map improvement
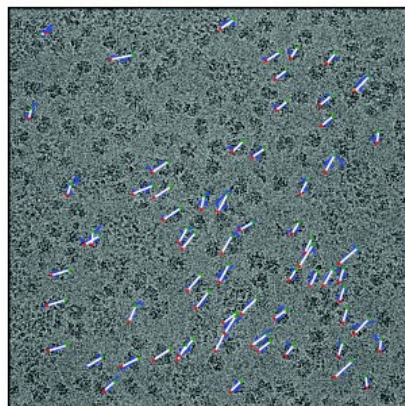
Sample distorsions during imaging

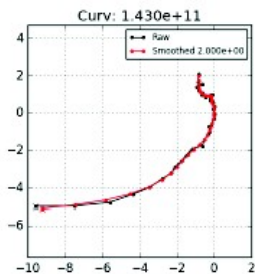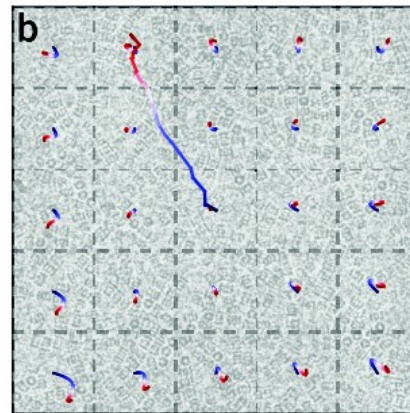- local motion different in distinct parts of the image

Compare particle in each frame to sum of frames

Compare particle in each frame to map

Compare patch from each frame to sum of frames

Alignparts_lmbfgs (Rubinstein & Brubaker, 2015, *JSB* 192, 188-95)
[improved version in cryoSPARC ver 2]

Relion Polishing (Scheres, 2014, *eLife 3:e03665*)
[improved version with Alignparts-like smoothing in Bayesian polishing]

MotionCor2 (Zheng…Agard, 2017, *Nat Meth* 14, 331-2)

Map improvement

Sample distorsions during imaging

- the information in each frame is damped by different B-factor due to distinct effects during data collection

- compensation for local motion (per particle) + per frame amplitude weighting with corresponding B-factor => particle polishing
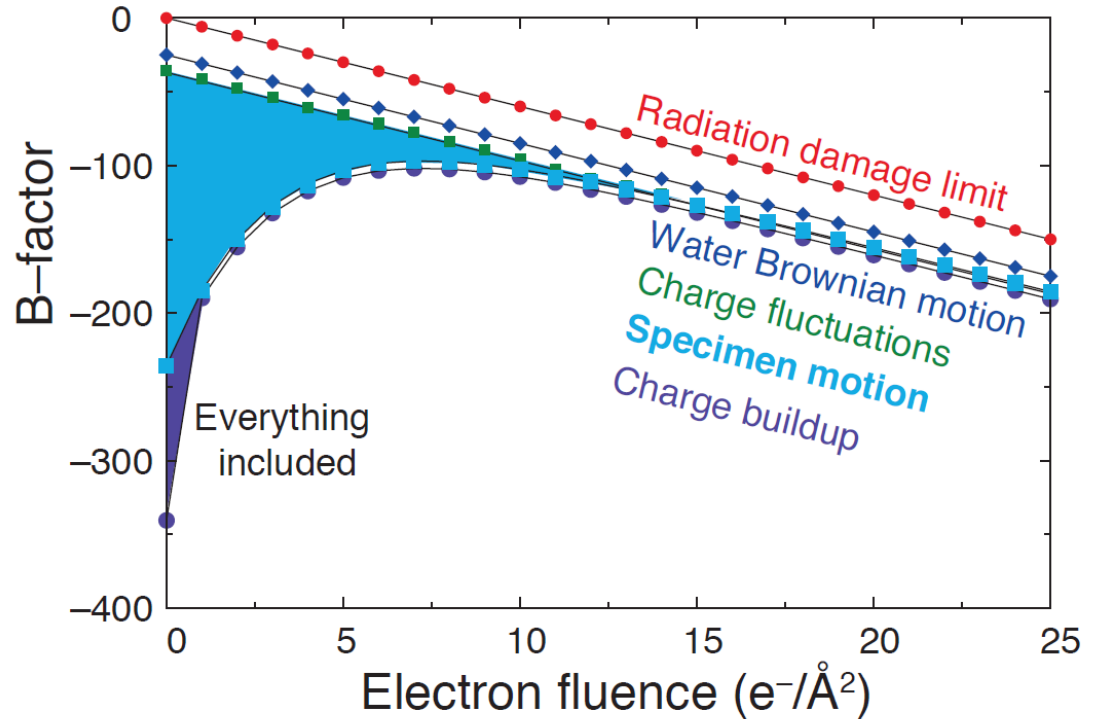
Map improvement

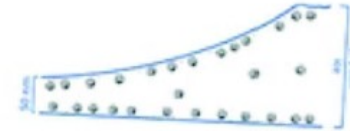## Sample distorsions during imaging

- distortion of sample surface due to illumination with electron beam

- particles located in different depth of the specimen layer

    → defocus variance for particles within single micrograph

- per particle defocus (astigmatism) estimation = ctf refinement
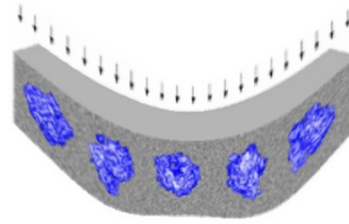
38*† Apoferritin (0.5 mg/mL)

39*† Apoferritin with 0.5 mM TCEP

40 Protein with Carbon Over Holes

41 Protein and DNA Strands with Carbon Over Holes
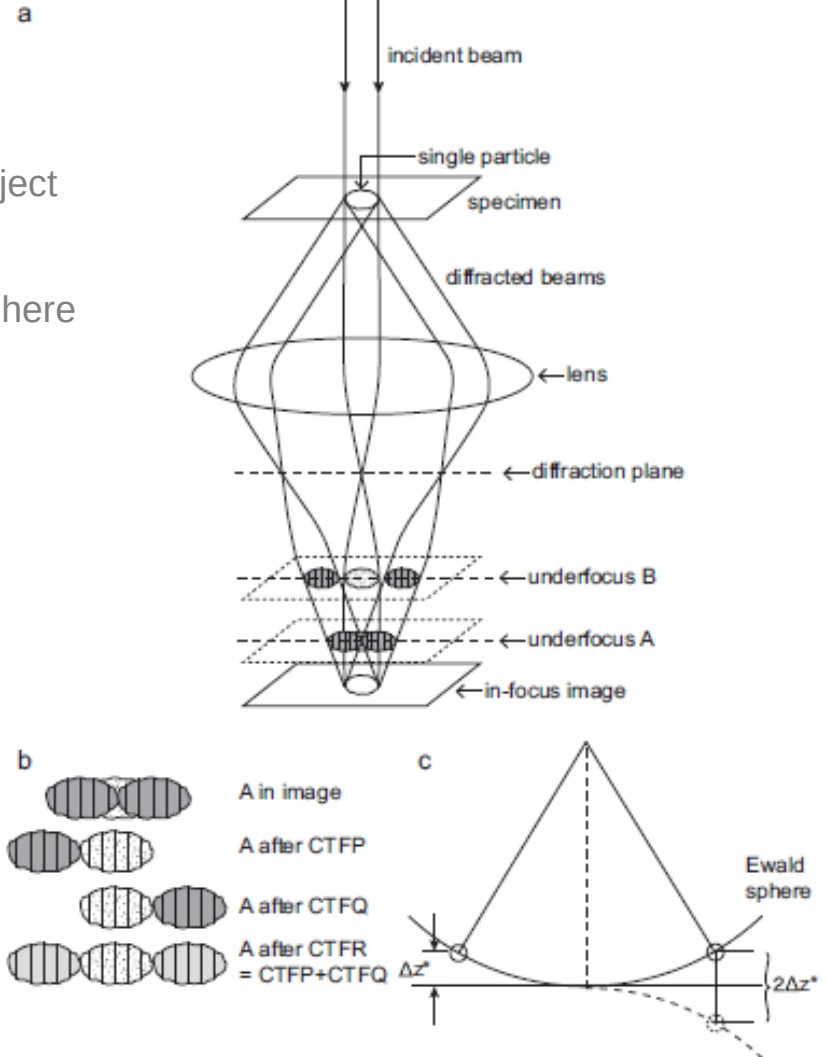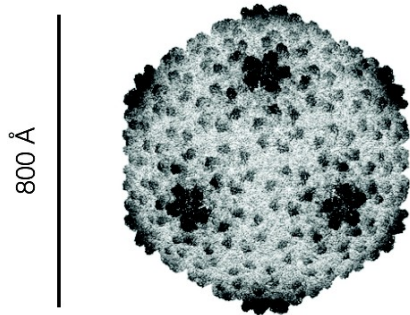
42*† T20S Proteasome

Noble et al eLife 2018

Map improvement

Ewald sphere correction

 - the assumption that the image is 2D projection of the 3D object does not hold for thick specimens

 - wave-function at the image plane samples surface of the sphere in 3D FT of the object

- Friedel symmetry is lost

- depends on electron wavelength – stronger effect (more important to consider) for 100keV than for 300keV

Russo & Henderson (2018), Ultramicroscopy

800 Å

c

$1/\lambda$

Ewald sphere

Δz^*

$2\Delta z^*$