CNRS
INSERM
Université de Strasbourg

Parc d'Innovation - 1, rue Laurent Fries - ILLKIRCH - C.U. DE STRASBOURG - FRANCE

Centre for Integrative Biology

Resolution estimation, cryo-EM map interpretation & Atomic model refinement into cryo-EM maps

Master M2S3 level, Brazil school & ISB workshop, Oléron school, Freiburg Structural Biology School

Bruno Klaholz,

Centre for Integrative Biology, IGBMC, Illkirch/Strasbourg

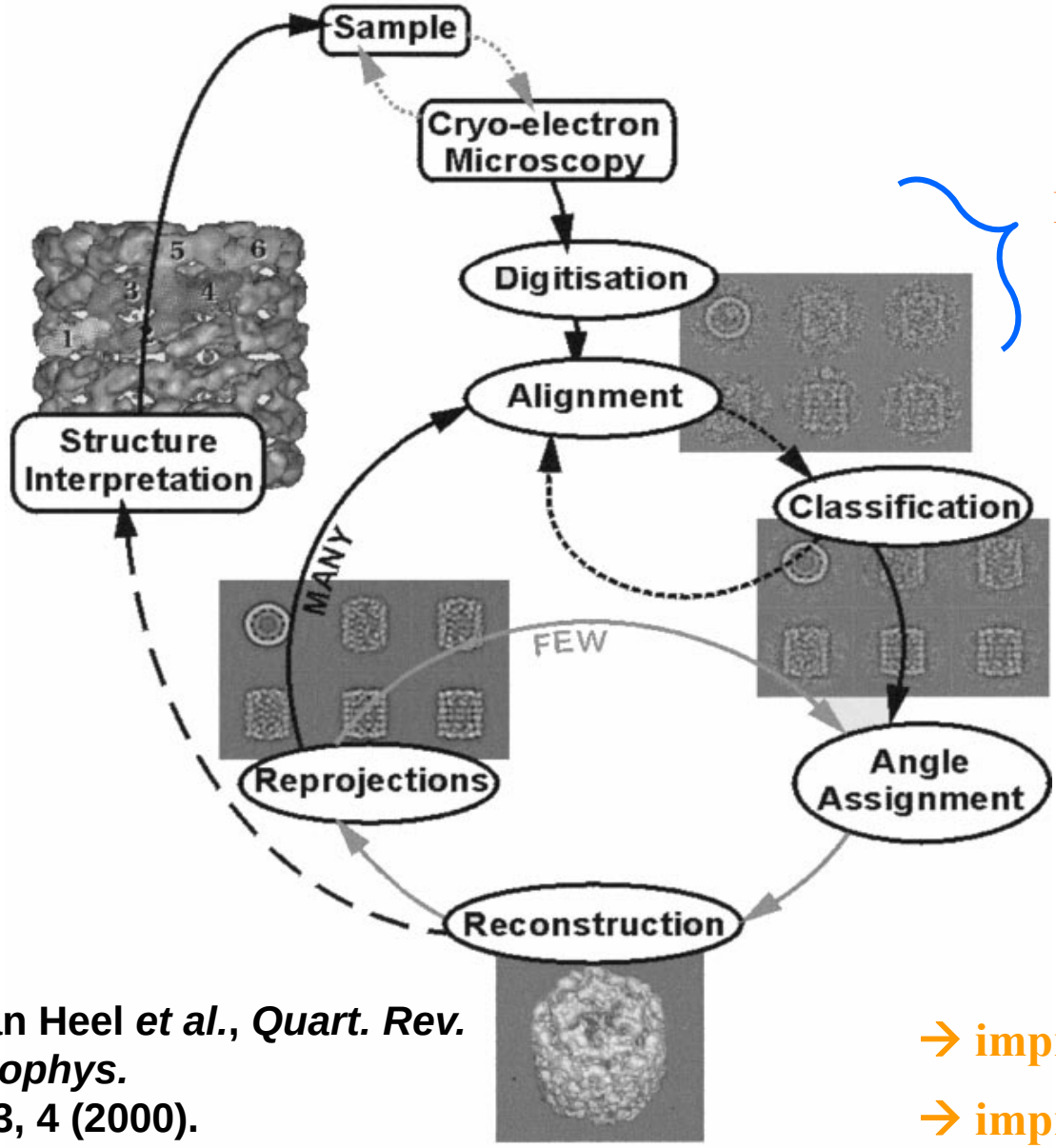
2019

http://igbmc.fr/Klaholz

Note: you are welcome to reuse slides elsewhere but please cite your source

Summary

Structure determination by single particle cryo-EM, overview:



van Heel *et al.*, *Quart. Rev. Biophys.* 33, 4 (2000).

I. Pre-processing

II. Structure determination

→ **improve quality of angle assignment**

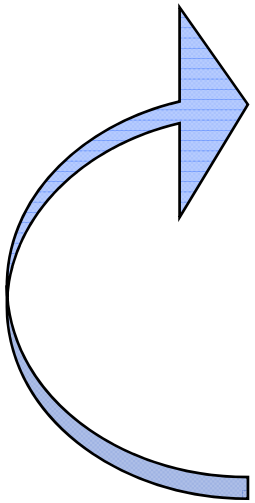
→ **improve quality of particle alignment**

IGBMC

II. Structure determination

- structure refinement

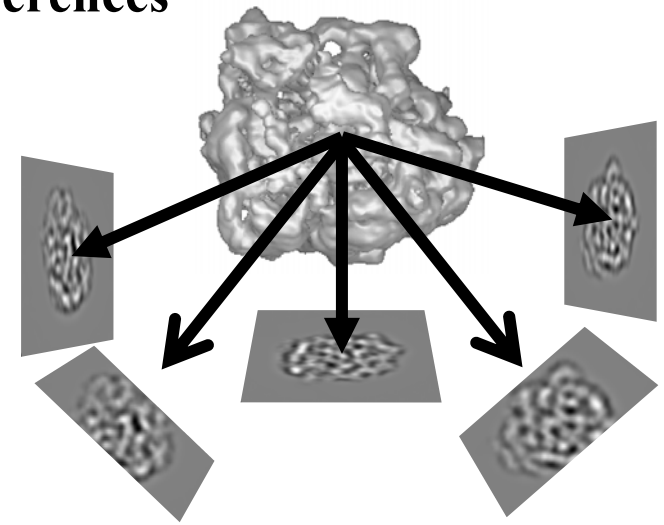
iteration



- centering/alignment
- variance analysis + classification
- angle assignment
- angular reconstitution → 3d-reconstruction
- reprojections = new references

"phase/put in register" the particles by aligning/classifying them

- improve quality of angle assignment
- improve quality of particle alignment



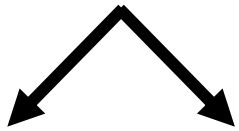
equally distributed forward-projections (re-projections)

IGBMC

II. Structure determination

- resolution assessment

Particle data set



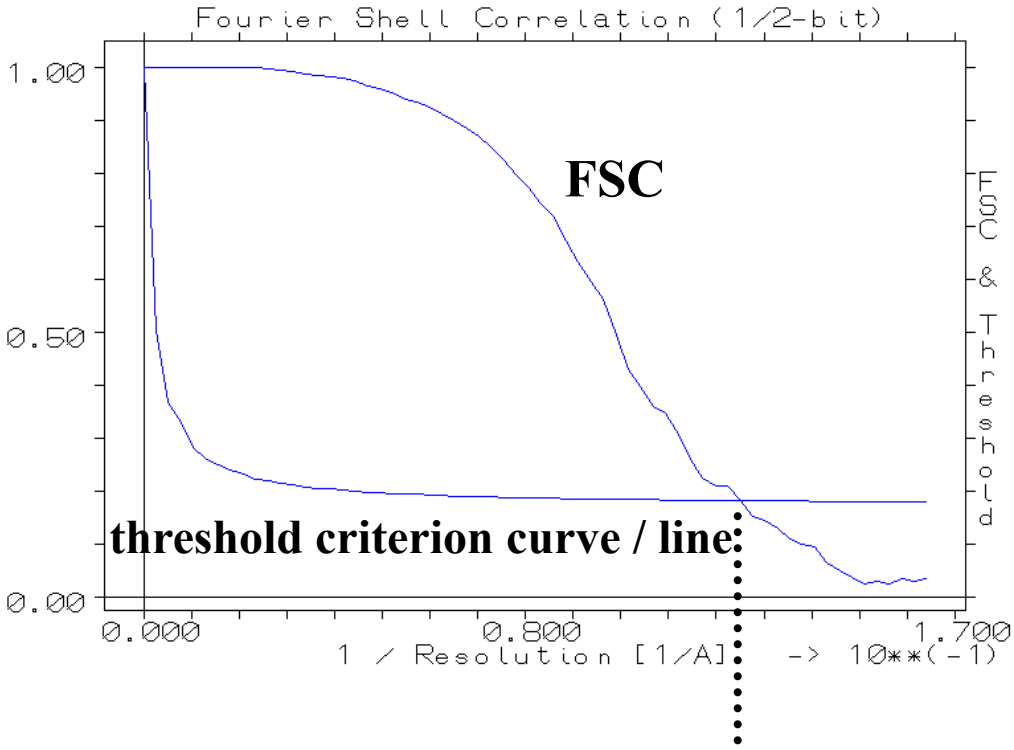
split into 2 halves

3D **3D**

calculate cross-correlation

by shells in Fourier space

Fourier Shell Correlation (FSC)



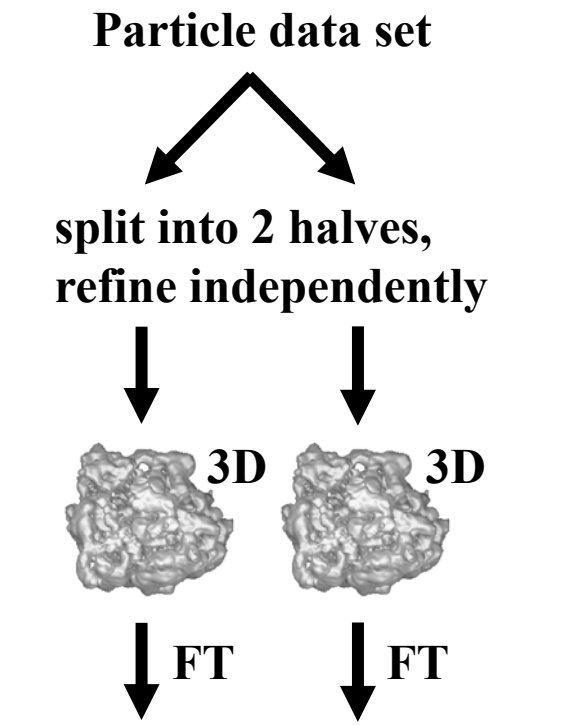
1 / 1.25 x 1 / 10 Å = 8 Å

Keep in mind: resolution is what you can resolve in the 3D map!

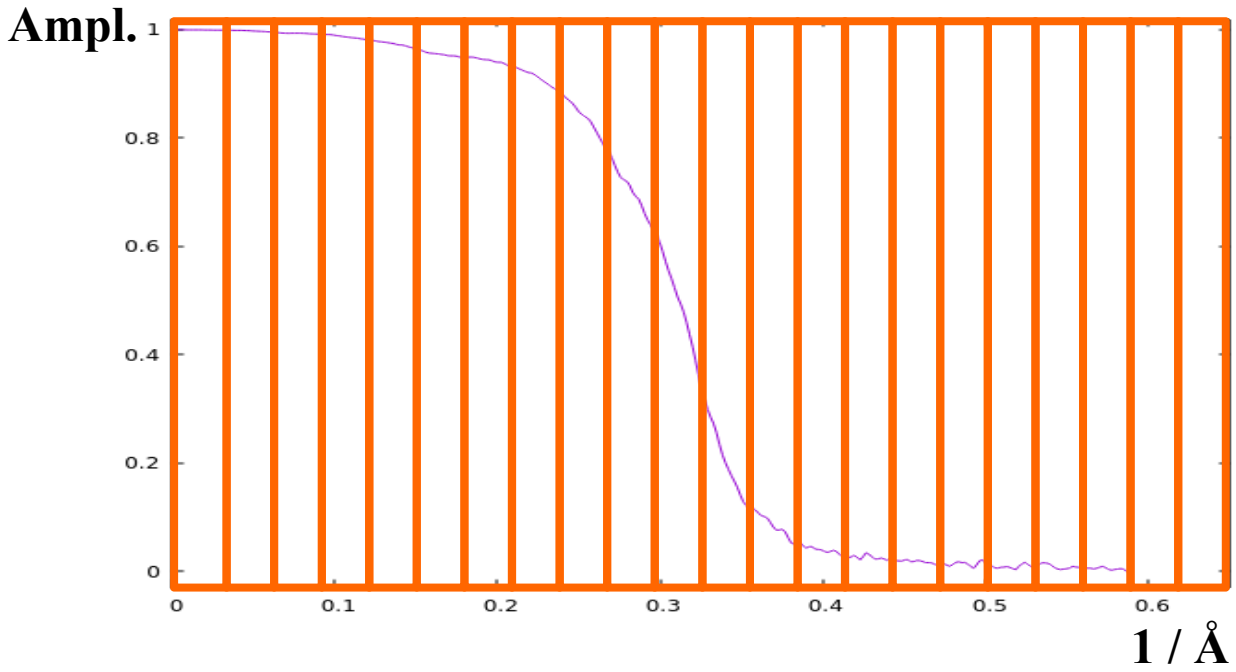
IGBMC

II. Structure determination

- resolution assessment



Fourier Shell Correlation (FSC)



Cross correlation is calculated in Fourier space as a function of frequency (i.e. resolution shells)

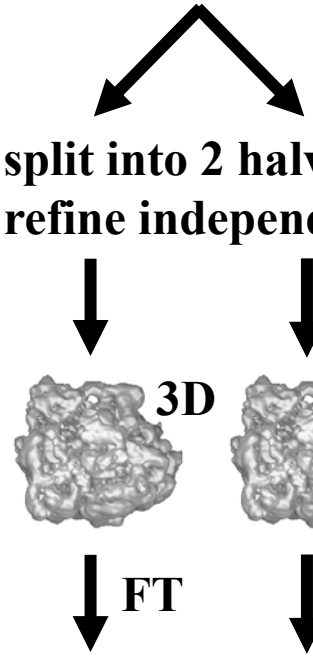
Fourier Shell Correlation (FSC): van Heel *et al.*, 1982, Saxton & Baumeister, 1982.

IGBMC

II. Structure determination

- resolution assessment

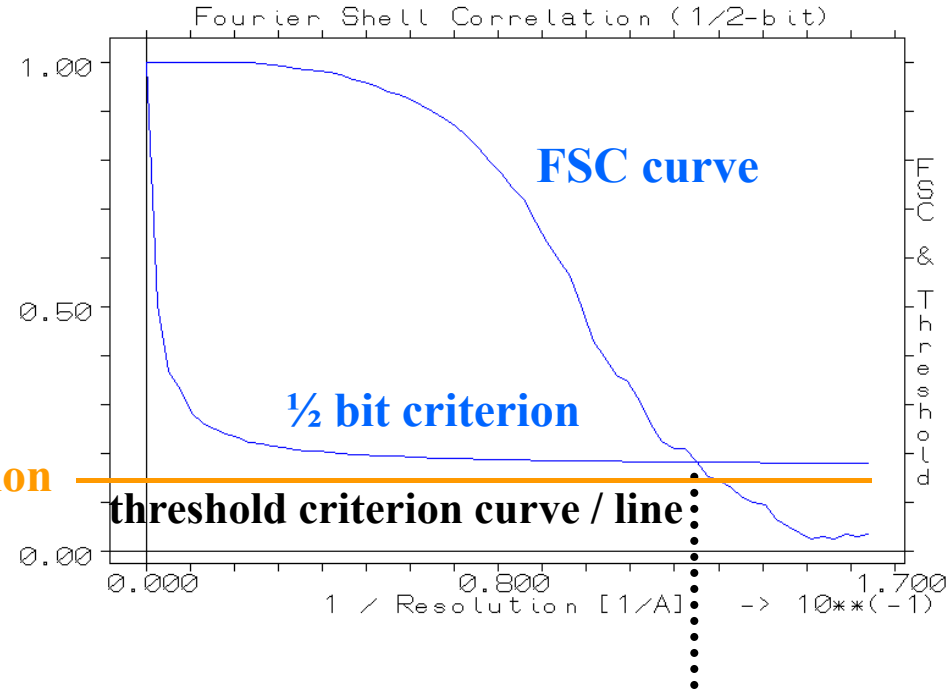
Particle data set



calculate cross-correlation

by shells in Fourier space

Fourier Shell Correlation (FSC)



1 / 1.25 x 1 / 10 Å = 8 Å

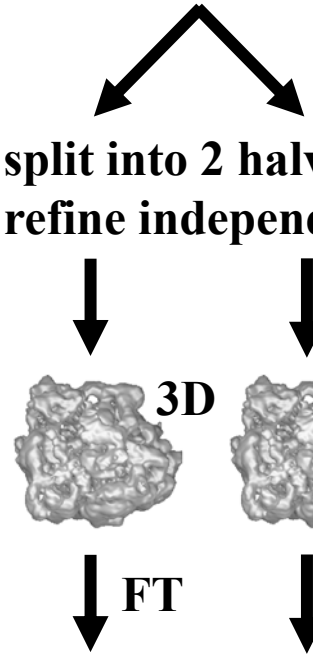
Rosenthal & Henderson, 2003 (0.143 criterion)
van Heel & Schatz, 2005 (1/2 bit criterion)

IGBMC

II. Structure determination

- resolution assessment

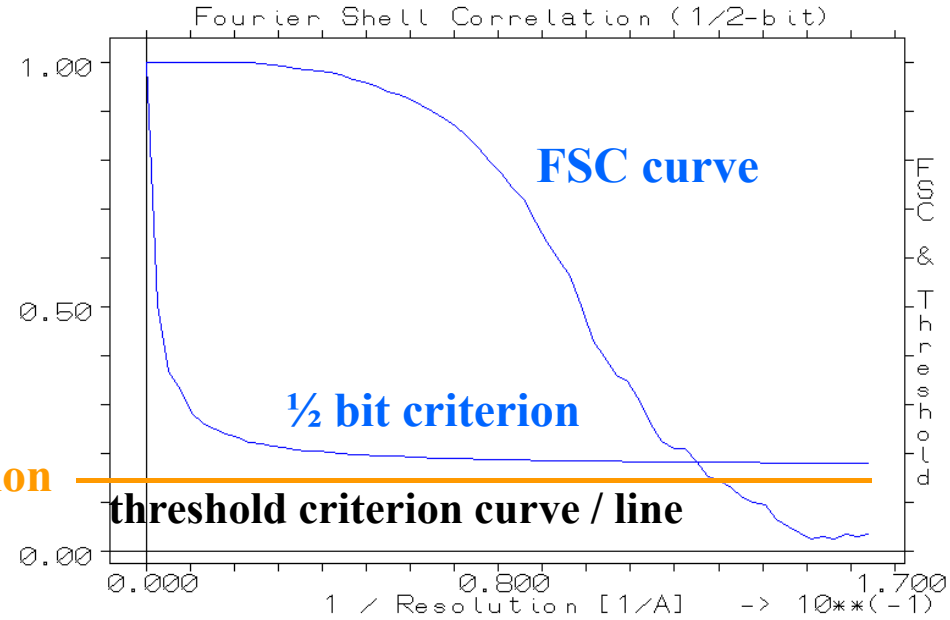
Particle data set



calculate cross-correlation

by shells in Fourier space

Fourier Shell Correlation (FSC)



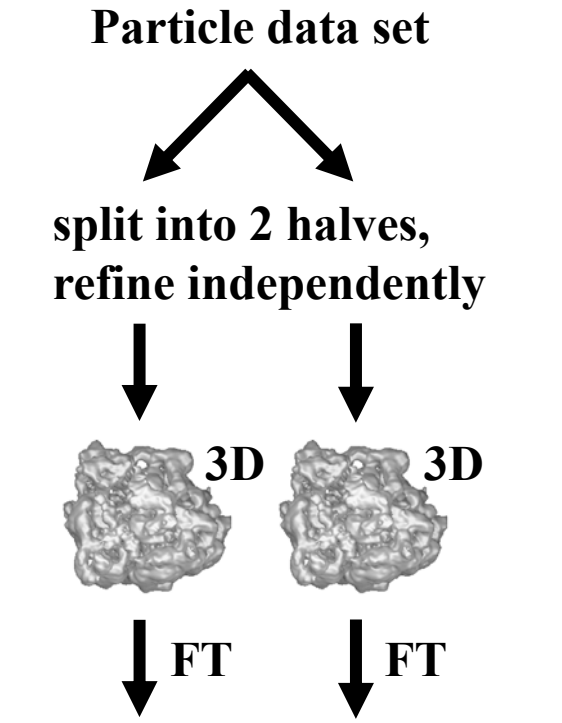
Take care to

- **do not cut off spectrum before Nyquist limit**
- **be generally careful with fixed threshold criteria**
- **masking can give artefacts at high frequencies**
- **maximum resolution possible ~2/3 Nyquist (unless over-sampling is used)**

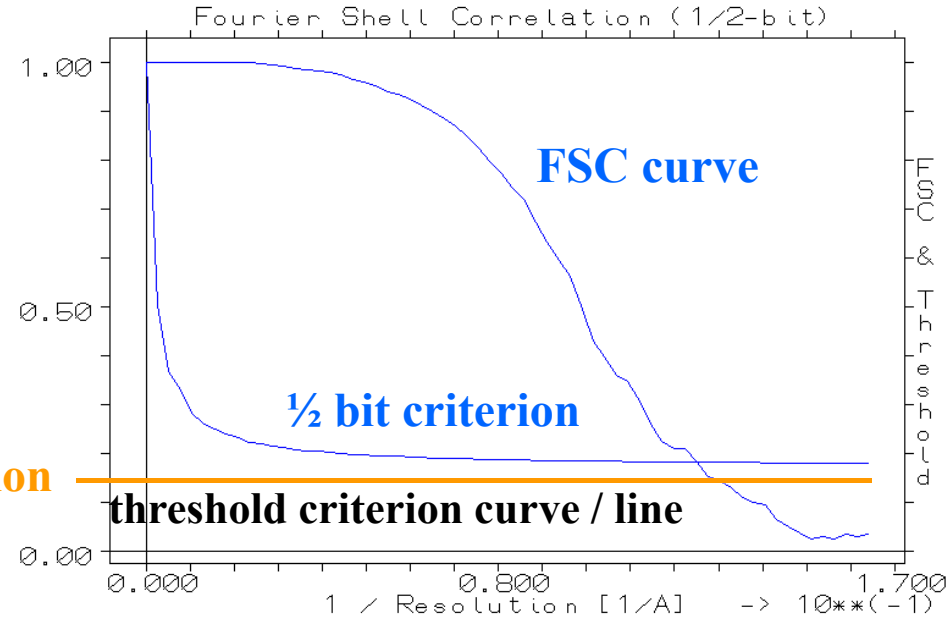
IGBMC

II. Structure determination

- resolution assessment



Fourier Shell Correlation (FSC)



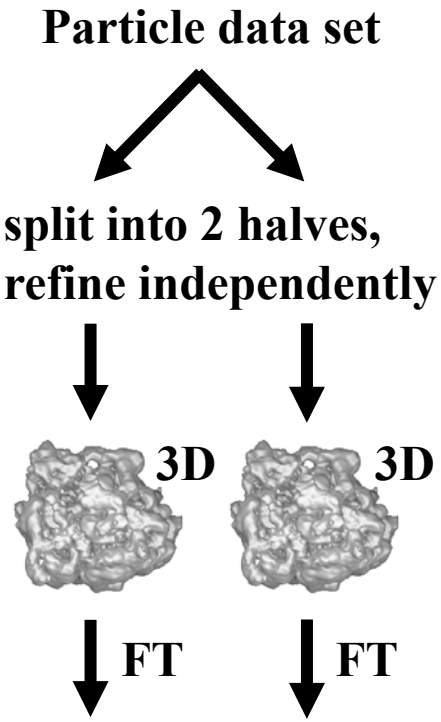
Due to the *Nyquist-Shannon theorem* and interpolation effects a 3D reconstruction can have a maximum resolution without distortions of ~2/3 Nyquist frequency, beyond this limit high-frequency components can be present but may be distorted

Keep in mind: resolution is what you can resolve in the 3D map!

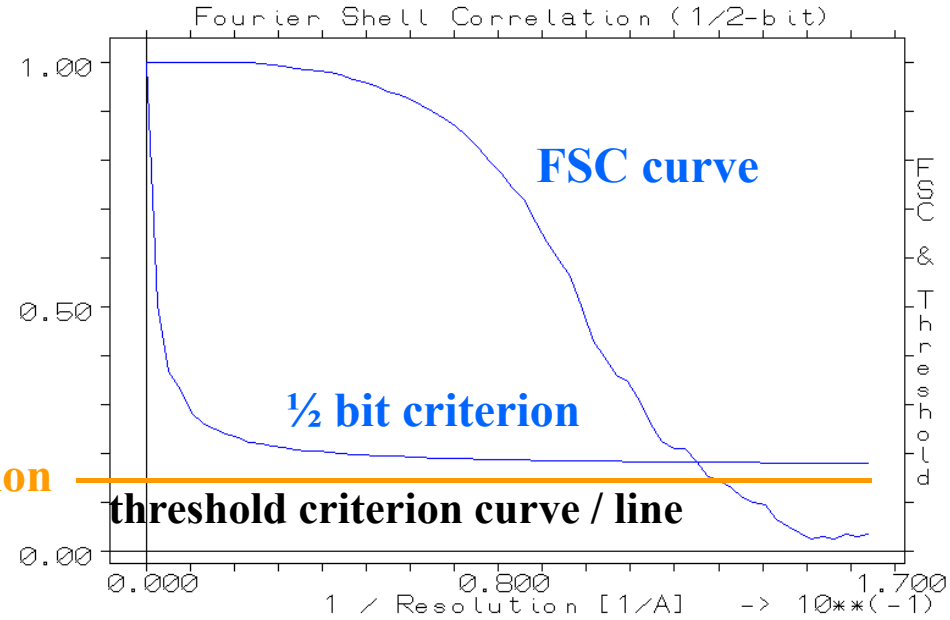
IGBMC

II. Structure determination

- resolution assessment



Fourier Shell Correlation (FSC)



Careful with artefacts: too tight masking (mask correlates with itself, gives high correlation at high frequencies, i.e. the curve goes up); careful with correlated noise at frequencies, use phase randomization (e.g. Relion postprocessing)

Keep in mind: resolution is what you can resolve in the 3D map!

IGBMC

Some basic concepts of cryo electron microscopy

Basic aspects:

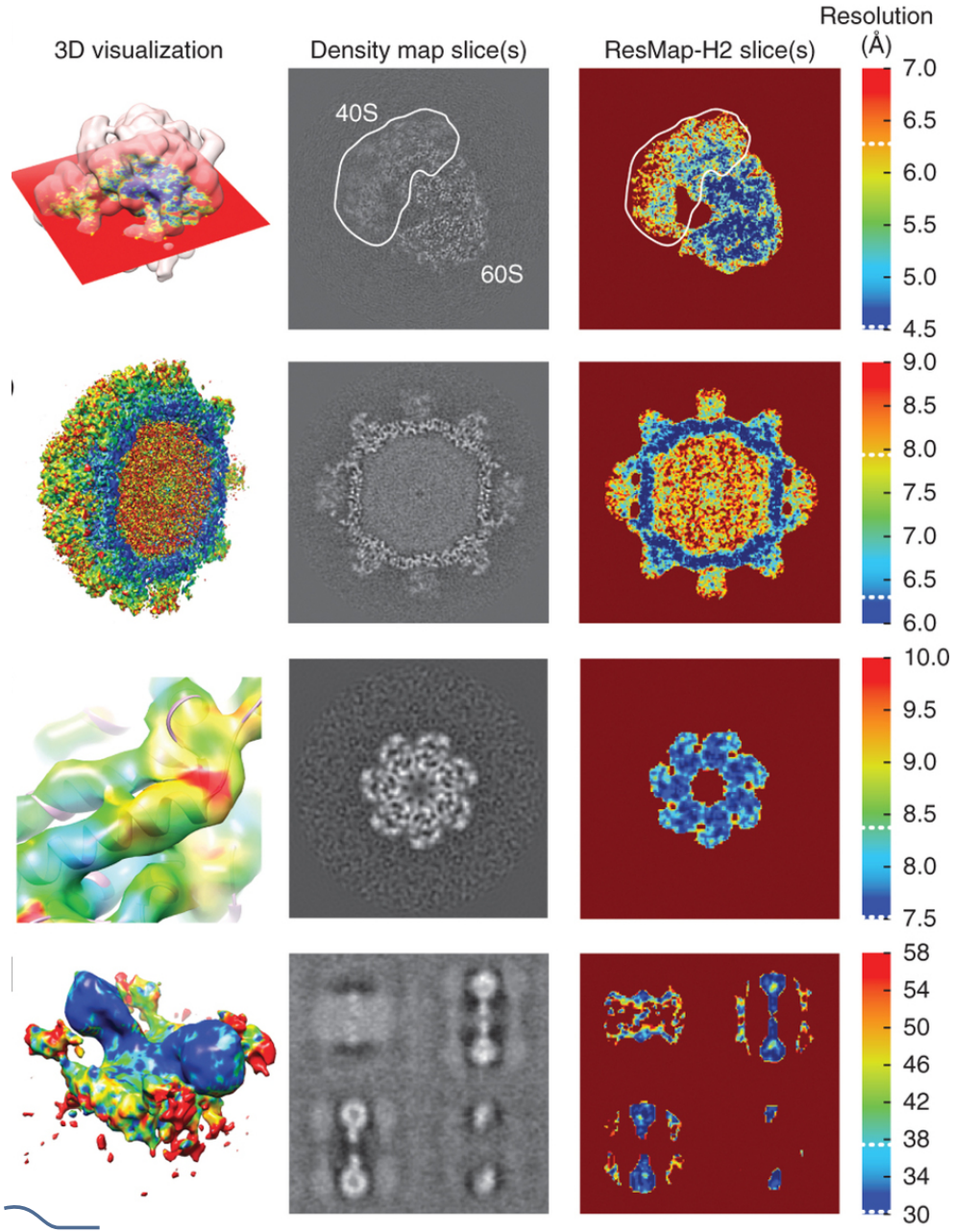
- "resolution" corresponds to "spatial frequency" in image processing (1/ Å)
- Nyquist frequency is = 1 / (2 x pixel size), e.g. 1 Å / pixel → Nyquist = 1 / (2 Å)
- interpolations during 2D image alignment and 3D reconstruction limit the possible resolution to about 2/3 of the Nyquist frequency, i.e. here ~ 3 Å *(exception: super-reso)*

pixels in 3D: "voxel"

Consider:

- any correlation calculation (e.g. alignment) is biased by the reference used
- resolution estimation, criteria used:
  - 0.5, arbitrary, historically from the virus field, tends to underestimate resolution
  - 0.143 (Henderson) and ½ bit (van Heel)
  - 3 σ, not used anymore (over-estimation; useful for noise estimation)
  - features in the map: can we see dsRNA helices (~10-12 Å resolution), α-helices (~8 Å), ß-sheets (~5 Å) or side chains (4-2.5 Å, depending on size)?

Local resolution analysis



Some softwares:

Bsoft (Heymann *et al.*, 2007)

ResMap (Kucukelbir *et al.*, 2014)

MonoRes (Vilas *et al.*, 2018)

Phenix (d99, phenix.autosharpen)

Afonine *et al.*, *Acta Cryst. D, 2018*.

Local resolution can be different from average resolution (FSC) due to structural disorder

IGBMC

II. Structure determination

- map interpretation

II. Structure determination

- map interpretation ; fitting of crystal or NMR structures

Fitting procedures:

- manual fitting (e.g. Coot, Emsley *et al.*, 2010)

- real space fitting

- reciprocal space fitting

1) global search

2) refinement

e.g. torsion-angle molecular dynamics

- fit complete structures, domains, factors;

Usually backbone is enough.

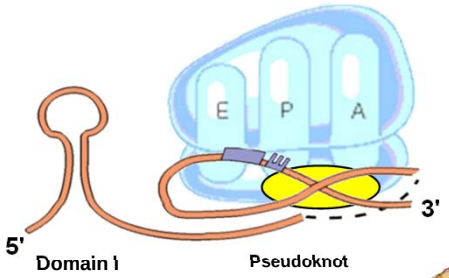
rigid body or flexible fitting

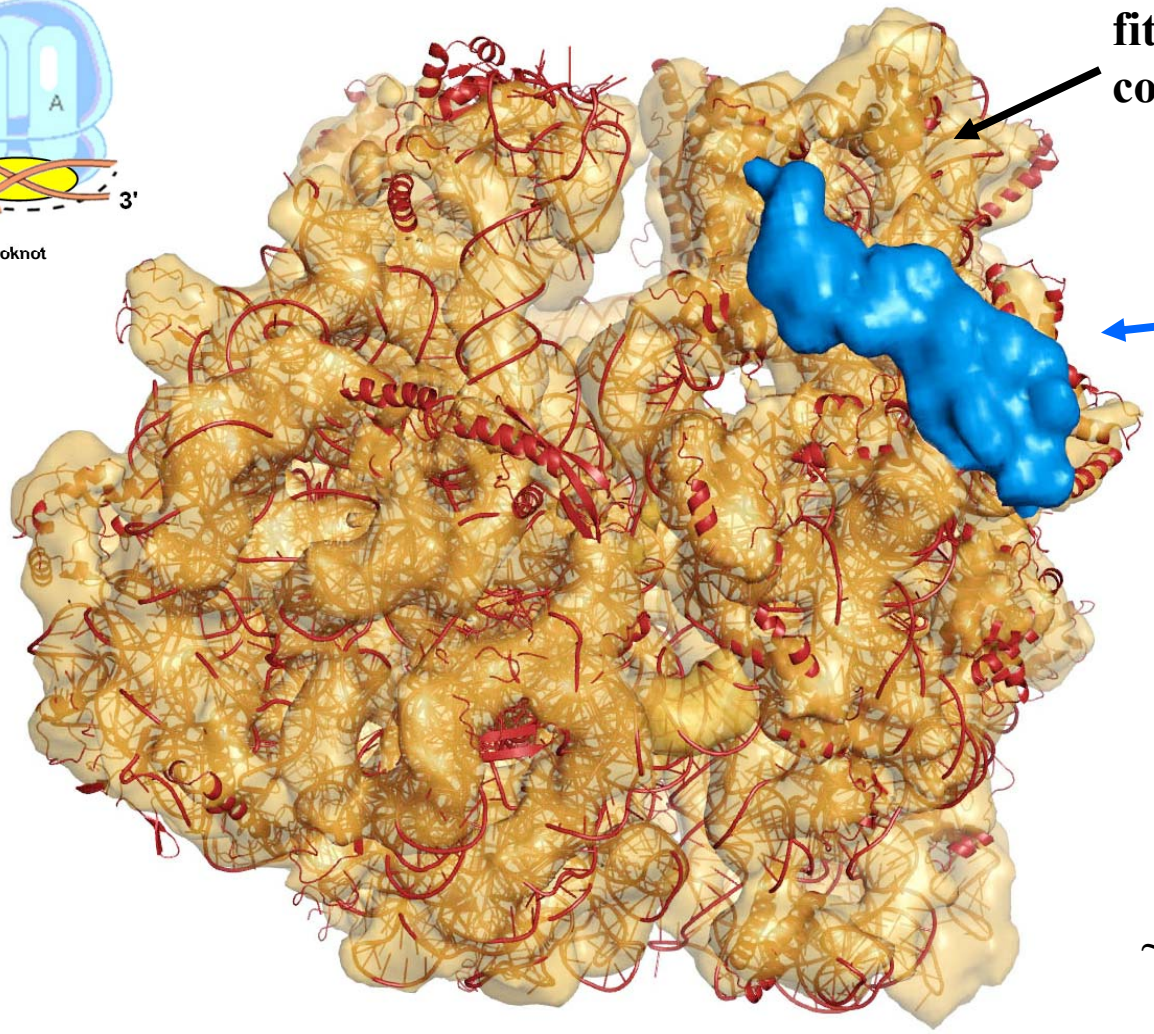
- use full maps or difference maps

Be careful with local minima and over-fitting!

II. Atomic model building into cryo-EM maps

II. Atomic interpretation of cryo-EM maps





fit atomic model of core complex

additional density, shows bound mRNA

~12 Å resolution

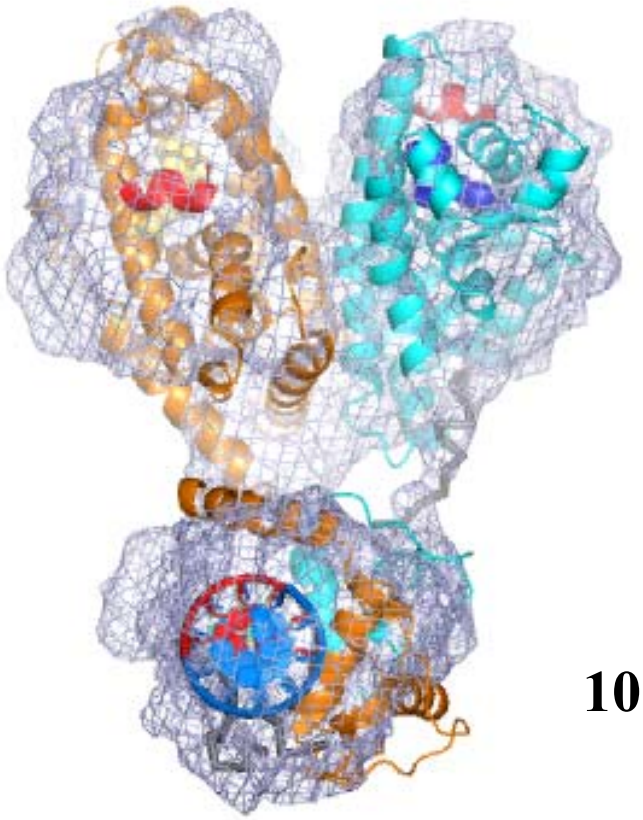
Marzi *et al.*, *Cell* 2007.

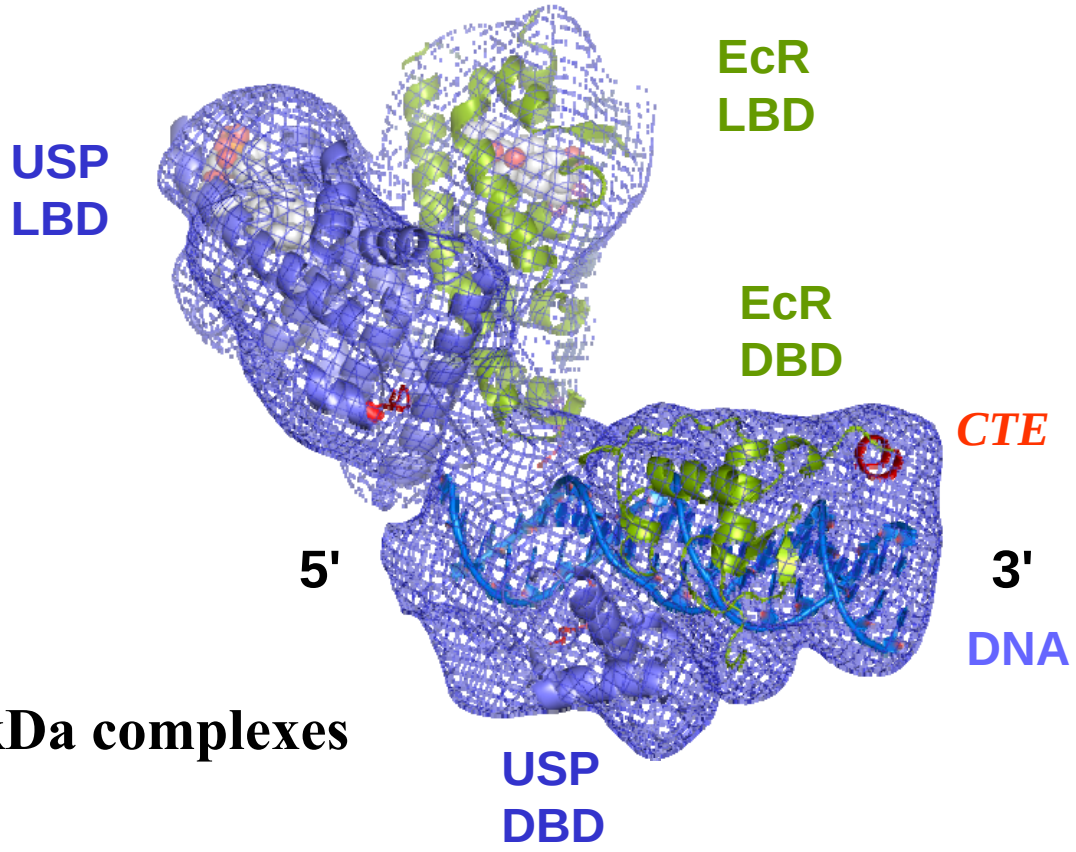
50S 30S

Fitting of 70S crystal structure from *E. coli* (Schuwirth *et al.*, 2005; *Science*, 310, 827-834)

II. Structure determination

- map interpretation





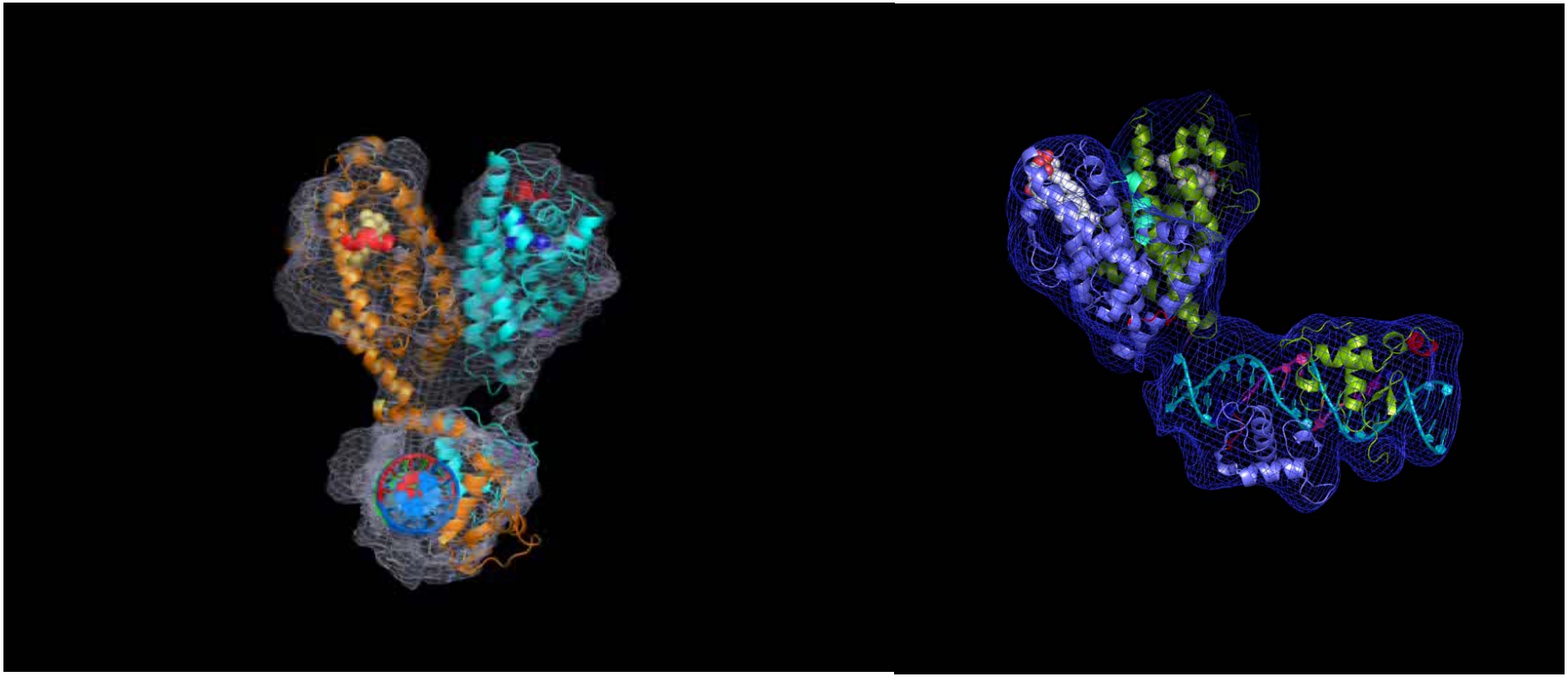
100 kDa complexes

Orlov *et al.*, *EMBO J.* 2012. Maletta *et al.*, *Nature Communications, 2014.*

Fitting of crystal structures from the various protein domains

IGBMC

Instrumentation & technical highlights towards multi-scale integration



100 kDa* nuclear receptor complexes: *RXR/VDR/DNA & USP/EcR/DNA

(on film or CCD camera, several years before CMOS cameras)

Orlov *et al.*, *EMBO J.* 2012. **Maletta *et al.*, *Nat. Commun., 2014.***

Important: medium-resolution maps give you the overall topology and conformational state, which are not a matter of resolution; i.e. filter a map to see the global domains etc.

IGBMC

Map interpretation, structure determination, atomic model building

Fitting procedures:

- manual fitting (e.g. O, Coot, Pymol, Chimera…)
- real space fitting
- reciprocal space fitting

1) global search

2) refinement

At ~8-20 Å resolution:

- fit complete structures, protein or RNA domains, factors; usually backbone is enough.

Rigid body or flexible fitting (e.g. Situs, MDFF, Flex-EM, iMODfit, …)

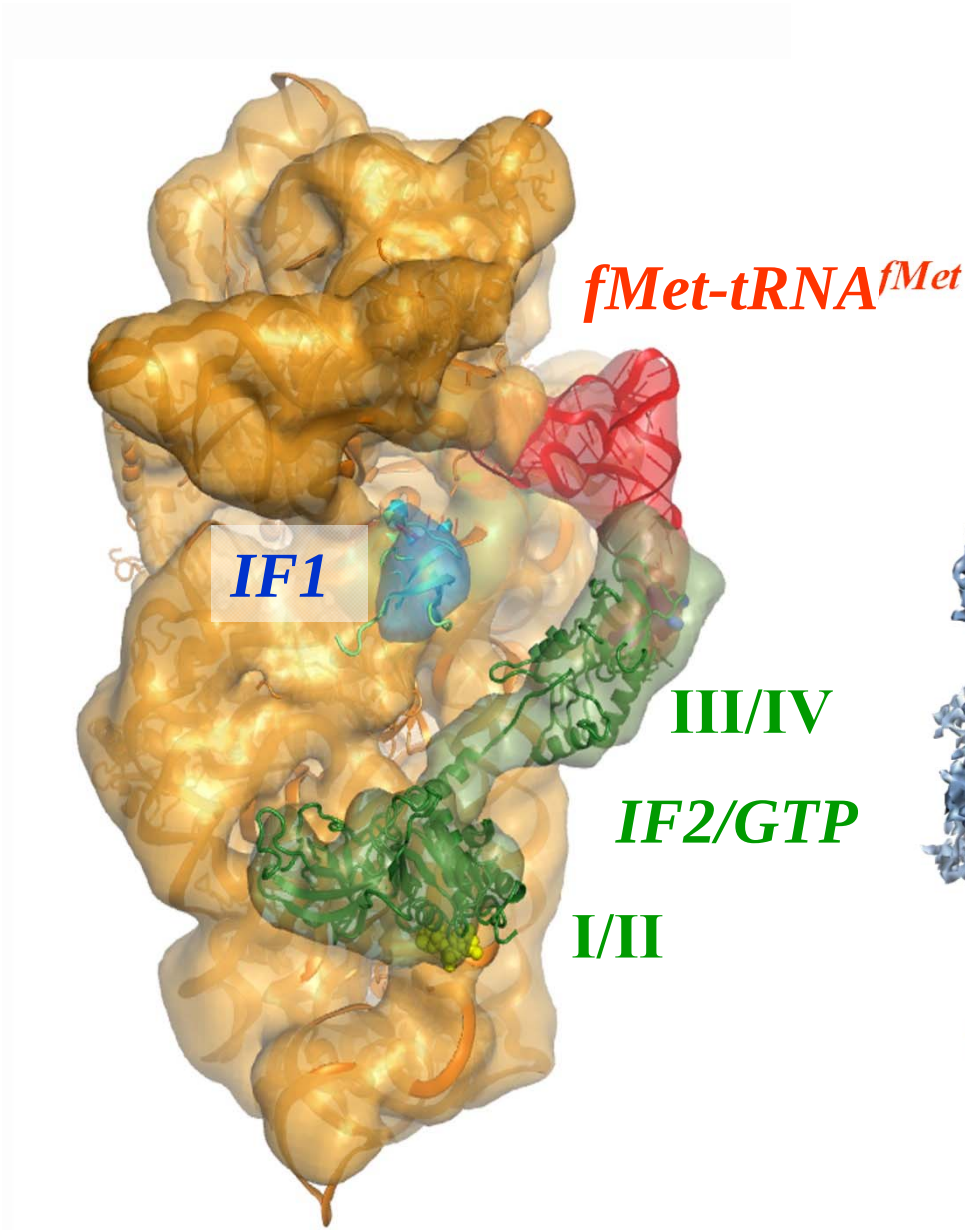
- use full maps or difference maps

At ~3-5 Å resolution:

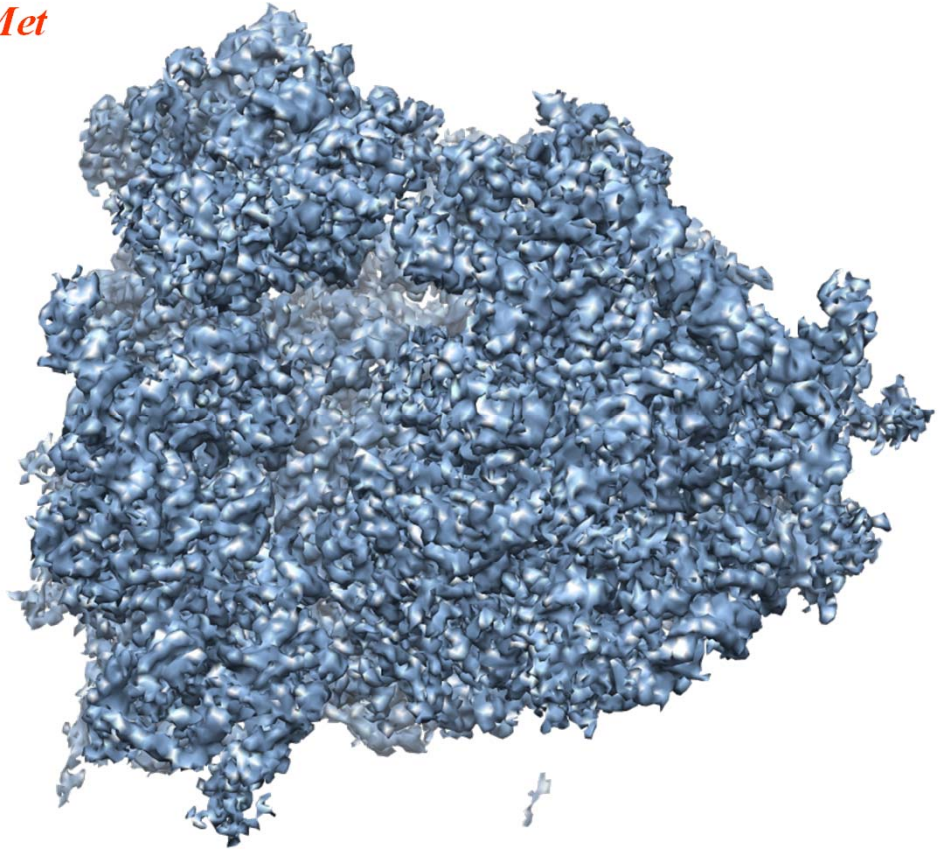
- atomic model building: start with poly-Ala model, check register (position of Cα atom), check secondary structure elements (e.g. direction of α-helices), refine with crystallography programs (CNS, Buster, Phenix, CCP4,…), add side-chains if clearly visible,

use information from multi-sequence alignments; check geometry with Ramachandran plot

In general: be careful with local minima and over-fitting/over-interpretation!



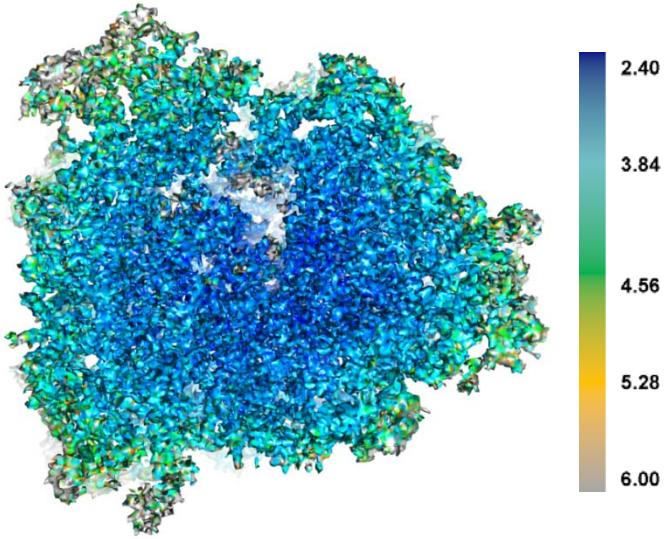
Simonetti *et al.*, *Nature, 2008.*

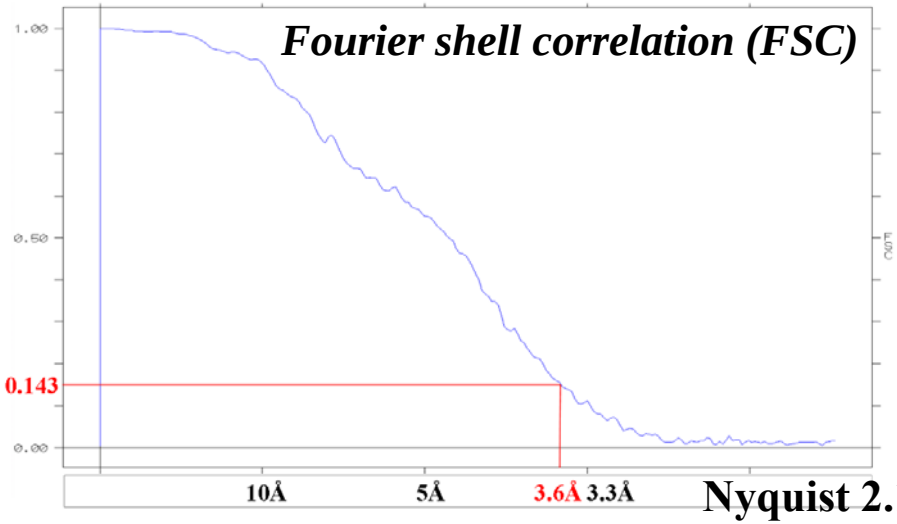


(was) ongoing structure refinement
(Oct. 2014)

IGBMC

Structure determination at ~ 3 Å resolution by single particle cryo-EM

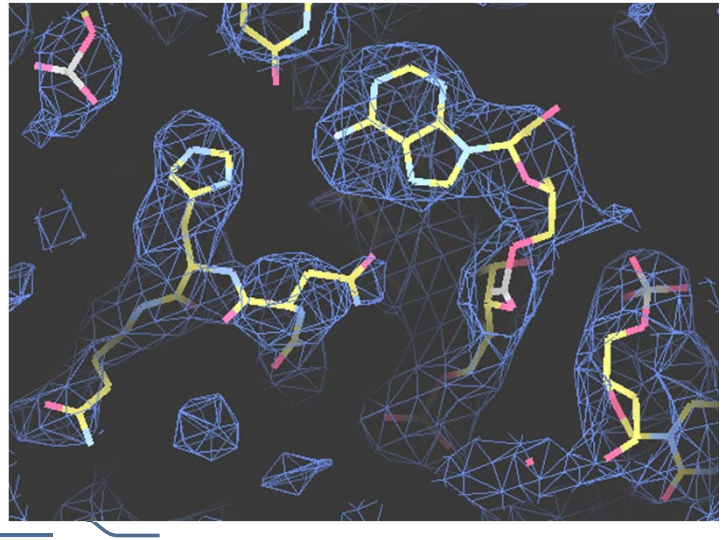


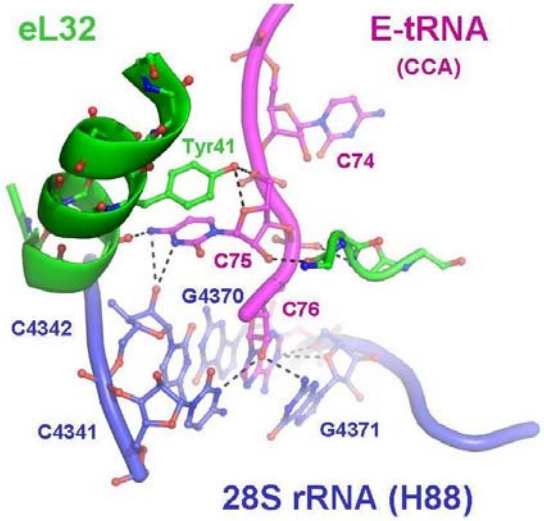


Nyquist 2.16Å

→ **local resolution can be notably better than average resolution**

cryo-EM map & fitted atomic model

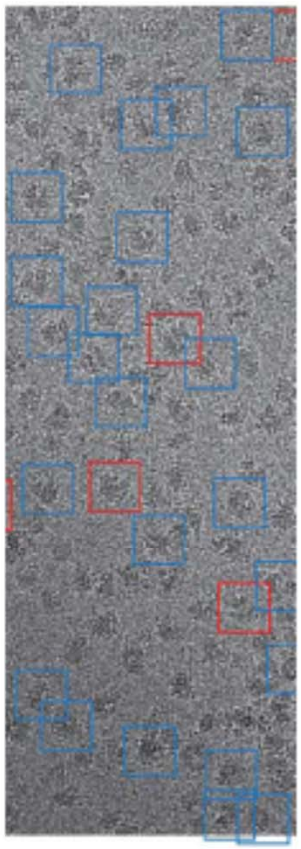


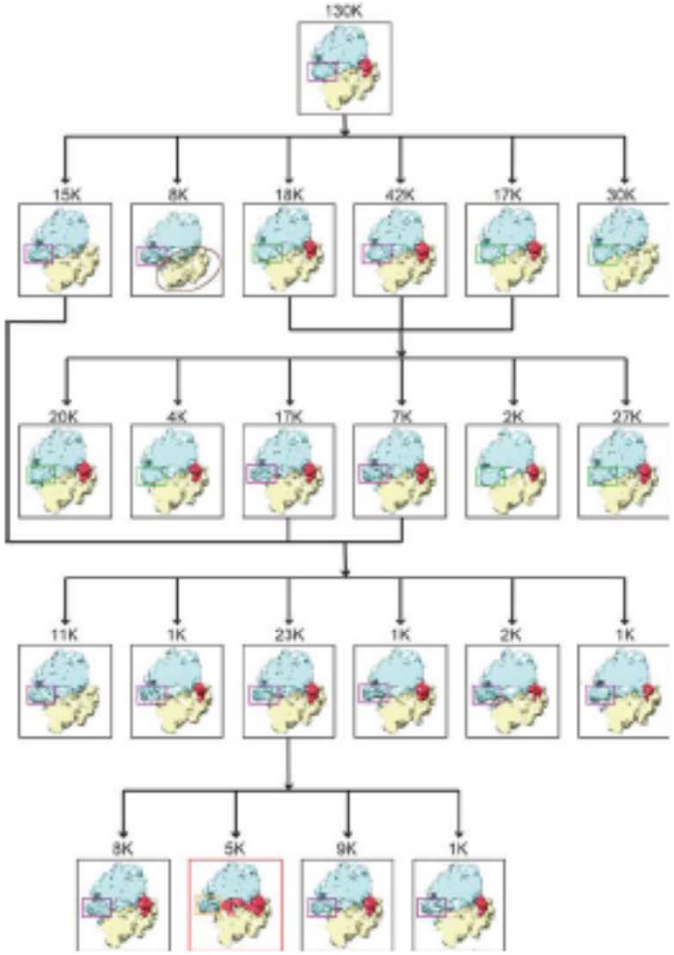


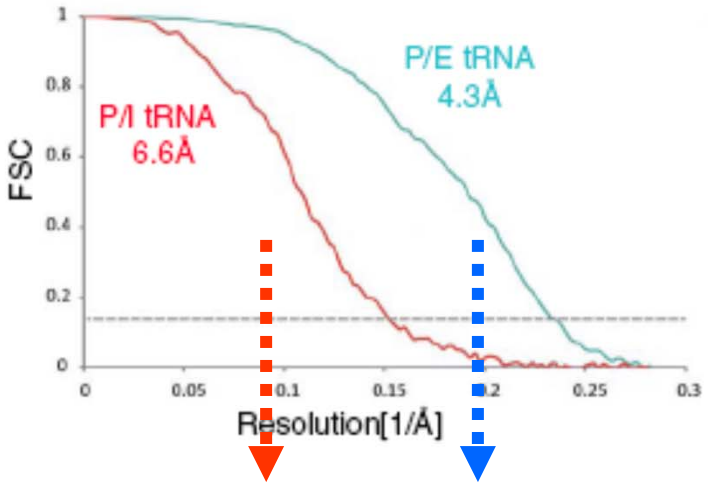
IGBMC

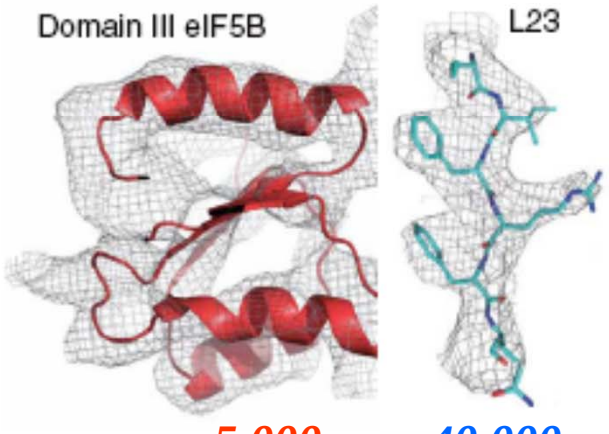
Khatter *et al.*, *Nature, 2015.*

II. Atomic interpretation of cryo-EM maps







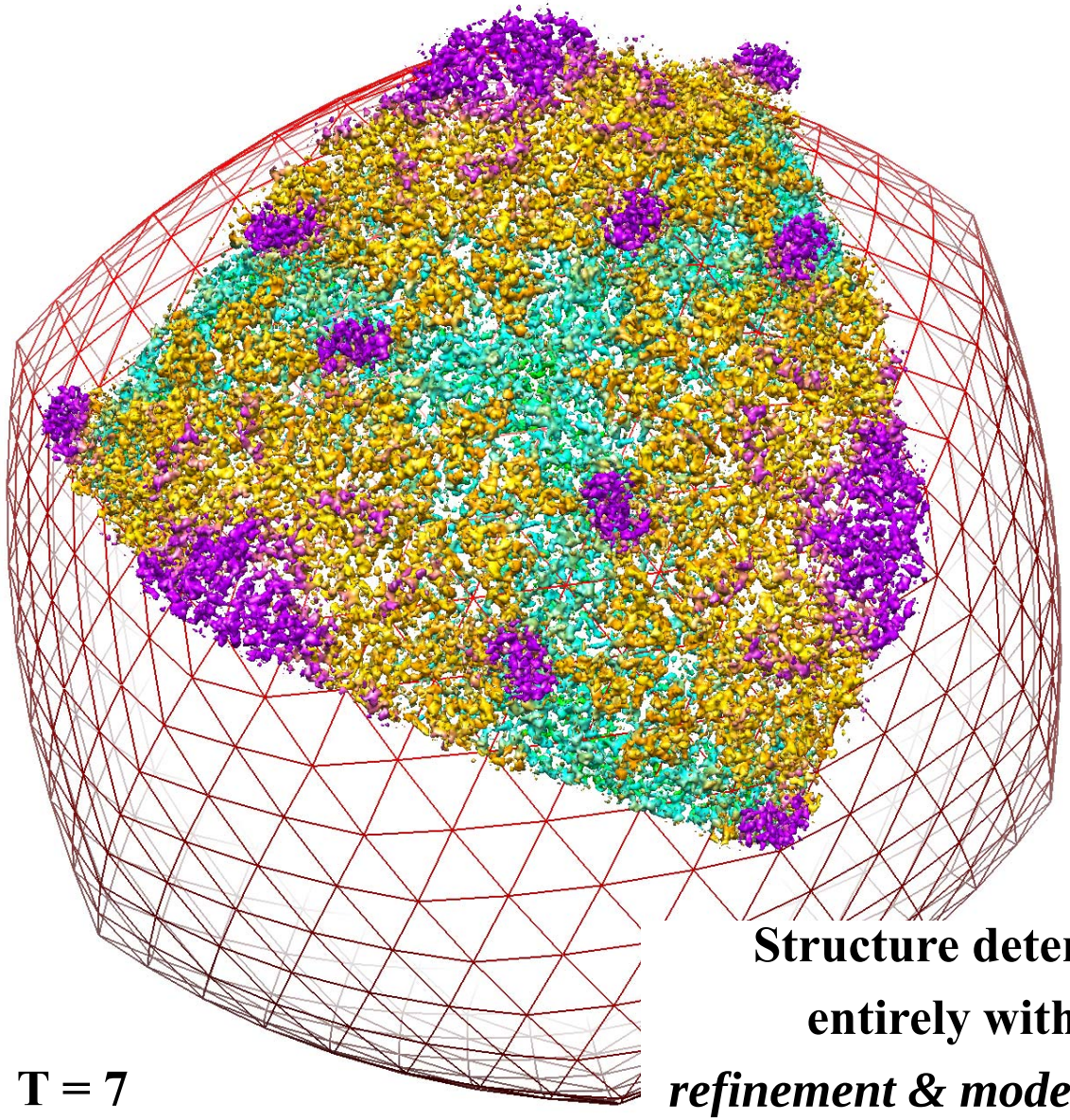


5 000 *40 000* ***particles***

Strong heterogeneity of a reconstituted eukaryotic translation initiation (eIF5B) complex: sorting → 5143 particles, representing 3% of the population in the sample, 6.6 Å reconstruction. Fernández *et al.*, Science 2013; V. Ramakrishnan & S. Scheres.

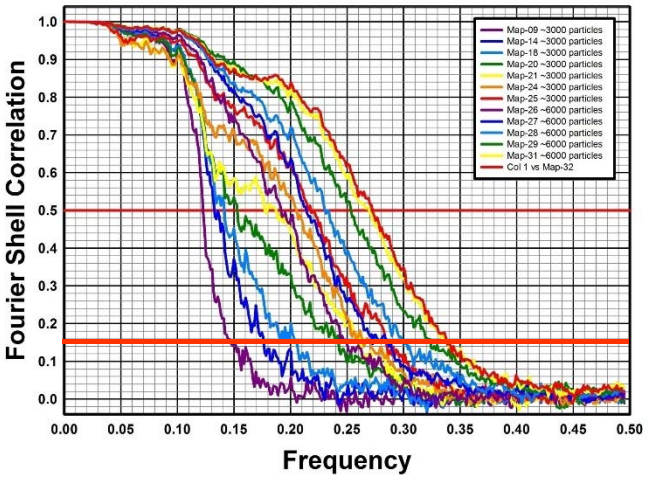
IGBMC

Mature FULL Capsid - 3.7Å resolution



T = 7

FSC curves during cryo-EM structure refinement

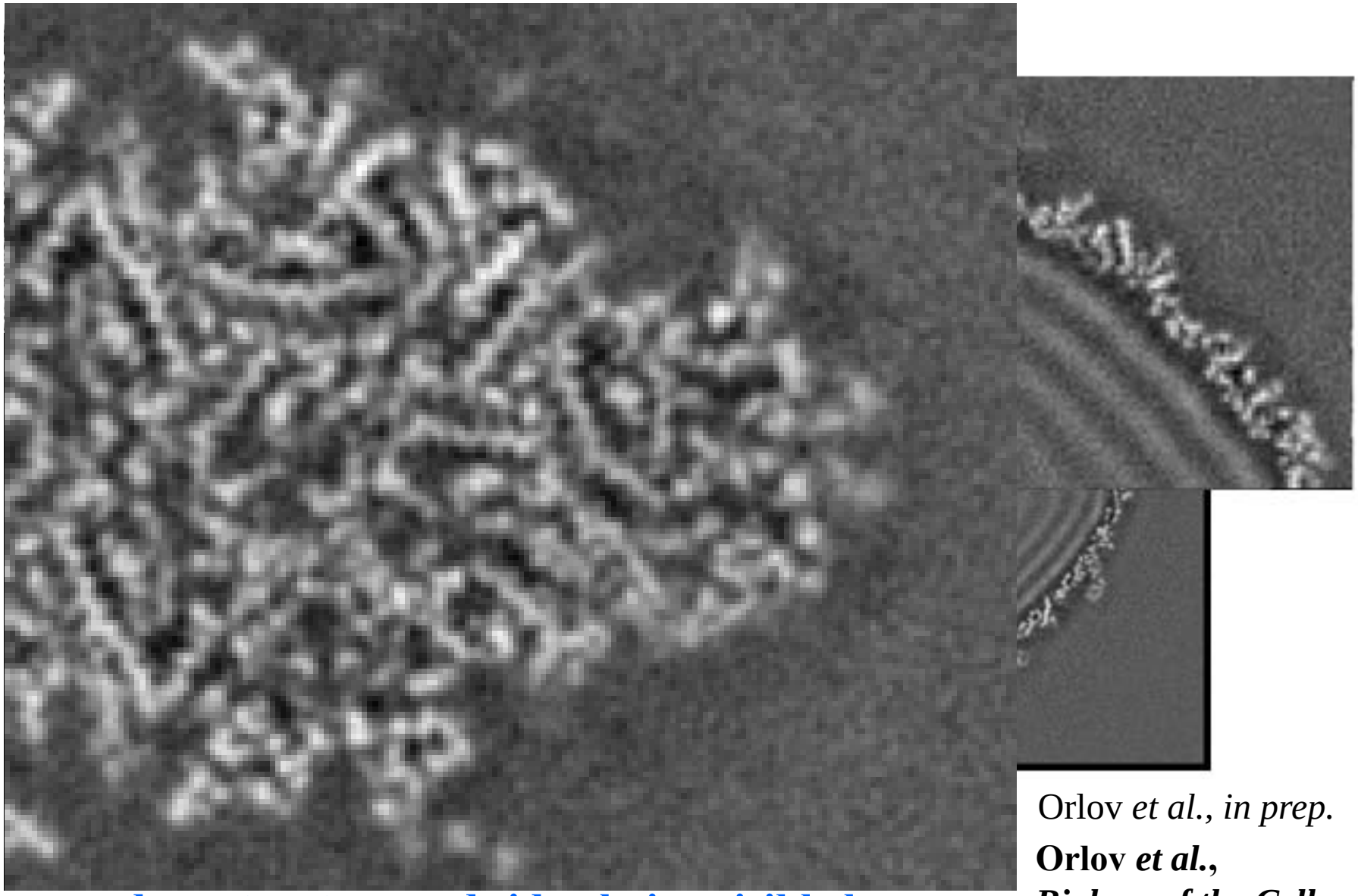


Structure determination and refinement:

entirely with common lines (Imagic)

refinement & model building ongoing*, Orlov *et al.

Instrumentation & technical highlights towards multi-scale integration

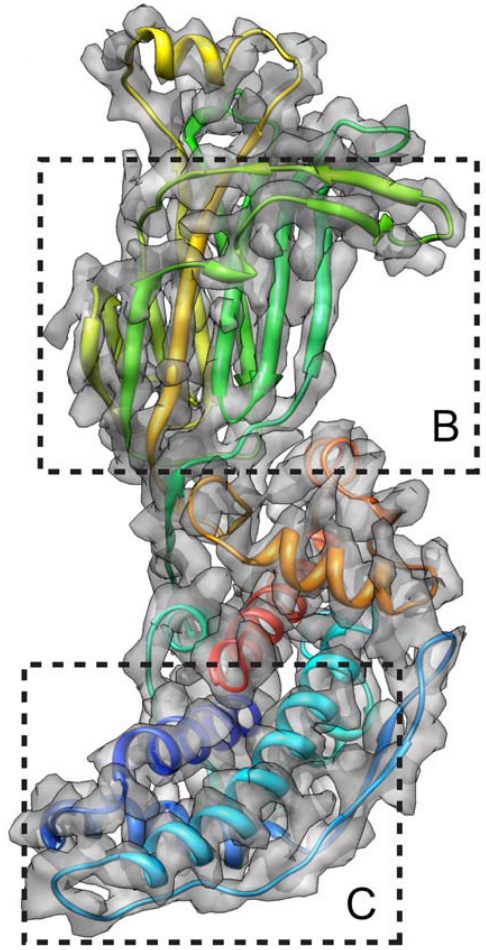


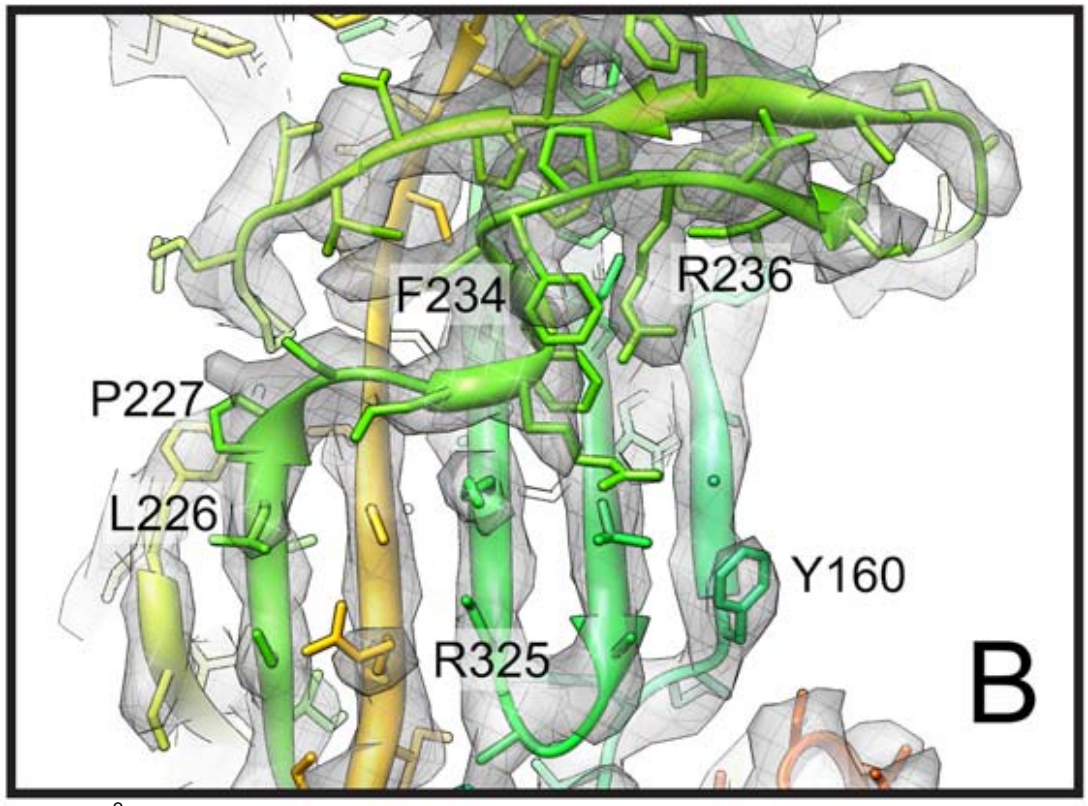
secondary structure and side-chains visible by eye…

Orlov *et al., in prep.*

Orlov *et al.*, *Biology of the Cell*, 2017.

Atomic model building examples in cryo-EM





Rotavirus VP6 cryo-EM structure; 3.8 Å resolution; α-helices, β-sheets, bulky side-chains;

Individual stands in the β-sheet region are separated, loops connecting the strands are defined.

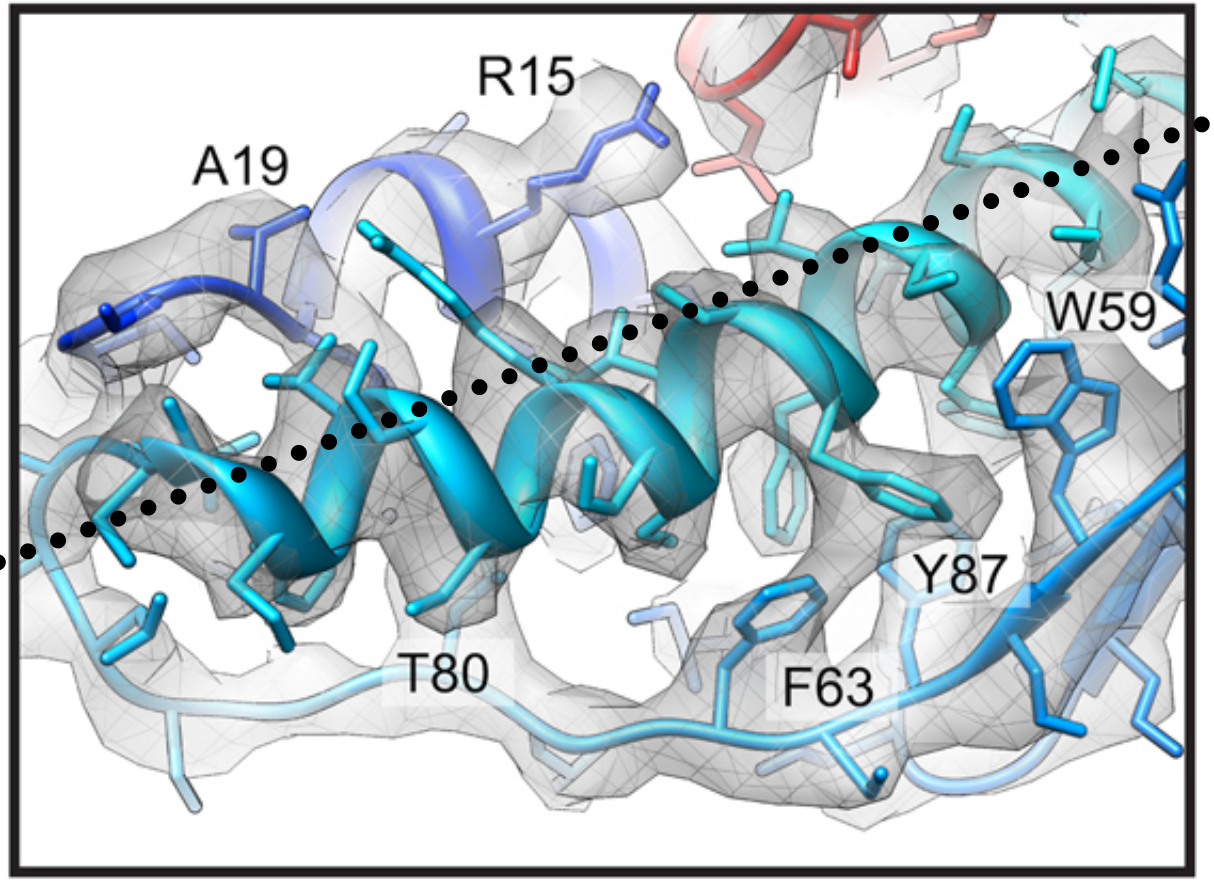
Near-atomic-resolution cryo-EM for molecular virology.

Hryc CF, Chen DH, Chiu W. *Curr Opin Virol.* 2011.

ICGBMC

Atomic model building examples in cryo-EM

high-resolution features: right-handed protein and DNA/ARN helices!



side-chains point to the N-ter

Rotavirus VP6 cryo-EM structure; 3.8 Å resolution; α-helices, β-sheets, bulky side-chains;

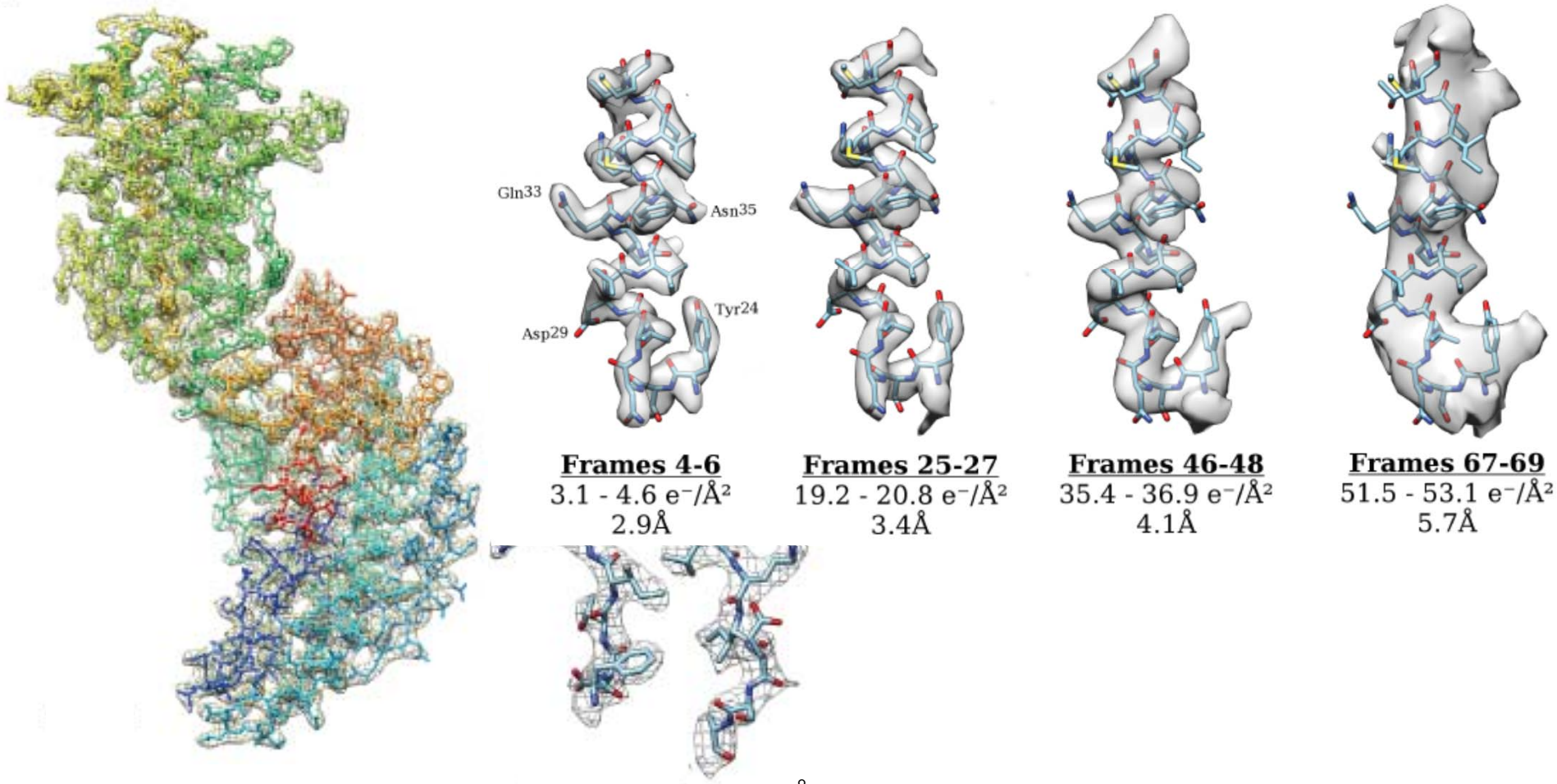
Individual stands in the β-sheet region are separated, loops connecting the strands are defined.

Near-atomic-resolution cryo-EM for molecular virology. Hryc, Chen, Chiu W., *Curr Opin Virol.* 2011.

See also tools from Wah Chiu lab on atomic model building: *Curr. Op. Struct. Biol.* 2015

IGBMC

Atomic model building examples in cryo-EM



Rotavirus VP6 cryo-EM structure; 2.6 Å resolution; side-chains are defined.

optimize exposure dose to select movie frames

Grant, T., Grigorieff, N., *eLife*, 2015.

IGBMC

Atomic model building:

combining cryo-EM and X-ray crystallography refinement procedures;

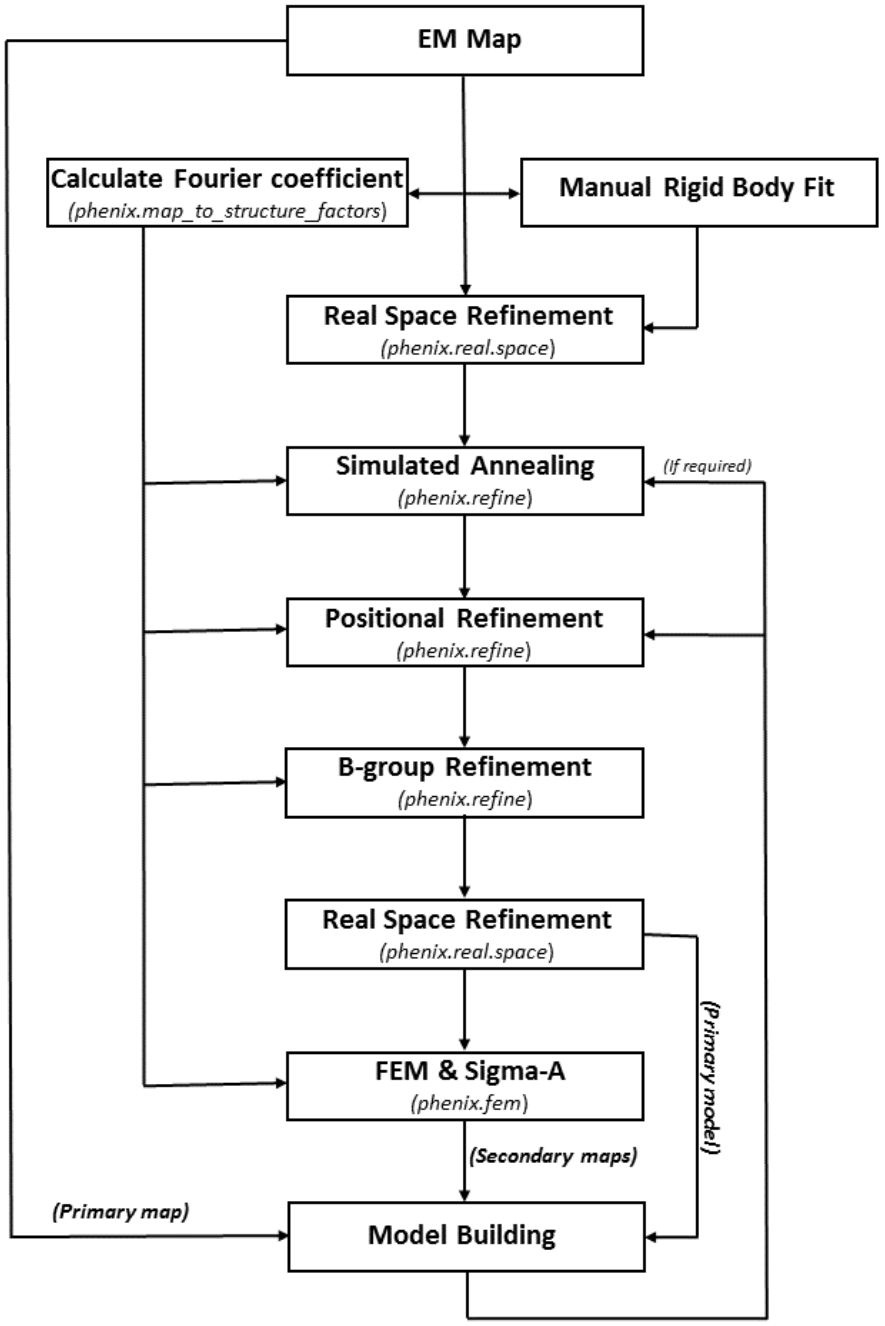
Validation of atomic models derived from cryo-EM maps:

d_{99} to estimate resolution,

phenix.autosharpen,

phenix.mtriage

Afonine *et al., Acta Cryst. D, 2018.*



Natchiar *et al., Nature Exchange Protocol, 2017.*

IGBMC

Atomic model building & refinement into cryo-EM maps:

Consider:

- in contrast to crystallography, the map is not modified during the atomic model refinement, because:

- the cryo-EM map is similar to a experimentally phased crystallography map (e.g. SAD, native sulfur phasing etc.)

- in other words, the quality of the cryo-EM map does not depend on the quality of the atomic model

- if during atomic model refinement one notices gaps or connectivity issues in the map, then this may indicate that the cryo-EM structure refinement (i.e. the 3D reconstruction refinement) is not good/finshed yet (e.g. inappropriate filtering / postprocessing. i.e. too much removal of low frequencies or overweighting of high frequencies; density distortions coming from preferential views; CTF estimation incorrect at high frequencies, weak CTF/thick ice, etc.)

Quality of the geometry of the atomic model

Data Collection	
Particles	139,234
Sampling (Å/pixel)	0.85
Defocus range (μm)	-0.4 to -2.5
Atomic model composition	
Non-hydrogen atoms	219,591
Amino acids	11,729
Nucleotides	5,863
Number of ligand atoms	75
Zn^{2+} / Mg^{2+} / H_2O	8 / 400 / 60
Refinement	
Average resolution (Å)	2.9 (60S) / 3.0 (40S body) / 3.1 (40S head)
Average B-factor ($Å^2$)	64.4
RMS Deviation	
Bond lengths (Å)	0.009
Bond angles (°)	1.057
Ramachandran statistics	
Favoured (%)	90.38
Allowed (%)	9.36
Outliers (%)	0.26

Natchiar *et al.*, *Nature* 2017.

IGBMC