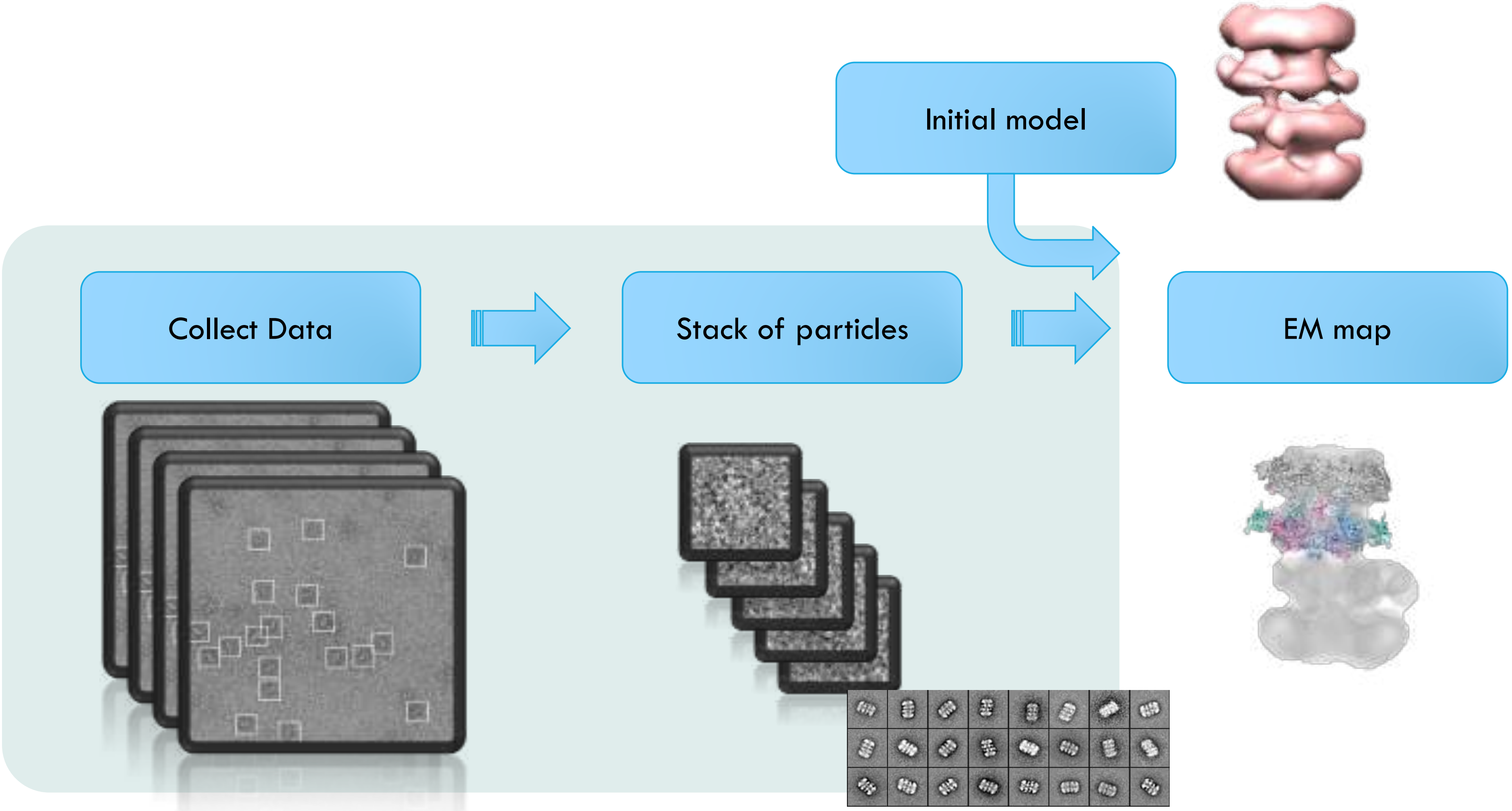
CRYOEM 001 :
SPA PROCESSING WORKFLOW, VALIDATION AND OTHER MATTERS

February 15, 2023

New York Structural
Biology Center

SIMONS ELECTRON
MICROSCOPY CENTER

SINGLE PARTICLE ANALYSIS



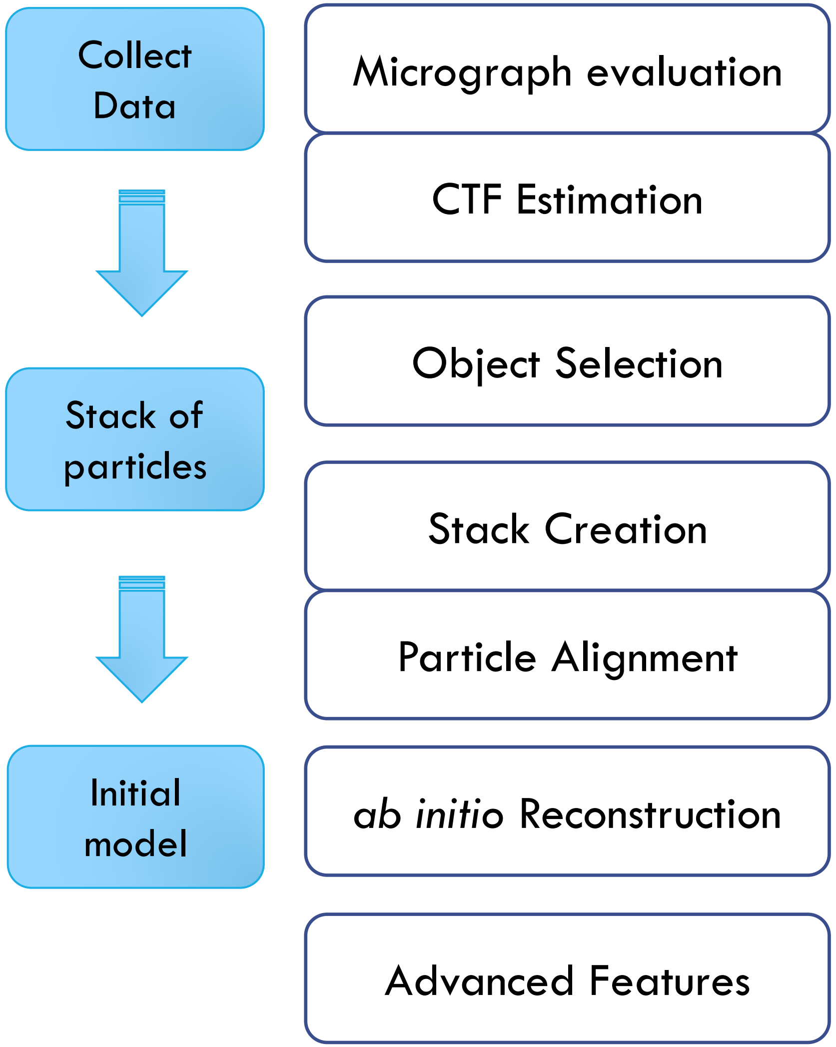
Initial model

Collect Data

Stack of particles

EM map

SINGLE PARTICLE ANALYSIS WORKFLOW



Collect Data

Stack of particles

Initial model

Micrograph evaluation

CTF Estimation

Object Selection

Stack Creation

Particle Alignment

ab initio Reconstruction

Advanced Features

EXAMPLE: SINGLE-PARTICLE WORKFLOW

During EM session

Leginon session

Setting up workflow

Data acquisition

Workflow Optimization *if needed*

Leginon

Appion session

Frame alignment

CTF estimation

Particle picking

Micrograph/ Particle curating

Appion

After EM session

SEMC computing

Micrograph/ Particle sorting

Initial 2D classification

Initial model generation

3D refinement

csLIVE

Home institution computing/cluster

2D classification

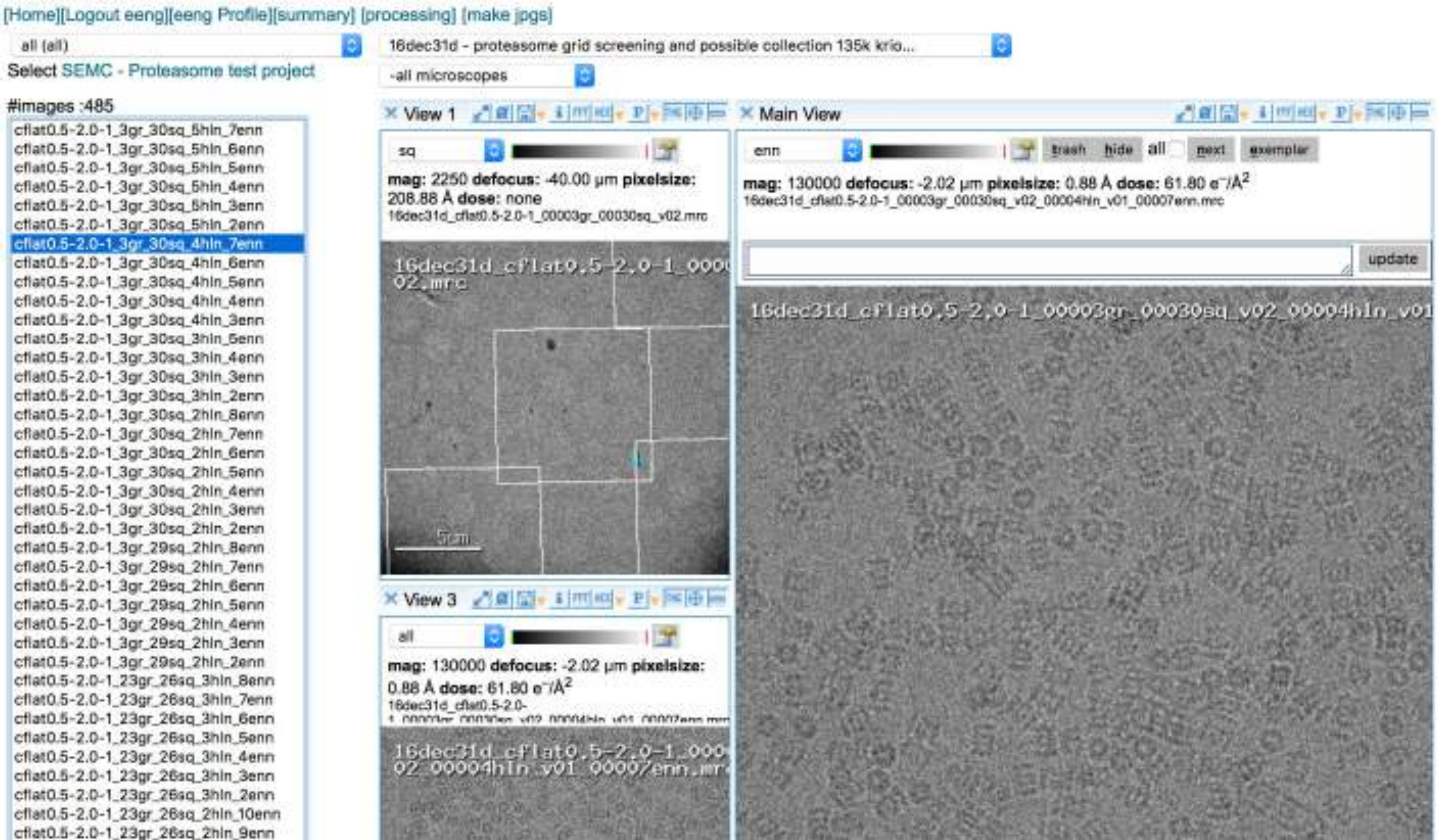
3D classification

3D refinement

Model building

cryoSPARC/RELION/cisTEM

MICROGRAPH EVALUATION



[Home][Logout eeng][eeng Profile][summary] [processing] [make jpgs]

all (all)

Select SEMC - Proteasome test project

16dec31d - proteasome grid screening and possible collection 135k krio...

-all microscopes

#images :485

cflat0.5-2.0-1_3gr_30sq_5hln_7enn
cflat0.5-2.0-1_3gr_30sq_5hln_6enn
cflat0.5-2.0-1_3gr_30sq_5hln_5enn
cflat0.5-2.0-1_3gr_30sq_5hln_4enn
cflat0.5-2.0-1_3gr_30sq_5hln_3enn
cflat0.5-2.0-1_3gr_30sq_5hln_2enn
cflat0.5-2.0-1_3gr_30sq_4hln_7enn
cflat0.5-2.0-1_3gr_30sq_4hln_6enn
cflat0.5-2.0-1_3gr_30sq_4hln_5enn
cflat0.5-2.0-1_3gr_30sq_4hln_4enn
cflat0.5-2.0-1_3gr_30sq_4hln_3enn
cflat0.5-2.0-1_3gr_30sq_3hln_5enn
cflat0.5-2.0-1_3gr_30sq_3hln_4enn
cflat0.5-2.0-1_3gr_30sq_3hln_3enn
cflat0.5-2.0-1_3gr_30sq_3hln_2enn
cflat0.5-2.0-1_3gr_30sq_2hln_8enn
cflat0.5-2.0-1_3gr_30sq_2hln_7enn
cflat0.5-2.0-1_3gr_30sq_2hln_6enn
cflat0.5-2.0-1_3gr_30sq_2hln_5enn
cflat0.5-2.0-1_3gr_30sq_2hln_4enn
cflat0.5-2.0-1_3gr_30sq_2hln_3enn
cflat0.5-2.0-1_3gr_30sq_2hln_2enn
cflat0.5-2.0-1_3gr_29sq_2hln_8enn
cflat0.5-2.0-1_3gr_29sq_2hln_7enn
cflat0.5-2.0-1_3gr_29sq_2hln_6enn
cflat0.5-2.0-1_3gr_29sq_2hln_5enn
cflat0.5-2.0-1_3gr_29sq_2hln_4enn
cflat0.5-2.0-1_3gr_29sq_2hln_3enn
cflat0.5-2.0-1_3gr_29sq_2hln_2enn
cflat0.5-2.0-1_23gr_26sq_3hln_8enn
cflat0.5-2.0-1_23gr_26sq_3hln_7enn
cflat0.5-2.0-1_23gr_26sq_3hln_6enn
cflat0.5-2.0-1_23gr_26sq_3hln_5enn
cflat0.5-2.0-1_23gr_26sq_3hln_4enn
cflat0.5-2.0-1_23gr_26sq_3hln_3enn
cflat0.5-2.0-1_23gr_26sq_3hln_2enn
cflat0.5-2.0-1_23gr_26sq_2hln_10enn
cflat0.5-2.0-1_23gr_26sq_2hln_9enn

View 1

sq

mag: 2250 **defocus:** -40.00 μm **pixelsize:** 208.88 Å **dose:** none
16dec31d_cflat0.5-2.0-1_00003gr_00030sq_v02.mrc

View 3

all

mag: 130000 **defocus:** -2.02 μm **pixelsize:** 0.88 Å **dose:** 61.80 e⁻/Å²
16dec31d_cflat0.5-2.0-

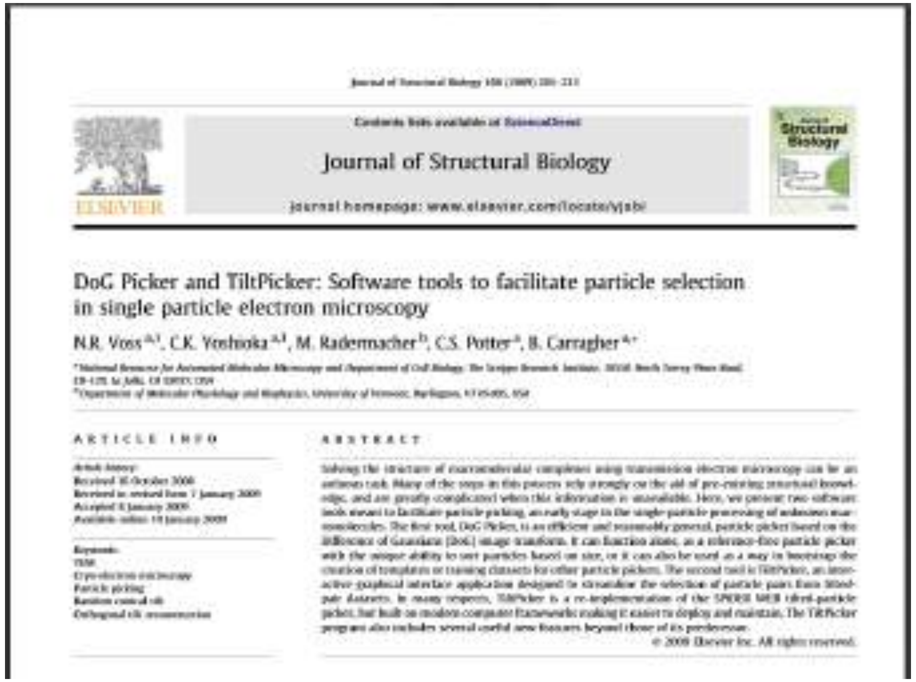
Main View

enn

mag: 130000 **defocus:** -2.02 μm **pixelsize:** 0.88 Å **dose:** 61.80 e⁻/Å²
16dec31d_cflat0.5-2.0-1_00003gr_00030sq_v02_00004hln_v01_00007enn.mrc

PARTICLE PICKING

Difference of Gaussian (DoG) Picker



Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy

N.R. Voss, C.K. Yoshioka, M. Radermacher, C.S. Potter, B. Carragher

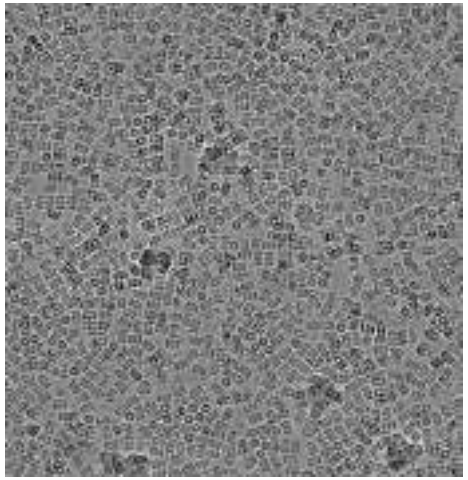
ARTICLE INFO

ABSTRACT

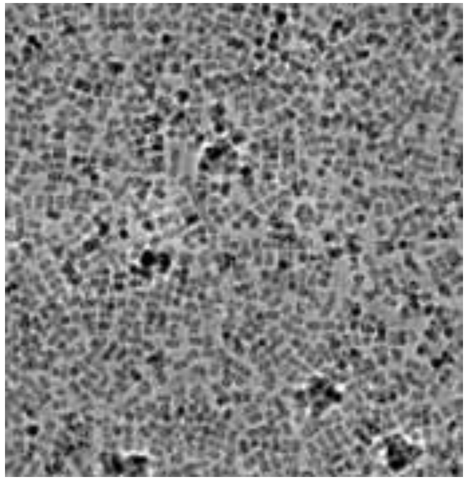
PARTICLE PICKING

Difference of Gaussian (DoG) Picker

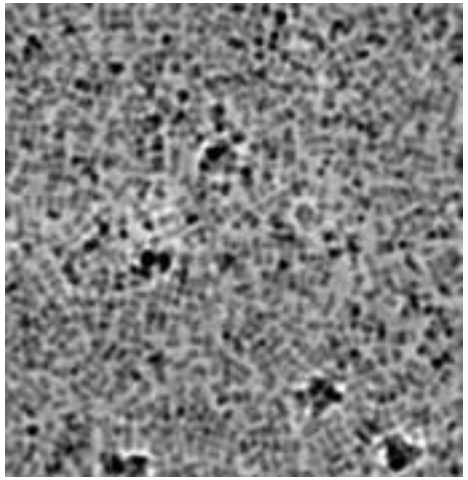
Low Pass Filtering



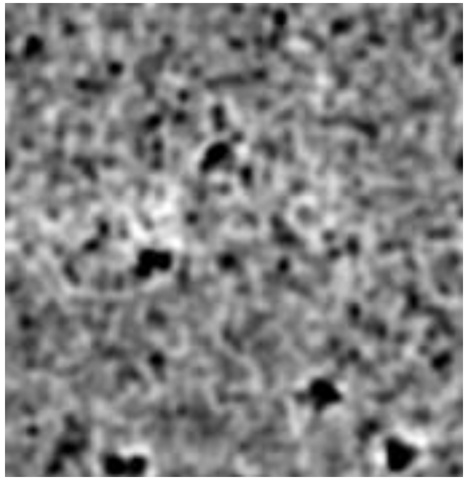
0 Å



20 Å



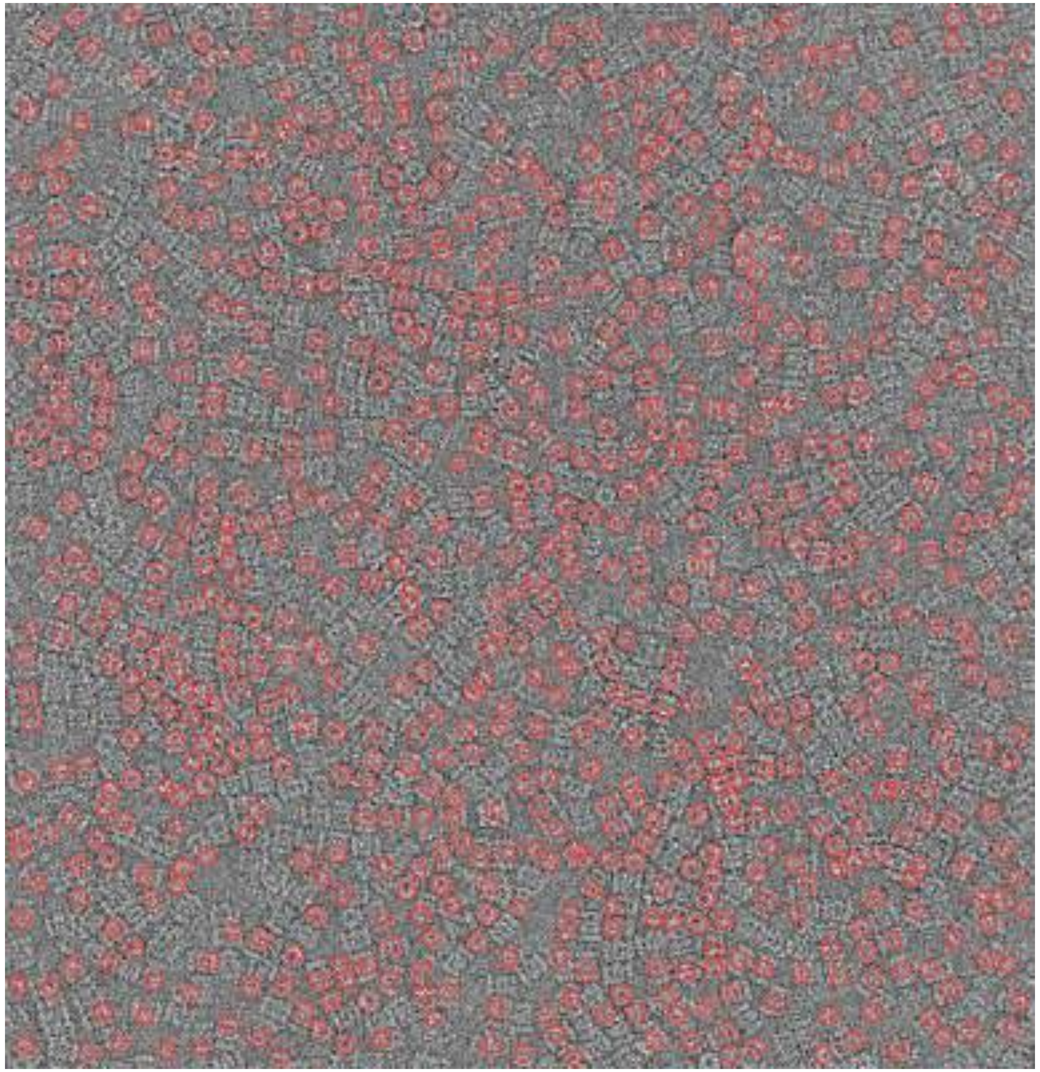
40 Å



60 Å

PARTICLE PICKING

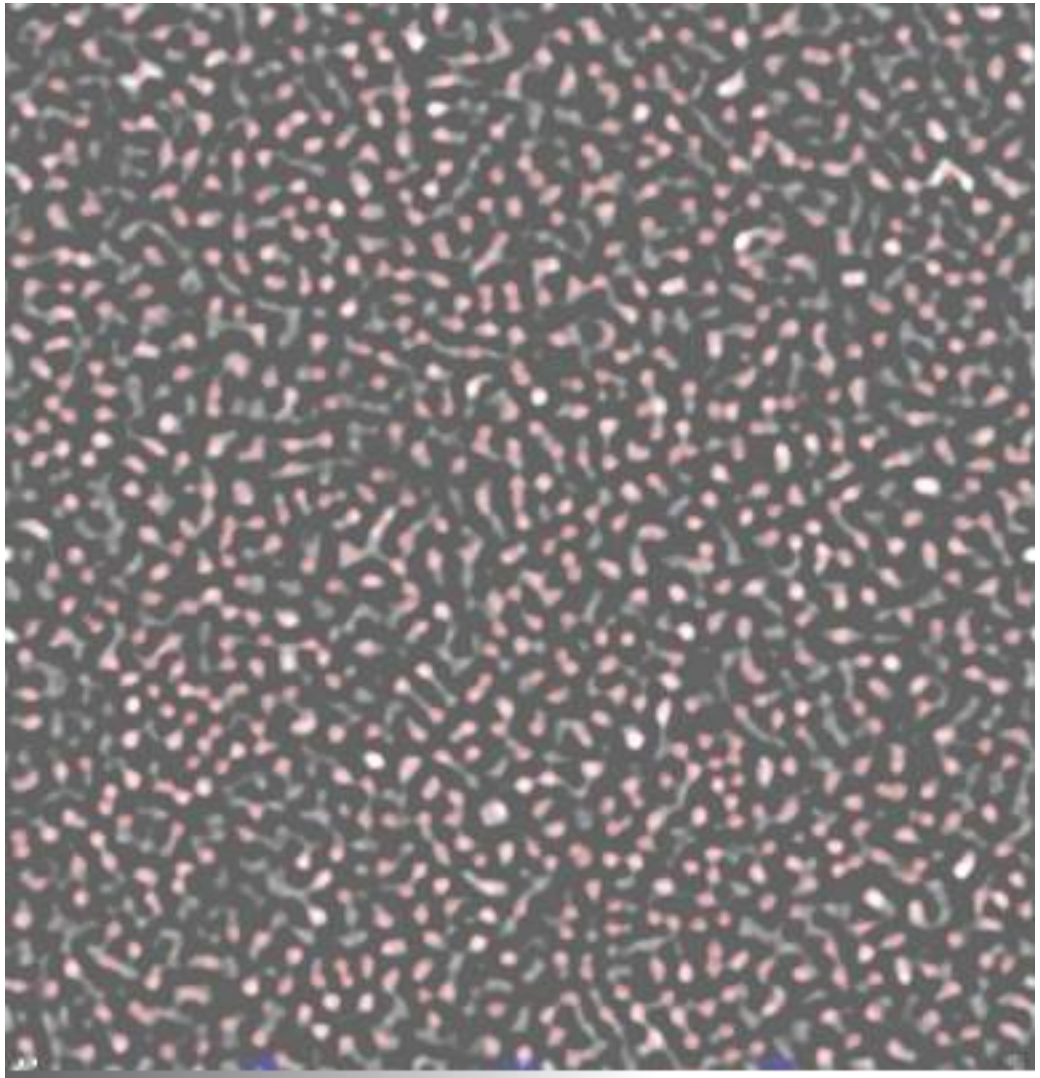
Difference of Gaussian (DoG) Picker



PARTICLE PICKING

Difference of Gaussian (DoG) Picker

DoG Map



ONE MORE NOTE ABOUT DEFOCUS

Highlights

- The spatial-coherence envelope function has almost no effect under the conditions used in cryo-EM.
- Even for defocus values as high as 8 µm, this envelope remains negligible compared to other factors.
- On the other hand, delocalization of high-resolution information can become an important factor.
- For many situations, it may thus be beneficial to use defocus values as high as ~5 µm or more.

ELSEVIER

Ultramicroscopy

Volume 222, March 2021, 113213

Defocus-dependent Thon-ring fading

Robert M. Glaeser [a], Wim J.H. Hagen [b], Bong-Gyoon Han [c], Richard Henderson [d], Greg McMullan [e], Christopher J. Russo [f]

https://doi.org/10.1016/j.ultramic.2021.113213

Under a Creative Commons license

ONE MORE NOTE ABOUT DEFOCUS

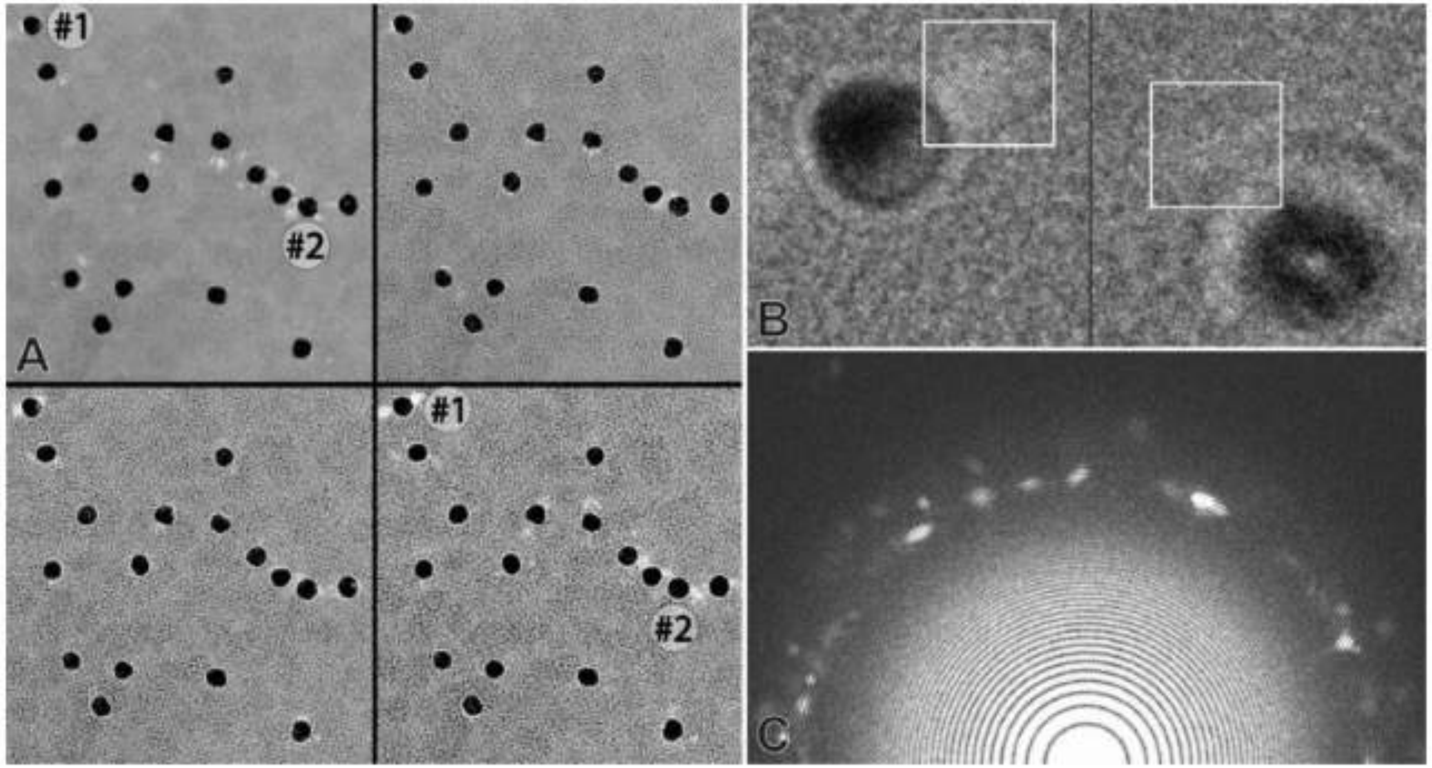
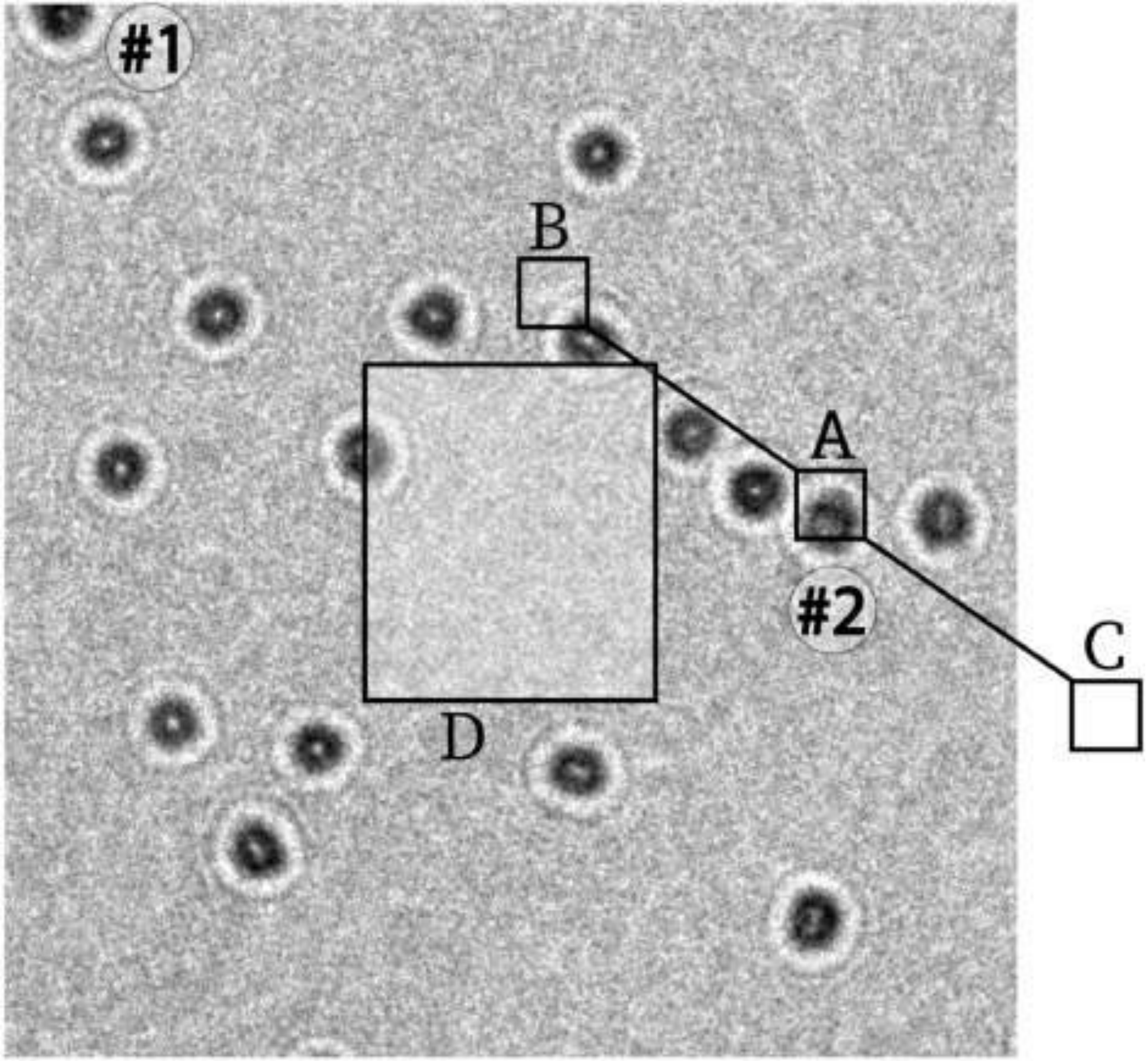


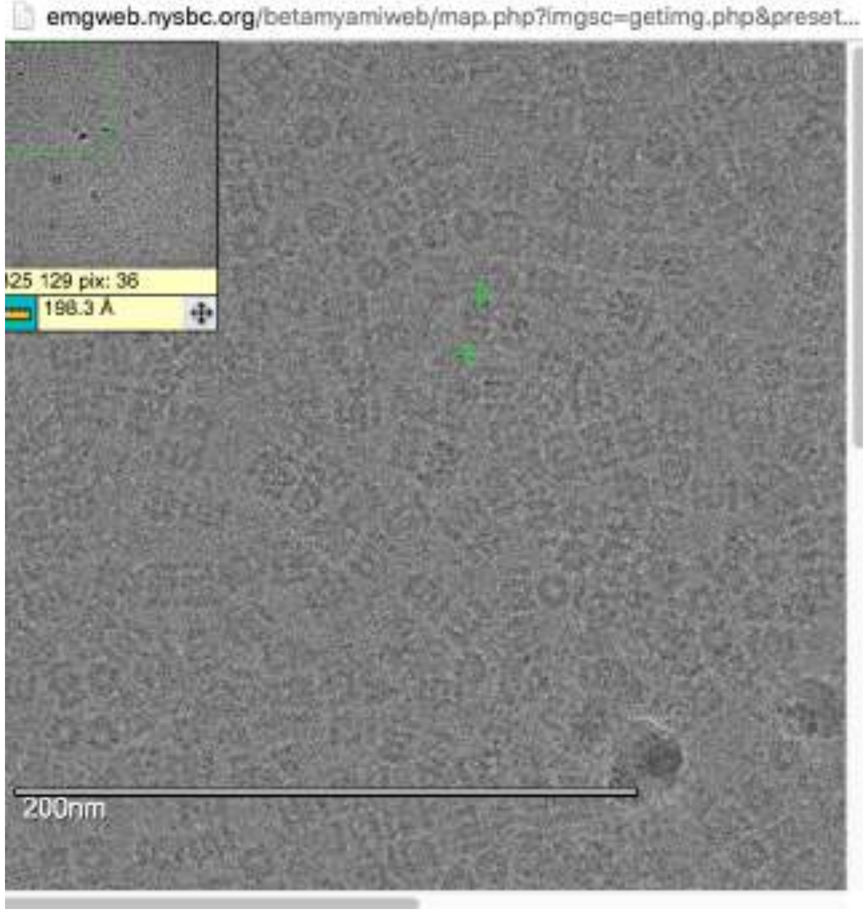
Fig. 1. Examples showing the focus-dependent delocalization of information carried by interference between scattered electrons and the unscattered, reference wave.

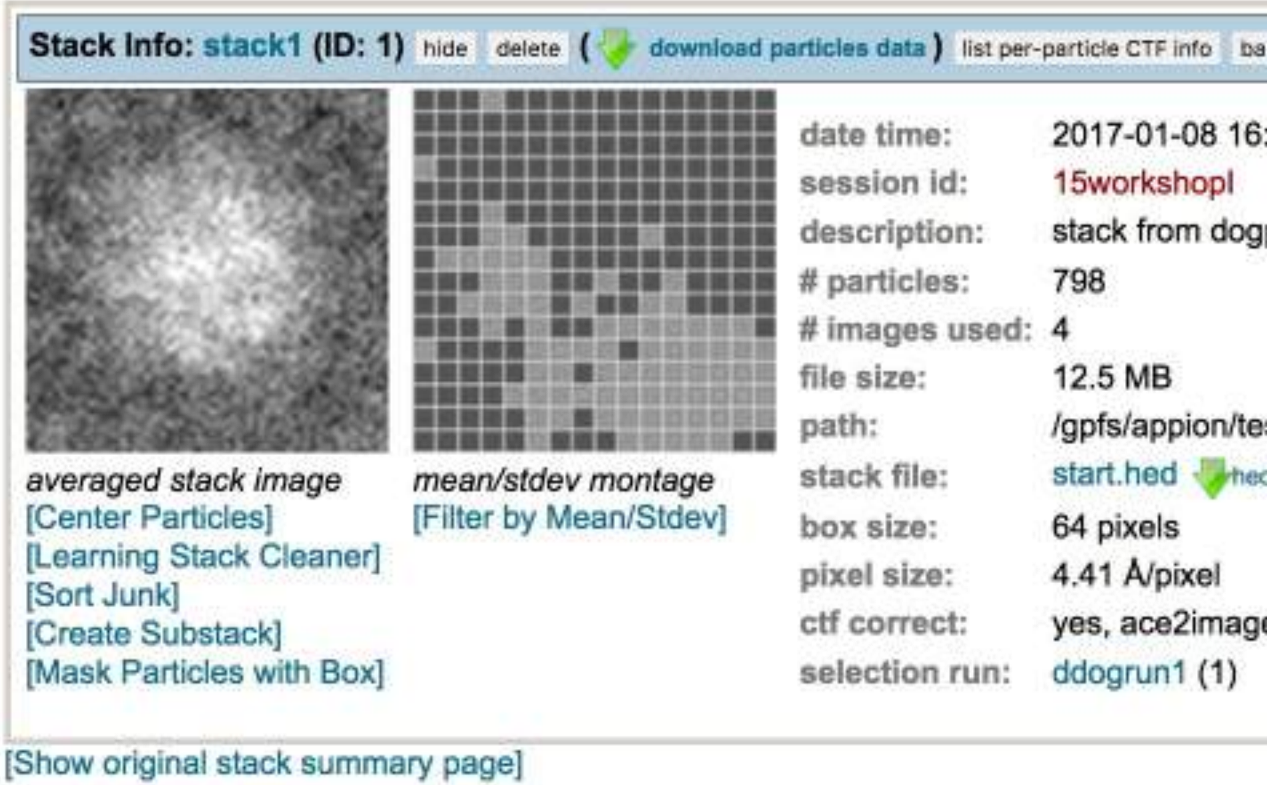
ONE MORE NOTE ABOUT DEFOCUS

Fig. 3. Image "gold-20", recorded at 8 μm defocus, showing how delocalization of one of the half-wavelets corresponding to the Au (111) lattice planes is moved outside the field of view of the camera. Gold nanoparticle #2 is labelled, with the diffracting crystallite labelled A, and the positions of 2.35 Å lattice fringe patches labelled B and C. Area C is outside the field of view, so this half-wavelet is not observed and it does not contribute to the power spectrum. Area B is shown at higher magnification in the right panel of Fig. 1B. The area outlined in the center would be the only region able to produce Thon-ring modulations at the 2.35 Å gold-lattice spacing in all directions, but, within this field of view, there is only about half of one of the 17 gold particles that is located within the outlined area. As a result, the power spectrum of this image shows no Thon-ring contribution to the power spectrum of such nanoparticles.



ALIGNMENT





Stack Info: stack1 (ID: 1)

averaged stack image
[Center Particles]
[Learning Stack Cleaner]
[Sort Junk]
[Create Substack]
[Mask Particles with Box]

mean/stdev montage
[Filter by Mean/Stdev]

date time: 2017-01-08 16
session id: 15workshopl
description: stack from dog
particles: 798
images used: 4
file size: 12.5 MB
path: /gpfs/appion/te
stack file: start.hed
box size: 64 pixels
pixel size: 4.41 Å/pixel
ctf correct: yes, ace2image
selection run: ddogrun1 (1)

[Show original stack summary page]

ALIGNMENT

stack: /gpfs/appion/test/15workshopI/align/maxlike1/part17jan08r24_average.hed

#images: 8

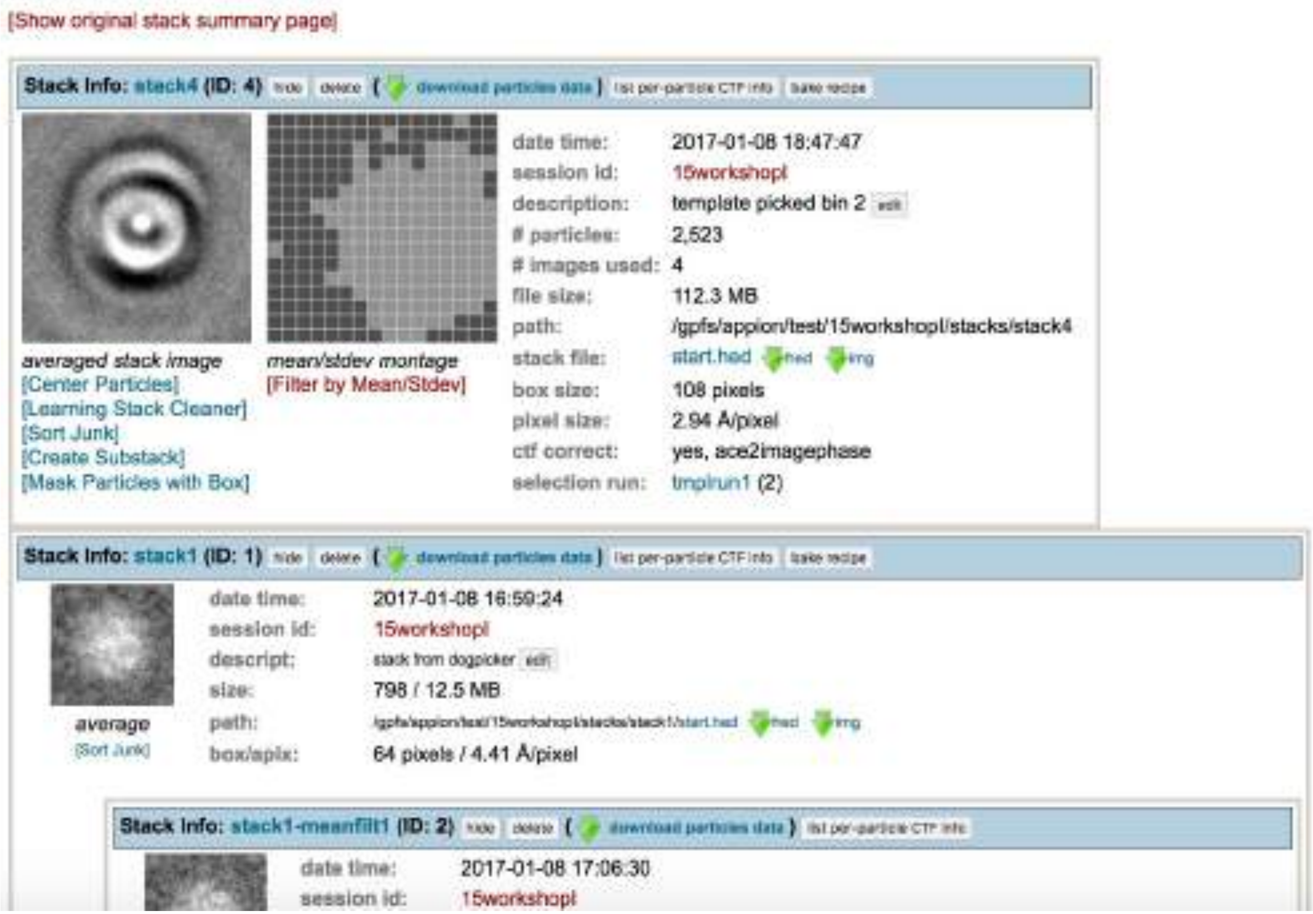
from: 0 to: 7 binning: 1 quality: jpeg 50 info: ☑ scale bar: ☐

Selection mode: select

Selected images: 0,5

Excluded images: 1,2,3,4,6,7

ALIGNMENT



Stack Info: stack4 (ID: 4)

averaged stack image
[Center Particles]
[Learning Stack Cleaner]
[Sort Junk]
[Create Substack]
[Mask Particles with Box]

mean/stdev montage
[Filter by Mean/Stdev]

date time:	2017-01-08 18:47:47
session id:	15workshopl
description:	template picked bin 2
# particles:	2,523
# images used:	4
file size:	112.3 MB
path:	/gpfs/appion/test/15workshopl/stacks/stack4
stack file:	start.hed
box size:	108 pixels
pixel size:	2.94 A/pixel
ctf correct:	yes, ace2imagephase
selection run:	tmplrun1 (2)

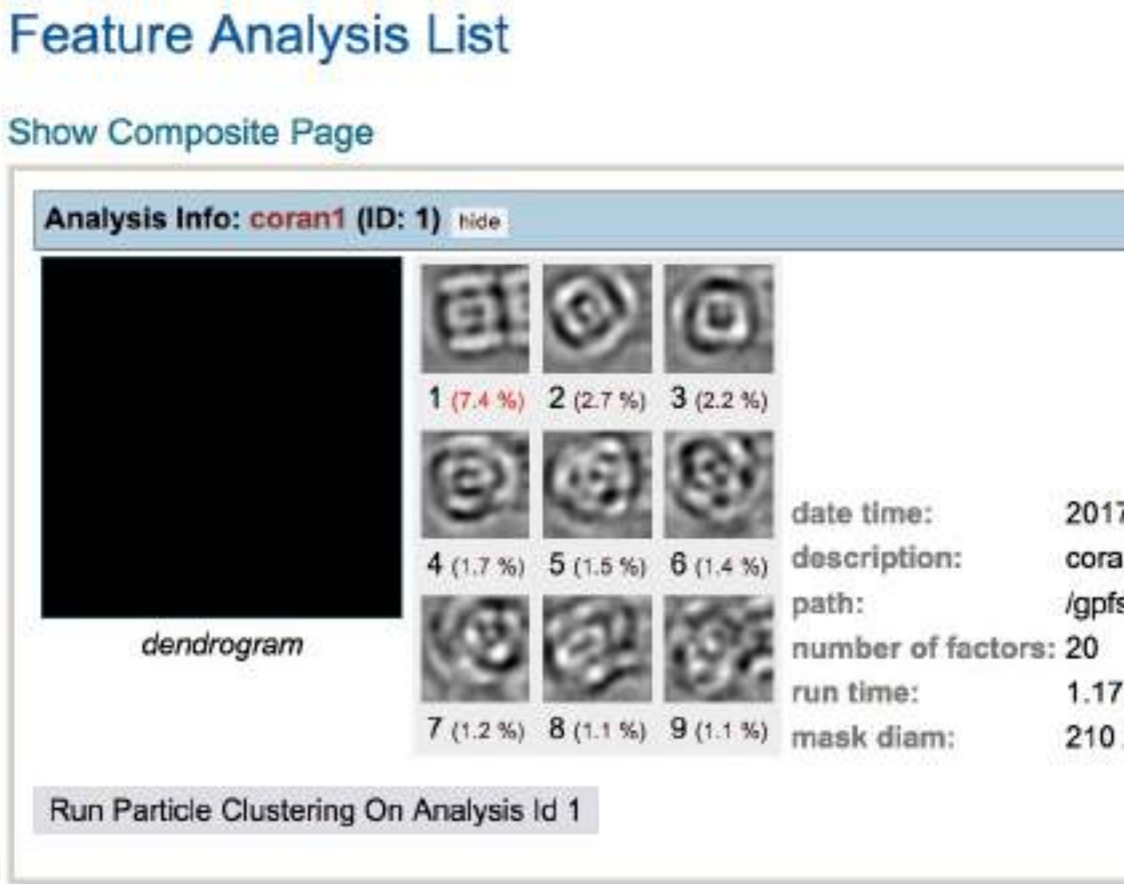
Stack Info: stack1 (ID: 1)

average
[Sort Junk]

date time:	2017-01-08 16:59:24
session id:	15workshopl
descript:	stack from dogpicker
size:	798 / 12.5 MB
path:	/gpfs/appion/test/15workshopl/stacks/stack1/start.hed
box/apix:	64 pixels / 4.41 Å/pixel

Stack Info: stack1-meanfilt1 (ID: 2)

date time:	2017-01-08 17:06:30
session id:	15workshopl



Feature Analysis List

Show Composite Page

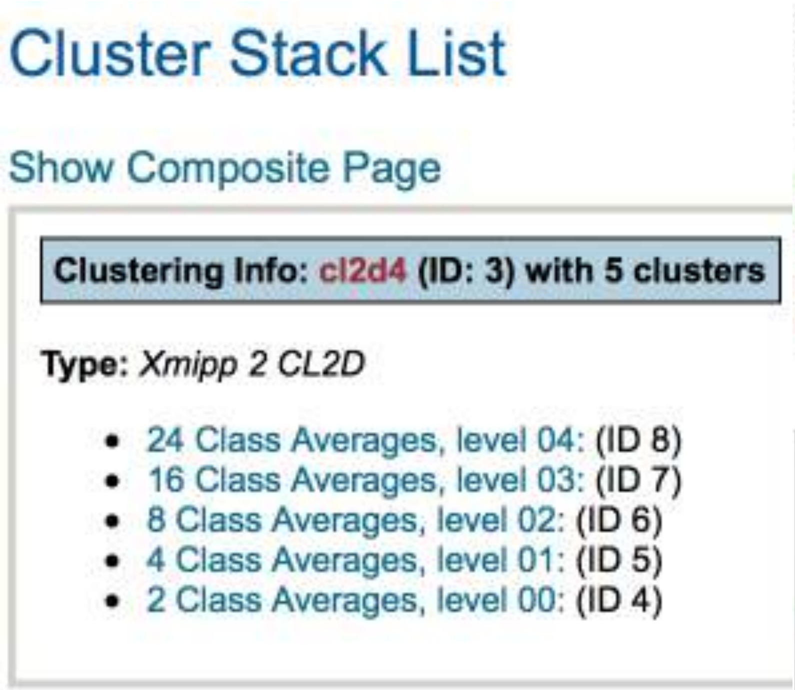
Analysis Info: coran1 (ID: 1)

dendrogram

1 (7.4 %) 2 (2.7 %) 3 (2.2 %)
4 (1.7 %) 5 (1.5 %) 6 (1.4 %)
7 (1.2 %) 8 (1.1 %) 9 (1.1 %)

date time:	2017
description:	cora
path:	/gpfs
number of factors:	20
run time:	1.17
mask diam:	210

CLUSTERING



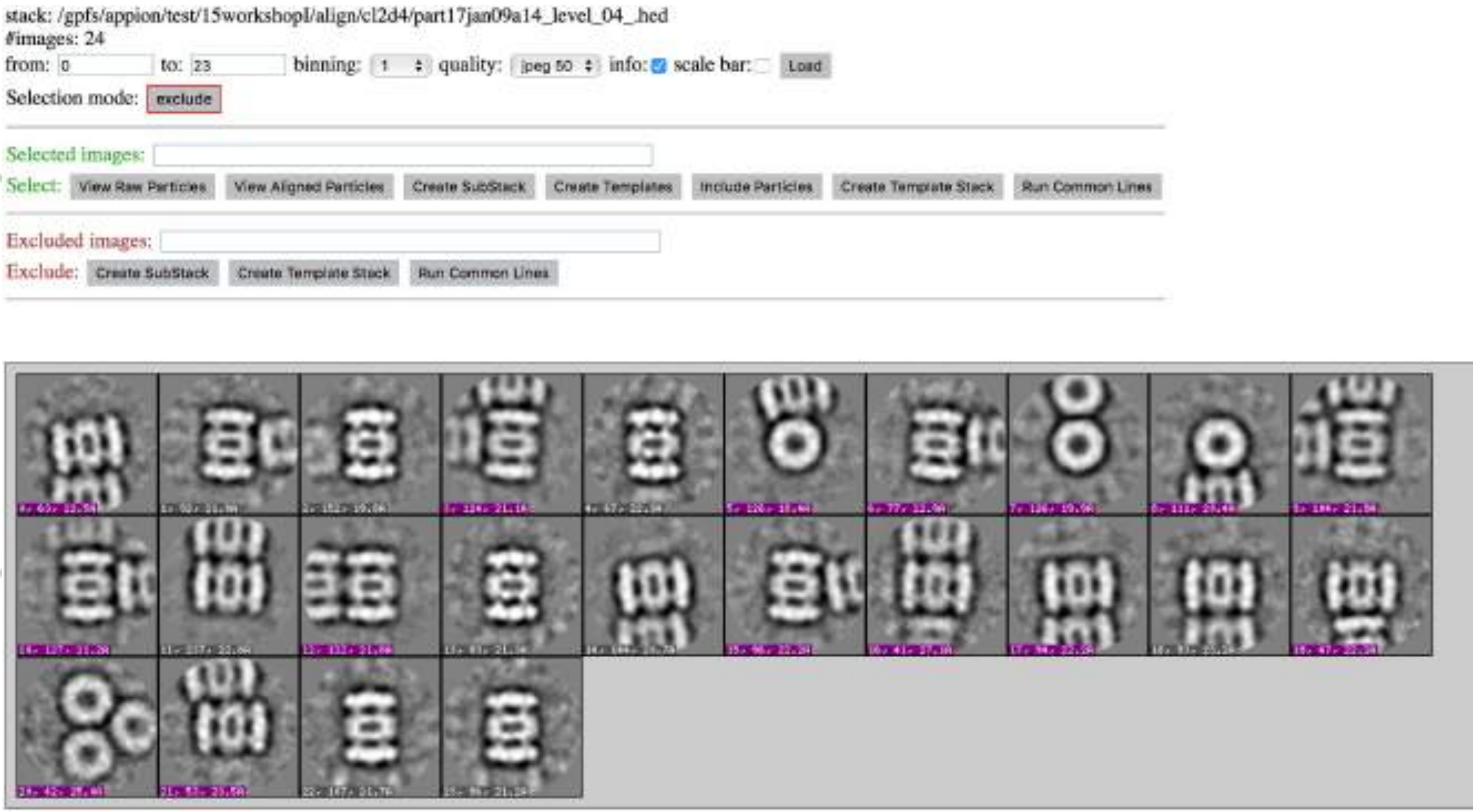
Cluster Stack List

Show Composite Page

Clustering Info: cl2d4 (ID: 3) with 5 clusters

Type: *Xmipp 2 CL2D*

- 24 Class Averages, level 04: (ID 8)
- 16 Class Averages, level 03: (ID 7)
- 8 Class Averages, level 02: (ID 6)
- 4 Class Averages, level 01: (ID 5)
- 2 Class Averages, level 00: (ID 4)



stack: /gpfs/appion/test/15workshopl/align/cl2d4/part17jan09a14_level_04_.hed
#images: 24
from: 0 to: 23 binning: 1 quality: jpeg 50 info: ☑ scale bar: ☐
Selection mode: exclude

Selected images:

Excluded images:

VALIDATION

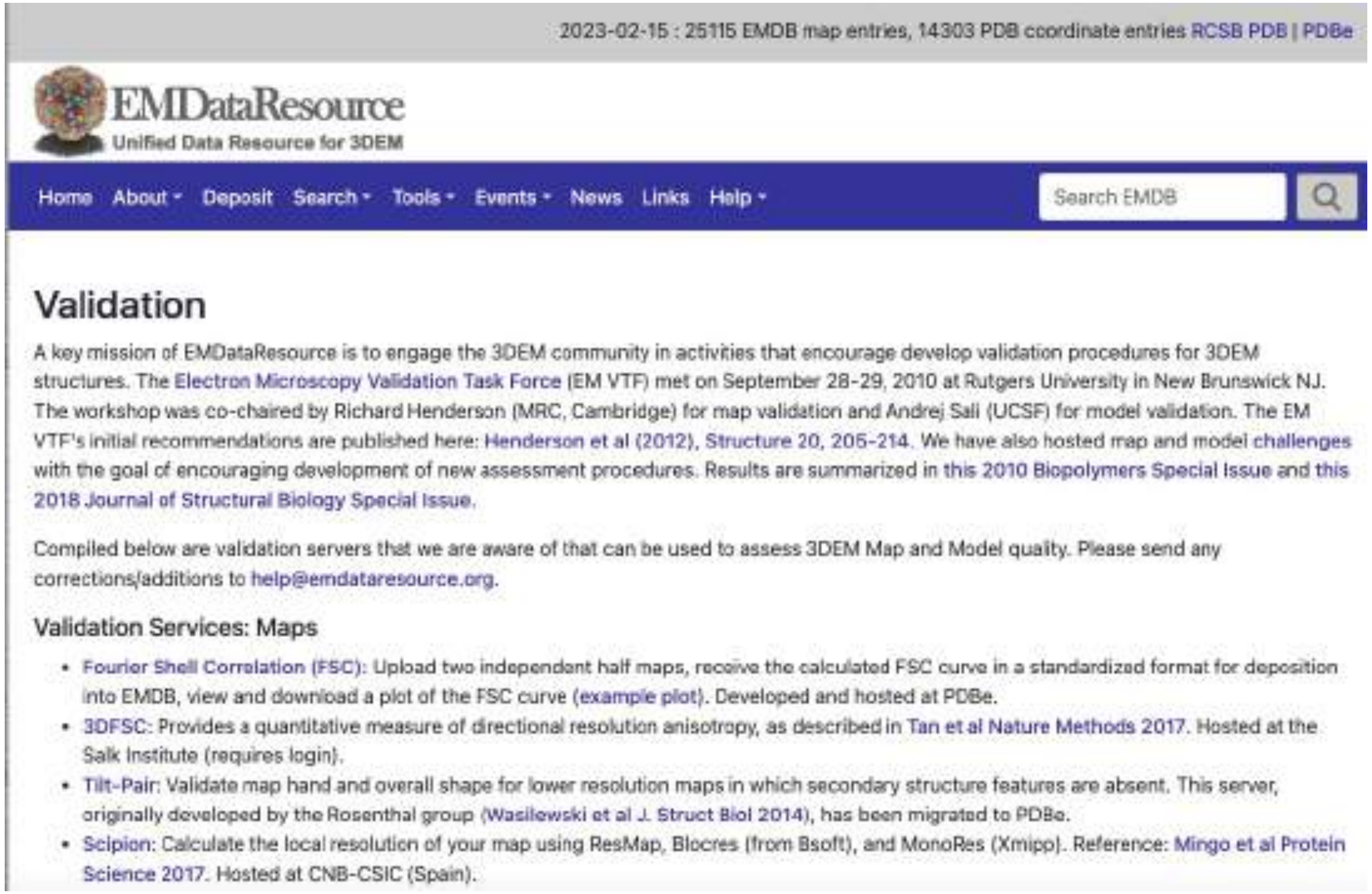
3D map generation –

Resolution estimation

- FSC between 2 half-maps
– Independent refinement of the two halves (gold standard)
– Independent high resolution shells

- Local resolution
– Resmap
– Bsoft LocRes
– Masking individual domains

- Variance estimation
– Resmap



Validation

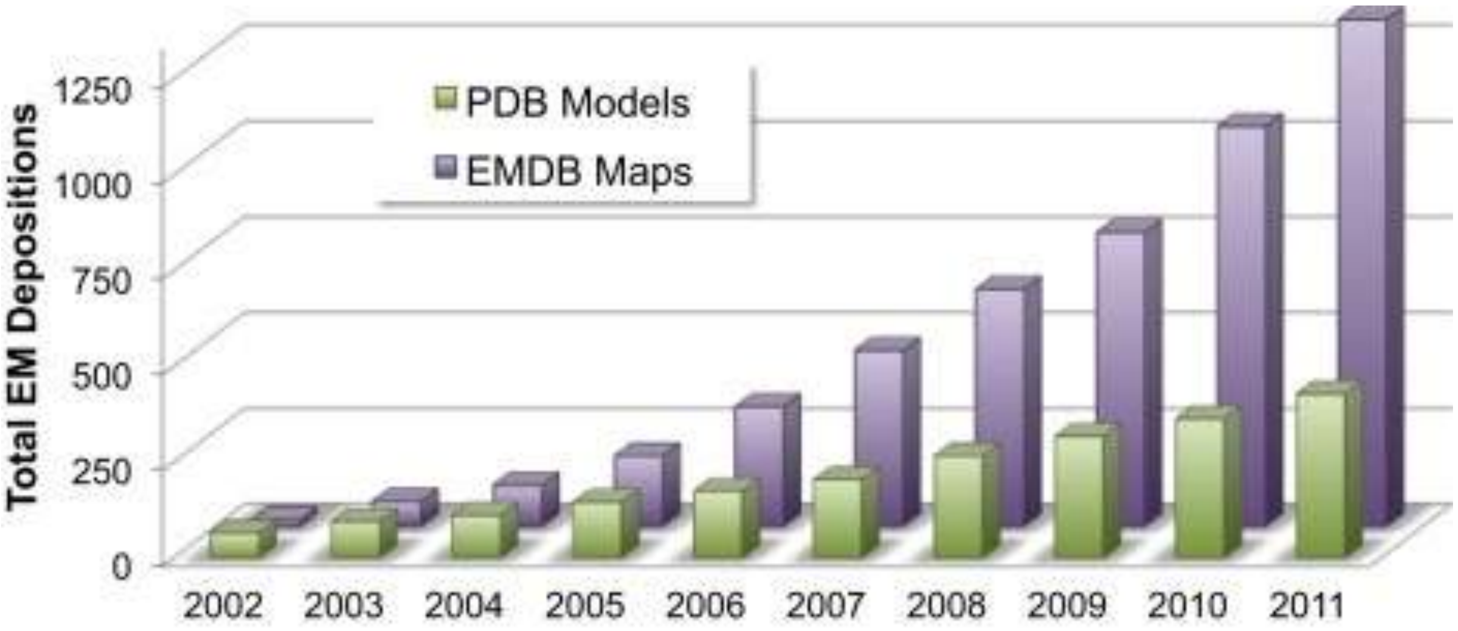
A key mission of EMDataResource is to engage the 3DEM community in activities that encourage develop validation procedures for 3DEM structures. The Electron Microscopy Validation Task Force (EM VTF) met on September 28-29, 2010 at Rutgers University in New Brunswick NJ. The workshop was co-chaired by Richard Henderson (MRC, Cambridge) for map validation and Andrej Sali (UCSF) for model validation. The EM VTF's initial recommendations are published here: Henderson et al (2012), Structure 20, 205-214. We have also hosted map and model challenges with the goal of encouraging development of new assessment procedures. Results are summarized in this 2010 Biopolymers Special Issue and this 2018 Journal of Structural Biology Special Issue.

Compiled below are validation servers that we are aware of that can be used to assess 3DEM Map and Model quality. Please send any corrections/additions to help@emdataresource.org.

Validation Services: Maps

- Fourier Shell Correlation (FSC): Upload two independent half maps, receive the calculated FSC curve in a standardized format for deposition into EMDB, view and download a plot of the FSC curve (example plot). Developed and hosted at PDBe.
- 3DFSC: Provides a quantitative measure of directional resolution anisotropy, as described in Tan et al Nature Methods 2017. Hosted at the Salk Institute (requires login).
- Tilt-Pair: Validate map hand and overall shape for lower resolution maps in which secondary structure features are absent. This server, originally developed by the Rosenthal group (Wasilewski et al J. Struct Biol 2014), has been migrated to PDBe.
- Scipion: Calculate the local resolution of your map using ResMap, Blocres (from Bsoft), and MonoRes (Xmipp). Reference: Mingo et al Protein Science 2017. Hosted at CNB-CSIC (Spain).

WHY IS IT SO CHALLENGING?



EINSTEIN FROM NOISE

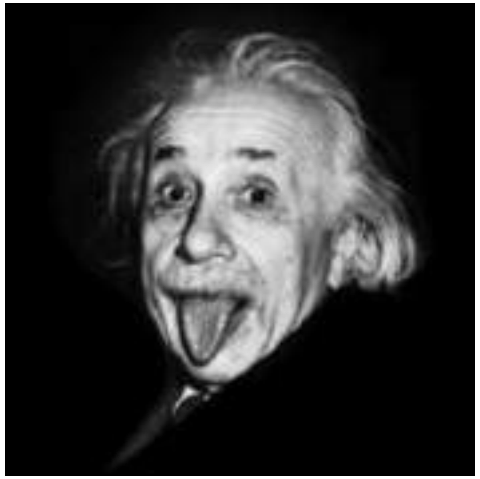
> J Struct Biol. 2009 Apr;166(1):67-78. doi: 10.1016/j.jsb.2008.12.008. Epub 2008 Dec 30.

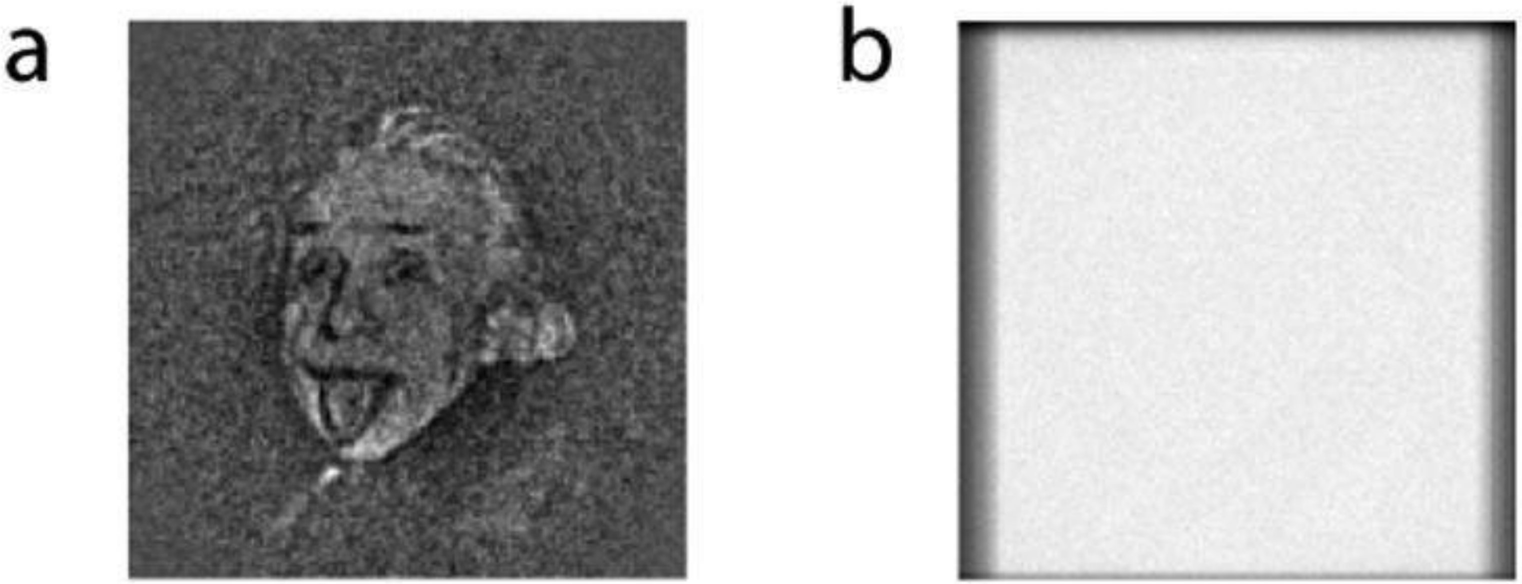
A method for the alignment of heterogeneous macromolecules from electron microscopy

Maxim Shatsky[1], Richard J Hall, Steven E Brenner, Robert M Glaeser

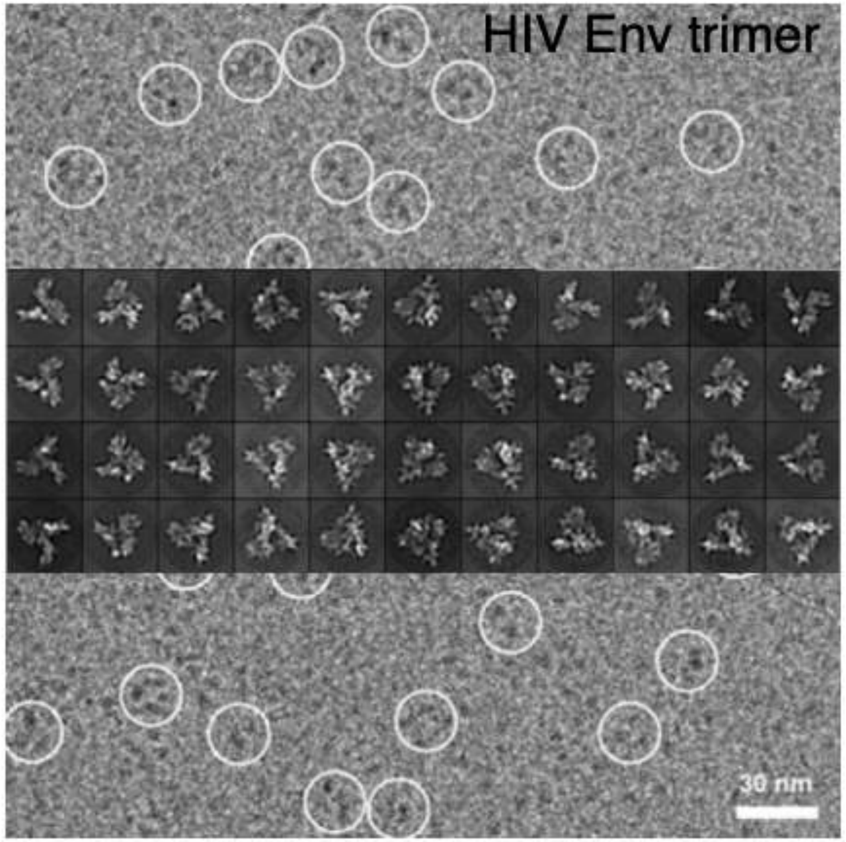
Affiliations + expand

PMID: 19166941 PMCID: PMC2740748 DOI: 10.1016/j.jsb.2008.12.008

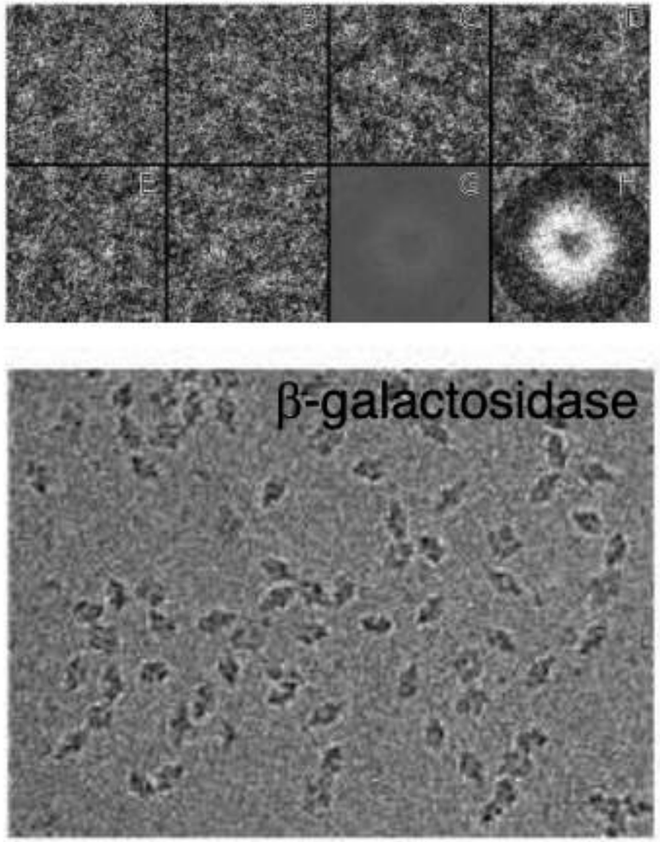




EINSTEIN FROM NOISE



Mao *et al.* (2013)
PNAS 110: 12438-12443



Henderson (2013)
PNAS 110: 18037-18041

-adapted from Tom Walz

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

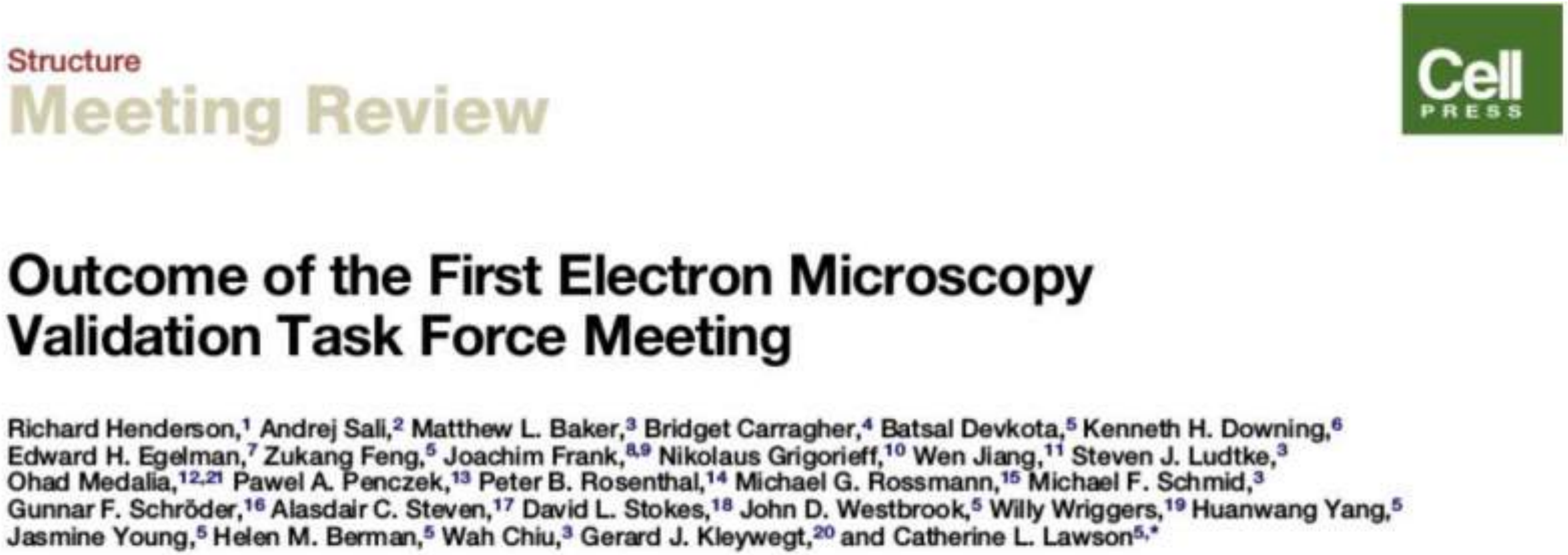
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



Structure

Meeting Review

Cell PRESS

Outcome of the First Electron Microscopy Validation Task Force Meeting

Richard Henderson,[1] Andrej Sali,[2] Matthew L. Baker,[3] Bridget Carragher,[4] Batsal Devkota,[5] Kenneth H. Downing,[6] Edward H. Egelman,[7] Zukang Feng,[5] Joachim Frank,[8,9] Nikolaus Grigorieff,[10] Wen Jiang,[11] Steven J. Ludtke,[3] Ohad Medalia,[12,21] Pawel A. Penczek,[13] Peter B. Rosenthal,[14] Michael G. Rossmann,[15] Michael F. Schmid,[3] Gunnar F. Schröder,[16] Alasdair C. Steven,[17] David L. Stokes,[18] John D. Westbrook,[5] Willy Wriggers,[19] Huanwang Yang,[5] Jasmine Young,[5] Helen M. Berman,[5] Wah Chiu,[3] Gerard J. Kleywegt,[20] and Catherine L. Lawson[5,*]

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

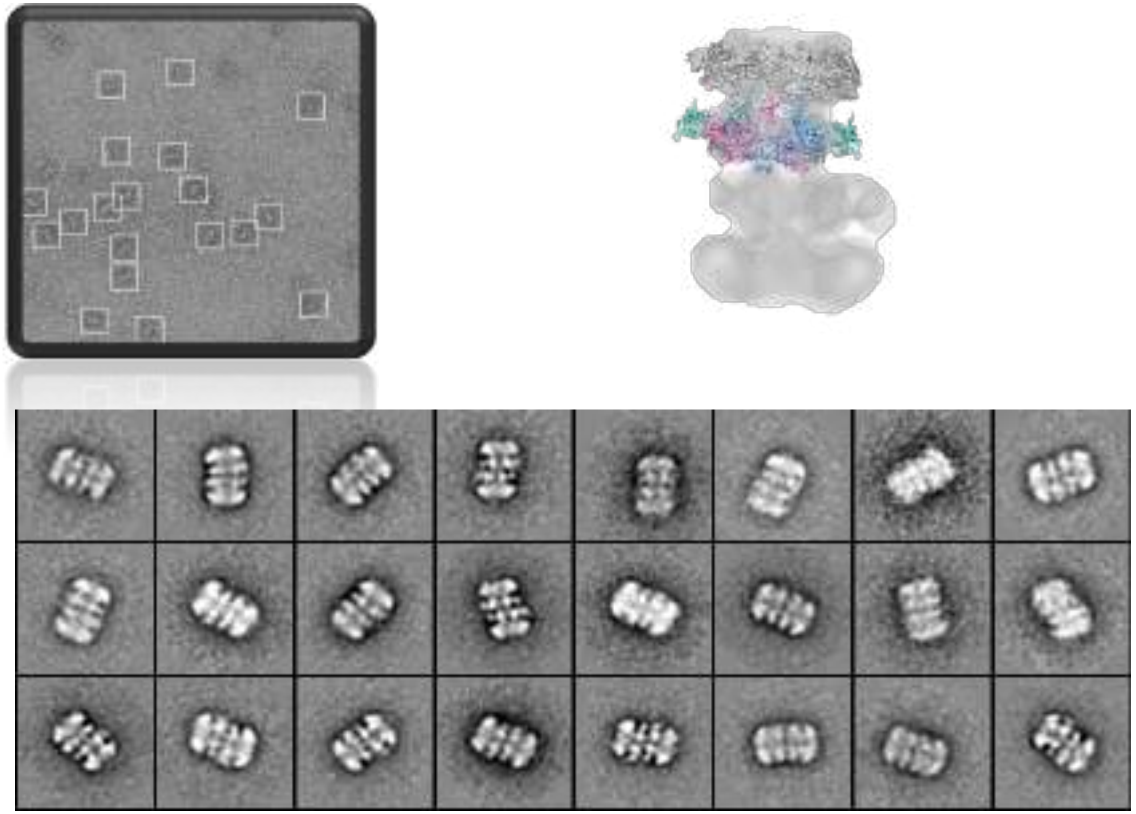
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



VALIDATION TIPS

Compare reference free average with projections

Tilt-pair analysis

“Gold standard” FSC

Randomize phases

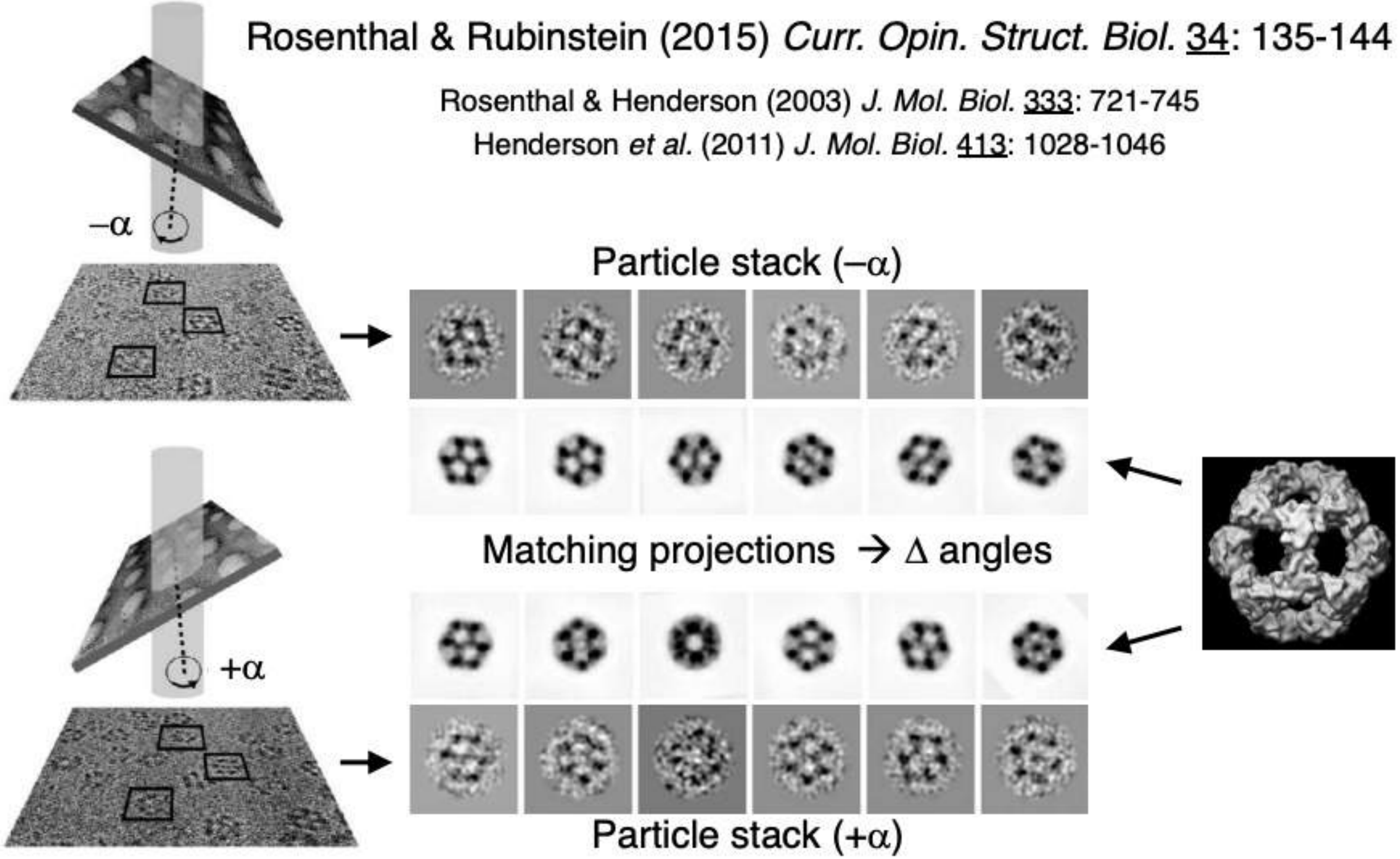
Appearance of expected secondary structure elements

only checks consistency of 3D ma with 2D data – excellent, also establishes handedness – not necessarily needed (but now pret much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz

Rosenthal & Rubinstein (2015) *Curr. Opin. Struct. Biol.* 34: 135-144

Rosenthal & Henderson (2003) *J. Mol. Biol.* 333: 721-745
Henderson *et al.* (2011) *J. Mol. Biol.* 413: 1028-1046



VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

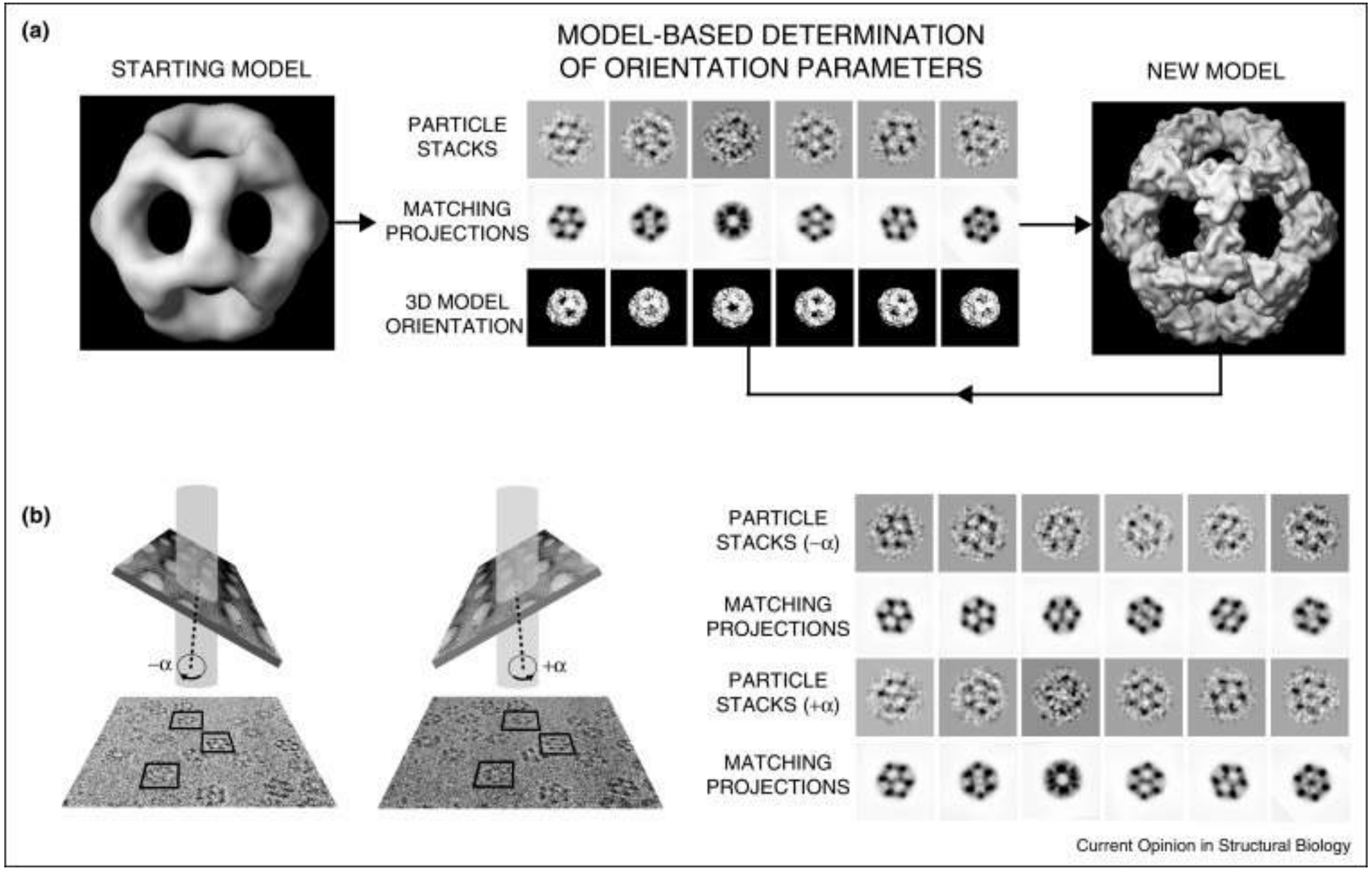
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



Current Opinion in Structural Biology

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

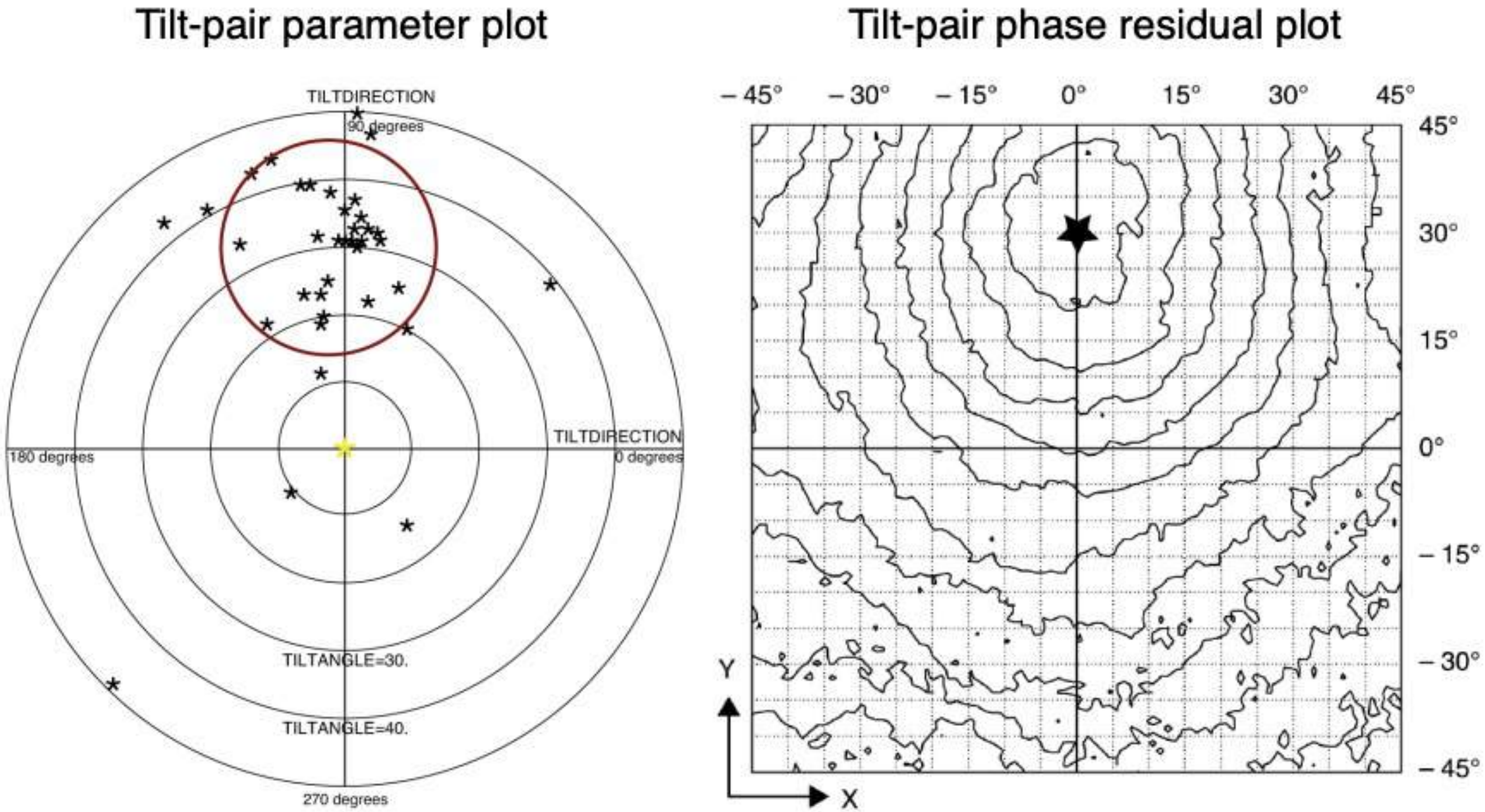
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



Rosenthal & Rubinstein (2015) *Curr. Opin. Struct. Biol.* 34: 135-144

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

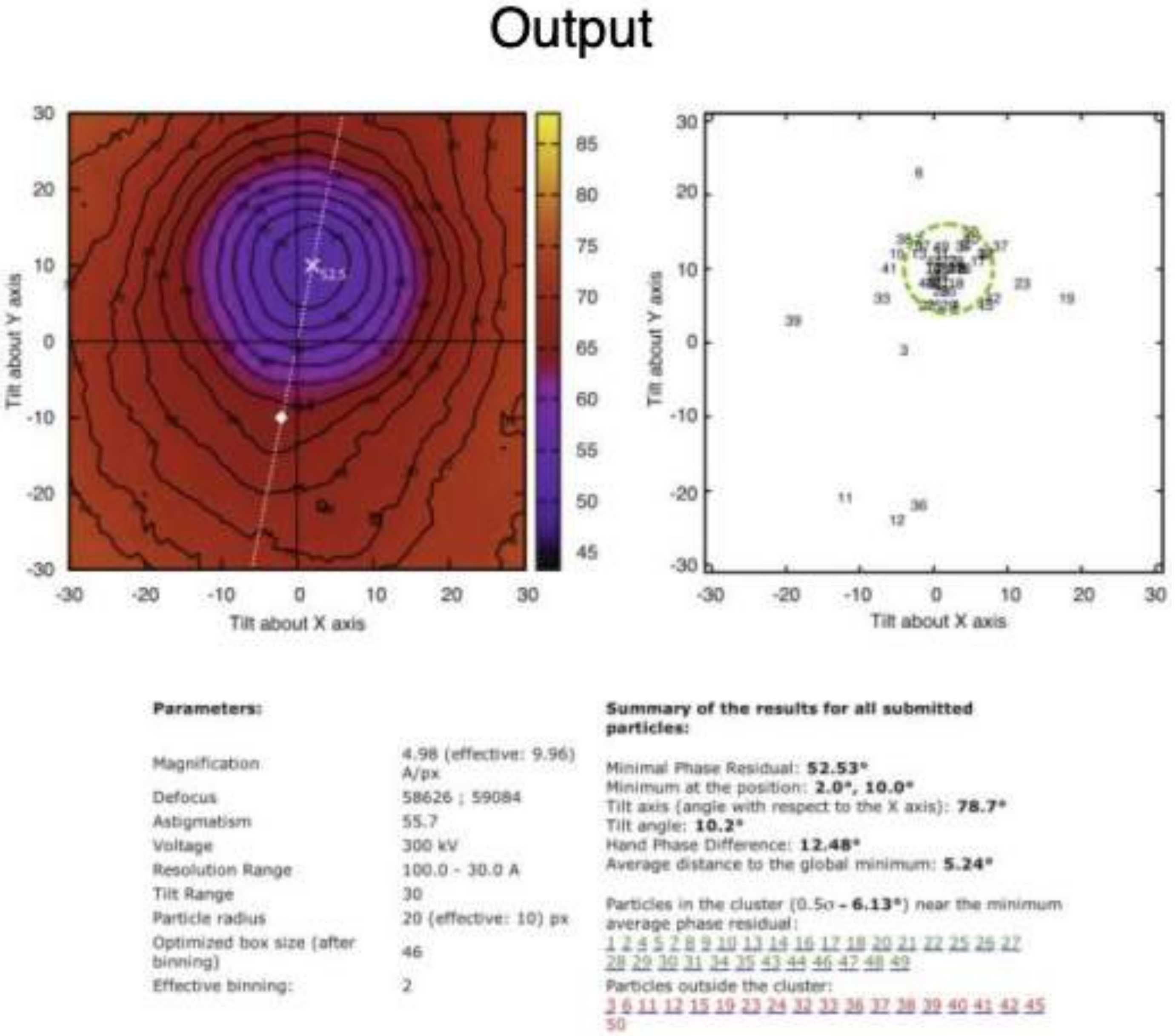
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



Parameters:

Magnification	4.98 (effective: 9.96) A/px
Defocus	58626 ; 59084
Astigmatism	55.7
Voltage	300 kV
Resolution Range	100.0 - 30.0 A
Tilt Range	30
Particle radius	20 (effective: 10) px
Optimized box size (after binning)	46
Effective binning:	2

Summary of the results for all submitted particles:

Minimal Phase Residual: **52.53°**
Minimum at the position: **2.0°, 10.0°**
Tilt axis (angle with respect to the X axis): **78.7°**
Tilt angle: **10.2°**
Hand Phase Difference: **12.48°**
Average distance to the global minimum: **5.24°**

Particles in the cluster (0.5σ - **6.13°**) near the minimum average phase residual:
1 2 4 5 7 8 9 10 13 14 16 17 18 20 21 22 25 26 27 28 29 30 31 34 35 43 44 46 47 48 49
Particles outside the cluster:
3 6 11 12 15 19 23 24 32 33 36 37 38 39 40 41 42 45 50

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

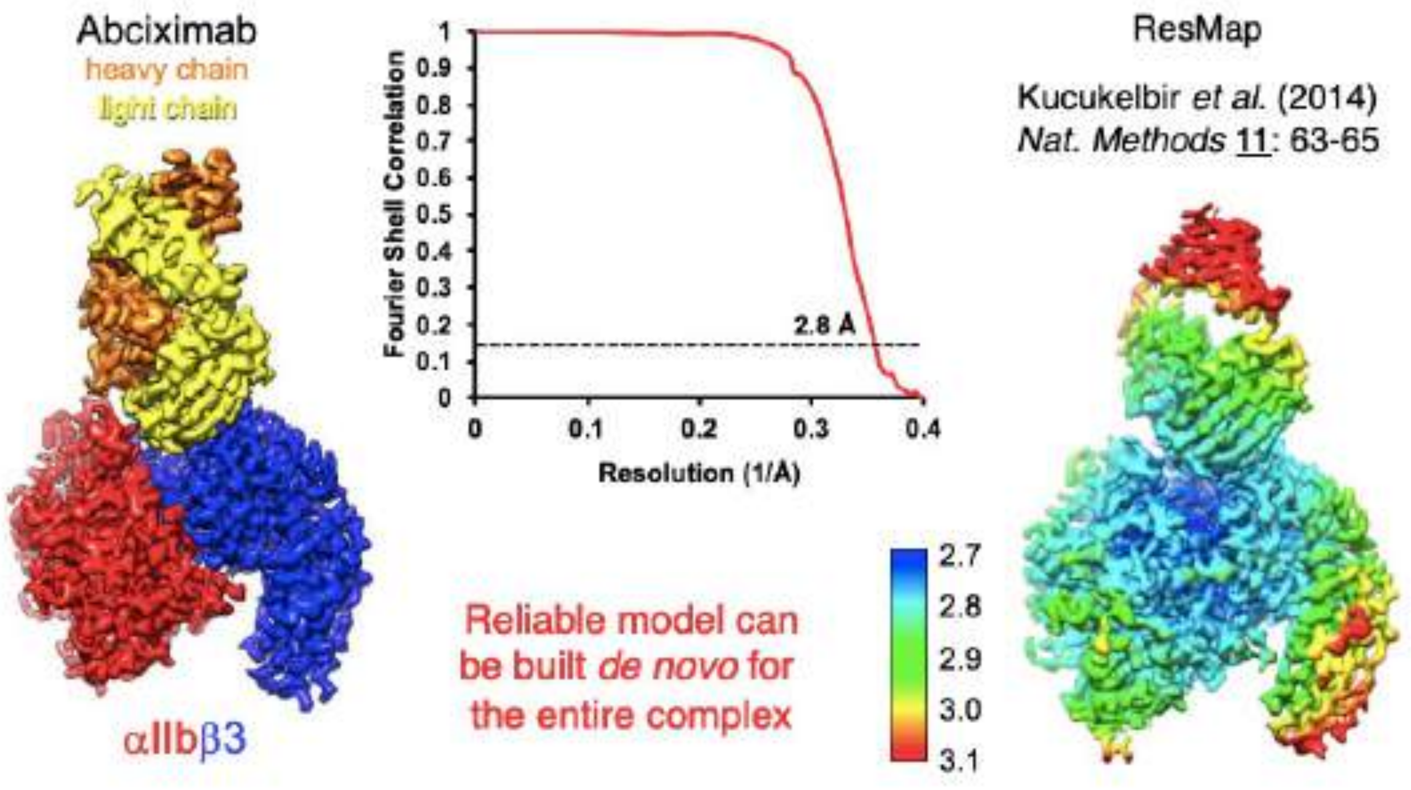
only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz

Voxel size: 1.63
Box length: 155
Degree of symmetry: 1
Object size ratio: 0.66

FSC

-adapted from Rich Hite



Kucukelbir *et al.* (2014)
Nat. Methods 11: 63-65

Reliable model can be built *de novo* for the entire complex

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

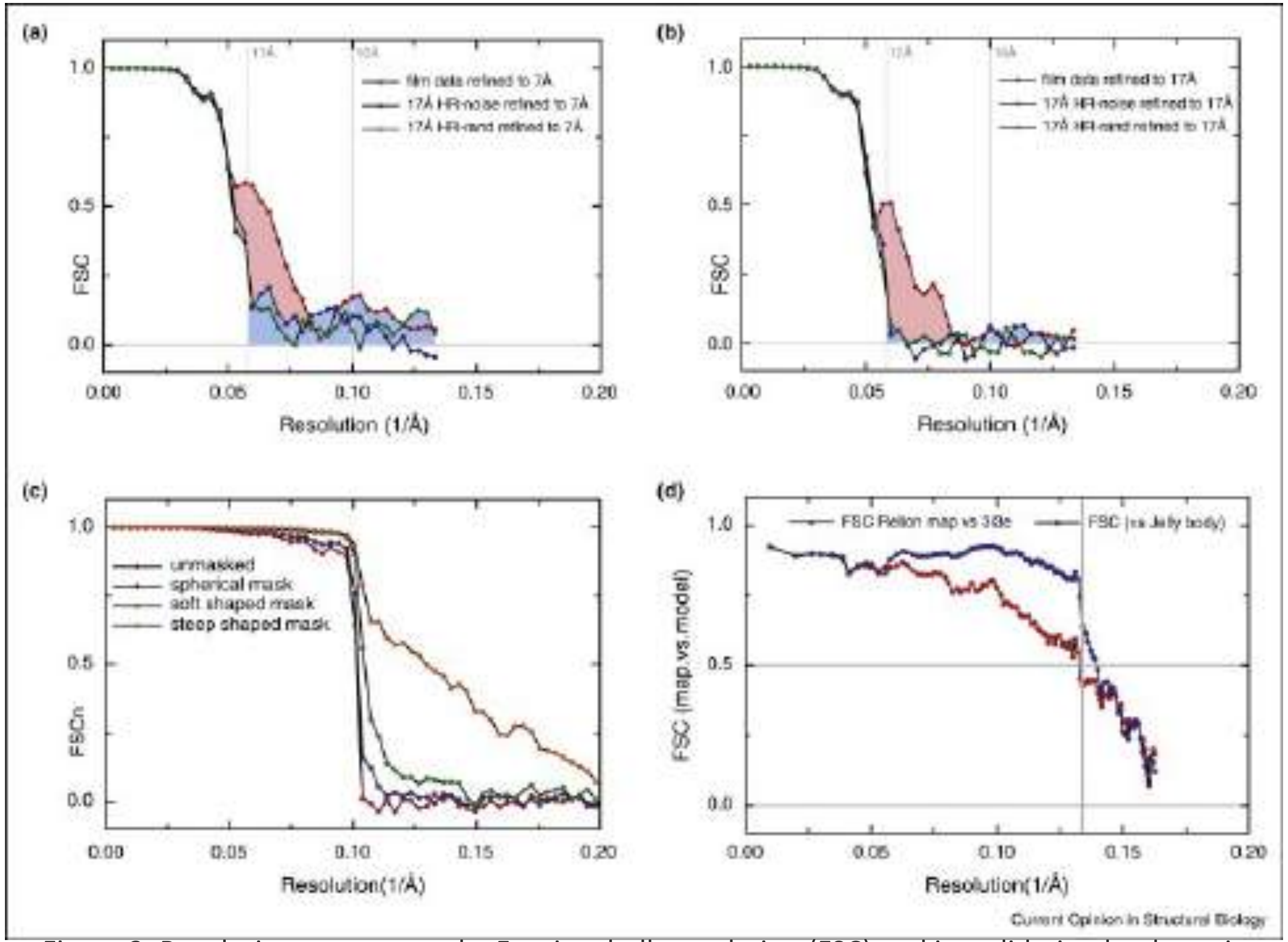
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



Current Opinion in Structural Biology

Figure 3. Resolution assessment by Fourier shell correlation (FSC) and its validation by the noise substitution test. **(a)** FSC between half-maps where particle orientation refinement uses image data to 7 Å resolution (red circle) or to 7 Å but where high resolution data beyond 17 Å has been substituted with data with randomized phases (HR-noise, green circle) or data from noise images (HR-noise, blue circle). Portion of FSC signal due to over-fitting is shown by the shaded blue region (FSCn), and the portion due to true structural signal is the shaded pink region (FSCtrue). **(b)** FSC curves as above but where particle orientation refinement has been limited to 17 Å resolution. The ‘free shells’ above 17 Å resolution show no over-fitting because they were not used in refinement. **(c)** FSCn between half-maps for noise-substituted data shows FSC artefacts that result from a tight mask. **(d)** FSC curve (red), also called Cref, between whole-map and coordinate model placed in the map by rigid body fitting crosses the 0.5 threshold at a point indicating 7.6 Å resolution. Flexible fitting of the model produces an overly optimistic FSC curve (blue). Panels are Figures 3a,b, 4b, and 5b of Ref. [6••].

VALIDATION TIPS

Compare reference free averages with projections

Tilt-pair analysis

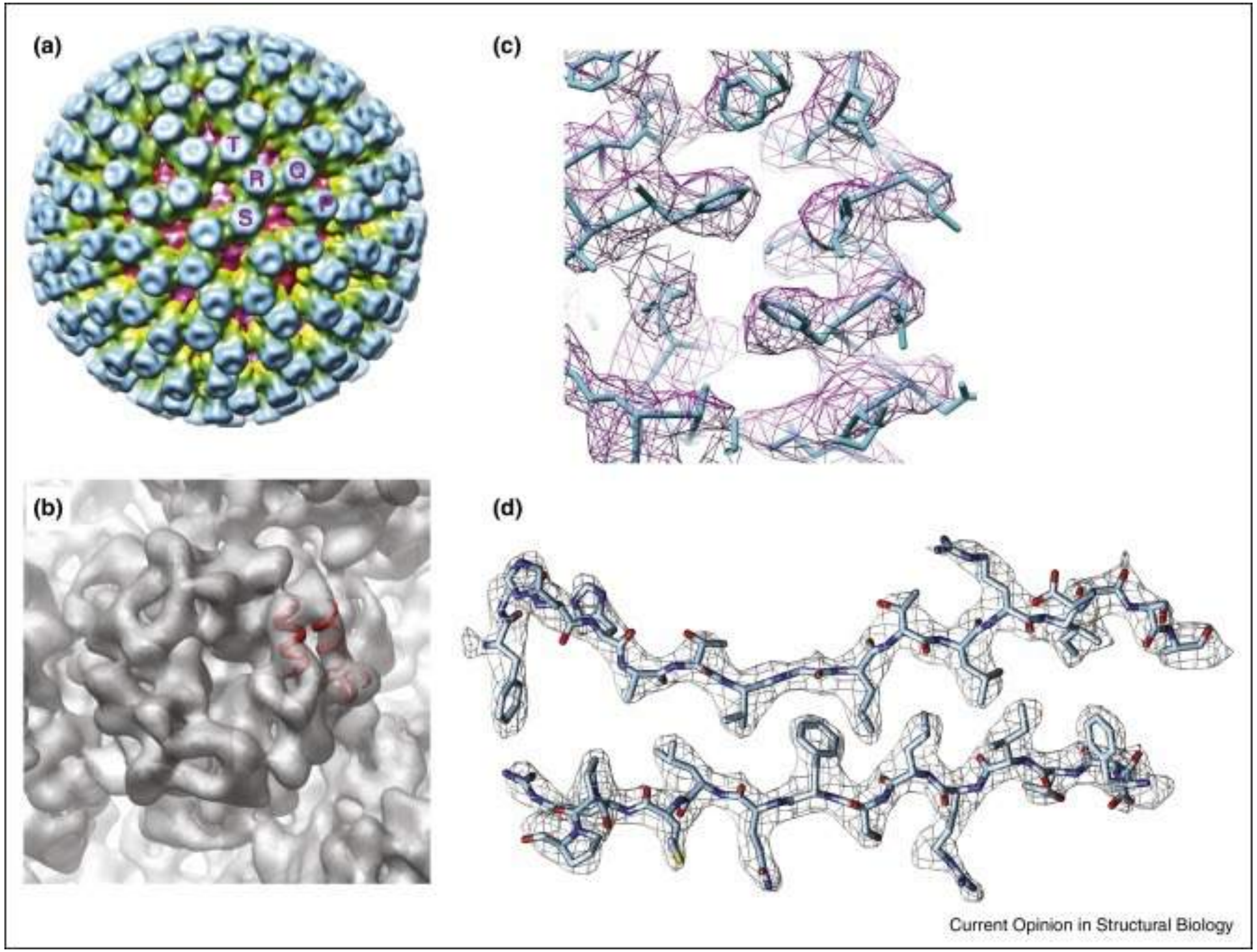
“Gold standard” FSC

Randomize phases

Appearance of expected secondary structure elements

only checks consistency of 3D map with 2D data – excellent, also establishes handedness – not necessarily needed (but now pretty much default) – excellent (implemented in software packages) – Evaluate with published information

-adapted from Tom Walz



Current Opinion in Structural Biology