SickKids RESEARCH INSTITUTE

Single Particle Cryo-EM

John Rubinstein

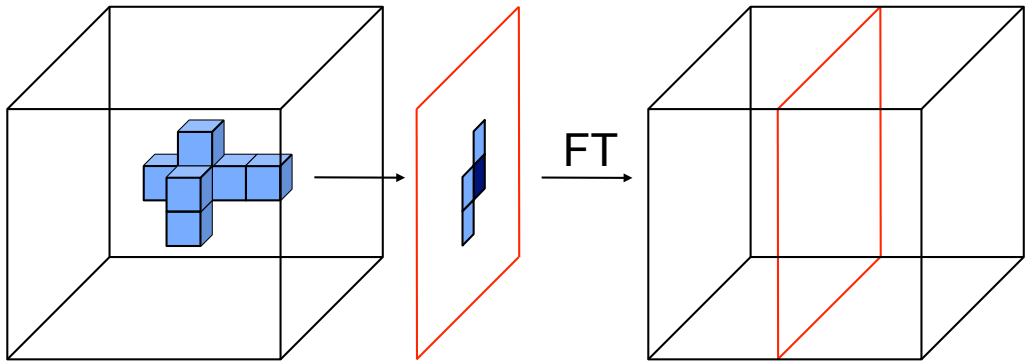
Molecular Medicine Program
The Hospital for Sick Children Research Institute

Departments of Biochemistry and Medical Biophysics
The University of Toronto

www.sickkids.ca/research/rubinstein
@RubinsteinJohn

UNIVERSITY OF TORONTO

The projection theorem

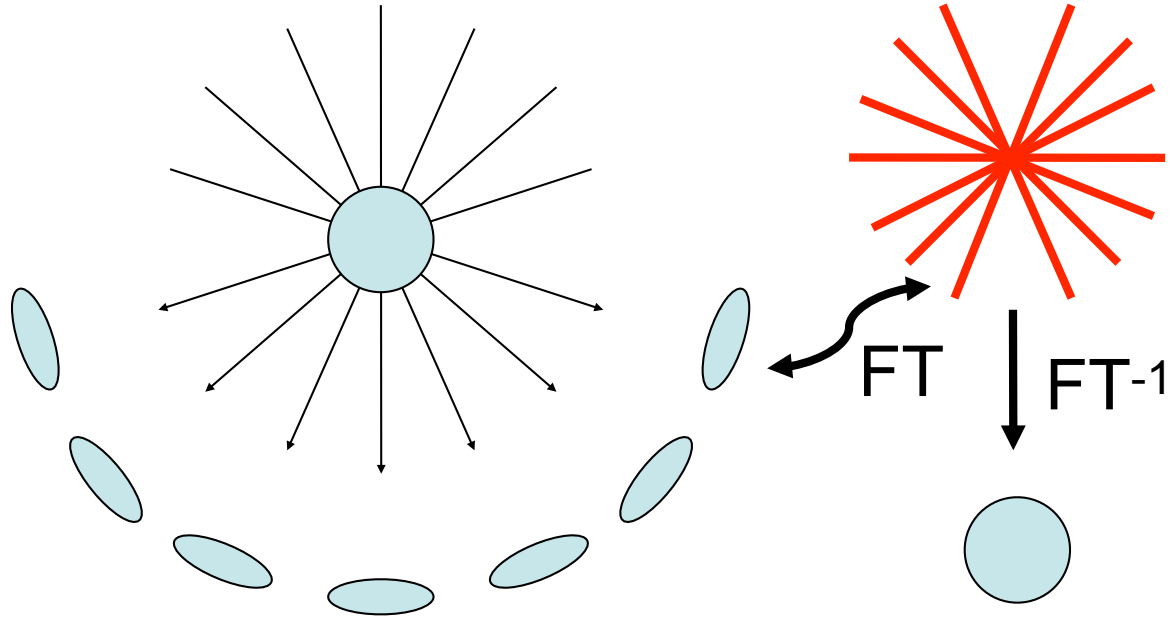


Projection

Central section through the 3-D FT of object

Building 3-D maps of proteins

1) Calculate the 2-D FTs of the projected views
2) Put them together to form the 3-D FT of the specimen
3) Calculate the structure of the specimen by performing a Fourier synthesis of the 3-D FT



Real Space Projections of Structure

De Rosier & Klug (1968). Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature 217, 130-4.

Nobel prize for chemistry, 1982

Aaron Klug *"for his development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes"*.

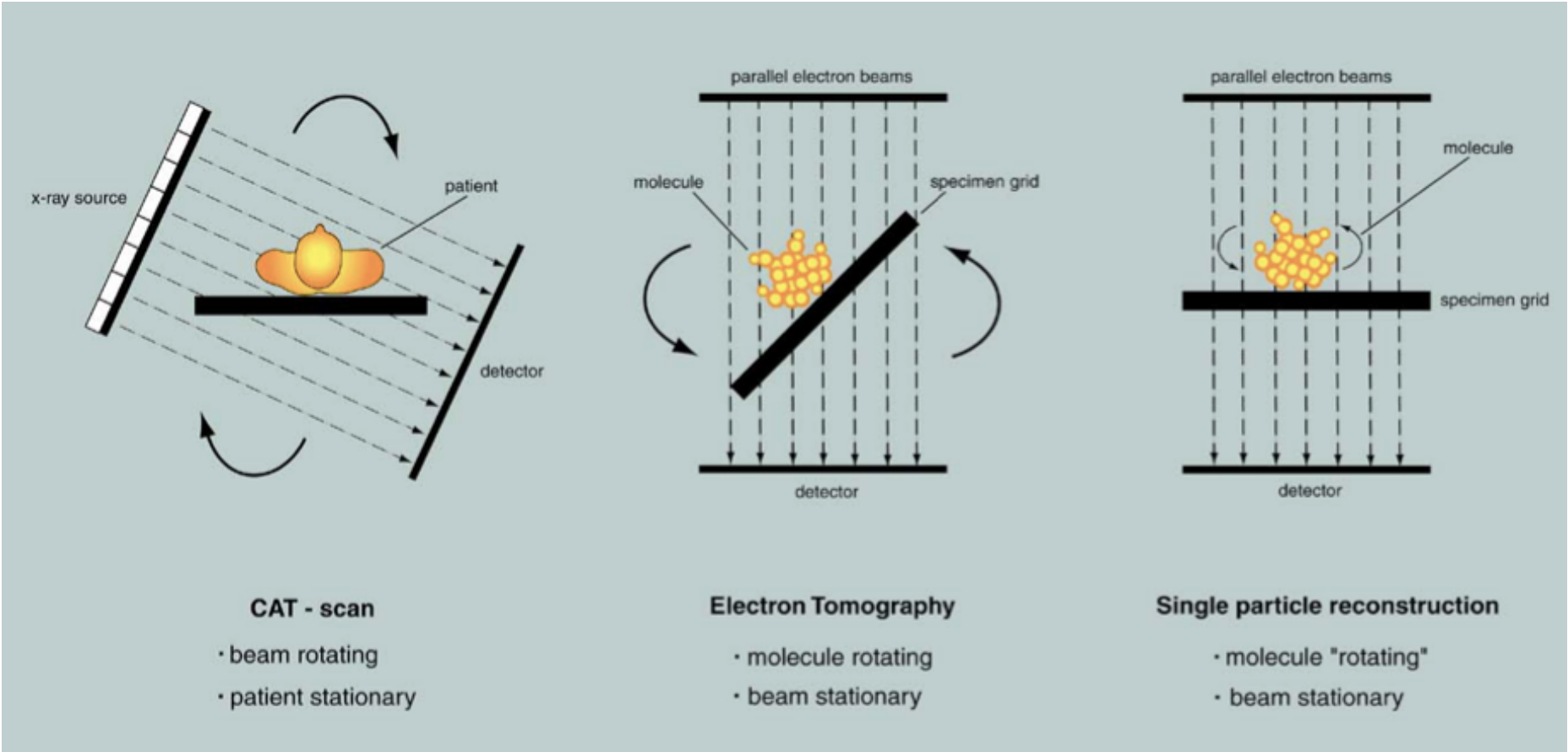
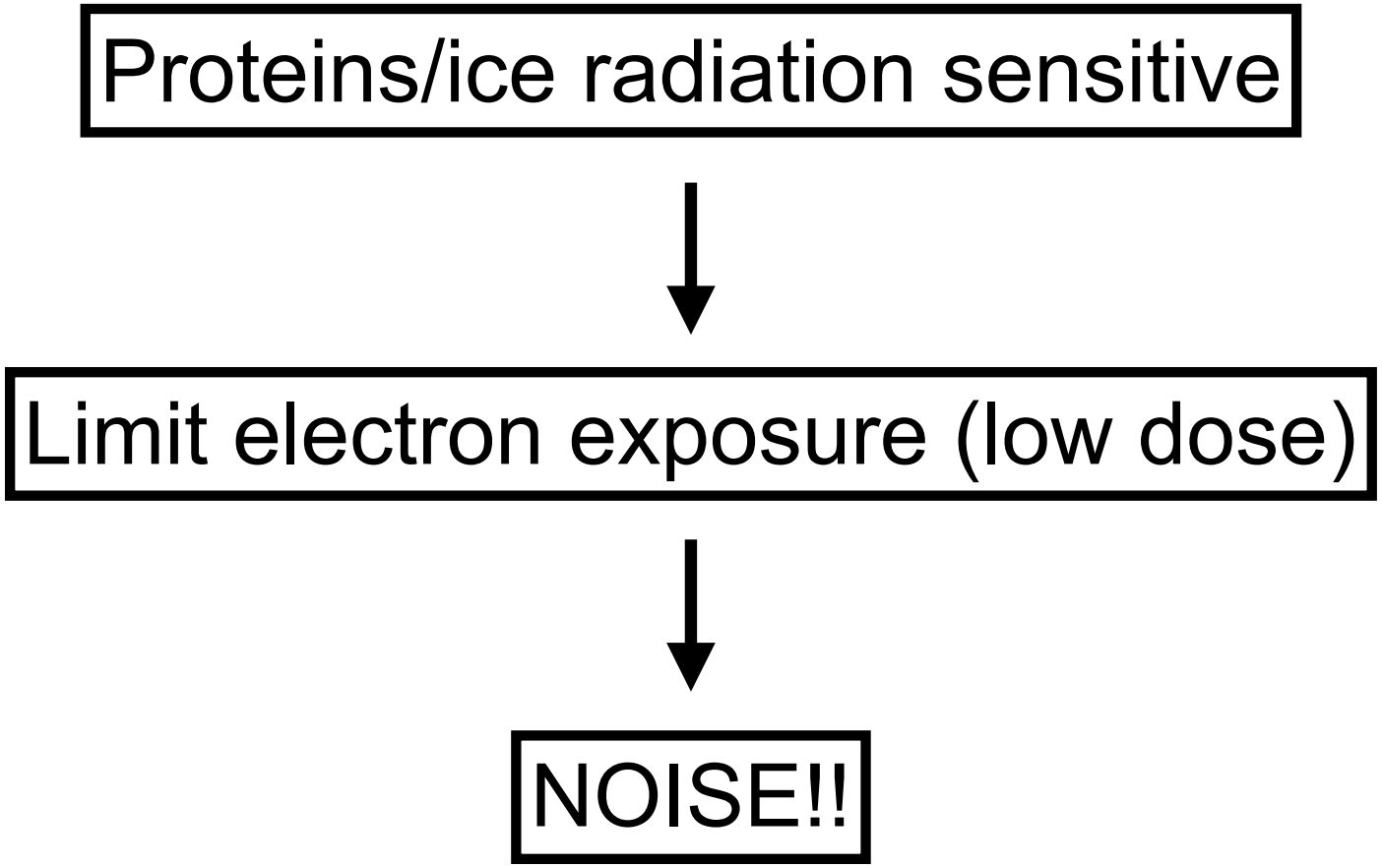
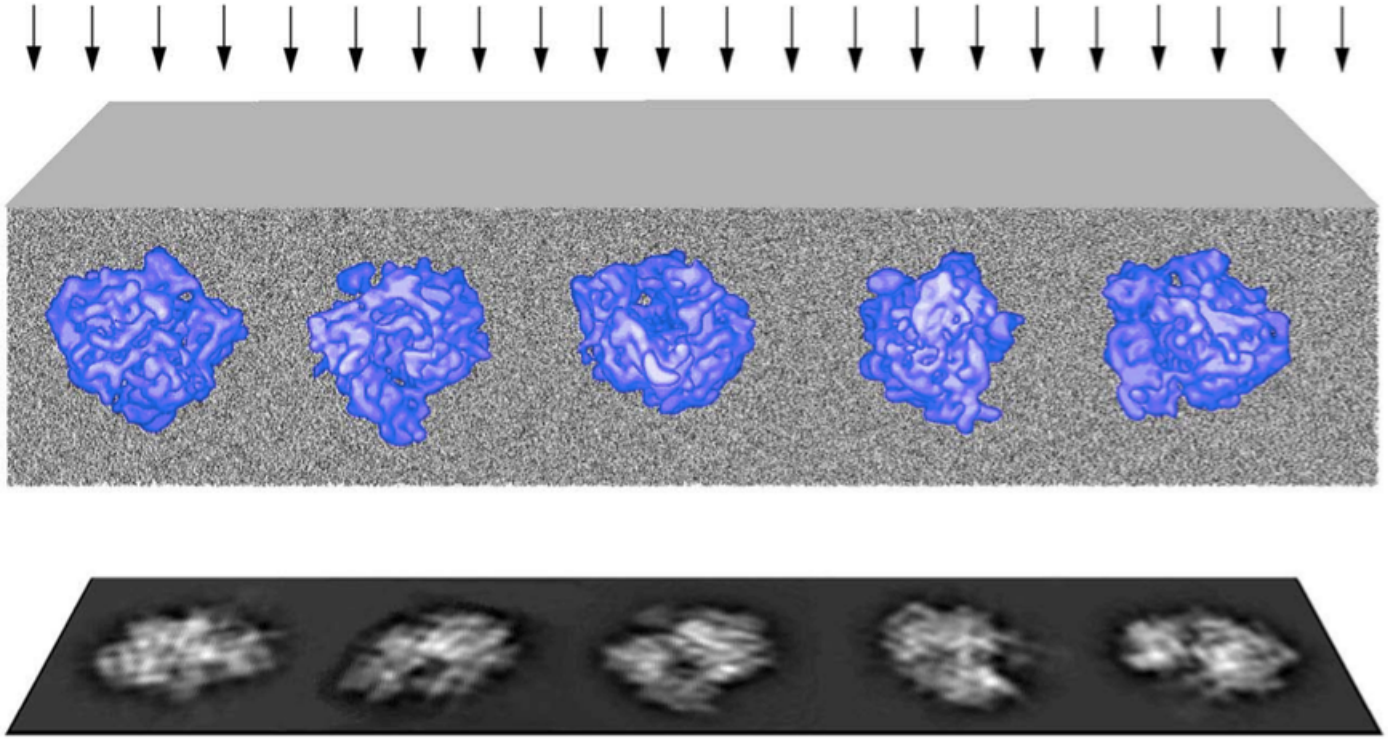
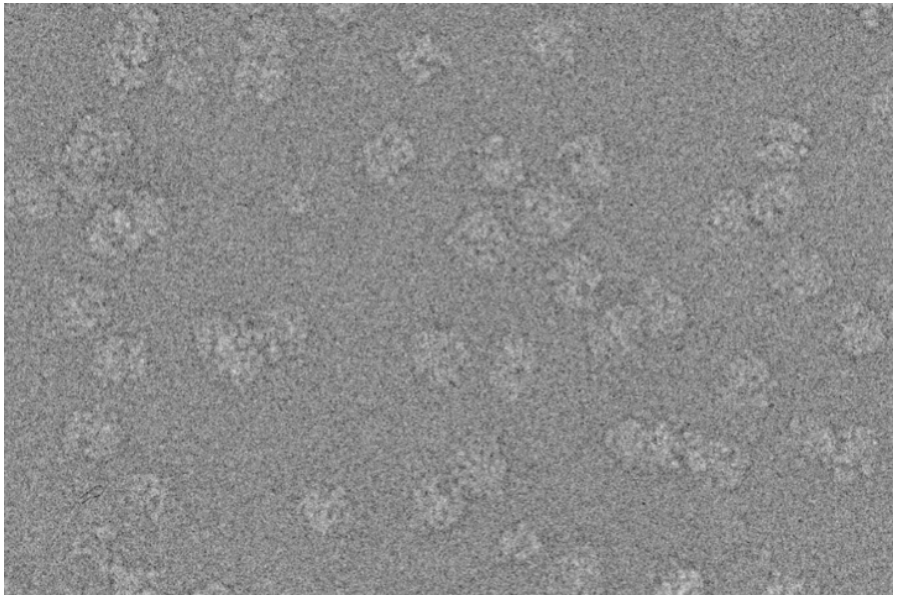


Image used with permission from Joachim Frank (Columbia University)


Proteins/ice radiation sensitive
Limit electron exposure (low dose)
NOISE!!





And one other important distortion (the Contrast Transfer Function)

J. Frank (2002). *Ann. Rev. Biophys Biochem. Struct.* **31** 303

Things we need to do in practice (not necessarily in this order):

1) Find the molecules in images (particle selection)
2) Correct for imperfections in the imaging system (contrast transfer function)
3) Correct specimen stage drift and beam-induced specimen movement (motion correction)
4) Determine which molecules come from the same structure (classification)
5) Determine the relative orientations of the images of molecules that belong together (alignment)
6) Calculate the 3D structure(s) (reconstruction)

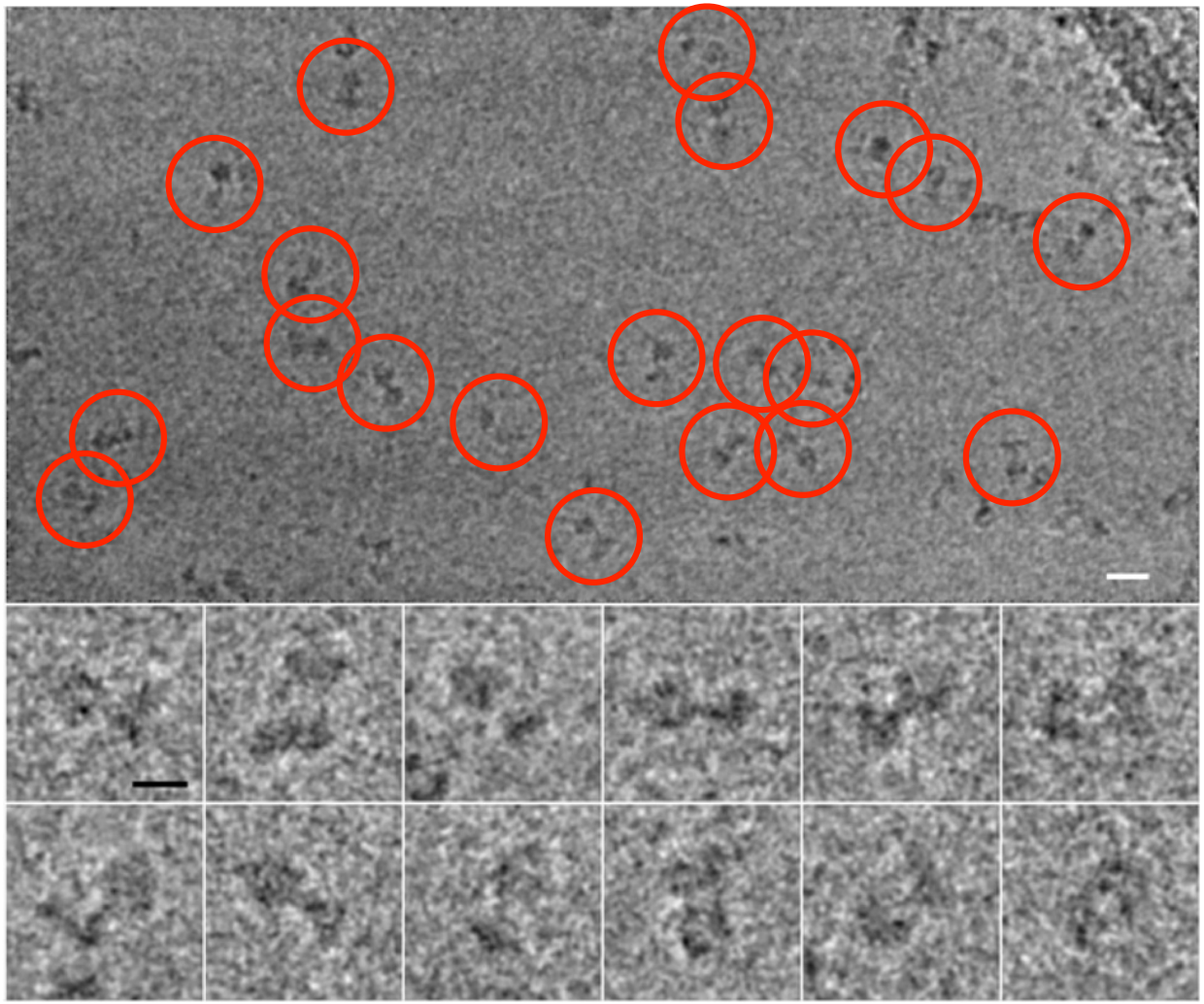
Particle selection

Particle selection

- extract from micrograph approximately centered images of the molecule of interest.
- Require a “stack” of between 10^2 and 10^6 images of individual particles

Micrograph of Detergent solubilized ATP synthase particles

Gallery of selected Particle images



Rubinstein, Walker, & Henderson (2003). *EMBO J.* 22, 6182-92.

Ways of picking particles in images

1) Template matching
2) Devise some method for distinguishing ‘objects’ from noise
3) Machine Learning methods

High-resolution template matching can get you into trouble…

A Cautionary Tale

Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer

Youdong Mao[a,b,1], Liping Wang[a,b], Christopher Gu[a,b], Alon Herschhorn[a,b], Anik Désormeaux[c], Andrés Finzi[c], Shi-Hua Xiang[d], and Joseph G. Sodroski[a,b,e,f,1]

[a]Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA 02215; [b]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115; [c]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Department of Microbiology and Immunology, Université de Montréal, Montréal, QC, Canada H3A 2B4; [d]Nebraska Center for Virology, School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, Lincoln, NE 68583; [e]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard, Cambridge, MA 02139; and [f]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115

Edited* by Beatrice H. Hahn, University of Pennsylvania, Philadelphia, PA, and approved May 15, 2013 (received for review April 19, 2013)

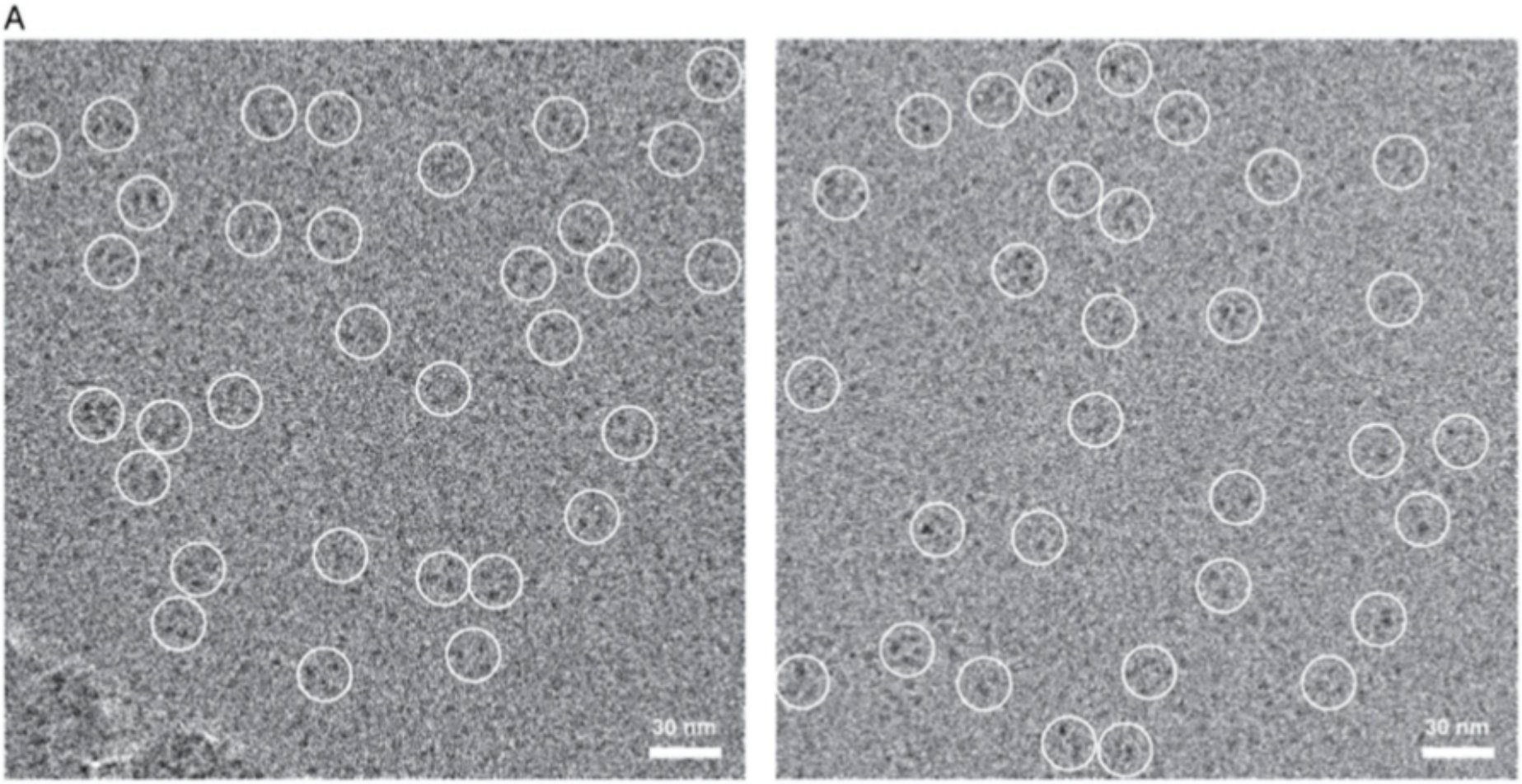
12438–12443 | PNAS | **July 23, 2013** | vol. 110 | no. 30

Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer

Youdong Mao[1], Liping Wang[1], Christopher Gu[1], Alon Herschhorn[1], Shi-Hua Xiang[1,5], Hillel Haim[1], Xinzhen Yang[2] & Joseph Sodroski[1,3,4]

NATURE STRUCTURAL & MOLECULAR BIOLOGY VOLUME 19 NUMBER 9 SEPTEMBER 2012

The authors acquired images:



Employed a template-based particle selection scheme:

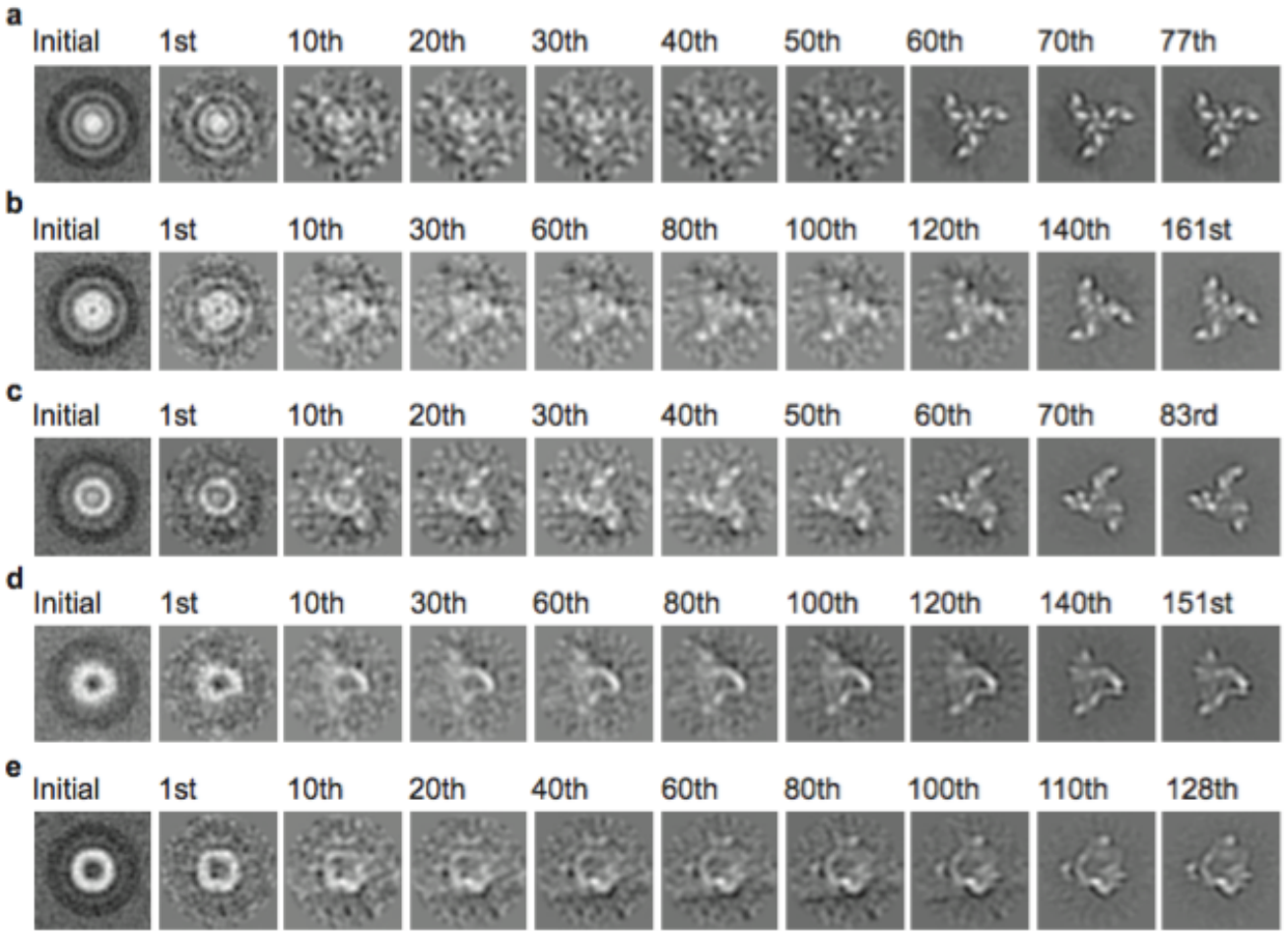
Dual-target function validation of single-particle selection from low-contrast cryo-electron micrographs

Youdong Mao, Luis R. Castillo-Menendez, Joseph Sodroski

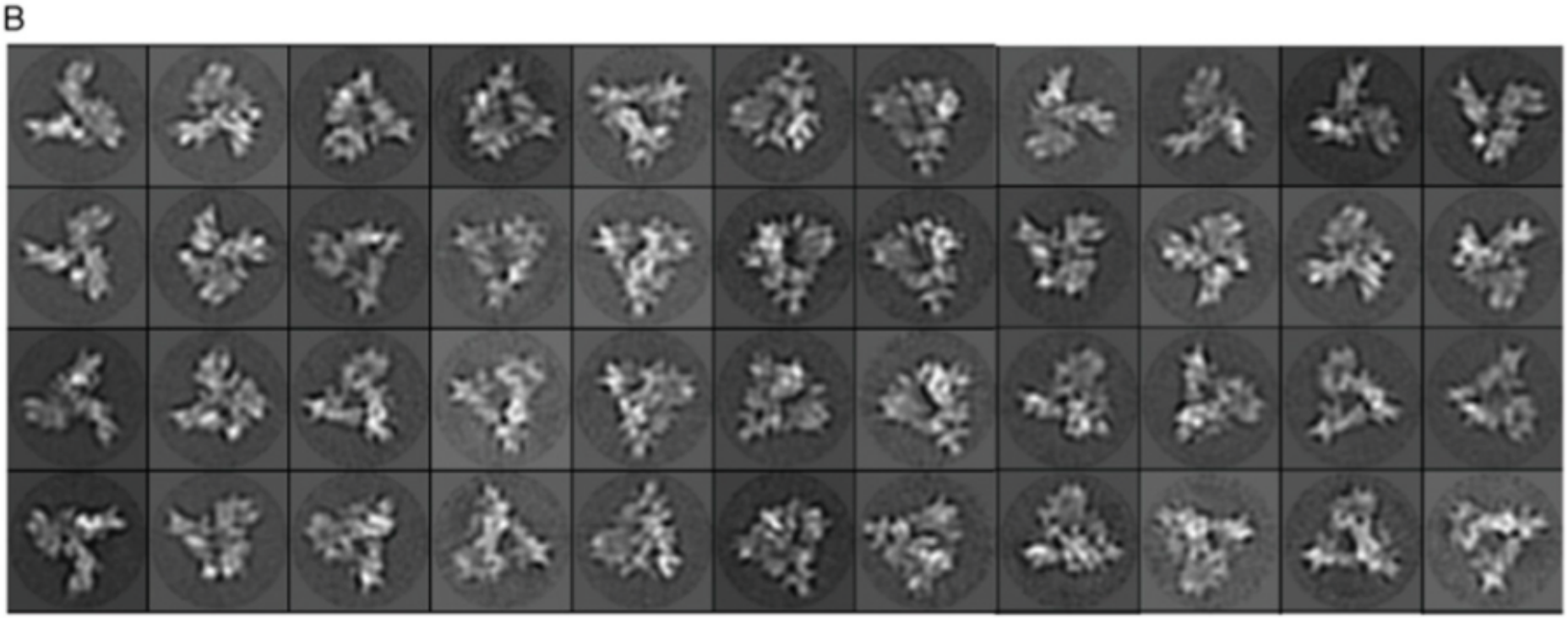
(Submitted on 10 Sep 2013)

Weak-signal detection and single-particle selection from low-contrast micrographs of frozen hydrated biomolecules by cryo-electron microscopy (cryo-EM) presents a practical challenge. Cryo-EM image contrast degrades as the size of biomolecules of structural interest decreases. When the image contrast falls into a range where the location or presence of single particles becomes ambiguous, a need arises for objective computational approaches to detect weak signal and to select and verify particles from these low-contrast micrographs. Here we propose an objective validation scheme for low-contrast particle selection using a combination of two different target functions. In an implementation of this dual-target function (DTF) validation, a first target function of fast local correlation was used to select particles through template matching, followed by signal validation through a second target function of maximum likelihood. By a systematic study of simulated data, we found that such an implementation of DTF validation is capable of selecting and verifying particles from cryo-EM micrographs with a signal-to-noise ratio as low as 0.002. Importantly, we demonstrated that DTF validation can robustly evade over-fitting or reference bias from the particle-picking template, allowing true signal to emerge from amidst heavy noise in an objective fashion. The DTF approach allows efficient assembly of a large number of single-particle cryo-EM images of smaller biomolecules or specimens containing contrast-degrading agents like detergents in a semi-automatic manner.

Generated reference free class averages:



Generated reference free class averages:



Calculated a 3D map (2012):

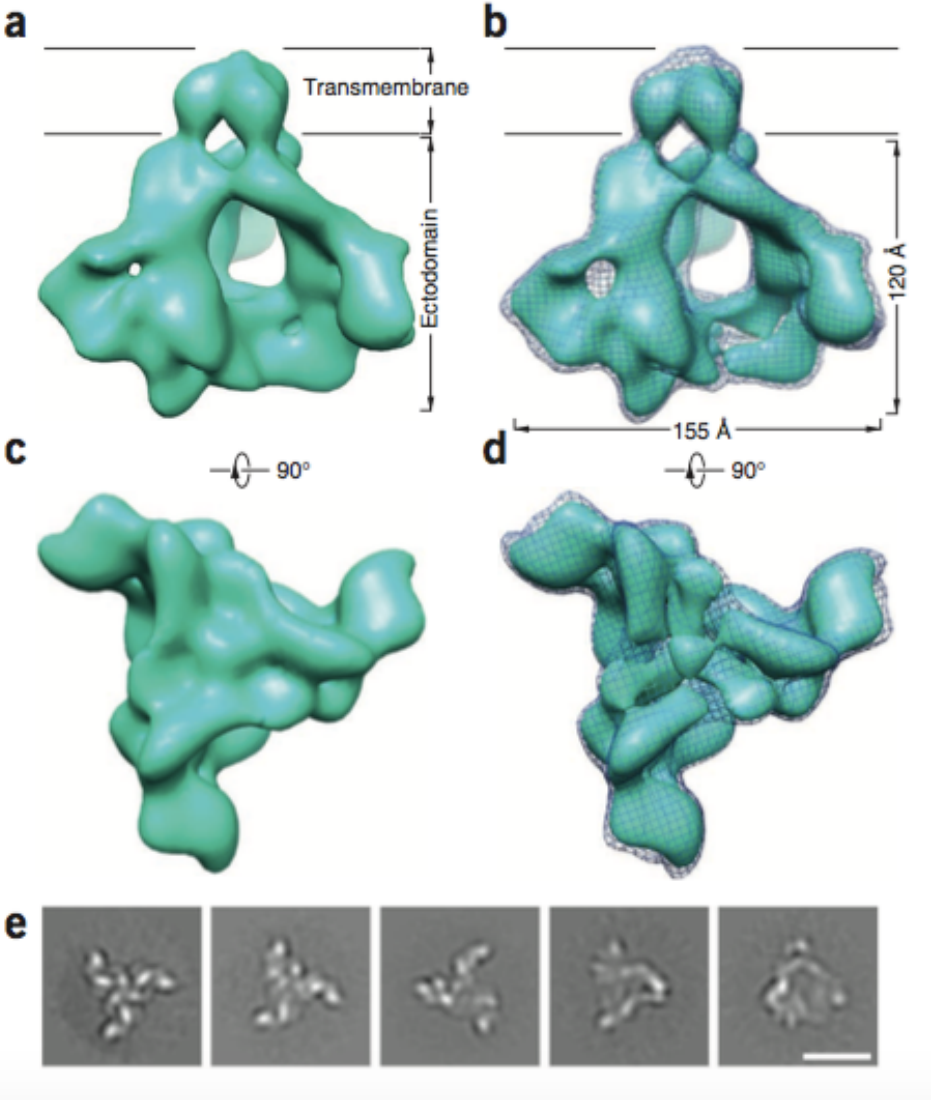


Figure 1 The cryo-EM structure of the membrane-bound HIV-1 Env trimer at ~11-Å resolution. (**a**) The reconstruction of the HIV-$1_{JR\text{-}FL}$ Env(–)ΔCT trimer is shown as a solid surface viewed from a perspective parallel to the viral membrane. The approximate boundaries of the transmembrane region and ectodomain are indicated. (**b**) The Env trimer reconstruction is visualized at two different levels of contour, illustrated as a meshwork (lower level) and a solid-surface representation (higher level). (**c**) The Env trimer reconstruction in a solid-surface representation is viewed from the perspective of the target cell, at the same contour level as that shown in **a**. (**d**) The Env trimer reconstruction is shown at two different levels of contour in the same way as in **b**, viewed from the perspective of the target cell. (**e**) The images show typical reference-free class averages produced by maximum-likelihood alignment with no C3 symmetry imposed. Scale bar, 10 nm. See **Supplementary Figures 3** and **4**, **Supplementary Movies 2** and **3** for more details.

Calculated a 3D map (2013):

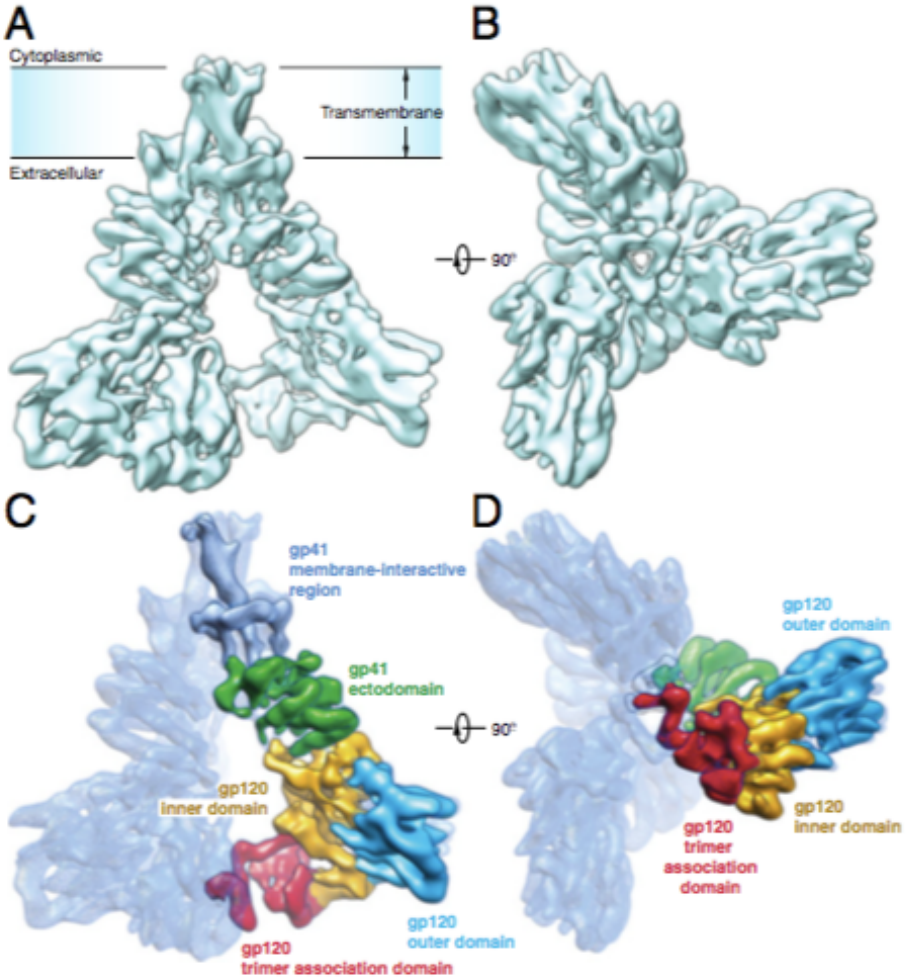


Fig. 1. Architecture of the HIV-1 Env trimer. (*A*) Cryo-EM map of the HIV-1_{JR-FL} Env trimer in a surface representation, viewed from a perspective parallel to the viral membrane. (*B*) Cryo-EM map of the HIV-1_{JR-FL} Env trimer, viewed from the perspective of the target cell. (*C* and *D*) Domain organization of the Env protomer, revealed by segmentation of the density map. The gp120 domains are colored as follows: outer domain, blue; inner domain, orange; and TAD, red. The gp41 domains are colored as follows: ectodomain, green; and transmembrane region, cyan.

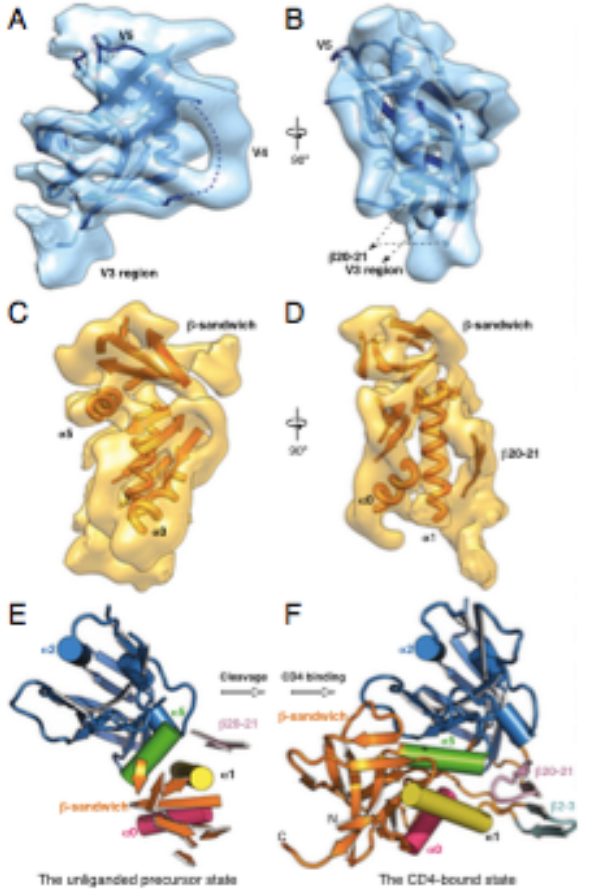
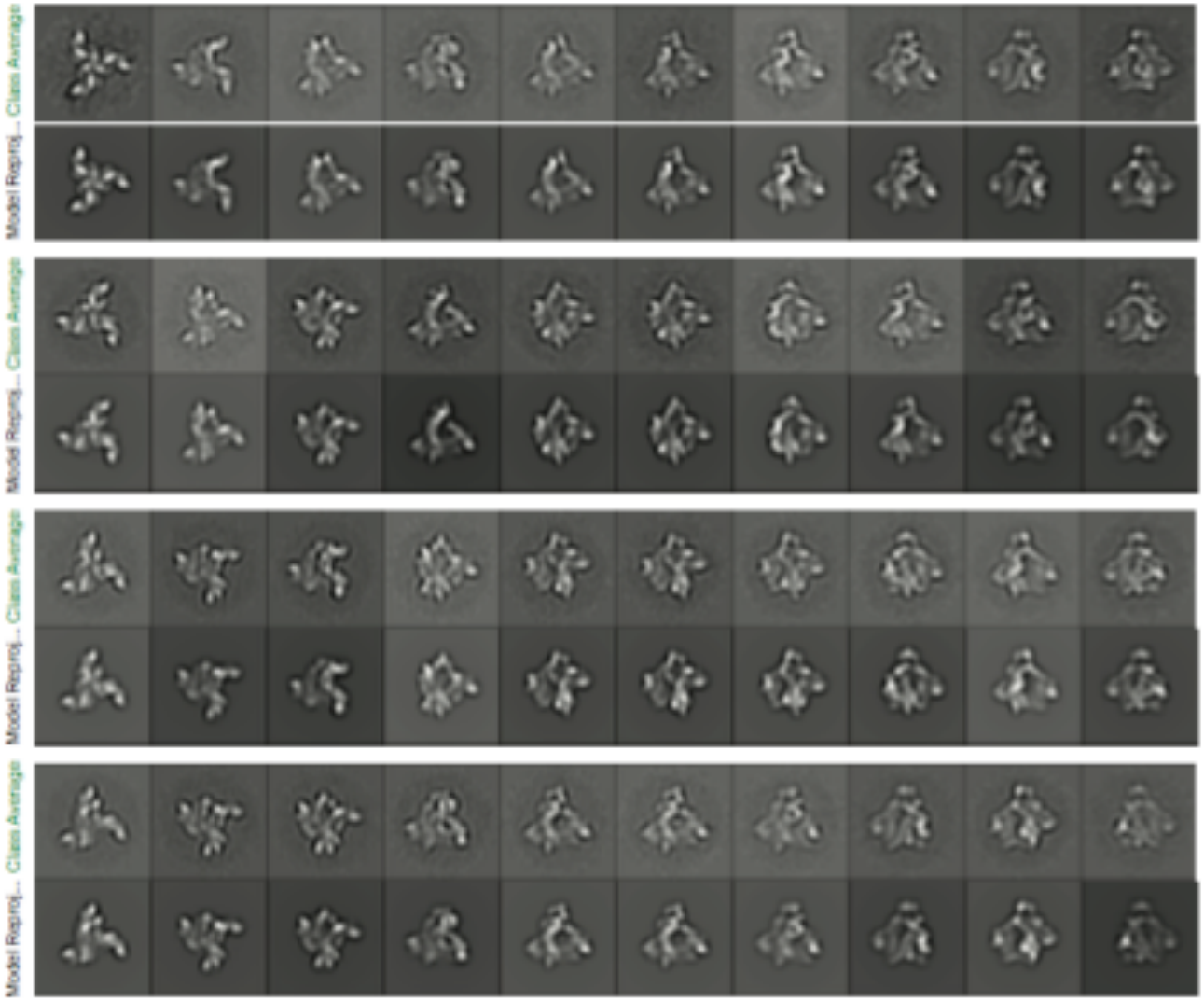


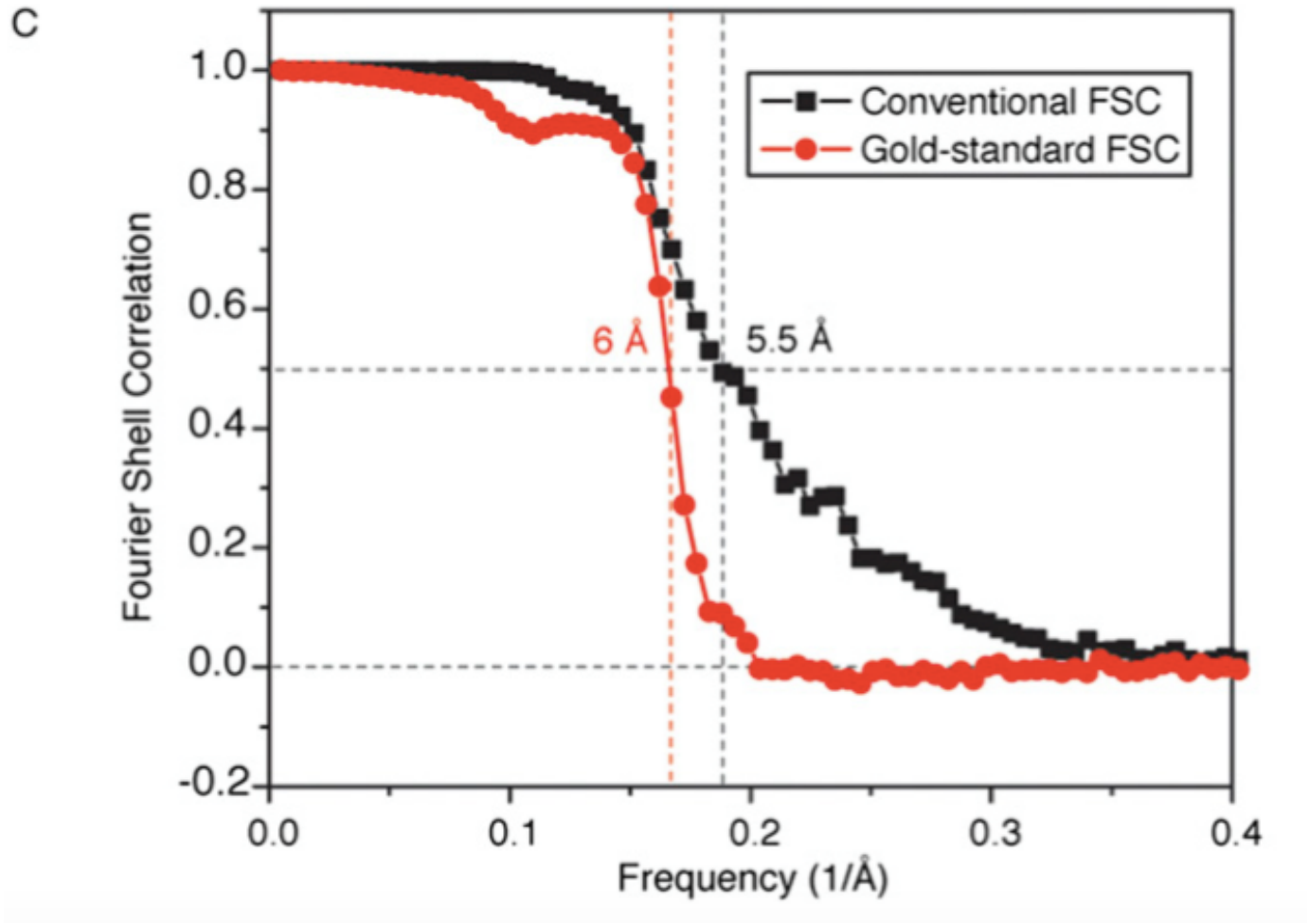
Fig. 2. The unliganded gp120 subunit and CD4-induced changes. (*A* and *B*) Crystal structure of the outer domain of the gp120 core (17) was fitted into the cryo-EM map of the unliganded Env trimer and is shown from two perspectives parallel to the viral membrane. The approximate location of the V4 variable region, which was not resolved in the crystal structure, is indicated by a broken line. (*C* and *D*) Secondary structure organization of the inner domain of gp120 was approximated in the cryo-EM map of the unliganded Env trimer, which is viewed from two perspectives parallel to the viral membrane. (*E* and *F*) For comparison of the unliganded precursor state (*E*) and the CD4-bound state (*F*) of the gp120 core, the gp120 outer domains (blue) are aligned in the same orientation. The gp120 core in the unliganded precursor state is derived as described in *A*–*D* above, and the CD4-bound gp120 core structure is from an X-ray crystal structure (Protein Data Bank ID: 3JWD).

Compared map projections to class averages:



Supplementary Figure 4. Comparison of the class averages refined by a maximum-likelihood (ML) approach and the re-projection of the reconstructed 3D model. The ML-aligned 2D class averages are shown in the 1st, 3rd, 5th, and 7th rows. The corresponding model re-projections are shown in the 2nd, 4th, 6th and 8th rows. The good agreement of the class averages and the model re-projections supports the validity of the 3D reconstruction.

Measured the resolution (2013):



Fit crystal structures (2013):

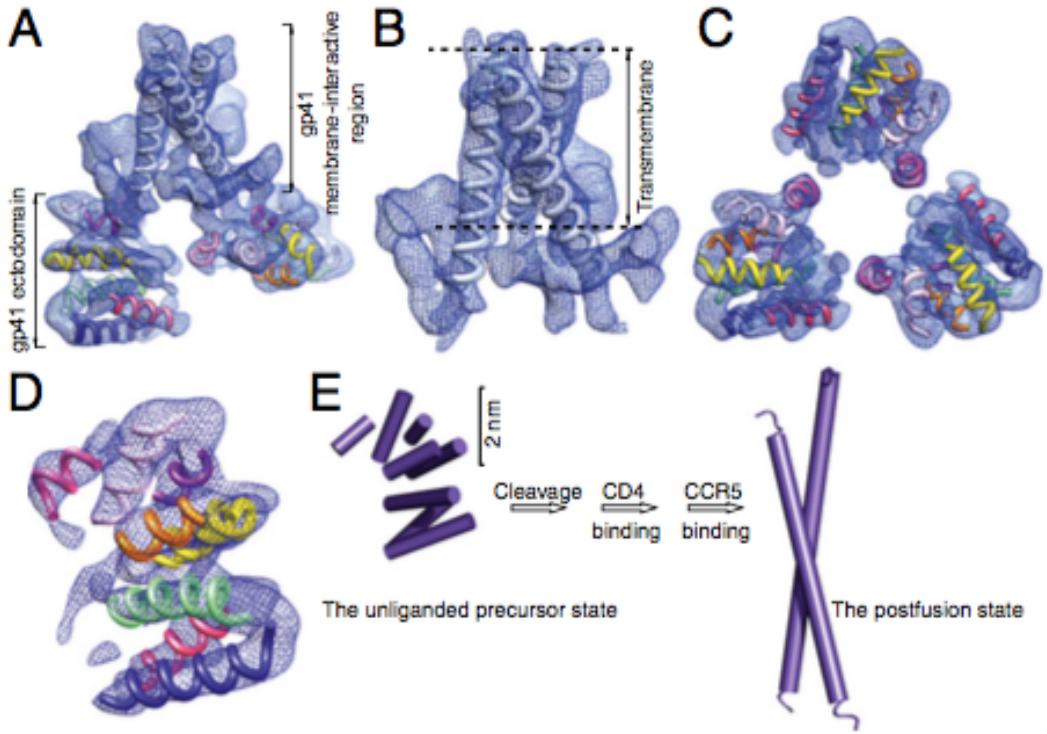


Fig. 4. gp41 subunit in the Env trimer. (*A*) Cryo-EM map of the three gp41 subunits in the Env trimer, viewed from an angle of ~30° with respect to the viral membrane. (*B*) gp41 membrane-interactive region viewed from a perspective parallel to the viral membrane. (*C*) gp41 ectodomain, viewed from the perspective of the virus. The torus-like topology of the three gp41 subunits in the Env trimer is evident from this perspective. (*D*) gp41 ectodomain, viewed from a perspective parallel to the viral membrane. (*E*) Comparison of the gp41 ectodomain in the unliganded precursor state and the postfusion state with respect to the tertiary organization of helices. The perspective is identical to that shown in *D*, with a reduction in scale. For simplicity, only one of the three gp41 glycoprotein subunits in the Env trimer is depicted.

…and were criticized extensively

PERSPECTIVE

Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise

Richard Henderson[1]

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

Edited by Michael G. Rossmann, Purdue University, West Lafayette, IN, and approved September 19, 2013 (received for review August 5, 2013)

Single particle cryo-electron microscopy is currently poised to produce high-resolution structures of many biological assemblies, but several pitfalls can trap the unwary. This critique highlights one problem that is particularly relevant when smaller structures are being studied. It is known as "Einstein from noise," in which the experimenter honestly believes they have recorded images of their particles, whereas in reality, most if not all of their data consist of pure noise. Selection of particles using cross-correlation methods can then lead to 3D maps that resemble the model used in the initial selection and provide the illusion of progress. Suggestions are given about how to circumvent the problem.

Fig. 1. Illustration taken from a paper describing model bias (23). The image is copied from figure 2*A* in that paper. The familiar photograph of Einstein emerged from 1,000 images of pure white noise, after alignment to the model using a cross-correlation function. Reprinted from *Journal of Structural Biology*, Vol. 166, M. Shatsky et al., A method for the alignment of heterogeneous macromolecules from electron microscopy, pp. 67–78, Copyright 2009, with permission from Elsevier.

From Henderson's paper

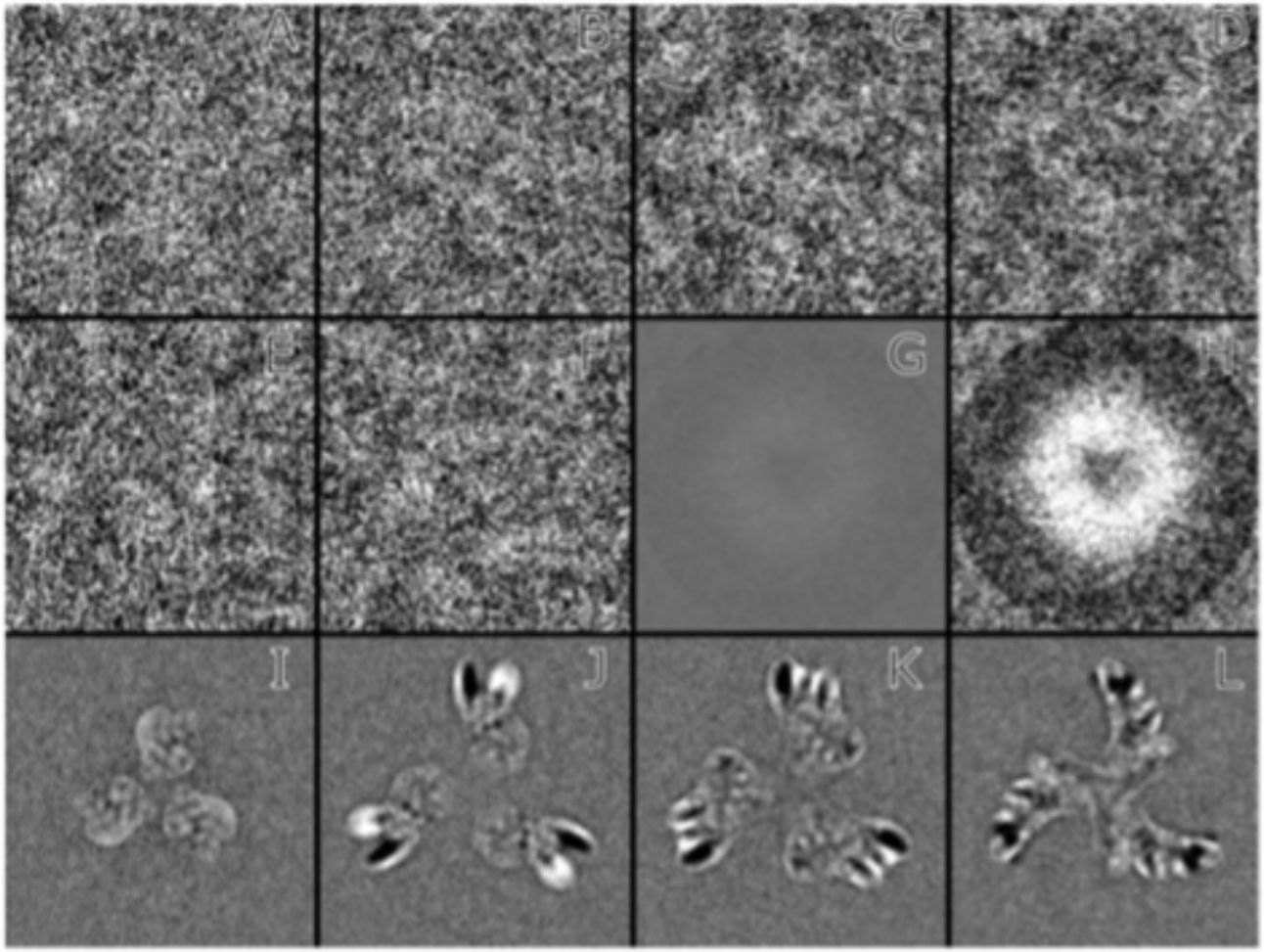


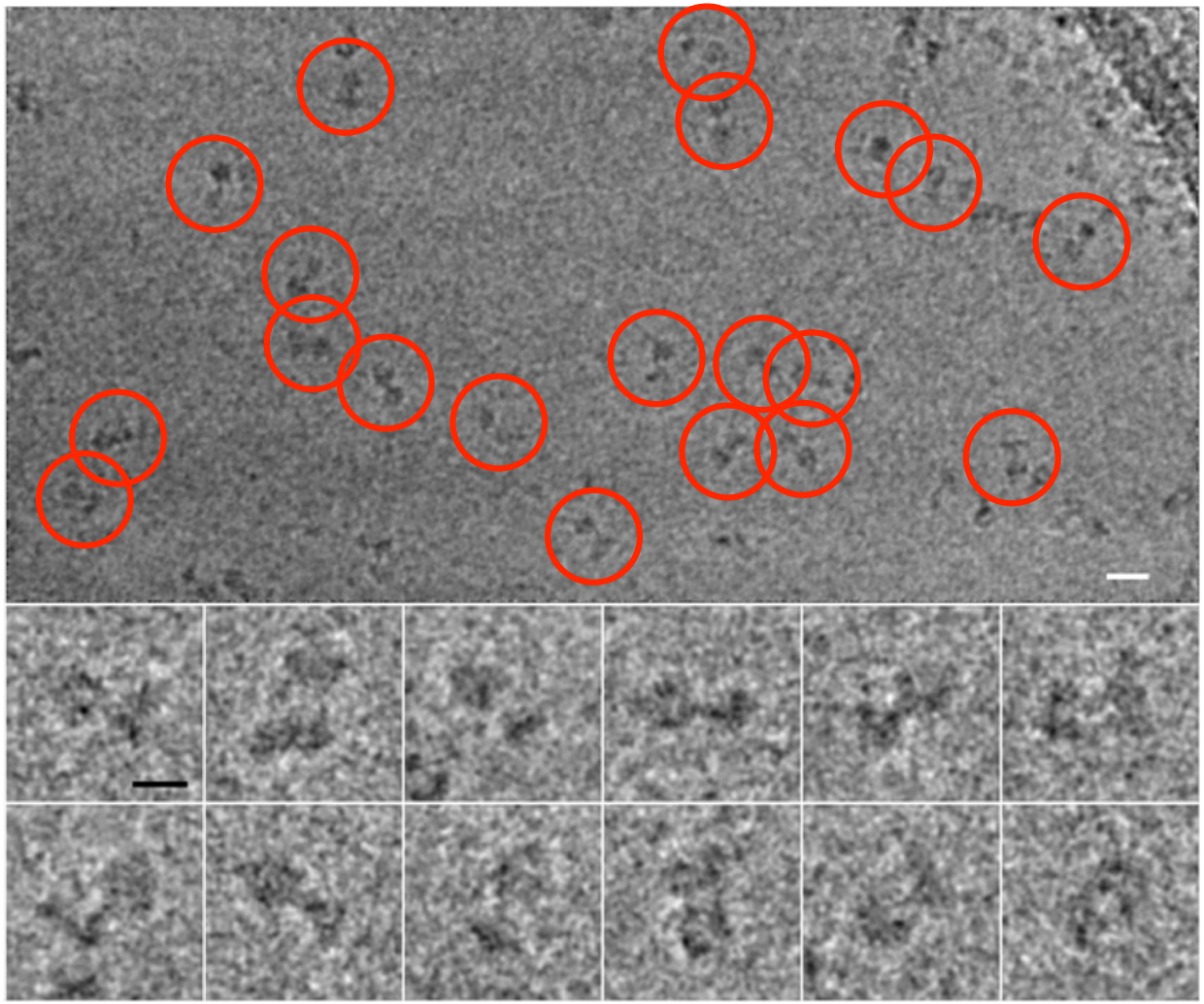
Fig. 2. (*A–F*) Six individual windowed images from the stack of 423 that was supplied by the authors (21). (*G*) Average of 423 windowed images using the same gray scale as *A–F*. (*H*) Average of 423 windowed images with 15x increased contrast. The density in the central region of *G* and *H* shows the average of the many views used in particle picking and verification. The circle of dark density round the edge of the average, seen more clearly in *H* should not be present in the raw images so must arise from masked projections from the 3D map or model used to extract the particles. (*I–L*) Difference maps obtained by subtraction of sections from the two independent half maps [i.e., maps calculated using only half the data, normally even and odd particles in the stack (18)] supplied by the authors (21). The four panels represent sections at different heights along the spike, viewed from the apex. The differences are confined to a sharply defined region with no gradation into the flat background. This clearly visible and relatively sharp mask serves to constrain any density to the region inside the mask during iterative refinement. Use of masking plus the same initial reference suggests how the apparent resolution was extended from 11 to 6 Å. All images are on the same scale, with a window size of 190 Å.

Henderson's recommendations:

- The experimentalist should be skeptical of their own results and consider all possible problems
- Make sure you can see particles (sufficient exposure and defocus)
- Pick particles by hand initially
- Ask whether different views of the particle can be identified
- If classification must be used *en route* to building an initial 3D, reference-free classification schemes should be employed

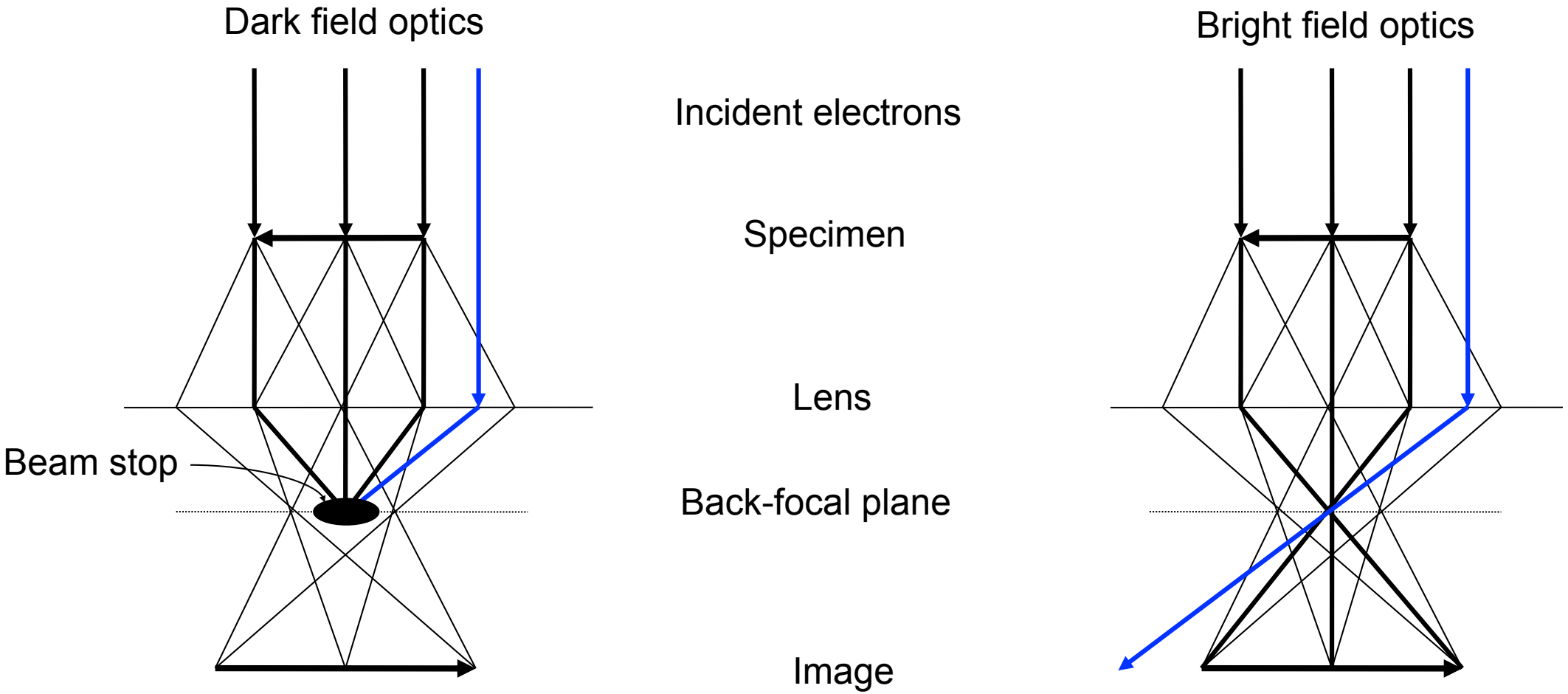
Back to our particle images…

Not just noisy projections of the structure…



The contrast transfer function (CTF)

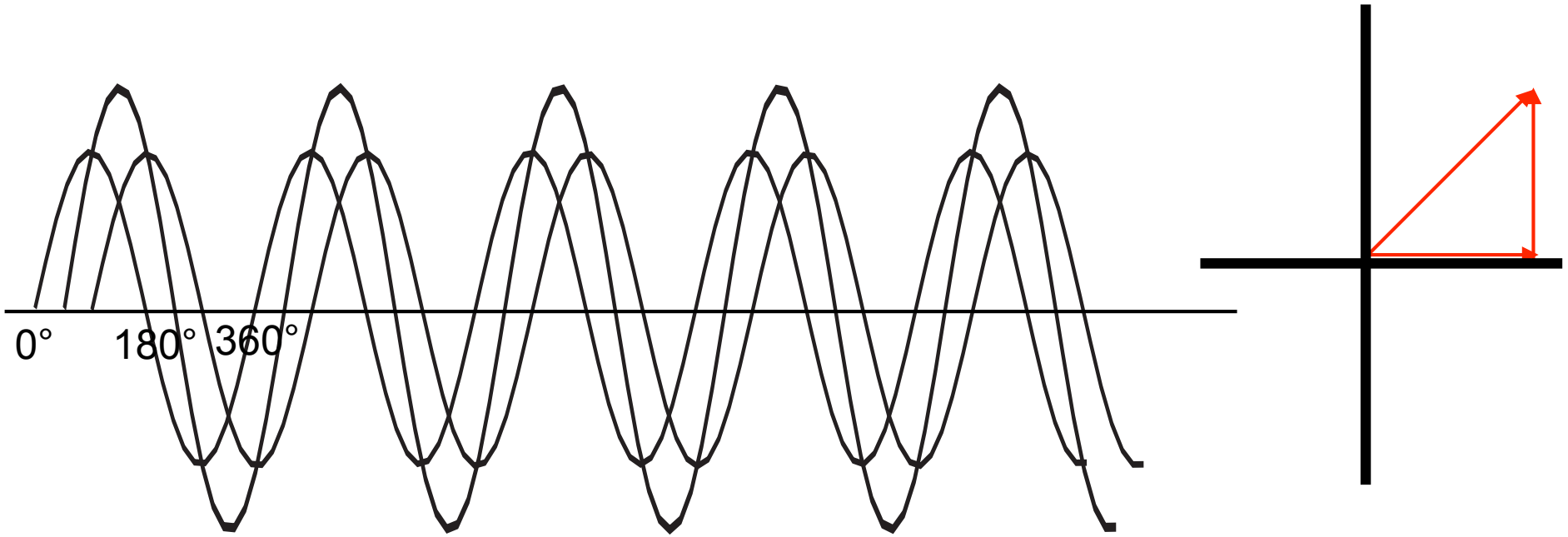
Dark field and bright field TEM



Dense specimen appears as a decrease in the number of electrons that contribute to the image.

Represented in this way, bright field TEM has **no contrast**

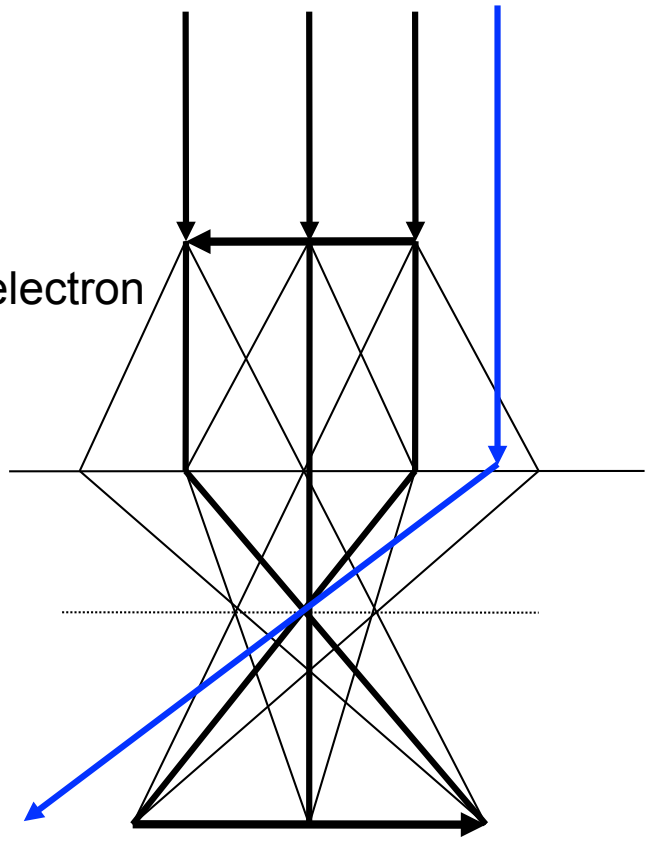
waves of the same wavelength can be added as vectors



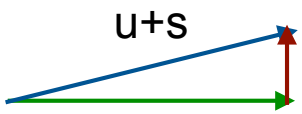
The sum of two waves with frequency *f* is another wave
also with frequency f

Bright field optics

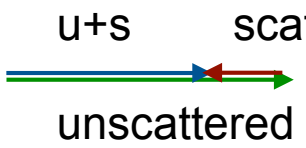
~10% of electron scattered



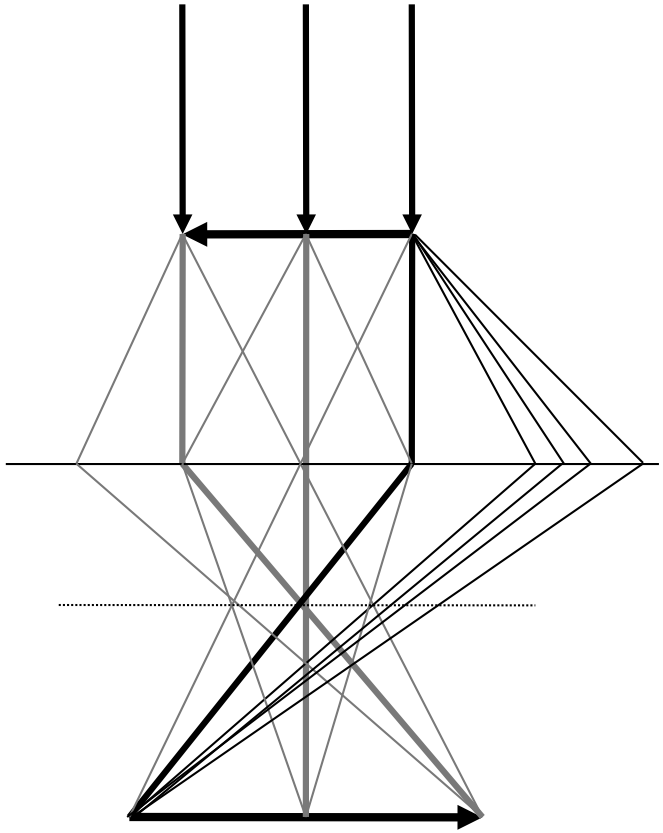
Starting:



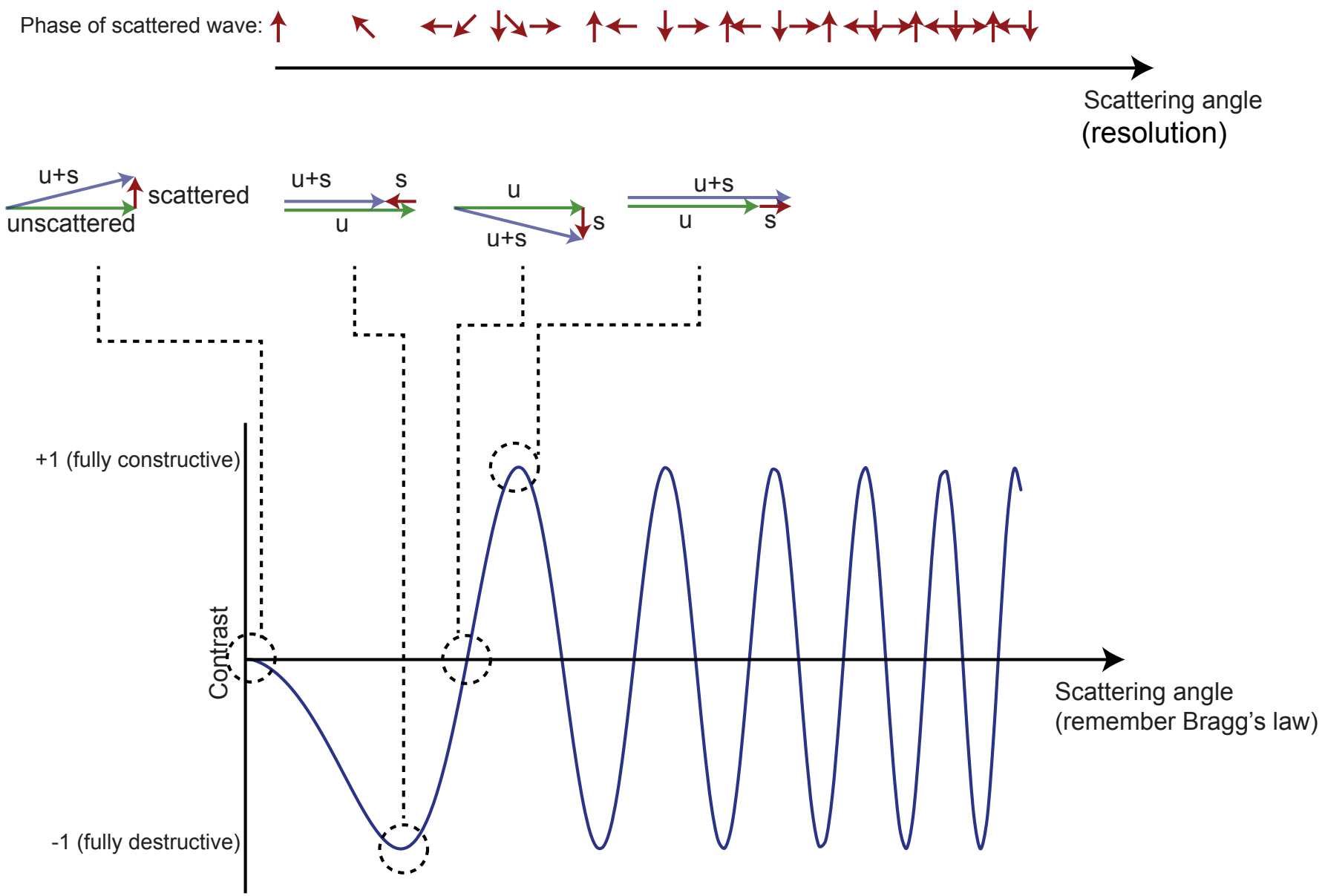
After phase shift: (ideal)



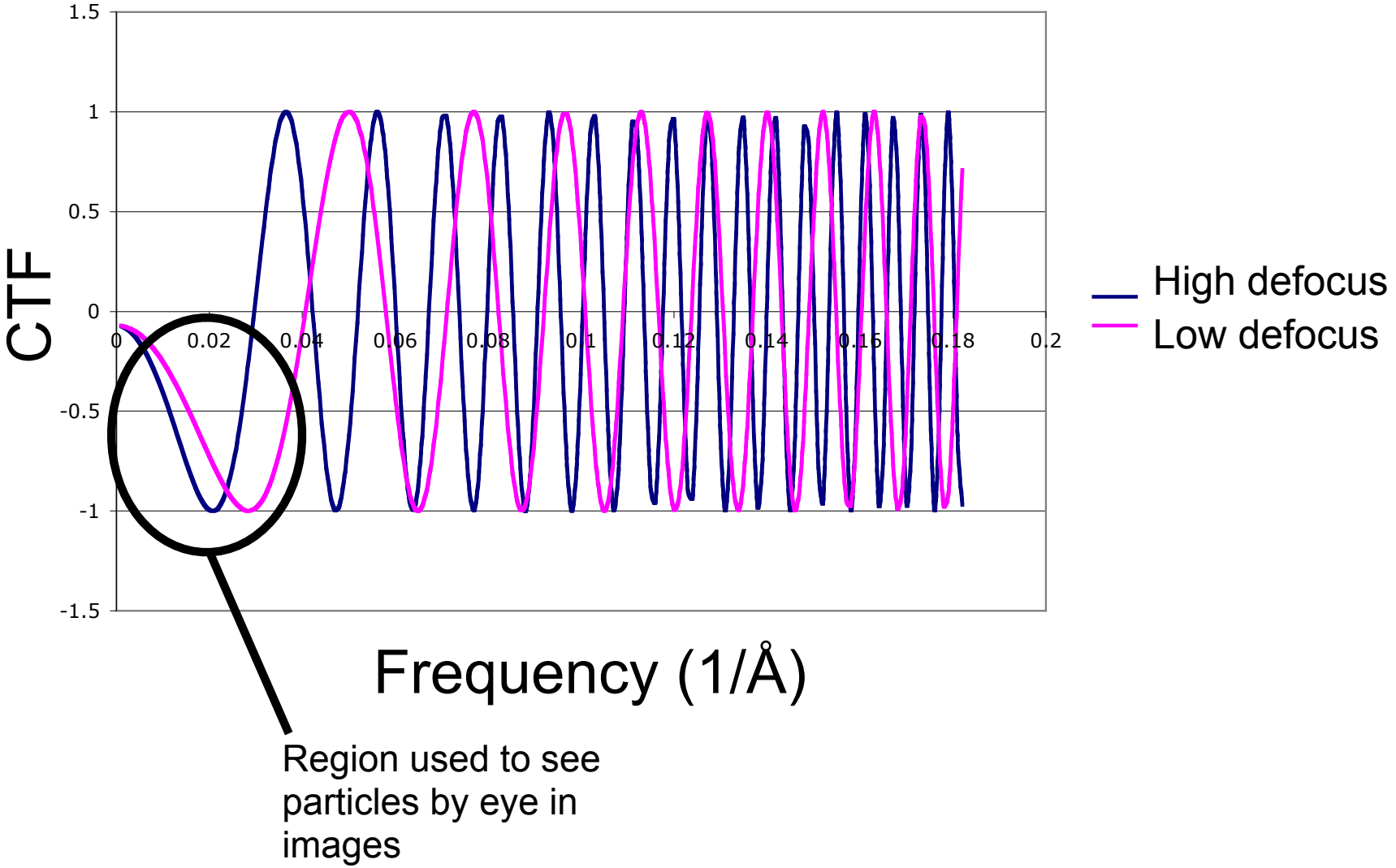
Bright field optics



Defocusing microscope adds a phase shift to the scattered waves… but unfortunately, the phase shift increases with increasing scattering angle!

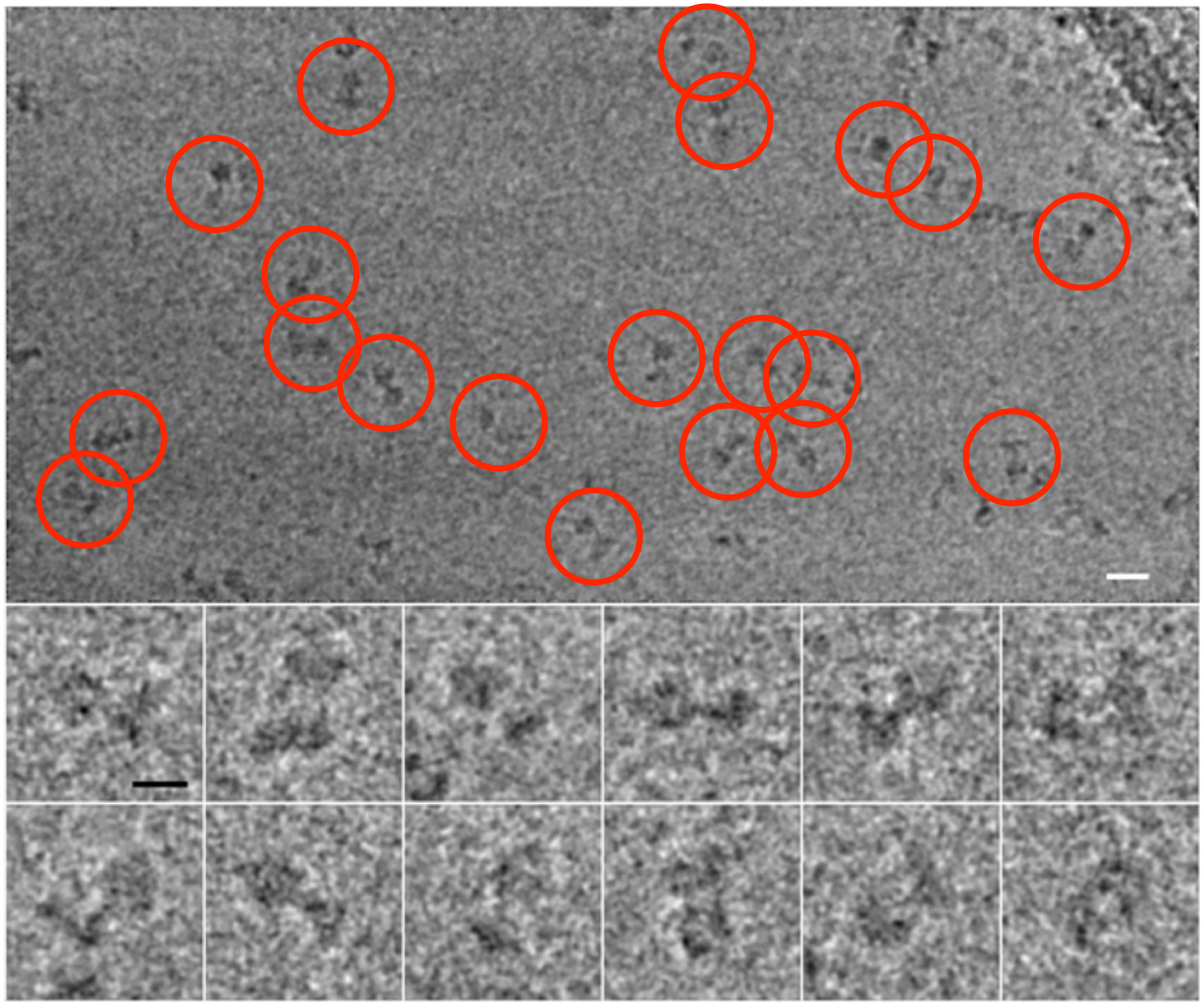

Phase of scattered wave:
Scattering angle
(resolution)
u+s
scattered
unscattered
u+s
s
u
u
u
u+s
u+s
u
s
s
+1 (fully constructive)
Contrast
-1 (fully destructive)
Scattering angle
(remember Bragg's law)

Effects of defocus:



Back to our particle images…

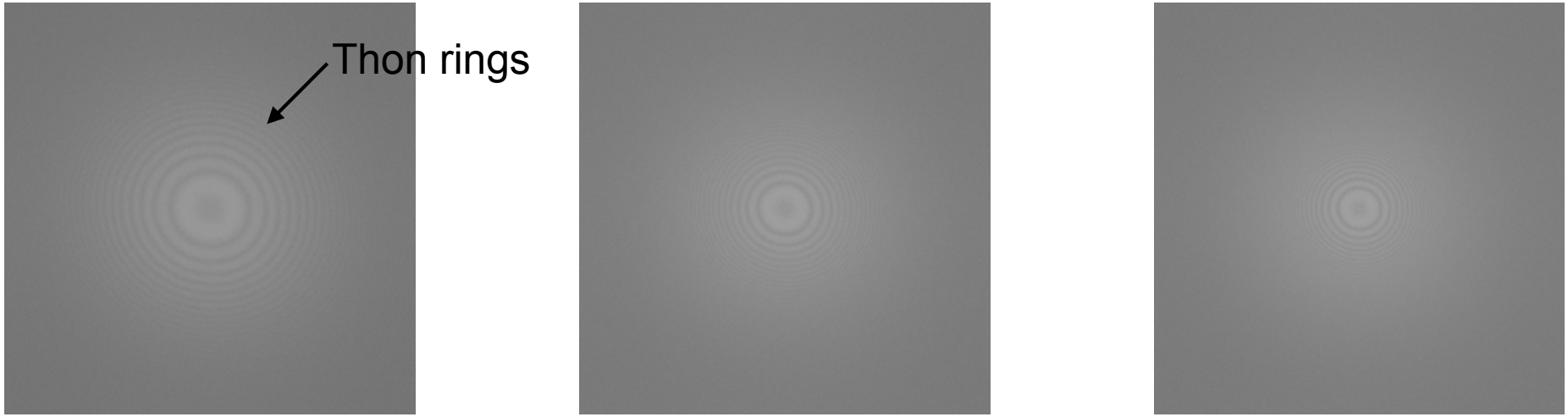
Noisy projections of the 3D structure that have been corrupted by the contrast transfer function



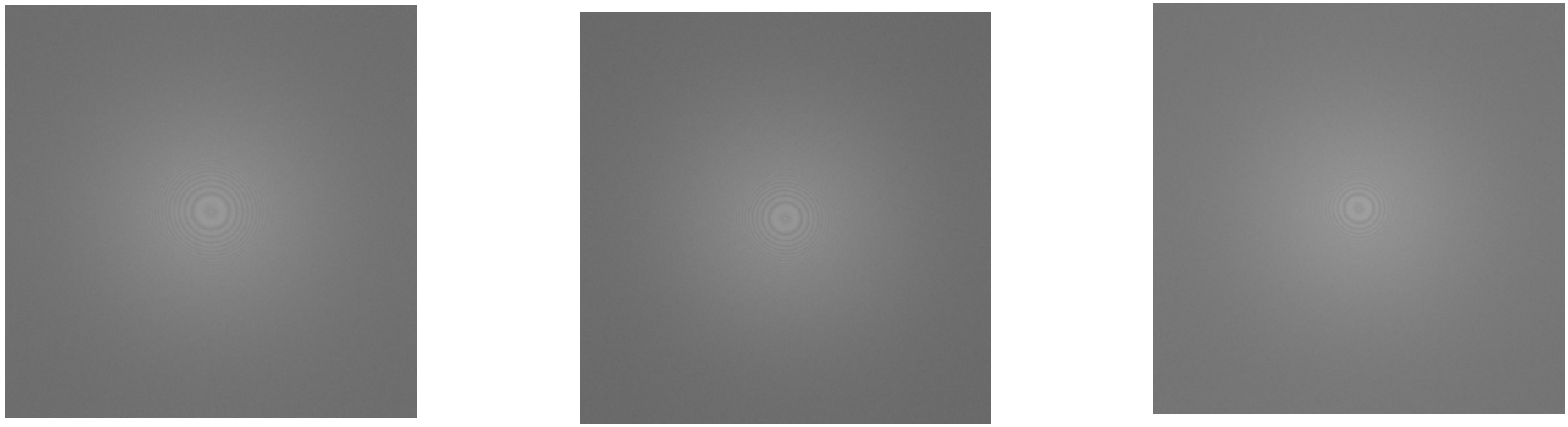
To correct for the contrast transfer function we have to be able to measure it.

Effects of defocus

Square of FT of images



~1 µm ~2 µm ~3 µm



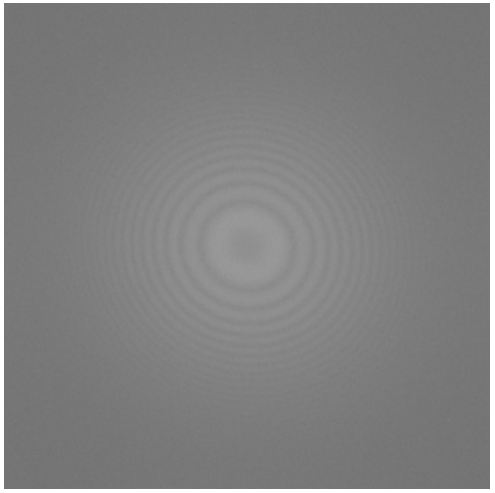
~4 µm ~5 µm ~6 µm

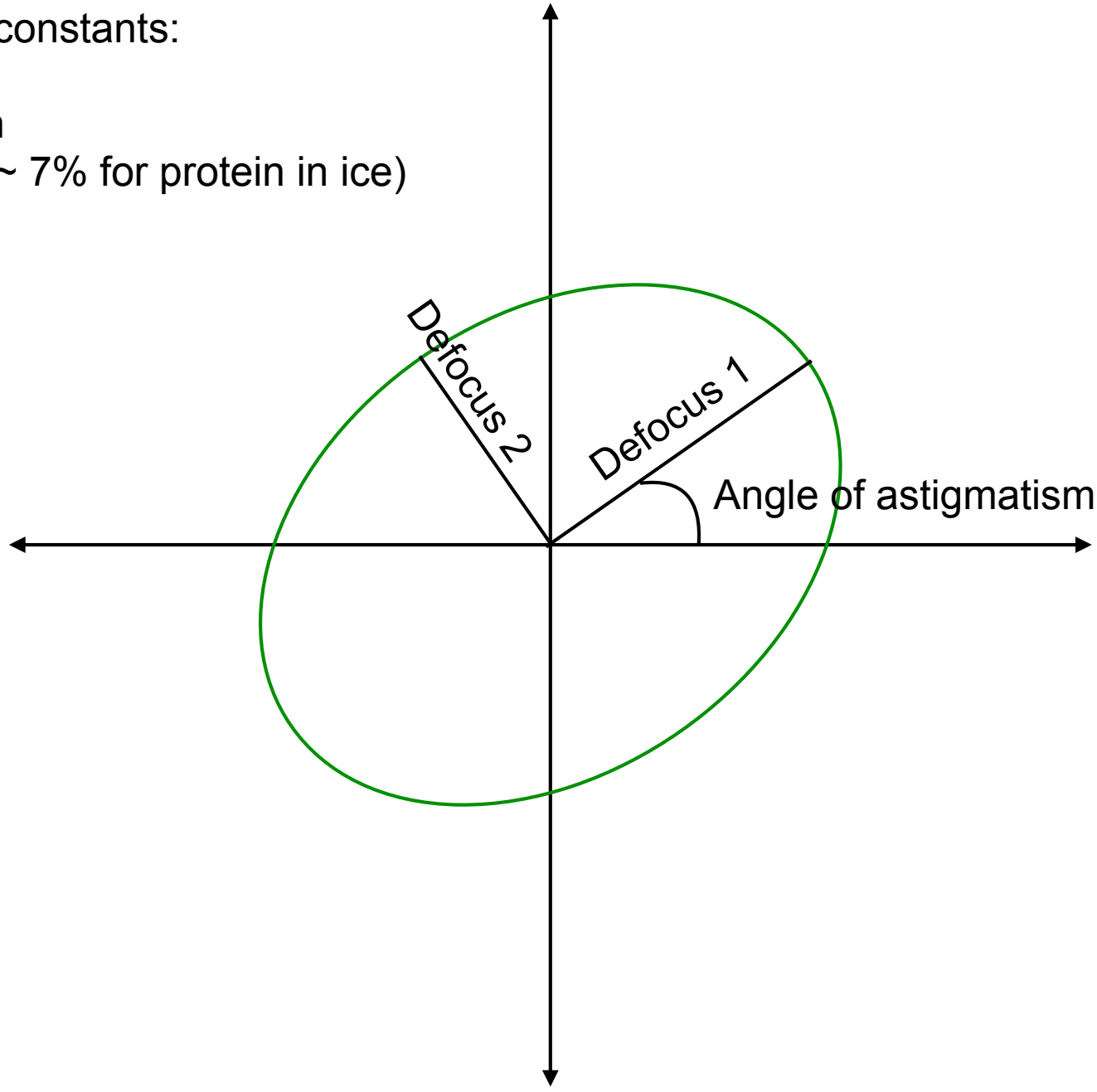
Image: exposure 12 $e^-/Å^2$, 1.4 Å/pixel, 24 e^-/pixel, 200 kV, record on F20 with DE-12
FFT: 2048x2048 pixels, compressed 5x5

Describing the CTF

Microscope/specimen constants:

- Wavelength
- Spherical abberation
- Amplitude contrast (~ 7% for protein in ice)

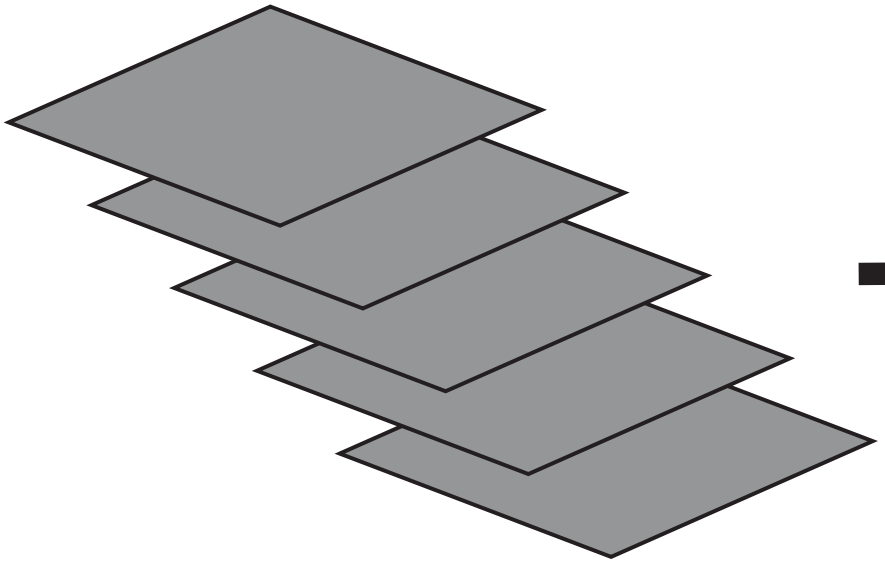


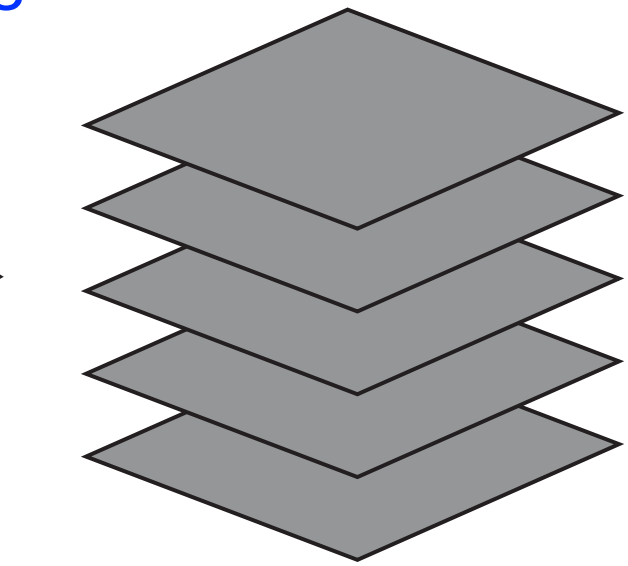


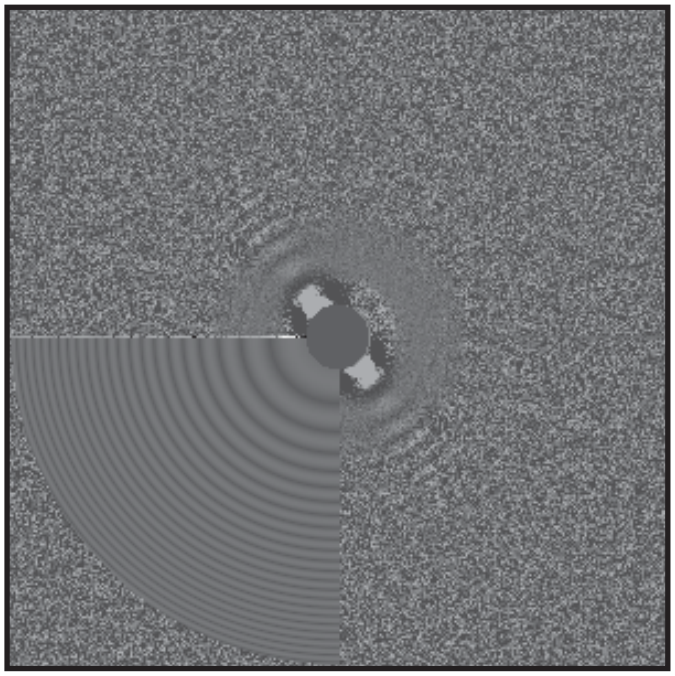
Frame alignment

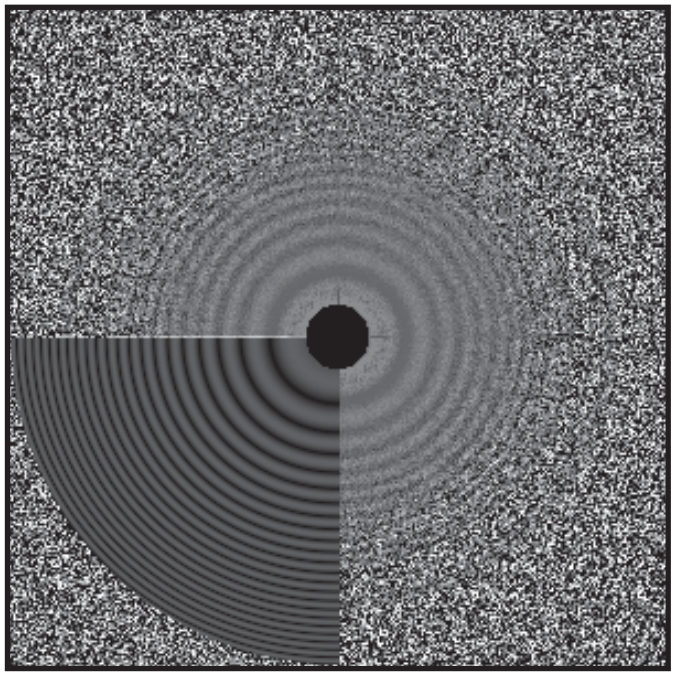
Enabled by the fact that direct electron detectors record movies

Whole frame alignment





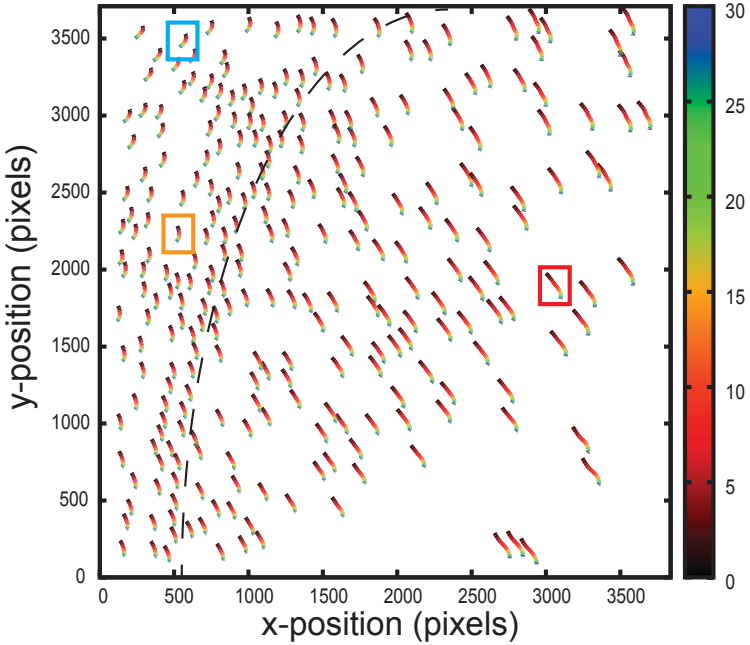




Motioncor (I), Unblur, alignframes_lmbfgs

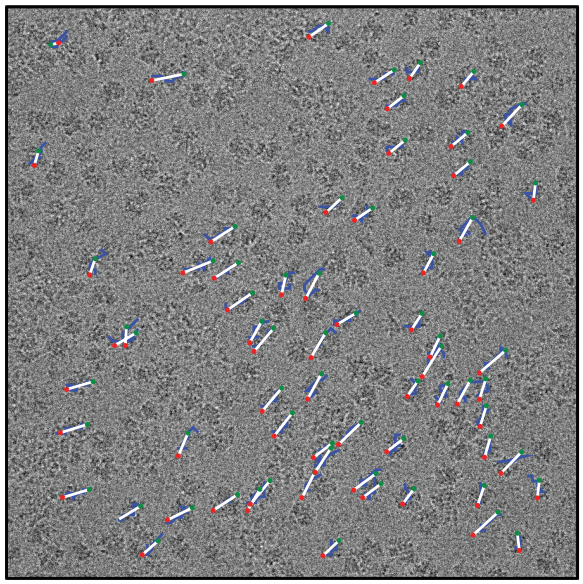
Particle/patch-based motion correction

Compare particle in each frame to sum of frames



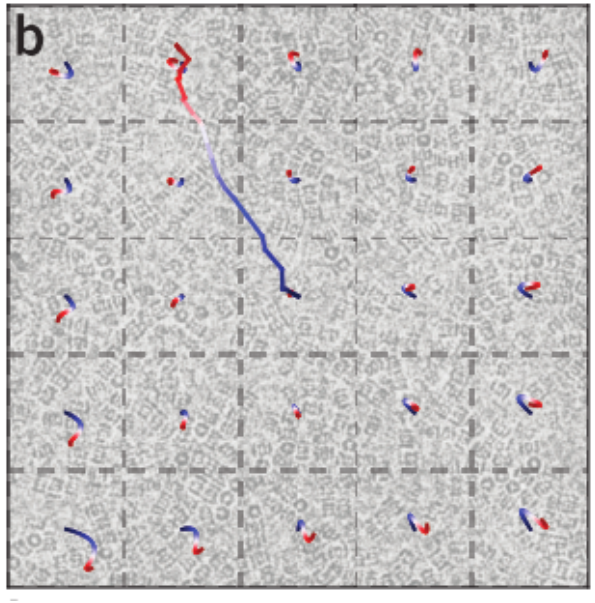
Alignparts_lmbfgs (Rubinstein & Brubaker, 2015, *JSB* 192, 188-95)
[improved version in cryoSPARC ver 2]

Compare particle in each frame to map



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

Compare patch from each frame to sum of frames

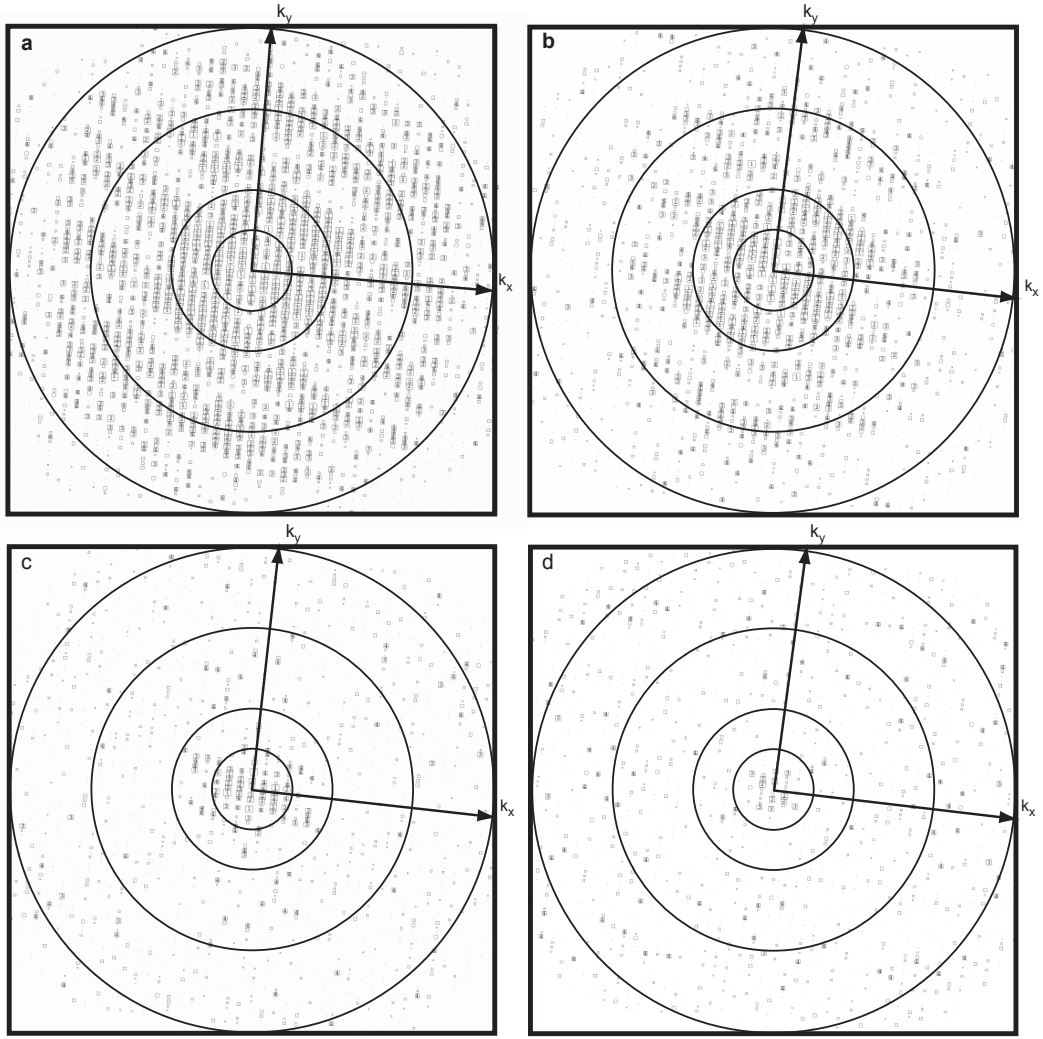


MotionCor2 (Zheng…Agard, 2017, *Nat Meth* 14, 331-2)

Accounting for radiation damage

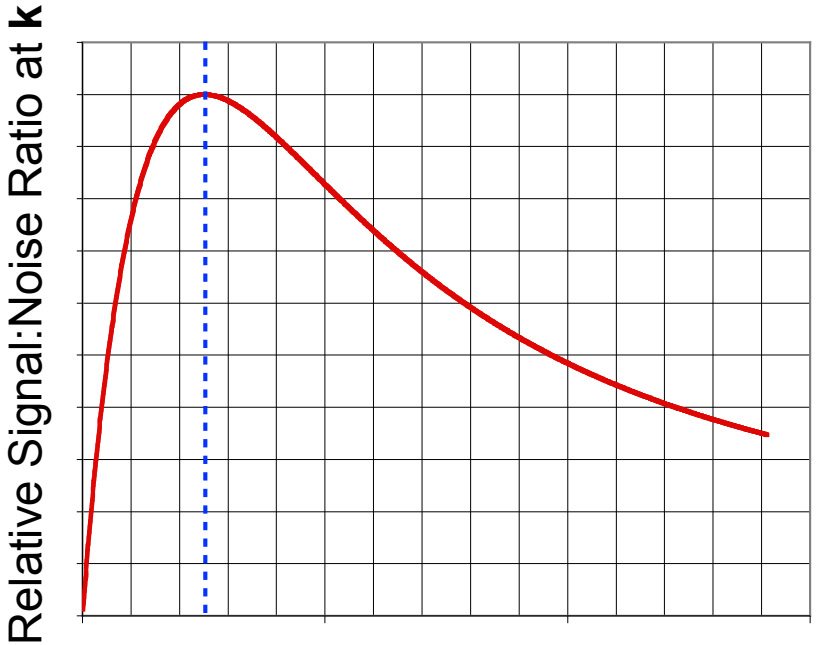
Optimizing signal-to-noise ratios in images

0-8 e$^-$/Å^2 8-16 e$^-$/Å^2



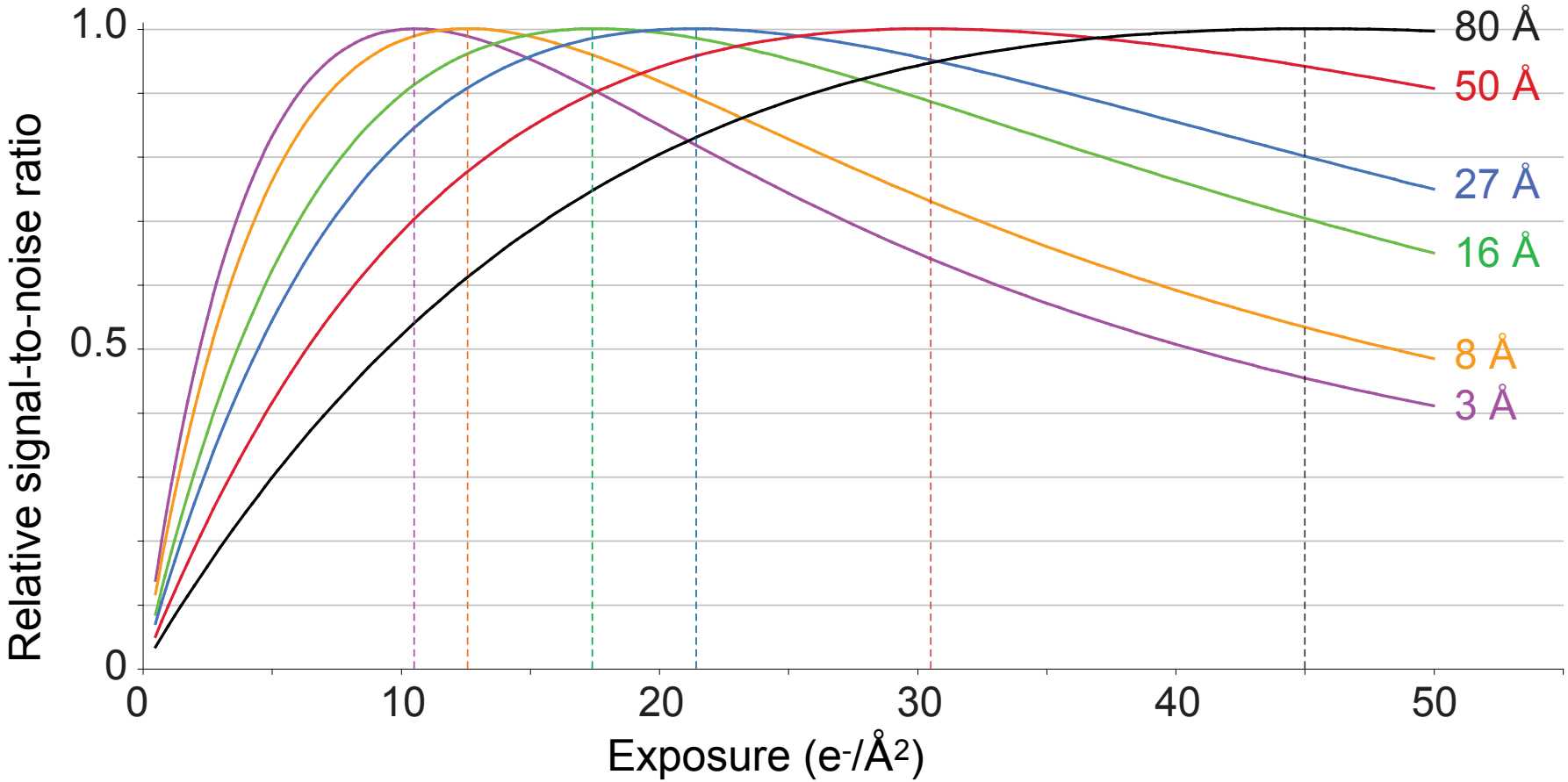
46-54 e$^-$/Å^2 78-96 e$^-$/Å^2

Baker, Smith, Bueler and Rubinstein (2010).
J Struct Biol **169**, 431-7.



Hayward and Glaeser (1979).
Ultramicroscopy **4**, 201-10.

Exposure weighting



Baker, Smith, Bueler, and Rubinstein (2010), *J. Struct. Biol.*, **169**, 431-7.
Baker and Rubinstein (2010), *Method Enzymol* **481**, 373-90.
Curves re-measured in Grant and Grigorieff (2015) eLife 4:e06980.

<u>Programs:</u>

- Alignparts_lmbfgs (cryoSPARC)
- Unblur
- Motioncorr2

<u>Not this approach</u>

- Relion (estimates damage from the data with a 3D reference)

Alignment in 2D

Shifts and a rotation are applied to images in order to bring them into register

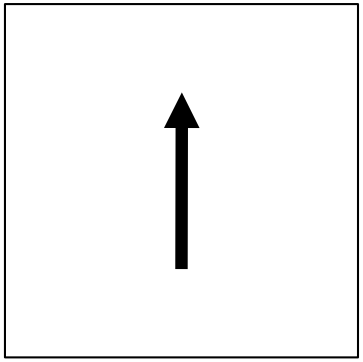


Image 1

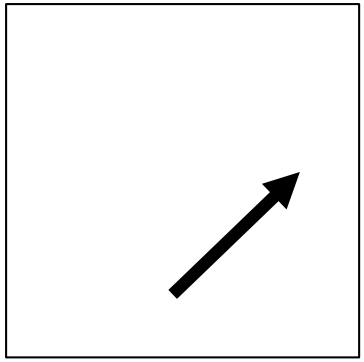
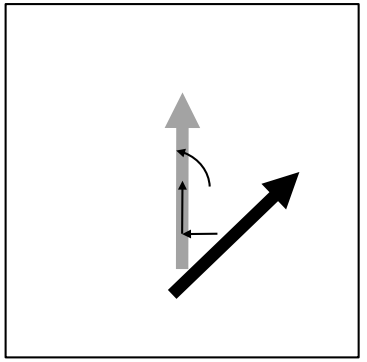


Image 2



Δx, Δy, $\Delta \phi$ required to bring Image 1 into register with image 2

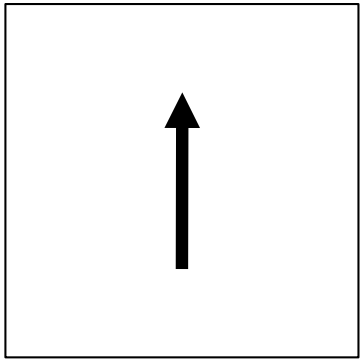
The average image (for 2 or more starting images) is given by:

$Pixel1_{ave}=(Pixel1_{image1}+Pixel1_{image2}\ldots)/N_{images}$

$Pixel2_{ave}=(Pixel2_{image1}+Pixel2_{image2}\ldots)/N_{images}$

$Pixel3_{ave}=(Pixel3_{image1}+Pixel3_{image2}\ldots)/N_{images}$

… up to the number of pixels in the image (e.g. 256x256)



Average of image 1 and 2 after image 2 has been brought into register.

Alignments - OLD SCHOOL; For illustration only

Optimum alignment of image with reference obtained by finding the maximum in a cross-correlation function

For two discretely sampled images f_1 and f_2 each with J pixels, their cross-correlation coefficient is given by

$$CC_{12} = \frac{\sum_{j=1}^{J} f_1(j) f_2(j)}{\sqrt{\sum_{j=1}^{J} \left(f_1(j)\right)^2 \sum_{j=1}^{J} \left(f_2(j)\right)^2}}$$

What is the cross correlation?

0	0	0
0	1	0
0	0	0

0	1	0
0	0	0
0	0	0

0

0	0	0
0	1	0
0	0	0

0	0	0
0	1	0
0	0	0

1

0	0	0
0	1	0
0	0	0

0	0	0
0	-1	0
0	0	0

-1

Cross correlation ***function***

The cross-correlation function describes the CC coefficient for all possible shifts in *x* and *y*

Reference

0	0	0
0	1	0
0	0	0

Image being aligned

0	0	0
0	0	1
0	0	0

Image after shifting Δx=-1

0	0	0
0	1	0
0	0	0

CCF_{12}

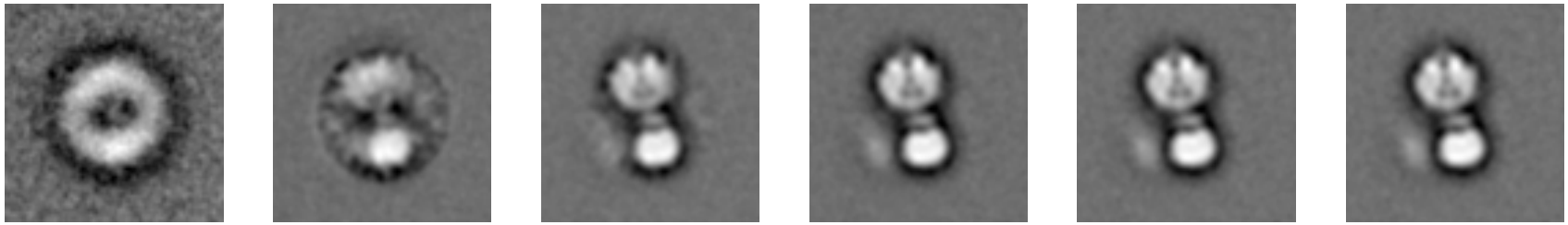
			Δy:
0	0	0	+1
1	0	0	0
0	0	0	-1

Δx: -1 0 +1

There is another handy Fourier space theorem that says that the cross-correlation function between image 1 and image 2 is given by

$$CCF_{12} = FT^{-1}\left(FT[f_1(x,y)]FT^{*}[f_2(x,y)]\right)$$

There are numerous algorithms for aligning large datasets.
e.g. 1: Pick a “typical” image and align everything to it
e.g. 2: Average everything together, align everything to the resulting image, average everything again, iterate



Itn #1 Itn #2 Itn #3 Itn #4 Itn #5 Itn #6

Simple 2-D alignment makes the assumption that all of the molecules are in the same orientation on the EM grid and have the same conformation. We can remove the assumption by classifying the images before averaging them.

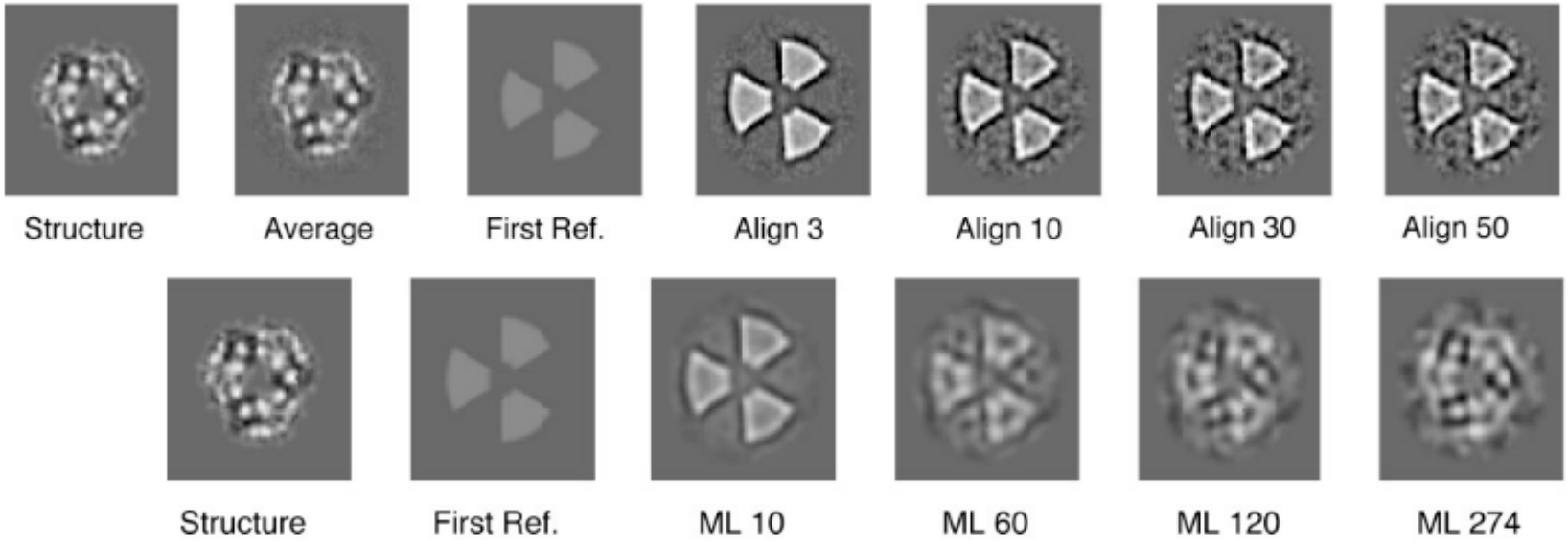
JOURNAL OF STRUCTURAL BIOLOGY **122,** 328–339 (1998)
ARTICLE NO. SB984014

A Maximum-Likelihood Approach to Single-Particle Image Refinement

F. J. Sigworth

Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06520-8026

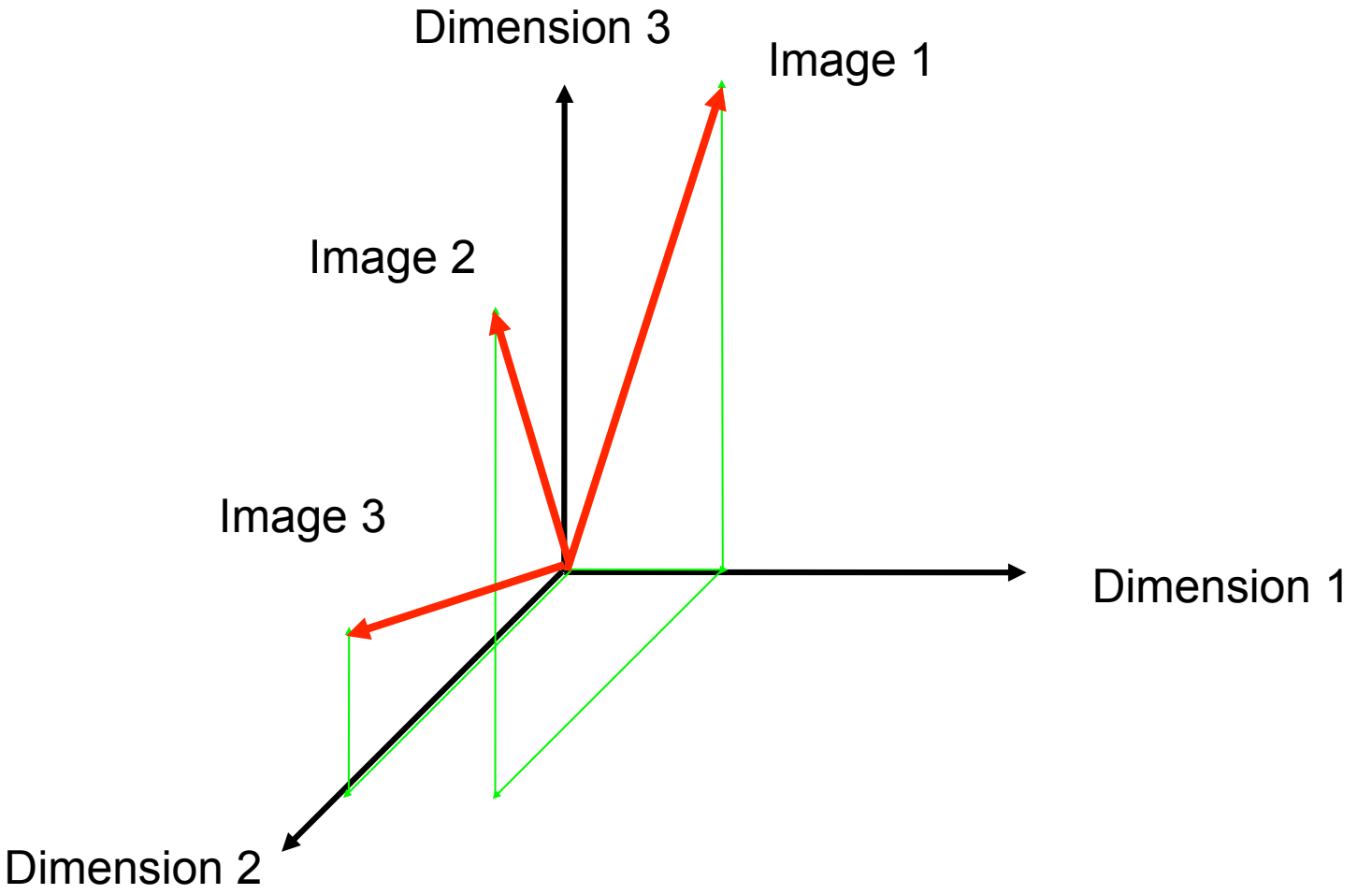
Received April 9, 1998



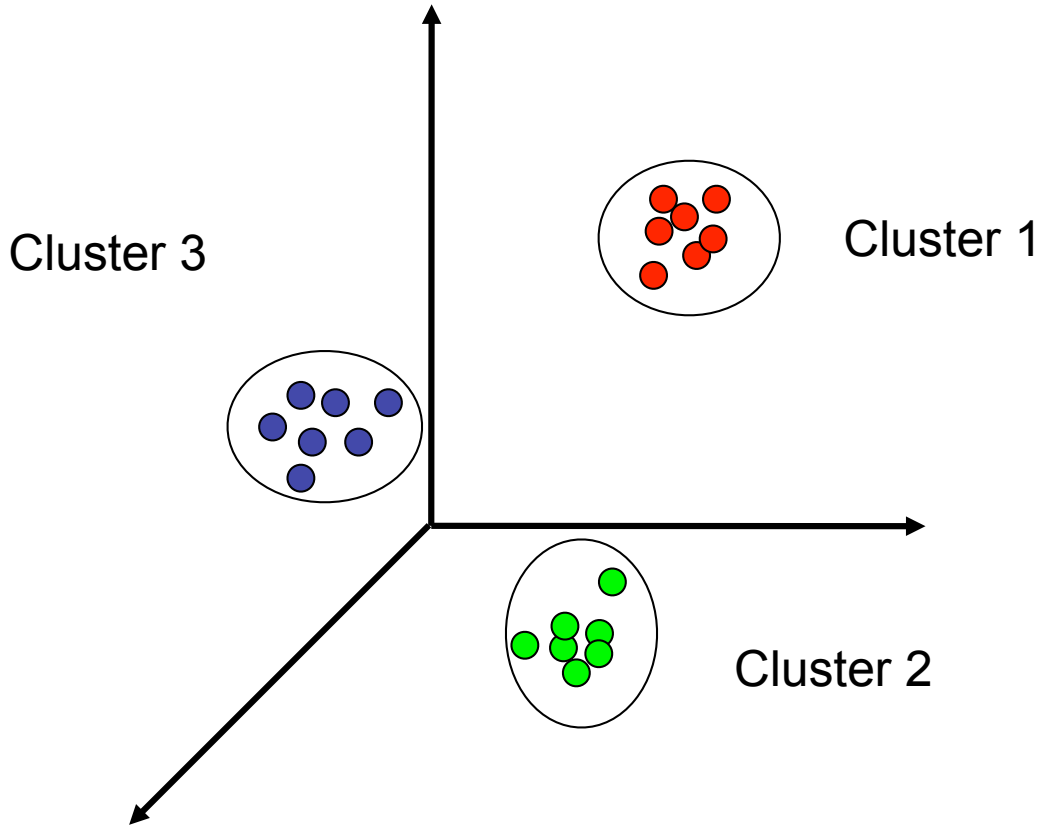
Classification (2D)

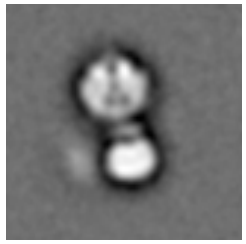
Multivariate statistical analysis - The old school way

An image of N pixels may be represented as a point (or vector) in N-dimensional space



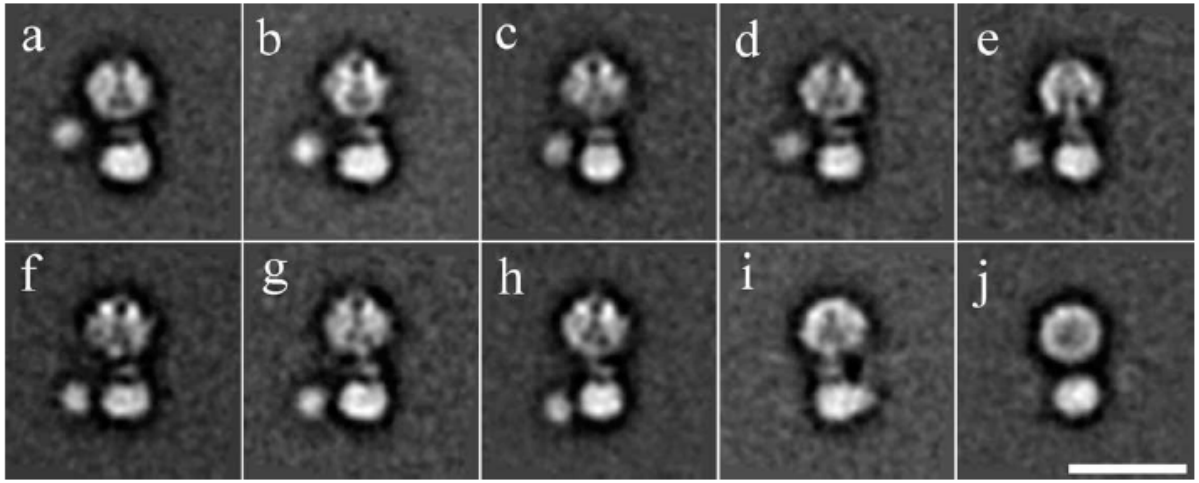
Determine which points are clustered together in N-dimensional space with the assumption that clustered images are similar but separated by noise. Average images in cluster to create a 'class average image'



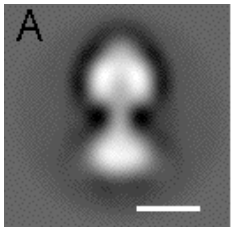


Average of aligned but flexible molecules
(stained ATP synthase particles labeled with avidin at the C terminus of a subunit)

Divided into 10 classes and the class-members averaged

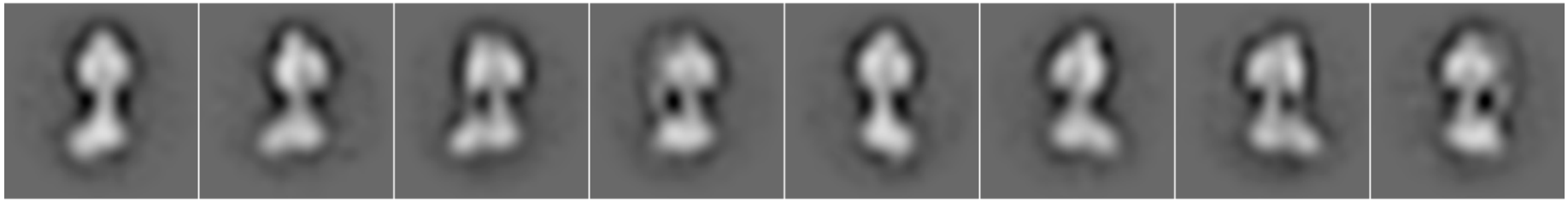


Here, the class-averages represent different conformations



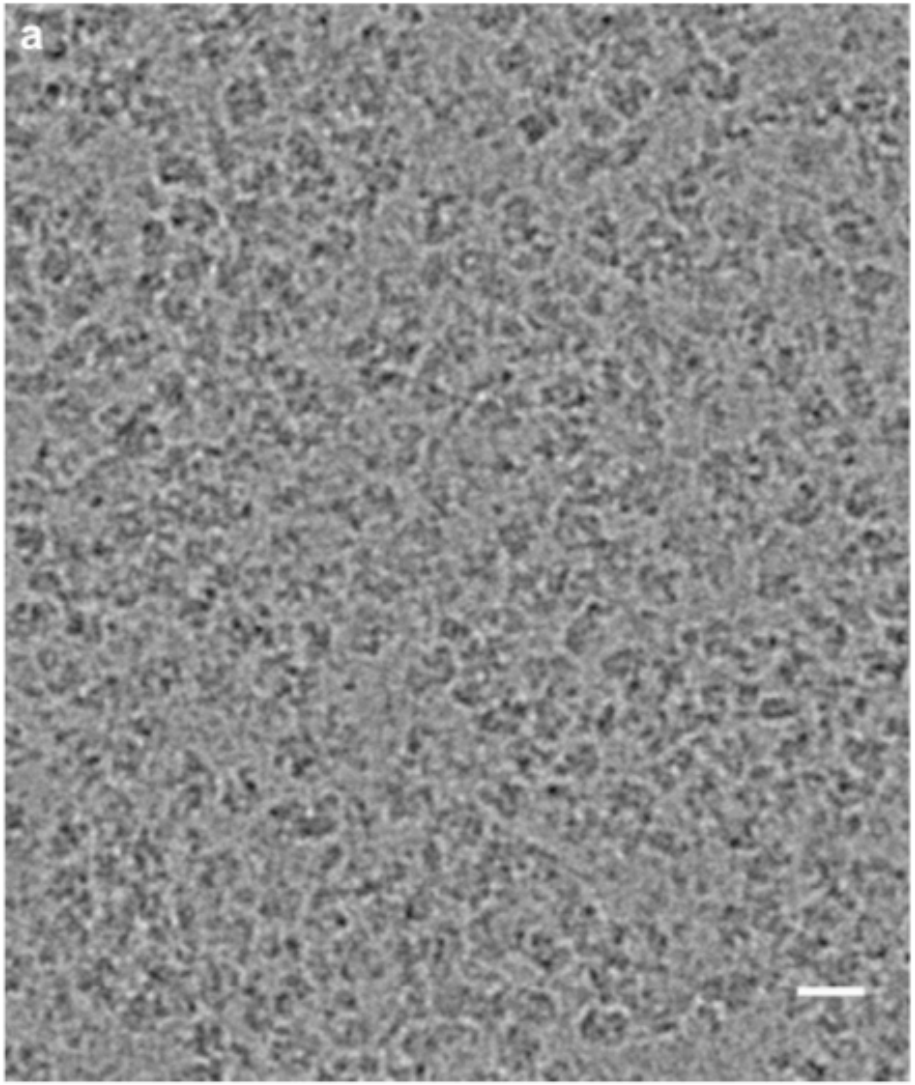
Average of incoherently aligned molecules
(cryo-EM images of bovine ATP synthase)

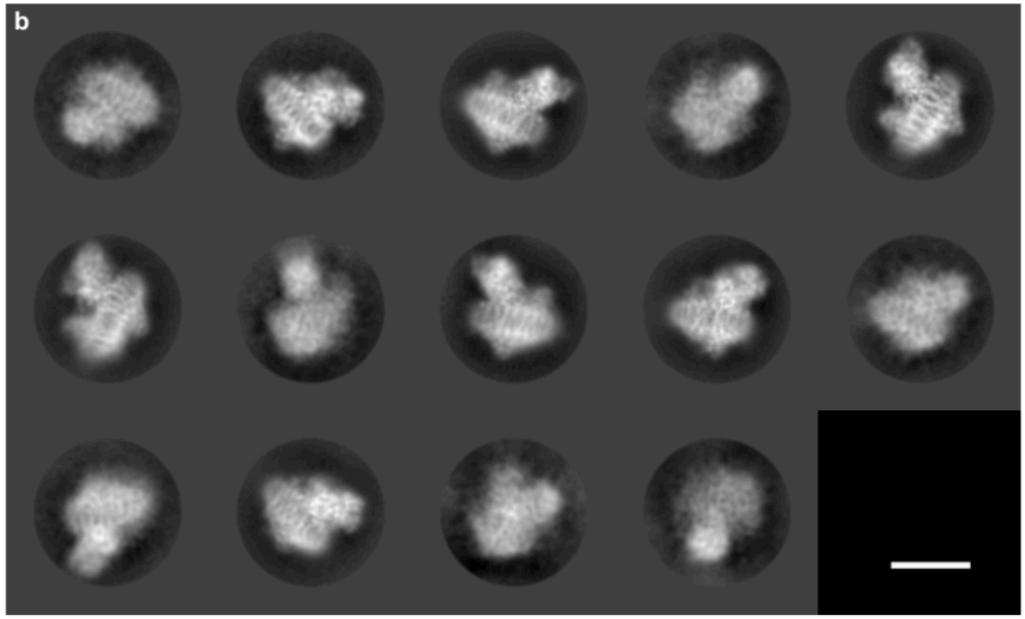
Divided into classes, class-members averaged and classes that represent views of the molecule identified



Here, the class-averages represent different views

2D classification to remove ‘junk particles’:

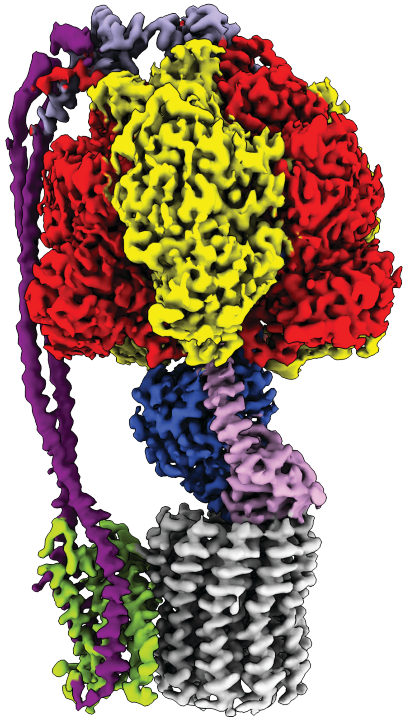




Sun, Benlekbir, Venkatakrishnan, Wang, Hong, Hosler, Tajkhorshid, Rubinstein, and Gennis (2018). Nature 557, 123-6.

Extending to 3D

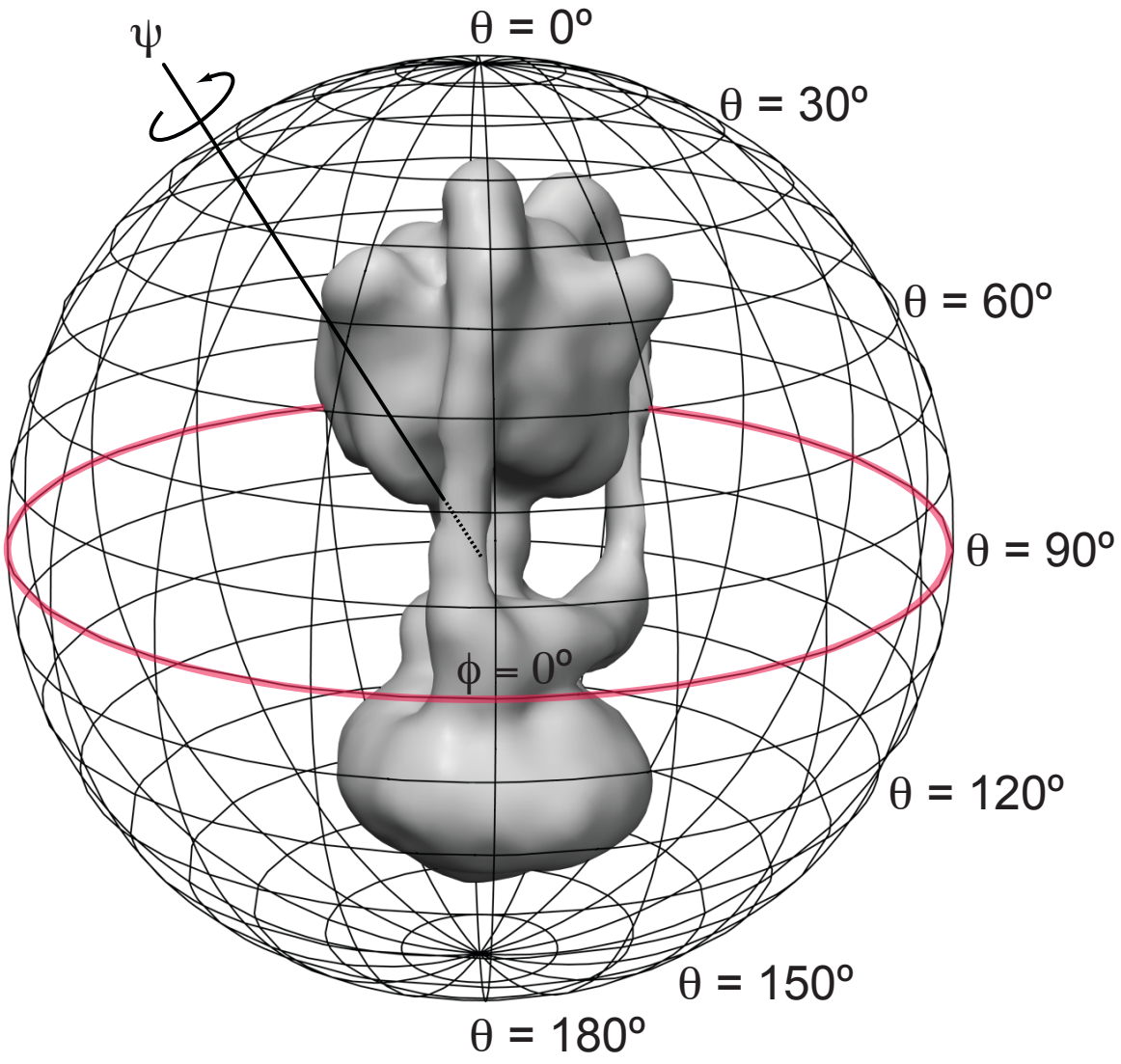
How many parameters do we need to describe a 'view' of this complex?

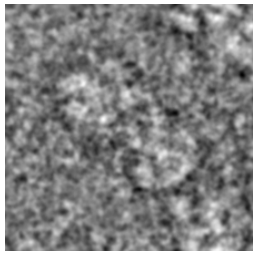


- 2 translations (Δx and Δy).
- 3 rotations.

1) A rotation about the *z*-axis by an angle ϕ
2) A rotation about the *new y*-axis (the one produced by the first rotation) by an angle θ
3) A rotation about the *new* z-axis (the one produced by the second rotation) by an angle ψ

Parameters needed to describe a view in 3D





For each particle image

1) Determine three angles (ϕ,θ,ψ)
2) Two shifts (x, y)

The Eulerian rotation convention

The three rotations are expressed in matrix form.

Matrix multiplication review (A, B etc. represent matrices):
AB ≠ BA
ABCD=A(B(CD))

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{vmatrix} b_1 \\ b_2 \\ b_3 \end{vmatrix} = \begin{vmatrix} c_1 \\ c_2 \\ c_3 \end{vmatrix}$$

where

$$c_1=a_{11}b_1+a_{12}b_2+a_{13}b_3$$
$$c_2=a_{21}b_1+a_{22}b_2+a_{23}b_3$$
$$c_3=a_{31}b_1+a_{32}b_2+a_{33}b_3$$

$$R_\psi = \begin{bmatrix} cos\psi & sin\psi & 0 \\ -sin\psi & cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad R_\theta = \begin{bmatrix} cos\theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\theta & 0 & cos\theta \end{bmatrix} \qquad R_\phi = \begin{bmatrix} cos\phi & sin\phi & 0 \\ -sin\phi & cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R = R_\psi R_\theta R_\phi$$

Eulerian rotations

The Eularian rotations are applied to the coordinate system to give a new coordinate system. The molecular does not change orientation and the projections are always generated along the new *z*-axis.

The axes are defined by the unit vectors:

$$\hat{x} \quad \hat{y} \quad \hat{z}$$

$$\begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} \quad \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix} \quad \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix}$$

The new axes are given by

$x' = R_\psi R_\theta R_\phi x$

$y' = R_\psi R_\theta R_\phi y$

$z' = R_\psi R_\theta R_\phi z$

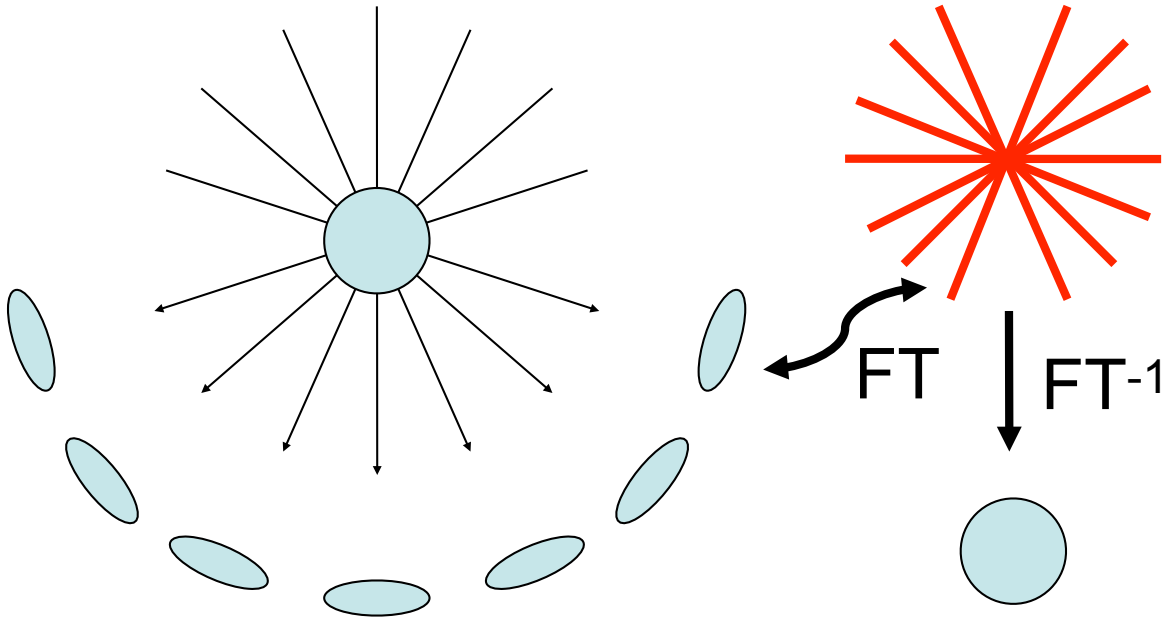
e.g.

$$\hat{x}' = \begin{vmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{vmatrix} \begin{vmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} \hat{x}$$

Building 3-D Models

We know how to build a 3-D model given a sufficient set of views of a molecule:

1) Calculate the 2-D FTs of the projected views
2) Put them together to form the 3-D FT of the specimen
3) Calculate the structure of the specimen by performing a Fourier synthesis of the 3-D FT



Real Space Projections of Structure

But how do we figure out the 3 rotations and 2 shifts?

Methods for Euler angle determination in single particle EM

Random Conical Tilting (RCT)

- Historically was done for proteins that adopt one or a few preferred orientations
- Rarely done now

Common-lines (a.k.a. angular reconstitution in real space)

- Developed (Tony Crowther, 1970) for icosohedral viruses (60-fold symmetry) and worked well for high-symmetry high-signal objects.
- It does not work well (possible not at all) for asymmetric single particles
- Rarely done now

Projection matching (in real or Fourier space)

- Original forms needed a pretty good starting map
- Used for refining low-resolution maps to high resolution
- Stochastic Gradient Descent form now the simplest way to get an initial map

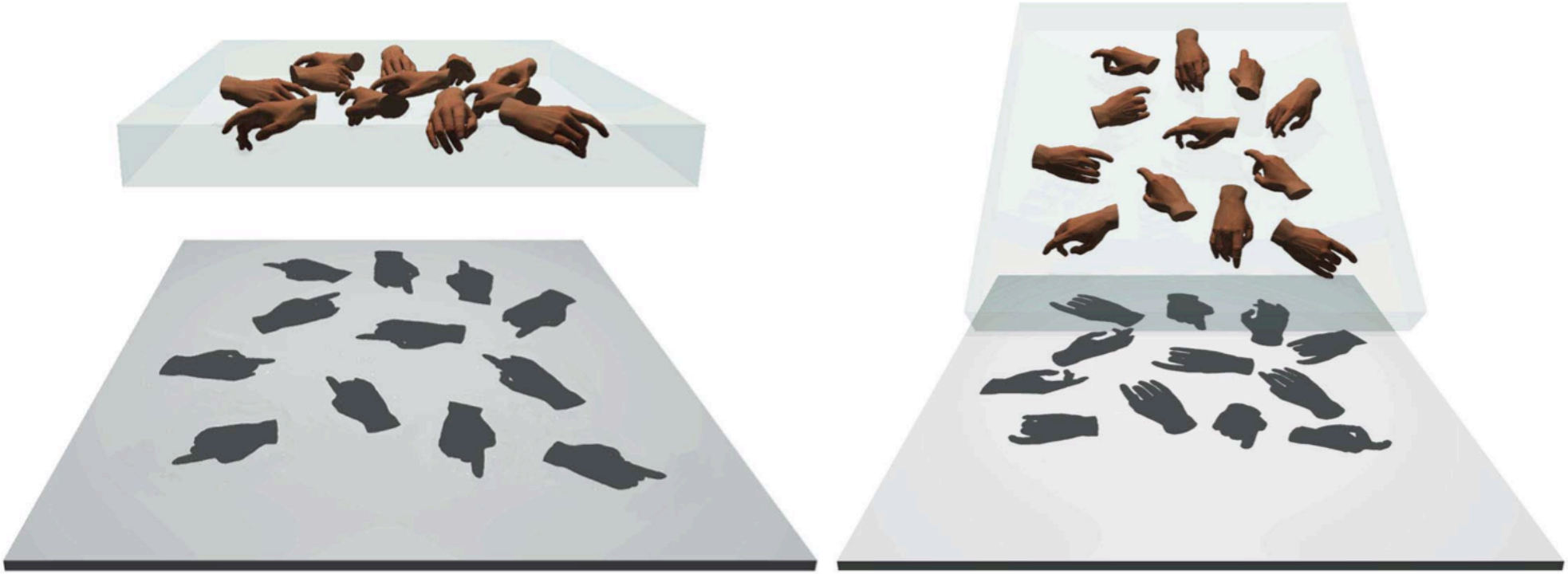
Approach 1: Random Conical Tilting

Procedure

• Obtain a pair of images with a tilt. In practice, the tilted image is obtained first since it is more important for the 3-D reconstruction and the first image suffers less radiation damage

• Determine the azimuthal angle of particles in the untilted image by 2-D alignment methods

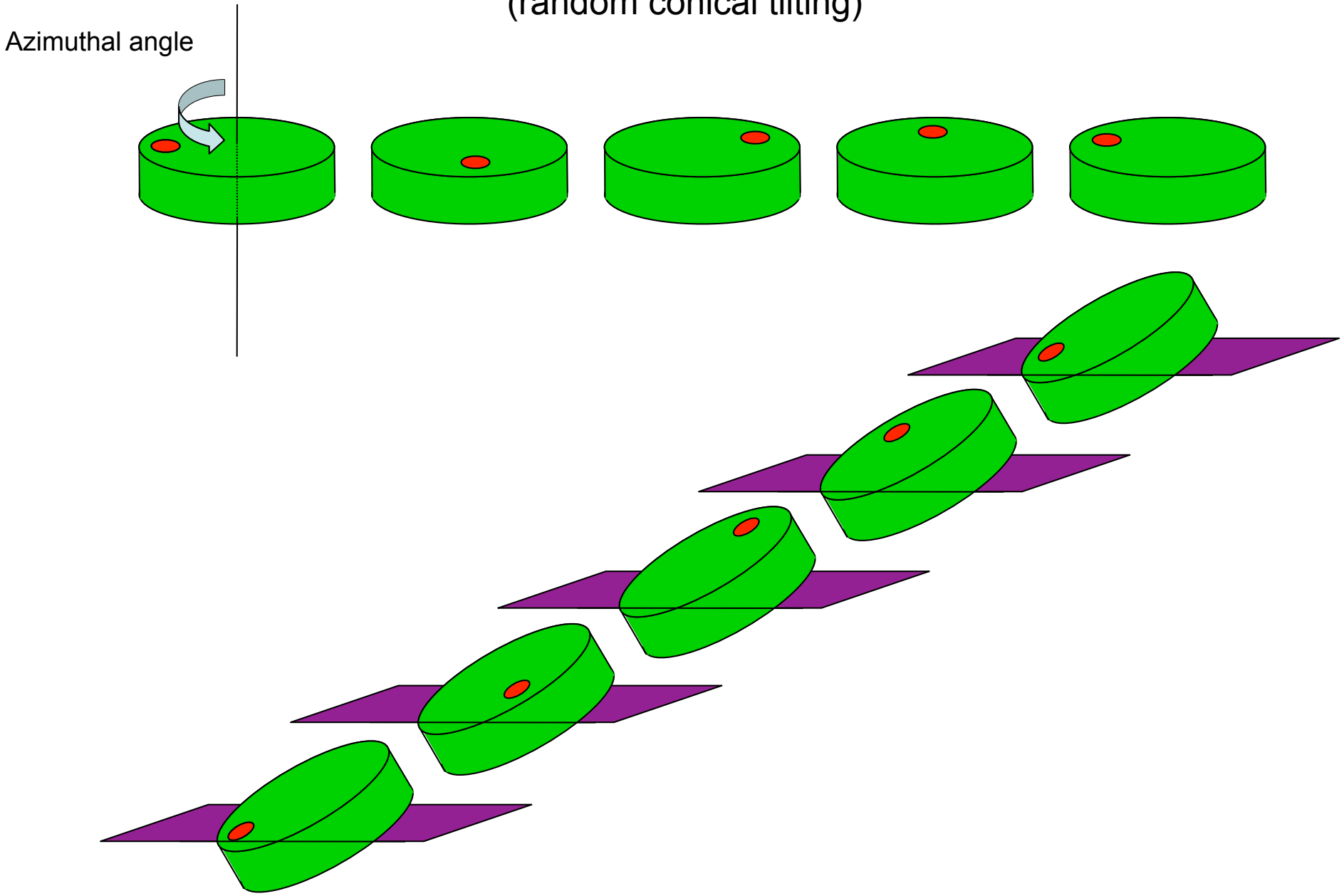
• Calculate the orientation of corresponding tilted views using azimuthal angle of the untilted view and the known tilt of the specimen holder

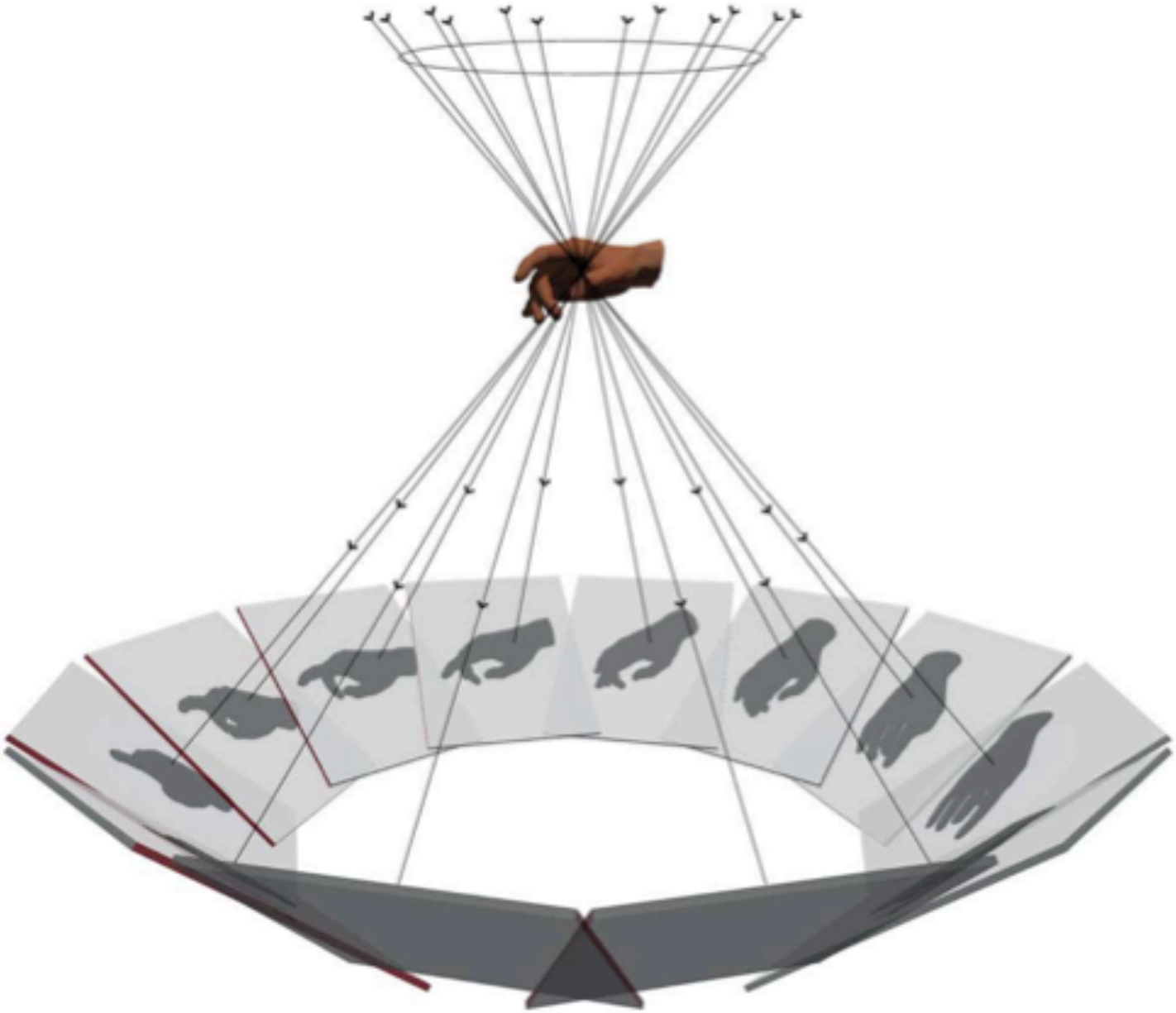


J. Frank, American Scientist 1998

Sampling of Fourier space using tilted, single orientation objects

(random conical tilting)

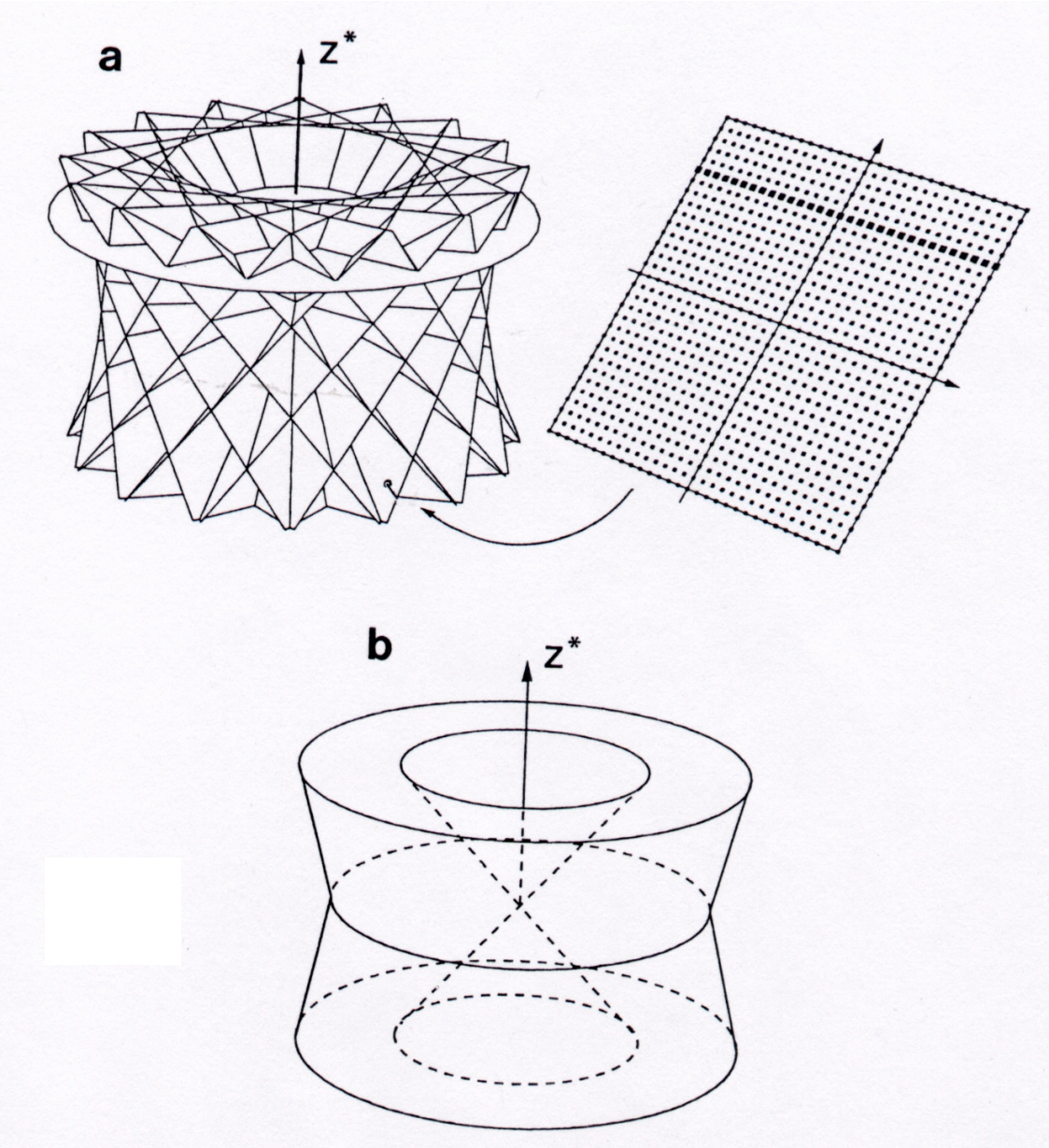




J. Frank, American Scientist 1998

Sampling of Fourier space using tilted, single orientation objects

(random conical tilting)



From: Electron Tomography: Three-Dimensional Imaging With the Transmission Electron Microscope (1992) Frank, J.(ed)

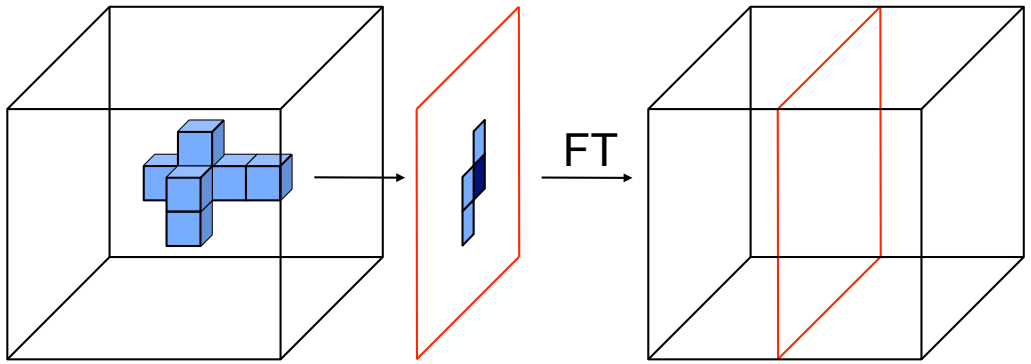
Approach 2: Common Lines

- Works best when particles do not take on any preferred orientations

- A common line (Fourier space) or common direction of projection (real space) can be found for any two images. If these common lines or projections can be found for any three images, the relative orientation of the three images can be determined.

- When signal to noise is low due to lack of symmetry, this approach should be treated with great suspicion.

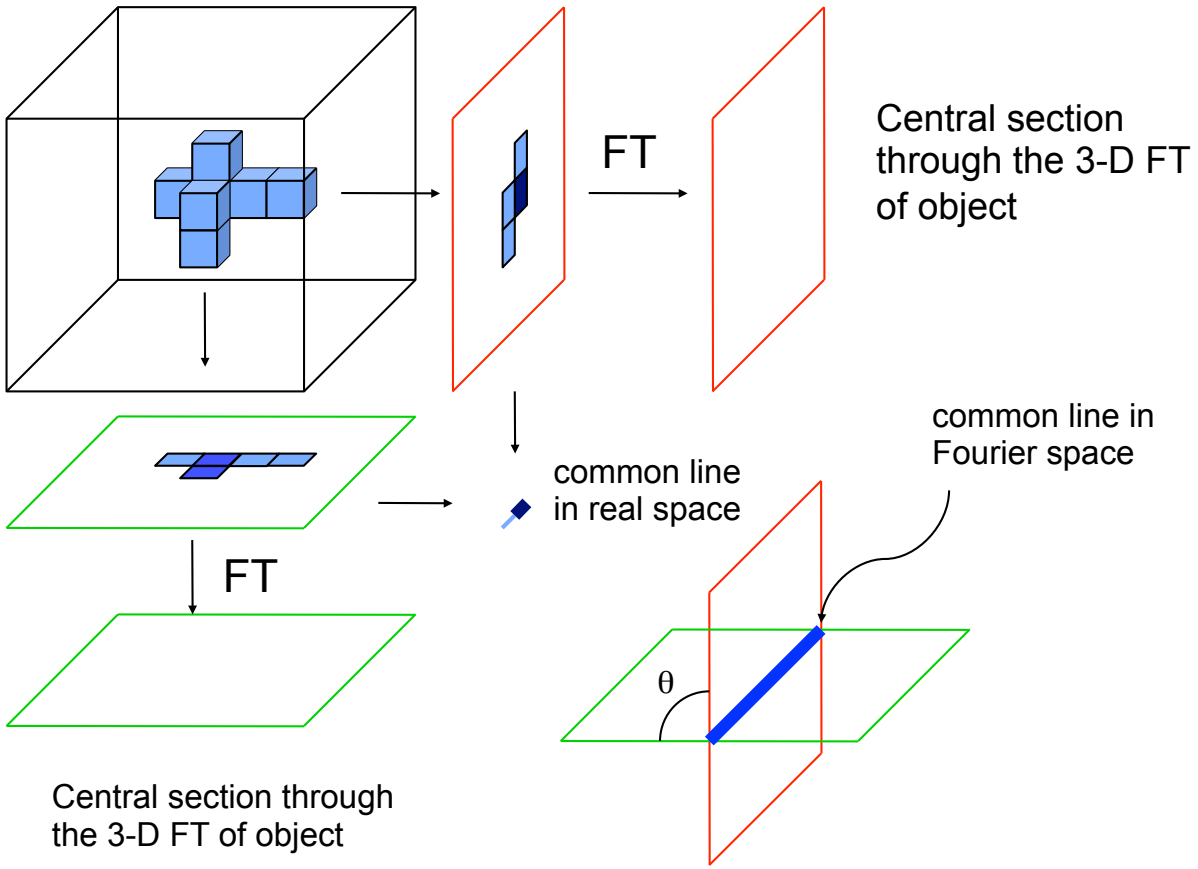
Remember the projection theorem



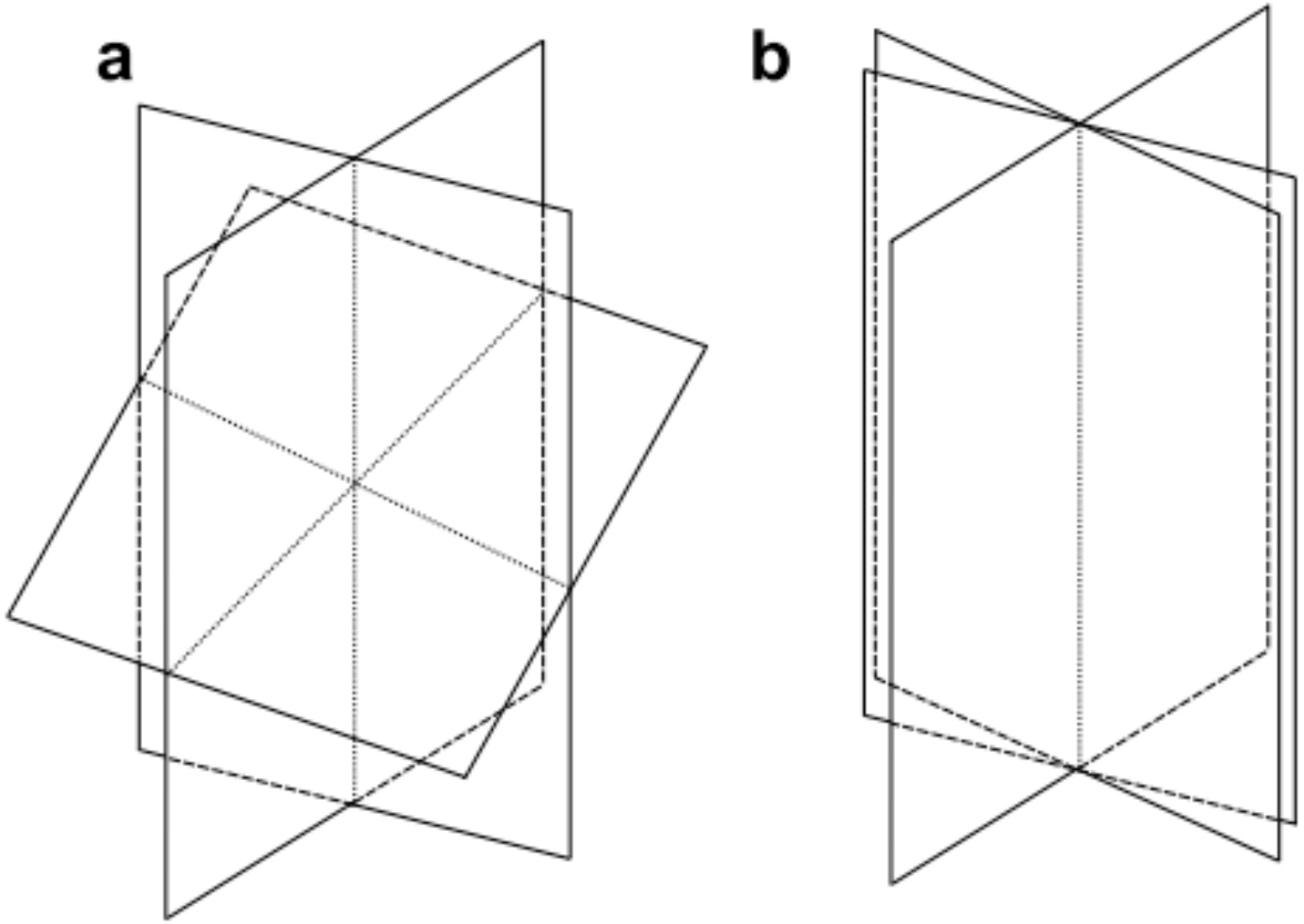
Projection

Central section through the 3-D FT of object

The common lines theorem



Three views are necessary to define orientations with common-lines



Approach 3: Projection matching

Requires a starting 3-D map

Projection matching can be implemented (exactly equivalently) in real space or Fourier space.

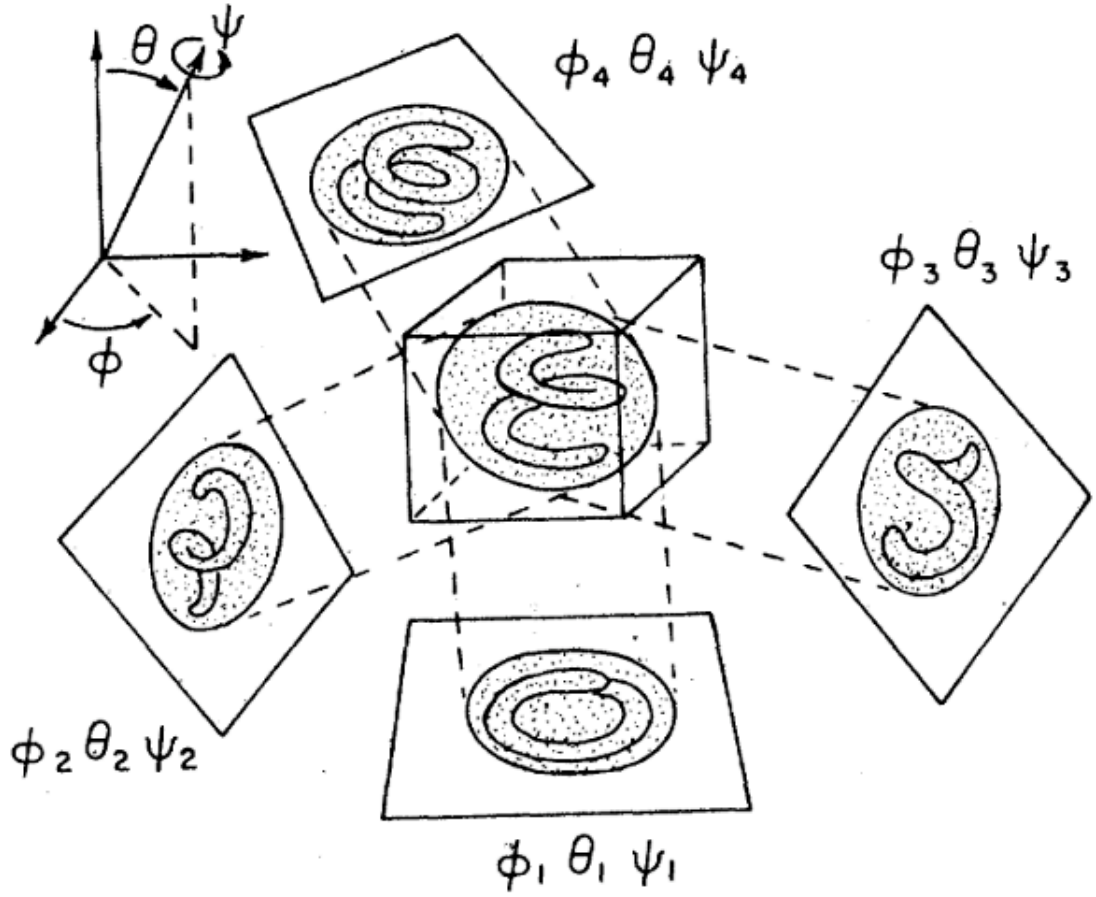
Real space: large number (~200,000?) projections generated from the 3-D map. A particle image is aligned against all of these projections and the best match is chosen. The experimental image is inserted into the map with the chosen angles.

Fourier space: the Fourier transform of a particle image searched against all of the possible central slices through the 3-D Fourier transform of the map. The best match is chosen and particle FT inserted into the map FT.

Nucleosome Reconstruction via Phosphorus Mapping

George Harauz and F. P. Ottensmeyer

Source: *Science*, New Series, Vol. 226, No. 4677 (Nov. 23, 1984), pp. 936-940



Angular Refinement by Projection Matching: real space

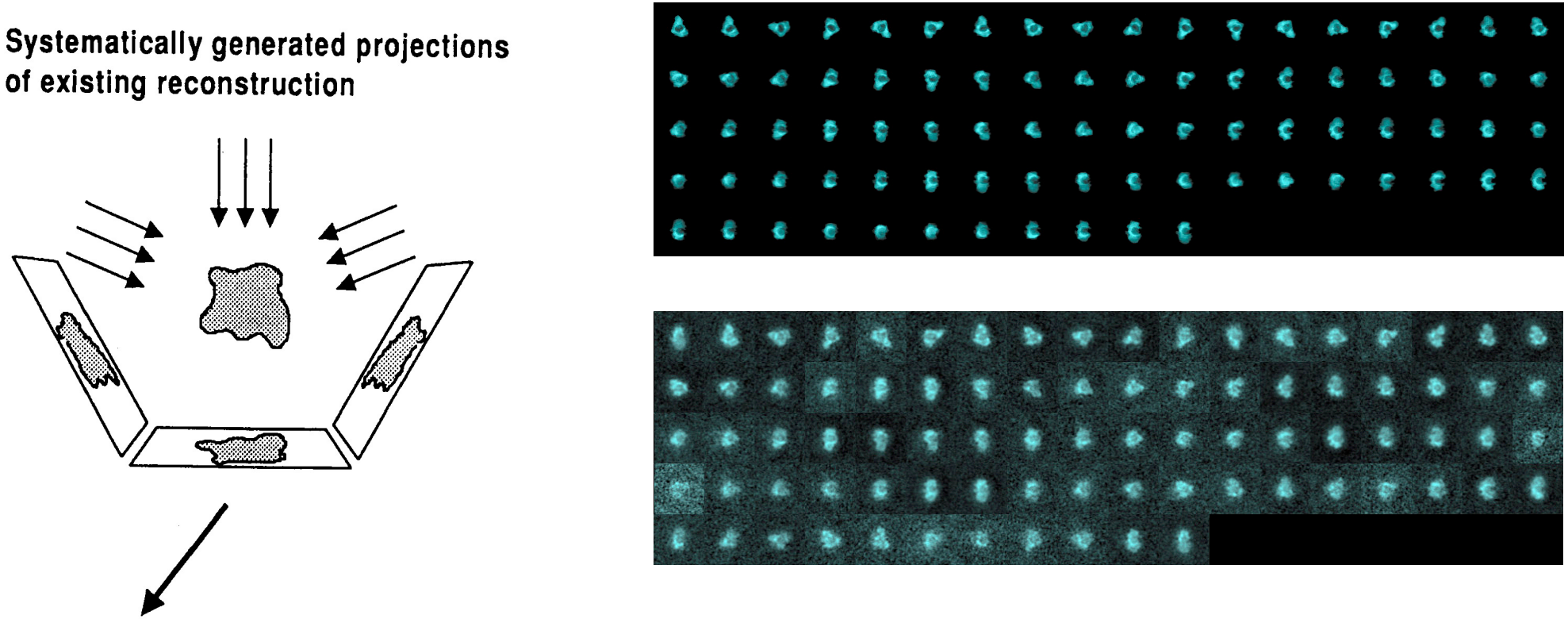


Image used with permission from Joachim Frank

‘Simple’ way of thinking about projection matching:

- Need to know those Euler angles!
- Iterate over angle assignment steps
- Better map -> better angles -> better map etc.

Another way of thinking

- We don’t care about Euler angles (they are nuisance parameters)
- Want the map that best explains the data

In second framework we can think of single particle cryoEM as an optimization problem:

A simple objective function:

Sum over all images

$$O(\mathcal{V}) = \frac{1}{N}\sum_{i=1}^{N} f_i(\mathcal{V})$$

Function to be minimized

Image

Projection

$$f_i(\mathcal{V}) = \min_{\mathbf{R},\mathbf{t}} \|\mathcal{I}^{(i)} - \mathbf{C}^{(i)}\mathbf{P}(\mathbf{R},\mathbf{t})\mathcal{V}\|^2$$

View of complex
(3D Rotation + 2D Translation)

Function used in cryoSPARC/Relion during marginalization

$$f_i(\mathcal{V}) = -\log \int_{\mathbf{R}} \int_{\mathbf{t}} \exp\left(-\frac{\|\mathcal{I}^{(i)} - \mathbf{C}^{(i)}\mathbf{P}(\mathbf{R},\mathbf{t})\mathcal{V}\|^2}{2\sigma^2}\right) d\mathbf{R}d\mathbf{t}$$

Minimize a function of the volume *O*(*V*) that describes the disagreement between each of *N* particles and the map/volume *V*.

Implications of considering as an optimization problem:

- May want to marginalize (not assign hard angles to each particle image)
- May not want to look at each particle image for each map update

Looking at every particle image every time

How do we optimize this objective function?

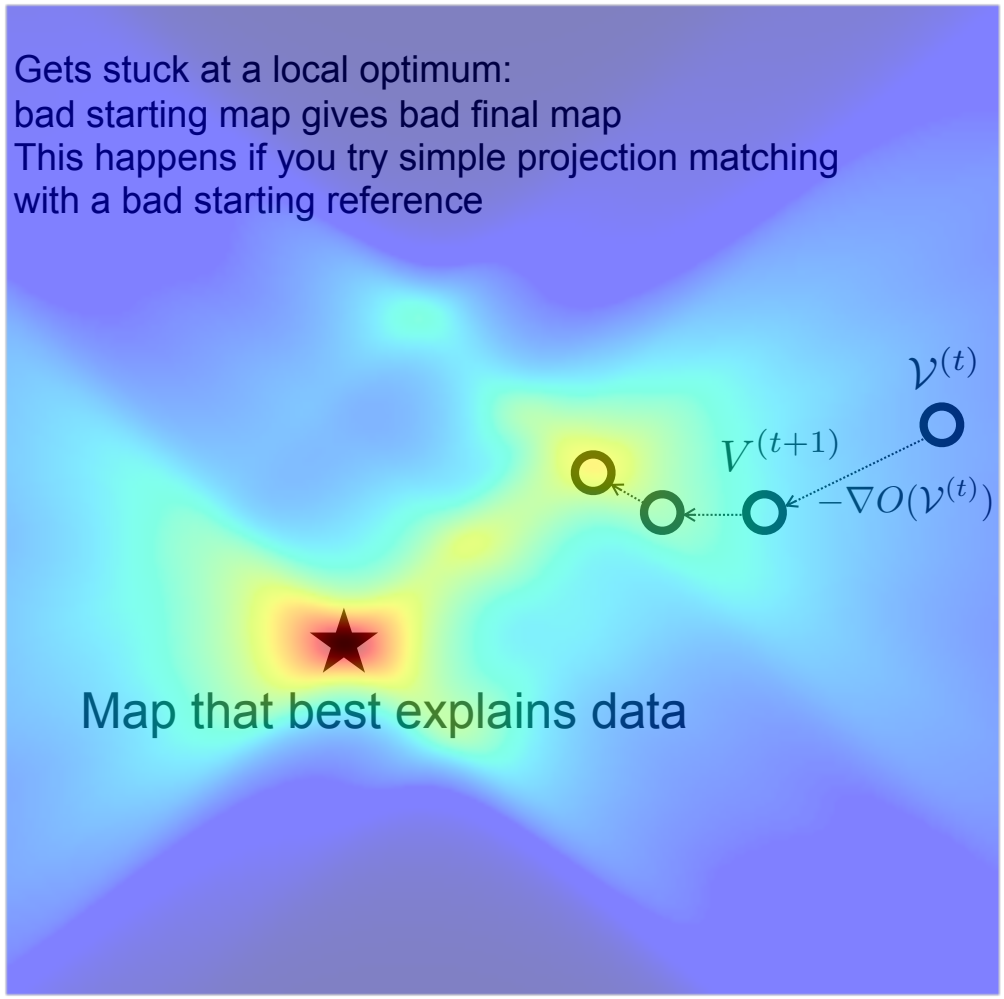
$$O(\mathcal{V}) = \frac{1}{N}\sum_{i=1}^{N} f_i(\mathcal{V})$$

Gradient based optimization

- At each iteration take a step toward best map you can

$$\mathcal{V}^{(t+1)} = \mathcal{V}^{(t)} - \epsilon_t \nabla O\left(\mathcal{V}^{(t)}\right)$$

Gradient of Objective Function



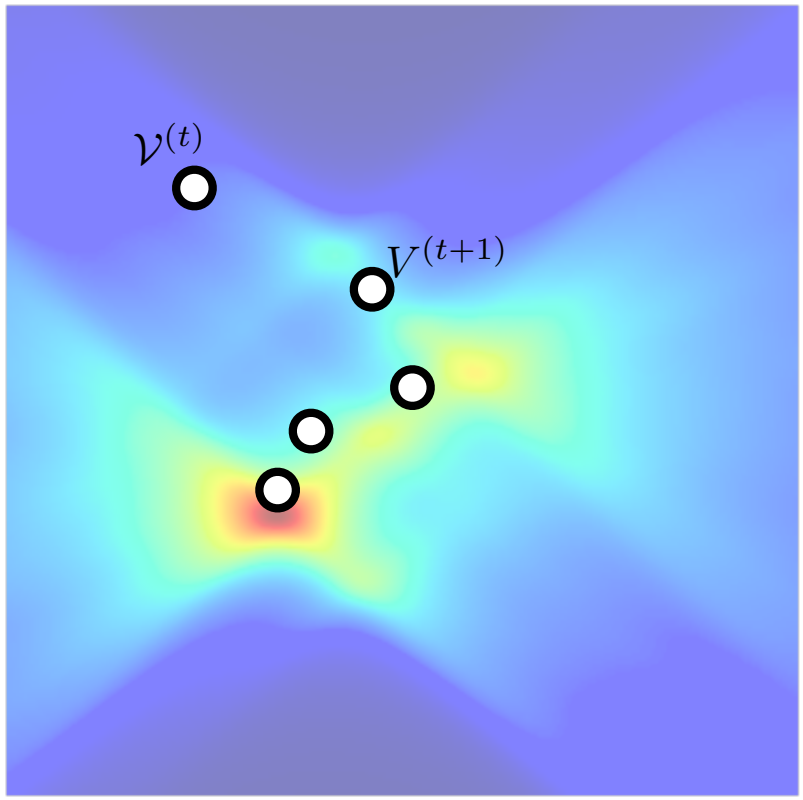
space of all possible maps

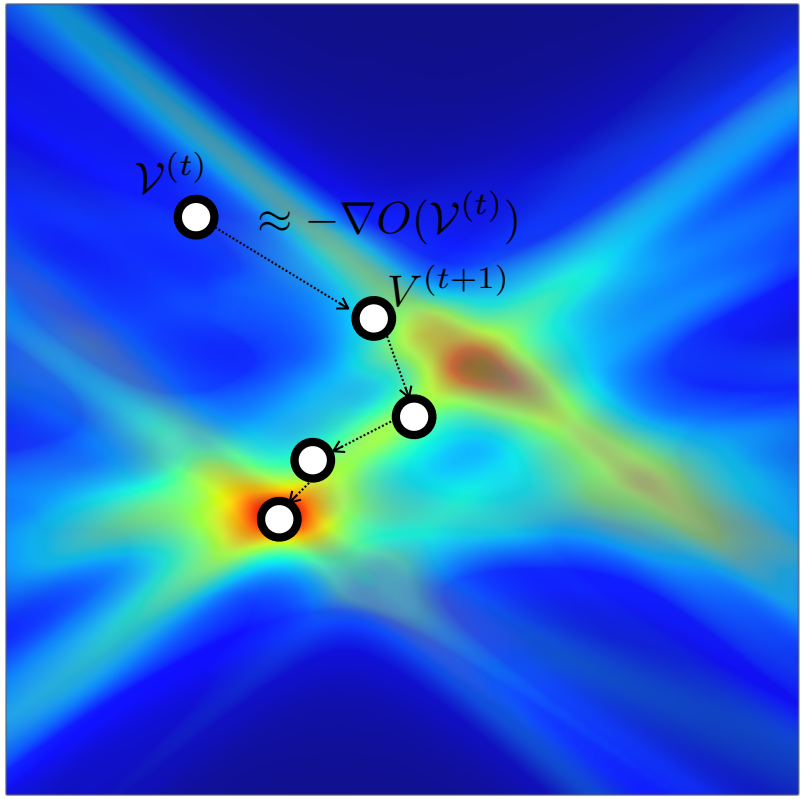
Stochastic Gradient Descent (don't look at each image at each iteration)

SGD approximates the objective and gradient with a random subset $\mathfrak{J}$

$$\nabla O(\mathcal{V}) \approx \frac{1}{|\mathfrak{J}|} \sum_{i \in \mathfrak{J}} \nabla f_i(\mathcal{V})$$

Exact Objective **Random Subset** Approximation

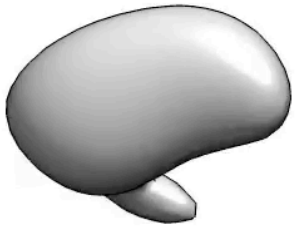


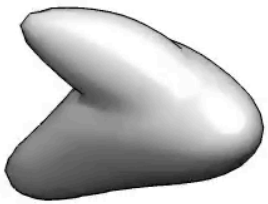


Punjani, Rubinstein, Fleet, and Brubaker (2017). *Nature Meth* **14**, 290-6.

Cryo-EM single particle *ab initio* reconstruction and classification

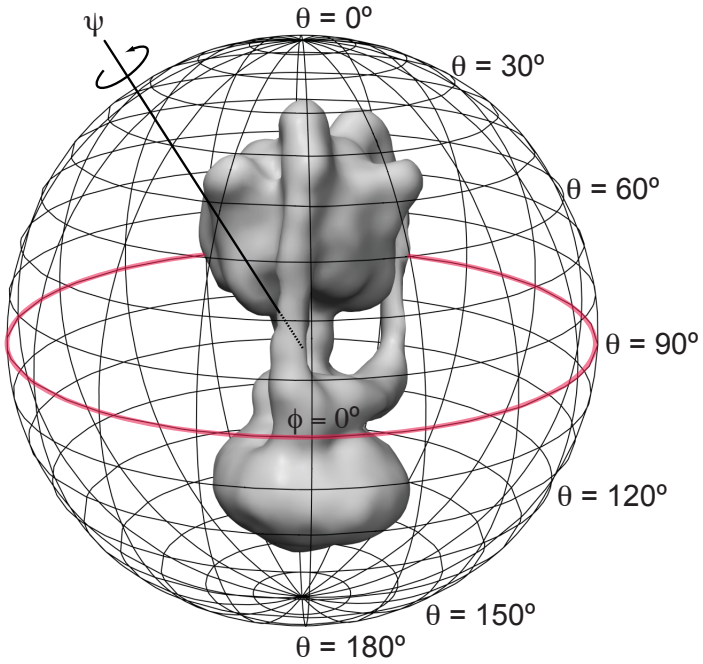
STRUCTURA BIOTECHNOLOGY

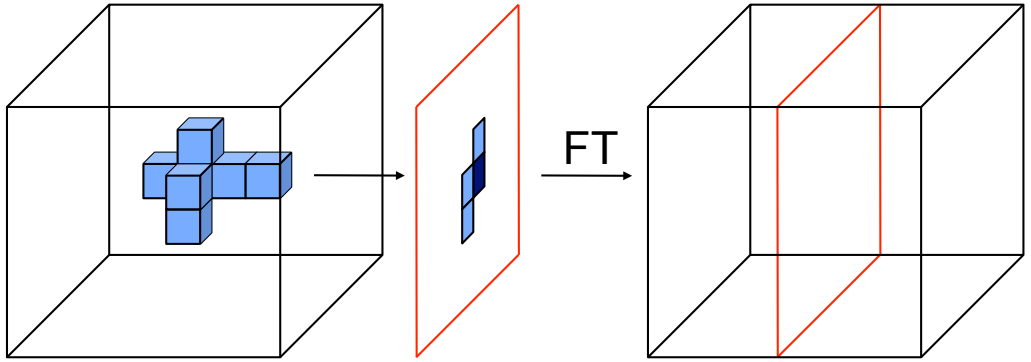




Punjani, Rubinstein, Fleet, and Brubaker (2017). *Nature Meth* **14**, 290-6.

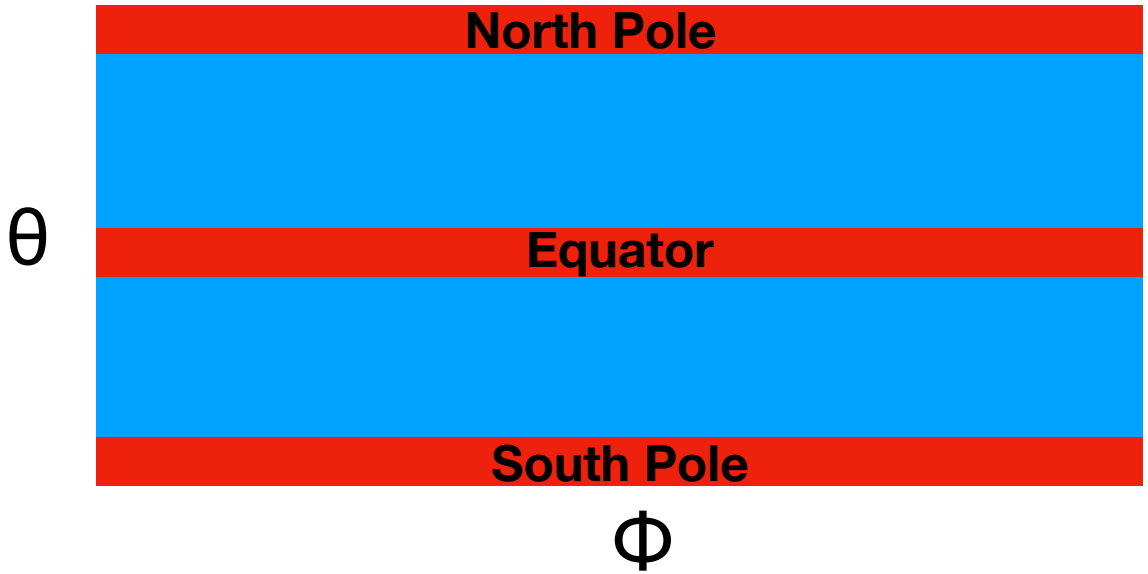
Do you have sufficient views to fill Fourier space?



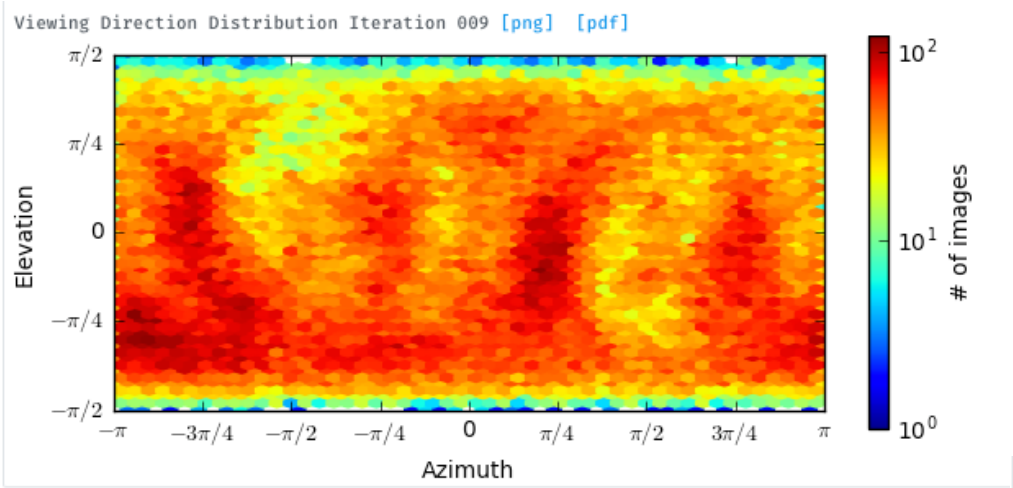


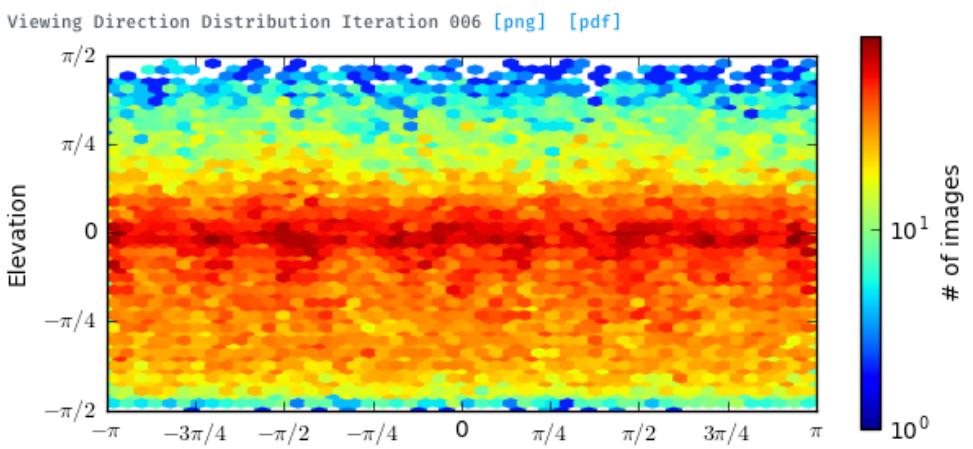
Projection

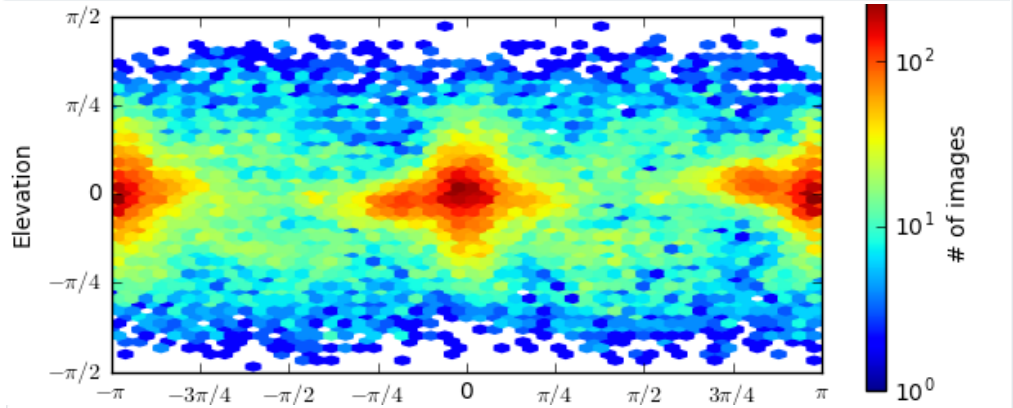
Central section through the 3-D FT of object

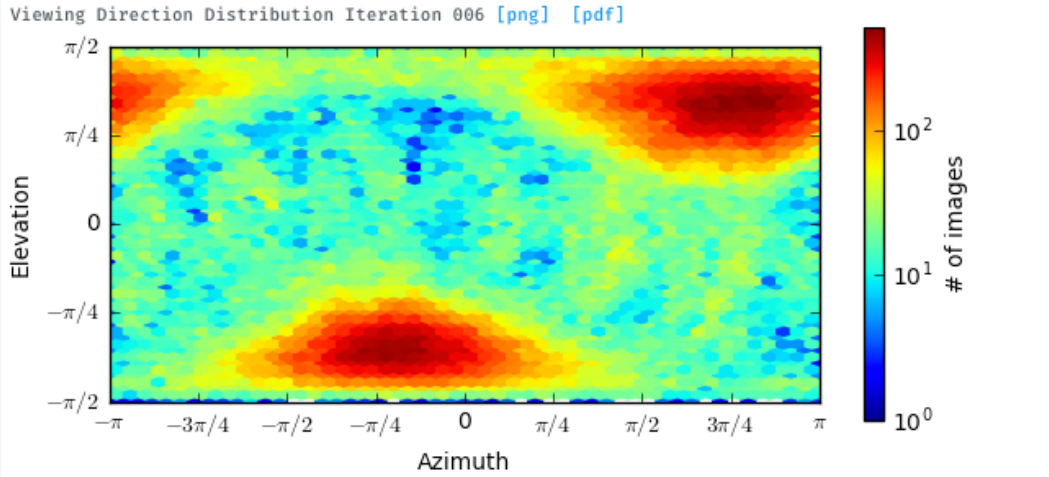


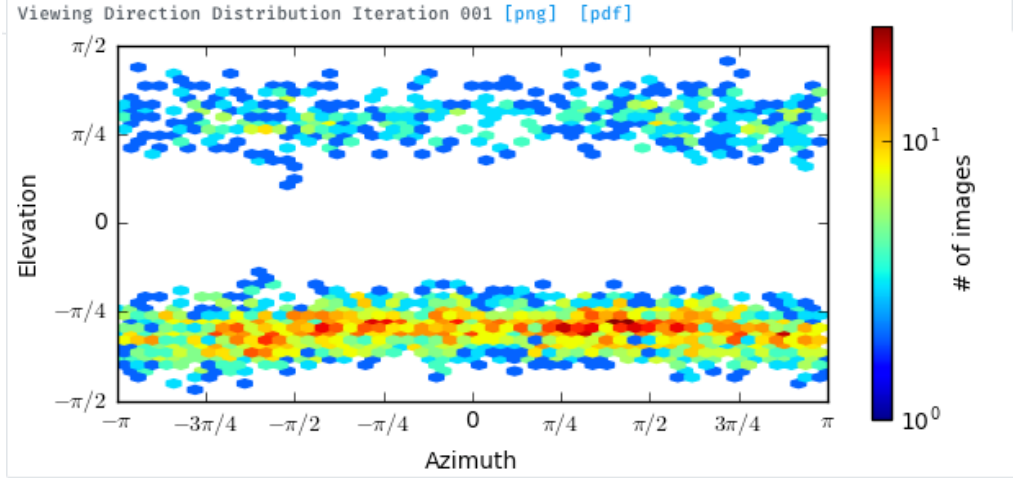
Do you have enough views to fill Fourier space?







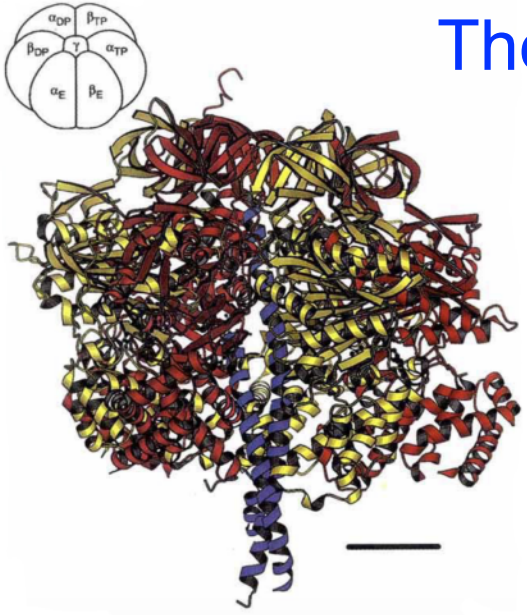




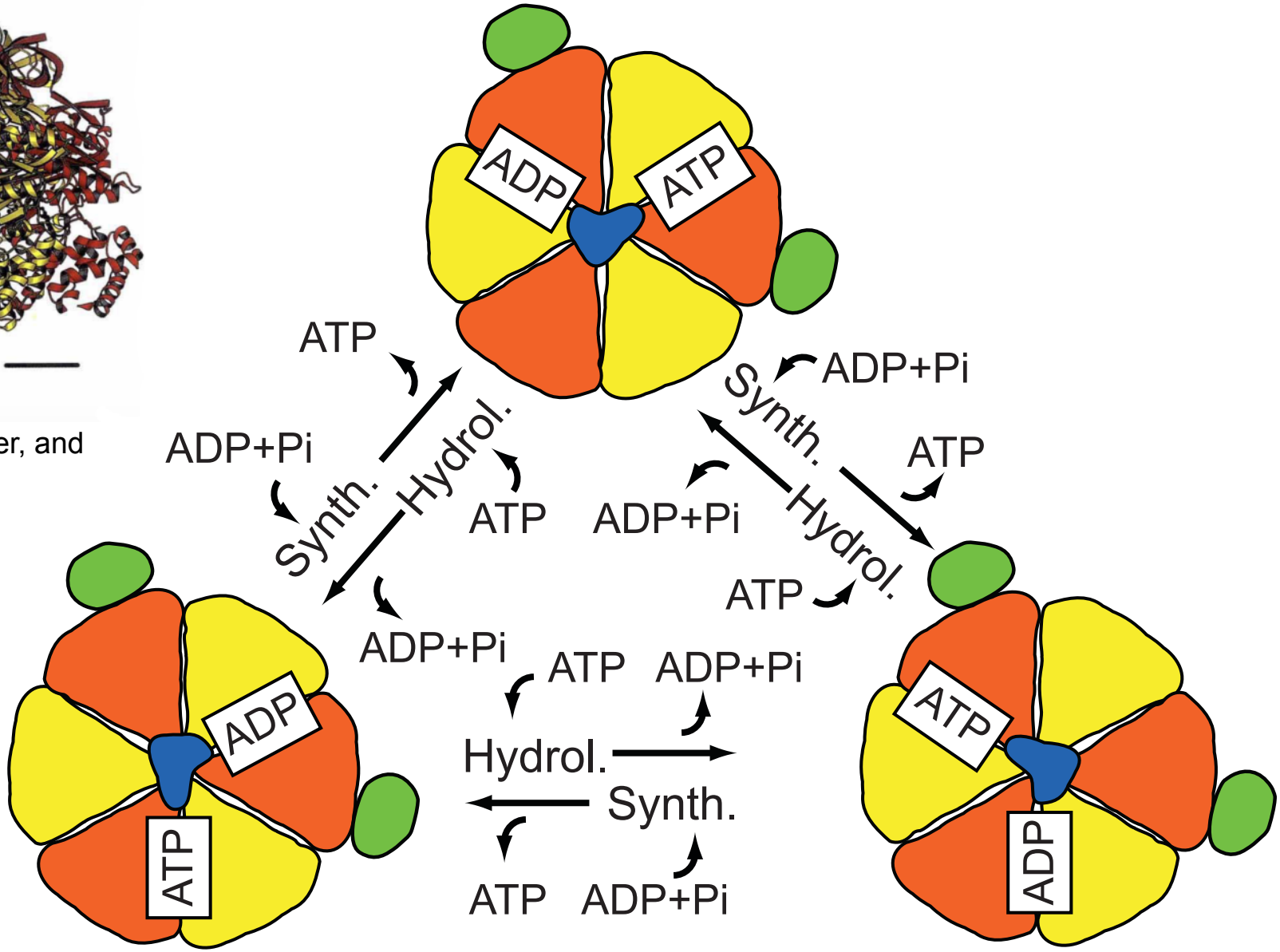
3D classification

3D classification to separate conformations:

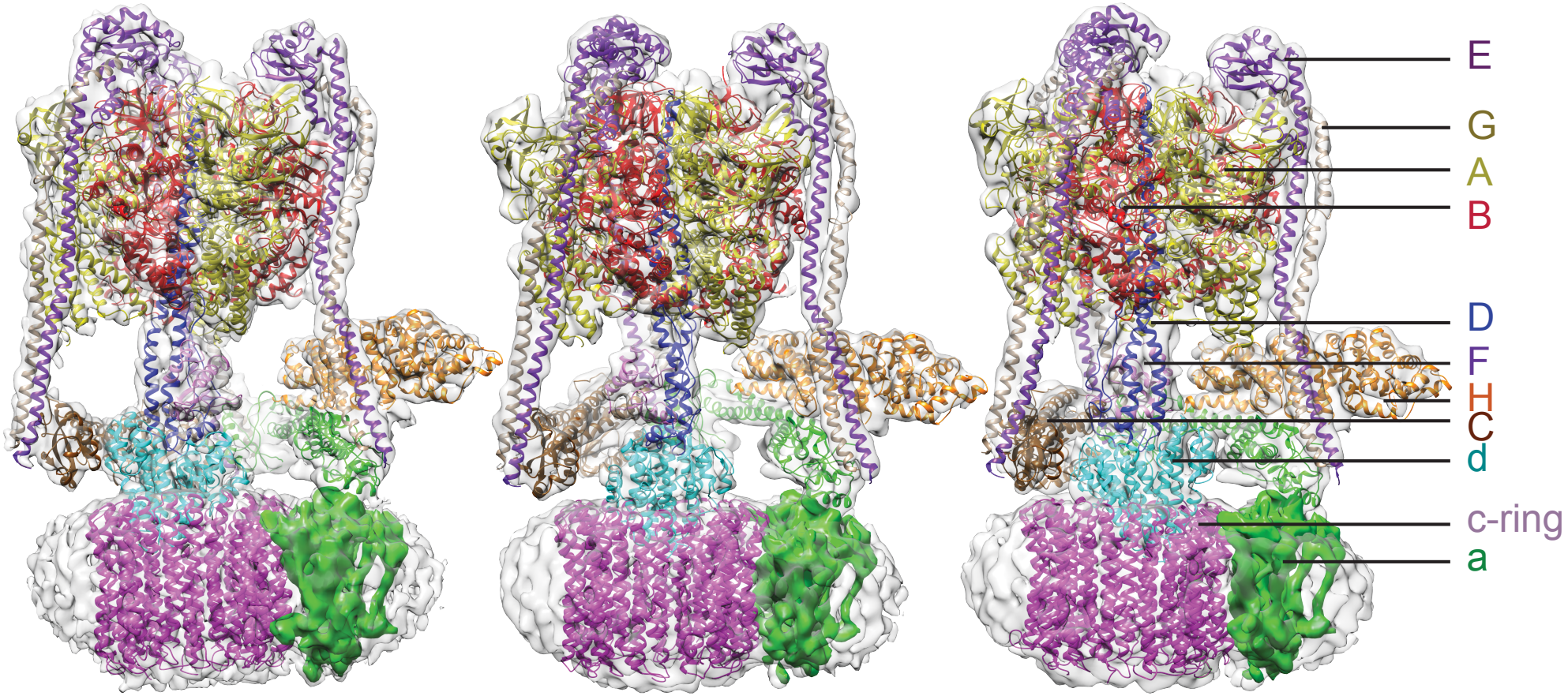
The binding change mechanism



Abraham, Leslie, Lutter, and Walker (1994)
Nature 370, 621-8



Separation of V-ATPase images into different classes (J. Zhao)



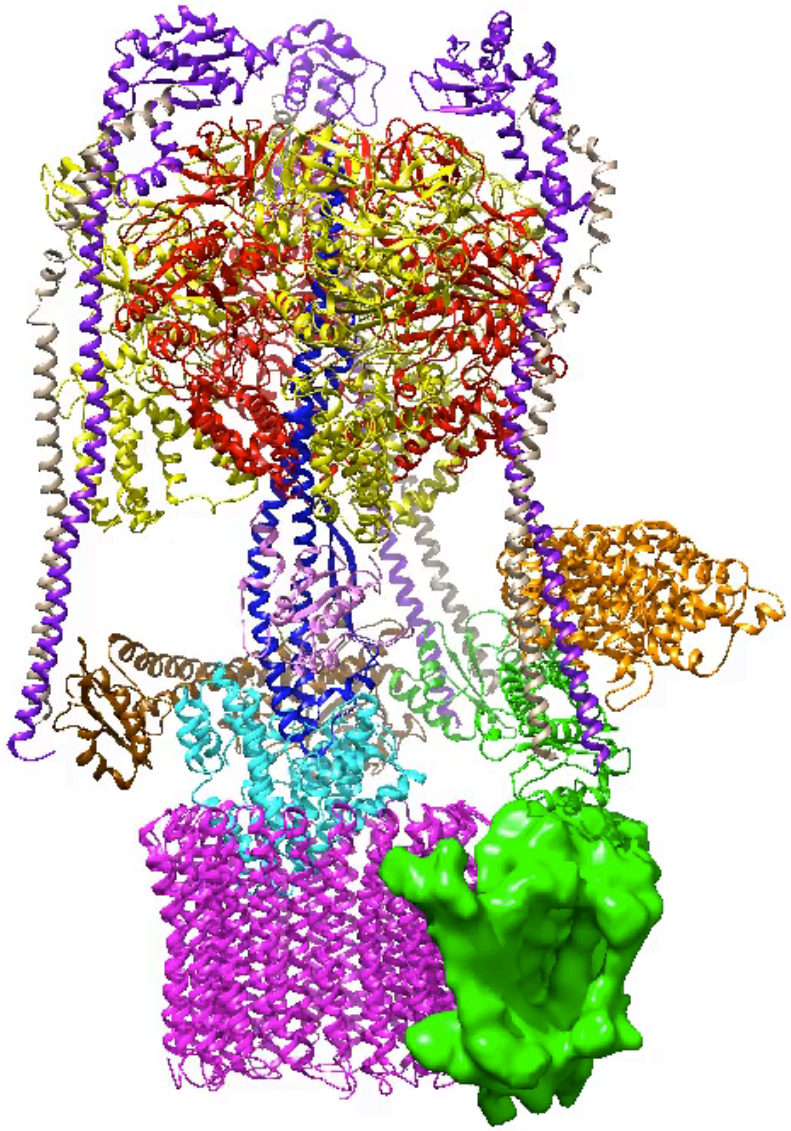
State 1
(47% of images)
6.9 Å

State 2
(36% of images)
7.7 Å

State 3
(17% of images)
8.1 Å

Zhao, Benlekbir, and Rubinstein (2015). *Nature* **521**, 241-5.

3-D classification to separate conformations:



Zhao, Benlekbir, and Rubinstein (2015). *Nature* **521**, 241-5.