De-demystifying 2017 Nobel Prize in Chemistry from a structural biologist view

Wei-hau Chang
Institute of Chemistry, Academia Sinica
weihau@gate.sinica.edu.tw

ISGC, March 21st, 2018
Academia Sinica, Taipei, Taiwan

Dubochet, Frank, Henderson

"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"

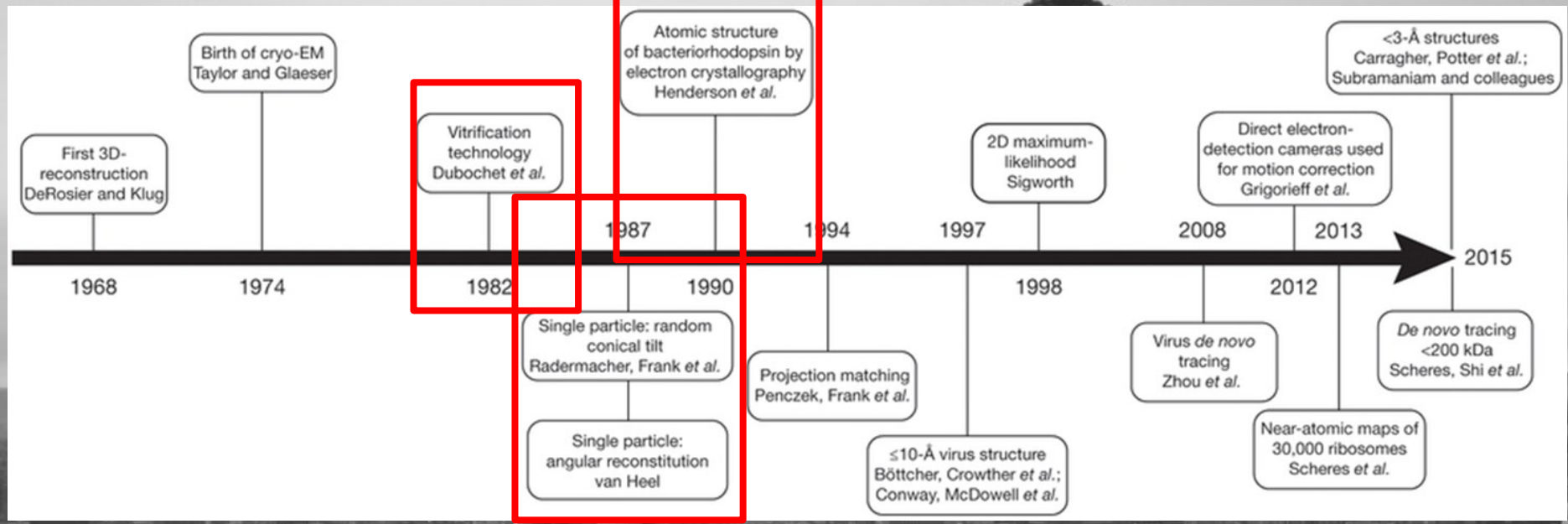
2017 Nobel Chemistry was for "single-particle cryo-EM"

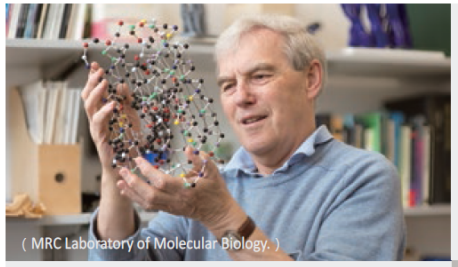
Inventor of sample preparation (1982-).

Inventor of algorithm and ruler (1987-).

Low-dose electron microscopy (1990-,1995-).

Milestones in Cryo-EM




（MRC Laboratory of Molecular Biology.）

韓德森（Richard Henderson, 1945~）
國籍｜英國
現任｜劍橋 MRC 分子生物實驗室
研究領域｜結構生物學

法蘭克（Joachim Frank, 1940~）
國籍｜德國、美國
現任｜哥倫比亞大學生物物理學教授
研究領域｜結構生物學、蛋白質生物合成

杜巴謝（Jacques Dubochet, 1942~）
國籍｜瑞士
現任｜洛桑大學生物物理學教授
研究領域｜結構生物學

諾貝爾化學獎—用低溫捕獲生命原態的原子細節

章為皓／中研院化學所副研究員，致力臺灣低溫電顯平台之建立。

現代化學的核心是關於物質的研究，因此，諾貝爾化學獎不外乎在合成、結構和分析方法 3 個領域的來回振盪。今（2017）年，化學獎頒給了看似物理學的低溫電子顯微術領域，卻是屬於結構和分析的

2017
諾貝爾獎特別報導

化學範疇，代表此領域獲獎的是分別在樣品製備、影像重組和低劑量電子成像 3 個關鍵有卓越突破的的杜巴謝（Jacques Dubochet）、法蘭克（Joachim Frank）以及韓德森（Richard Henderson），獲獎原因爲「發展低溫電子顯微術，可應用於溶液中生物分子的高解析度結構測定」。了解生物分子的結構到底有什麼重要性？細胞是生命的基本單元，細胞的運作是靠著裡面許許多多的蛋白質機器不斷運轉。了解這些分子機器如何運動，使得人類能一窺細胞的奧秘，並通過控制或改造這些生物分子，使人類有機會脫離疾病的束縛。而如何了解這些分子機器如何運動，最直接的方法就是對這些分子機器攝像。

這樣的夢想因爲英國醫學研究中心的比魯茲（Max Perutz）在 1950 年代初解決了 X 射線蛋白質晶體繞射圖的相位問題而實現。從 1980~2000 年間，由於同步輻射、大面積相機和結構測定軟體分享網絡的興起，X 射線蛋白質晶體學逐漸成爲結構生物學的主流工具，許多大分子的晶體結構紛紛被解出，最有名的爲光合作用中心、ATP 合成酶、鉀離子通道、核醣核酸聚合酶和核醣體的原子結構，囊括了 5 次諾貝爾化學獎，而其中有 3 次是被英國醫學研究中心的研究員或校友抱走。然而，嚴格來說，晶體結構與能表徵生理功能的溶液結構還是不同。如今，通過杜巴謝等人發展的「電子照妖鏡」，科學家終於能在茲卡病毒（Zika）甫一出現之際，不需用結晶學就能破解球殼蛋白的原子圖譜，爲藥物設計提供精確藍圖。本文將分別對韓德森、杜巴謝和法蘭克的研究做深度報導。

韓德森：走出結晶學的先知

韓德森在低溫電顯的關鍵貢獻是引入低劑量欠焦電子成像法。筆者有幸在 1993 年於加州大學柏克萊求學時巧遇韓德森，便先從韓德森談起。時年 48 歲的他，爲人謙和卻又風

作，我了解到若要藉電子

結構，需透過二維結晶提

平均而提升訊噪比。韓德

了一套電子結晶學的標準

純化一個轉錄複合物並長

晶體，靠著韓德森方法分

二維結晶是必要的？其實

年前已臻完美，是材料科

利器，然而，一旦把金屬

清晰度有如飛蚊症患者看

這是由於蛋白質對電子束

劑量的電子（每平方埃 10

一分子的影像中的原子細

在跨入電顯這一行之前，

射大師布洛（David Blow

位後，旋赴耶魯大學，打

蛋白結構。礙於純化膜蛋白

屬於超級困難的挑戰，韓德

研究中心，由克里克（Fra

得主）和克盧格（Aaron K

共同領導的團隊，當時組

（David DeRosier）、克勞

芬區（John Finch）和昂溫

爾和克勞德長於三維重建

顯微鏡攝像。韓德森基於 19

與其學生泰勒（Taylor）

輻射傷害的首度觀察，並

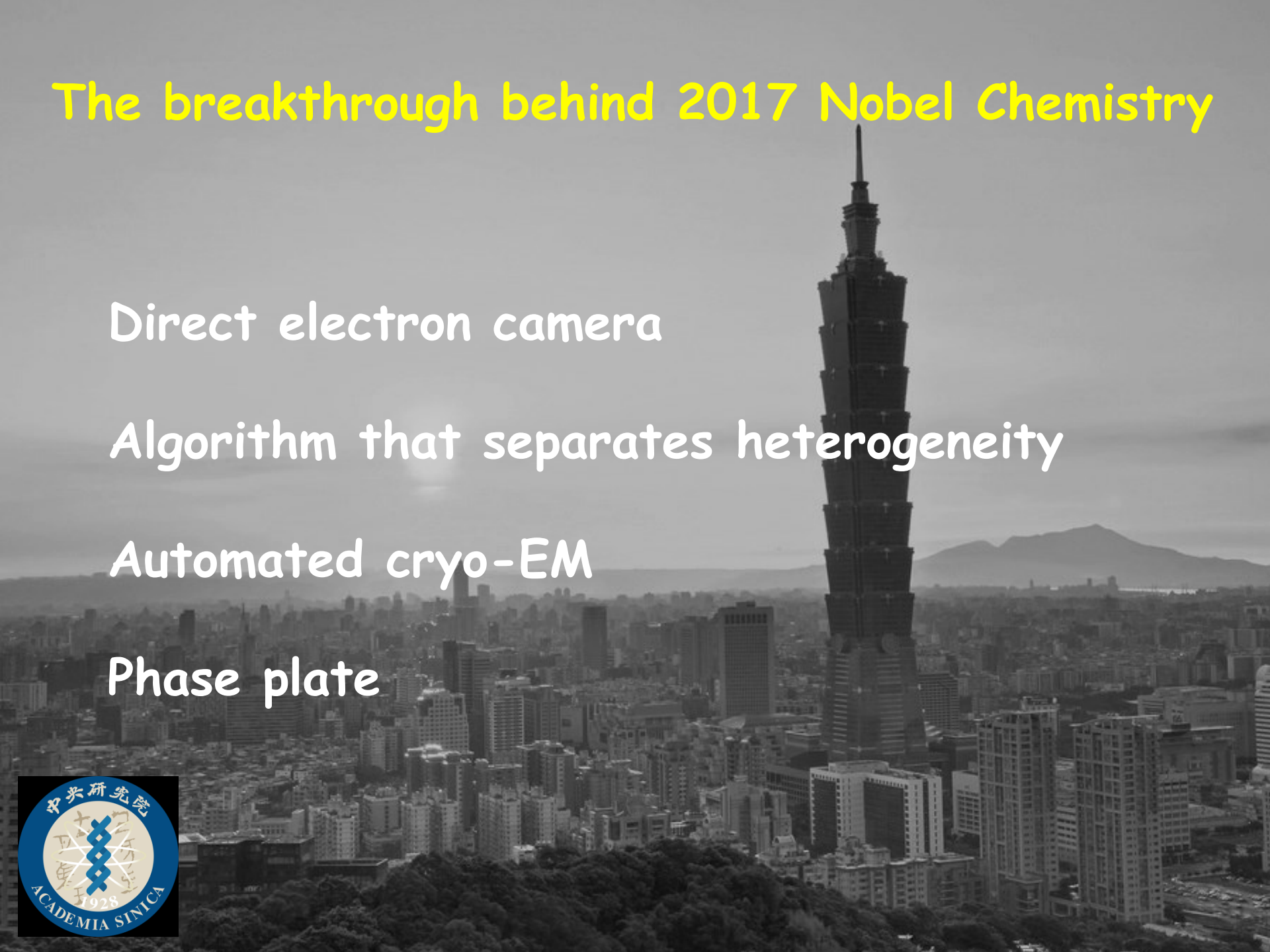
電子顯微術，在 1975 年

（7Å），觀察到紫質蛋白

級結構是 α 螺旋（alpha h

雖然這個工作所獲得的是

The breakthrough behind 2017 Nobel Chemistry



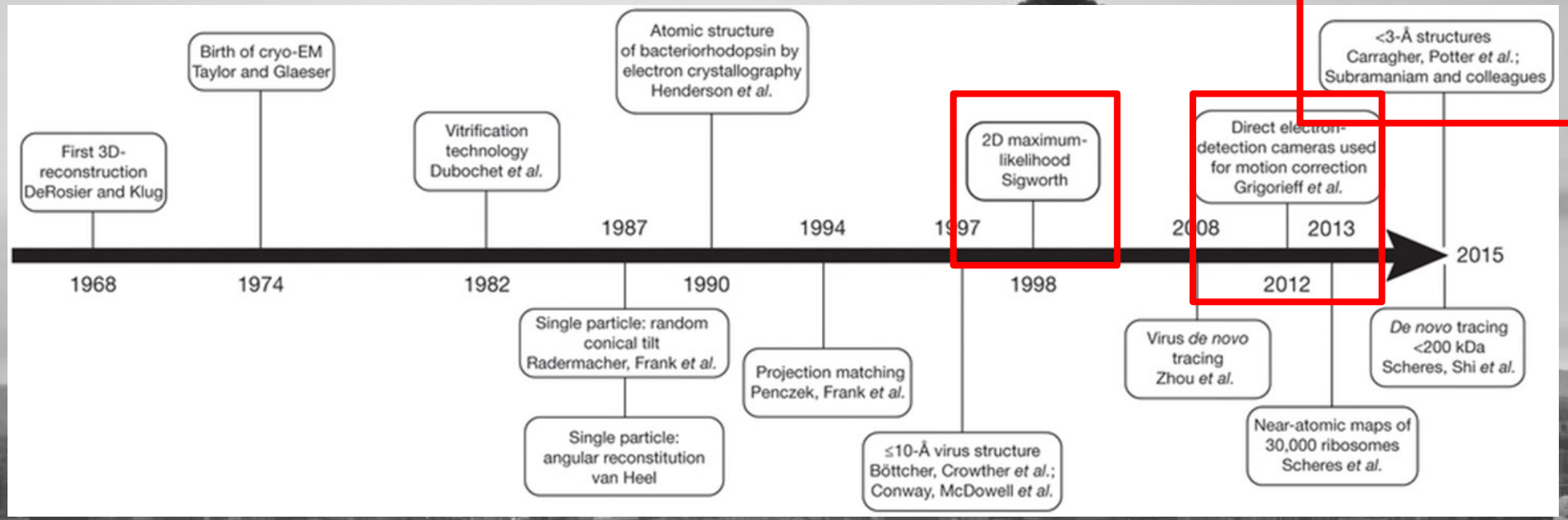
Direct electron camera

Algorithm that separates heterogeneity

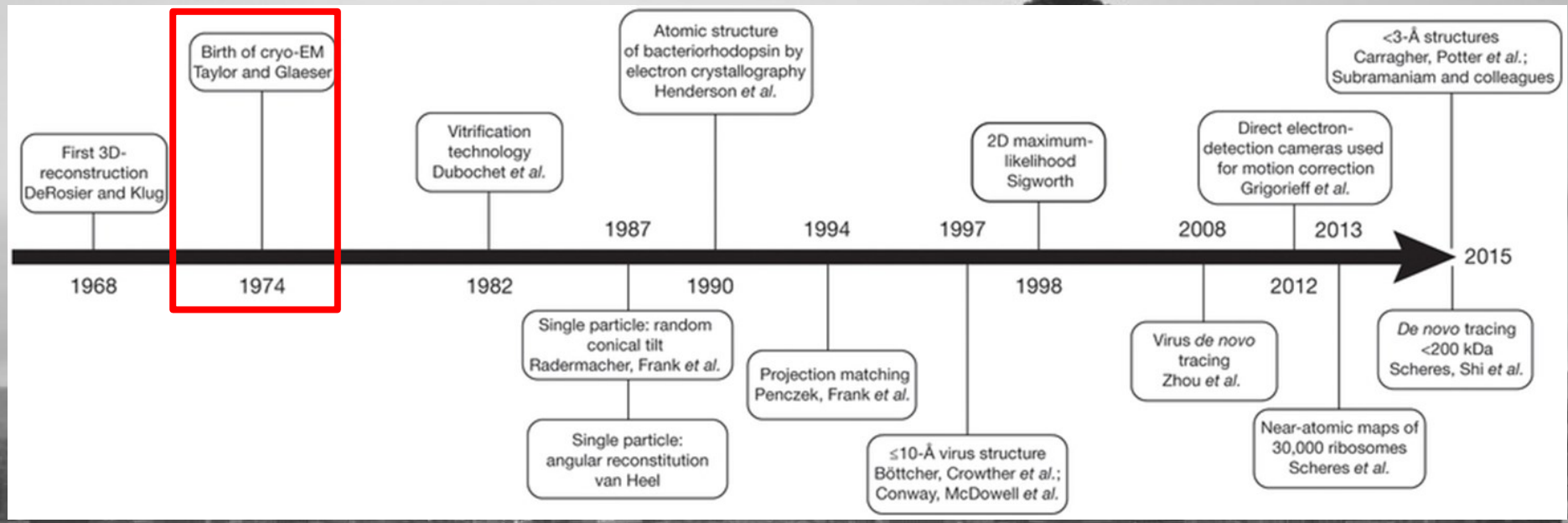
Automated cryo-EM

Phase plate

Milestones in Cryo-EM



Milestones in Cryo-EM



Frozen catalase crystal to 3.4 Angstrom

Taylor and Glaeser (1974)

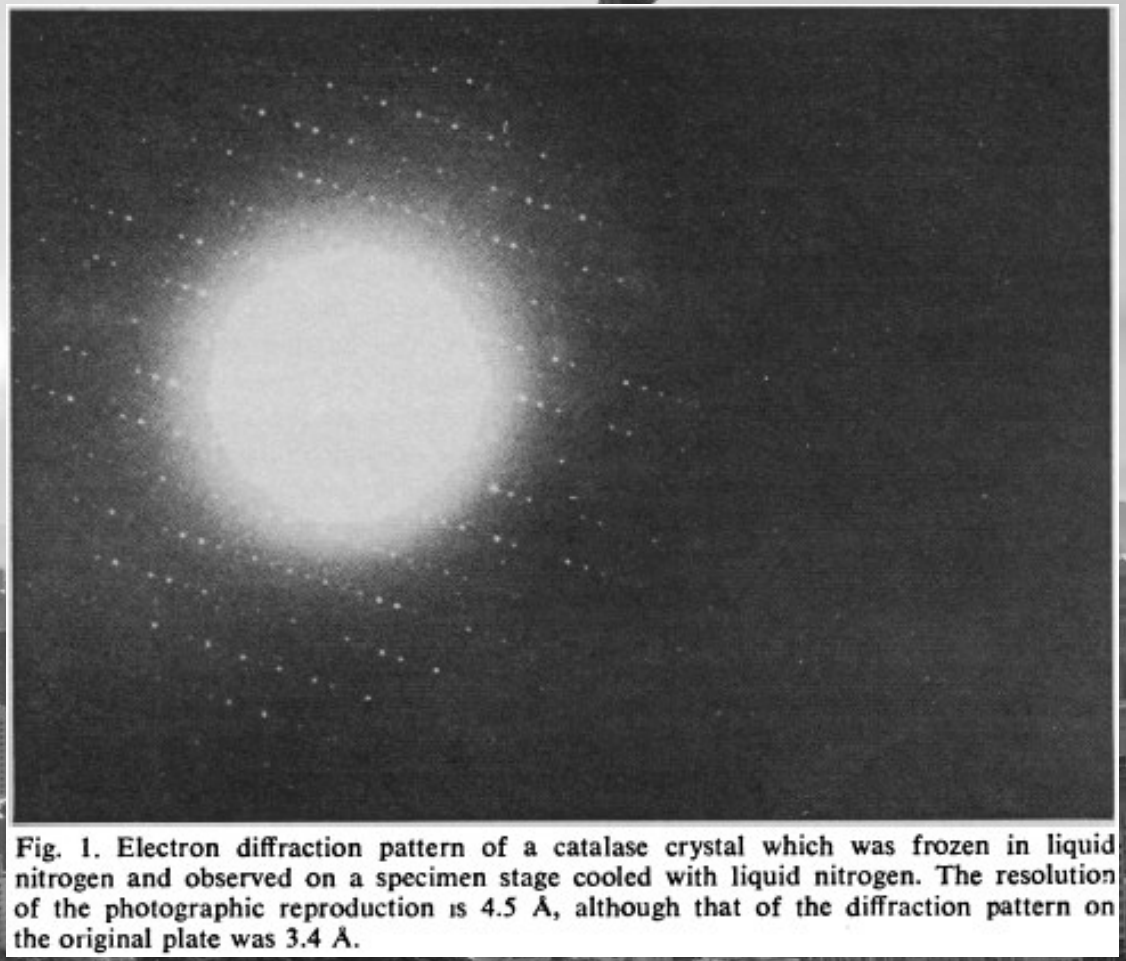
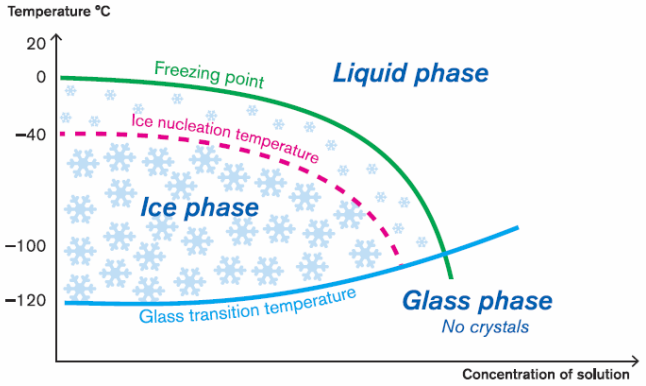
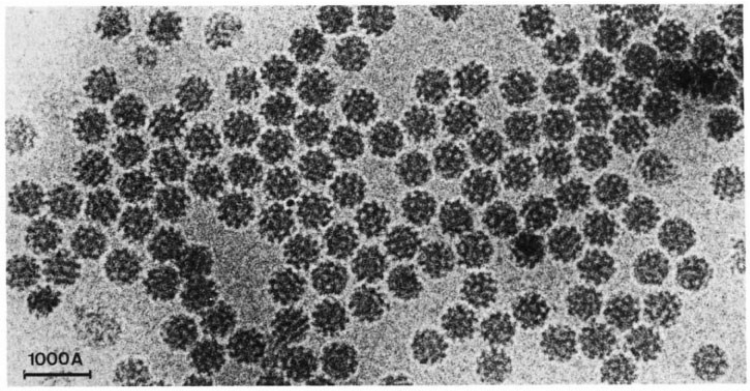


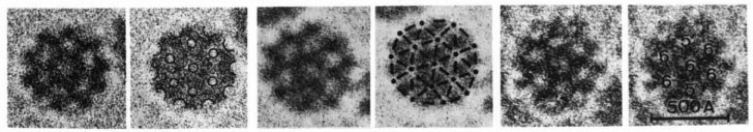
Fig. 1. Electron diffraction pattern of a catalase crystal which was frozen in liquid nitrogen and observed on a specimen stage cooled with liquid nitrogen. The resolution of the photographic reproduction is 4.5 Å, although that of the diffraction pattern on the original plate was 3.4 Å.

Rapid flash-freezing

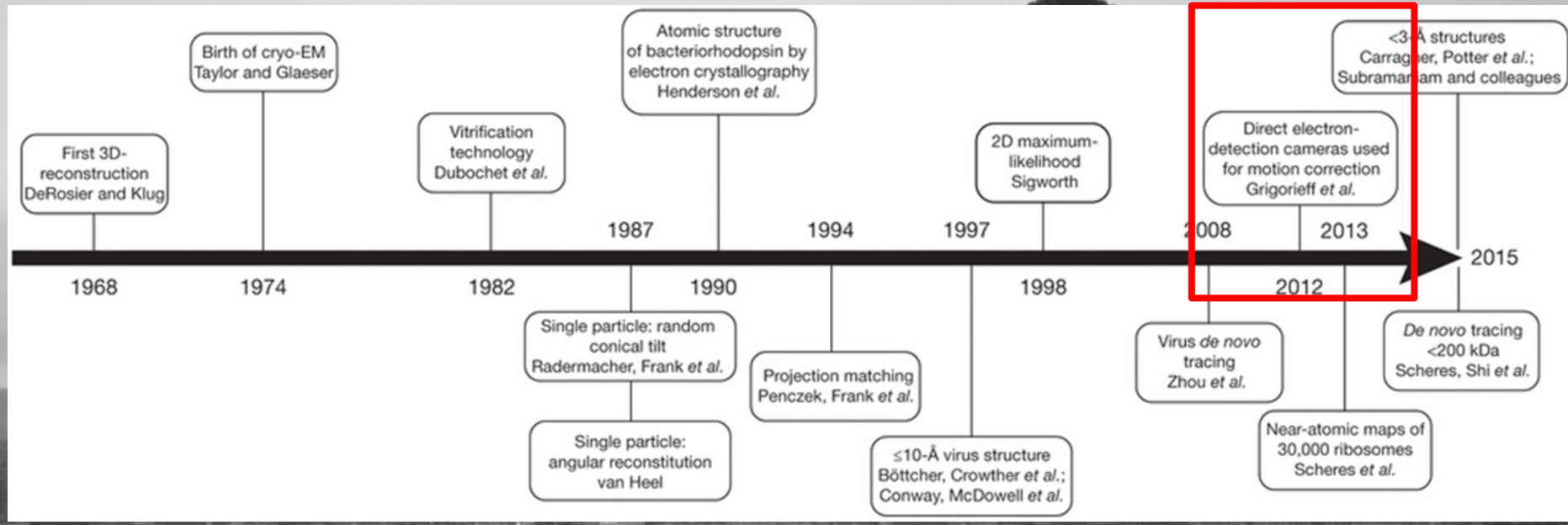
Dubochet (1984)





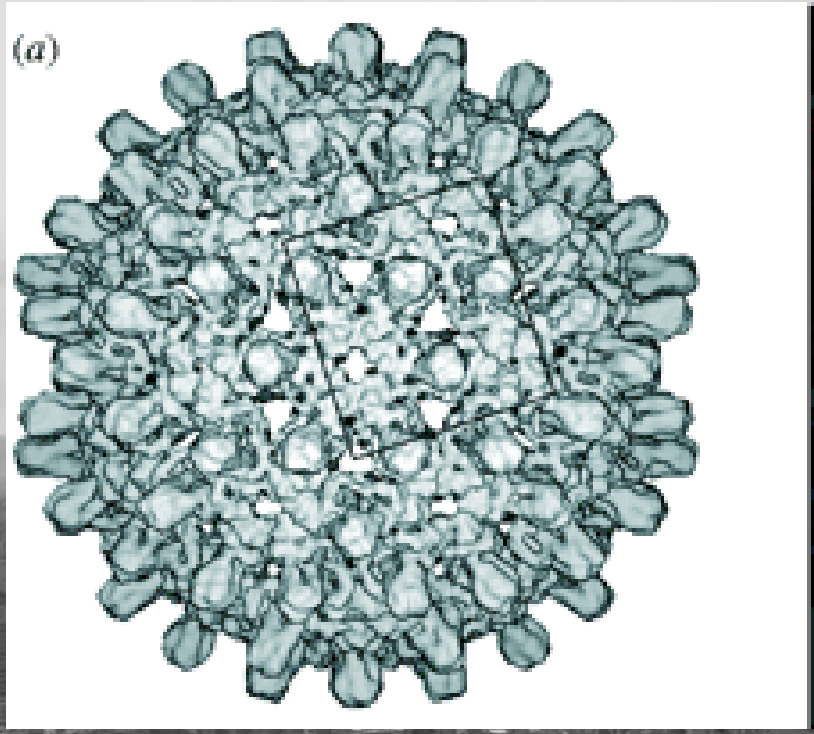


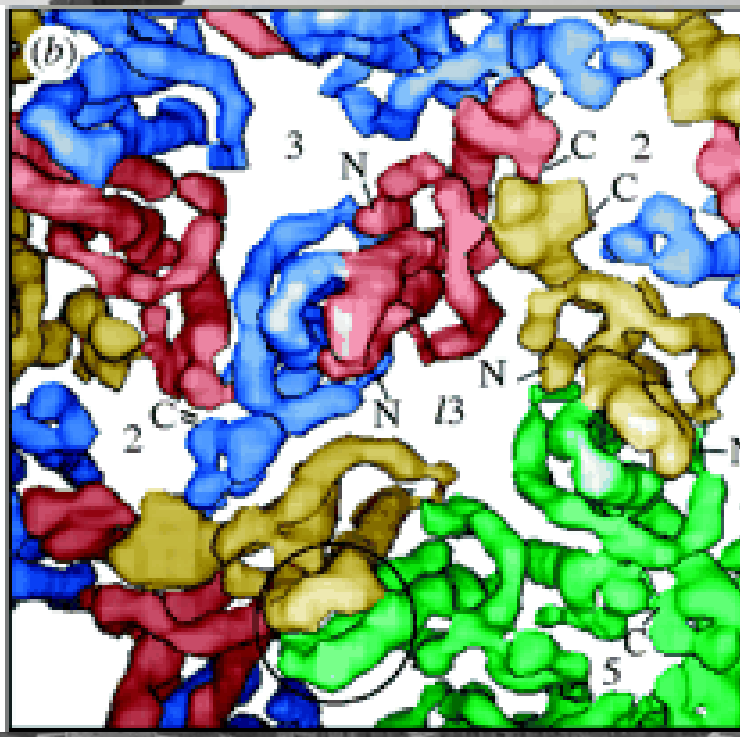
The power of direct electron camera



HBV core particle at 7 Angstrom

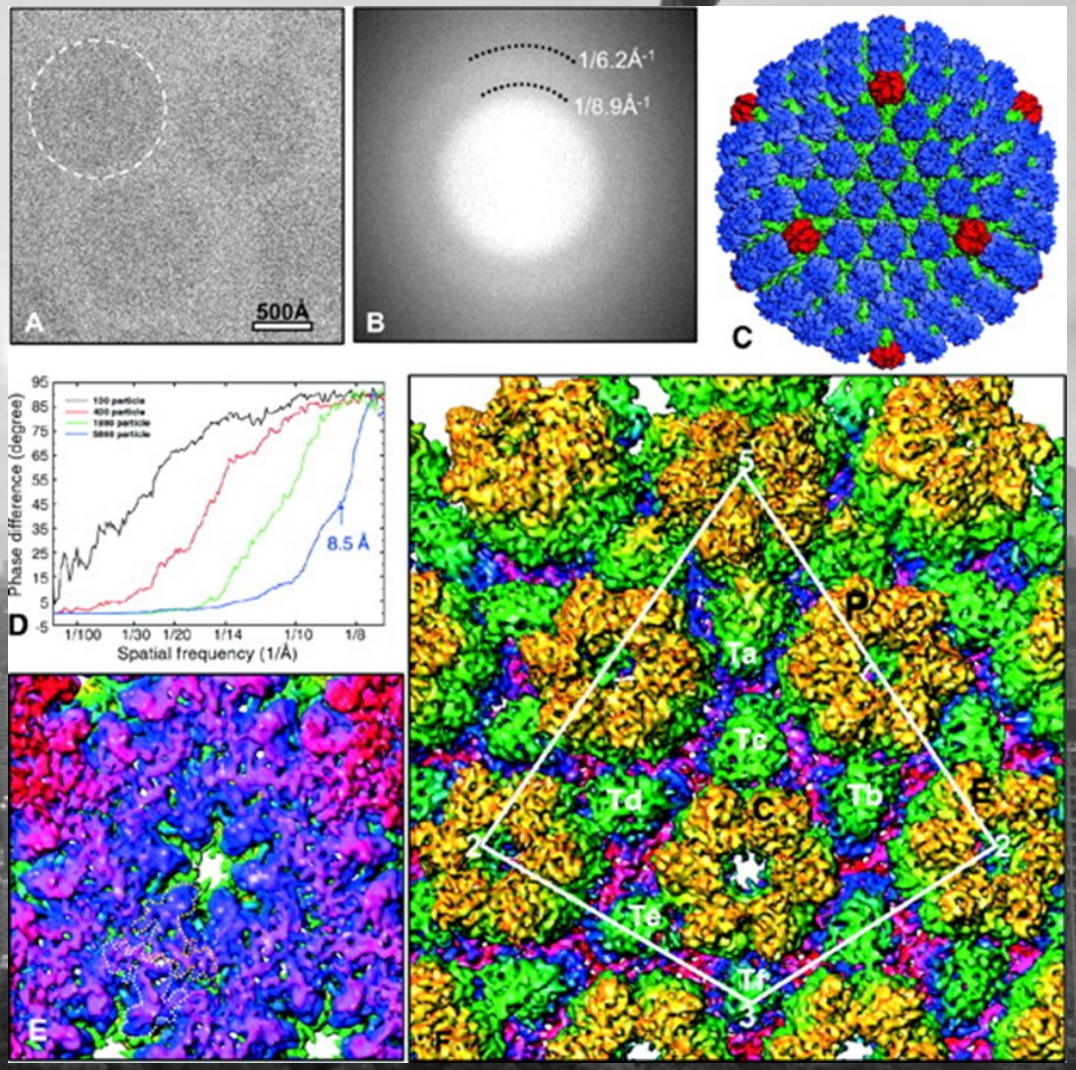
Bottcher and Crowther (1997)



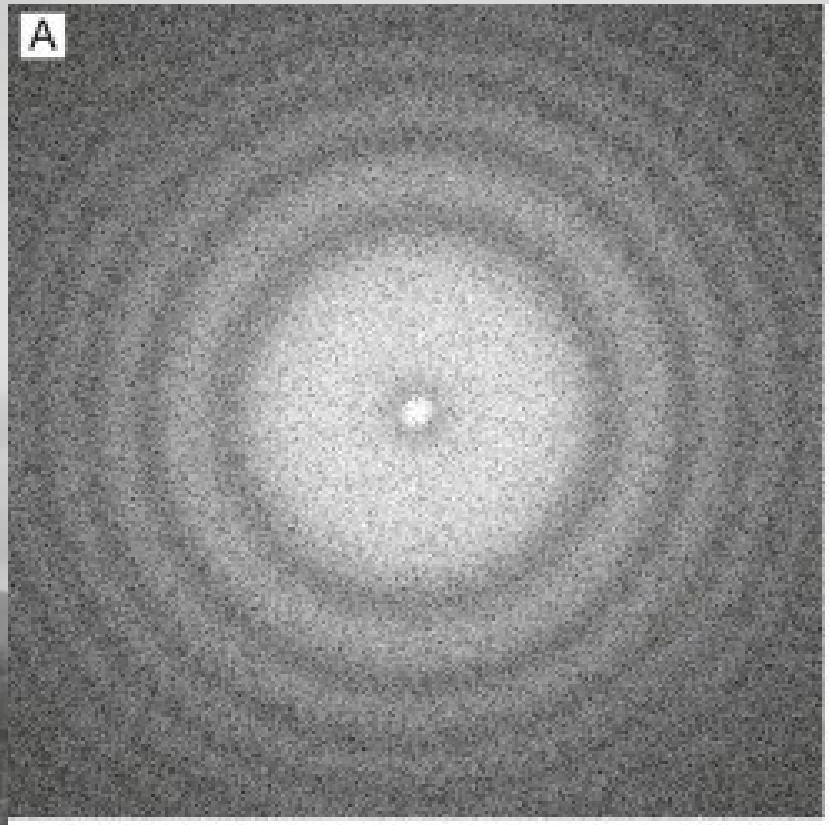


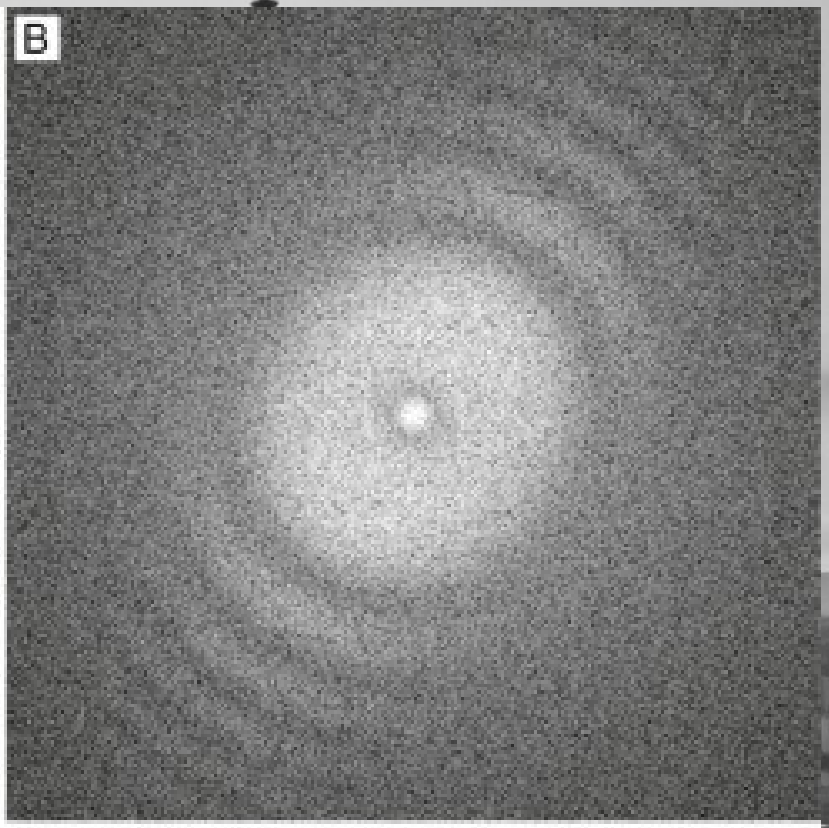
HSV-1 B capsids at 8.5 Angstrom

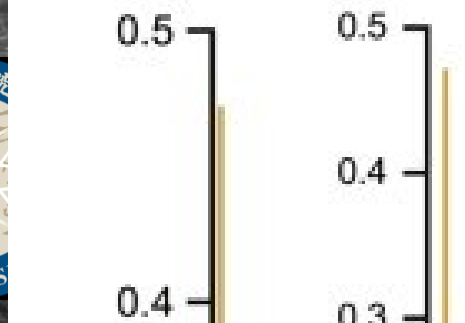
Zhou et al., and Chiu (2000)

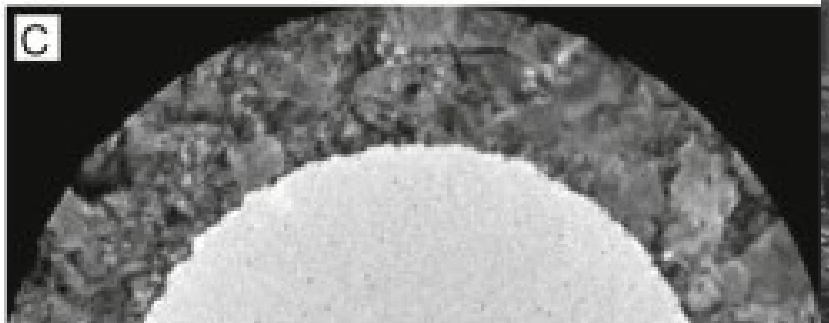


Charge-induced motion (Passmore and Russo)

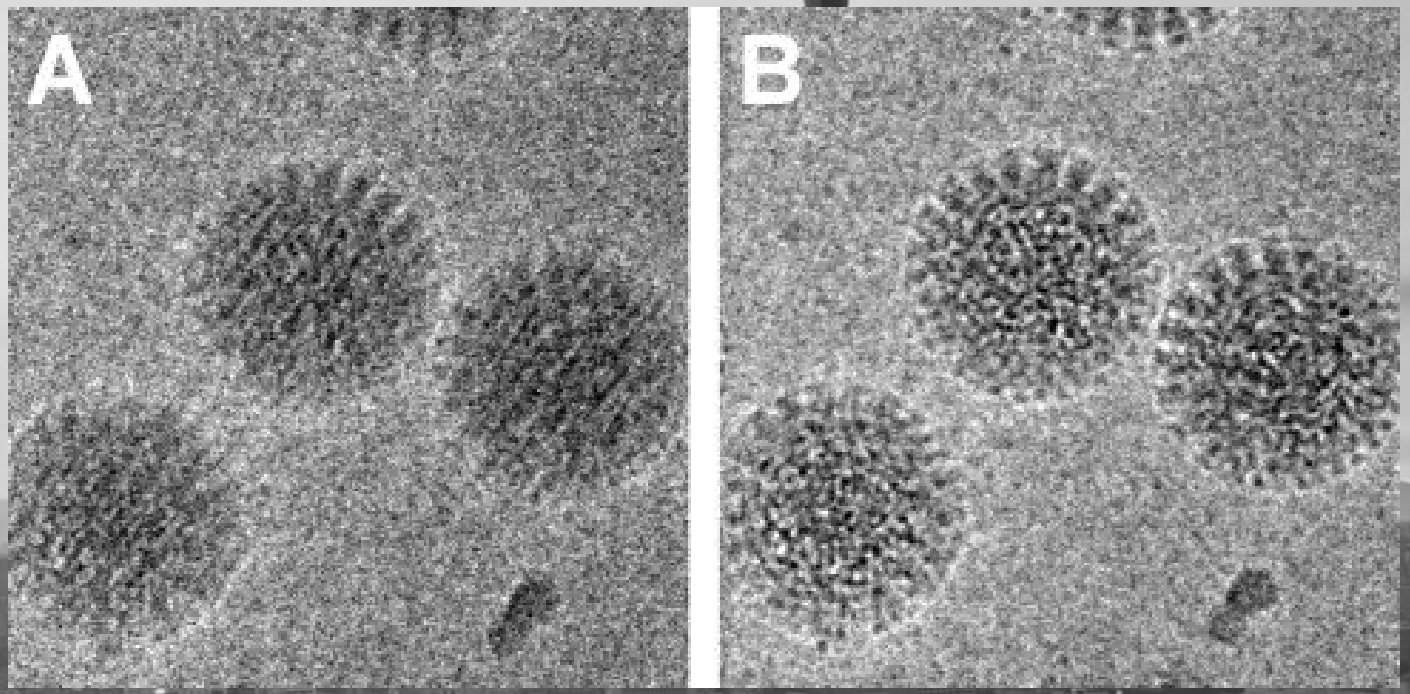






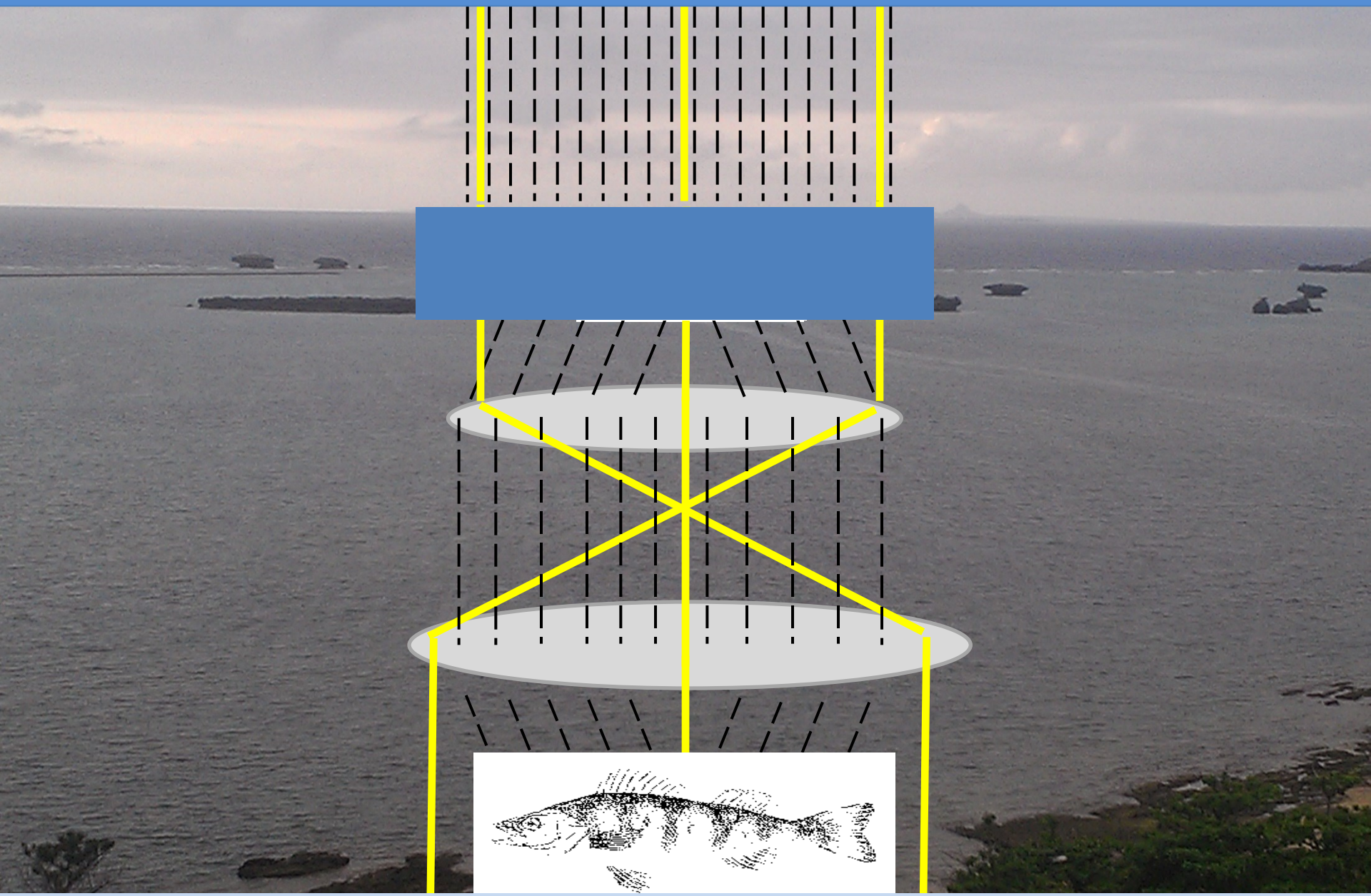


Correcting the mtion of cryo specimen Grigorieff (2012)

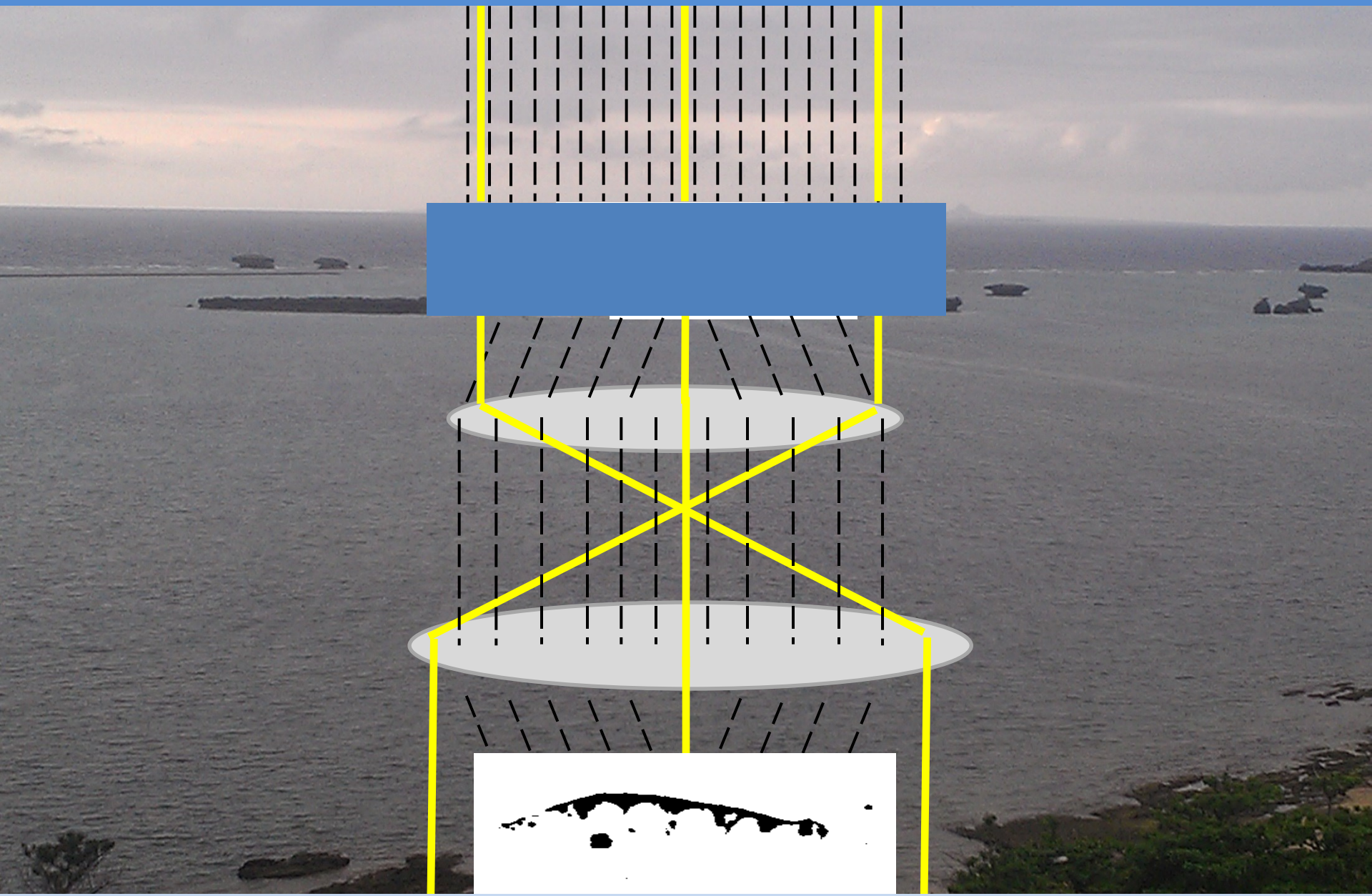


CMOS camera Since 2012

Ideal case: no sample motion

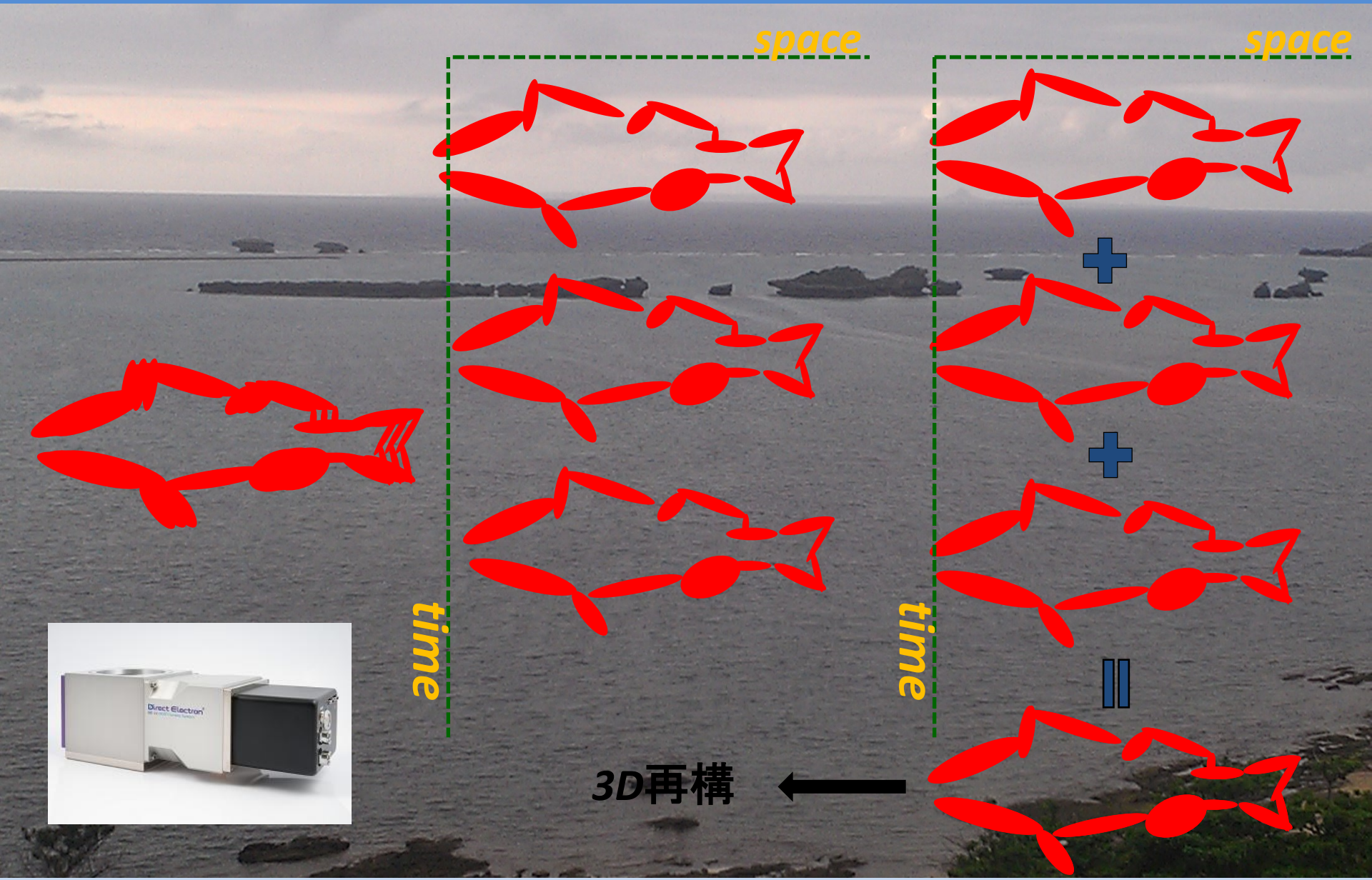


In-plane motion



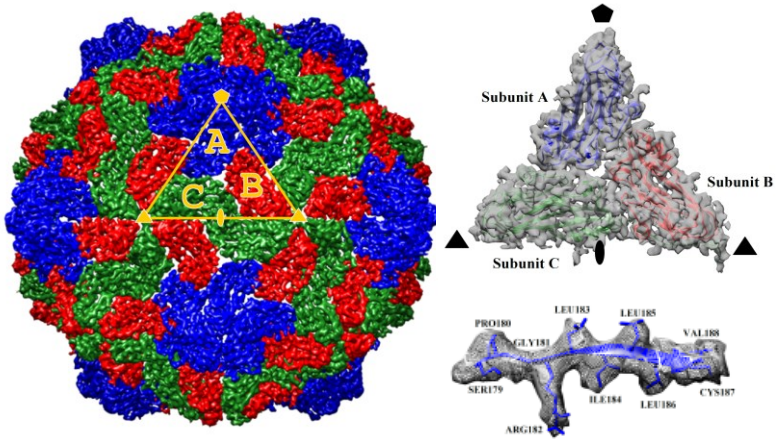
De-blurring by correcting in-plane motion

(Li and Cheng, 2013)

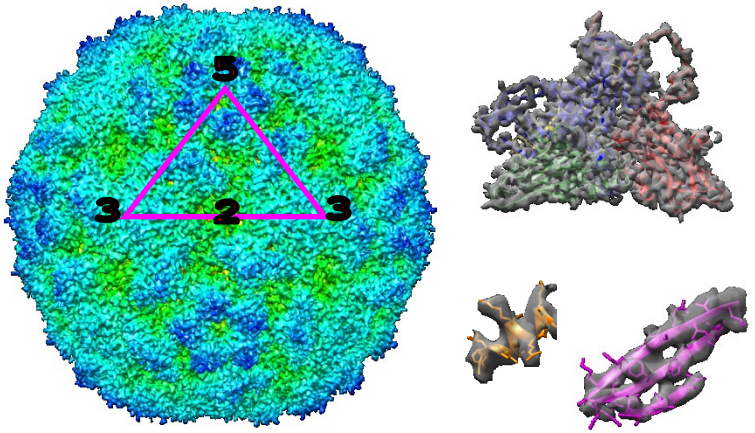


Achieving near atomic resolution with a direct electron camera (DE-20)

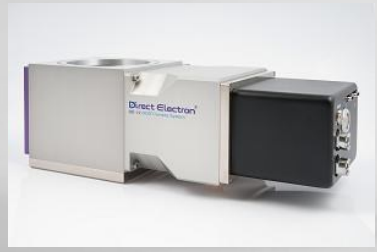
2013-2015: 3.6 Å DGNNV structure by cryo-EM

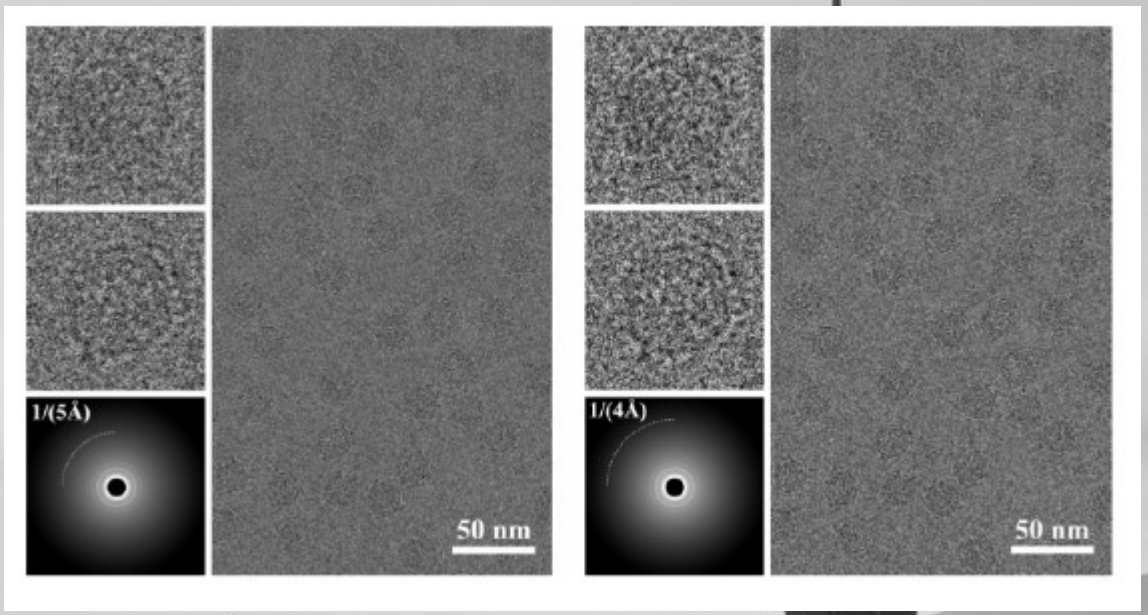


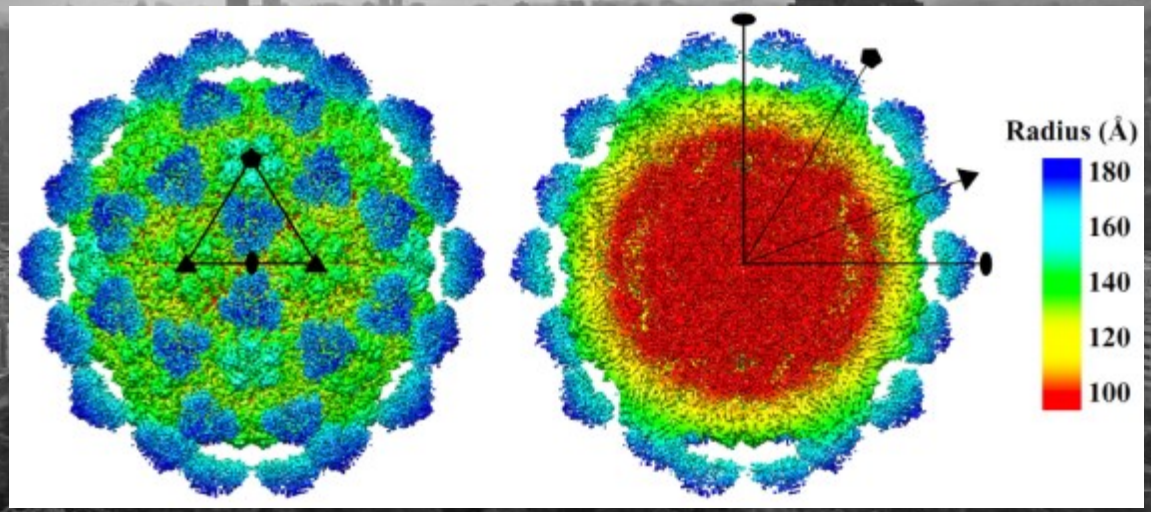
201 -2015: 3.6 Å EV71 structure by cryo-EM

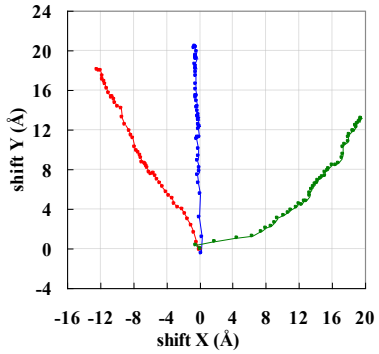


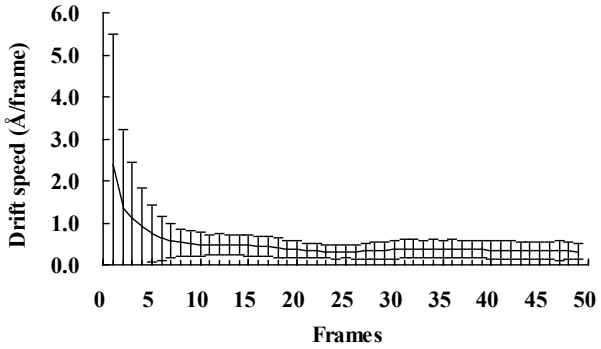
Reaching *better than 4 Å for icosahedral particle* (Wang May 2015, unpublished)

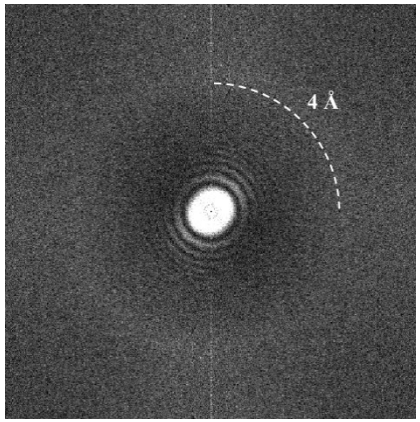


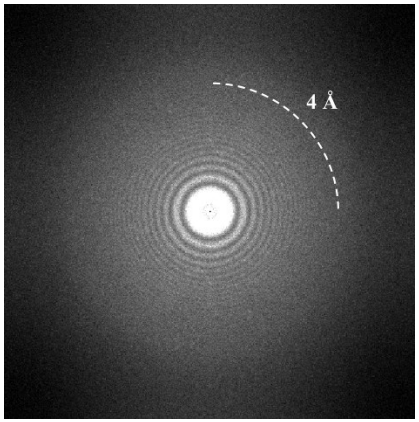


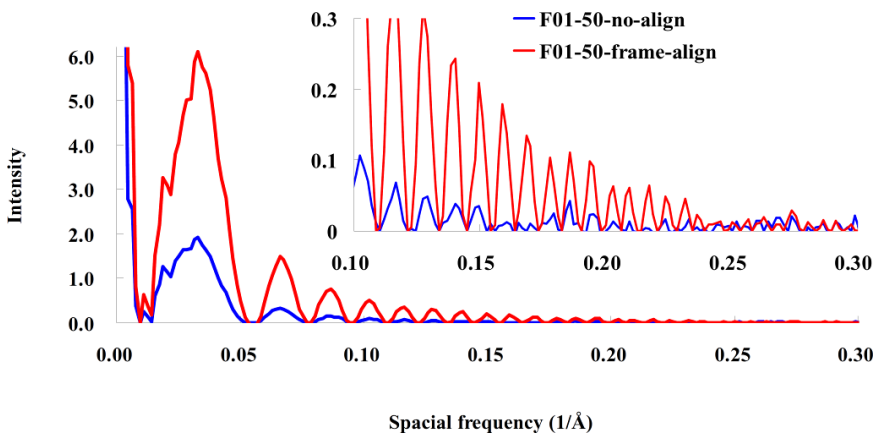



shift Y (Å)
shift X (Å)


Drift speed (Å/frame)
Frames


4 Å


4 Å


F01-50-no-align
F01-50-frame-align
Intensity
Spacial frequency (1/Å)

The power of sorting heterogeneity: Likelihood-based algorithm

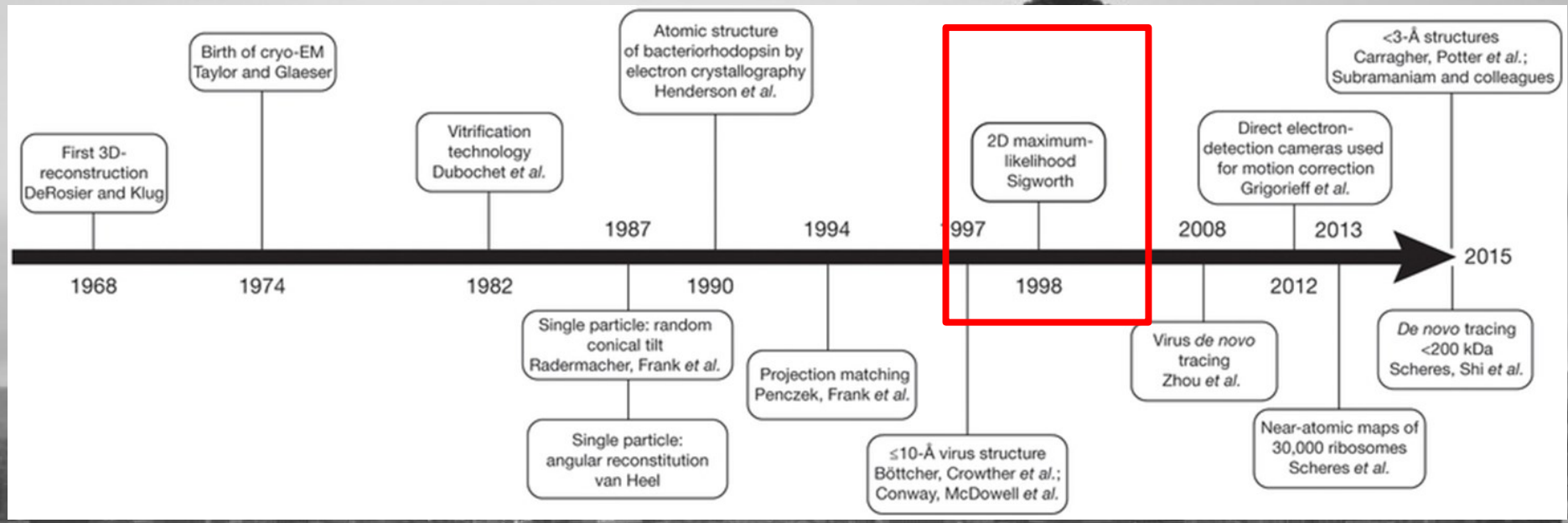
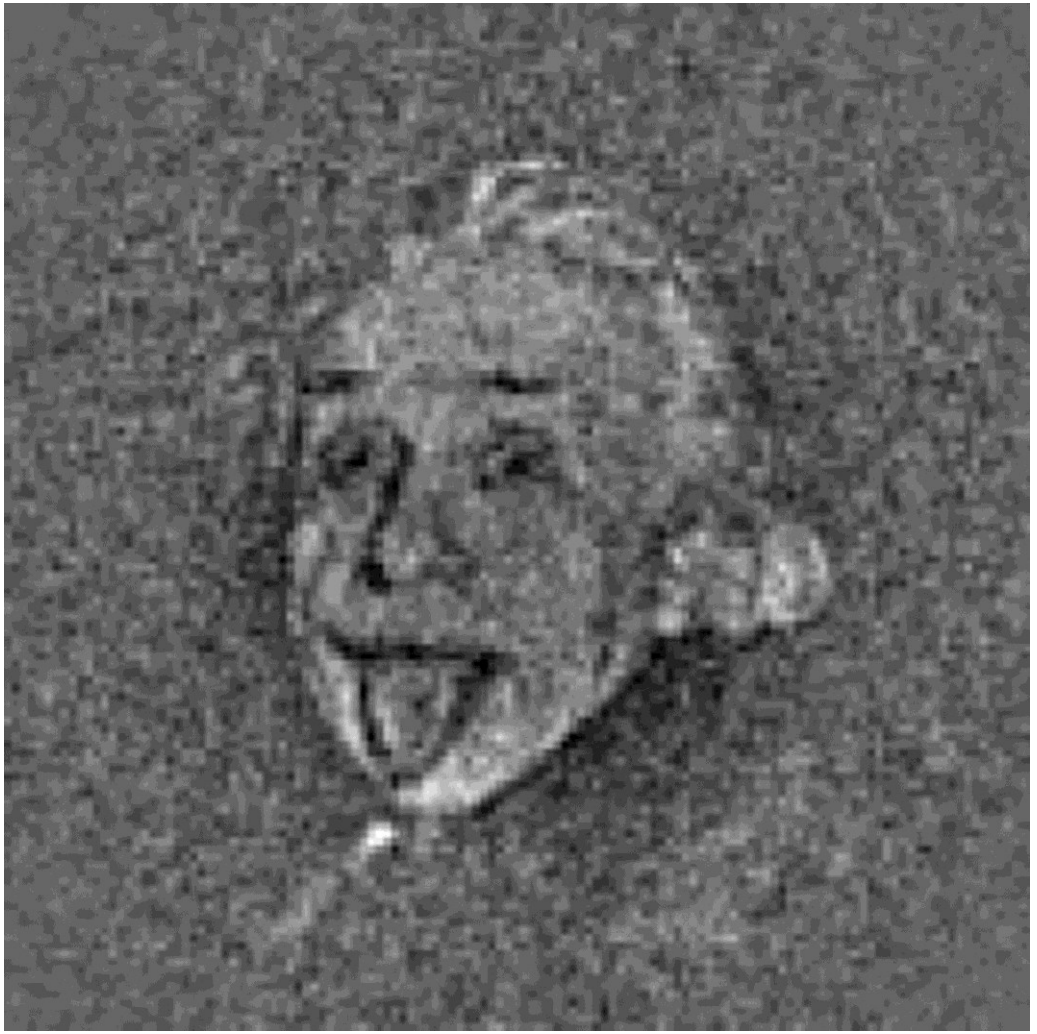


Illustration taken from a paper describing model bias



Richard Henderson PNAS 2013;110:18037-18041

the classical demonstration that the portrait of Einstein can be extracted from a few thousand images of pure random noise

©2013 by National Academy of Sciences

CHAPTER TEN

An Introduction to Maximum-Likelihood Methods in Cryo-EM

Fred J. Sigworth,[*] Peter C. Doerschuk,[†] Jose-Maria Carazo,[‡] *and* Sjors H. W. Scheres[‡,1]

[*] Department of Cellular and Molecular Physiology, Yale University, New Haven, Connecticut, USA
[†] Department of Biomedical Engineering, Cornell University, Weill Hall, Ithaca, New York, USA
[‡] Biocomputing Unit, Centro Nacional de Biotecnología – CSIC, Cantoblanco, Madrid, Spain
[1] Current address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK.

Methods in Enzymology, Volume 482
ISSN 0076-6879, DOI: 10.1016/S0076-6879(10)82011-7

© 2010 Elsevier Inc.
All rights reserved.

One example of straightforward likelihood optimization in cryo-EM is the estimation of the underlying signal from a series of noisy, structurally homogeneous, and aligned 2D images. Let us assume the following data model:

$$X_i = A + \sigma G_i, \quad \text{with } i = 1, \ldots, N, \tag{10.10}$$

$$P(X_i|\Theta) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{(X_{ij} - A_j)^2}{-2\sigma^2}\right\} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^J \exp\left\{\frac{\| X_i - A\|^2}{-2\sigma^2}\right\}, \tag{10.11}$$

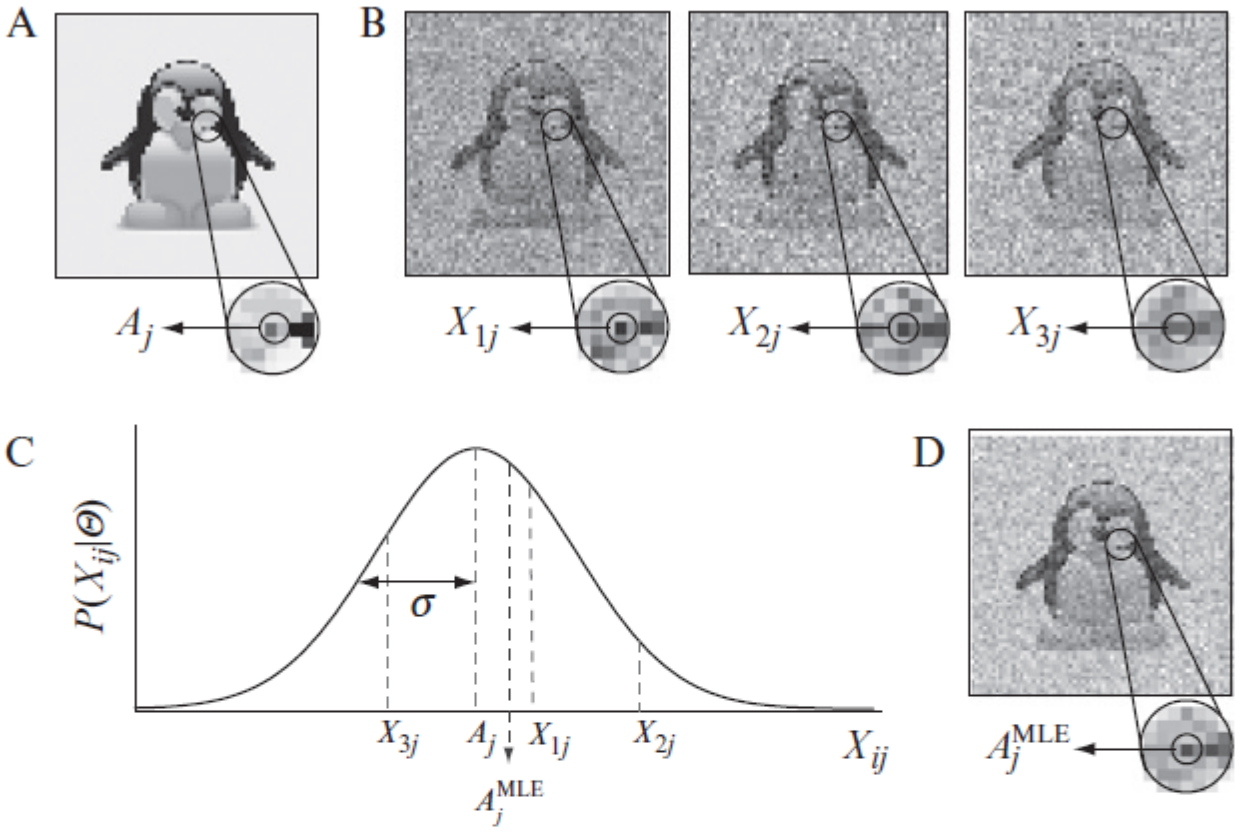


Figure 10.3 An example of direct MLE calculation. (A) An image A with a zoom window centered on one individual pixel value with value A_j is shown. (B) Three copies of image A with different instances of white Gaussian noise are shown together with zoom windows on the same pixel j with pixel values X_{1j}, X_{2j}, and X_{3j}. (C) the PDF of the jth pixel in the noisy images is shown as a Gaussian curve centered at A_j and with SD σ. Direct calculation of A_j^{MLE} is performed by averaging over X_{1j}, X_{2j}, and X_{3j}. (D) The MLE of the entire image A is shown together with a zoom window, centered on A_j^{MLE}. Note that the MLE of the image will approach the image in (A) if larger numbers of noisy images are available.

For any given transformation ϕ and parameter set Θ, the conditional probability of observing image X_i is again expressed as a multiplication over J Gaussian distributions, this time centered at the correspondingly oriented reference image $R_\phi A$:

$$P(X_i|\phi,\Theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^J \exp\left\{\frac{\| X_i - R_\phi A \|}{-2\sigma^2}\right\}. \qquad (10.17)$$

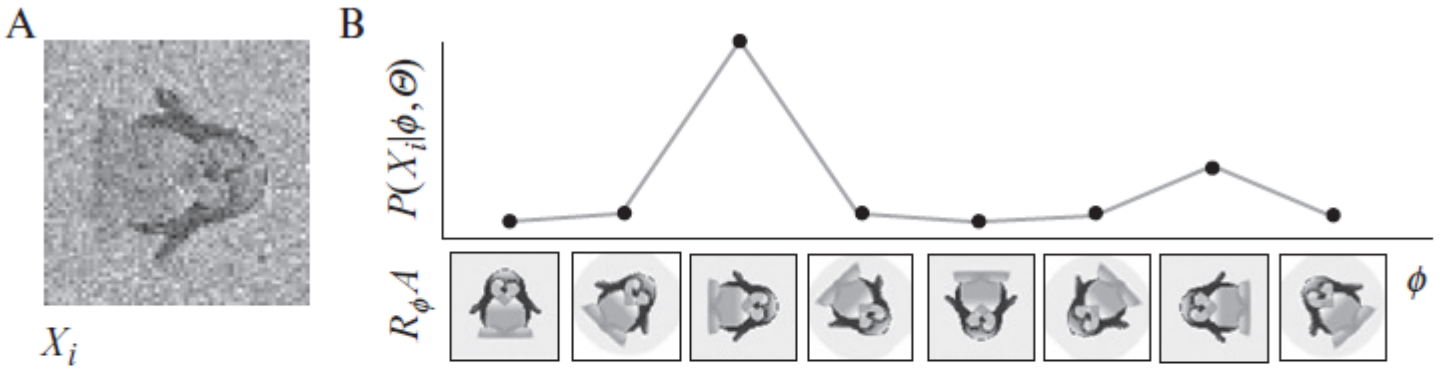


Figure 10.4 The PDF of a noisy image as a function of its relative orientation with respect to a reference image. (A) X_i, a rotated and noisy version of reference image A from Fig. 10.3A is shown. (B, top panel) The probability $P(X_i|\phi, \Theta)$ of observing X_i given a model Θ that comprises image A is shown (on an arbitrary scale) as a function of the relative orientation ϕ. The bottom panel shows $R_\phi A$, that is, image A rotated according to ϕ. Note that $P(X_i|\phi, \Theta)$ is highest when ϕ corresponds to the correct orientation of X_i.

The incompleteness of the observed data lies in the fact that the relative orientations of all images have remained unobserved in the experiment. The complete data set would be $(\mathcal{X}, \mathcal{Y})$, with $\mathcal{Y} = (\phi_1, \phi_2, \ldots, \phi_N)$, and finding the MLE for the complete data set would be as trivial as described in the simple example above.

For the incomplete case, the marginal log-likelihood function, cf. Eq. (10.14), is given by:

$$L(\Theta) = \sum_{i=1}^{N} \log \int_{\phi} P(X_i|\phi, \Theta) P(\phi|\Theta) \mathrm{d}\phi. \tag{10.16}$$

Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization

Sjors H W Scheres[1], Haixiao Gao[2], Mikel Valle[1,5], Gabor T Herman[3], Paul P B Eggermont[4], Joachim Frank[2] & Jose-Maria Carazo[1]

Although three-dimensional electron microscopy (3D-EM) permits structural characterization of macromolecular assemblies in distinct functional states, the inability to classify projections from structurally heterogeneous samples has severely limited its application. We present a maximum likelihood–based classification method that does not depend on prior knowledge about the structural variability, and demonstrate its effectiveness for two macromolecular assemblies with different types of conformational variability: the *Escherichia coli* ribosome and Simian virus 40 (SV40) large T-antigen.

Conformation and composition heterogeneity problem

"Cryo-EM will Make a very big Difference"

(February 6th, 2015) **It is rumoured that structural biology is undergoing a revolution. Once dominated by X-ray crystallography methods, cryo-electron microscopy (cryo-EM) is now transforming the field. Sjors Scheres is one of the driving forces behind this revolution.**

LT: *How did **RELION** change the field of structural biology?*
The field of cryo-EM is undergoing a revolution. There are two reasons for that: one is the development of a new detector for electrons, and that I think is the most important one. Before people were taking pictures on photographic film, or on CCD cameras, which would introduce quite a bit of noise in the imaging process. The other thing that has improved is better image processing and RELION fits into that category. What's special about RELION is that it can find an optimal way of filtering the data automatically, whereas with older programmes, you had to be quite experienced in image processing to get good results. With RELION you don't need to be an expert, that's the main contribution. Also you can separate distinct 3D structures from a single dataset with RELION, that's very important. At the moment, we're working on making the technique work for smaller protein complexes, and also for those that are floppy, things that have not just two or three conformations, but a whole continuous spectrum of conformations

Likelihood estimator formalism *(extracted from Lecture 7)*

The incompleteness of the observed data lies in the fact that the relative orientations of all images have remained unobserved in the experiment. The complete data set would be $(\mathcal{X}, \mathcal{Y})$, with $\mathcal{Y} = (\phi_1, \phi_2, \ldots, \phi_N)$, and finding the MLE for the complete data set would be as trivial as described in the simple example above.

For the incomplete case, the marginal log-likelihood function, cf. Eq. (10.14), is given by:

$$L(\Theta) = \sum_{i=1}^{N} \log \int_{\phi} P(X_i|\phi, \Theta) P(\phi|\Theta) \mathrm{d}\phi. \qquad (10.16)$$

X: experimental observation;

ϕ: the best angles to be determined;

θ: the best 3D reconstruction to be obtained

Extending the "Likelihood estimator" *to cope with multiple conformation classes*

$$L(\Theta) = \sum_{i=1}^{I} \ln \sum_{k=1}^{K} \int_{\varphi} P(X_i | k, \varphi, \Theta) P(k, \varphi | \Theta) d\varphi$$

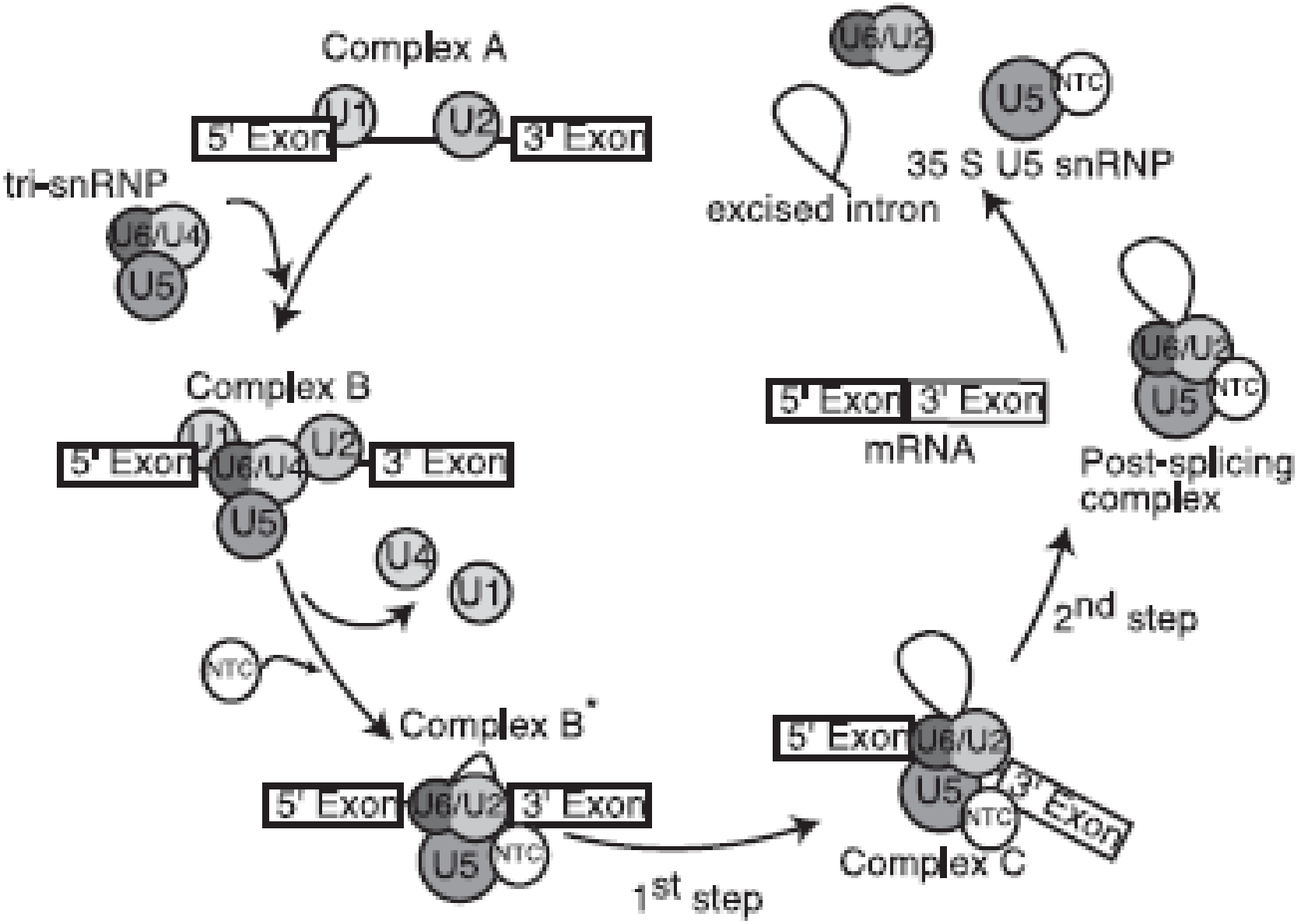
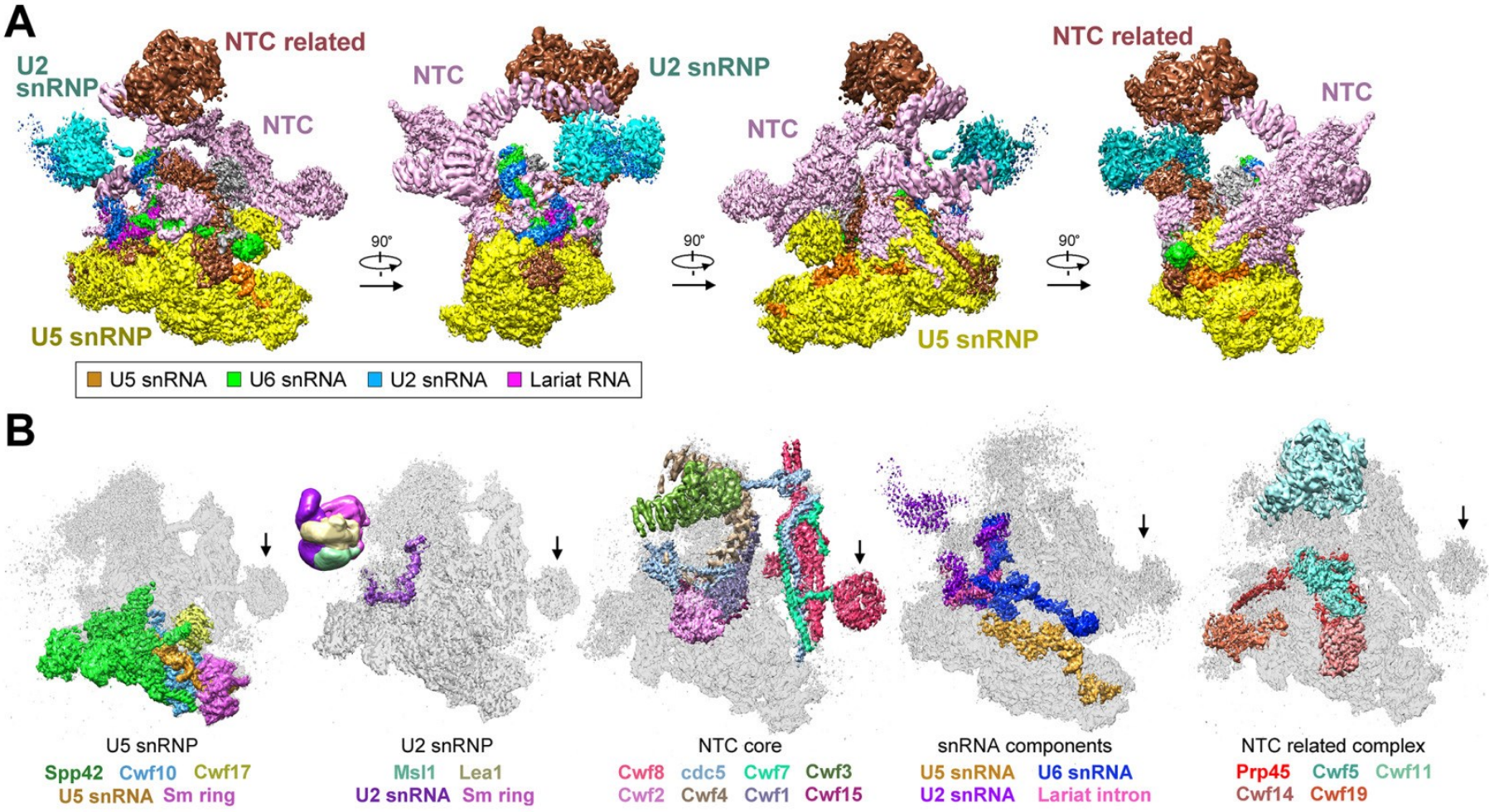
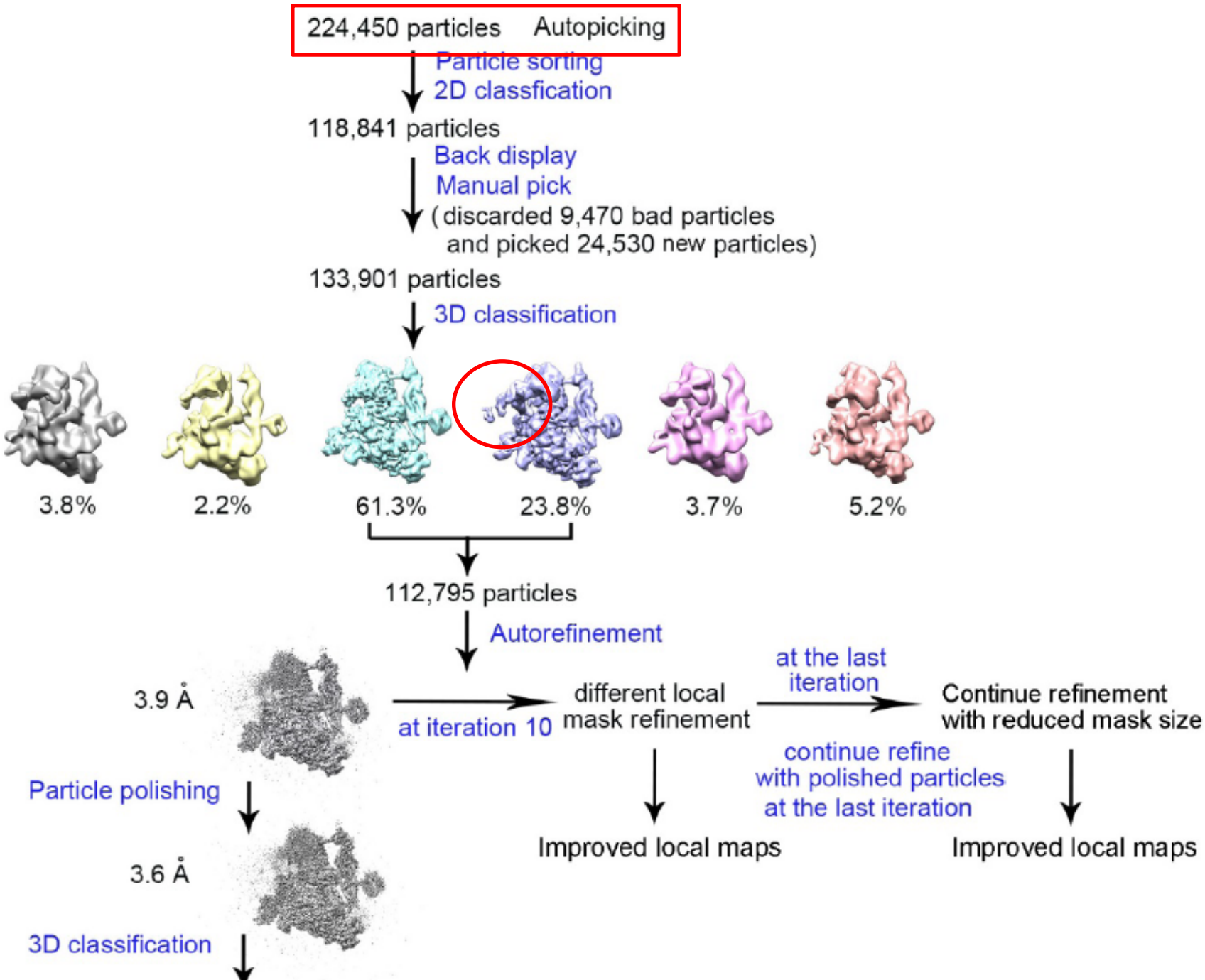


Fig. 1. Schematic model of pre-mRNA splicing.

China has risen: Yigong Shi and the world largest cryo-EM center at Beijing!

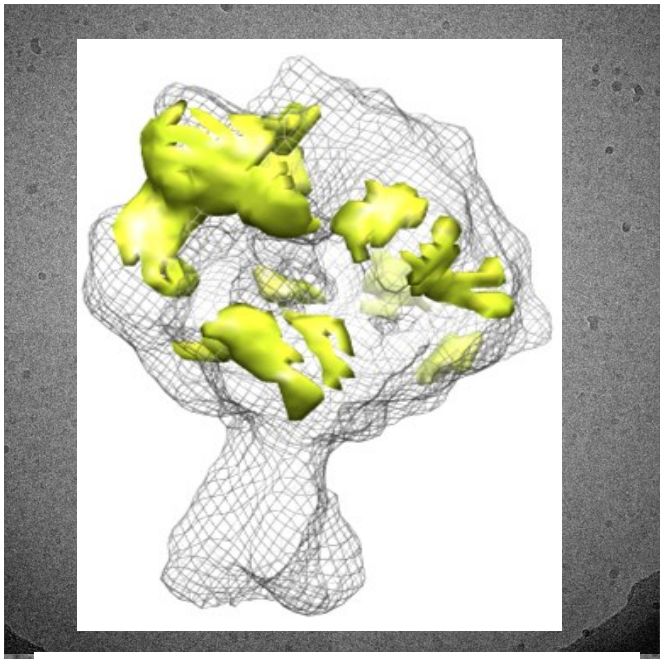
Structure of multi-component sub-complexes

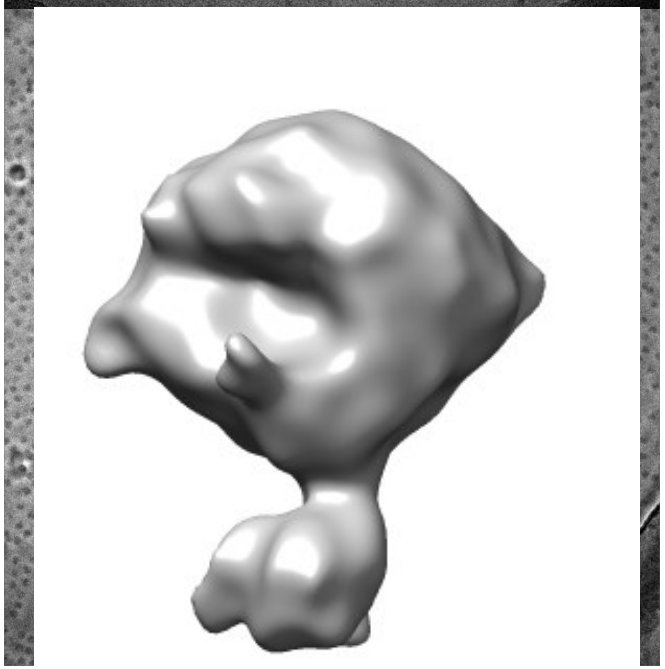



224,450 particles
Autopicking
Particle sorting
2D classfication
118,841 particles
Back display
Manual pick
(discarded 9,470 bad particles
and picked 24,530 new particles)
133,901 particles
3D classification
3.8%
2.2%
61.3%
23.8%
3.7%
5.2%
112,795 particles
Autorefinement
3.9 Å
at iteration 10
different local
mask refinement
at the last
iteration
Continue refinement
with reduced mask size
continue refine
with polished particles
at the last iteration
Particle polishing
3.6 Å
3D classification
Improved local maps
Improved local maps

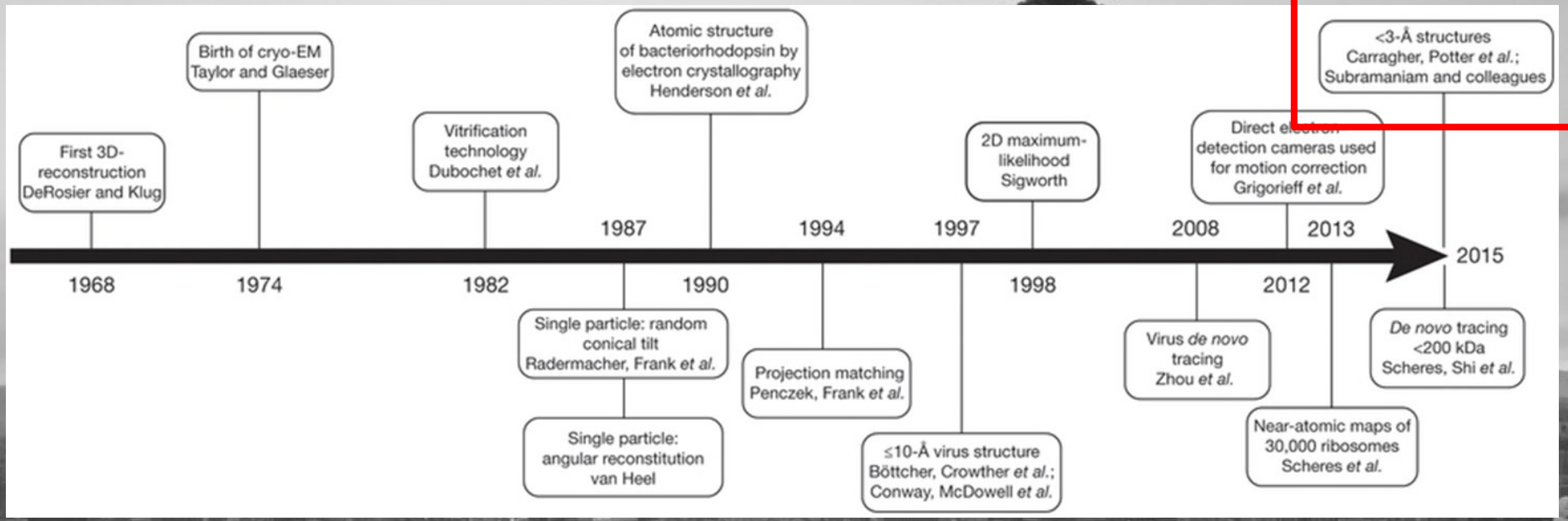
A 200 kV Field-emission cryo-EM

Wu et al., EMBO J 31 (2012)

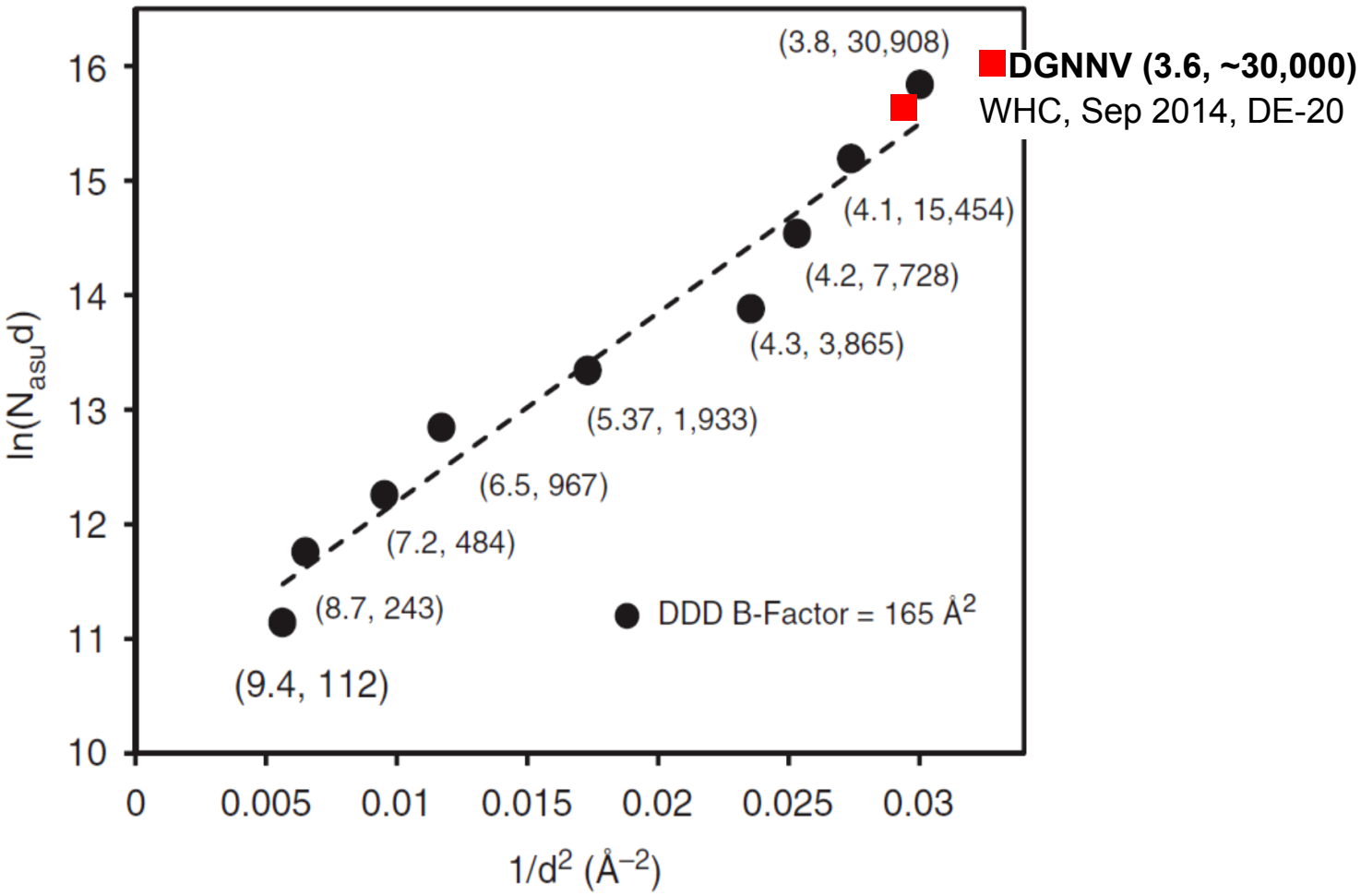




The power of "automated" cryo-EM



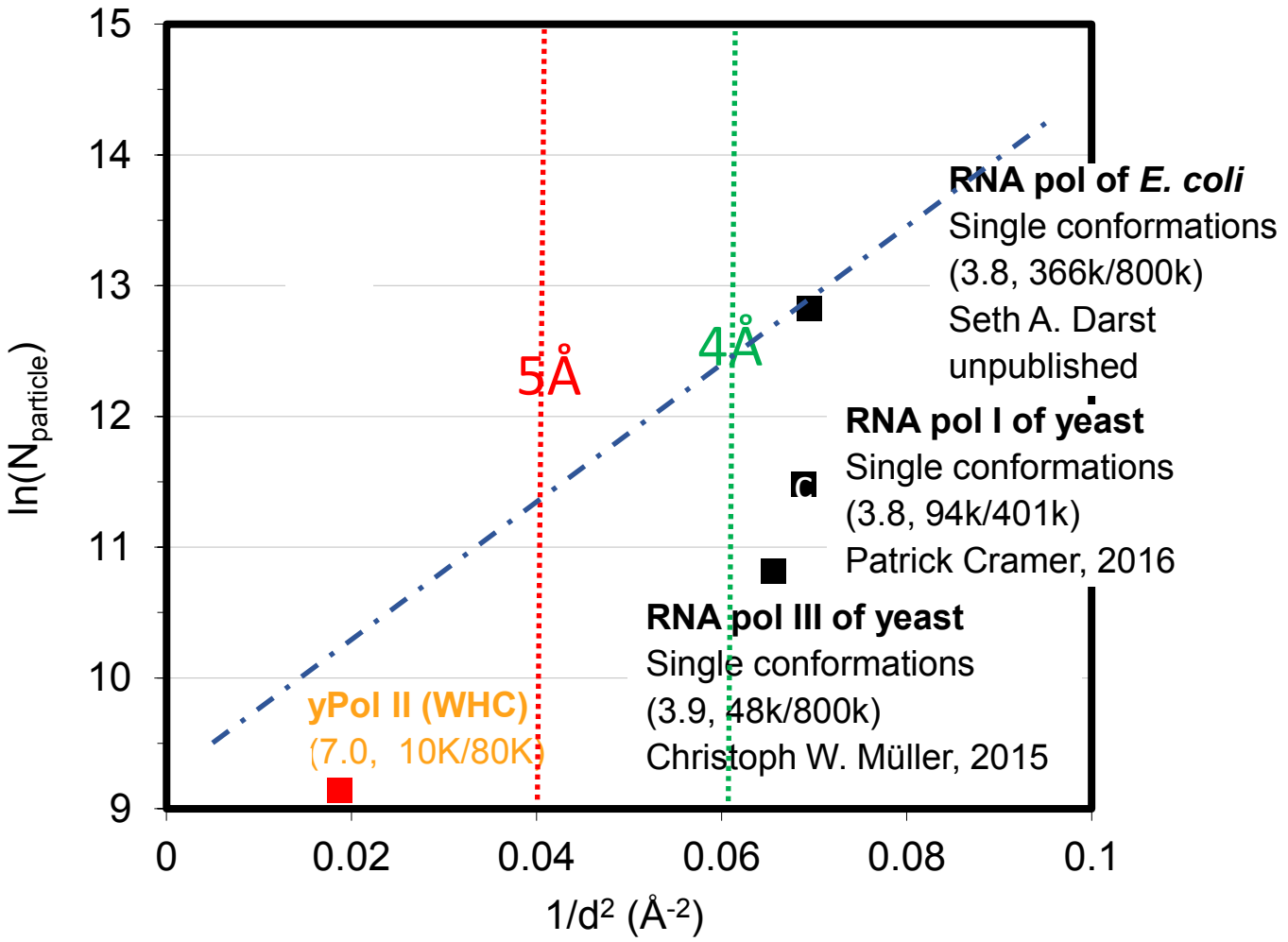
Resolution *vs*. particle number (Case for icosahedral particle)



DE-12, JEM3200FSC. Wang Z… Chiu W. Nat Comm. 2014

Life without an automated machine:
Two weeks to get 30,000 icosahedral particles

Resolution *vs*. ASU number (Case for protein complex)



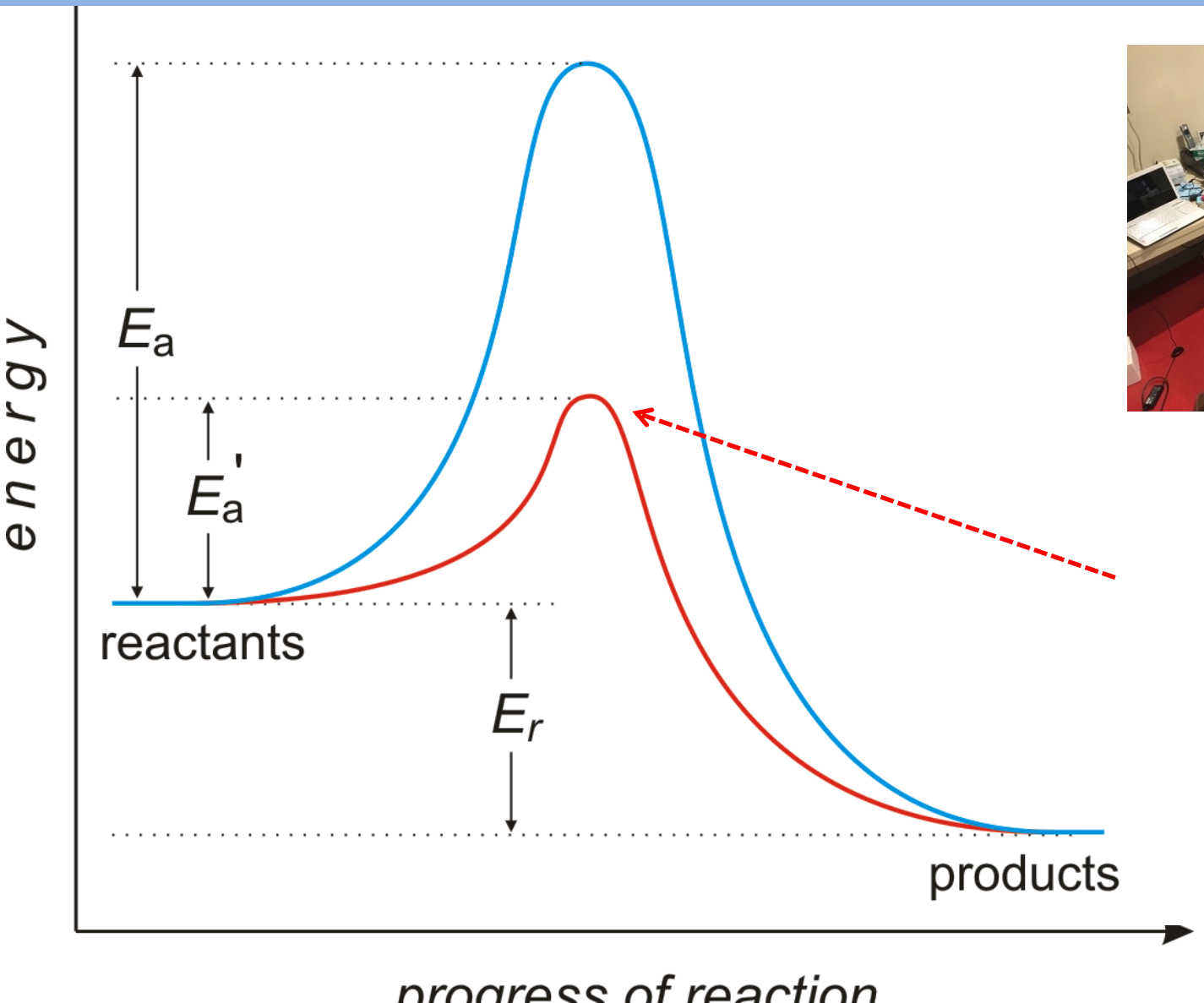
Do we still need automated cryo-EM?
(Chang et al., 2010)

Table 2. Total Number of Particles Required to Reach a Target Resolution

				Defocus (μm) (CEM)			
	Resolution (Å)	Ideal	ZEM (30% loss)	0.25	0.5	1	2
pol II	11	~800	~1,570	NA	~3,200	~2,200	~3,200
	4.5	~4,800	~15,000	<500,000	~38,000	~48,000	~240,000
	3.3	~13,000	~60,000	<500,000	~260,000	~400,000	>500,000
TfR	11	~800	~1,400	NA	NA	~3,900	~6,400
	4.5	~5,000	~16,000	NA	NA	>600,000	>600,000
	3.3	~21,000	~48,000	NA	NA	>1,000,000	>1,000,000
T7 pol-lys	11	~4,000	~6,400	NA	NA	~63,000	~160,000
	4.5	~21,000	~62,000	NA	NA	>1,000,000	>1,000,000
	3.3	~66,000	~160,000	NA	NA	>1,000,000	>1,000,000

Prediction made on the basis of single conformation & 90 degree phase shift (Chang et al., 2010)

Atomic cryo-EM requires huge amount of image DATA: "Titan Krios" has catalyzed the DATA FLUX



“Titan Krios” brings robot-revolution and fierce competition follows suit

Very stable optics

Auto-loading

Multiple specimens

Remote control

High-throughput:
2000 movies/day (10T)

24/7

Conclusions

Direct electron camera broken cryo-EM resolution.

Likelihood algorithm disentangles the mixed conformation or composition problem in solution.

The more the better: Titan Krios is a robot or beamline!

Cryo-EM demands huge computation resources for processing.