Image Alignment and Application of 2D Fast Rotational Matching in Single Particle cryo-Electron Microscopy

Yao Cong

Baylor College of Medicine

deus ex biomachina

biomachina.org

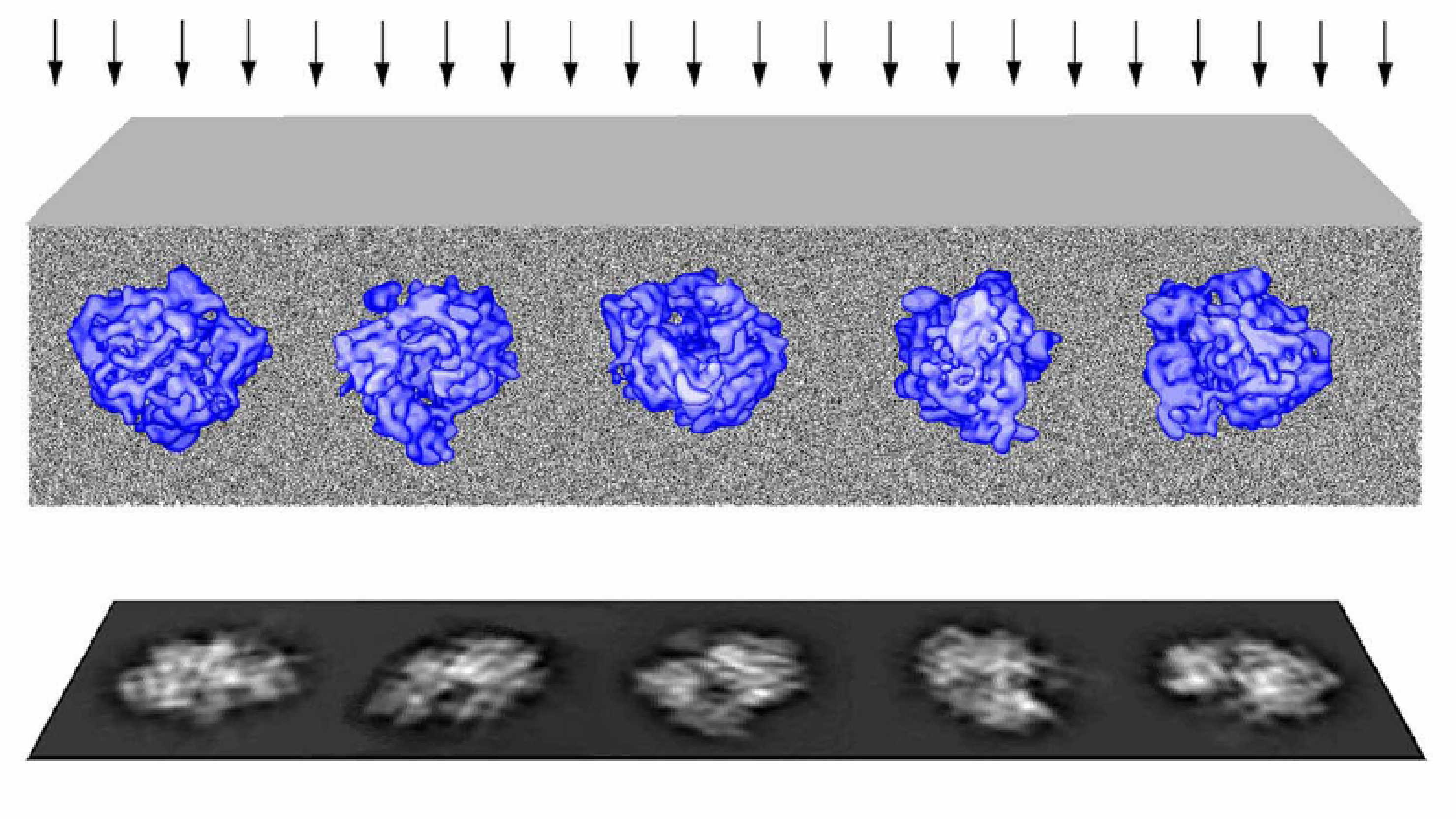
NCMI

National Center for Macromolecular Imaging

cryo-EM & single particle analysis

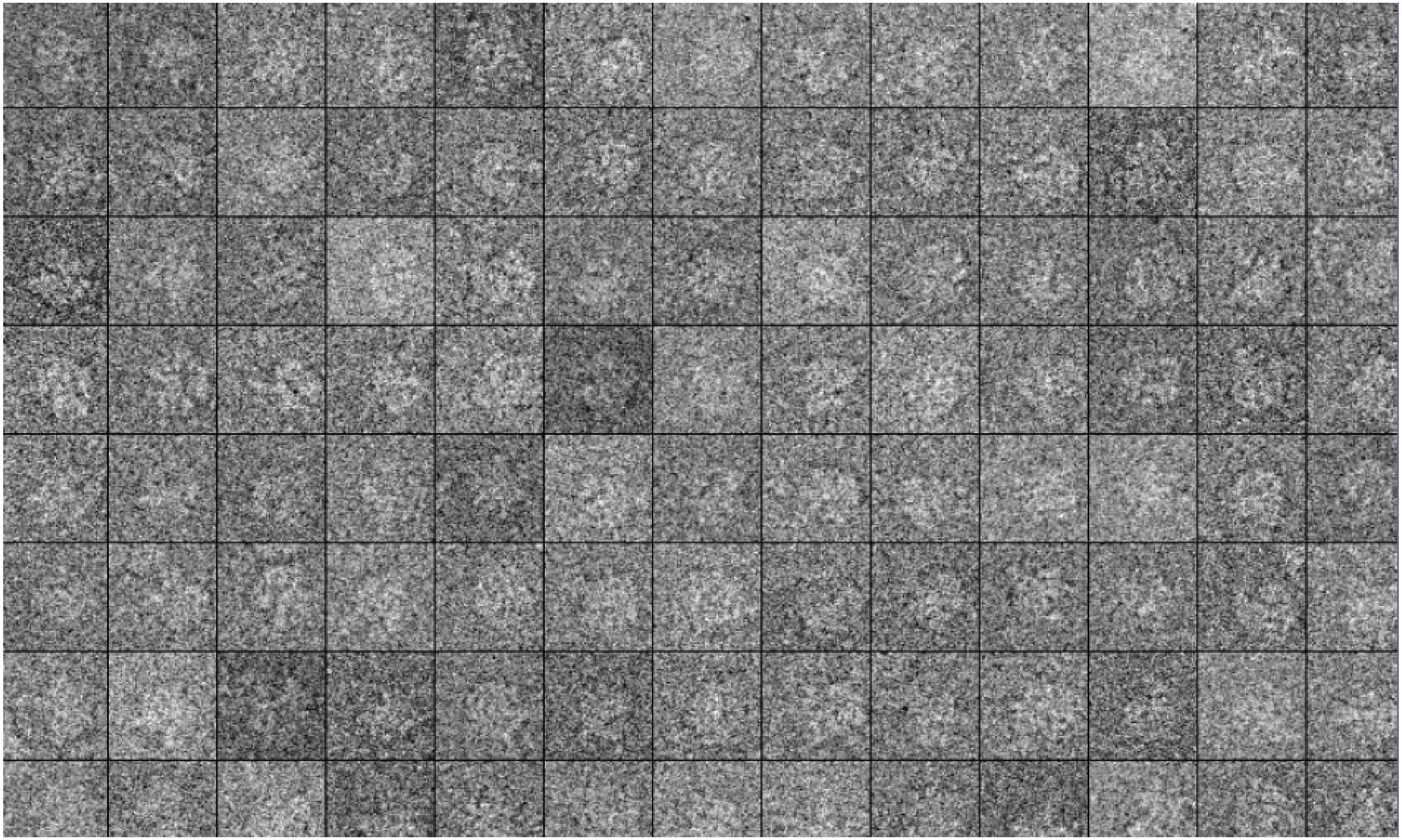
- Single-particle electron microscopy is an excellent way to characterize the overall structure of multiprotein complexes
- Identify relative positions of individual components
- Study the dynamics of macromolecules
- With resolution still improving, single-particle analyses are already depicting secondary structures. (4-5Å)

What is Observed in Single Particle Imaging



- Single particle: particles assume random orientations in vitrified ice
- 2D views of mass density of individual proteins in random orientations

cryo -EM Micrograph of Single Particles



electron beam is damaging to biological samples → to minimize the dose on the specimen
→ the amount of signal present in any individual image is low

Classification of Images

- The primary process of single-particle analysis is the classification of images according to their Euler angles

- Since the low SNR, the images in each classified group then being averaged to enhance the single

- Classification methods are divided into “supervised” and “unsupervised”

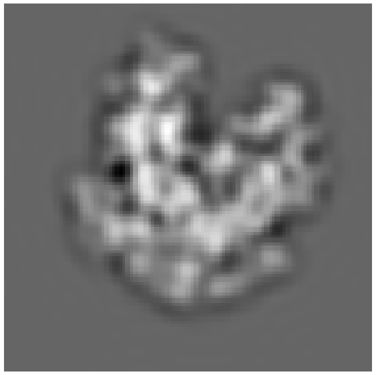
- Supervised: divide according to similarity with “template” or “reference”. → homogeneous

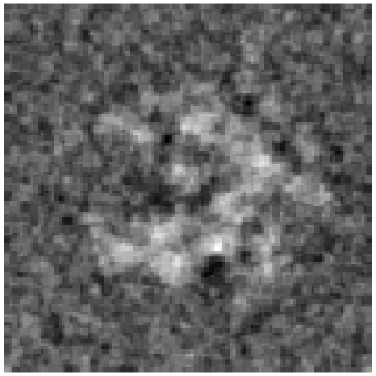
- Unsupervised: divide according to intrinsic properties. →heterogeneous

Frank et al., 1978; van Heel and Frank, 1981

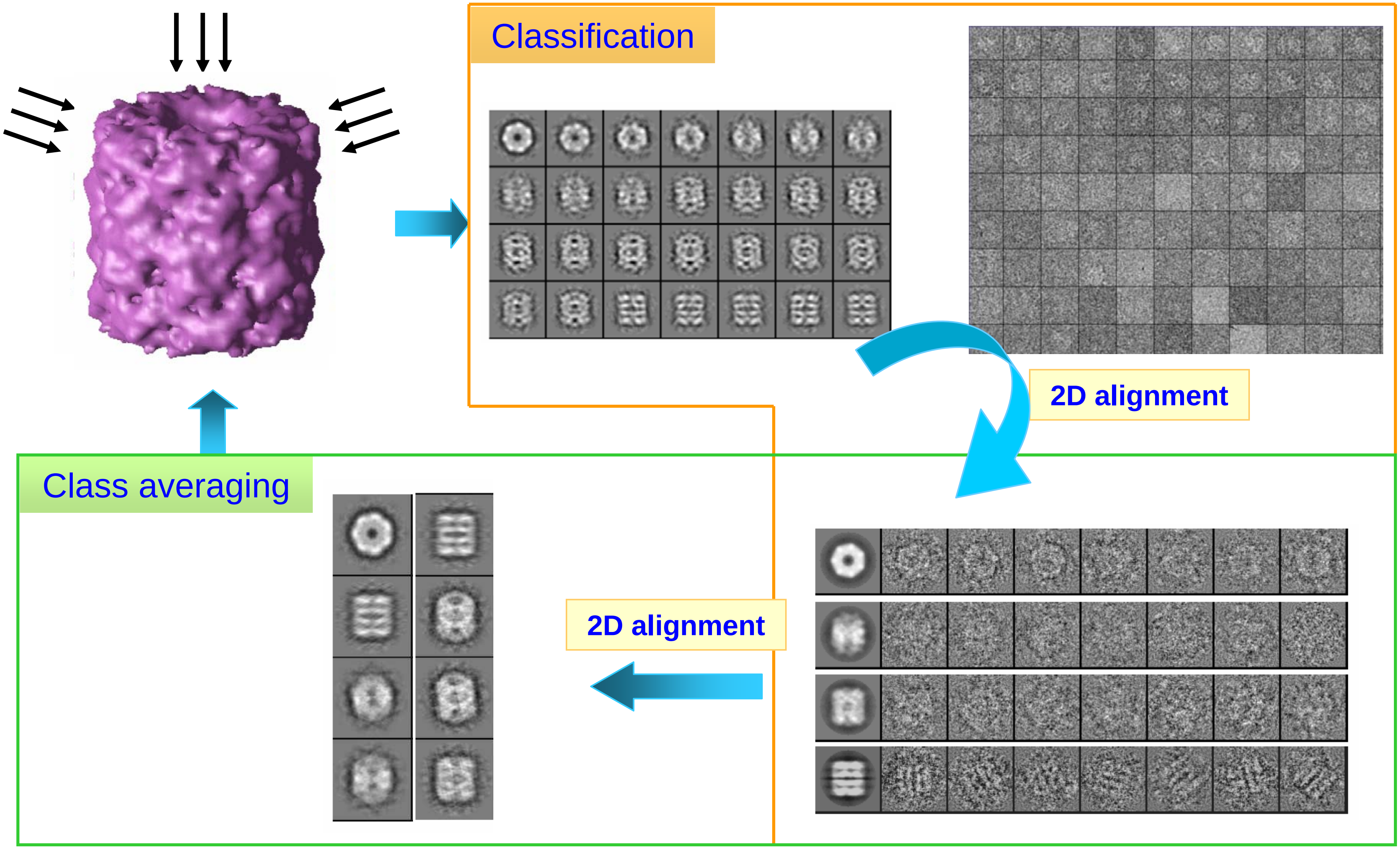
2D Image Alignment Introduction

- **2D alignment**: determines the 3 relative transformation parameters of two images (1 rotat. & 2 trans.)





Model-Based refinement strategy



Classification

2D alignment

Class averaging

2D alignment

2D Image Alignment Introduction

- **Application:** in model-based single particle analysis
    - ***Classification***
    - ***Class averaging***
    - ***3D reconstruction***
- **Accuracy:** a critical factor to obtain a faithful 3D reconstruction structure
- **Efficiency:** a limiting factor for the improvement of resolution (cryo-EM, 1/4Å resolution → 10^6 particle images)

2D Image Alignment Introduction

- Various 2D alignment methods
  - *direct alignment in real space*
  - *direct alignment using 2D FFT*
  - *sinograms*
- Commonly used 2D alignment methods:

→ *Resampling to Polar Coordinates (RPC)*

Polar Fourier Transform (PFT)

FFT accelerate 1D rotation search

→ *Self-Correlation Function (SCF):*

SCF: inverse FT of the amplitude spectrum of the image

SCF of an image is invariant with respect to the translation of this image

L. Joyeux, P. A. Penczek, Ultramicroscopy 92(2002) 33

2D Image Alignment Introduction

- Current state of the art

  → **RPC/PFT** is the most accurate method.

  → **RPC/PFT** is the most efficient one if the translational search can be restricted to a small value;

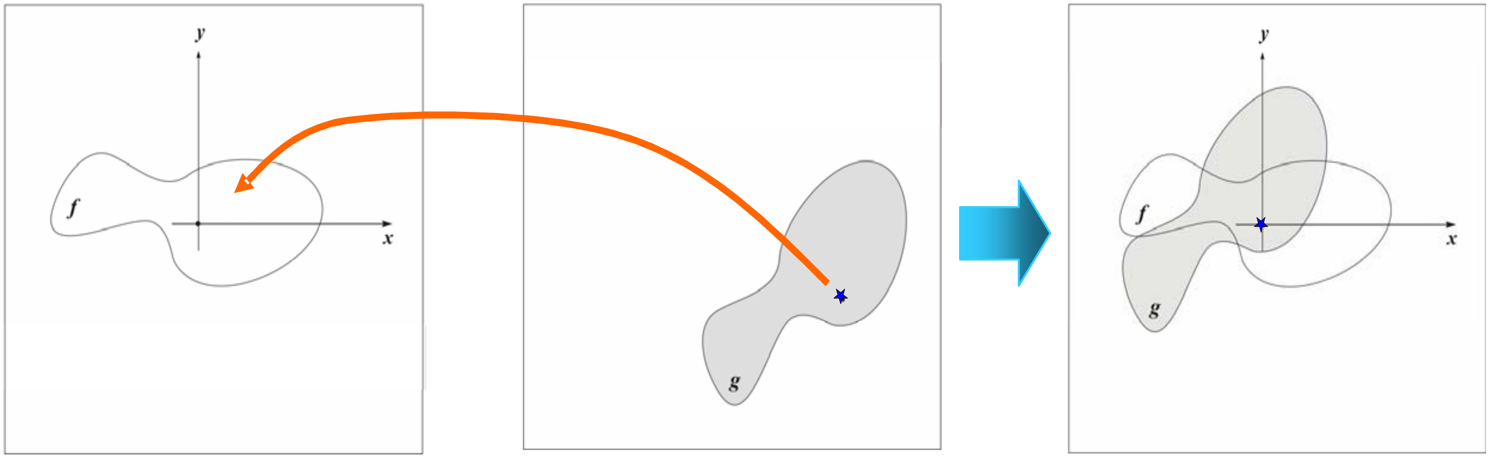
  otherwise, **SCF** ranks the most efficient one.

- Classical methods describe the 3 transformation parameters of the images by 1 rota. + 2 trans. parameters ➡ fix one image & (rota. + trans.) the other

L. Joyeux, P. A. Penczek, Ultramicroscopy 92(2002) 33

FRM2D Description

Idea: Rotate both objects around their own center of mass, while translate one object along the positive x axis, until find the best matching position.



FRM2D: 2 rota.+1 trans.

- 2D FFT accelerate 2 rota. param. search
- Avoid expensive zero padding

movie

FRM2D Description

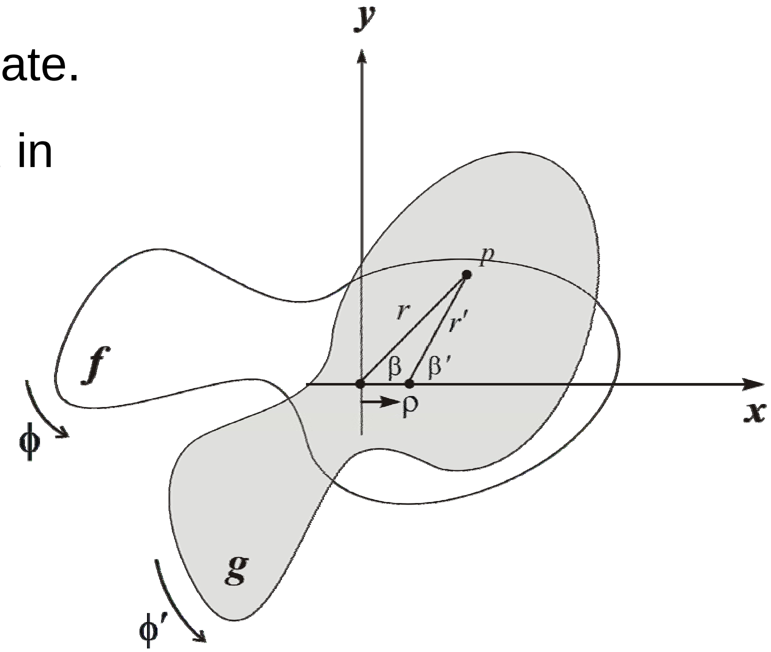
Resampling the objects in Polar coordinate.

The given density objects are expanded in Fourier series:

$$f(r,\beta) = \sum_m \hat{f}_m(r) e^{im\beta}$$

$$g(r,\beta) = \sum_n \hat{g}_n(r) e^{in\beta}$$

$\hat{f}_m$ are *the Fourier coefficients*



FRM2D Description

Rotated and translated objects:

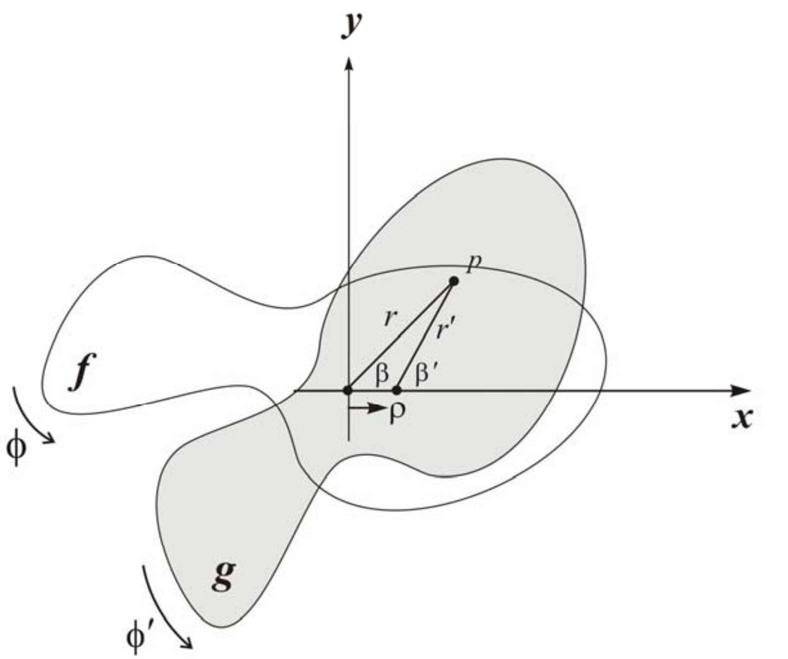
$$f(\phi)(r,\beta) = \sum_m \hat{f}_m(r) e^{im(\beta-\phi)}$$

$$g(\phi';\rho)(r,\beta) = \sum_n \hat{g}_n(r') e^{in(\beta'-\phi')}$$

The correlation function is a function of 2

rotations and 1 distance:

$$c(\phi,\phi';\rho) = \int_{\mathbf{R}^2} f(\phi)\cdot g(\phi';\rho)$$



The criterion for matching of the two objects is to maximize the correlation between them.

FRM2D Description

Correlation function:

$$c(\phi,\phi';\rho) = \sum_{m,n} e^{i(m\phi+n\phi')} I_{mn}(\rho)$$

$$\Big\downarrow$$

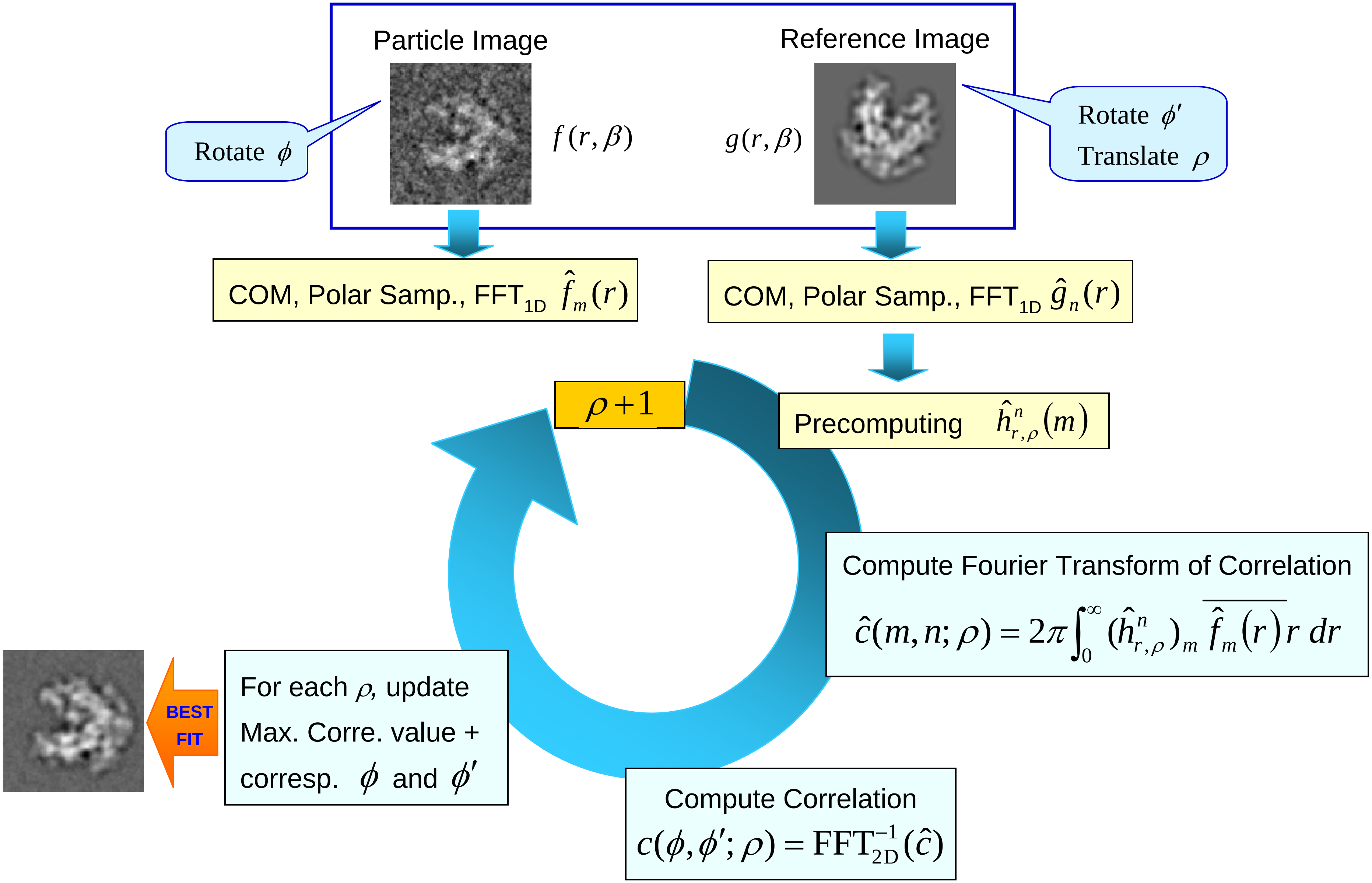
2D FT of correlation function:

$$\hat{c}(m,n;\rho) = I_{mn}(\rho) = 2\pi \int_0^\infty (\hat{h}^n_{r,\rho})_m \overline{\hat{f}_m(r)}\, r\, dr$$

$$\boxed{\hat{c}(m,n;\rho) \xrightarrow{\text{Inverse FFT}} c(\phi,\phi';\rho)}$$

Cong, Y., Kovacs, J. A., Wriggers, W., 2003. J. Struct. Biol. 144, 51-60

How can we implement the FRM2D search?



Particle Image

$f(r,\beta)$

Reference Image

$g(r,\beta)$

Rotate ϕ

Rotate ϕ'

Translate ρ

COM, Polar Samp., FFT$_{1D}$ $\hat{f}_m(r)$

COM, Polar Samp., FFT$_{1D}$ $\hat{g}_n(r)$

Precomputing $\hat{h}^n_{r,\rho}(m)$

$\rho+1$

Compute Fourier Transform of Correlation

$$\hat{c}(m,n;\rho)=2\pi\int_0^\infty (\hat{h}^n_{r,\rho})_m \overline{\hat{f}_m(r)}\, r\, dr$$

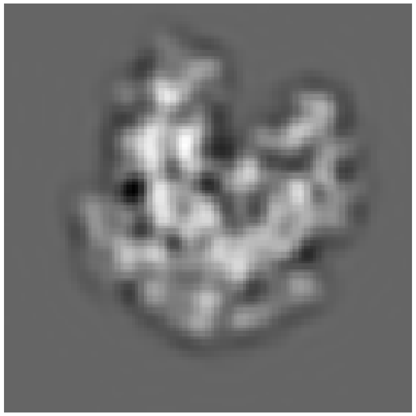
Compute Correlation

$$c(\phi,\phi';\rho)=\mathrm{FFT}^{-1}_{2D}(\hat{c})$$

For each ρ, update Max. Corre. value + corresp. ϕ and ϕ'

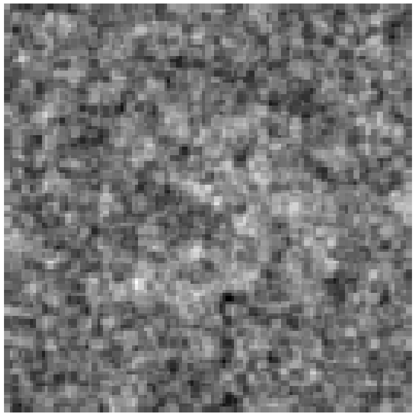
FRM2D Accuracy and Efficiency Performance Test

Efficiency test images (RNA polymerase)



(a) reference image

104 x 104 pixels



(b) raw image

(100º; 6, -4)

SNR=0.1477

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

L. Joyeux, P. A. Penczek, Ultramicroscopy 92(2002) 33

What effects the efficiency of 2D alignment?

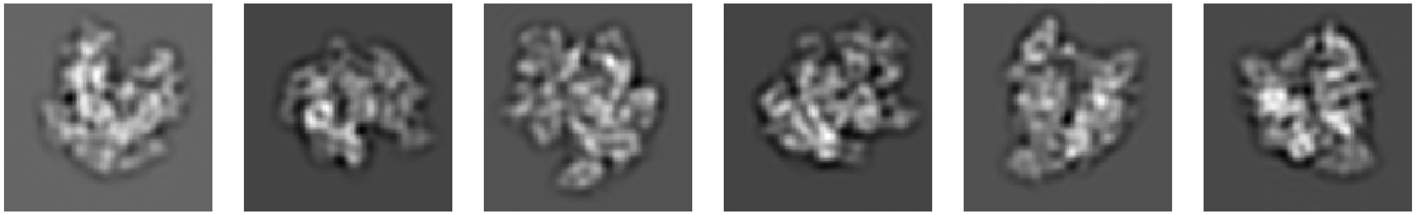
	Angular sampling	4000 particle images		
		FRM2D	RPC	SCF
K=4	11°	79.98	447.02	193.39
	6°	174.89	736.52	201.19
	3°	542.93	1552.72	214.87
	1.4°	2044.64	2875.23	238.34
K=10	11°	104.16	1958.55	193.39
	6°	291.69	3253.71	201.19
	3°	1063.13	6708.58	214.87
	1.4°	4183.73	13694.25	238.34

* timings in sec.

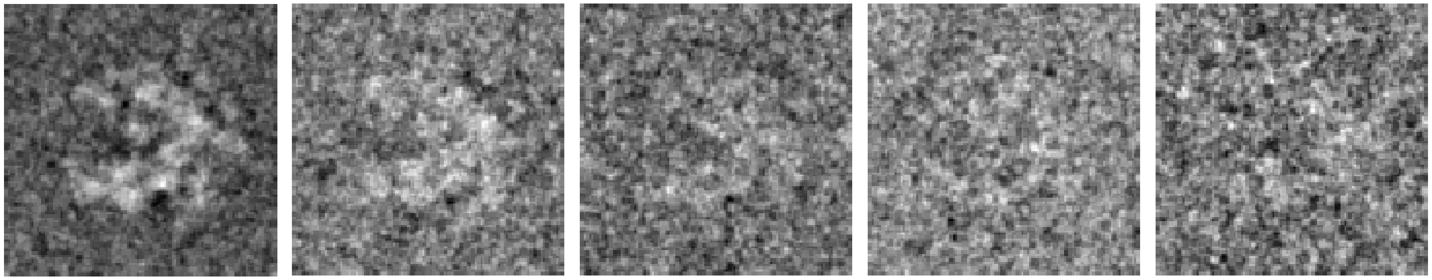
$$\boxed{l = (2k+1)^2}$$

linear scan parameter was set to 4 or 10 pixels

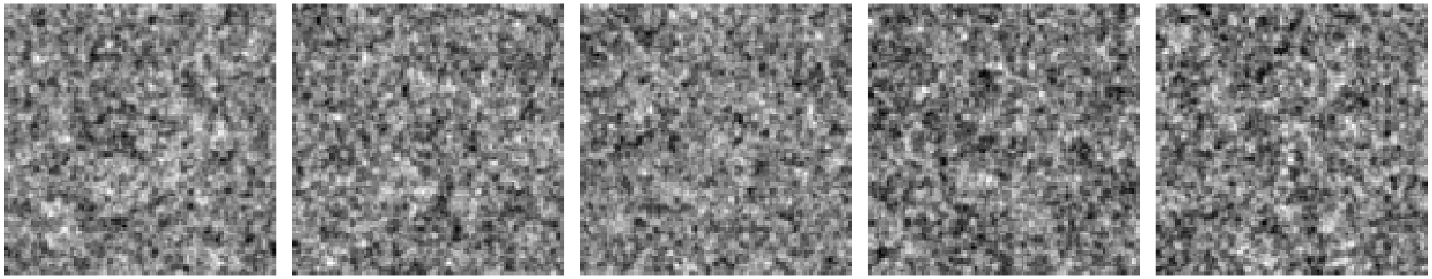
Class averaging accuracy test images



(a)



(g) (SNR=1.3716) **(h) (SNR=0.3350)** **(i) (SNR=0.1477)** **(j) (SNR=0.0847)** **(k) (SNR=0.0544)**



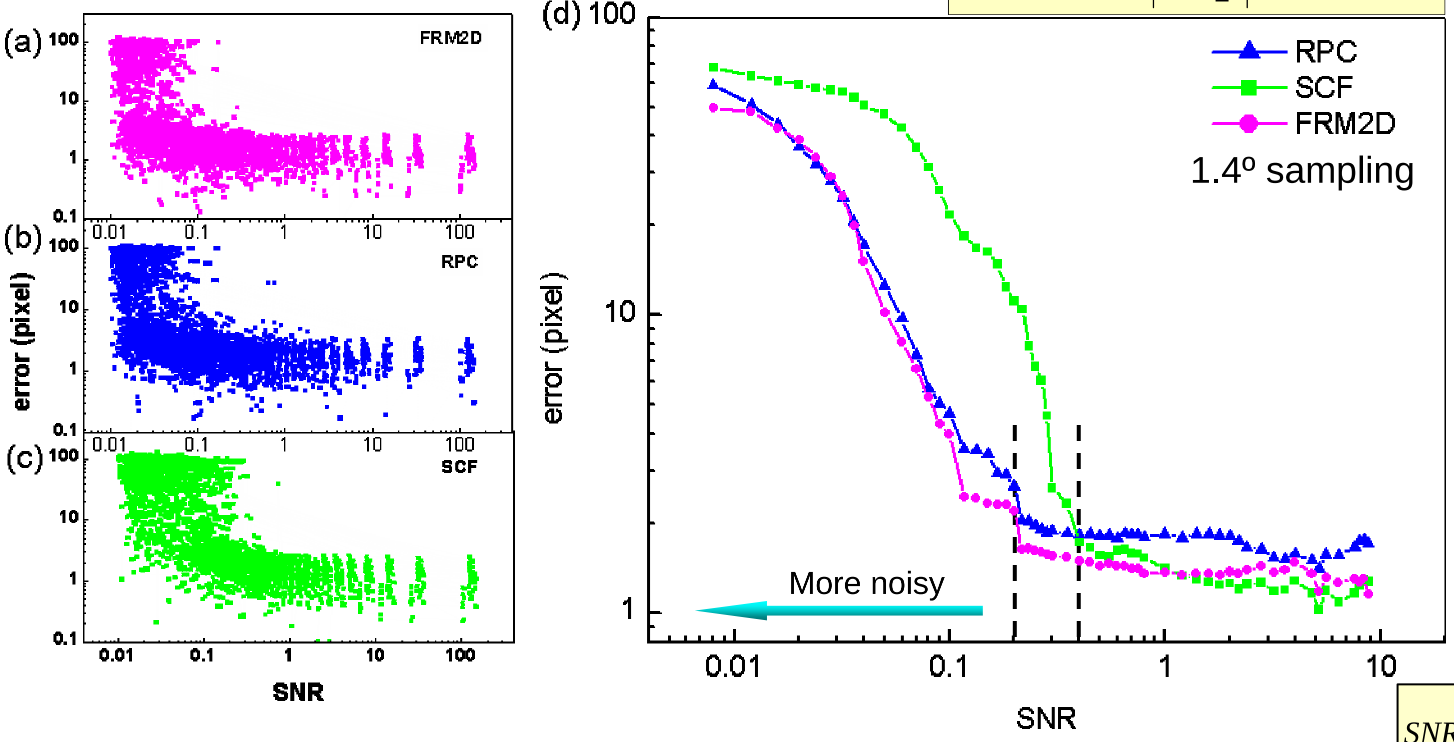
(l) (SNR=0.0373) **(m) (SNR=0.0282)** **(n) (SNR=0.0216)** **(o) (SNR=0.0173)** **(p) (SNR=0.0138)**

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

Total 60,000 test particle images. SNR (0.008~153.12).

Class averaging accuracy comparison

$$\text{Pixel Error} = d\left|\sin\frac{\Delta\phi}{2}\right| + \sqrt{\Delta x^2 + \Delta y^2}$$



$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

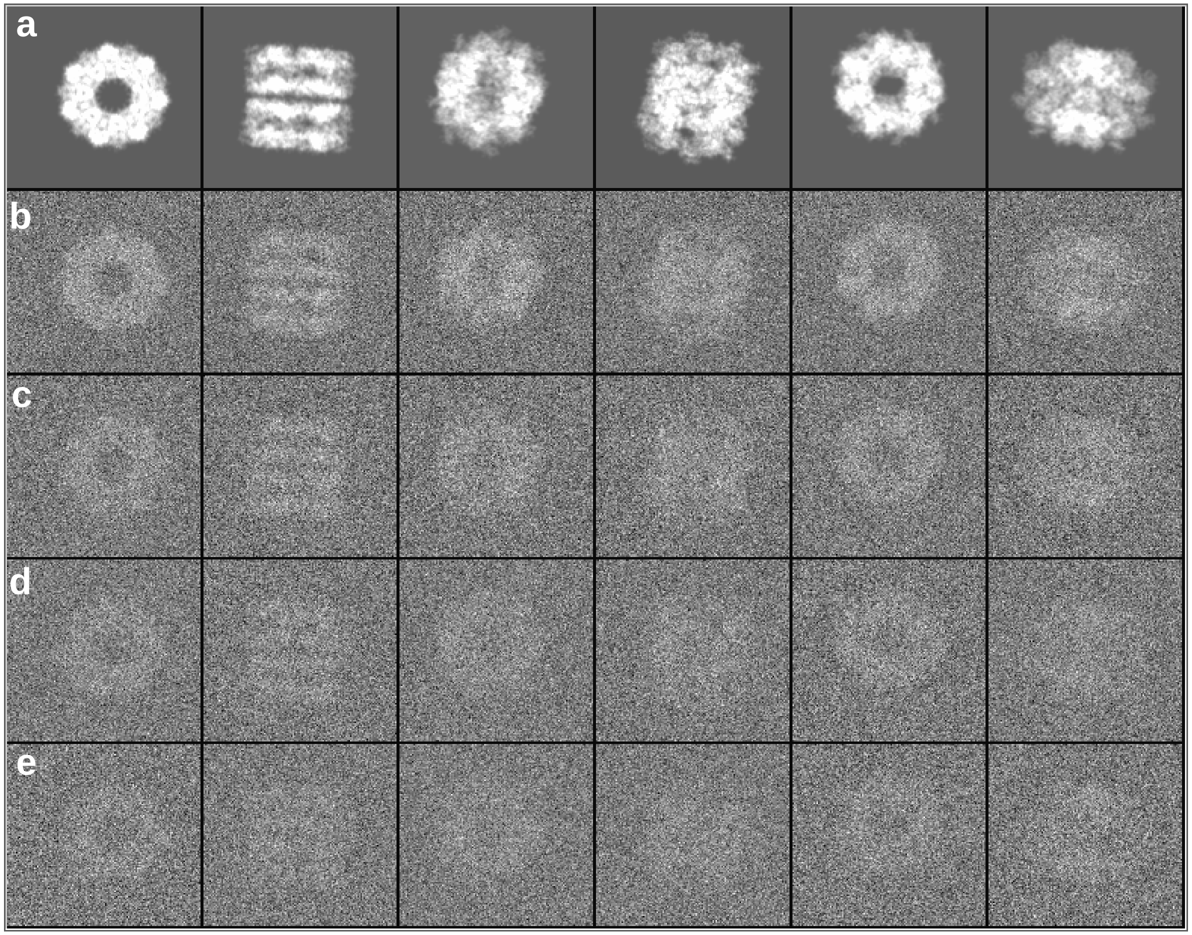
(a-c) Scatter plot of pixel error as a function of SNR for FRM2D, RPC and SCF

(d) Average pixel error as a function of SNR for the three methods.

- The accuracy of FRM2D and RPC is comparable especially in low SNR region**.**
- SCF exhibits more intrinsic sensitivity to noise than FRM2D and RPC.

Cong, Y., Kovacs, J. A., Wriggers, W., 2003. J. Struct. Biol. 144, 51-60

Classification accuracy test images (GroEL)



SNR=0.15

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

SNR=0.06

SNR=0.04

SNR=0.03

93 projections,
930 raw images

- GroEL Gaussian low-pass filtered to ~5Å,
- 93 projections covering all projection directions at an interval of 4º were generated
- random rotation, translation, flip → 930 test images
- 128 × 128 pixels, 1.9 Å/pixel.

Classification accuracy test

93 projections,
930 raw images

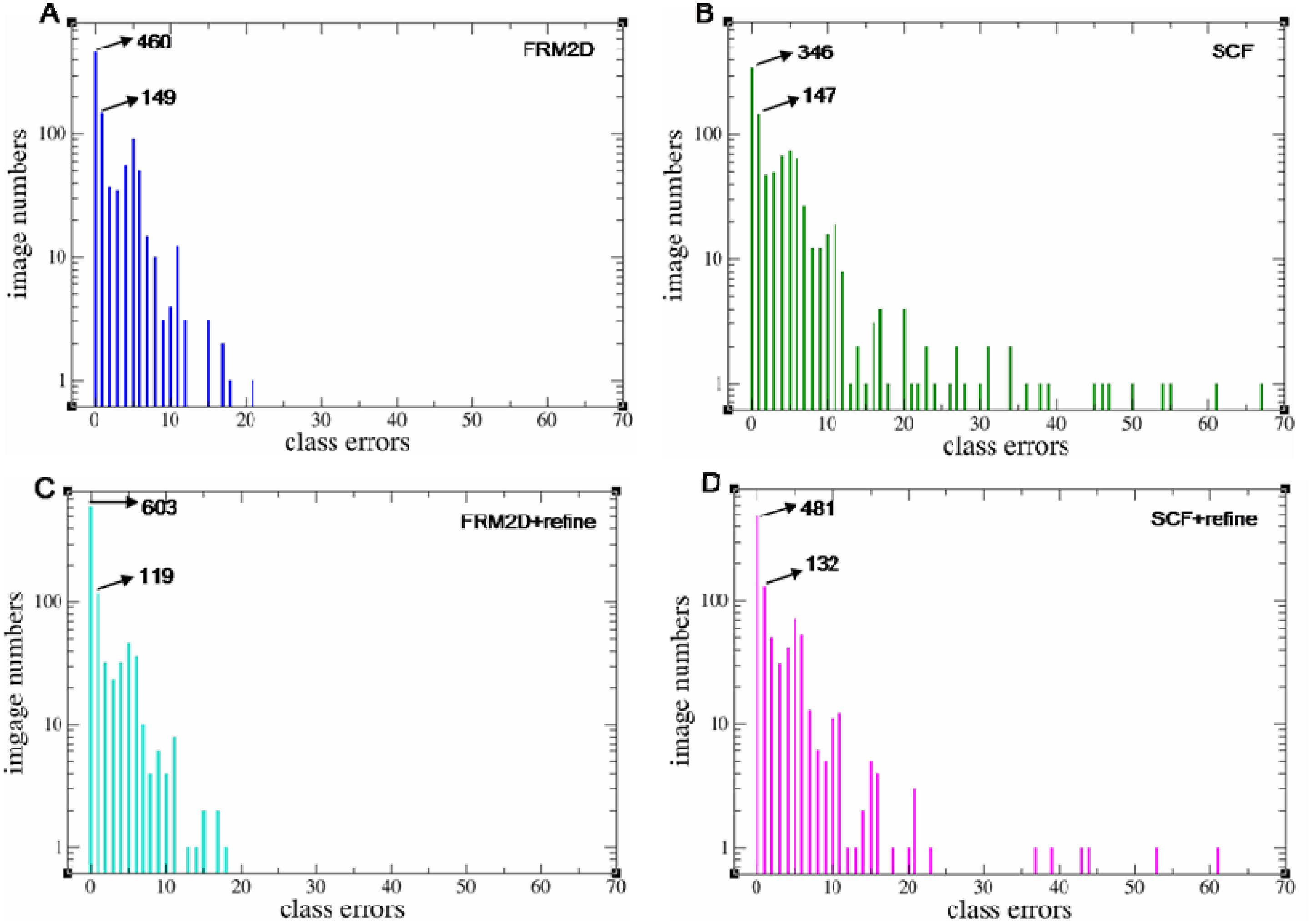
SNR	Correctly classified image numbers			
	0.15	0.06	0.04	0.03
FRM2D	759 (82%)	646 (69%)	536 (58%)	460 (49%)
SCF	777 (84%)	570 (61%)	455 (49%)	346 (37%)
FRM2D+refine	892 (96%)	768 (83%)	696 (75%)	603 (65%)
SCF+refine	872 (94%)	742 (80%)	626 (67%)	481 (52%)

- FRM2D improves over SCF especially in cases of low SNR data (12%)
- Adding a sub-pixel refinement step is beneficial to both methods

Cong, Jiang, Birmanns, Zhou, Chiu, Wriggers, J. Struct. Biol, 152 (2005) 104

Classification error distribution

SNR=0.03

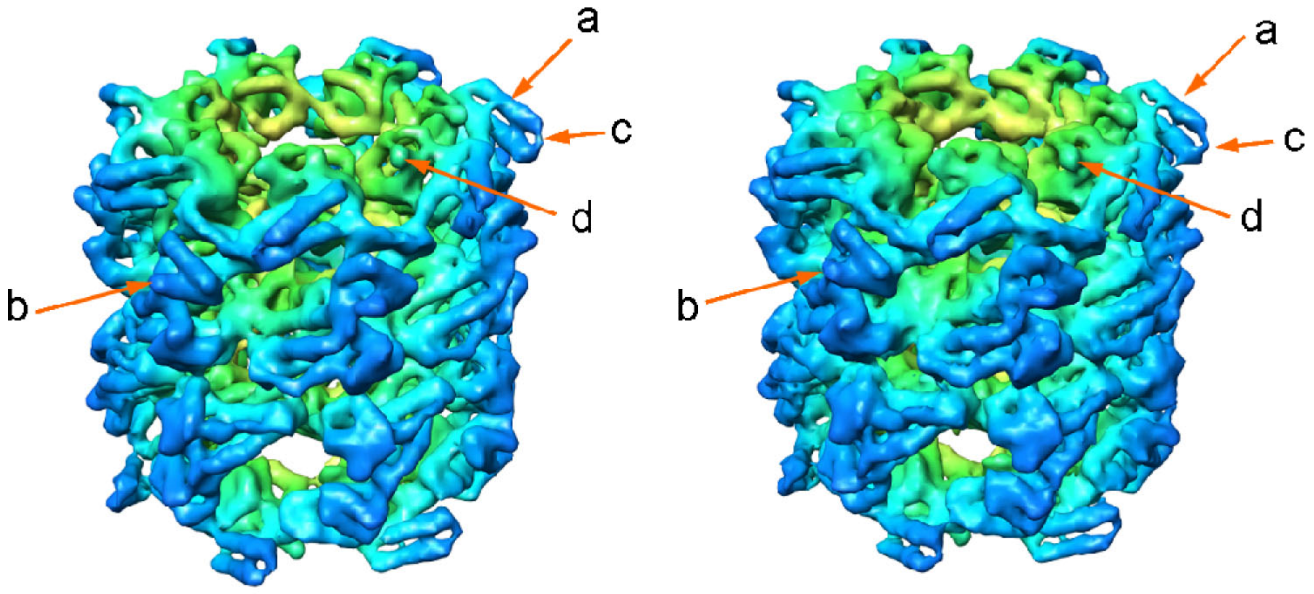


FRM can considerably improve the classification accuracy, makes better use of available raw images

Application of FRM2D as an alignment kernel in cryo-EM 3D reconstruction

GroEL 3D reconstruction test

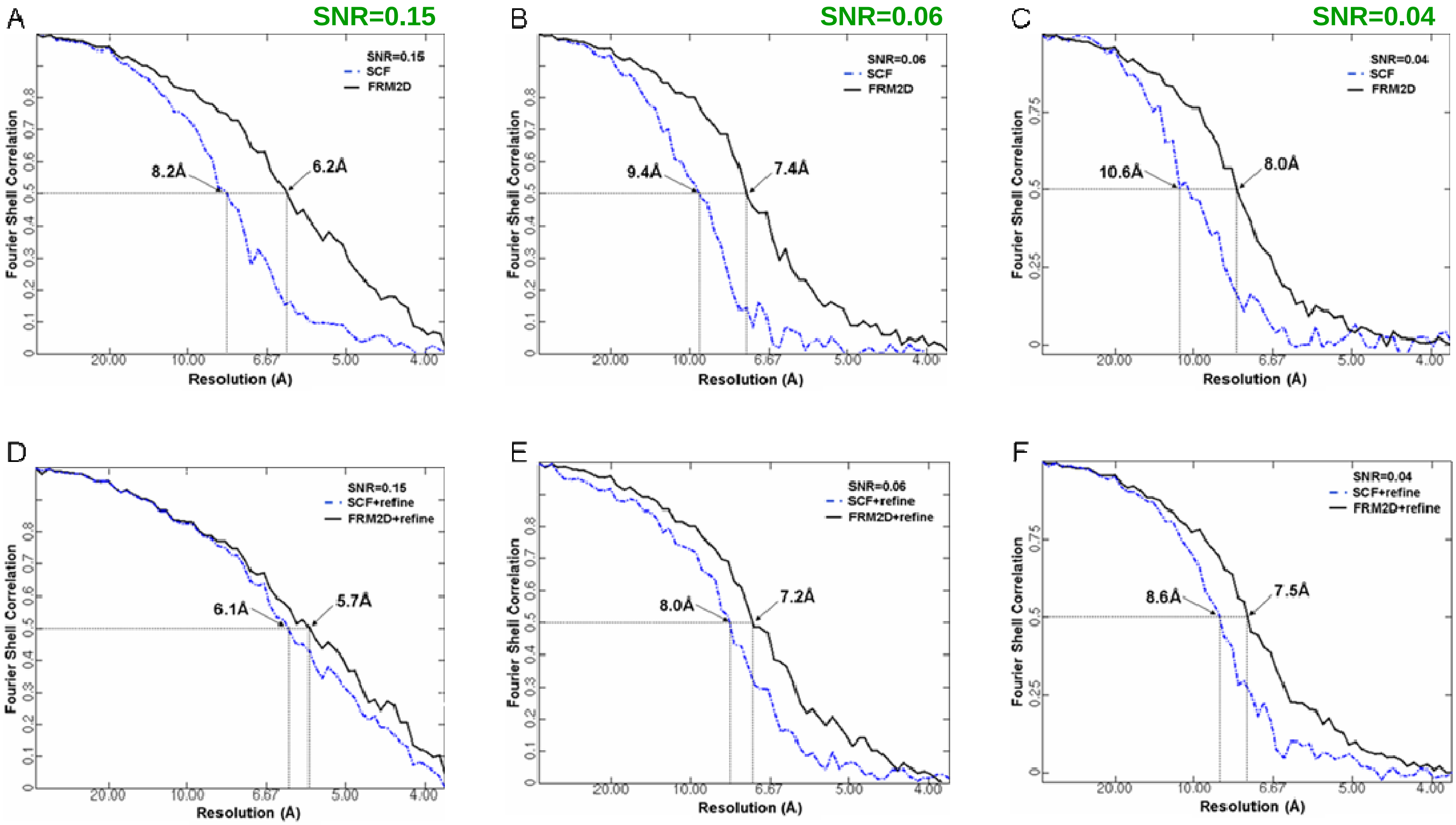
- 4443 raw particles, 3 date sets at different SNR level
- initial model: X-ray structure blurred to ~30 Å
- 1.4º sampling, 5 iterations
- 0.5 FSC between the 3D reconstruction and the 5A blurred X-ray structure



X-ray
low-pass filtered to ~6.3 Å

FRM2D SNR=0.15
~6.3 Å

GroEL 3D reconstructions at different noise levels



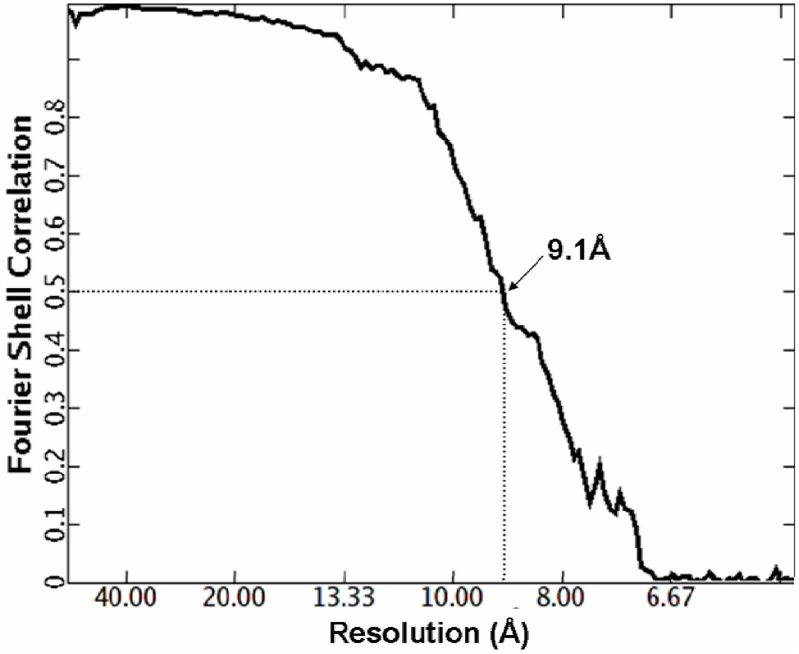
Rice Dwarf Virus (RDV) 3D reconstructions

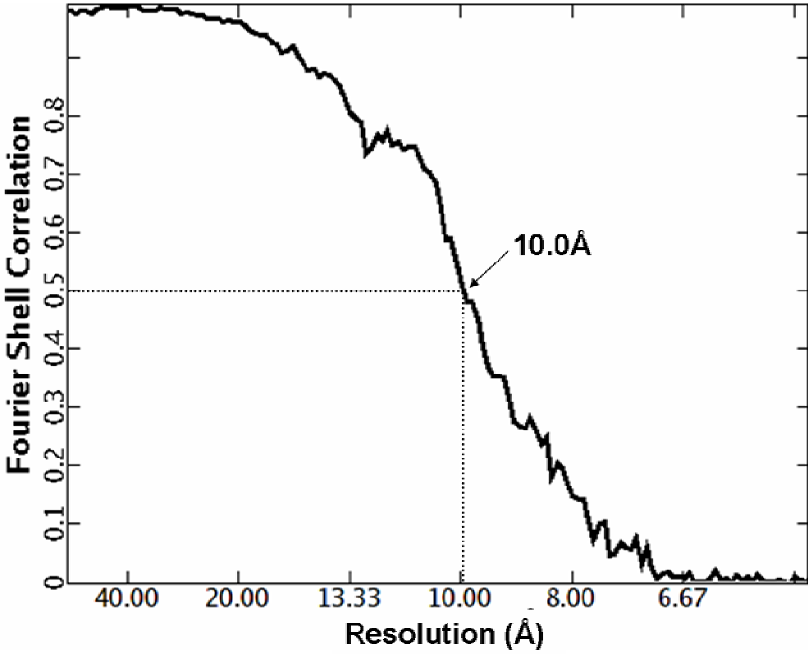
- RDV: a major pathogen of the rice plants in Southeast Asia
- Cryo-EM reconstruction: 6.8 Å with 3261 unique particles (Zhou el al. 2001)
- X-ray structure: 3.5 Å resolution (Nakagawa et al., 2003)

- RDV raw images from close-to-focus micrographs
- Collected in a JEOL 4000 electron cryomicroscopy
- Image size is 300×300 pixels with 2.8 Å /pixel
- a data set containing 3500 close-to-focus particles used in the test
- other than the 2D alignment function, all the other functions and conditions are the same
- 1.4º sampling

RDV 3D reconstruction resolution evaluations

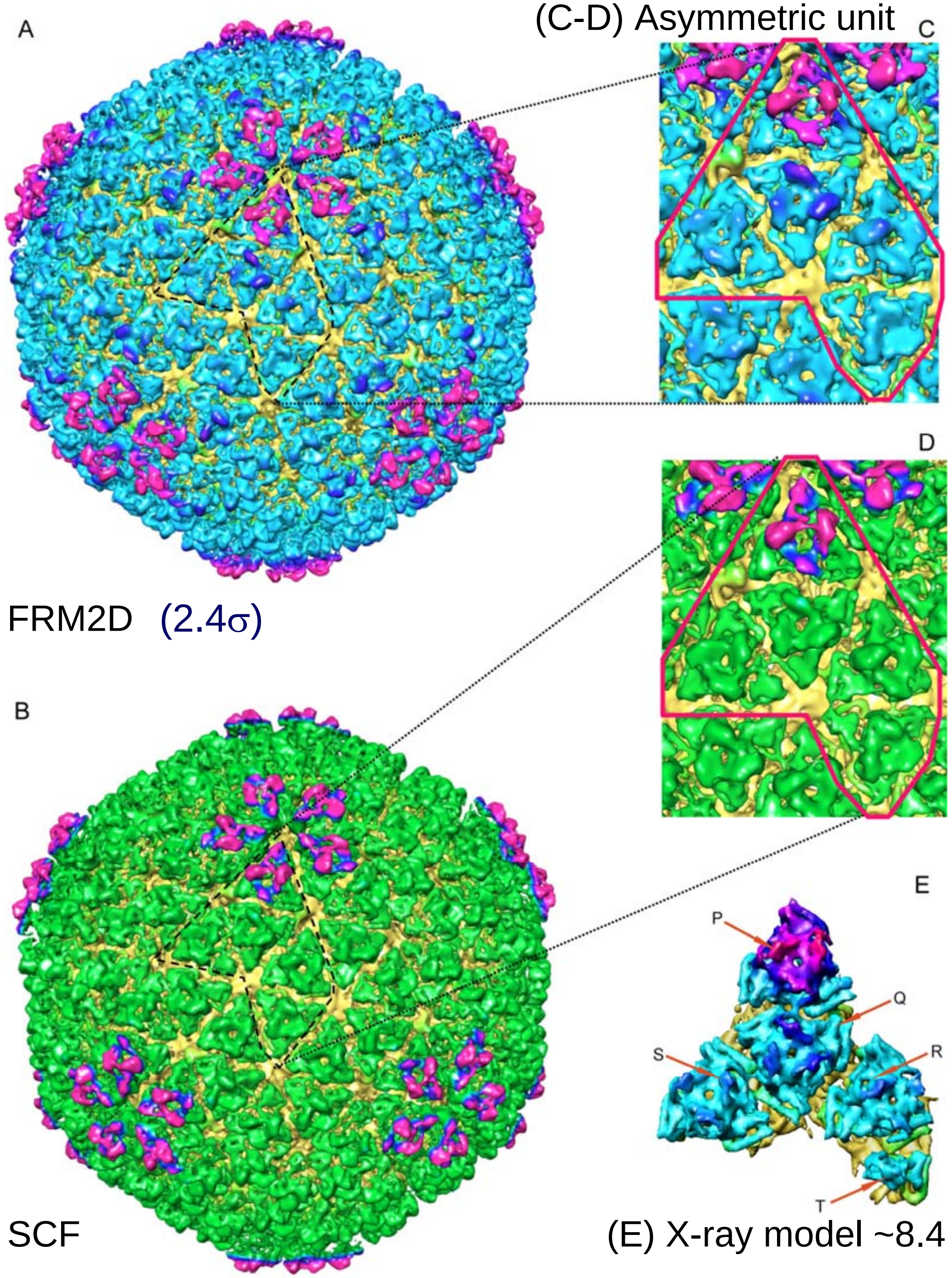
1.4º sampling


FRM2D+refine vs. 6.8Å map


SCF+refine vs. 6.8Å map

Cong, Jiang, Birmanns, Zhou, Chiu, Wriggers, J. Struct. Biol, 152 (2005) 104

RDV 3D reconstruction validation

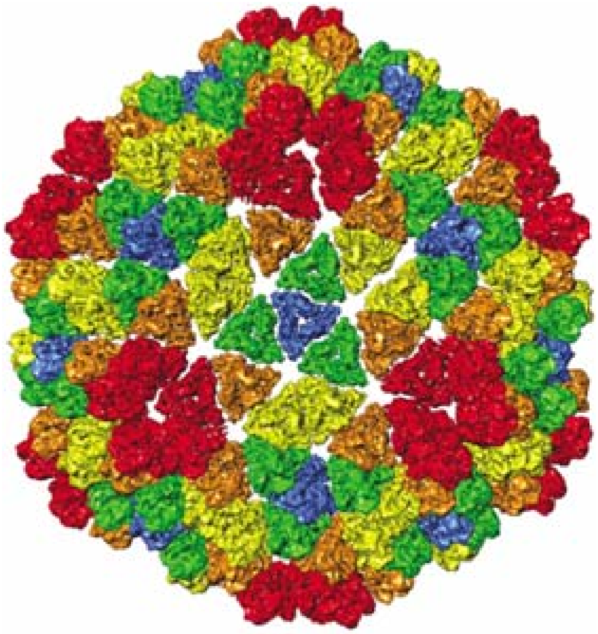


(C-D) Asymmetric unit

(A) FRM2D (2.4σ)

(B) SCF

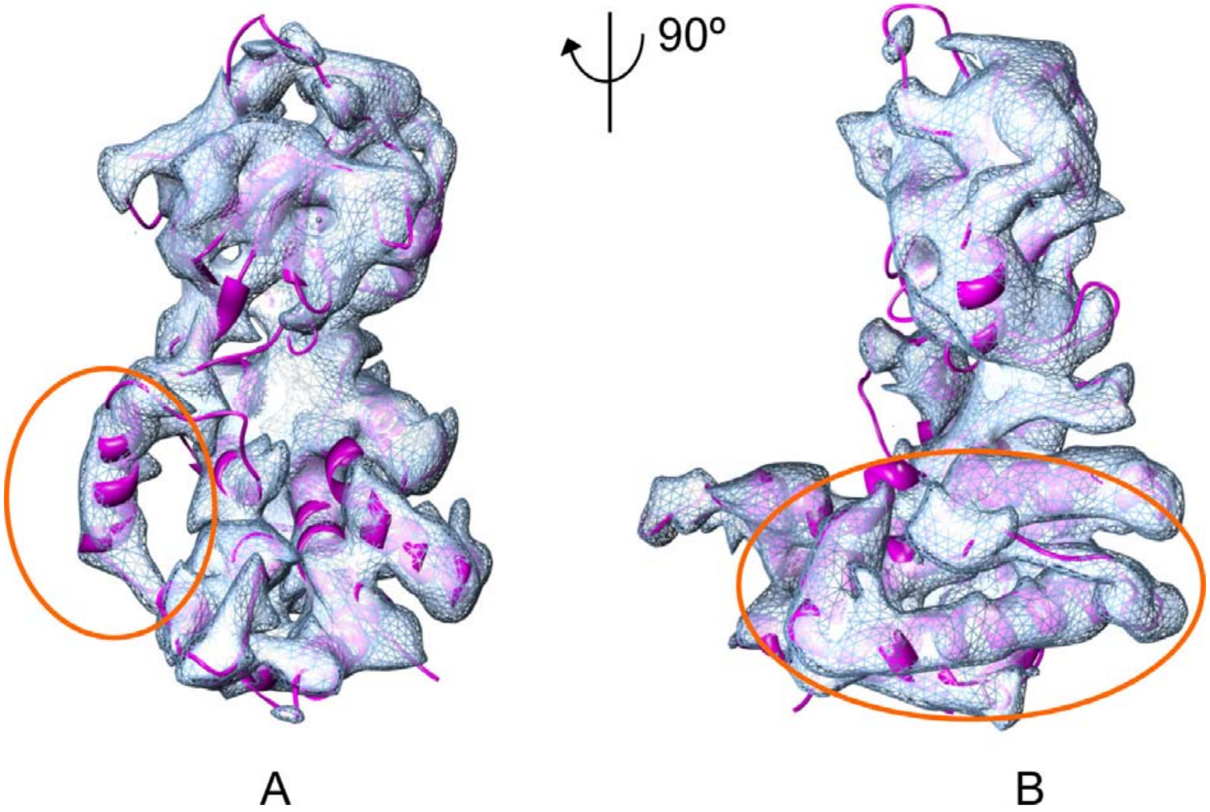
(E) X-ray model ~8.4 Å



(F) 6.8 Å model (Zhou)

T=13/ icosahedral lattice

RDV 3D reconstruction validation



- The outlines of helical regions can still be recognized down to 1/7 to 1/10 Å^{-1}
- Effective resolution of FRM2D reconstruction is below 10 Å
- Utilizing FRM2D alignment kernel can accomplish reliable 3D reconstructions directly from extremely noisy experimental data sets

Summarization

- Advantage: using 2D FFT to accelerate the two rotational parameters search, avoiding the costly zero padding
- Accuracy: FRM2D accuracy comparable to RPC, and outperforms SCF especially in low SNR cases.
- Efficiency: FRM2D has the potential to outperform the traditional 2D alignment methods depending on the desired fineness of angular sampling and exhaustive search range.
- 3D reconstruction: FRM2D is a reliable and robust alignment kernel with observable resolution improvement than the classical method especially in low SNR cases.

Future directions

- Better way to determine the COM to reduce the x-axis direction scan range.
- Better way of sampling
- Release soon with Situs and EMAN

Acknowledgements

People

Willy Wriggers
Julio Kovacs
Stefan Birmanns

Wah Chiu
Wen Jiang
Steve Ludtke
Grant Tang

Z. Hong Zhou

Pawel Penczek

Fundings

W.M. Keck Center
NIH
Human Frontier Foundation
Alfred P. Sloan Foundation