instruct
Integrating
Biology

I2PC

Instruct Image Processing Centre

JM CARAZO

Image Processing in SPA: Principles and Workflows

José María Carazo (carazo@cnb.csic.es)
Spanish National Center for Biotechnology
Instruct Image Processing Center

Cryo-EM Course Lectures Outline

Monday	
13:00-13:45	JO: General introduction to three-dimensional cryo microscopy
13:45-14:30	SdC: Sample preparation and vitrification
15:00-16:30	MW: Modelling/Fitting
Tuesday	
09:00-9:45	JMC: Single Particle Analysis (SPA): The basics
09:45-10:30	JMC: Image processing workflows for SPA
Thursday	
09:00-09:45	JO: Image processing workflows for Tomography
09:45-10:30	SdC: New EM technologies

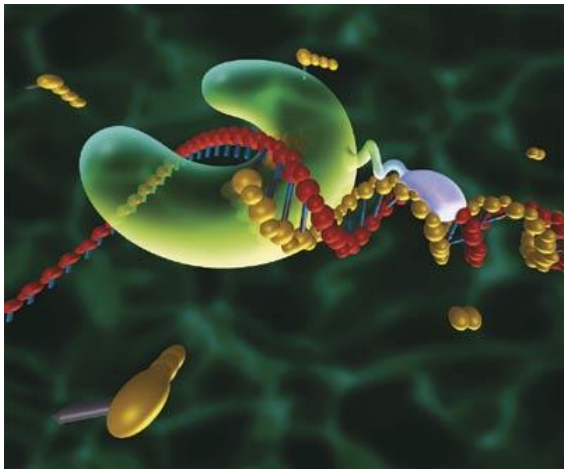
CSIC

instruct Integrating Biology

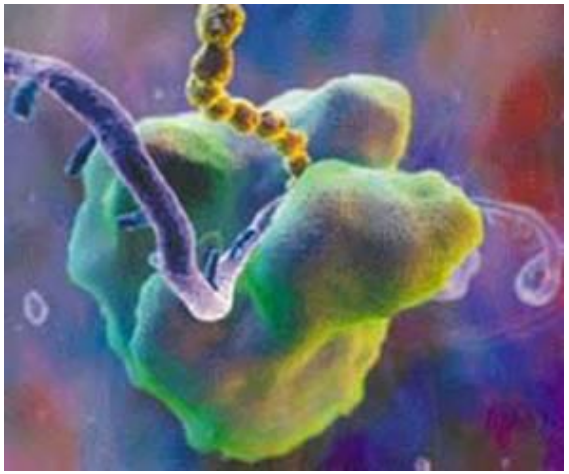
BIO COMPUTING Unit

instruct image Processing Center

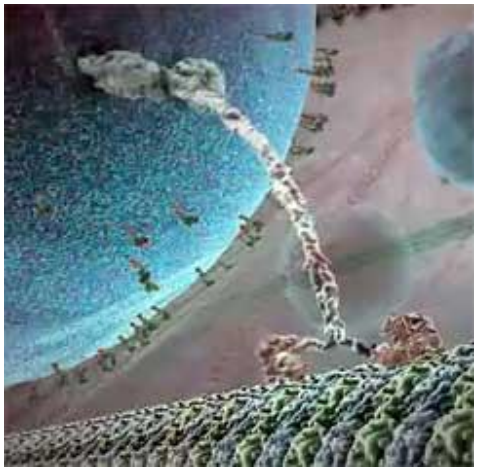
Life is based on macromolecular machines



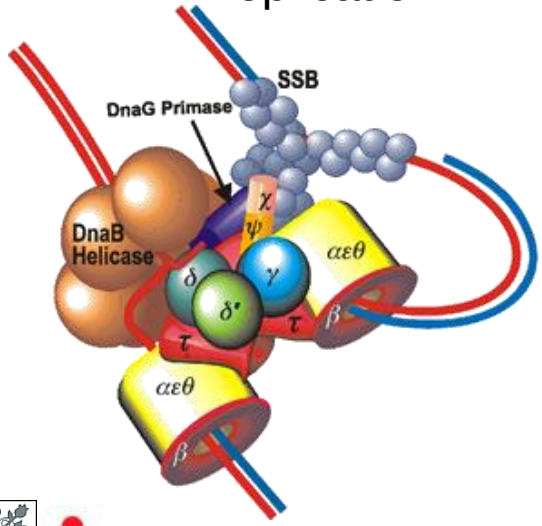
DNA replication

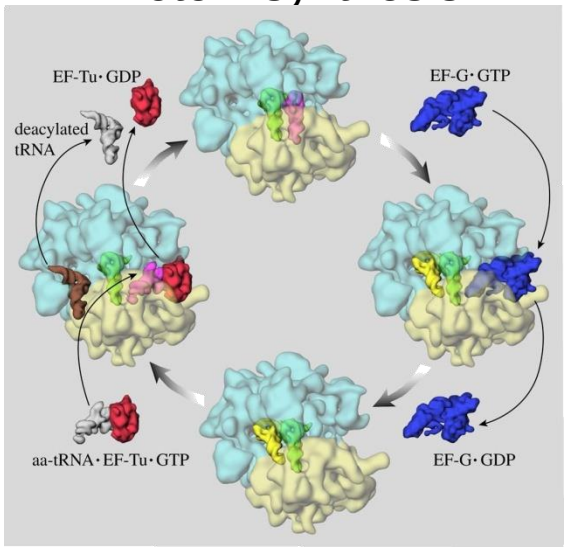


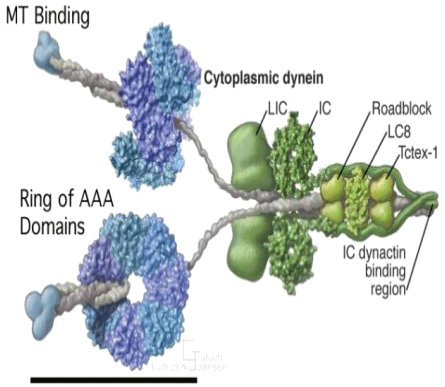
Protein synthesis



Dynein motion







CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

This is our objective!

ARTICLE

doi:10.1038/nature12822

Structure of the TRPV1 ion channel determined by electron cryo-microscopy

Maofu Liao[1]*, Erhu Cao[2]*, David Julius[2] & Yifan Cheng[1]

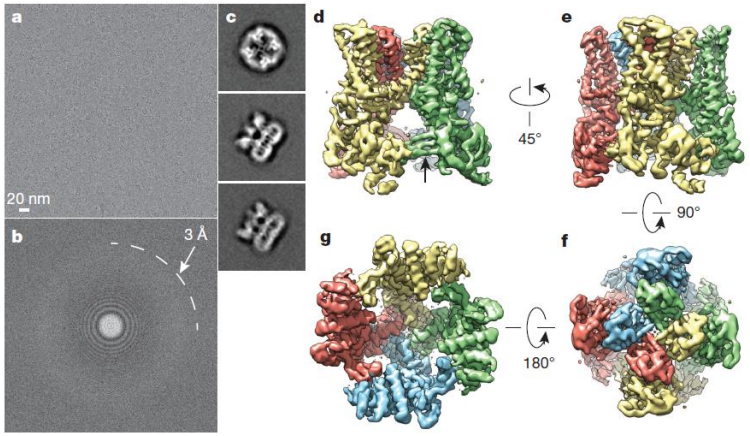


Figure 1 | 3D reconstruction of TRPV1 determined by single-particle cryo-EM. **a**, Representative electron micrograph of TRPV1 protein embedded in a thin layer of vitreous ice recorded at a defocus of 1.7 μm. **b**, Fourier transform of micrograph shown in **a**, with Thon rings extending to nearly 3 Å. **c**, Enlarged views of three representative 2D class averages show fine features of tetrameric channel complex. **d**–**g**, 3D density map of TRPV1 channel filtered to a resolution of 3.4 Å (scaled to atomic structure) with each subunit colour-coded. Four different views of the channel are shown, from side (**d**, **e**), top (**f**) and bottom (**g**). The arrow in panel **d** indicates β-sheet structure in the cytosolic domain of TRPV1.

CSIC

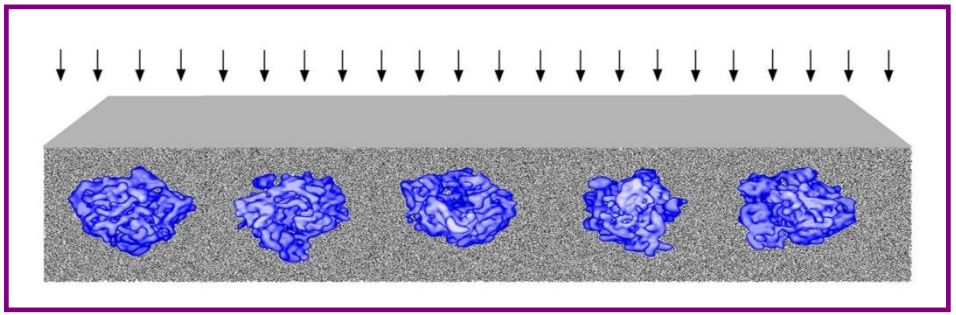
instruct Integrating Biology

BO Computing Unit

instruct image Processing Center

The cryo-EM SPA pledge

- In 3D Electron Microscopy ***individual macromolecules*** are visualized down to ***atomic resolution***.
- Trapped in ice, these molecules are free to expose their internal flexibility/plasticity.



The cryo-EM SPA pledge

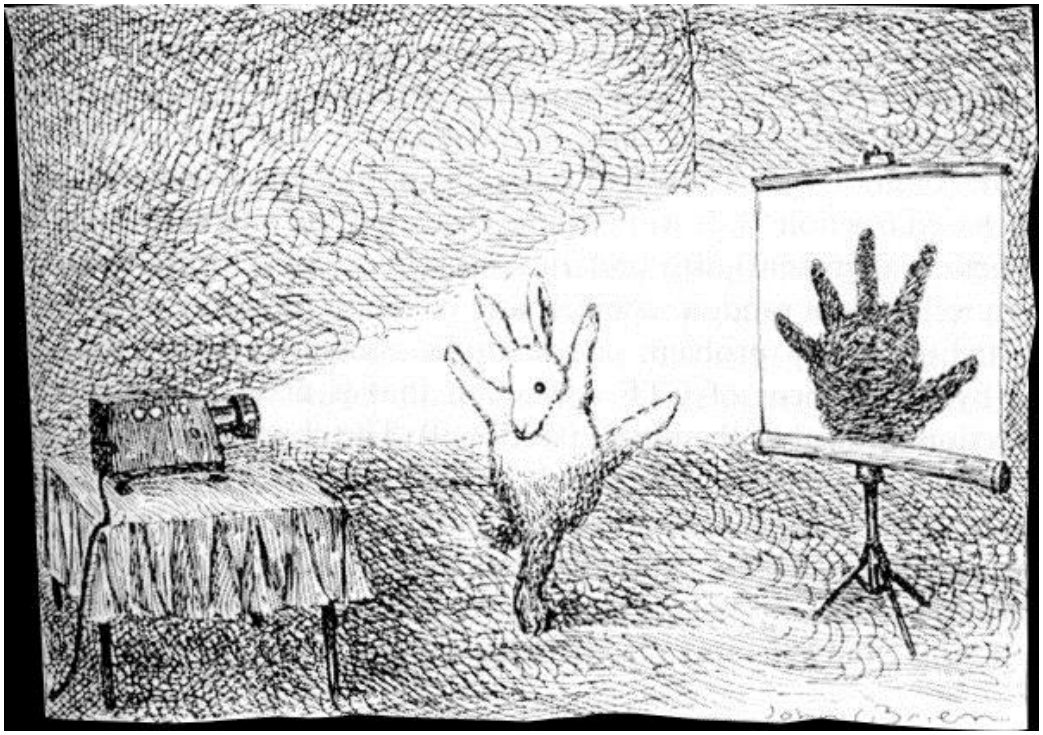
- In 3D Electron Microscopy ***individual macromolecules*** are visualized down to ***atomic resolution***
  - ***Attention to every detail of the image formation process***
  - ***Very precise image processing***
- Trapped in ice, these molecules are free to expose their internal flexibility/plasticity
  - ***Need to classify individual images in 3D***

The cryo-EM SPA pledge

Attention to every detail!

- ***Characterization of each image***
- ***A posteriori characterization of the Projection Geometry***
- ***3D reconstruction process***
- ***3D classification***
- ***Validation***

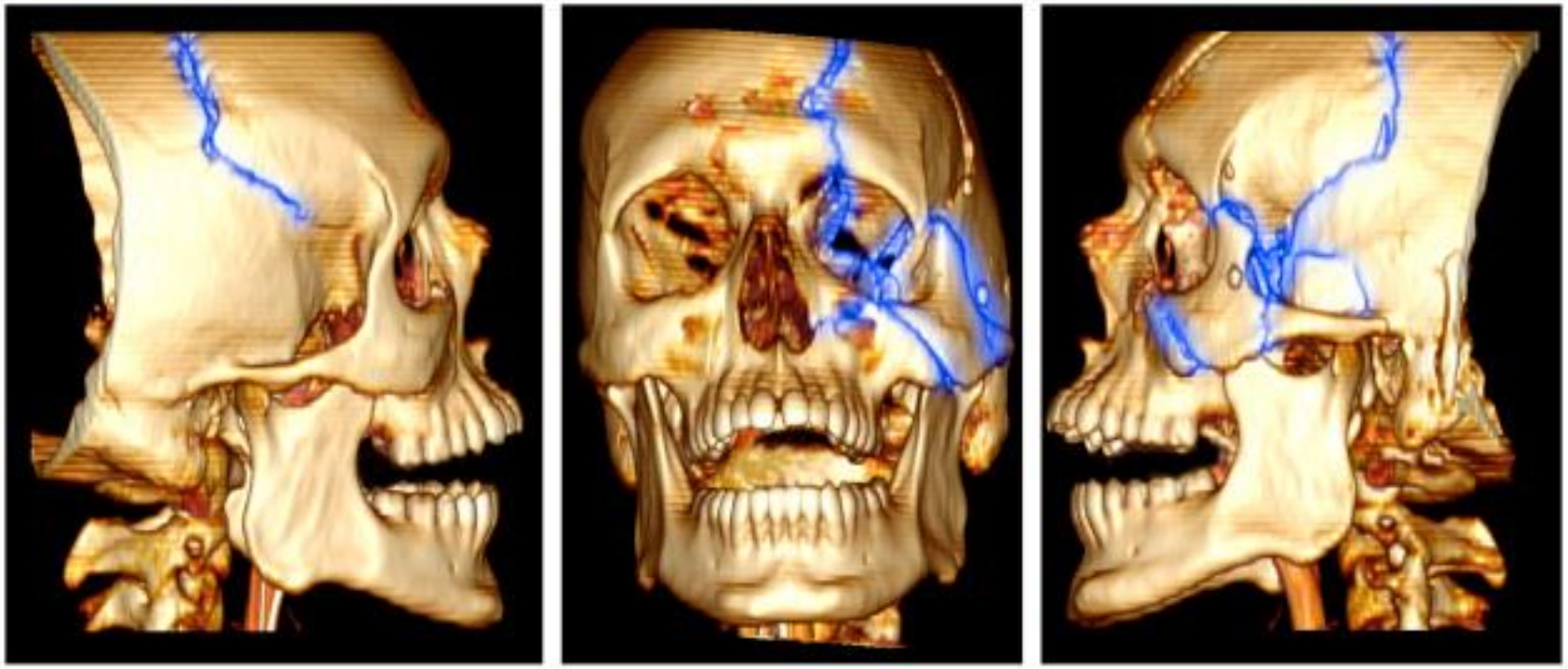
As we were saying yesterday …..



2D projections : lack of information → Limits the comprehension of complex objects

The value of a «radiography»

Compared to full 3D CT



CSIC

instruct
Integrating
Biology

instruct
image
Processing
Center

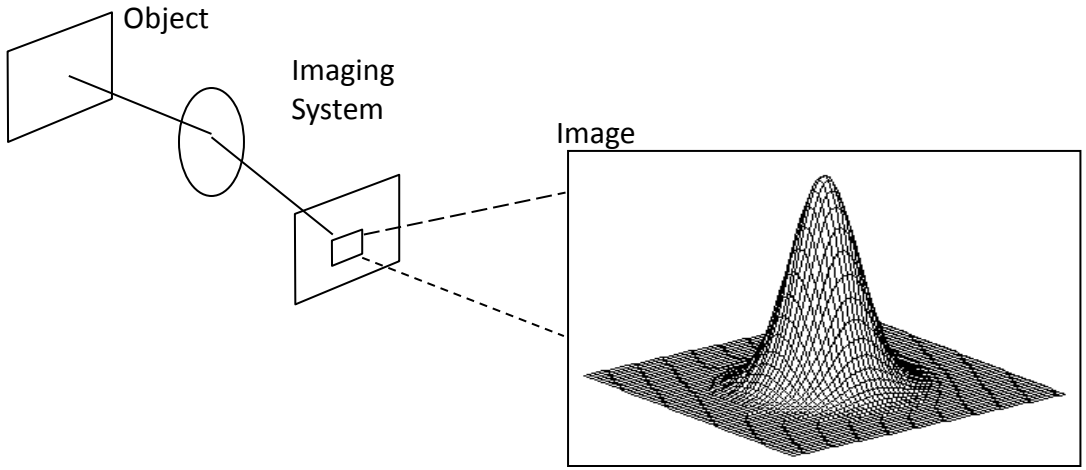
Compared to full 3D CT

CSIC

instruct
Integrating
Biology

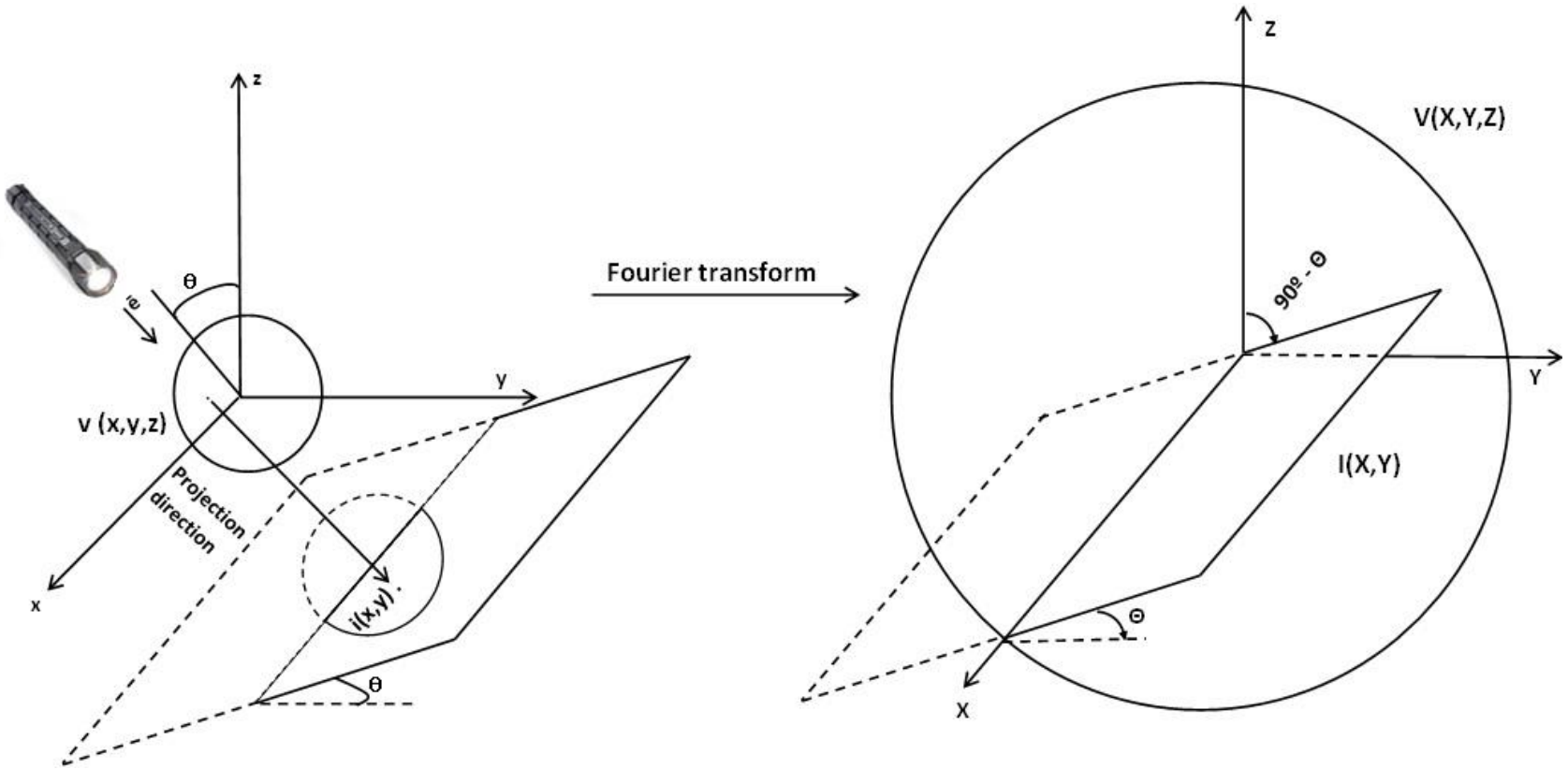
instruct
image
Processing
Center

Realistic Image Formation Model: CTF characterization

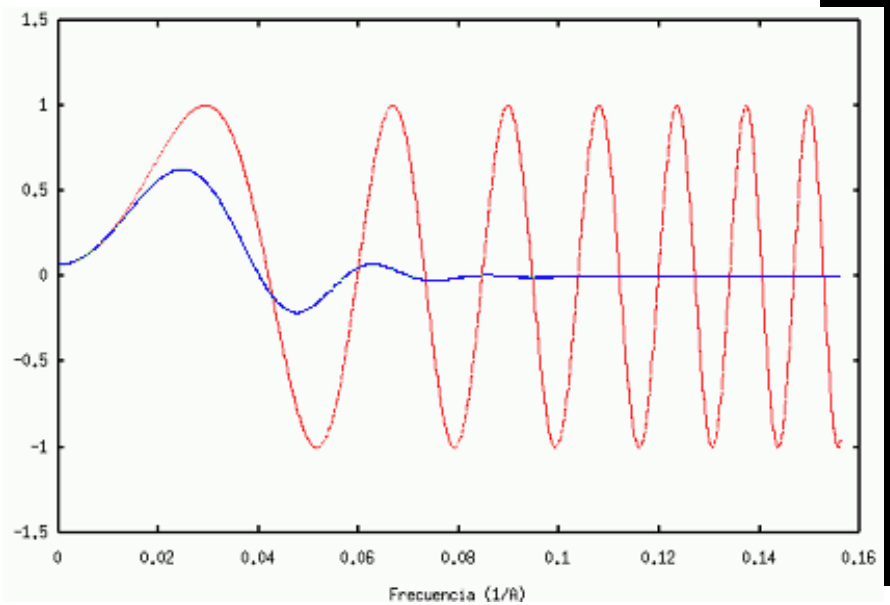


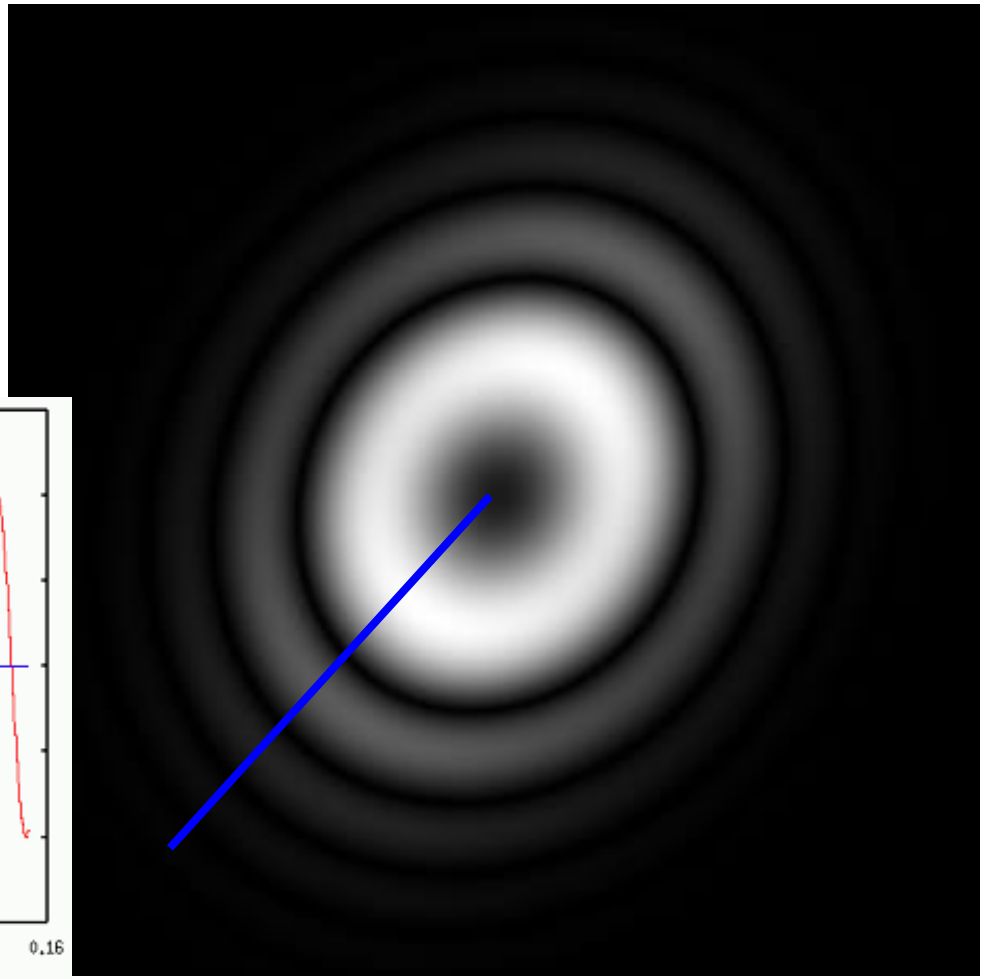
PSF:
Point Spread Function

Real space and Fourier Space



CTF Profile in Fourier Space





CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

CTF Challenge

CTF Challenge: Result Summary

Roberto Marabini[(1)], Bridget Carragher[(2)], Shaoxia Chen[(9)], James Chen[(13)], Anchi Cheng[(2)], Kenneth H. Downing[(4)], Joachim Frank[(5)], Robert A. Grassucci[(5)], Bernard Heymann[(12)], Wen Jiang[(6)], Slavica Jonic[(10)], Hstau Y. Liao[(5)], Steven J. Ludtke[(3)], Shail Patwari[(11)], Angela L. Piotrowski[(11)], Adrian Quintana[(7)], Carlos O.S. Sorzano[(7)], Henning Stahlberg[(8)], Javier Vargas[(7)], Neil R. Voss[(11)], Wah Chiu[(3)], Jose M. Carazo[(7)]

August 4, 2014

[(1)]Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain.

[(2)] The National Resource for Automated Molecular Microscopy, The Scripps Research Institute, La Jolla, CA 92037, USA

[(3)] Baylor College of Medicine, Houston, Texas 77030, USA

CTF Challenge

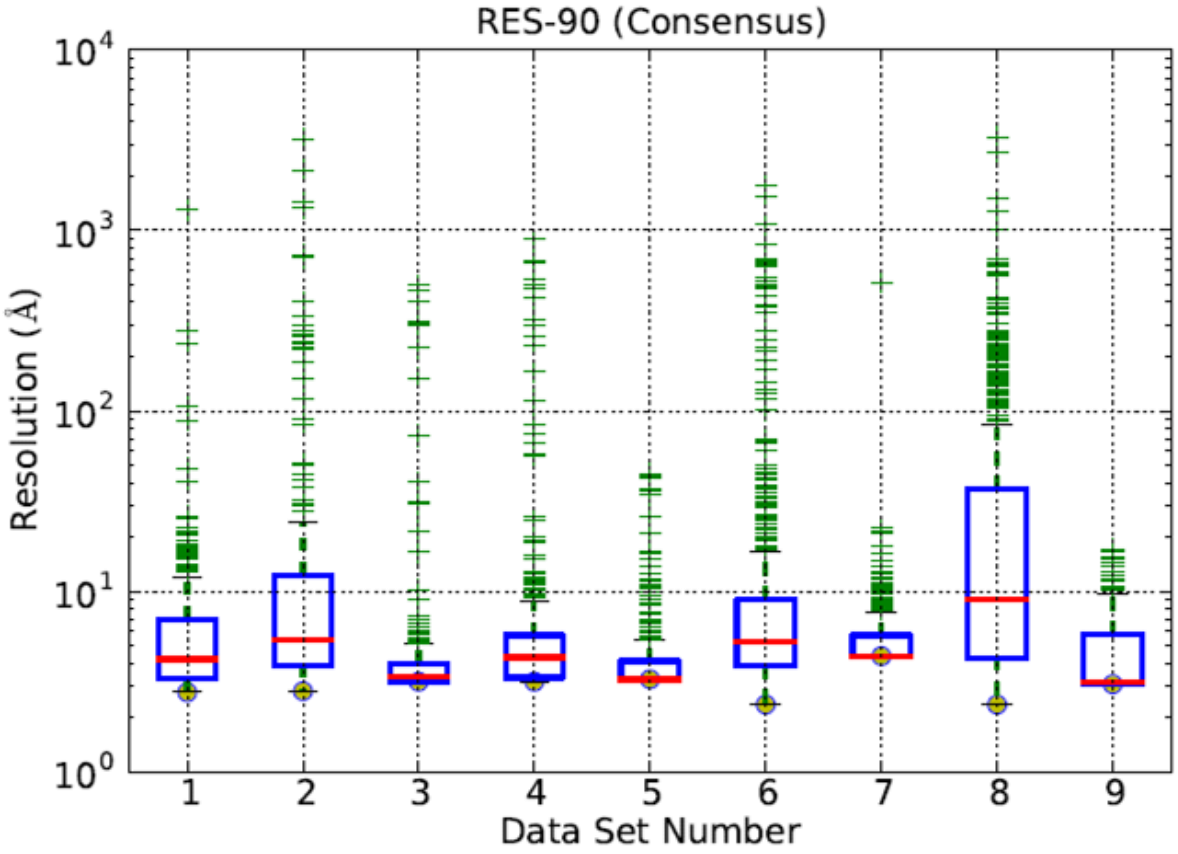


Figure 9: RES-90 analysis. X-axis ticks refer to data sets, *Y*-axis represents an estimation of the resolution limit imposed by the accuracy in the CTF determination. Yellow circles show the Nyquist frequency for each data set.

CSIC

instruct Integrating Biology

BO Computing Unit

instruct image Processing Center

How precise should we be?

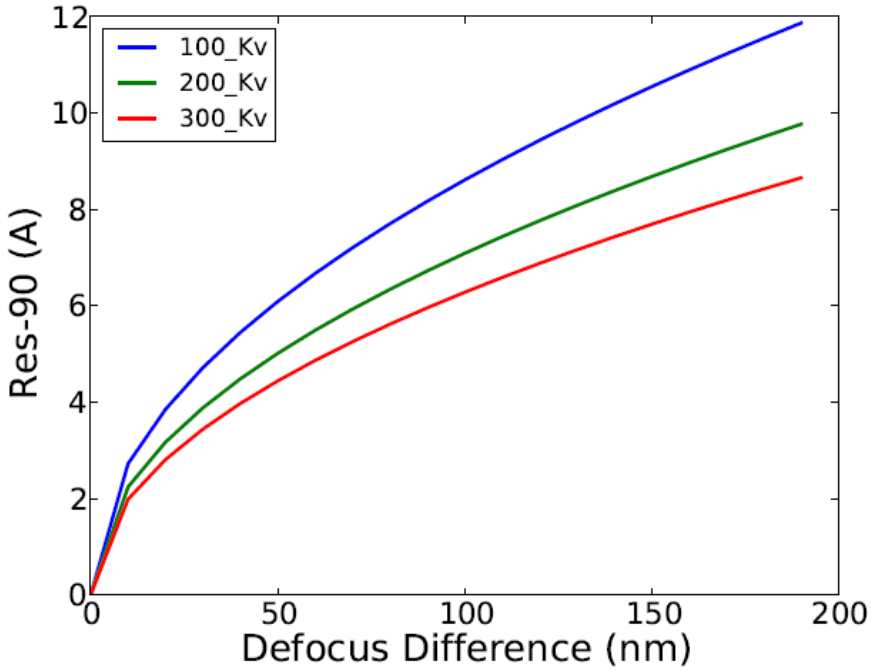


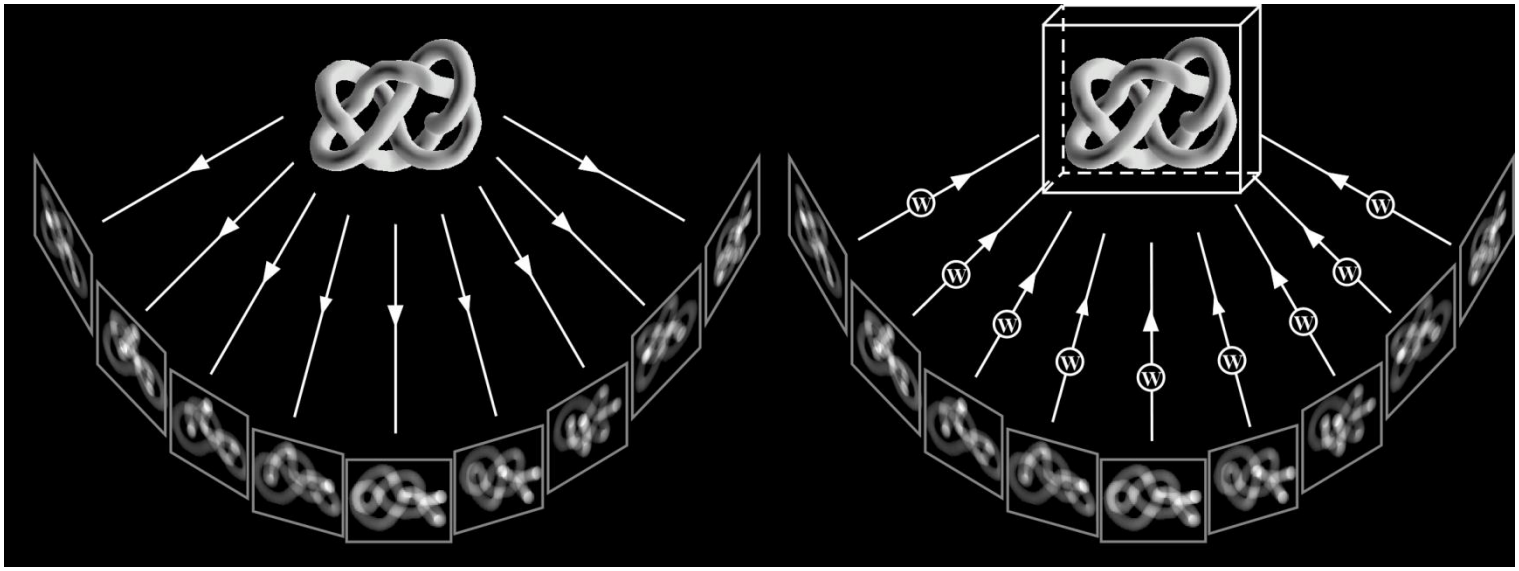
Figure 8: Resolution at which the wave aberration shift introduced by a given defocus error is 90°. Note that this magnitude depends only on the defocus error and not on the actual amount of defocus. Additionally, note that the plot would be the same if instead of considering two non-astigmatic CTF estimations we would consider the defocus difference between the two astigmatic axis, assuming no errors in astigmatic angle estimation.

The cryo-EM SPA pledge

Attention to every detail!

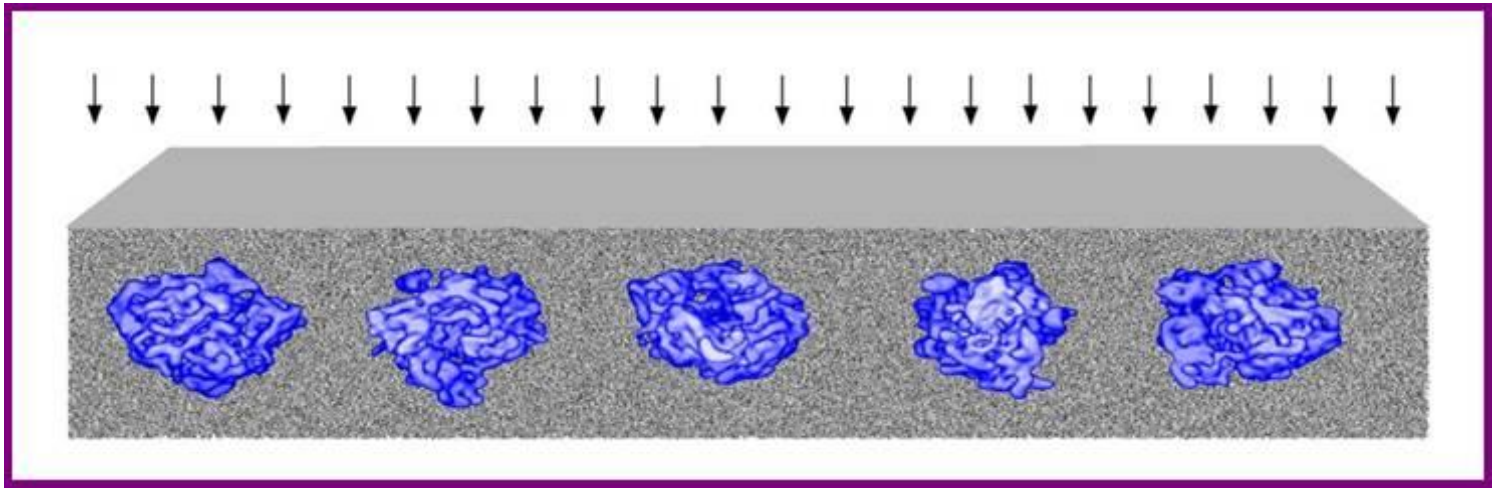
- ***Characterization of each image***
- ***A posteriori characterization of the Projection Geometry***
- ***3D reconstruction process***
- ***3D classification***
- ***Validation***

Tomography Principle



Acquisition of tilted image series → Correction of microscope default (mechanical drift, CTF...) → Reconstruction

That does not apply to SPA



Parameter space

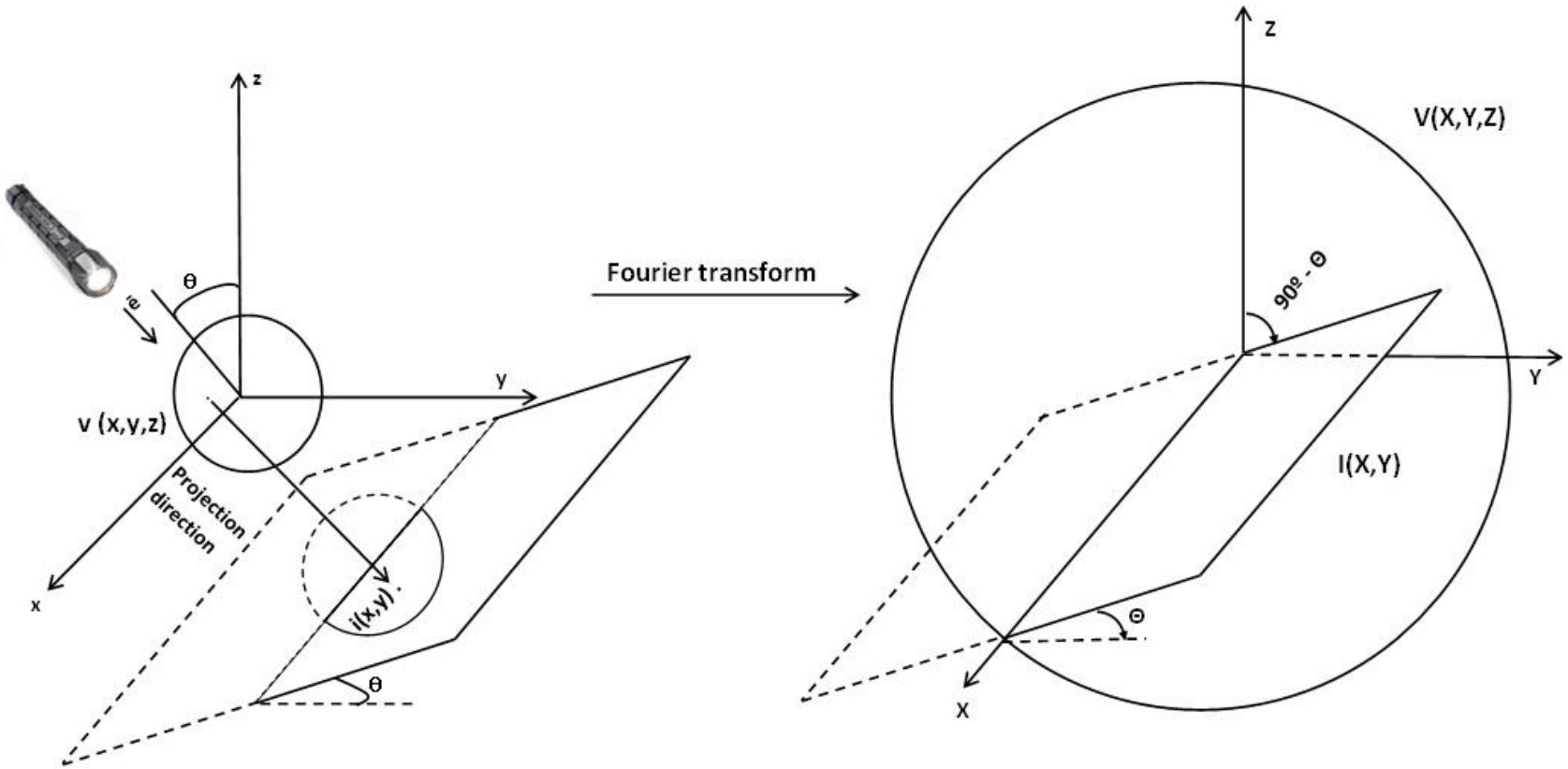
- For each particle we need to determine 3 angles and 2 shifts. FIVE parameters.
- If we have 100.000 particle images.
- We then have a space of 500.000 parameters!

The cryo-EM SPA pledge

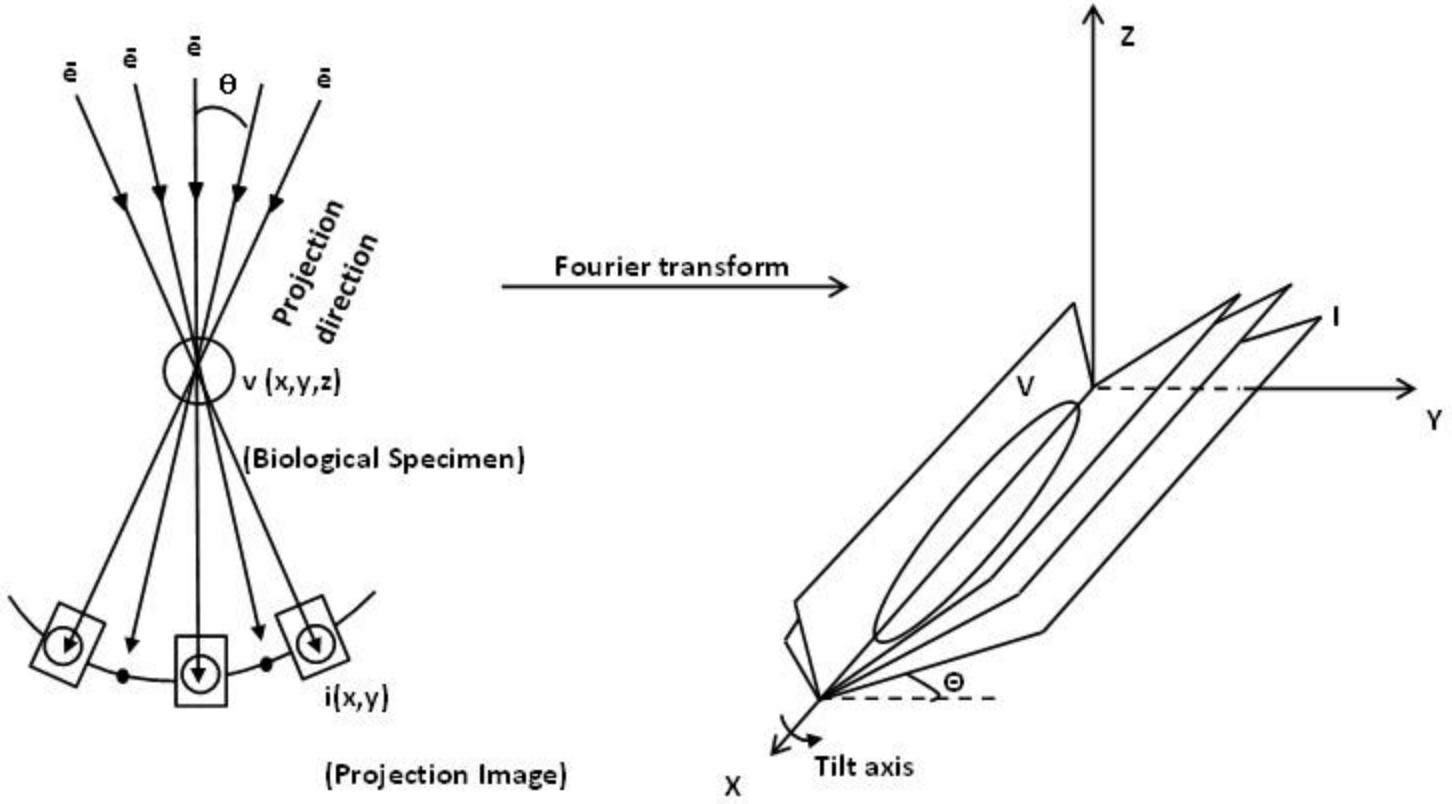
Attention to every detail!

- ***Characterization of each image***
- ***A posteriori characterization of the Projection Geometry***
- ***3D reconstruction process***
- ***3D classification***
- ***Validation***

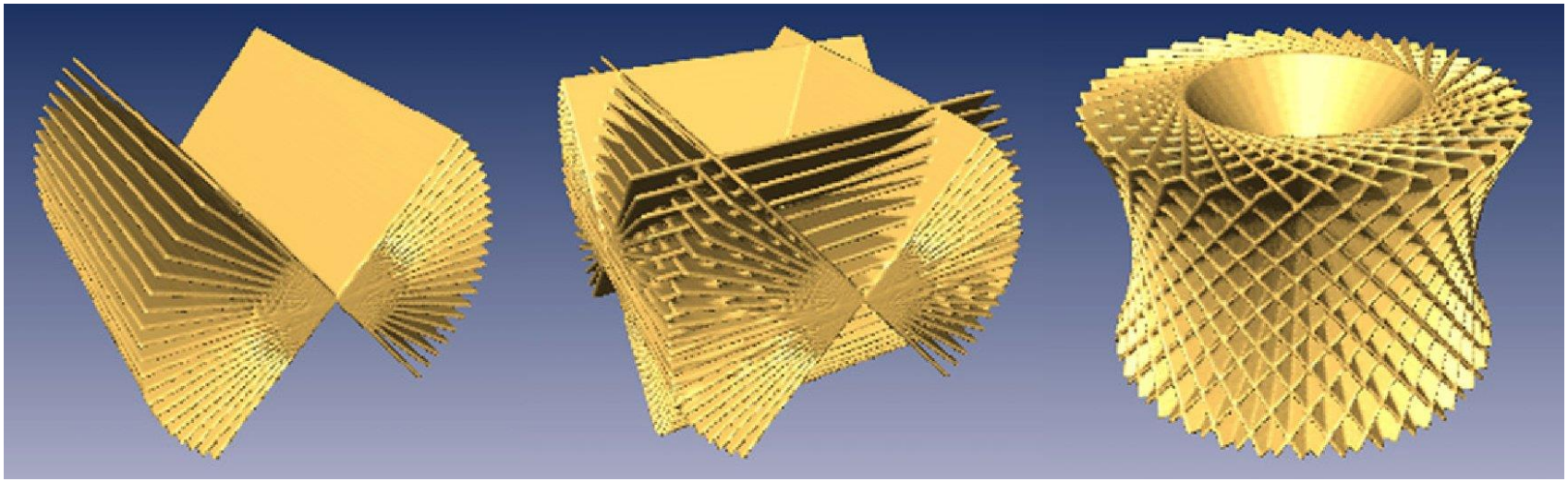
Real space and Fourier Space



Principles of Fourier Reconstruction Method



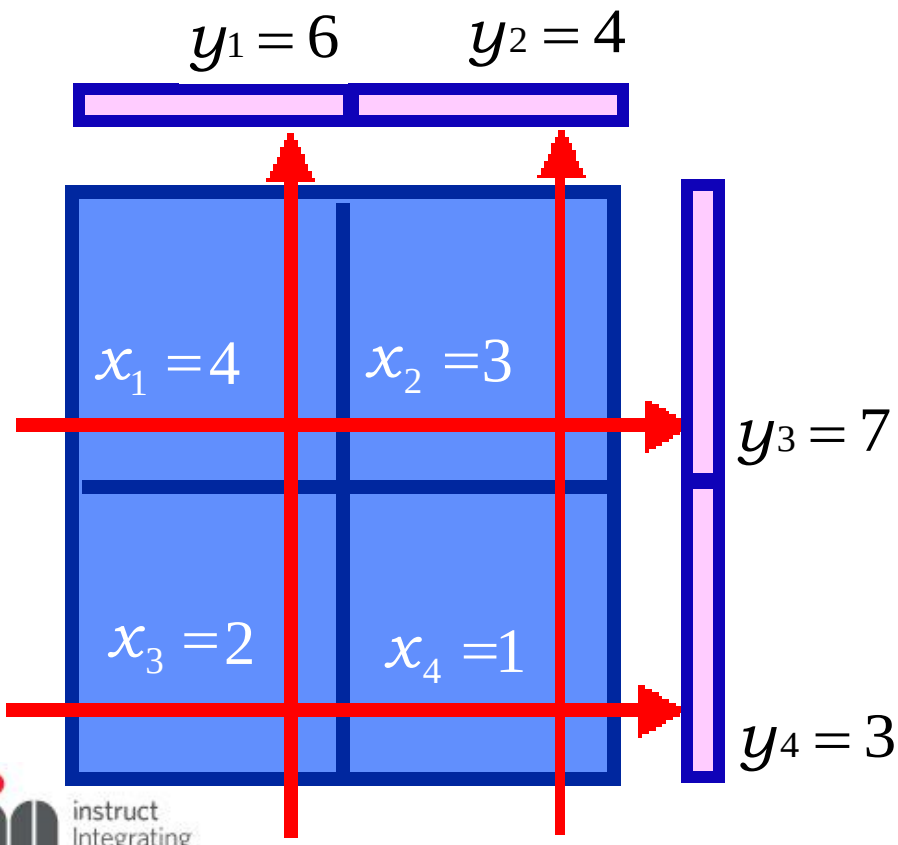
More complex geometries in Fourier space



$$f(\mathbf{r}) \approx \sum_{j=1}^{J} x_j b_j(\mathbf{r})$$

Reconstruction as a linear set of equations

$$y_i \approx \sum_{j=1}^{J} l_{i,j} x_j \qquad l_{i,j} = 1,0$$



$$\begin{cases} x_1 + x_3 = 6 \\ x_2 + x_4 = 4 \\ x_1 + x_2 = 7 \\ x_3 + x_4 = 3 \end{cases}$$

CSIC

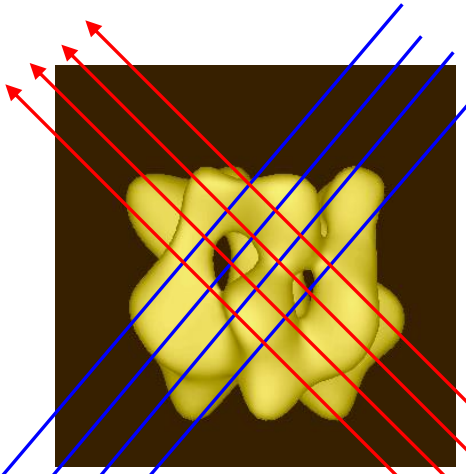
instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

ART, the “basics”

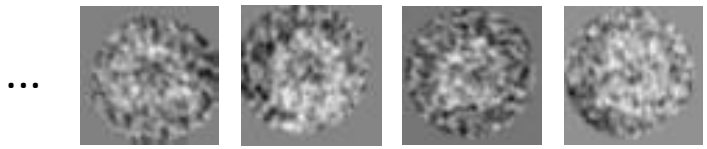
Conceptual schema



Project

Reproject

Iterate (n times)



−

x

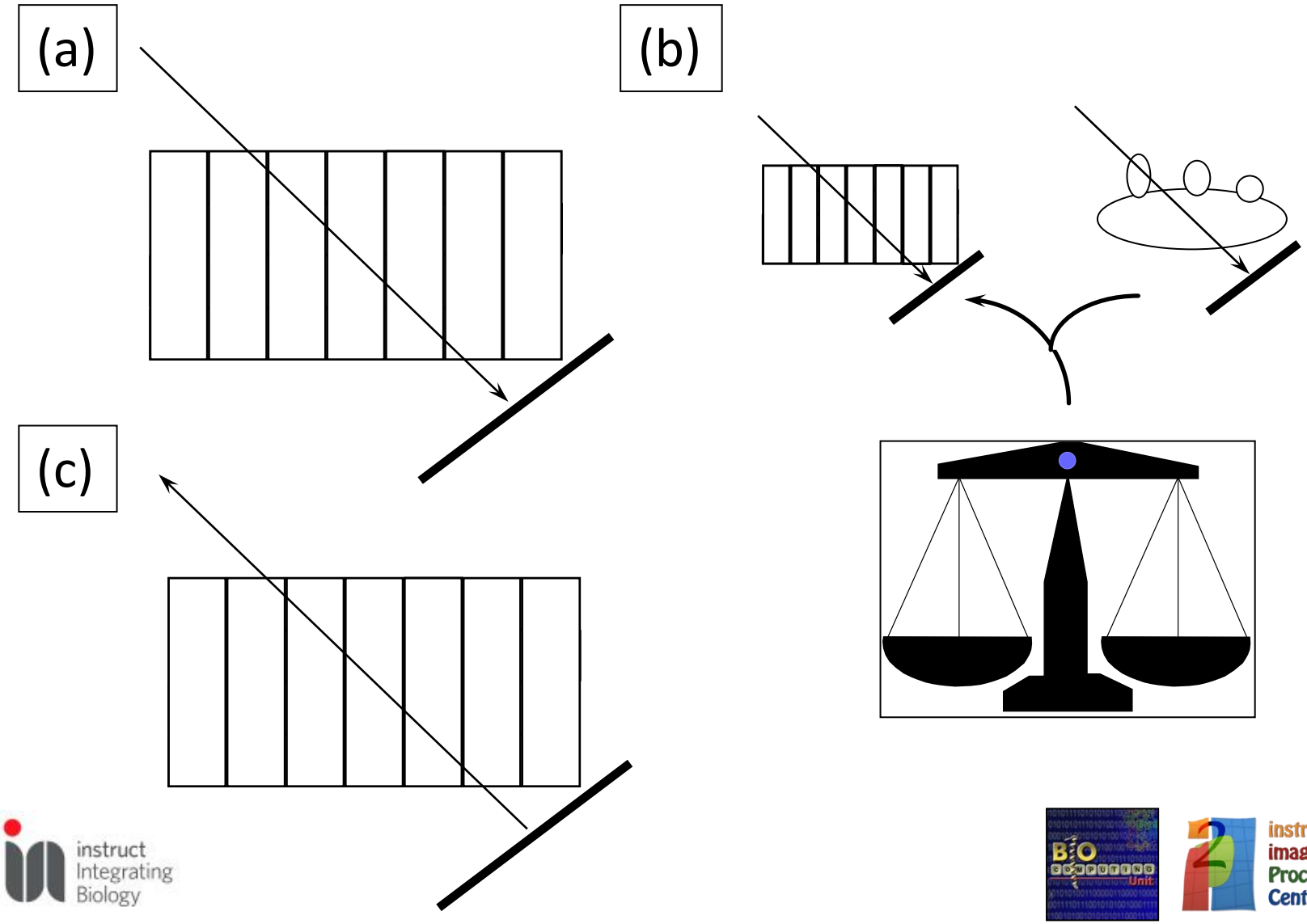
λ

CSIC

instruct Integrating Biology

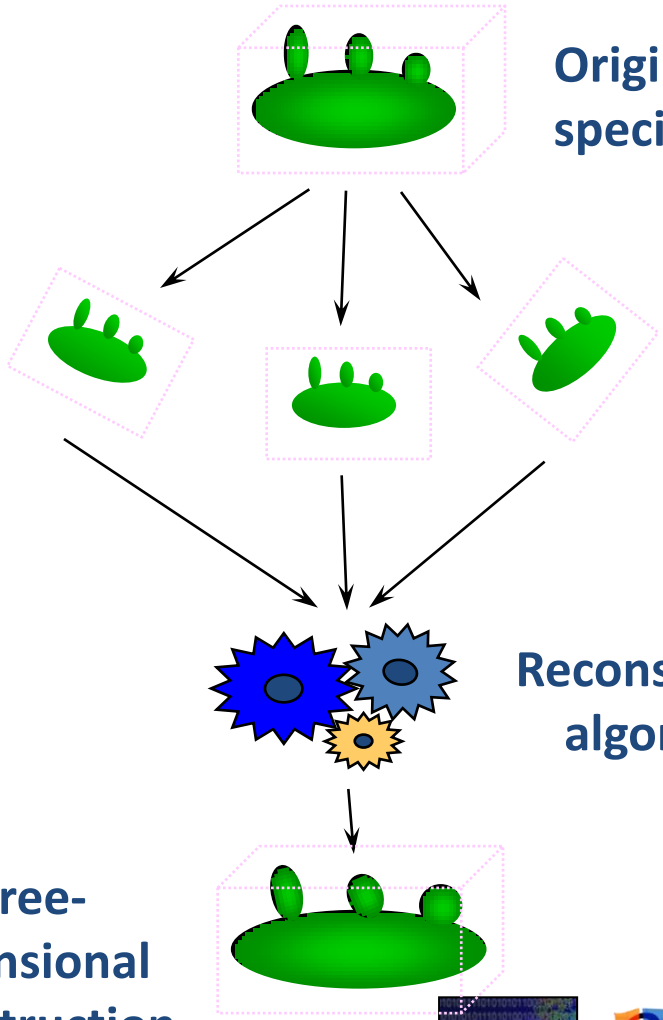
instruct image Processing Center

ART, the “basics”



3D reconstruction approach:

- Limited number of projections
- Image noise
- Partial lack of control over particle homogeneity
- Particle lack of control over data collection geometry



The cryo-EM SPA pledge

Attention to every detail!

- ***Characterization of each image***
- ***A posteriori characterization of the Projection Geometry***
- ***3D reconstruction process***
- ***3D classification***
- ***Validation***

The 3D flexibility challenge

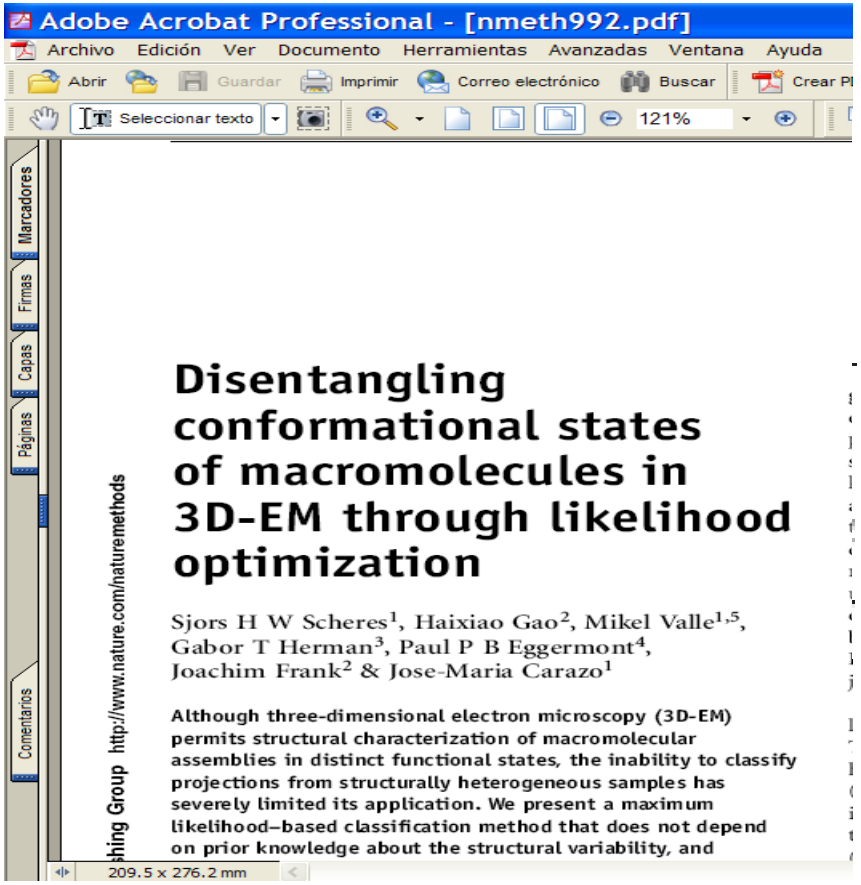
CSIC

instruct
Integrating
Biology

BIO COMPUTING Unit

instruct
image
Processing
Center

The 3D flexibility challenge



position in each image. The computational effort, using some 4,000 CPU hours on a computer cluster, is perhaps the most audacious application of the expectation-maximization algorithm ever performed. It also showcases an extremely powerful new tool for structural biology.

NOW in RELION

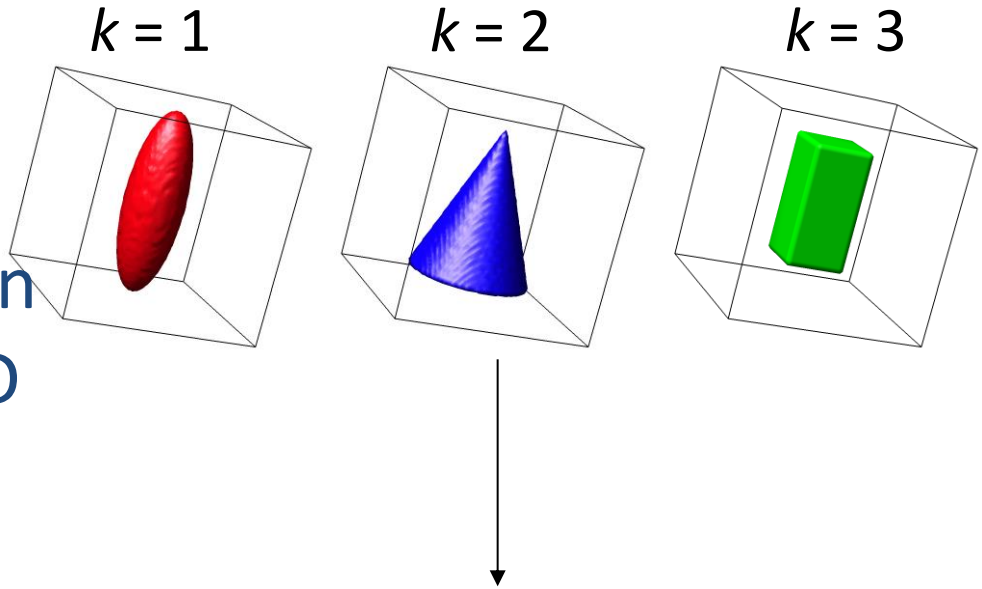
CSIC

instruct Integrating Biology

BIO COMPUTING Unit

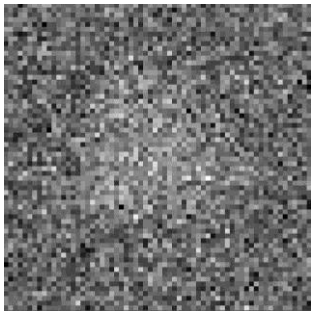
instruct image Processing Center

Statistical model



Each image is a projection of one of *K* underlying 3D objects *k*

with addition of noise

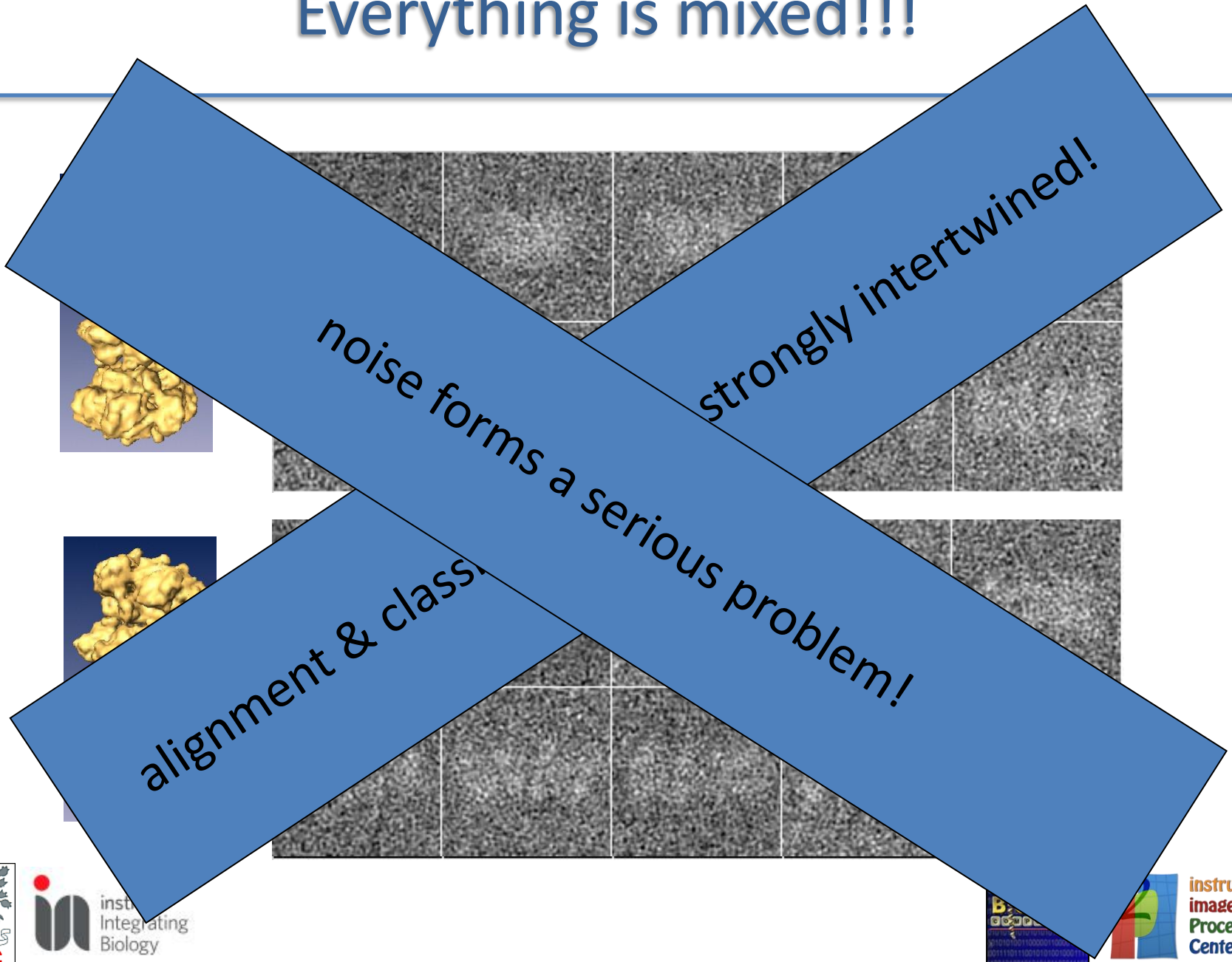


Unknowns: the 3D objects *k*, orientations

Parameter space of cryo EM SPA

- Target (the X's): A volume of (for example) 100 x 100 x 100 voxels = 10**6 variables
  - (Plus 500.000 = 5 x 10**5 geometry variables)
  - (plus 100.000 x k (classes))
- Measurements (the Y's): 100.000 particle images of 100 x 100 pixels = 10**9
- But we have noise!: 2 + 2 = 5 (or 3, or 6 …)

Everything is mixed!!!

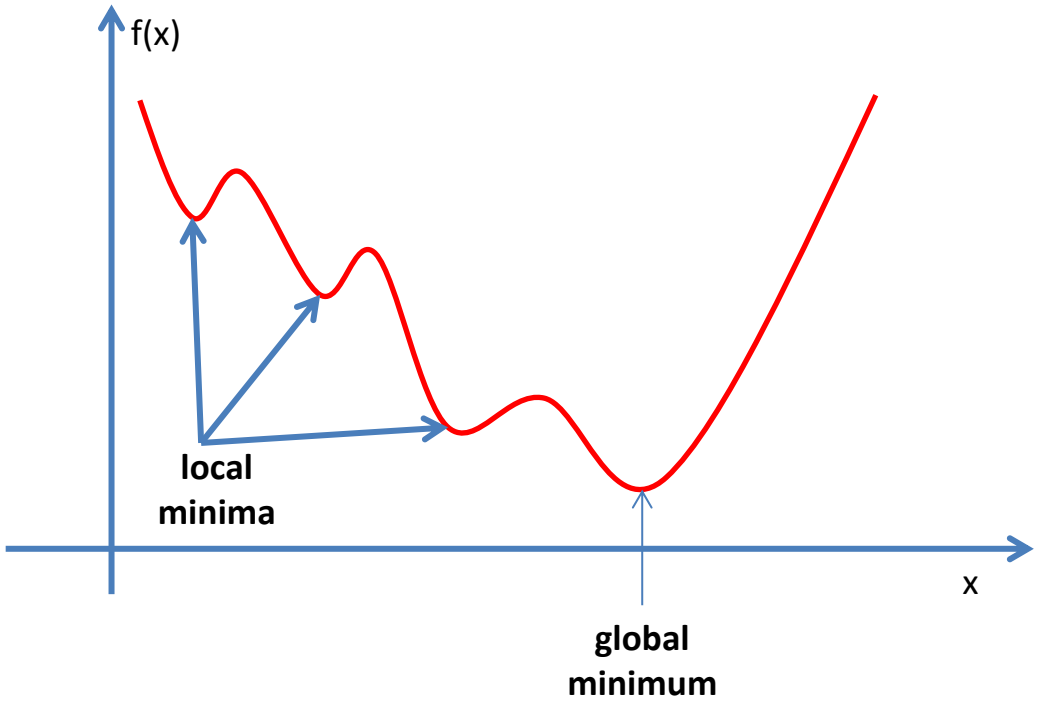


CSIC

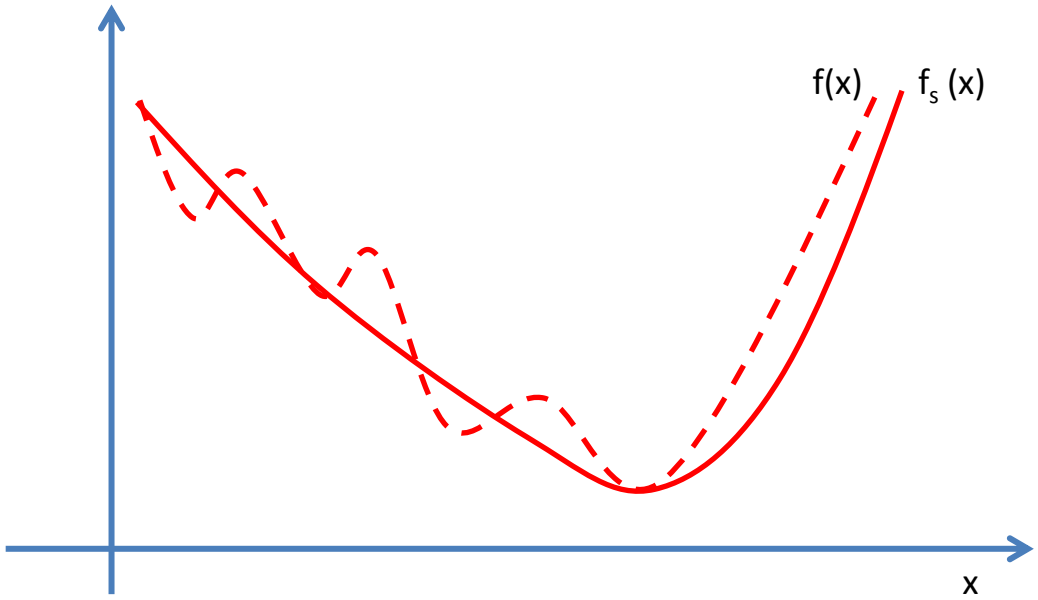
Integrating Biology

instruct image Processing Center

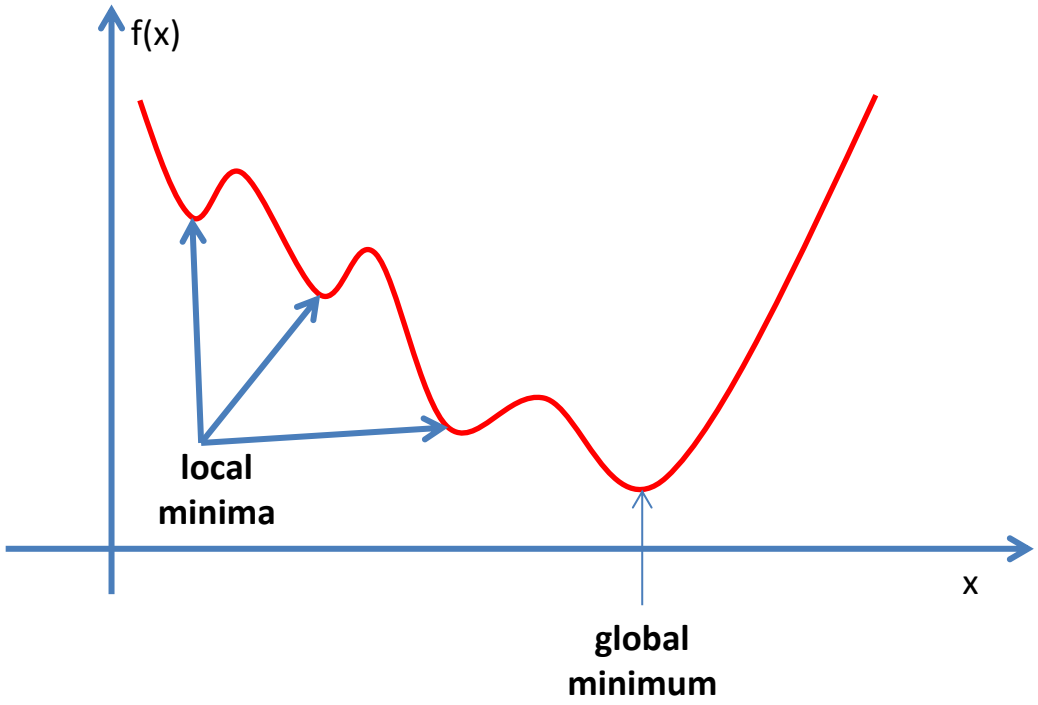
Parameter space of cryo EM SPA



Parameter space of cryo EM SPA



The Initial Volume Problem in SPA



The cryo-EM SPA pledge

Attention to every detail!

- ***Characterization of each image***
- ***A posteriori characterization of the Projection Geometry***
- ***3D reconstruction process***
- ***3D classification***
- ***Validation***

Validation

doi:10.1016/j.jmb.2011.09.008 J. Mol. Biol. (2011) 413, 1028–1046

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Tilt-Pair Analysis of Images from a Range of Different Specimens in Single-Particle Electron Cryomicroscopy

Richard Henderson[1]*, Shaoxia Chen[1], James Z. Chen[2], Nikolaus Grigorieff[2], Lori A. Passmore[1], Luciano Ciccarelli[3], John L. Rubinstein[4], R. Anthony Crowther[1], Phoebe L. Stewart[5] and Peter B. Rosenthal[6]

[1]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK*
[2]*Rosenstiel Basic Medical Science Research Center, Howard Hughes Medical Institute, Brandeis University, Waltham, MA 02154, USA*
[3]*Max Planck Institute for Biophysics, D-60438 Frankfurt, Germany*
[4]*Molecular Structure and Function Program, The Hospital for Sick Children Research Institute, Toronto, ON, Canada M5G 1X8*
[5]*Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN 37232, USA*
[6]*Division of Physical Biochemistry, MRC National Institute for Medical Research, London NW7 1AA, UK*

VALIDATION:
results of our validation method on controversial HIV data

CSIC

instruct
Integrating
Biology

BIO COMPUTING Unit

instruct
image
Processing
Center

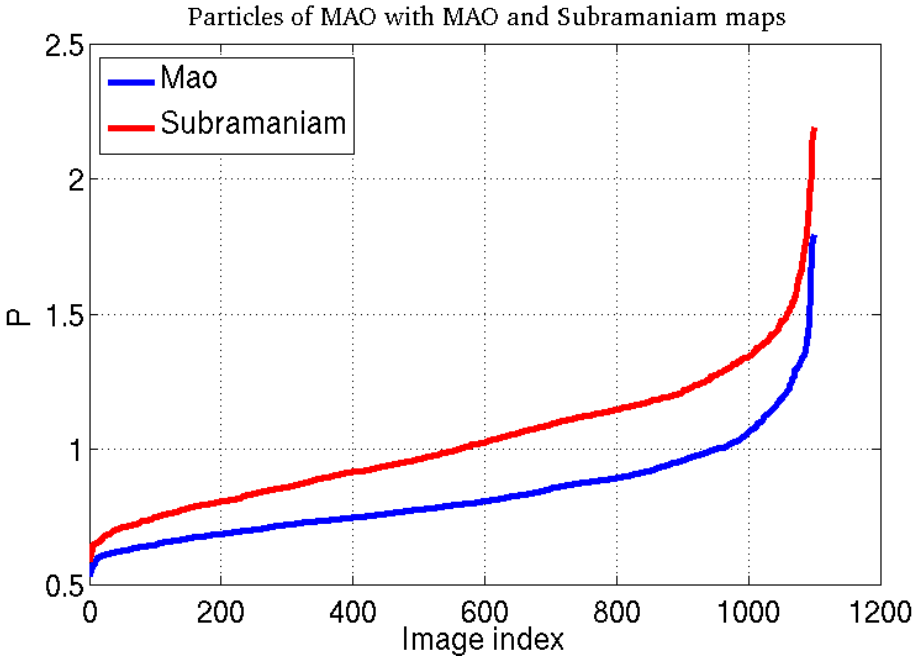
We have applied our approach to validate the map presented in the controversial work of MAO (1) and we have compared the results with the map reported by Subramaniam (2)

(1) Mao Y, et al. (2013) Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proc Natl Acad Sci USA* 110(39: 12428-12433

(2) Subramaniam S (2013) Structure of trimetric HIV-1 envelope glycoproteins. *Proc natl Acad Sci USA* 110(E4172-E4174)

We have used the data deposited at EMDB EMPIAR 10008 (MAO) & EMPIAR 10004 Subramaniam, for the validation we have used approximately 1000 particles and the EMDB maps

For the validation approach we have used first the extracted particles of MAO, which has been compared with the two maps (MAO & Subramaniam)



The quality parameters obtained are of

Q0 = 0.48 (Subramaniam)

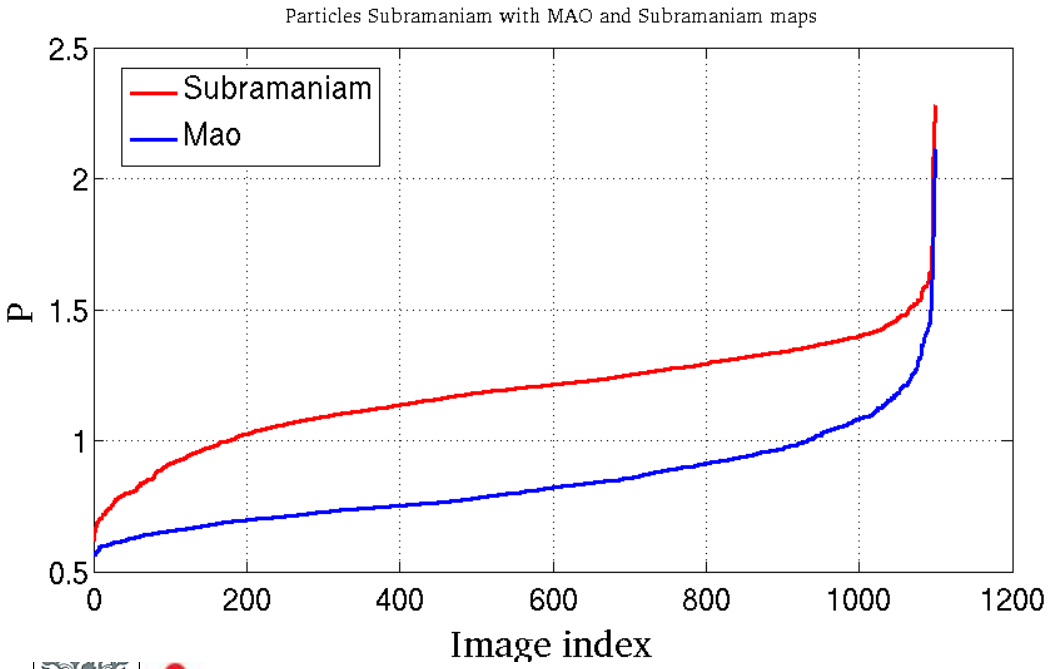
Q0 = 0.13 (Mao)

CSIC

instruct Integrating Biology

instruct image Processing Center

For the validation approach we have used first the extracted particles of MAO, which has been compared with the two maps (MAO & Subramaniam)



The quality parameters obtained are of

Q0 = 0.84 (Subramaniam)

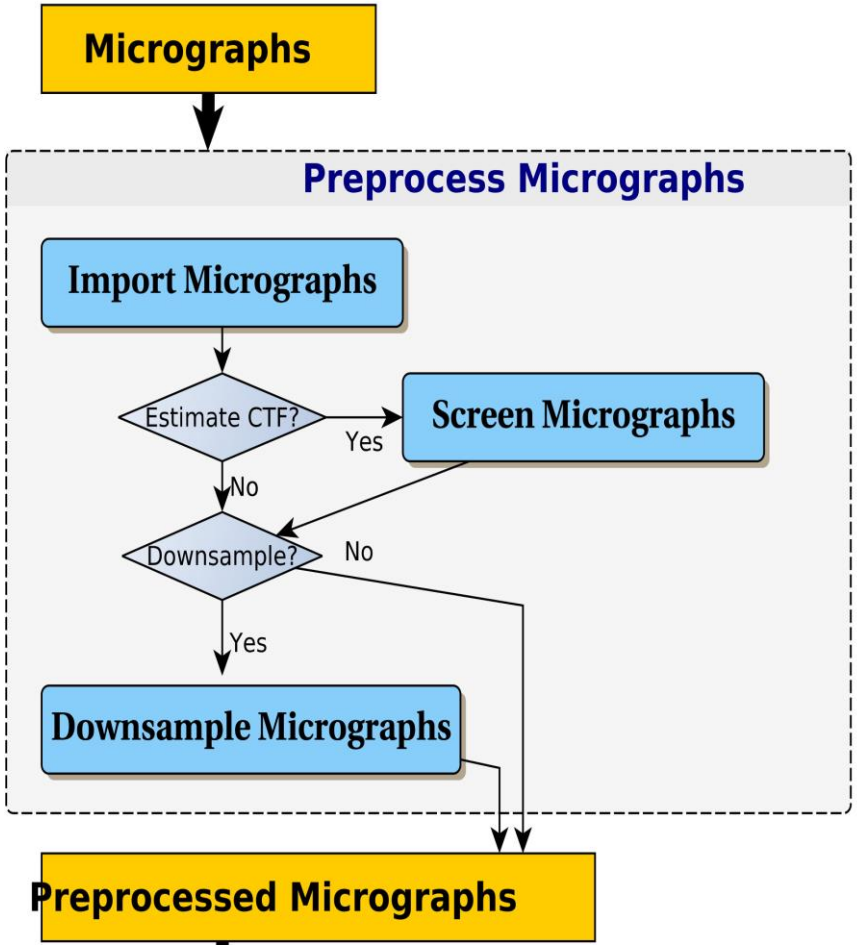
Q0 = 0.15 (Mao)

CSIC

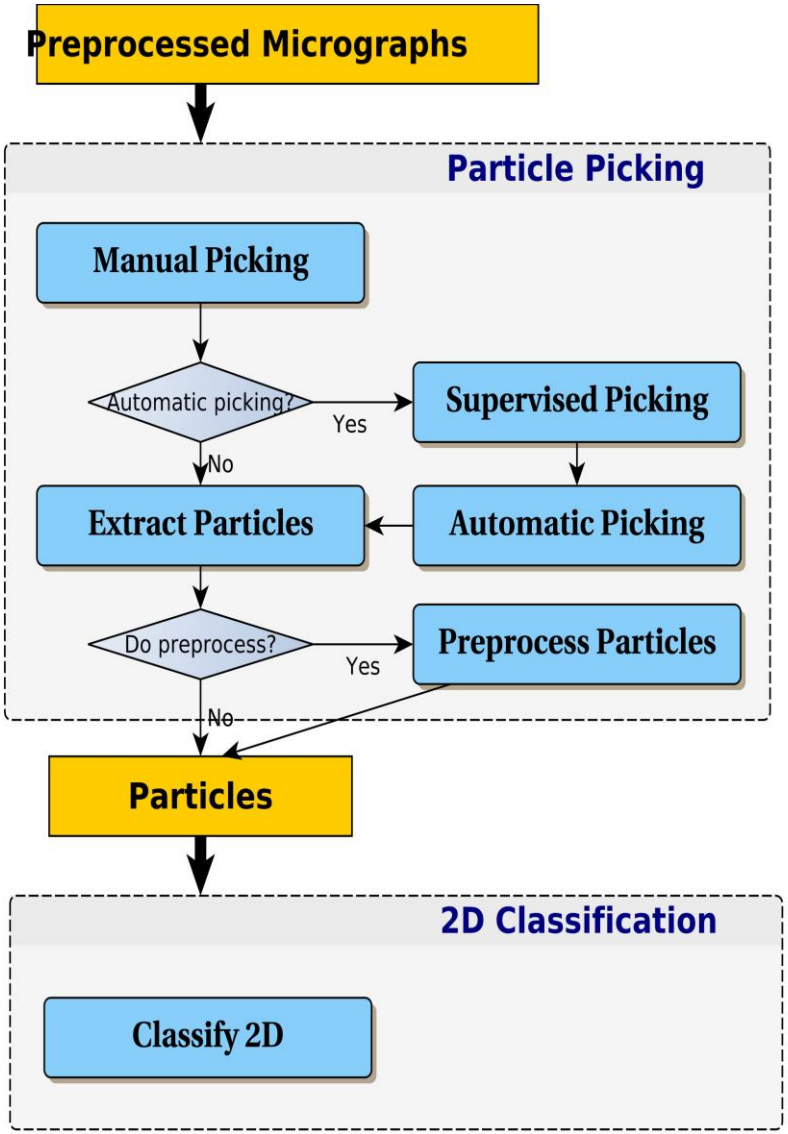
instruct Integrating Biology

instruct image Processing Center

Workflows: How do we do it in practice?



Workflows: How do we do it in practice?

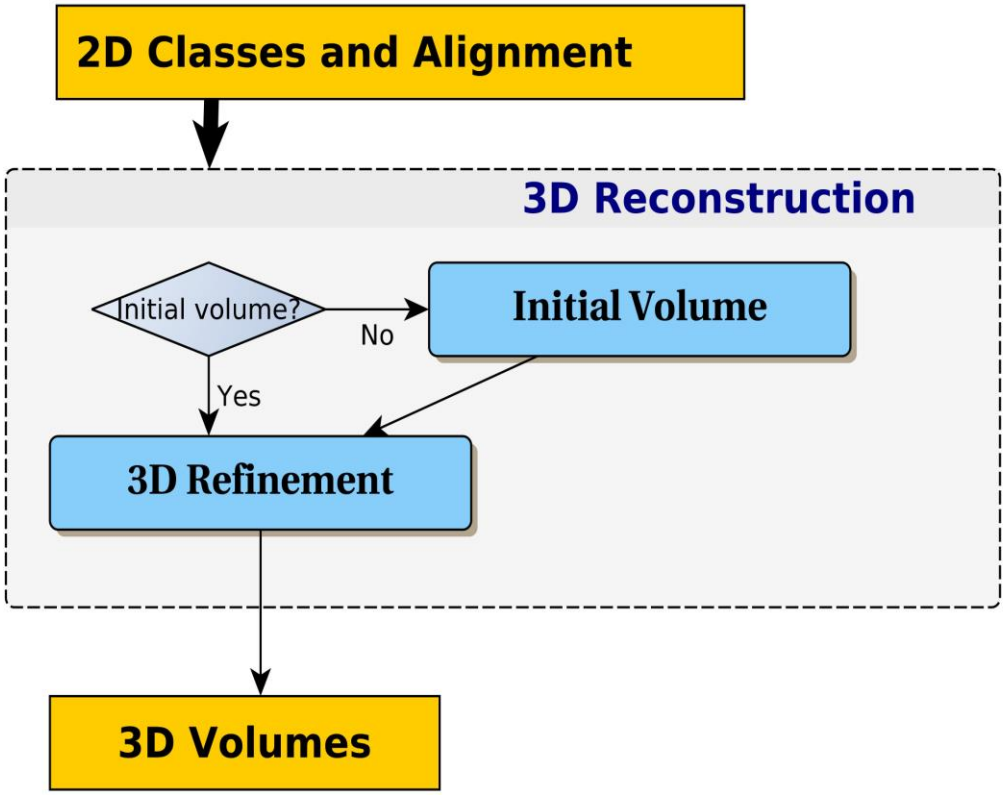


CSIC

instruct Integrating Biology

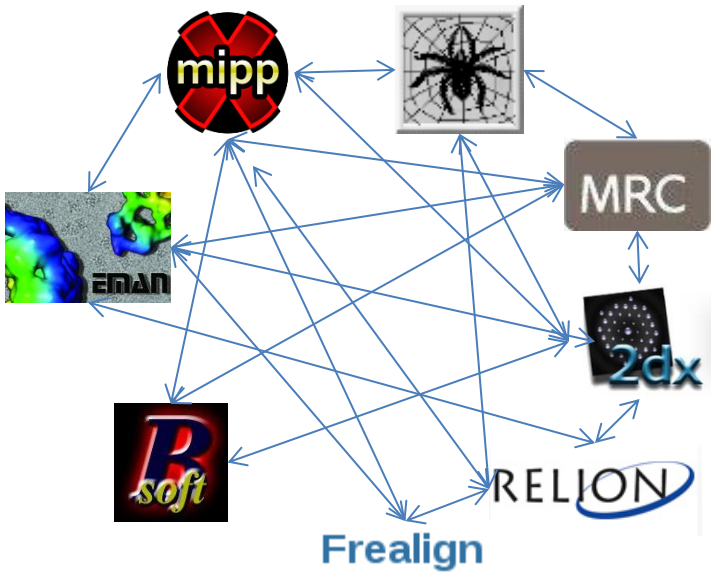
instruct image Processing Center

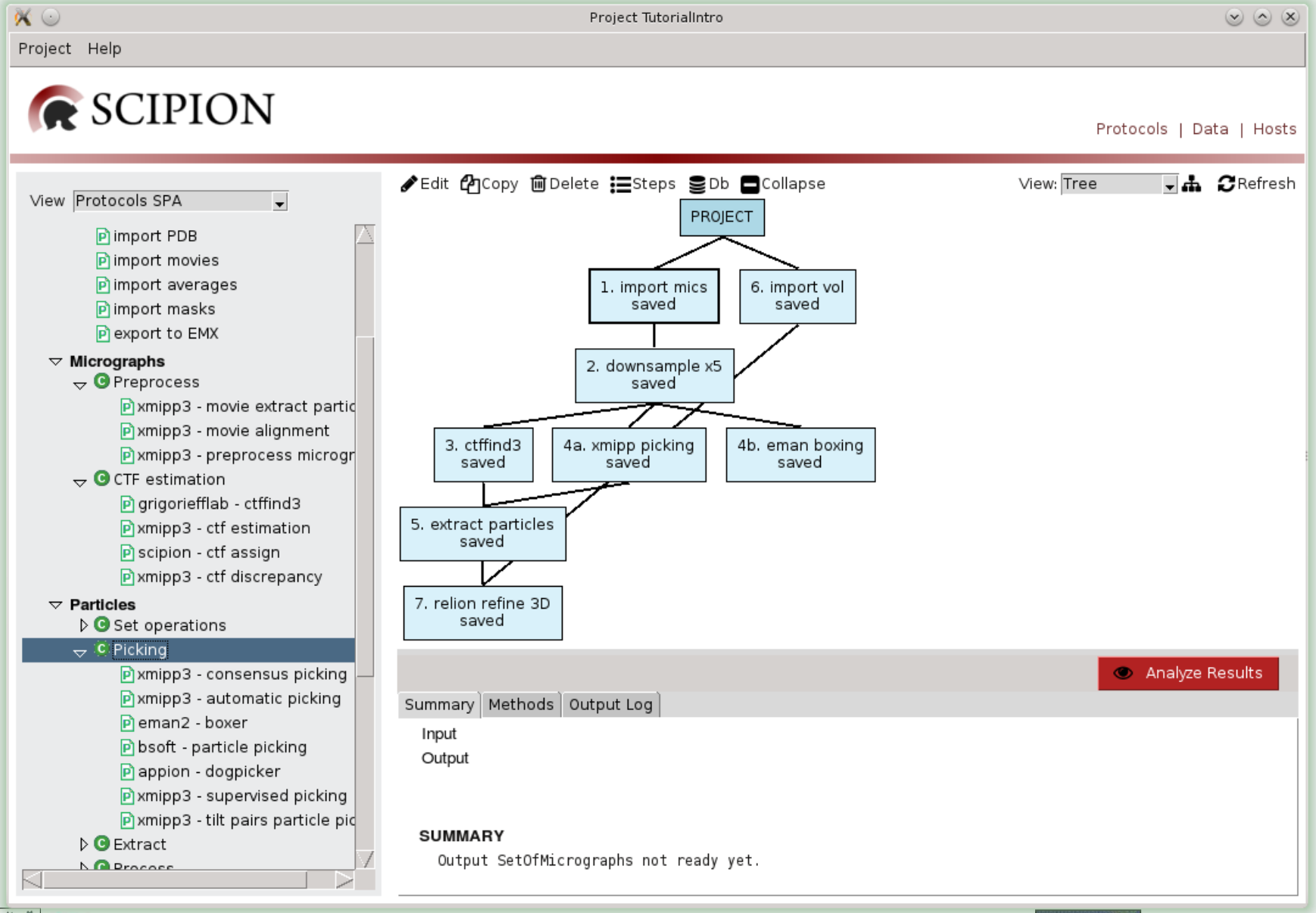
Workflows: How do we do it in practice?



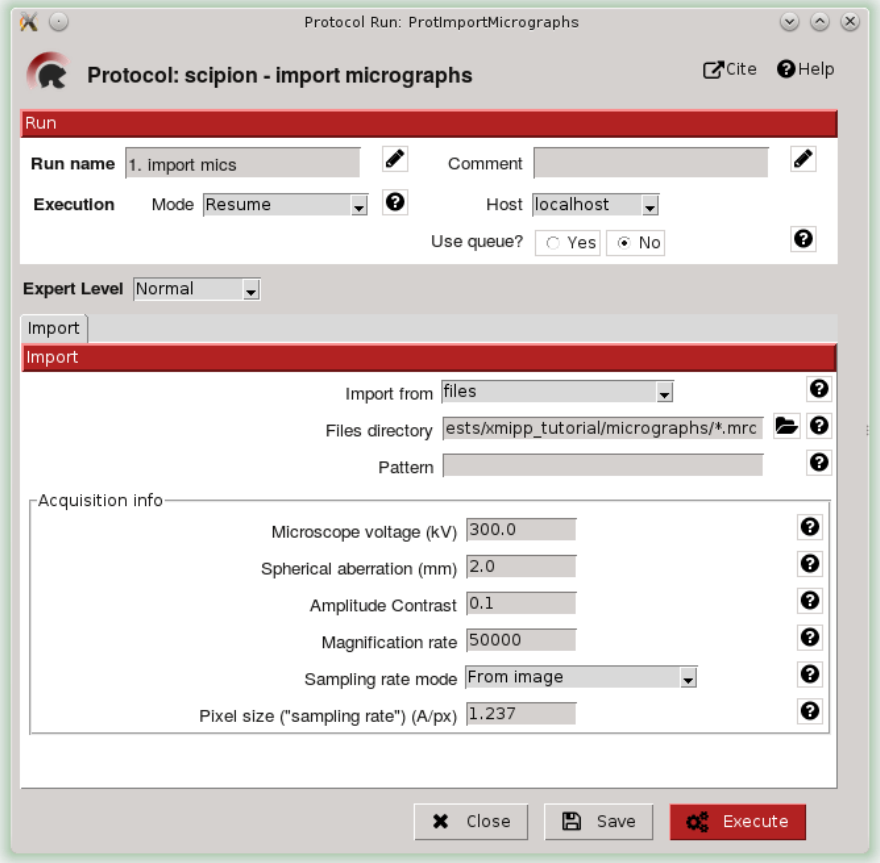
Workflows: How do we do it in practice?

Using different EM software packages is now like the tower of Babel




Project TutorialIntro
Project Help
SCIPION
Protocols | Data | Hosts
View Protocols SPA
import PDB
import movies
import averages
import masks
export to EMX
Micrographs
Preprocess
xmipp3 - movie extract partic
xmipp3 - movie alignment
xmipp3 - preprocess microgr
CTF estimation
grigoriefflab - ctffind3
xmipp3 - ctf estimation
scipion - ctf assign
xmipp3 - ctf discrepancy
Particles
Set operations
Picking
xmipp3 - consensus picking
xmipp3 - automatic picking
eman2 - boxer
bsoft - particle picking
appion - dogpicker
xmipp3 - supervised picking
xmipp3 - tilt pairs particle pic
Extract
Edit
Copy
Delete
Steps
Db
Collapse
View: Tree
Refresh
PROJECT
1. import mics saved
6. import vol saved
2. downsample x5 saved
3. ctffind3 saved
4a. xmipp picking saved
4b. eman boxing saved
5. extract particles saved
7. relion refine 3D saved
Analyze Results
Summary
Methods
Output Log
Input
Output
SUMMARY
Output SetOfMicrographs not ready yet.

Importing micrographs

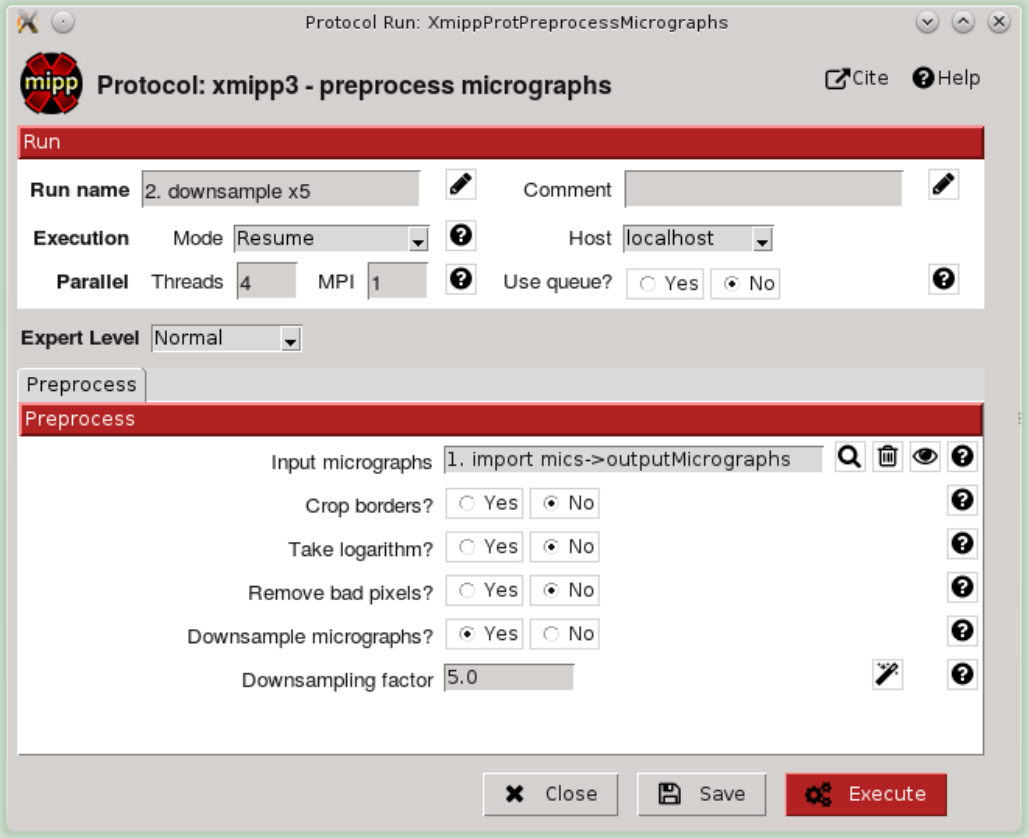


CSIC

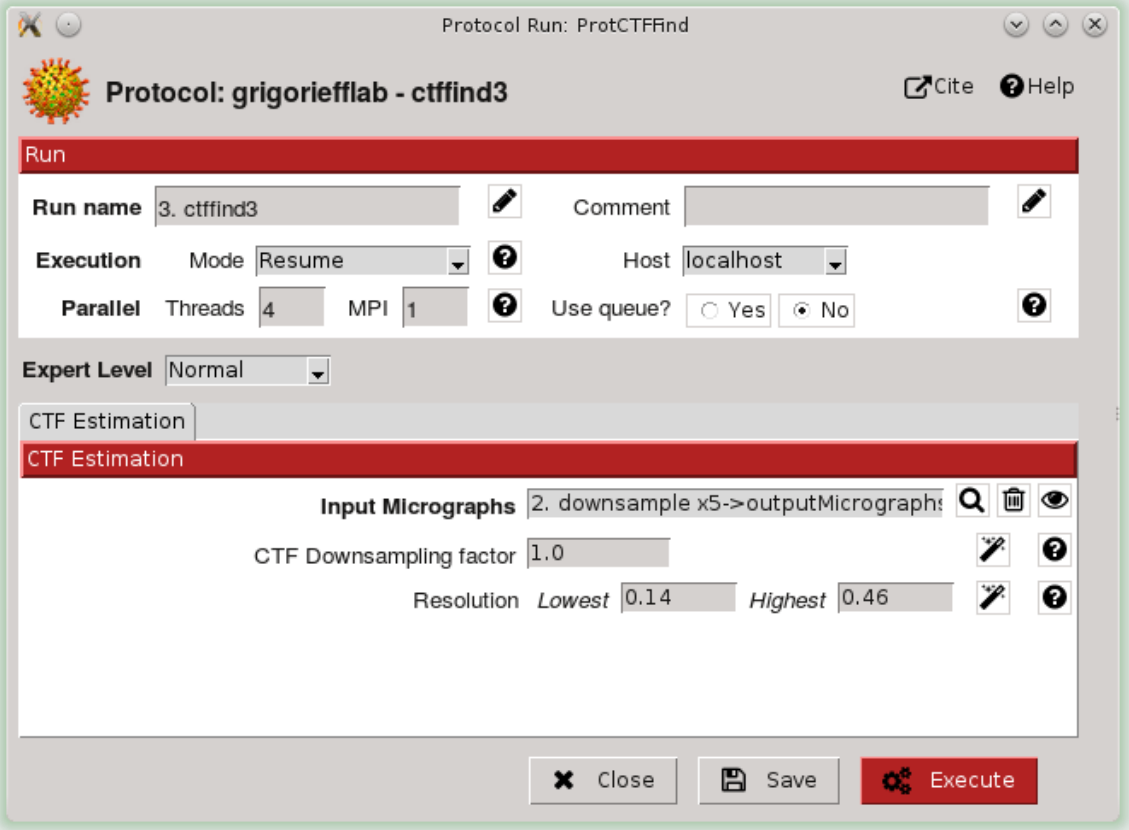
instruct Integrating Biology

instruct image Processing Center

Preprocessing micrographs



Calculating CTF



CSIC

instruct Integrating Biology

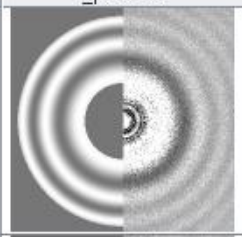
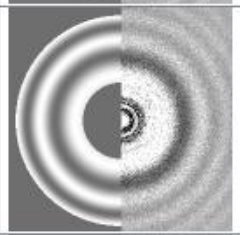
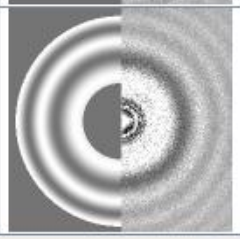
instruct image Processing Center

Typical CTF displays

File Display Help

50 1 Cols 8 Rows 3

Block CTFModels

	id	enabled	comment	_psdFile	_defocusU	_defocusV	_defocusAngle	
1	1	☑			23751,9395	23901,1504	-36,8300	Runs/000070_XmippPr
2	2	☑			21755,3906	21923,6309	-31,6300	Runs/000070_XmippPr
3	3	☑			22199,0195	22382,5801	-33,5800	Runs/000070_XmippPr

Close Micrographs Recalculate CTFs

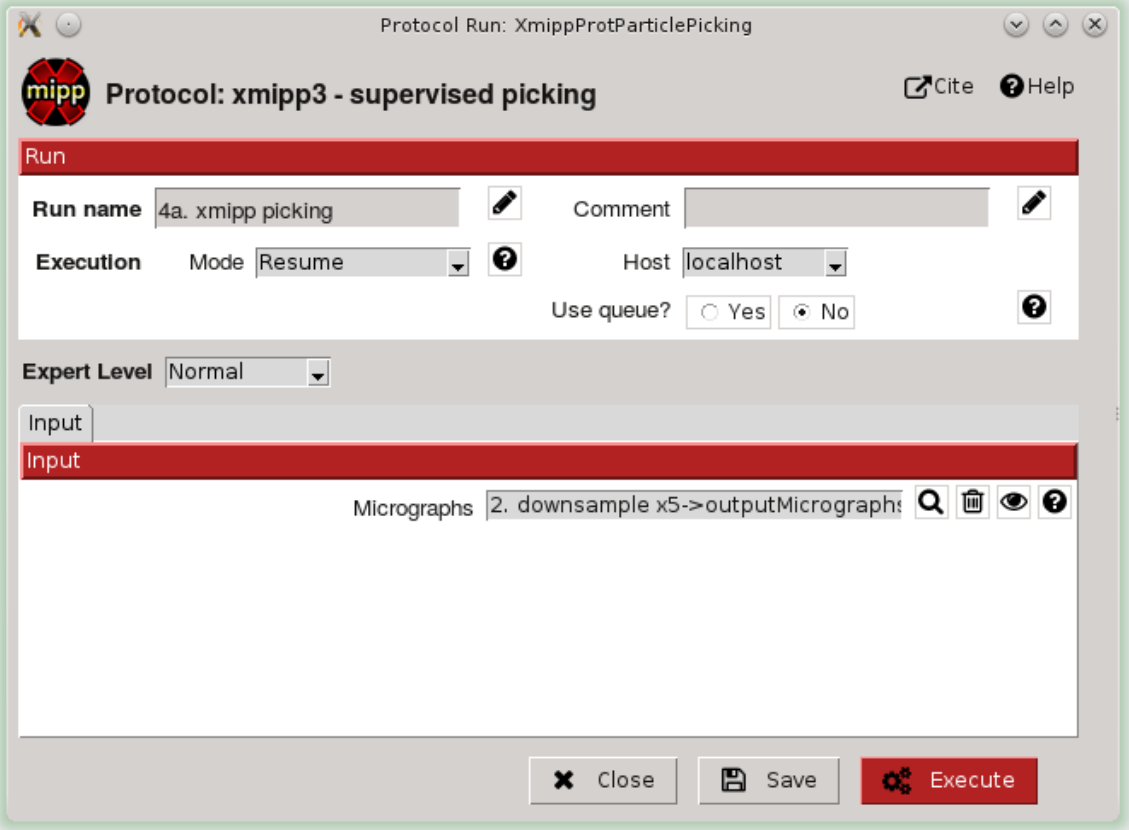
CSIC

instruct Integrating Biology

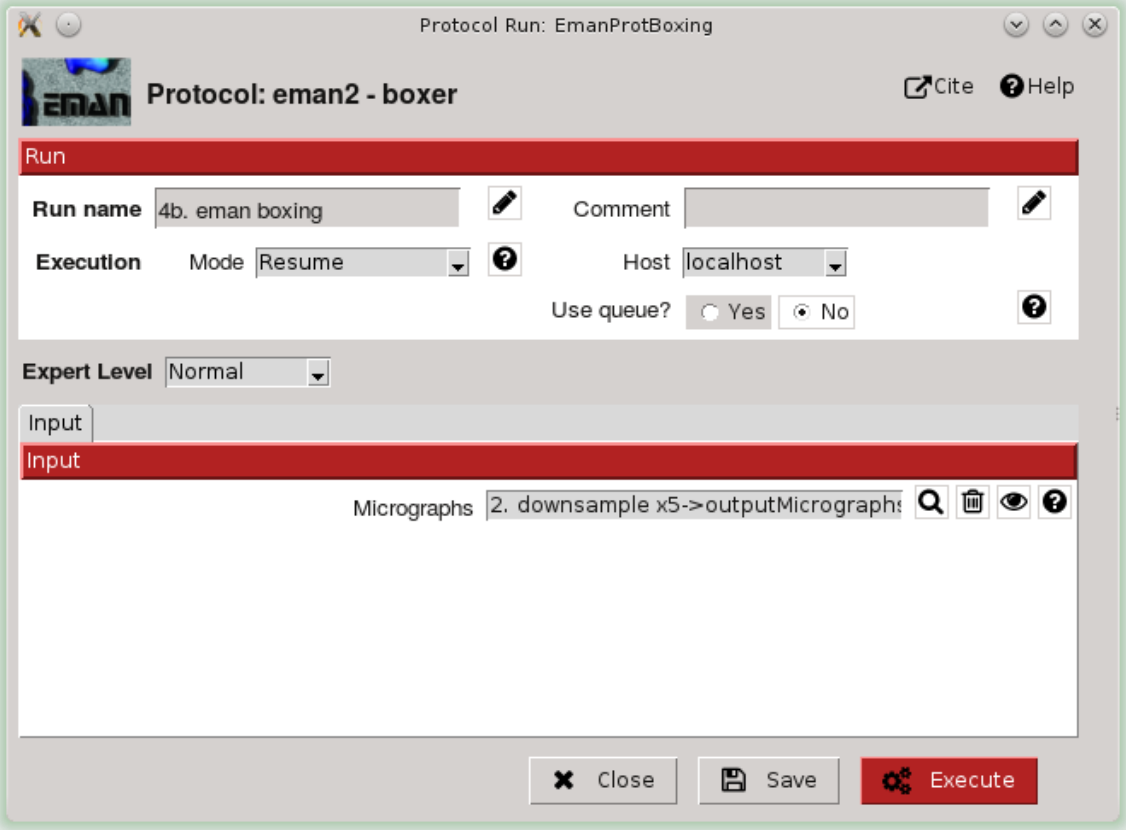
BIO COMPUTING Unit

instruct image Processing Center

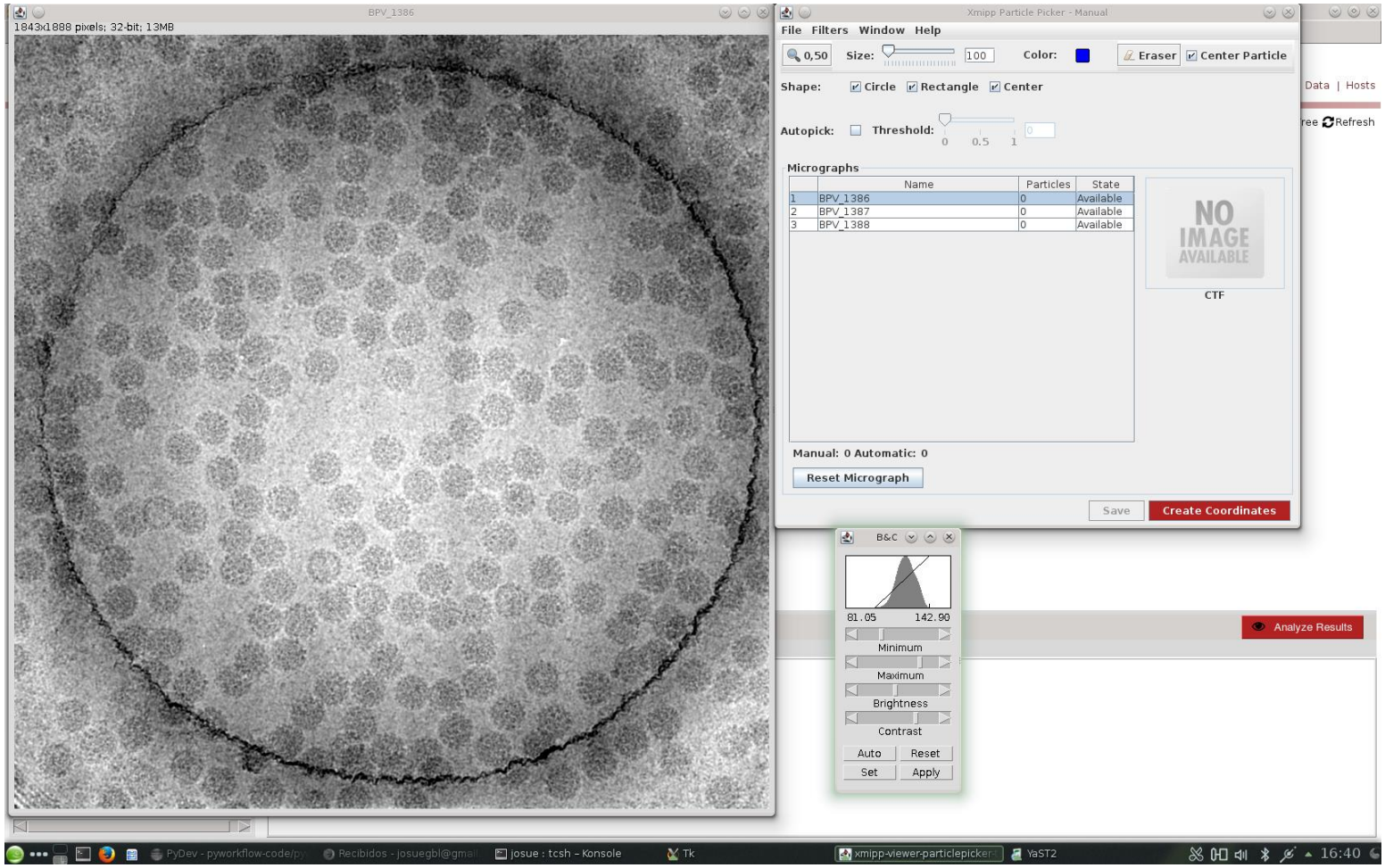
Particle picking



Particle picking



Typical particle picking display



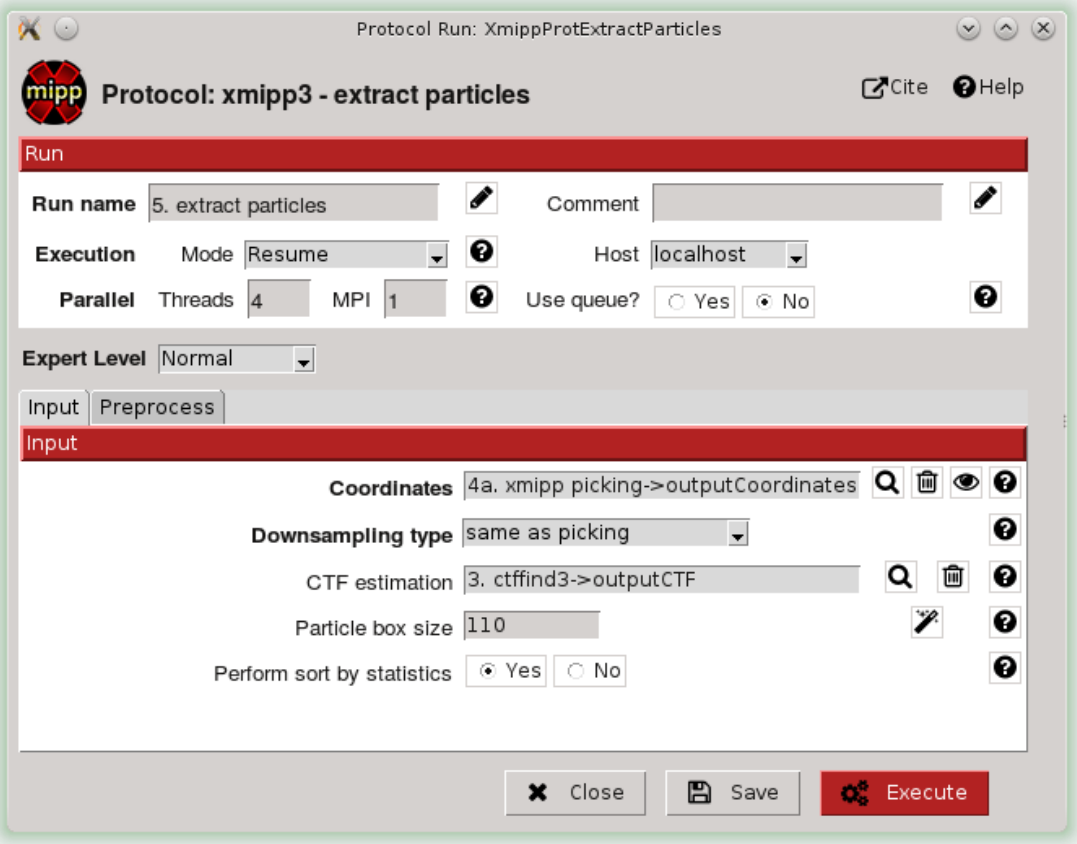
CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

Extracting particles

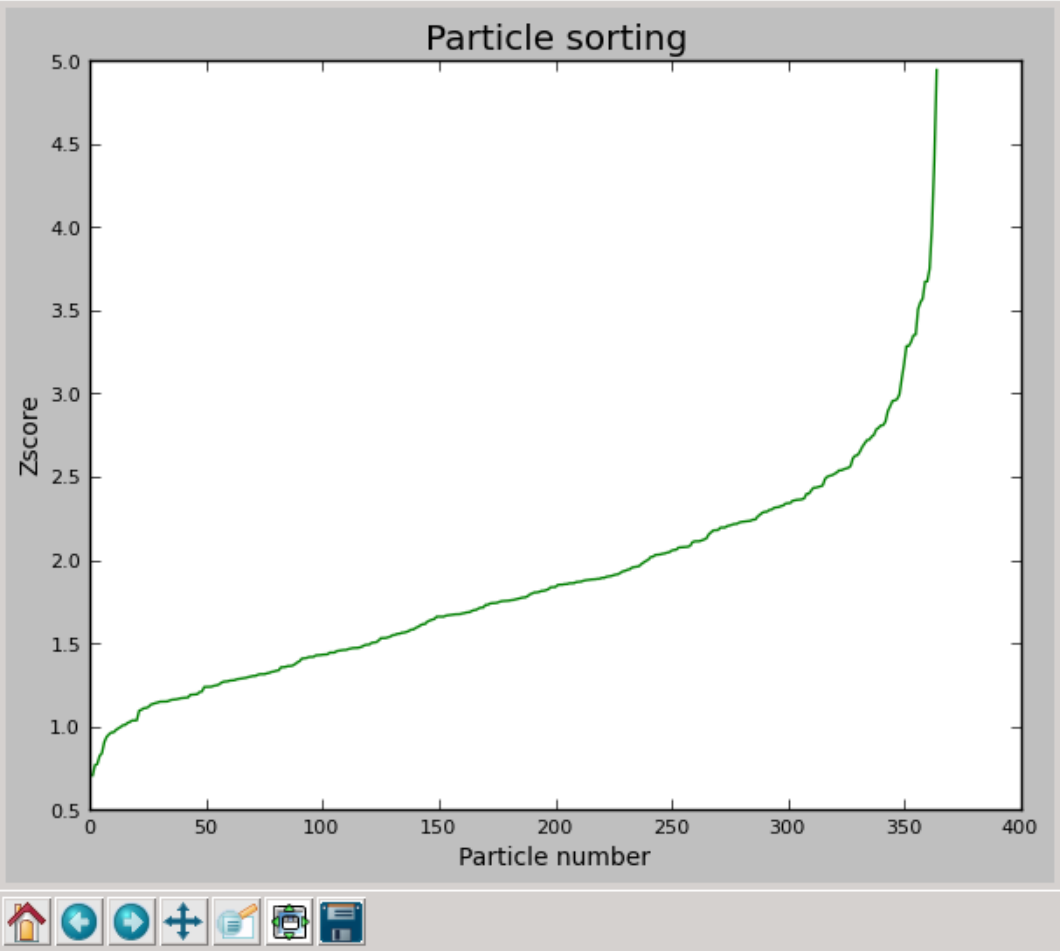


CSIC

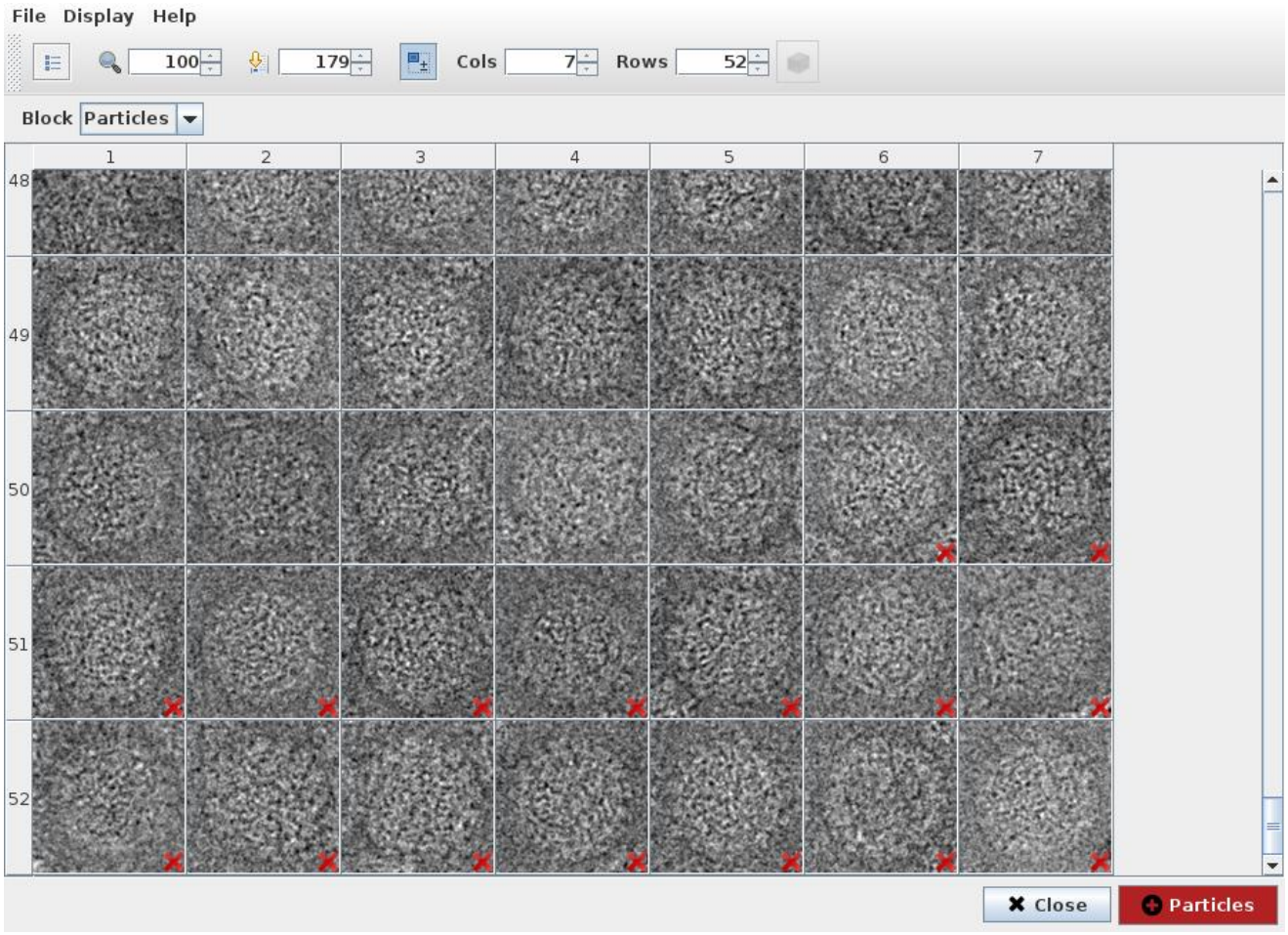
instruct Integrating Biology

instruct image Processing Center

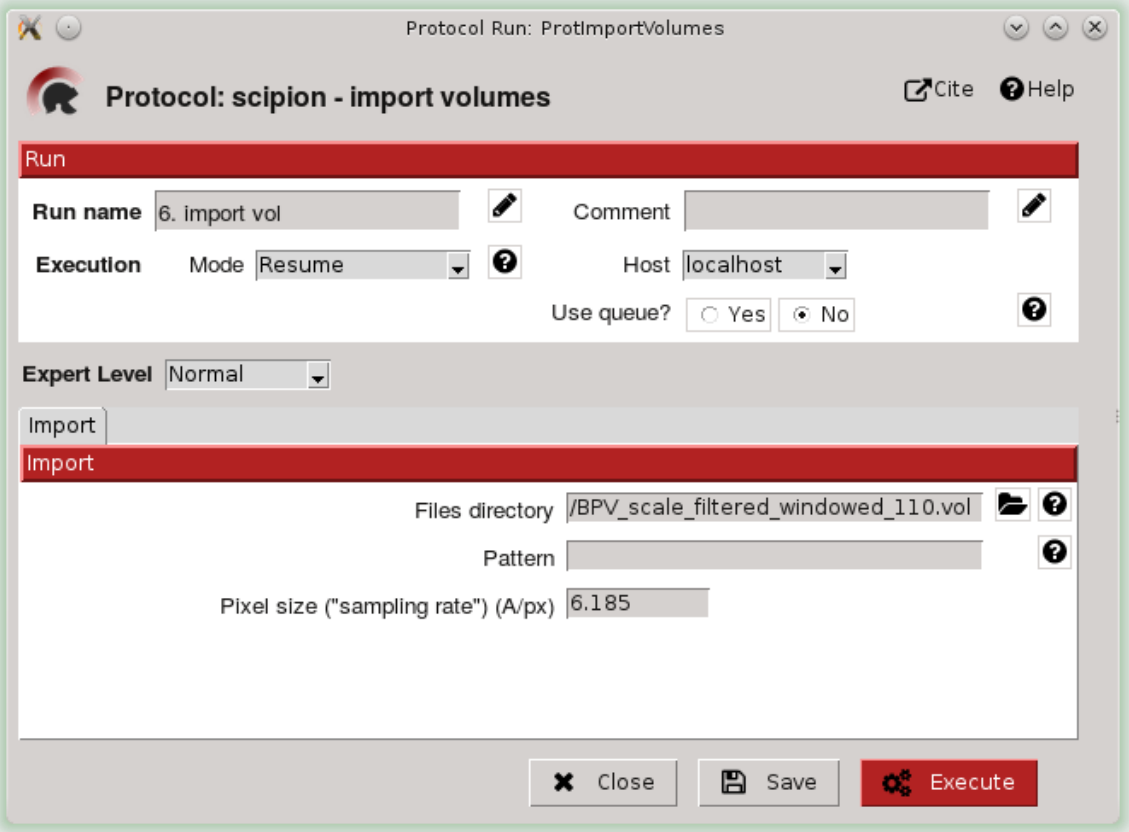
Sorting particles



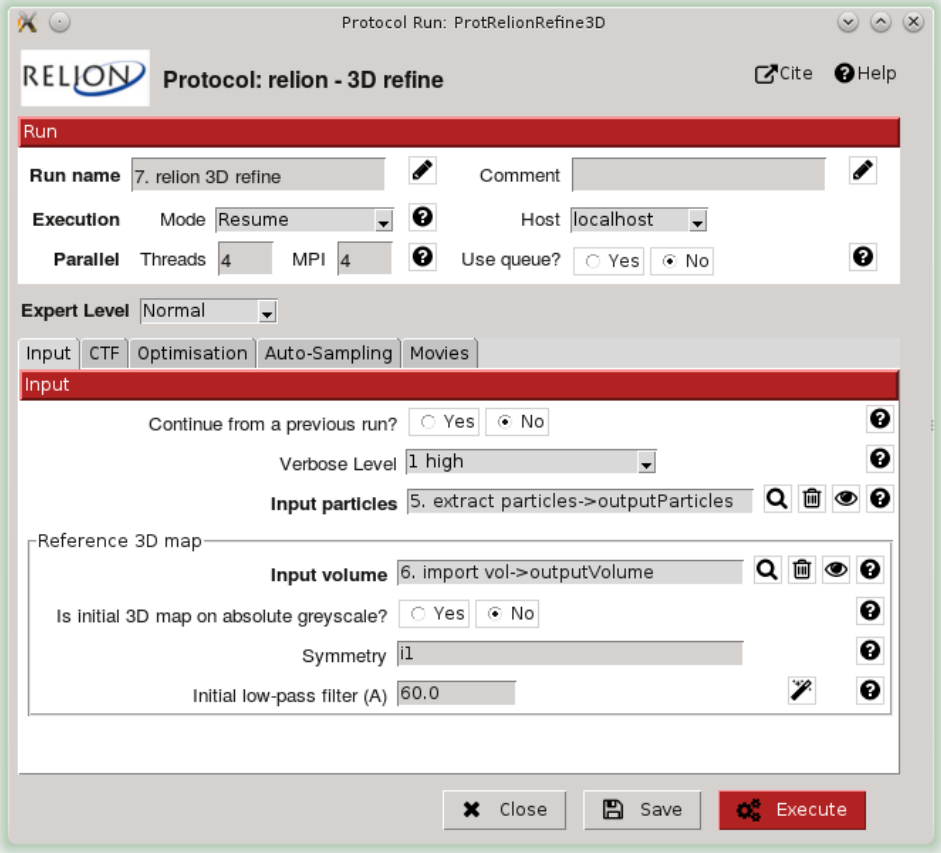
Typical sorting display



Importing Initial Volume (and the "Problem"?)



Finding Projection Geometry



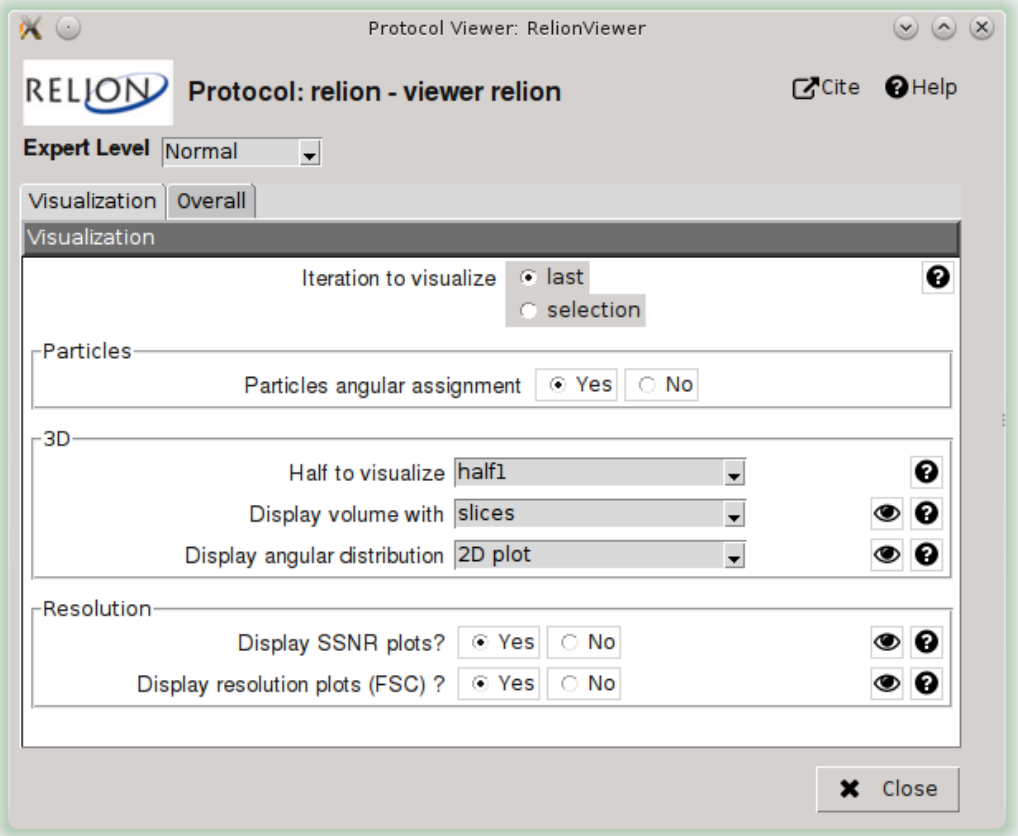
CSIC

instruct Integrating Biology

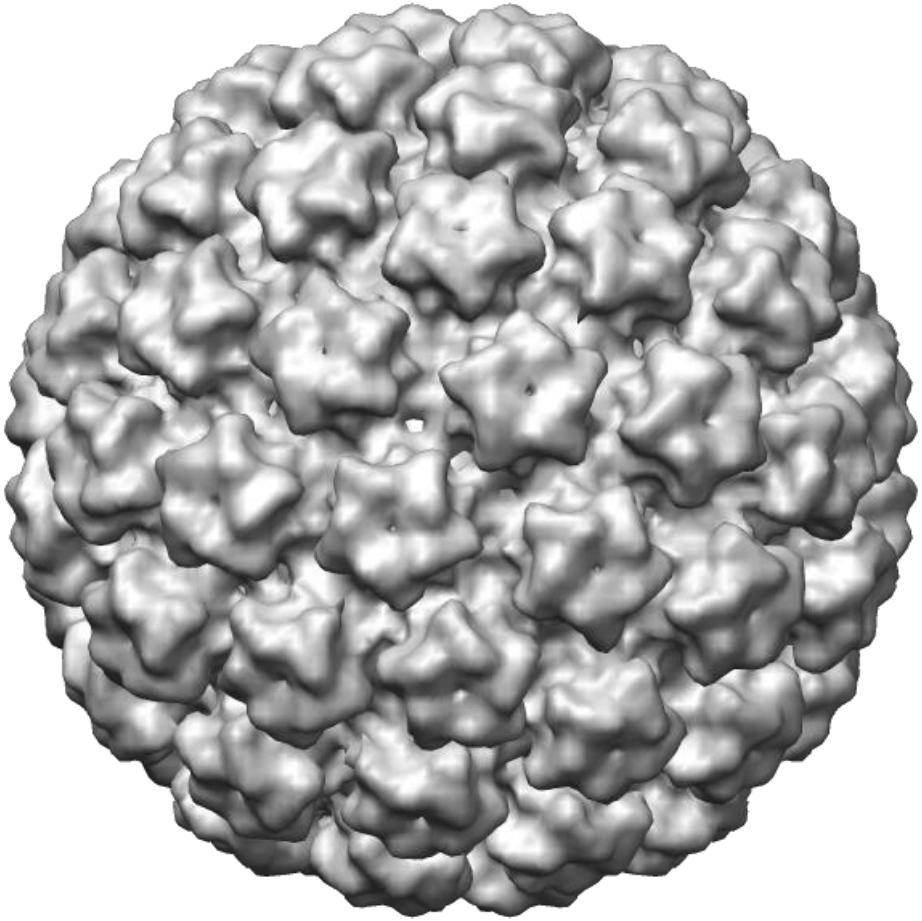
BIO COMPUTING Unit

instruct image Processing Center

Typical Relion 3D display



Typical Volume display

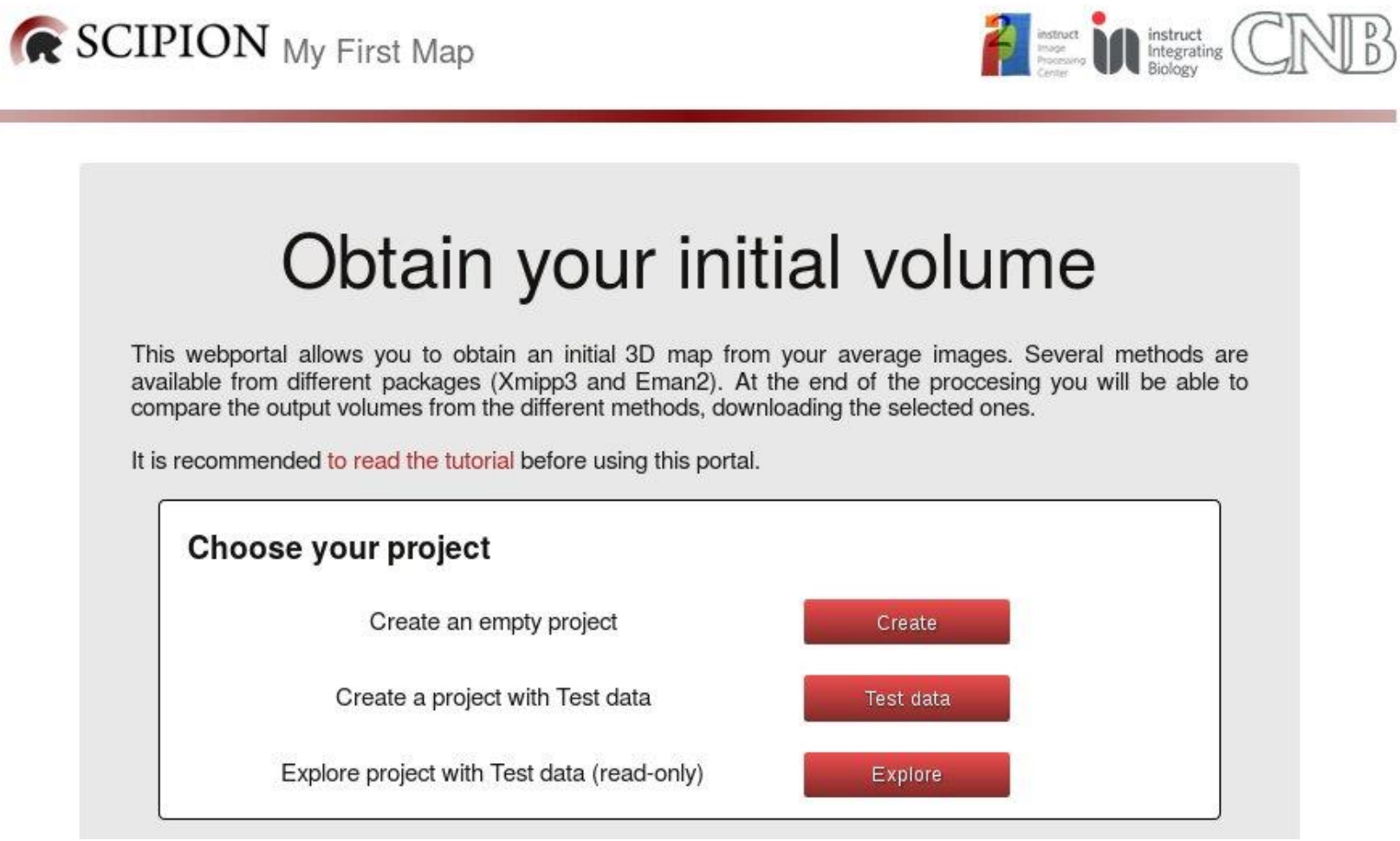


CSIC

instruct
Integrating
Biology

instruct
image
Processing
Center

The Initial Volume Problem (in the Web)



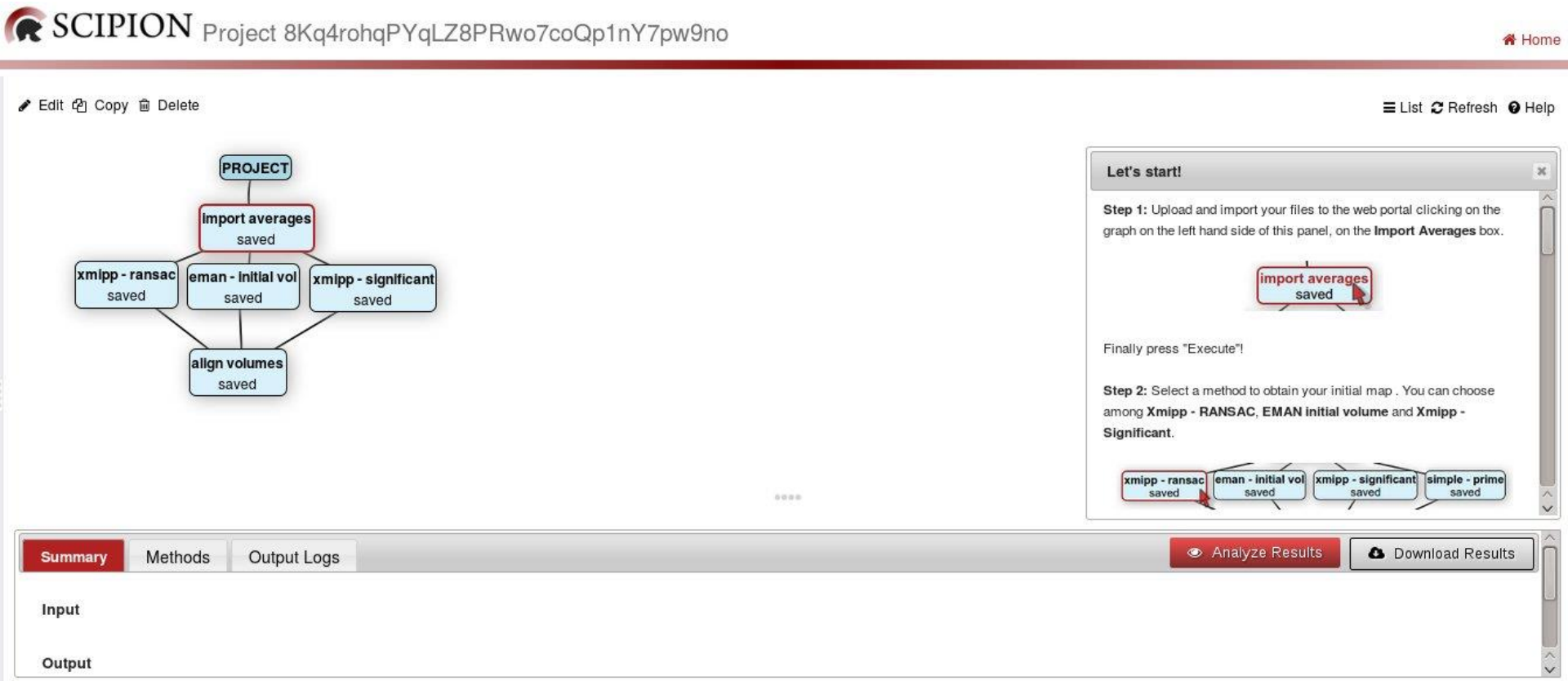
CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

The Initial Volume Problem (in the Web)



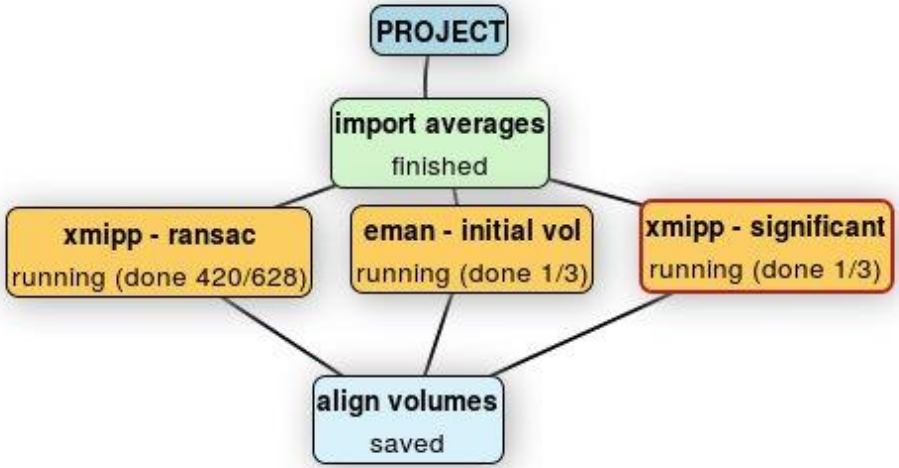
CSIC

instruct Integrating Biology

BIO COMPUTING Unit

instruct image Processing Center

The Initial Volume Problem (in the Web)

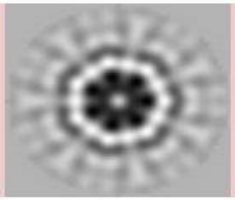
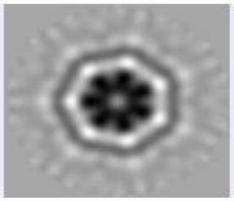
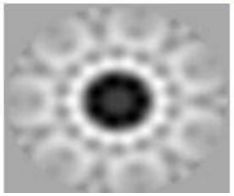


The Initial Volume Problem (in the Web)

SCIPION Xmipp Showj Web

Select Block Volumes

Search:

id	enabled	comment	_filename
1	☑	ransac volume 01	
2	☑	ransac volume 02	
3	☑	ransac volume 03	

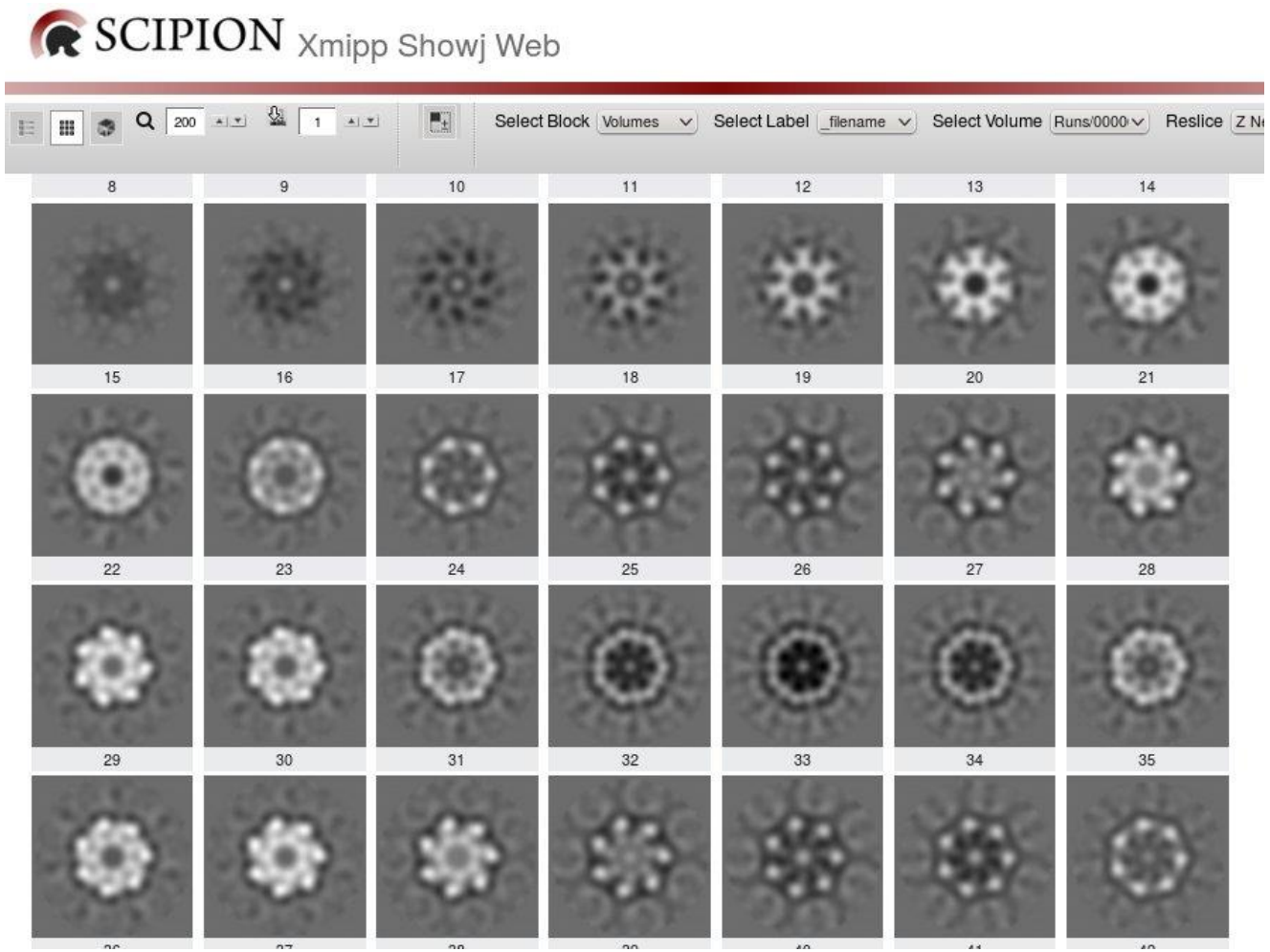
CSIC

instruct Integrating Biology

BO Computing Unit

instruct image Processing Center

Typical 3D displays

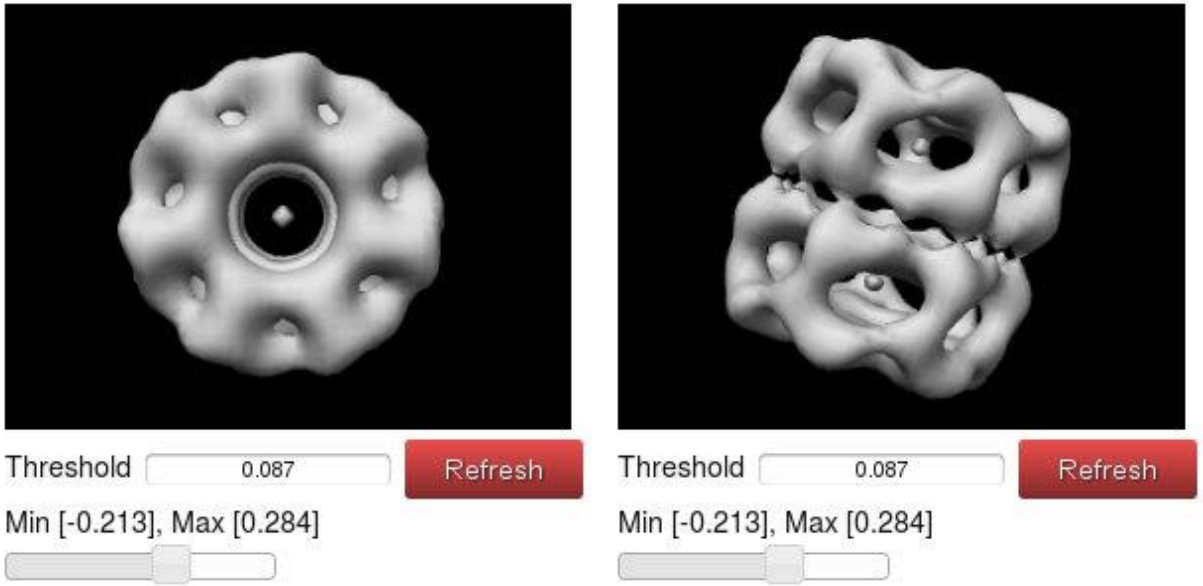


CSIC

instruct Integrating Biology

instruct image Processing Center

Typical 3D displays



Instruct Open call for Access

instruct Integrating Biology

Search...

Home | Access | Training | Information | Network | Register | Login | Submit Proposal

Home > Support > What's on > News > Call for proposals utilising Instruct-funded structural biology techniques

Call for proposals utilising Instruct-funded structural biology techniques

13th Feb 2014

Call for proposals utilising Instruct-funded structural biology techniques

Information on molecular structure provides insight in many areas of research including functional pathways.

Are you new to structural biology research? This fast-developing technology field may offer new approaches to further your work.

Do you have local or national access to ALL the structural biology instrumentation you need for your research? You might qualify for free access to some of the best structural biology facilities in Europe.

Instruct has 16 Instruct Centres around Europe and in Israel offering access to a full range of advanced technologies for obtaining structural and dynamic information on systems in various dimensions, complexity and resolution. Instruct also ensures expert support at the facility for the best research outcome.

Access to the Instruct Centre technologies is FREE to scientists from Instruct member countries.

Academic students and scientists who are working in laboratories of any of the Instruct member countries (presently BE, CZ, ES, FR, DE, IL, IT, NL, PT, SE, UK) are eligible to apply for access. Access is free and we pay your travel and consumables. Proposals should outline an integrated programme of research that requires one or

Latest Jobs

Protein Crystallographer
United Kingdom

Post-doctoral position in structural biology
Ghent University, Belgium

Post doctoral position
Copenhagen, Denmark

PhD Position
Dublin, Ireland

View all (4)

Join

Already a member?
Log in here

instruct
Integrating
Biology

www.structuralbiology.eu

A distributed infrastructure for integrated structural biology

CSIC

instruct
Integrating
Biology

instruct
image
Processing
Center