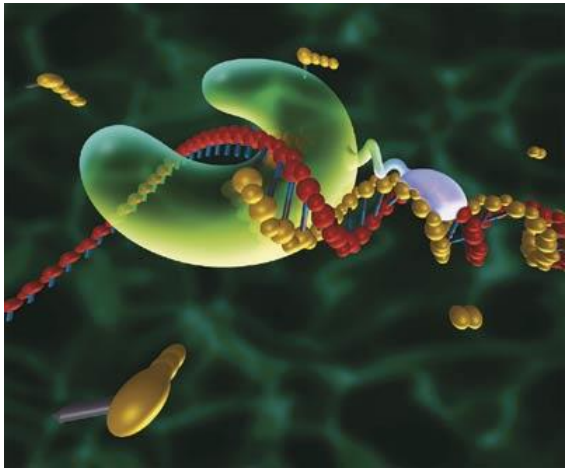
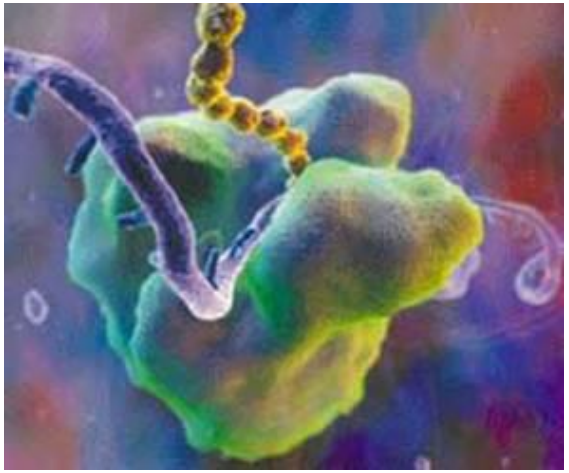
Macromolecular flexibility: The Zernike3D approach

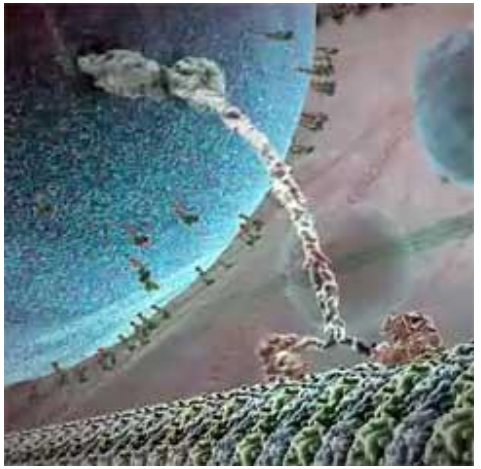
José-Maria Carazo

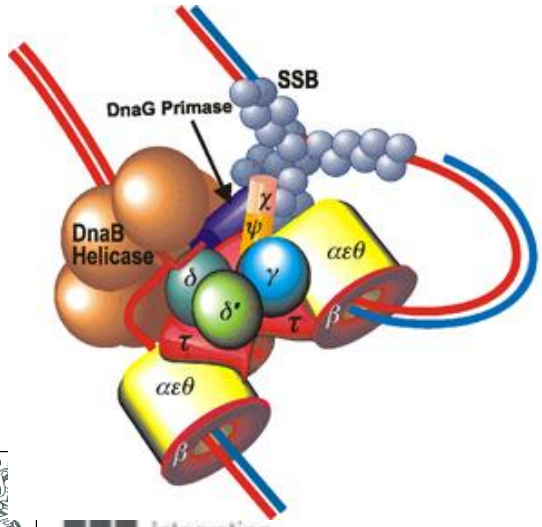
Biocomputing Unit, CNB-CSIC
INSTRUCT-ES. Instruct Image Processing Center, Madrid

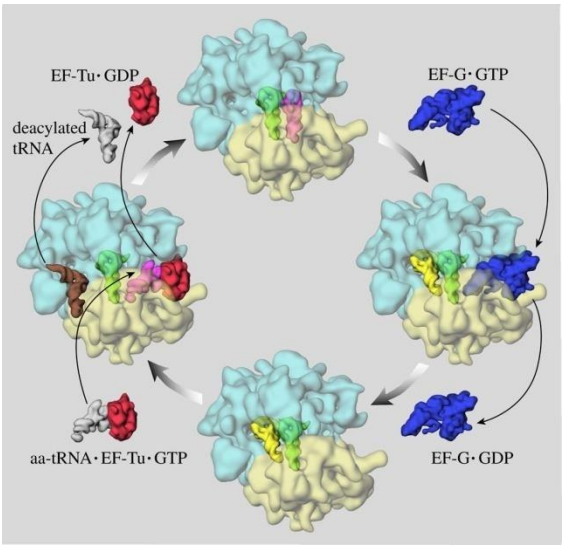
Life …. and atomic interactions

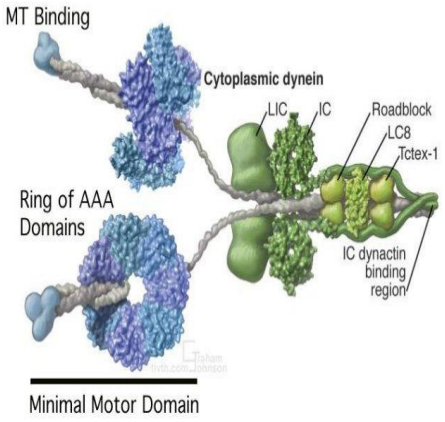










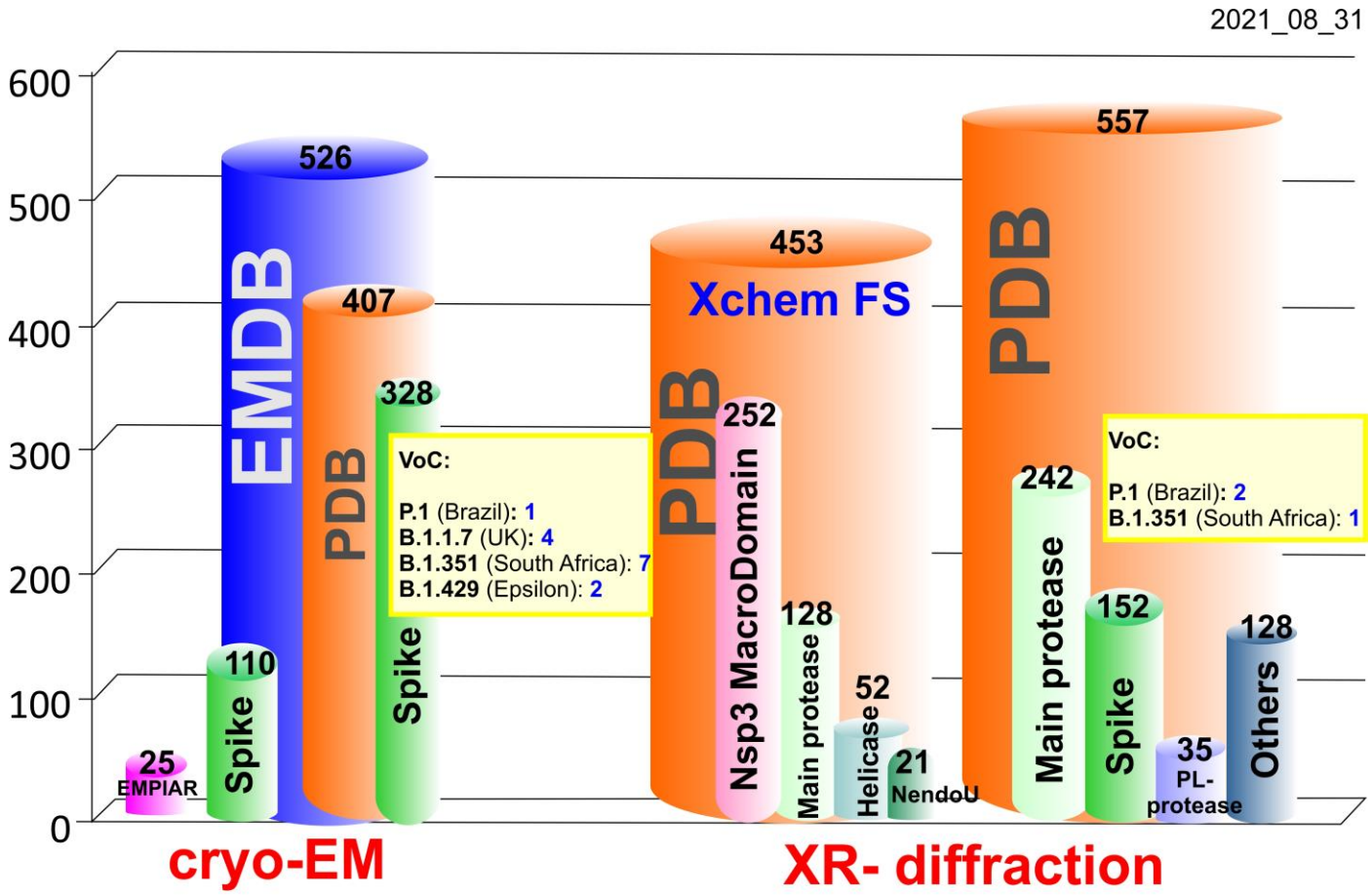


CSIC

Integrating Biology

Processing Center

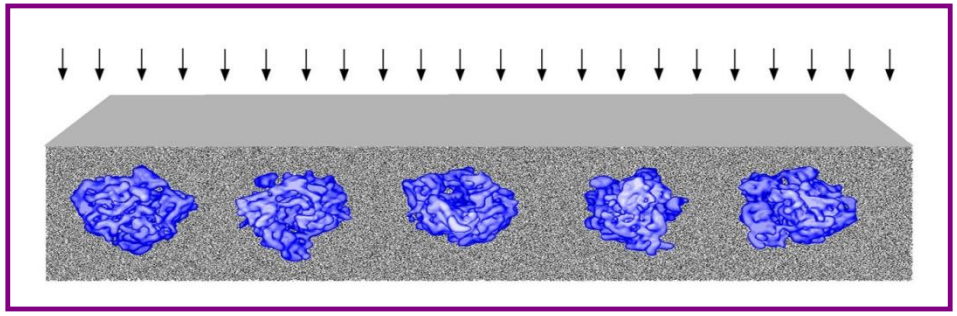
The role of cryo-EM: Statistics on SARS-CoV-2



Other experimental methods: NMR: 6 (PDB)
Neutron diffraction: 3 (PDB)

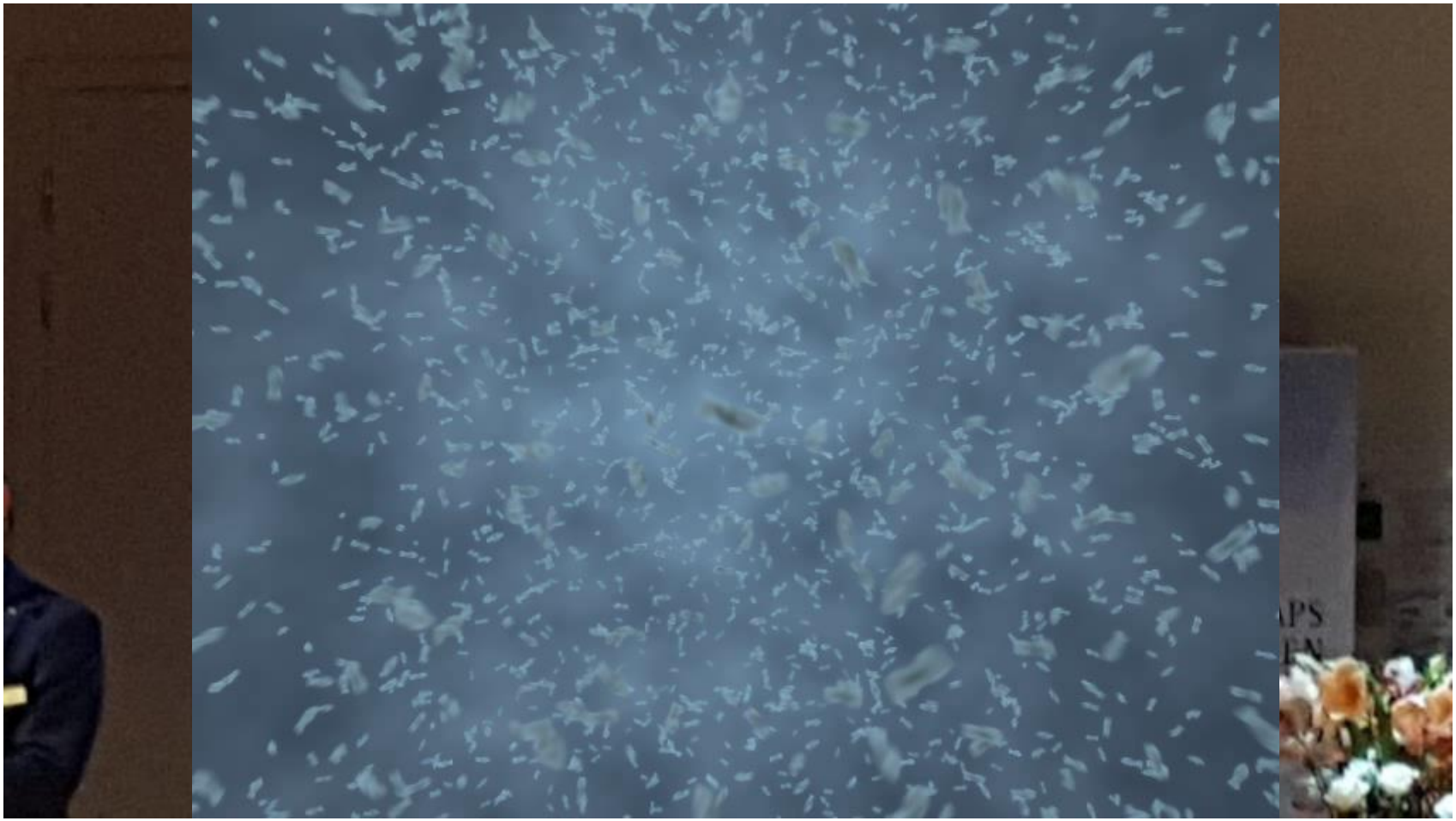
The cryo-EM SPA pledge

- **Trapped in ice, these molecules are close to their native state**
- **In 3D Electron Microscopy *individual molecules* are visualized**
- ***Therefore, flexible macromolecular machines in solution can be studied directly!***



CSIC instruct Integrating Biology BIO Unit instruct image Processing Center

The principles of cryoEM



CSIC

instruct
Integrating
Biology

instruct
image
Processing
Center

Starting from “first principles”….

Nat Methods. 2007 Jan;4(1):27-9. doi: 10.1038/nmeth992. Epub 2006 Dec 10.

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

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

“These probabilities calculations are based on the assumption that all Xi’s are projections of one of K underlying tree-dimensional objects ……..”

(So, continuous flexibility was not addressed)

CSIC

instruct Integrating Biology

instruct image Processing Center

And where we are nowadays!

Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials

D. Herreros[1], R.R. Lederman[2], J.M. Krieger[1], A. Jiménez-Moreno[1], M. Martínez[1], D. Myška[3], D. Strelak[1,4], J. Filipovic[3], C.O.S. Sorzano*[1], and J.M.Carazo*[1]

Herreros et al., submitted

IUCrJ
ISSN 2052-2525
CRYO|EM

Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials

David Herreros,[a]* Roy R. Lederman,[b] James Krieger,[c] Amaya Jiménez-Moreno,[a] Marta Martínez,[a] David Myška,[d] David Strelak,[a,e] Jiri Filipovic,[d] Ivet Bahar,[c] Jose Maria Carazo[a] and Carlos Oscar S. Sanchez[a]

Received 26 March 2021
Accepted 25 August 2021

CSIC

instruct Integrating Biology

erc European Research Council

instruct image Processing Center

Our aims:

- 1) Address continuous flexibility, not imposing discrete classes
  - But incorporating discrete classification information as well

- 2) Obtain reduced-dimensionality conformational landscapes, and analyze *interactively* (or by any sort of postprocessing) the flexibility they convey
  - We aim at subtle changes that may be masked by discrete classification
  - We also aim at simplifying the analysis process, minimizing decisions

- 3) Reduce map flexibility-induced blurring by "deforming back" images during the reconstruction process
  - This increases local resolution in flexible regions

How to study flexibility?

Classical Algorithms

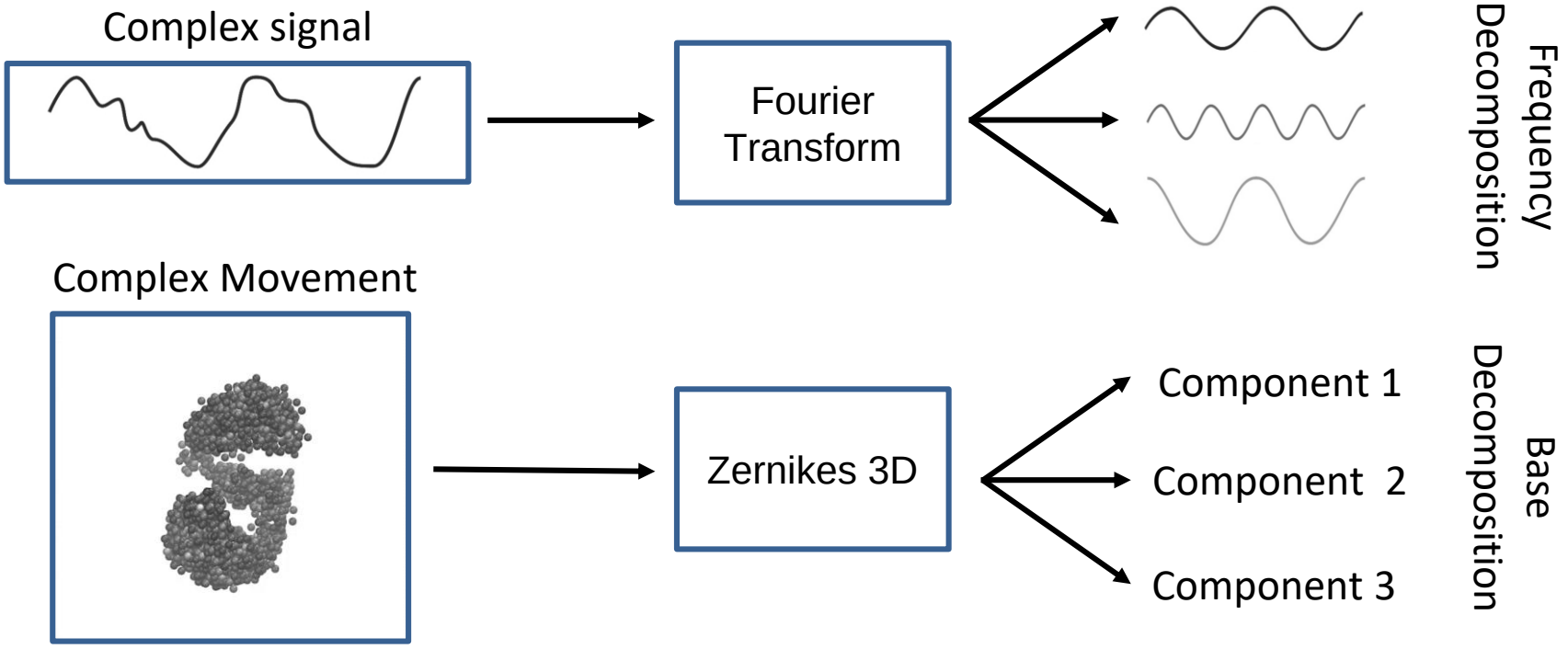
- Normal Mode Analysis
- Molecular Dynamics
- Principal Components Analysis

New Algorithms

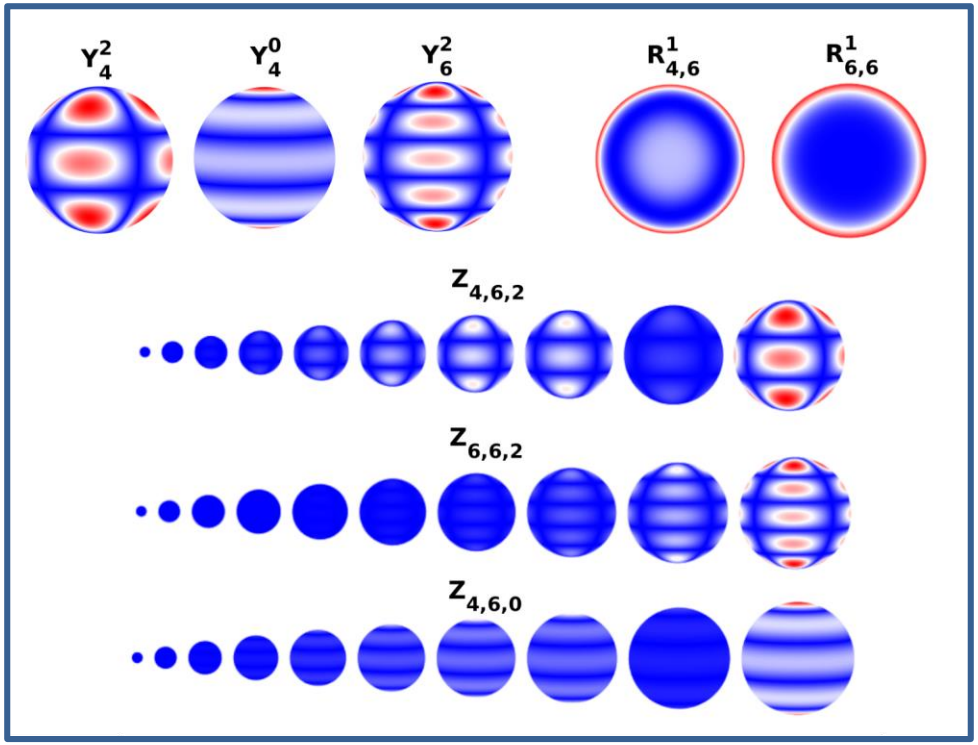
- REFINED Manifold embedding
- Normal Mode Analysis"Plus"
- CryoDRGN
- Gaussian Mixtures
- Relion (CCPEM-Relion)
- CrioSpark BioXriv on elastic deformations
- **Zernikes 3D**

"Hot topic!!!!!"

Zernikes3D: A completely different approach. Projecting onto a 3D base



Mathematical principles: How the basis look



IUCrJ
ISSN 2052-2525
CRYO|EM

Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials

David Herreros,[a]* Roy R. Lederman,[b] James Krieger,[c] Amaya Jiménez-Moreno,[a] Marta Martínez,[a] David Myška,[d] David Strelak,[a,e] Jiri Filipovic,[d] Ivet Bahar,[c] Jose Maria Carazo[a] and Carlos Oscar S. Sanchez[a]

Received 26 March 2021
Accepted 25 August 2021

Mathematical Basis: Expressing the deformation field ONLY

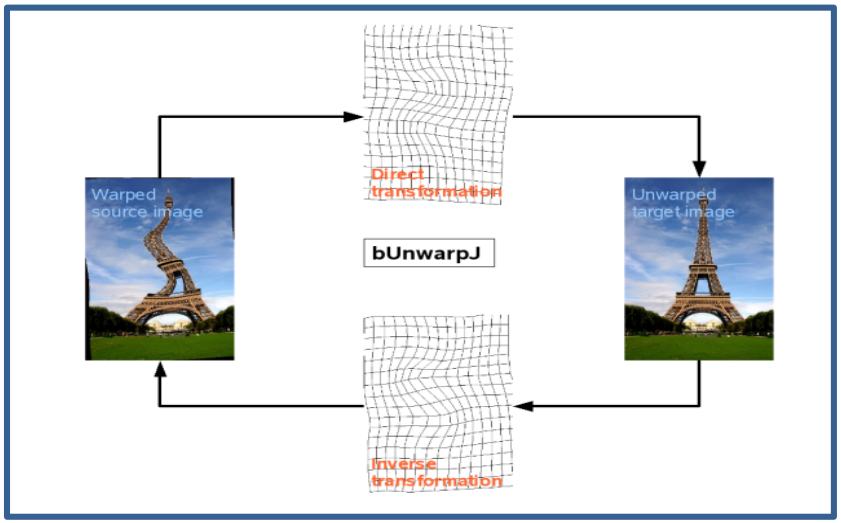
$$\min_{g_L} \int \left| V_1(\vec{r}) - V_2\left(\vec{r} + \vec{g}_L(\vec{r})\right) \right| d\vec{r}$$

$$g_L(\vec{r}) = \sum_l \sum_n \sum_m \vec{c}_{N_1,N_2,m} Z_{l,n,m}(r, \theta, \phi)$$

Purpose of the tool: to find a per-point displacement (deformation) that minimizes the distances between between two images

Parameters

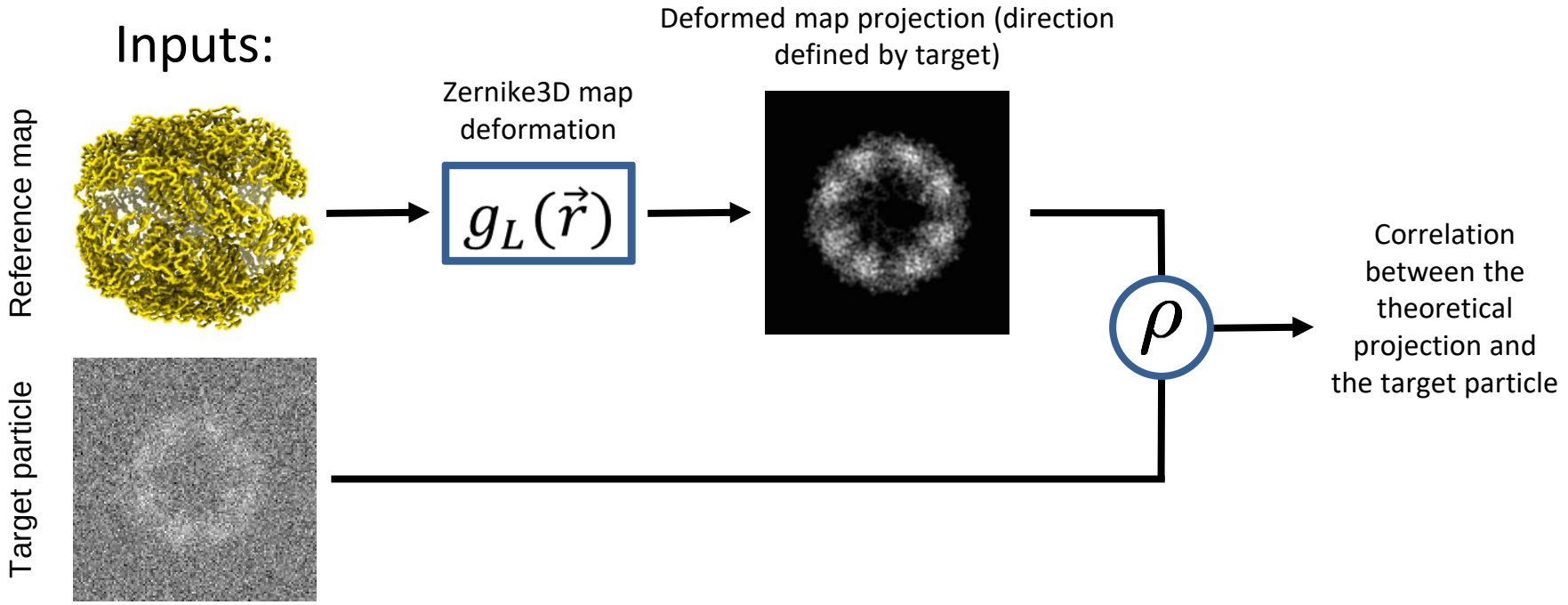
- L and N → Maximum degree of the basis (movement frequency)
- R_{max} → Maximum radius of the sphere (enclosing the molecule)



CSIC

instruct Integrating Biology

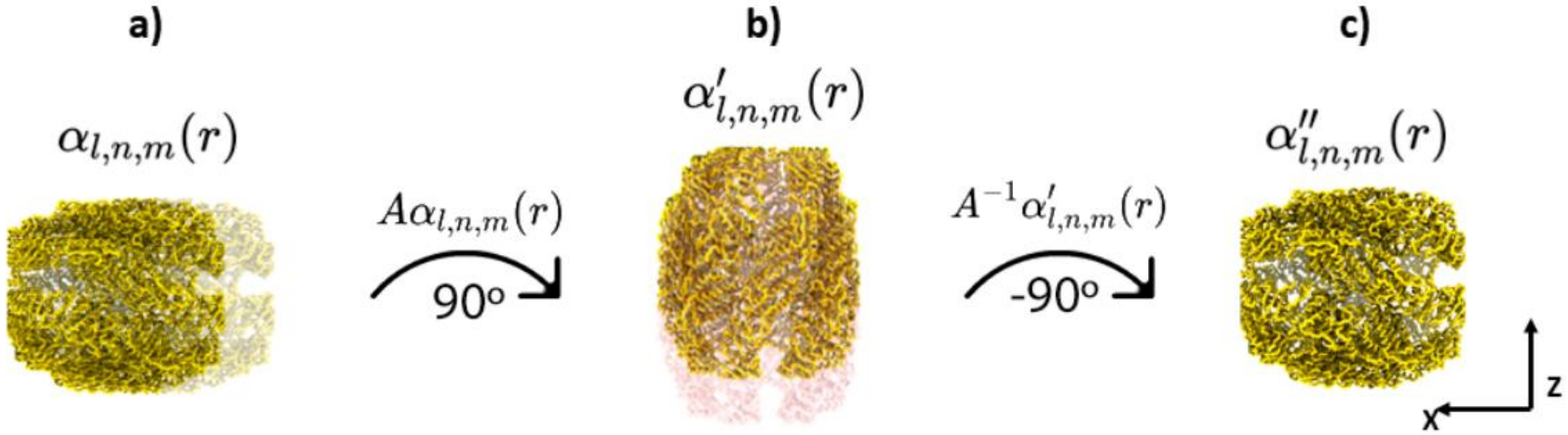
Starting from 2D cryoEM images, model the 3D deformation field



Purpose of the tool: to find a per-point displacement (deformation) that minimizes the distances between between **two images (+ volume)**

CSIC instruct Integrating Biology BIO instruct image Processing Center

Formally addressing the different information content along projection directions



$$\overrightarrow{g_l}(r'_i) = \overrightarrow{g_l}(Rr_i) = R\overrightarrow{g_l}(r_i)$$

$$\overrightarrow{g_l}(r_i) = R^{-1}\overrightarrow{g_l}(Rr_i) = \sum_l \sum_n \sum_m R^{-1} c_{l,n,m} Z_{l,n,m}(Rr_i)$$

IUCrJ
ISSN 2052-2525
CRYO|EM

Received 26 March 2021
Accepted 25 August 2021

Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials

David Herreros,[a]* Roy R. Lederman,[b] James Krieger,[c] Amaya Jiménez-Moreno,[a] Marta Martínez,[a] David Myška,[d] David Strelak,[a,e] Jiri Filipovic,[d] Ivet Bahar,[c] Jose Maria Carazo[a] and Carlos Oscar S. Sanchez[a]

CSIC

instruct Integrating Biology

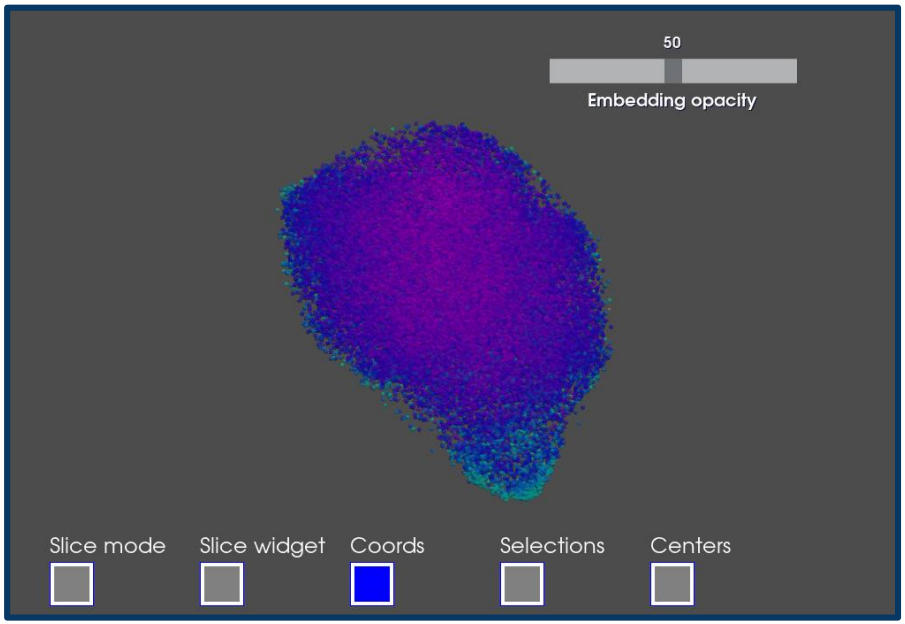
instruct image Processing Center

Zernike3D coefficient embedding

Low-dimensional representation of the coefficient space defined by a set of CryoEM particles

EMPIAR-10080

- Representative dataset of the Pf80S rotation state
- Two well differentiated areas found in the embedding
  - Unrotated (reference) state in purple
  - Rotated state in green



CSIC

instruct Integrating Biology

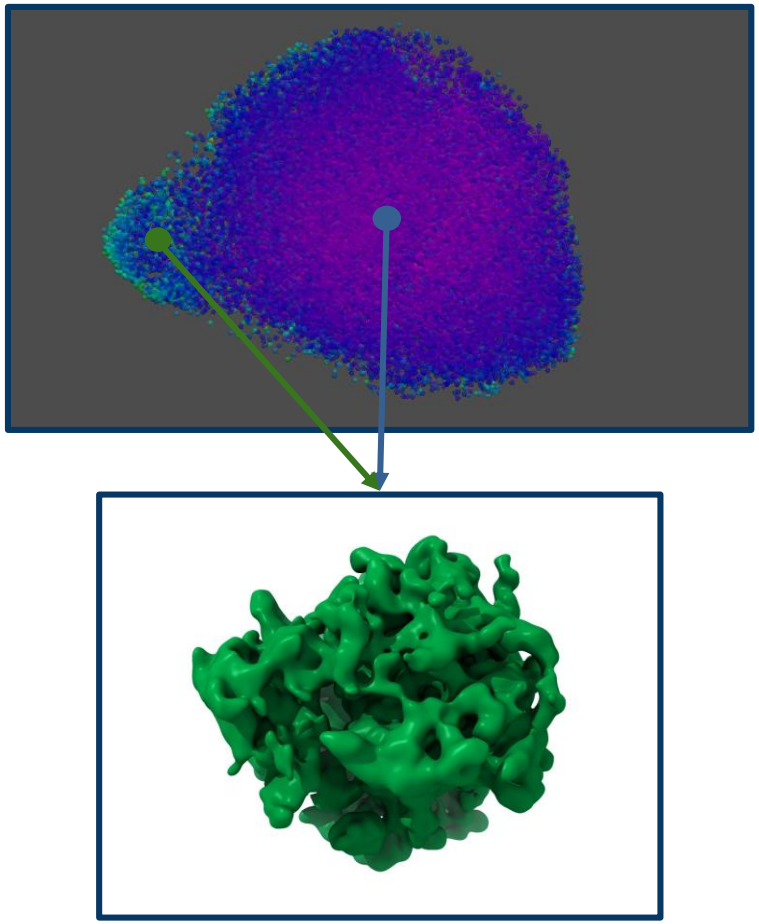
instruct image Processing Center

Zernike3D coefficient embedding

Low-dimensional representation of the coefficient space defined by a set of CryoEM particles

EMPIAR-10080

- Representative dataset of the Pf80S rotation state
- Two well differentiated areas found in the embedding
  - Unrotated (reference) state in purple
  - Rotated state in green



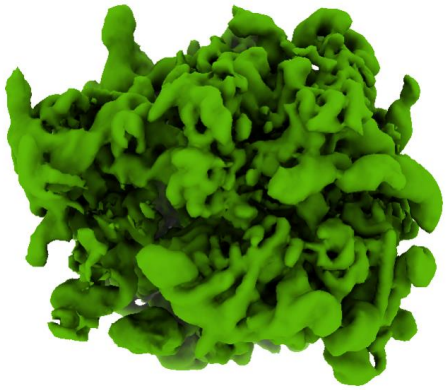
CSIC

instruct Integrating Biology

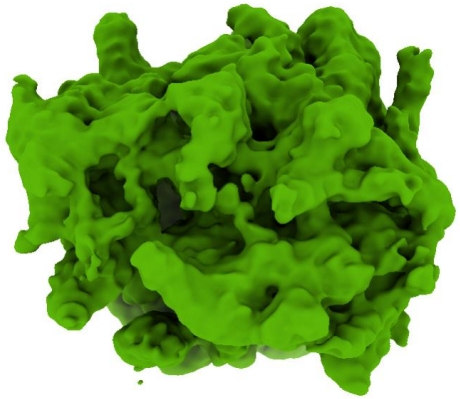
instruct image Processing Center

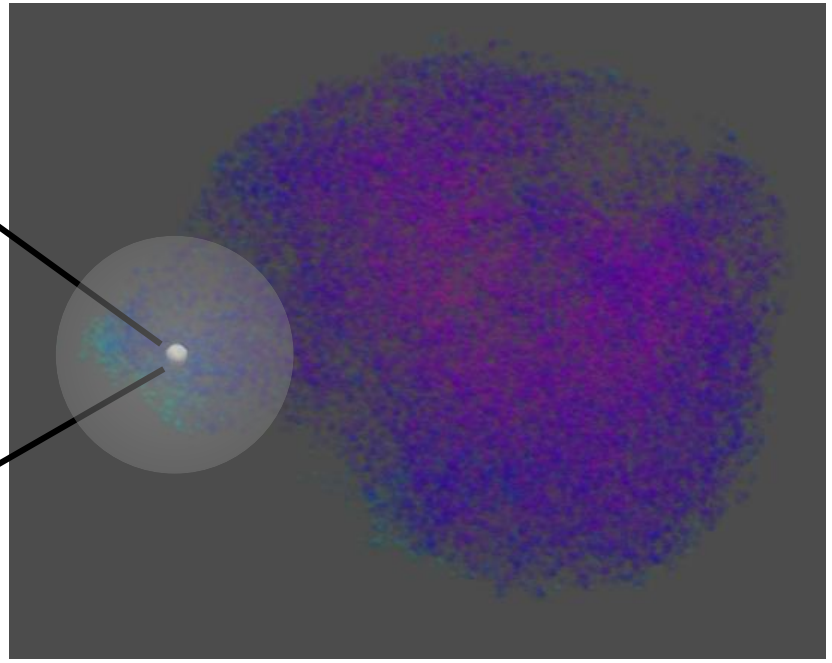
Zernikes3D representations are very accurate

Zernike3D



Refinement





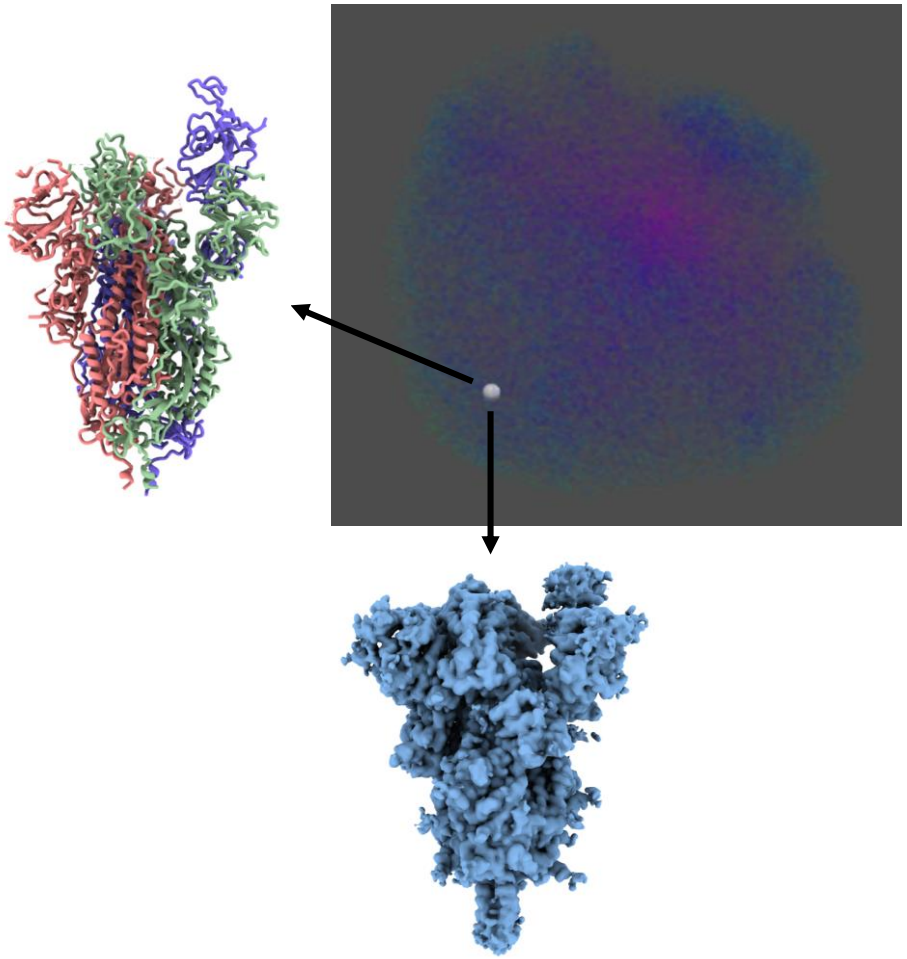
CSIC

instruct Integrating Biology

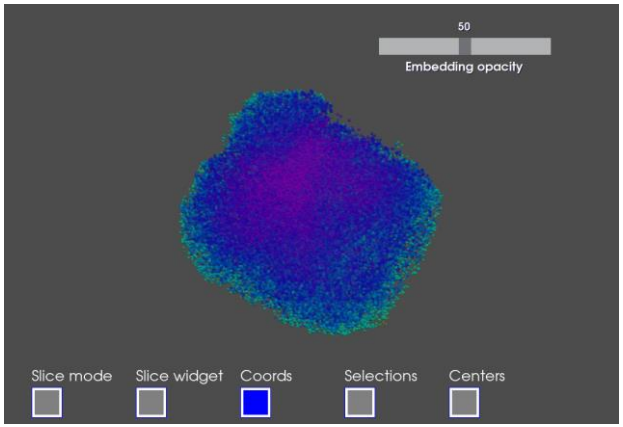
instruct image Processing Center

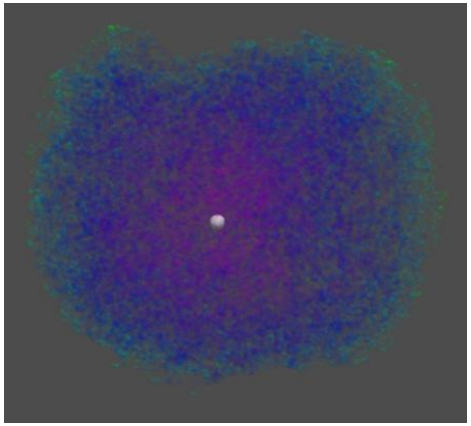
Working with maps and models

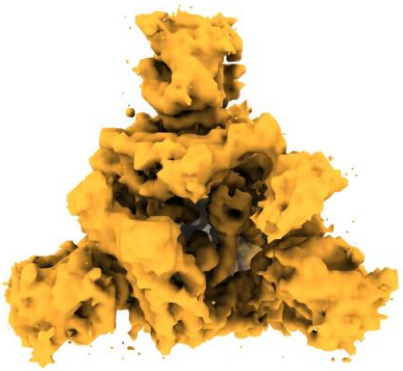
Zernike3D coefficients can be applied simultaneously to CryoEM maps and structures



Focused heterogeneity analysis







Flexibility correction: ZART reconstruction

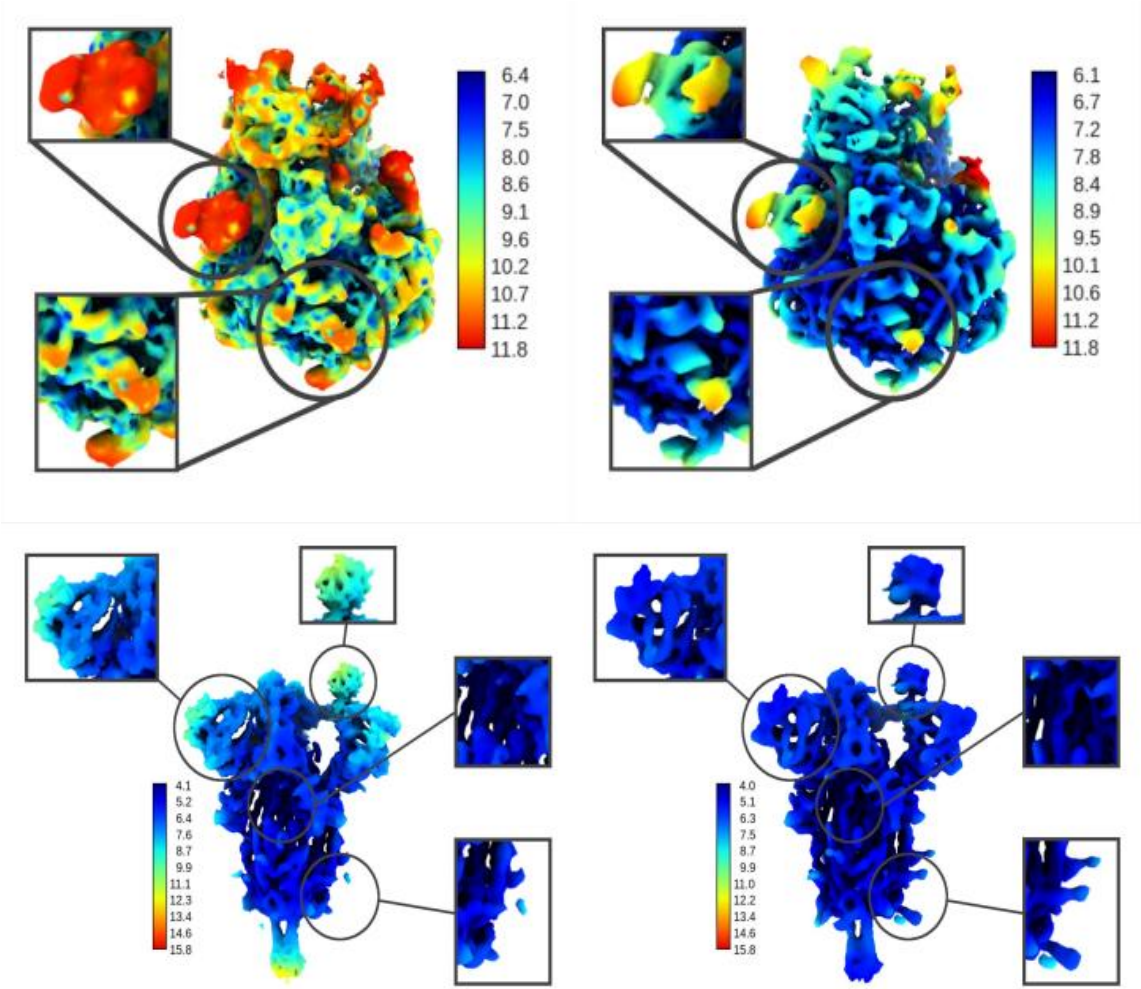
- We have develop a new ART based reconstruction algorithm that incorporates the per-particle deformation field information of the Zernikes3D coefficients in the reconstruction process
  - Improves resolvability of moving areas in the maps by decreasing the motion blurring

ART $\longrightarrow$ $$V(r)^{k+1} = V(r)^{k} + \lambda(P_H(V(r)) - I(P_H(r)))$$

$\downarrow$

ZART $\longrightarrow$ $$V(r)^{k+1} = V(r)^{k} + \lambda(P_H(V(r + g_L)) - I(P_H(r)))$$

“Undoing” flexibility: ZART reconstruction



CSIC

instruct Integrating Biology

instruct image Processing Center

In a nutshell

- 1) We do obtain continuous macromolecules deformation fields
- 2) We get very informative low dimensional representations of these deformation fields
- 3) We can “Undo” deformations and increase local resolution
- 4) We are opening a new way to interface with Molecular Dynamics

Acknowledgments

And the economical support received from:

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

erc European Research Council Established by the European Commission

Instituto de Salud Carlos III

CSIC Consejo Superior de Investigaciones Científicas

instruct ERIC