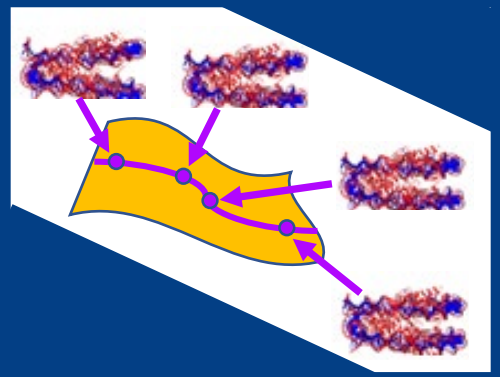
Hybrid methods for analyzing conformational variability in cryo-EM and cryo-ET data

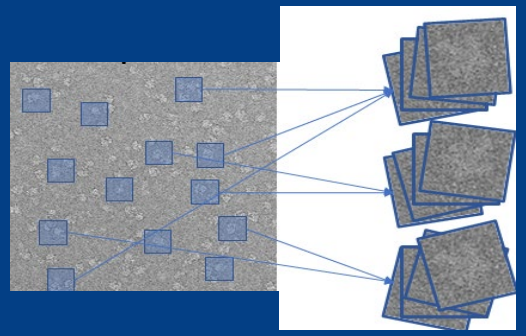
Slavica Jonic

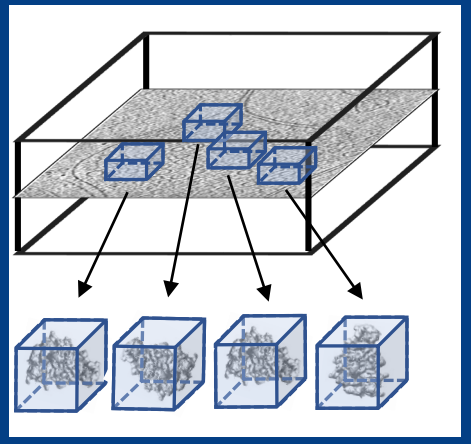
Hybrid methods for analyzing conformational variability in cryo-EM and cryo-ET data

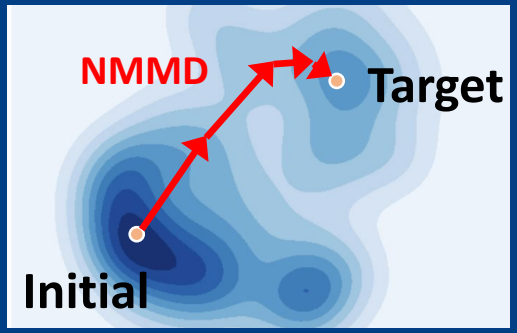
Slavica Jonic

CNRS & Sorbonne University, Paris, France
IMPMC – UMR 7590





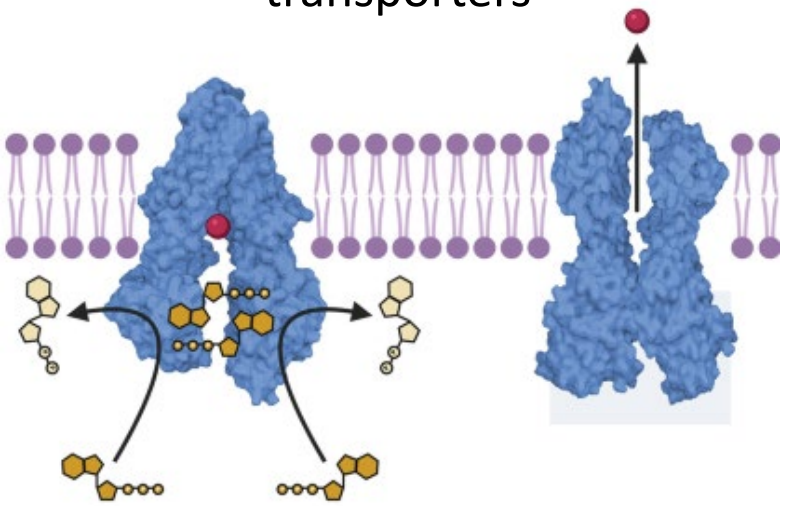




CNRS Thematic School AlgoSB (Algorithms in Structural Bioinformatics), Cargese, France, November 20-24, 2023

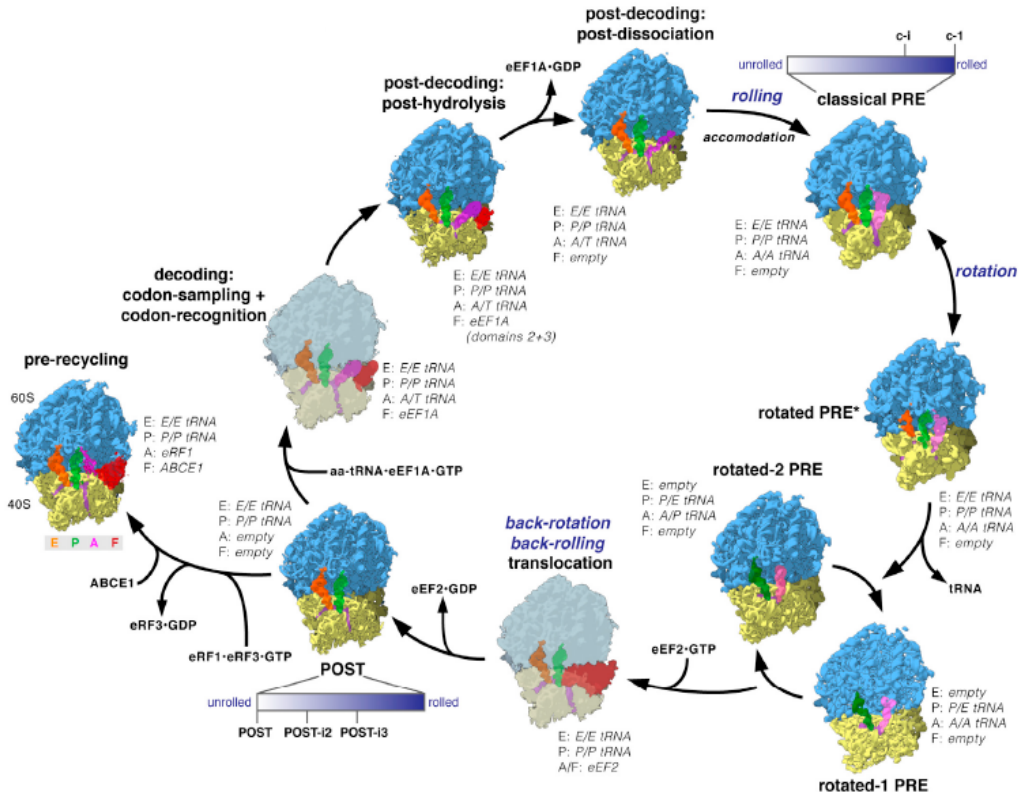
Conformational dynamics of biomolecules:
Link to biological function

Dynamics of membrane transporters



Koehn, *ADME Encyclopedia*, 2021

Ribosomal dynamics



Behrmann et al., *Cell 2015*

Approaches to study conformational dynamics

Experimental (cryo-EM/ET), theoretical (simulation), hybrid (combination)

PART 1 :

INTRODUCTION

PART 1: INTRODUCTION

- Cryo-EM (Single Particle Analysis) *vs.* Cryo-ET
- Terminology
- Conformational heterogeneity in cryo-EM and cryo-ET data
- Conformational landscapes and Free-Energy landscapes
- A brief review on methods for extracting continuous conformational landscapes from data
  - Based on machine learning
  - Based on molecular conformational dynamics simulation (hybrid methods)

Cryo-EM (Single Particle Analysis) *vs.* Cryo-ET

Principles, Advantages, Limitations

Cryo-EM Single Particle Analysis (SPA) – Principles

Sample preparation and imaging

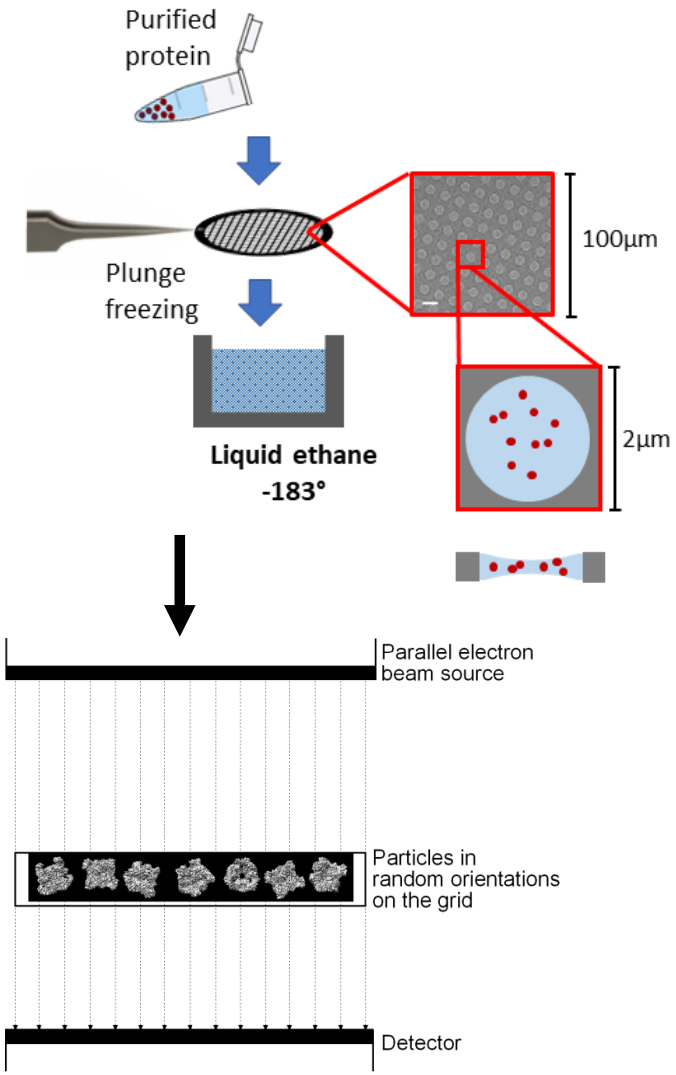
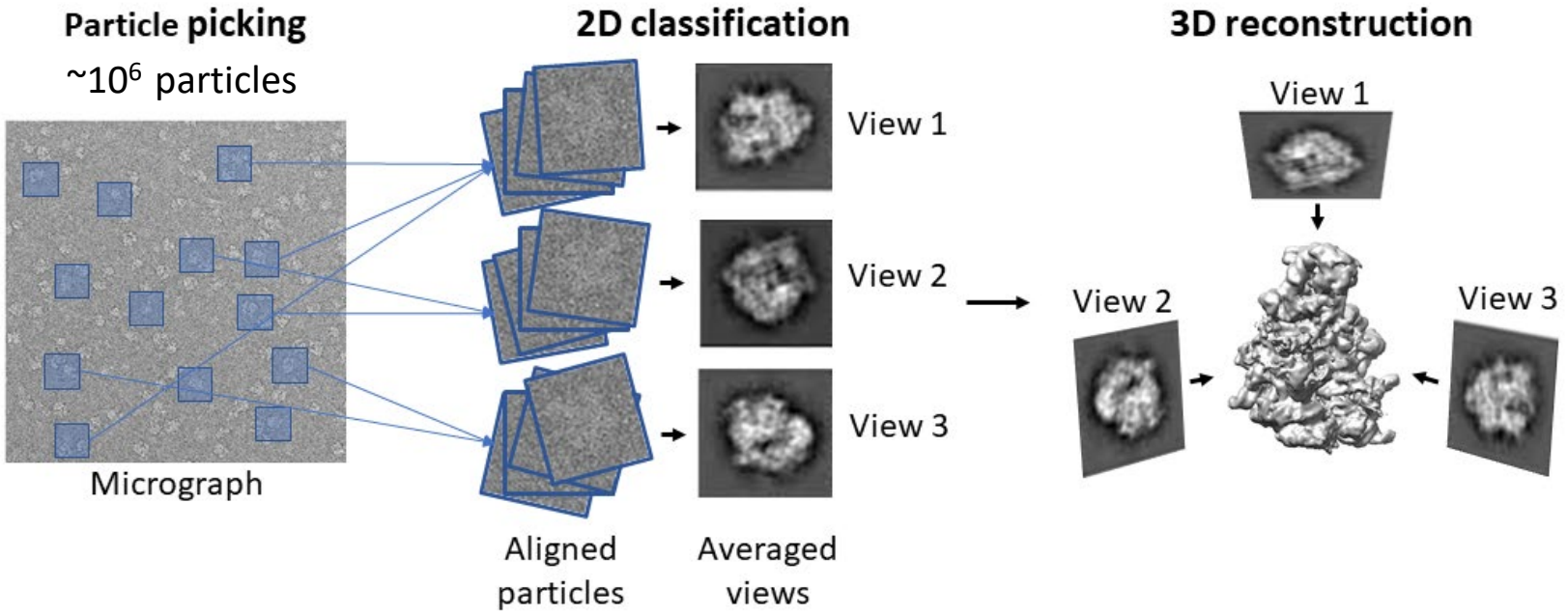


Image processing



Cryo-Electron Tomography (cryo-ET) - Principles

Sample preparation and imaging

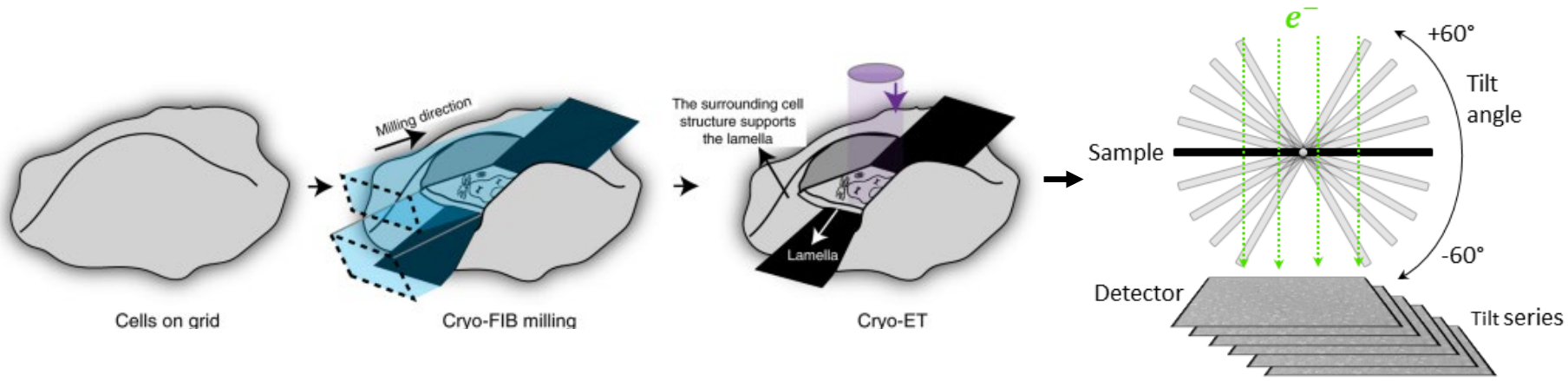
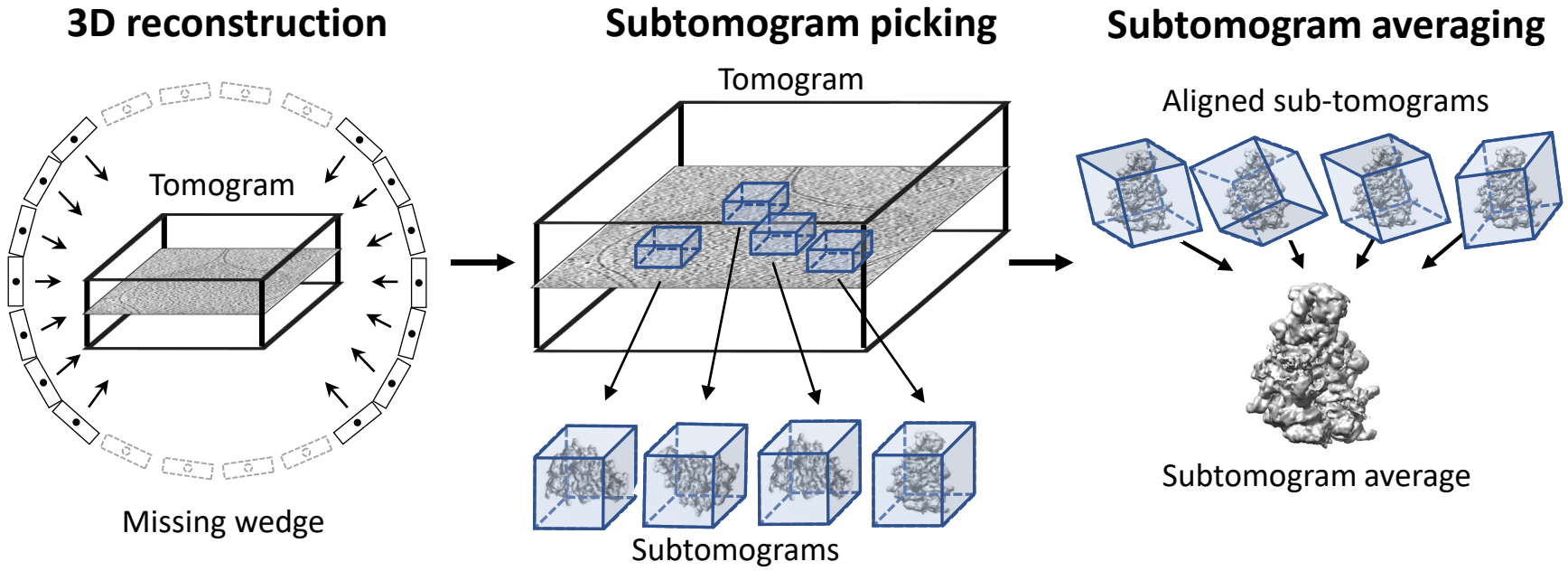
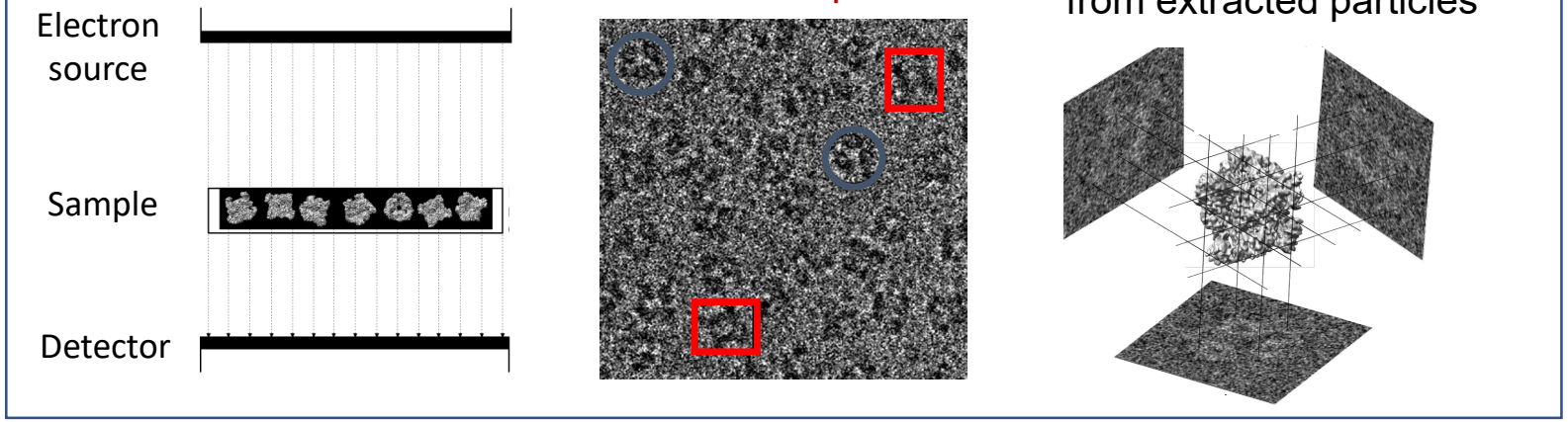
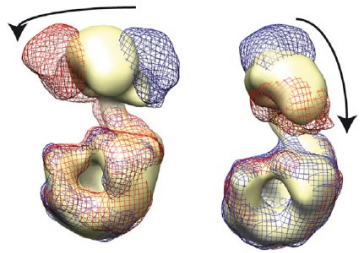


Image processing



Cryo-EM SPA advantages and limitations



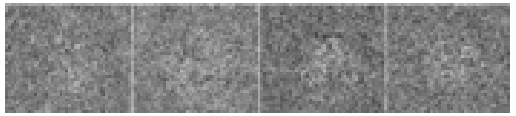


DNA replication

DNA Pol α - B complex

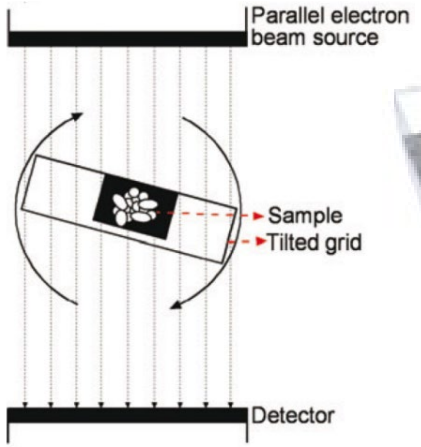
Jin et al., Structure 2014

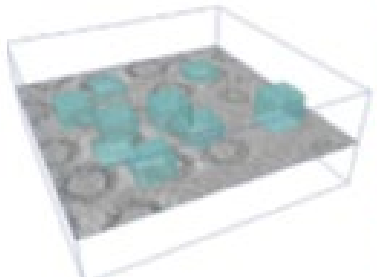
- ❑ No sample crystallization required
- ❑ Studying purified complexes
- ❑ Allows studying flexible complexes (multiple conformations can be obtained from the same sample)
  - ❑ Heterogeneity of orientations
  - ❑ Heterogeneity of conformations (continuous: uncountable states)
- ❑ Noise due to a low electron dose, used to minimize the sample damage
- ❑ Low SNR but still higher than with cryo-ET data



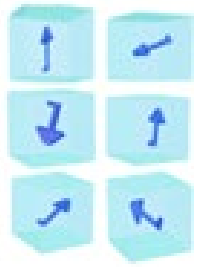
Different orientations ?
Different conformations ?

Cryo-ET advantages and limitations

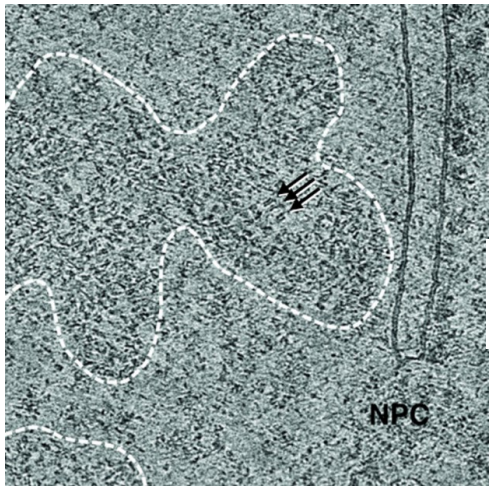


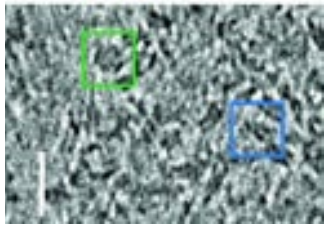


Reconstructed tomogram

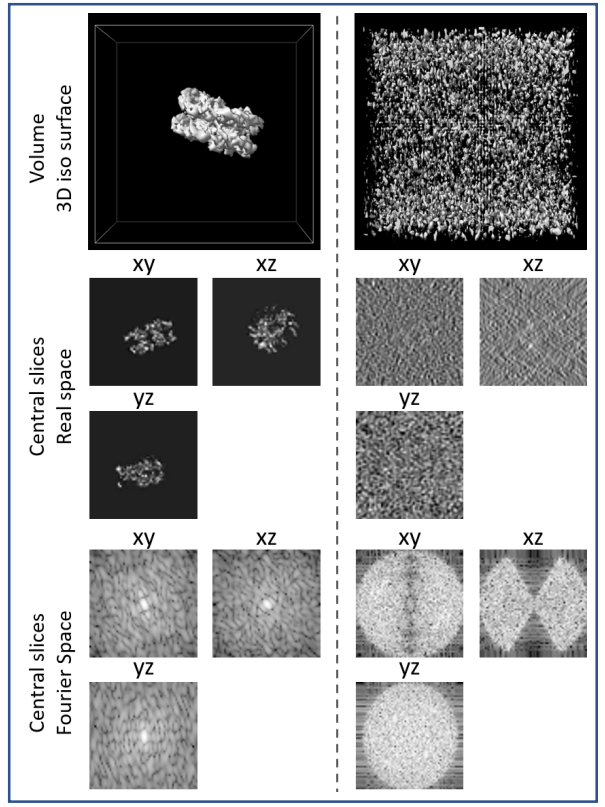


Extracted subtomograms (“particles”)

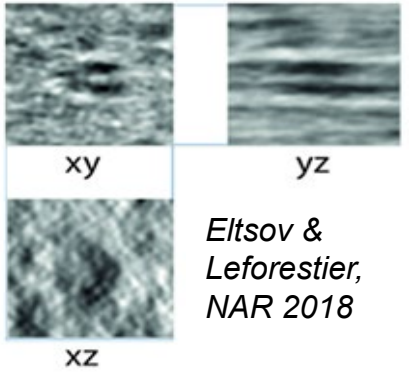




Eltsov & Leforestier NAR 2018



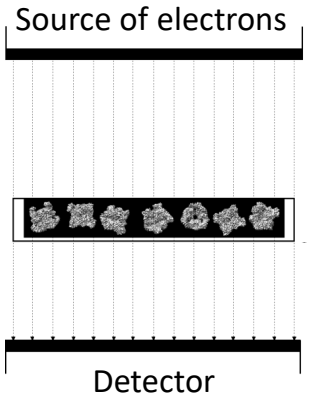
- ❑ Allows *in situ* studies of dynamic molecular complexes
- ❑ But, crowded environment
- ❑ Low SNR (electron dose split over multiple tilts)
- ❑ “Missing Wedge” problem due to the limited range of the tilt angle (induces deformations of the reconstructed object in real space)
- ❑ Heterogeneity of conformations and orientations of molecules

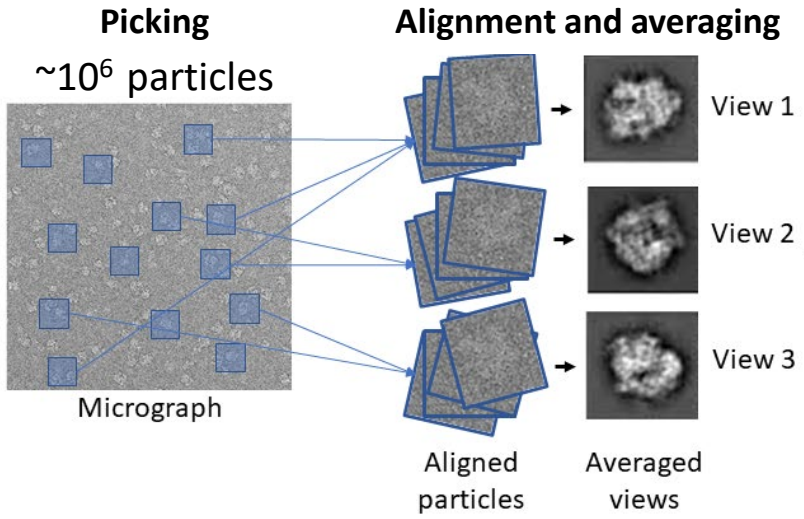


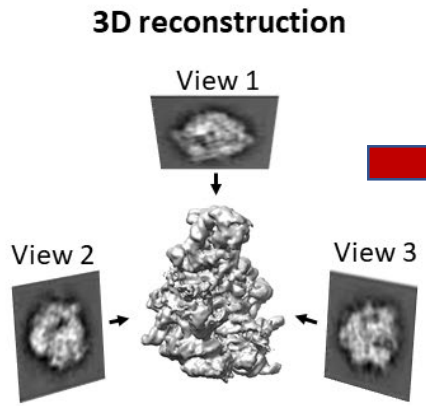
Eltsov & Leforestier, NAR 2018

Combined heterogeneity of conformations, orientations, and translations in highly noisy data (particle images or subtomograms)

Single Particle Analysis (SPA)

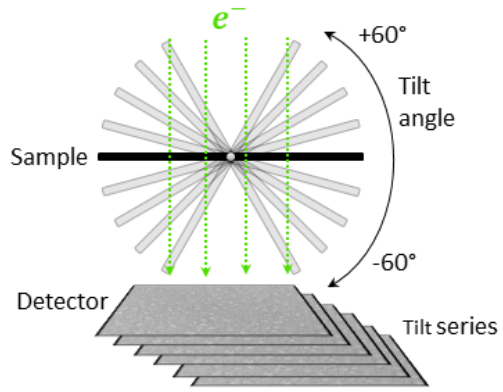


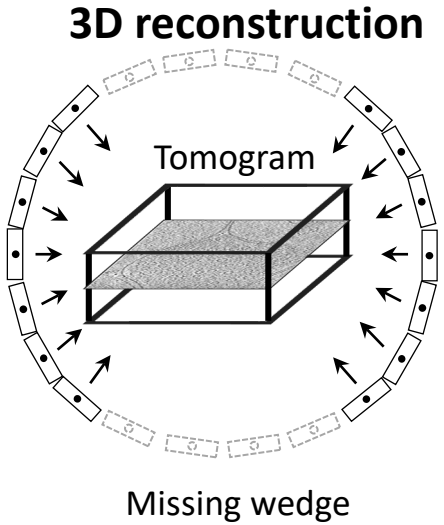


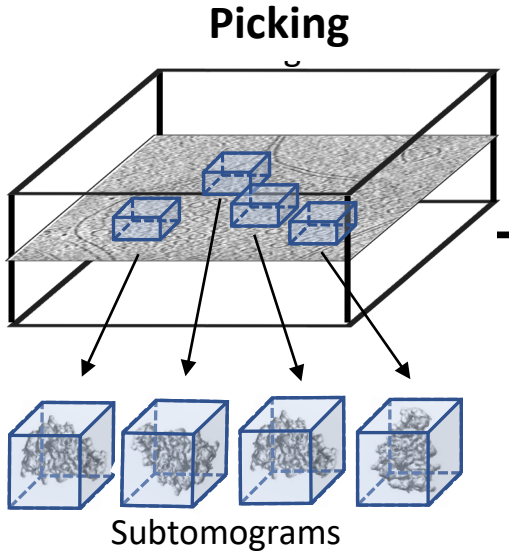


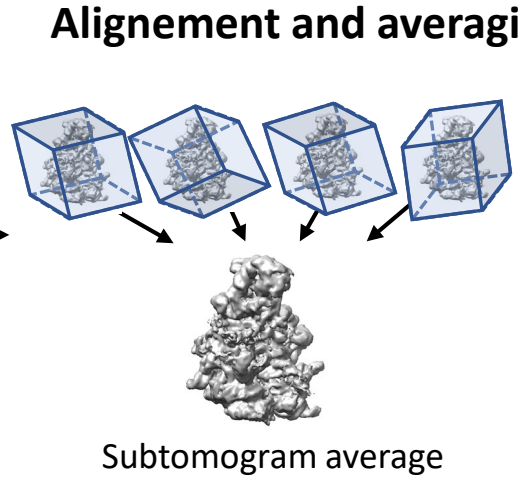
Heterogeneity analysis by discrete classification

Electron Tomography (ET)









Discrete-classification approaches

Principle

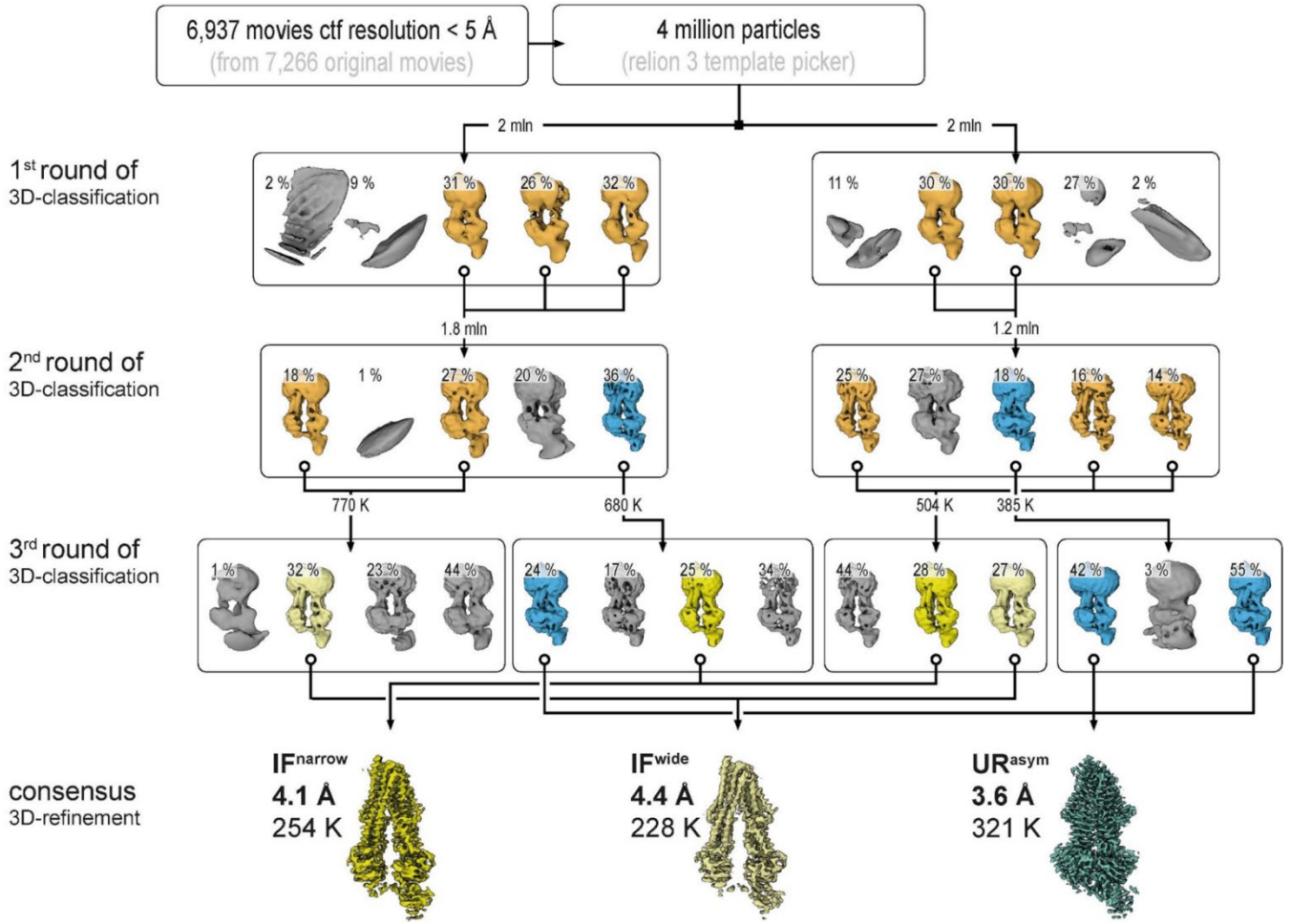
- Aim at sorting particles into a given number of homogeneous classes
- Maximum likelihood-based classification (available in RELION, cryoSPARC, cisTEM, …)

Advantages

- Achieve high resolution of a small number of conformational states

Limitations

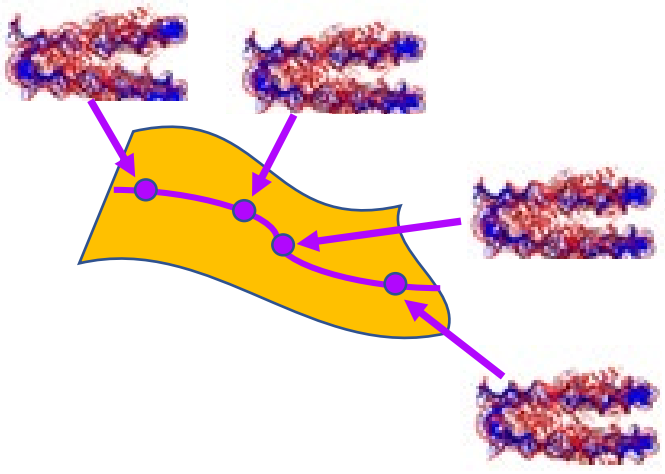
- Discard many particles (low-resolution classes, heterogeneous, continuous heterogeneity)
- Rare states cannot be elucidated

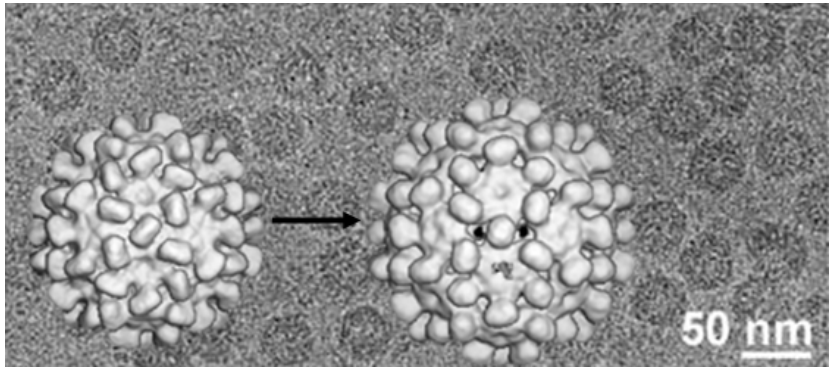


Hoffman et al. *Nature* 2019

Terminology

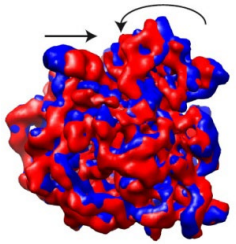
- Conformational changes/variability (related to biological function)
- Conformational heterogeneity (related to data)
- Discrete *vs.* Continuous



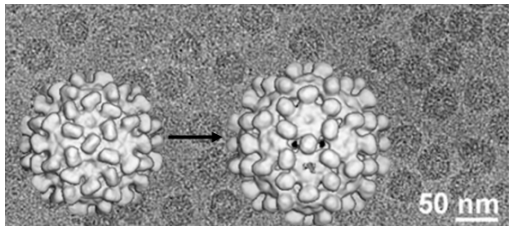


Discrete *vs.* Continuous conformational changes/variability (Function)

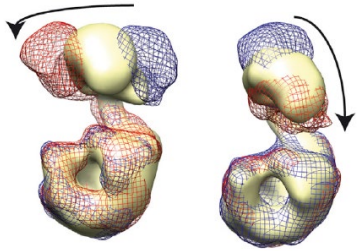
Continuous



Protein synthesis
(70S ribosome)

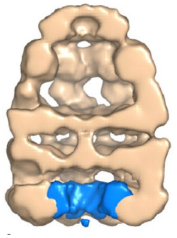


Virus maturation
(Tomato Bushy Stunt Virus)



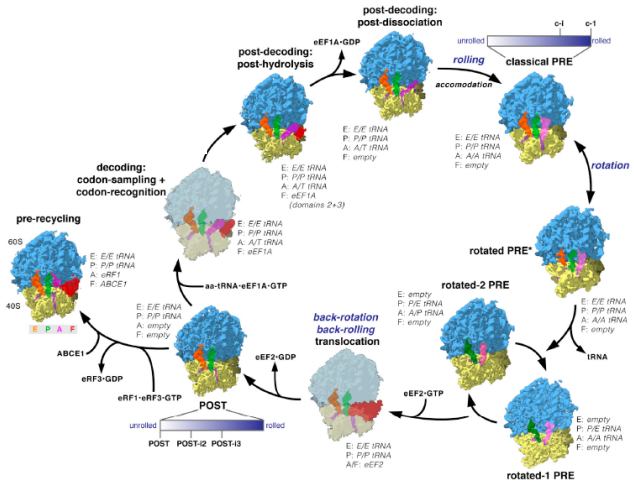
DNA replication
(DNA Pol α - B complex)

Discrete



Substrate binding or unbinding
(GroEL–GroES vs.
GroEL–GroES–rhodanese)
Elad et al, JSB 2008

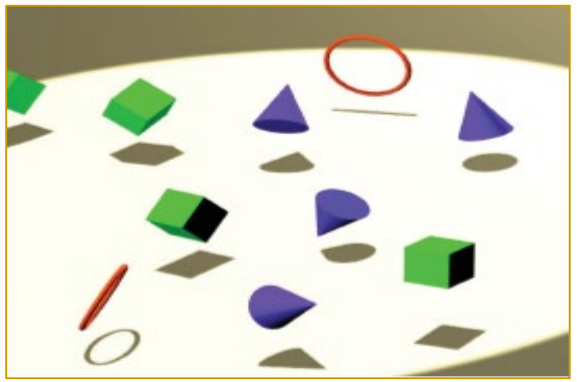
Combination of discrete and continuous

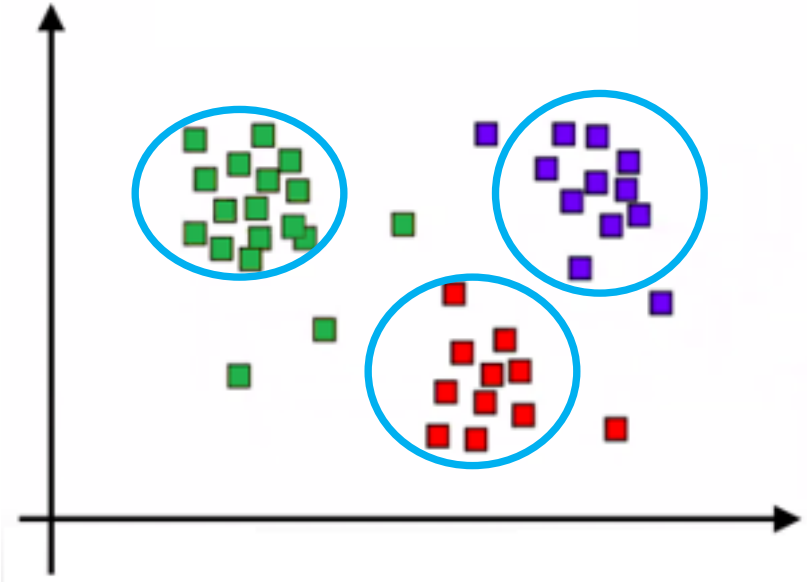


80S ribosome elongation cycle
Behrmann et al., Cell 2015

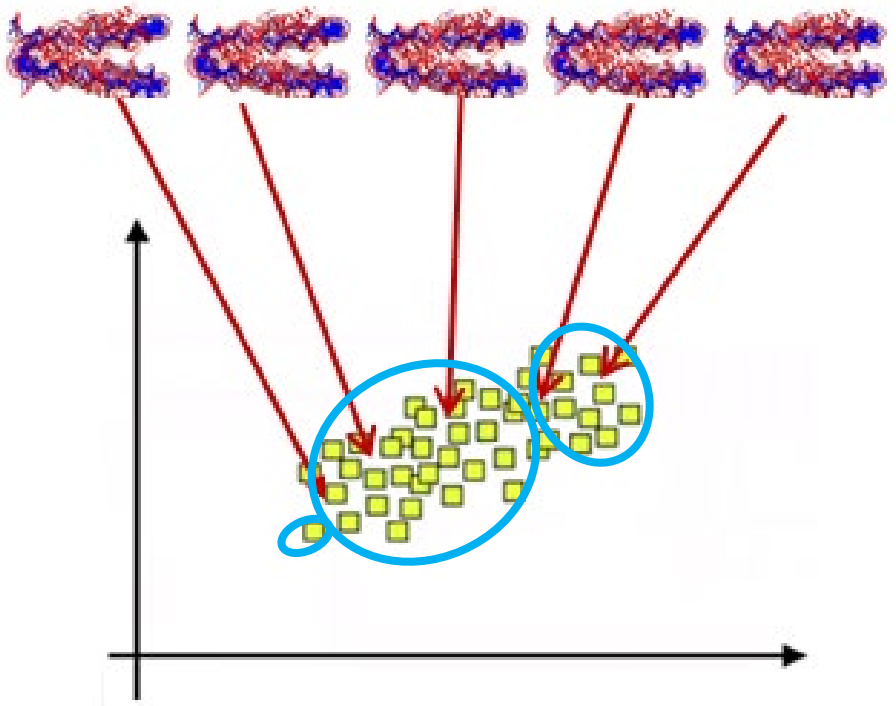
Discrete *vs.* Continuous conformational heterogeneity (Data)

Discrete





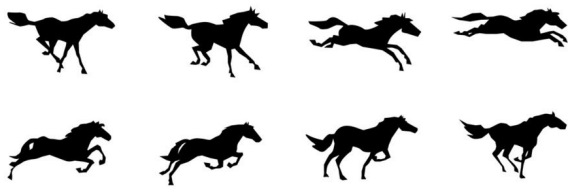
Continuous



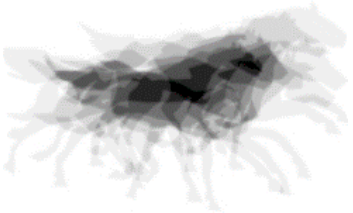
Conformational heterogeneity in cryo-EM and cryo-ET data

Obstacle and Opportunity

Heterogeneity: Obstacle to high resolution reconstruction, but Opportunity to get multiple conformations from the same sample

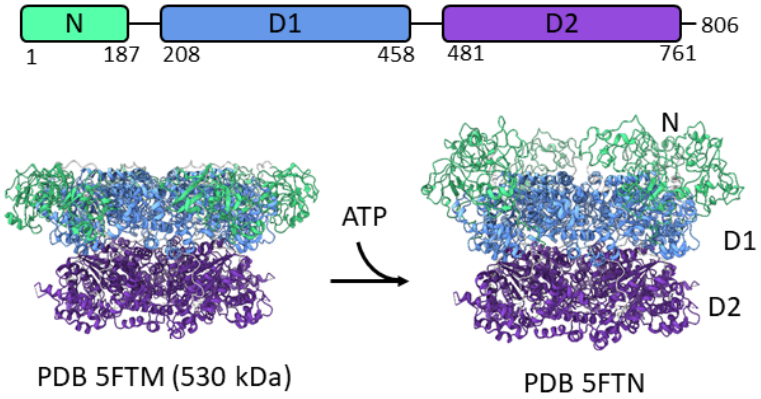


Averaging different conformations



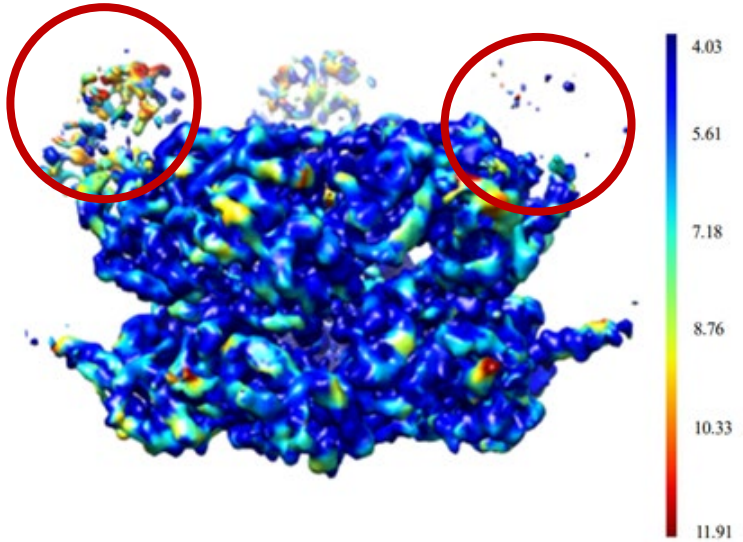
- Low resolution
- Losing flexible domains
- Losing variability information

AAA ATPase p97



Hexamer formed of domains N, D1, and D2

Cryo-EM of p97 disease-associated mutant (R155P)

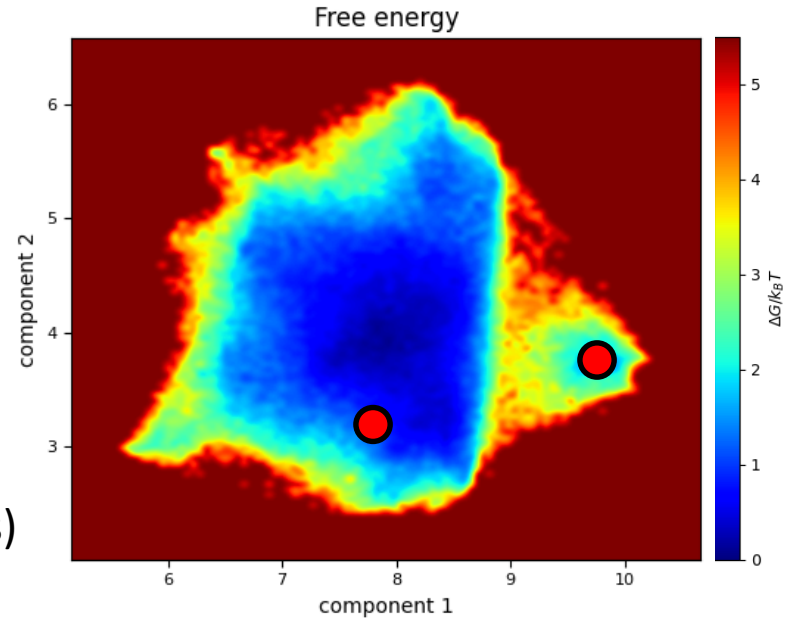


S. Valimehr & I. Rouiller
The University of Melbourne

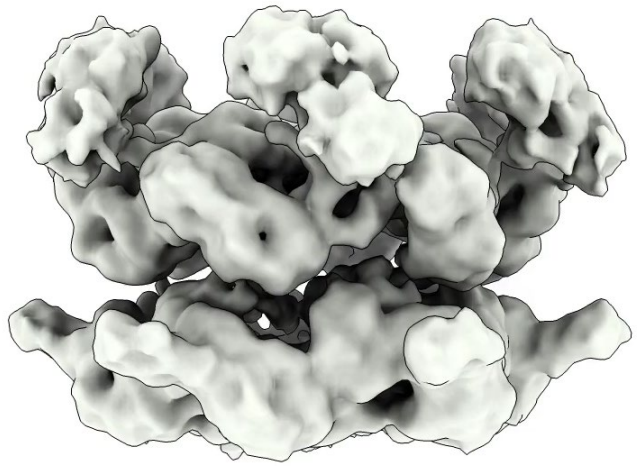
Continuous conformational heterogeneity SPA with MDSPACE

MDSPACE analysis of p97 dataset of 274,640 particles

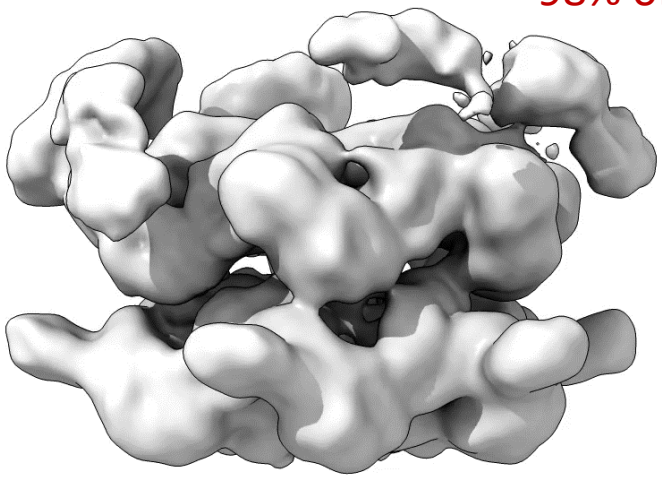
- Initial model: Cα Gō model of PDB:5FTN (N domains up)
- 2 iterations of MDSPACE
  - Iter 1: 5,000 particles, 5 NMs
  - Iter 2: 274,640 particles, 10 PCs
- 50 ps NMMD simulations (time step : 1 fs)

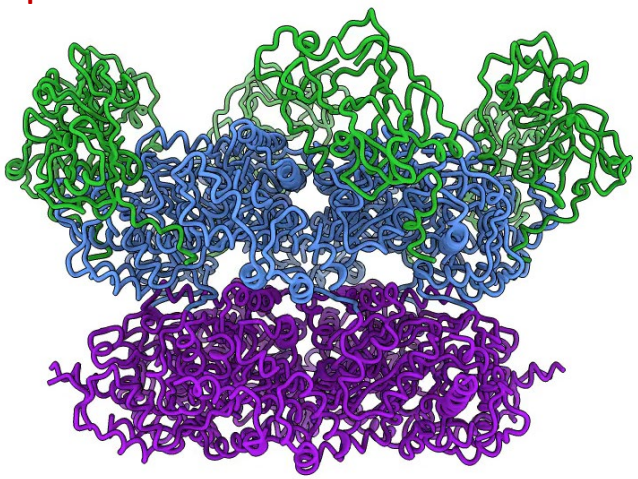


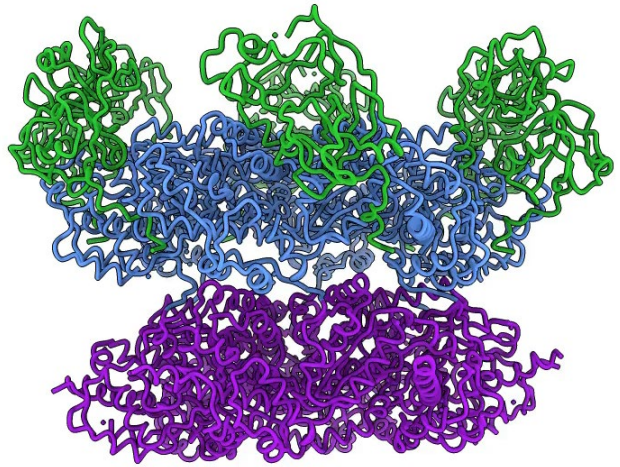
2% of the particles



98% of the particles





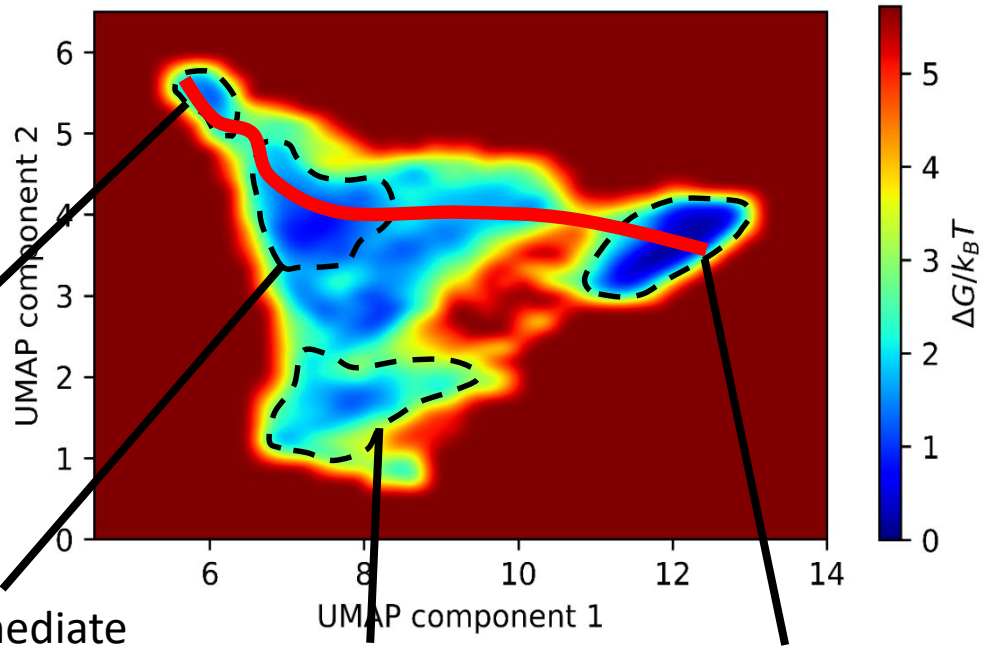


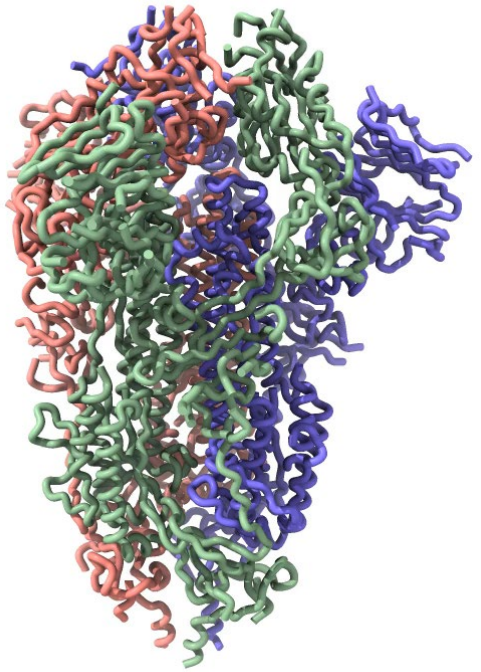
Manuscript in preparation (PhD thesis of Rémi Vuillemot 2023, Collab. Rouiller team, The University of Melbourne)

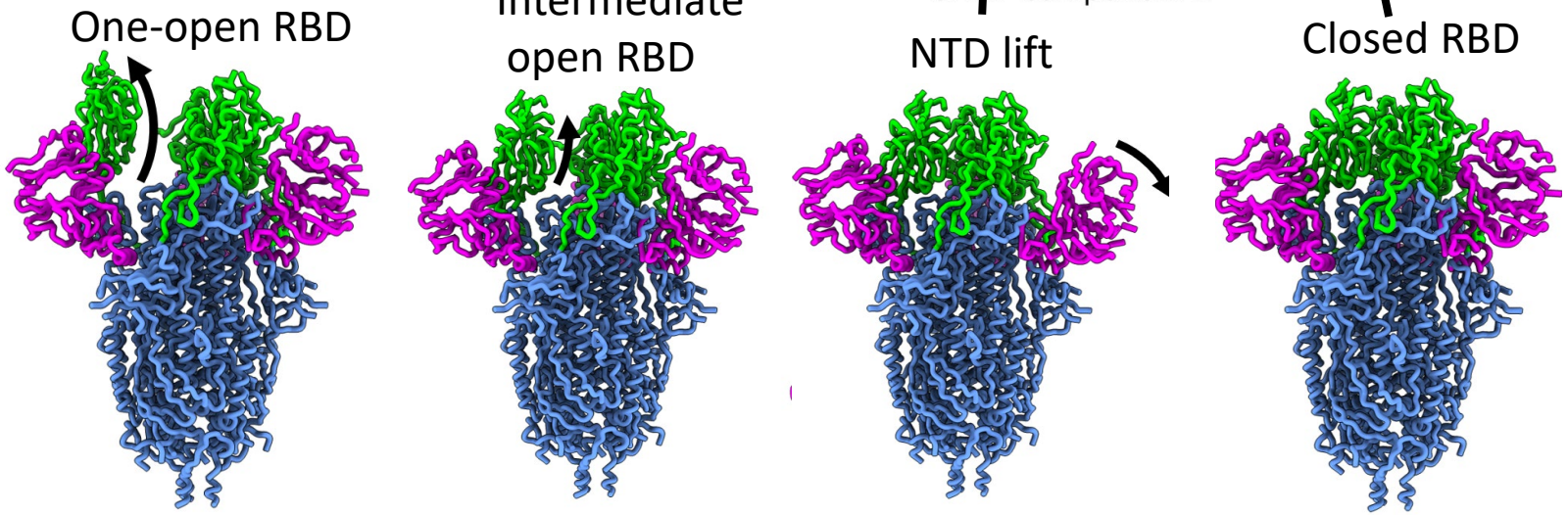
Conformational heterogeneity subtomogram analysis with MDTOMO

MDTOMO analysis of SARS-CoV2 spike dataset of 20,080 subtomograms

- Initial model: Cα Gō model of PDB:6VXX (closed RBDs)
- 1 iter of MDTOMO
- 100 ps NMMD simulations
- Force constant: 7000 kcal/mol





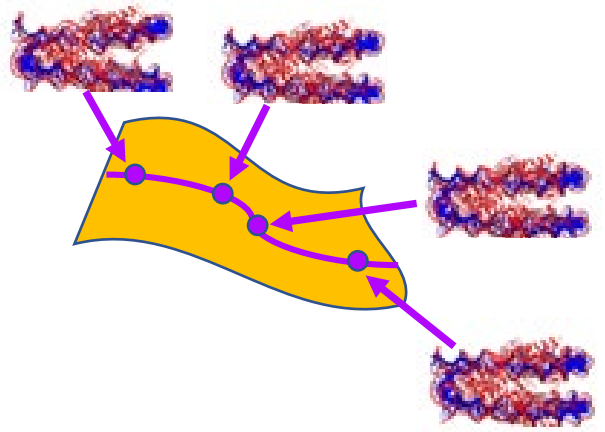


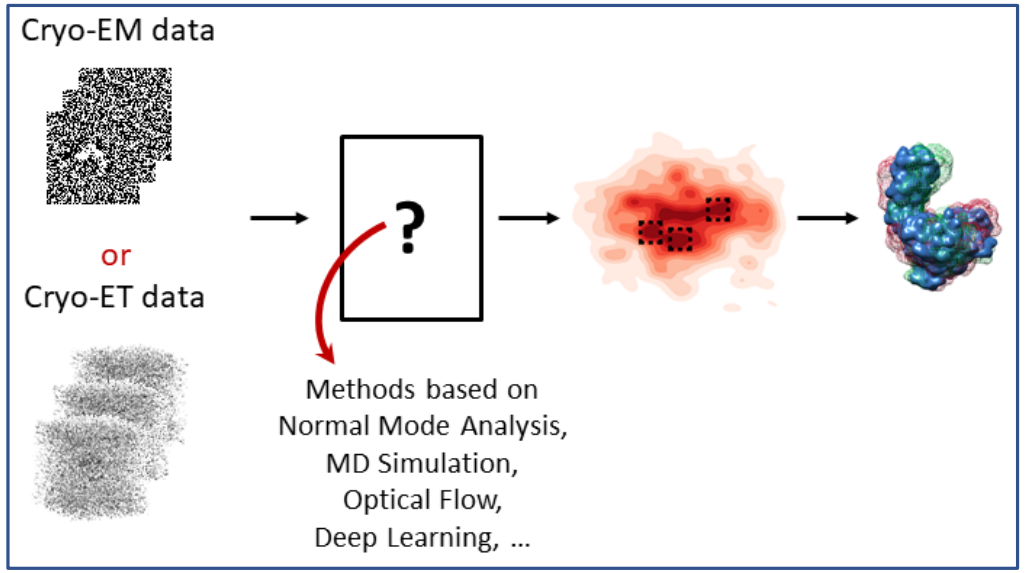
Vuillemot et al. *Sci Rep* 2023, *data provided by B. Turonova, Max Planck Inst. Biophys, Germany*

Conformational landscapes and Free-Energy landscapes

Terminology and Examples

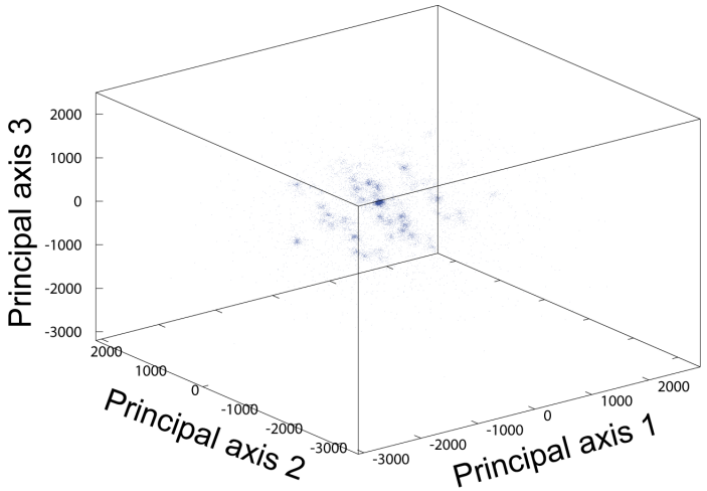
Conformational space (also called landscape or manifold)



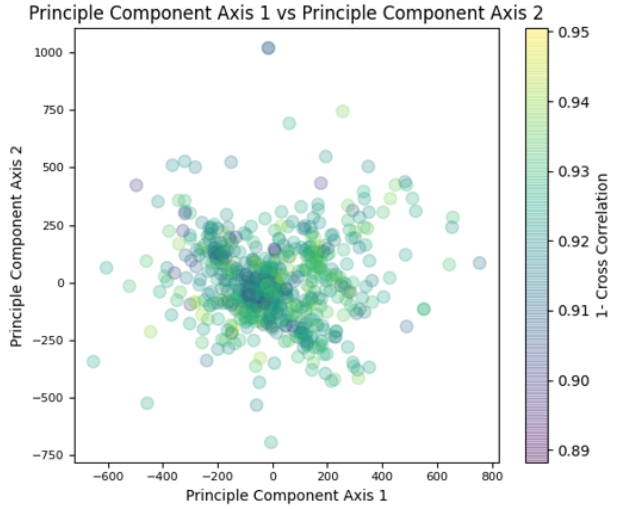


Conversion from Point Density to Free-Energy difference:

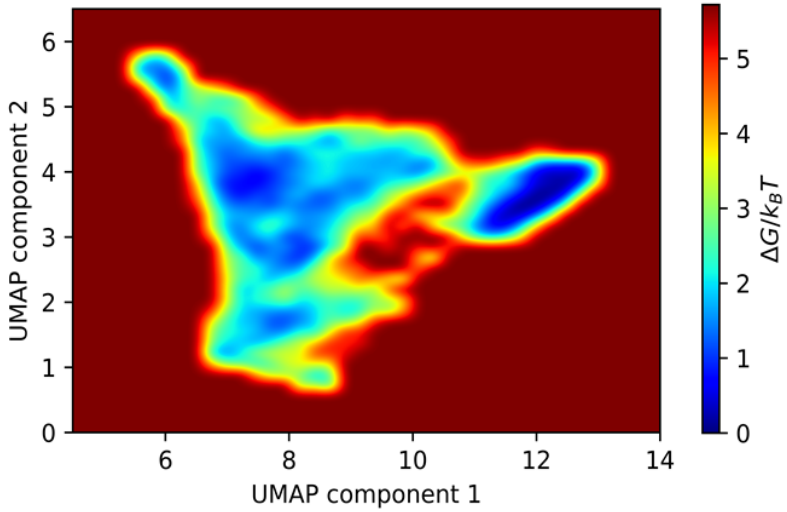
$$\Delta G/k_B\,T = -\ln(n/n_0)$$



Jin et al. *Structure* 2014



Harastani et al. *Front Mol Biosc* 2021

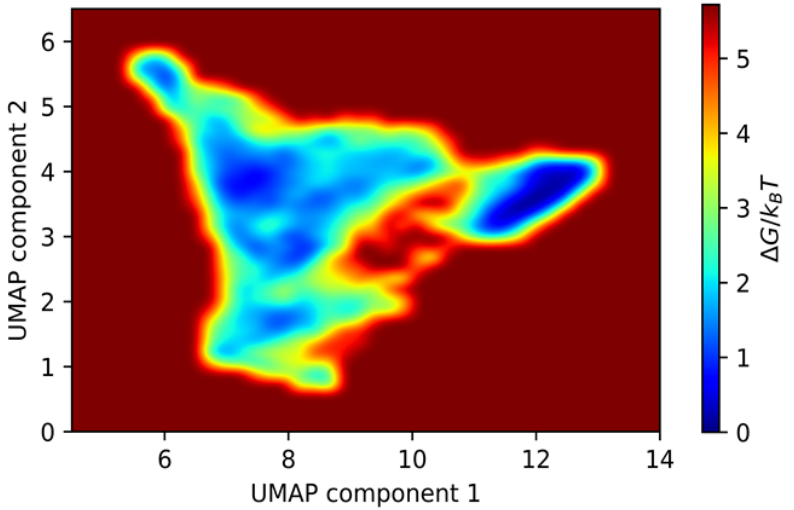


Vuillemot et al. *Sci Rep* 2023

Relationship between Point Density and Free-Energy Difference

$$\exp\left(-\frac{\Delta G}{k_B T}\right) = \frac{n_i}{n_0}$$

$$\Rightarrow \quad \Delta G / k_B\, T = -\ln(n_i / n_0)$$



Vuillemot et al. *Sci Rep* 2023

ΔG: Free-energy difference with respect to a reference state with population n_0

n_0 : Number of particles in the most populated region

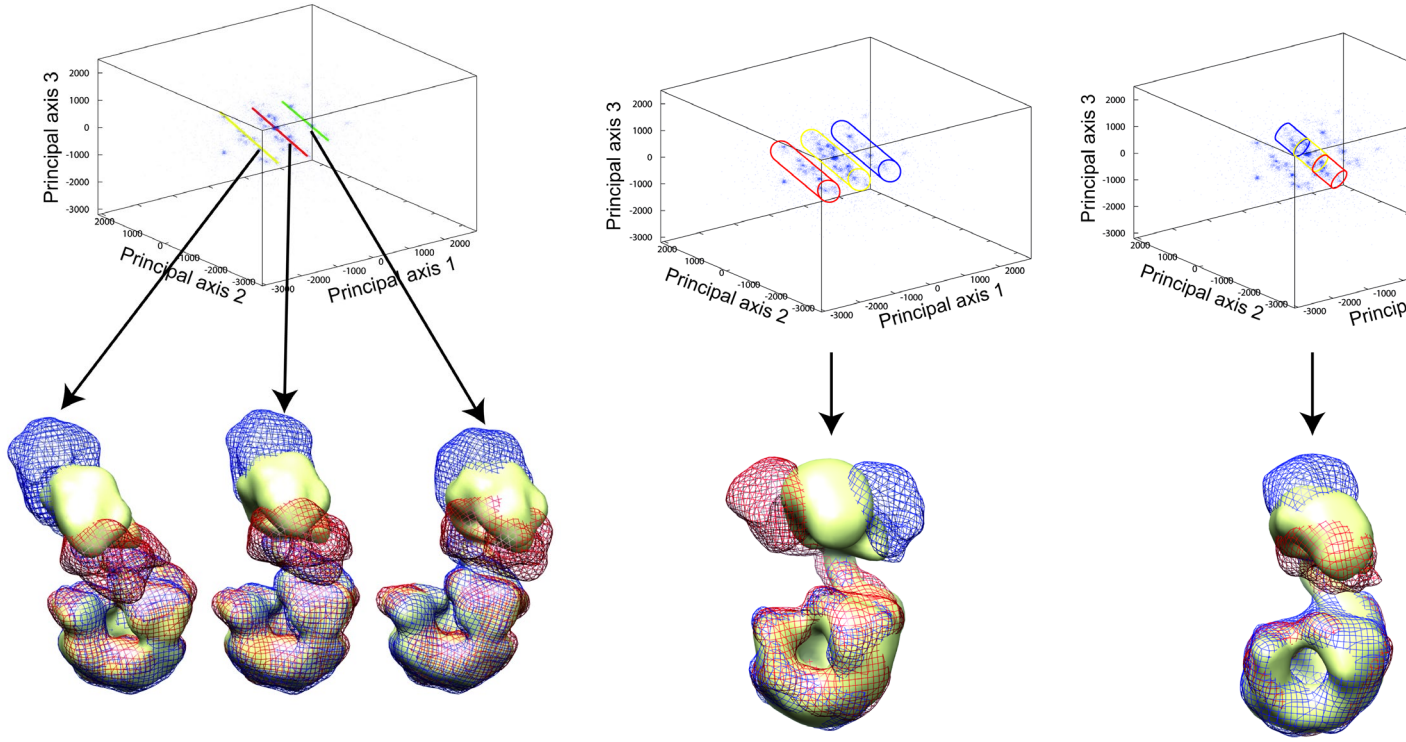
n_i: Number of particles in region *i*

k_B : Boltzmann constant

T : Temperature of the system

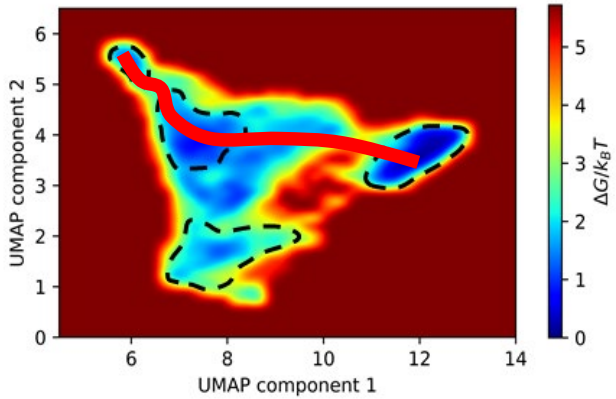
Interpretation of the conformational space

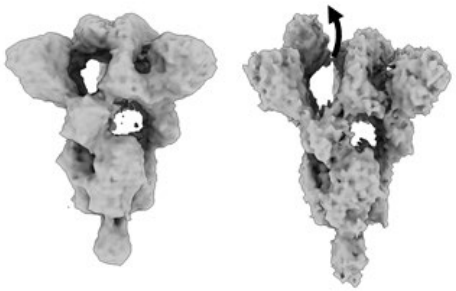
HEMNMA, Jin et al. *Structure* 2014 *SPA, DNA Pol α - B complex*

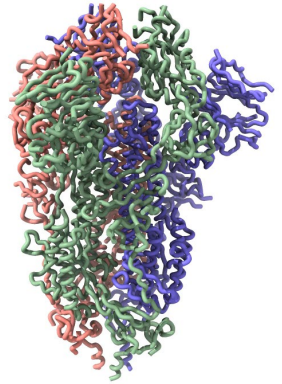


MDTOMO, Vuillemot et al. *Sci Rep* 2023

Cryo-ET, SARS-CoV-2 spike protein

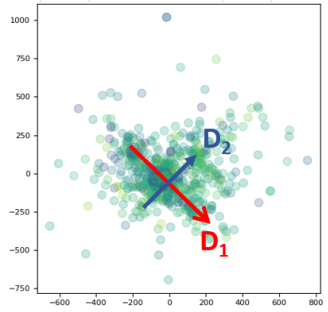


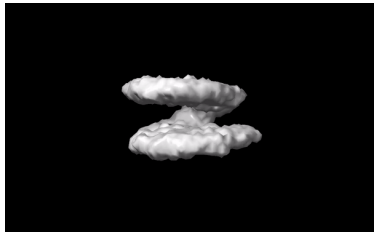




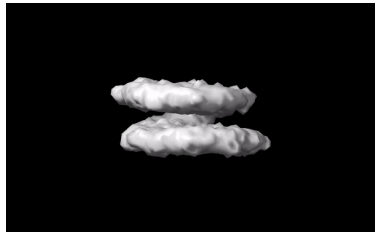
HEMNMA-3D
Harastani et al.
Front Mol Biosc 2021

Cryo-ET,
Nucleosome in situ





D_1



D_2

But, how to obtain
conformational space from experimental data ?

A brief review on methods for extracting continuous conformational landscapes from data

- Methods based on machine learning (statistical or deep learning)
- Methods based on molecular conformational dynamics simulation (hybrid methods)

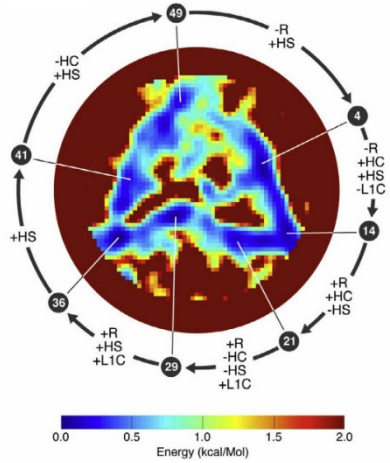
Both types of methods can be considered as data driven

Machine learning methods

- Linear, statistical machine learning methods (e.g., 3DVA in CryoSPARC, Punjani and Fleet 2021)
  - Assume that each conformation can be represented with a sum of a reference conformation and a linear combination of principal conformations, as those that can be obtained by PCA of the covariance matrix

- Non-linear, statistical machine learning methods

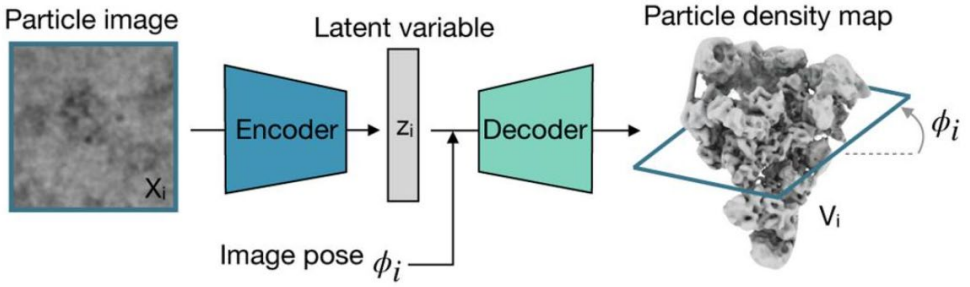
  (e.g., Manifold Embedding, Dashti et al. 2014)



- Non-linear, deep learning methods

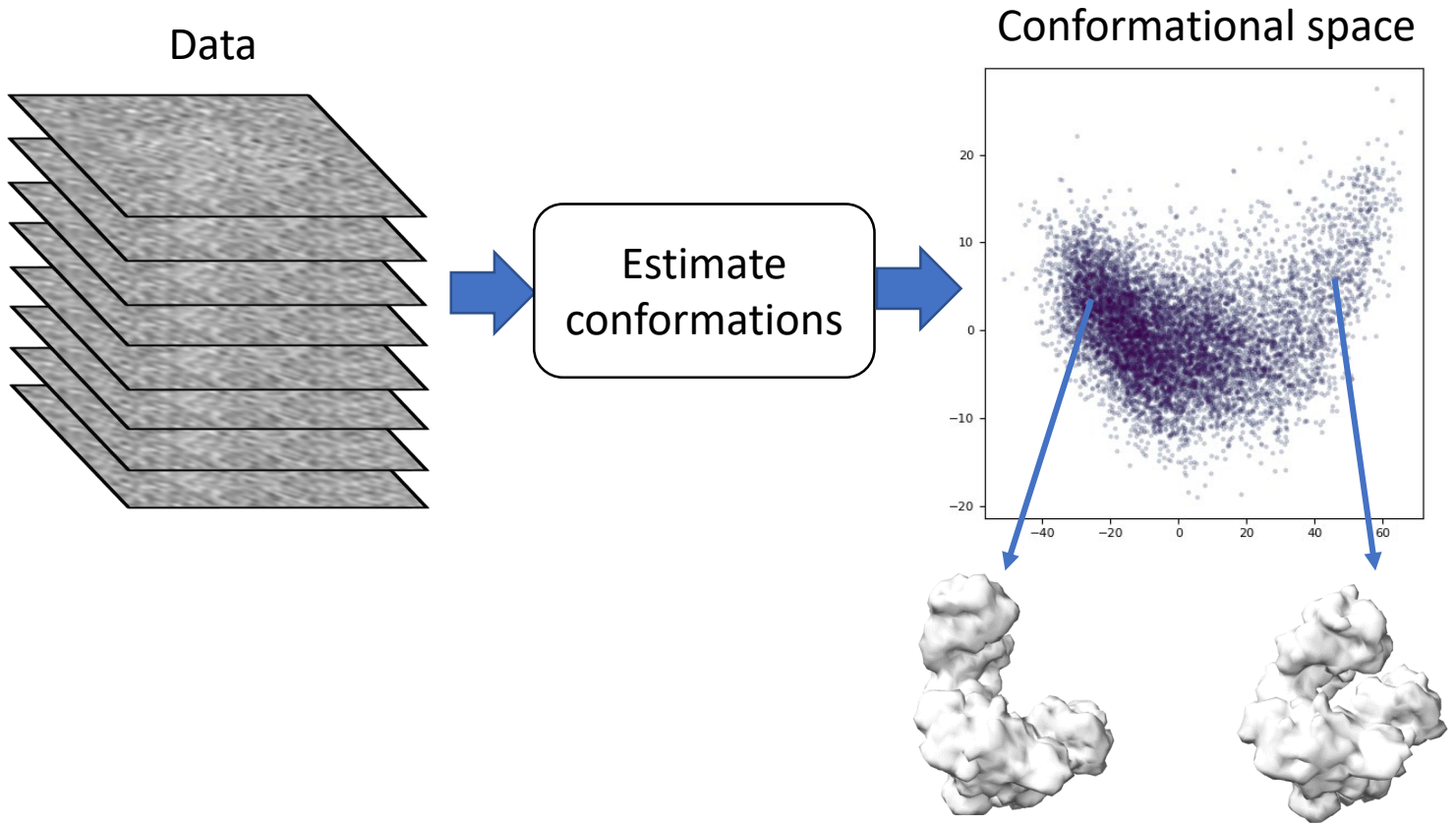
  (e.g., CryoDRGN, Zhong et al. 2021;

  3DFlex, Punjani & Fleet 2023)



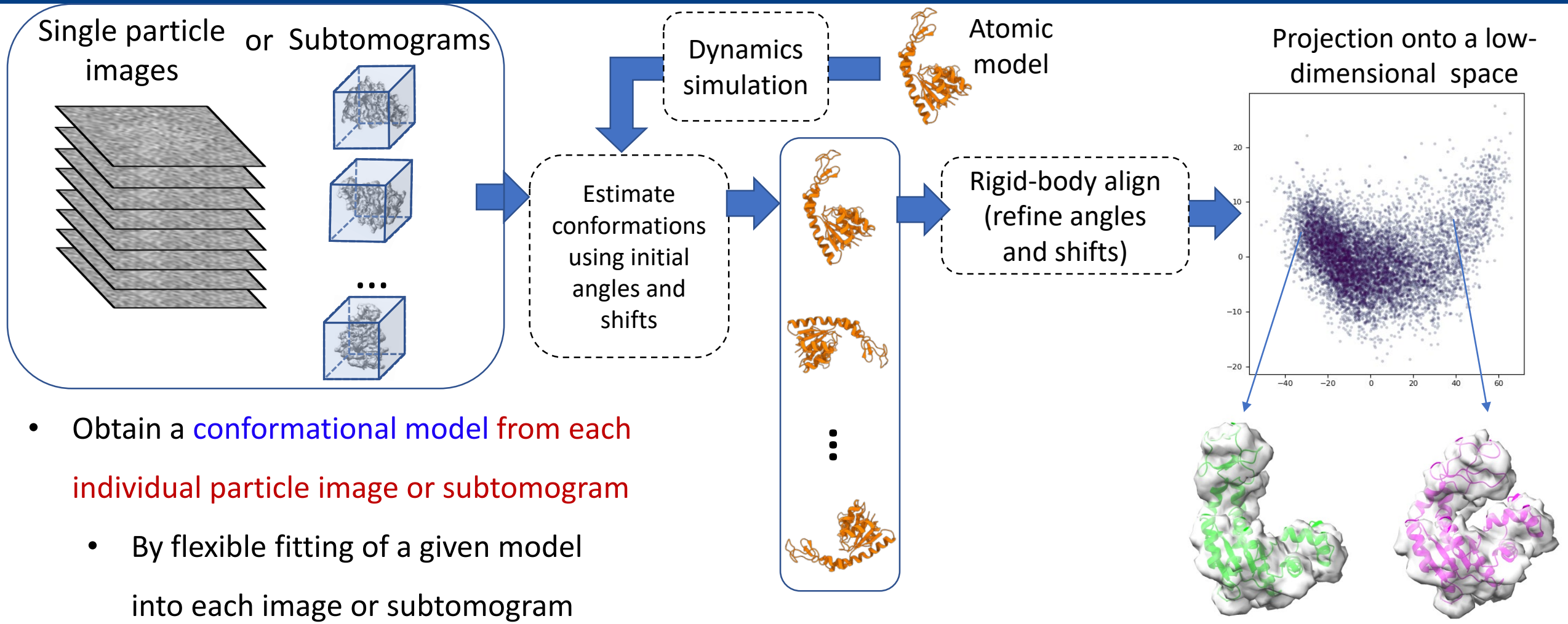
Machine learning methods – Cont'd

- Often use pre-estimated angles and shifts
- Produce density-map representations of the conformational spaces
- Deep learning methods produce models of density maps
  - Prediction can be correct or incorrect



Hybrid methods: Use a prior structural knowledge (e.g., atomic model)

Integrate dynamics simulation using this model into data analysis

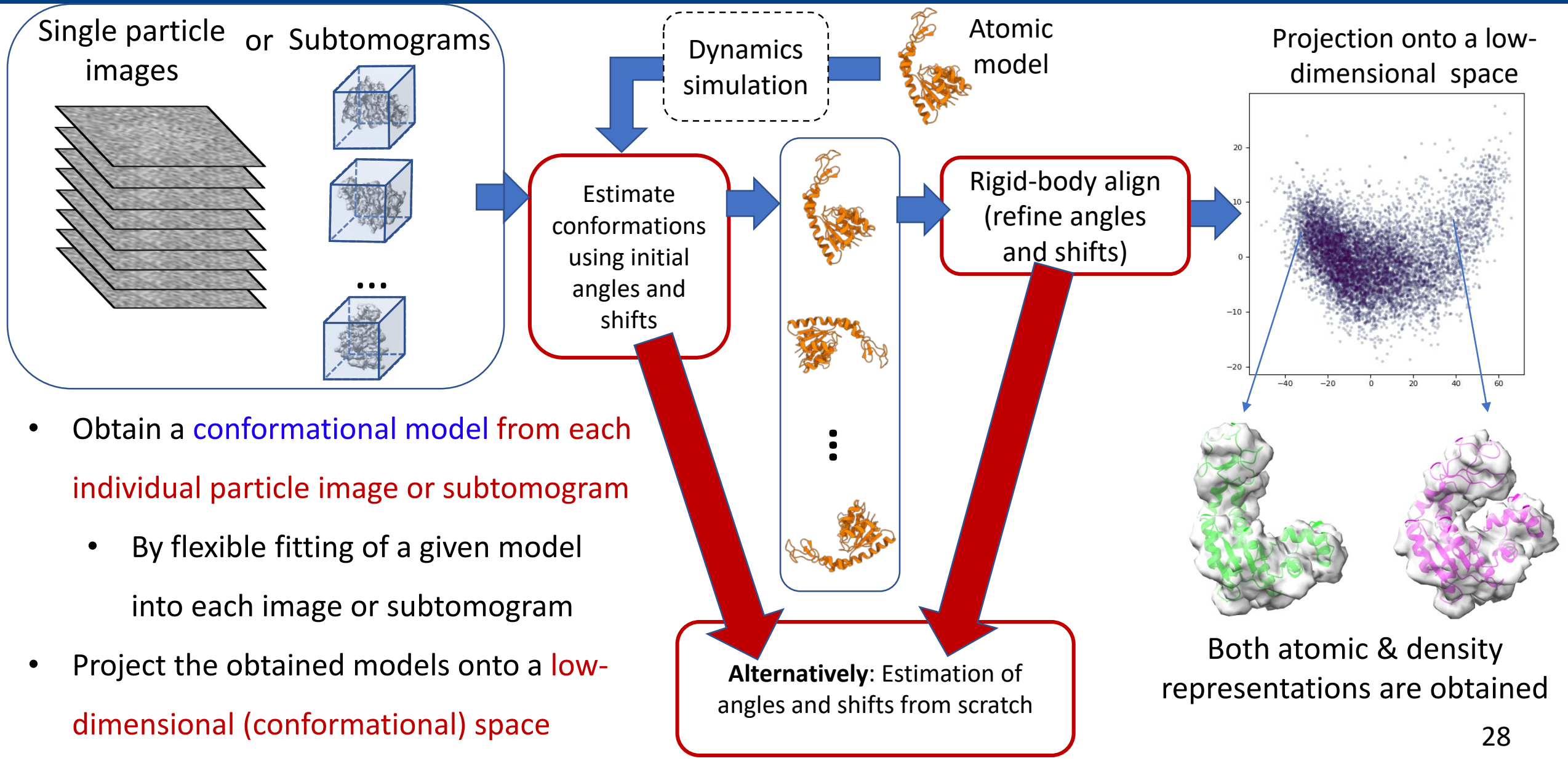


- Obtain a conformational model from each individual particle image or subtomogram
  - By flexible fitting of a given model into each image or subtomogram
- Project the obtained models onto a low-dimensional (conformational) space

Both atomic & density representations are obtained

Hybrid methods: Use a prior structural knowledge (e.g., atomic model)

Integrate simulation of dynamics using this model into data analysis



- Obtain a conformational model from each individual particle image or subtomogram
  - By flexible fitting of a given model into each image or subtomogram
- Project the obtained models onto a low-dimensional (conformational) space

Hybrid methods

- Advantages
  - Use prior structural information if available
  - Obtain conformational space at atomic level
  - Elucidate rare states at atomic level

- Challenges
  - Computational requirements
    - Realistic simulations (e.g., all atoms or longer simulations sometimes required)
    - Large number of particles should be analyzed

PART 2 :

GOING FURTHER INTO DETAILS OF HYBRID METHODS

…. and combining simulation, image processing,
and even deep learning for analyzing conformational variability

PART 2: GOING FURTHER INTO DETAILS OF HYBRID METHODS

- Standard approaches to simulate conformational dynamics
  - Molecular Dynamics (MD) simulation
  - Normal Mode Analysis (NMA)
- Earlier hybrid approaches
  - 3D-to-3D flexible fitting of an initial atomic model into a single EM map
  - Using MD simulation or Normal Modes
- How to efficiently combine NMA and MD simulations ?
  - Normal Mode Molecular Dynamics (NMMD) approach
  - Use of NMMD in 3D-to-3D flexible fitting

Standard approaches to simulate conformational dynamics

- Molecular Dynamics (MD) simulation
- Normal Mode Analysis (NMA)

Molecular Dynamics (MD) simulation

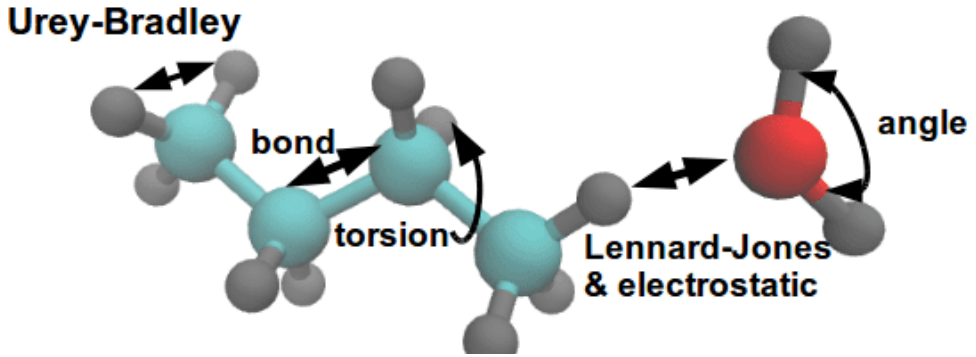
- Mimics the behavior of the atoms in the real world
- Integrates Newton‘s equations of motion

$$F = ma$$

$$F = -\nabla U(x) \qquad a = \frac{d^2x}{dt^2}$$

$$\frac{d^2x}{dt^2} = -\frac{1}{m}\nabla U(x)$$

↳ Solved using numerical integrator (e.g. Verlet, Leapfrog, ...)



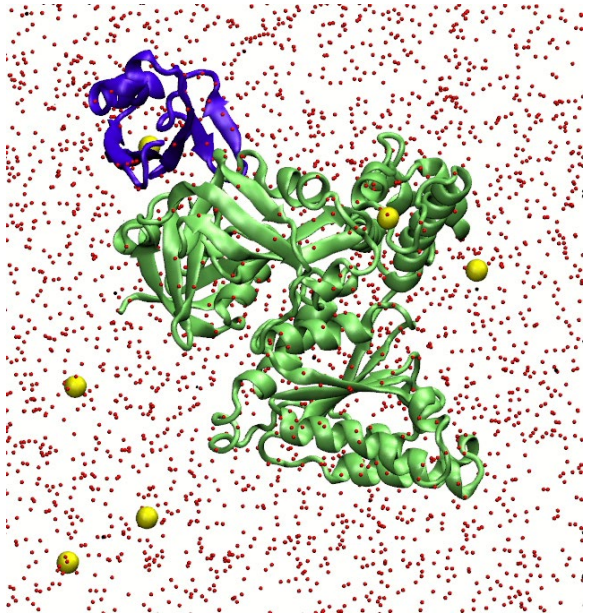
$$U = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[1 + cos(n\phi - \delta)]$$
$$+ \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u(u - u_0)^2$$
$$+ \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_r r_{ij}} \right)$$

CHARMM forcefield

Applications of MD simulations

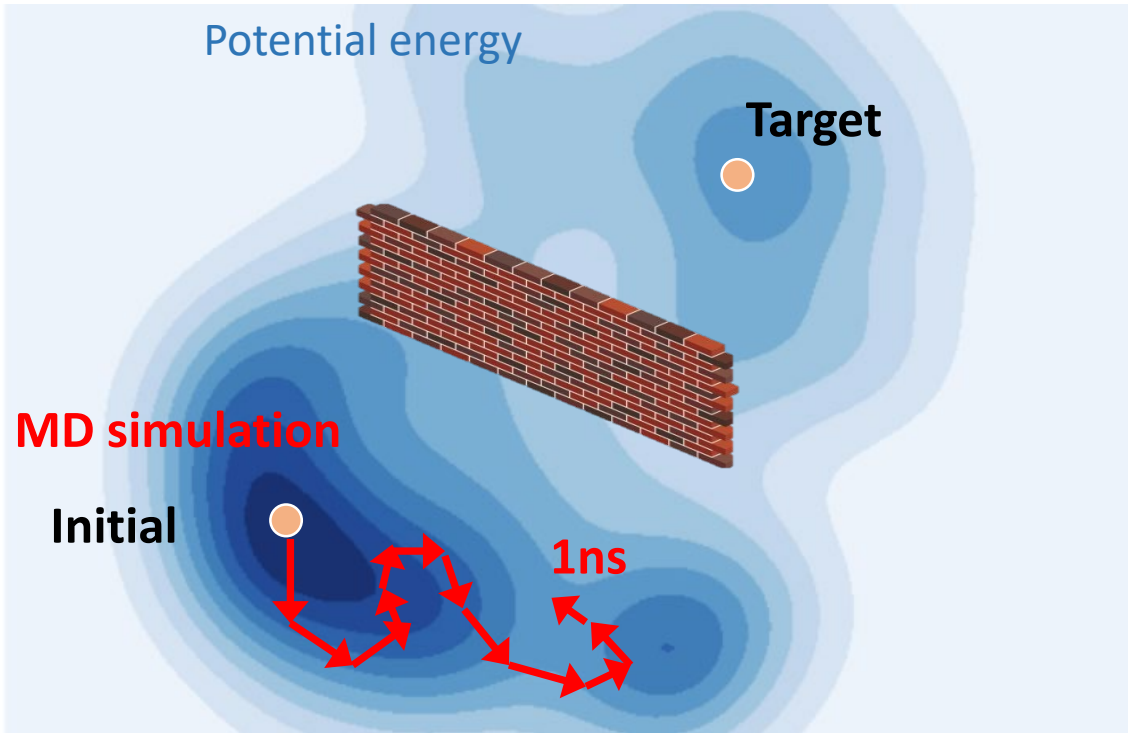
MD simulation allows studying:

- Protein folding
- Protein interactions with other proteins
- Ligand binding
- Conformational flexibility



MD simulation of p97 in complex with p37
Abolfazl et al. *BMC Mol. and Cell Biol.* 2022

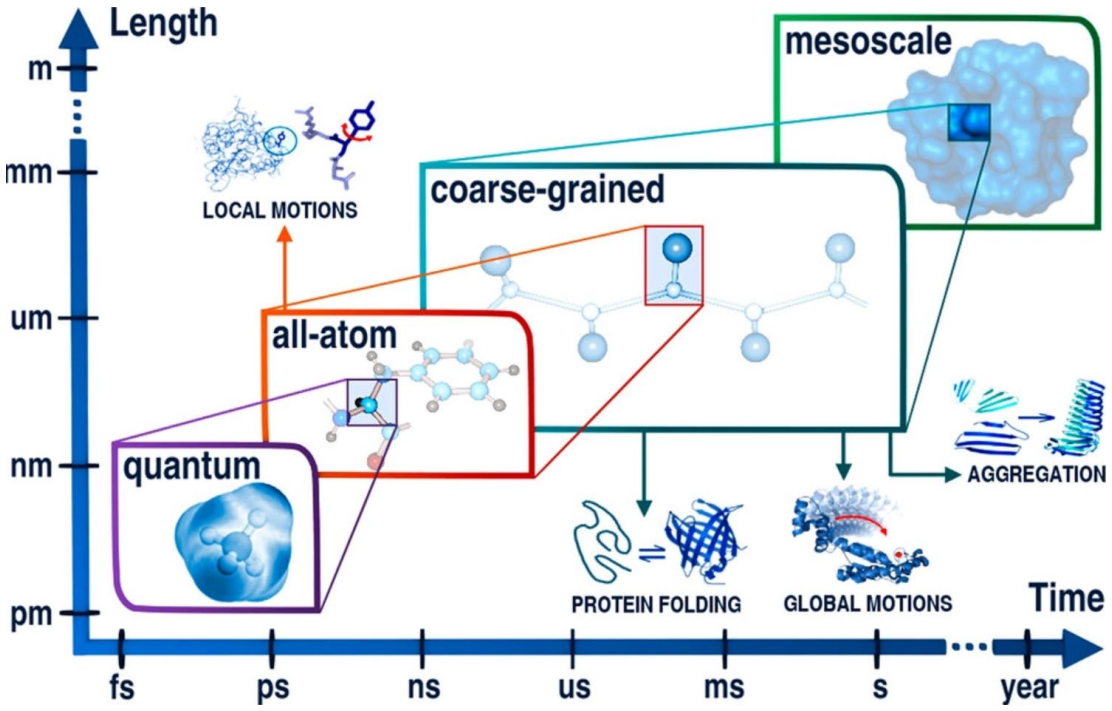
Collab. Rouiller team, The University of Melbourne

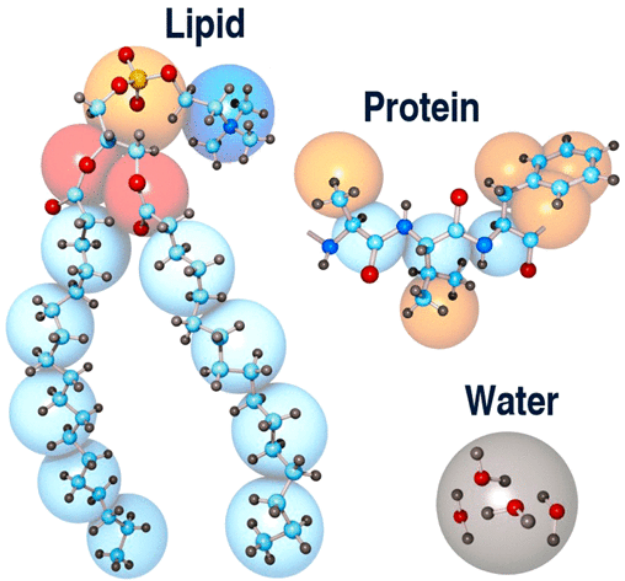


$$U = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+cos(n\phi-\delta)]$$
$$+ \sum_{impropers} k_\omega(\omega-\omega_0)^2 + \sum_{Urey-Bradley} k_u(u-u_0)^2$$
$$+ \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_r r_{ij}} \right)$$

Limitations of MD simulation

- High computational cost :
  - Small time step
  - Large number of degrees of freedom
- Could be reduced with coarse-graining





All-atom *vs*. coarse-grained representation (MARTINI model)

Kmiecik et al. *Chem. Rev.* 2016

Brief conclusions on MD simulation

MD simulation

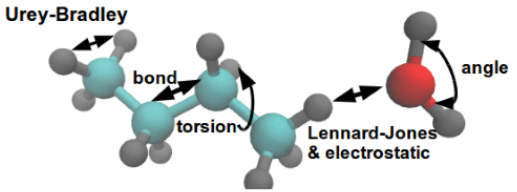
Newton‘s equations of motion

$$F = ma$$

$$F = -\nabla U(x) \qquad a = \frac{d^2x}{dt^2}$$

$$\frac{d^2x}{dt^2} = -\frac{1}{m}\nabla U(x)$$

↳ Solved using numerical integrator (e.g. Verlet, Leapfrog, ...)



$$U = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+cos(n\phi-\delta)]$$
$$+ \sum_{impropers} k_\omega(\omega-\omega_0)^2 + \sum_{Urey-Bradley} k_u(u-u_0)^2$$
$$+ \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_r r_{ij}} \right)$$

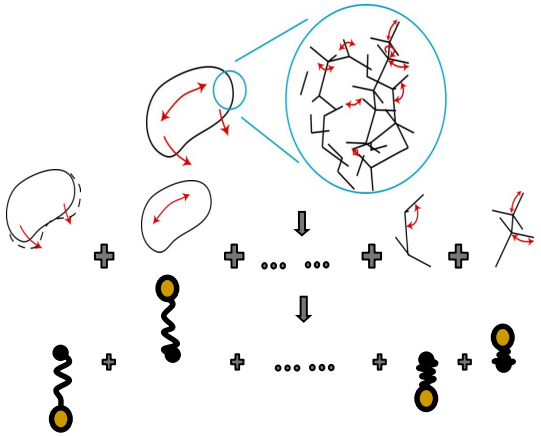
CHARMM forcefield

⇒High-quality atomic models

⇒Slow

Normal Mode Analysis (NMA)

Decomposes the total motion into harmonic oscillator motions

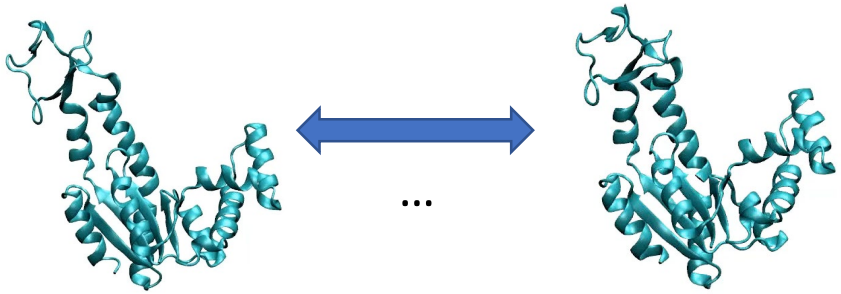


Low frequencies
Collective motions

High frequencies
Localized motions

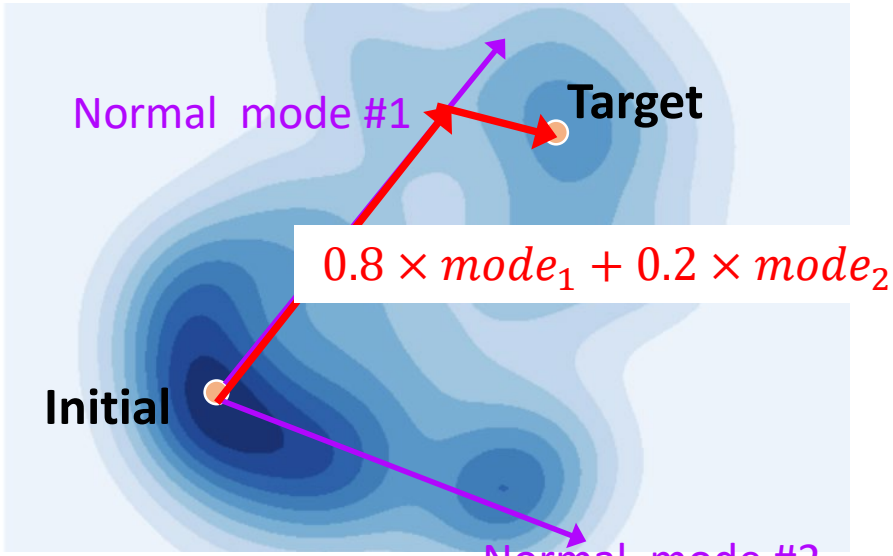
$$x = q \cdot A + r_0$$

Total motion: Linear combination of normal-mode motions



Low frequency

High frequency

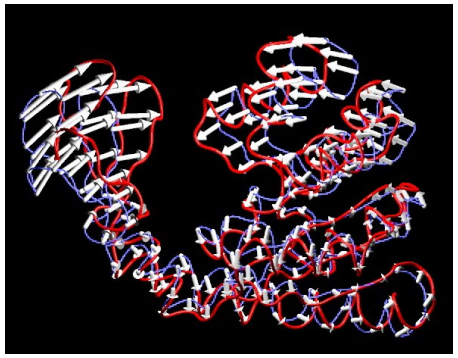


How to obtain normal modes ?

Tirion's Elastic Network Model (ENM)

Tirion, Phys Rev Lett 1996

- Allows obtaining normal modes of atomic structures or EM maps
- Before NMA, EM maps must be converted into "pseudoatoms" (3D Gaussians)
- Initial conformation is assumed to be at the energy minimum
- Atoms or "pseudoatoms" are connected with elastic springs (within a given radius)

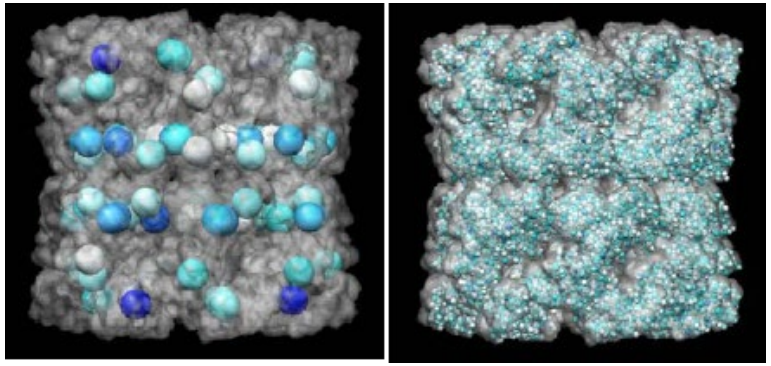


Atomic displacement with one normal mode

R: Radius of interaction (within which atoms are connected)

$$E_p = \sum_{r^0_{a,b}<R} E\left(r_a, r_b\right)$$

$$E\left(r_a,\ r_b\right) = \frac{C}{2}\left(\ \left|r_{a,b}\right|\ -\ \left|r^0_{a,b}\right|\ \right)^2$$



EM map and its two coarse-grained models (spheres: 3D Gaussian functions called pseudoatoms)

Jonic & Sorzano, IEEE J STSP 2016

Harmonic approximation of the potential energy and Diagonalization of the Hessian matrix

Coordinate change

$$E_p(r) \cong \frac{1}{2}\sum_{ij}\left.\frac{\partial^2 E_p}{\partial r_i \partial r_j}\right|_{r=r_o}(r_i - r_i^0)(r_j - r_j^0) \longrightarrow \frac{1}{2}\sum_n \omega_n^2 q_n^2$$

$$\mathbf{H} = \left(\frac{\partial^2 E_p}{\partial r_i \partial r_j}\right)_{r=r_o}$$

Hessian: 2nd derivative of the potential → Size of the system: *3N* x *3N*

N : Number of atoms or pseudoatoms

Eigenvalue problem

$$\mathbf{A}^t\mathbf{H}\mathbf{A} = \mathbf{L}$$

$$\mathbf{A} = \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots\end{pmatrix}$$

Eigenvector = normal mode

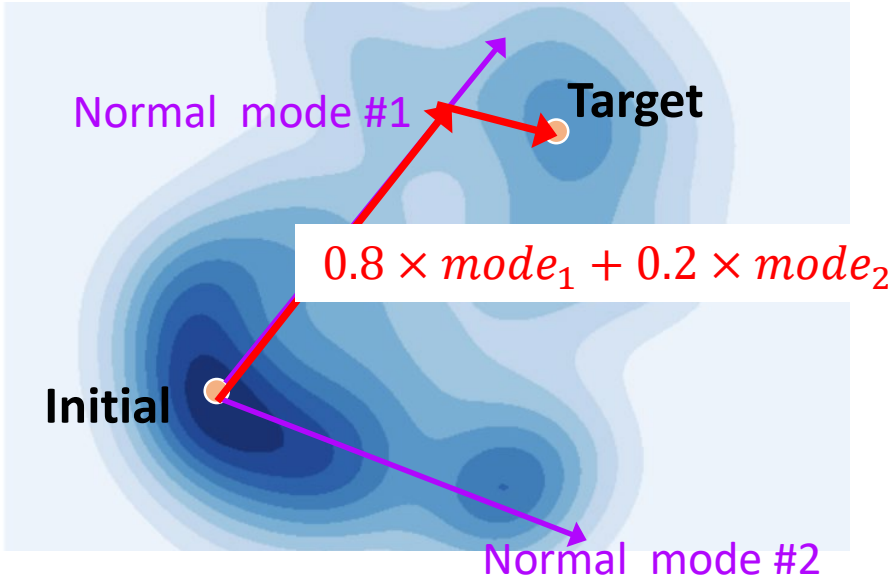
$$\mathbf{L} = \begin{pmatrix}\omega_1^2 & & 0 \\ & \omega_2^2 & \\ 0 & & \ddots\end{pmatrix}$$

Eigenvalue = frequency^2

$$x_i(\mathbf{q}) = \sum_{j=7}^{3N} a_{ij} q_j + x_i^0$$

Amplitudes (no physical units) can be obtained by correlating models with EM data

Challenges and Limitations of the use of normal modes



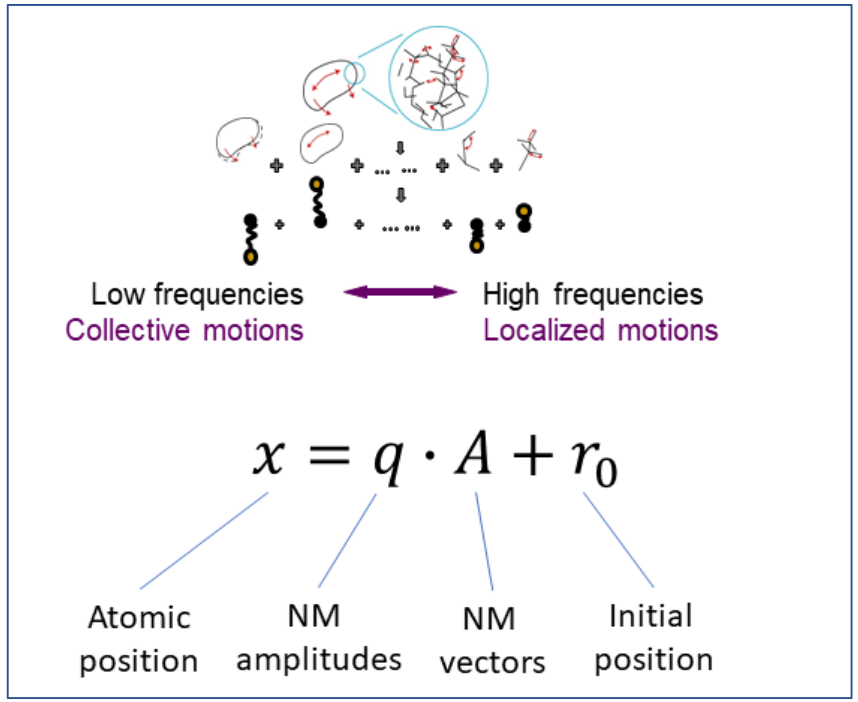
$$x = q \cdot A + r_0$$

- How to select normal modes?
  - Usual use of low-frequency collective normal modes (fitting of global, large-scale motions)
- Validity of NMA for larger motion amplitudes around the initial conformation
  - May generate structural distortions

Brief conclusions on NMA

NMA

Linear combination of harmonic oscillator motions
obtained by harmonic approximation of potential energy

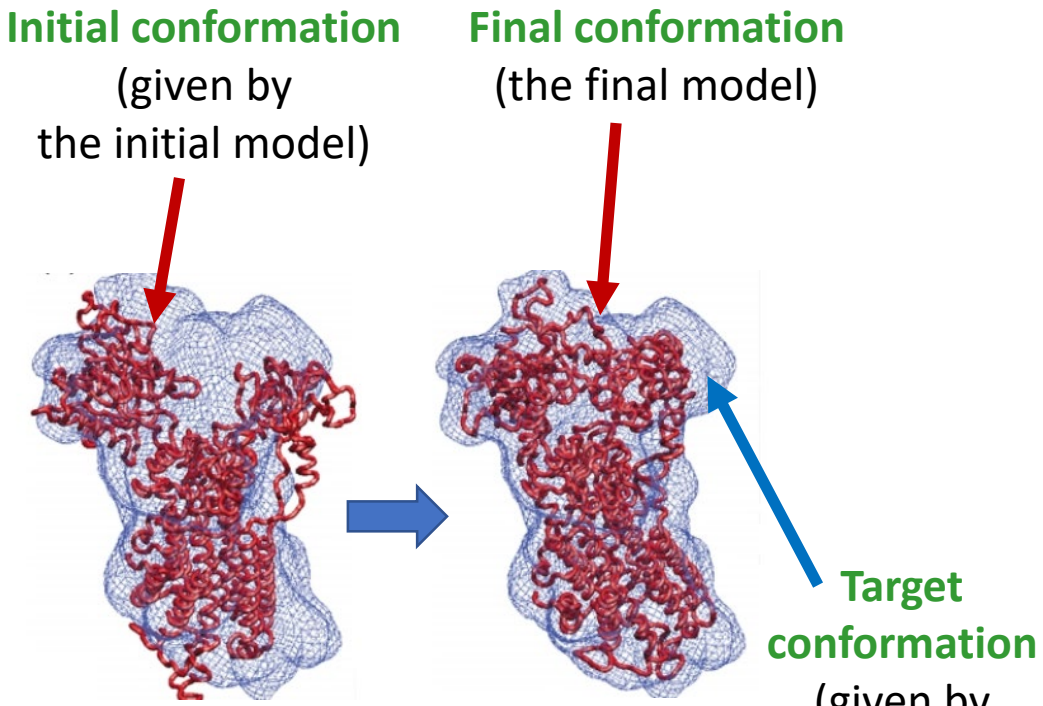


⇒Fast

⇒Large amplitudes of normal modes may induce model distortions

Earlier hybrid approaches

3D-to-3D flexible fitting of an initial atomic model into a single EM map

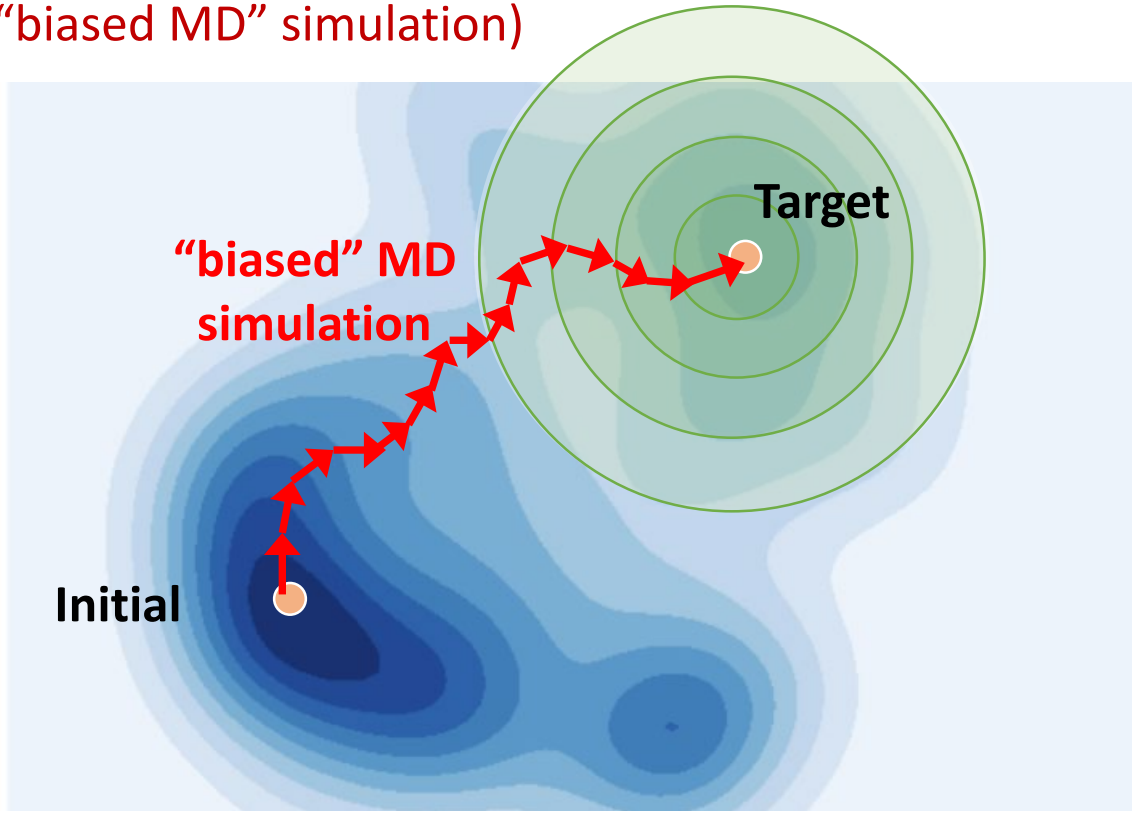


Tama et al. 2004

- Fitting using MD simulation
- Fitting using Normal Modes

Use of MD simulation for 3D-to-3D flexible fitting of an EM map with an atomic model

Adding a biasing potential using experimental data
("biased MD" simulation)



$$U = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+cos(n\phi-\delta)] + \sum_{impropers} k_\omega(\omega-\omega_0)^2 + \sum_{Urey-Bradley} k_u(u-u_0)^2 + \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_r r_{ij}} \right) + U_{biased}$$

$$U_{biased} = k(1-CC)$$

k: Force constant

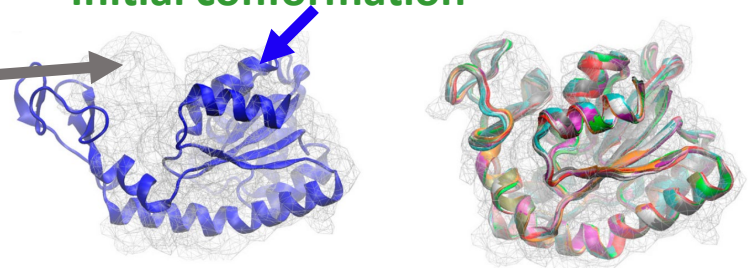
$$CC = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2} \sqrt{\sum_i (y_i - y_m)^2}}$$

CC : Pearson correlation coefficient

"3D biasing potential"

Initial conformation

Target conformation (given by the EM map)



Orzechowski & Tama 2008

Use of MD simulation for 3D-to-3D flexible fitting of an EM map with an atomic model (Cont'd)

$$g(x,y,z;x_n,y_n,z_n,) = \exp\left[-\frac{3}{2\sigma^2}\{(x-x_n)^2 + (y-y_n)^2 + (z-z_n)^2\}\right],$$

$$\rho^{sim}(i,j,k) = \sum_{n=1}^{N} \int_{V_{iik}} dxdydz\, g(x,y,z,x_n,y_n,z_n),$$

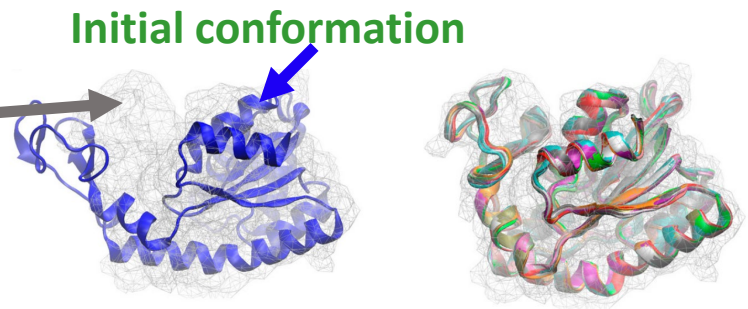
$$c.c. = \frac{\sum_{ijk} \rho^{exp}(i,j,k)\rho^{sim}(i,j,k)}{\sqrt{\sum_{ijk} \rho^{exp}(i,j,k)^2 \sum_{ijk} \rho^{sim}(i,j,k)^2}},$$

Fitting using MD simulation :

⇒High-quality atomic models

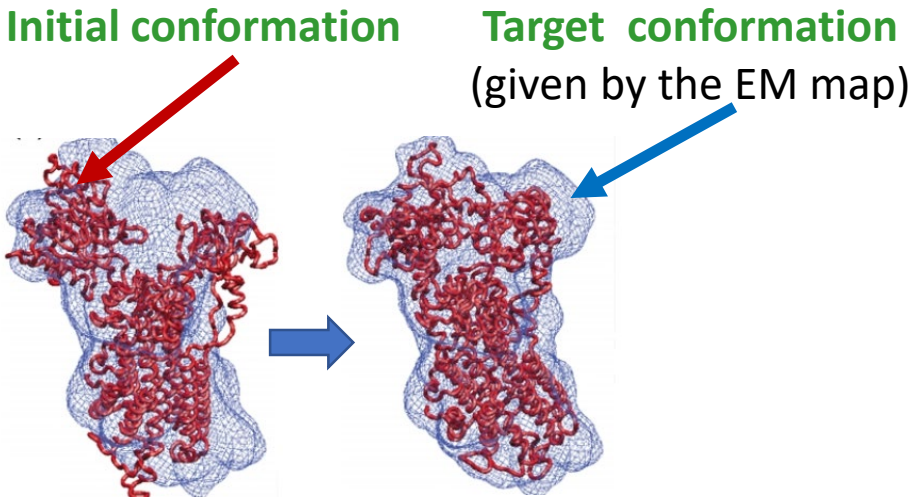
⇒Local motions

⇒Slow, especially for large complexes



Orzechowski & Tama 2008

Use of Normal Modes for 3D-to-3D flexible fitting of an EM map with an atomic model



Tama et al. 2004

Optimizing q to match the target data

$$x = q \cdot A + r_0$$

NMA-based fitting: *M* parameters, *M* << *3 N*
(**compared to MD-based fitting:** *3 N* parameters)

Fast, but …

Challenges and Limitations:

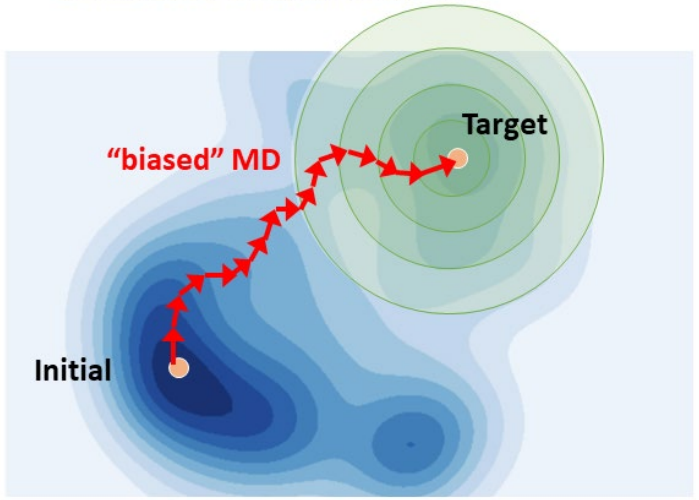
- How to select normal modes?
  - Usual use of low-frequency collective normal modes (fitting of global, large-scale motions)
- Validity of NMA for larger motion amplitudes around the initial conformation
  - May generate structural distortions

Brief conclusions on earlier hybrid approaches

Fitting using biased MD simulation

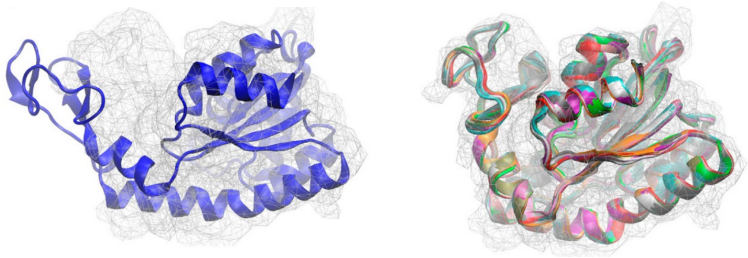
$$U_{biasing} = 1 - CC$$

⇒ High-quality atomic models
⇒ Local motions
⇒ Slow



$$U = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+cos(n\phi-\delta)] + \sum_{impropers} k_\omega(\omega-\omega_0)^2 + \sum_{Urey-Bradley} k_u(u-u_0)^2 + \sum_{nonbonded} \left(\epsilon_{ij}\left[\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon_r r_{ij}}\right) + U_{biasing}$$

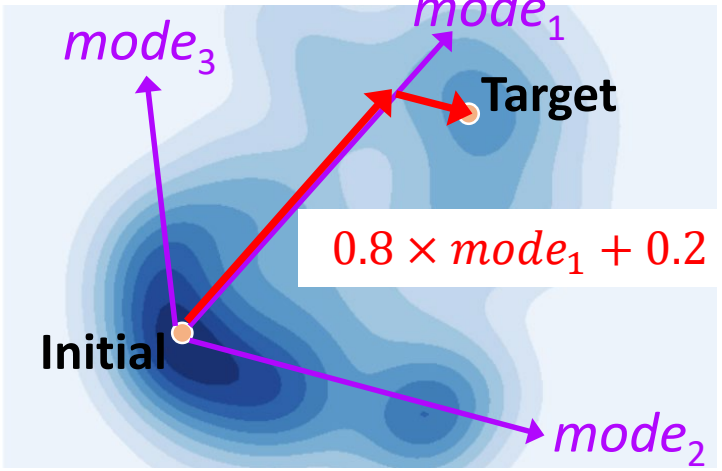
Initial Final



3 N parameters, *N* atoms

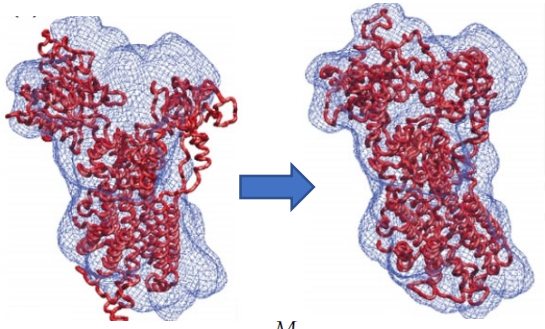
Fitting using normal modes

⇒ Fast
⇒ Global motions
⇒ Not all conformations possible
⇒ Distortions for large amplitudes of conformational change



$$0.8 \times mode_1 + 0.2 \times mode_2$$

Initial Final



$$x_n(\mathbf{q}) = \sum_{l=1}^{M} a_{nl} q_l + x_n^0$$

M parameters, *M* << *N*

How to efficiently combine NMA and MD simulations ?

- Normal Mode Molecular Dynamics (NMMD) approach (Vuillemot et al. *J Mol. Biol.* 2022)
- Incorporates displacements along normal modes (NMs) into MD simulation (into the computation of the potential energy function for MD simulation)
- Simultaneously determines both NM and MD parameters
- Speeds up MD simulations using NMs while preserving high quality of the fitted models

Normal Mode Molecular Dynamics (NMMD) approach

Incorporates NM-displacements in the computation of the potential energy function $U(r)$ during MD simulation

$$r(t) = \underbrace{q(t) \cdot A}_{\text{NM-based displacement}} + \underbrace{x(t)}_{\text{Atomic displacement}} + r_0$$

$$F_x = m_x \ddot{x} \qquad F_q = m_x \ddot{q}$$

$$F_x = -\frac{\partial U}{\partial x} \qquad F_q = -\frac{\partial U}{\partial q}$$

$$F_q = A F_x$$

$$\mathbf{x}(t+\Delta t) = \mathbf{x}(t) + \dot{\mathbf{x}}(t)\Delta t + \frac{1}{2}\ddot{\mathbf{x}}(t)\Delta t^2$$

$$\dot{\mathbf{x}}(t+\Delta t) = \dot{\mathbf{x}}(t) + \frac{\ddot{\mathbf{x}}(t) + \ddot{\mathbf{x}}(t+\Delta t)}{2}\Delta t,$$

$$\mathbf{q}(t+\Delta t) = \mathbf{q}(t) + \dot{\mathbf{q}}(t)\Delta t + \frac{1}{2}\ddot{\mathbf{q}}(t)\Delta t^2$$

$$\dot{\mathbf{q}}(t+\Delta t) = \dot{\mathbf{q}}(t) + \frac{\ddot{\mathbf{q}}(t) + \ddot{\mathbf{q}}(t+\Delta t)}{2}\Delta t,$$

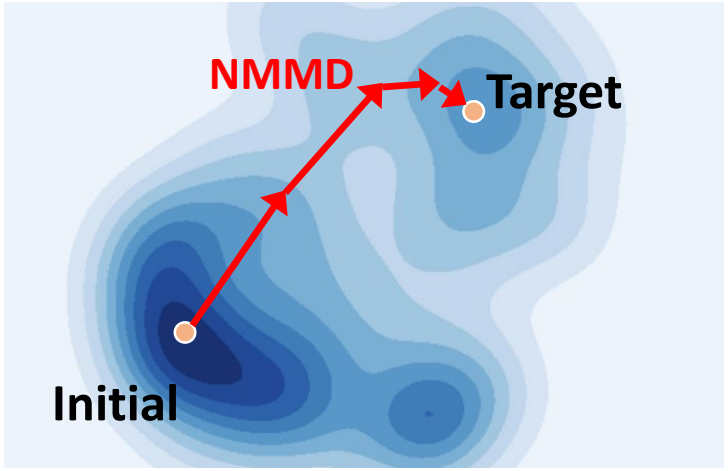
$$\begin{aligned}
\mathbf{F}_q(t) &= -\frac{\partial U}{\partial \mathbf{q}}(t) \\
&= -\frac{\partial \mathbf{r}}{\partial \mathbf{q}}(t) \cdot \frac{\partial U}{\partial \mathbf{r}}(t) \\
&= -\frac{\partial \mathbf{r}}{\partial \mathbf{q}}(t) \cdot \left(\frac{\partial \mathbf{r}}{\partial \mathbf{x}}(t)\right)^{-1} \cdot \frac{\partial U}{\partial \mathbf{x}}(t) \\
&= \mathbf{A} \cdot \mathbf{F}_x(t).
\end{aligned}$$

Rémi Vuillemot
PhD, Oct 2023
(now Pdoc
INRIA, Grenoble)

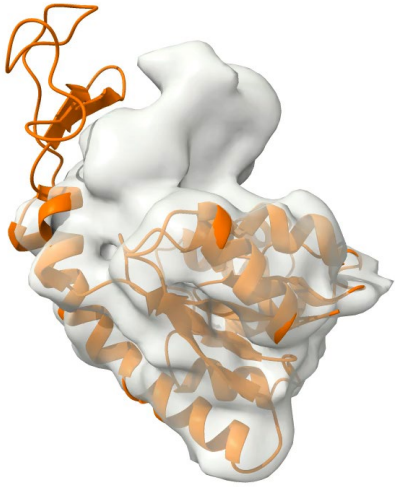
Vuillemot et al.
J Mol. Biol. 2022

NMMD approach – Cont'd

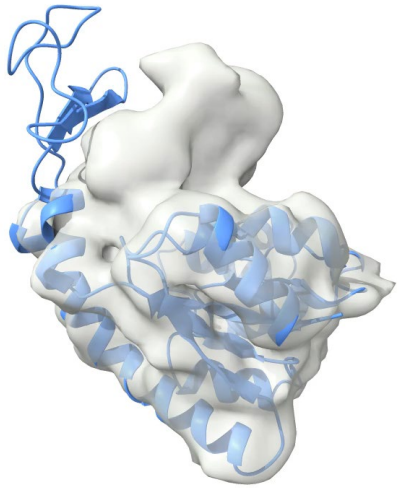
$$r(t) = \underbrace{q(t) \cdot A}_{\text{NM-based displacement}} + \underbrace{x(t)}_{\text{Atomic displacement}} + r_0$$



Faster than MD alone



NMMD



MD

Vuillemot et al. *J Mol. Biol.* 2022

Use of NMMD in 3D-to-3D flexible fitting

+40% in average

Faster fitting with NMMD than with MD

Biomolecular complex	Fitting method	CC	RMSD (Å)	Total time (min)	Convergence time (min)	Speed increase	MolProbity score
LAO binding protein	MD	0.83	1.04	10.5	6.2	+32%	2.84
	NMMD	0.83	1.06	10.6	4.2		2.79
Adelynate kinase	MD	0.83	1.88	9.8	5.3	+62%	2.61
	NMMD	0.84	1.58	10.0	2.0		2.59
Lactoferrin	MD	0.79	1.57	132.2	122.2	+46%	2.33
	NMMD	0.79	1.45	133.2	64.9		2.37
Elongation factor 2	MD	0.73	3.67	402.5	358.2	+10%	2.67
	NMMD	0.75	2.13	405.0	322.6		2.49
ABC exporter	MD	0.79	2.18	229.9	179.3	+66%	2.21
	NMMD	0.79	1.86	235.3	61.1		2.21
p97 ATPase	MD	0.83	7.17	1754.8	1526.7	+17%	2.00
	NMMD	0.81	6.57	1796.0	1257.2		1.99

- NMMD uses MD simulation package GENESIS 1.4 *(Kobayashi et al., JCC 2017)*
- Can be used with all-atom models (CHARMM and Go) or coarse-grained models (Cα Gō)
- Available in our ContinuousFlex software package

Vuillemot et al. *J Mol. Biol.* 2022

PART 3 : NEW HYBRID APPROACHES

Flexible fitting of large sets of 3D and 2D data

- Analysis of many volumes (cryo-ET subtomograms)
- Analysis of many images (cryo-EM single particle images)

For application in determination of the entire conformational space from a given data set

$\Rightarrow$ Fast approaches are required

PART 3: NEW HYBRID APPROACHES (Flexible fitting of large sets of 3D and 2D data)

- Challenge of flexible fitting of a model into 2D data
- 3D-to-2D flexible fitting of large sets of cryo-EM images
  - HEMNMA: First approach for 3D-to-2D flexible fitting (based on NMs)
  - DeepHEMNMA: Recent approach for HEMNMA speed-up
  - MDSPACE: First approach for 3D-to-2D flexible fitting based on MD simulation
- 3D-to-3D flexible fitting of large sets of cryo-ET subtomograms
  - HEMNMA-3D, MDTOMO
- Open-source ContinuousFlex software package

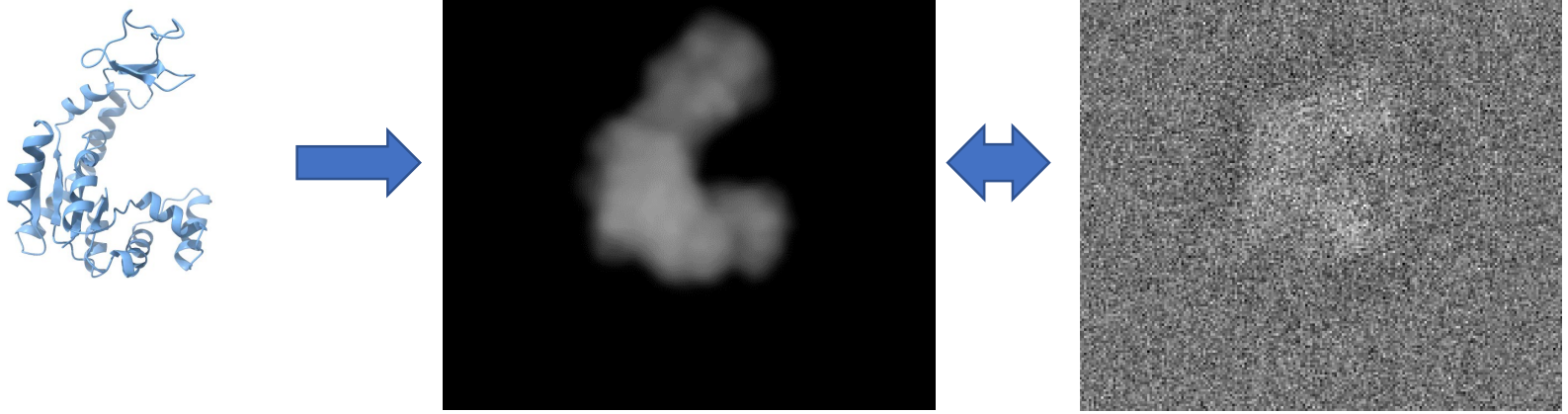
Challenge of flexible fitting of a model into 2D data (missing 3rd dimension and low SNR)

- Going from 3D-to-3D flexible fitting to 3D-to-2D flexible fitting
- Fitting a large set of images with a given prior model (atomic model or EM map)
- At each fitting step, a 2D projection is calculated and compared with a given image

Simulation
(Starts from a given initial conformation)

2D projection
(Test conformation)

Given image
(Target conformation)

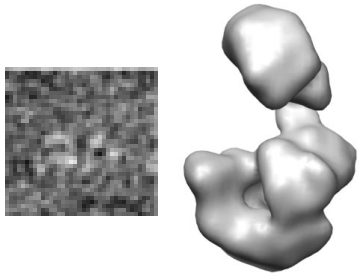


3D-to-2D flexible fitting of PDB 4AKE to a synthetic particle image

3D-to-2D flexible fitting of large sets of cryo-EM images

HEMNMA: First approach for 3D-to-2D flexible fitting (based on NMs)
Obtains simultaneously orientation, position, and conformation from each image

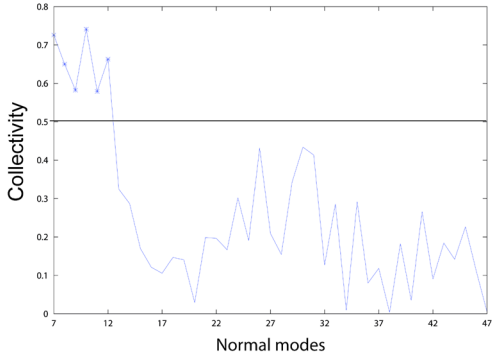
Image Reference



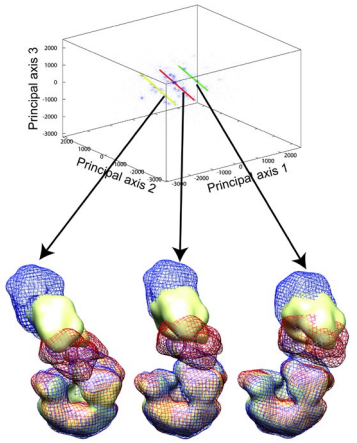
DNA polymerase Pol α – B

Pseudoatoms and NMA

Selection of a subset of normal modes
(low frequency & high collectivity)

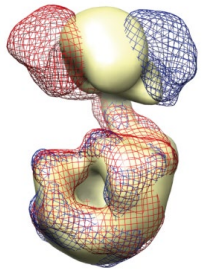


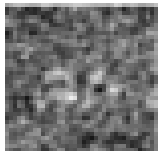
Animated "trajectories"
Displacement of the reference in selected directions



$$x_n(\mathbf{q}) = \sum_{l=1}^{M} a_{nl} q_l + x_n^0$$

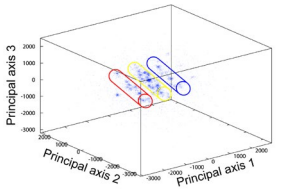
Determine q_l through matching with EM images

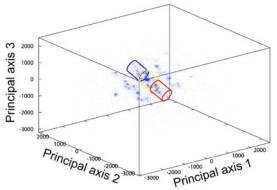


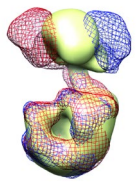


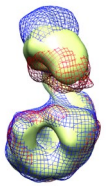
For each single particle image:

Iterative 3D-to-2D elastic and rigid-body alignment to find conformation, orientation, and translation **quasi-simultaneously**

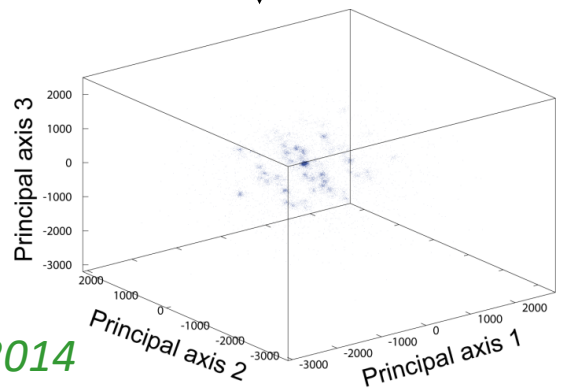








3D reconstructions
Interactive grouping of images with similar conformations



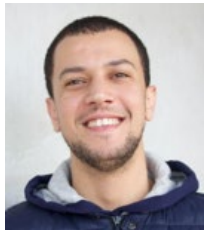
Low-dimensional conformational space:

Image mapping in a low-dimensional space of normal-mode displacement amplitudes

➢Each point: 3 angles, 2 shifts, and M normal-mode amplitudes (conformation)

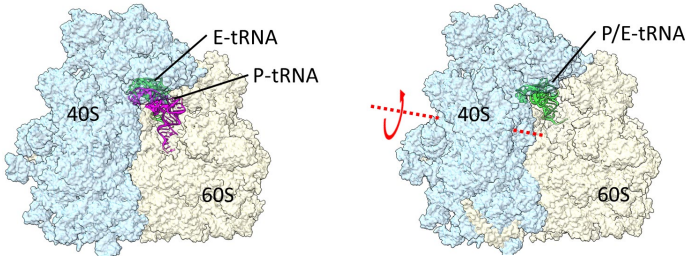
Jin et al., *Structure 2014*

DeepHEMNMA: Recent approach for HEMNMA speed-up
Based on Deep Learning of rigid-body and elastic parameters

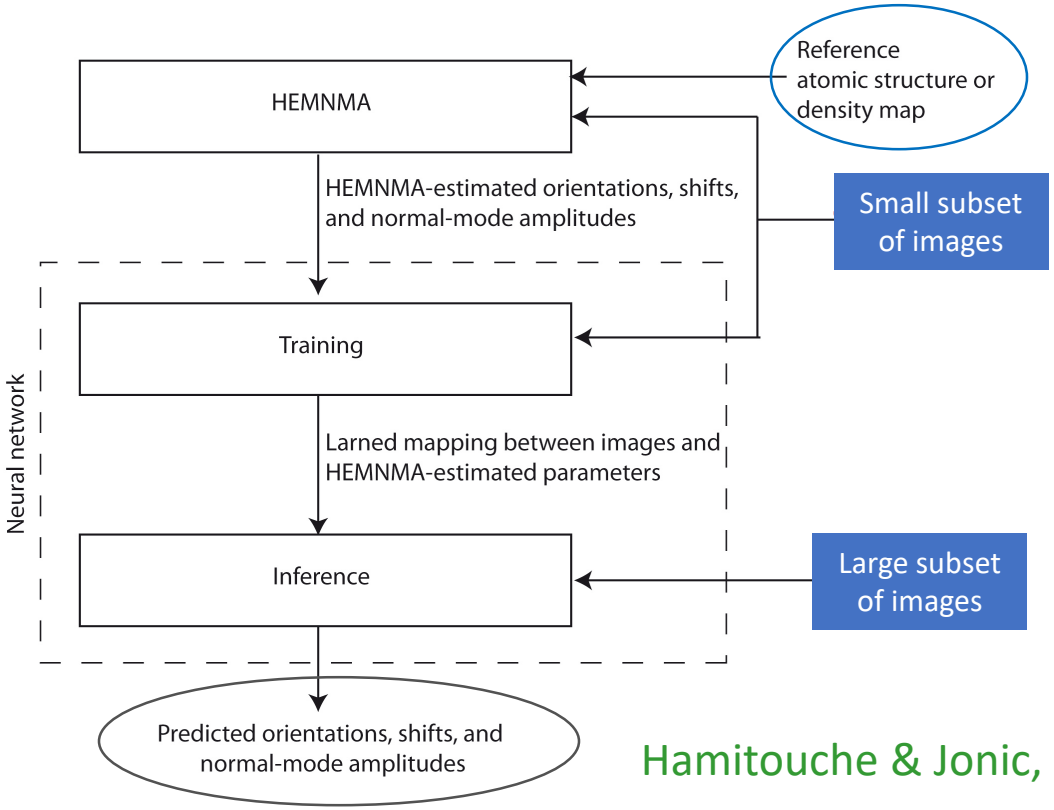


Ilyes Hamitouche
PhD, March 2023
(now Pdoc Curie, Paris)

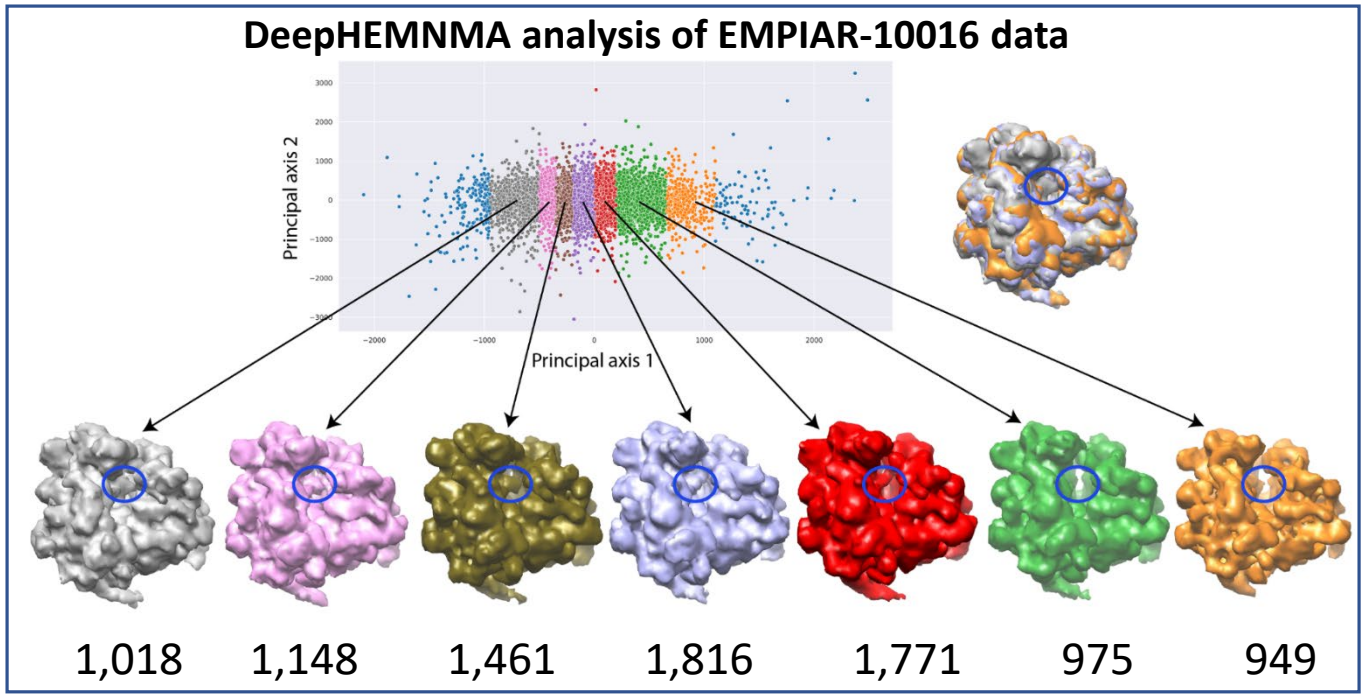
Resnet 34 (34-layer residual net) followed by a 4-layer MLP
Regression learning:
- small number of conformational parameters
  (normal-mode amplitudes)
- rotations and translations



Yeast 80S ribosome-tRNA complexes



DeepHEMNMA analysis of EMPIAR-10016 data



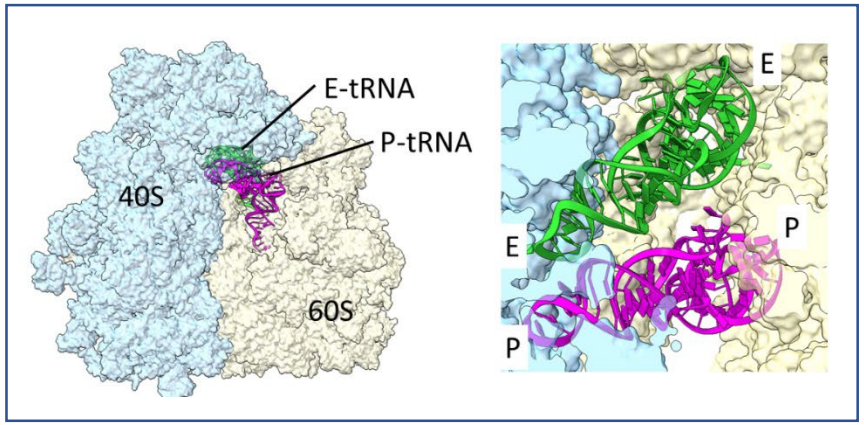
Reconstruction of 7 maps showing variable degree of inter-subunit rotation and presence/absence of the second tRNA

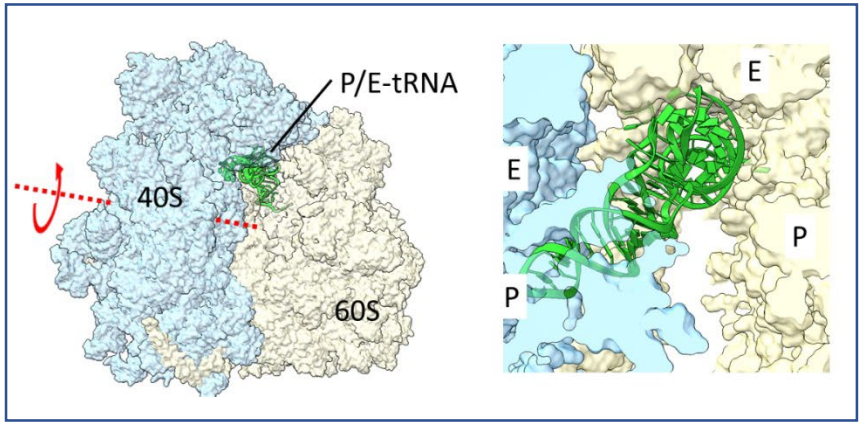
Hamitouche & Jonic, *Front Mol Biosci 2022*

DeepHEMNMA using cryo-EM data of yeast 80S ribosome-tRNA EMPIAR-10016 data (Svidritskiy et al., 2014)

Data available:

- Set of around 90,000 particle images was classified in 5 classes with FREALIGN
- Two classes with 23,726 and 22,369 images resulted in two 3D reconstructions (EMD-5976 & EMD-5977, about 6.3 Å resolution)
- EMD-5977: “non-rotated”, two tRNA in the classical P-P and E-E states
- EMD-5976: inter-subunit rotation of around 9°, one tRNA in a hybrid P/E state
- Two atomic models derived (PDB-3J77 from EMD-5976; PDB-3J78 from EMD-5977)

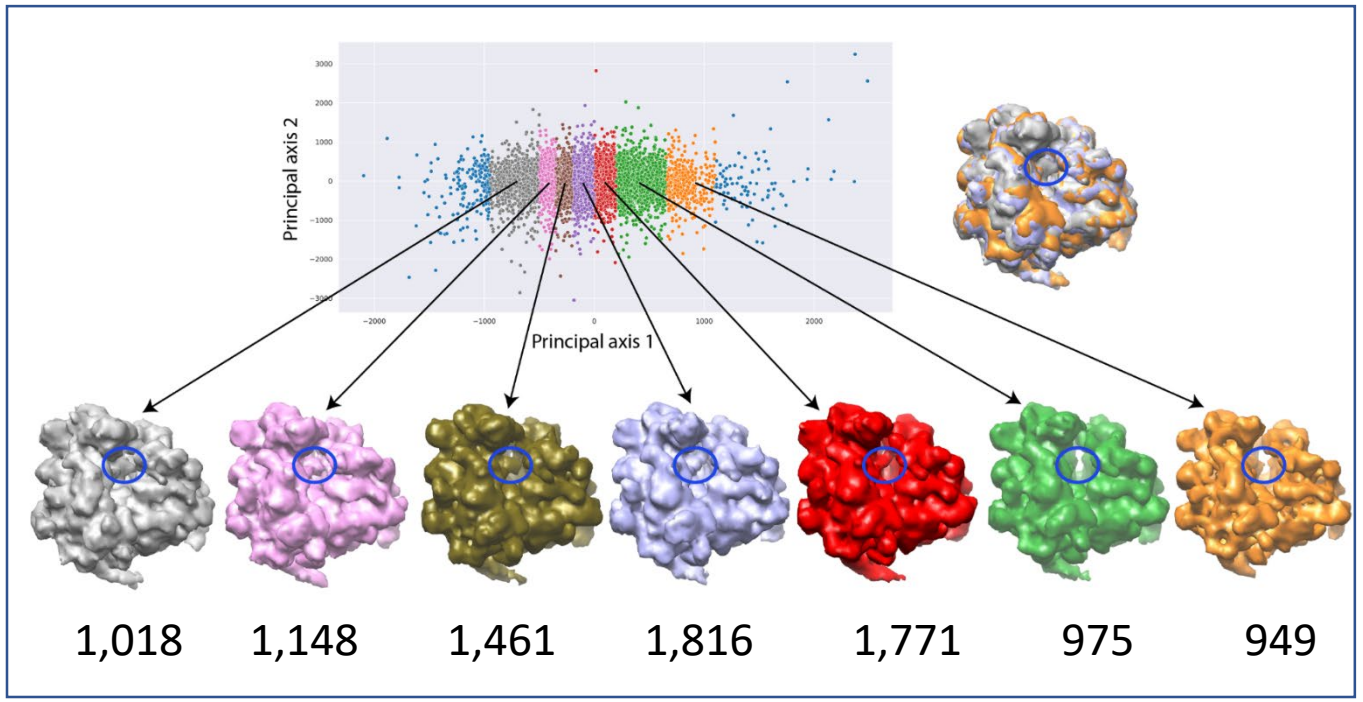


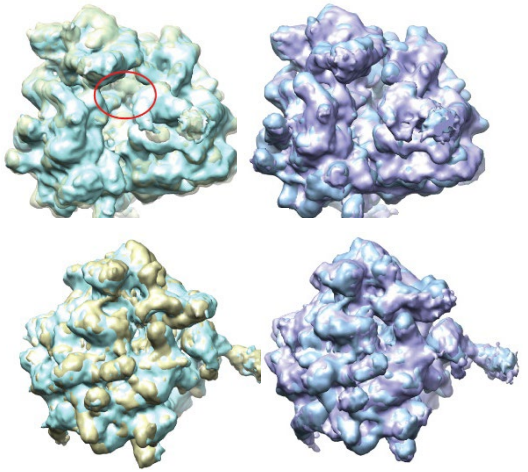


Hamitouche & Jonic, *Front Mol Biosci 2022*

Experiment with DeepHEMNMA using EMPIAR-10016 data:

- Use of 46,095 images (mixture of 2 classes, with 1 tRNA and 2 tRNAs)
- Downscaling from 360 × 360 pixels (pixel: 1.05 Å) to 128 × 128 pixels (pixel: 2.95 Å)
- Splitting images into 3 subsets: training (32,000), validation (2,000), and test (12,095)
- PCA space split quasi-uniformly to get at least 900 images per group
- Reconstruction of 7 maps with variable degree of inter-subunit rotation and presence/absence of the second tRNA





Superposition with reconstructions from two FREALIGN classes

Processing time (wall-clock):
DeepHEMNMA *vs.* HEMNMA

Synthetic data, 3 normal modes, DeepHEMNMA faster than HEMNMA more than 40 times

HEMNMA 160 INTEL 2.6 GHz CPU cores	1 image 1 core	20,000 images 160 cores (time per core)	10^6 images 160 cores (time per core)
256x256 pixels	8 min	15.6 h	800 h
128x128 pixels	4 min	7.7 h	400 h

Training 4 NVIDIA V100 / 5120 CUDA cores	6,000 images	14,000 images	50,000 images
256x256 pixels	15 h	28 h	75 h
128x128 pixels	11 h	19 h	55 h

Prediction 1 NVIDIA V100 / 5120 CUDA cores	2 images	2,000 images	50,000 images	10^6 images
256x256 pixels	36 ms	0.3 min	7.5 min	2.5 h
128x128 pixels	6 ms	0.2 min	5 min	1.7 h

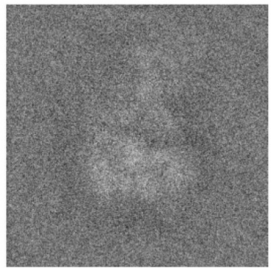
But, large amplitudes of NMs may induce model distortions

- NMs can be combined with MD simulation to avoid model distortions and accelerate simulations
  (NMMD in MDSPACE)

Hamitouche & Jonic, *Front Mol Biosci 2022*

DeepHEMNMA tests with synthetic data

Adenylate kinase Chain A (PDB:4AKE)
Pixel size: 0.325 Å
Size: 256 x 256 pix



SNR=0.1,
Defocus = -0.5 µm

Random parameters:

- 3 Euler angles and 2 shifts
- Amplitudes of 3 normal modes (modes 7-9):

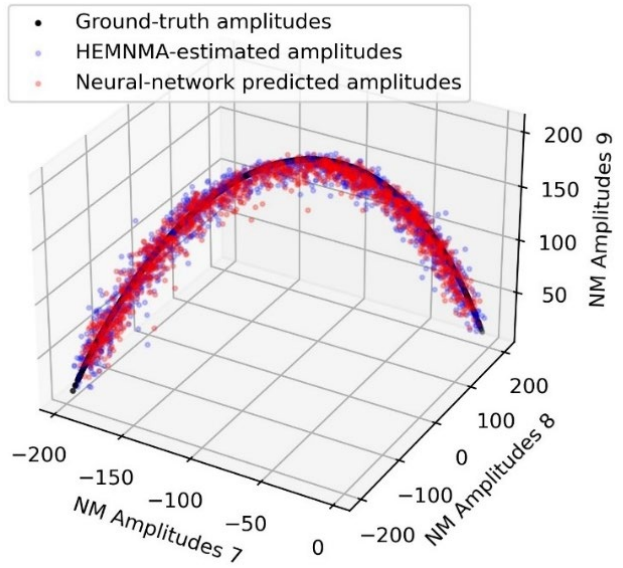
$$q_7(r) = -200 \cdot r, \quad q_8(r) = 200 \cdot \sin(\pi \cdot r),$$
$$q_9(r) = 200 \cdot \cos(\pi \cdot r),$$

r : random uniform, between 0 and 1

Training: 14,000 images

Validation: 2,000 images

Prediction: 2,000 images

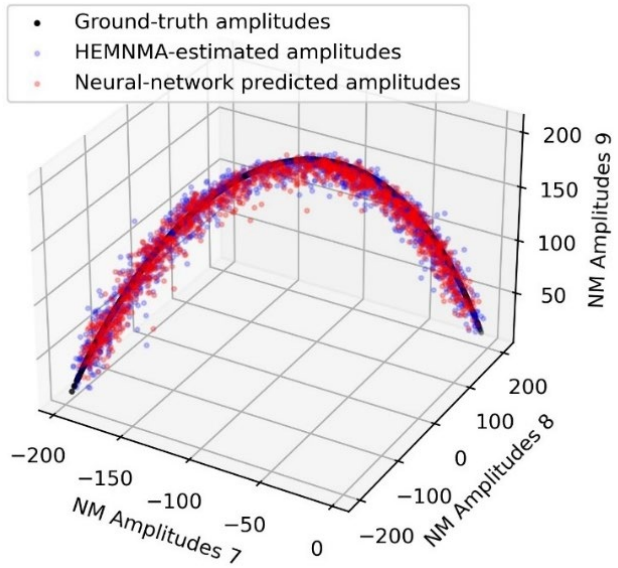


128 x 128 pix images

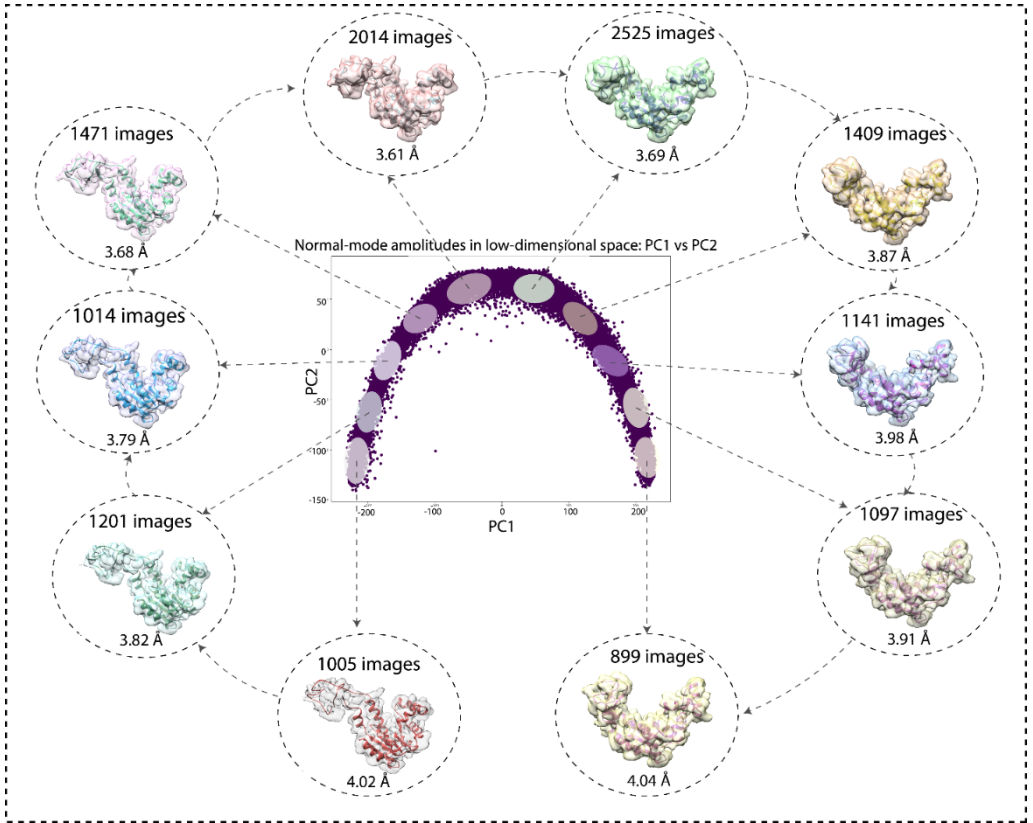
Amplitude error
Pred. *vs.* GT: 7.5
RMSD: 0.4 Å

Hamitouche & Jonic, *Front Mol Biosci 2022*

Errors	Normal-mode amplitudes							Angles [°]		Shifts X [Å]		Shifts Y [Å]	
	Mean over modes 7-9	Mode 7		Mode 8		Mode 9							
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Predicted vs. Ground-truth	7.5	5.4	6.5	8.2	9.2	8.9	10.5	2.5	3.3	0.2	0.1	0.2	0.1
Predicted vs. HEMNMA	6.9	5.4	6.7	7.3	9.0	7.9	9.6	1.9	3.4	0.2	0.1	0.2	0.1
HEMNMA vs. Ground-truth	6.6	5.7	8.4	6.2	7.2	7.8	7.2	1.0	0.9	0.2	0.2	0.2	0.2



RMSDs	Normal-mode amplitudes [Å]		Angles [Å]		Shifts [Å]	
	Mean	SD	Mean	SD	Mean	SD
Predicted vs. Ground-truth	0.4	0.2	0.9	1.0	0.3	0.2



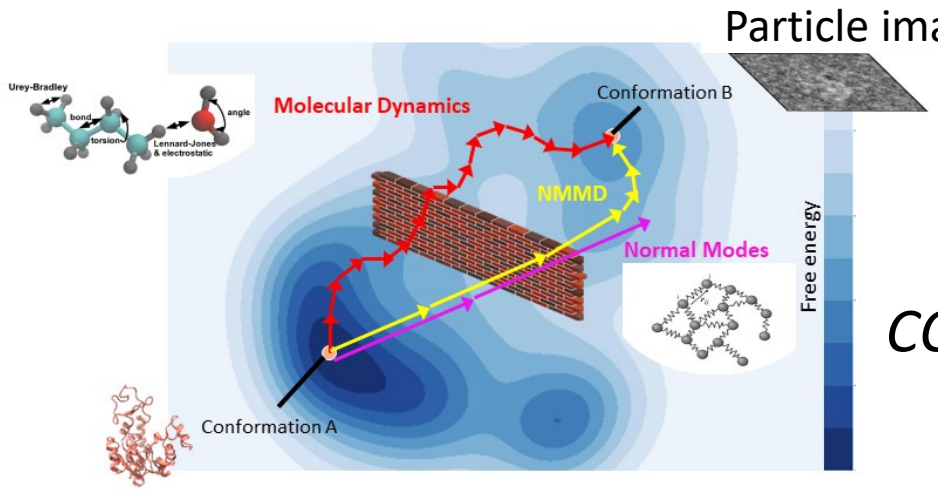
DeepHEMNMA tests with synthetic data

Hamitouche & Jonic, *Front Mol Biosci 2022*

MDSPACE: First approach for 3D-to-2D flexible fitting based on MD simulation
Based on NMMD for the purpose of speed and obtaining high quality models

$$r(t) = \underbrace{q(t) \cdot A}_{\text{NM-based displacement}} + \underbrace{x(t)}_{\text{Atomic displacement}} + r_0$$

Vuillemot et al., J Mol Biol 2023



CC : Pearson correlation coefficient

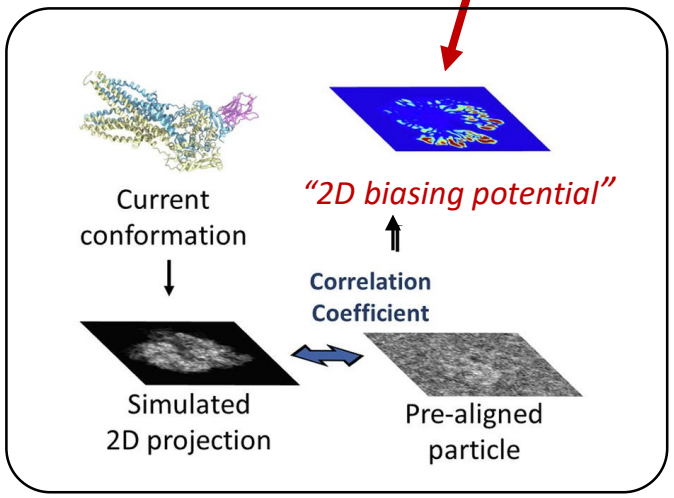
$$CC = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2}\sqrt{\sum_i (y_i - y_m)^2}}$$

$$U = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[1 + cos(n\phi - \delta)] + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u(u - u_0)^2 + \sum_{nonbonded} \left(\epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_r r_{ij}} \right)$$

Biasing potential

$$+ k\,(1 - CC)$$

Force constant

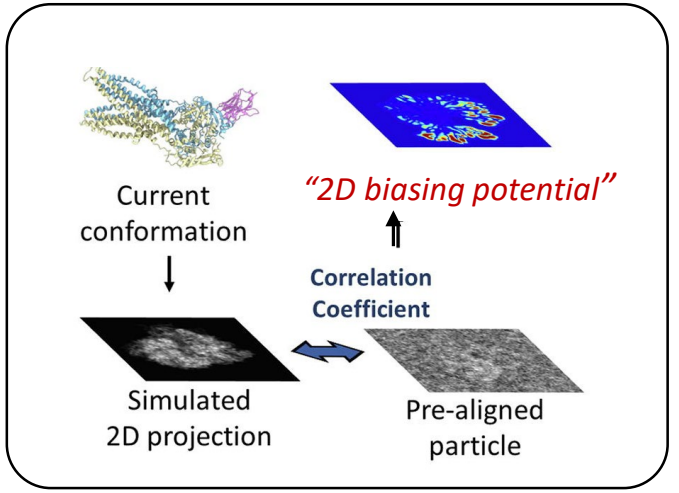


MDSPACE: First approach for 3D-to-2D flexible fitting based on MD simulation (Cont'd)

$$p_{sim}(x,y) = \sum_{n=1}^{N} p_{sim}^{n}(x,y),$$

$$p_{sim}^{n}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left((x-r_x^n)^2+(y-r_y^n)^2\right)}$$

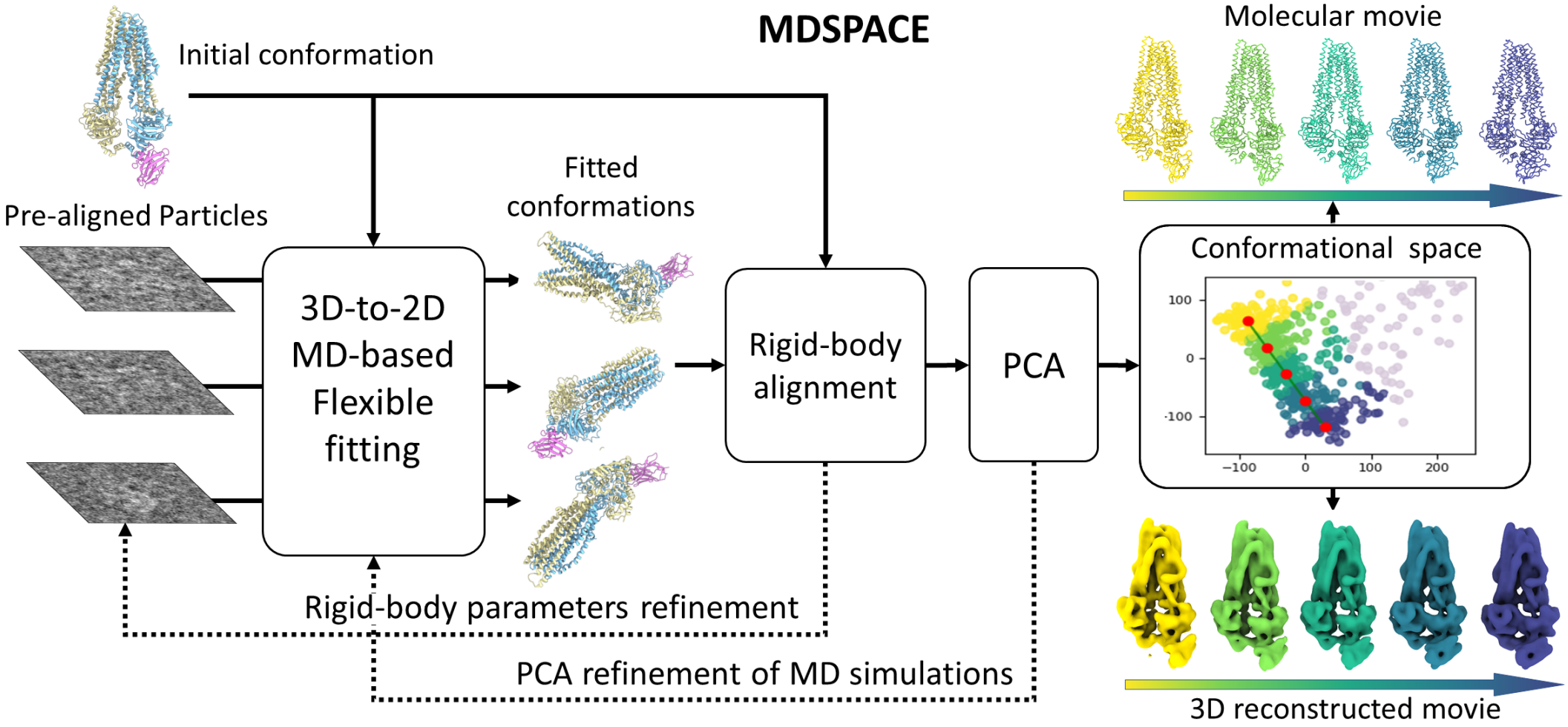
$$CC = \frac{\sum_{i,j} p_{sim}(i,j)p_{exp}(i,j)}{\sqrt{\sum_{i,j} p_{exp}^2(i,j) \sum_{i,j} p_{sim}^2(i,j)}},$$



- NMMD is faster than MD alone, thanks to normal modes
- MDSPACE produces high-quality models (better than those obtained with normal modes only)
- Selection of normal modes is less critical in MDSPACE than in HEMNMA

Vuillemot et al., J Mol Biol 2023

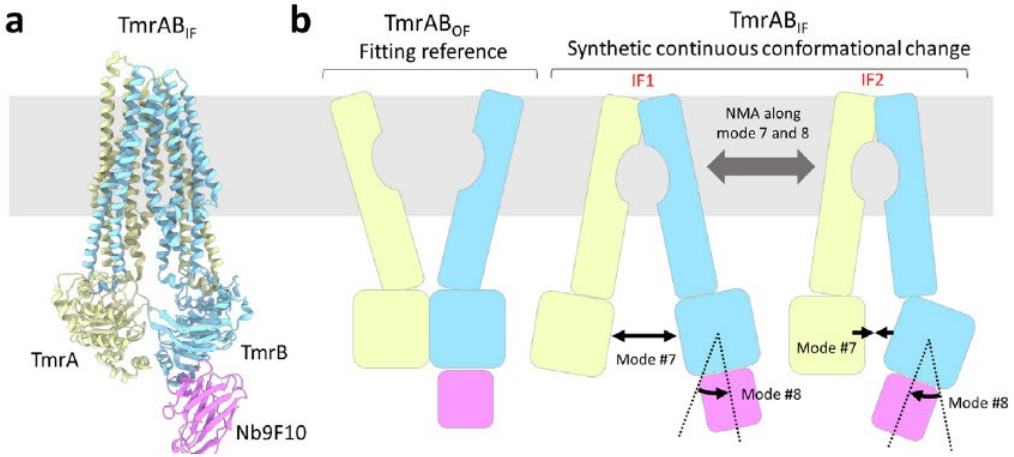
Iterative MDSPACE procedure for refining conformational space

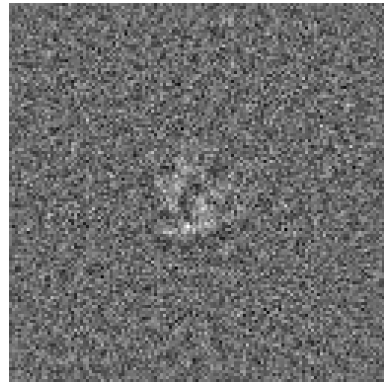


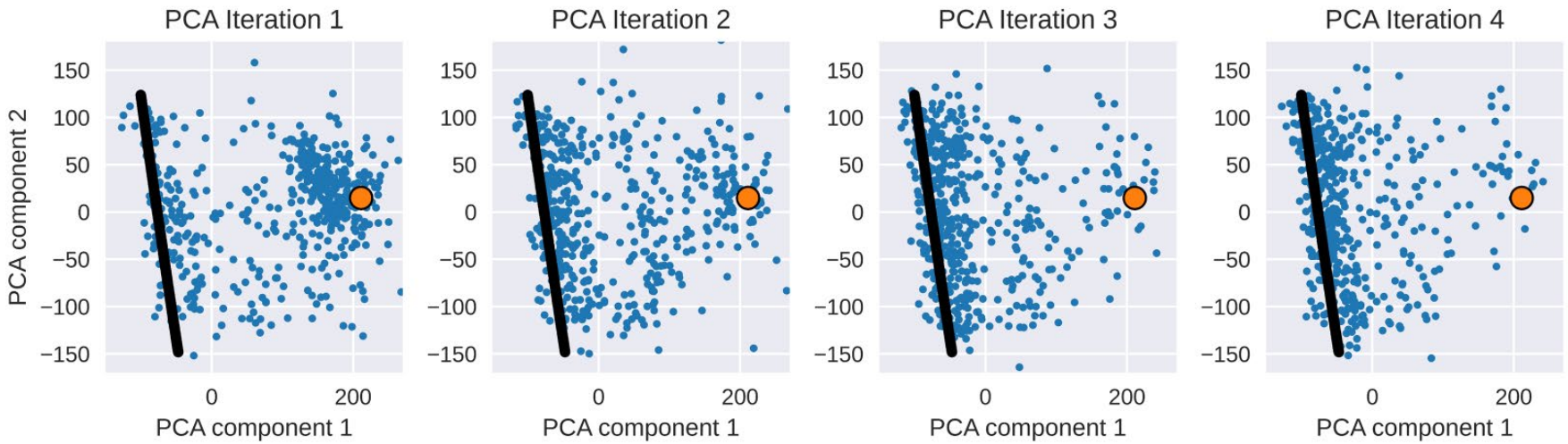
The initial rigid-body alignment parameters are refined during the fitting in each MDSPACE iteration and updated before each new iteration *Vuillemot et al., J Mol Biol 2023*

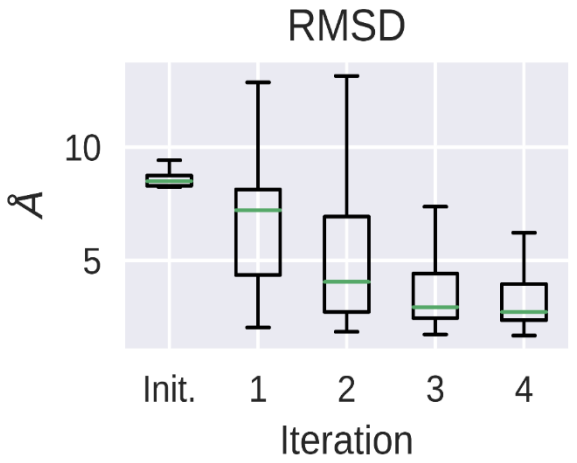
MDSPACE validation with synthetic data of an ABC transporter

Synthetic dataset of 3 000 particle images of ABC transporter PDB-6RAF (Hoffman et al. *Nature* 2019)







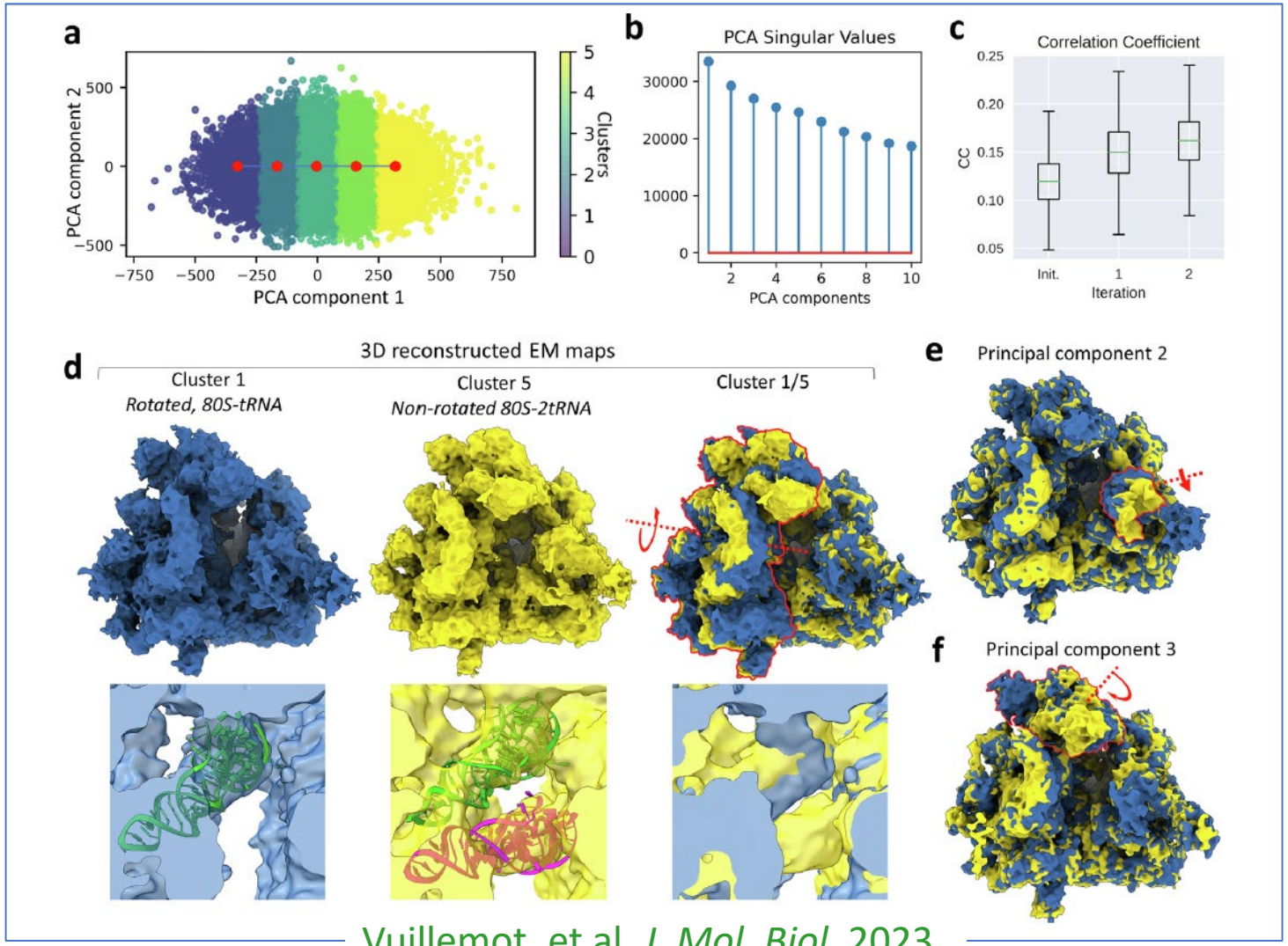


RMSD between the ground truth models and the produced models

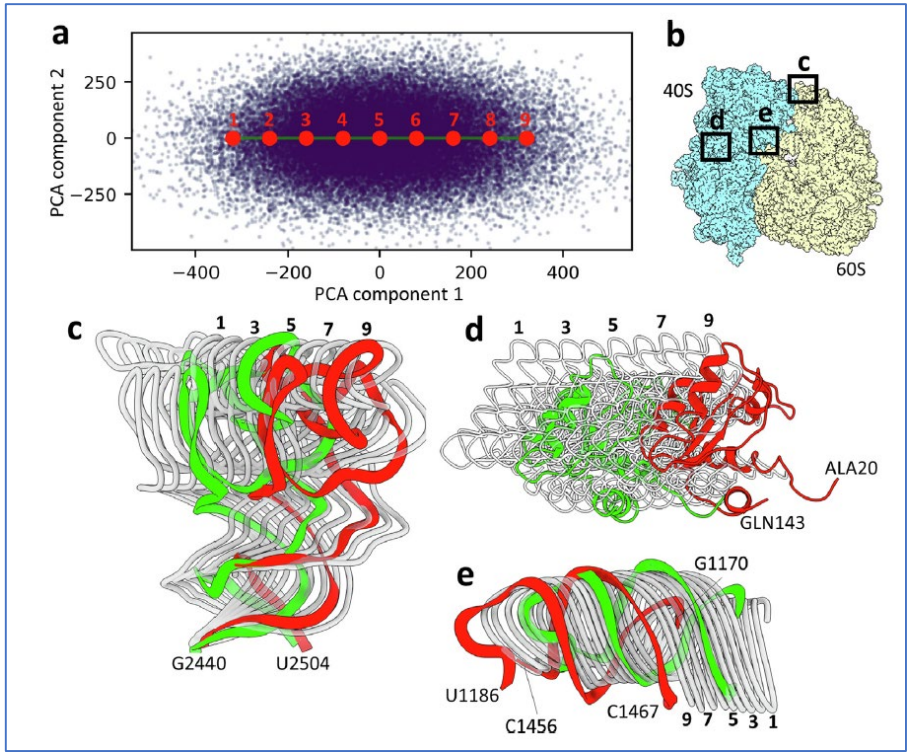
Vuillemot, et al. *J. Mol. Biol.* 2023

MDSPACE analysis of EMPIAR-10016 set (80S ribosome-tRNA complexes)

46,095 particles, 180^2 pixels, 3 days on 640 CPU cores (Intel Xeon 6248 processors, 2.5 GHz)



Vuillemot, et al. *J. Mol. Biol.* 2023



Green : 80S-tRNA (rotated, hybrid P/E tRNA)

Red : 80S-2tRNA (nonrotated, P-P & E-E tRNAs)

Initial model : Cα Gō model of 80S-tRNA

Two iterations of MDSPACE

- Iter 1: 10,000 particles, 10 NMs
- Iter 2: 46,095 particles, 3 PCs

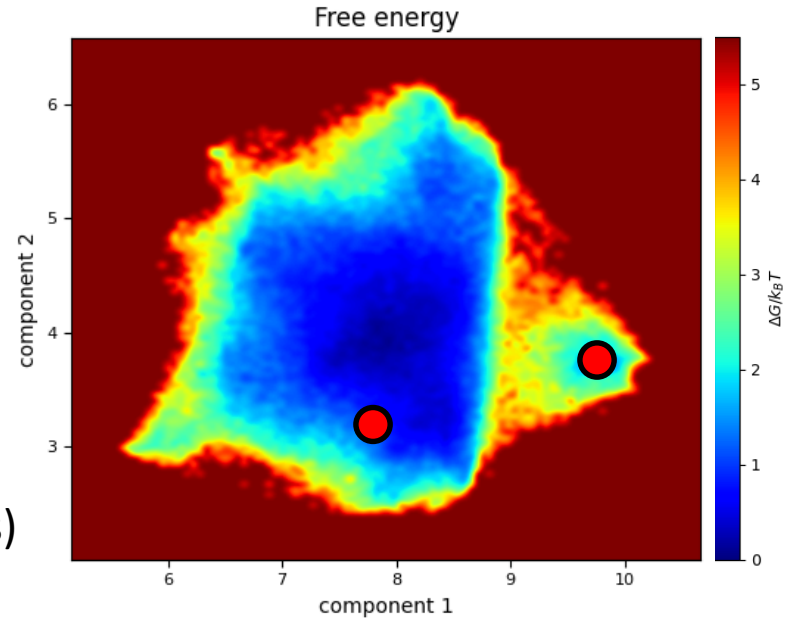
30 ps NMMD simulations

Force constant: 10,000 kcal/mol

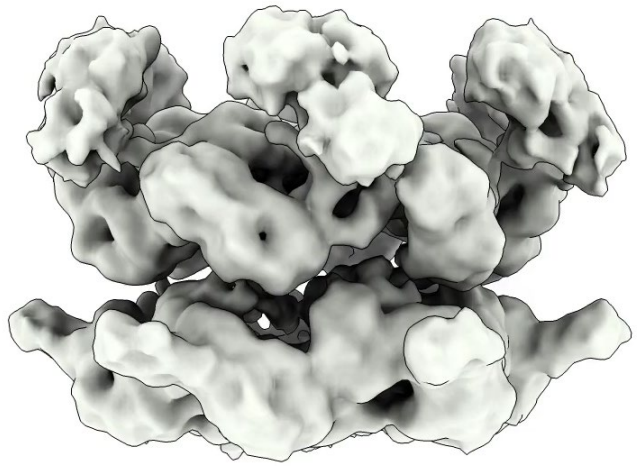
MDSPACE analysis of ATPase p97 dataset

MDSPACE analysis of p97 dataset of 274,640 particles

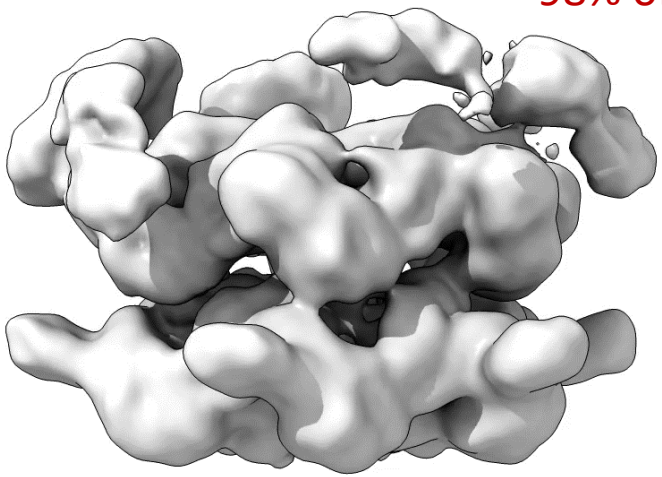
- Initial model: Cα Gō model of PDB:5FTN (N domains up)
- 2 iterations of MDSPACE
  - Iter 1: 5,000 particles, 5 NMs
  - Iter 2: 274,640 particles, 10 PCs
- 50 ps NMMD simulations (time step : 1 fs)

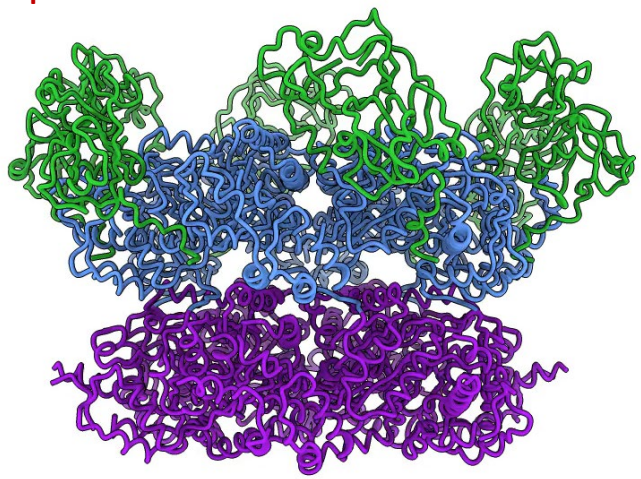


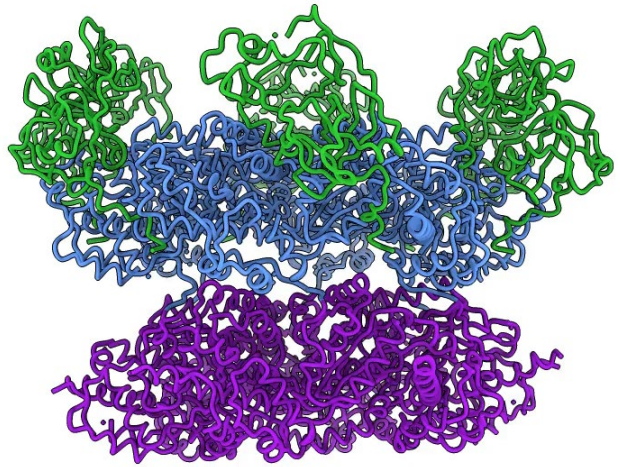
2% of the particles



98% of the particles







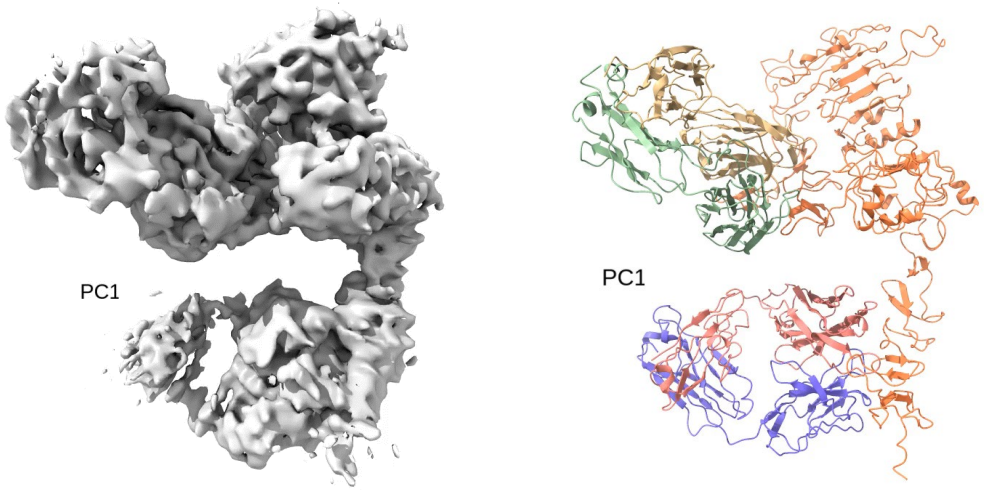
Manuscript in preparation (PhD thesis of Rémi Vuillemot 2023, Collab. Rouiller team, The University of Melbourne)

MDSPACE in brief

- First method for analyzing conformational variability of cryo-EM images using MD-based fitting
- Obtains the conformational space at atomic level
- Tested using synthetic and experimental data

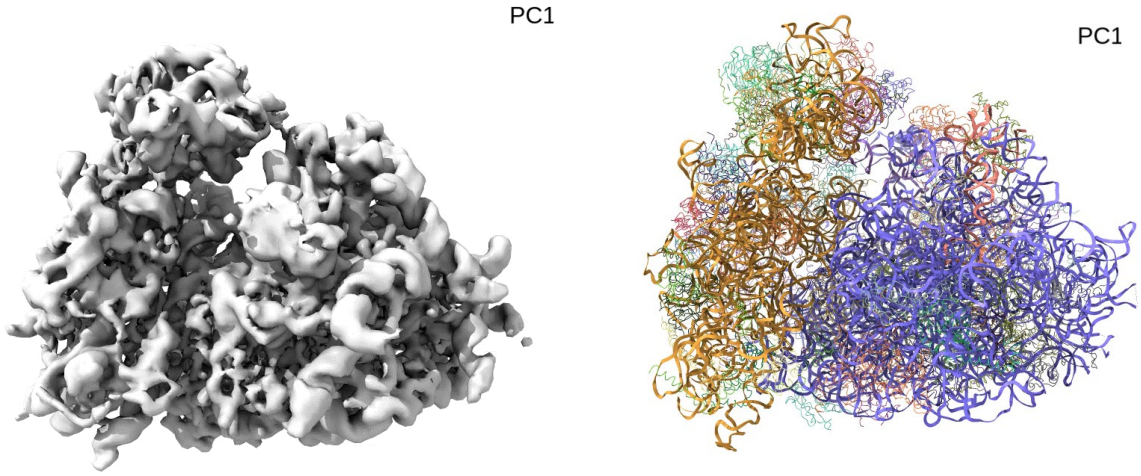
Rémi Vuillemot

Other examples of studies using MDSPACE ...



HER2-Trastuzumab Fab-Pertuzumab Fab (164.51 kDa)

Collab. S. Bressaneli & R. Ruedas (I2BC, Gif-sur-Yvette)



Human 80S ribosome (3.2 MDa)

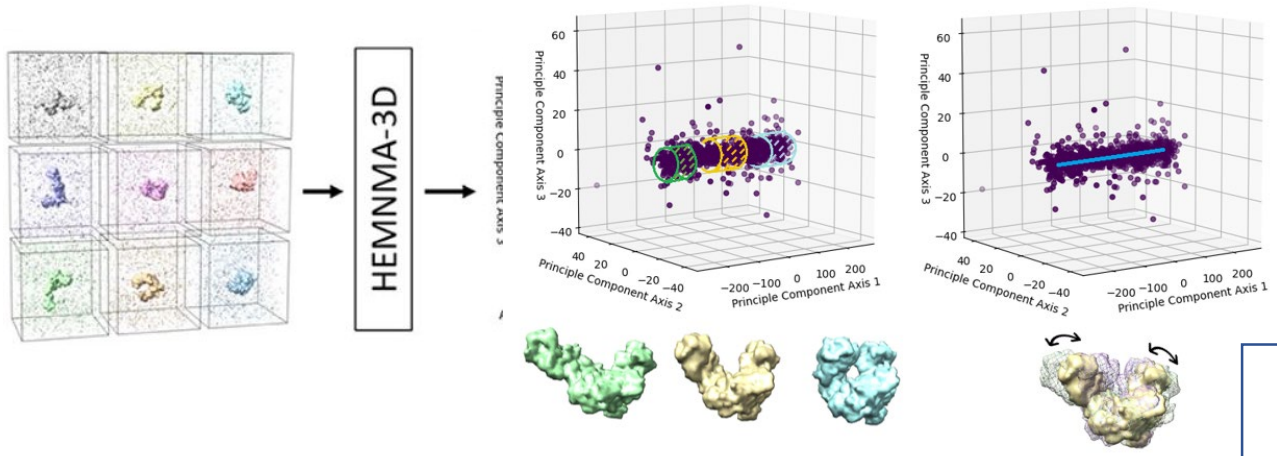
Collab. B. Klaholz & L. Fréchin (IGBMC, Strasbourg)

3D-to-3D flexible fitting of large sets of cryo-ET subtomograms

Extension of new hybrid approaches to large volumetric datasets

3D-to-3D fitting of large sets of EM maps or subtomograms

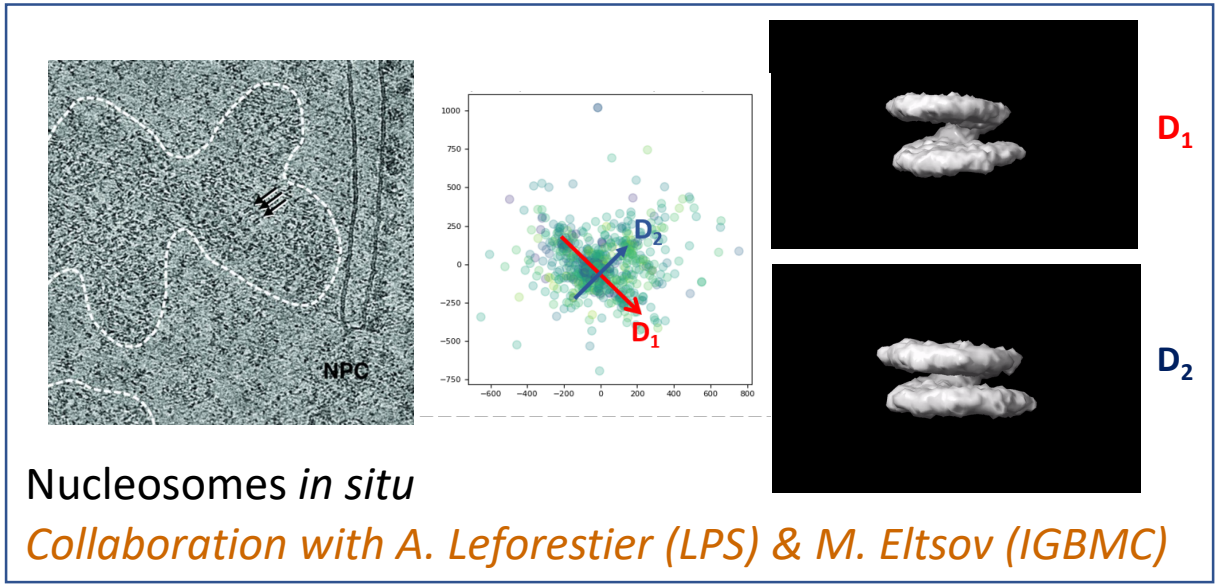
HEMNMA-3D: Extension of HEMNMA method to 3D-to-3D flexible fitting
(fitting of an atomic model or an EM map to sets of subtomograms using normal modes)



Mohamad Harastani
PhD, Oct. 2022
(now Pdoc IGBMC, Illkirch)

Recalling that :

- NMs are fast
- But, large amplitudes of NMs may induce model distortions
- NMs be combined with MD simulation to avoid model distortions and accelerate simulations (NMMD in MDTOMO)



Nucleosomes *in situ*

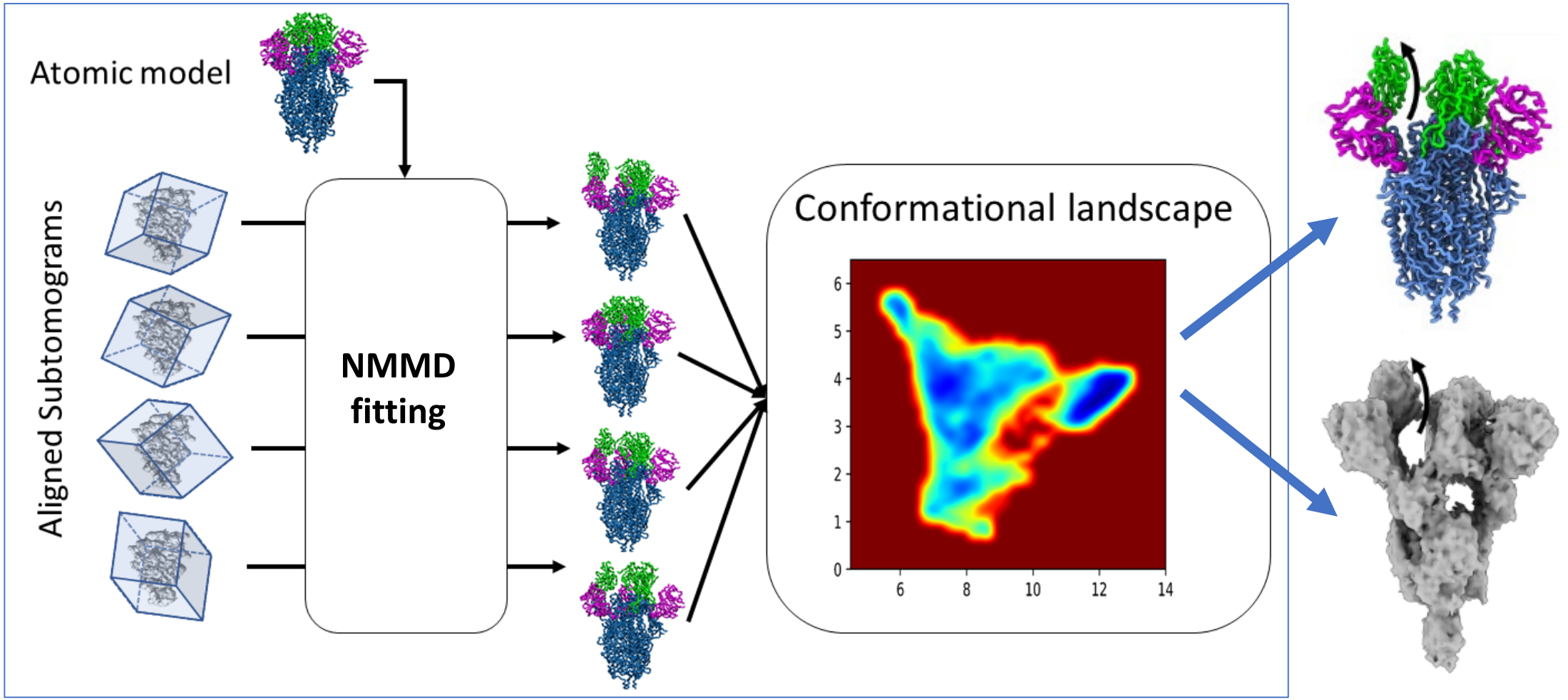
Collaboration with A. Leforestier (LPS) & M. Eltsov (IGBMC)

Harastani et al., Front Mol Biosci 2021

MDTOMO : NMMD simulation integrated into analysis of subtomograms

The same strategy of refining the initial rigid-body alignment as in MDSPACE

Vuillemot, et al. *Sci Rep* 2023

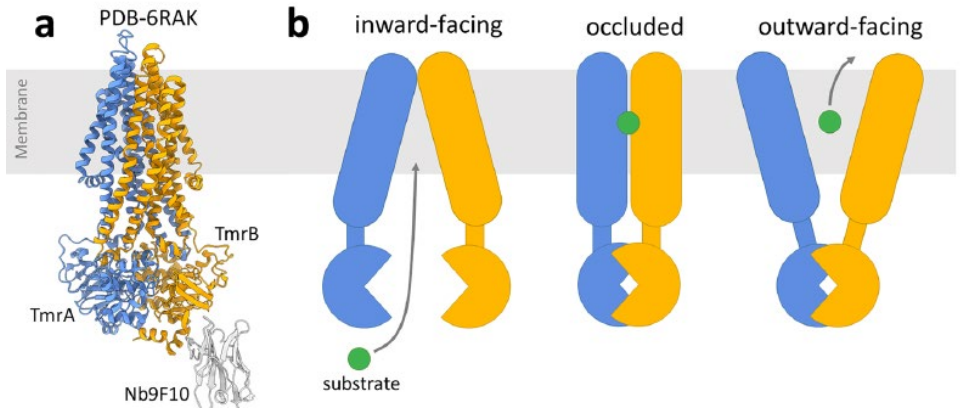


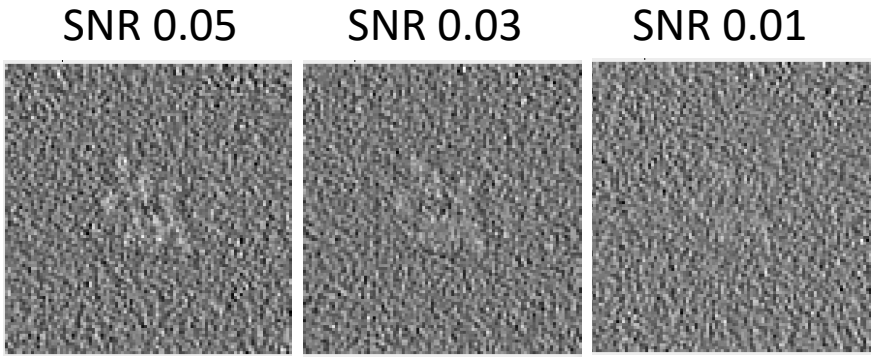
Atomic models

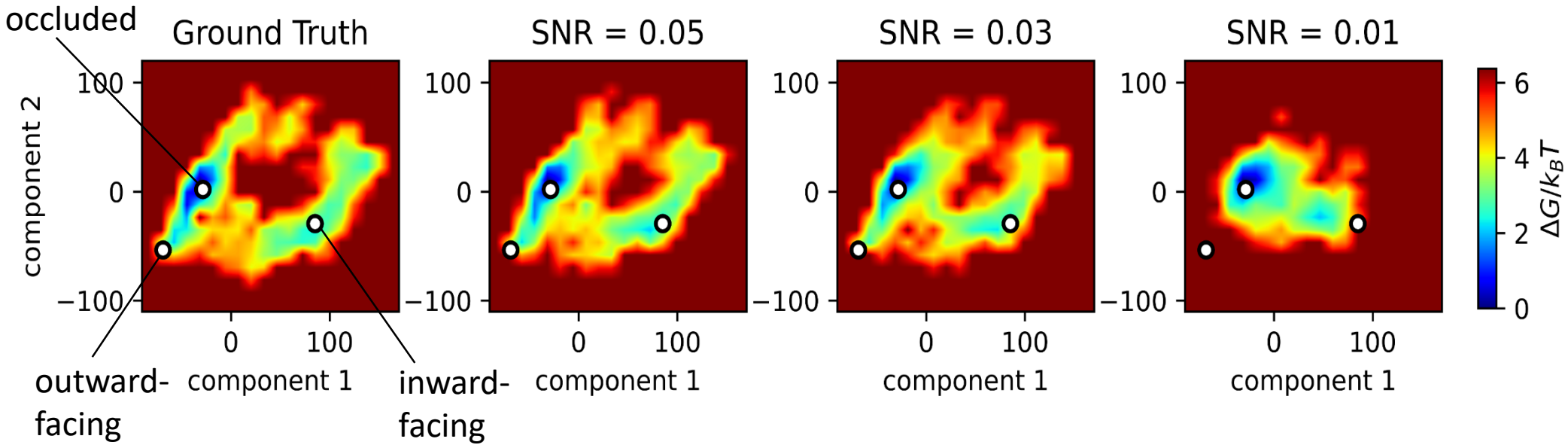
Subtomogram averages

MDTOMO validation using synthetic dataset of an ABC transporter

3 synthetic datasets of 3,000 subtomograms of ABC transporter PDB-6RAK (Hoffman et al. *Nature* 2019)



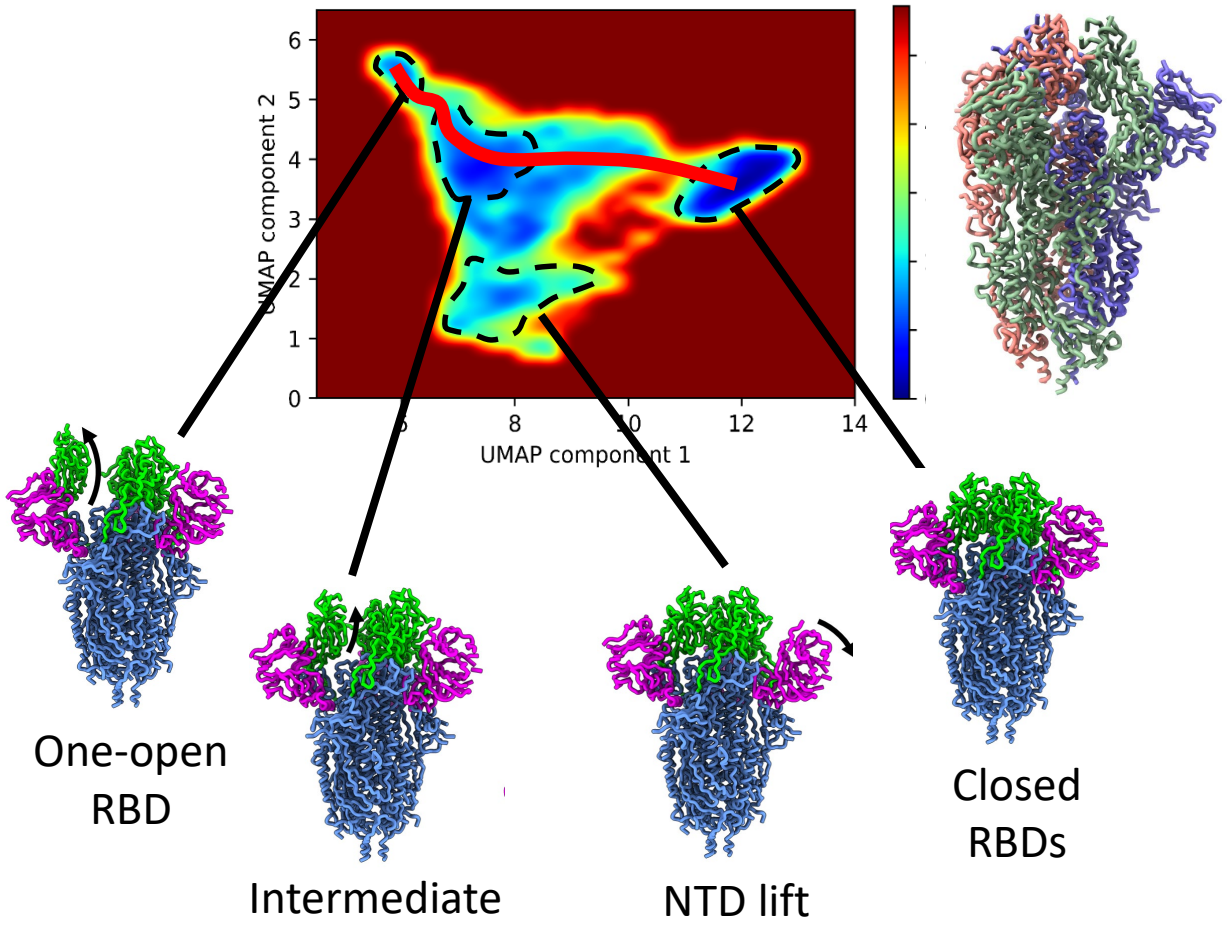


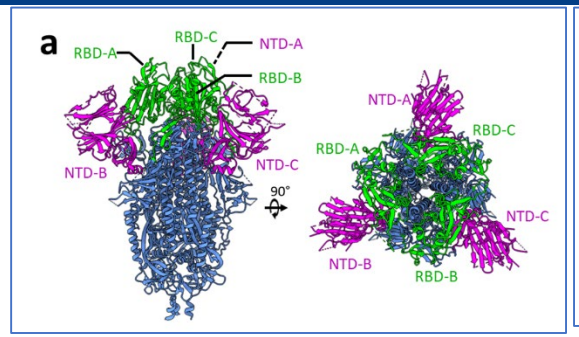


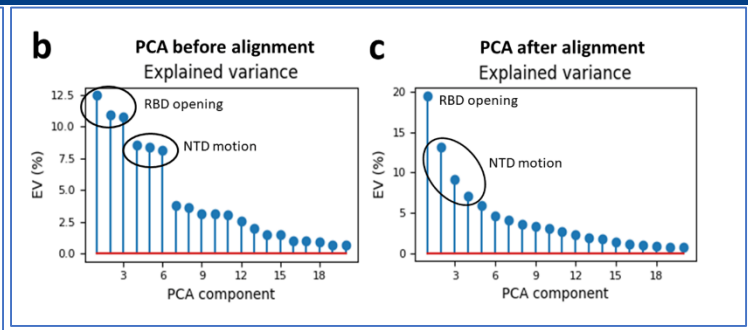
Vuillemot, et al. *Sci Rep* 2023

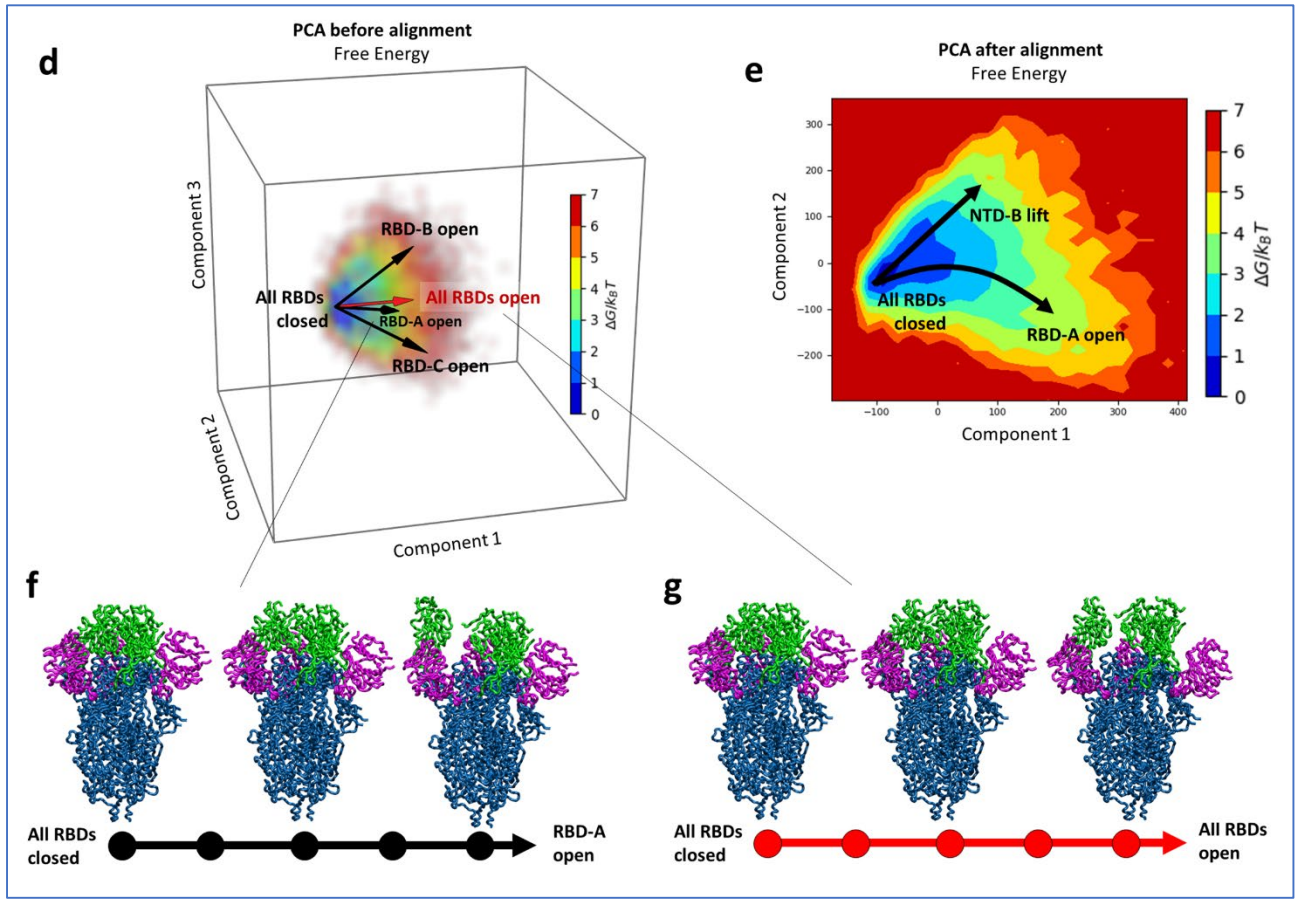
MDTOMO analysis of EMPIAR-10453 set (SARS-CoV-2 spike subtomograms)

20 080 subtomograms *(B. Turonova, Max Planck Inst. Biophys)*
Cα Gō model of PDB:6VXX (closed RBDs), 100 ps NMMD simulations, Force constant: 7000 kcal/mol, 1 iter of MDTOMO, 17.8 h on 320 CPU cores (Xeon 6248 processors), 128^3 voxels

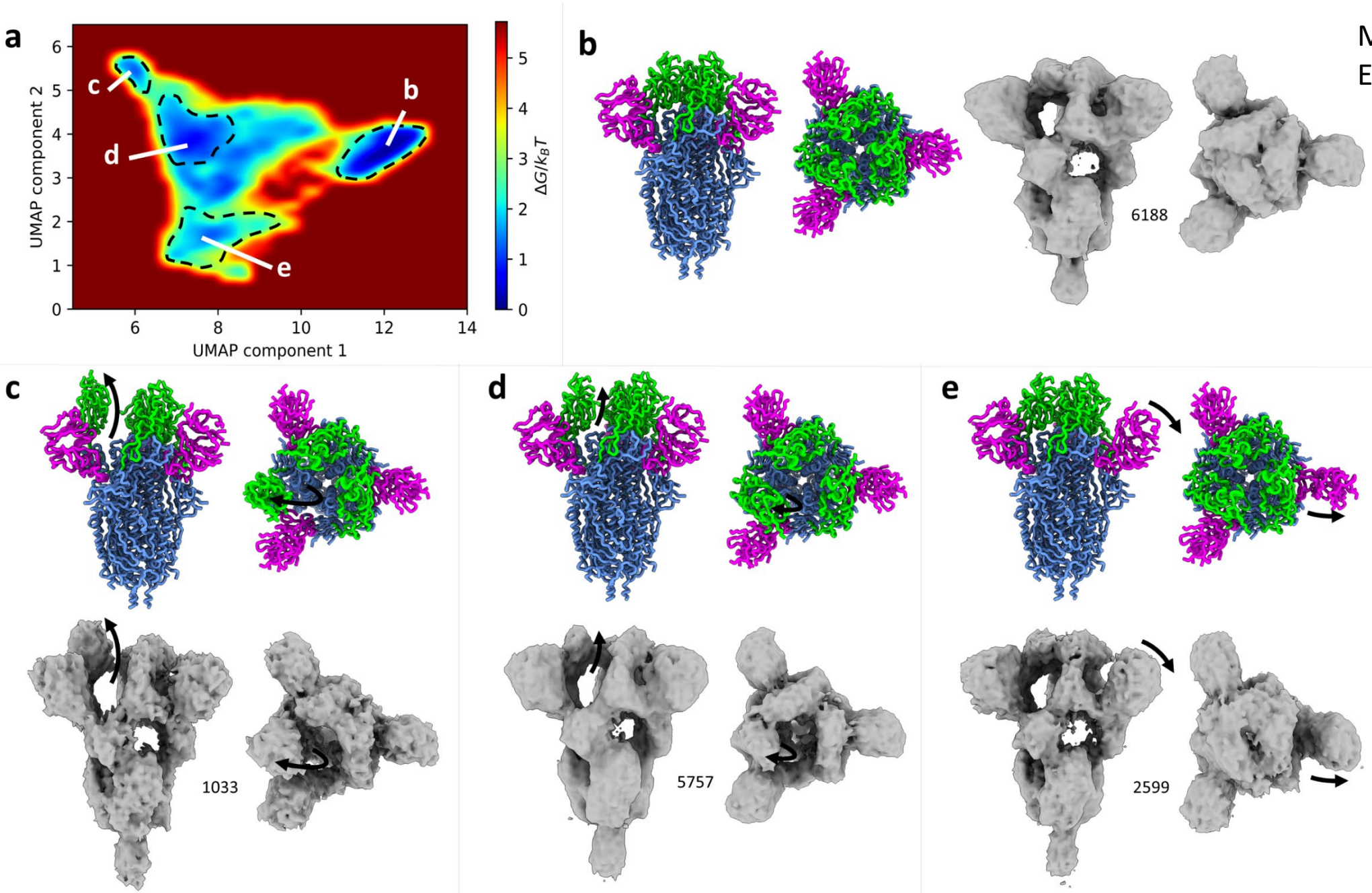








Vuillemot, et al. *Sci Rep* 2023



MDTOMO analysis of EMPIAR-10453 data

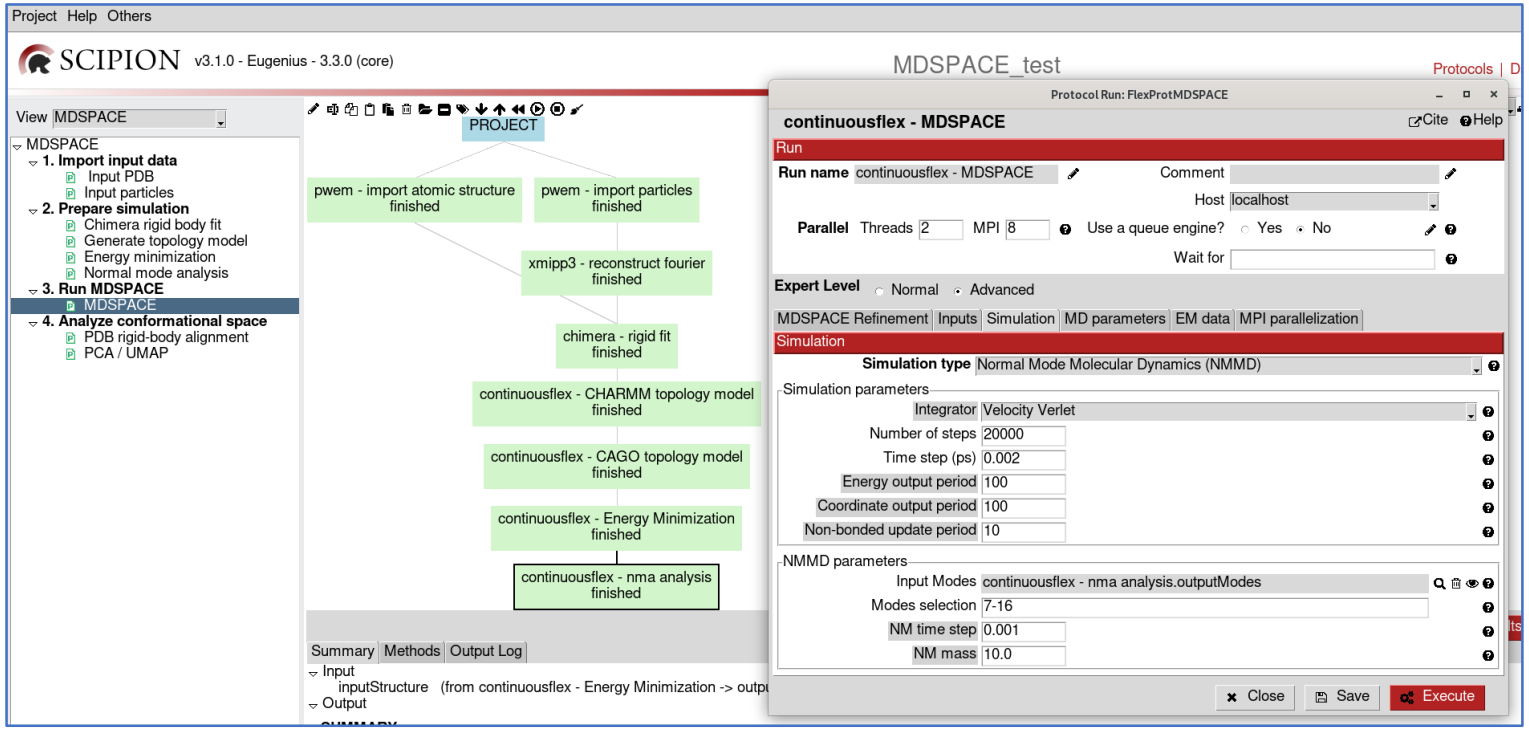
Vuillemot, et al. *Sci Rep* 2023

Open-source ContinuousFlex software package (GitHub, Scipion)

Journal of Structural Biology
Volume 214, Issue 4, December 2022, 107906

ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data

Mohamad Harastani, Rémi Vuillemot, Ilyes Hamitouche, Nima Barati Moghadam, Slavica Jonic



SPA:
HEMNMA / DeepHEMNMA
MDSPACE

Cryo-ET:
HEMNMA-3D
MDTOMO
TomoFlow

MD/NMA:
Genesis for MD *(Kobayashi et al., JCC 2017)*
ElNemo for NMA *(Suhre et al., NAR 2004)*

- **Automated tests**
- **Automated workflow templates**
- **Open test data** (Zenodo)

preprints.org > computer science and mathematics > software > doi: 10.20944/preprints202311.0460.v1

Preprint **Article** *Version 1* **Preserved in Portico** **This version is not peer-reviewed**

MDSPACE and MDTOMO Software for Extracting Continuous Conformational Landscapes from Datasets of Single Particle Images and Subtomograms Based on Molecular Dynamics Simulations: Latest Developments in ContinuousFlex Software Package

Rémi Vuillemot , Mohamad Harastani , Ilyes Hamitouche , Slavica Jonic *

Version 1 : Received: 6 November 2023 / Approved: 7 November 2023 / Online: 7 November 2023 (14:56:57 CET)

Take-home messages

- Several methods developed for analyzing conformational variability in cryo-EM single particle images and cryo-ET subtomograms
- Hybrid methods integrate dynamics simulation into data analysis
- Reference model required (atomic model or EM map)
- For NMA-based methods, the reference can be an EM map (HEMNMA, HEMNMA-3D)
- For MD-based methods, the reference should be atomic model (MDSPACE, MDTOMO)
  - Advantage of using a reference atomic model: atomic-scale conformational landscape
- Require less data to interpret conformational heterogeneity than other methods
- Can be used in combination with supervised deep learning to speed up analysis (e.g., DeepHEMNMA)

Summary: 3D-to-2D flexible fitting of each particle image with a given model

HEMNMA *(Jin et al., Structure 2014)*

- First approach for analyzing continuous conformational heterogeneity in large sets of particle images, which integrates simulated motion directions of a given model into image analysis
- Simulations performed using NMA

DeepHEMNMA (*Hamitouche & Jonic, Front Mol Biosci 2022*)

- First approach integrating motion simulation into deep learning (training data obtained with HEMNMA)
- Accelerates HEMNMA

MDSPACE (*Vuillemot et al., J Mol Biol 2023*)

- First approach integrating MD simulation into SPA for continuous conformational heterogeneity analysis, where each image is analyzed independently of others
- Uses NMMD (MD simulation empowered with normal modes to speed up the analysis while preserving high quality resulting models)

Summary: 3D-to-3D flexible fitting of each subtomogram with a given model

HEMNMA-3D *(Harastani et al., Front Mol Biosc 2021)*

- First approach for analyzing continuous conformational heterogeneity in a set of cryo-ET subtomograms, which integrates simulated motion directions of a given model into subtomogram analysis
- Extension of HEMNMA (simulations performed using NMA)

MDTOMO (*Vuillemot et al., Sci Rep 2023*)

- First approach integrating MD simulation into subtomogram analysis for continuous conformational heterogeneity studies
- Uses NMMD (MD simulation empowered with normal modes to speed up the analysis while preserving high quality resulting models)

References

- Vuillemot R, Harastani M, Hamitouche I, **Jonic S.** *MDSPACE and MDTOMO Software for Extracting Continuous Conformational Landscapes from Datasets of Single Particle Images and Subtomograms Based on Molecular Dynamics Simulations: Latest Developments in ContinuousFlex Software Package*. Preprints **2023**, 2023110460. https://doi.org/10.20944/preprints202311.0460.v1
- Vuillemot R, Rouiller I, **Jonic S**. *MDTOMO method for continuous conformational variability analysis in cryo electron subtomograms based on molecular dynamics simulations*. Sci Rep. **2023**; 13 :10596. [Journal]
- Vuillemot R, Mirzaei A, Harastani M, Hamitouche I, Frechin L, Klaholz BP, Miyashita O, Tama F, Rouiller I, **Jonic S**. *MDSPACE: Extracting Continuous Conformational Landscapes from Cryo-EM Single Particle Datasets Using 3D-to-2D Flexible Fitting based on Molecular Dynamics Simulation*. J Mol Biol. **2023**;435:167951. [Journal][HAL]
- Harastani M, Vuillemot R, Hamitouche I, Moghadam NB, **Jonic S**. *ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data*. J Struct Biol. **2022**;214:107906. [Journal][HAL]
- Hamitouche I, **Jonic S**. *DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images*. Front Mol Biosci. **2022**;9:965645. [Journal]
- Vuillemot R, Miyashita O, Tama F, Rouiller I, **Jonic S**. *NMMD: Efficient Cryo-EM Flexible Fitting Based on Simultaneous Normal Mode and Molecular Dynamics atomic displacements*. J Mol Biol. **2022**;434:167483. [Journal][HAL]
- Harastani M, Eltsov M, Leforestier A, **Jonic S**. *HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes*. Front Mol Biosci. **2021**;8:663121. [Journal]
- Jin Q, Sorzano CO, de la Rosa-Trevín JM, Bilbao-Castro JR, Núñez-Ramírez R, Llorca O, Tama F, and **Jonic S.** *Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes*. Structure 2014;22:496-506 **[**Journal]

Thank You