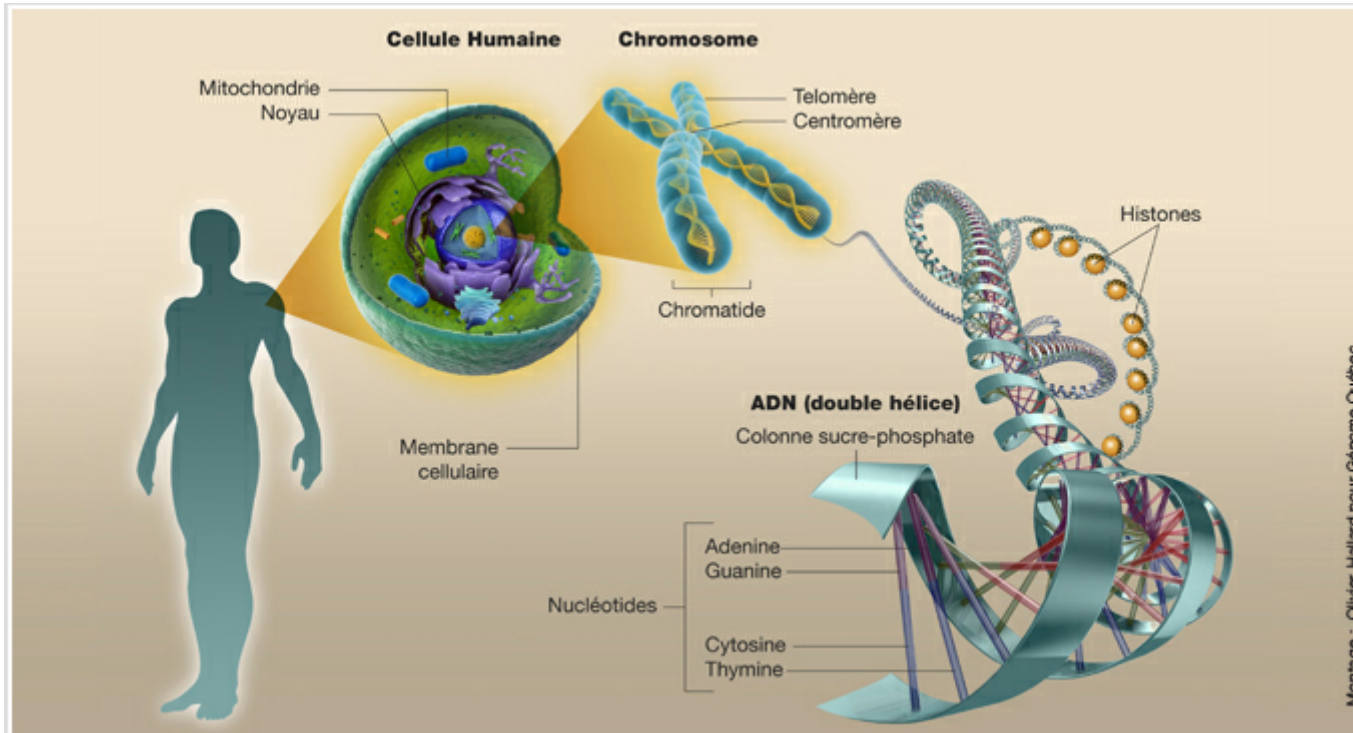
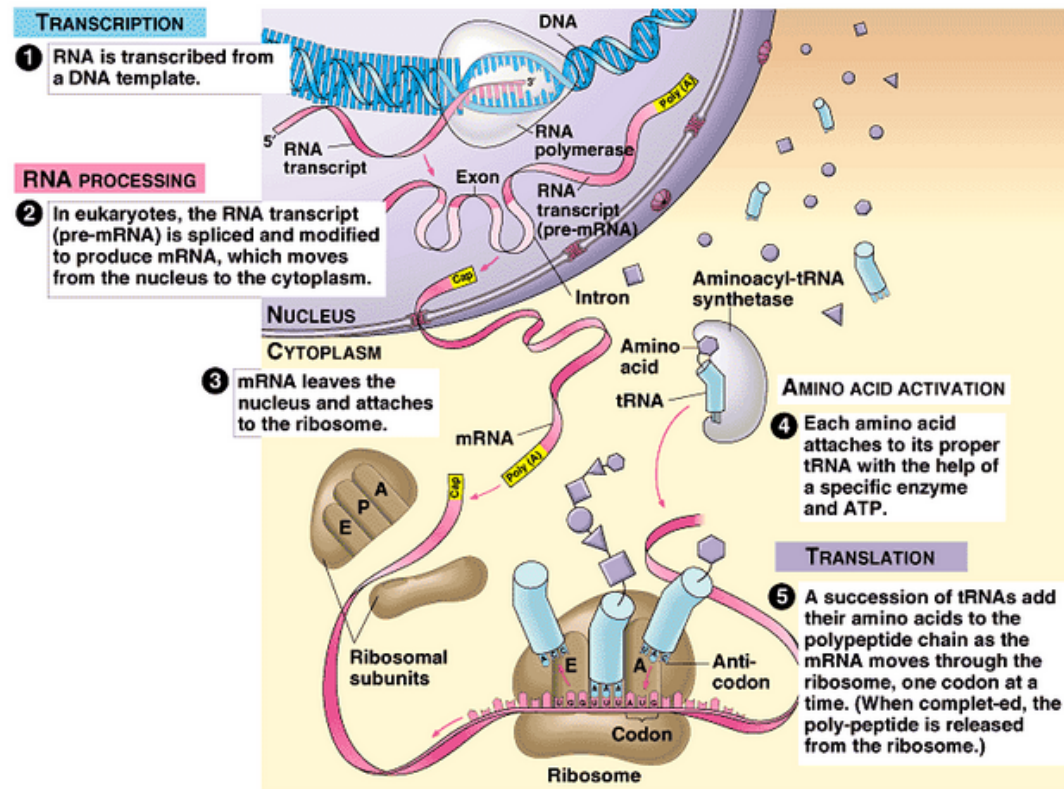


# Le corps humain, la cellule, l'ADN



# Le ribosome, maître d'oeuvre de la synthèse protéique

## Traduit le code génétique en protéines

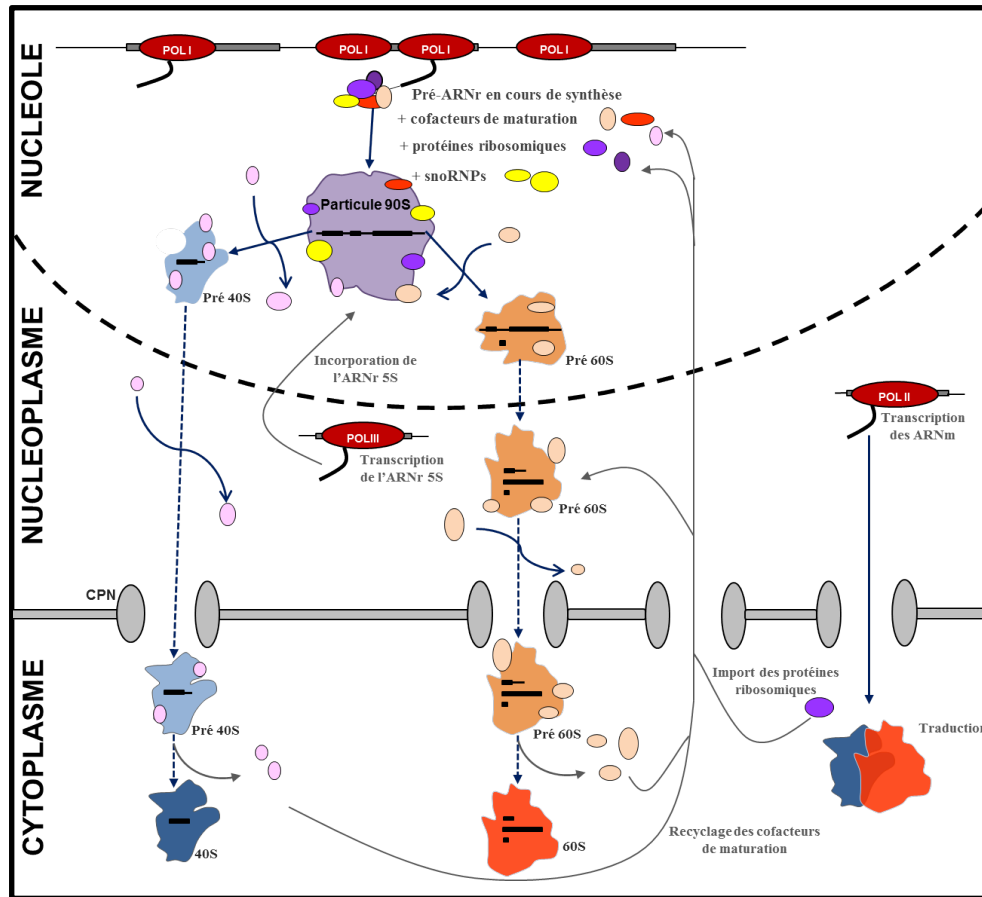




# Ribosome biogenesis : overview

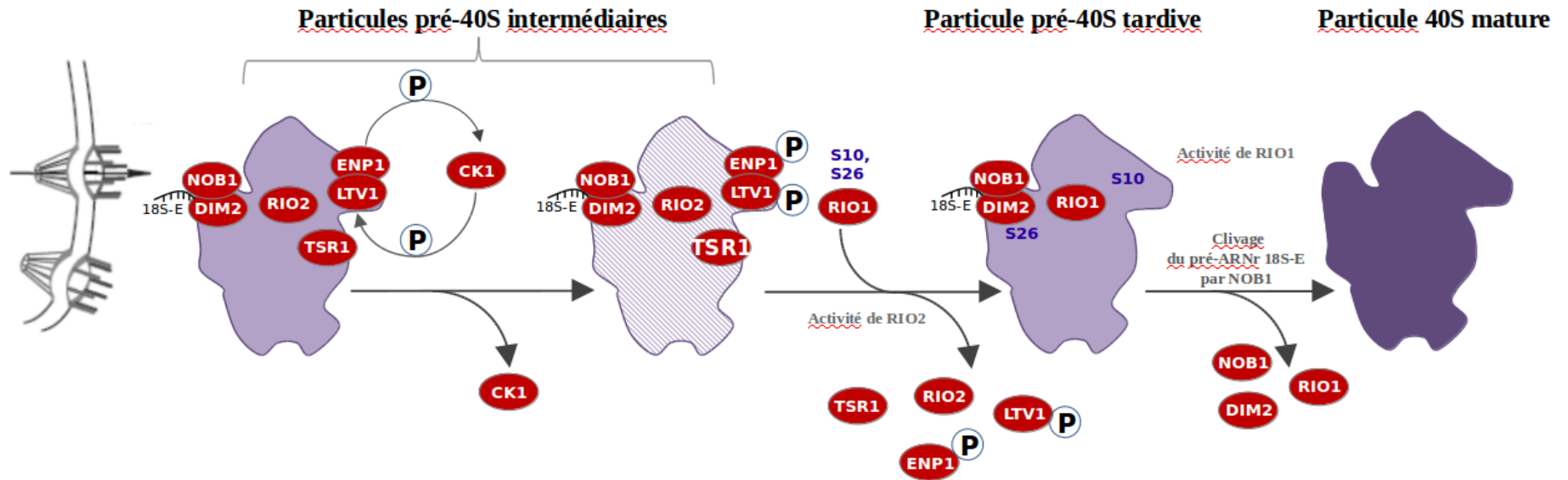
Ribosome biogenesis - is the most energy-consuming process in growing cells  
 - must be finely regulated to adapt to cellular needs

Defects in ribosome assembly linked to human pathologies (cancers, ribosomopathies)

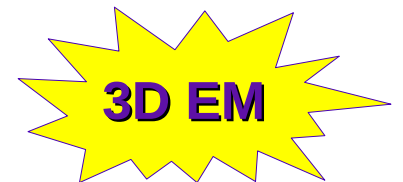


- Production of pre-rRNAs  
transcription, cleavages, modifications
- Correct folding with RPs
- Nuclear export
- Final maturation steps in the cytoplasm
- Requires > 200 Ribosome Biogenesis Factors  
precise function(s) mostly unknown

## Last (cytoplasmic) steps of the small ribosomal subunit maturation

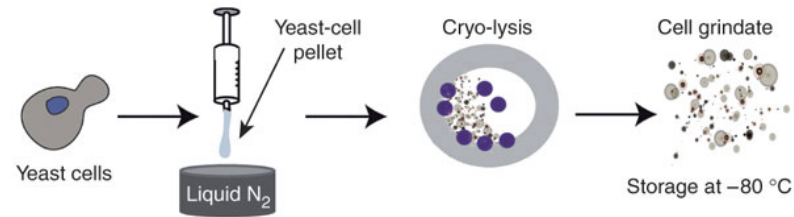


- What are the 3D structures of the cytoplasmic pre-40S ribosomes ?
- How do they relate to the mature 40S subunit ?
- Where are maturation factors on these pre-40S structures ?
- Are the cytoplasmic pre-40S particles structures different in yeast and human ?

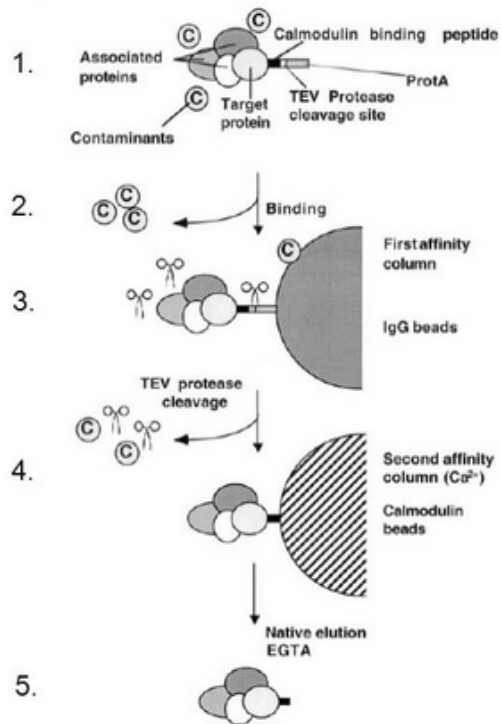


# Purification de particules pré-ribosomiques

## 1. Culture (levures ou cellules humaines), broyage et récolte de l'extrait total



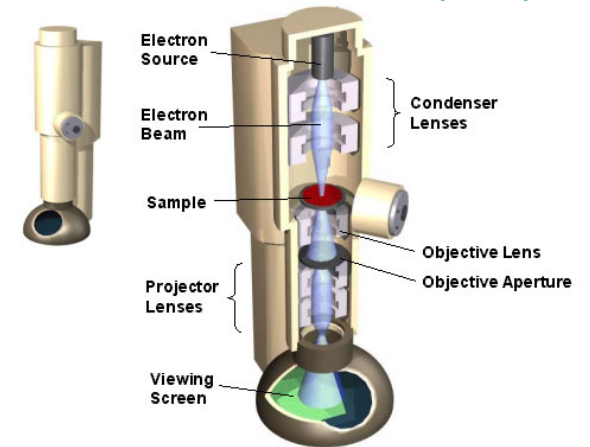
## 2. Purification d'affinité



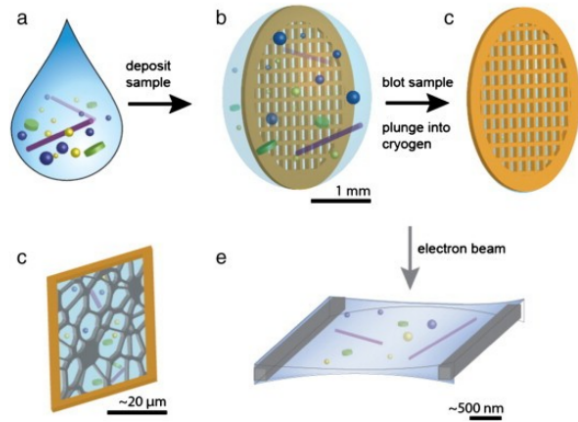
## 3. Dépôt de la solution de particules pré-ribosomiques sur grille de microscopie électronique



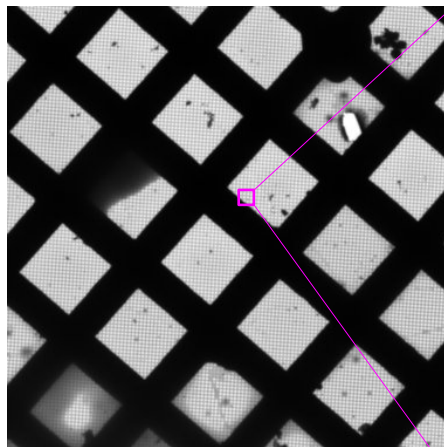
## 4. Observations au microscope électronique en transmission (TEM)



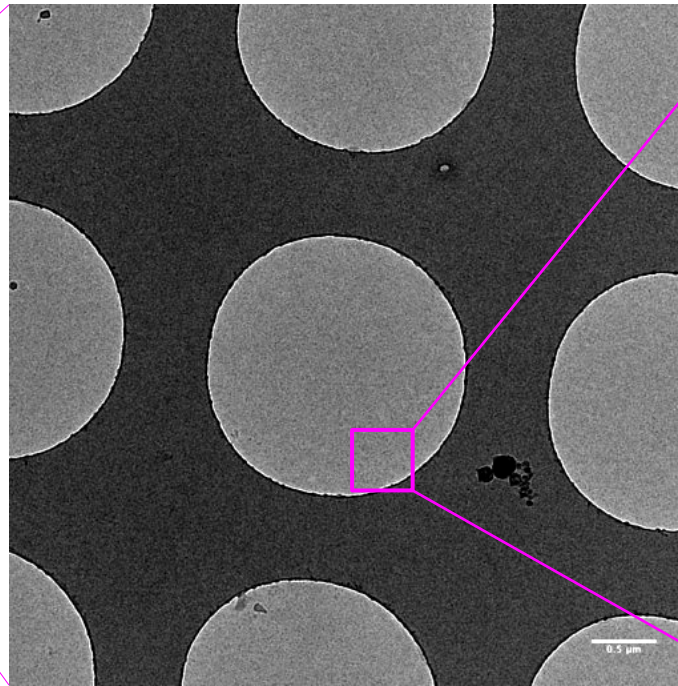
# Observations en cryo-TEM



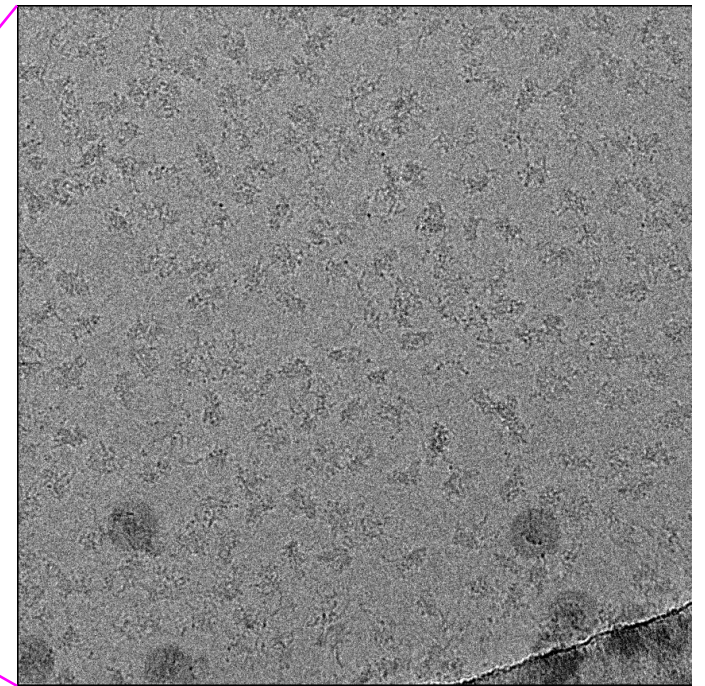
Sample embedded into vitreous ice :  
+ Quasi native state  
- Very low contrast



x 250



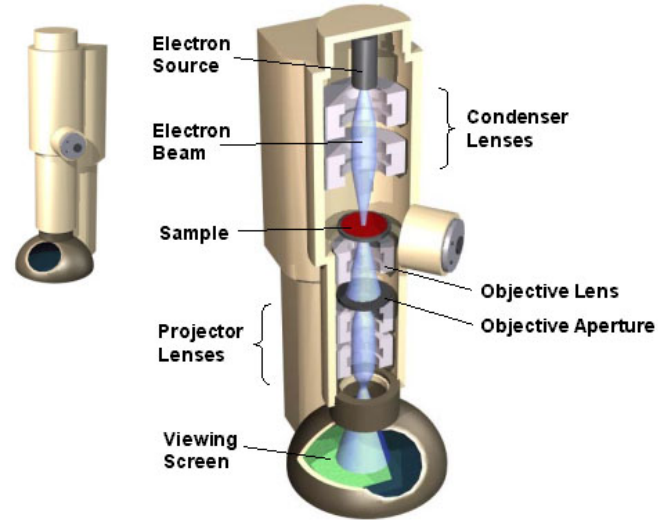
x 5000



x 50 000



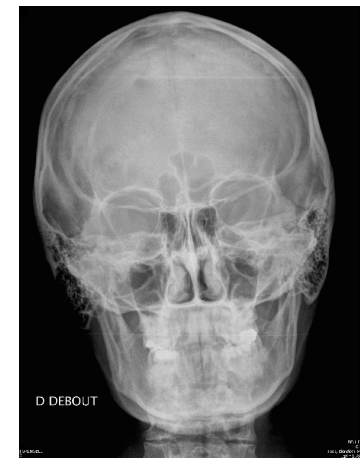
# Principles of 3D (T) EM



- Signal from electrons transmitted throughout the object
- Images = 2D projections of a 3D object

© Elena Orlova

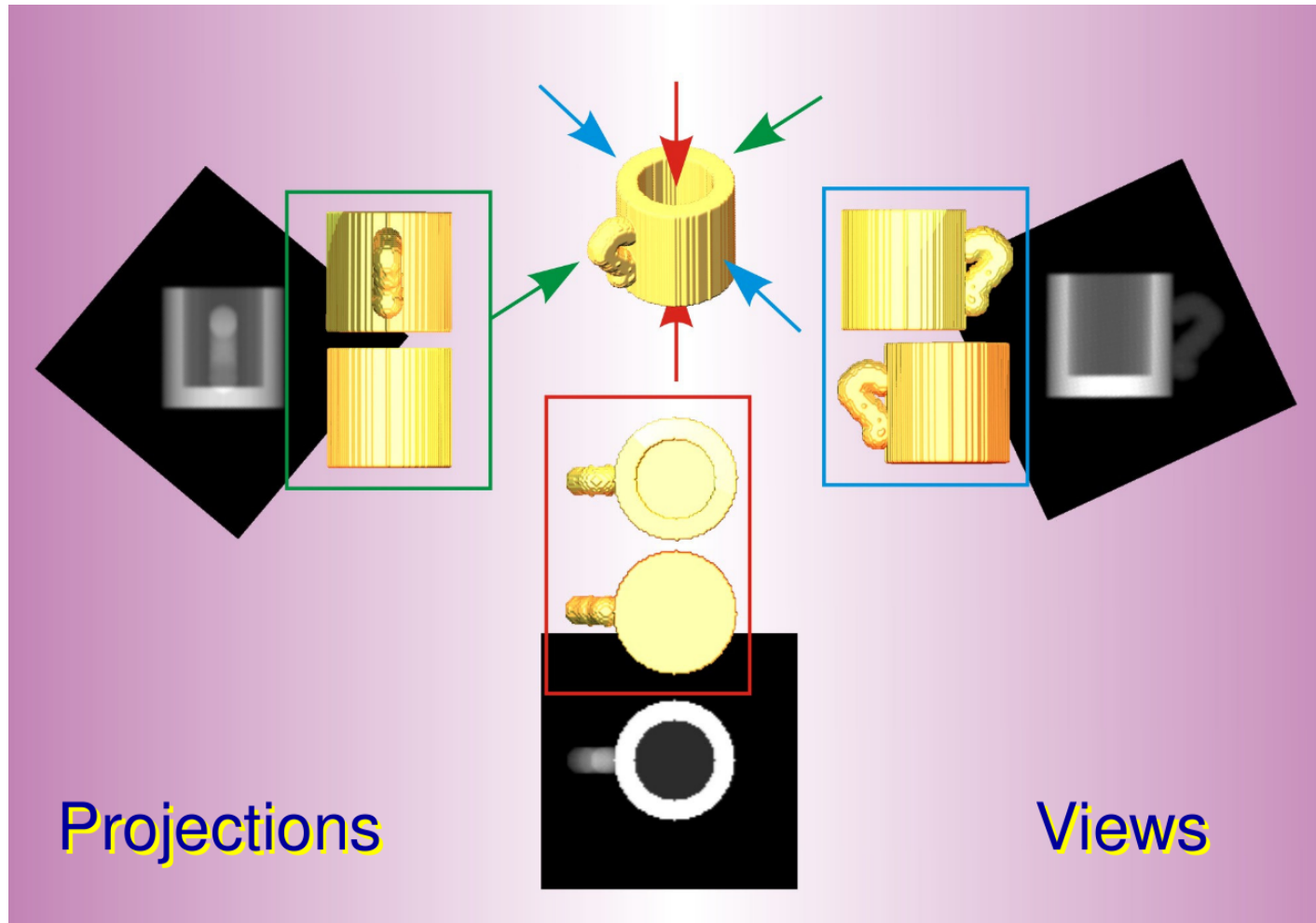
Another 2D projection :  
X-Ray radiograph



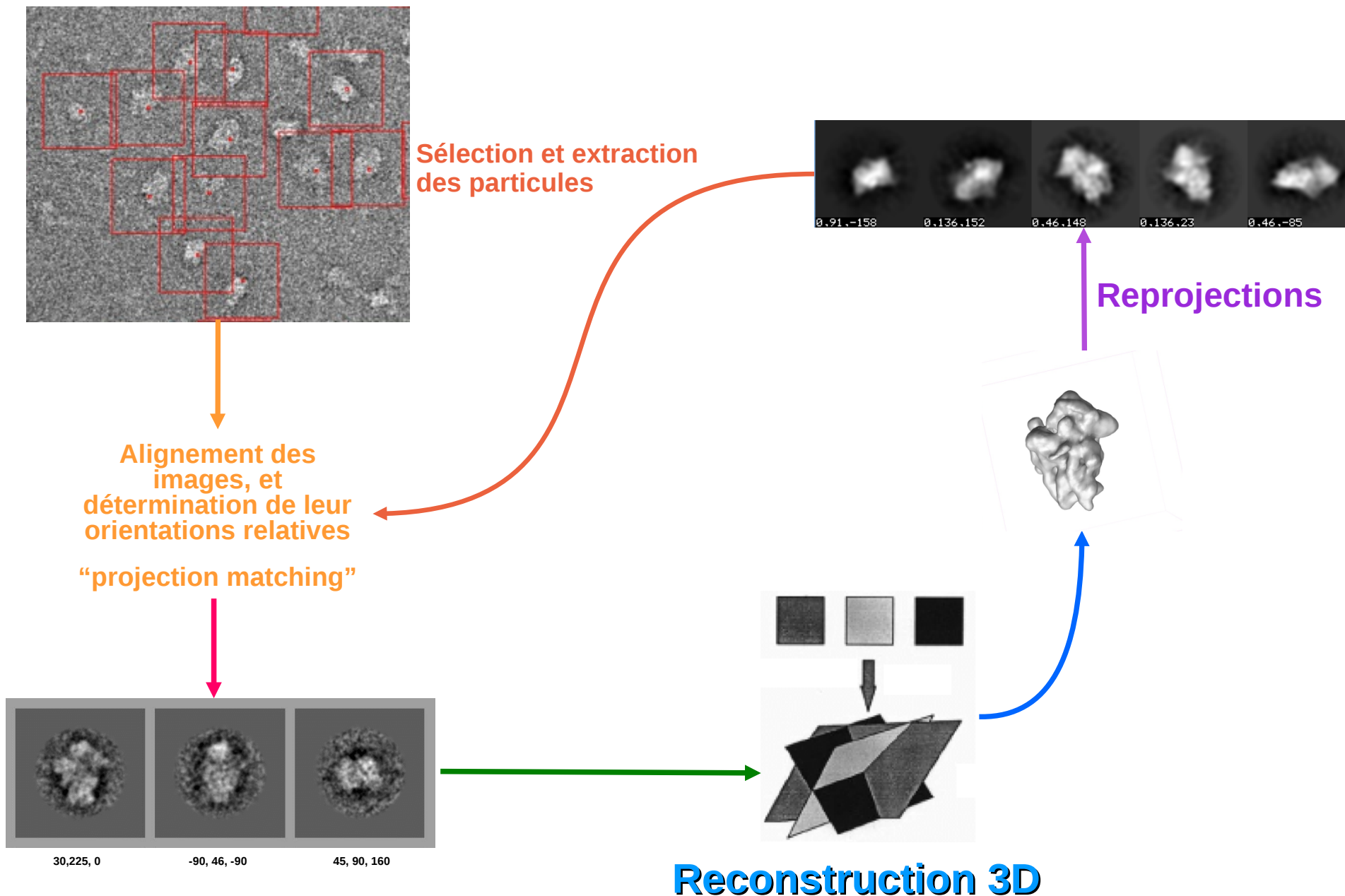
© F. Bing, CHU Grenoble

# MET 3D comment faire?

- Images = projections 2D d'objets 3D
- En théorie, combinaison de 3 vues différentes => structure 3D

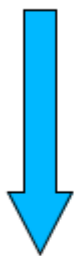


# Analyse d'images de particules isolées



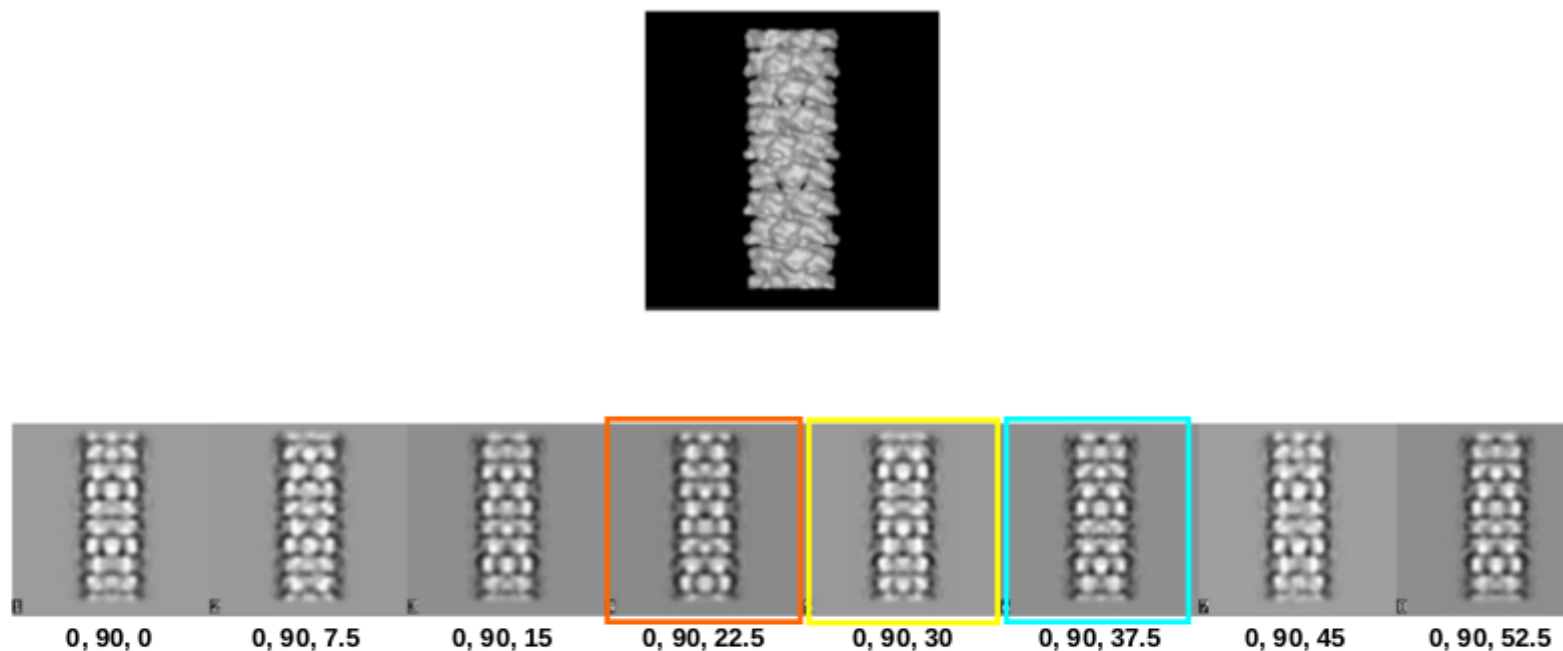
# Détermination de l'orientation des images

Modèle 3D



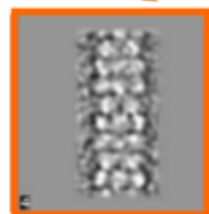
Projections  
2D  
du modèle

(Orientations connues)

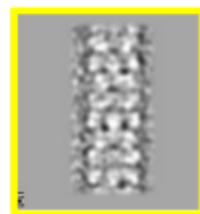


Projection  
matching

Images



0, 90, 22.5



0, 90, 30



0, 90, 37.5

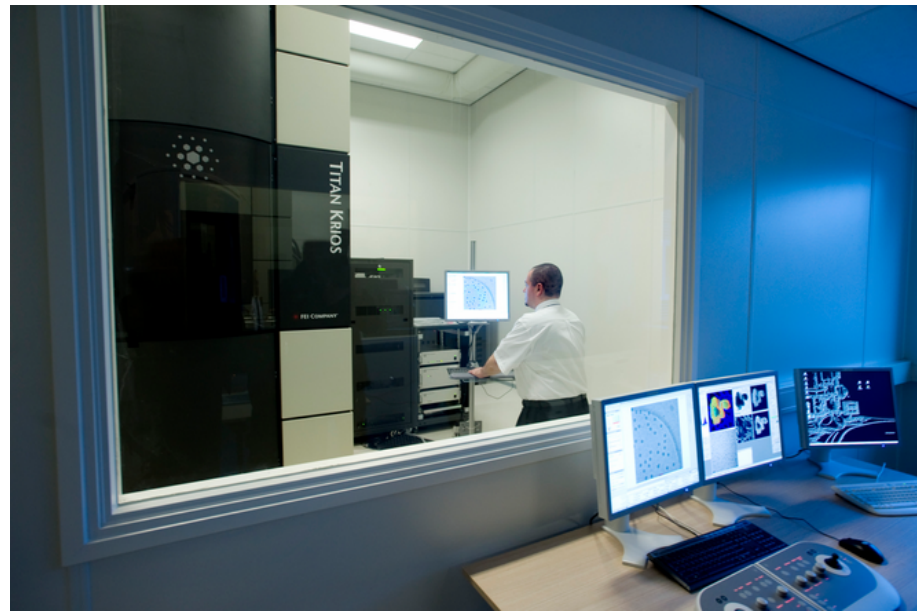
# The "resolution revolution" (1)

Développements récents en cryo-TEM => Structures 3D à une résolution atomique

Correction des aberrations des lentilles => augmentation de la résolution des images

Acquisition des images entièrement automatisée => augmentation du nombre d'images

Caméras à détecteurs directs d'électrons => Augmentation du rapport signal / bruit des images



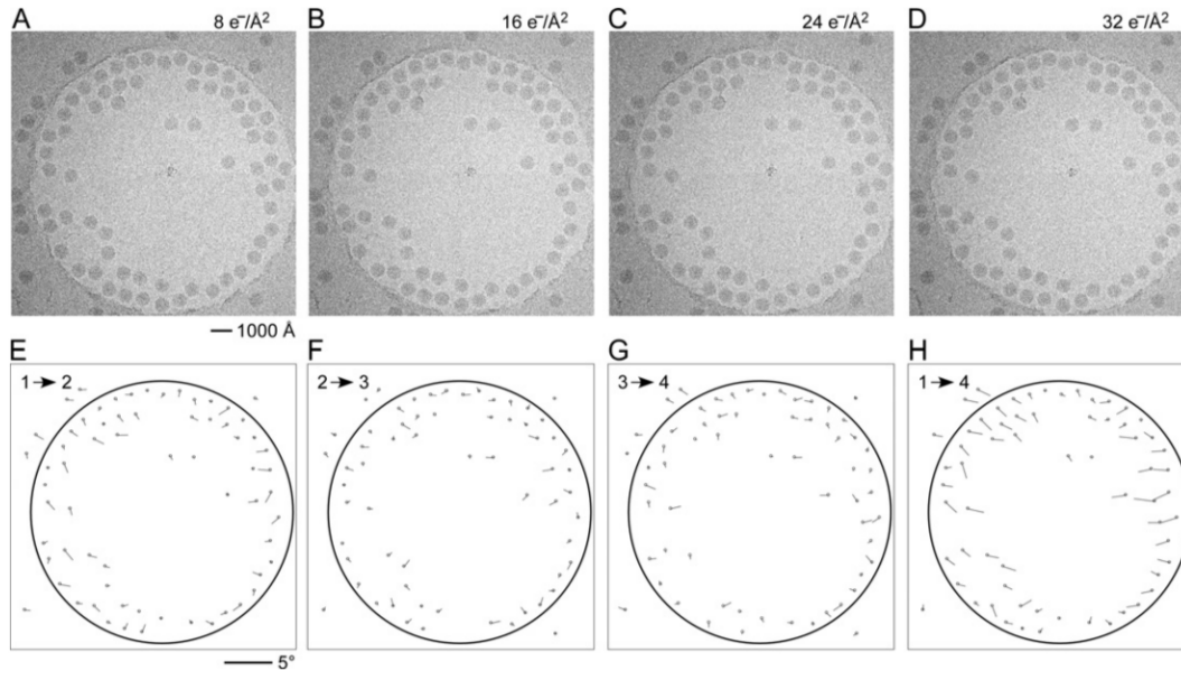
Titan Krios (FEI), NeCEN, Leiden (Pays Bas)

Tension d'accélération : 300 kV

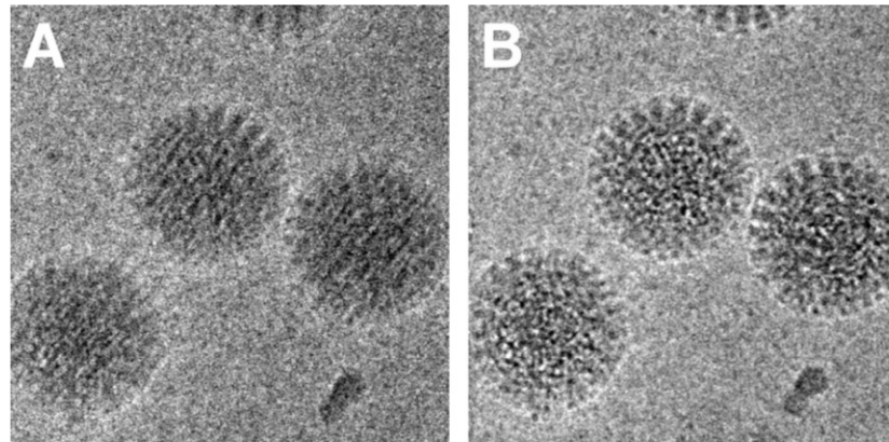
Source d'électrons : pointe FEG

Correction d'aberration sphérique des lentilles électromagnétiques

# Movie alignment



Beam induced  
movements of ice-  
embedded particles

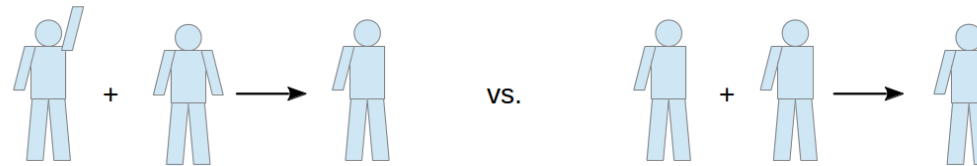


Raw frames average

Average after translational  
alignment of frames

# The "resolution revolution" (2)

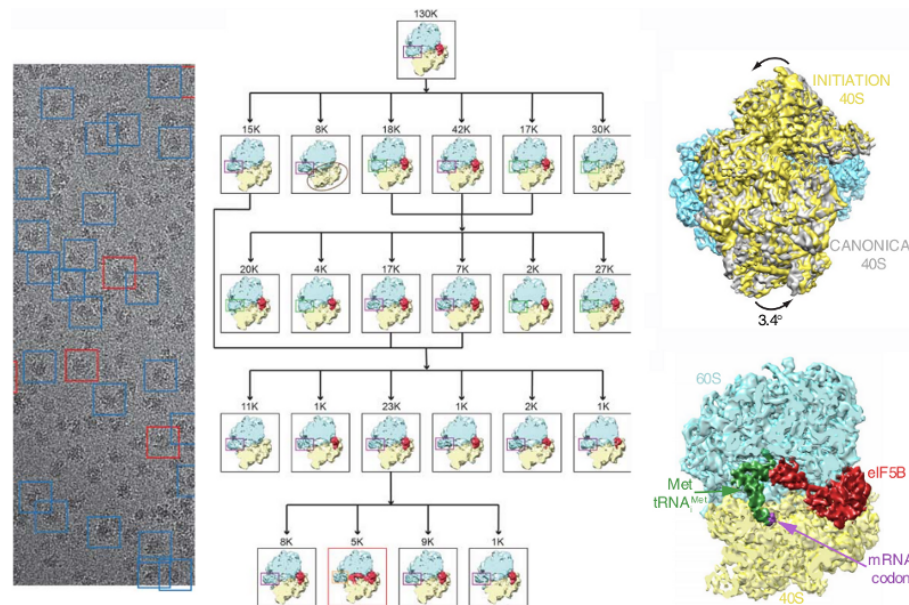
The trouble with (classical) image analysis methods based on averaging data...



Averaging all particles into a single 3D structure limits the resolution

Improved image analysis methods : sorting out 3D structural heterogeneity

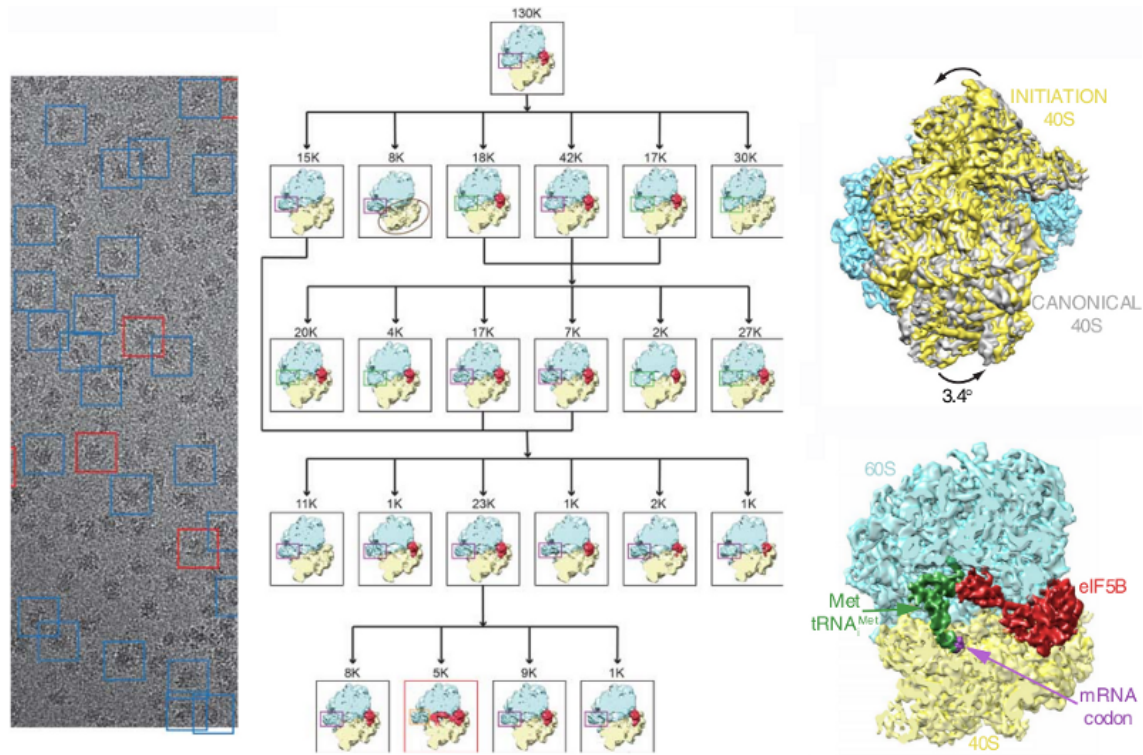
Projection Matching against several 3D references



Structure of yeast translation initiation complex, 6.6Å resolution (Fernandez *et al.*, Science 2013)

# RELION 1.3 (Sjors Scheres, 2012)

## Statistique Bayesienne



Projection matching + prise en compte du rapport signal / bruit de chacune des images

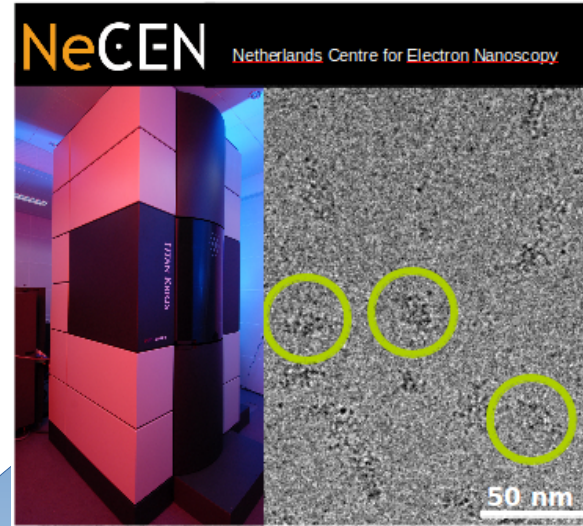
Calcul d'1 structure 3D consensus et / ou de n variants structuraux

« memory-intensive calculations »

Parallélisation hybride mpirun / pthread



### 3)- Automated images acquisition



2x96h beamtime on FEI KRIOS + DED  
Automatic acquisition of ~6000 frames

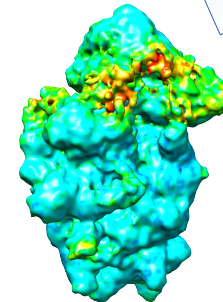
The « big data » era :  
Cryo-EM large scale study

### 4)- Image Analysis on computing clusters



Eos, Calmip, Toulouse  
612 nodes :  
64Gb mem, 20 cores  
300 000 hours cpu

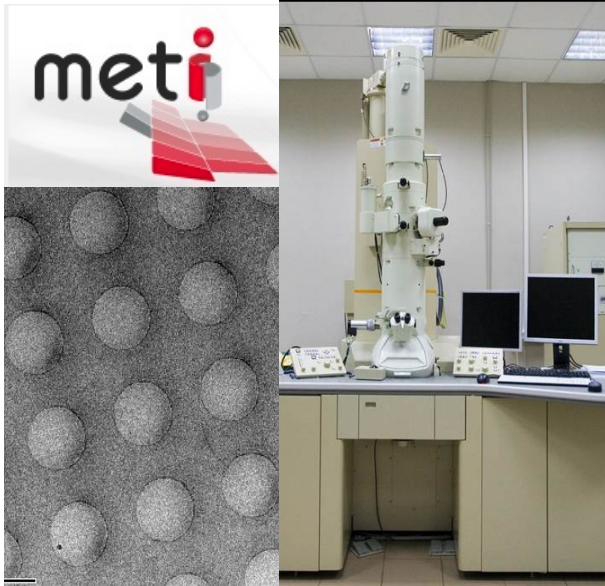
### 5) 3D Structure determination and analysis



### 1)- Pre-40S particles purification



### 2)- Cryo-EM grid production and selection

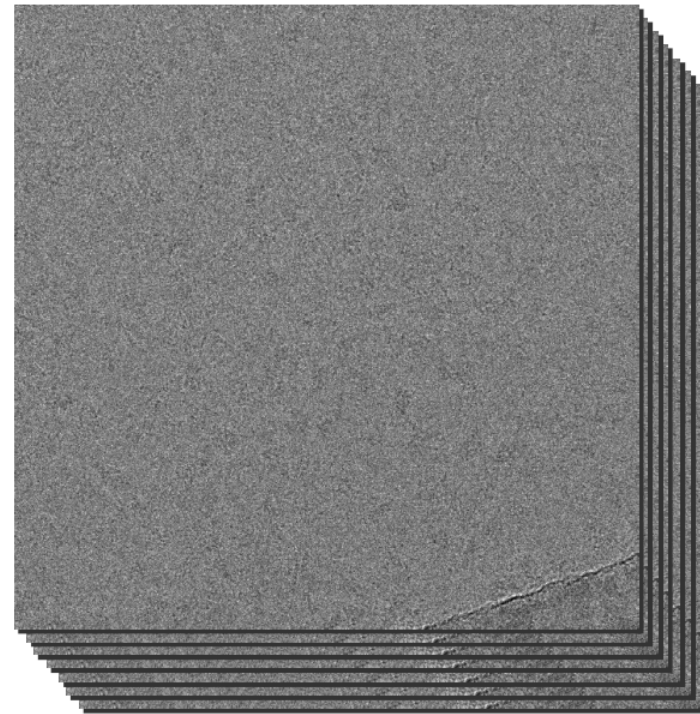
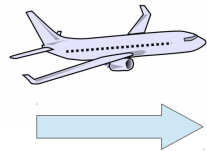
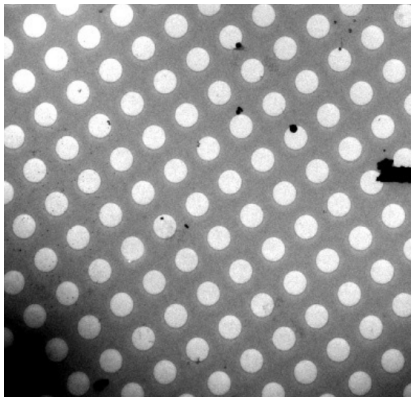


# Structural study of pre-40S pre-ribosomal particles

cryo-TEM images acquisition on a last generation microscope



Cryo-EM grids  
preparation and checking

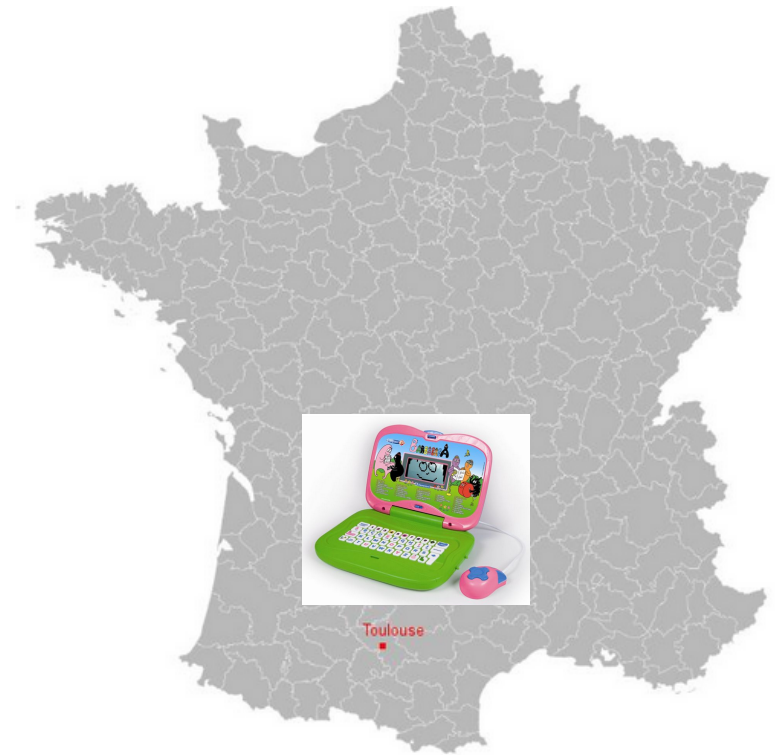
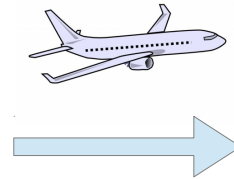
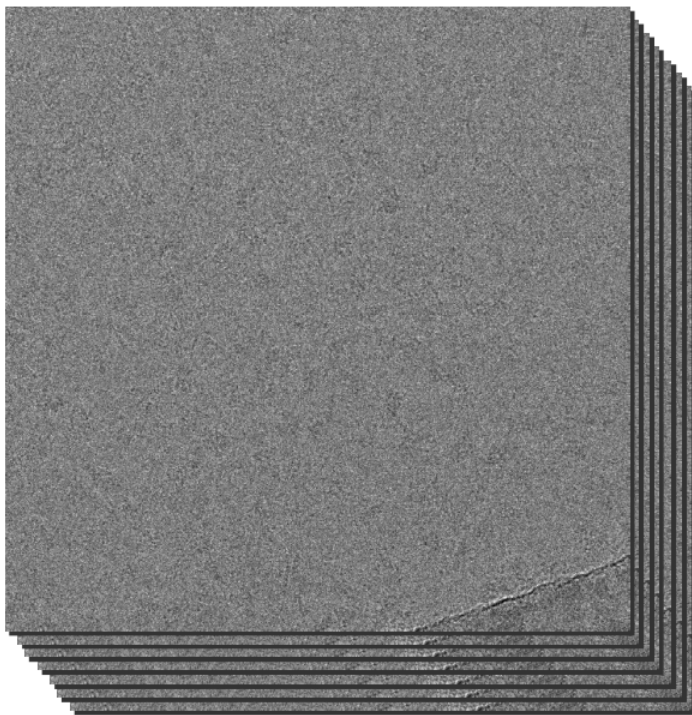


February 2016  
96h beamtime on FEI Titan KRIOS + DED  
Automatic acquisition of ~4000 stacks,  
(7 frames each)

# Acquisition / Transfert des images de cryo-microscopie électronique

96h temps de faisceau  
4000 stacks de 7 frames  
4096 x 4096 pixels ; 33 M / image + fichiers annexes

**750G pour le jeu de données initial**



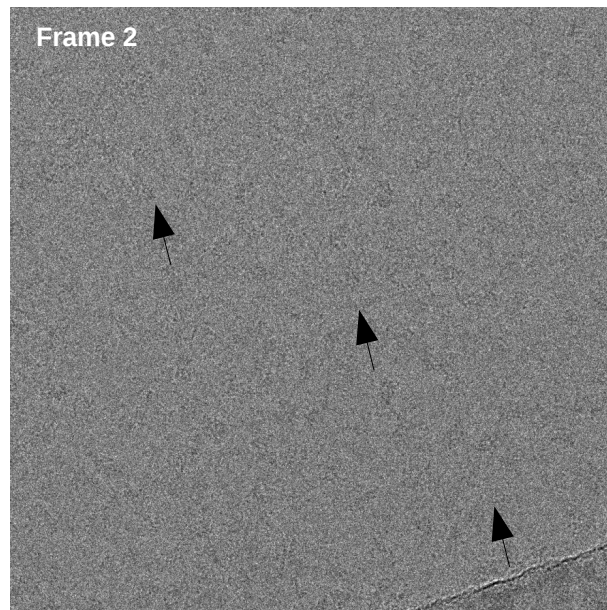
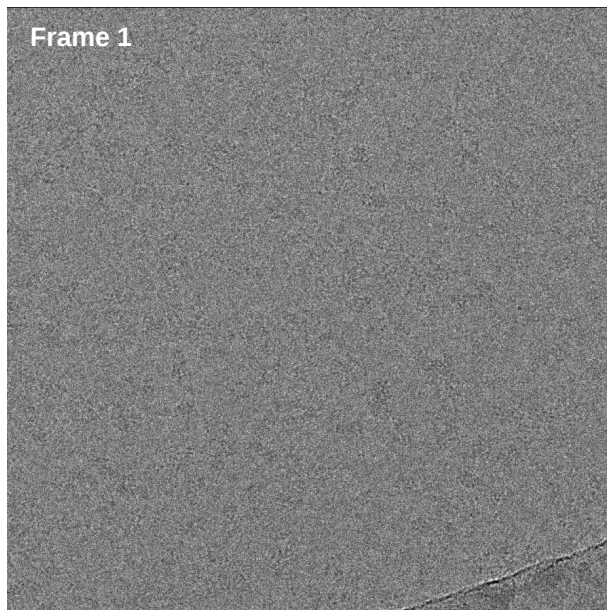
Transfert / archivage du dataset brut



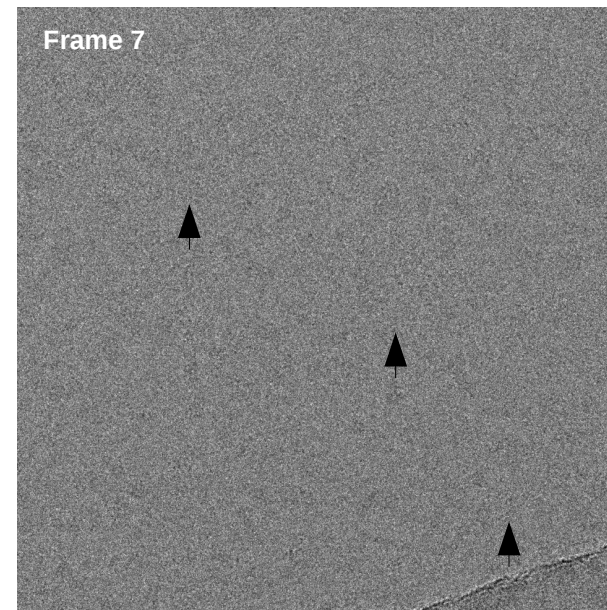
# Pre-processing (1) : Frame alignment and averaging



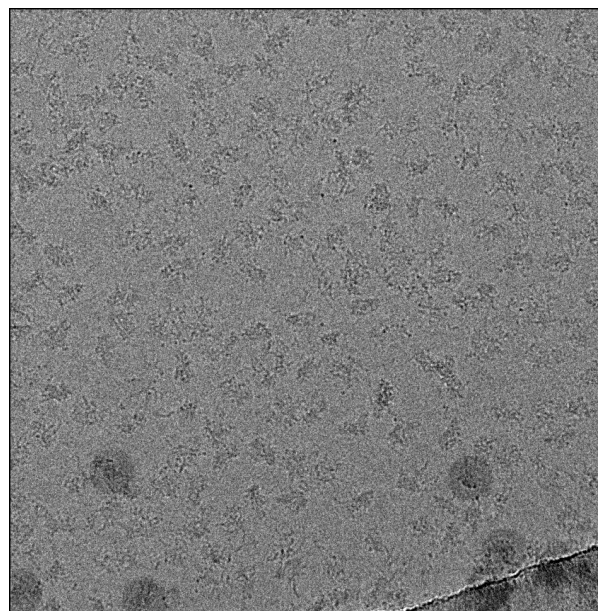
4000 x 7 frames



(...)



Sum after  
realignment

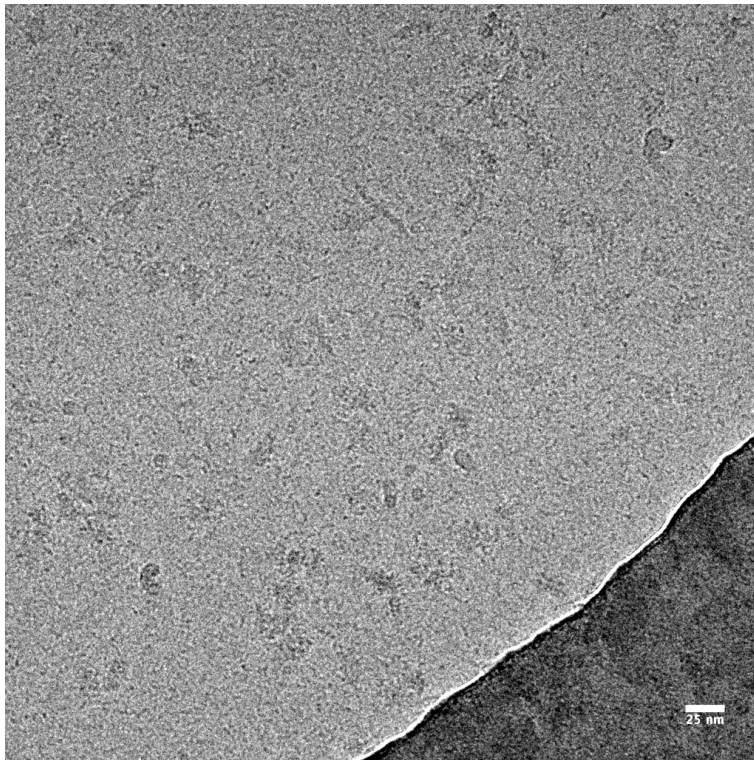


Embarrassingly parallel process  
Output : 4000 images  
~200 Gb  
Needs to be kept at hand

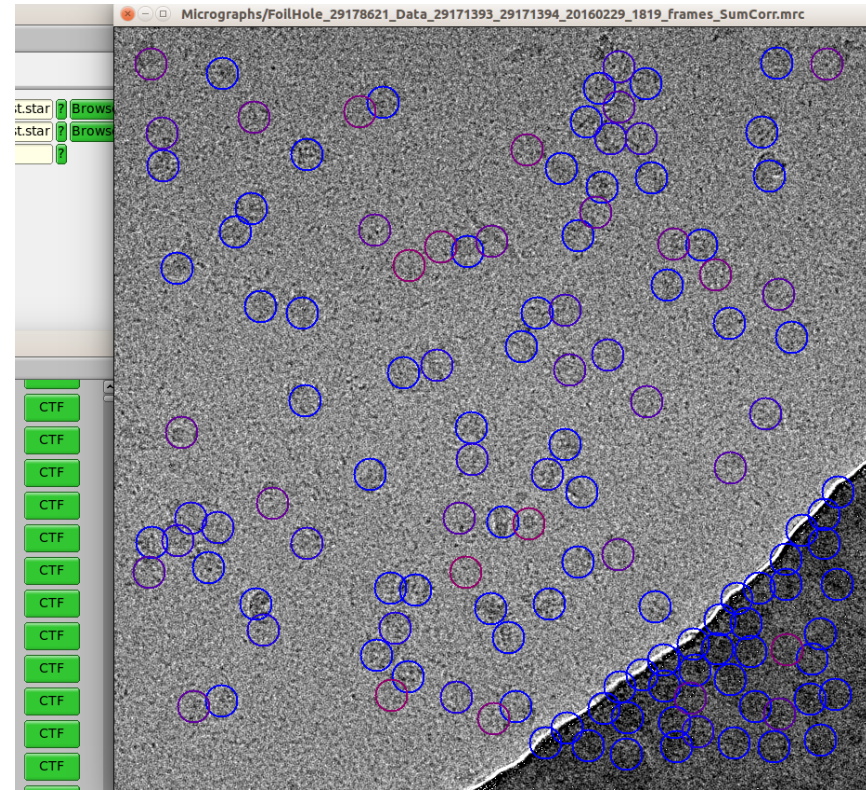


## Pre-processing (2)

- Sélection des images :
  - (1) visuelle
  - (2) estimation de la ctf des images
- Picking & Extraction des particules



Realigned image



Autopicking

4 000 images, 3 220 kept for further analysis  
324 799 extracted particles (auto-picking : Relion)

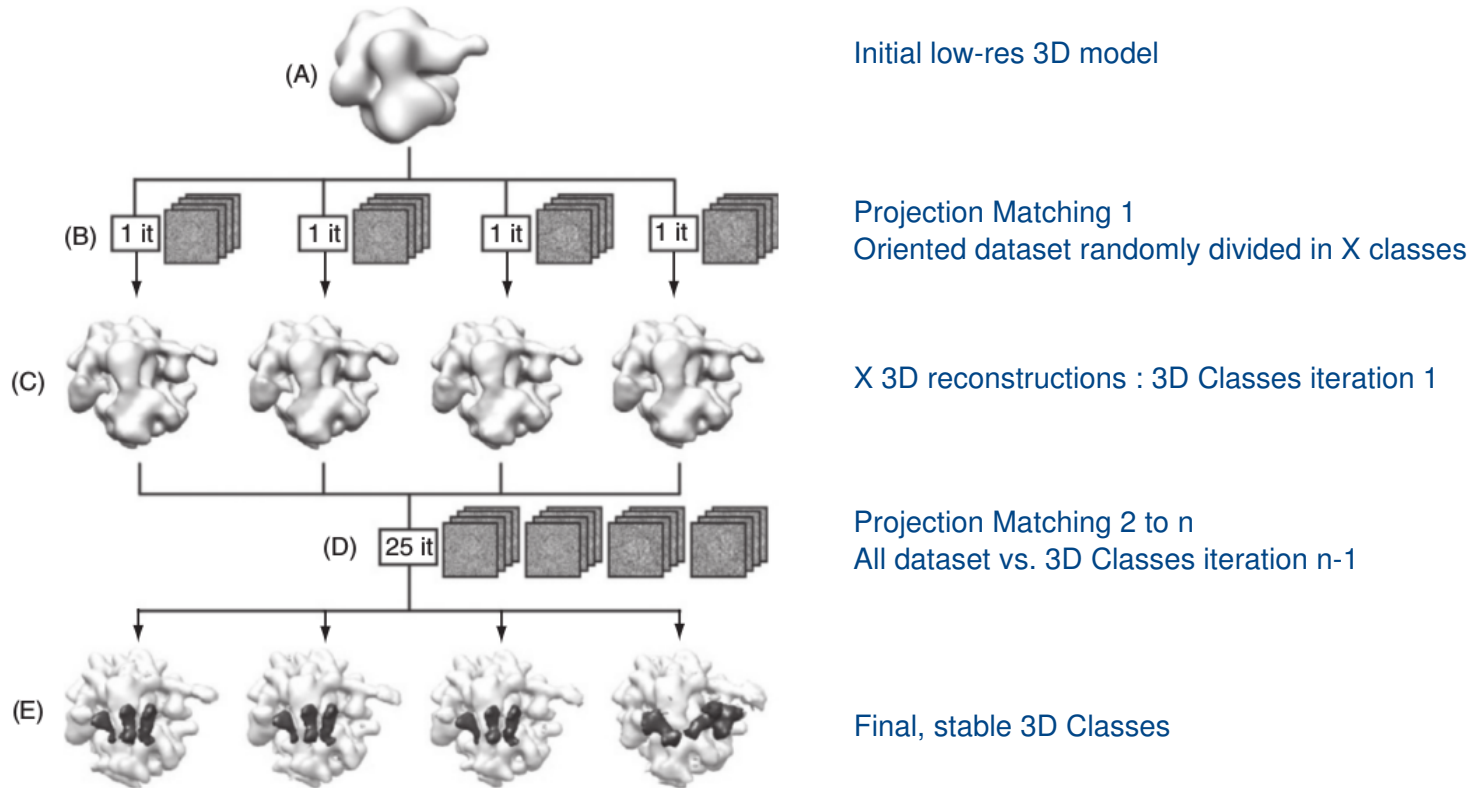
Total time of the pre-processing step : ~6-8 weeks (wall time), user time & input highly needed  
Cpu time : ~10 000 hrs



# 3D Classification : how?

Several methods to sort out structural heterogeneity : bootstrap method, supervised classification, maximum likelihood classification...

## ML 3D Classification



Scheres, 2010

All the current 3D classification methods give good results for compositional and conformational variabilities, but sorting out flexible 3D structures is still an issue to be addressed.

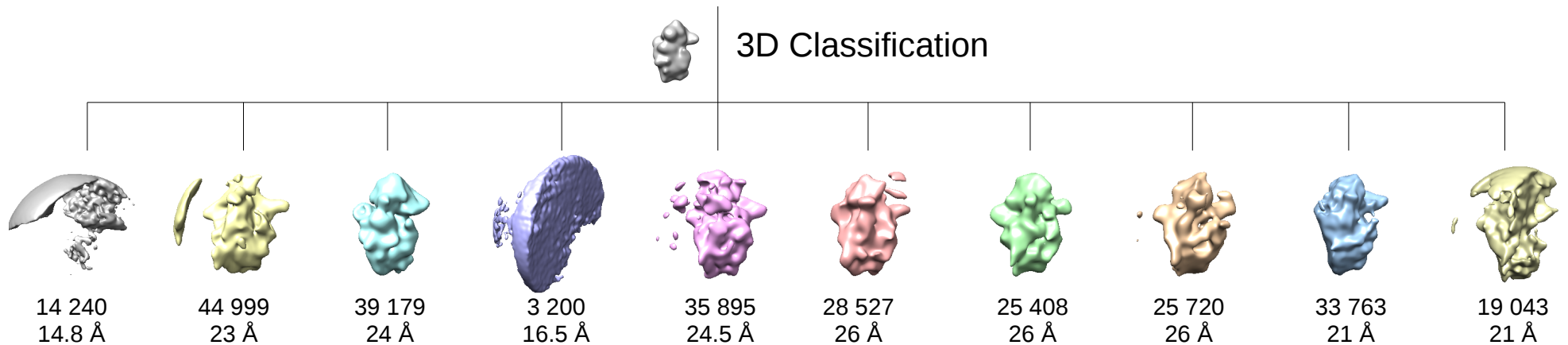
# Processing – Classification 3D

Programme Relion 1.3



Obtention de n variants structuraux, attribution de chacune des particules à l'un des variants

324 799 particles  $\xrightarrow{\text{Visual selection}}$  269 976 particles



RAM requise : 3Gb per task ;  
20 nœuds (20 cœurs / 64 Gb /noeud),  
4 mpi process / nœud (5 cpus per task), threads per cpu = 2 ;  
Walltime = ~ 120 h (48 000 h cpu)

**Augmente avec la taille et le nb de particules, la finesse d'orientation et le nb d'itérations demandées par l'utilisateur**

**Output = fichiers textes + structures 3D (mrc format ; ~28 Mb)**



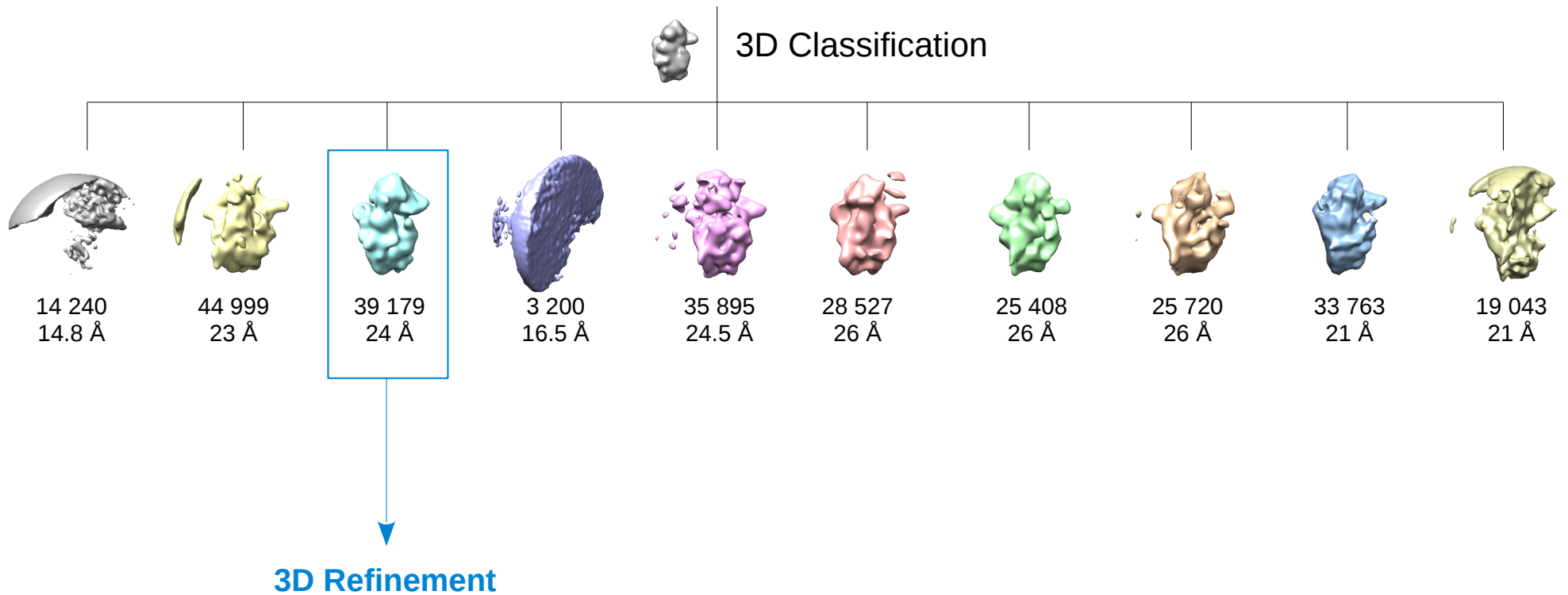
# Processing – Classification 3D

Programme Relion 1.3



Obtention de n variants structuraux, attribution de chacune des particules à l'un des variants

324 799 particles  $\xrightarrow{\text{Visual selection}}$  269 976 particles



détermination de l'orientation des particules le plus finement possible pour une reconstruction 3D à la plus haute résolution possible

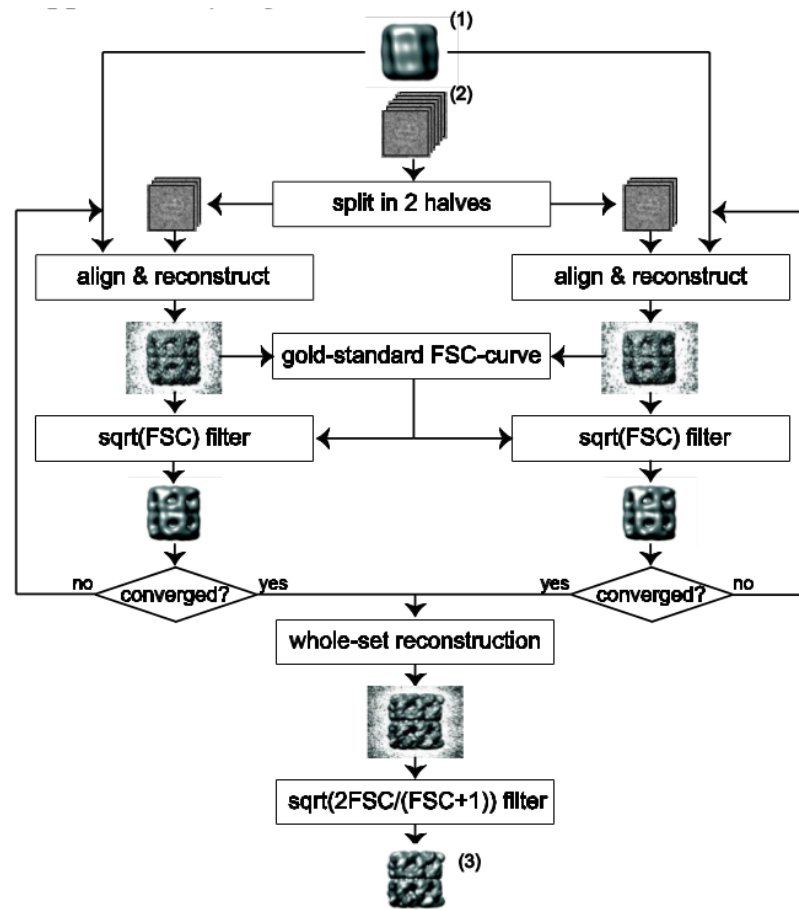
## 3D Refinement

Iterative ML projection matching from an initial, low resolution model

Goal : align and determine orientation of particles with the highest possible precision degree

At the beginning, the dataset is randomly divided into two groups, refined separately from each other

The 3D reconstruction resulting from a cycle of MLPM will be used as reference for the next one

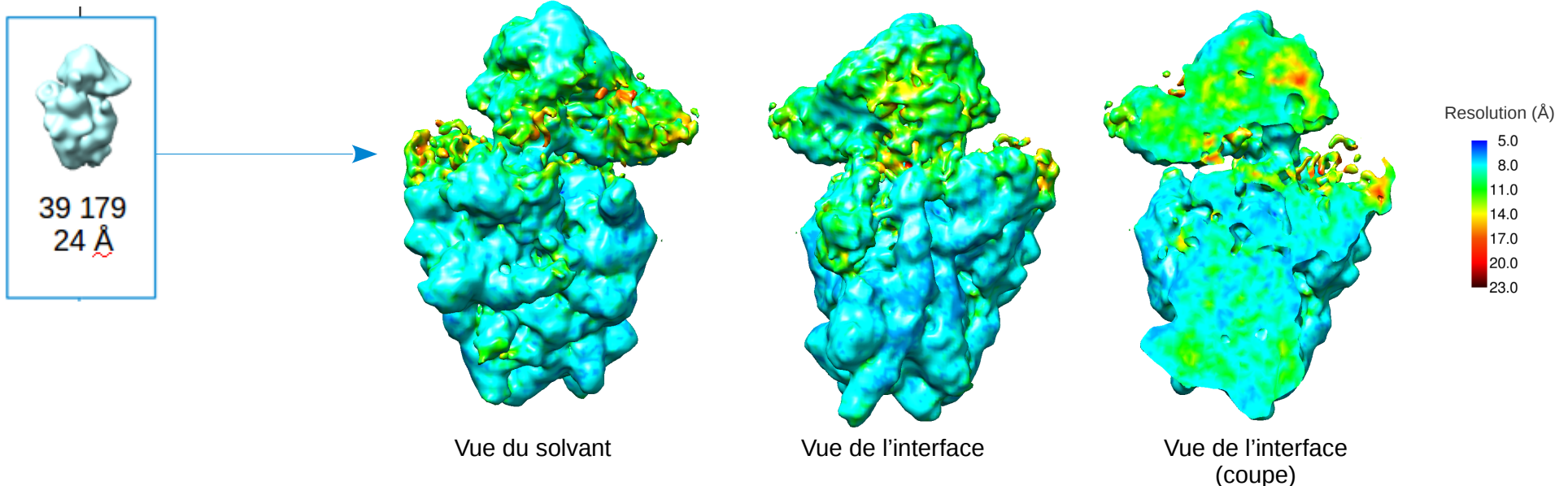


# Processing - Refine

Programme Relion 1.3



Obtention d'un seul modèle 3D « consensus », à la plus haute résolution possible  
(~40 000 particules)



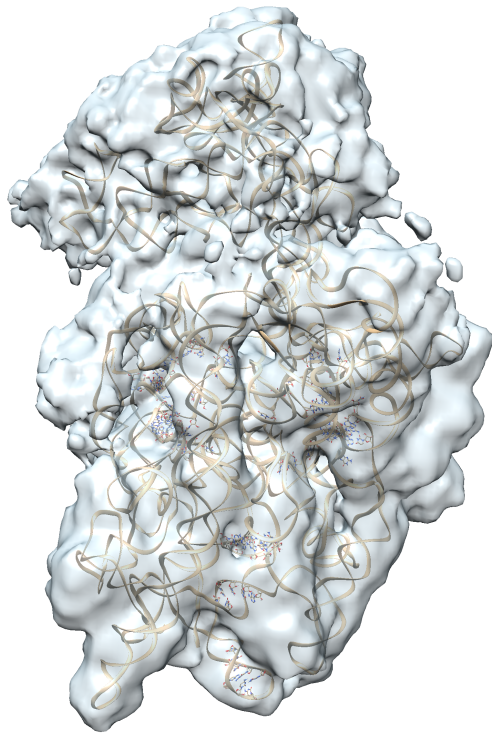
RAM requise : 3Gb per task ;  
15 nœuds (20 cœurs / 64 Gb /noeud),  
4 mpi process / nœud (5 cpus per task), threads per cpu = 2 ;  
Walltime = ~ 2 h (600 h cpu)

**Augmente avec la taille et le nb de particules, et le nb d'itérations (ie., la résolution finale)**

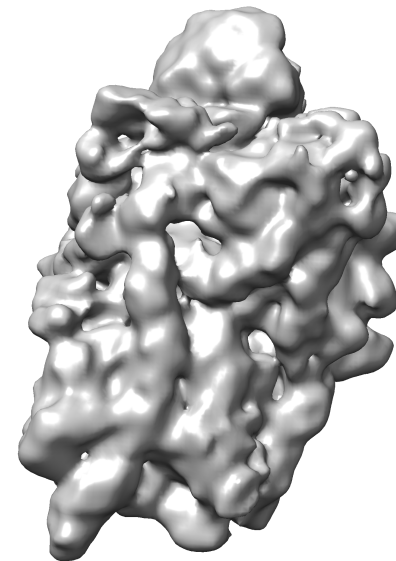
**Output = fichiers textes + structures 3D (mrc format ; ~28 Mb)**

# The joys of image processing and/or model analysis are endless...

Molecular dynamics  
(MDFF)



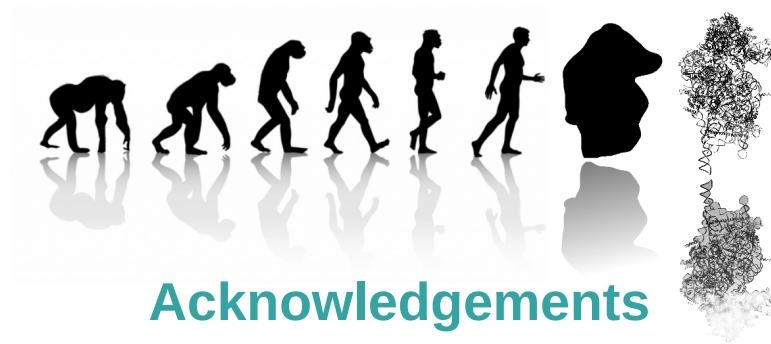
Local Refinement/Classifications



## Perspectives / Besoins



- Gestion des fichiers intermédiaires produits lors de l'analyse  
(Dataset 15G => n Structure(s) 3D 28 M + 120 G fichiers « temporaires »)
- Relion 1.3 optimisé pour le calcul sur Eos (Re-compilé avec Intel : merci les Calmipiens!) ;  
Goulot d'étranglement actuel : temps de transfert des fichiers
- Gestion de projets d'équipe (compte / permissions)
- Relion 2.0 fonctionne sous GPU...  
Classification 3D jusqu'à 40x plus rapide que sous cpu (Kimanius et al., eLife Nov 2016)



## Acknowledgements

**Plateau METI, Toulouse**  
Stéphanie Balor  
Vanessa Soldan

**Ramtin Shayan**  
**Natacha Larburu**

Marie-Françoise O'Donohue  
Dana Rinaldi  
Nathalie Montel  
Franck Delavoie  
Léo Gagliardi  
*Marlène Faubladièr*  
Pierre-Emmanuel Gleizes

**CALMIP, Toulouse**  
Nicolas Renon  
Emmanuel Courcelle  
Pierrette Barbaresco

**ETH, Zurich**  
Christian Montellese  
Ulrike Kutay

**LBME, Toulouse**  
Anthony Henras

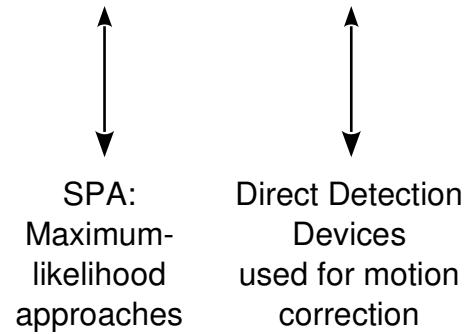
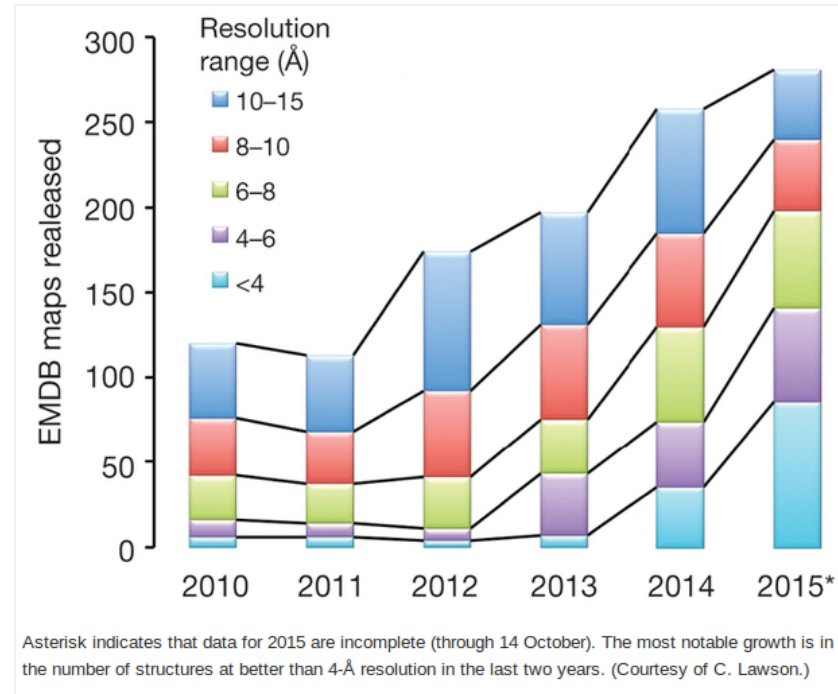
**IBMC, Strasbourg**  
Yaser Hashem

**Service Informatique**  
Alain Kamgoué  
Christophe Carles

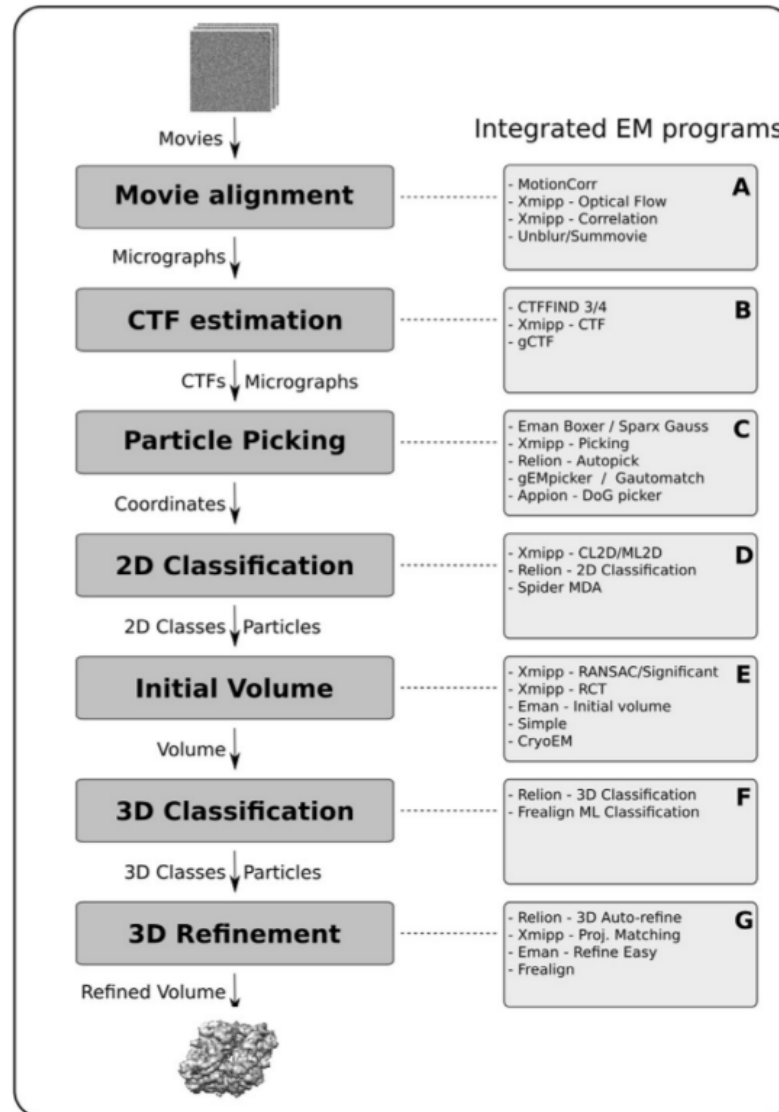


# The “resolution revolution”

Figure 1: Increase in the number of EM map depositions in the Electron Microscopy Data Bank (<http://emdatabank.org/>) over the last six years.



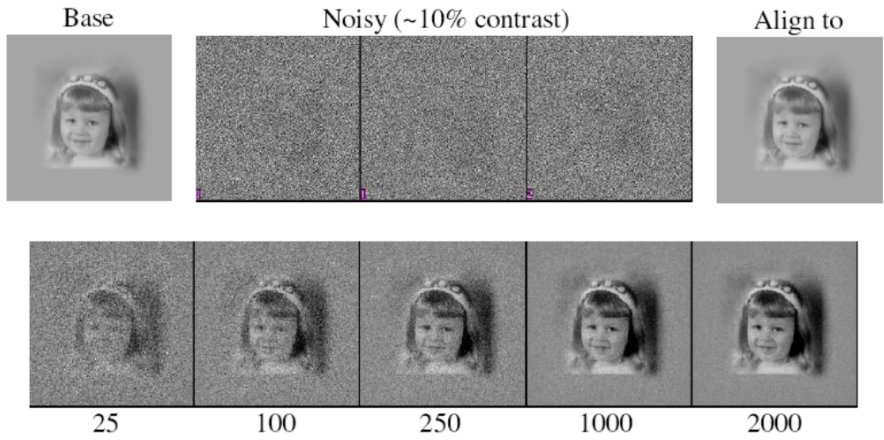
# Single particle analysis : workflow



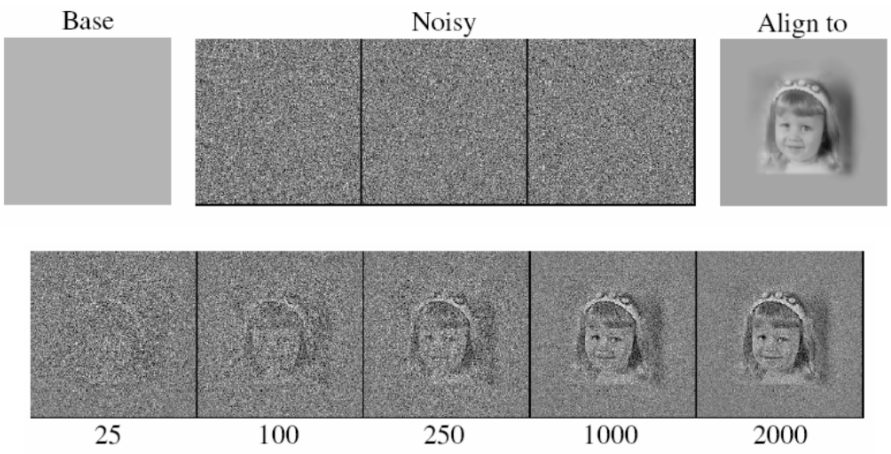


# Data Overfitting / Model Bias

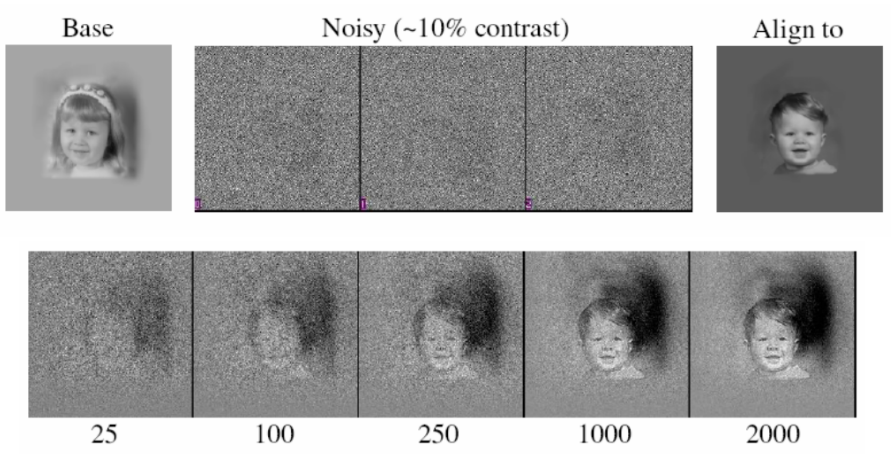
## « Ideal » case



## White noise

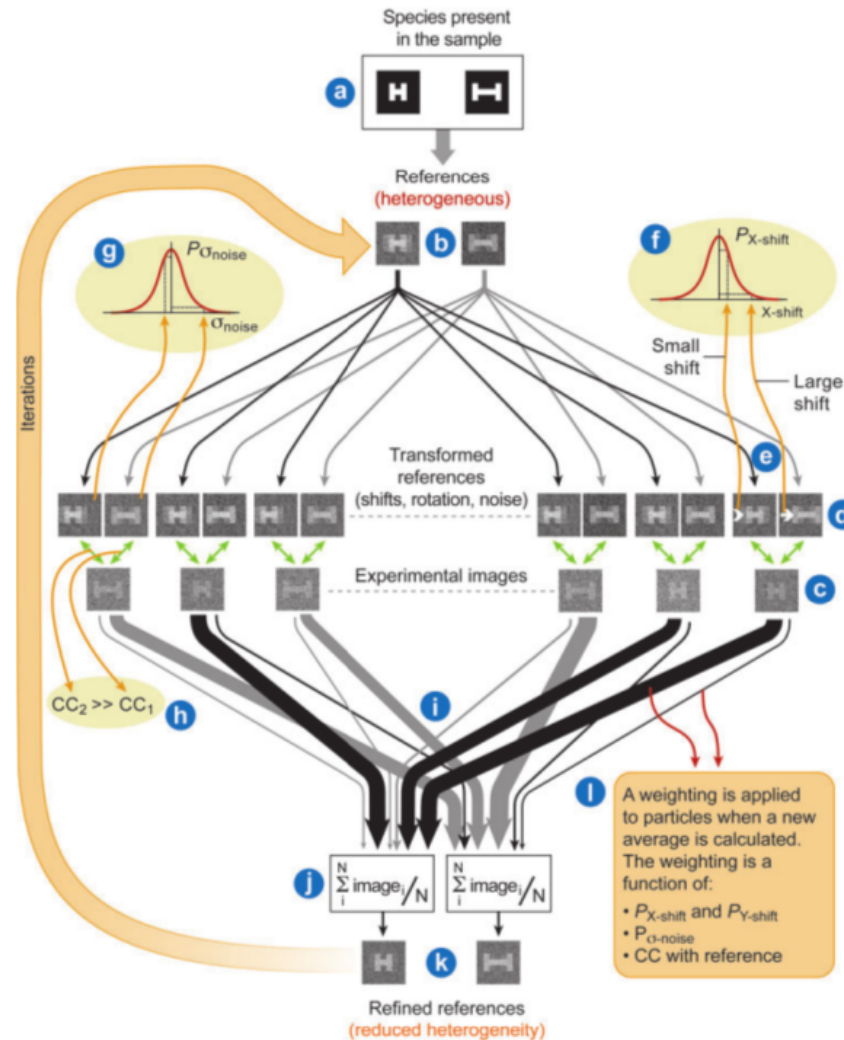


## Wrong model



# Orientation determination: PM + probabilistic approaches (maximum likelihood)

Particles orientation are given by projection matching, but a weighting is applied to each particle before 3D reconstruction. “Good” particles have more weight than “bad” ones. This can at least partly prevent overfitting.

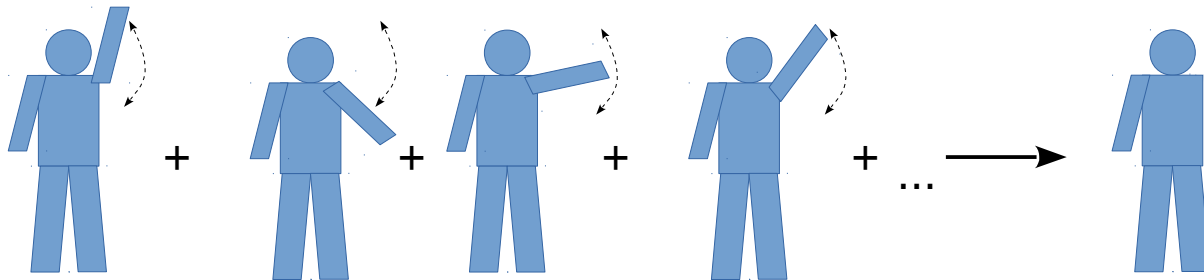


# 3D Classification : why?

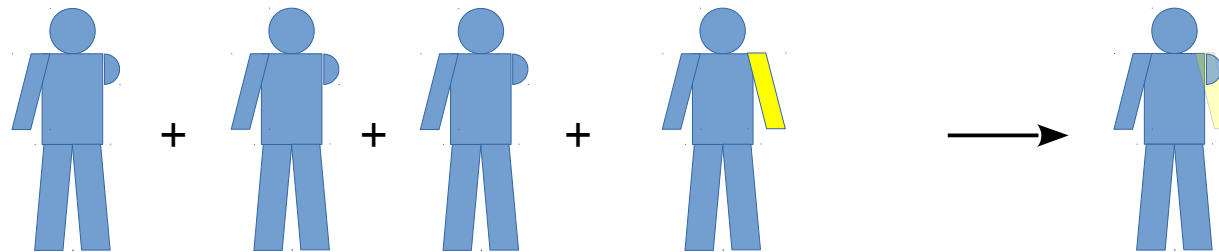
All datasets are heterogenous! If we can sort out conformers from the particles' population, we can increase the resolution of each of them instead of averaging them in a single consensus 3D structure.

Structural heterogeneity can result from: flexibility (and/or) compositional variability (and/or) conformational variability

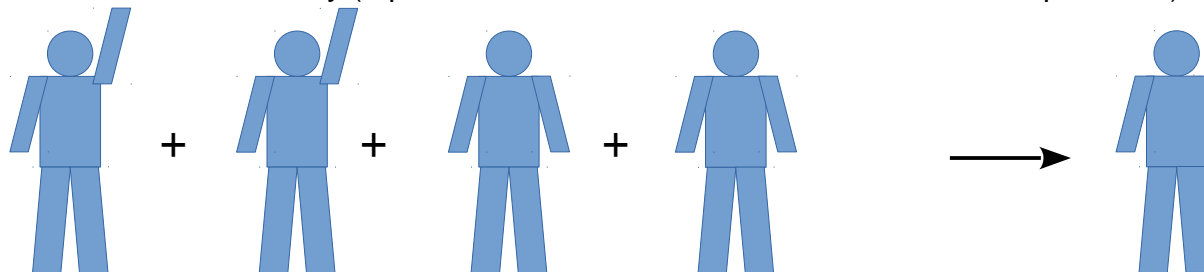
Flexibility (constant movement of a part of the 3D structure: n particles, n conformers)



Compositional variability



Conformational variability (a part of the 3D structure has a fixed number of positions)



Particles

Consensus 3D structure