

# **Molecules in motion** ***towards four-dimensional cryo-EM***

**Pawel A. Penczek**

**The University of Texas – Houston Medical School,  
Department of Biochemistry.**



**THE UNIVERSITY *of* TEXAS**

**HEALTH SCIENCE CENTER AT HOUSTON**

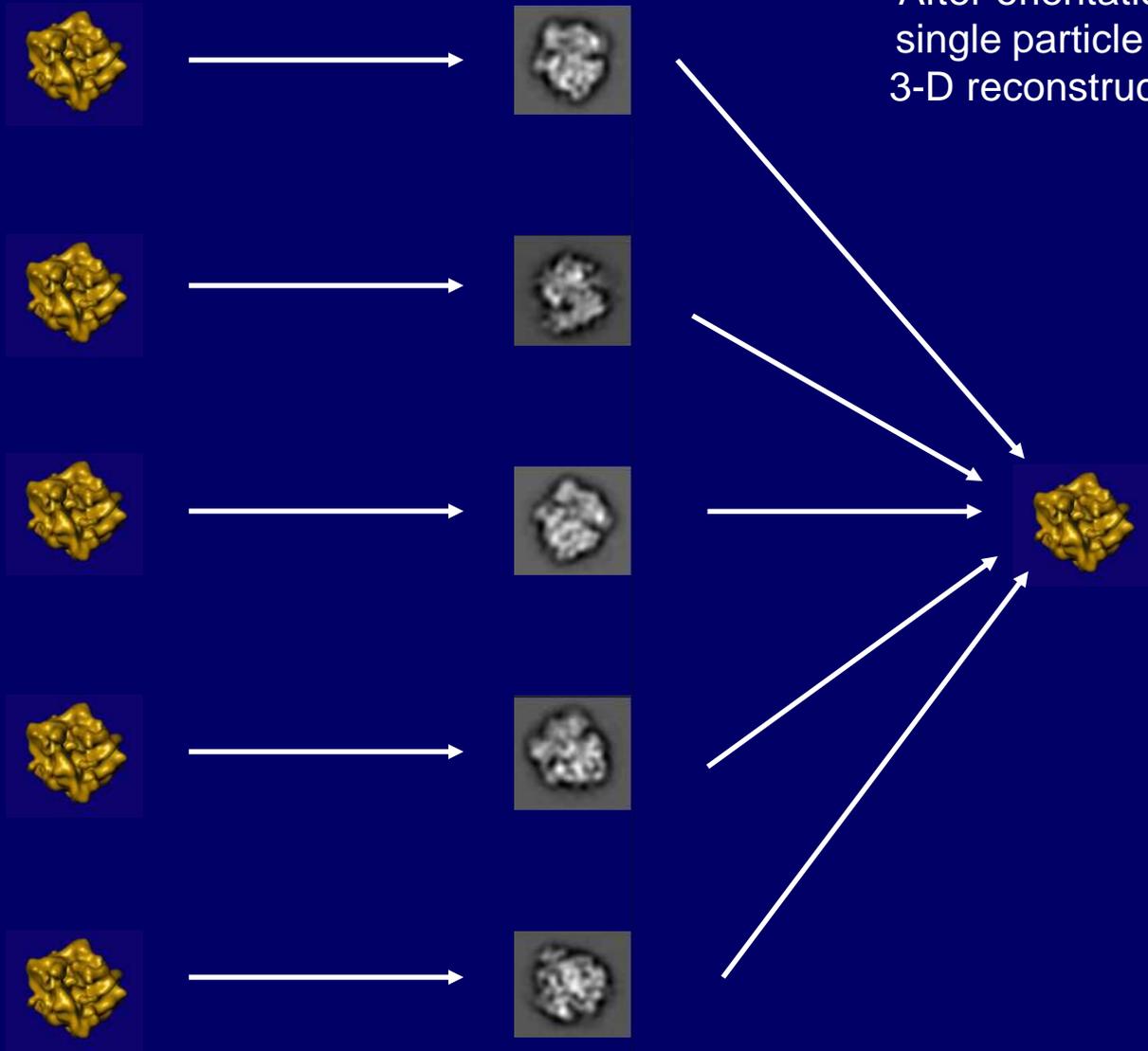
**MEDICAL SCHOOL**

# Premise of single particle cryo-EM

*“single particles” stands for “isolated, unordered particles with -- in principle -- identical structure”*

If this premise is fulfilled, it is possible to align and superimpose the individual images for the purpose of forming an average or 3D reconstruction of the mass density of the biological macromolecule.

In electron microscope, single particles of the same macromolecule (3-D) are imaged from different directions. In principle, 2-D projections of observed macromolecules that in principle have the same structure.



After orientation parameters of single particle views are found, 3-D reconstruction is calculated.

There is mounting evidence that macromolecules occur naturally in a mixture of conformational states:

- ribosome
- RNA polymerase
- human transcription factor
- pyruvate dehydrogenase complex (breathing core)

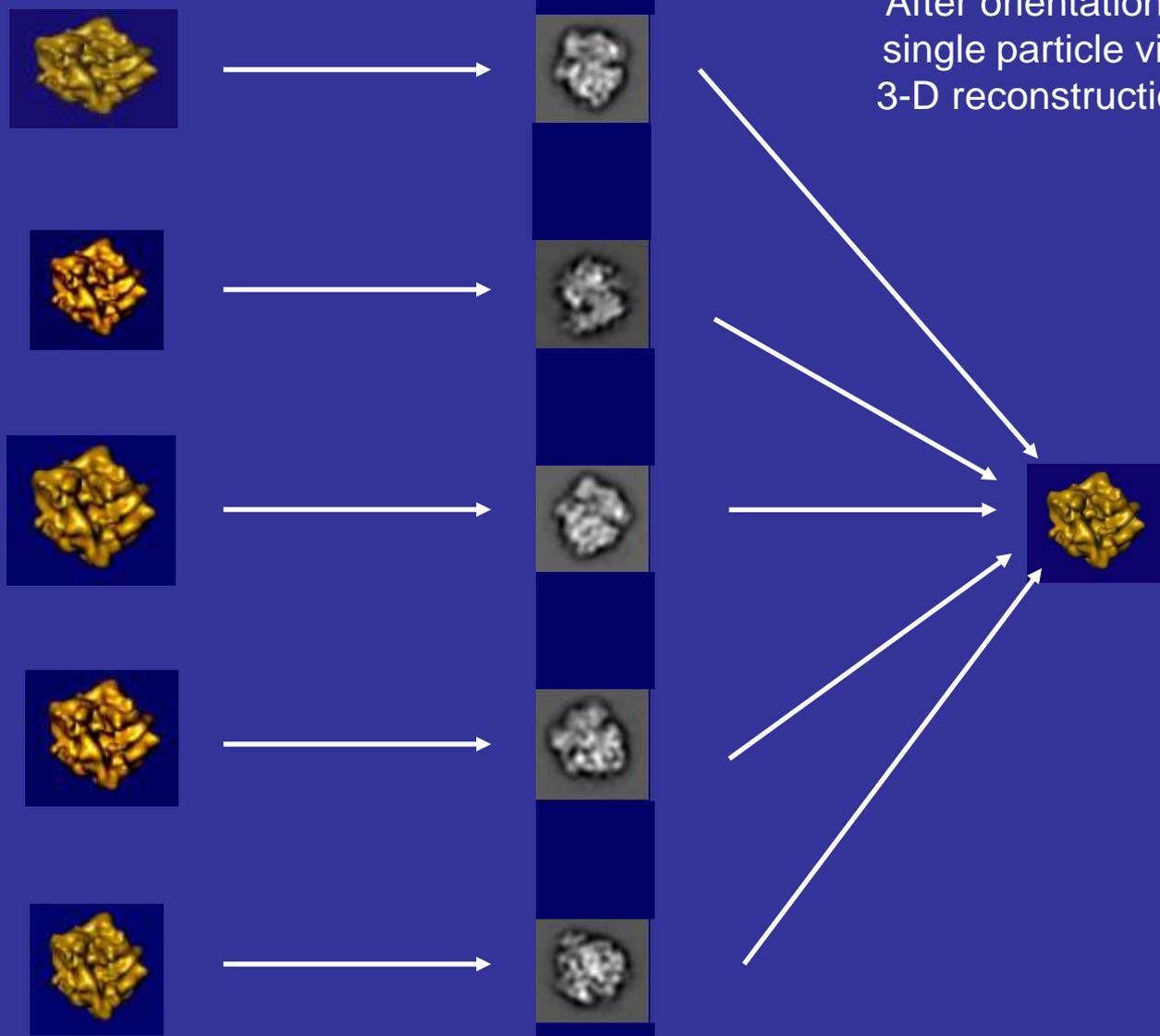
**In addition to the expected conformational heterogeneity of the assemblies that is due to fluctuations of the structure around the ground state, one can expect to capture molecules in different functional states, especially if the binding of a ligand induces a conformational change in the macromolecular assembly.**

***Therefore, data set of images from an EM experiment must be interpreted as a mixture of projections from similar but not identical structures.***

# In single particle analysis (cryo-EM), projections may originate from different 3D structures.

Different states of the same macromolecule (3-D).

In electron microscope, 2-D projections of observed macromolecules are formed.



After orientation parameters of single particle views are found, 3-D reconstruction is calculated.

# Computational time-resolved cryo-EM

- Multi-reference alignment
- Focused classification
- Multiple particle analysis

Structures of various conformers are determined using cryo-EM data that are taken at successive times from a system that is known to be developing in time.

# Real-space variance in single particle analysis

Images from an EM experiment must be interpreted as a mixture of projections from similar but not identical structures

- Detection of different functional states (caused by binding of a ligand)
- Significance of small details in 3-D reconstructions
- Conformational heterogeneity of the assemblies due to fluctuations of the structure around the ground state
- Significance of details in difference maps
- Fitting (docking) of known structural domains into EM density maps

# Calculation of a real space variance in 3-D reconstruction from projections is a difficult problem.

- The data is available in form of projections, i.e., information is partial.
- In single particle analysis (cryo-EM), the projections originate from different 3D structures.
- The main difficulty is that *there is only one data set*. In addition, even if we know that some macromolecules on the grid are identical, *we do not know which particle view corresponds to which macromolecule*.
- Exact inversion of the projection process is impossible. Thus, the step of 3D reconstruction itself is a source of noise.

$$g_{x,y} = \int f_{x,y,z} dz$$

## Variance of a 3-D structure calculated from the set of its line projections: closed-form solution.

Ray transform

$$\mathbf{g} = \mathbf{P}\mathbf{f}$$

We assume that an inverse transformation exists and that it is regularized by a 'smoothing' operator  $\mathbf{S}$ :

$$\hat{\mathbf{f}} = \mathbf{S} \mathbf{P}^T \mathbf{P}^{-1} \mathbf{P}^T \mathbf{g} = \mathbf{S} \mathbf{P}^\dagger \mathbf{g} = \mathbf{R}\mathbf{g}$$

$\mathbf{R}$  denotes any linear reconstruction algorithm.

The variance/covariance matrix of the reconstructed object :

$$\mathbf{C}_{\hat{\mathbf{f}}} = \mathbf{R}^T \mathbf{C}_{\mathbf{g}} \mathbf{R}$$

We need covariance matrix of the projection data  $\mathbf{g}$ !

# Resampling strategies

- Bootstrap
- Jackknife-d
- Jackknife

3-D reconstruction – weighted sum of the input projections with the weights dependent on the number and distribution of projections.

Backprojection

*(in real space)*

Voxel = algebraic (weighted) sum  
of projection pixels



Weighting

*(in Fourier space)*

Compensation for uneven  
distribution of projections in Fourier  
space

# Bootstrap technique

Resampling with replacements

Original data set of nine 2-D projections  
(k=9)    1 2 3 4 5 6 7 8 9



Resampled data sets of 2-D projections,  
each contains nine projections.    1 1 3 4 4 4 4 8 8    →

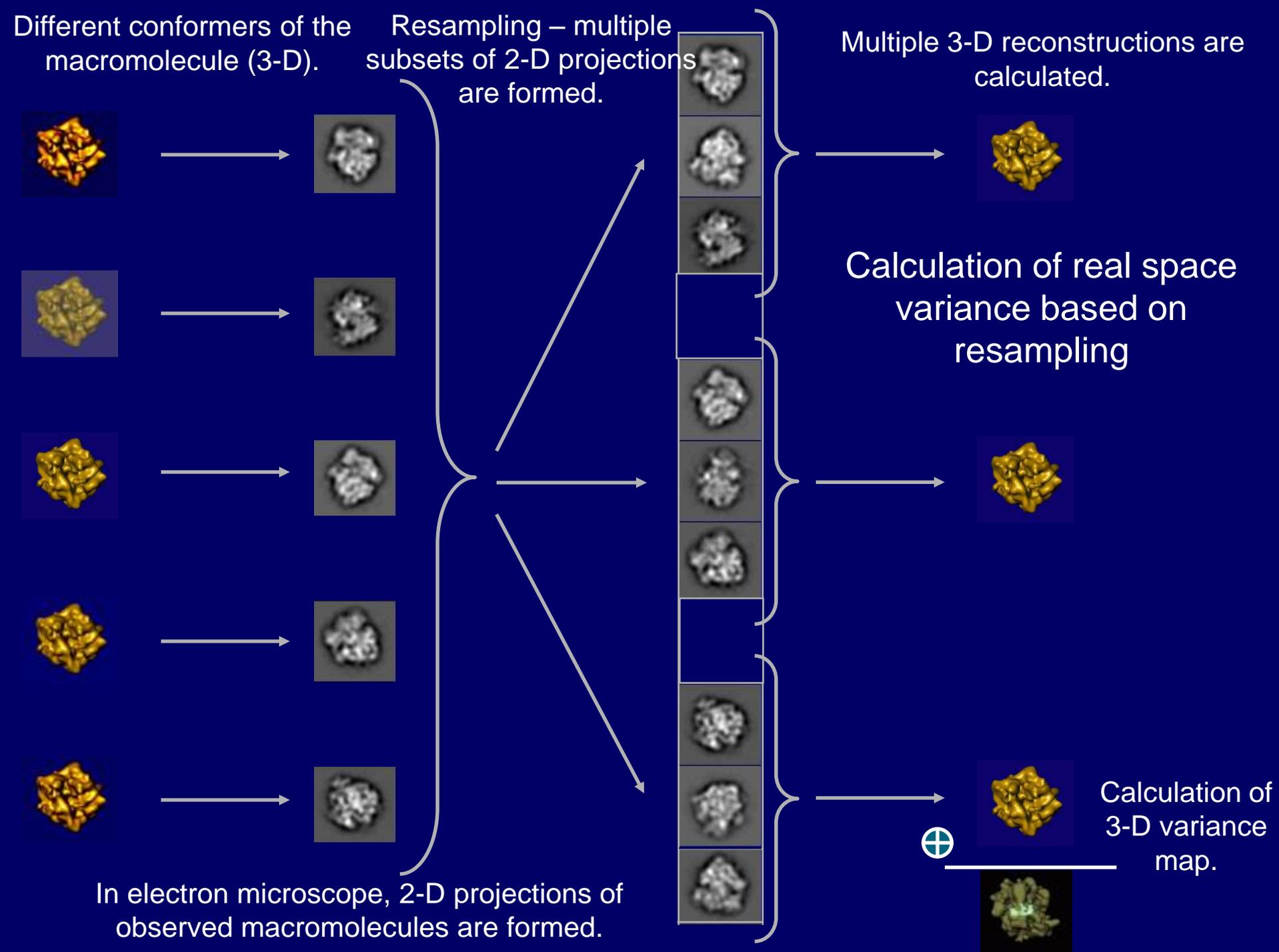
2 3 4 5 6 8 8 9 9    →

3-D reconstruction  
Large number of “different” volumes

- 
- 
- 
- 
- 
- 
- 

---

Variance/covariance!



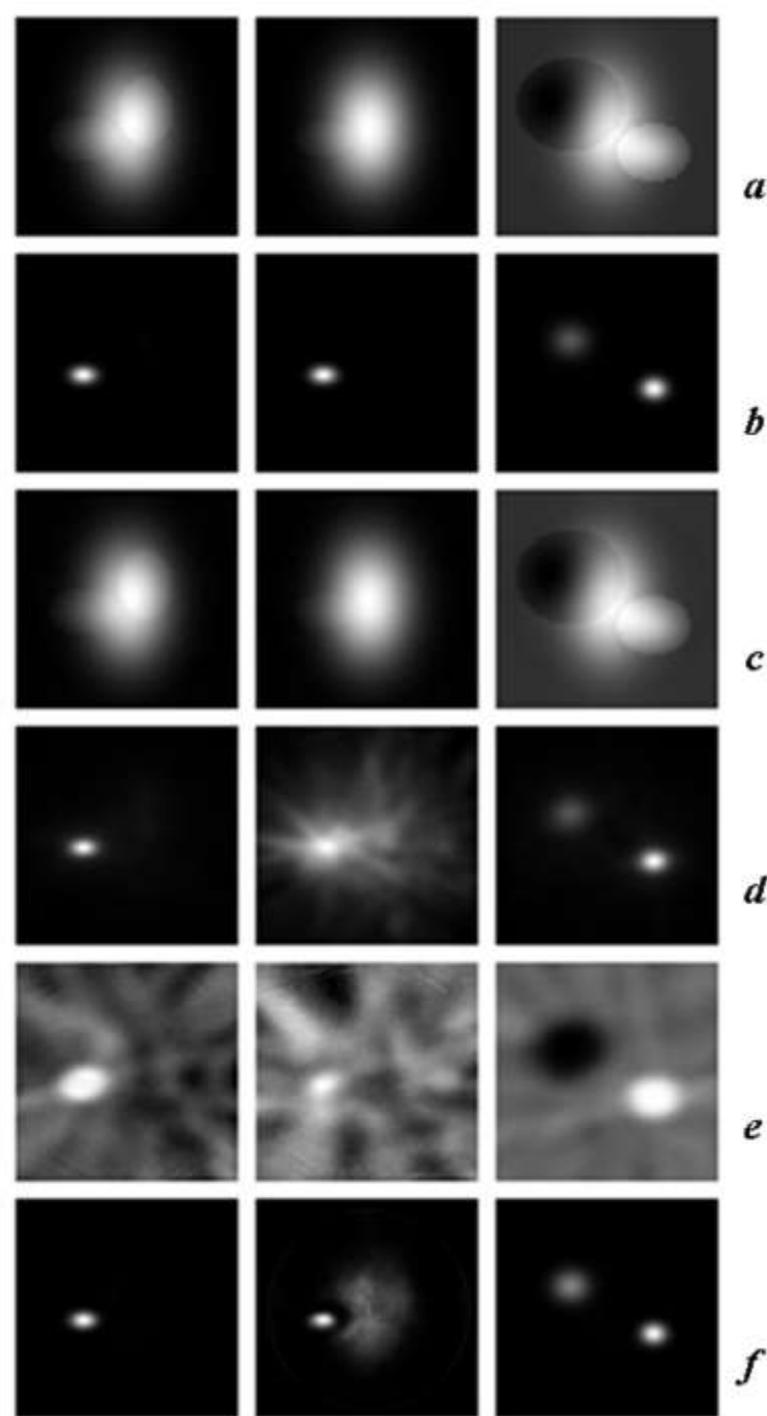
# Sources of variance in 3-D reconstructions

- Variability of the structure
- Noise in projection data
- Uneven distribution of projections
- Normalization errors in projections
- Numerical accuracy of the reconstruction algorithm

## Test of the bootstrap estimation of the structure variance in a noise-free case.

Contrast within each slice was adjusted independently, so the intensities do not reflect absolute values in respective slices.

- Average of model structures.
- The variance calculated using 1,253 simulated model structures.
- The average bootstrap structure.
- Structure variance calculated using the bootstrap method.
- Correlation map between the center of the feature *A* and the remaining voxels calculated using sample volumes.
- Variance calculated using the solvent variance estimation method, i.e., the expectation maximization algorithm.



# Components of bootstrap variance

$$\sigma_{SVar}^2 = \sigma_B^2 = \sigma_{Conf}^2 + \sigma_{Ali}^2 + \sigma_{Rec}^2 + \sigma_{Back}^2$$

- *Back* - background noise in projections
- *Rec* - reconstruction algorithm and distribution of projections
- *Ali* - alignment errors
- *Conf* - conformational variability of the 3-D structure (for example due to structural variability or non-stoichiometric binding of ligands).

$\sigma_f^2$ 

A-D: ROIs within respective variance/correlation maps. *Center*: ROI defined as a centrally located ball with radius 2 pixels. (1) variance of test structures. (2) variance of structures estimated using the bootstrap technique. (3) variance calculated using the method for the solvent variance estimation.

Correlation coefficients between the central center of the ROI A and the centers of all ROIs for the test (4) and bootstrap results (5), respectively. Correlation coefficients were averaged within respective ROIs.

1

2

3

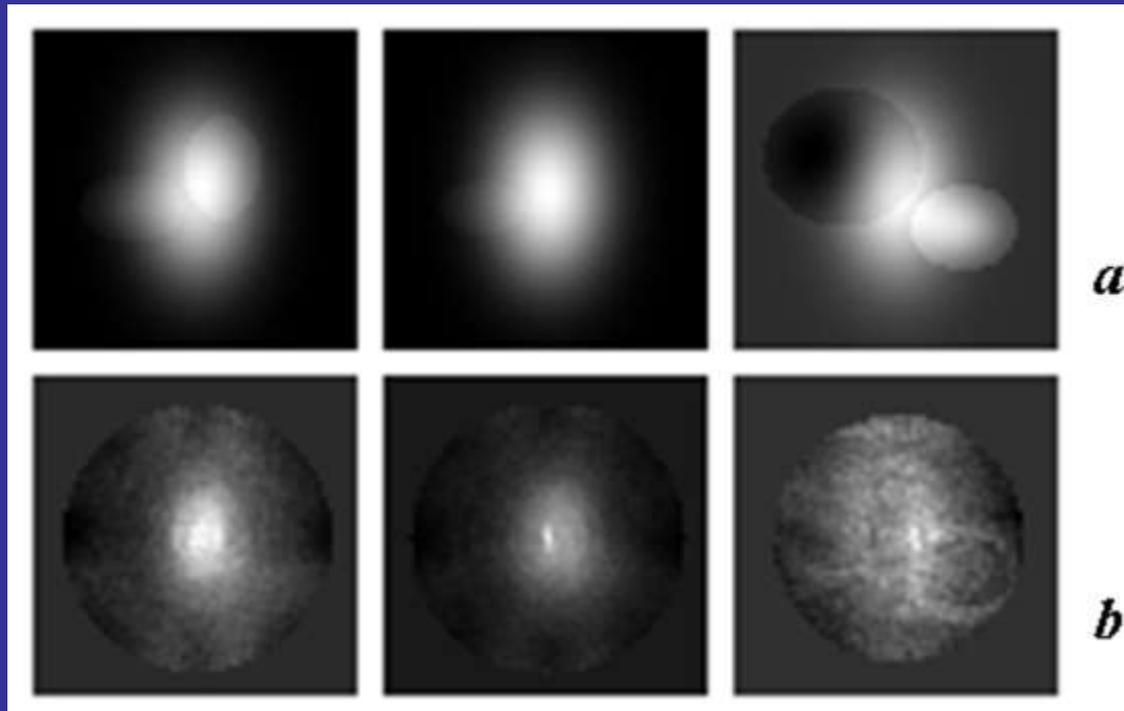
4

5

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>Center</i>
$\sigma_f^2$	2.25	1.49	0.68	2.37	$10^{-8}$
$\sigma_{Struct}^2$	2.51	1.79	0.93	2.76	0.55
$\sigma_{Solvent}^2$	15.5	12.2	7.78	17.7	0.77
$r_f$	1.00	0.02	-0.71	0.48	$10^{-7}$
$r_{Struct}$	0.86	0.02	-0.55	0.37	$-7 \times 10^{-3}$

# The variance due to the reconstruction process

*(Fourier inversion with NN interpolation and 4x padding)*



**Test volume  $64^3$  voxels.**

**1,253 quasi-evenly distributed 1,253 projections (4 degs angular step ).**

**The bootstrap sample size of the sample  $B = 500$ .**

**The CCC between the average bootstrap structure and the model structure 0.99996.**

**The densities of the structure are between -2.85 and 10.2, and their variance 4.24.**

**The variance values are between  $4.67 \times 10^{-6}$  and  $5.35 \times 10^{-5}$  with an average of  $1.17 \times 10^{-5}$ .**

**The SNR of the reconstruction is  $\sim 10^5$ . It could be increased if appropriate low-pass filtration of bootstrap volumes was applied.**

## Calculation of the background variance using micrograph noise samples and the bootstrap technique

1. Select samples of the background noise from micrographs. Their number has to be the same as the number of available projection images.
2. Apply the bootstrap technique to calculate the 3-D variance map of the background noise, it will also contain the reconstruction variance.
3. Calculate the average level of the background variance within the 3-D region corresponding to the support of the structure.

$$\overline{\sigma}_{Back}^2$$

# Calculation of the variance of structures

$$\sigma_{Struct}^2 = K \sigma_B^2 - \overline{\sigma}_{Back}^2$$

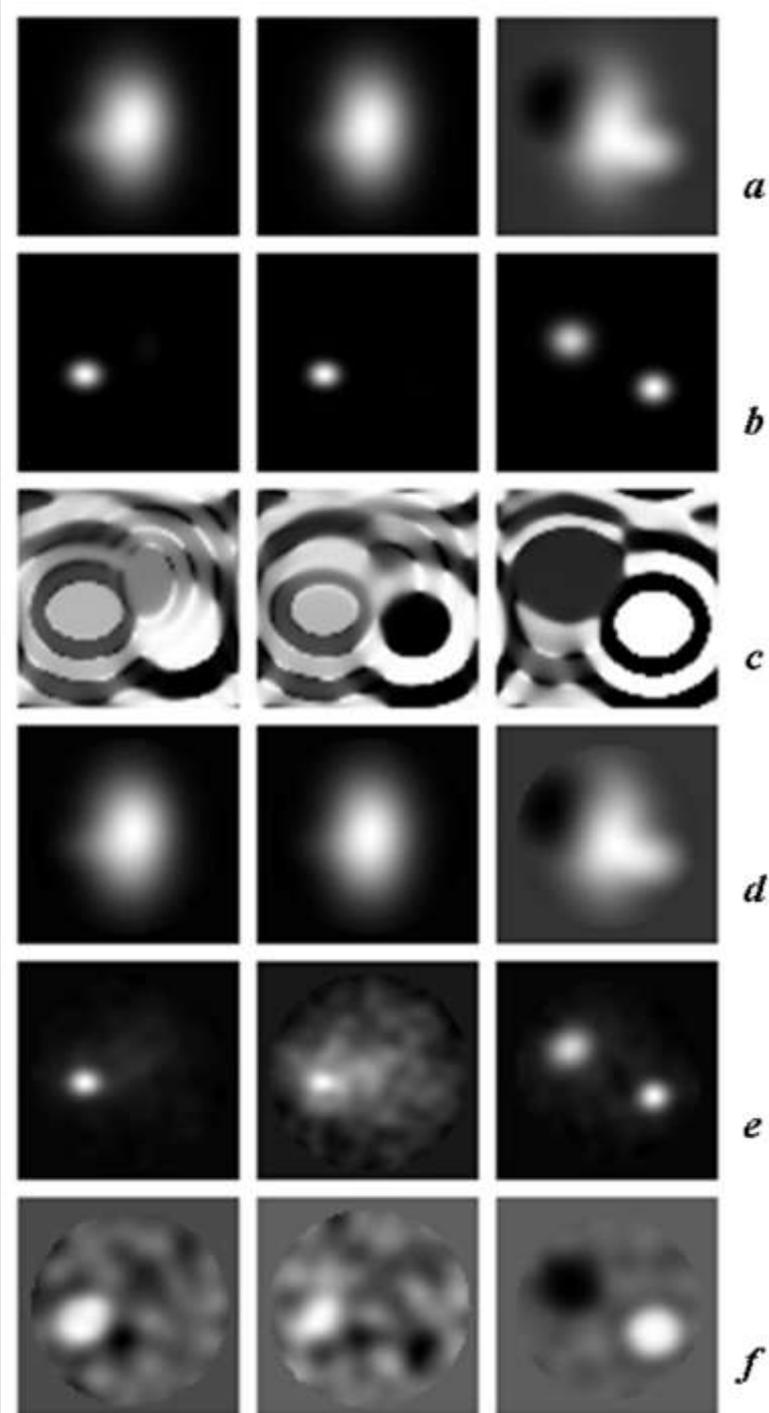
We disregard the variance arising from alignment errors, as there is no method to estimate it independently.

Test in the presence of additive noise  
 $N(0,30)$ ,  $\text{SNR} = 2.3$  in the projection data.

$B = 500$  bootstrap volumes

- (a) Average of low-passed model structures.
- (b) The variance calculated using 1,253 simulated low-passed model structures.
- (c) Correlation map between the center of the feature  $A$  and the remaining voxels calculated for simulated low-passed volumes. The unusual pattern is due to correlations introduced into the volumes by the process of low-pass filtration.
- (d) The average of low-passed bootstrap structures.
- (e) Structure variance calculated using the bootstrap method and estimated from low-passed sample volumes.
- (f) Correlation map between the center of the feature  $A$  and the remaining voxels calculated using low-passed bootstrap volumes.

Contrast within each slice adjusted independently, so the intensities do not reflect absolute values in respective slices.



Test of the estimation of the structure variance using the bootstrap method in the presence of additive independent Gaussian in projections.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>Center</i>
$\sigma_f^2$	1.17	1.24	0.65	1.35	$4 \times 10^{-2}$
$K\sigma_{SVar}^2$	1.51	1.59	1.14	1.83	0.77
$\sigma_{Struct}^2$	1.19	1.28	0.82	1.52	0.46
$r_f$	1.00	0.00	-0.70	0.44	0.19
$r_{SVar}$	0.87	-0.02	-0.44	0.34	$-7 \times 10^{-3}$

$$\sigma_{Struct}^2 = K \sigma_{SVar}^2 - \bar{\sigma}_{Back}^2$$

# 3-D reconstruction algorithm

**A direct Fourier inversion 3-D reconstruction algorithm incorporates full per-image CTF correction using the Wiener filter approach.**

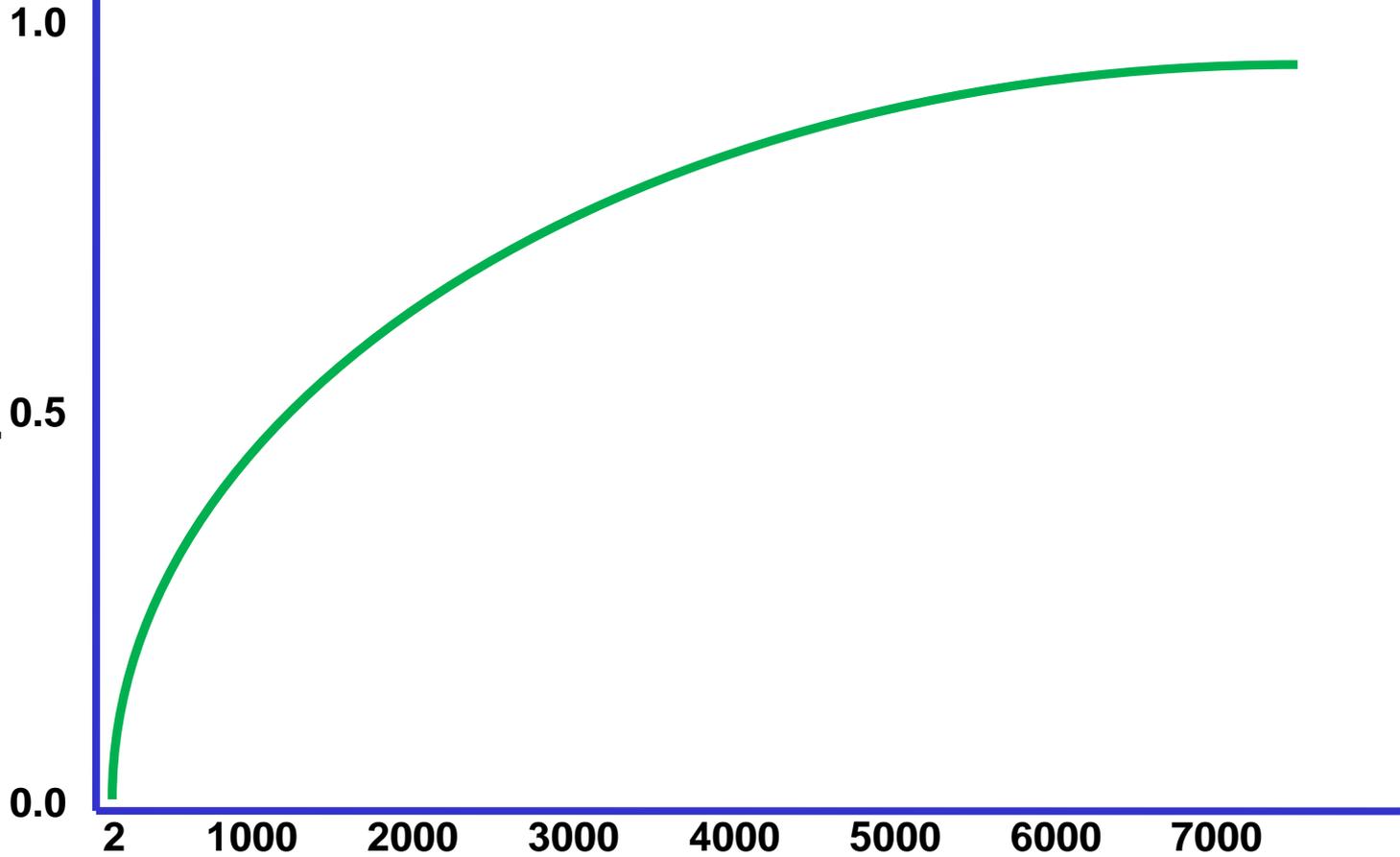
**2D images are four times padded with zeros and Fourier transformed. Next, these Fourier 2D images are multiplied by respective CTFs and inserted into 3D volume using a nearest neighbour interpolation. After all projections are processed, uneven coverage of 3D Fourier space by contributions from 2D projections is accounted for by weighting voxels according to the fraction of the volume they occupy. The precise weights are computed using tessellation of the 3D Fourier volume (computed using Voronoi diagram approach). After 3D inverse Fourier transform, the relevant part of the volume is windowed in real space yielding CTF-corrected 3D reconstruction.**

**It takes ~3 minutes of a single CPU time to produce a resampled volume for a set of 30,000 projection images  $75^2$ .**

**Thus, using 128 nodes of a cluster and a set of we can generate ~25,000 bootstrap volumes in 10 hours.**

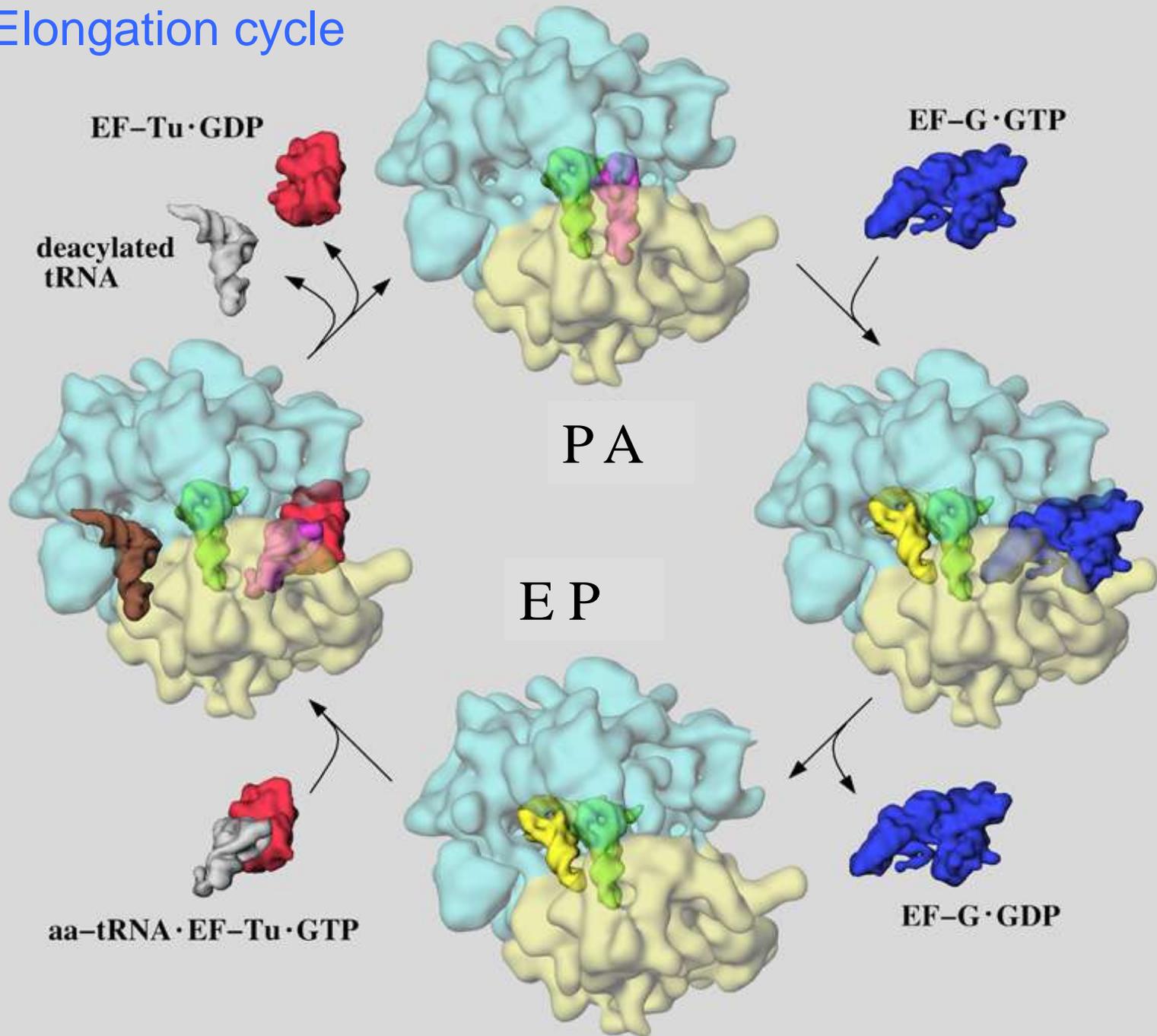
# Determination of the number of bootstrap volumes

ccc between variance maps calculated  
using even- and odd-numbered  
bootstrap volumes



Conformational modes of ribosomal  
*T. thermophilus*  
70S EF-G complex

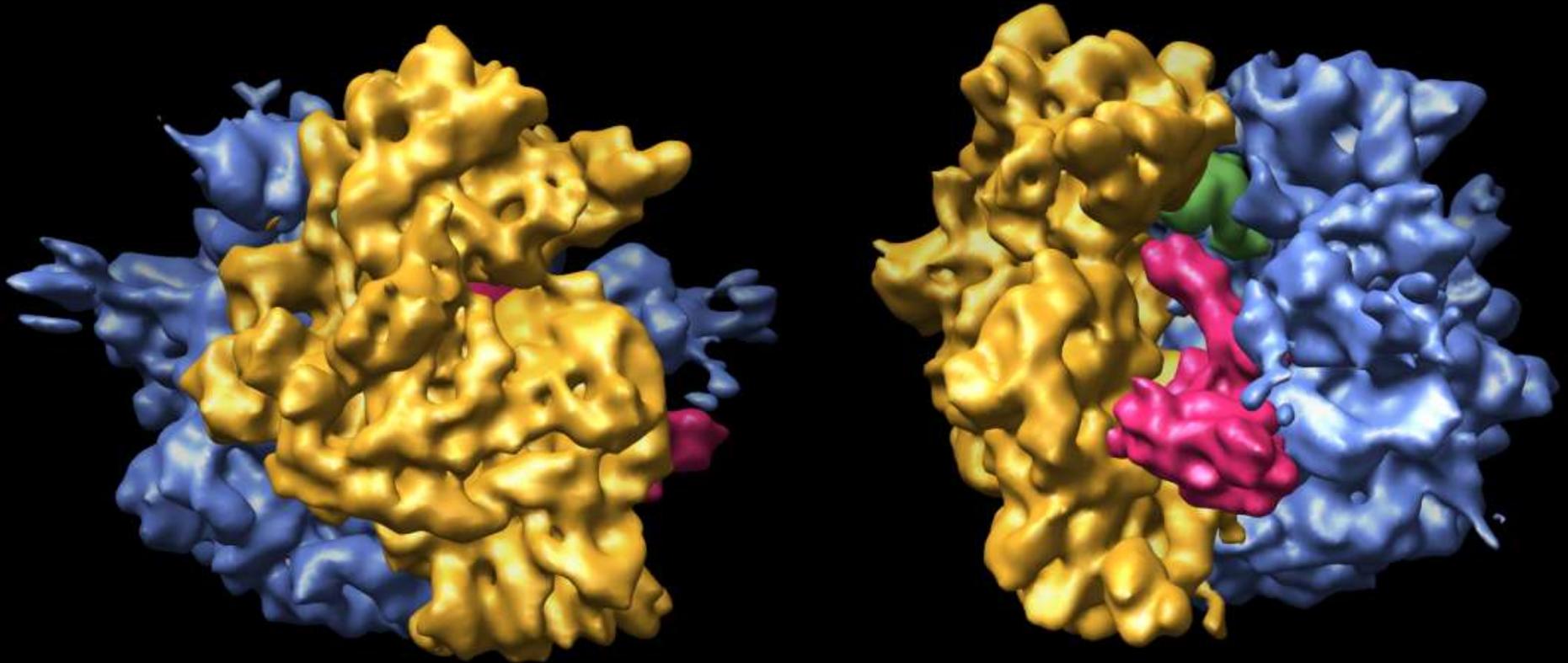
# Elongation cycle



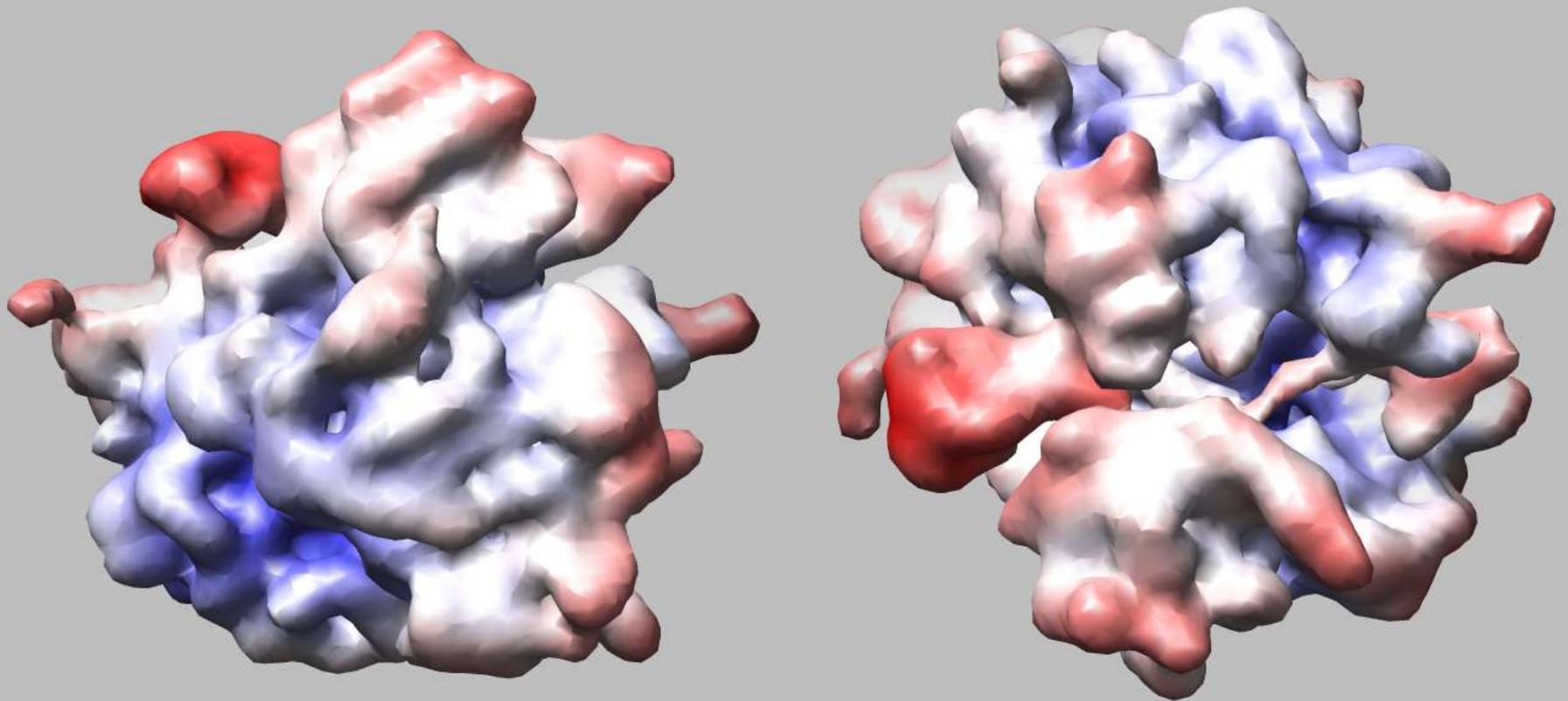
# 70S *Thermus thermophilus* ribosome complexed with G protein EF-G in the presence of the non-hydrolyzable GTP analogue GMPPNP.

- *T. thermophilus* 70SEF-G complex in which EF-G was stalled on the ribosome using the non-hydrolysable GTP analogue GMPPNP, The occupancy of EF-G in the complexes was ~60-70%.
- Tecnai G2 Polara at 300 kV, 39,000x magnification, pixel size 1.22 Å, 362,361 particle projections
- The original structure determined to 7.5 Å resolution (FSC @0.5).

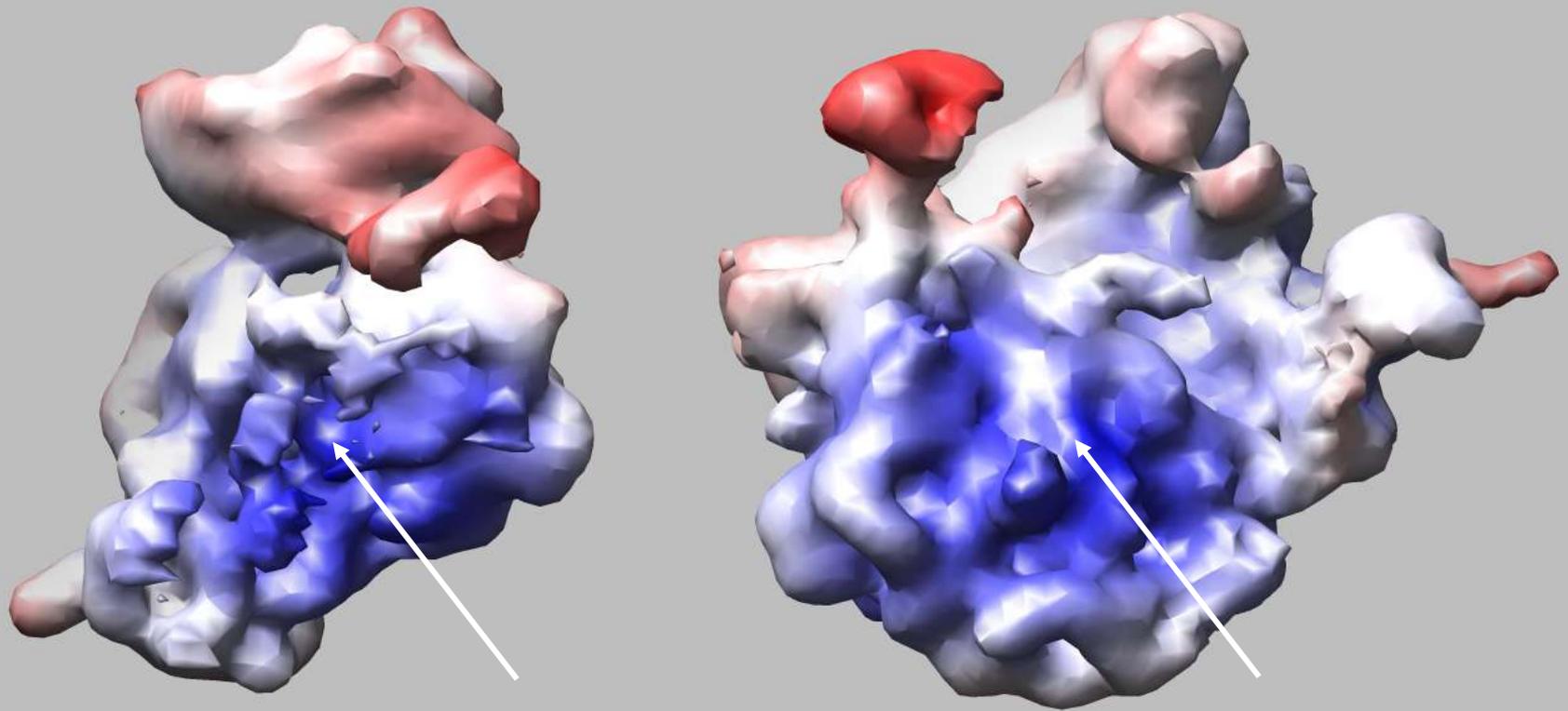
# 70S•EF-G•GMPPNP complex

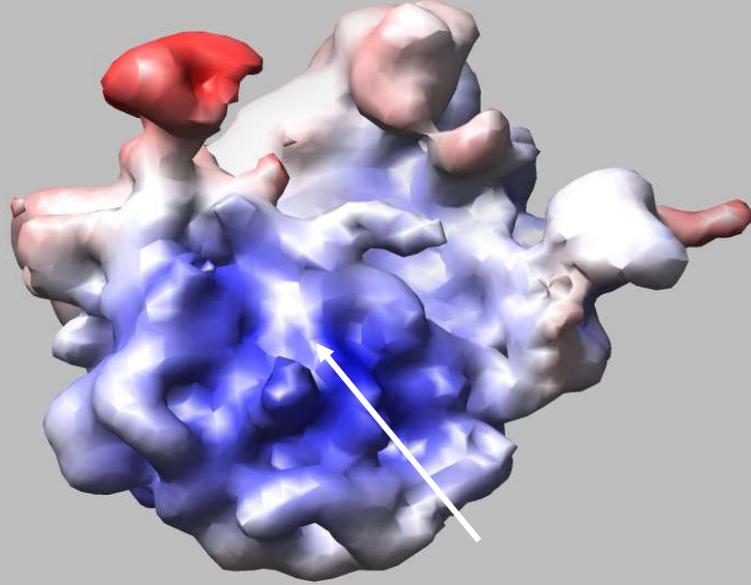
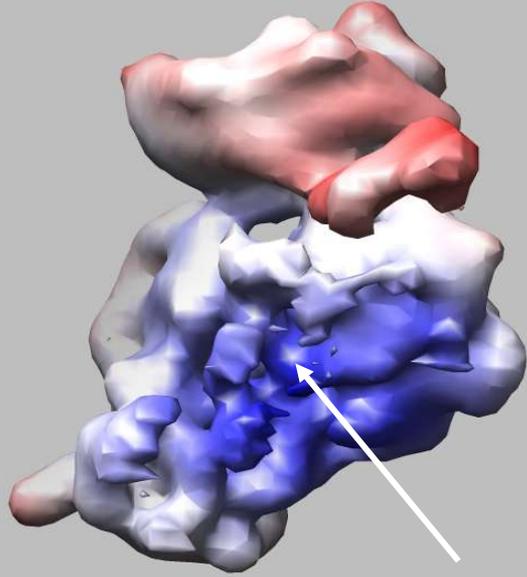
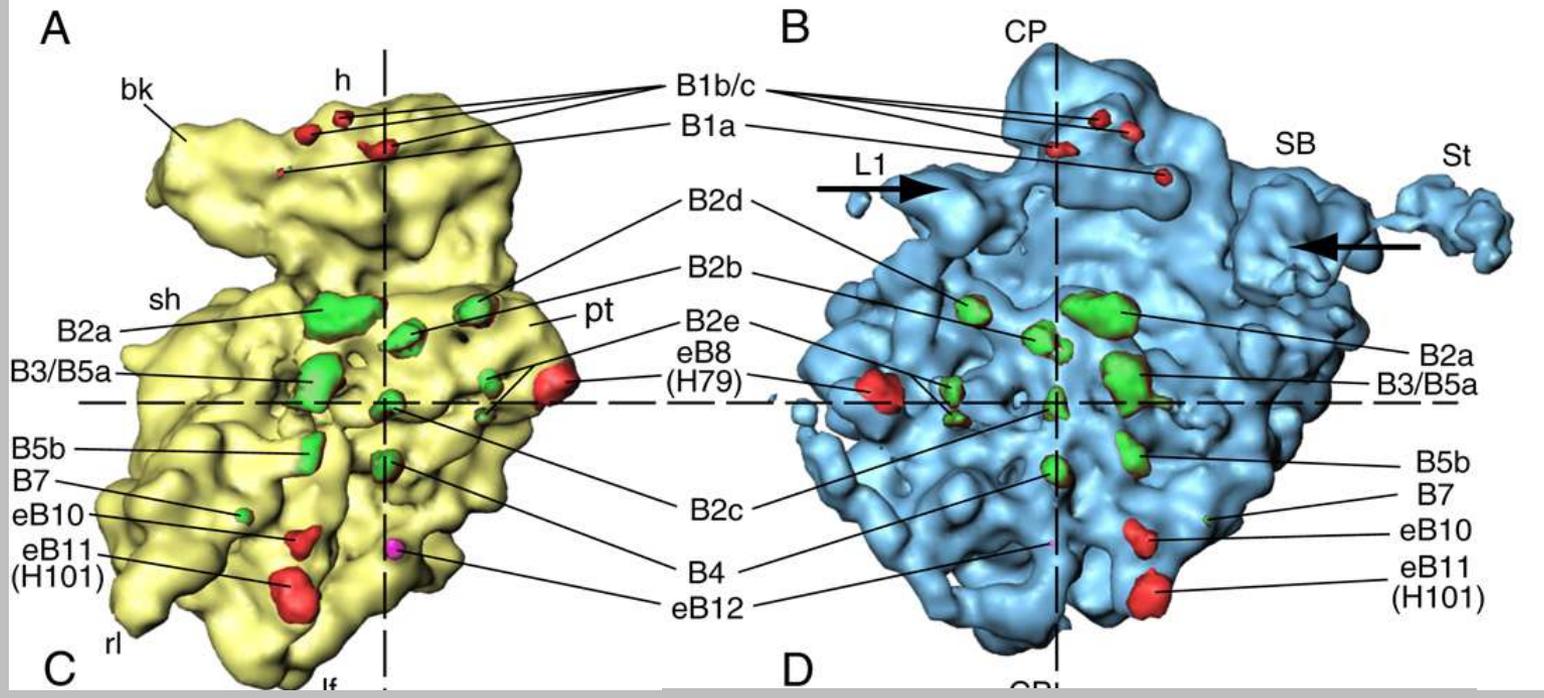


# Variance map



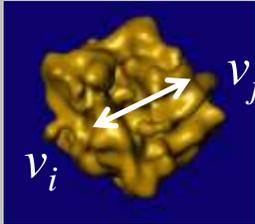
# Variance map



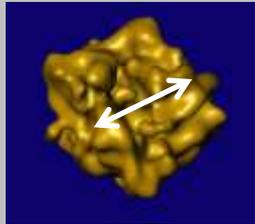


**Spahn, C. M., [...] Frank, J. Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. EMBO J. 2004.**

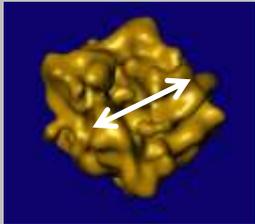
## **B** bootstrap 3-D reconstructions, pair-wise correlations



For a volume size  $n^3$ ,  
there are  $\sim n^6$  pair-wise correlations ( $\sim 10^{12}$ )!



Impossible to visualize/analyze.



Perform eigenanalysis of bootstrap volumes:  
eigenvectors (eigenvolumes) provide  
information about variability of the structure,  
i.e., **conformational modes** of the  
structure.

---

$$C_{ij} = \sum_{l=1}^B (v_i^l - \bar{v}_i) (v_j^l - \bar{v}_j)$$

## 3-D classification of conformations using bootstrap technique

1. Calculation of the eigenvalues and eigenvectors using the entire available dataset.
2. Calculation of the eigenvalues and eigenvectors using bootstrap technique.
3. Calculation of the eigenvalues and eigenvectors using bootstrap technique.
4. Eigenanalysis of the covariance matrix yields eigenvolumes.
5. Calculation of the eigenvalues and eigenvectors using a small subset of the data.
6. Cluster analysis in the reduced factorial space facilitates determination of plausible number of distinct conformers and rapid determination of seed structures.
7. Calculation of the eigenvalues and eigenvectors using bootstrap technique.
8. 3-D multivariate analysis yields coordinates yields.
9. Within-group comparison of structures yields resolution nearly homogeneous structures.

*For mildly heterogeneous datasets, particle projection that originate from similar molecules.*

*Cluster analysis in the reduced factorial space facilitates determination of plausible number of distinct conformers and rapid determination of seed structures.*

# 3-D classification of projections using bootstrap technique

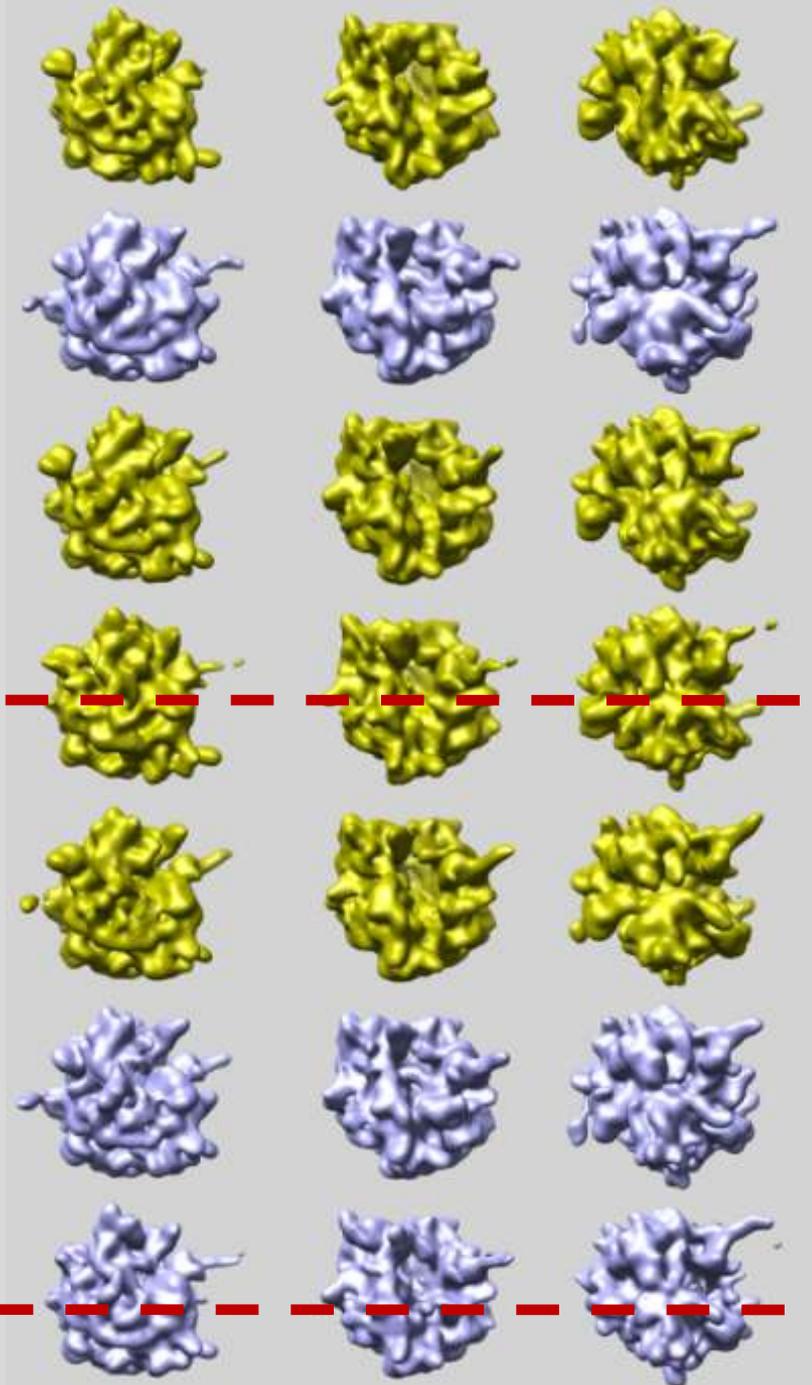
Computational challenges:

1. If precise initial classification is sought, a very large number of resampled volumes might be required ( $\sim 100,000$ ): the problem is trivially parallelizable and the time of calculation depends only on the number of available nodes of an MPI cluster. However, the required disk space might be very large.
2. If the number of structures  $K$  is large, the 3-D multi-reference alignment is time consuming. A better strategy is to separate the problem into within-class 3-D projection alignment and 3-D  $K$ -means (reassignment of projection data based on projections of reference structures without correction of angles).

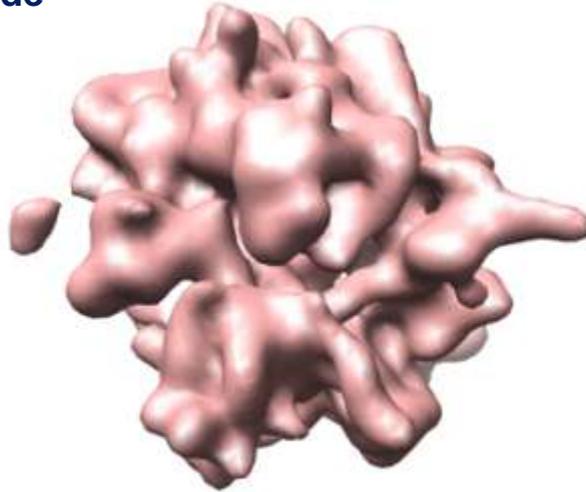
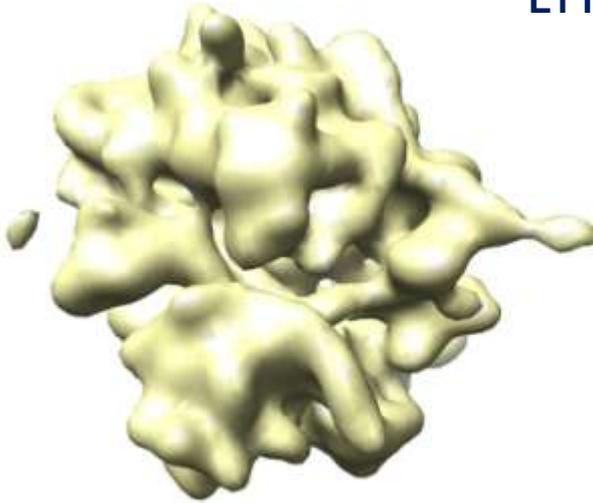
**Results of the multi-reference  
projection refinement**

1.	31,376	8.7%
2.	123,304	34.0%
3.	57,753	15.9%
<del>4.</del>	<del>21,346</del>	<del>5.9%</del>
5.	53,975	14.9%
6.	50,731	14.0%
<del>7.</del>	<del>23,876</del>	<del>6.6%</del>
total:	362,361	100.0%

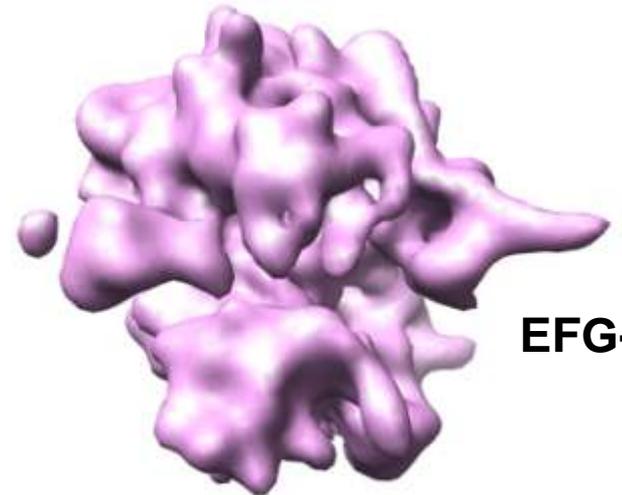
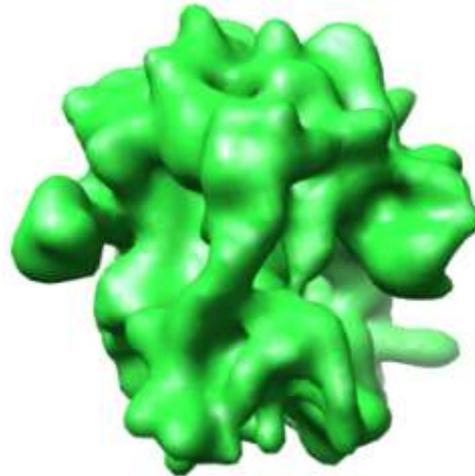
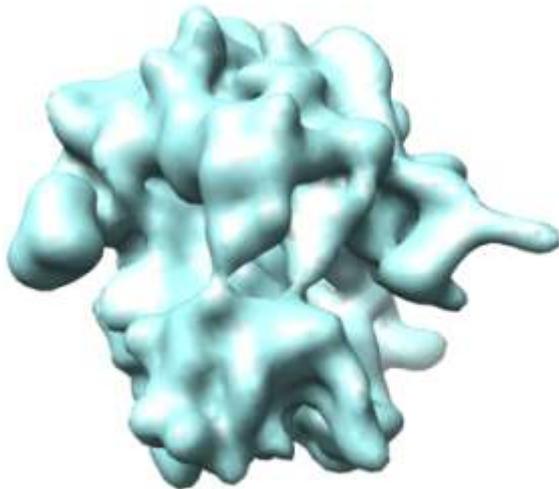
*All structures low-passed filtered to the  
lowest common denominator ~20Å.*



**tRNA in E-site hybrid state  
L1 inside**



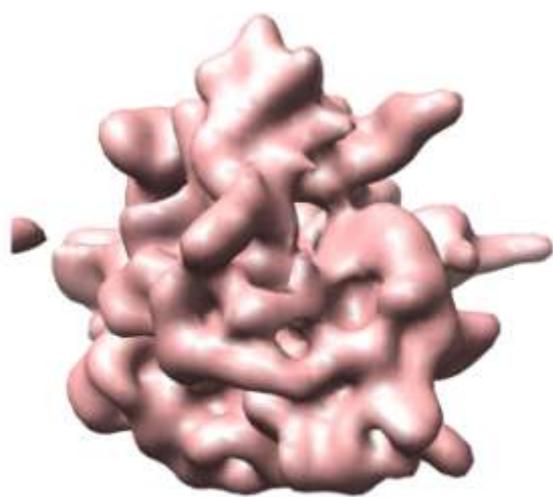
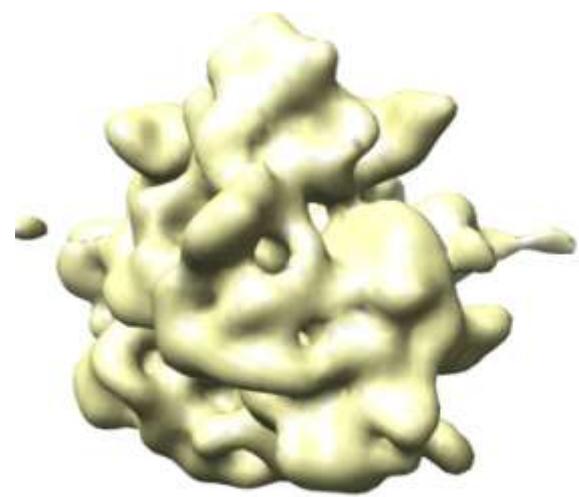
**EFG+**



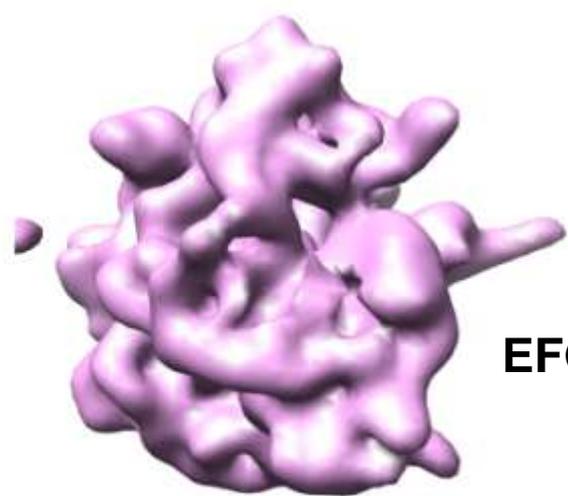
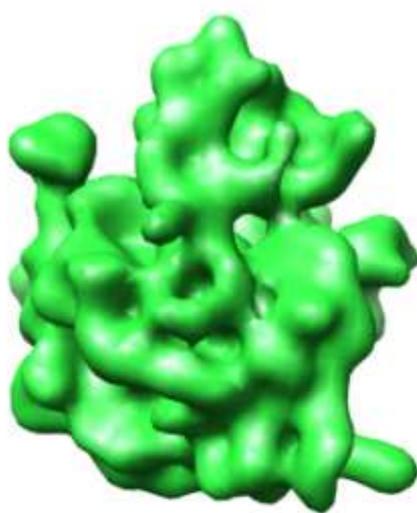
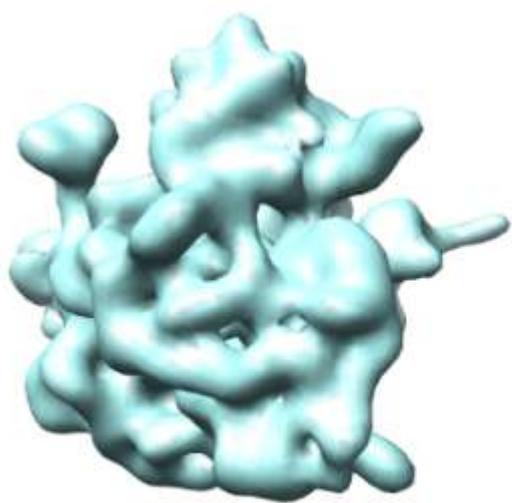
**EFG-**

**No extended stalk and stalk-base in IN  
position.  
Strong interaction between 50S CP  
and 30S head.**

**E-site tRNA present.  
Causes inward-movement of the  
L1 protuberance.**

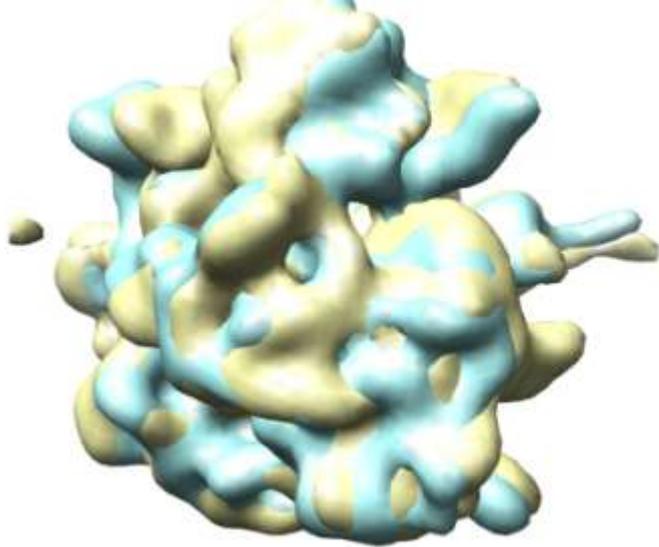


**EFG+**

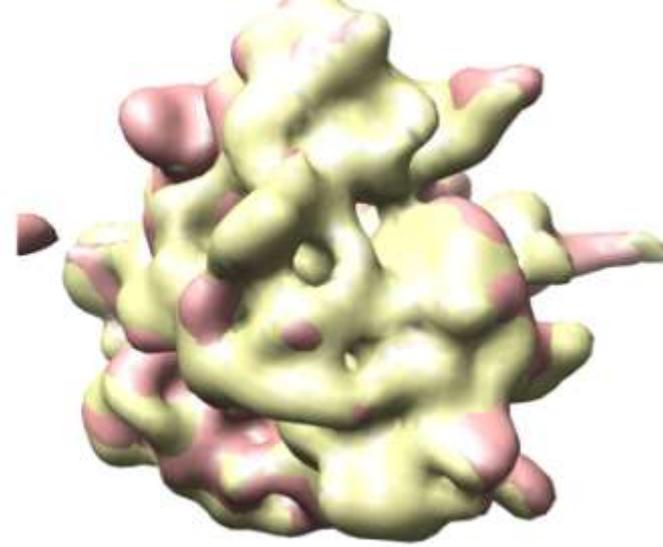


**EFG-**

**Large ratchet movement  
between EFG+ and EFG-**

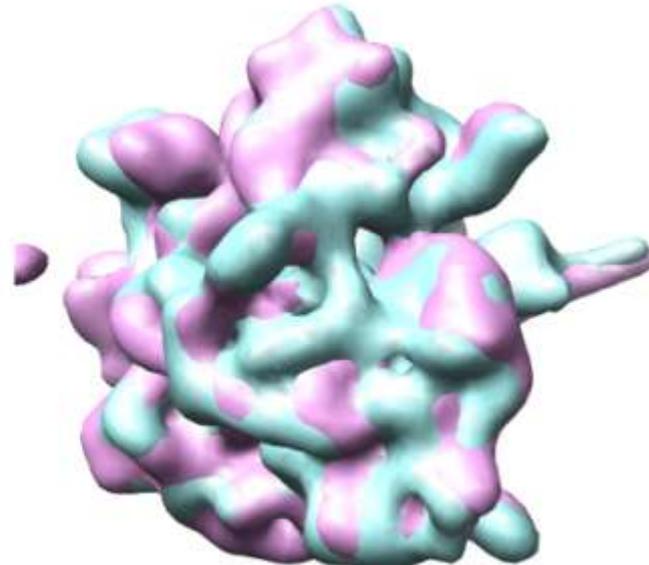
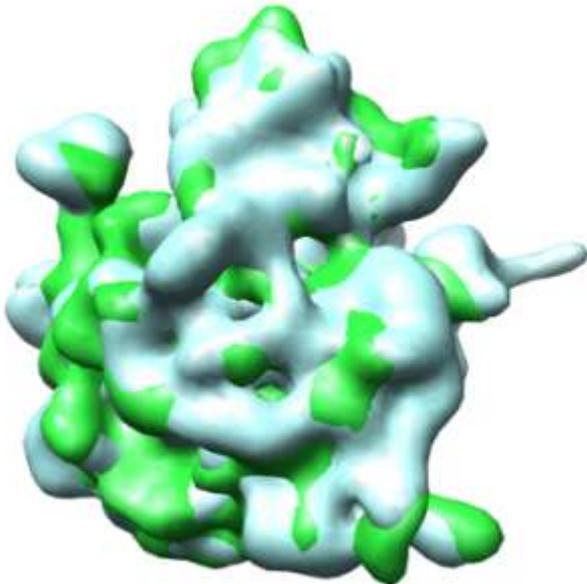


**Small 30S rotation in  
EFG+ structures**



**EFG+**

**Four macrostates: different positions of 30S subunit**

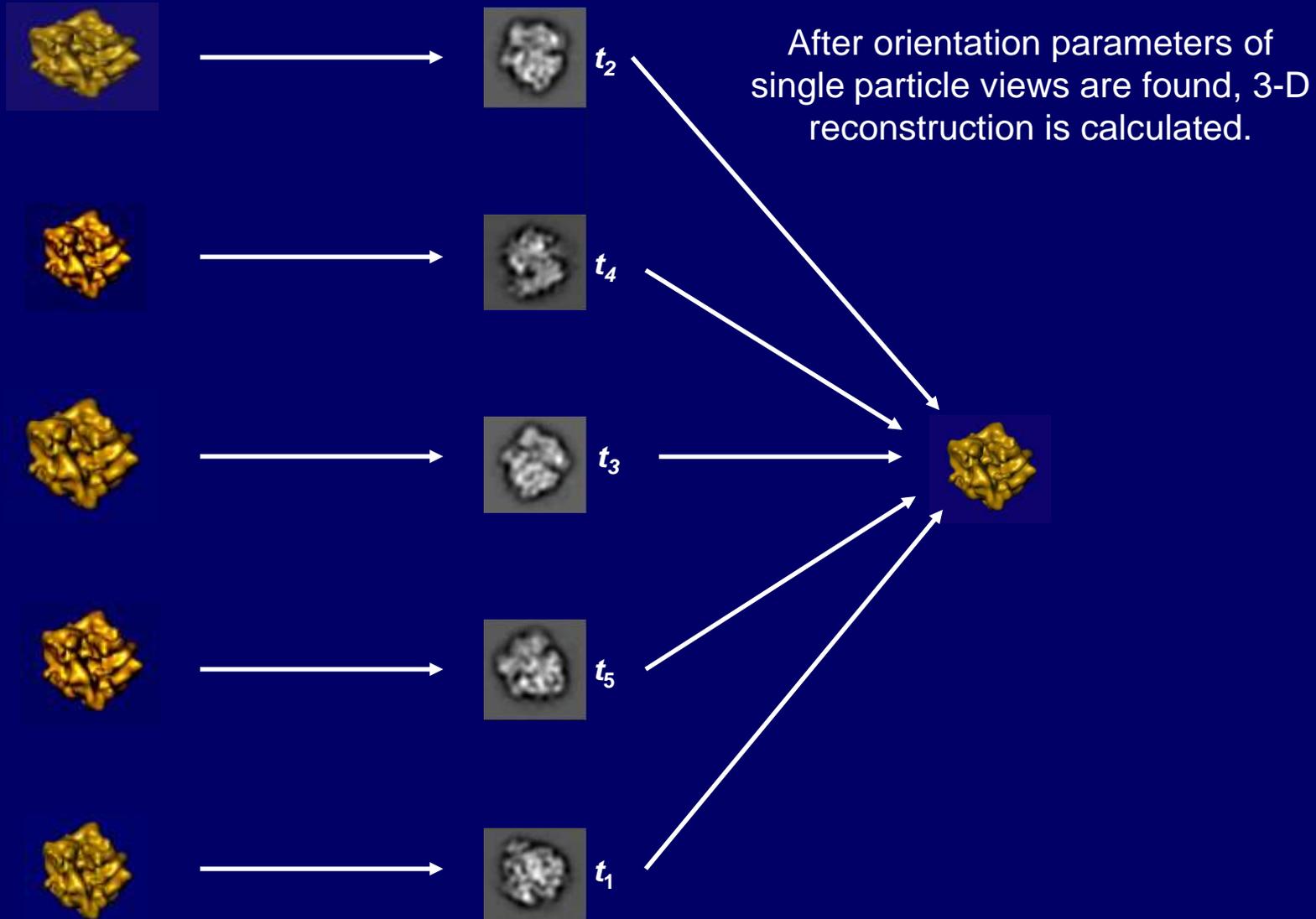


**EFG-**

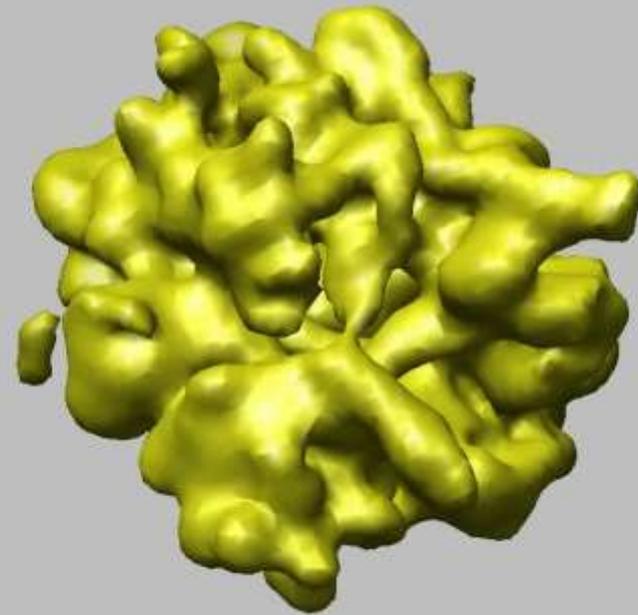
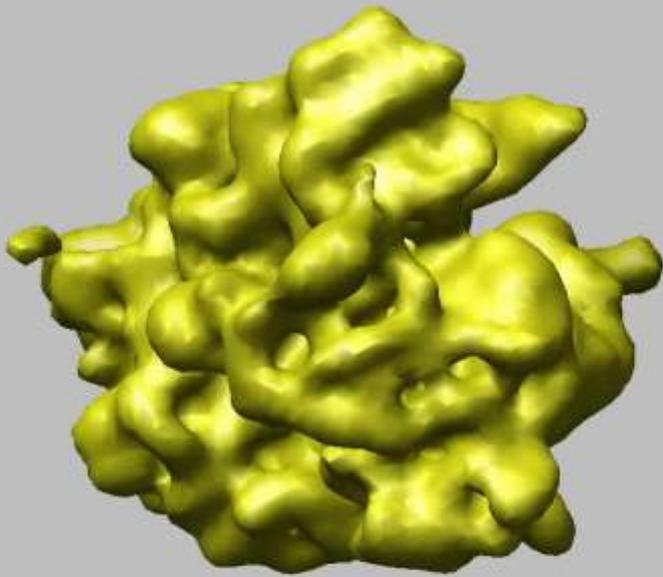
Inward movement of L1 due to E-tRNA and 30S subunit rotation that is smaller but in the same direction as the RSR.

Towards 4-D single particle analysis....

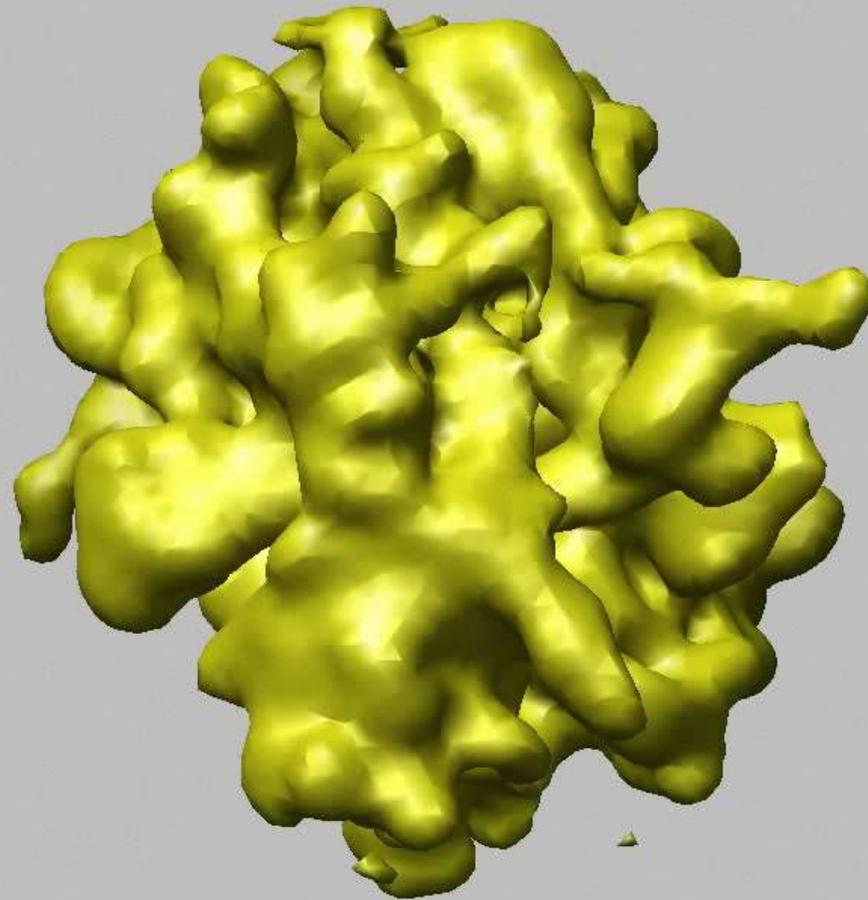
Different conformations of the same macromolecule (3-D).  
In electron microscope, 2-D projections of observed atoms are captured.  
In electron microscope, projections originate from different conformers of the macromolecule.



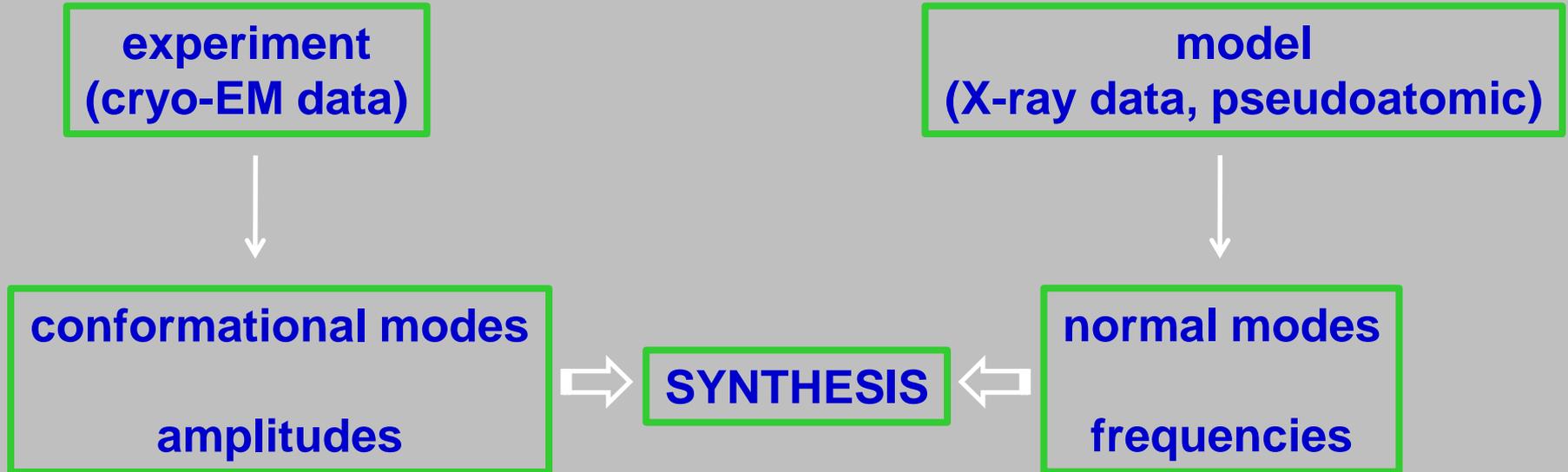
# Ratchet movement



# Bridge formation



# Validation of conformational modes



# Conclusions/Future work

- ✓ Resampling yields per-voxel variance (error) in cryo-EM structures reconstructed from sets of their projections: reliability of structure determination, sample homogeneity.
- ✓ Conformational modes: classification of projections, structure dynamics, truly 4-D cryo-EM.
- ✓ In general, the method works, but...
- ? Resampling does not work if distribution of projections is uneven, i.e., for all practically encountered cases: symmetry?
- ? Resolution limits.
- ? Interpretability limits:  
distinction between discrete states and genuine subunit movements;  
physically/biochemically plausible conformational changes;  
relation to MD modeling.
- ? Any method will yield results only as good as the quality of the sample permits...

# Acknowledgments

**Wei Zhang, UTH**



**Christian M.T. Spahn, Charité, Berlin.**



**NIH**