ANALYSIS OF CONFORMATIONAL HETEROGENEITY OF MACROMOLECULES IN CRYO-ELECTRON MICROSCOPY

Pawel A. Penczek

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

Outline

- Real-space analysis of 3-D variance/covariance in macromolecules reconstructed from a set of their projections
- 3-D classification of projections
- Functional states and conformational variability in the ribosomal elongation cycle
- Analysis of the 70S *E. coli* ribosome complexed with the elongation factor G (EFG)-GDP

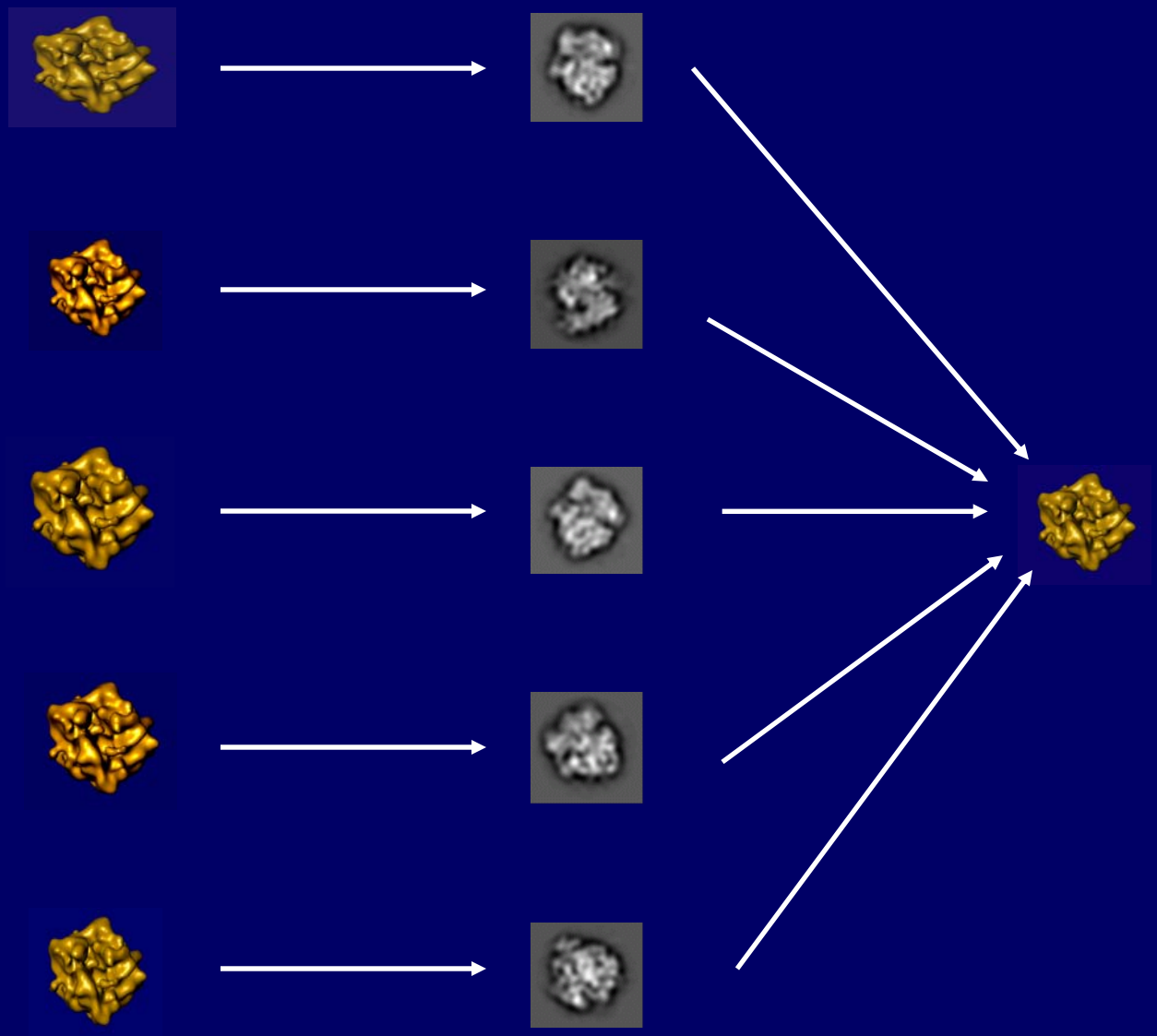
Real-space variance in single particle analysis

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

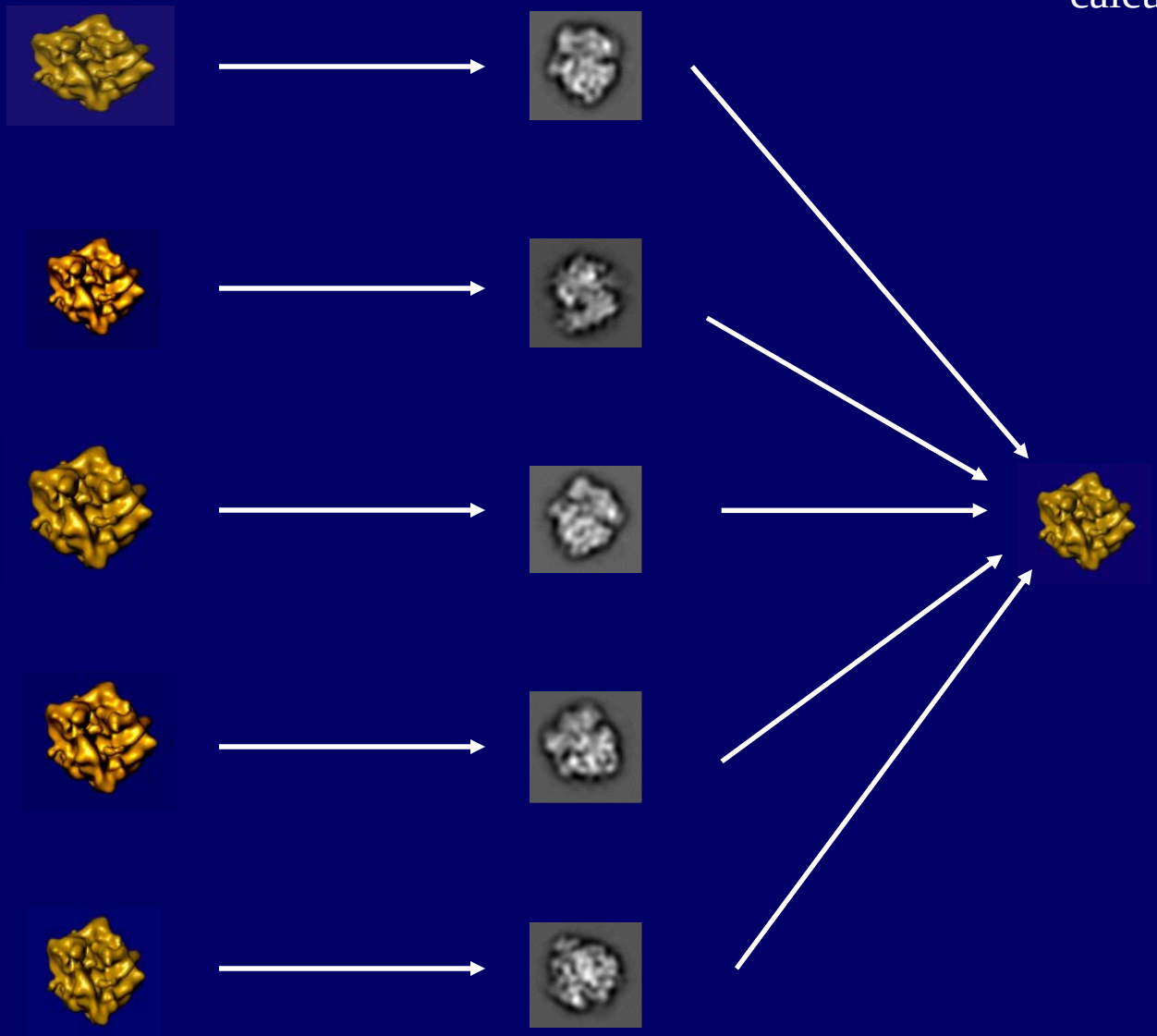
In single particle analysis (cryo-EM) ,projections originate from different 3D structures.



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

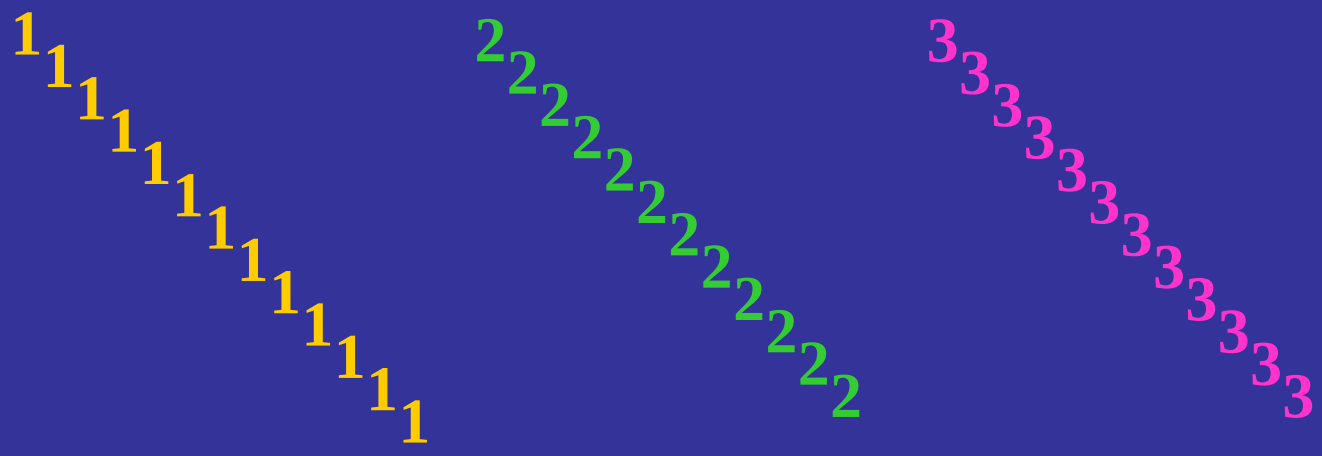
In electron microscope, 2-D projections of observed macromolecules are formed.

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



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., the 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 process of 3D reconstruction is a source of noise.



3-D reconstruction

Total average

2

3-D reconstructions

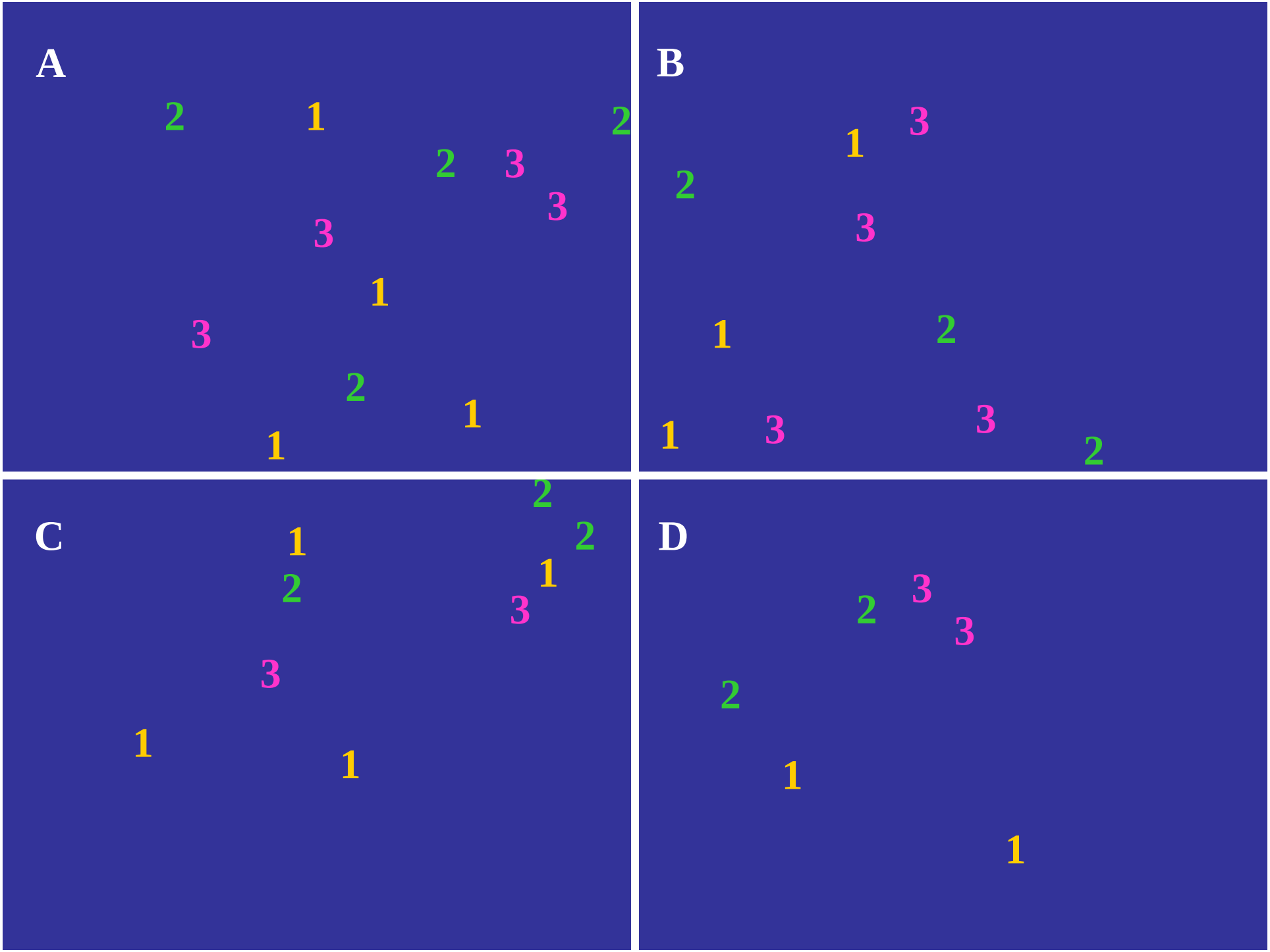
averages

1 2 3

Variance of structures

$$\sigma^2 = \frac{1}{3}\sum_{i=1}^{3}\left(x_i - \bar{x}\right)^2 = \frac{1}{3}\left[\left(1-2\right)^2 + \left(2-2\right)^2 + \left(3-2\right)^2\right] = \frac{2}{3} = 0.66$$

$$\sigma = \sqrt{0.66} = 0.82$$


A
2
1
2
2
3
3
3
1
3
2
1
1
B
3
1
2
3
1
2
1
3
3
2
C
2
1
2
2
1
3
3
1
1
D
3
2
3
2
1
1

3-D reconstruction
Total average

2

A	B	C	D
2.3	1.7	2.1	1.8

3-D reconstructions
averages

Variance?

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(x_i - \bar{x})^2 = \frac{1}{4}\left[(2.3-2)^2 + (1.7-2)^2 + (2.1-2)^2 + (1.8-2)^2\right] = \frac{0.23}{4} = 0.058$$

$$\sigma = \sqrt{0.058} = 0.24$$

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

Resampling strategies

- Bootstrap
- Jackknife-d
- Jackknife

Bootstrap technique

Resampling with replacements

Original data set of nine 2-D projections (k=9)

1 2 3 4 5 6 7 8 9

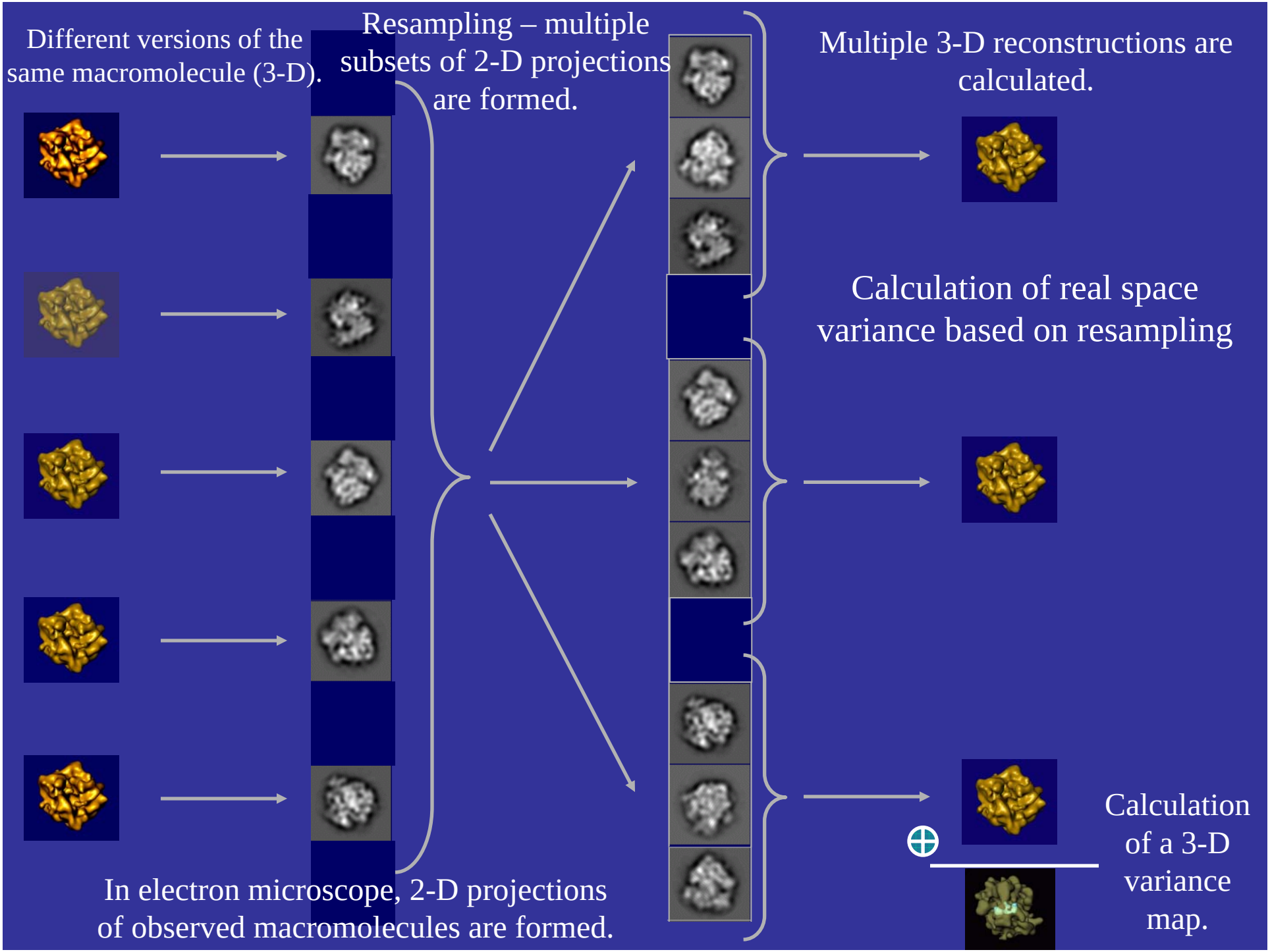
Resampled data sets of 2-D projections, each contains nine projections.

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!


Different versions of the same macromolecule (3-D).
Resampling – multiple subsets of 2-D projections are formed.
Multiple 3-D reconstructions are calculated.
Calculation of real space variance based on resampling
Calculation of a 3-D variance map.
In electron microscope, 2-D projections of observed macromolecules are formed.

Comparison of resampling techniques

Resampling technique	Sampling	Number of possible volumes	Multiplicative factor K for the variance	Number of projections changed per volume
Bootstrap	with replacements	$\frac{(2k-1)!}{(k-1)!k!}$	k	~63%
Jackknife-d	without replacements (m selected from a set of k)	$\frac{k!}{m!(k-m)!}$	$\frac{km}{k-m}$	MIN(m, k-m)
Jackknife	leave one out	k	k(k-1)	1

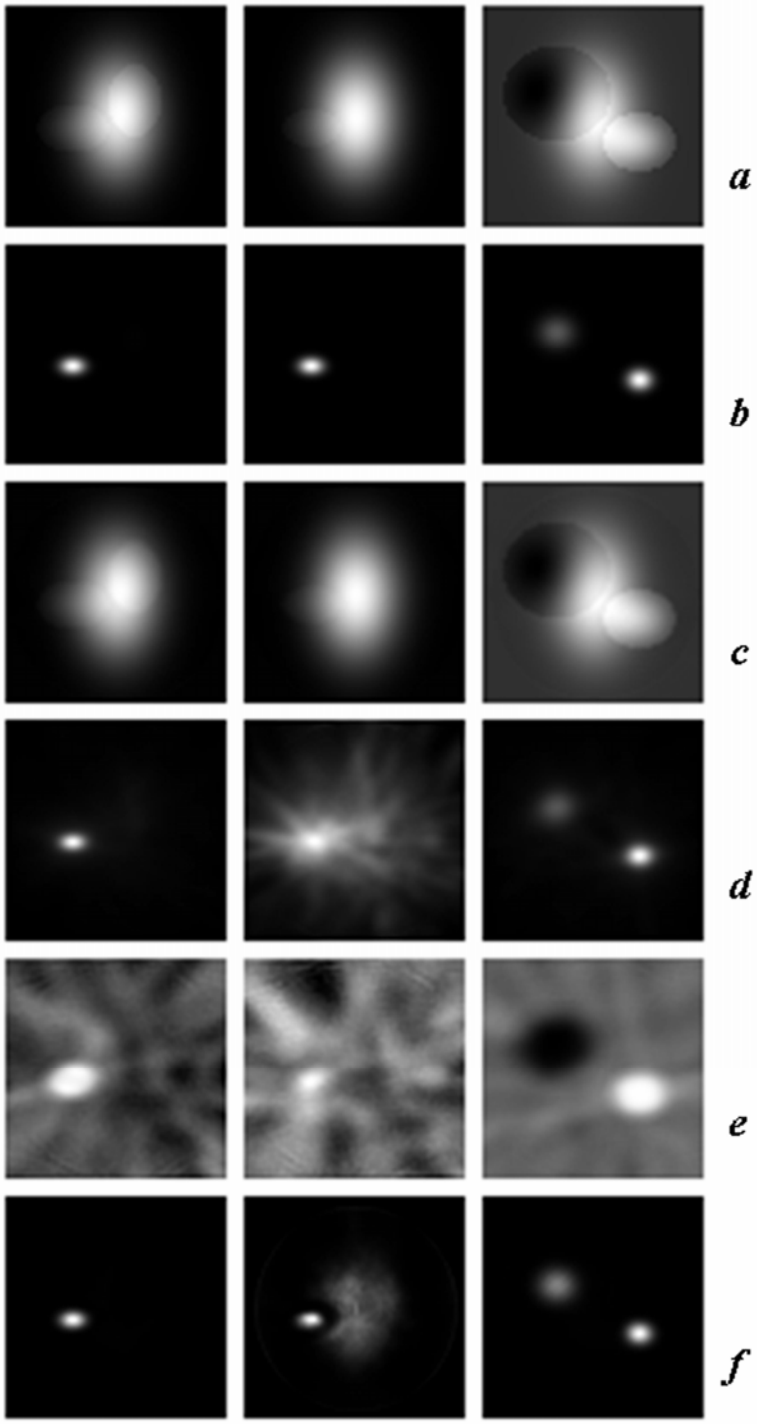
$$\sigma^2 = K\sigma_{\bar{x}}^2$$

Sources of variance in 3-D reconstructions

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

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



σ_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.

		A	*B*	*C*	*D*	*Center*
1	σ_f^2	2.25	1.49	0.68	2.37	10^{-8}
2	σ_{Struct}^2	2.51	1.79	0.93	2.76	0.55
3	$\sigma_{Solvent}^2$	15.5	12.2	7.78	17.7	0.77
4	r_f	1.00	0.02	-0.71	0.48	10^{-7}
5	r_{Struct}	0.86	0.02	-0.55	0.37	$-7\text{x}10^{-3}$

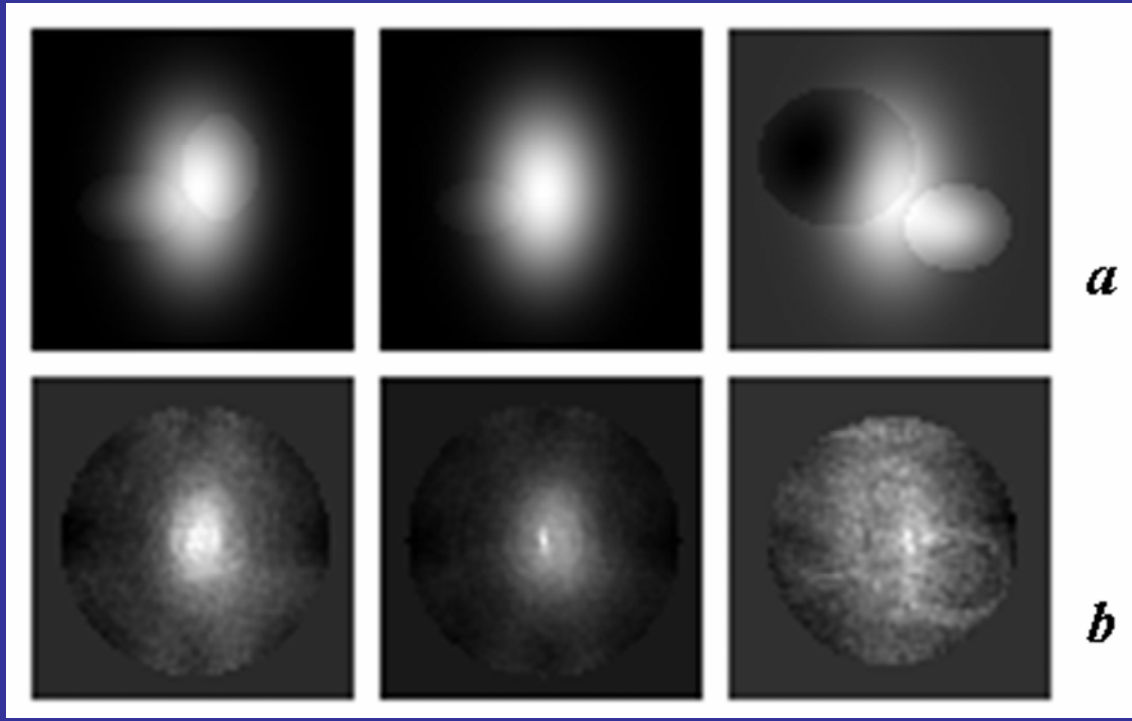
Components of bootstrap variance

$$\sigma^2_{SVar} = \sigma^2_B = \sigma^2_{Conf} + \sigma^2_{Ali} + \sigma^2_{Rec} + \sigma^2_{Back}$$

- ***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).

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 4.67x10-6 and 5.35x10-5 with an average of 1.17x10-5.
The SNR of the reconstruction is ~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.

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

Calculation of the variance of structures

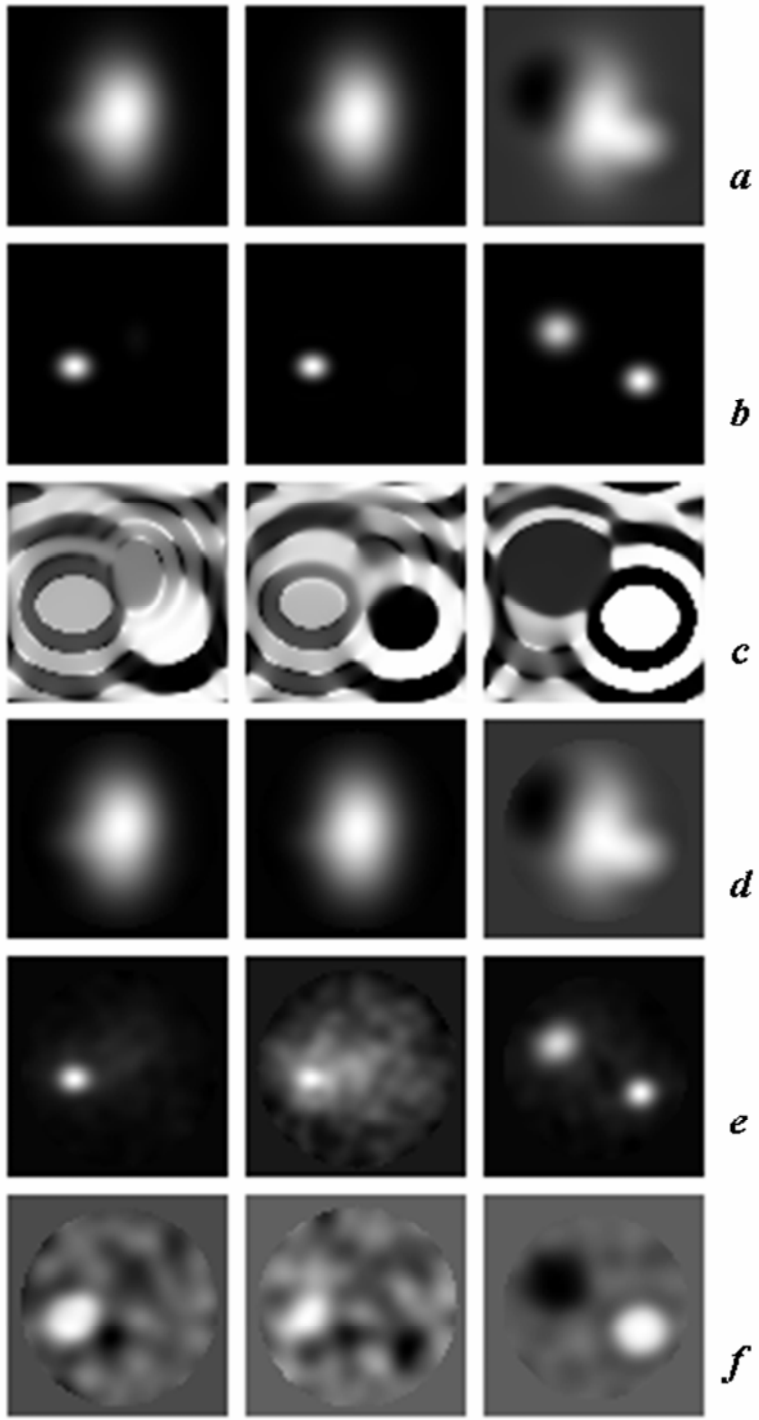
$$\sigma^2_{Struct} = K\left(\sigma^2_B - \bar{\sigma}^2_{Back}\right)$$

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

Test in the presence of additive noise $N(0,30)$, 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.

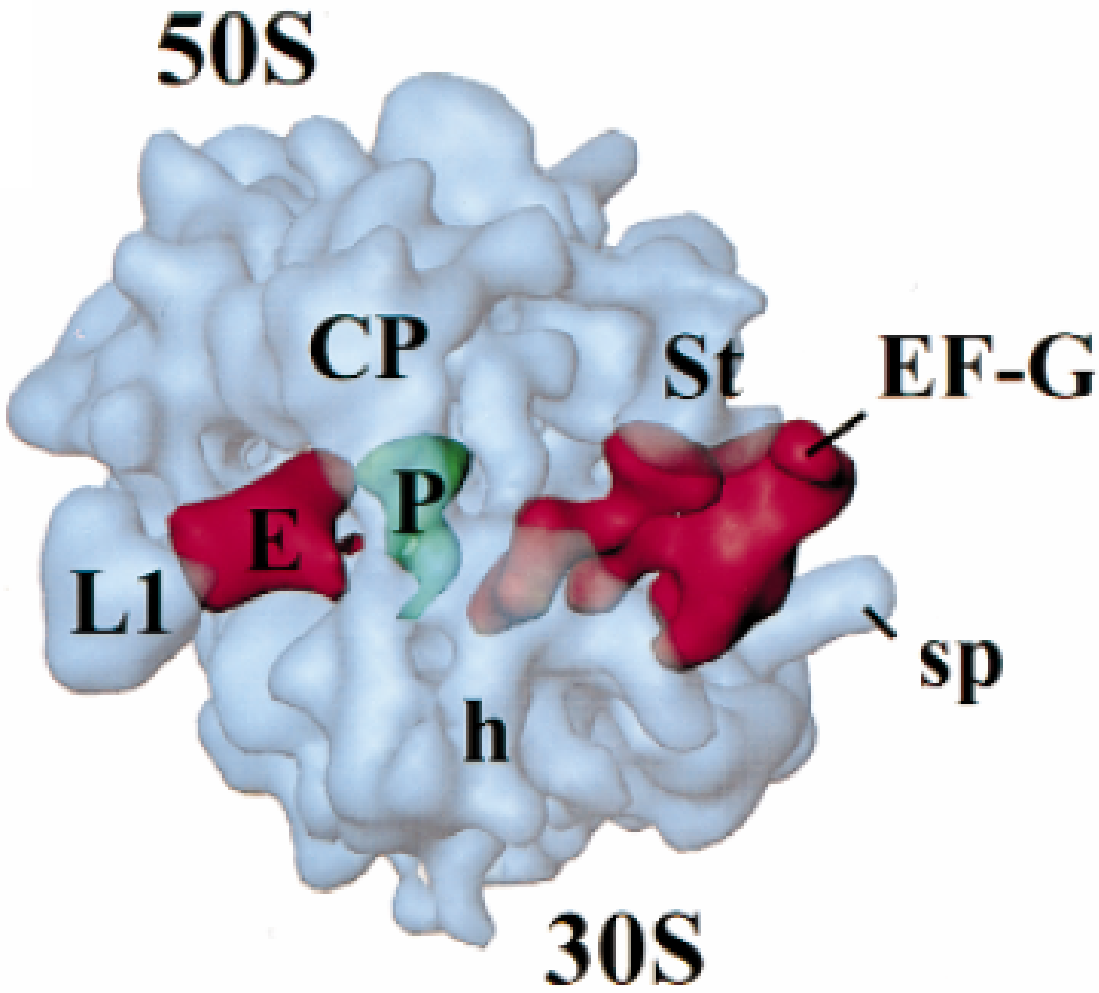


Test of the estimation of the structure variance using the bootstrap method in the presence of additive independent Gaussian in projections.

	A	*B*	*C*	*D*	*Center*
σ_f^2	1.17	1.24	0.65	1.35	$4\text{x}10^{-2}$
$K\sigma_{SVar}^2$	1.51	1.59	1.14	1.83	0.77
σ_{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	$-7\text{x}10^{-3}$

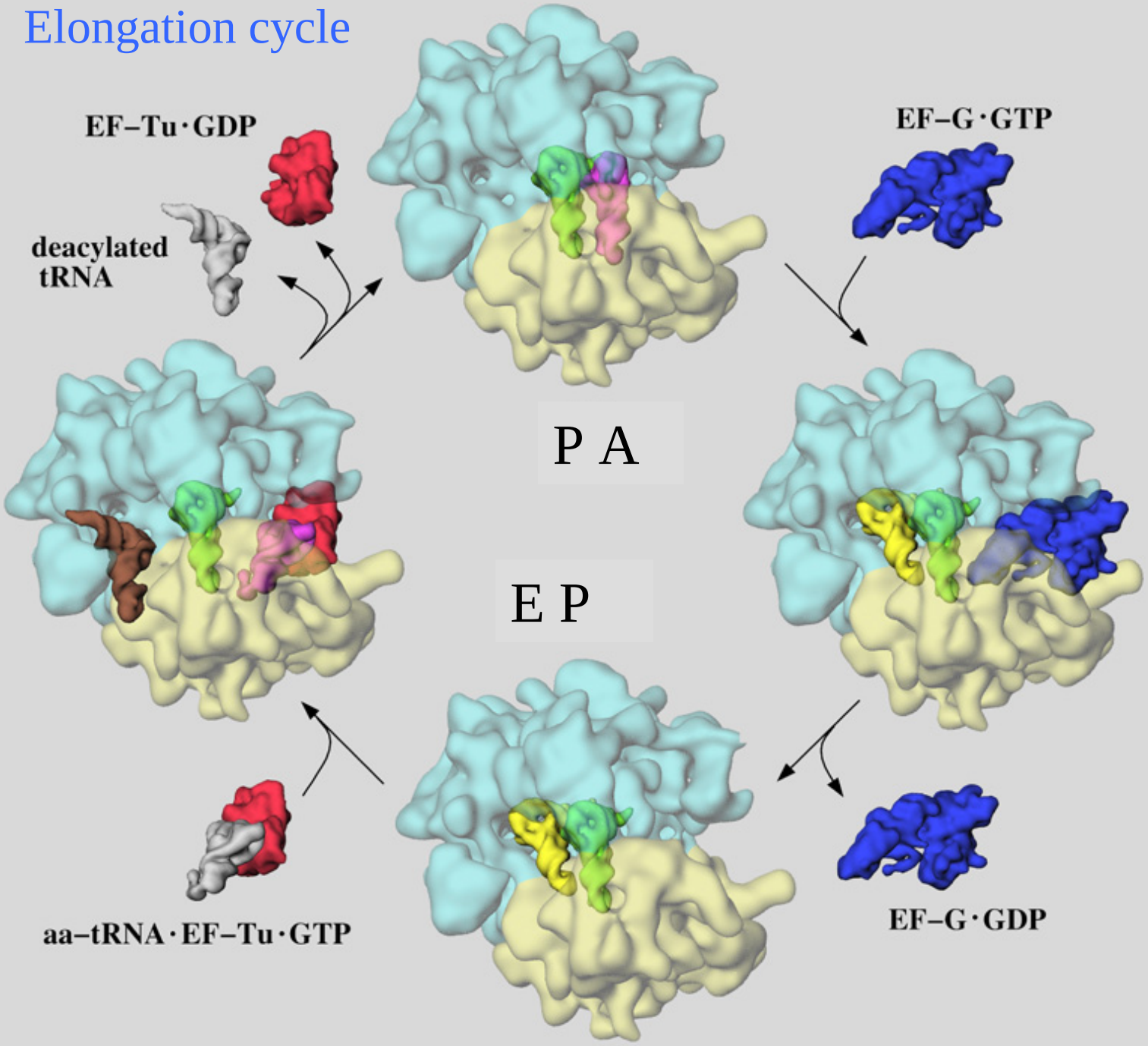
$$\sigma_{Struct}^2 = K\left(\sigma_{SVar}^2 - \bar{\sigma}_{Back}^2\right)$$

Analysis of an *E. coli* 70S ribosomal complex containing EF-G and tRNAs



Agrawal, R. K., Heagle, A. B., Penczek, P., Grassucci, R. A., Frank, J., 1999. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. Nat. Struct. Biol. 6, 643-7.

Elongation cycle



Sample preparation

Pretranslocational 70S *E. coli* complex programmed with poly(U) was incubated with a 1.6 molar excess of $tRNA^{Phe}$ to fill the P site.

The occupancy of $tRNA^{Phe}$ was checked in a parallel reaction by nitrocellulose filter binding assay and it was ~76%.

Next, the mixture was incubated with a 1.6 molar excess of $[^{14}C]$Phe-$tRNA^{Phe}$ to fill the A site.

The occupancy of $[^{14}C]$Phe-$tRNA^{Phe}$ was ~58%.

Finally, the complex was incubated with EF-G to obtain the 70S-$(tRNA)_2$-EF-G-GMPP(CH_2)P complex.
The occupancy of EF-G was ~36%.

Agrawal, R. K., Heagle, A. B., Penczek, P., Grassucci, R. A., Frank, J., 1999. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. Nat. Struct. Biol. 6, 643-7.

Electron microscopy

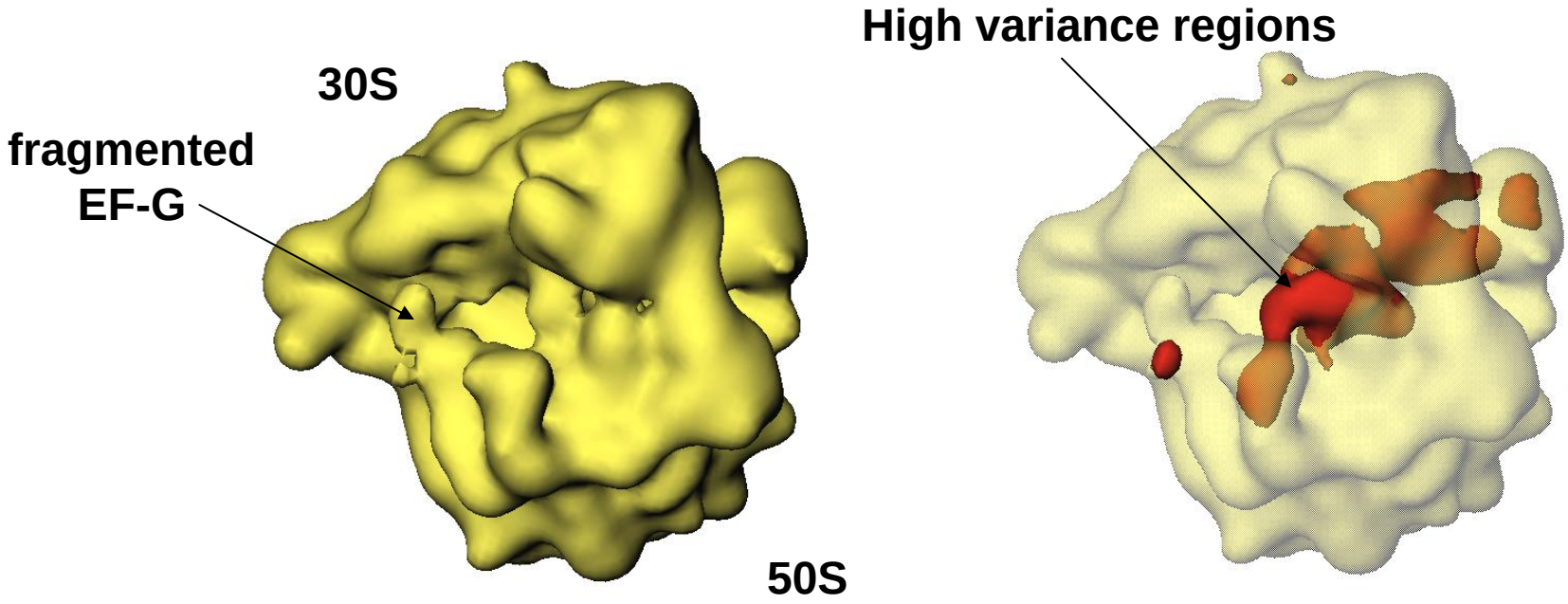
The sample was applied on a grid covered with a thin layer of carbon film and rapidly plunged into liquid ethane.

The images were collected in several defocus groups on a Philips EM420.

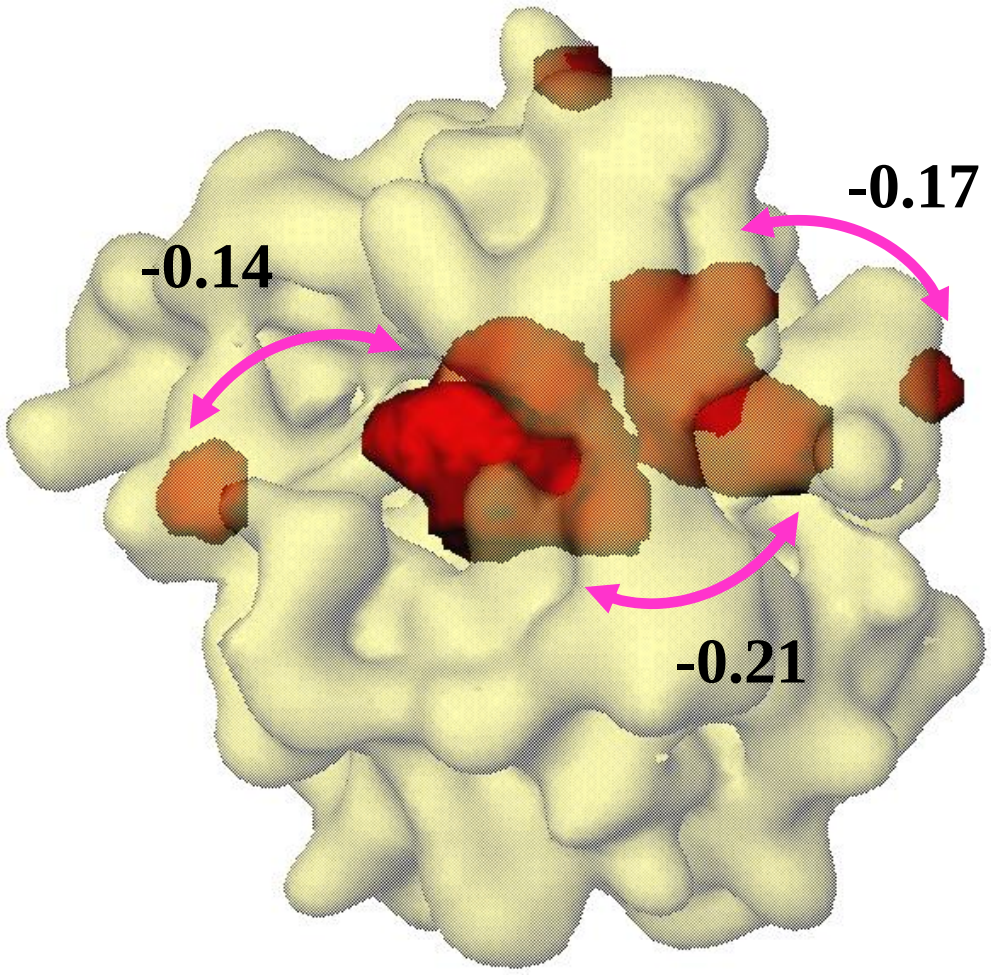
A total of 10,927 particle images were collected.

The 3-D structure of the complex was solved at 17 Å resolution.

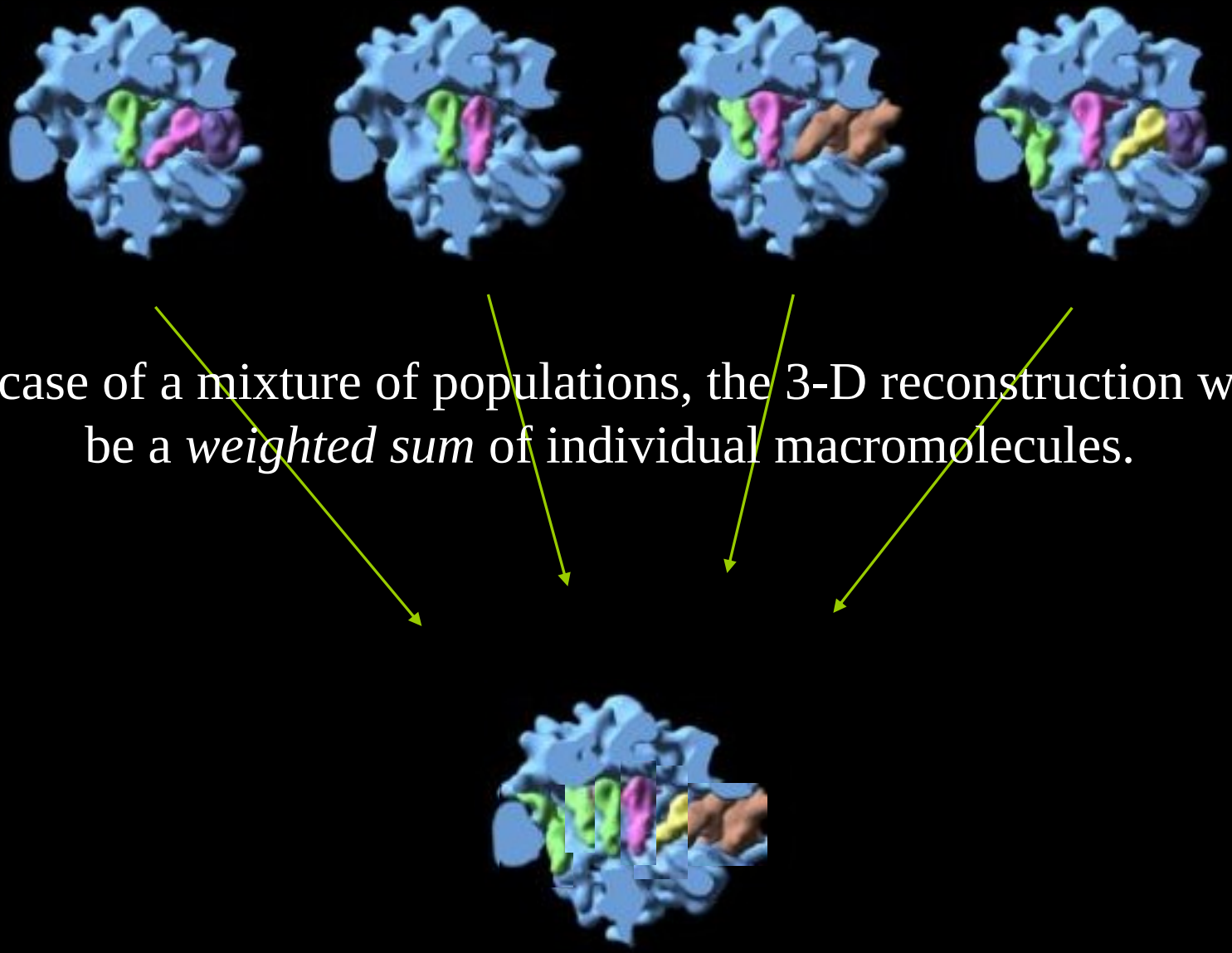
Calculation of real space variance for the *E. coli* 70S ribosomal complex containing EF-G and tRNAs



Calculation of real space correlations for the *E. coli* 70S ribosomal complex containing EF-G and tRNAs

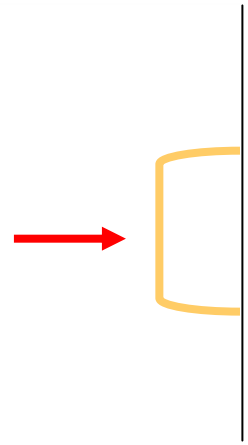


3-D classification of projections

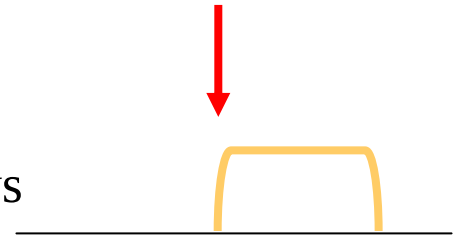


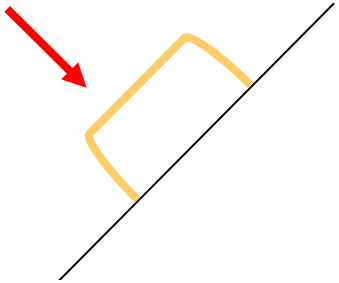
In case of a mixture of populations, the 3-D reconstruction will be a *weighted sum* of individual macromolecules.

1. Create a spherical mask around the expected location of the EFG (indicated by the analysis of the variance)

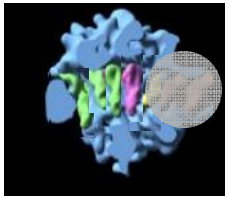


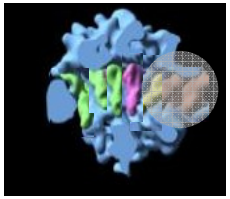
2. Project the mask in the directions of the particle views



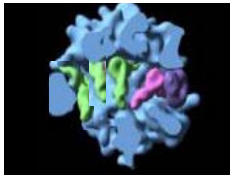


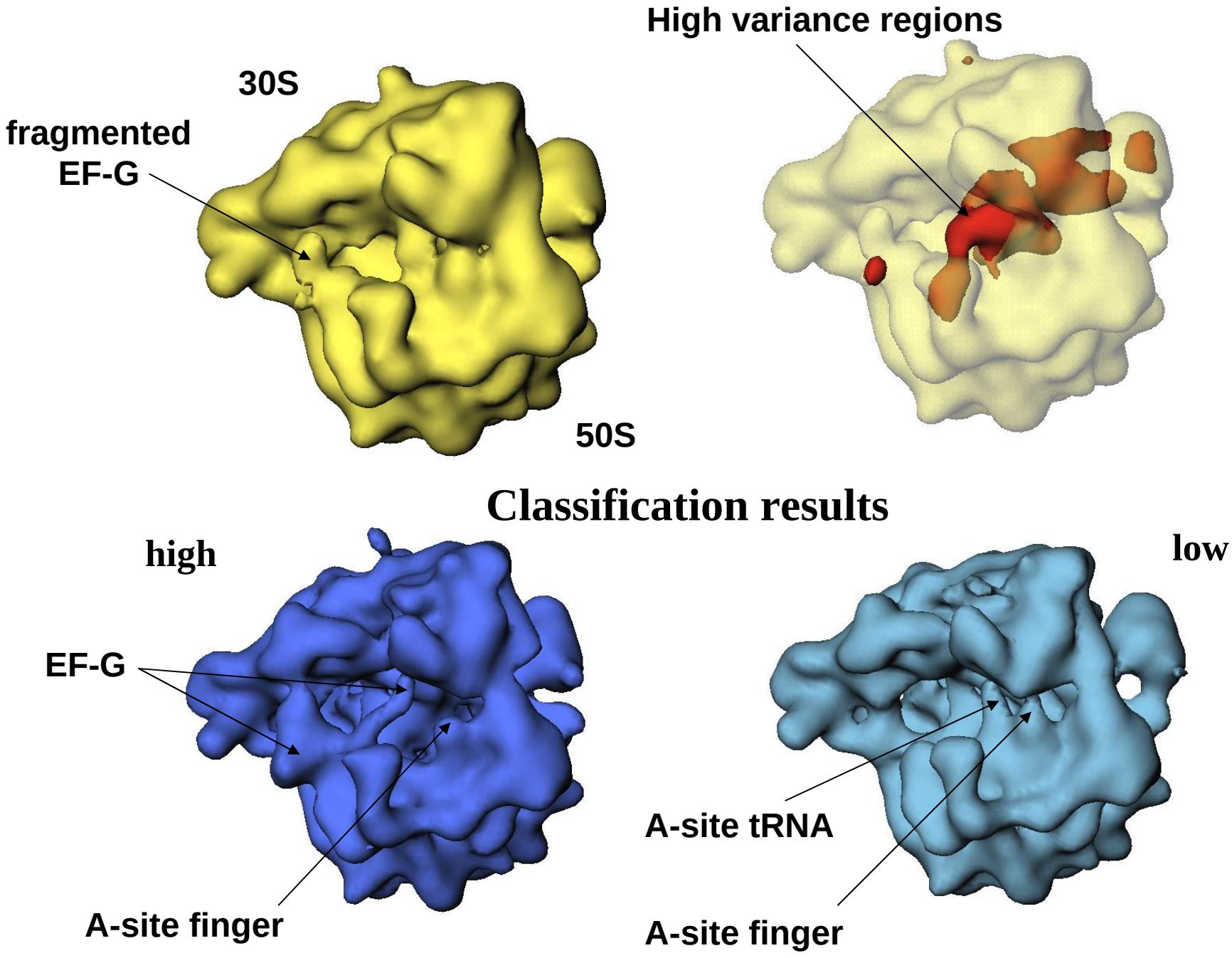
3. Divide particle views into two classes: high mass and low mass under the mask and calculate two respective “seed” structures.





4. Proceed with the “*3-D K-means procedure*”, in which the particle views are reassigned to either of the two classes based on the similarity to projections of the current 3-D structures.

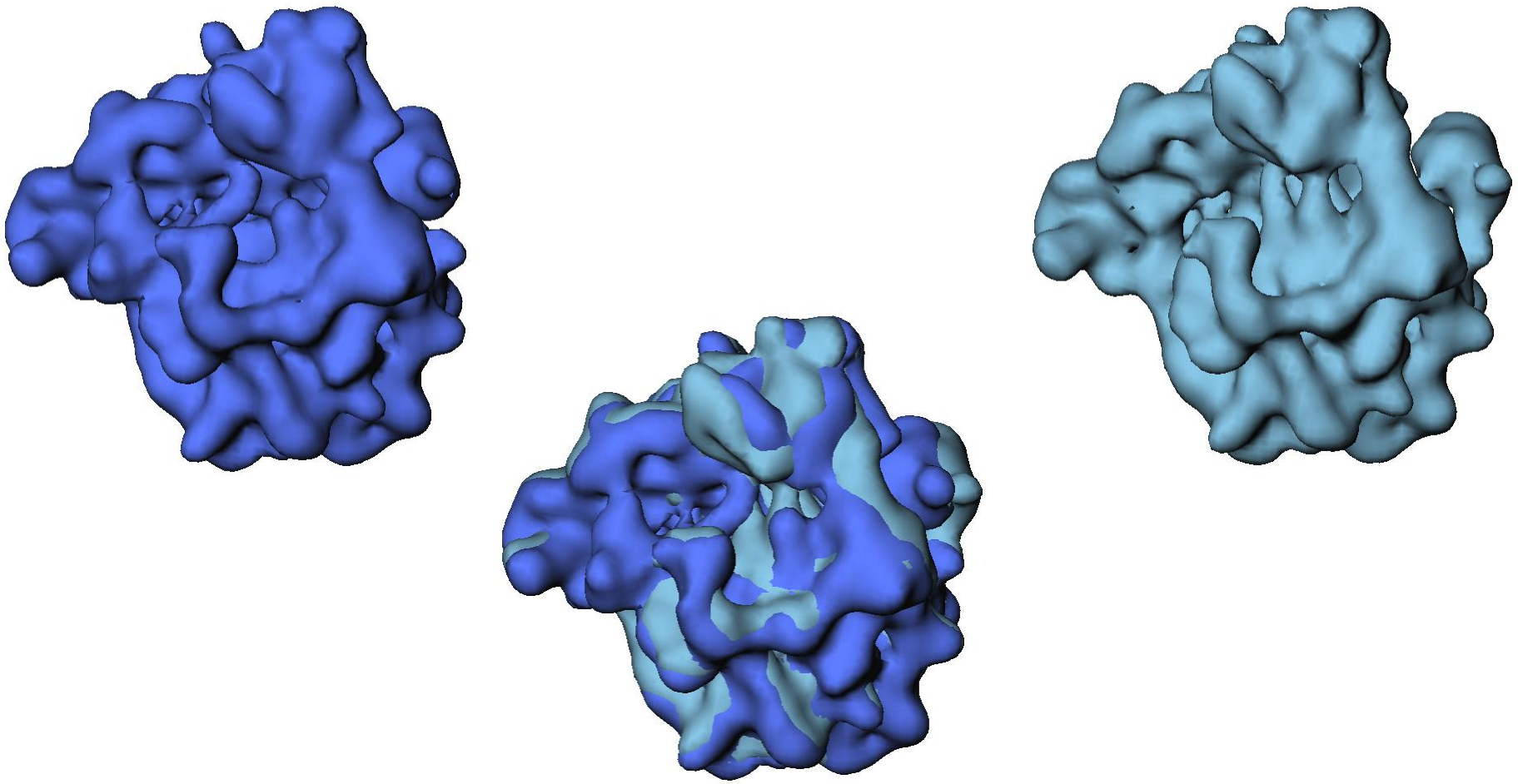



High variance regions
30S
fragmented
EF-G
50S
Classification results
high
low
EF-G
A-site tRNA
A-site finger
A-site finger

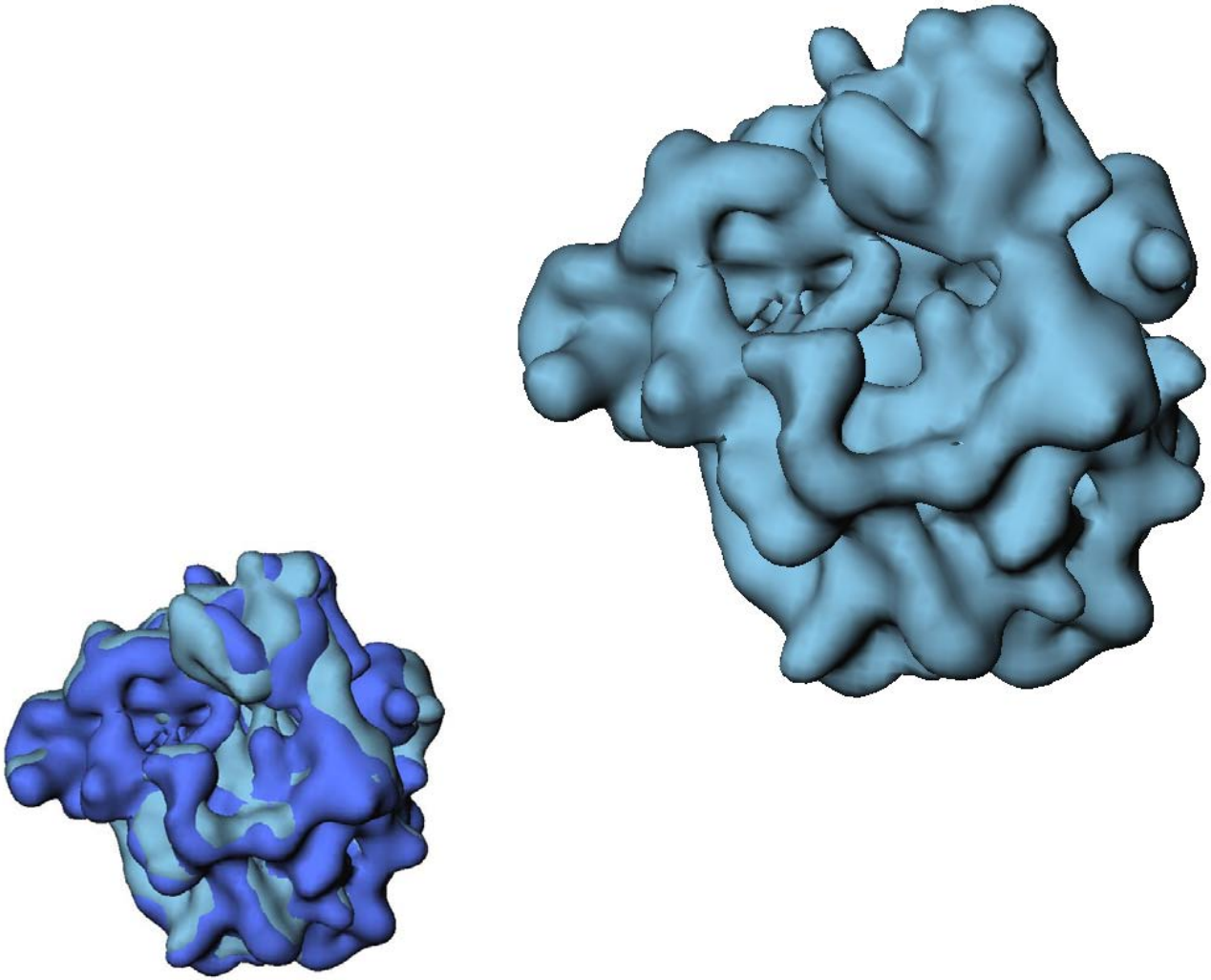
Conformational changes upon EF-G binding

high

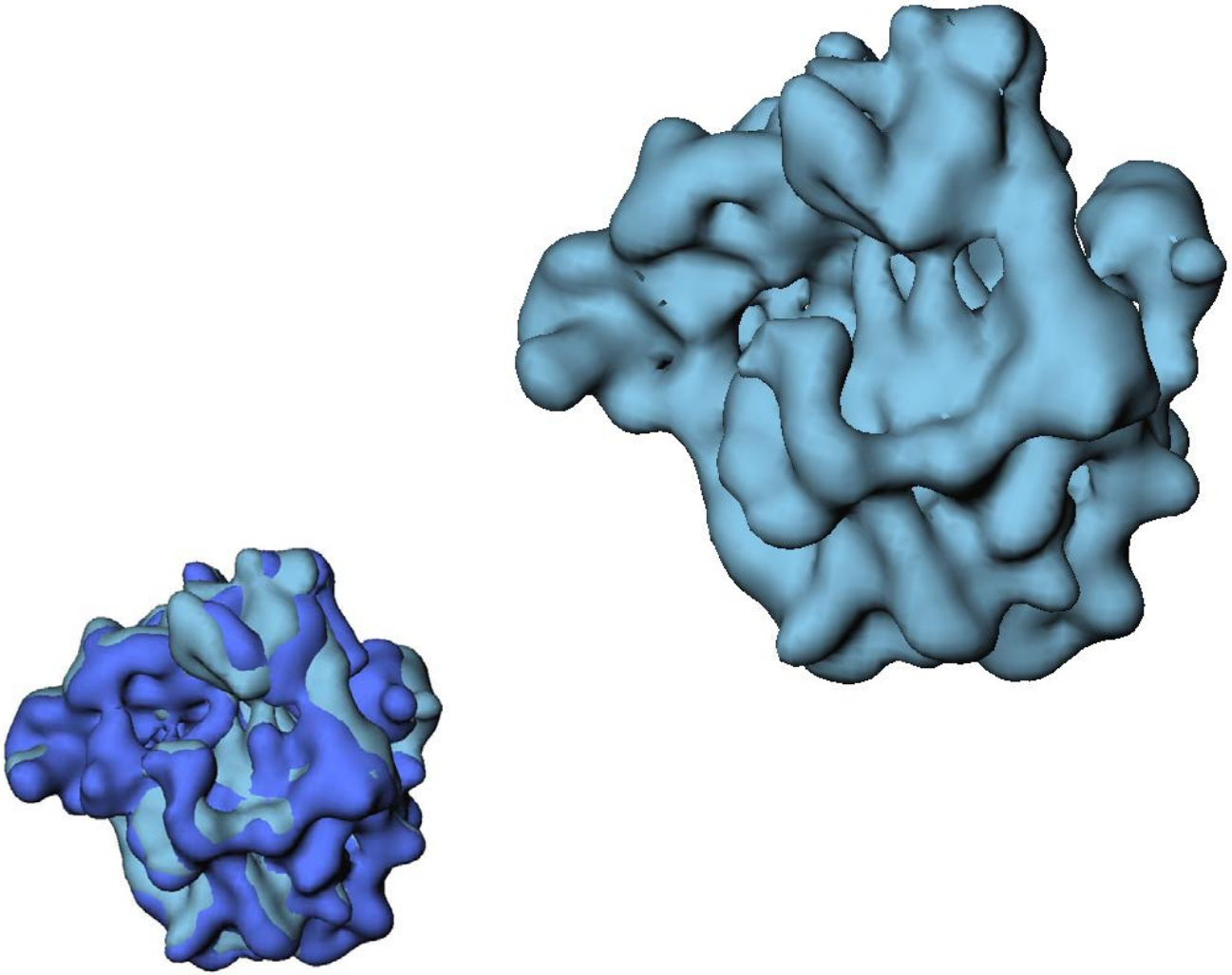
low



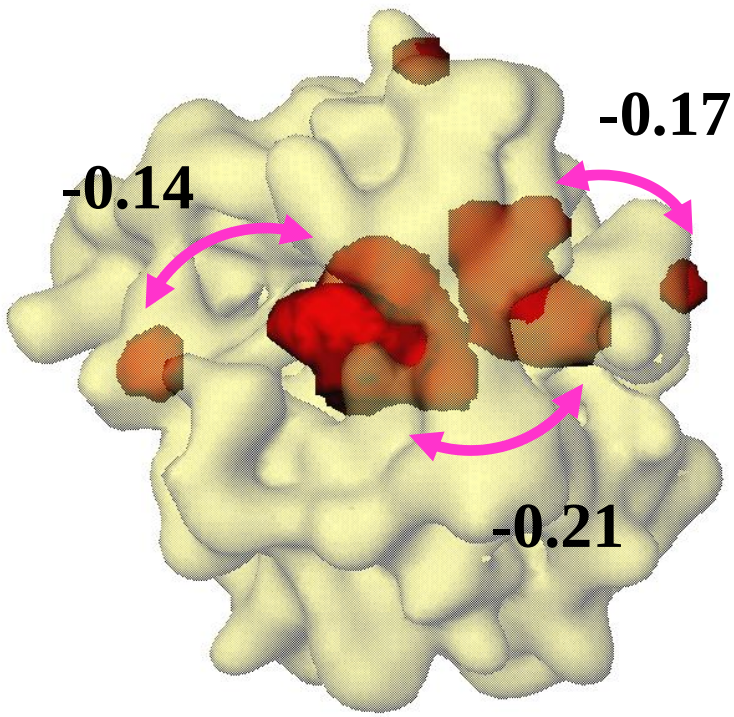
Conformational changes upon EF-G binding



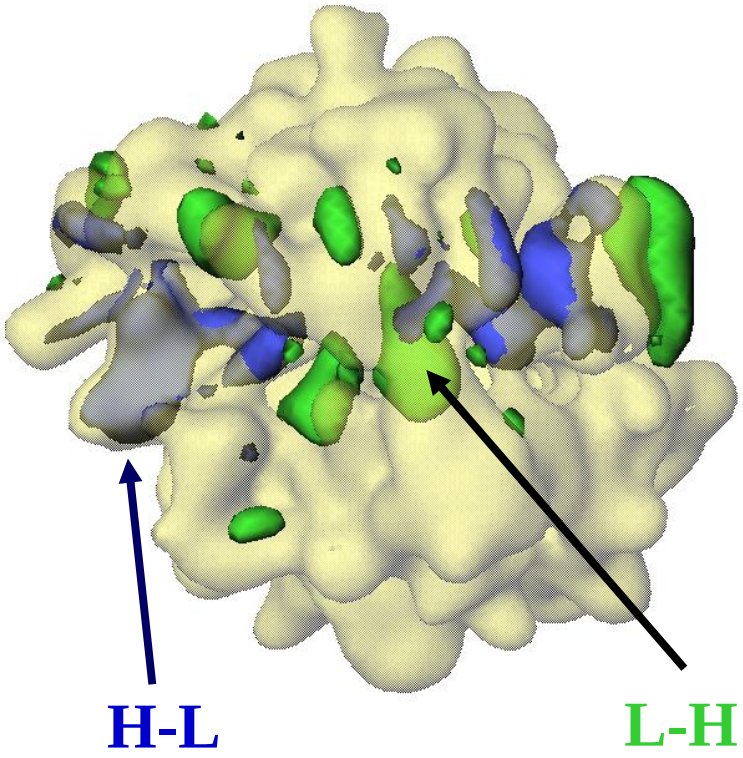
Conformational changes upon EF-G binding



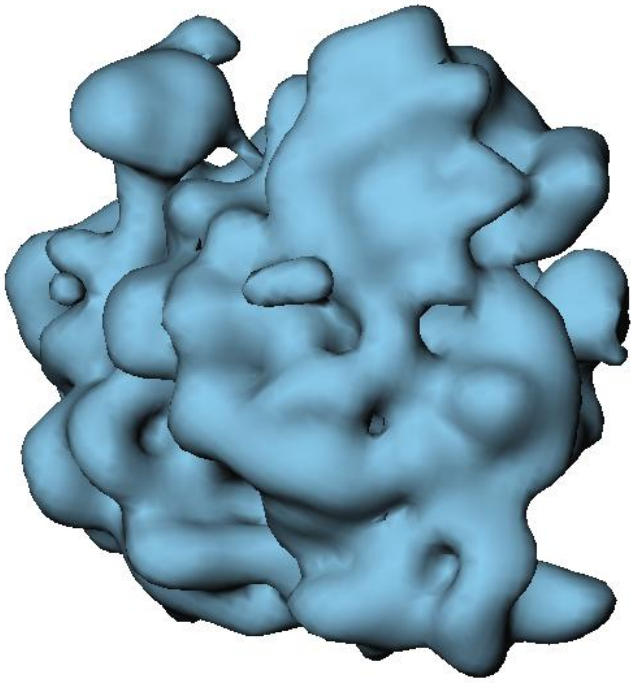
Calculation of real space correlations for the *E. coli* 70S ribosomal complex containing EF-G and tRNAs



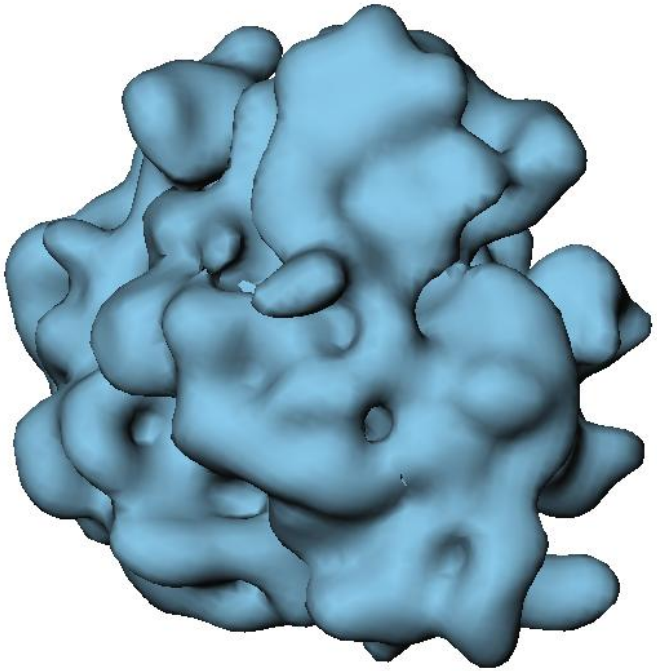
Difference maps



Conformational changes upon EF-G binding



Conformational changes upon EF-G binding

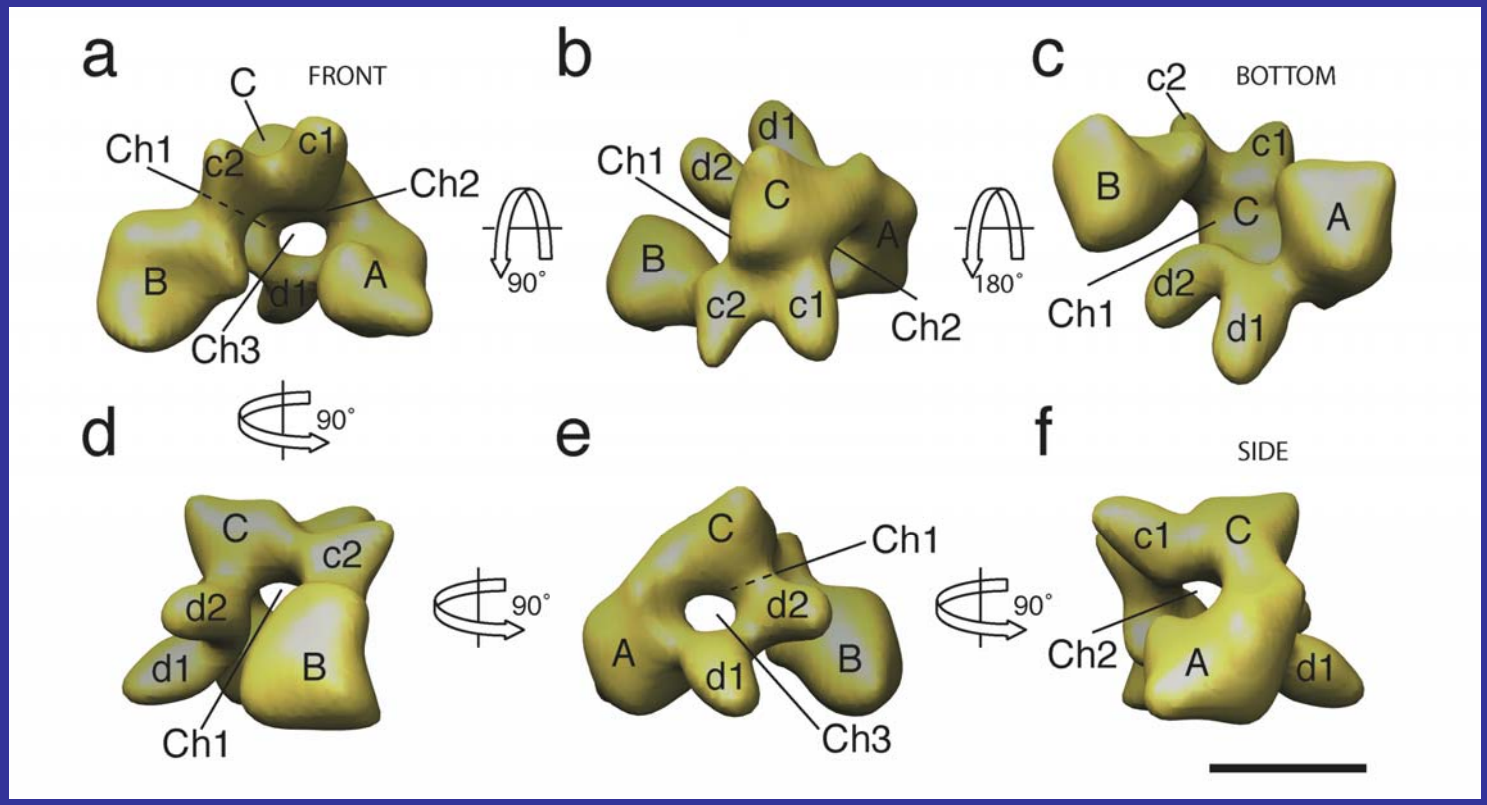


Conformational flexibility of human TFIID revealed by cryo-electron microscopy studies

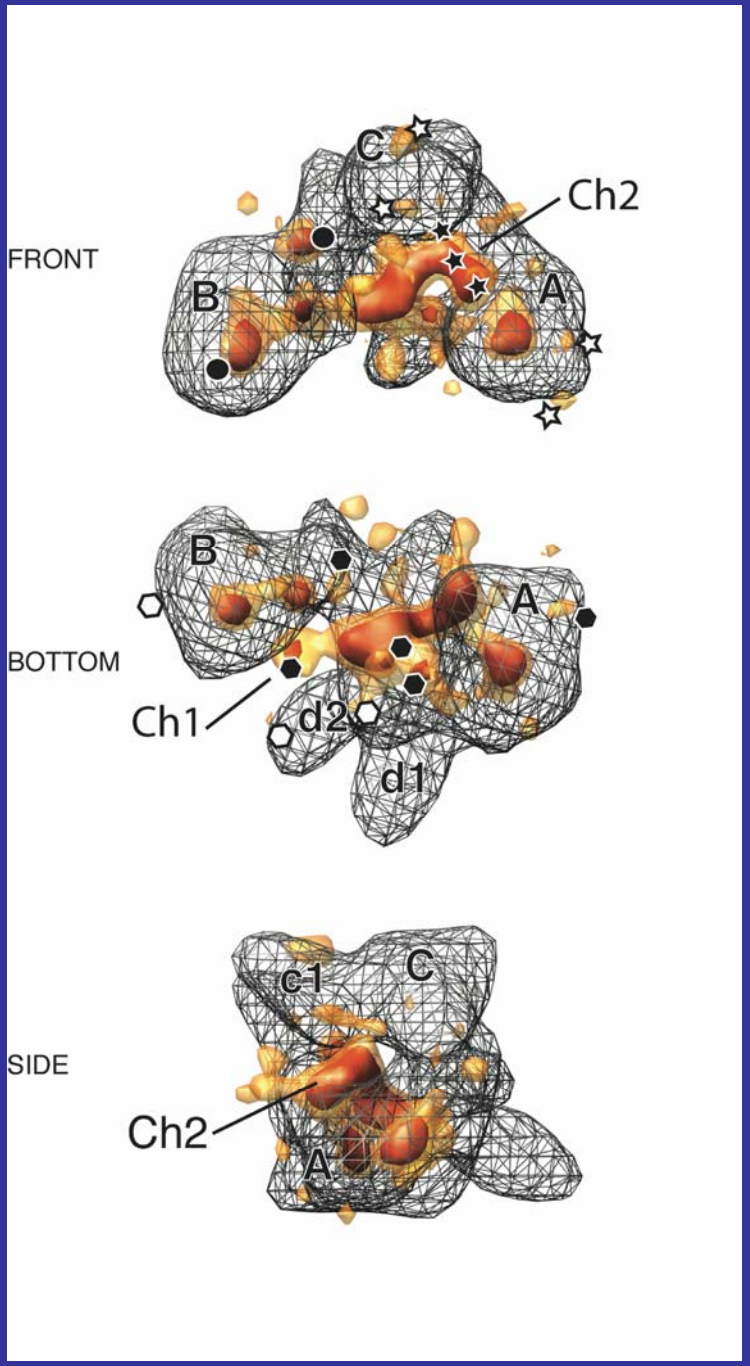
Structure, in press.

- **Patricia Grob[1], Michael J. Cruse[2],[3], Carla Inouye1,**
- **Marian Peris[4], Pawel A. Penczek[5], Robert Tjian4, Eva Nogales2,4**
- [1] Molecular and Cell Biology Department, University of California, Berkeley, CA 94720, USA.
- [2] Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
- [3]Presently at University of California, San Francisco at Genentech Hall, 600 16th Street, San Francisco, CA 94107, USA.
- [4] Howard Hughes Medical Institute, Molecular and Cell Biology Department, University of California, Berkeley, CA 94720, USA.
- [5] The University of Texas – Houston Medical School, Department of Biochemistry and Molecular Biology, 6431 Fannin, MSB 6.218, Houston, TX 77030, USA.

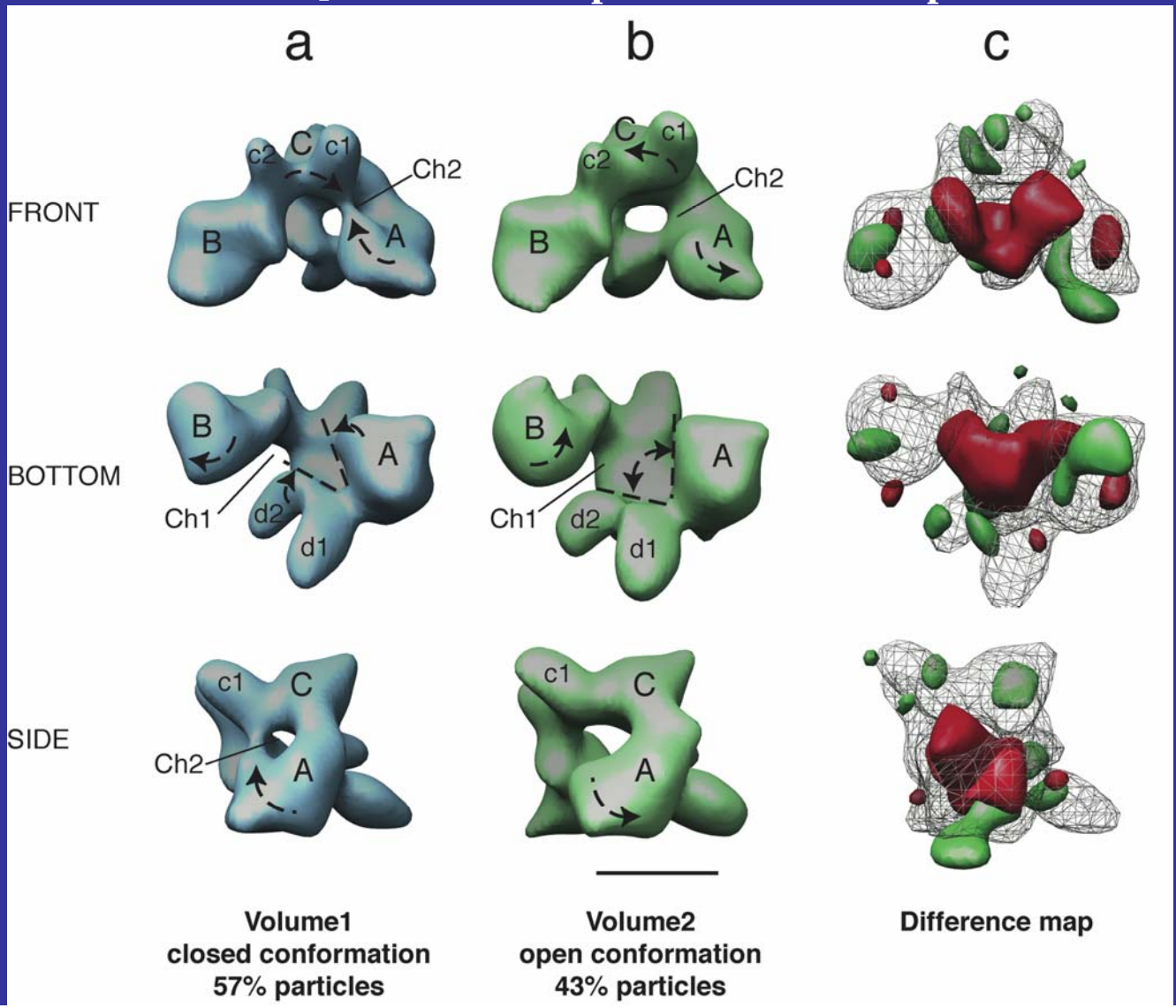
3D reconstruction of human TFIID at 33Å in ice.
The density threshold has been chosen for a 1MDa complex. The main lobes of density have been labeled A, B, C and the smaller ones c1-2, d1-2. The channels and cavities have been labeled Ch1-3. The scale bar represents 100Å.



3-D variance map and covariance analysis. The high density (red) and lower density (yellow) of the 3-D variance map is superposed on the 3-D reconstruction (mesh). The high variance is mainly localized inside the central channel, Ch1 and Ch2. A few spots are also visible in the tips of lobes A and B, around their hinges and around c1, c2, d1, d2. The results of the covariance analysis are represented by the black and white symbols. Each symbol (star, hexagon, dot) corresponds to a group of variance points that have a significantly high correlation between each other. Black-white represents a negative correlation, white-white and black-black represent a positive correlation.



3-D reconstructions of groups, after classification in the high variance region. (a) The first group has higher density in the central region and has narrower channel. (b) The second group has wider opening. The numbers and arrows point out to the putative "movements". (c) Positive part (red) and negative part (green) of the [Volume1-Volume2] difference map. The scale bar represents 100Å.



Conclusions/Future work

- ✓ In general the bootstrap method works, but…

?We ignored the CTF

?Outliers have adverse effect on the bootstrap variance map, weighted bootstrap…

?Better reconstruction algorithm for bootstrap