Enhanced diffractograms for image sorting and improvement of robustness and accuracy of CTF estimation: Application on high-resolution reconstruction of glutamate synthase and phosphorylase kinase

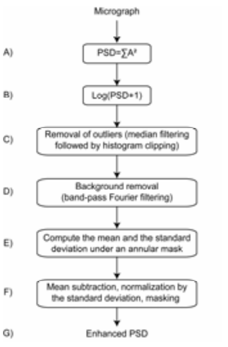
S. Jonić[1], C.O.S. Sorzano[2,3], M. Cottevieille[1], E. Larquet[1], C. Vénien-Bryan[4], N. Boisset[1]

[1]Université Pierre et Marie Curie, IMPMC-UMR7590, Paris, F-75005 France; CNRS, IMPMC-UMR7590, Paris, F-75005 France; Université Paris 7; IPGP.
[2]Bioengineering Lab. Escuela Politécnica Superior. Univ. San Pablo CEU. Campus Urb. Montepríncipe s/n. 28668 Boadilla del Monte, Madrid, Spain.
[3]Unidad de Biocomputación, Centro Nacional de Biotecnología (CSIC), Campus Universidad Autónoma s/n, 28049 Cantoblanco, Madrid, Spain.
[4]Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, United Kingdom.

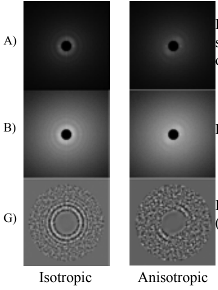
slavica.jonic@impmc.jussieu.fr

Abstract:
In a recent work, we developed a method for improving the visibility of diffraction rings computed from cryo-electron micrographs of vitreous ice (without carbon film or high concentration of diffracting material) [1]. We used these enhanced power spectrum densities (enhanced PSDs) to semi-automatically detect and remove micrographs with a strong envelope function or with a low signal-to-noise ratio (e.g., non-diffracting micrographs or their areas) or those that introduce errors in the global 3D map (e.g., drifted micrographs or charged areas). This sorting is based on the normalized cross-correlation between enhanced PSDs and their copies rotated by 90 degrees. In the present work, we show single-particle reconstruction of two macromolecular complexes, using this image sorting method. The two complexes are: 1) bacterial glutamate synthase at resolution of 9.8 Å according to the $FSC_{0.5}$ criterion or 8.3 Å according to the 1/2-bit criterion, and 2) mammalian phosphorylase kinase at resolution of 9.8 Å according to the $FSC_{0.5}$ criterion or 8.4 Å according to the 1/2-bit criterion. We also show the use of enhanced PSDs in improving robustness and accuracy of contrast transfer function (CTF) estimation on images that are kept after sorting. We show that the use of enhanced PSDs can improve the estimation of a full two-dimensional PSD model that takes into account the envelope, such as the one proposed in [2]. Determination of defocus parameters (maximum defocus, minimum defocus, angle of astigmatism) is the most difficult part of the PSD model estimation. This is why the algorithm first estimates roughly as many parameters of the PSD model as possible (i.e., background and envelope parameters), and then, estimates very accurately defocus parameters based on a weighted correlation between the enhanced experimental PSD and the corresponding model [3]. In a final step, all parameters are refined. The source code of the algorithms for PSD enhancement and CTF estimation is available for implementing in any image processing package, upon request from the authors or as a part of the open-source package Xmipp (http://xmipp.cnb.csic.es/).

1. Improvement of visibility of diffraction rings



Examples of application of the enhancement algorithm (left) on two cryo-EM images without carbon film (JEOL JEM 2100F, ultra high-resolution pole piece, no tilt, voltage: 200 kV, Cs: 0.5 mm, magnification: X50.000, defocus: from -1.7 to -3.2 µm, pixel size: 1.59 Å × 1.59 Å)



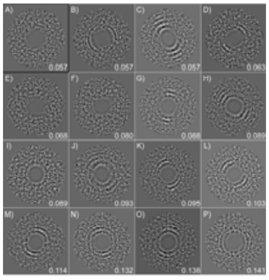
A) PSD=∑A² (Diffractogram: averaging of spectra of about 1000 overlapped sub-areas of size 512 × 512 pixels)

B) Log(PSD+1)

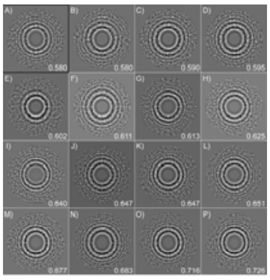
G) Diffractogram of an improved visibility (Enhanced PSD)

2. Sorting of 151 micrographs based on sorting of their enhanced PSDs

Sorting based on normalized cross-correlation (NCC) between the enhanced PSDs and their copies rotated by 90°



Sixteen lowest NCC values: Drifted (B-D,G-M) , non-diffracting (A,E,F), and strongly astigmatic (N-P) images.

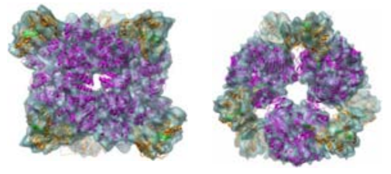


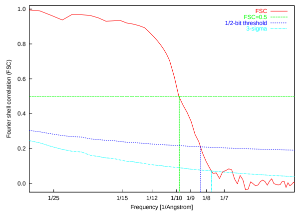
Sixteen highest NCC values: Isotropic images

3. Examples of 3D cryo-EM maps computed after image sorting based on enhanced-PSD sorting

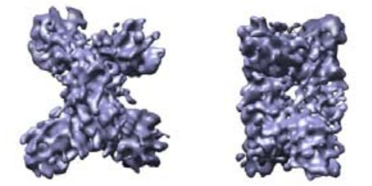
Bacterial Glutamate synthase

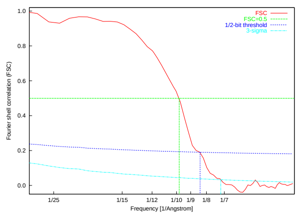
Resolution: 9.8 Å ($FSC_{0.5}$), 8.3 Å (1/2-bit), 7.7 Å (3σ)





Mammalian Phosphorylase kinase
Resolution: 9.8 Å ($FSC_{0.5}$), 8.4 Å (1/2-bit), 7.2 Å (3σ)





4. Power spectrum density model

PSD: $PSD_{theoretical}(\boldsymbol{\varpi}) = K^2 \left|H(\boldsymbol{\varpi})\right|^2 + S_N(\boldsymbol{\varpi})$

CTF: $H(\boldsymbol{\varpi}) = -E_{\Theta_{1,2,3}}(\boldsymbol{\varpi})(\sin(\chi_{\Theta_{1,2}}(\boldsymbol{\varpi})) + Q_0 \cos(\chi_{\Theta_{1,2}}(\boldsymbol{\varpi})))$

$\Theta_1 = (V, C_s)$ ………………………………(voltage, spherical aberration)

$\Theta_2 = (\Delta f_M, \Delta f_m, \theta)$ ………………………(maximum defocus, minimum defocus, astigmatism angle)

$\Theta_3 = (C_a, \frac{\Delta V}{V}, \frac{\Delta I}{I}, \alpha, \Delta F, \Delta R)$ …………….(chromatic aberration, electron energy spread,
lens current instability, angular aperture,
vertical sample displacement, in-plane sample displacement)

Defocus vector: $\boldsymbol{\Delta f}(\angle\boldsymbol{\varpi}) = (\Delta f_M \cos(\angle\boldsymbol{\varpi} - \theta), \Delta f_m \sin(\angle\boldsymbol{\varpi} - \theta))$

Envelope: $E(\boldsymbol{\varpi}) = E_{spread}(\boldsymbol{\varpi}) E_{coherence}(\boldsymbol{\varpi}) E_{drift}(\boldsymbol{\varpi})$

$$E_{spread}(\boldsymbol{\varpi}) = e^{-\frac{\left(\frac{\pi}{4}C_a\lambda\left(\frac{\Delta V}{V}+2\frac{\Delta I}{I}\right)\right)^2}{\log(2)}|\boldsymbol{\varpi}|^4}, E_{coherence}(\boldsymbol{\varpi}) = e^{-\pi^2\alpha^2\left(C_s\lambda^2|\boldsymbol{\varpi}|^3 + |\boldsymbol{\Delta f}(\angle\boldsymbol{\varpi})||\boldsymbol{\varpi}|\right)^2}$$

$$E_{drift}(\boldsymbol{\varpi}) = J_0\left(\pi\Delta F\lambda|\boldsymbol{\varpi}|^2\right)\mathrm{sinc}\left(|\boldsymbol{\varpi}|\Delta R\right)$$

λ : Electron wavelength depending on voltage

Background: $S_N(\boldsymbol{\varpi}) = b + K_s e^{-|\mathbf{r}_{\Theta_4}(\angle\boldsymbol{\varpi})|\sqrt{|\boldsymbol{\varpi}|}} + K_G e^{-|\mathbf{r}_{\Theta_5}(\angle\boldsymbol{\varpi})|(|\boldsymbol{\varpi}| - |\mathbf{r}_{\Theta 6}(\angle\boldsymbol{\varpi})|)^2} - K_g e^{-|\mathbf{r}_{\Theta_7}(\angle\boldsymbol{\varpi})|(|\boldsymbol{\varpi}| - |\mathbf{r}_{\Theta_8}(\angle\boldsymbol{\varpi})|)^2}$

$\Theta_i = (R_i, r_i, \beta_i), i = 4, 5, \ldots, 8$

$\mathbf{r}_{\Theta_i}(\angle\boldsymbol{\varpi}) = (R_i \cos(\angle\boldsymbol{\varpi} - \beta_i), r_i \sin(\angle\boldsymbol{\varpi} - \beta_i))$

Note: Unknown parameters are in red

5. PSD model estimation taking into account enhanced PSDs

Note:

Currently estimated parameters are in red

Θ_1

1. Background estimation

$(\Theta_1, b, K_s, K_G, \Theta_4, \Theta_5, \Theta_6)$

2. Envelope estimation

$(\Theta_1, b, K_s, K_G, \Theta_4, \Theta_5, \Theta_6, K, \Theta_3)$

3. Defocus estimation

$(\Theta_1, b, K_s, K_G, \Theta_4, \Theta_5, \Theta_6, K, \Theta_3, \Theta_2)$

4. Parameter refinement

$(\Theta_1, b, K_s, K_G, \Theta_4, \Theta_5, \Theta_6, K, \Theta_3, \Theta_2)$

5. Refinement of parameters estimated in step 4 with additional estimation of K_g,Θ_7, and Θ_8 for a better fit at low frequencies

$(\Theta_1, b, K_s, K_G, \Theta_4, \Theta_5, \Theta_6, K, \Theta_3, \Theta_2, K_g, \Theta_7, \Theta_8)$

Determination of Θ_2 is the most difficult part of the algorithm. Robustness and accuracy of this determination is improved using enhanced PSD in steps 3 and 4. In these steps, the following cost function is optimized:

$$L = \frac{1}{\mathrm{card}\{\Omega\}} \sum_{\boldsymbol{\varpi}_i \in \Omega} \left|PSD'_{experimental}(\boldsymbol{\varpi}_i) - PSD'_{theoretical}(\boldsymbol{\varpi}_i)\right| - \mathrm{w}\rho(PSD'_{enhanced}(\boldsymbol{\varpi}_i), H(\boldsymbol{\varpi}_i))$$

$$PSD'(\boldsymbol{\varpi}) = \frac{PSD(\boldsymbol{\varpi}) - S_N(\boldsymbol{\varpi})}{K^2 E^2(\boldsymbol{\varpi})}$$

: Normalized PSD

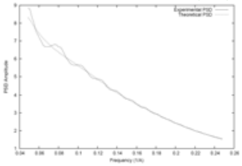
$\rho(x,y)$: Correlation coefficient between x and y

w : Weight determining the influence of the PSD enhancement term on the optimization algorithm (w>0)

Ω : Annular mask in the frequency domain

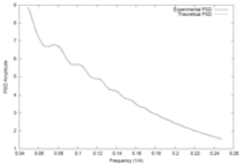
6. Results of PSD estimation on a single micrograph

Enhanced PSD switched off (w=0):

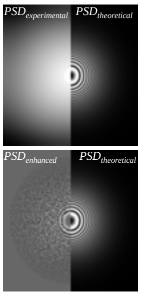


Radial averages of experimental and estimated PSDs. The estimated PSD does not fit well the experimental PSD since the algorithm fails to estimate Θ_2 accurately.

Enhanced PSD switched on (w=5):



Radial averages of experimental and estimated PSDs. The estimated PSD fits better the experimental PSD than in case w=0 (estimation of Θ_2 is more accurate).



Literature

[1] Jonic S, Sorzano CO, Cottevieille M, Larquet E, Boisset N (2007). A novel method for improvement of visualization of power spectra for sorting cryo-electron micrographs and their local areas. J Struct Biol 157(1): 156-67.

[2] Velazquez-Muriel JA, Sorzano CO, Fernandez JJ, Carazo JM (2003). A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment. Ultramicroscopy 96(1): 17-35.

[3] Sorzano CO, Jonic S, Nuñez R, Boisset N, Carazo JM. Fast, robust and accurate determination of transmission electron microscopy contrast transfer function. (submitted to J Struct Biol, N° JSB-07-64).