Journal of Structural Biology 203 (2018) 90–93

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Blind estimation of DED camera gain in Electron Microscopy

C.O.S. Sorzano[a,b,*], E. Fernández-Giménez[b], V. Peredo-Robinson[b], J. Vargas[a], T. Majtner[a], G. Caffarena[b], J. Otón[a], J.L. Vilas[a], J.M. de la Rosa-Trevín[a], R. Melero[a], J. Gómez-Blanco[a], J. Cuenca[a], L. del Cano[a], P. Conesa[a], R. Marabini[c], J.M. Carazo[a]

[a] Centro Nac. Biotecnología (CSIC), c/Darwin, 3, 28049 Cantoblanco, Madrid, Spain
[b] Univ. San Pablo – CEU, Campus Urb. Montepríncipe, 28668 Boadilla del Monte, Madrid, Spain
[c] Univ. Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain

ARTICLE INFO

Keywords:
Single particle analysis
Electron microscopy
Direct detectors

ABSTRACT

The introduction of Direct Electron Detector (DED) videos in the Electron Microscope field has boosted Single Particle Analysis to a point in which it is currently considered to be a key technique in Structural Biology. In this article we introduce an approach to estimate the DED camera gain at each pixel from the movies themselves. This gain is needed to have the set of recorded frames into a coherent gray level range, homogeneous over the whole image. The algorithm does not need any other input than the DED movie itself, being capable of providing an estimate of the camera gain image, helping to identify dead pixels and cases of incorrectly calibrated cameras. We propose the algorithm to be used either to validate the experimentally acquired gain image (for instance, to follow its possible change over time) or to verify that there is no residual gain image after experimentally correcting for the camera gain. We show results for a number of DED camera models currently in use (DE, Falcon II, Falcon 3, and K2).

1. Introduction

Direct Electron Detector (DEDs) videos have revolutionized the way images are acquired and processed in cryo Electron Microscopy (cryoEM) (Kühlbrandt, 2014). Indeed, they have provided access to unprecedented quality images, allowing to obtain 3D maps of a broad range of macromolecules at quasi-atomic resolution. Still, and, as any other image acquisition device, particularly electronic ones, DEDs have non-negligible differences between the gain of the different sensor areas. These differences need to be corrected to produce a reliable image in which all readouts are mapped to the same gray level range. This latter effect is normally corrected by measuring a gain image (G) that, in its simplest form, relates the ideal image and the actually recorded one by the expression below:

$$I_{recorded} = I_{ideal} G \quad (1)$$

The gain image is experimentally obtained by using an image recorded without a specimen (note that as such, it is not possible to separate the camera gain from the microscope flat field). Note that this relationship assumes a simplified image formation model at the camera level (the process is assumed to be linear, it can be explained as a simple multiplication, and deterministic, there is no noise). Eq. (1) is certainly a simplification of the real image formation process, which can be further enhanced considering non-linearities and noise. However, we show in this work that for the sake of the estimation of the gain image for validation purposes, it is a good enough approximation. Ideally, the gain image should be constant and equal to 1 (or, at least, constant everywhere). In practice, it may take values significantly different from 1 and not be constant at all. Many cameras are calibrated and internally perform the correction before writing the recorded images.

In this work we will start from Eq. (1) and elaborate on the possibility to estimate the gain image directly from a single recorded movie. The algorithm starts from a gain image which is iteratively refined. Note that this estimation is not intended to replace an independent experimental measure. Indeed, its main proposed practical use is not as an initial gain estimation, but as a second gain estimation performed after initial correction by the supposedly correct gain image (we will refer to this as "residual gain estimation"). Ideally, residual gain estimation should result in a constant image, so that any deviation from it should be understood as an indication of possible problems in the initial gain estimation. In this way, we introduce here a complementary validation tool that could help to discern cases in which the experimental gain image used for correction may not be the proper one (for instance, as a result of its change during acquisition time). Whatsoever, the possibility of using the estimated gain image to correct the movies is

* Corresponding author at: Centro Nac. Biotecnología (CSIC), c/Darwin, 3, 28049 Cantoblanco, Madrid, Spain.
E-mail address: coss@cnb.csic.es (C.O.S. Sorzano).

https://doi.org/10.1016/j.jsb.2018.03.007
Received 3 July 2017; Received in revised form 13 March 2018; Accepted 14 March 2018
Available online 15 March 2018
1047-8477/

still open and it should be the subject of further research. We present cases extracted from EMPIAR (Iudin et al., 2016) in which the estimate gain image is qualitatively compared to the experimentally measured one resulting in a good match, and cases in which the initial gain image used for correction was clearly inadequate.

2. Algorithm

The need to estimate the gain image from the recorded image is also shared by other imaging techniques, like infrared imaging (Tendero et al., 2012). The idea is based on the local smoothness of the histograms of rows and columns. Let us first introduce the Smooth Midway Equalization (SME) operator (Tendero et al., 2012). We denote the estimate of the ideal image corresponding to the i-th frame of a DED movie at iteration n as $I^{(n)}_{ideal,i}$. For each column c of this ideal image, we calculate the cumulative density function of $I^{(n)}_{ideal,i}$, which will be denoted as $H^{col,(n)}_{ideal,i,c}$. The inverse of this function is denoted as $(H^{col,(n)}_{ideal,i,c})^{-1}$. Then, the inverse midway σ-smoothed histogram of that frame is calculated as

$$(\widetilde{H}^{col,(n)}_{ideal,\sigma,i,c})^{-1} = \frac{\sum_{k=-K}^{K} g_{\sigma,k}\,(H^{col,(n)}_{ideal,i,c+k})^{-1}}{\sum_{k=-K}^{K} g_{\sigma,k}} \tag{2}$$

where c is a column of the frame, $g_{\sigma,k} = \exp(-0.5k^2/\sigma^2)$ is a Gaussian weight, and $K = \lceil 3\sigma \rceil$ represents the effective extent of the Gaussian in the spatial domain. At the micrograph borders the sum above only considers those columns for which there is data, i.e., no wrapping is performed nor the image is extended with zeros. Note that the σ-smoothed histogram above produces a different histogram for each column. Additionally, in cryoEM we may assume that in most cases all columns should have similar histogram characteristics (unless there are large contaminations or the foil hole occupies a significant portion of the micrograph). With this assumption, we may average over all columns in order to increase the accuracy with which the underlying histogram is constructed

$$(\widetilde{H}^{col,(n)}_{ideal,\sigma,i})^{-1} = \frac{1}{C}\sum_{c=0}^{C-1} (\widetilde{H}^{col,(n)}_{ideal,\sigma,i,c})^{-1} \tag{3}$$

The implementation in Xmipp de la Rosa-Trevín et al., 2013; Sorzano et al., 2004 supports both operating modes and results below have been calculated using this latter option.

The column σ-equalized frame is defined as

$$\widetilde{I}^{col,(n)}_{ideal,\sigma,i} = (\widetilde{H}^{col,(n)}_{ideal,\sigma,i,c})^{-1} \circ H^{col,(n)}_{ideal,i,c} \circ I^{(n)}_{ideal,i} \tag{4}$$

where $\circ$ denotes function composition. Intuitively, this equation means that for each pixel of the current estimate of the ideal frame, $I^{(n)}_{ideal,i}$, we calculate the percentile of pixels in its column, c, that are below the pixel value being considered. Then, we look for the gray value in the smoothed histogram having the same percentile, and substitute the original pixel value by the pixel value given by the smoothed histogram.

The column SME operator applied to the image $I^{(n)}_{ideal,i}$ is defined as

$$SME^{col}\{I^{(n)}_{ideal,i}\} = \widetilde{I}^{col,(n)}_{ideal,\sigma_{TV},i} \tag{5}$$

where

$\sigma_{TV} = \arg\min_\sigma TV^{col}(\widetilde{I}^{col,(n)}_{ideal,\sigma,i}), TV^{col}(I) = \sum_{r=0}^{R-1}\sum_{c=1}^{C-1} |I(r,c) - I(r,c-1)|$ is the total variation by columns of the image I, and R is the number of rows in the frame; in other words, for each frame we look for the σ value that minimizes the total variance of the output image (typical values in the examples below give σ estimates between 0.5 and 1.5). We have presented the theory with column histograms. However, there is no reason to favor columns over rows (or actually any other orientation, although this latter choice would imply image interpolations, which should be normally avoided for speed and data quality issues). For this reason, we define the SME operator as

$$SME\{I^{(n)}_{ideal,i}\} = \frac{1}{2}(SME^{col}\{I^{(n)}_{ideal,i}\} + SME^{row}\{I^{(n)}_{ideal,i}\}) \tag{6}$$

where SME^{row} is the row SME operator (defined analogously to the column SME operator). Note that the vertical and horizontal σ's do not need to coincide and that separate optimizations are performed in each direction. Also, note that Eq. (1) holds true for each frame, but the gain image is the same for all of them. Adding the equations for every frame and solving for G, we come to the iterative step

$$G^{(n+1)} = \frac{\sum_{i=0}^{N-1} I_{recorded,i}}{\sum_{i=0}^{N-1} SME\{I^{(n)}_{ideal,i}\}} \tag{7}$$

where N the total number of frames and n the current iteration (the equation above is applied pixel-wise, e.g., at every pixel we calculate a different gain value). The estimate of the ideal i-th frame is then updated with $I^{(n+1)}_{ideal,i} = \frac{I_{recorded,i}}{G^{(n+1)}}$ (in the first iteration $G^{(1)} = 1$). This recursive algorithm can be executed until convergence (which can be measured through the norm of the gain difference between an iteration and the next; however, we have observed that normally just one iteration is enough to get a good estimate of the gain image).

To the best of our knowledge, the only article that addresses a similar problem is that of Afanasyev et al. (2015). In their work, they assimilate the gain of the camera to the standard deviation of each pixel over a large number of movies, and they prove this is a successful way of identifying dead pixels. However, our results show that this approach does not provide a consistent gain estimation (Fig. 1).

3. Results and discussion

We have applied the presented algorithm to estimate the camera gain of all datasets in EMPIAR (Iudin et al., 2016) that have deposited movies (21 out of 72, i.e., 29.2% of EMPIAR). The cameras covered by these 21 movies are K2 (11 movies), Falcon II (7 movies), and Falcon I, DE12, DE20 (1 movie each). Additionally, FEI has kindly provided data from a FEI Falcon 3EC camera.

We first analyzed the correctness of our method as a tool to estimate the gain image "per se". Most EMPIAR entries do not have their camera gain image. However, EMPIAR 10010 and 10025 are examples of studies with submitted gain image. In Fig. 1 we show the experimental gain, the estimated gain, and the gain estimated by the algorithm of Afanasyev et al. (2015) (currently, the only published alternative to our algorithm). We can see that our estimated gain resembles the experimentally measured one, while the alternative estimate is measuring the pixel standard deviation (which, as can be seen in Fig. 1, it is not always similar to the gain). In both cases, we report the determination coefficient, R^2, for the two algorithms.

Once the gain is corrected on the movie, the residual gain image should look as a relatively flat image. Indeed, this is the case for most EMPIAR entries. As an example, Table 1 shows a statistical summary of pixel values corresponding to residual gain images from representative entries obtained by different types of cameras. As can be seen from the table, the gain is well concentrated around 1. These studies represent examples of well calibrated cameras.

Subsequently, we examined whether the common assumption that initial gain estimation is relatively constant during a single acquisition holds in practice. We tested it by analyzing the 600 movies of the entry EMPIAR 10028 (Falcon II). As shown in Fig. 2 top, the gain distribution for each movie is almost symmetric and centered around 1. The 2.5% and 97.5% percentiles are relatively fixed around the values 0.99387 and 1.00598 (that is, the gain is effectively concentrated in a range

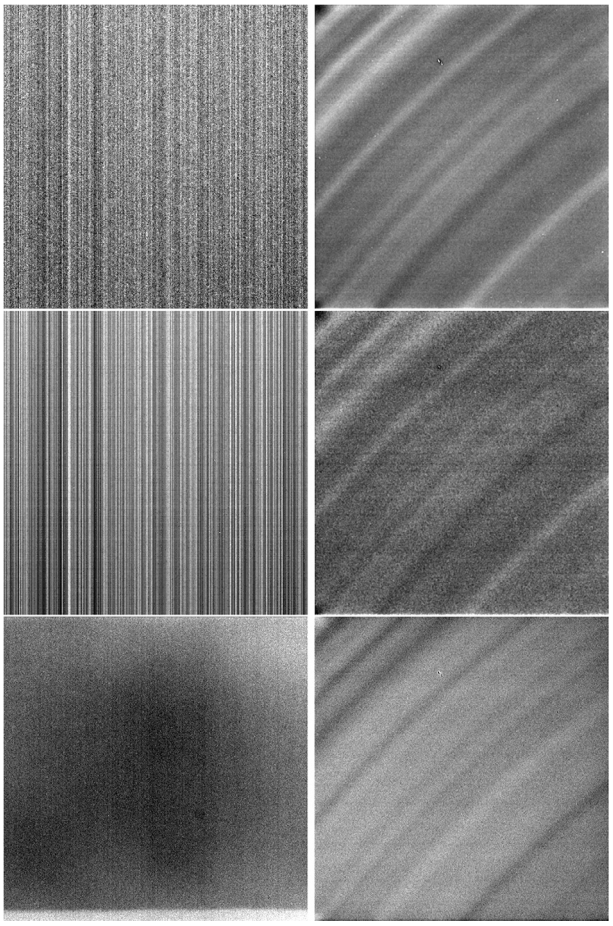

Fig. 1. Top: Gain image experimentally determined for 10010 (DE12, left) and 10025 (K2, right). Middle: Gain image estimated by the proposed algorithm ($R^2 = 0.91$, left, and $R^2 = 0.72$, right). Bottom: Gain image estimated by the algorithm of Afanasyev et al. (2015) ($R^2 = 0.30$, left, and $R^2 = 0.53$, right).

Table 1
Minimum, maximum, and standard deviation for the gains found for each of the datasets. The mean of the gain is set to 1.

Dataset	Min.Gain	2.5%	97.5%	Max.Gain	Std.Dev.
10025 (K2)	0.8332	0.9999	0.9999	3.1706	0.0052
10026 (Falcon II)	0.9756	0.9962	1.0038	1.0423	0.0020
Falcon 3EC	0.8290	0.9724	1.0262	2.0820	0.0138

approximately between −0.6% and 0.6% of its nominal value of 1). We have checked that to estimate the effective range (the 2.5 and 97.5 percentiles) it is not needed to perform the gain estimation on the whole movie, because 4 or 5 frames per movie normally produce reliable estimates of these quantities.

Then, we analyzed the hypothesis of no time drift of the gain on a separate dataset. In Fig. 2 middle, we have analyzed the behavior of the camera gain over long acquisitions (4.5 days) in our own Thalos Arctica microscope equipped with a Falcon II camera. During the acquisition, the camera gain was experimentally determined only at the beginning. We estimated the camera gain from movies every 12 h using our algorithm. As can be seen from the figure, there is a progressive flattening (Fig. 2 middle and bottom) of the gain histogram indicating that there is a drift over time of the camera gain. Interestingly the last gain histograms are heavily skewed indicating that this drift is different for bright and dark gray values.

In order to understand the real value of a systematic use of our proposed algorithm as an additional quality control procedure, we

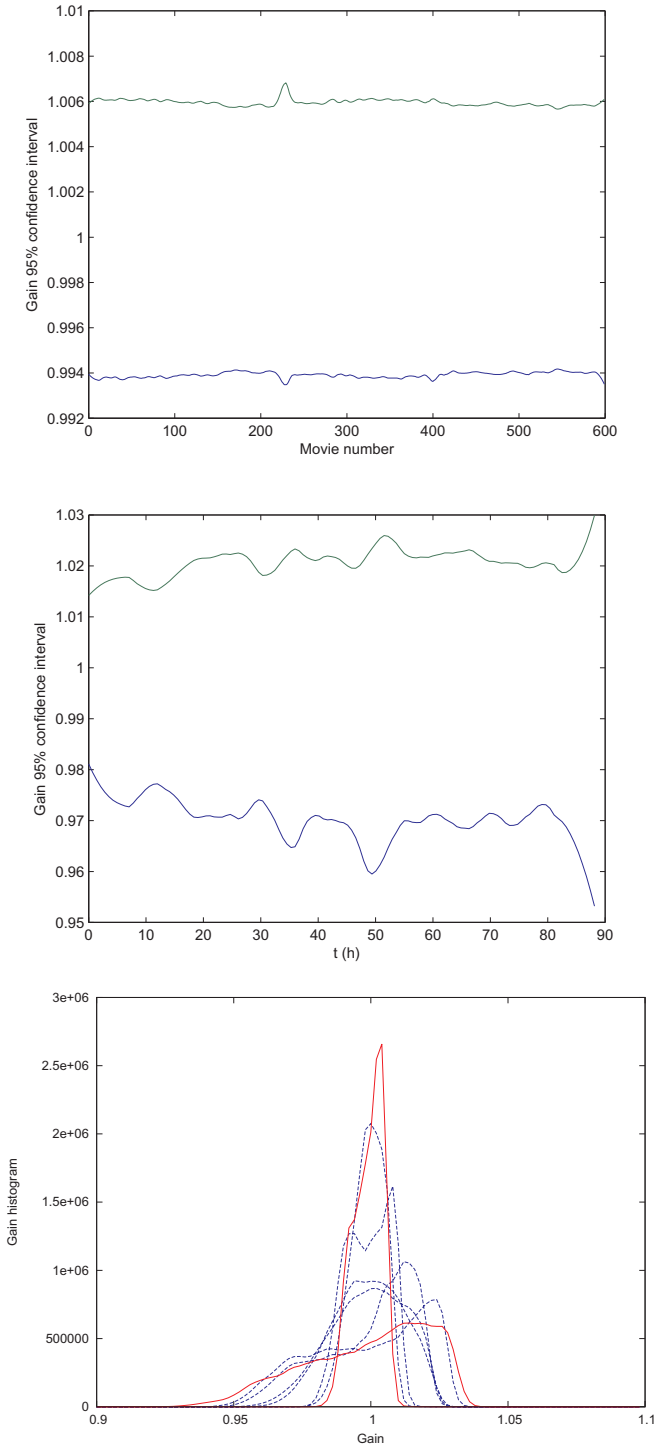

Fig. 2. Stability of the gain image along an acquisition. Top: The Y axis represents the 2.5 and 97.5 percentiles of the gain image estimated for each movie of the entry EMPIAR 10028 (Falcon II). Middle: Same plot for an in-house experimental acquisition spanning over 4.5 days in a Falcon II. Bottom: Gain histograms of representative movies of the in-house data from the beginning to the end of the acquisition (the first and last histograms are shown in red, and some intermediate time points are shown in blue). In this figure we can see a progressive flattening of the residual gain estimate. All curves have been smoothed with Rloess.

concentrated in those entries of EMPIAR that could be considered "outliers" from the point of view of producing residual gain estimations more different from a constant image. In this way, we could identify cameras with broken rows or columns (Fig. 3, top; EMPIAR 10023) or studies in which the gain image used for initial calibration had

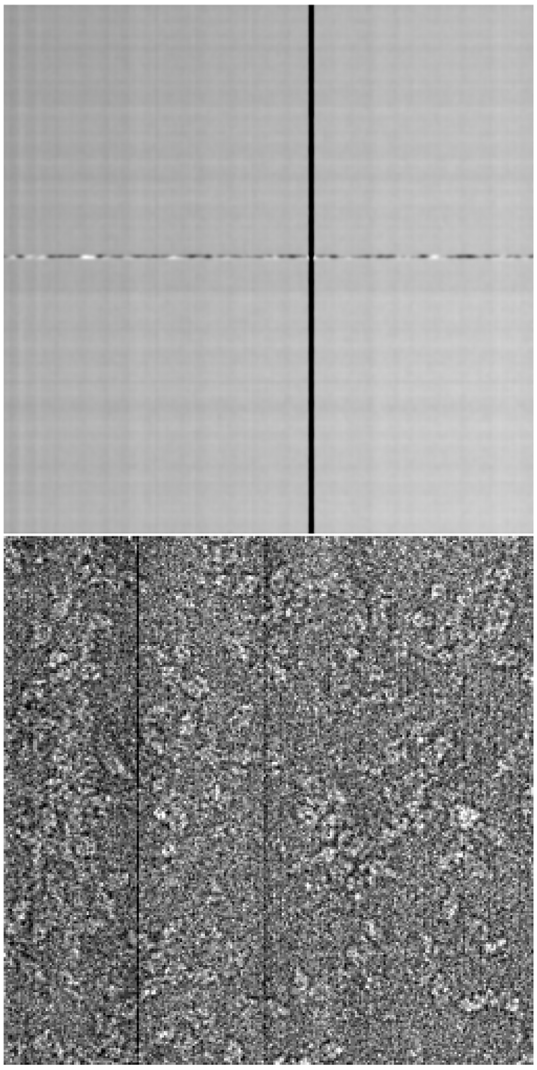

Fig. 3. Residual gain image for EMPIAR 10023 (Falcon II, top) and 10013 (K2, bottom).

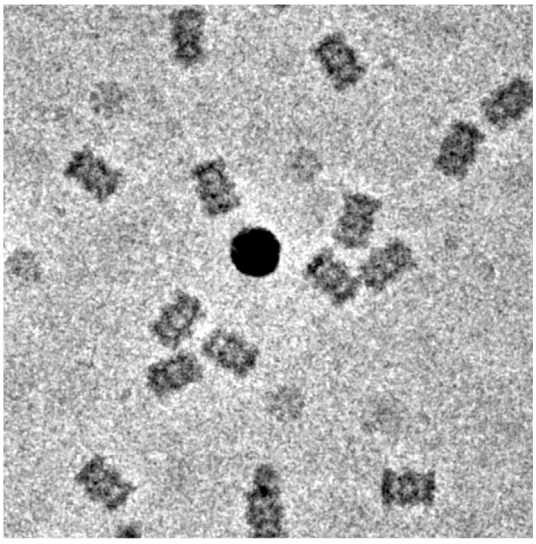

Fig. 4. Phase plate micrograph (EMPIAR 10057).

considerably drifted from the ideal one (Fig. 3, bottom; EMPIAR 10013). In particular, EMPIAR 10013 was supposed to be already corrected at acquisition; however, particles are seen in the residual gain image, which clearly should not be the case. We have observed this behavior in several EMPIAR studies, specifically, 10005 (K2), 10013 (K2), 10030 (K2), 10058 (K2), 10061 (K2), 10071 (DE20)) (a non-negligible subset of all entries with movies, close to 30% of the total). Phase plate movies also show particles in the residual gain image (EMPIAR entries 10050 (K2) and 10057 (K2)). However, the reason for this behavior is that these images strongly violate the assumption that the histogram of all rows and columns is basically the same (Eq. (3)), as can be seen in Fig. 4. Consequently, our algorithm is not valid for phase plate images.

Given the current body of EMPIAR entries, *ad hoc* thresholds that could be used to automatically flag out potentially incorrect calibrations are cases for which (a) residual gain standard deviation is larger than 0.04, or (b) the ratio between the 97.5 and 2.5 percentiles is larger than 1.15, or (c) the ratio between the maximum gain value and the 97.5 percentile is larger than 4.5.

This algorithm is publicly available in Xmipp (`xmipp_movie_estimate_gain`, de la Rosa-Trevín et al. (2013)) and Scipion (`estimate gain` protocol, de la Rosa-Trevín et al. (2016)), assuring simplicity and readiness of use. Additionally, it has been implemented to work in streaming mode so that the gain image can be regularly estimated from the movies as they are acquired by the microscope (that is, the estimate gain protocol can be part of the so-called Scipion Box, a system that tackles the preliminary image processing steps as soon as movies are acquired at the microscope). The algorithm has been implemented in both CPU and GPU and the execution time, in GPU, goes from 15 s to 3 min depending on the frame size and number of frames (4Kx4K, 8 frames in 15 s; 4Kx4K, 40 frames, 3 min; the execution time in CPU is 10 times larger). A monitor of the gain estimate can be set to warn the user if the residual acquisition gain goes beyond certain limits (defined by the user as thresholds on its standard deviation, the ratio between percentiles and the maximum and percentiles).

Acknowledgements

We thank FEI for supplying Falcon 3EC datasets, and Drs. Cuéllar and Valpuesta for the Thalos Arctica data. The authors would like to acknowledge economical support from: the Comunidad de Madrid through grant CAM (S2010/BMD-2305), the Spanish Ministry of Economy and Competitiveness through Grants AIC-A-2011-0638 and BIO2016-76400-R, and the Fundación General CSIC (Programa ComFuturo).

References

Afanasyev, P., Ravelli, R.B.G., Matadeen, R., De Carlo, S., van Duinen, G., Alewijnse, B., Peters, P.J., Abrahams, J.-P., Portugal, R.V., Schatz, M., van Heel, M., 2015. A posteriori correction of camera characteristics from large image data sets. Sci. Rep. 5, 10317.

de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184 (2), 321–328.

de la Rosa-Trevín,, J.M., Quintana, A., Del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J.L., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2016. Scipion: a software framework toward integration, reproducibility and validation in 3d electron microscopy. J. Struct. Biol. 195, 93–99.

Iudin, A., Korir, P.K., Salavert-Torres, J., Kleywegt, G.J., Patwardhan, A., 2016. Empiar: a public archive for raw electron microscopy image data. Nat. Methods 13, 387–388.

Kühlbrandt, W., 2014. The resolution revolution. Science 343 (6178), 1443–1444.

Sorzano, C.O.S., Marabini, R., Velázquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H.W., Carazo, J.M., Pascual-Montano, A., 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. 148, 194–204.

Tendero, Y., Landeau, S., Gilles, J., 2012. Non-uniformity correction of infrared images by midway equalization. Image Process. Line 2, 134–146.