Journal of Structural Biology 212 (2020) 107579

Contents lists available at ScienceDirect

Journal of Structural Biology

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

# Permutation testing of Fourier shell correlation for resolution estimation of cryo-EM maps

Maximilian Beckers[a,b,c,d,*], Carsten Sachse[a,c,d,e,*]

[a] European Molecular Biology Laboratory (EMBL), Structural and Computational Biology Unit, Meyerhofstraße 1, 69117 Heidelberg, Germany
[b] Candidate for Joint PhD degree from EMBL and Heidelberg University, Faculty of Biosciences, Germany
[c] Ernst-Ruska Centre for Microscopy and Spectroscopy with Electrons (ER-C-3/Structural Biology), Forschungszentrum Jülich, 52425 Jülich, Germany
[d] JuStruct: Jülich Center for Structural Biology, Forschungszentrum Jülich, 52425 Jülich, Germany
[e] Department of Chemistry, Heinrich Heine University, Universitätsstraße 1, 40225 Düsseldorf, Germany

ARTICLE INFO

*Keywords:*
Cryo-EM
Fourier shell correlation
Resolution criteria
Permutation testing
False discovery rate

ABSTRACT

Fourier shell correlation (FSC) has become a standard quantity for resolution estimation in electron cryo-microscopy. However, the resolution determination step is still subjective and not fully automated as it involves a series of map interventions before FSC computation and includes the selection of a common threshold. Here, we apply the statistical methods of permutation testing and false discovery rate (FDR) control to the resolution-dependent correlation measure. The approach allows fully automated and mask-free resolution determination based on statistical thresholding of FSC curves. We demonstrate the applicability for global, local and directional resolution estimation and show that the developed criterion termed FDR-FSC gives realistic resolution estimates based on statistical significance while eliminating the need of any map manipulations. The algorithms are implemented in a user-friendly GUI based software tool termed SPoC (https://github.com/MaximilianBeckers/SPOC).

## 1. Introduction

Electron cryo-microscopy (cryo-EM) is becoming established as one of the methods of choice for macromolecular structure determination. The technique has undergone major improvements in hardware and software, allowing to determine 3D structures routinely at close-to-atomic resolution (Kühlbrandt et al., 2014; Bartesaghi et al., 2015; Weis et al., 2019). These resolutions provide the landmark features for atomic model building and the resulting atomic coordinates rationalize biological mechanism and function. At the core of all these developments are improvements in the resolvability of molecular detail that can be achieved. Therefore, the resolution is a reported value that makes the data comparable with other structural biology methods, and more importantly, gives guidance how confidently the density of a given feature in 3D space can be interpreted. Fourier shell correlation (FSC) curves are the standard metric for resolution estimation of cryo-EM maps. FSCs were originally introduced to the field of cryo-electron microscopy (Harauz and Van Heel, 1986) and more recently were also applied to related fields such as super-resolution microscopy (Nieuwenhuizen et al., 2013; Banterle et al., 2013).

The FSC measures the correlation between Fourier coefficients within every resolution shell of two independently determined half-maps. A typical curve shows high correlations at low resolutions until it drops to zero at higher resolutions when noise starts to dominate the signal. In order to report a resolution value of structures based on FSC curves, a threshold value needs to be selected. A fixed threshold of 0.143 has been proposed (Rosenthal and Henderson, 2003), which is widely used for resolutions better than 10 Å when map features can be used to validate the obtained resolution. For lower resolutions, a more conservative 0.5 threshold has typically been used (Spahn et al., 2004). The 0.5 threshold is also favored when local FSCs within small windows are computed and local resolution is estimated (Cardone et al., 2013). Nevertheless, fixed value thresholds ignore the effective number of independent Fourier coefficients in the respective resolution shell (Van Heel and Schatz, 2005). This effect becomes particularly relevant at local resolution estimation with small window sizes (Cardone et al., 2013) and for 3D FSC with small number of Fourier pixels (Zi Tan et al., 2017). Alternatively, other criteria like $\sigma$ (Saxton and Baumeister, 1982; Orlova et al., 1997) as well as the half-bit criterion have been proposed that compensate for these effects (Van Heel and Schatz, 2005). More generally, applying thresholds such as the 0.143 and 0.5 criteria are to be interpreted as identifying the highest resolution shell that still

* Corresponding authors at: Ernst-Ruska Centre for Microscopy and Spectroscopy with Electrons (ER-C-3/Structural Biology), Forschungszentrum Jülich, 52425 Jülich, Germany.
*E-mail addresses:* maximilian.beckers@embl.de (M. Beckers), c.sachse@fz-juelich.de (C. Sachse).

https://doi.org/10.1016/j.jsb.2020.107579
Received 6 March 2020; Received in revised form 1 July 2020; Accepted 15 July 2020
Available online 18 July 2020

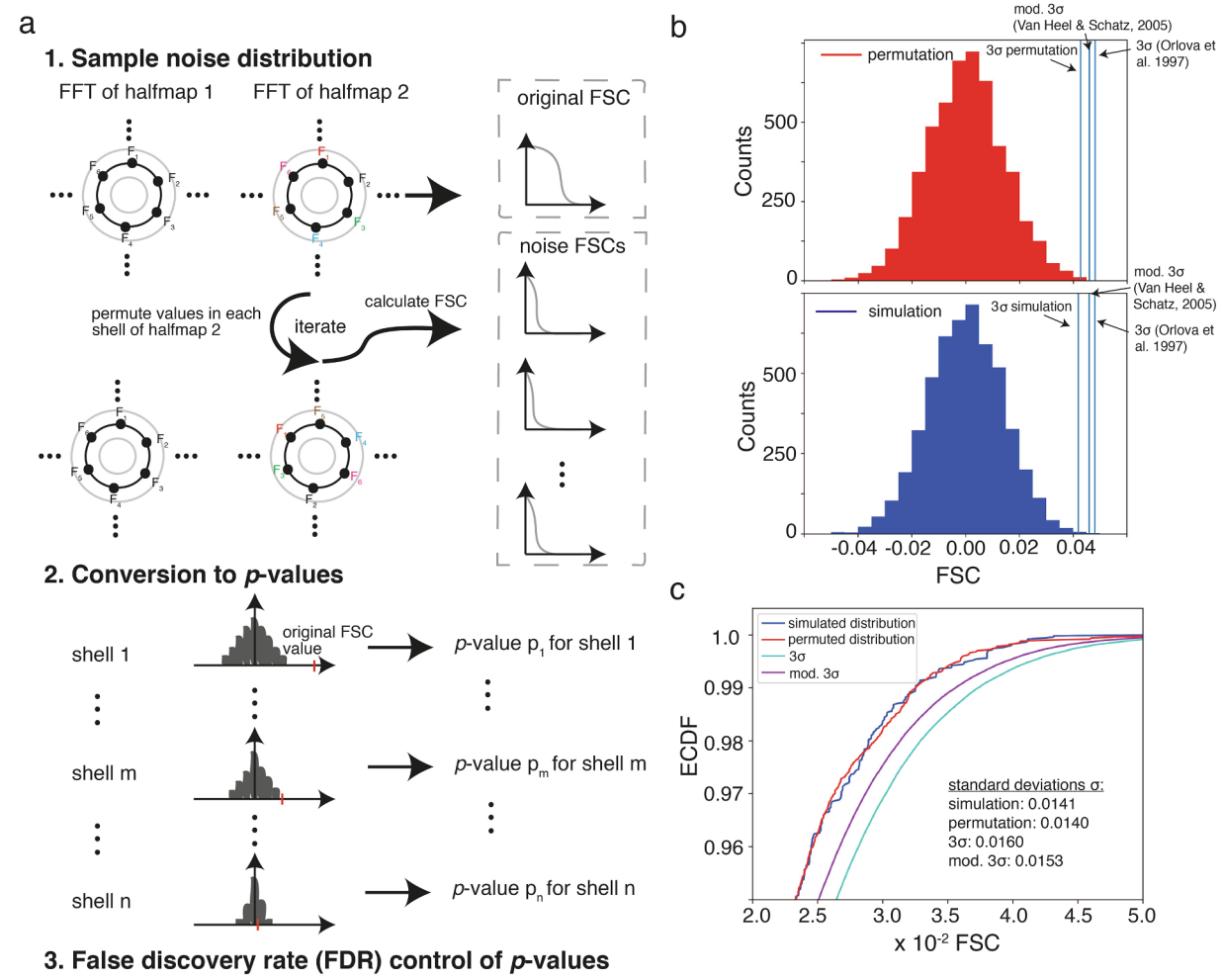


**Fig. 1.** Resolution estimation by permutation-based FDR-FSC. (a) Samples of random noise FSCs are generated by permuting Fourier coefficients in half-map 2, creating newly paired Fourier coefficients (1.). The resulting random noise FSC distribution for each resolution shell is used to estimate the significance, i.e. a *p*-value for the original FSC value of each resolution shell (2.). The resulting *p*-values per resolution shell are subjected to false discovery rate (FDR) control (3.). (b) Histograms of permuted and simulated FSC values of independent half-maps from the half-Nyquist resolution shell (1/4 pixel) together with the respective 3$\sigma$ cutoffs. (c) Comparison of the empirical cumulative distribution functions (ECDFs) in a zoomed view. The true distribution sampled by 5000 simulations of two noise maps and subsequent FSC calculation is shown (blue). The distribution obtained by the permutation approach is shown (red). ECDFs with standard deviations from the $\sigma$-threshold criteria are shown in blue (modified 3$\sigma$ criterion) and cyan (original 3$\sigma$ criterion). The true and the permutation-based distributions follow each other closely, especially at the tail of the distributions. Moreover, the modified 3$\sigma$ is closer to the simulated distribution than the original 3$\sigma$..

has a certain signal-to-noise ratio (SNR) or information content present.

As an alternative to fixed thresholds, $\sigma$-thresholds were proposed to provide cutoffs when the FSC exceeds random noise correlations (Saxton and Baumeister, 1982; Van Heel and Schatz, 2005). The $\sigma$ curve gives a significance threshold level for every resolution shell to compare it with the correlation level. It is simply estimated as a function of voxels $n(r_i)$ in the Fourier shell with resolution $r_i$ by:

$$\sigma(r_i) = \frac{\sigma_{factor}}{\sqrt{n(r_i)/2}} \quad (1)$$

$\sigma_{factor}$ specifies the multiple of the standard deviation used for thresholding, i.e. for a 3$\sigma$ cutoff we have $\sigma_{factor} = 3$. A modified version, $\sigma_{mod}(r_i)$, has been proposed later that takes into account the symmetry (Orlova et al., 1997) and the filling degree of the map (Van Heel and Schatz, 2005) and is given as

$$\sigma_{mod}(r_i) = \frac{\sigma_{factor}}{\sqrt{n_{eff}(r_i)/2}},$$
$$\text{with } n_{eff} = \frac{n(r_i) * \left(\frac{3}{2}\frac{D}{L}\right)^2}{2 * n_{asym}} \quad (2)$$

where $n_{asym}$ is the number of asymmetric volume units in the map, D the linear object size and L the size of the volume. However, statistics based on simple $\sigma$-thresholds suffer from several drawbacks as linking $\sigma$-levels to significance requires strict assumptions about the underlying correlation distributions, which are usually not known.

In order to get correct estimates of resolution-dependent information measures of the molecular density within a 3D reconstruction, solvent noise needs to be removed or flattened from the volume, as big parts of the reconstructed structure correspond to solvent noise only and thus decreases the resolution values. Therefore, application of a user-defined mask that encloses the particle shape is critical to remove solvent noise, yet poses the danger of introducing artificial correlations. Often, this effect is compensated for by mask deconvolution (Chen et al., 2013). In practice, this correction approach still involves empirical testing of several masks to improve the resolution while avoiding artificial correlations. The computation of the FSC in the absence of such a user-defined mask will be referred to as unmasked FSC throughout the manuscript. Although the calculation of unmasked and corrected FSC curves considering the molecular mass have been proposed to circumvent this problem (Sindelar and Grigorieff, 2012), such an approach still requires the selection of an FSC threshold. Here, we propose an approach of FSC thresholding for resolution determination

using a combination of permutation testing and *p*-value correction by false discovery rate (FDR) control. We show that the procedure is able to detect signal reliably over a wide range of noise levels that eliminates the need of solvent noise flattening. We validate the approach by comparing a large batch of structures with reported resolution values obtained from the EM databank and further extend the approach successfully to challenging cases of local and directional resolution determination.

## 2. Results

### 2.1. The procedure of permutation testing of Fourier shell correlations

In order to circumvent principal and practical problems of thresholding FSC curves, we developed a procedure for identifying the highest resolution shell of interpretable signal based on parameter-free permutation sampling and subsequent statistical inference. Permutation tests are statistical procedures for estimating the distribution of a quantity of interest from the data itself and thus do not require any prior knowledge about the underlying distributions (Lehmann and Romano, 2005). Permutation sampling of the FSC for each resolution shell is straightforward: we generate new samples by changing the order of the Fourier coefficients of the second half-map shell and compute a large series of FSCs (Fig. 1a). By this approach, any pre-existing correlations between Fourier coefficients will be canceled and, therefore, yield a sample of the noise distribution of the FSC for each resolution shell. When applied to every resolution shell, the distributions together with the original FSC-values can then be statistically tested and conveniently transformed into *p*-values. This approach is invariant to the color of the noise as we sample every resolution shell independently and we do not impose any assumptions on the FSC distribution. When we test multiple resolution shells in parallel, there is a higher risk of falsely identifying resolution shells as significant, i.e. to make a false positive discovery, which is also known by the multiple testing problem in statistics. In order to reduce the risk of false positive errors, *p*-values are further corrected by means of FDR control (Benjamini and Hochberg, 1995) control and thresholded at 1%. We term this approach the FDR-FSC method of resolution determination.

### 2.2. Permutation sampling of Fourier shell correlation coefficients

In order to verify whether the permutation approach captures the principal distribution of Fourier correlations per shell, we simulated two pure noise reference half-maps with an average of 0 and a standard deviation of 1 and applied the permutation sampling as described. To assess the true distribution of Fourier correlations per shell, we simulated 5000 noise half-maps (again with an average of 0 and a standard deviation of 1) and calculated FSC curves for each pair of 5000 half-maps. Comparison of the simulated with the permuted histograms generated from the FSC values of the half-Nyquist resolution shell (1/4 pixel) and the corresponding empirical cumulative distribution functions (ECDF) show that the distributions from the permutation approach and the simulation follow each other closely, including the tail of the distributions (Fig. 1b and c). Furthermore, we compared the standard deviations from the true simulation FSC distribution with the one based on the $\sigma$ or modified $\sigma$ criteria obtained from the two pure noise reference half-maps (Van Heel and Schatz, 2005). The $\sigma$ or modified $\sigma$ distributions yield systematic deviations of the true FSC distribution from the simulation whereas the permutation matches the FSC distribution from the simulation well. Next, we wanted to assess the performance over all resolution shells and plotted the right-sided 10, 5.0, 1.0 and 0.15% cutoff values (percentiles) for each resolution shell. The percentiles of the distributions from the permutation and the simulation follow each other closely for both white (Supplementary Fig. 1a) and colored noise (Supplementary Fig. 1b). In order to assess whether signal present in maps affects the results of permutation sampling, we simulated three maps with increasing signal-to-noise ratios of 0.3, 3.0 and 30. The ECDF between noise and signal-containing FSC distribution at half-Nyquist show that they match very closely regardless of the presence of higher signal (Supplementary Fig. 2). Moreover, the percentiles of the FSC distributions of the entire set of Fourier shells confirm that increasing levels of signal over noise does not affect the permutation sampling. Together, the permutation approach is able to accurately sample the noise distribution of the FSC in the presence of signal as well in the absence of signal with both white and colored noise in the maps. The standard deviations computed from permutations are more accurate than the standard deviations obtained from the $\sigma$ criteria.

### 2.3. Resolution estimation using the FDR-FSC threshold

Based on the proposed FDR-FSC approach, the resolution estimate will be assigned to the highest spatial frequency that contains significant signal at 1% FDR (from now on referred to as FDR-FSC). For conventional FSC computation, the half-maps are masked by the user with a shape closely following the outline of the particle for the purpose of solvent flattening. In contrast, for the here proposed FDR-FSC approach the untreated half-maps are placed in the soft reconstruction sphere and directly used for FSC determination. Due to the well-known shape of the FSC curve, the resolution value will be assigned to the first resolution shell that crosses the significance threshold when progressing along the FSC in reciprocal resolution (Fig. 2a). Comparison of the conventional 0.143 cutoff values thresholds for a $\gamma$-secretase map (Bai et al., 2015) (EMD3061) at 3.4 Å obtained in the presence of a user-defined mask with the 3.4 Å FDR-FSC value obtained in the absence of solvent flattening indicates a very similar resolution. Moreover, the 1% FDR threshold crosses the significance level two shells earlier than a simple *p*-value threshold of 1%. For a series of eight EM databank (EMDB) entries (EMD2677, EMD0587, EMD4589, EMD0043, EMD3061, EMD0415, EMD6287, EMD8908), we used the raw half-maps for FSC computation and assessed the influence of the chosen FDR threshold on the resolution measurement. As for commonly used significance levels between 0.1 and 10% FDR, the estimated resolution is robust with respect to varying FDR thresholds (Supplementary Fig. 3a), we continue to use the 1% FDR threshold for the FDR-FSC resolution estimation. In order to test the influence of the *p*-value correction by FDR control, we also determined the resolution values using simple *p*-value thresholding to the above structure series of eight EMDB entries (Supplementary Fig. 3b). We observe that the final resolution cutoff changes only marginally within the first decimal place. To recapitulate the thresholding step in more detail, we plotted the sorted *p*-values and FDR controlled *p*-values derived from the FSCs of the human TFIIH core complex (EMD0587) and TMEM16 lipid scramblase (EMD4589) (Supplementary Fig. 3c). We find that FDR-control renders the *p*-values closer to a binary distribution of zeros and ones and therefore, minimizes the number of resolution shells to be considered for significance discrimination. The thresholding can also be alternatively presented by a horizontal cutoff bar in the case of *p*-values or a cutoff line in the case of applied significance testing of FDR control thereby separating the significant *p*-values from the insignificant ones. The zoomed inset of the plot reveals that FDR controlled *p*-values is less dependent on the chosen cutoff value in comparison with the simple *p*-value threshold (Supplementary Fig. 3d). In order to make the thresholding of resolution based on permutation testing as robust as possible, we routinely correct *p*-values by FDR control and assign the resolution cutoff to the maximum resolution shell of a large continuous resolution band showing significant Fourier correlations.

### 2.4. Effects of symmetry and reconstruction sphere on FSC permutation sampling

Map processing operations like masking and symmetrization

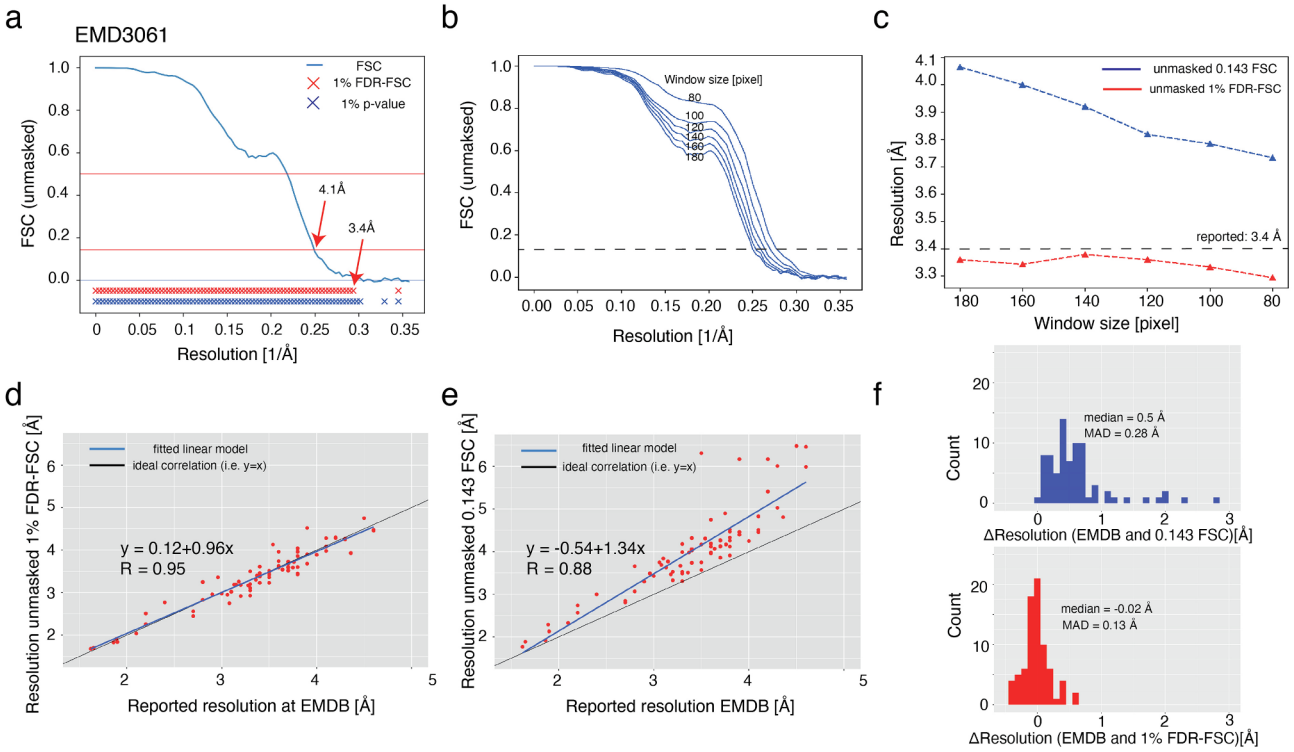


**Fig. 2.** Effects of surrounding solvent noise on FDR-FSC and benchmarking resolution estimation using 77 maps from the EMDB. (a) Example of a Fourier shell correlation (FSC) curve for γ-scretase map (EMD3061) computed from two half-maps without solvent flattening. Resolutions shells with significant correlations beyond random fluctuations at 1% FDR-FSC are marked with red crosses and for a simple 1% *p*-value threshold with blue crosses. (b) Effect of removing solvent noise on FSC curve by decreasing window sizes in steps of 20 voxels and thereby incrementally excluding solvent. (c) Effect of removing solvent noise on resolution estimates at 0.143 FSC (blue) compared with 1% FDR-FSC (red). (d) Scatter plot of reported resolutions against resolutions determined by the 1% FDR-FSC method. Fitted line (R = 0.95) is shown in blue. For comparison, ideal correlation is shown in black. (e) Scatter plot of reported resolutions against unmasked 0.143 FSC values. (f) Two histograms of resolution differences of the 1% FDR-FSC (red) and the unmasked 0.143 cutoff with respect to the reported resolutions (blue).

applied to both half-maps lead to higher correlations between Fourier shells. These principal effects on the FSC computation with consequences for threshold determination have been raised previously and were proposed to be compensated by correction factors (Van Heel and Schatz, 2005). Introduced dependencies between Fourier coefficients in a shell affect the distribution estimation of Fourier correlations by the outlined permutation approach. For example, when symmetry is applied to the image reconstruction, symmetry-related Fourier voxels will be correlated regardless of whether they contain signal or noise. When standard FSC permutation is performed with equally symmetrized half-maps, dependencies between Fourier coefficients will give rise to a narrower distribution of Fourier correlations than can really occur with such symmetrized half-maps. In statistics, such a configuration can be described by the reduced number of independent variables that are involved in a statistic or, in other words, the number of degrees of freedom of a system is decreased. Therefore, as the sample of Fourier corefficients behaves like a sample of smaller size, subsampling is required to better estimate the true variation of Fourier correlations per shell.

In order to verify the requirement of permutation subsampling and identify the most efficient implementation, we tested 2-fold, 4-fold and 10-fold rotationally averaged maps from the EMDB (EMD0408, EMD5778, EMD10924). First, we performed permutation of Fourier correlations of the complete shell resulting in the nominal sample size for reference computation. Second, in order to account for the reduced number of degrees of freedom, we limited permutation of Fourier correlations per shell to the asymmetric unit, i.e. one half, quarter or tenth of the Fourier transform, respectively divided along the symmetry axis. Third, we reduced the number of degrees of freedom using random subsampling of the complete shell resulting in an effective sample size. In contrast to the nominal sample size, the effective sample size approach reduces the number of Fourier coefficients per shell used for permutation by the symmetry averaging factor, e.g. C4 symmetry has four asymmetric volume units and reduces the effective sampling to 1/4 or 25% of the initial number of Fourier coefficients. For each permutation approach, we plotted the resulting right-sided 10, 5.0, 1.0 and 0.15% percentiles for each resolution shell (Supplementary Fig. 4). The resulting assessment of the distributions of Fourier correlations per shell shows that, as expected, when the complete shell is permuted the distribution is narrower than when limited to the asymmetric unit. This effect is more pronounced for higher degrees of symmetry such as C10. Notably, the percentiles obtained from permutation results by limiting the Fourier coefficients to the asymmetric unit and by effective sample sizes overlap very closely. The results show that random subsampling by effective sample sizes present a computationally efficient way to take into account dependencies between Fourier coefficients for the FSC permutation sampling approach. Effective sample sizes are commonly used in Monte Carlo simulations, a class of methods that the proposed permutation belongs to. As the number of degrees of freedom are reduced by introduced dependencies, effective sample sizes are used when the sample behaves like a sample with smaller size (Brooks et al., 2011). Finally, the determined 1% FDR-FSC resolution results match the reported EMDB values closely: EMD0408 with C2 symmetry: 2.9 (3.2 Å reported), EMD5778 with C4 symmetry: 3.5 (3.3 Å reported) and EMD10924 with C10 symmetry: 4.5 (4.4 Å reported). The latter test of the C10 symmetry also shows that when effective sample size correction is ignored, incorrect resolutions claiming better resolution than 4.2 Å are obtained. Together, the results show that taking into account dependencies between Fourier coefficients used for FSC computation is important for the proposed permutation testing and can be efficiently implemented by effective sample size correction in the permutation

procedure.

Similar effects of introducing correlations between Fourier coefficients occur during the reconstruction procedure within a sphere. This step is equivalent to multiplication of the structure by a mask and corresponds to a convolution of the shape with the structure in Fourier space. Most reconstruction software programs already apply variants of such a spherical mask with approx. the diameter of the volume size to limit the structure to the reconstruction sphere. In order to introduce a constant effective sample size correction factor in our computation, we decided to place each structure inside the same reconstruction sphere. For this purpose, we designed a soft reconstruction sphere of volume dimension diameter including a Gaussian falloff of two voxel standard deviation. Unfortunately, it is not straightforward to derive an effective sample size for this map operation. Therefore, we estimated the effective sample size for the effect reconstruction sphere empirically by minimizing the deviation between the true distribution obtained from 5000 noise half-maps (with reconstruction sphere applied) and the distribution calculated from two noise half-maps (with reconstruction sphere applied) by permutation including effective sample sizes. By systematically testing different effective sample sizes and computing the Kolmogrow-Smirnow distance of the two distributions (Massey, 1951) (see Methods for more details), we found that our reconstruction sphere reduces the effective sample size to 70% of the initial sample size (Supplementary Fig. 5a). We also tested a common Hann window, as it is used for local resolution FSC (below), and identified 23% as the sample size correction factor (Supplementary Fig. 5b). The utility of effective sample sizes is demonstrated by the comparison of permuted and simulated percentiles using a series of different window sizes: the resulting percentile lines follow each other closely only when effective sample sizes are considered (Supplementary Figs. 6 and 7). It should be noted that the tests also show that the application of any user-defined mask for solvent flattening the structure of interest will affect permutation sampling of the FDR-FSC procedure and is, therefore, not recommended in the absence of a mask-specific sample size correction factor.

After assessing the cases of symmetry and reconstruction sphere with regards to sample size correction separately, in deposited half-maps containing symmetry, both effects on the effective sample sizes need to be considered. As a simple approximation, we combine both corrections, e.g. in case of C4 symmetry the effective sample size becomes 1/4 times 0.7 times the nominal sample size. In order verify the approach, we tested permutation-based sampling of two symmetry-imposed half-maps in the presence and absence of the effective sample sizes. In analogy to the test on the permutation approach above, we compared the permuted distributions with the ones obtained 5000 noise symmetrized half-maps by plotting the 10, 5.0, 1.0 and 0.15 percentiles as a function of spatial frequency. In the absence of any effective sampling size correction, the permuted distribution gives rise to tail probabilities that are too small in the presence of a D4 and D7 symmetry. The combined effect of symmetry and reconstruction sphere, however, can be conservatively corrected by consideration of the effective sample size (Supplementary Fig. 8a and b). In conclusion, operations that introduce additional correlations between half-maps such as volume symmetrization and reconstruction sphere require consideration of reduced degrees of freedom for permutation by effective sample sizes.

### 2.5. FDR-FSC estimation and surrounding solvent noise

To test the sensitivity of the FDR-FSC approach with respect to noise, we successively reduced the size of the volume of the γ-secretase map (EMD3061) in steps of 20 voxels, thereby successively removing surrounding solvent noise. It is evident that FSC curves show higher correlations towards the Nyquist frequency as more solvent noise is removed (Fig. 2b). While the estimated resolution of 0.143 FSC threshold decreases from 4.1 Å to 3.7 Å (Fig. 2c), the 1% FDR threshold practically remains constant and only fluctuates at the second decimal digit. In order to validate this noise tolerance of the FDR-FSC approach, we simulated two noise-free reference half-maps of β-galactosidase (PDB ID 5a1a (Bartesaghi et al., 2015)) at 2.2 Å resolution at a pixel size of 0.637 Å and scaled their signal to a standard deviation of 1. Next, we added colored Gaussian noise to the half-maps including a B-factor falloff of 100 Å$^2$ resulting in a final SNR of approx. 1.5 (Supplementary Fig. 9a). We computed the FSC and determined the resolution by the FDR-FSC approach (Supplementary Fig. 9b). To test the effect of additional solvent noise on the resolution determination, we increased the volume size from 300 to 700 pixels leading to a larger content of solvent noise in the map. In support of the results from the experimental γ-secretase map, we found that the resolutions obtained by the FDR-FSC criterion practically remain constant at 2.2 Å when increasing the solvent content, whereas the resolutions obtained from the 0.143 FSC threshold are highly sensitive to the content of solvent noise in the map (Supplementary Fig. 9c) and thereby increase in numbers from 2.3 to 3.3 Å. Taken together, the FDR-FSC criterion sensitively estimates resolution even in cases of large amounts of solvent noise present in the volume and therefore eliminates the need of solvent flattening of the half-maps by a user-defined mask.

### 2.6. Performance comparison of FDR-FSC with common fixed threshold FSC resolution determination

In order to benchmark the proposed algorithm, we compared the reported resolutions for a high-resolution and a low-resolution set of deposited half-maps from the EMDB. For the high-resolution comparison, we used 77 reference half-maps with resolutions ranging from 1.6 to 4.5 Å, determined the resolution using the FDR-FSC approach, which does not require any solvent flattening by a user-defined mask, and compared them with the reported resolutions. The scatter plot indicates high correlation (R = 0.95) between the 77 resolution estimates with the reported resolutions at the EMDB (Fig. 2d), while the 0.143 FSC threshold correlation deteriorates in the absence of solvent flattening due to the presence of solvent noise (Fig. 2e). The resolution differences to the reported EMDB resolutions are minimal in the case of 1% FDR-FSC whereas the differences to the reported resolutions for the 0.143 threshold without any solvent flattening have a much higher median deviation of 0.5 Å (Fig. 2f, Supplementary Table 1). Although for lower-resolution structures, deposited half-maps are rare and resolutions are more difficult to validate based on the visible features, results of our 1% FDR-FSC procedure also show good agreement with the reported resolutions in such cases (Table 1). In conclusion, the determined resolution estimates using 1% FDR-FSC are in close agreement with previously reported resolution values from the EMDB and can be computed in a fully automated fashion without any free parameters that have to be optimized by the user.

**Table 1**
FDR-FSC resolution estimates of low-resolution maps.

| EMDB-ID | Symmetry | Reported in EMDB [Å] | Pixel size [Å] | 0.143 unmasked FSC [Å] | 1% FDR-FSC [Å] |
|---|---|---|---|---|---|
| EMD9625 | C1 | 6.78 | 2.64 | 7.71 | 6.48 |
| EMD9306 | C1 | 7.5 | 1.32 | 9.39 | 8.05 |
| EMD9779 | C1 | 16 | 2.24 | 16.87 | 11.03 |
| EMD4748 | C2 | 5.14 | 0.85 | 6.07 | 5.0 |
| EMD0074 | C1 | 18 | 5.24 | 31.44 | 21.68 |
| EMD0075 | C1 | 21 | 5.24 | 31.44 | 21.68 |
| EMD0086 | C1 | 14 | 2.62 | 20.96 | 13.1 |
| EMD0087 | C1 | 16 | 2.62 | 32.25 | 23.29 |
| EMD20913 | O | 6.3 | 1.58 | 6.77 | 6.32 |

### 2.7. Application of FDR-FSC to local resolution estimation

Cryo-EM maps usually exhibit local variations of resolutions and estimating local resolutions has become critical in order to evaluate the biological structures (Cardone et al., 2013; Kucukelbir et al., 2014; Vilas et al., 2018). We reasoned that the FDR-FSC thresholding approach could provide benefits over fixed FSC-threshold approaches. Therefore, we extended our approach by computing local FSC curves and tested it on cases with large resolution variations. In order to design a test for assessing local resolution determination, we simulated a composite map from $\beta$-galactosidase (PDB ID 5a1a (Bartesaghi et al., 2015)) of filtered subunits at resolutions of 2.0, 3.0, 4.0 and 5.0 Å, scaled the signal to a standard deviation of 1 and included additional Gaussian white noise (Supplementary Fig. 9d). Next, we computed the local FSC's from overlapping cubes including a Hann window of half the cube dimension followed by permutation testing including the respective sample size correction factor. Using local 1% FDR-FSC thresholding, we determined local resolutions measures at different noise levels that ranges in standard deviations between 0.5 and 1.5. In comparison, the commonly used 0.5 local FSC threshold shows stronger differences to the simulated resolutions at higher noise levels and tends to give lower resolutions in comparison with the 1% FDR-FSC threshold at low noise levels (Supplementary Fig. 9e). Application to an experimental 3.4 Å γ-secretase map (Bai et al., 2015) (EMD3061) shows that the 1% FDR-FSC criterion assigns higher resolutions to the best-resolved core of the membrane protein when compared with the 0.5 FSC cutoff, while at the same time, lower resolutions between 10 and 15 Å are found at peripheral glycosylations and the detergent belt (Fig. 3a top left). Both observations can be supported by the available visual density features and are thus favorable for the 1% FDR-FSC resolution estimation method. An important aspect is the limited number of resolution shells included in the FSC computation due to the limited window size. In order to test the window size effect on the resolution, we computed the local resolutions by the 1% FSC-FDR as well as the 0.5 local threshold FSC method of the experimental 3.4 Å $\gamma$-secretase map (Bai et al., 2015) (EMD3061) using increasing window sizes from 10 to 50 pixels (Supplementary Fig. 10a). For very small windows (10 and 15), we observe too high resolutions, presumably due to undersampling, whereas for windows larger than 20 pixels we observe constant resolution values in the high-resolution transmembrane domain (Supplementary Fig. 10b left). Using a 0.5 FSC threshold, the resolution is estimated worse than expected at 3.8–4.0 Å, whereas the 1% FDR-FSC threshold yields a resolution of 3.4 Å, which is supported visually by the features of the cryo-EM density map. In the low-resolution region corresponding to the detergent micelle, however, the resolution with respect to window size is difficult to assess by visual features. In principle, the window size needs to be sufficiently large to allow sampling of lower spatial frequencies. At same time, larger windows loose locality of resolution information and will include different resolution information from other map parts. In this case, the inclusion of the better resolved transmembrane region in the window explains why resolution gets better beyond sizes of 35 pixels (Supplementary Fig. 10b right). Taken together, for local resolution estimation using the FDR-FSC approach, the window size remains a free parameter that needs to be chosen as a compromise between locality of resolution information and sufficient sampling of lower spatial frequencies.

Further comparisons with available software packages computing local resolution show that using standard settings applied to the γ-secretase map, ResMap (Kucukelbir et al., 2014) assigns more optimistic resolutions, whereas MonoRes (Vilas et al., 2018) gives estimates of more conservative resolution (Fig. 3a right). The local resolution histogram of ResMap lacks a low-resolution peak from the detergent whereas MonoRes assigns too conservative estimates around 4–5 Å to the high-resolution parts, similar to 0.5 FSC (Fig. 3a bottom). The local resolution histograms show that the FSC methods cover both high and low-resolution features better than the other two approaches. Additional analyses of a 3.9 Å ATP synthase (Guo et al., 2019) (EMD9333) and a eukaryotic ribosome (Juszkiewicz et al., 2018) (EMD0194) map confirms this notion that the 1% FDR-FSC criterion is able to better resolve high-resolution differences compared with 0.5 FSC and MonoRes (Fig. 3b and 3c). In these cases, the 1% FDR-FSC estimated resolution measures can be validated by visual features appearance such as the low-resolution L1-stalk domain > 7 Å and the clearly visible side-chain density of the α-helical segment at 3–4 Å. Despite the sensitivity to window sizes inherent to any local FSC determination, the here presented 1% local FDR-FSC approach can robustly assign local resolution in cases of variable local noise levels within the map.

### 2.8. Application of FDR-FSC to directional resolution estimation

In addition to local resolutions, directional resolutions have been recently evaluated to assess the effect of preferred particle orientations (Zi Tan et al., 2017). In practice, the FSC curve from a conical 3D Fourier transform is calculated including voxels of a specified angle. In analogy to local resolution windows, due to the limited number of Fourier coefficients inside a conical volume, directional FSCs suffer from poor statistics and therefore the FDR-FSC approach could provide similar benefits as demonstrated in the case of local FSCs. We tested the approach in more detail using three different cases: a map of the soluble portion of the small influenza hemagglutinin (HA) trimer with highly preferred orientations (EMD8731, Fig. 4a and 4d) (Zi Tan et al., 2017), a highly symmetric apoferritin map (EMD0144, Fig. 4b) (Zivanov et al., 2018) and an asymmetric map of γ-secretase (EMD3061, Fig. 4c) (Bai et al., 2015). Inspection of the directional resolution plots reveals that the 0.143 FSC-thresholds tend to give more optimistic resolution estimates compared with the 1% FDR-FSC approach. This effect is more pronounced at lower resolutions. As a result, the 1% FDR-FSC criterion displays stronger resolution differences with lower resolutions up to 8 Å in the vertical direction and higher resolutions up to 4.2 Å in the horizontal direction (Fig. 4a and 4c). Due to highly preferred orientations, such a result can be expected in this sample (Zi Tan et al., 2017). In contrast, the directional values derived by the 0.143 FSC threshold are not able to resolve the differences in directional resolution apparent in the HA map. For the 1.6 Å resolution map of apoferritin, clear directional resolution differences are not displayed due to high symmetry and homogenous particle orientations (Fig. 4b). Similarly, the asymmetric map of γ-secretase only exhibits minor directional resolution differences (Fig. 4c). We attribute the larger detected resolution differences by the FDR-FSC approach to the more sensitive resolution determination in analogy to the benefits observed for local resolution. In conclusion, FDR-FSC is suitable for local resolution determination as well as for directional resolution measurements as it estimates resolution at different noise levels more robustly in comparison with fixed threshold approaches.

## 3. Discussion

Resolution estimation is one of the essential tasks to assess the experimental quality and confidence for the interpretation of cryo-EM maps. Therefore, an automated procedure with least user input delivering robust results is desirable. Here, we present an approach of thresholding of FSC curves by non-parametric permutation sampling followed by FDR control. Previously, the statistical FDR approach was applied to a very different type of problem of cryo-EM map thresholding (Beckers et al., 2019). The FDR-FSC approach does not have any free parameters and only requires knowledge of the volume symmetry that is generally known from the 3D image reconstruction. This way, the approach enables sensitive resolution estimation without solvent flattening and thus eliminates the requirement of generating tight user-defined masks including the optional deconvolution process (Chen et al., 2013). Although mask-free approaches have been proposed

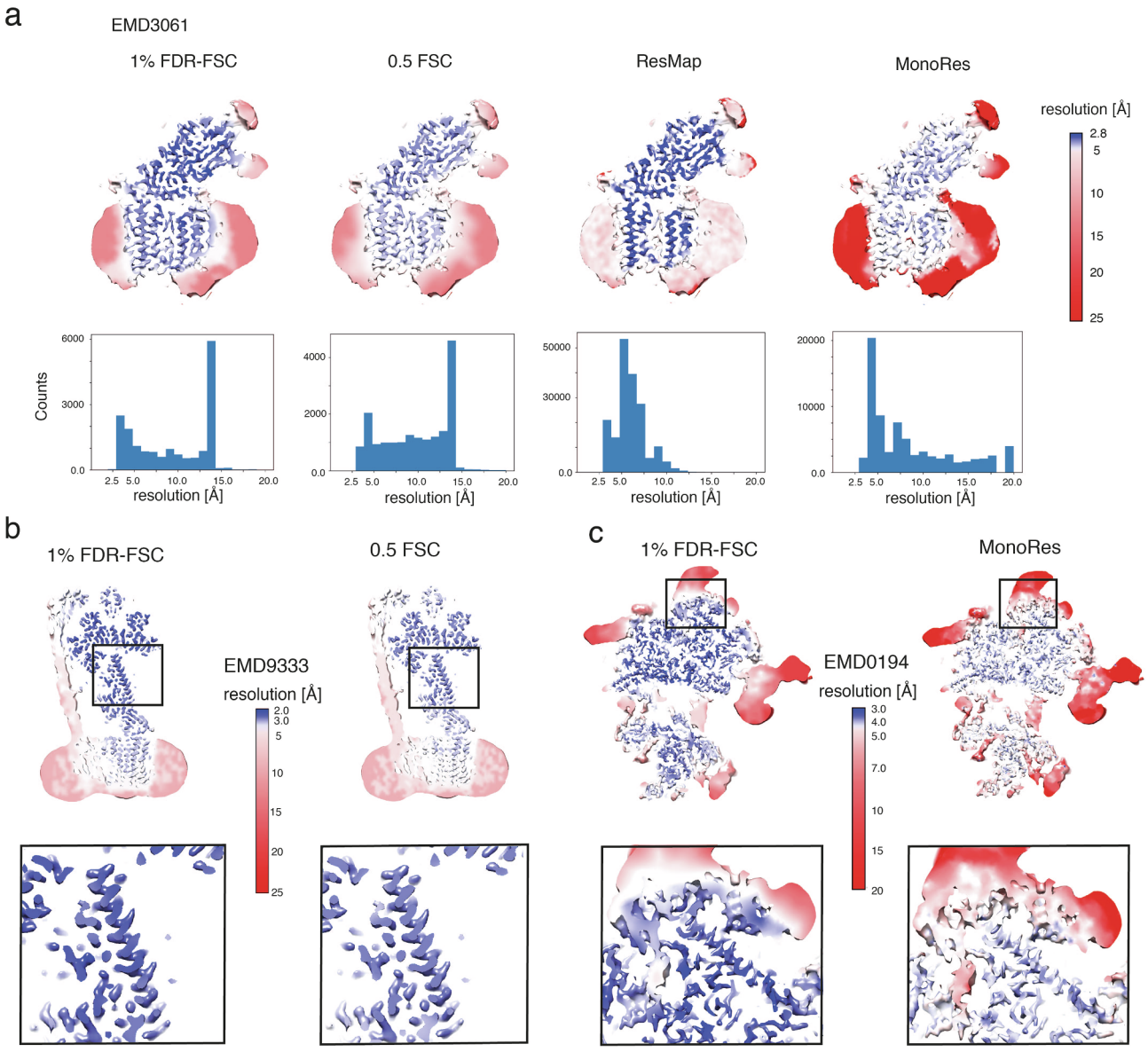


**Fig. 3.** Application of FDR-FSC to local resolution estimation. (a) Comparison of local resolutions estimated for the γ-secretase map (EMD3061) by 1% FDR-FSC, 0.5 FSC, ResMap and MonoRes grouped with corresponding resolution histograms below. (b) Surface mapped local resolutions determined using 1% FDR-FSC (left) of a 3.9 Å map of a bacterial ATP synthase (EMD9333) compared with 0.5 FSC mapping (right). (c) Surface mapped local resolutions determined using 1% FDR-FSC of a 3.8 Å map of a eukaryotic ribosome (EMD0194) compared with MonoRes mapping.

(Sindelar and Grigorieff, 2012), they require estimates of the expected molecular volume, which corresponds to information that may not be routinely available from the cryo-EM experiment in particular when heterogeneity is involved. Further application to local resolution estimation showed that judged by visual map features the FDR-FSC method captures locally high-resolution features in the core of protein structures as well as lower resolution in the periphery of a protein complex. Complementary to local resolutions, directional FSCs can also be analyzed by FDR-FSC and resulted in well-estimated resolutions in cases of anisotropic reconstructions. The algorithm is implemented in a Python program that takes a few seconds for small maps (< 200 pixels volume size) up to a few minutes for big maps (> 400 pixels volume size).

A conceptual advantage of the proposed FDR-FSC approach is that the method only assesses whether any significant correlations beyond random noise can be detected in a given resolution shell, which represents a paradigm shift to the conventional resolution determination of using the FSC as an estimator of the spectral SNR. In this way, the FDR-FSC approach is very sensitive to signal and less affected by the presence of solvent noise (see Fig. 2) and therefore does not require any solvent flattening. In addition, inference of statistically significant signal in the resolution shells only requires the distribution of random Fourier correlations determined by permutation and does not require the consideration of complicated correlations between signal and noise (van Heel and Schatz, 2017). The permutation sampling of FSCs per resolution shell has the advantage that signal dependencies have little influence on determining the principal FSC distributions per shell. Therefore, under the assumption that noise is statistically independent from signal, permutation testing is straightforward. The experimentally measured FSC values per shell, however, are affected by introduced dependencies between Fourier coefficients inside a single resolution shell, which in turn changes the noise distribution of the FSC. Common volume operations such as symmetry or masking within the reconstruction sphere introduce such dependencies, and as a result the number of degrees of freedom are notably reduced and the true FSC distribution needs to be compensated for by subsampling. We implemented this subsampling by the effective sample size to capture the

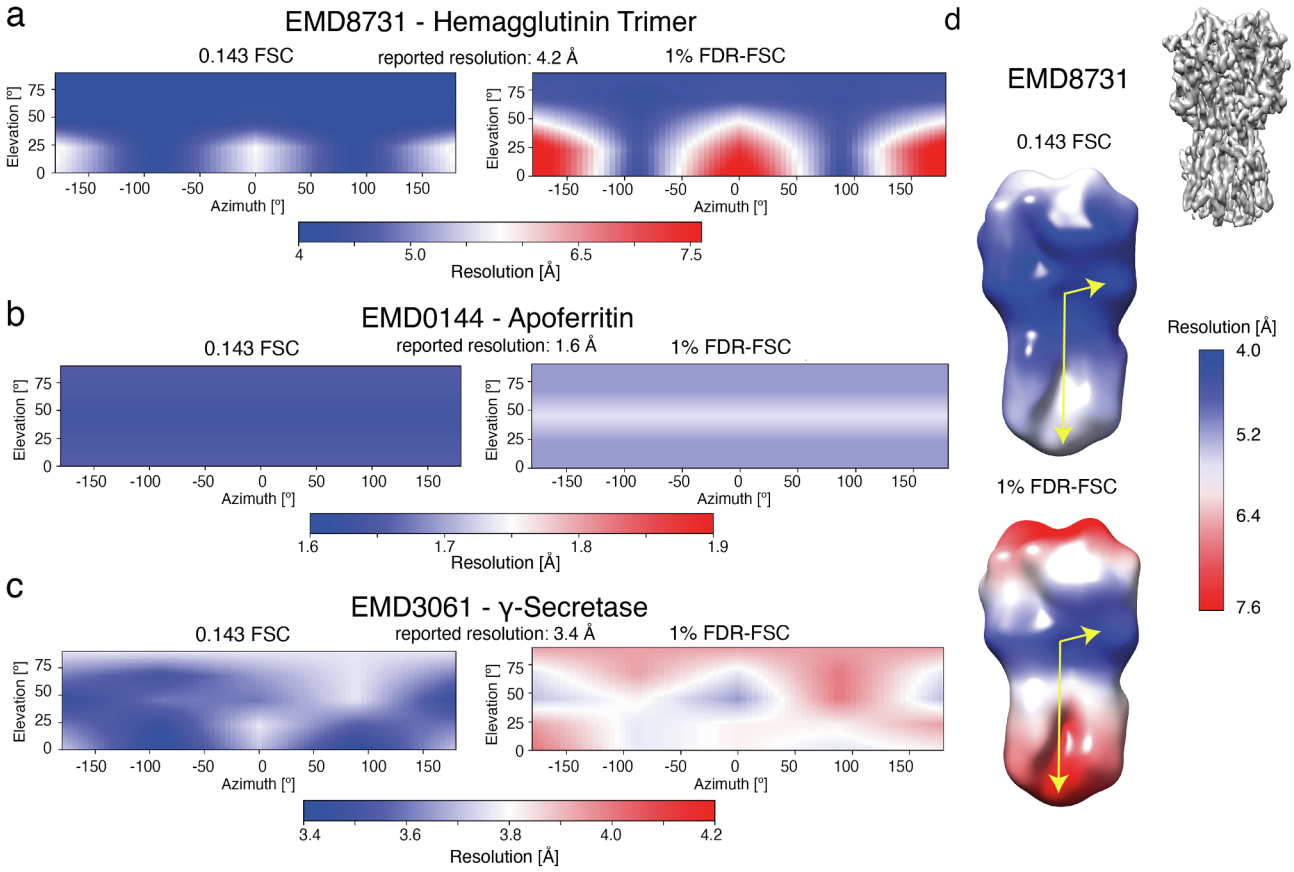


**Fig. 4.** Application of FDR-FSC to directional resolution estimation. Comparison of directional resolution plots of EMDB entries (a) hemagglutin trimer (EMD8731), (b) apoferritin (EMD0144) and (c) γ-secretase (EMD3061) for 0.143 FSC (left) and 1% FDR-FSC thresholds (right). Resolutions are shown in colors for the respective directions corresponding to angles of azimuth and elevation. (d) Directional resolutions mapped on the low-pass filtered surface of EMD8731. The resolution at each voxel specifies the resolution in the direction given by the vector from the center to the respective voxel (c.f. yellow arrows). Stronger directional resolution differences are reported using the 1% FDR-FSC measurement.

true FSC value distribution per shell in contrast to permutations if the nominal sample size was used. In addition, dependencies between half-maps are introduced by: mis-alignment of noisy particle images using masks, mis-alignment of noisy particle images with different signal components such as alternative conformations, interpolations inherent to the 3D reconstruction procedure and 3D alignment of determined half-maps. These contributions are currently not considered in our FSC permutation sampling as they are not easy to quantify and specific to the applied image processing pipeline. Commonly used image processing work-flows are working with generous and soft masks, separating different particle populations using classification, aligning particles by omitting high-resolution information or using independent half-set refinement in an attempt to minimize the introduction of such dependencies between Fourier coefficients of half-maps.

Once Fourier correlation distributions have been estimated, *p*-values of the measured FSC value for each resolution shell can be derived. Finally, FDR control of these *p*-values is followed by the thresholding step. Although our comparison with simple *p*-values shows only small differences in the determined resolution cutoffs, controlling the FDR adds to the robustness of the thresholding step. FDR control is a statistically well-established routine for identifying thresholds when multiple tests are performed simultaneously (Benjamini and Yekutieli, 2001). The FDR control procedure automatically adjusts the effective *p*-value threshold to the number of available resolution shells. The controlling procedure will have a stronger effect when evaluating very large box sizes in the presence of high noise levels and could, therefore, be more advantageous in challenging cases such as tomograms. Besides, it is not known that the FDR controlling procedure would introduce any drawbacks to the proposed permutation testing approach. An additional restraint to the thresholding step is imposed by limiting the resolution cutoff to the maximum spatial frequency of a large resolution band that shows continuous presence of significant signal. We want to stress, however, that the FDR-FSC approach, like any other FSC based approach, will only deliver reliable resolution estimates if half-maps with independent noise are provided (Cheng et al., 2015). For example, when wrong references are aligned against non-particle images (Mao et al., 2013) or incorrect symmetries imposed on the 3D reconstruction, the resulting half-maps will have correlated noise that can not be captured by the permutation approach. Moreover, excessive Fourier shell correlations arise from 3D structures of non-uniform orientation distribution, in which case directional resolution assessment should be consulted (see Fig. 4). Together, when independent half-maps from cryo-EM reconstructions are available, the FDR-FSC criterion provides robust resolution estimates for single-particle image reconstructions.

When we applied FDR-FSC to local resolution estimation, the procedure shows improved detection of molecular detail over the commonly used 0.5 FSC cutoff or alternative approaches. We attribute this performance to the high sensitivity of signal in the presence of solvent noise. Nevertheless, for these calculations, the window size remains a free parameter that affects the obtained values of local resolutions regardless of which thresholding procedure is being used (see Supplementary Fig. 10). This is a general property when local FSC calculations are computed and is independent from the respective threshold criterion as it is already elaborated in detail previously (Cardone et al., 2013). Irrespective of the window size, the here presented FDR-FSC method has advantages over 0.5 FSC, as it is less

sensitive to differences in noise levels in the center and periphery of the protein that can affect the resolution determination.

Given its sensitivity, robustness and minimal input requirements, it is conceivable that the proposed FDR-FSC approach can be used to re-evaluate many deposited EMDB structures when raw half-maps were available. In our evaluation, we found very good agreement between the reported resolution values, mostly based on the 0.143 FSC criterion including solvent flattening in the presence of a user-defined mask, and the FDR-FSC resolution values. This correlation shows that the overwhelming majority of depositors submits their resolution value with care in correspondence to the presented visual features such as β-strand separation and side-chain densities. For future EMDB depositions, the FDR-FSC method could be used as an additional resolution assessment in the context of other standard validation tools. Due to the increased statistical robustness, we see particular use for extending resolution determination into more challenging applications such as local and directional resolution where resolution reporting is often critical for interpretation but technically less standardized. In the current manuscript, we presented the FDR-FSC method with a focus on cryo-EM map evaluation. It has to be noted that the proposed framework is applicable to 2D Fourier ring correlations as used in super-resolution microscopy (Nieuwenhuizen et al., 2013). Due to the demonstrated ease of use and robustness, we anticipate a common applicability of the FDR-FSC method to image analysis and processing tools in and beyond the cryo-EM field.

## 4. Methods

### 4.1. Permutation sampling of Fourier shell correlation coefficients

We consider $X_{r_i}$ and $Y_{r_i}$ to be complex random variables and we denote the corresponding Fourier coefficients at the specific location $r_i$, $i = 1, \cdots, N$, in resolution shell $r$ of half-map 1 and 2, where $N \in \mathbb{N}$ is the number of Fourier coefficients in the respective shell. Due to Friedel symmetry, there are $n = N/2$ independent Fourier coefficients. Correlations between $X_{r_i}$ and $Y_{r_i}$ are introduced by common signal at position $r_i$, which we denote by $S(r_i) \in \mathbb{C}$ being identical in both half-maps. Thus, Fourier coefficients are usually modelled as

$$X_{r_i} = S(r_i) + N_X(r_i) \text{and} Y_{r_i} = S(r_i) + N_Y(r_i) \tag{3}$$

where $N_X(r_i)$ and $N_Y(r_i)$ are complex valued noise variables (Van Heel and Schatz, 2005). No assumptions about the distribution of the noise in resolution shell $r$ are made. Furthermore, we denote by $X_R = (X_{r_1}, \cdots, X_{r_n})$ and similarly $Y_R = (Y_{r_1}, \cdots, Y_{r_n})$ the complete set of Fourier coefficients in the resolution shell with spatial frequency $R$.

We design a statistical test that assesses whether there are significant Fourier shell correlations beyond noise levels. Mathematically, for two half-maps that are aligned in register we test whether the Fourier coefficients in the respective resolution shell of half-map 1 are correlated with the Fourier coefficients of half-map 2 according to their location $r_i$ in Fourier space. Expressed in statistical terms: the null hypothesis $H_0$ is that $X_{r_i}$ and $Y_{r_i}$ are uncorrelated, and the alternative hypothesis $H_1$ is that there are significant correlations. These correlations are typically measured by the Fourier shell correlation $FSC$, which is, using the Friedel symmetry inherent to Fourier transforms of real valued data, given as

$$FSC(X_R, Y_R) = \frac{\sum_{i=1}^{n} X_{r_i} Y_{r_i}^* + Y_{r_i} X_{r_i}^*}{\sqrt{\sum_{i=1}^{n} 2|X_{r_i}|^2 \sum_{i=1}^{n} 2|Y_{r_i}|^2}} \tag{4}$$

and it becomes clear that $FSC(X_R, Y_R)$ is a real valued random variable. Statistically, $FSC(X_R, Y_R)$ is the estimator of the true value of the Fourier shell correlation, which we denote from now on by FSC.

In order to test for significant correlations, we calculate a $p$-value as follows. The $p$-value $p$ of the observed value of the Fourier shell correlation, $fsc$, is then given as the probability that $FSC$ is bigger than $fsc$ given the null hypothesis $H_0$, i.e.

$$p = \mathbb{P}(FSC \geq fsc | H_0) \tag{5}$$

A permutation test is performed as follows. Under the null hypothesis (see above), the Fourier coefficients $X_{r_i}$ and $Y_{r_i}$ are uncorrelated for $i = 1, \cdots, n$. Newly paired samples of Fourier coefficients can thus be generated by permutation of $Y_R$. Therefore, the sampled distribution is the distribution of the $FSC$ under the null hypothesis, i.e. the noise distribution of the $FSC$. For a more detailed discussion regarding permutation of correlation coefficients we refer to (DiCiccio and Romano, 2017). From each generated sample, we then calculate the Fourier shell correlation, which results in a sample of $FSC$ under the null hypothesis. Denoting $S_n$ as the set of all permutations $\pi$ of $\{1, \cdots, n\}$, where $n$ is the sample size as above, we calculate the $p$-value $p$ by

$$p = \frac{1}{n!} \sum_{\pi \in S_n} \mathbb{I}\{FSC(X_R, Y_{\pi(R)}) \geq fsc\} \tag{6}$$

where $\mathbb{I}\{\}$ denotes the indicator function, i.e.

$$\mathbb{I}\{FSC(X_R, Y_{\pi(R)}) \geq fsc\} = \begin{cases} 1 & if \quad FSC(X_R, Y_{\pi(R)}) \geq fsc \\ 0 & if \quad FSC(X_R, Y_{\pi(R)}) < fsc \end{cases} \tag{7}$$

As the number of possible permutations grows rapidly with the sample size, technically often just a random subset $H \subset S_n$ of all $n!$ permutations is used. Thus, $p$ is replaced by its Monte-Carlo estimator $\hat{p}_{MC}$, given as

$$\hat{p}_{MC} = \frac{1}{|H|} \sum_{\pi \in H} \mathbb{I}\{FSC(X_R, Y_{\pi(R)}) \geq fsc\}, \tag{8}$$

where $|H|$ is the cardinality of the set $H$, i.e. the number of permutations selected for the Monte-Carlo estimation. It is important to note that the distribution of the $FSC$ under the null hypothesis, which we estimate by the permutation approach, is not necessarily the distribution of the $FSC$ in the absence of any signal. As we are permuting in the presence of possible signal, which adds additional variation, the permutation distribution of the $FSC$ will have larger tail probabilities than the $FSC$ distribution without any signal. Consequently, this could give rise to more conservative resolution estimates. However, simulations of permutation in the presence of signal at high and low signal-to-noise ratios showed that this effect seems to have less practical relevance (Supplementary Fig. 2). Moreover, the amount of signal in the most critical resolution shells close to the true resolution of the structure is low and will thus have limited influence on the actual distribution. As a compromise between computational efficiency and statistical accuracy, we restrict the number of permutations to a maximum of 1000. Permutations are only performed for resolution shells with an effective sample size > 10, which allows for more than 1000 permutations. As the initial two resolution shells can have effective sample sizes smaller than 10 Fourier coefficients, they will not be sampled by permutation. In cases, when the following sampled resolution shells already drop below the significance level, the conservative 0.9 FSC threshold is used as the resolution cutoff of the volume.

### 4.2. Multiple testing correction

As we are testing the complete number of resolution shells separately, there is a risk of obtaining positive tests accidentally simply because of the larger number of involved tests. This is commonly referred to as the multiple testing problem and requires the significance levels of the individual tests to be adjusted. With respect to FSC thresholding such a correction is important, as the number of tested resolution shells commonly varies from as low as 20 for windows of local resolution estimation to several hundred for complete maps. As a workaround to this problem, approaches that put individual significance levels in the context of a global significance level have been proposed (Benjamini and Hochberg, 1995). In particular, the false

discovery rate (FDR) is a widely used approach for addressing the multiple testing problem. The FDR corresponds to the fraction of false positive tests when all positives are considered. In the here proposed approach, we apply the FDR by the Benjamini-Yekutieli procedure (Benjamini and Yekutieli, 2001). This implementation is known to control the FDR under arbitrary dependencies of the *p*-values. A more detailed treatment of the FDR control process including proofs of correctness is beyond the scope this manuscript and we refer the interested reader to the original reports (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

### 4.3. Effects of symmetrization and masking

As mentioned in the Results section, the reconstruction sphere or masking and symmetrization of the volumes lead to dependencies between Fourier coefficients and thus reduces the number of degrees of freedom. Due to dependencies, the sample behaves statistically like a sample of independent realizations of smaller size. In the field of Monte Carlo simulations this smaller sample size is typically referred to as the effective sample size $n_{eff}$. The effective sample size of imposed symmetry during reconstruction can be given as

$$n_{eff} = \frac{n}{n_{as}} \tag{9}$$

where $n$ is the number of Fourier coefficients in the respective resolution shell and $n_{as}$ is the number asymmetric volume units at the given symmetry (Supplementary Fig. 4). Effective sample sizes are incorporated in the permutation framework by subsampling of Fourier coefficients in the respective shells. Effects of the reconstruction sphere or masking on the effective sample size are more complicated and depend on the specific shape and volume of the mask. When the nominal sample size of the complete resolution shell is given by $n$, we estimate the effective sample size $n_{eff} \approx \alpha * n$ by finding the factor $\alpha \in (0, 1]$ through minimization of the mean Kolmogorov-Smirnov distance $D_\alpha$ over the resolution shells. The two-sample Kolmogorov-Smirnov distance (Massey, 1951), which we denote by $D_{\alpha,r}$, is a measure of similarity of two probability distributions. It is defined as the maximum difference between the two empirical cumulative distribution functions (ECDF). The ECDF itself specifies the fraction of samples that are smaller or equal to a specified value $x$, i.e. it estimates the cumulative probability of observing a sample smaller than $x$. In other words, the Kolmogorov-Smirnov distance specifies the maximum difference between these cumulative probabilities of two distributions that can be observed for any $x$.

In our case for resolution shell $r$ and with the effective sample size correction factor $\alpha$ from above, the Kolmogorov-Smirnov distance is then given as:

$$D_{\alpha,r} = \sup_{x\in[-1,1]} |F_{1,\alpha,r}(x) - F_{2,sim,r}(x)|, \tag{10}$$

where in our setting $F_{1,\alpha,r}(x)$ denotes the ECDF of the permutation approach of resolution shell $r$, which is estimated using an effective sample size $N_{eff} \approx \alpha * n$, and $F_{2,sim}(x)$ an estimate of the true cumulative distribution function of resolution shell $r$ in the presence of the respective mask. $F_{2,sim}(x)$ can be obtained by simulation of two noise maps, placed in the reconstruction sphere, and subsequent FSC calculation, followed by another simulation of such two maps with subsequent FSC calculation and so forth. This allows sampling of the true distribution of the FSC when there are no correlations beyond random noise. $\bar{D}_\alpha$ is then calculated as the mean of the Kolmogorov-Smirnov distance $D_{\alpha,r}$ over all the resolution shells $r$, i.e.

$$\bar{D}_\alpha = \frac{1}{m}\sum_{r=1}^{m} D_{\alpha,r}, \tag{11}$$

where $m$ is the number of resolution shells. An estimate $\hat{\alpha}$ for $\alpha$ is then given by

$$\hat{\alpha} = \arg\min_{\alpha\in(0,1]} \bar{D}_\alpha. \tag{12}$$

In the presented algorithm, we apply a soft spherical mask of volume dimension diameter including a Gaussian falloff of two voxel standard deviation for global resolution estimation by default, because most half-maps already reside within a reconstruction sphere applied by the reconstruction software. The resulting effect can be corrected by an effective sample size of $n_{eff} \approx 0.7n$, i.e. $D_\alpha$ was minimized for $\alpha \approx 0.7$ (Supplementary Fig. 5a), which allows accurate calculation of tail probabilities for various box sizes (Supplementary Fig. 5, Supplementary Fig. 6). Application of windowing functions, as used for local resolution estimation (see below), also leads to dependencies similar to masking, which have to be taken into account via effective sample sizes. In a similar way as for the soft circular mask, we found that a Hann-window leads to an effective sample size $n_{eff} \approx 0.23n$ (Supplementary Fig. 5b, Supplementary Fig. 7).

### 4.4. Local resolution estimation

Local resolutions are estimated by a moving window across both half maps and subsequent calculation of local FSC thresholds (Cardone et al., 2013). In order to account for high-resolution artifacts introduced through spectral leakage, a Hann-function is used as a windowing function. The masking effects, which are introduced by the Hann window, are considered via effective sample sizes as described above. Moreover, effects of symmetry do not have any influence on the effective sample sizes when the size of the sliding windows is smaller than the size of the asymmetric volume unit in the map. In order to speed up the calculations, a step-size option is implemented allowing movement of the sliding window of more than a single voxel. Moreover, in order to avoid repeating permutations of the same map, the permutations are only performed at 10 random locations of the sliding window. The resulting samples of the null distribution are merged and subsequently used for *p*-value calculation at all locations. Presented experiments were performed using a window size of 25 pixels.

### 4.5. Directional resolution estimation

The implementation of the directional resolution estimation follows Lyumkis and colleagues (Zi Tan et al., 2017). For each direction, the FSC is calculated by taking those voxels from each half-map that are included by a specified angle at the respective direction. This results in rotating an inverse cone over one half of the 3D Fourier transforms, thereby accounting for the Friedel symmetry, and calculating the FSC only from samples inside the inverse cone. In analogy to (Zi Tan et al., 2017), an angle of 20° was used for the presented experiments providing a good compromise between the number of Fourier coefficients per shell and the preservation of local directionality. Moreover, in analogy to local resolutions, locality of the directional resolution does not require the correction of symmetry effects. As implemented in the case of local resolution estimation, in order to accelerate the algorithm, the resolutions are only calculated for a limited set of directions and results are interpolated in order to avoid repetitive FSC calculations of overlapping cones.

### 4.6. Software

The procedures are implemented together with the previously published Confidence Map tools (Beckers et al., 2019) in a GUI based software named SPoC – Statistical Processing of cryo-EM maps (Supplementary Fig. 11). Code is written in Python3 based on NumPy (Van Der Walt et al., 2011), matplotlib (Hunter, 2007), SciPy (Oliphant, 2007), mrcfile (Burnley et al., 2017) and parallelized by the Python multiprocessing module. The software is available at https://github.com/MaximilianBeckers/SPOC. Figures were prepared with UCSF Chimera (Pettersen et al., 2004).

## Author contributions

M.B. and C.S. designed the research. M.B. developed and implemented the algorithm and conducted the experiments. M.B. and C.S. wrote the manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We are grateful to Thomas Hoffmann and Jurij Pecar (IT Services) for maintenance of the high-performance computing at EMBL.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2020.107579.

## References

Bai, X.C., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S.H.W., Shi, Y., 2015. Nature 525, 212–217.
Banterle, N., Bui, K.H., Lemke, E.A., Beck, M., 2013. J. Struct. Biol. 183, 363–367.
Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L.S., Subramaniam, S., 2015. Science (80-.) 348, 1147–1151.
Beckers, M., Jakobi, A.J., Sachse, C., 2019. IUCrJ 6, 18–33.
Benjamini, Y., Hochberg, Y., 1995. J. R. Stat. Soc. B 57, 289–300.
Benjamini, Y., Yekutieli, D., 2001. Ann. Stat. 29, 1165–1188.
Brooks, S., Gelman, A., Jones, G.L., Meng, X.L., 2011. Handbook of Markov Chain Monte Carlo. Chapman & Hall.
Burnley, T., Palmer, C.M., Winn, M., 2017. Acta Crystallogr. Sect. D Struct. Biol. 73, 469–477.
Cardone, G., Heymann, J.B., Steven, A.C., 2013. J. Struct. Biol. 184, 226–236.
Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H.W., Henderson, R., 2013. Ultramicroscopy 135, 24–35.
Cheng, Y., Grigorieff, N., Penczek, P.A., Walz, T., 2015. Cell 161, 438–449.
DiCiccio, C.J., Romano, J.P., 2017. J. Am. Stat. Assoc. 112, 1211–1220.
Guo, H., Suzuki, T., Rubinstein, J.L., 2019. Elife 8, e43128.
Harauz, G., Van Heel, M., 1986. Optik (Stuttg) 73, 146–156.
Heel, M. van, Schatz, M., 2017. BioRxiv.
Van Heel, M., Schatz, M., 2005. J. Struct. Biol. 151, 250–262.
Hunter, J.D., 2007. Comput. Sci. Eng. 9, 90–95.
Juszkiewicz, S., Chandrasekaran, V., Lin, Z., Kraatz, S., Ramakrishnan, V., Hegde, R.S., 2018. Mol. Cell 72, 469–481.
Kucukelbir, A., Sigworth, F.J., Tagare, H.D., 2014. Nat. Methods 11, 63–65.
Kühlbrandt, W., Amunts, A., Liao, M., Cao, E., Julius, D., Cheng, Y., Allegretti, M., Li, X., Ban, N., Wimberly, B.T., Faruqi, A.R., Henderson, R., Daum, B., Walter, A., Horst, A., Osiewacz, H.D., Kühlbrandt, W., Schur, F.K., Scheres, S.H., 2014. Science 343, 1443–1444.
Lehmann, E., Romano, J., 2005. Testing Statistical Hypotheses.
Mao, Y., Wang, L., Gu, C., Herschhorn, A., Désormeaux, A., Finzi, A., Xiang, S.H., Sodroski, J.G., 2013. Proc. Natl. Acad. Sci. U.S.A. 110, 12438–12443.
Massey, F.J., 1951. J. Am. Stat. Assoc. 46, 68–78.
Nieuwenhuizen, R.P.J., Lidke, K.A., Bates, M., Puig, D.L., Grünwald, D., Stallinga, S., Rieger, B., 2013. Nat. Methods 10, 557–562.
Oliphant, T.E., 2007. Comput. Sci. Eng. 9, 10–20.
Orlova, E.V., Dube, P., Harris, J.R., Beckman, E., Zemlin, F., Markl, J., Van Heel, M., 1997. J. Mol. Biol. 271, 417–437.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. J. Comput. Chem. 25, 1605–1612.
Rosenthal, P.B., Henderson, R., 2003. J. Mol. Biol. 333, 721–745.
Saxton, W.O., Baumeister, W., 1982. J. Microsc. 127, 127–138.
Sindelar, C.V., Grigorieff, N., 2012. J. Struct. Biol. 180, 26–38.
Spahn, C.M.T., Jan, E., Mulder, A., Grassucci, R.A., Sarnow, P., Frank, J., 2004. Cell 118, 465–475.
Vilas, J.L., Gómez-Blanco, J., Conesa, P., Melero, R., Miguel de la Rosa-Trevín, J., Otón, J., Cuenca, J., Marabini, R., Carazo, J.M., Vargas, J., Sorzano, C.O.S., 2018. Structure. 26, 337–344.
Van Der Walt, S., Colbert, S.C., Varoquaux, G., 2011. Comput. Sci. Eng. 13, 22–30.
Weis, F., Beckers, M., Hocht, I., Sachse, C., 2019. EMBO Rep.
Zi Tan, Y., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B., Lyumkis, D., 2017. Nat. Methods 14, 793–796.
Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E., Scheres, S.H.W., 2018. Elife. e42166.