Journal of Structural Biology 204 (2018) 527–542

Contents lists available at ScienceDirect

Journal of Structural Biology

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

This article is part of the Special Issue on the 2016 CryoEM Challenges

# Map challenge: Analysis using a pair comparison method based on Fourier shell correlation

R. Marabini[a,*], M. Kazemi[b], C.O.S. Sorzano[c], J.M. Carazo[c]

[a] Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain
[b] Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, The University of Melbourne, VIC 3010, Australia
[c] Biocomputing Unit, National Center for Biotechnology (CSIC), C/ Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain

ARTICLE INFO

*Keywords:*
Structural biology
Electron microscopy
3D reconstruction
High-resolution
Benchmarking
Challenge

ABSTRACT

This document presents the analysis performed over the Map Challenge dataset using a new algorithm which we refer to as *Pair Comparison Method*. The new algorithm, which is described in detail in the text, is able to sort reconstructions based on a figure of merit and assigns a level of significance to the sorting. That is, it shows how likely the sorting is due to chance or if it reflects real differences.

## 1. Introduction

As image processing in 3D electron microscopy (3DEM) advances, it becomes more difficult to compare the performance of the different algorithms just by looking at their specifications. Benchmarks are a valuable resource for measuring the performance of different image processing pipelines, but unfortunately good datasets and standardized procedures do not exist for most of the problems in 3DEM. The Map Challenge is a step in the right direction, since provides a very interesting collection of datasets to be reconstructed. Unfortunately, interpretation of benchmarking data is extraordinarily difficult, and 3DEM researchers are far from offering a clear and standard way to assess the performance of the different software packages.

In this work we describe a new algorithm able to sort the 3D maps using a quality criteria based on a figure of merit (FOM) and provide a significance value for the claim that two maps are different. In plain English, *significant* means important, while in Statistics it means not due to chance. The new algorithm will provide two values for each comparison. The first value tells us if the result is highly significant, that is, if it is very likely to be true; while the second gives an idea of how important the difference is between the 3D maps, since highly significant differences are not always important.

The organization of this document is as follows. First, we present a brief description of the new algorithm, then the results of applying the algorithm to the 3D maps uploaded to the Map Challenge is presented. Finally, a full description of the algorithm including all the tests performed to validate it is shown in Appendix A.

## 2. Methods

In this section we first summarize the *pair comparison method*, and then describe the preprocessing applied to the data before analyzing it with the sorting algorithm.

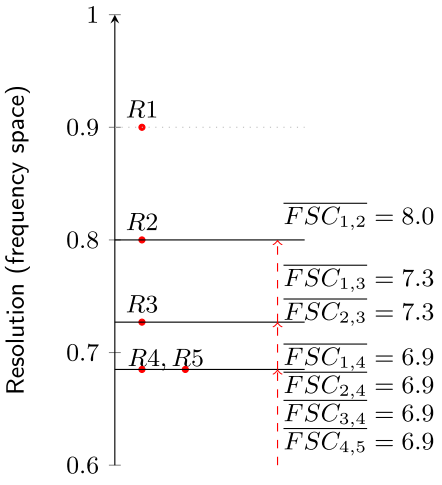
### 2.1. Brief description of the sorting algorithm

The main idea behind the *pair comparison method* is to create, for a given experimental dataset, all possible pairs of reconstructions and estimate the FSC between the members of each pair. In the absence of systematic bias, the distribution of all these FSCs should reveal the reconstruction for which its FSC are better than the rest. Oversimplifying, the pair comparison method can be summarized by the following steps:

- Compute a *weighted integrated FSC* between each pair of reconstructions (all possible pairs should be computed). We will denote this magnitude as $\overline{FSC_{i,j}} = \int_{\frac{1}{200\,\text{Å}}}^{\nu_{max}} FSC_{i,j}(\nu)\nu d\nu$ where $i$ and $j$ refer to the $i$ and $j$ reconstructions respectively, $\nu$ is the frequency in Fourier space and $\nu_{max}$ is the higher resolution reported for the dataset under analysis. Therefore, $\overline{FSC_{i,j}}$ is related with the area below the *FSC* curve calculate for maps $i$ and $j$.
- Group the $\overline{FSC_{i,j}}$ by $i$, that is, the first 3D map involved in the FSC

* Corresponding author.
*E-mail address:* roberto.marabini@uam.es (R. Marabini).

https://doi.org/10.1016/j.jsb.2018.09.009
Received 19 March 2018; Received in revised form 13 September 2018; Accepted 20 September 2018
Available online 28 September 2018
1047-8477/ 




| | R1 | R2 | R3 | R4 | R5 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|
| **R1** | | 8.0 | 7.3 | 6.9 | 6.9 | $\overline{FSC_1} = 29.1$ |
| **R2** | 8.0 | | 7.3 | 6.9 | 6.9 | $\overline{FSC_2} = 29.1$ |
| **R3** | 7.3 | 7.3 | | 6.9 | 6.9 | $\overline{FSC_3} = 28.4$ |
| **R4** | 6.9 | 6.9 | 6.9 | | 6.9 | $\overline{FSC_4} = 27.6$ |
| **R5** | 6.9 | 6.9 | 6.9 | 6.9 | | $\overline{FSC_5} = 27.6$ |

**Fig. 1.** Resolution, $\overline{FSC_{i,j}}$ and $\overline{FSC_i}$ values for the five reconstructions **R1**, **R2**, etc.

estimation

- For each group, sum. We will denote the $\overline{FSC_{i,j}}$ sum for a given $i$ as $\overline{FSC_i} = \sum_{j,i\neq j} \overline{FSC_{i,j}}$.

The sum will be used to sort the reconstructions. Then we perform a Wilcoxon signed-rank test (Wilcoxon, 1945) to verify if the mean ranks of the $\overline{FSC_{i,j}}$ values observed for any two maps $i$ and $j$ are different. A more precise description of the *pair comparison method* is available in Appendix A.

A thought experiment will help to clarify how the algorithm works. Let us assume that we have computed five 3D maps reconstructed from the same dataset using different algorithms. We will refer to these reconstructions as **R1**, **R2**, ···, **R5**. Let us assume that the ideal reconstruction maximum frequency is 1 and that (in frequency space) **R1** is equal to this ideal reconstruction up to frequency 0.9, **R2** is equal to it up to frequency 0.8, etc (see Fig. 1 for the resolution of the other reconstructions). If we compute $\overline{FSC_{i,j}}$, the value will be proportional to the resolution of the worst of the two reconstructions. Possible values for this magnitude are shown in Fig. 1. After grouping by $i$ (the first reconstruction involved in the computation of $\overline{FSC_{i,j}}$), we obtain the $\overline{FSC_i}$ values shown in the last column of the table in Fig. 1. We see that indeed, using $\overline{FSC_i}$, we can sort the reconstructions and place the best ones at the top of the sorting and the worst ones at the bottom. Note that we do not claim here that the best reconstruction will be the first one in the sorted list, but only that good reconstructions will be at the top and bad ones at the bottom.

The Wilcoxon signed-rank test will help to determine if the difference between any two volumes is significant. For instance, we could compare **R2** with **R3** by checking if the mean rank of the **R2** sequence (8.0, 7.3, 6.9, 6.9; see table in Fig. 1) is significantly different from the **R3** sequence (7.3, 7.3, 6.9, 6.9).

### 2.2. Data preprocessing

In order to compute the statistics described in the previous section the 3D maps need to be registered, that is, they should be aligned and sampled at the same sampling rate. We describe here the exact preprocessing workflow followed:

For each specimen:

- Create a reference volume by randomly rotating the first uploaded 3D map with sampling rate equal to the input data.
- Align each 3D map with respect to the reference volume using Chimera
  - –Load reference and problem volume: `chimera Ref.mrc emcd $NUM_$SPECIMEN_unfiltered.mrc`
  - –Place origin of coordinates in 3D map center: `viewer - > coordinates - > center`
  - –Manually align the different 3D maps with the reference.
  - –Refine alignment with command `viewer - > tools - > fit in map` (3D maps with $CC < 0.9$ are dropped)
  - –Interpolate aligned map in the reference system of coordinates with the reference sampling rate: `vop resample #1 onGrid #0`
  - –Save interpolated 3D map:`viewer - > file - > save_as`
- Apply a soft spherical mask to the 3D maps. We applied this masking because several unfiltered 3D maps presented spherical masks introduced by the reconstruction workflow.
- Apply the new comparison method.
  - – Compute magnitudes $\overline{FSC_{i,j}}$ and $\overline{FSC_i}$ (fully described in Appendix A)

Once $\overline{FSC_{i,j}}$ and $\overline{FSC_i}$ have been computed:

- Sort the 3D maps based on $\overline{FSC_i}$
- Test the null hypothesis "two 3D maps can be distinguished" based on $\overline{FSC_{i,j}}$. We will assume that two volumes can be distinguished if the P-value resulting from applying the Wilcoxon signed-rank test to them is smaller than 0.05.

## 3. Results

For each one of the datasets provided by the Map Challenge, two tables and a dendrogram are shown. The first table presents the 3D maps sorted by the feature $\overline{FSC_i}$ (see for example Table 1). The higher the value of $\overline{FSC_i}$, the better the reconstruction. The second table highlights reconstruction pairs where the hypothesis "that the two maps in the pair are different" cannot be accepted with a P-value smaller than 0.05 (see for example Table 2). The election of P-value = 0.05 as threshold is a quite standard practice but nevertheless arbitrary. We recall here that a P-value is the probability of making the wrong decision when the null hypothesis is true. In this way, a P-value = 0.05 does not mean that 5% of the times we are wrong but that 5% of the times we are wrong because we think that two volumes are similar when they are not. In order to compute the total number of wrong decisions we must add to these false positives the number of times we are wrong because we think that two volumes are not similar but they are.

Cells in the above mentioned second table contain the P-values obtained from comparing the two 3D maps associated to the corresponding row and column using the Wilcoxon signed-rank test. P-values greater or equal to 0.05 will be outlined using a red color and they mark those pairs formed by two reconstructions which are similar. In this way, if the reconstruction **R1** is sorted above the reconstruction **R2** in the first table, we can say that **R1** is better than **R2** if the P-value assigned to the pair (**R1**,**R2**) is smaller that 0.05 in the second table. The magnitudes $\overline{FSC_i}$ and $\overline{FSC_{i,j}}$ used for sorting and for computing the P-values are available at the end of this document (see Appendix C).

The second table allows us to know if one particular reconstruction is truly better than another but, can we go further? Is it possible to cluster the reconstructions using the information available in this table? The second table can be understood as a similarity matrix, that is, the larger the value assigned to the pair of reconstructions $(i, j)$, the more probable is that both reconstructions are equivalent. We have used the

implementation of hierarchical clustering provided in the Python package SciPy to build clusters from similarity matrices and make dendrogram plots as the one shown in Fig. 3. In these dendrograms, the vertical axis represents the distance or dissimilarity between clusters (1−P-value). The horizontal axis represents the different reconstructions which are identified by their ID together with the position. In this way, **123–3** stands for the 3D map with ID **123** which is the third best reconstruction. Note that reconstructions close in the horizontal axis may not be similar and it is the vertical length of the path that joins two 3D maps (or clusters) the real distance.

Except where noted, for each dataset we present a dendrogram with four labels in the *x*-axis. The first line of the label has been already explained. The second, third and fourth lines refers to the 3D refinement algorithm, dose weighting scheme and movie alignment algorithm used to create that particular 3D map respectively. The dose weighting label has the structure *X*, *Y*-*Z*; if a weighting algorithm has been applied $X = T$ otherwise $X = F$, $Y$ and $Z$ refers to the first and last movie frame used to produce the particles. All labels belonging to the best cluster use a green font while labels belonging to the worst cluster use a red font.

It should be noted here that the Map Challenge original deadline was April 15th 2016. After this deadline, the data were analyzed by each of the assesors and the results discussed in a two day Workshop opened for all challenge participants held on October 6th 2017. Subsequent to this workshop, a second submission period was opened in which 3D maps could be revised and resubmitted. We have applied the algorithm to both the old and the revised versions of each dataset. Since we have not found statistically significant variations, we will show the data corresponding to the first submission period. Nevertheless, when presenting the data collected for each specimen, we will comment on the small variations introduced by the revised 3D maps. 3Dmaps **emcd130** and **emcd134** had the wrong hand, consequently, we have flipped them. Finally, we are using the second submission of **emcd146** because we could not align the first submission with the reference volume with a cross correlation coefficient greater than 0.9.

To clarify the meaning of the different tables and dendrograms instead of just presenting the data we will explain and interpret it for the first dataset. The other datasets will be further discussed in the Discussion section.

### 3.1. First dataset: GroEL in silico

As mentioned before, Table 1 lists the different reconstructions sorted by $\overline{FSC_i}$. The first relevant question is if this sorting is significant. That is, given two consecutive reconstructions, for example **emcd143** and **emcd132**, is the first one better than the second one or are they just equivalent. The answer to this question is in Table 2. This table shows the P-value resulting from applying the Wilcoxon test and, as mentioned before, we assume that two maps are different if the P-value is smaller than 0.05. Therefore, the answer to the question, "are reconstructions **emcd143** and **emcd132** distinguishable" is No. Table 2 allows us to answer this question for any pair of reconstructions.

The GroEL *in silico* dendrogram is shown in Fig. 2. Since GroEL *in silico* data have been generated *in silico* without simulating beam

**Table 1**
Sorting based on $\overline{FSC_i}$ for specimen GroEL *in silico*. The main two clusters are surrounded by rectangles.

| $\overline{FSC_i}$ | 3Dmap |
|---|---|
| 47.65 | emcd143 |
| 46.1 | emcd132 |
| 45.75 | emcd165 |
| 43.35 | emcd169 |
| 42.6 | emcd104 |
| 42.5 | emcd120 |
| 36.8 | emcd168 |
| 35.65 | emcd153 |
| 35 | emcd158 |

**Table 2**
P-values resulting of comparing all 3Dmap pairs for specimen GroEL *in silico*.

| | emcd143 | emcd132 | emcd165 | emcd169 | emcd104 | emcd120 | emcd168 | emcd153 | emcd158 |
|---|---|---|---|---|---|---|---|---|---|
| **emcd143** | 0. | 0.4 | 0.01 | 0.09 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| **emcd132** | 0.4 | | 0.3 | 0.1 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| **emcd165** | 0.01 | 0.3 | | 0.6 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| **emcd169** | 0.09 | 0.1 | 0.6 | | 0.6 | 1 | 0.01 | 0.01 | 0.01 |
| **emcd104** | 0.01 | 0.01 | 0.01 | 0.6 | | 0.2 | 0.02 | 0.01 | 0.01 |
| **emcd120** | 0.01 | 0.01 | 0.01 | 1 | 0.2 | | 0.02 | 0.01 | 0.01 |
| **emcd168** | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | | 0.4 | 0.01 |
| **emcd153** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.4 | | 0.6 |
| **emcd158** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.6 | |

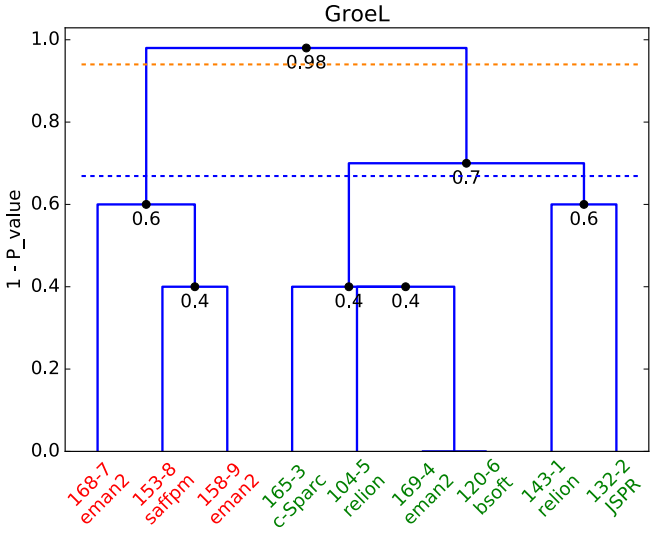



**Fig. 2.** GroeL hierarchical clustering rendered as a dendrogram. X-labels show: 3D map ID and ranking position (first line), software used for 3D angular refinement (second line). Non standard acronyms used in the X-axis labels are listed in B. (CI = 0.73.). Orange dashed line shows the recommended threshold for clustering, blue dashed line shows an alternative, non recommended, threshold that produces three clusters.

induced movement or radiation damage, the labels related to movie alignment and dose weighting algorithms are not present. Looking at the dendrogram in Fig. 2, we can see two clusters clearly different. The first cluster is formed by reconstructions with IDs {**168**, **153**, **158**} and the second by the rest of the reconstructions. This division can also be easily extracted from Table 2, because it is possible to draw two lines (orange dashed lines in the figure) that create two sets of reconstructions with the following property: all members of the first set are distinguishable from members of the second set.

Coming back to the dendrogram, we may lower the threshold and cluster our dataset in three classes. Note that this is equivalent to increasing the P-value from 0.05 (orange dashed line) to more than 0.3 (blue dashed line) and may produce claims that are not correct. Finally, it is worth mentioning that dendrograms are an easy way to visualize hierarchical classifications and cluster data, but contain less information than the similarity matrix used to create them. The extent to which a dendrogram represents a similarity matrix can be measured by the cophenetic index (CI). The closer this value is to 1, the more reliable is the dendrogram. The value of this coefficient can be seen in the caption of all dendrograms.

### 3.1.1. Revised maps

There is a revised version of map **emcd158**. Recalculating the sorting table with the revised volume produced a change of order between the maps **emcd165**, **emcd169**, **emcd104** and **emcd120** which is irrelevant (see dendrogram in Fig. 2) since these four volumes are very close in terms of similarity.

We now proceed to show the results obtained with the rest of the of specimens.

## 3.2. T20S proteasome

Tables 3 and 4, and Fig. 3 summarize the results for T20S *proteasome*. Note: since unfiltered 3Dmaps **emcd130** and **emcd131** are identical we have ignored **emcd131**.

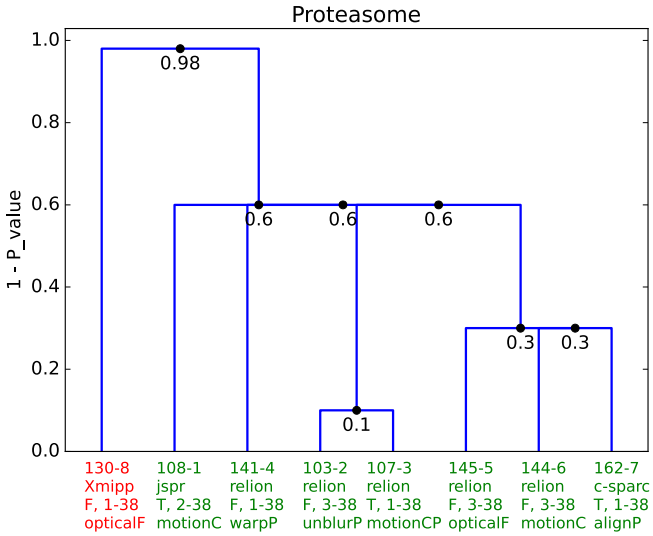



**Fig. 3.** *Proteasome* hierarchical clustering rendered as a dendrogram. X-labels show: 3D map ID and ranking position (first line), software used for 3D angular refinement (second line), if dose weighting has been applied and which frames has been used to create the particle images (third line) and software used for movie alignment (fourth line). Non standard acronyms used in the X-axis labels are listed in B. (.CI = 0.73.).

### 3.2.1. Revised maps

There is a revised version for maps **emcd103** and **emcd130**. Sorting table recalculation using the revised maps produced a change of order between **emcd108**, **emcd103** and **emcd107**, as well as maps **emcd141**, **emcd145** and **emcd144**. This new sorting is equivalent to the old one given by Table 3.

## 3.3. Apo-Ferritin

Tables 5 and 6, and Fig. 4 summarize the results for Apo-Ferritin.

**Table 3**
Sorting based on $\overline{FSC_i}$ for specimen T20S *Proteasome*. The main two clusters are surrounded by rectangles.

| $\overline{FSC_i}$ | 3Dmap |
|---|---|
| 114.94 | emcd108 |
| 114.72 | emcd103 |
| 114.55 | emcd107 |
| 112.96 | emcd141 |
| 112.35 | emcd145 |
| 111.26 | emcd144 |
| 106.49 | emcd162 |
| 101.79 | emcd130 |

**Table 5**
Sorting based on $\overline{FSC_i}$ for specimen Apo-Ferritin. The main three clusters are surrounded by rectangles.

| $\overline{FSC_i}$ | 3Dmap |
|---|---|
| 58.85 | emcd166 |
| 57.15 | emcd118 |
| 53.89 | emcd121 |
| 52.46 | emcd112 |
| 51.64 | emcd124 |
| 32.74 | emcd155 |
| 27.72 | emcd147 |
| 27.19 | emcd122 |

**Table 4**
P-values resulting of comparing all 3Dmap pairs for specimen T20S *Proteasome*.

| | emcd108 | emcd103 | emcd107 | emcd141 | emcd145 | emcd144 | emcd162 | emcd130 |
|---|---|---|---|---|---|---|---|---|
| **emcd108** | | 0.3 | 0.1 | 0.3 | 0.4 | 0.4 | 0.02 | 0.02 |
| **emcd103** | 0.3 | | 0.9 | 0.02 | 0.3 | 0.3 | 0.02 | 0.02 |
| **emcd107** | 0.1 | 0.9 | | 0.2 | 0.4 | 0.3 | 0.02 | 0.02 |
| **emcd141** | 0.3 | 0.02 | 0.2 | | 0.4 | 0.4 | 0.02 | 0.02 |
| **emcd145** | 0.4 | 0.3 | 0.4 | 0.4 | | 0.02 | 0.7 | 0.02 |
| **emcd144** | 0.4 | 0.3 | 0.3 | 0.4 | 0.02 | | 0.7 | 0.02 |
| **emcd162** | 0.02 | 0.02 | 0.02 | 0.02 | 0.7 | 0.7 | | 0.02 |
| **emcd130** | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | |

**Table 6**
P-values resulting of comparing all 3Dmap pairs for specimen Apo-Ferritin.

| | emcd166 | emcd118 | emcd121 | emcd112 | emcd124 | emcd155 | emcd147 | emcd122 |
|---|---|---|---|---|---|---|---|---|
| **emcd166** | | 0.02 | 0.6 | 0.6 | 0.04 | 0.02 | 0.02 | 0.02 |
| **emcd118** | 0.02 | | 0.7 | 0.7 | 0.04 | 0.02 | 0.02 | 0.02 |
| **emcd121** | 0.6 | 0.7 | | 0.3 | 0.6 | 0.02 | 0.02 | 0.02 |
| **emcd112** | 0.6 | 0.7 | 0.3 | | 0.7 | 0.02 | 0.02 | 0.02 |
| **emcd124** | 0.04 | 0.04 | 0.6 | 0.7 | | 0.02 | 0.02 | 0.02 |
| **emcd155** | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | | 0.04 | 0.02 |
| **emcd147** | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.04 | | 0.6 |
| **emcd122** | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.6 | |

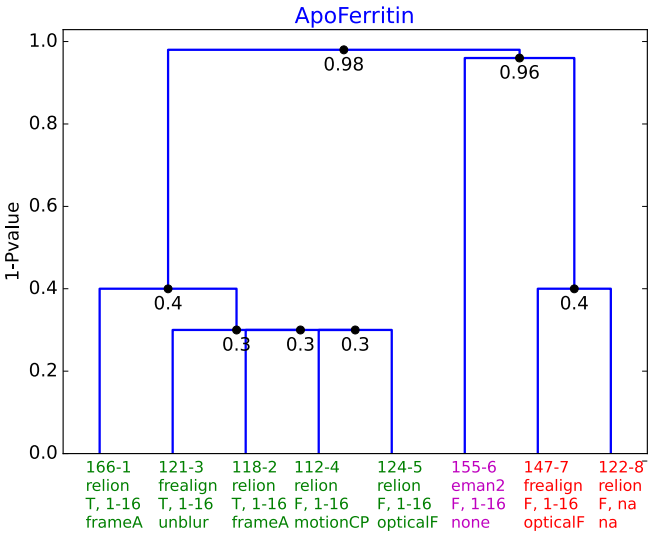



**Fig. 4.** ApoFerritin hierarchical clustering rendered as a dendrogram. X-labels show: 3D map ID and ranking position (first line), software used for 3D angular refinement (second line), if dose weighting has been applied and which frames has been used to create the particle images (third line) and software used for movie alignment (fourth line). Non standard acronyms used in the X-axis labels are listed in B. (CI = 0.78.).

### 3.3.1. *Revised maps*

A revised version of map **emcd155** was uploaded. In the new sorting table, maps **emcd112** and **emcd124** interchange positions but since they are consecutive and their corresponding P-value is 0.7 the new table is equivalent to the old one (Table 5).

## 3.4. *TRPV1 channel*

Tables 7 and 8, and Fig. 5 summarize the results for TRPV1 channel.

**Table 7**
Sorting based on $\overline{FSC_i}$ for specimen TRPV1 Channel. The main three clusters are surrounded by rectangles.

| $\overline{FSC_i}$ | 3Dmap |
|---|---|
| 47.93 | emcd161 |
| 47.56 | emcd135 |
| 46.91 | emcd133 |
| 46.46 | emcd115 |
| 44.46 | emcd101 |
| 36.74 | emcd156 |
| 36.56 | emcd163 |
| 30.80 | emcd146 |

**Table 8**
P-values resulting of comparing all 3Dmap pairs for specimen TRPV1 Channel.

| | emcd161 | emcd135 | emcd133 | emcd115 | emcd101 | emcd156 | emcd163 | emcd146 |
|---|---|---|---|---|---|---|---|---|
| **emcd161** | | 0.5 | 0.5 | 0.1 | 0.04 | 0.04 | 0.04 | 0.02 |
| **emcd135** | 0.5 | | 0.04 | 0.6 | 0.5 | 0.04 | 0.04 | 0.02 |
| **emcd133** | 0.5 | 0.04 | | 0.5 | 0.6 | 0.04 | 0.04 | 0.02 |
| **emcd115** | 0.1 | 0.6 | 0.5 | | 0.04 | 0.04 | 0.04 | 0.02 |
| **emcd101** | 0.04 | 0.5 | 0.6 | 0.04 | | 0.04 | 0.04 | 0.02 |
| **emcd156** | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | | 0.8 | 0.02 |
| **emcd163** | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.8 | | 0.02 |
| **emcd146** | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | |

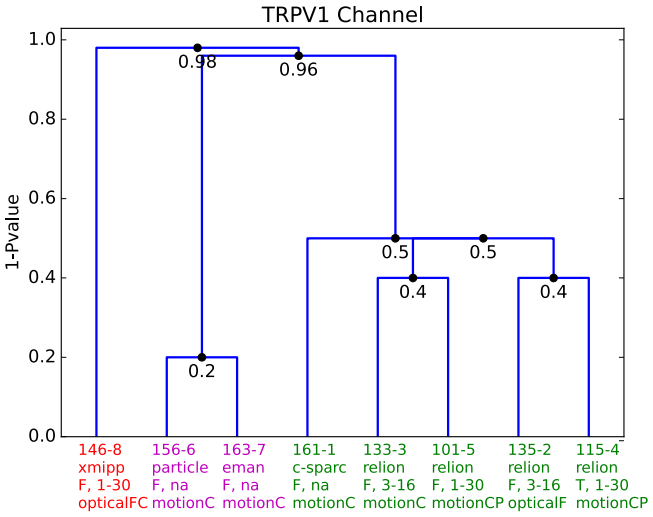



**Fig. 5.** TRPV1 channel hierarchical clustering rendered as a dendrogram. X-labels show: 3D map ID and ranking position (first line), software used for 3D angular refinement (second line), if dose weighting has been applied and which frames has been used to create the particle images (third line) and software used for movie alignment (fourth line). Non standard acronyms used in the X-axis labels are listed in B. (CI = 0.77.).

**Table 9**
Sorting based on $\overline{FSC_i}$ for specimen 80S Ribosome. The main six clusters are surrounded by rectangles.

| $\overline{\mathbf{FSC_i}}$ | 3Dmap |
|---|---|
| 218.89 | emcd123 |
| 208.88 | emcd151 |
| 204.22 | emcd150 |
| 203.06 | emcd149 |
| 198.06 | emcd126 |
| 193.87 | emcd114 |
| 190.91 | emcd125 |
| 180.48 | emcd127 |
| 177.90 | emcd148 |
| 176.32 | emcd128 |
| 168.14 | emcd119 |
| 164.88 | emcd111 |
| 141.11 | emcd129 |

### 3.4.1. Revised maps

There is a revised version of maps **emcd146** and **emcd163**. Recalculating the sorting table using the revised volumes produced a change of order between the maps **emcd135**, **emcd133** and **emcd115** and between the maps **emcd156** and **emcd163**. This new sorting is equivalent to the old one given by Table 7.

## 3.5. 80S ribosome

Tables 9 and 10, and Fig. 6 summarize the results for *80S ribosome*.

This is the only specimen in which small variations of the P-Value produce different numbers of clusters. Results are shown for P-Value = 0.05 and P-value = 0.1. A particularity of this dataset is that five 3D maps have been submitted by the same author. This author classified the images in 4 groups and uploaded the reconstruction form each class and from all the four classes together. See 3D maps **emcd126**, **emcd127**, **emcd128**, **emcd129** and **emcd123**.

**Table 10**
P-values resulting of comparing all 3Dmap pairs for specimen 80S Ribosome. In order to make the table fit in the page the name of the 3Dmaps has been shortened from the canonical form **emcdXXX** to **eXXX**.

| | e123 | e151 | e150 | e149 | e126 | e114 | e125 | e127 | e148 | e128 | e119 | e111 | e129 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **e123** | | 0.07 | 0.03 | 0.02 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| **e151** | 0.07 | | 0.07 | 0.09 | 0.06 | 0.09 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 |
| **e150** | 0.03 | 0.07 | | 0.02 | 0.4 | 0.2 | 0.07 | 0.00 | 0.02 | 0.00 | 0.03 | 0.01 | 0.00 |
| **e149** | 0.02 | 0.09 | 0.02 | | 0.4 | 0.2 | 0.1 | 0.00 | 0.02 | 0.00 | 0.03 | 0.01 | 0.00 |
| **e126** | 0.00 | 0.06 | 0.4 | 0.4 | | 0.37 | 0.1 | 0.00 | 0.05 | 0.00 | 0.02 | 0.02 | 0.00 |
| **e114** | 0.01 | 0.09 | 0.2 | 0.2 | 0.37 | | 0.28 | 0.09 | 0.06 | 0.02 | 0.03 | 0.03 | 0.00 |
| **e125** | 0.00 | 0.01 | 0.07 | 0.1 | 0.1 | 0.28 | | 0.09 | 0.06 | 0.01 | 0.03 | 0.01 | 0.00 |
| **e127** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.09 | | 0.4 | 0.00 | 0.06 | 0.06 | 0.00 |
| **e148** | 0.01 | 0.00 | 0.02 | 0.02 | 0.05 | 0.06 | 0.06 | 0.4 | | 0.7 | 0.1 | 0.07 | 0.00 |
| **e128** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.7 | | 0.1 | 0.07 | 0.00 |
| **e119** | 0.00 | 0.01 | 0.03 | 0.03 | 0.02 | 0.03 | 0.03 | 0.06 | 0.1 | 0.1 | | 0.01 | 0.00 |
| **e111** | 0.00 | 0.01 | 0.01 | 0.01 | 0.02 | 0.03 | 0.01 | 0.06 | 0.07 | 0.07 | 0.01 | | 0.01 |
| **e129** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | |

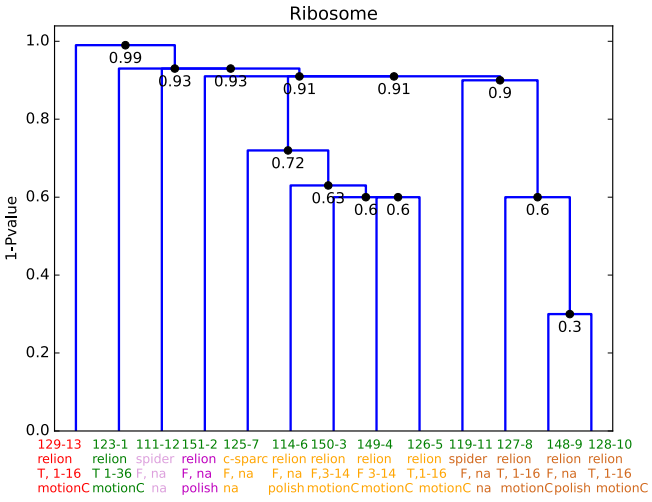



**Fig. 6.** Ribosome hierarchical clustering rendered as a dendrogram. X-labels show: 3D map ID and ranking position (first line), software used for 3D angular refinement (second line), if dose weighting has been applied and which frames has been used to create the particle images (third line) and software used for movie alignment (fourth line). Non standard acronyms used in the X-axis labels are listed in B. (CI = 0.83.) Threshold used for clustering P.-value = 0.05.

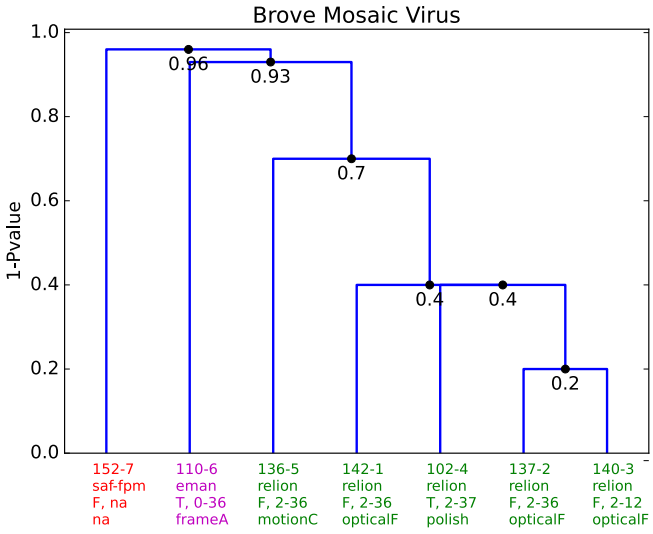



**Fig. 7.** Brove Mosaic Virus hierarchical clustering rendered as a dendrogram. X-labels show: 3D map ID and ranking position (first line), software used for 3D angular refinement (second line), if dose weighting has been applied and which frames has been used to create the particle images (third line) and software used for movie alignment (fourth line). Non standard acronyms used in the X-axis labels are listed in B. (CI = 0.86.).

### 3.5.1. Revised maps

A revised version of map **emcd111** was uploaded. Sorting table recalculation using the revised map did not alter the old Table 9 order.

## 3.6. Brome mosaic virus

Tables 11 and 12, and Fig. 7 summarize the results for *Brome mosaic virus*. Alignment for this specimen was made based on symmetry. First, the orientation of the symmetry axes was detected and all maps were rotated so that they present i3 orientation (for example, 3D maps with i1 orientation were rotated using the Euler angles (0,63.43494882,0)). We note here that no 3D map presented originally i3 orientation. (Our notation for symmetry orientation is summarized inhttps://github.com/I2PC/xmipp-portal/wiki/Symmetry.) A particularity of this dataset is that four 3D maps out of seven have been submitted by the same author with small variations in the way movies are aligned (see datasets **emcd136**, **emcd137**, **emcd140** and **emcd142).**

**Table 11**
Sorting based on $\overline{FSC_i}$ for specimen Brome Mosaic Virus. The main three clusters are surrounded by rectangles.

| $\overline{FSC_i}$ | 3Dmap |
|---|---|
| 92.04 | emcd142 |
| 90.33 | emcd137 |
| 89.69 | emcd140 |
| 87.11 | emcd102 |
| 79.89 | emcd136 |
| 69.92 | emcd110 |
| 56.12 | emcd152 |

### 3.6.1. Revised maps

A revised version of map **emcd110** was uploaded. Sorting table recalculation using the revised map did not alter the old Table 11 order.

## 3.7. β-Galactosidase

Tables 13 and 14, and Fig. 8 summarize the results for *β-Galactosidase*. This is the only dataset in which all reconstructions belong to the same cluster.

**Table 12**
P-values resulting of comparing all 3Dmap pairs for specimen Brome Mosaic Virus

| | emcd142 | emcd137 | emcd140 | emcd102 | emcd136 | emcd110 | emcd152 |
|---|---|---|---|---|---|---|---|
| **emcd142** | | 0.6 | 0.6 | 0.2 | 0.04 | 0.04 | 0.04 |
| **emcd137** | 0.6 | | 0.8 | 0.6 | 0.04 | 0.04 | 0.04 |
| **emcd140** | 0.6 | 0.8 | | 0.2 | 0.1 | 0.04 | 0.04 |
| **emcd102** | 0.2 | 0.6 | 0.2 | | 0.3 | 0.04 | 0.04 |
| **emcd136** | 0.04 | 0.04 | 0.1 | 0.3 | | 0.07 | 0.04 |
| **emcd110** | 0.04 | 0.04 | 0.04 | 0.04 | 0.07 | | 0.04 |
| **emcd152** | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | |

**Table 13**
Sorting based on $\overline{FSC_i}$ for specimen *β-Galactosidase*. There is a single cluster.

| $\overline{FSC_i}$ | 3Dmap |
|---|---|
| 93.44 | emcd138 |
| 93.17 | emcd139 |
| 91.57 | emcd159 |
| 91.5 | emcd164 |
| 89.76 | emcd106 |
| 88.2 | emcd134 |
| 87.63 | emcd167 |
| 86.57 | emcd113 |
| 83.18 | emcd160 |
| 79.77 | emcd157 |
| 76.41 | emcd154 |

### 3.7.1. Revised maps

Revised versions of maps **emcd134**, **emcd157** and **emcd160**, were uploaded. In the new sorting table, maps **emcd106 emcd159** and **emcd164** change order as well as maps **emcd134 emcd167**, **emcd113** and **emcd160**. Nevertheless, since there is a single class the new table is equivalent to the old one (Table 13).

## 4. Discussion

The goal of this work is to uncover patterns in the data collected by the Map Challenge. In the Results section we have ranked the different 3D Maps and now the task is to relate good reconstructions with a particular image processing workflow (IPW). This is not an easy goal, since the data collected are not ideal. To start with, the number of reported reconstructions is small, and although there is not a rule of thumb to determine the optimal sample size, some researchers suggest that there should be around 10 observations per variable (that is, per step in the IPW where different algorithms be used). Another sources of difficulties analyzing the data sets is that the IPWs followed by the different participants are not always carefully described, and the coverage of the different datasets is very uneven. For example, for the Brome Mosaic Virus, seven reconstructions have been uploaded, four of them made by the same author (see data sets **emcd136**, **emcd137**, **emcd140** and **emcd142**). A similar situation occurs for the ribosome data set (see 3D maps **emcd126**, **emcd127**, **emcd128**, **emcd129** and **emcd123**). From a more methodological point of view, the Pair Comparison method, as most of the statistical tools, assumes that the samples -reconstructions in our case- are taken at random, but in the Map Challenge one package -Relion, (Scheres, 2012)- is predominant. If Relion presents a systematic bias in its reconstructions, our statistical test may be biased. A second source of bias is the tendency to be selective reporting outcomes, that is, researches report only the best results and hide the rest (see Table 14).

Let us start the analysis with a few almost self evident results. The first lesson we learn from the challenge is that all algorithms work properly if the data is good. The two datasets with higher resolution, T20S *Proteasome* and *β-Galactosidase*, produce a set of reconstructions that cannot been distinguished (with P-value = 0.05). (The only exception to this rule is the reconstruction **emcd130**.) On the other hand Apo-Ferritin, the worst specimen in terms of final resolution, presents three clear clusters meaning that for challenging data different algorithms perform differently. Unfortunately, no clear pattern seems to emerge from these clusters. For example, for 3D map refinement, Relion has produced some of the best and worst reconstructions. Another challenging specimen that produces many clusters is the 80S Ribosome

**Table 14**
P-values resulting of comparing all 3Dmap pairs for specimen *β-Galactosidase*. In order to make the table fit in the page the name of the 3Dmaps has been shortened from the canonical form **emcdXXX** to **eXXX**.

| | e138 | e139 | e159 | e164 | e106 | e134 | e167 | e113 | e160 | e157 | e154 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **e138** | | 0.3 | 0.5 | 0.5 | 0.5 | 0.1 | 0.5 | 0.06 | 0.1 | 0.01 | 0.007 |
| **e139** | 0.3 | | 0.5 | 0.5 | 0.5 | 0.08 | 0.5 | 0.06 | 0.1 | 0.01 | 0.007 |
| **e159** | 0.5 | 0.5 | | 0.1 | 0.5 | 0.5 | 0.007 | 0.5 | 0.5 | 0.5 | 0.5 |
| **e164** | 0.5 | 0.5 | 0.1 | | 0.5 | 0.5 | 0.007 | 0.5 | 0.5 | 0.5 | 0.5 |
| **e106** | 0.5 | 0.5 | 0.5 | 0.5 | | 0.6 | 0.5 | 0.05 | 0.1 | 0.007 | 0.007 |
| **e134** | 0.1 | 0.08 | 0.5 | 0.5 | 0.6 | | 0.5 | 0.4 | 0.3 | 0.02 | 0.01 |
| **e167** | 0.5 | 0.5 | 0.007 | 0.007 | 0.5 | 0.5 | | 0.5 | 0.5 | 0.5 | 0.5 |
| **e113** | 0.06 | 0.06 | 0.5 | 0.5 | 0.05 | 0.4 | 0.5 | | 0.3 | 0.02 | 0.007 |
| **e160** | 0.1 | 0.1 | 0.5 | 0.5 | 0.1 | 0.3 | 0.5 | 0.3 | | 0.007 | 0.01 |
| **e157** | 0.01 | 0.01 | 0.5 | 0.5 | 0.007 | 0.02 | 0.5 | 0.02 | 0.007 | | 0.08 |
| **e154** | 0.007 | 0.007 | 0.5 | 0.5 | 0.007 | 0.01 | 0.5 | 0.007 | 0.01 | 0.08 | |

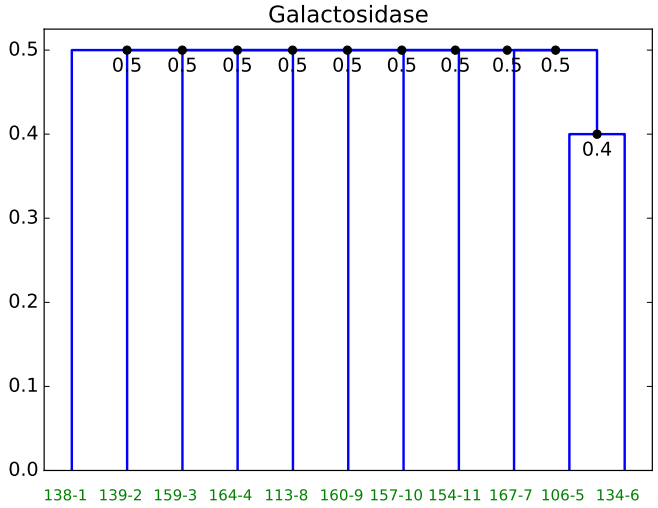



**Fig. 8.** *β-Galactosidase* hierarchical clustering rendered as a dendrogram. X-label show: 3D map ID and ranking position (first line). The results show that all reconstructions are equivalente with a P-value = 0.05. (CI = 0.19.).

because it presents data heterogeneity. As in the previous case, Relion has produced some of the best and the worst volumes.

In a typical IPW there are many steps. We will comment here only on those for which we have been able to obtain some conclusions. In particular, we will skip the CTF estimation since the performance of the different CTF estimation algorithms under different conditions was analyzed in the CTF Challenge (Marabini et al., 2015) and one of the conclusions was: "…when a data set is *good*, most packages provide similar results." Since all experimental Map Challenge datasets qualify as good, we do not expect to find here an algorithm that surpasses the others. We will neither comment on classification, since the Map Challenge is not oriented toward heterogeneity, nor on initial model. The generation of an initial model is still an open and challenging problem but the Map Challenge design is not the adequate one to tackle it. The main reason is that good solutions (high resolution maps) are provided for the different datasets and, therefore, the process of creating an initial reference is trivial.

In most cases, the first step in the IPW is to align the frames within a given movie. Many algorithms have been applied to this task including: DE script (Spear et al., 2015), motioncor (Li et al., 2013), polishing (Scheres, 2012), optical flow (Abrishami et al., 2015), unblur (Grant et al., 2015), imod (Kremer et al., 1996), warp ( https://github.com/dtegunov/warp), etc. In all cases, the reconstructions that belong to the winner group use movies that have been aligned. Therefore, as it was already suspected, movie alignment produces improved reconstructions. The fourth line in the dendrograms shows the frame alignment algorithms used by the different reconstructions. Unfortunately, from this set of dendrograms it is not possible to pick an algorithm that is significantly better than the rest. Nevertheless, it is interesting to examine the pairs of reconstructions: (**emcd144**, **emcd145**), (**emcd149**, **emcd150**), (**emcd136**, **emcd137**) and (**emcd133**, **emcd135**). These pairs of reconstructions have been performed by the same author and, for each pair, they are identical except in the alignment step. In all cases, the movies have been aligned using motioncorr and in half of then optical-flow has also been executed. The 3D maps produced by optical flow consistently are better ranked than those produced using motioncorr alone. If we are strict, we can only conclude that for the given IPW followed by a given author, to use optical-flow is better than to use motioncorr alone, but it is tempting to conclude than algorithms that can perform local alignments are potentially better than algorithms that work globally. We point out here that a program similar to motioncorr, not available when the Map Challenge was active, and known as motioncor2 (Zheng et al., 2017) includes a local alignment step. It is also important to comment that although the reconstructions performed with optical flow are consistently better than the reconstructions performed without it, the magnitude of the improvement is small.

During or after movie alignment, frames may be dose weighted to limit the effect of radiation damage in the average images. This effect may be achieved either by applying a weight to each frame or simply by rejecting some of the last acquired frames. Dendrograms displayed in Figs. 3–7 show if some weighting mechanism has been applied to the input data. There is a strong correlation between no weighting and being among the worst cluster but not the other way around. That is: the worst reconstructions are never weighted, but non-weighted reconstructions many times are among the best results. The only exception to the correlation between non weighted data and being in the worst group is 3D map **emcd129** but this is one of the five maps submitted by the same author testing the four different subsets obtained after classifications, and very likely what we are seeing here is the effect of processing a data set with a small number of particles.

Finally, we arrive to one of the most delicate steps, 3D map refinement that usually is coupled with 3D classification. In this section Relion is specially over-represented, since it has been used in 30 out of 55 3D maps. As mentioned before, over-representation of a particular package might result in $\overline{FSC_i}$ being a biased estimator. In the following comparison we keep using the terms bad and good for reconstructions with high and low $\overline{FSC_i}$ but, in purity, if a large number of reconstructions are biased we would be reporting on how similar or different a reconstruction is from the average of the reconstructions uploaded by all participants. In the next paragraph, the first time a software package is cited we indicate the number of times that it has been used in parenthesis, to give an idea of package over-representation. Since all reconstructions created from the *β-Galactosidase* data perform equally good, this specimen is not taken into account to compute representation.

Dendrograms in Figs. 2–7 show the map refinement method used for each reconstruction. Here, a clear pattern emerges. At least one of the 3D maps in the best group have been produced by Relion, although in two occasions Relion has produced 3D maps that belong to the worst group. SAF-FPM (2, Estrozi et al., 2010) produced the worst results, followed by XMIPP new 3D map refinement algorithm *highres* (4, unpublished). Other software packages used, sorted by alphabetical order are: bsoft(2, Heymann, 2001), cryoSparc (4, Punjani et al., 2017), eman2 (6, Tang et al., 2007), freealign (2, Grigorieff, 2007), jspr(2, Guo and Jiang, 2014), particle (1,http://www.image-analysis.net/EM/), and spider (2, Shaikh et al., 2008). These methods are underrepresented and are difficult to sort. JSPR is always among the best, but it has only been applied to two datasets. CryoSparc has always produced 3D maps that belong to the best group but for the ribosome case, but has only been applied to 4 datasets.

In summary, results are not very conclusive but it seems that movie alignment improves the final results and very likely local movie alignment makes this improvement slightly higher. From the results, it is not unambiguous if dose weighting is beneficial, but researchers that use it never produce the worst results. Finally, it is clear that Relion is the software selected by most of the participants for angular refinement. Relion is able to produce good results and it is being widely applied, but from this work we cannot conclude that it is the best option.

## 5. Conclusions

We have presented the analysis performed over the 3D maps created for the Map Challenge. To perform the analysis, a new algorithm called Pair Comparison Method has been developed. The algorithm is able to sort reconstructions and assign a level of significance to the sorting.

The authors of this work have not been able to propose an ideal image processing workflow because, with the available data, several algorithms produce results that are not significantly different. We believe that a more focused challenge, or set of challenges, in which each image processing step would be isolated and analyzed would have provided more useful information than the present challenge covering the whole image processing workflow.

We believe that the more important output of the Map Challenge is not the ranking of the reconstructions and software packages, which is always a matter of controversy, but the production of an impressive collection of curated data sets that, no doubt, will be used as reference in the future.

We would like to end this article by thanking the challenge organization who has worked hard during the organization of the event and to all the participants that have spent a lot of time and CPU producing reconstructions.

## Acknowledgements

The authors would like to acknowledge economical support from: The Spanish Ministry of Economy and Competitiveness through Grants BIO2013-44647-R, BIO2016-76400-R(AEI/FEDER, UE) and AEI/FEDER BFU 2016 74868P, the Comunidad Autónoma de Madrid through Grant: S2017/BMD-3817, European Union (EU) and Horizon 2020 through grant CORBEL (INFRADEV-1-2014-1, Proposal: 654248). This work used the EGI Infrastructure and is co-funded by the EGI-Engage

project (Horizon 2020) under Grant No. 654142. European Union (EU) and Horizon 2020 through grant West-Life (EINFRA-2015-1, Proposal: 675858) European Union (EU) and Horizon 2020 through grant Elixir - EXCELERATE (INFRADEV-3-2015, Proposal: 676559) European Union (EU) and Horizon 2020 through grant iNEXT (INFRAIA-1–2014-2015, Proposal: 653706). The authors acknowledge the support and the use of resources of Instruct, a Landmark ESFRI project.

## Appendix A. Description of the pair comparison method

In this section we describe in detail the pair comparison method including the test performed prior to its application to the Map Challenge.

### A.1. Algorithm description

For a given specimen the proposed method requires:

- Compute all possible pairs of reconstructions. That is, if 4 reconstructions were uploaded ($r_1$, $r_2$, $r_3$ and $r_4$)$6 \times 2$ pairs will be created (($r_1$, $r_2$), ($r_1$, $r_3$), ($r_1$, $r_4$), ($r_2$, $r_3$), ($r_2$, $r_4$) and ($r_3$, $r_4$) plus six redundant extra pairs in which the first and second member are interchanged ($r_2$, $r_1$), ($r_3$, $r_1$), ($r_4$, $r_1$), ($r_3$, $r_2$), ($r_4$, $r_2$) and ($r_4$, $r_3$).)
- For each pair formed by the $i$th and $j$th reconstructions, compute the Fourier Shell Correlation ($FSC_{i,j}$) between the first member of the pair and the second member of the pair. (Obviously ($FSC_{i,j} = FSC_{j,i}$)).
- Then compute a weighted Fourier shell correlation integral defined as $\overline{FSC_{i,j}} = \int_{200\text{Å}}^{\nu_{max}} FSC_{i,j}(\nu)\nu d\nu$ where $\nu$ is the frequency in Fourier space and $\nu_{max}$ is the higher resolution reported for the dataset under analysis. (The integral, that is, the region under the FSC curve has been approximated by a set of rectangles and then added up the area of these rectangles. The rectangle width is equal to the sampling rate.)
- For each reconstruction $r_i$ compute $\overline{FSC_i} = \sum_{j=1,i\neq j}^{j=J} \overline{FSC_{i,j}}$ where $J$ is the number of reconstructions.
- $\overline{FSC_i}$ will be used for sorting the 3D maps.

In the absence of systematic bias, the higher the resolution of the reconstruction $r_i$, the higher will be the value of $\overline{FSC_i}$. Therefore this magnitude may be used to sort the reconstructions. Unfortunately, even if we can rank our 3D maps, we do not know if two consecutively ranked reconstructions $r_\alpha$ and $r_\beta$ are statistically different. This is an important question because if they are statically different we could claim that the image processing workflow (IPW) used to produce $r_\alpha$ is superior to the IPW used to produce $r_\beta$ (for the particular specimen under study). On the other hand, if $r_\alpha$ and $r_\beta$ are not statistically different we cannot reject the hypothesis that both IPWs perform equally well.

To answer the question whether two reconstructions $r_\alpha$ and $r_\beta$ are statistically different we follow this approach:

- Let $\overline{FSC_{\alpha,k}}$ and $\overline{FSC_{\beta,k}}$ be the set of weighted Fourier correlation integrals related with $r_\alpha$ and $r_\beta$ respectively.
- For a given $k$, $\overline{FSC_{\alpha,k}}$ and $\overline{FSC_{\beta,k}}$ are correlated, and therefore we may use a paired test to compare the two population means.
- The best known paired test is *paired t-test*. However, the *paired t-test* assumes that the sample is normally distributed, which very likely will not be the case. Therefore, we will use the Wilcoxon signed rank test that does not require this assumption.

### A.2. Test on the performance of the pair comparison method

A collection of experiments has been performed in order to judge how reliable and robust is the method. Two different phantoms have been used. The first is totally asymmetric (a ribosome) while the second one presents high symmetry (an icosahedral virus). Since results are very similar for both cases, in the following we present in detail the experiments performed with the second phantom. This phantom is based on the quasi-atomic model of bacteriophage T7 procapsid shell described in Agirrezabala et al. (2007) and deposited in the PDB with accession number 3IZG. A surface rendering can be seen in Fig. 9.

In a nutshell the design of the experiments is as follows. A large set of projections is created. These projections are divided in subsets and reconstructed. The pair comparison method is applied to the reconstructions in order to sort them based on a figure of merit. Since we are working with phantoms, we can compare this sorting with a control one and check if the new algorithm is working properly. Finally, for those reconstructions that are in different positions in the sorting produced by the new algorithm and by the control, we test if this disagreement is statistically significant or not.20,000 noisy unaligned projections were created with a sampling rate of 1.5 Å/px. From this data set, 13 independent subsets of projections were generated with: 700, 1020, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000 and 10000 projections respectively. These subsets were reconstructed using Relion (Scheres, 2012) and sorted applying the described algorithm. In the rest of the article we use the symbol $r_{xxxx}$ to denote a reconstruction obtained from *xxxx* projections.

The first step before analyzing the results is to establish which is the correct reconstruction order. *A priori* we expect that there should be a strong correlation between the reconstruction position and the number of projections. Nevertheless, it is possible that reconstructions obtained from fewer projections may have higher “quality”. Using the phantom as reference, we define as correct order the one given by sorting the value $\overline{FSC_{i,phan}}$. Table 15 (fourth column) shows the results of sorting by this control magnitude and confirm our suspicion that, in a few cases for example $r_{4000}$ and $r_{5000}$, a reconstruction obtained from a higher number of projections present a lower “quality” than a reconstruction obtained from less projections.

Table 15 (second column) shows the reconstructions sorted by the pair comparison method. We see that the sorting provided by this method and the control one, although similar, is not identical (see for example $r_{6000}$ and $r_{5500}$). The question that arises now is if both sortings are equivalents, that is, can we claim that $r_{6000}$ and $r_{5500}$ (or $r_{4500}$, $r_{4000}$ and $r_{5000}$) are different?. To answer this question, we apply the Wilcoxon test. In Table 16 we show the P-value obtained from comparing the set of values $\overline{FSC_{\alpha,j}}$ and $\overline{FSC_{\beta,j}}$ related with the reconstruction $r_\alpha$ and $r_\beta$. We define that two reconstructions are statistically distinguishable if the P-value between them is smaller than 0.05. In Table 16, where P-values higher than 0.05 are marked in red, we see that we cannot claim that pairs ($r_{4000}$, $r_{4500}$), ($r_{4000}$, $r_{5000}$), ($r_{4500}$, $r_{5000}$) and ($r_{5500}$, $r_{6000}$) are different with a 0.05 statistical significance. In Table 15 reconstructions with the same background color are equivalent (from the point of view of the pair comparison method). From these data, we conclude that the order provided by the pair comparison method and the reference one are statistically equivalent.

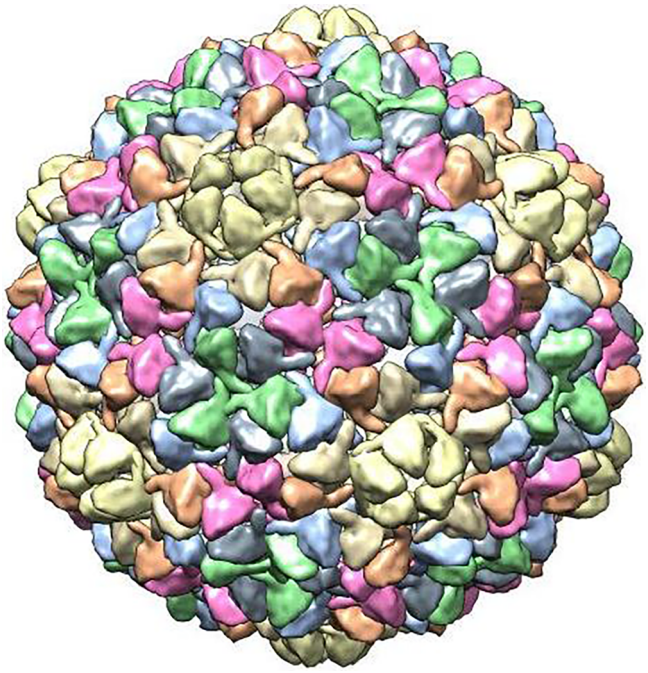


**Fig. 9.** Surface rendering of bacteriophage T7 procapsid shell (PDB Id = 3IZG).

**Table 15**
Comparison of the sorting provided by the pair comparison method (first and second columns) vs the control (third and fourth columns). Cell with the same color contain subsets of reconstructions for which we can not calim the null hypothesis "the reconstructions are different" with an statistical significance of 0.05.

| $\overline{\mathbf{FSC_{i,j}}}$ | sort using pair comparison method | $\overline{\mathbf{FSC_{i,phan}}}$ | sort using Phantom |
|---|---|---|---|
| 182.41 | $r_{10000}$ | 19.67 | $r_{10000}$ |
| 176.63 | $r_{6000}$ | 18.34 | $r_{5500}$ |
| 176.55 | $r_{5500}$ | 18.28 | $r_{6000}$ |
| 172.59 | $r_{4500}$ | 17.78 | $r_{4000}$ |
| 172.22 | $r_{4000}$ | 17.71 | $r_{4500}$ |
| 171.33 | $r_{5000}$ | 17.46 | $r_{5000}$ |
| 167.81 | $r_{3500}$ | 16.88 | $r_{3500}$ |
| 162.94 | $r_{3000}$ | 16.22 | $r_{3000}$ |
| 156.23 | $r_{2500}$ | 15.15 | $r_{2500}$ |
| 153.56 | $r_{2000}$ | 14.86 | $r_{2000}$ |
| 142.44 | $r_{1500}$ | 13.53 | $r_{1500}$ |
| 134.85 | $r_{1020}$ | 12.60 | $r_{1020}$ |
| 125.04 | $r_{700}$ | 11.37 | $r_{700}$ |

**Table 16**
Wilcoxon test. Red colored cells mark pairs of reconstructions that cannot be distinguish with a P-value greater than 0.05.

| | 700 | 1020 | 1500 | 2000 | 2500 | 3000 | 3500 | 4000 | 4500 | 5000 | 5500 | 6000 | 10000 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **700** | NA | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| **1020** | 0.02 | NA | 0.02 | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| **1500** | 0.02 | 0.02 | NA | 0.05 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 |
| **2000** | 0.02 | 0.03 | 0.05 | NA | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| **2500** | 0.02 | 0.03 | 0.03 | 0.02 | NA | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| **3000** | 0.02 | 0.03 | 0.03 | 0.02 | 0.02 | NA | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| **3500** | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | NA | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| **4000** | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | NA | 0.9 | 0.2 | 0.02 | 0.02 | 0.02 |
| **4500** | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.9 | NA | 0.4 | 0.02 | 0.02 | 0.02 |
| **5000** | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.2 | 0.4 | NA | 0.02 | 0.02 | 0.02 |
| **5500** | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | NA | 0.7 | 0.02 |
| **6000** | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.7 | NA | 0.02 |
| **10000** | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | NA |

### *A.3. Using alternative experimental setups*

As a way to further validate the proposed pair comparison method, we decided to modified it incorporating two variants. In the first case we used instead of $\overline{FSC_i}$ the magnitude known as R-factor (see Eq. A.1b) for computing the control sorting. In the second variant we check the influence of applying a tight mask to the reconstructed 3D map.

#### *A.3.1. R-factor*

In this experiment R-factors were calculated applying the macromolecular refinement program REFMAC (Murshudov et al., 1997) in the frequency range 20–3.5 Å. Using these values a new control sorting was generated. The results, which are presented in Table 17 (second column), are partially in disagreement with the sorting computed using the pair comparison method (Table 17, fourth column). This discrepancy cannot be justified even if we take into account the information provided by the Wilcoxon test that points out which reconstructions are similar. On the other hand, if the algorithm was executed using as meassurement of quality another magnitude produced by REFMAC called *average Fourier shell correlation*; the results are quite close (Table 17, sixth column) to the ones produced by the original pair comparison method. The only significant difference is that the reconstruction in the ninth and tenth positions are swapped. Finally, we were able to reconcile the sorting produced by all magnitudes if the R-factor was computed only for high frequencies in the range 5–3.5 Å(Table 17, eight column). In this case, both variants of the algorithm, the one based on $\overline{FSC_i}$ and the one based on R-Factor produce equivalent results except for the reconstructions in the last two positions which are swapped.

The divergence between both measures of quality (R-factor and $\overline{FSC_i}$) is due to the fact that R-factor depends more heavily on the low frequency values than the FSC. FSC is computed by rings (see Eq. A.1a), and the value of each ring does not depend on the absolute magnitude of the Fourier

**Table 17**
Comparison of the sorting provided by R-factor (refmac, first and second columns), $\overline{FSC_i}$ (new method, third and fourth columns) and Average Fourier Shell Correlation (refmac, fifth and sixth columns). Cell with the same color contain subsets of reconstructions that are statistically indistinguishable.

| R-factor (20 to 3.5) | sort using R-factor (20 to 3.5) | $\overline{FSC_i}$ | sort using pair comparison method | averag. FSC | sort using averg. FSC | R-factor (5.0 to 3.5) | sort using R-factor (5.0 to 3.5) |
|---|---|---|---|---|---|---|---|
| 0.1885 | $r_{10000}$ | 182.41 | $r_{10000}$ | 0.861 | $r_{10000}$ | 0.321 | $r_{10000}$ |
| 0.1991 | $r_{5000}$ | 176.63 | $r_{6000}$ | 0.8231 | $r_{5500}$ | 0.3422 | $r_{5500}$ |
| 0.2024 | $r_{4500}$ | 176.55 | $r_{5500}$ | 0.8182 | $r_{6000}$ | 0.3465 | $r_{6000}$ |
| 0.2034 | $r_{6000}$ | 172.59 | $r_{4500}$ | 0.8089 | $r_{4500}$ | 0.3515 | $r_{4000}$ |
| 0.2054 | $r_{5500}$ | 172.22 | $r_{4000}$ | 0.8017 | $r_{4000}$ | 0.3567 | $r_{4500}$ |
| 0.2068 | $r_{4000}$ | 171.33 | $r_{5000}$ | 0.7995 | $r_{5000}$ | 0.3615 | $r_{5000}$ |
| 0.2131 | $r_{3000}$ | 167.81 | $r_{3500}$ | 0.7805 | $r_{3500}$ | 0.3772 | $r_{3500}$ |
| 0.2163 | $r_{1020}$ | 162.94 | $r_{3000}$ | 0.7609 | $r_{3000}$ | 0.3907 | $r_{3000}$ |
| 0.2184 | $r_{3500}$ | 156.23 | $r_{2500}$ | 0.709 | $r_{2000}$ | 0.3909 | $r_{2500}$ |
| 0.2193 | $r_{2500}$ | 153.56 | $r_{2000}$ | 0.7087 | $r_{2500}$ | 0.4008 | $r_{2000}$ |
| 0.2326 | $r_{2000}$ | 142.44 | $r_{1500}$ | 0.6799 | $r_{1500}$ | 0.4339 | $r_{1500}$ |
| 0.2345 | $r_{1500}$ | 134.85 | $r_{1020}$ | 0.6402 | $r_{1020}$ | 0.4507 | $r_{700}$ |
| 0.2389 | $r_{700}$ | 125.04 | $r_{700}$ | 0.6204 | $r_{700}$ | 0.4526 | $r_{1020}$ |

**Table 18**
Comparison of the sorting provided by weighted Fourier shell correlation (first and second columns). R-factor (refmac, third and fourth columns).

| $\overline{FSC_{i,j}}$ | sort usingpair comparisonmethod | R-factor(5.0 to 3.5) | sort using R-factor (5.0 to 3.5) |
|---|---|---|---|
| 228.33 | $rm_{10000}$ | 0.3118 | $rm_{10000}$ |
| 227.03 | $rm_{5500}$ | 0.3265 | $rm_{5500}$ |
| 224.86 | $rm_{4500}$ | 0.3404 | $rm_{4500}$ |
| 223.27 | $rm_{3500}$ | 0.3465 | $r_{6000}$ |
| 219.09 | $rm_{2500}$ | 0.3515 | $r_{4000}$ |
| 209.77 | $rm_{1500}$ | 0.3571 | $rm_{3500}$ |
| 195.98 | $rm_{700}$ | 0.3615 | $r_{5000}$ |
| 190.26 | $r_{6000}$ | 0.3694 | $rm_{2500}$ |
| 186.02 | $r_{4000}$ | 0.3907 | $r_{3000}$ |
| 183.93 | $r_{5000}$ | 0.4008 | $r_{2000}$ |
| 173.73 | $r_{3000}$ | 0.4107 | $rm_{1500}$ |
| 162.25 | $r_{2000}$ | 0.412 | $rm_{700}$ |
| 139.98 | $r_{1020}$ | 0.4526 | $r_{1020}$ |

components at that ring but on the similarity between the compared 3D maps at that frequency. On the other hand R-factor (see Eq. A.1b) is a summation over the whole Fourier space and even after applying a $\beta$-factor to the reconstruction, it is more sensible to similarities at low frequency than FSC. We end this subsection with the equations that define *FSC* and *Rfactor*

$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}} \tag{A.1a}$$

$$Rfactor = \frac{\sum_{r_{min}}^{r_{max}} \left| \left| F_{\text{obs}} \right| - \left| F_{\text{calc}} \right| \right|}{\sum_{r_{min}}^{r_{max}} \left| F_{\text{obs}} \right|} \tag{A.1b}$$

where, $F_1$ is the Fourier transform of the first 3D map, $F_2^*$ is the complex conjugate of the Fourier transform of the second 3D map 2, and $r_i$ is the individual voxel element at radius $r$. $F_{\text{obs}}$ and $F_{\text{cal}}$ are the Fourier transforms of the reconstructed and the reference 3D map from the PDB file respectively, the sum extends over all the space between a range of frequencies $r_{min}$ and $r_{max}$.

### *A.3.2. Applying tight masks*

In our last experiment we wanted to check the behavior of the proposed method when a tight mask was applied to some of the reconstructions. In this way, before performing the sorting, we applied to half of the reconstructions a mask obtained using the algorithm *post-process* provided by Relion. *post-process* was executed with the default parameters except for the *binarization threshold*, *mask pixel extension* and *add soft edge* that were set to 0.02, 3px and 3px respectively. For each reconstruction its corresponding mask was calculated and applied. That is, the masks applied to the different volumes are similar but not identical. The control sorting computed for this data set using $\overline{FSC_i}$ is shown in Table 18 (second column). In this table we differentiate the reconstructions with and without masks by adding the character *m* to the reconstruction name, in this way $rm_{10000}$ is a reconstruction from 10,000 projections that has been masked while $r_{6000}$ is a reconstruction from 6000 projections that has not been masked. The table clearly shows that all masked reconstructions are in the first positions. Therefore, we may conclude that, as it is well known, applying a tight mask has a major impact comparing reconstructions. If we form two subgroups containing the masked and unmasked reconstructions we see that within each group the higher is the number of projections the better is the reconstruction. One of the obvious conclusions is that $\overline{FSC_i}$ is not a robust magnitude for sorting data sets in which mask and unmasked reconstruction are mixed together but works properly if all reconstruction have been masked with similar masks.

## Appendix B. Abbreviations and acronyms used in the main text

Dendrogram labels contain the name of the algorithms applied to the different 3D maps. Due to space limitations, in many cases it is not possible to use the full algorithm name and we have been forced to created an acronym. In this appendix we show a list of the used acronyms.

| Abbreviation | Full Name |
|---|---|
| alignP | Alignparts_lmbfgs followed by relion polish |
| c-Sparc | cryoSPARC |
| frameA | Frame alignment script provided by Direct Electron |
| motionC | motioncorr |
| motionCP | motioncorr followed by relion polish |
| na | not available |
| none | no information available |
| opticalF | motioncorr followed by optical flow |

| | |
|---|---|
| polish | relion polish |
| saffpm | Fast Projection Matching with Symmetry Adapted Functions |
| unblurP | unblur followed by relion polish |
| warpP | warp followed by relion polish |

## Appendix C. Values of the feature $\overline{FSC_{i,j}}$

In this Appendix we show the value of the feature $\overline{FSC_{i,j}}$ for each specimen and pair of reconstructions. These values have been used to compute the sorting and P-values.

### C.1. GroEL in Silico

**Table C.1**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen GroEL *in silico*.

| | emcd104 | emcd120 | emcd132 | emcd143 | emcd153 | emcd158 | emcd165 | emcd168 | emcd169 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| emcd104 | 0 | 5.7 | 6.2 | 6.6 | 4.2 | 4.3 | 5.9 | 4.4 | 5.3 | 42.6 |
| emcd120 | 5.7 | 0 | 6.2 | 6.1 | 4.5 | 4.3 | 5.9 | 4.5 | 5.3 | 42.5 |
| emcd132 | 6.2 | 6.2 | 0 | 6.7 | 4.7 | 4.8 | 6.3 | 4.9 | 6.3 | 46.1 |
| emcd143 | 6.6 | 6.1 | 6.7 | 0 | 5.1 | 4.6 | 7.7 | 5 | 5.85 | 47.65 |
| emcd153 | 4.2 | 4.5 | 4.7 | 5.1 | 0 | 3.7 | 5.05 | 4 | 4.4 | 35.65 |
| emcd158 | 4.3 | 4.3 | 4.8 | 4.6 | 3.7 | 0 | 4.5 | 3.7 | 5.1 | 35 |
| emcd165 | 5.9 | 5.9 | 6.3 | 7.7 | 5.05 | 4.5 | 0 | 4.8 | 5.6 | 45.75 |
| emcd168 | 4.4 | 4.5 | 4.9 | 5 | 4 | 3.7 | 4.8 | 0 | 5.5 | 36.8 |
| emcd169 | 5.3 | 5.3 | 6.3 | 5.85 | 4.4 | 5.1 | 5.6 | 5.5 | 0 | 43.35 |

### C.2. T20S proteasome

**Table C.2**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen T20S *Proteasome*.

| | emcd103 | emcd107 | emcd108 | emcd130 | emcd131 | emcd141 | emcd144 | emcd145 | emcd162 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| emcd103 | 0 | 16.3 | 16.16 | 11.92 | 11.92 | 15.52 | 14.05 | 14.31 | 14.54 | 114.72 |
| emcd107 | 16.3 | 0 | 16.82 | 11.74 | 11.74 | 15.85 | 13.59 | 13.82 | 14.69 | 114.55 |
| emcd108 | 16.16 | 16.82 | 0 | 11.63 | 11.6 | 16.06 | 13.77 | 13.97 | 14.93 | 114.94 |
| emcd130 | 11.92 | 11.74 | 11.63 | 0 | 20.39 | 11.66 | 11.74 | 11.78 | 10.93 | 101.79 |
| emcd131 | 11.92 | 11.74 | 11.63 | 20.39 | 0 | 11.66 | 11.74 | 11.78 | 10.93 | 101.79 |
| emcd141 | 15.52 | 15.85 | 16.06 | 11.66 | 11.66 | 0 | 13.92 | 14.1 | 14.19 | 112.96 |
| emcd144 | 14.05 | 13.59 | 13.77 | 11.74 | 11.74 | 13.92 | 0 | 19.38 | 13.07 | 111.26 |
| emcd145 | 14.31 | 13.82 | 13.97 | 11.78 | 11.78 | 14.1 | 19.38 | 0 | 13.21 | 112.35 |
| emcd162 | 14.54 | 14.69 | 14.93 | 10.93 | 10.93 | 14.19 | 13.07 | 13.21 | 0 | 106.49 |

### C.3. Apo-Ferritin

**Table C.3**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen Apo-Ferritin.

| | emcd112 | emcd118 | emcd121 | emcd122 | emcd124 | emcd147 | emcd155 | emcd166 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|---|
| emcd112 | 0 | 9.41 | 11.05 | 4.5 | 7.8 | 4 | 5.88 | 9.82 | 52.46 |
| emcd118 | 9.41 | 0 | 9.69 | 4.2 | 11.48 | 4.28 | 4.99 | 13.1 | 57.15 |
| emcd121 | 11.05 | 9.69 | 0 | 4.44 | 8.23 | 4.27 | 6 | 10.21 | 53.89 |
| emcd122 | 4.5 | 4.2 | 4.44 | 0 | 3.65 | 2.73 | 3.29 | 4.38 | 27.19 |
| emcd124 | 7.8 | 11.48 | 8.23 | 3.65 | 0 | 4.73 | 4.12 | 11.63 | 51.64 |
| emcd147 | 4 | 4.28 | 4.27 | 2.73 | 4.73 | 0 | 3.23 | 4.48 | 27.72 |
| emcd155 | 5.88 | 4.99 | 6 | 3.29 | 4.12 | 3.23 | 0 | 5.23 | 32.74 |
| emcd166 | 9.82 | 13.1 | 10.21 | 4.38 | 11.63 | 4.48 | 5.23 | 0 | 58.85 |

### C.4. TRPV1 channel

**Table C.4**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen TRPV1 Channel.

| | emcd101 | emcd115 | emcd133 | emcd135 | emcd156 | emcd161 | emcd163 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|
| emcd101 | 0 | 8.58 | 7.54 | 7.66 | 6.27 | 8.49 | 5.92 | 44.46 |
| emcd115 | 8.58 | 0 | 7.89 | 8.02 | 6.48 | 9.21 | 6.28 | 46.46 |
| emcd133 | 7.54 | 7.89 | 0 | 10.74 | 5.96 | 8.54 | 6.24 | 46.91 |
| emcd135 | 7.66 | 8.02 | 10.74 | 0 | 6.06 | 8.75 | 6.33 | 47.56 |
| emcd156 | 6.27 | 6.48 | 5.96 | 6.06 | 0 | 6.56 | 5.41 | 36.74 |
| emcd161 | 8.49 | 9.21 | 8.54 | 8.75 | 6.56 | 0 | 6.38 | 47.93 |
| emcd163 | 5.92 | 6.28 | 6.24 | 6.33 | 5.41 | 6.38 | 0 | 36.56 |

### C.5. 80S ribosome

**Table C.5**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen 80S Ribosome. In order to make the page fit in the page the name of the 3Dmaps has been shorten from the canonical form *emcd129* to *e129*.

| | e-111 | e-114 | e-119 | e-123 | e-125 | e-126 | e-127 | e-128 | e-129 | e-148 | e-149 | e-150 | e-151 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| emcd111 | 0 | 14.12 | 18.3 | 15.18 | 15.12 | 14.01 | 12.5 | 12.27 | 9.52 | 12.28 | 13.65 | 13.81 | 14.12 | 164.88 |
| emcd114 | 14.12 | 0 | 14.2 | 18.6 | 17.6 | 16.83 | 14.61 | 14.14 | 10.46 | 19.2 | 16.35 | 16.62 | 21.14 | 193.87 |
| emcd119 | 18.3 | 14.2 | 0 | 15.73 | 15.29 | 14.48 | 12.86 | 12.62 | 9.33 | 12.39 | 14.1 | 14.3 | 14.54 | 168.14 |
| emcd123 | 15.18 | 18.6 | 15.73 | 0 | 19.57 | 22.13 | 18.76 | 18.19 | 13.51 | 16.46 | 19.99 | 20.17 | 20.6 | 218.89 |
| emcd125 | 15.12 | 17.6 | 15.29 | 19.57 | 0 | 17.52 | 15.04 | 14.57 | 10.86 | 14.37 | 16.63 | 16.87 | 17.47 | 190.91 |
| emcd126 | 14.01 | 16.83 | 14.48 | 22.13 | 17.52 | 0 | 15.01 | 14.58 | 11.8 | 15.57 | 18.56 | 18.65 | 18.92 | 198.06 |
| emcd127 | 12.5 | 14.61 | 12.86 | 18.76 | 15.04 | 15.01 | 0 | 14.05 | 11.57 | 15 | 16.96 | 16.98 | 17.14 | 180.48 |
| emcd128 | 12.27 | 14.14 | 12.62 | 18.19 | 14.57 | 14.58 | 14.05 | 0 | 11.58 | 14.62 | 16.51 | 16.51 | 16.68 | 176.32 |
| emcd129 | 9.52 | 10.46 | 9.33 | 13.51 | 10.86 | 11.8 | 11.57 | 11.58 | 0 | 13.19 | 13.11 | 12.96 | 13.22 | 141.11 |
| emcd148 | 12.28 | 19.2 | 12.39 | 16.46 | 14.37 | 15.57 | 15 | 14.62 | 13.19 | 0 | 15.04 | 14.99 | 14.79 | 177.9 |
| emcd149 | 13.65 | 16.35 | 14.1 | 19.99 | 16.63 | 18.56 | 16.96 | 16.51 | 13.11 | 15.04 | 0 | 22.13 | 20.03 | 203.06 |
| emcd150 | 13.81 | 16.62 | 14.3 | 20.17 | 16.87 | 18.65 | 16.98 | 16.51 | 12.96 | 14.99 | 22.13 | 0 | 20.23 | 204.22 |
| emcd151 | 14.12 | 21.14 | 14.54 | 20.6 | 17.47 | 18.92 | 17.14 | 16.68 | 13.22 | 14.79 | 20.03 | 20.23 | 0 | 208.88 |

### C.6. Brome mosaic virus

**Table C.6**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen Brome Mosaic Virus.

| | emcd102 | emcd110 | emcd136 | emcd137 | emcd140 | emcd142 | emcd152 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|
| emcd102 | 0 | 12.87 | 13.03 | 15.74 | 18.05 | 17.32 | 10.1 | 87.11 |
| emcd110 | 12.87 | 0 | 10.75 | 12.48 | 12.64 | 12.62 | 8.56 | 69.92 |
| emcd136 | 13.03 | 10.75 | 0 | 17.71 | 14.08 | 15.86 | 8.46 | 79.89 |
| emcd137 | 15.74 | 12.48 | 17.71 | 0 | 16.59 | 17.98 | 9.83 | 90.33 |
| emcd140 | 18.05 | 12.64 | 14.08 | 16.59 | 0 | 18.71 | 9.62 | 89.69 |
| emcd142 | 17.32 | 12.62 | 15.86 | 17.98 | 18.71 | 0 | 9.55 | 92.04 |
| emcd152 | 10.1 | 8.56 | 8.46 | 9.83 | 9.62 | 9.55 | 0 | 56.12 |

### C.7. β-Galactosidase

**Table C.7**
$\overline{FSC_{i,j}}$ (columns: 2nd-next to last) and $\overline{FSC_i}$ (last column) for specimen *β-Galactosidase*.

| | emcd106 | emcd113 | emcd134 | emcd138 | emcd139 | emcd154 | emcd157 | emcd159 | emcd160 | emcd164 | emcd167 | $\overline{FSC_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| emcd106 | 0 | 10.47 | 10.54 | 10.63 | 10.7 | 8.77 | 9.48 | 6.66 | 9.49 | 6.66 | 6.36 | 89.76 |
| emcd113 | 10.47 | 0 | 9.18 | 10.47 | 10.24 | 9.43 | 8.72 | 6.5 | 8.79 | 6.5 | 6.27 | 86.57 |
| emcd134 | 10.54 | 9.18 | 0 | 11.61 | 11.64 | 7.98 | 8.37 | 6.82 | 8.78 | 6.82 | 6.46 | 88.2 |
| emcd138 | 10.63 | 10.47 | 11.61 | 0 | 14.73 | 8.4 | 8.77 | 6.64 | 9.22 | 6.64 | 6.33 | 93.44 |
| emcd139 | 10.7 | 10.24 | 11.64 | 14.73 | 0 | 8.33 | 8.75 | 6.65 | 9.16 | 6.65 | 6.32 | 93.17 |
| emcd154 | 8.77 | 9.43 | 7.98 | 8.4 | 8.33 | 0 | 7.72 | 6.04 | 7.87 | 6.04 | 5.83 | 76.41 |
| emcd157 | 9.48 | 8.72 | 8.37 | 8.77 | 8.75 | 7.72 | 0 | 6.34 | 9.22 | 6.34 | 6.06 | 79.77 |
| emcd159 | 6.66 | 6.5 | 6.82 | 6.64 | 6.65 | 6.04 | 6.34 | 0 | 6.99 | 20.23 | 18.7 | 91.57 |
| emcd160 | 9.49 | 8.79 | 8.78 | 9.22 | 9.16 | 7.87 | 9.22 | 6.99 | 0 | 6.99 | 6.67 | 83.18 |
| emcd164 | 6.66 | 6.5 | 6.82 | 6.64 | 6.65 | 6.04 | 6.34 | 20.23 | 6.99 | 0 | 18.63 | 91.5 |
| emcd167 | 6.36 | 6.27 | 6.46 | 6.33 | 6.32 | 5.83 | 6.06 | 18.7 | 6.67 | 18.63 | 0 | 87.63 |

## References


Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.O., Carazo, J.M., 2015. Alignment of direct detection device micrographs using a robust Optical Flow approach. J. Struct. Biol. 189 (3), 163–176.

Agirrezabala, X., Velazquez-Muriel, J.A., Gomez-Puertas, P., Scheres, S.H., Carazo, J.M., Carrascosa, J.L., 2007. Quasi-atomic model of bacteriophage t7 procapsid shell: insights into the structure and evolution of a basic fold. Structure 15 (4), 461–472.

Estrozi, L.F., Navaza, J., 2010. Ab initio high-resolution single-particle 3D reconstructions: the symmetry adapted functions way. J. Struct. Biol. 172 (3), 253–260.

Grant, T., Grigorieff, N., 2015. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. Elife 4, e06980.

Grigorieff, N., 2007. FREALIGN: high-resolution refinement of single particle structures. J. Struct. Biol. 157 (1), 117–125.

Guo, F., Jiang, W., 2014. Single particle cryo-electron microscopy and 3-D reconstruction of viruses. Methods Mol. Biol. 1117, 401–443.

Heymann, J.B., 2001. Bsoft: image and molecular processing in electron microscopy. J. Struct. Biol. 133 (2–3), 156–169.

Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., 1996. Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. 116 (1), 71–76.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., Cheng, Y., 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods 10 (6), 584–590.

Marabini, R., Carragher, B., Chen, S., Chen, J., Cheng, A., Downing, K.H., Frank, J., Grassucci, R.A., Bernard Heymann, J., Jiang, W., Jonic, S., Liao, H.Y., Ludtke, S.J., Patwari, S., Piotrowski, A.L., Quintana, A., Sorzano, C.O., Stahlberg, H., Vargas, J., Voss, N.R., Chiu, W., Carazo, J.M., 2015. CTF Challenge: result summary. J. Struct. Biol. 190 (3), 348–359.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53 (Pt 3), 240–255.

Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14 (3), 290–296.

Scheres, S.H., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180 (3), 519–530.

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., Frank, J., 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat Protoc 3 (12), 1941–1974.

Spear, J.M., Noble, A.J., Xie, Q., Sousa, D.R., Chapman, M.S., Stagg, S.M., 2015. The influence of frame alignment with dose compensation on the quality of single particle reconstructions. J. Struct. Biol. 192 (2), 196–203.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., Ludtke, S.J., 2007. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157 (1), 38–46.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bull. 1 (6), 80–83.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., Agard, D.A., 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14 (4), 331–332.