Work Package 7, Milestone 28: The EM Quality Assurance workflow defined

Progress in single particle cryo-EM, most recently due to the introduction of direct detector devices, has made the high-resolution structure determination of biological assemblies smaller than 500 kDa more routine, but has also increased attention on the need for tools to demonstrate the validity of single particle maps. Although map validation is a continuing subject of research, some consensus has been reached on procedures that reduce model bias and over-fitting during map refinement as well as specific tests that demonstrate map validity. Tilt-pair analysis may be used as a method for demonstrating the consistency at low resolution of a map with image data. For higher-resolution maps, new procedures for more robust resolution assessment and for validating the refinement of atomic coordinate models into single particle maps have been developed (from http://dx.doi.org/10.1016/j.sbi.2015.07.002).

Workflow[2] definition

We propose a workflow to test how reliable is a cryoEM map compared to the images of the macromolecules that were used to generate it (referred to as "particles" [3]). We are using 2 methods [1] in the proposed workflow, referred to as "validate against overfitting" (1) and "alignment reliability" (2). The rationale behind providing two different methods in the workflow is to add an extra level of confidence in the consistency between the particles and the map obtained. In both cases, result are necessary but not sufficient, in the sense that a bad result in any of them will tell that the reconstruction is not very consistent. On the contrary good results do not imply that the reconstruction is good. The two of them make use of a set of aligned particles and assess the consistency between particles and maps. The overall workflow will look like the one in Figure 1:
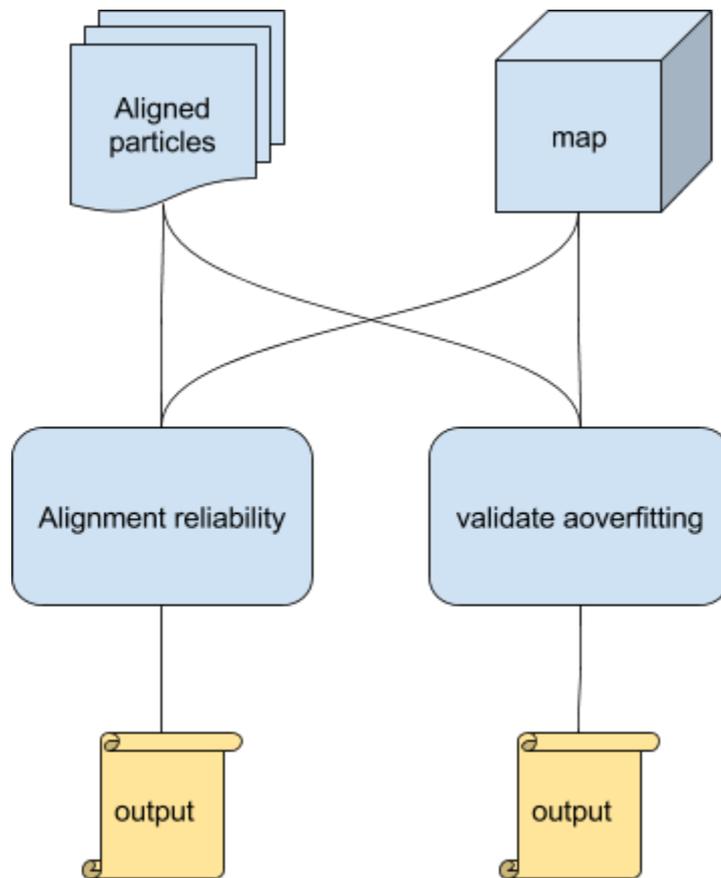
Figure 1. Reliability workflow diagram

Input data: aligned particles [5].

Set of aligned particles used to reconstruct the map. Ideally the whole set used in the experiment is the best option, but a subset can be used to reduce processing time. Valid for both methods.

Input data: map[4]

Final map reconstruction. Valid for both methods

Alignment reliability :

Provides a statistical analysis, without using tilt-pairs, with the capability to provide objective information about the consistency between the reconstructed 3D map and a set of 2D projections (or 2D classes) used in the map reconstruction process (2).

Parameters definition in the proposed implementation

Parameter	Description
Symmetry [7]	Symmetry of the volume
Image alignment method	Either significant or projection matching
Resolution to filter	Range of resolutions to be filtered
Angular sampling	Angular distance (in degrees) between neighboring projection points
Significance	Significance of the alignability with respect to a set of uniformly distributed random points

Output

The alignment reliability method has two outputs. One is assigned to each of the image particles, and the other is an ensemble quality measure. The former accounts for the cluster tendency of each image orientation with respect to the map (as described in detail in Vargas et al., 2016), and the other is an integral over these individual measures, providing a global reliability index.
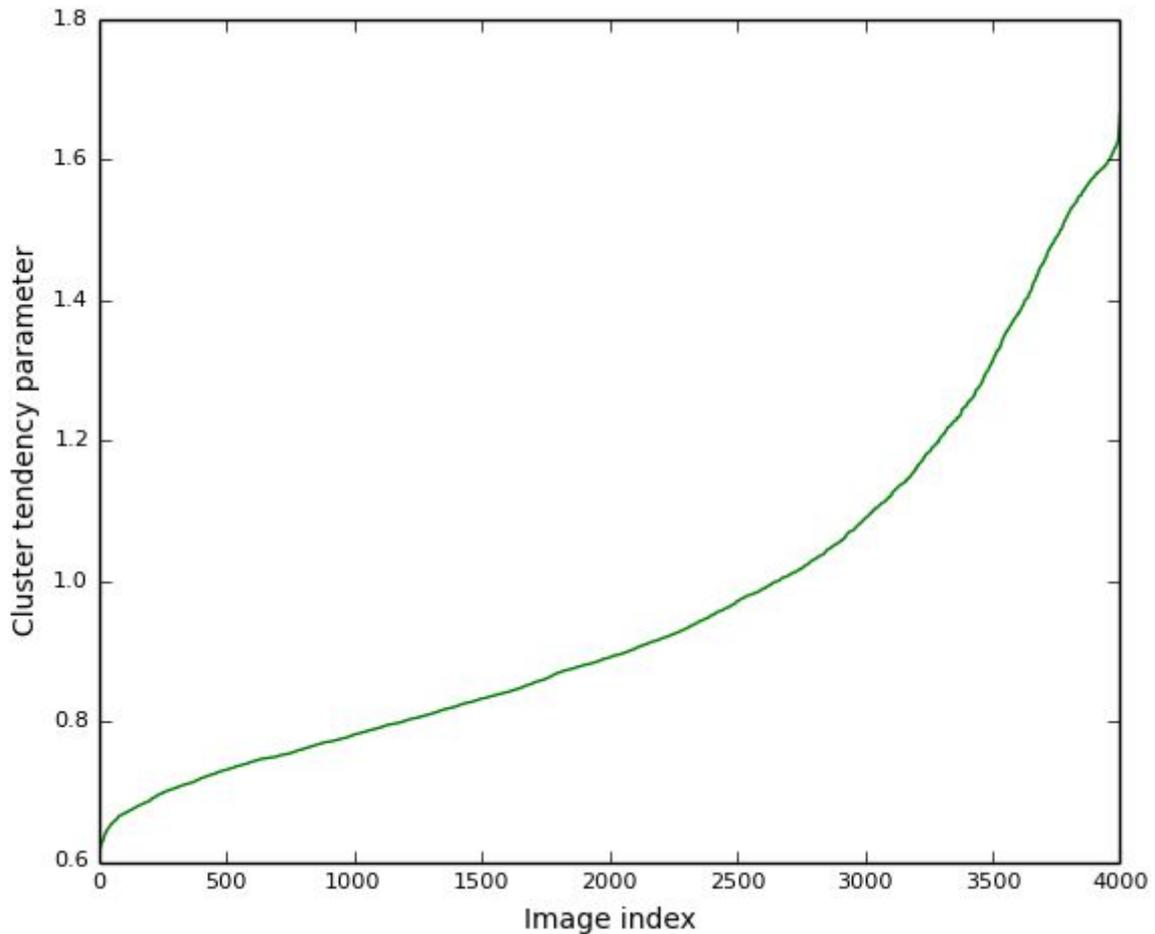
Validate against overfitting :

Given a map from a current project or from the EMDB and a small subset of images used for it, the images are aligned to the map without masking or resolution limits, and compared to a reconstruction [6] from an equivalent number of noise images aligned to the same map. The expectation is that the reconstructions from micrograph images should give better resolutions than noise-derived reconstructions
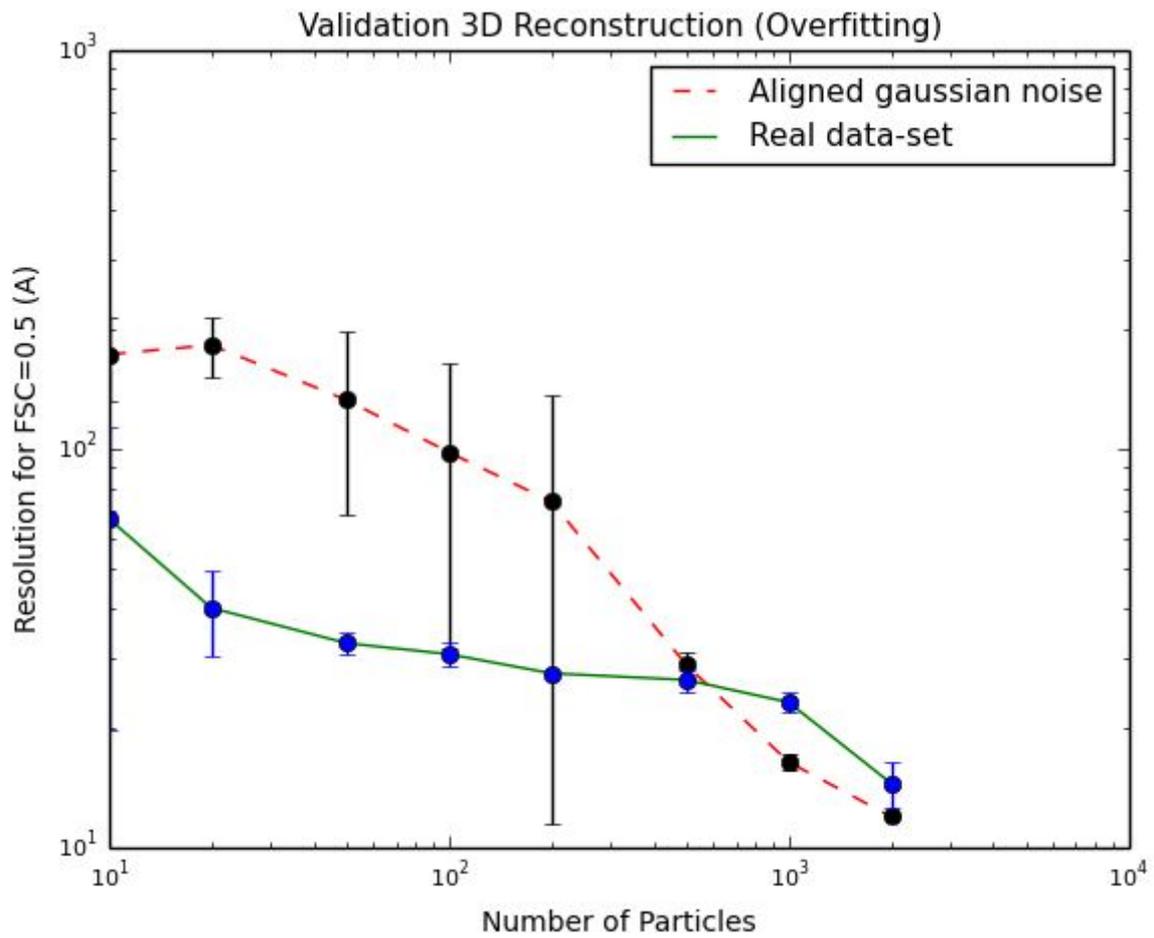
Parameters definition

Parameter	Description
Symmetry [7]	Symmetry of the volume
Number of sets and size	List with the size of each set, progressively increasing values: 10 20 50 …
Angular sampling	Angular distance (in degrees) between neighboring projection points

Milestone M28

Chart as described in Heymann, 2015 (1).

Milestone M28

Ontologies annotations
Ontologies are referenced with a number surrounded by brackets.

REF	Name	Description	Candidate ontologies
1	process, node	Each of the parts of a workflow…	PROV-O: p-step
2	Workflow	Execution plan	PROV-O:p-plan
3	particles	data type	
4	map	data type	
5	aligned particles	attribute of a set of particles meaning particles are aligned.	
6	volume reconstruction	process of reconstruction of a volume	
7	symmetry	Symmetry of a volume	and all its possible values

References
(1) Heymann, 2015
(2) Vargas et al., Sci. Rep., 2016