19.10. SINGLE-PARTICLE RECONSTRUCTION WITH *EMAN*

19.10.3. Single-particle reconstruction

The single-particle reconstruction process is broken down into a sequence of subtasks in the *EMAN2* workflow. Each subtask is documented in the GUI, with a block of text and mouse-over popups. While the workflow is the method of choice for performing reconstructions, all of the individual subtasks can also be completed using direct command-line programs. The workflow is launched with the `e2workflow.py` program. For each stage, the alternative command-line program is also shown in parentheses.

19.10.3.1. Particle selection (`e2boxer.py`)

Locating images of individual particles in the micrograph/charge-coupled device (CCD) frame is arguably the most critical and labour-intensive step in single-particle processing. Owing to the high noise levels in typical cryo-EM images, combined with various types of contamination, correctly identifying all relevant particles within a micrograph is challenging (Zhu *et al.*, 2004). Despite many years of effort, no reliable and generally applicable algorithms for fully automated particle picking exist. The present approach of *EMAN2* is to streamline the process of semi-automated particle picking. Multiple CCD frames/micrographs can be opened simultaneously and are handled by the user as a group. At the time of writing, there is a single autopicking algorithm available, based on Woolford *et al.* (2007), but there are plans to expand this in the future.

19.10.3.2. Contrast transfer function/image evaluation (`e2ctf.py`)

EMAN1 contained a very accurate contrast transfer function (CTF) correction methodology, which was also very labour-intensive. While the mathematical basis for correction remains effectively unchanged, the correction methodology is now entirely automatic. The only substantial change to the mathematical formulation is that the spectroscopic profile of the background noise, which was originally characterized using four empirical parameters, is now specified as a data-derived non-parametric curve. In addition, this methodology permits the computation of a one-dimensional structure factor from sets of images without user intervention, a process which was quite challenging in *EMAN1* (Ludtke *et al.*, 1999). The automated process is supplemented by a fully featured GUI for manual image assessment.

19.10.3.3. Grouping particles into sets

Once images have been preprocessed, the user is given another, optional, opportunity to examine the images of individual particles, now in Wiener-filtered form, to identify any 'bad' particles which should be eliminated from further processing. In addition, several values are provided to the user which permit assessment of particles on a per-micrograph basis, such as the integrated signal-to-noise ratio, defocus and *B* factor. Particles from user-selected micrographs are then combined into a set for further processing.

19.10.3.4. Reference-free two-dimensional classification (`e2refine2d.py`)

This fully automated principal-component-analysis-based (Frank, 2006) process produces two-dimensional averages representative of the various particle views present in the raw images. As with the workflow as a whole, while the user can provide a large number of different options for this subtask, sensible default values are provided, which should produce good results in most cases. The two-dimensional classification process is quite similar to the corresponding process in *EMAN1* (Chen *et al.*, 2006), with subtle improvements to increase speed and accuracy.

The two-dimensional classification process generates class averages with improved signal-to-noise, characteristic of the different orientations and dynamic states of the particles. These can be used to help identify the particle's symmetry and any problems with preferred orientation or particle flexibility/heterogeneity.

19.10.3.5. Initial model generation (`e2initialmodel.py`)

There are a variety of opinions in the single-particle reconstruction community about how initial models should be generated. Briefly, the most widely used methods are: common lines with reference-free class averages (van Heel, 1987), random conical tilt (Radermacher *et al.*, 1987), orthogonal tilt (Leschziner & Nogales, 2006) and single-particle tomography with three-dimensional averaging (Walz *et al.*, 1997). While not the primary method we promote for traditional single-particle processing, the *EMAN2* workflow includes a task for three-dimensional alignment and averaging of tomographic subvolumes.

The primary method provided to generate initial models in *EMAN2* is based on the robustness of the iterative refinement strategy used for the final high-resolution refinement (Section 19.10.3.6). For any given three-dimensional structure, there are a small number of stable structures which can result from refinement, regardless of starting model. The correct model represents the global minimum in the energy space defined by requiring that projections of the reconstruction match the particle images.

To make use of this concept in *EMAN2*, we downsample the two-dimensional class averages (Section 19.10.3.4) and make a large number of completely random starting models, then quickly iteratively refine each. While the speedups used in this process make it more susceptible to local minima than the full high-resolution refinement method, it will generally still produce the correct model some fraction of the time, without having to resort to additional experiments. The resulting models are easily assessed by comparing the original class averages with projections of the model, so the best answer can be readily identified. The primary risk associated with this process is with respect to particles with structural heterogeneity in solution or those with strongly preferred orientations, in which case the standard refinement algorithm is not robust. In such cases, the tomographic method can provide further insights.

19.10.3.6. Refinement (`e2refine.py`)

Once particles have been prepared and an initial model has been produced, the next step is three-dimensional refinement. This process is largely unchanged from *EMAN1* (Ludtke *et al.*, 1999), apart from minor improvements designed to speed up the process and provide additional flexibility. As with two-dimensional refinement, there are a large number of options for the user to specify, but, where possible, sensible defaults are provided. Documentation for the various parameters is provided through the workflow interface.

19.10.4. Evaluating the reconstruction

Evaluating the final three-dimensional reconstruction remains one of the most problematic aspects of single-particle processing from a quantitative standpoint, and there have been numerous

references