

Chapter 19.10. Single-particle reconstruction with *EMAN*

S. LUDTKE

19.10.1. Introduction

Single-particle reconstruction is a technique for determining the three-dimensional structure of identical nanoscale objects without requiring crystallization (Frank, 2006). Typical targets are large proteins or macromolecular assemblies, with ~ 200 kDa regarded as the lower size limit for this technique due to high image noise levels. Typical non-viral targets are in the size range from 500 kDa to 5 MDa, with icosahedral viruses extending to hundreds of megadaltons. Several examples of structures being reconstructed to ~ 4 Å resolution using this technique have been published (Ludtke *et al.*, 2008; Jiang *et al.*, 2008; Zhang *et al.*, 2008; Yu *et al.*, 2008) and subnanometre resolution is frequently achieved, but in negative stain, or with unusually difficult specimens, resolutions may still be limited to 12–30 Å.

19.10.2. Overview of *EMAN*

EMAN is a suite of scientific image-processing tools with a particular focus on cryogenic electron microscopy (cryo-EM) and single-particle reconstruction. The first stable version was released in 1999 and it has undergone continuous development since that time. Since roughly 2005, development has focused on *EMAN2*, with a completely redesigned modular image-processing library, a workflow system, and a new OpenGL-based graphical user interface (GUI). *EMAN2* supports virtually all image file formats used in the cryo-EM community, as well as various conventions for particle-orientation specification. The image-processing library is completely modular, so new algorithms can be made available throughout the system without any modification of end-user programs. Metadata, which refer to both experimental and derived information describing the image data, are archived using an embedded database system. These metadata can be harvested to produce information for manuscripts or for deposition to centralized databases such as the EMDB (<http://www.emdatabank.org>), operated by the PDBe (Europe) and RCSB PDB (USA).

EMAN2 has a tiered structure. The GUI interface and workflow represent the top of the hierarchy. Next are individual command-line programs, providing the user with a greater level of control and a wide range of utility functions. Below this is the Python (<http://www.python.org>) layer, for interactive use or high-level programming. Finally, we have the C++ core image-processing library, which handles all computationally intensive image processing. Full documentation is provided through the *EMAN* wiki (<http://blake.bcm.tmc.edu/eman/eman2>).

19.10.2.1. GUI layer and workflow

The workflow interface (Fig. 19.10.2.1) is designed to guide the end user through entire processes, while storing a complete record of all processing completed during any given task. While this interface is gradually being expanded with new tasks, such as tomographic particle averaging, for the purposes of this discussion we will focus on the single-particle reconstruction task, which is broken down into a number of discrete subtasks as discussed in detail in Section 19.10.3. The workflow is designed so

that users wishing to perform only a subset of any task can easily import data into any subtask, proceed through additional subtasks, then export the results at any point. There are future plans to integrate other cryoEM software into the workflow interface to enable users to try alternative algorithms for specific subtasks. The workflow is accessed using either `e2workf1ow.py` or `e2desktop.py`.

The GUI interface (Fig. 19.10.2.2) includes a number of visualization modules for displaying individual two-dimensional images, stacks of two-dimensional images, three-dimensional volumetric models, two-dimensional plots, three-dimensional plots and symmetry information. These modules are used by the interactive file/database browser available in the workflow or *via* the program `e2display.py`, but may also be used directly from the Python interface.

19.10.2.2. Command-line programs

EMAN2 (Tang *et al.*, 2007) currently contains ~ 60 command-line programs for specific tasks, plus an additional ~ 60 programs as part of the related *SPARX* project (Hohn *et al.*, 2007), included with *EMAN2* releases. High-level programs such as `e2refine2d.py`, `e2initialmodel.py`, `e2eotest.py` and `e2refine.py` are essentially wrapper scripts which call the lower-level Python scripts to complete specific tasks.

In addition, there are a number of general-purpose utility programs, such as `e2proc2d.py` for two-dimensional image processing, `e2proc3d.py` for three-dimensional image processing and `e2bdb.py` for access to the embedded database. Both `e2proc2d.py` and `e2proc3d.py` take similar arguments and



Figure 19.10.2.1 The overall workflow interface, and the dialogue for the three-dimensional refinement subtask.