Picking particles faster than one at a time

December 1 2005

Computer scientists and biologists at the Department of Energy's Lawrence Berkeley National Laboratory have developed software that can select tens of thousands of high-quality images of biological molecules from electron microgaphs, rapidly and automatically, with accuracy approaching that of experienced human analysts.

The new algorithm, described as "particle picking by segmentation," promises to greatly increase the speed and power of methods for determining biological structures at high resolution, based on data from electron microscopy. The researchers report their results in the forthcoming issue of the *Journal of Structural Biology* in an article now available to subscribers online.

When what's needed is a high-resolution structure of a large and complicated biological molecule -- a ribosome, say, which combines protein and RNA, or a membrane protein that readily falls apart in water and is hard to crystallize -- biologists often turn to cryo-electron microscopy (cryo-EM) to perform single-particle reconstruction.

Understanding structure is often the key to devising antibiotics and other therapies that can interfere with unwanted biological activity -- for example, the ability of infectious bacteria to synthesize proteins can be wrecked by jamming their ribosomes, if the ribosome structure is known in detail. Single-particle reconstruction with cryo-EM holds the promise of providing many high-resolution structures which may be difficult or impossible to obtain otherwise.
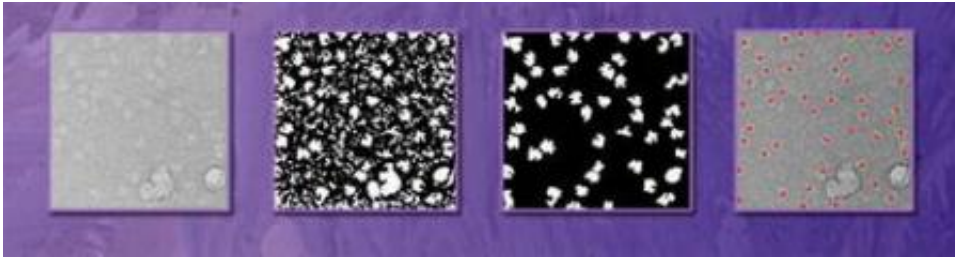
Image: Some of the steps in the process of particle picking by segmentation: on the left, the original low-contrast micrograph of ribosomes; second, with background subtracted and contrast greatly enhanced; third, segmentation has retained only particles of the right size and brightness; finally, the chosen individual ribosome images have been marked (red spots) on the original micrograph.

Instead of trying to coax molecules to arrange themselves in a repeating crystalline structure, as is necessary for x-ray crystallography, cryo-EM uses individual molecules frozen in random orientations. Capturing two-dimensional images of the molecule from many different angles allows powerful computers to recreate the structure in three dimensions, a process molecular biologist Robert Glaeser of Berkeley Lab's Physical Biosciences and Life Sciences Divisions, who is also a professor of biochemistry and molecular biology at the University of California at Berkeley, calls "crystallization in silico."

"In theory, you need twice as many particles as the molecular weight of what you want to image," explains Umesh Adiga, a member of Glaeser's laboratory and a staff scientist in the Physical Biosciences Division. Molecular weight roughly corresponds to the number of atoms in the molecule. "So for a molecule with half a million atoms, you need a

million particle images -- thousands for each orientation."

These must be chosen from many millions of candidates, and each must show the whole particle and nothing but the particle. A typical micrograph may show fifteen hundred or more particles, but picking them out isn't easy. The microscope's electron beam has to be kept at low power to prevent radiation damage, so the signal-to-noise ratio is low and the particles are barely perceptible shapes in a field of gray.

"It's hard to find good candidates even with an expert eye," says Adiga. "Having to choose hundreds of thousands of particles is a bottleneck in the process of single-particle reconstruction."

Automatic particle-picking methods have been devised to meet this challenge, but until now even the best yield more than 30 percent false positives -- either poor-quality images of particles or something else altogether, like debris or background noise. Therefore "a human still has to go through them and pick out the good ones," Adiga says.

Adiga and his colleagues decided that concentrating too much attention on the particle itself in the early stages of picking -- for example, approximating its shape and creating a template into which real images are forced to fit, a process common to all previous automatic methods -- simply added to the difficulty. "We decided that if there's noise, there's noise, so at first let's not deal with the particle but with the noise," he says. "If the particle is the foreground, we deal with the background."

By first establishing the average gray-scale range of the particles of interest, contrast can be maintained while the fine texture of the background is smoothed out. The smoothed-out background is then subtracted.

The next steps involve a procedure called segmentation, developed by

Adiga and his colleagues. After the background is subtracted, the micrograph is rendered in high contrast. Only shapes of a certain size and brightness are retained; all the rest are thrown away in a step called binarization, or thresholding. "You need not know how the particle looks before you set out to pick good images of it, only how big it is," says Adiga.

The thresholding procedure is iterative, but eventually the processed high-contrast particle images can be matched unambiguously with their originals in the more highly detailed, low-contrast micrograph. Some images may still remain problematic -- for example, some particles may be so close together they appear to be touching; in these cases, an additional procedure called "pinch-off" separates candidates that aren't actually connected and discards those that are. Boxes are drawn around the final picks and their image quality is enhanced by an operation called "shrink-wrapping."

If a portion of an adjacent particle protrudes into the box, it is automatically discarded and replaced with a pattern textured like the rest of the background. At this end stage of the procedure -- although not at the beginning -- it may be advantageous to use templates (which include shape information about the particle) to refine identifications.

Scores of micrographs are needed to supply the hundreds of thousands of particles in a typical large-molecule reconstruction, but a program user needs to set parameters like particle size and gray-scale range only once, on a single micrograph. Thereafter the program runs on its own, sorting through each micrograph in about ten minutes.

Adiga and his colleagues tested the new algorithm by using it to pick images from among over 130,000 ribosome particles in 55 micrographs provided by the Wadsworth Center of the New York State Department of Health in Albany. Adiga separately inspected the 55 micrographs by

eye and "manually" selected particles, well over 80 percent of which turned out to be the same as those picked by the program. Fewer than 10 percent of the images chosen by the program were false positives.

A coauthor of the paper, William Baxter, independently inspected 14 of the same micrographs, chosen at random. On his first pass, intending to select only particles of the highest quality -- a "gold standard" -- he chose roughly two-thirds of the same particles picked by the software. When the program's additional candidates were inspected more closely, however, many turned out to be true positives of good quality; only about 10 percent of the program's picks were false positives.

Similar results were obtained when the segmentation program was used to pick particles from a smaller and more difficult molecule, a convex or "boat-shaped" enzyme labeled TPP-II, isolated from the fruit-fly. Although an initial comparison between manual selection and automatic selection indicated that 15 percent of the program's nominations were false positives, when the program was run again -- using a template after segmentation to filter out incompatible shapes -- false positives dropped to a mere 7 percent.

Beyond the demonstrated goal of selecting the same particles an expert would select with a low error rate, future refinement of the segmentation algorithm aims higher. By concentrating on the highest quality particles, crystallization in silico may need far fewer than hundreds of thousands of particles.

"Jacqueline Milne of the National Cancer Institute has demonstrated that high-quality structural maps can be achieved with a few hundred particles or less -- better than those using tens of thousands of particles -- provided the picks are good enough," Adiga says. "'Good enough' is a completely qualitative term, unfortunately, but if we can define it so that image-processing software makes only the best choices, we will have a

powerful new tools for biology."

Adiga says, "Particle-picking algorithms are a small part of a larger goal of mapping the healthy constituents of cells against diseased cells, from cellular organelles right down to interactions among atoms in a protein." Together with work initiated by Adiga and his colleagues in confocal image analysis, electron microscopy of cell sections, and electron tomographic image analysis, he says that being able to model the whole range of morphological and functional changes in cellular constituents, from the microscale to the nanoscale, comes ever closer to reality.

Source: Lawrence Berkeley National Laboratory