Microsc. Microanal. 8 (Suppl. 2), 2002
DOI 10.1017.S1431927602100353
© Microscopy Society of America 2002

Challenges in the Automation of Cryo-microscopy of Macromolecular Structure

C.S. Potter, D. Fellmann, R. Milligan, J. Pulokas, C. Suloway, Y. Zhu and B. Carragher

Department of Cell Biology, The Scripps Research Institute, La Jolla, CA, 92037

Although the methodology of molecular microscopy has enormous potential, it is time consuming and labor intensive. The techniques required to produce a three dimensional (3D) electron density map of a macromolecular structure normally require manual operation of the electron microscope by a skilled operator and manual supervision of the sometimes complex software needed for analysis and calculation of 3D maps. Typically it will take weeks to months to collect and analyze a dataset in order to reconstruct a map at 10-20Å resolution. It is generally agreed that in order to increase the resolution to a range where secondary structure will be discernable in the map (~7Å) will require an order of magnitude increase in the amount of data collected and analyzed. It is clear that this will only be practical as a mainstream technique if we can automate the imaging and analysis processes and greatly improve the overall throughput.

We will report on our system, called Leginon [1], which provides a high level of automation by locating a specimen on the grid using a multi-scale image analysis strategy. During a 24 hour experiment, Leginon can typically collect 500-1000 image defocus pairs without the necessity of operator intervention. Further automation is achieved by integrating Leginon with automated filament selection algorithms and a helical reconstruction software package so that a map is computed in parallel with data collection. We have shown the feasibility of this approach by automatically calculating an electron density map of tobacco mosaic virus to a resolution of ~10Å within 20 hours of inserting the grid in the microscope. The map is shown in figure 1. To produce the map, 167 grid squares were analyzed and 411 high magnification image pairs were collected. From these images, 725 TMV filaments were extracted and 100 of these were used in the final map. We estimate that approximately 130,000 subunits (particles) from the 100 filaments were used in constructing the average. We propose that the throughput of a data collection system be defined by (i) the *rate* at which data can be collected (e.g. particles/hour), (ii) *yield* - the fraction of the data that is useful and (iii) *sustainability* - a measure of the duration over which data can be continuously acquired. For our TMV experiment, data was acquired at a rate of approximately 48,000 particles/hour, with an overall yield of 14%, sustained through 20 hours of continuous operation. The total throughput was 130,000 particles for the entire experiment. Since minimal operator intervention is required for collecting the data and creating the map, we could feasibly perform 3-5 such experiments/week or collect approximately 0.5 million particles/week.

One of the principal factors that currently limits data throughput is the geometry and resolution of the CCD cameras that are available. We collect data using a 2Kx2Kx12bit CCD camera that has a bin size of 24x24μm^2. At the voltages that we typically use to collect images (120kV) the camera performs satisfactorily at about 1/3 of the Nyquist frequency, i.e. if the pixel size is 2Å at the specimen, we are able to collect data to a resolution of about 6Å. To achieve this resolution we need to work at a magnification of about 88,000x. In contrast a piece of film can be digitized to yield approximately 7Kx7K pixels at an equivalent resolution. Thus there is a factor of ~12 between the effective collection area of film vs. the 2Kx2K CCD camera when acquiring a single image. Since CCD cameras in cryoEM have been mostly used for manual data collection, this factor of 12 has

https://doi.org/10.1017/S1431927602100353 Published online by Cambridge University Press

clearly had a major dampening effect on the enthusiasm of the community for CCD's and they have consequently have not been generally adopted for data collection in spite of the advantage they offer in negating the need to develop and scan films.

If the TMV experiment we describe above had been performed using film and traditional manual data collection methods, an experienced microscopist could have collected perhaps 3 cassettes of film (~150 micrographs) in a reasonable day's work (~8 hours). Given the factor of 12 in effective collection area for film over CCD's, the corresponding data collection rate would be ~576,000 particles/hour. If we assume the same overall yield of 14%, the total throughput for the 8 hour data collection is on the order of 650,000 particles/experiment. However at the end of the data collection session, the films must be developed and scanned and then processed further. Given the time consuming nature of these tasks it might be reasonable to expect that one such experiment of this type could be performed each week on a sustained basis by a single investigator, which would make the overall throughput of manual and automated data collection very comparable (~0.5 million particles/week).

Our challenge is to substantially improve both the sustained throughput and the yield of automated data collection. We will report on the current performance and status of the Leginon system.

[1] Carragher, B., et al. (2000) Journal of Structural Biology, 132, 33-45.
[2] This work was supported by the National Science Foundation (DBI-0296063) and the National Institutes of Health (GM61939).

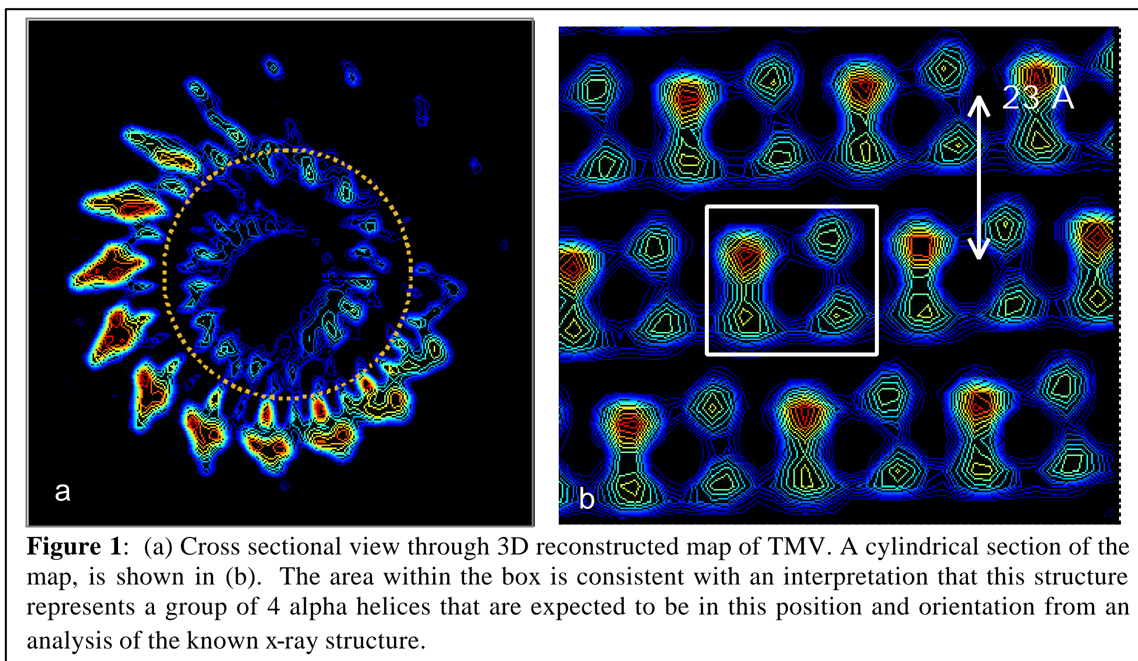

Figure 1: (a) Cross sectional view through 3D reconstructed map of TMV. A cylindrical section of the map, is shown in (b). The area within the box is consistent with an interpretation that this structure represents a group of 4 alpha helices that are expected to be in this position and orientation from an analysis of the known x-ray structure.

https://doi.org/10.1017/S1431927602100353 Published online by Cambridge University Press