Using Scipion for Stream Image Processing at CryoEM Facilities

J.M. de la Rosa-Trevín(c), J. Gómez-Blanco(b), P. Conesa(b), Y. Rancel(b), L. del Cano(b), R. Arranz(b), F.J. Chichón(b), J. Martín-Benito(b), M. Carroni(c), S. Fleichman(c), E. Lindahl(c), M. Basham(e), D.K. Clare (e), K. Savage(e), C.A. Siebert(e), A.W, Ashton(e), R. Marabini (a), C.O.S. Sorzano(b), J.M. Carazo (b)

(a) Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain.
(b) Biocomputing Unit, National Center for Biotechnology (CSIC), C/ Darwin, 3, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain.
(c) Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden
(e) Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, England

Three dimensional electron microscopy (3DEM) is becoming a very data-intensive field in which vast amounts of experimental images are acquired at high speed. To manage such large-scale projects, we have previously developed a modular workflow system called Scipion (de la Rosa-Trevín et al., 2016). We present here a major extension of Scipion that allows stream processing of EM images while the data is being acquired. This approach helps to detect problems at early stages, saves computing time and provides users with an in deep evaluation of the data quality before the acquisition is finished. Different EM facilities have implemented their stream processing solution using custom-made scripts, but this script based approach is not flexible and present a slow response time to new feature requests. The advantages of our proposed solution are: (1) flexibility: users can customize their workflows from a large collection of preloaded algorithms; (2) repeatability and traceability: workflows are stored and can be re-executed later, either with the same data or a different data set; (3) monitorization: there are utilities that constantly check how the execution of the algorithms is going and inform about possible anomalies; (4) reports: we have developed several graphical interfaces that are updated periodically producing a summary (e.g, CTF defocus values, system load, etc). This summary may be generated in HTML format and published though a public web-site to provide access for external users; (5) backup: both local and remote backup of the data and image processing is provided. At present, Scipion has been installed and is in production mode in seven Cryo-EM facilities and it is being considered in many more.

Keywords: electron microscopy

References
de la Rosa-Trevín, J., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C., Carazo, J., 2016. Scipion: A software framework toward integration, reproducibility and validation in 3d electron microscopy 195 (1), 93-99.