

## **XMIPP 3.0: A SPECIALIZED SUITE FOR IMAGE PROCESSING IN MICROSCOPY**

**R. Marabini<sup>2</sup>, J.M. Rosa-Trevin<sup>1</sup>, J. Cuenca<sup>1</sup>, A.E. Rey<sup>2</sup>, J. Vega<sup>1</sup>, J.Oton<sup>1</sup>, C. Sorzano-Sanchez<sup>1</sup> and J.M.Carazo<sup>1</sup>**

1 Instruct Image Processing Center-Biocomputing Unit, Centro Nacional de Biotecnología CNB-CSIC, Campus Universidad Autónoma, 28049 Madrid.

2 Escuela Politécnica Superior | Universidad Autónoma de Madrid | Francisco Tomás y Valiente, 11 | 28049 Madrid

*e-mail: roberto.marabini@uam.es*

Xmipp is a specialized suite of image processing programs, primarily aimed at obtaining the 3D reconstruction of biological specimens from large sets of projection images acquired by transmission electron microscopy. It is a comprehensive suite of image processing algorithms with a strong emphasis in the analysis of single particles although it is extending towards electron tomography and X-ray tomography. Xmipp accumulates more 226 citations from JCR journal papers, has 752 different registered users and has been downloaded more than 1900 times since 2008 (an average of 1.6 downloads/day; all these downloads have been started manually so that this number does not include Internet robots).

This package was introduced more than a decade ago [1], written in C and based on the X-Windows graphical interface. In 2004 a big re-implementation (completely done in C++) of the package was done [2], improving the hierarchical structure of well-documented data structures and functions. In 2008 a series of protocols were introduced using Python-Tkinter guiding the user in the reconstruction work-flow [Scheres2008]. A particular emphasis was placed on the treatment of structurally heterogeneous data through maximum-likelihood refinements and self-organizing maps. These protocols represented a major standardization of numerous existing scripts and recipes that circulated among the Xmipp user community, thereby representing years of experience by multiple researchers and providing novice users with a convenient tool to quickly obtain useful results with minimum efforts in learning about the details of the package.

Xmipp has proved to be useful for the EM-community, reflected in many structural studies over the last decade. Nevertheless, to obtain better 3D structures, the reconstruction process is demanding more computer power each day. Also the exchange of data between different packages and the standardization of procedures are becoming very important. Another important goal of most packages in the field is to develop more intuitive GUI's to reduce the learning curve and improve the usability of the command line programs. With these thoughts in mind and looking to the future, the last year a big re-organization of the package was started.

Next version of Xmipp is format-independent and support for reading/writing several image (see [http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/Image\\_formats](http://xmipp.cnb.csic.es/twiki/bin/view/Xmipp/Image_formats)). The image-handling functions have been modularized and images/volumes can be mapped on disk,

allowing working with big images files even on computers with less RAM memory and faster processing. A large effort has been dedicated to have flexible input/output files and developing functions and tools to work with them. The internal implementation has been done using a SQL engine, but the files are read/written in the STAR format, to be more user-friendly. This allows the exchange data in a seamless way. The programs have been arranged hierarchically, sharing some common behaviour. The program input is defined in a more semantic way, allowing automatic validation of command line arguments and also the generation of a simple GUI for each program. Special attention has also been put in developing some functions to ease the implementation of new parallel programs using threads and MPI library. Libraries bindings for popular programming languages like Java and Python have been added for providing Xmipp functions for those languages. Also a group of new GUIs tools are in development using Java and ImageJ libraries that will.

Although the main focus of Xmipp has traditionally been the analysis of single particles, Xmipp has extensive image processing capabilities applicable to other imaging modalities, such as electron tomography. Additionally, It also has some features typical of electron tomography, such as a new fiducial-free alignment of single axis tilt series [Sorzano2009] as well as 3D reconstruction algorithms that explicit mask “interfering regions” (like gold beads). Even more, many of the algorithmic developments of a specialized electron tomography package like TomoJ [Messaoudi2007] are translations into Java of algorithms in Xmipp. Another tomography specialized package like TOM Toolbox [Nickell2005] directly binds some of the Xmipp functions from Matlab.

This work was funded by the Spanish Ministerio de Ciencia e Innovación, (BIO2007-67150-C01, BFU2007-60228, BFU2009-09331), the Comunidad Autónoma de Madrid (S-GEN-0166-2006, CCG08-CSIC/SAL-3442), the Spanish National Institute of Bioinformatics (a project funded by the Instituto de Salud Carlos III) and by Award Number R01HL070472 from the National Heart, Lung, And Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, And Blood Institute or the National Institutes of Health.

[1] Marabini, R.; Masegosa, I. M.; San Martn, M. C.; Marco, S.; Fernández, J. J.; de la Fraga, L. G.; Vaquerizo, C. & Carazo, J. M. J. Structural Biology, 1996, 116, 237-240

[2] Sorzano, C. O. S.; Marabini, R.; Velázquez-Muriel, J.; Bilbao-Castro, J. R.; Scheres, S. H. W.; Carazo, J. M. & Pascual-Montano, A. J. Structural Biology, 2004, 148, 194-204

[3] Scheres, S. H. W.; Núñez-Ramírez, R.; Sorzano, C. O. S.; Carazo, J. M. & Marabini, R. Nature Protocols, 2008, 3, 977-990

Type of Presentation: ORAL

Topic: Microscopy in Life Sciences