A computational chemistry application of the Kepler scientific workflow platform

J. M. Campanera Alsina[1], X. Barril[1,2] & F. J. Luque[1]

[1] Dept. Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII, s/n. 08028 Barcelona, Catalonia, Spain; [2] ICREA

Scientific workflows are platforms able to amalgamate and orchestrate a range of problem-solving techniques, from databases to data mining tools, in an integrated and comprehensive environment. Workflows platforms on the business world (for instance, BPEL and XPDL) have already appeared and succeeded as disrupting new technologies that boost efficiency and automation. Although scientific and business workflows share the underlying philosophy of automation, integration and simplification, certainly the former are data-oriented while the later are action-oriented. Further more, scientific workflows require computational intensive resources such as grid technologies. The Kepler scientific workflow system is an open-source application that among other features allows: seamless (yet accurately) access to resources and services, service composition, reusability, configurability, workflow design, scalability, detached execution, reliability & fault-tolerance, user-interaction and data provenance.

Here we present a computational chemistry application developed using such technology. A computational chemistry project has to consider three main steps: (1) data preprocessing, (2) job submission and (3) data postprocessing and analysis. The study of the MST solubility model for aminoacids brings us an excellent opportunity to apply this new organizational paradigm: the scientific workflows. The first stage comprises the selection and creation of the molecules with the right format to be calculated and also their storage in a database. The second part manages the submission of the Gaussian script inputs. Finally the last part analyses the results by means of visualisation tools, data mining and machine learning algorithms.