MOLECULAR MODELLING WITH SOMA2

Life scientists have only recently turned to supercomputing to help them model molecules, proteins and cellular activities and their interactions.

The Finnish CSC-IT Center for Science Ltd provides scientists working in all disciplines with software and databases alongside a supercomputing environment. Many computer applications in this supercomputing environment have old-fashioned command-driven interfaces using UNIX commands. Seasoned computer users have no problem with that, but for newcomers like life scientists this poses an obstacle. Hence SOMA2 - the Open Source molecular modelling workflow environment.

WHAT WAS THE PROBLEM?

Consider you are a biomolecular scientist working on a research question involving the binding of two substances: a so-called ligand, usually a small molecule, which forms a new structure with a biomolecule, for instance a protein. The resulting binding might lead to a biological response - for example, one which helps lower your blood pressure. What would a biomolecular scientist normally have to do to solve this research question?

1. Search various databases to relevant molecular structures
2. Retrieve relevant results – in which structures are often shown in two dimensions
3. Convert the 2D molecular structures into a three-dimensional presentation
4. Enter these 3D structures in a ‘docking program’ that calculates the molecules’ options for binding to each other. This often requires a supercomputing environment
5. Retrieve the results from the ‘docking program’ in various formats for further processing, for example, for visualisation of the resulting molecules

Life scientists are spending a considerable amount of time solving technical problems instead of focusing on the scientific tasks at hand. For example, between the steps, the user has to take care of conversions from one format to the other and has to understand and apply various command languages of the different programs.

HOW WAS THE PROBLEM SOLVED?
This is where SOMA2 comes in: it takes care of the conversions and various command languages of the specific programmes, thereby improving the scientists’ workflow. The big challenge was to create a framework that could fit almost any scientific application.

The SOMA2 developers created a standardised internal data format for data exchange between the various applications using chemical mark-up language or CML, which includes information about molecular structures, their associated properties and other metadata. The creators of different applications install a plug-in within their application to allow this improved data exchange, which contains an XML document describing the scientific applications and execution schemes, such as pre-and post-processing scripts and data file templates.

It is estimated that the development time took approximately two person-years. Its initial development was funded by the National Technology Agency of Finland as part of the DRUG2000 programme (2002-2006) and additional development was financed as part of the EGI-INSPIRE project (2010-2012). No further development has taken place in the last year.

**WHAT ARE THE MAIN OUTCOMES?**

Approximately 80-100 biomolecular scientists and chemists now make regular use of SOMA2, the workflow environment for molecular modelling which can be connected to almost any molecular modelling program. Some calculations can take considerable time - even making use of a supercomputer. It is therefore possible to enter your molecules and describe the desired workflow on a Friday afternoon, close the programme and go home. After the weekend, you come back and find the results, as supercomputers also work weekends!

**KEY LESSONS LEARNED**

Firstly, the release of SOMA2 to the wider community took extra effort because the team had to check all the licences of the software models that were used by SOMA2 but developed by other parties.

Secondly, SOMA2’s product manager Tapani Kinnunen now regrets that they did not open up the software development at an earlier stage in order to create a community of users and developers at an early stage. If they had been successful in creating such a community, the development of SOMA2 might have been less dependent on funding - and its upgrade to more modern technical software applications might have been achieved in a gradual way.

**FUTURE DEVELOPMENTS**

What does the future hold for SOMA2? The short answer is that any further development will depend on new funding. While SOMA2 was very innovative at the time of its development, now new applications have been developed by commercial providers that offer similar possibilities. SOMA2 has been developed using technologies that are becoming somewhat outdated and therefore less relevant for developers.
Find out more

Explore the tool and find out more on the [CSC website](http://www.csc.fi).

---

**References**

Image of molecule model taken by [Kennysarmy](https://www.flickr.com/photos/kennysarmy/) and shared on Flickr under a creative commons license.

Image of researcher on computer shared on Flickr by the [University of Georgia](https://www.flickr.com/photos/universityofgeorgia/) under a creative commons license.


SOMA--workflow for small molecule property calculations on a multiplatform computing grid.

Lehtovuori PT1, Nyrönen TH.

Chemistry Central Journal

March 2008, 2:P4,

Open Access

SOMA2 – open source framework for molecular modelling workflows Tapani Kinnunen, TH Nyrönen, P Lehtovuori

Computer generated image used courtesy of the SOMA2 project team.

Based on an interview with Tapani Kinnunen by Maurits van der Graaf on 6 February 2014.

---

March 2014

With thanks to Tapani Kinnunen, and to Maurits van der Graaf and Nicola Yeeles for writing the text.

This work is made available under a [Creative Commons Attribution 3.0 Unported licence](http://creativecommons.org/licenses/by/3.0/).

**Knowledge Exchange** is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research.

[www.knowledge-exchange.info](http://www.knowledge-exchange.info)