

SOMA2 – Open Source Framework for Molecular Modelling Workflows

Tapani Kinnunen, Tommi Nyrönen, Pekka Lehtovuori

CSC – Finnish IT Center for Science

P.O.Box 405, FI-02101 Espoo, Finland

E-Mail Addresses: tapani.kinnunen@csc.fi, tommi.nyronen@csc.fi, pekka.lehtovuori@csc.fi

Introduction

SOMA2 environment [1,2,3] is a web server based system offering a framework for integrating molecular modelling applications, including molecular data exchange. SOMA2 allows users to combine software available in the computing system into unique workflows, which are automatically executed. Recently, the SOMA2 source code was released under GPL license [4].

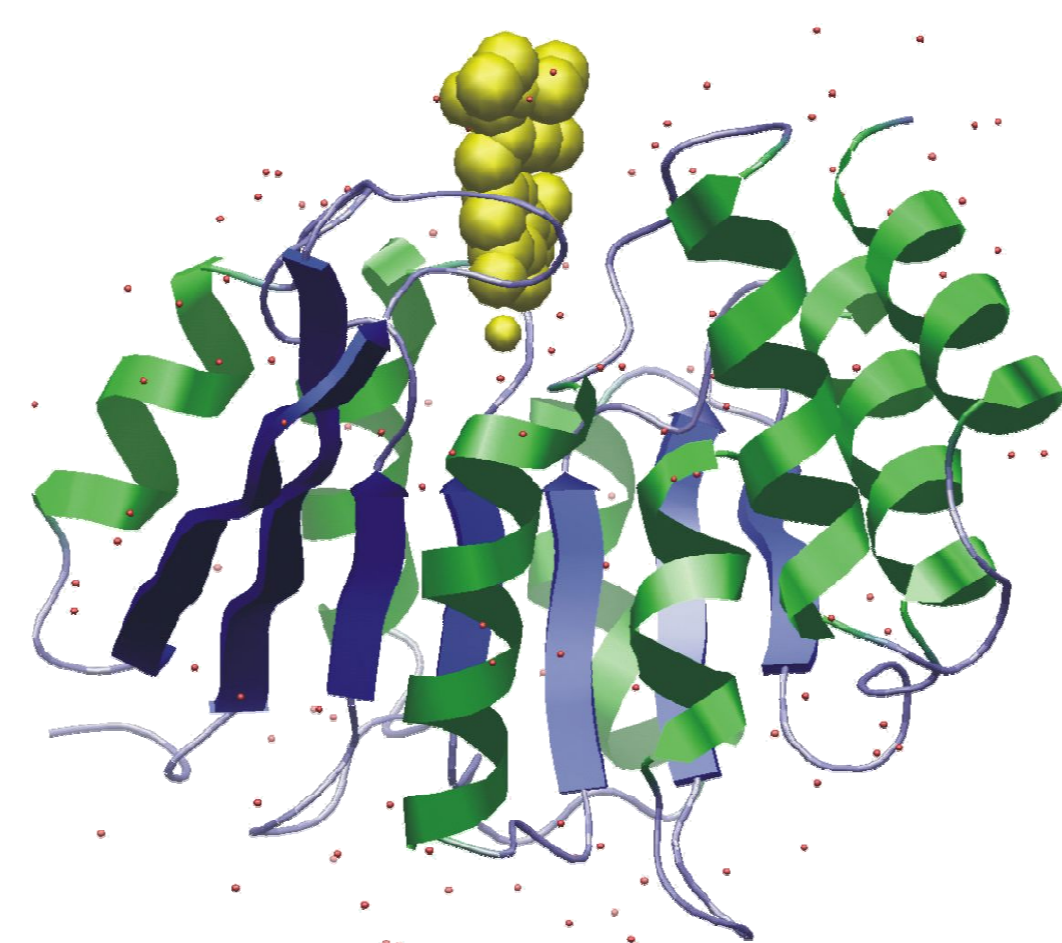
For end users, SOMA2 offers a secure and personalised web browser operated environment for inputting molecular data, submitting and controlling jobs and analysing the results. Parameters of the scientific applications are presented and configured via uniform web forms. The web forms guide the users to correctly configure a program by supplying default values, thresholds, runtime help and content verification. For experts, SOMA2 provides a framework to make virtually any command-line application or combination of applications easily accessible to the (naïve) users. The system enables communication and data exchange between molecular modelling programs on different computing platforms by employing a general data exchange format, CML (Chemical Markup Language) [5].

SOMA2 framework has a modular design, third-party scientific applications are described as pluggable capsules with generic interfaces to manage the data. The description of a scientific application in SOMA2 facilitates the transfer of technical know-how from experts to service users. A SOMA2 capsule consists of an XML description, used e.g. to automatically generate a web form for the program, and scripts & configuration file templates to enable automatic program execution and processing of the program output. The XML descriptions are based on a schema, which we have developed. No programming skills are required to create a SOMA2 capsule.

A demo installation of SOMA2 with limited features is available [6].

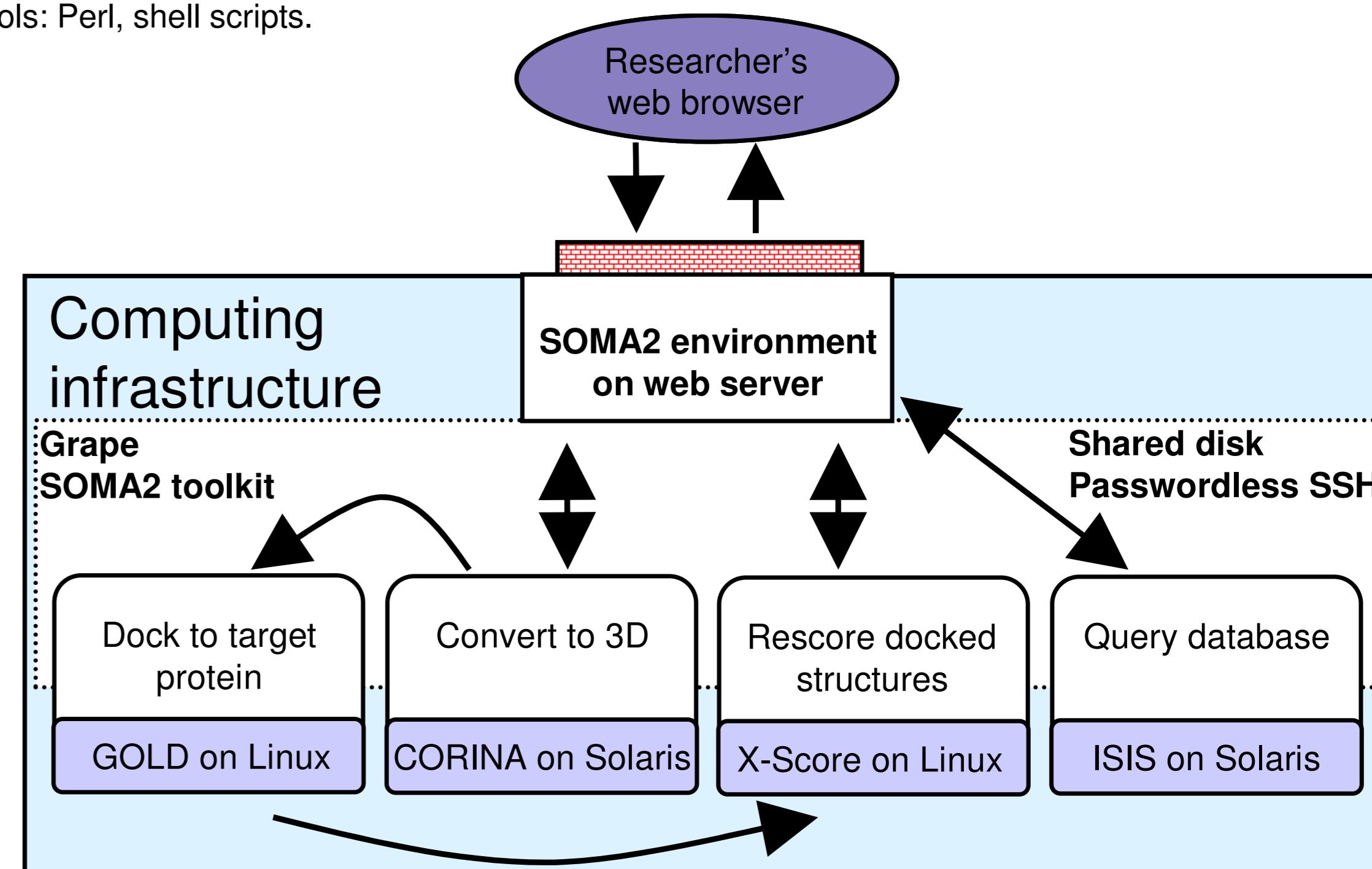
SOMA2 is open source

- **Open source released in May 2007**
 - SOMA2 source code is licensed under GNU General Public License (GPL).
 - Downloads available from SOMA2 WWW –pages: <http://www.csc.fi/soma>.
 - SOMA2 demo installation with limited features available at: <http://soma2demo.csc.fi>.
- **Distribution contains example SOMA2 capsules**
 - Can be used as examples in creating own capsules.
 - obenergy (Open Babel single point energy calculator, <http://openbabel.sourceforge.net>).
 - obprop (Open Babel molecular property calculator, <http://openbabel.sourceforge.net>).
 - identity (SOMA2 test capsule).

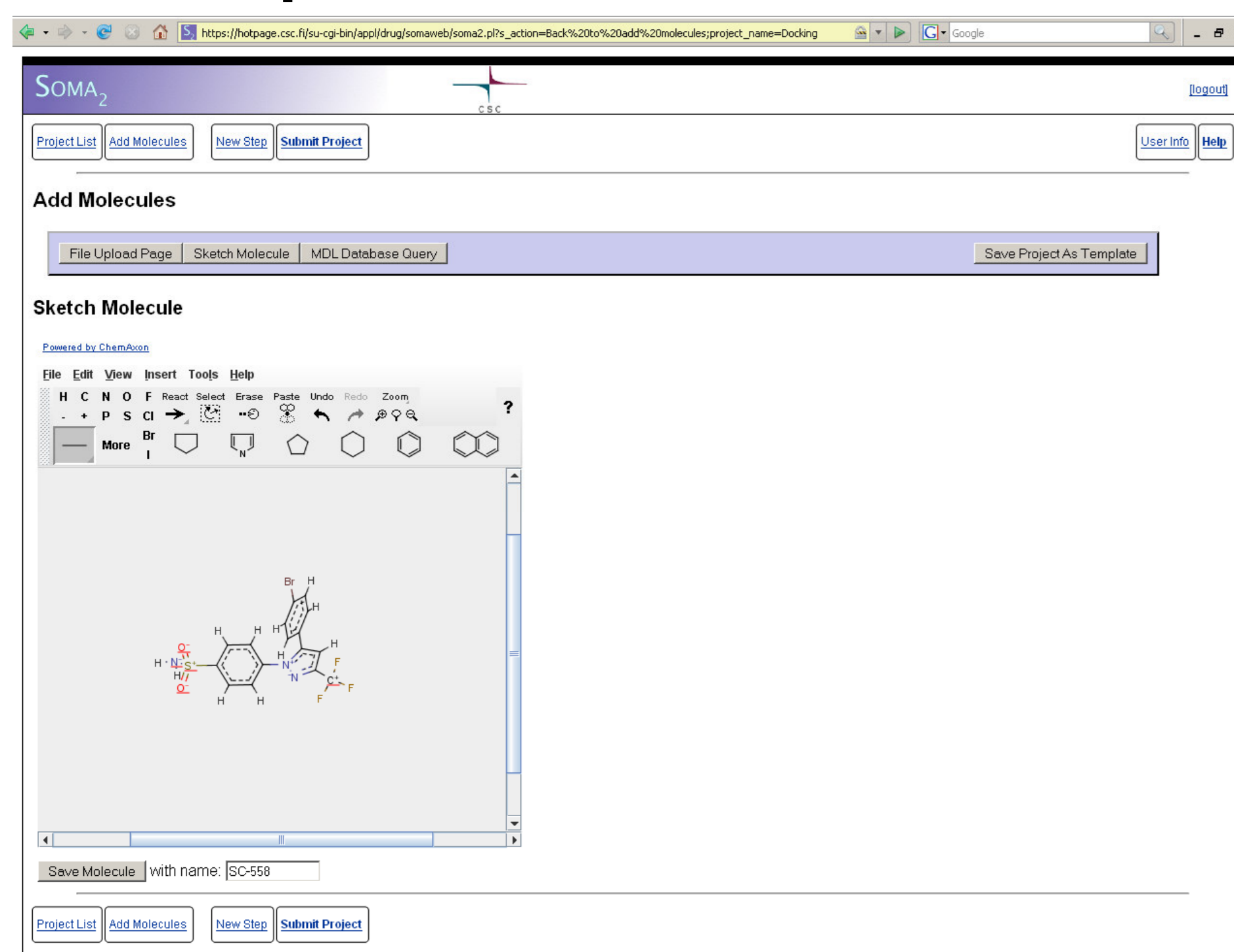


Components of SOMA2

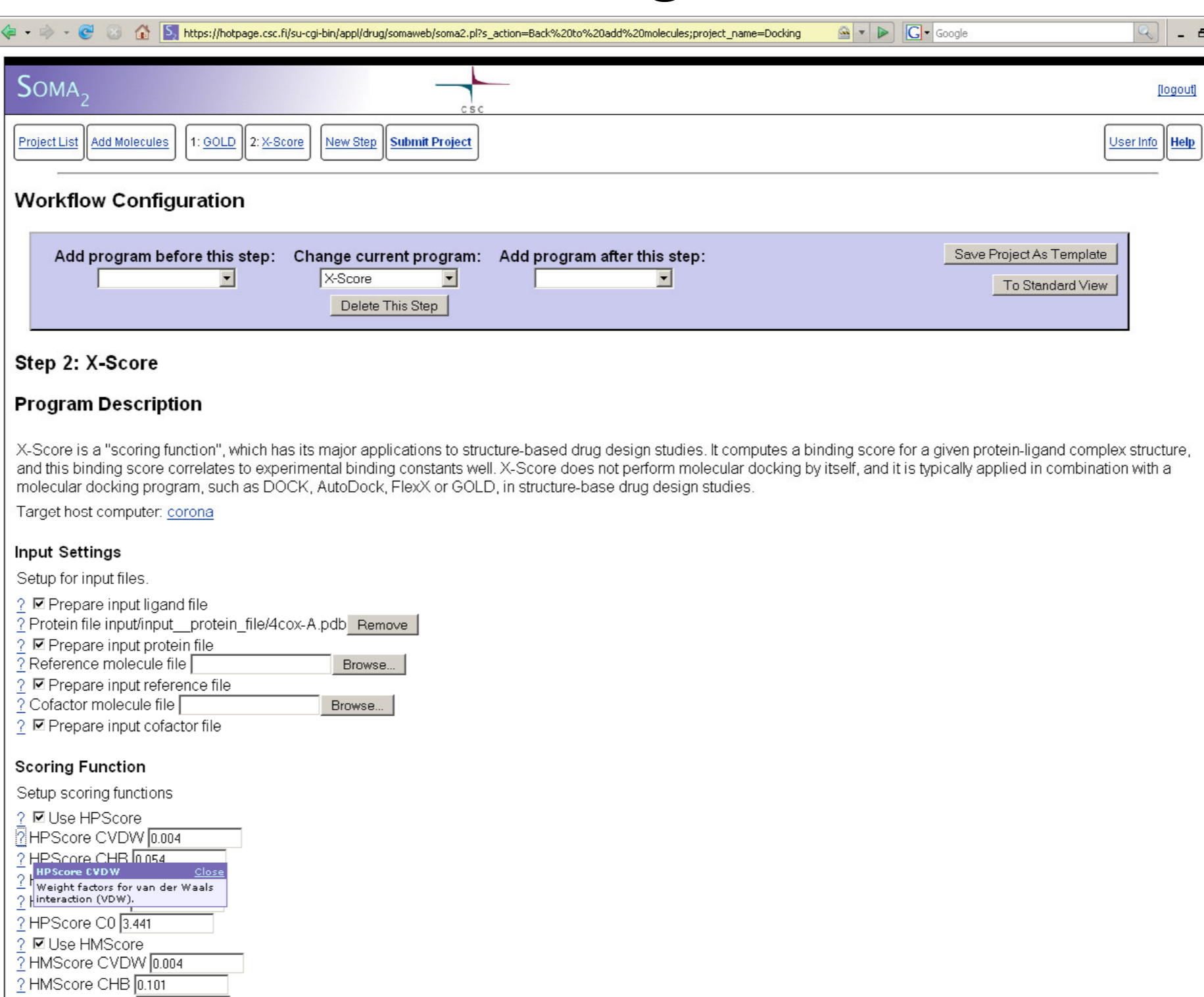
- **WWW interface**
 - User authentication, input of molecular data, building the program configurations, performing database queries, creating a workflow and analysing the results.
 - Tools: Perl, JavaScript, HTML, CSS.
- **Workflow manager program Grape**
 - Execution, logistics and monitoring of program execution (2D XML graph).
 - Tools: Java.
- **SOMA2 capsules**
 - eXtended Markup Language (XML) description for attaching a scientific program to be used via SOMA2.
 - Templates of program configuration files, command scripts for executing programs, batch queue system scripts and program output parsers.
 - Tools: XML, shell scripts.
- **Toolkit of helper applications**
 - Programs for molecule format conversions, building the execution files from the templates and managing the internal data.
 - Tools: Perl, shell scripts.



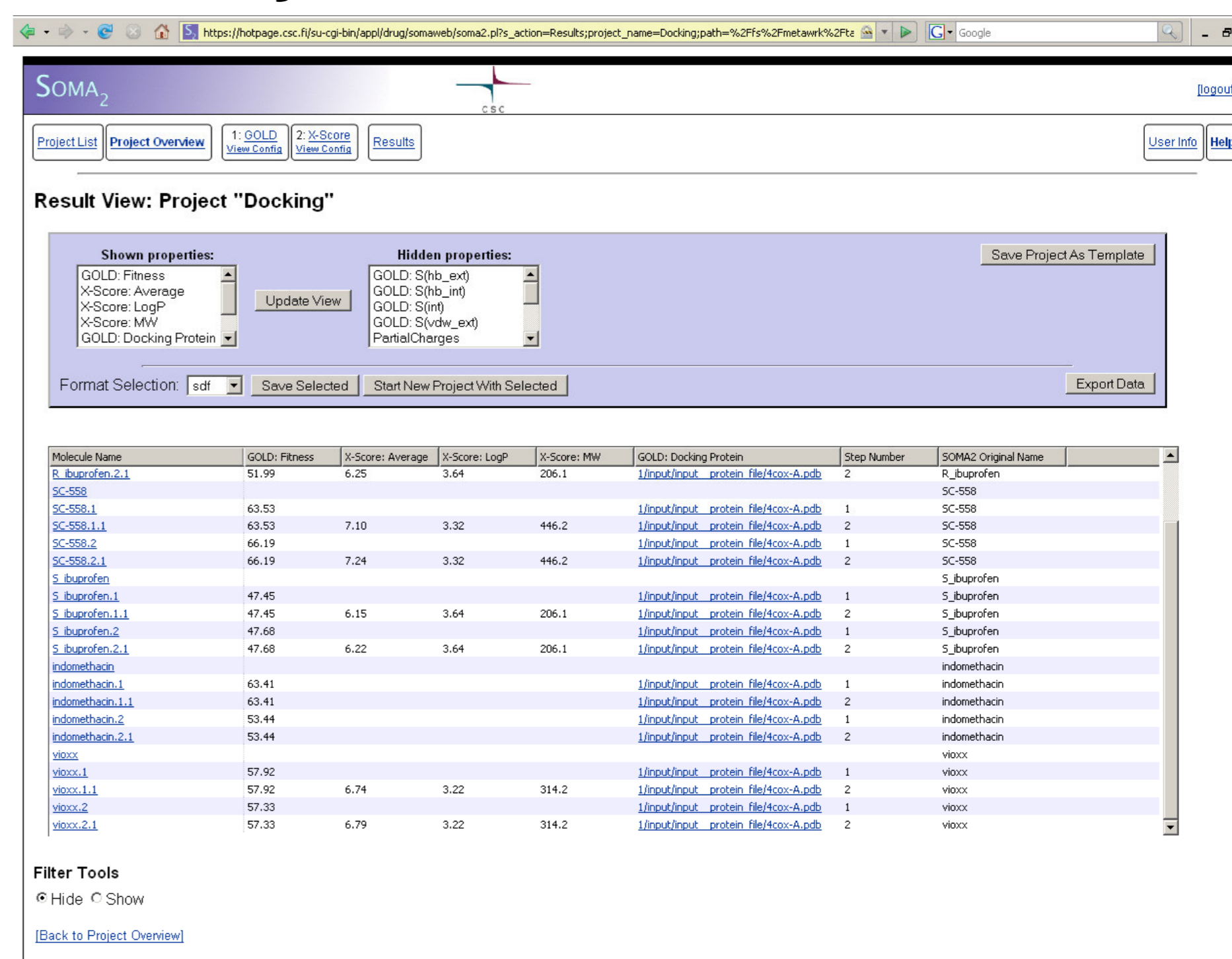
From input molecules



With workflow configuration



To analysis of the results



SOMA2 capsules at CSC

- **Capsules for 12 different scientific programs**
 - 2D-Property (Sybyl module) (<http://www.tripos.com>).
 - 3D-Property (Sybyl module) (<http://www.tripos.com>).
 - AutoDock (<http://autodock.scripps.edu>).
 - BRUTUS (<http://www.uku.fi/farmasia/fake/modelling/>).
 - CORINA (<http://www.molecular-networks.com/software/corina>).
 - Gaussian 03 (<http://www.gaussian.com>).
 - GOLD (http://www.ccdc.cam.ac.uk/products/life_sciences/gold/).
 - obenergy (<http://openbabel.sourceforge.net>).
 - obprop (<http://openbabel.sourceforge.net>).
 - Overlay (<http://www.uku.fi/farmasia/fake/modelling/>).
 - ROTATE (<http://www.molecular-networks.com/software/rotate>).
 - Sybyl (<http://www.tripos.com>).
 - Tanimoto similarity (Sybyl module) (<http://www.tripos.com>).
 - TURBOMOLE (<http://www.turbomole.com>).
 - Volsurf (Sybyl module) (<http://www.tripos.com>).
 - X-Score (<http://sw16.im.med.umich.edu/software/xtool/>).

Third-Party components

- **ActiveWidgets JavaScript library** (<http://www.activewidgets.com>)
 - All tables within SOMA2 user interface.
- **overLIB JavaScript library** (<http://www.bosrup.com/web/overlib/>)
 - Popup information boxes within SOMA2 user interface.
- **Open Babel** (<http://openbabel.sourceforge.net>)
 - Molecule file format conversions and property calculation.
- **JDB** (<http://www.isi.edu/~johnh/SOFTWARE/JDB/index.html>)
 - ASCII data filtering tools.
- **ChemAxon Marvin Java applets** (<http://www.chemaxon.com>)
 - Tools for building and visualising molecular structures.

Acknowledgements

- CSC – Finnish IT Center for Science (<http://www.csc.fi>)
- Tekes (National Technology Agency of Finland) (<http://www.tekes.fi>)
- ChemAxon Ltd. (<http://www.chemaxon.com>)

References

- [1] P. Lehtovuori, T. Nyrönen, *J. Chem. Inf. Model.* 2006, 46(2), 620.
- [2] T. Nyrönen, T. Kinnunen, P. Lehtovuori, *CSCnews* 2006, 2, 36. <http://www.csc.fi/english/csc/publications/cscnews>
- [3] <http://www.csc.fi/soma>.
- [4] <http://www.gnu.org/copyleft/gpl.html>.
- [5] P. Murray-Rust, H.S. Rzepa, *J. Chem. Inf. Model.* 2003, 43(3), 757.
- [6] <http://soma2demo.csc.fi>.