



SOMA₂

SOMA2 User Manual

Tapani Kinnunen
CSC – IT Center for Science Ltd.

SOMA2 User Manual

Version: 1.4 Aluminium

Written by: Tapani Kinnunen

Copyright (C) 2006-2011 CSC – IT Center for Science Ltd.

All rights reserved.

Table of Contents

Table of Contents	3
1. Introduction.....	4
2. Getting Started	4
3. Creating and Configuring a New Project	5
3.1. Project Name and Model Template	5
3.2. Pre-Commit Navigation.....	6
3.2.1. Project List.....	6
3.2.2. Add Molecules.....	6
3.2.3. New Step.....	6
3.2.4. Submit Project.....	7
3.2.5. User Info.....	7
3.2.6. Help.....	7
3.3. Input Molecules.....	7
3.3.1. Uploading Molecule File	8
3.3.2. Sketching a Molecule	8
3.3.3. Molecule Files List.....	10
3.3.4. File Details.....	10
3.4. Workflow and Program Configuration	11
3.4.1. Starting a Workflow	11
3.4.2. Saving Workflow As Template	11
3.4.3. Workflow Management Tools.....	12
3.4.4. Program Configuration	12
3.4.4.1. Outlook of the Interface	12
3.4.4.2. Parameter Specific Help Texts	13
3.4.4.3. Connected Parameters and Preset Values.....	14
3.4.4.4. Requirements for the Parameters.....	14
3.5. Submitting Project.....	14
4. Project Overview and Results	14
4.1. Post-Commit Navigation	15
4.1.1. Project Overview	15
4.1.2. Program / View Config	15
4.1.3. Results	15
4.2. Project Status View.....	15
4.3. Viewing Results	17
4.3.1. Result Table	18
4.3.2. Property Data Fields.....	18
4.3.3. Using Filter Tools	19
4.3.4. Exporting Data Table.....	19
4.3.5. Visualisation of a Result	19
4.3.6. Visualising Additional Structure Files	20
4.3.7. Viewing Images and Data Files.....	21
4.3.8. Saving Result Molecules	21
4.3.9. Starting a New Project with Result Molecules.....	21
5. SOMA2 Citation.....	21
6. Technical Requirements.....	22

1. Introduction

SOMA2 modelling environment operated through WWW-browser is a versatile tool that aids researchers in the field of molecular modelling, especially in the computational drug discovery. The SOMA2 environment offers a full scale modelling environment from inputting molecular data to analysing and visualising the final results a computational project. The system includes user interface tools for a variety of programs in the fields of chemistry and drug discovery.

Software from the different vendors are typically not compatible and getting the output from one program to some other can require file conversions, file editing and other technical procedures requiring special skills. The SOMA2 environment offers not only the user interfaces for the individual programs but also acts as integrator between the programs and the data that they produce. Programs that are available through the SOMA2 are fully integrated and can be run in virtually any order within the system. This provides a possibility to construct unique computational workflows that are automatically run within the SOMA2 environment.

In the workflows of SOMA2 the data is transferred from one program to another so that the original input is input for the first program to be executed. Output of the first program is then the input for the second one, and so on. This way, the users can perform meaningful total tasks like 2D-3D coordinate conversion, docking simulation for the generated 3D structures and rescoring the docked structures with other scoring functions.

The SOMA2 environment hides the technical details beneath the surface providing the end user an easy and straightforward access to computational tools without the requirement for any technical knowledge of computer platforms and programs that are run in them. In addition to the workflow control and the interface tools, for the advanced users SOMA2 environment offers ready made platform and program specific configuration files that are usable also outside the SOMA2 environment. The SOMA2 system is operated with the existing user accounts of the computing system. This enables an easy user authentication and fully secured data management.

The purpose of this document is to work as end user documentation for SOMA2 environment. This document assumes that SOMA2 is properly installed, user authentication procedure has been dealt with and the web application has been started. The aim of the document is then to describe all the features that SOMA2 offers for the end user. For complete client requirements, see chapter 6.

2. Getting Started

As SOMA2 application starts, the first page loaded is "Project List" page that shows a list of existing user projects (Figure 1). If user has no projects, a brief welcome message is shown instead. From this page user can proceed to start a new project by using "New Project" button in toolbox of the page. In case of existing projects, the user can select one from the list by clicking a corresponding project name and proceeding to see the overview and the results of committed project or continue configuring project that has not yet been committed.

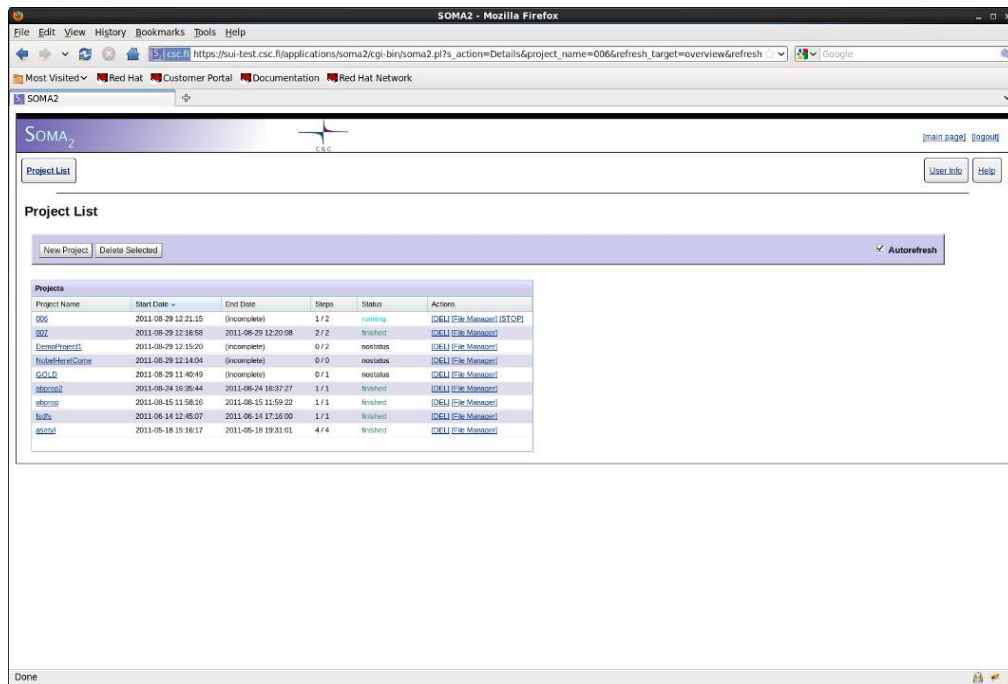


Figure 1. Welcome page.

The project list itself offers current status information of each project that user has. Interactive table (Ext JS, <http://www.extjs.com>) includes tools for sorting the projects according to user selected column or deletion of a selected project. The table is sorted by clicking a column header. “[STOP]” link in the “Actions” column can be used to stop execution of the corresponding project. “[STOP]” link is visible only for projects that are running. Using “[DEL]” link first stops project, if it is running, and then deletes the project. Deleting a project deletes all the files associated with the project. “[File Manager]” link can be used to access files of the project directly. Using “[File Manager]” link opens “File Manager” page, which contain directory and file view shown in a tree menu (Ext JS, <http://www.extjs.com>). This menu can be used to browse the filesystem, open files and save files. In addition, file details such as owner and file size are shown. It should be noted that file manager utility may not be available in all SOMA2 installations. Multiple projects can be deleted by first selecting corresponding rows and then pressing “Delete Selected” button in the toolbox of the page. Using control/shift+mouse button, will enable multiple selections of rows. “Project List” page contains automatic update procedure of project states which will update the page, if changes in project states occur. This feature can be controlled by using “Autorefresh” checkbox in the toolbox of the page.

3. Creating and Configuring a New Project

This chapter describes creation of a new project in SOMA2 environment. The chapter covers all the possible functions of the SOMA2 environment from starting a new project to the actual submission of the project.

3.1. Project Name and Model Template

When “New Project” button on the Welcome page is pressed, a configuration of a new project is started and “New Project” page is loaded. First, the project is given a unique name. After this a model template is chosen. The default is to proceed with an empty template that has no predefined programs or workflow settings in it. In this case program selection and configuration is up to the user.

The template list also shows “Factory Templates” from which user can select a predefined workflow for some specific task. In the latter case, program or programs,

with a preset configuration, are already included in the workflow and user only needs to give original input and check each of the program configurations. Selected “Factory Template” can still be customised to suit user’s purpose. User customised template based either on empty or Factory templates can be stored as a user template that will be also shown on the template list.

Not all of the “Factory Templates” are directly available for all the users. The templates including programs that the user has no access to are disabled. These templates can be activated by acquiring access to the computing platforms where the corresponding programs are installed.

After setting the project name and the model template user can proceed by using “Initialise Project” button in toolbox of the page.

3.2. Pre-Commit Navigation

On every page of SOMA2 environment, there is a navigation menu tool that can be used to explore between different project phases and steps (Figure 2). The basic “pre-commit” navigation links are “Project List”, “Add Molecules”, “New Step”, “Submit Project”, “User Info” and “Help”. During the workflow and program configuration more links will be added, according to program selections. User’s current view in SOMA2 user interface is shown in boldface in the navigation menu.



Figure 2. Navigation menu.

In the SOMA2 environment, moving between different phases of the project using the navigation menu is completely free; any of the navigation links can be used at any point of the project configuration. Also if user closes the WWW-browser or the browser crashes, configuring the project can be continued by selecting the project from the project list; SOMA2 automatically saves the current project as the settings change.

3.2.1. Project List

Using the “Project list” link of the navigation, the initial page described in chapter 2 is loaded. Getting back into the project where navigation link was used, the user needs to select the corresponding project from the table.

3.2.2. Add Molecules

Clicking “Add molecules” will always take to “File Upload Page” where user can input molecules in different ways. See chapter 3.3 for the details.

3.2.3. New Step

When this link is used for the first time, a workflow and program configurations will be started. Afterwards, the link can be used to add a new step to the workflow. See chapter 3.4 for the details.

3.2.4. Submit Project

Clicking “Submit Project” link will take to page showing a brief review about the current project. On this page, user can actually commit the project. See chapter 3.5 for the details.

3.2.5. User Info

Selecting “User Info” from the navigation menu loads a page that shows information on the programs, services and computer platforms that are connected to SOMA2 environment. On this page detailed information of the resources that the user has or has not access to, are shown. This page also shows user’s CPU quota. It should be noted that CPU quota information may not be available in all SOMA2 installations. In case grid computing is enabled in the installation, several grid related tabs are shown. These include Certificates, Delete Certificate and Proxies, but their visibility depends on the installation.

The data shown on “User Info” page is not updated in real time and to refresh all the information the user needs to click “Update User Info” button shown on the page. Updating the data is particularly important if the user account is changed. For example, if user has just gained access to some computer platform that includes programs connected to SOMA2 environment, the additional resources are available only after the “Update User Info” button is clicked.

3.2.6. Help

Moving a mouse pointer over this navigation item and clicking mouse button, will open a text box popup (Ext JS, <http://www.extjs.com>) which shows a runtime help for the current page (Figure 3). The help system in SOMA2 is built so that the actual text is changed while moving between different pages of the system.

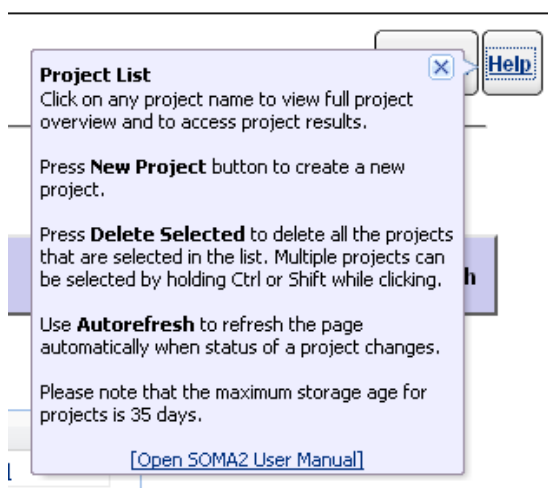


Figure 3. Help system of SOMA2.

While opened, the text briefly explains most important functions of the current page. To close the help text box user must click “close” icon in upper right corner of the popup. The help text always includes a link to the online version of the SOMA2 user manual. Clicking the link opens the user manual in a new browser window.

3.3. Input Molecules

This chapter describes the different ways of inputting molecular data in SOMA2 environment. Also the general features of handling the inputted molecular data is described here.

When the project is started or the “Add Molecules” navigation link is used, default view will always be the “File Upload Page”. User can then switch the “Add Molecules” view by using the different buttons of the toolbox that is shown on the page. Possible views are “File Upload Page” and “Sketch Molecule”. In addition, the toolbox includes “Save Template” button that can be used to save user template. Each of the previous are explained in the following chapters. Regardless of source and format of molecule files that are inputted to SOMA2 environment, the files are converted to internal format and prepared automatically. All methods for inputting molecular data feature options “Add Hydrogens” and “Clean Structure” which are controlled by using provided checkboxes in the page. “Add Hydrogens” option will automatically add explicit hydrogen atoms to all input molecules and “Clean Structure” option will automatically create initial 3D-structure for all input molecules. “Add Hydrogens” and “Clean Structure” options as well as all internal file format conversions in SOMA2 are performed with Open Babel file format conversion tool (<http://openbabel.sourceforge.net>).

3.3.1. Uploading Molecule File

From the “File Upload Page” a molecule data file can be uploaded to SOMA2 environment from user’s own workstation. The SOMA2 environment accepts also files including multiple molecules.

To perform the upload it is recommended that “[Supported Filetypes]” list is checked by clicking the provided link on the “File Upload Page”. If the user has a file that is in suitable format for the SOMA2, clicking “Browse...” button will open a file upload window where the user can search for a file and then select it. Selecting the file automatically starts the upload process. After a successful upload, “Molecule Files:” list appears to the page. See chapter 3.3.4 for the details. In this view, “Add Hydrogens” and “Clean Structure” checkboxes are located below the file upload field.

3.3.2. Sketching a Molecule

In SOMA2 environment it is also possible to draw and save a molecular structure. To perform this user needs to press “Sketch Molecule” button from the toolbox. This will load an applet tool that can be used to sketch a molecule (Figure 4 and Figure 5). The SOMA2 environment currently supports two distinct applet tools for sketching, Marvin Sketch (<http://www.chemaxon.com>) and JChemPaint (<http://sourceforge.net/apps/mediawiki/cdk/index.php?title=JChemPaint>), which are loaded here depending on the installation. These applets provide a different set of features. JChemPaint won't provide for example 3D sketching support but on the other hand, is free software. Details and help for using the applet can be accessed by selecting “Help” from the applet menu.

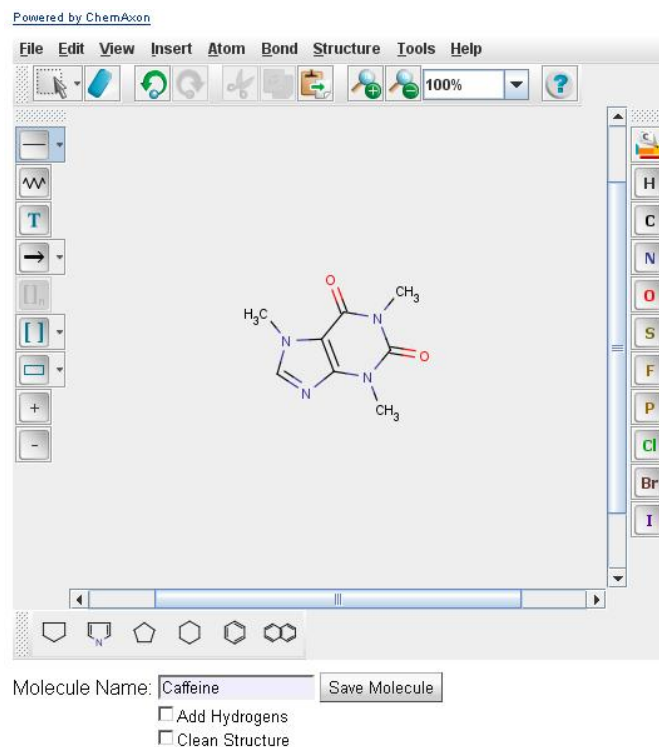


Figure 4. Marvin Sketch tool.

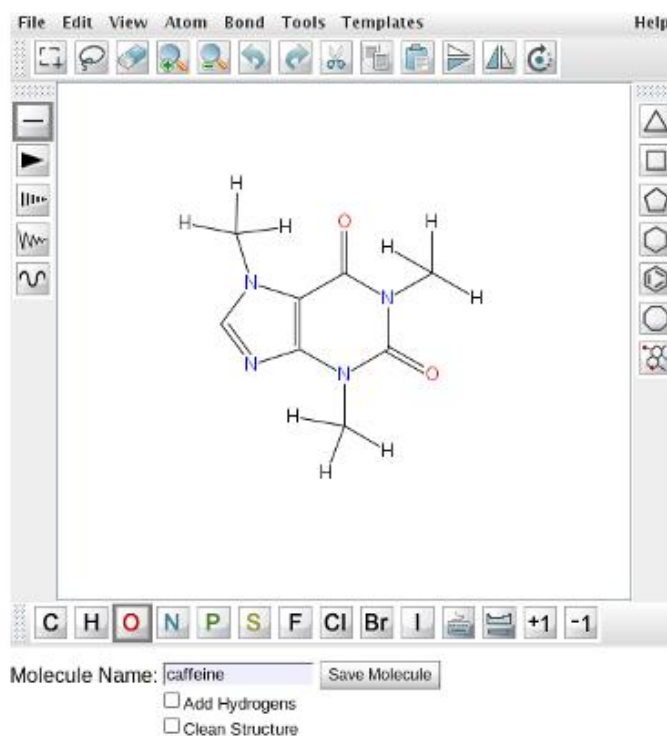


Figure 5. JChemPaint Sketch tool.

After molecule is sketched it can be saved to SOMA2 system by providing a name for the structure and then clicking “Save Molecule” button. By default, the applet tool sketches and saves 2D-structure that has no explicit hydrogen atoms added. In this view, “Add Hydrogens” and “Clean Structure” checkboxes are located below molecule name text field. Notice that if the installation is using JChemPaint, 3D structure is flattened to 2D if the file is opened again for modifications. Furthermore, if the user is using Marvin, in addition to provided checkboxes, explicit hydrogen atoms and/or 3D-structure can be added to the sketched molecule with the applet tool. Adding explicit

hydrogen atoms is done by selecting “Structure/Add/Add Explicit Hydrogens” from the applet menu. 3D-structure is generated by choosing “Structure/Clean 3D/Clean in 3D” from the applet menu. These actions override the defaults when the sketch is saved.

The applet tools can also be used to directly load/save molecule files on the user’s workstation. For this, the user needs to select “File/Open” or “File/Save As” from the applet menu. If the user saves the molecule in the SOMA2 environment, the molecule is added to “Molecule Files” list described in chapter 3.3.3.

3.3.3. Molecule Files List

Storing molecules to SOMA2 environment from the different sources described in the earlier chapters, will add entries to “Molecule Files” list that appears after the first molecule file addition (Figure 6).

Molecule Name / File Name	Molecules	Status	Actions
- / 1.mol2	-	Failed	[DEL]
Caffeine / caffeine.sdf	1	ok	[DEL] [Save] [Edit/View]
- / 29997.sdf	929	processing ~60%	[STOP]

Autorefresh Files

Please note that you cannot submit the project before all the input molecules are prepared successfully.

Figure 5. “Molecule Files:” list.

Each item on the “Molecule Files” list consists of “Molecule Name / File Name” showing the molecule name and/or the file name, number of molecules that the file includes and status of the file. In addition, each item has action links. If a molecule file contains a single molecule, both the molecule name and the file name are shown. Otherwise, only the file name is shown. During the automatic preparation of the input files, the status of the file is “processing”. The status changes to “ok” if the initial conversions and preparations succeed. File preparation statuses data is refreshed automatically once preparation process proceeds. No manual refreshing of the page is required. To disable automatic status refreshing, “Autorefresh Files” checkbox can be used. Checkbox is visible only when at least one of the files is being processed. The initial preparation is performed as a background process and it can take some time. Despite the fact that the preparation process for a file is unfinished and status of the file is still “processing”, user can continue working normally with the SOMA2 environment. During the preparation, the only available action link for the file is “[STOP]”. The “[STOP]” link is used to terminate the automatic preparation process. If the status of the file is “ok”, the actions links are “[DEL]”, “[Save]” and “[Edit/View]”. The prepared input file can be deleted by clicking “[DEL]” link. Using “[Save]” link saves the file to user’s local computer. “[Edit/View]” link changes the view to the “Sketch Molecule” displaying the corresponding molecule in the sketcher applet described in chapter 3.3.2. In this case user can also modify the shown structure and then overwrite the previous contents of the file or save the modified structure as a new input file. Currently “[Edit/View]” link is available only for files containing a single molecule.

As mentioned above, status “ok” of the file indicates a successful automatic preparation process. If the status is missing or the status is not “ok”, the automatic preparation has failed and user is required to delete the file from the system. It should be noted that the project can be committed only when all the input files have “ok” status.

3.3.4. File Details

Even though SOMA2 environment includes automatic format conversions and preparations for the input files, in some cases it is recommended for the user to check the input file manually. This applies especially to situations where the input file includes molecular properties.

After successful automatic preparation process, each entry on “Molecule Files” list has an embedded link in “Molecule Name / File Name” column. Clicking this link loads “File Details” view where the user sees all the properties that system detects in the input molecule file. The auto detection tries to set the molecule name and data types of the different properties. In the “File Details” view user can override the automatic settings by changing the title field, property names and the data types by using the provided form elements. The user can also delete the properties that are not needed by checking the “delete” option on the page.

Some or all of the properties can also be checked to be inherited. Checking the “inherit” option of a property will cause the property to be transferred to all the descendants of the input molecules that are produced during the workflow of SOMA2. The user should consider carefully when using the “inherit” option. It is safe to inherit properties that for example depend only on the molecular 2D-structure or the molecular formula since these do not change during the SOMA2-workflows. However, properties like docking score or molecular total energy depend on the 3D-structures of the molecules. The 3D-structures of the input molecules are usually expected to change and therefore this kind of properties are not recommended to be inherited.

All the changes are applied to the input file by clicking “Apply Changes” button on the page whereas “Cancel” button will ignore the changes. Clicking either one of the buttons will take back to the view where user was before link was used.

3.4. Workflow and Program Configuration

As mentioned in the introduction, SOMA2 environment offers not only the interface tools but also a possibility to construct computational workflows that can consist of several different programs. This enables the construction of meaningful tasks where output of a program is the input of a second program to be executed, and so on.

In the current version of SOMA2, a sequential execution of the programs is supported. This chapter describes the workflow and program configuration in the SOMA2 environment.

3.4.1. Starting a Workflow

A workflow construction can be started at any point of the project configuration. Clicking “New Step” link in the navigation menu adds a step into workflow and loads a page showing a toolbox from which user can choose an appropriate configuration detail level. Choosing the “Advanced” mode requires a more detailed knowledge of the programs but offers more configuration options than the “Standard” mode. In the “Standard” mode most of the configurable options are preset with reasonable default values. The configuration level setting can be switched between “Advanced” and “Standard” modes at any point of the workflow generation.

In the “New Step” page, there is also a popup menu from which the user selects a program, which will define newly added step of the workflow. Selecting a program will load the corresponding program configuration page and add the program name to the navigation links as mentioned in chapter 3.2. Added program links are fully integrated to the internal navigation system of SOMA2; clicking these will take the user directly to the corresponding program configuration page.

3.4.2. Saving Workflow As Template

As mentioned in chapter 3.1, user can store own templates in SOMA2 environment. Whenever “Save Template” button is clicked, “Save Template” page is loaded. On this page the user needs to supply a name and short description for the template. Also template author credentials can be stored. The template is saved by clicking “Save” button on the page.

The saved template includes the current workflow with the effective settings at the save moment. However, the input molecules and additional files are not saved in the template. After this, saved template is available in the model templates list. If user wishes, any of the user’s templates can be deleted by using “[DEL]” link on interactive table that shows the model templates.

3.4.3. Workflow Management Tools

In SOMA2 environment every program configuration page shows a toolbox that includes tools for modifying the workflow (Figure 7). With these tools, user can freely customise a workflow that is suitable for the current requirements.

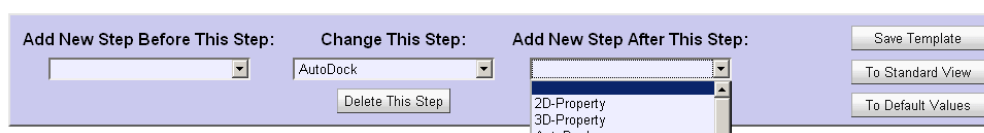


Figure 6. Workflow management toolbox.

Using popup menu “Change This Step:” or clicking “Delete This Step” button, the user can change or delete the current step of the workflow. With “Add New Step Before This Step:” and “Add New Step After This Step:” popup menus, the user can add a new step to the workflow before or after the current one. Using the change or add options will directly load configuration page of the newly selected program. After the workflow construction is started, using the “New Step” link in the navigation will add a new step after the last step in the current workflow. Otherwise operation of the “New Step” link is the same as described in chapter 3.4.1.

3.4.4. Program Configuration

This chapter describes the general features of program configuration interfaces in SOMA2 environment. When configuring a program, user needs to be aware that the same configuration will apply to all the molecules that are input of the program. So if the program configuration includes parameters related specifically to the molecular content of the input, the user should verify that only chemically similar molecules are inputted to the SOMA2.

As already described in chapters 3.4.1 and 3.4.3, adding a program to the workflow adds the selected program to the navigation menu. The navigation links can then be used to switch between the different program interfaces at any point; SOMA2 environment automatically saves the current configuration when the view is switched. In the program configuration interfaces of the SOMA2 environment the user do not need to know on which computation server a program is run or the technical details of running the program on the target platform. All the technical information is built in to the SOMA2 and all the program and platform specific configuration files are generated automatically.

3.4.4.1. Outlook of the Interface

Basically, every program configuration interface in SOMA2 environment shows the adjustable parameters and settings of the program as a tab panel form. Currently, parameters are presented with typical web form elements such as popup menus, checkboxes, textfields, textboxes and input file fields (Figure 8).

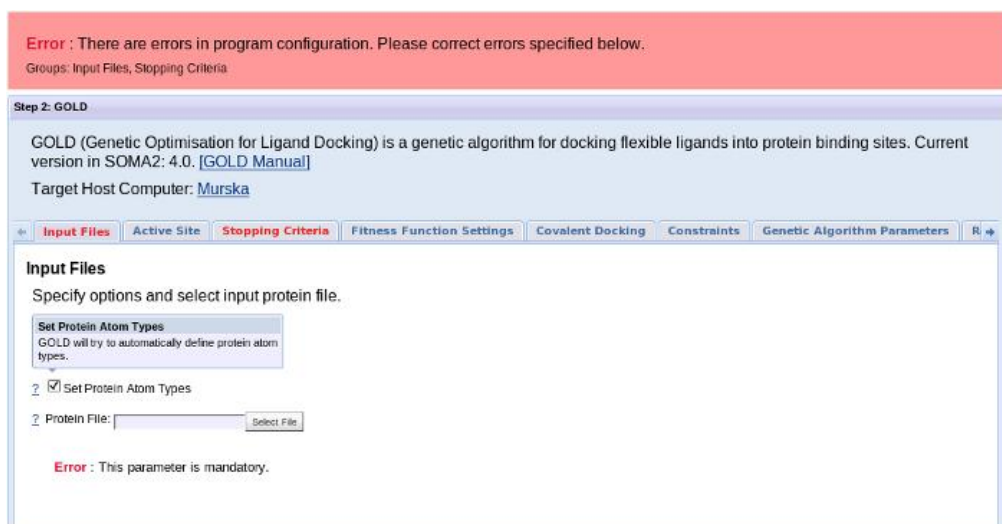


Figure 7. Program configuration interface.

The configuration area consists of two elements, of which the error window is only showed if fixes are required somewhere in the form. The main panel on the other hand is constantly visible and is vertically separated in two parts. The upper area of the main panel shows the program description. This description is accompanied with a link to the online manual of the program if such a thing exists. The target platform where a program is run, is shown in the program description as a link. The link can be used to access detailed information of the target platform. Under the program description is a tabbed area, which shows the parameters of the current program divided in tabs based on the groups they belong to.

Each parameter group is displayed in a distinct tab named after the group title. Typically these parameter groups are associated to the function of the parameters in them, for example parameter group “Input” will consist of parameters related to input options of the current program. Under each parameter group title, there is a text that briefly explains the function of the group. The configuration pages of some programs can contain user adjustable property inherit settings. If available, these are presented in a parameter group titled “Inherit Properties”. Choosing the properties to be inherited functions similarly as already described in chapter 3.3.5. The only difference is that it is also possible to set the properties to be inherited to the ancestors of the result molecules.

The toolbox containing the workflow control tools also contains “To Standard View” or “To Advanced View” button that can be used at any point to switch between advanced and standard configuration modes as described in chapter 3.4.1. When using standard view, the tabs available only in advanced view are shown as disabled and are enabled for modification when changing view mode.

3.4.4.2. Parameter Specific Help Texts

Also on the program configuration page, navigation item “Help” provides general help as described in chapter 3.2.6. To make a program configuring a bit easier, SOMA2 environment includes also a parameter specific help texts.

All adjustable parameters in the SOMA2 are equipped with help texts. These help texts are accessed by moving a mouse pointer over the question mark (?) characters on the page (Figure 8). Moving the mouse pointer away will hide the help text popup. The parameter specific help texts provide runtime aid for configuring a specific program but usually they are not complete. In cases where additional guidance is needed, user should consult the user manual of the program.

3.4.4.3. Connected Parameters and Preset Values

As the program configuration page is initially loaded some of the parameters are directly accessible but some can also be disabled. This takes place if function of a parameter affects the function of some other parameter.

For example, choosing a method releases its sub options to be configured. Deselecting the method disables the sub options. These connections between the parameters are determined automatically and the configuration page is refreshed when user modifies the parameters.

In addition, almost all of the adjustable parameters have some default values already included when a program configuration page is first shown. These are typically set as reasonably as possible to ensure operation of a program. The default configuration settings are restored by using "To Default Values" button, which is located in toolbox containing the workflow control tools.

3.4.4.4. Requirements for the Parameters

Typically the parameters of a program have some requirements. For example, some integer may have only positive values or number of processors used in the computation can be set from one to eight, and so on. In addition, parameters may be mandatory so they must not be left empty.

The program configuration pages in SOMA2 environment also have internal checking routines to verify a valid program configuration. Validity of a parameter is automatically verified when modifying its value. If any incorrect settings are found, an error message and instructions indicating the corrections are shown on the page.

3.5. Submitting Project

When user is finished with inputting molecules and configuring the workflow, the project is submitted for computation by clicking "Submit Project" from the navigation menu. This loads "Submit Project" page where the user sees a brief summary of the project.

Entering the page will also launch automatic checking of the input molecules, the configured programs and the workflow. The internal data management in SOMA2 is designed so that all the integrated programs, in principle can be run in any order. Even this is technically possible it might not be physically meaningful to do so. If the SOMA2 environment, according to its internal knowledge of the programs, finds any inconsistency in the current workflow, a warning or error message is shown. For example, if user configures a workflow where solely 3D-structure dependent properties would be calculated to 2D input structure, a warning message will be displayed. If any incorrect parameters from the program configurations are found, an error message and instructions for making the corrections are shown.

The user can actually commit the project only after the errors are corrected. If no errors are found, "Commit Project" button in page's toolbox is enabled and pressing it will send the project for computation. The user can also enable an e-mail notification upon project completion. This is done by using "E-Mail Notification" checkbox in the page's toolbox before using the "Commit Project" button. It should be noted that e-mail sending feature may not be available in all SOMA2 installations.

4. Project Overview and Results

This chapter describes the functions of SOMA2 environment after a project is submitted for the computation.

4.1. Post-Commit Navigation

After the project is submitted the navigation menu changes a little (Figure 9). Instead of “Add Molecules”, “Submit Project” and the workflow dependent program links, “Project Overview”, “Results” and “Program name / View Config” links are shown. The details of these altered navigation items are explained in the following chapters. Navigation menu items “Project List”, “User Info” and “Help” and their functions remain the same.



Figure 8. Post-commit navigation menu.

The “post-commit” navigation is also shown on every page of the committed project and can be used similarly as the “pre-commit” navigation; using the navigation menu items to move from one view to another is possible at any time. Also here, current view is shown in boldface in the navigation menu.

4.1.1. Project Overview

Using the “Project Overview” link always loads a page showing status data of the current project. From this page user also accesses the results of the project. See chapter 4.2 for the details.

4.1.2. Program / View Config

After a project is committed, workflow dependent program links in the navigation menu are divided in two parts. Clicking the program name link loads results of the program step. Result link is active only when the program step is in “finished” status. For the details of viewing the results, see chapter 4.3.

Using the “View Config” link loads the program configuration page showing the parameters that were configured for the program before committing the project. The program configuration page is loaded in view only mode so all the parameters are disabled.

4.1.3. Results

Using the “Results” navigation menu item will load results of project. The link is active only when the project status is “finished”. For the details of viewing the results, see chapter 4.3.

4.2. Project Status View

The first page that is automatically loaded after “Commit Project” button is pressed on the “Submit Project” page is “Project Overview” page (Figure 10). The “Project Overview” page operates as the main page of a project after the project is committed. It also is the first page that is shown if user selects a committed project from the project list as mentioned in chapter 2.

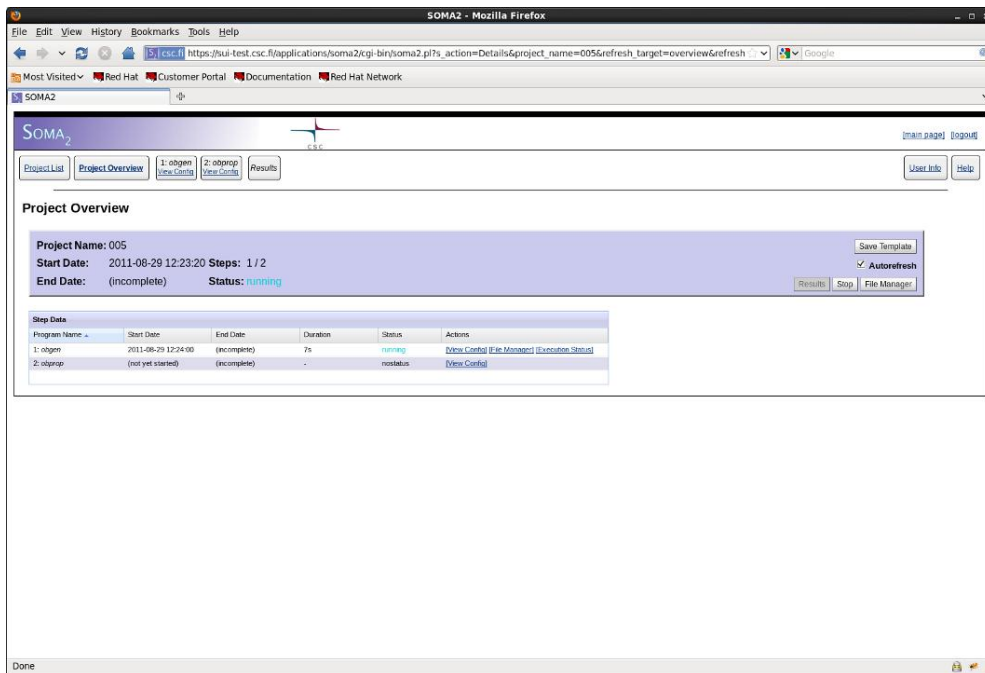


Figure 9. “Project Overview” page.

The main functions of the “Project Overview” page are to show the total project status, statuses of each individual program steps that the workflow includes and offer access to the results. For the details of viewing the results, see chapter 4.3. The general design of the “Project Overview” page is to first show the overall project info in a toolbox followed by detailed information of the program steps included in the workflow. The program step information is in a table, presented one program per row under title “Step Data:”. The project info consists of project name, overall project status, current program number that is being executed out of all and start/end date info. The program step data is very similar, only the status info is for an individual program and executed program number info is excluded. In addition, step data contains duration column for a program. The duration info contains all pre- and post processing operations and computation time for a program.

In addition, a step data for a program includes action links. “[View Config]” link is always available for all the program steps. The link can be used to load the program configuration page of the step, similarly as used from the navigation. “[File Manager]” link appears instantly when execution of a program step is started. The link can be used to directly access the file system and see the files associated with the program step. It should be noted that file manager utility may not be available in all SOMA2 installations. During the program execution, also “[Execution Status]” link may be available. The availability of this link depends on a particular program. Using the “[Execution Status]” link provides real time information of the program execution on the target computing platform. As in the navigation menu, also in the step data table the program names contain links. The links can be used to access results of the program step. The link in a program name is active only when the program step status is “finished”.

The possible statuses for the project include “nostatus”, “initialising”, “running”, “finished”, “failed” and “stopped”. A program step can have statuses “nostatus”, “initialising”, “queued”, “running”, “finalising”, “finished”, “failed” and “stopped”. Meaning of the statuses mentioned in previous are, respectively: program step or project not yet started, program input is being prepared for the execution, program is queuing for the execution, program or project is running, program execution has ended and output data is being processed, program step or project is finished successfully, program step or project is failed and program step or project is stopped. The statuses of the program steps and the project are automatically updated once

project data changes, and there is no need for a manual page refreshing. The automatic page refresh procedure can be controlled by using “Autorefresh” checkbox in the toolbox of the page.

While the project has status “initialising” or “running”, the toolbox contains an active “Stop” button that can be used to stop the running project. Using the “Stop” button stops the project and all the programs steps. If the project ends with “finished” status, the “Stop” button changes inactive and button “Results” becomes active in the toolbox. Similarly as the “Results” navigation menu link, the button can be used to access all the results from the workflow. In case of “failed” project, a link in the name of the last successful program step can be used to access the existing results. If project has only one program step that ends with “failed” status, “Project Overview” page shows “Molecule Files:” list which can be used to access initial project input molecule files. If the project or a program step ends with “failed” status, the “[File Manager]” links in the table or the corresponding button in the toolbox can be used to access the file system and possible error log files of the project or program. These log files can point out the reason for failed program run. It should be noted that file manager utility may not be available in all SOMA2 installations.

4.3. Viewing Results

In this chapter, details of viewing the results of a project are explained. The contents of the “Results” page depend on if results of the whole workflow or a step results are requested by using the navigation menu items or the action links in the “Project Overview” page, as described in the earlier chapters. A step results include all the results produced so far, including the original input molecules (Figure 11). The results of the whole workflow include all the data produced, including the original input and output of all the program steps.

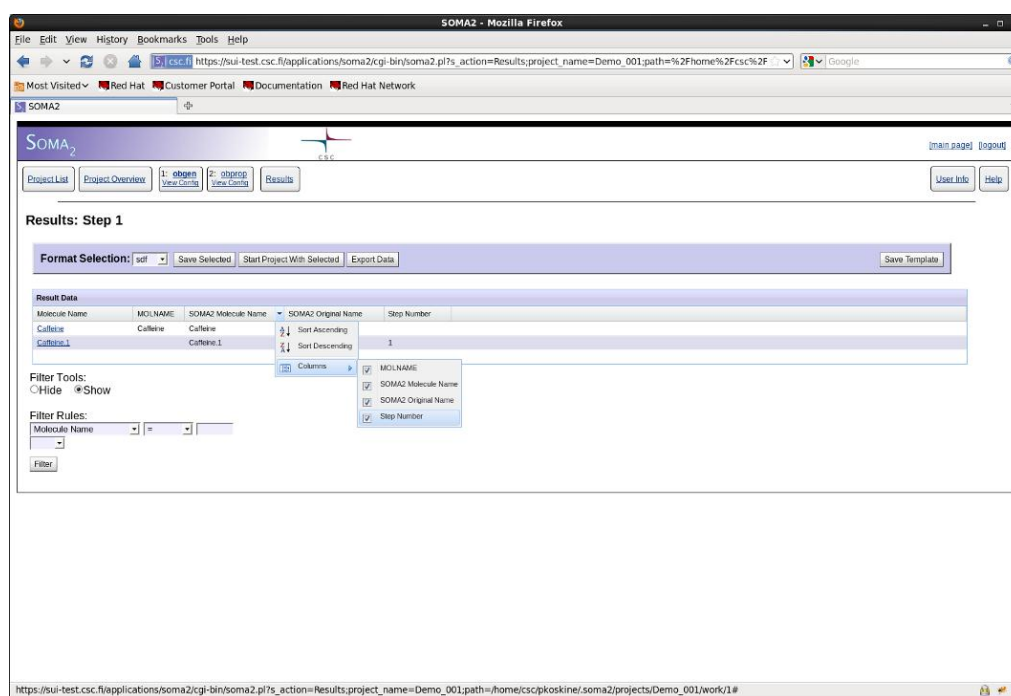


Figure 10. “Results” page.

The result molecules are presented in an interactive table that is described in detail in chapter 4.3.1. The “Results” page also shows a toolbox including tools for, saving the result molecules, starting a new project with selected results and exporting contents of data table. Below the result table, there is a radio group titled “Filter Tools:” which can be used to filter the result data. All the previous functions are described in the following chapters.

4.3.1. Result Table

“Results” page of SOMA2 environment offers a spreadsheet like interactive result table that includes the result molecules, one on each row. The result table is titled “Result Data”. The result table has the same features as the other data tables in the SOMA2; the table can be sorted according to property data field by clicking the corresponding column header and if needed, the table has its own scrollbars.

The first property data field of the table is always “Molecule Name” that is the molecule name record. The molecule name also acts as the identifier of a result. In the default view the results are sorted in alphabetical order according to the “Molecule Name” field. In the result table, the original input molecules are presented with their original names. Since the original input is the input for the first program and its output is the input of the second one and so on, names of all the result molecules of the different programs are derived from the name of the corresponding original input molecule. The name of the descendant molecule is always formed by adding a suffix “.1”, “.2”, “.3” and so on, according to number of the descendants, to the name of the parent molecule. This way, the results are presented as generations of different original input molecules which act as the ancestor molecules.

4.3.2. Property Data Fields

In addition to “Molecule Name”, the default property data fields for all the results include “Step Number” and “SOMA2 Original Name”. The default property data fields in the table are included to help in identifying and filtering the results.

The “Step Number” indicates the program from which the result originates. It can be used to easily sort the results according to the program. The “SOMA2 Original Name” property data field can be used to identify the ancestor of the result. The rest of the property data fields in the table depend on the programs that are included in the workflow and the original input molecules.

If a program produces data it is shown in the table as a property data field that is always named in format “Program name: Property Name”. If the original input molecule files contain properties, they are shown in the result table. Usually these properties are shown only on the original input molecules and the properties have no program name included.

The property data fields in the table can be hidden or shown according to user selection (Figure 11). A downwards pointing arrow icon will appear on any of the column headers of the data table once mouse pointer is moved over one. Clicking this icon will bring up a menu showing “Columns” option. Choosing this option will show a list of all property data fields available for the table. Using provided checkboxes, visible data columns can be changed interactively. By default, all the property data fields are shown.

Property data fields of the result table can be adjusted to desired order by using drag and drop mechanism. This is done by moving mouse pointer over column header. After this, mouse button is kept pressed and column is dragged to desired position.

Some of the molecules in the result table can also include property data fields that the program from which the results are from, does not provide. This phenomenon is due to the inheriting of the properties. If some of the properties are checked to be inherited in “File Details” view described in chapter 3.3.5 or a program, according to its internal knowledge in the SOMA2, produces inheritable data, the corresponding property data fields are automatically transferred to correct results. If a program that is integrated to the SOMA2 environment produces inheritable data, the data can be inherited both up and/or down in the molecule generations, according to content of the data.

4.3.3. Using Filter Tools

“Results” page of SOMA2 environment also offers a possibility to filter the result data shown in interactive table. There is a radio group form element titled “Filter Tools:” below the result table. By the default, the “Filter Tools:” is in “Hide” mode. User can enable the filtering by selecting “Show” from the radio group.

After enabling the “Filter Tools:”, a filter query is built using the form elements that appear to the page below title “Filter Rules:”. The filter query is built in a form: “property operator value/range”. Selecting logical operator “AND” or “OR” from smaller popup menu reveals additional rows of form elements that can be used to make a query more complex. The first popup menus automatically include the property data fields that are currently shown in the result table. Currently, six query sentences can be combined by using the logical operators.

Using “Filter” button executes the actual filtering task (JDB ASCII data filtering tools, <http://www.isi.edu/~johnh/SOFTWARE/JDB/index.html>) and the result table is reformed according to the filtered data. After the first filtering task, page also includes “Reset View” button that can then be used to reset the result table to its original content.

4.3.4. Exporting Data Table

The result data shown in interactive table can be exported to user as ASCII data. This option is useful, if the user wishes to post process the data, for example, in a spreadsheet calculator application.

Toolbox of “Results” page includes “Export Data” button that can be used to export the data table. The download of the data file starts automatically when the button is used. The data in the exported file is a plain text ASCII data with tabulator as a field separator (TSV data), and it is directly importable to the common spreadsheet applications.

Using the “Export Data” button stores the current view of the result data table. If the user has employed “Filter Tools:”, the data table must be reset in order to export all the available data.

4.3.5. Visualisation of a Result

Each individual result molecule that is shown in the result table of SOMA2 environment can also be visualised (Figure 12 and Figure 13). Clicking the link that is embedded to the property data field “Molecule Name” of a result loads “Result Details” page showing the molecular structure and summary of all the property data fields of the selected result.

Result Details: "Caffeine.1"

Parent: [Caffeine](#)

Select Viewer:

Viewer Sketcher



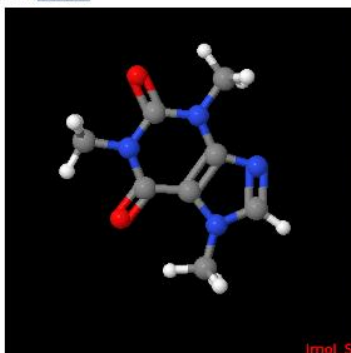
Property	Value
SOMA2 Molecule Name:	Caffeine.1
SOMA2 Original Name:	Caffeine
Step Number:	1

Powered by ChemAxon

Figure 11. Visualisation of a result molecule using Marvin

Result Details: "Caffeine.1"

Parent: [Caffeine](#)



Property	Value
obprop: Canonical SMILES:	Cn1cnc2c1c(=O)n(C)c(=O)n2C
obprop: Exact Mass:	194.08
obprop: InChI:	InChI=1S/C8H10N4O2/c1-10-4-9-6-5/107(13)12(3)8(14)11(6)2/h4,1-3H3
obprop: logP:	-1.0293
obprop: Molecular Formula:	C8H10N4O2
obprop: Molecular Weight:	194.191
obprop: MR:	52.039
obprop: Number of Atoms:	24
obprop: Number of Bonds:	25
obprop: Number of Residues:	0
obprop: Number of Rings:	2
obprop: PSA:	61.82
obprop: Sequence:	-
SOMA2 Molecule Name:	Caffeine.1
SOMA2 Original Name:	Caffeine
Step Number:	1

Figure 13. Visualisation of a result molecule using Jmol

The SOMA2 interface currently supports two programs, Marvin (<http://www.chemaxon.com>) and Jmol (<http://jmol.sourceforge.net/>). If Marvin is the viewer applet program in the current installation, it is possible to switch viewing mode between Viewer and Sketcher Applets using the radio group titled "Select Viewer:". The default is to show 3D molecular structure in the Viewer. Jmol provides only a viewer applet, so no switching is provided in this case. All of the visualising tools include a possibility to directly save the current result molecule to the local computer. In Marvin the saving is performed by clicking right mouse button and selecting "Save As" from the menu (Viewer) or choosing "File/Save As" directly from the menu (Sketcher). In Jmol the molecule can be saved from lower right corner's Main Menu (appears when clicking Jmol_S) and choosing "File/Save...", where Save is accompanied by several different saving options. The "Result Details" page also shows a whole molecular generation and the parent and children molecules of the current result molecule. These are presented as links associated with titles "Parent:", "Siblings:" and "Children:". Using the links directly load the "Result Detail" page of the selected molecule. If there are many Siblings or Children for a molecule, they are shown with a popup menu. Choosing a molecule from the menu will load the "Result Detail" page of the selected molecule.

4.3.6. Visualising Additional Structure Files

In case of certain applications, SOMA2 environment offers a possibility to visualise additional structure data files. Currently visualisation of optimisation trajectory files from geometry minimisation, protein-ligand complexes from docking simulation and template molecules from alignment search are supported if installation uses Marvin. Only trajectory files apply to installations using Jmol. If the used viewer application is Marvin, there can be additional visualisations available and the corresponding

property data fields are supplied with links. These links are also available in the property data fields summary shown on the “Result Details” page. Clicking the link loads a page similar to the “Result Details” page described in chapter 4.3.5. The difference in the page functionality is that user is not able to switch between the viewer and sketcher tools. Also above “Parent”, “Siblings” and “Children” links there is a title “Original Molecule” accompanied with a link that is used to load the “Result Details” of the actual result molecule.

In case of visualising the additional structure files, the viewer of the “Result Details” page depend on the application. The optimisation trajectories are visualised similarly as the actual result molecules. In case of the docking simulation or alignment search, a viewer applet capable of displaying several molecules simultaneously is used (Marvin Space, <http://www.chemaxon.com>). Details and help for using the applet can be accessed by selecting “Help” from the applet menu.

4.3.7. Viewing Images and Data Files

Many scientific applications produce data that can be presented graphically. In SOMA2 environment, it is possible to view images from certain property data fields. In some cases, also output files of programs or some parts of them can be viewed directly from the SOMA2 results. If property data fields containing images or data files exist, they are also presented as links in the main result table. When the link is clicked, the content is opened in a separate window. This also enables viewing multiple images or data files simultaneously.

4.3.8. Saving Result Molecules

After evaluation of the result molecules, user can save some or all of them to the local computer in selected molecule format. To perform the save, first the molecules are selected from the result table by selecting the rows. Using control/shift+mouse button enables the multiple selections. Then a desired molecule file format is selected from the toolbox of “Results” page. Clicking “Save Selected” button from the toolbox starts a direct download of the selected results. All the selected results are saved in a single file.

4.3.9. Starting a New Project with Result Molecules

The selected result molecules can also be used to directly start a new project in the SOMA2 environment. First the desired molecules are selected from the table by selecting the rows similarly as saving the result molecules. Then using “Start Project With Selected” button from the toolbox of “Results” page switches the view to the “New Project” that is described in chapter 3.1.

The only difference on the shown page is that it includes the “Project Overview” navigation menu item that can be used to return back to the current project. If user chooses to proceed, a fully new project creation is initiated.

5. SOMA2 Citation

If SOMA2 environment is used, the users are kindly asked to acknowledge it by using the following citations:

Kinnunen, T., Nyrönen, T., Lehtovuori, P., SOMA2 - Open Source Framework for Molecular Modelling Workflows, *Chemistry Central Journal*, 2(Suppl 1):P4, (2008).

Lehtovuori, P., Nyrönen, T., SOMA - Workflow for Small Molecule Property Calculations on a Multiplatform Computing Grid, *J. Chem. Inf. Model*, 46(2) (2006) 620-625.

In addition, the integrated programs used are cited by the instructions provided by the program vendors.

6. *Technical Requirements*

For proper operation of SOMA2 modelling environment, the client computer needs to satisfy the following requirements:

WWW-browser must be available.

JavaScript support must be enabled in the used browser.

WWW-browser Java Plug-in needs to be installed.

Cookie support in browser must be enabled (if server uses cookie authentication).

The SOMA2 modelling environment is tested to be compatible with Internet Explorer 8/9, Firefox 3 and later, Opera 11 and Chrome 14 and later web browsers.