# QuickStart Guide to

# *eTOXlie: A framework for iterative Linear Interaction Energy calculations created at VU University Amsterdam*

This document is intended to provide a quickstart guide and a brief update on the advanced features introduced in the framework developed by VU Amsterdam within the eTOX project. The Implementation was previously described in the document *eTOXsys model implementation by VU University Amsterdam*.

The workflow described in the above mentioned document has been revised and made independent from the eTOXlab framework to allow for higher flexibility. Although some workflows have been modified, the main concepts described there are still valid.

01/06/17

Luigi Capoferri

l.capoferri@vu.nl

# Main changes

1. The code has been completely revised since the previous version, and reorganized in two independent parts: an interface that is used to submit the jobs and a job manager that computes the properties (interaction energies) for each compound submitted to screening. Representation of the architecture is presented in Fig 3.
2. A web interface has been created to allow for an easy handling of the jobs.

The new interface allows for implementation of advanced features, of which some have been already implemented (e.g. monitoring of the jobs, creation of new models, etc), others are in the pipeline (e.g. stop jobs, download most probable binding pose, etc).

1. Modules for the configuration of new models has been implemented in order to create all the files required for the execution starting only from the coordinates of the protein (off)target provided as pdb.
2. Simulations for the single poses are performed in multiple replicas with different starting velocities for evaluation of statistical convergence.

# Macintosh HD:Users:luigi:Science:DOTTORATO:AMSTERDAM:writing:VUALIE:untitled folder:Capoferri_LIE:Slide09.tiff

Figure 3 eTOXlie software architecture

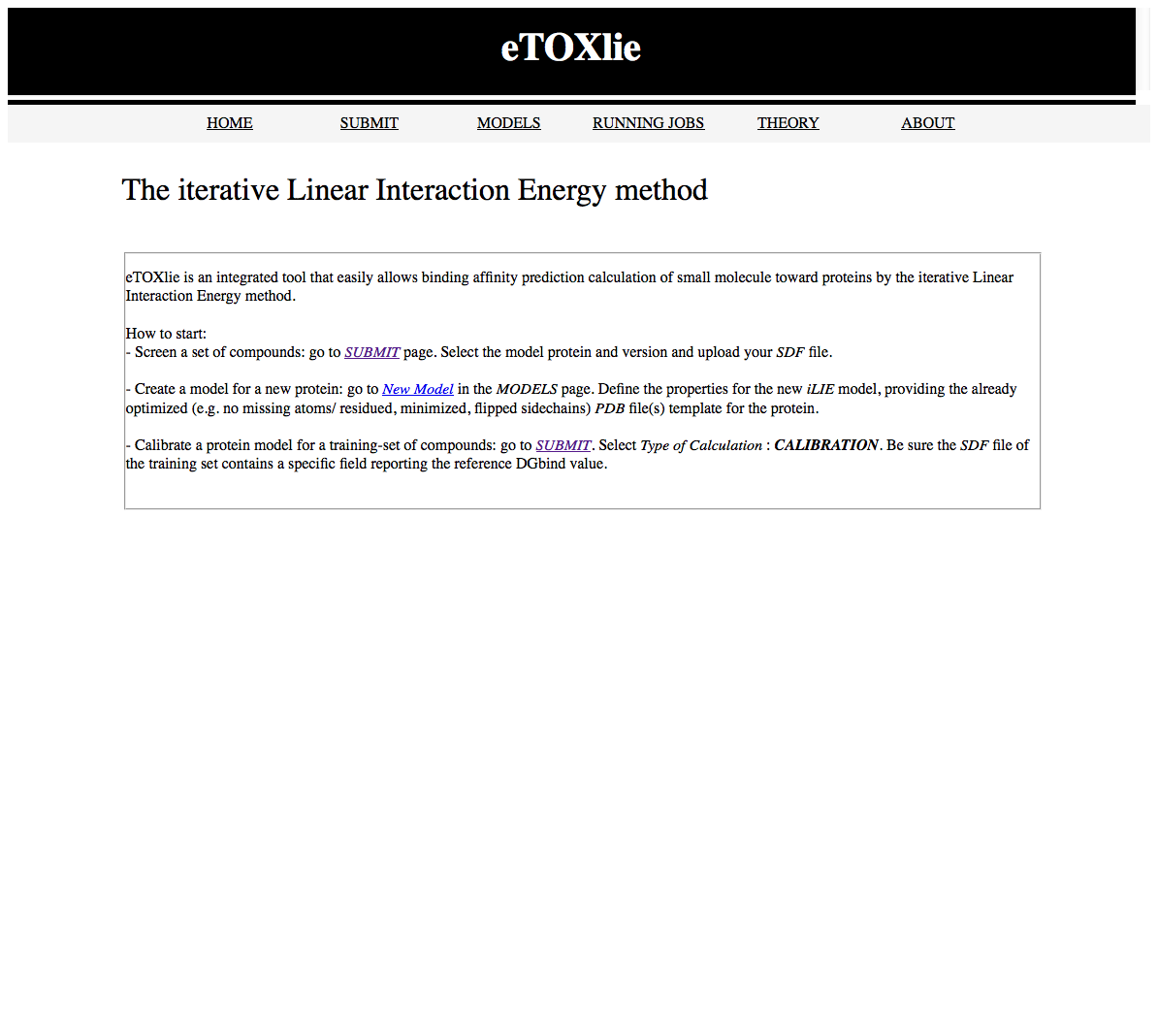
# Brief Guide to eTOXlie interface

It is possible to create new models and submit and monitor jobs from a web interface. The web interface is accessible with a normal web browser at the page:

*Localhost:5000* on the VM. If the VM is configured as suggested (port forwarding configuration, see above), from the machine hosting the VM, it is possible to access the interface via the address: http://localhost:5555/

## HOME

From the home page links to the common operations are directly available also through the main text



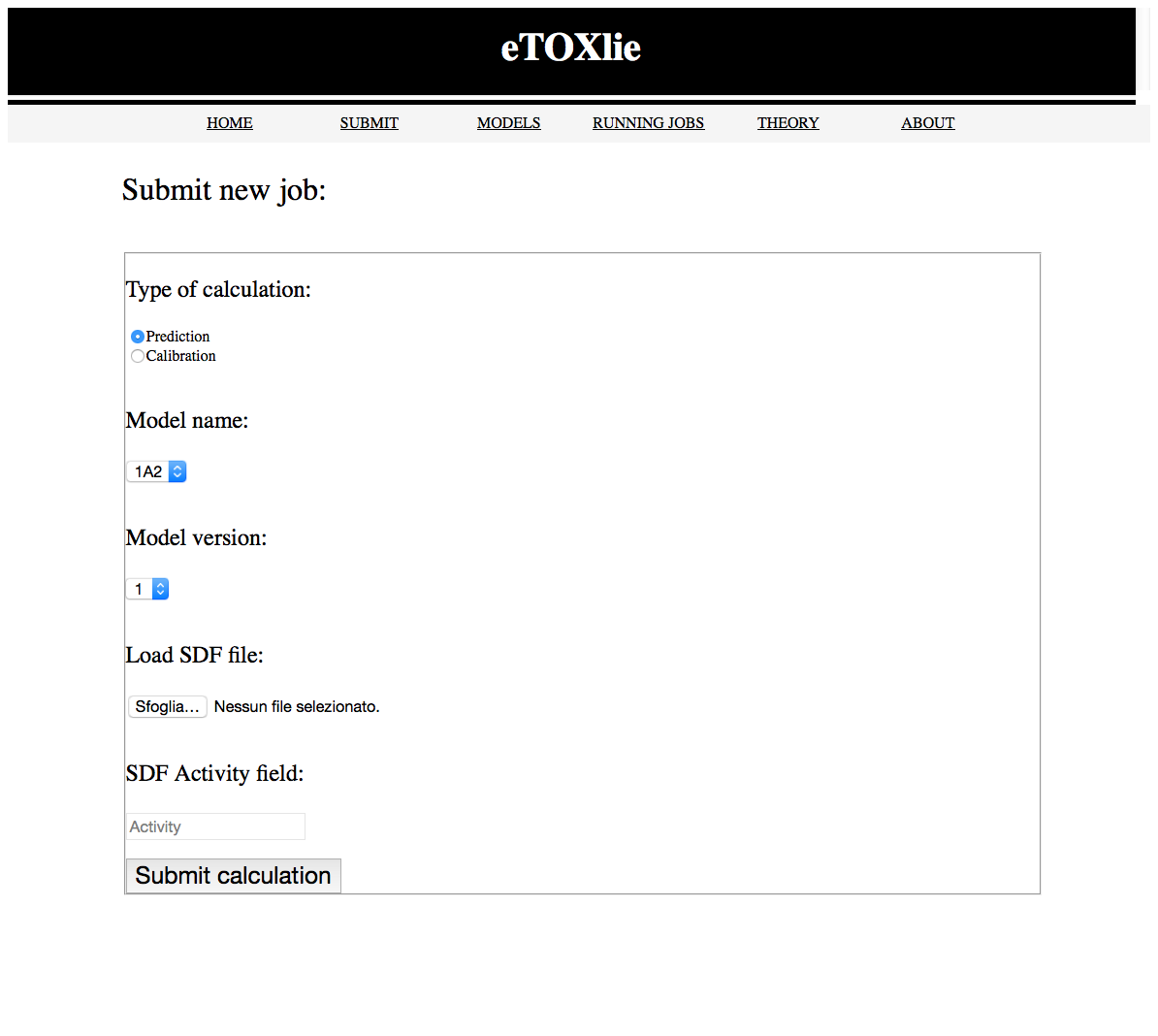
## SUBMIT

Submission of new jobs are here available.

The input must be provided via SDF file and loaded through the window *Load SDF file*.

Two *Type of Calculation* are available:

* *Calibration*: calibrate a new version of a model. The model must be previously configured (protein files, docking center, etc) and must be selected in the list *Model Name*. To configure a new model see *MODELS 🡪 Create New*. The new version will be created automatically according to the number of versions already present. Each molecule in the input sdf file should contain a field including the experimental ΔGbind. The name of this attribute MUST be defined in *SDF Activity Field* form on the submission page.
* *Prediction*: compute a prediction according to the selected *Model name* and *Model version*. The models available are listed at loading of the page. Selecting a model, the version available for that model will appear. If no versions are available the model needs to be calibrated first.

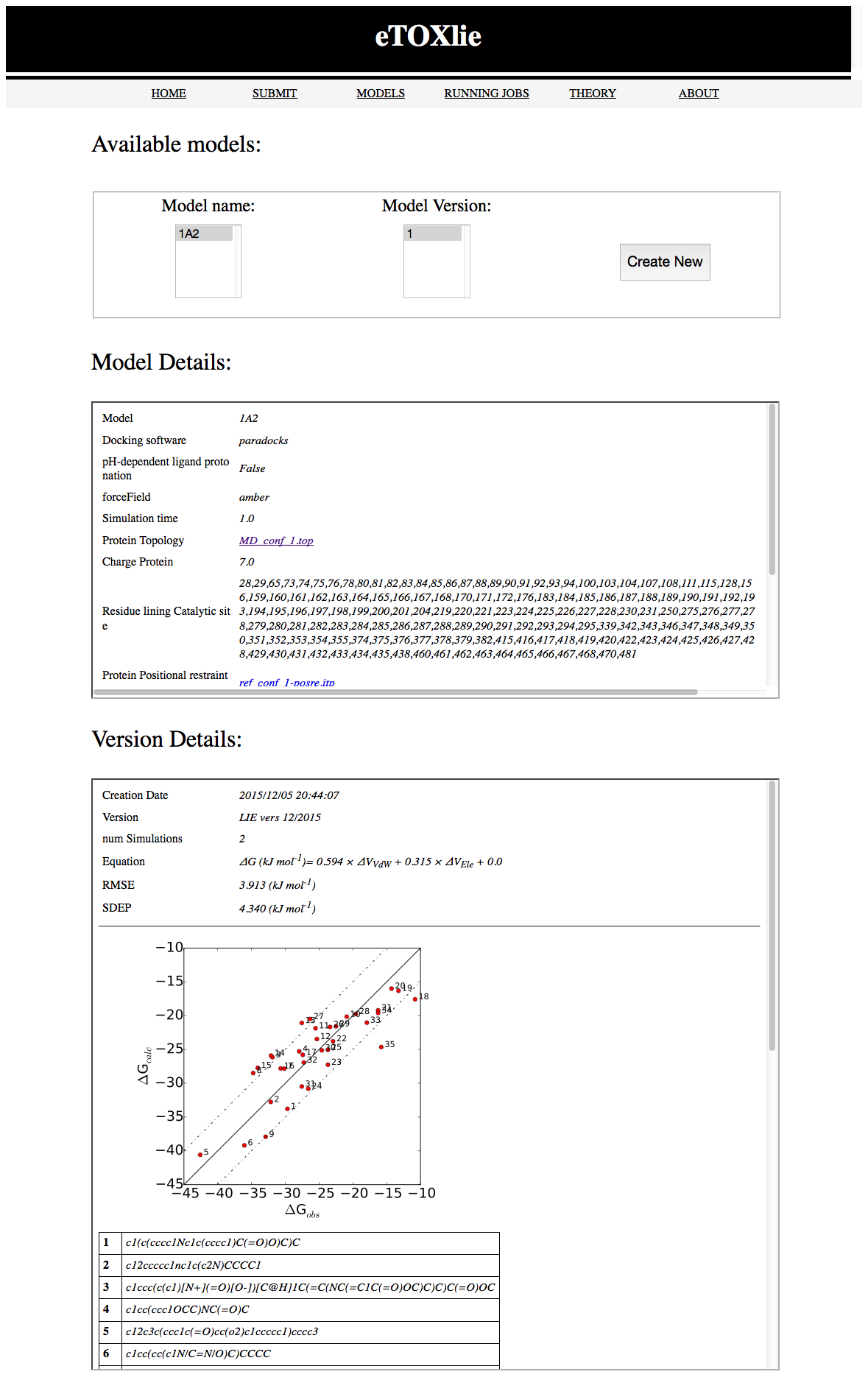


## MODELS

Models available can be inspected. Selecting the *Model name* its configuration will be shown in the *Model details* form. Configuration files (e.g. topology file, docking template, etc) used by the model can be downloaded and inspected here.

Selecting a version of the model, the Version Details form will show statistics of the model and the training set used to train the model (displayed as SMILES). Chemical structure of each single molecule can be visualized clicking on the SMILES representation.

Clicking on the *Create New* button a new model can be configured.



## Create New

From this page new models can be created.

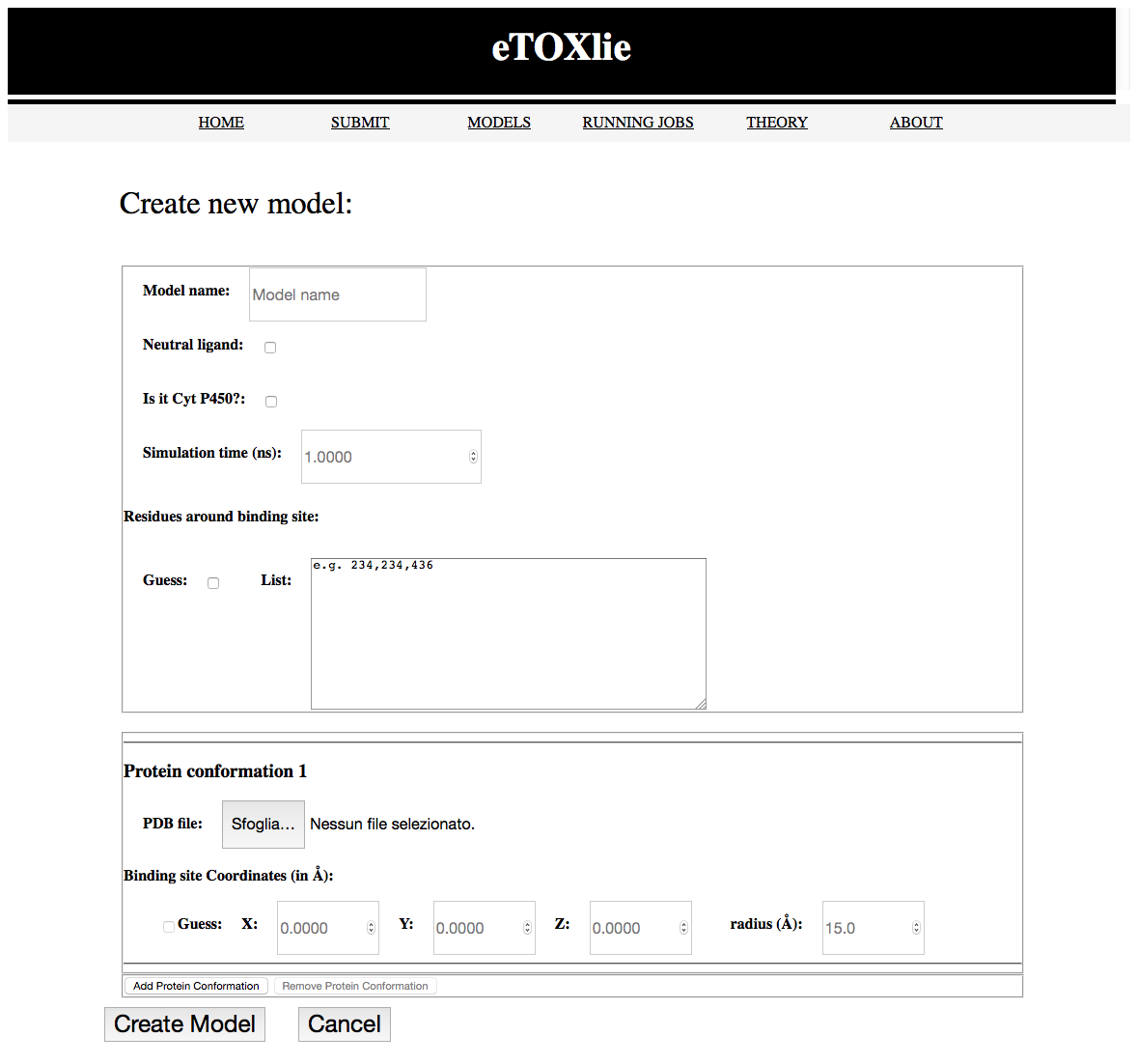
Every field should be filled.

If *Neutral ligand* is selected, the ligand will be optimized in the non ionic form. Otherwise, the ligand will be protonated according to pH 7.4.

If the protein is a Cytochrome P450, the box *Is it Cyt P450?* can be checked. This allows for automated detection of the center of the catalytic site (for docking) based on the position of the heme residue.

Simulation time indicates the length of each single simulation. The total number of simulations depends on the number of representative poses obtained during the automated analysis of the poses obtained during docking.

The list of residues around the binding site is used for the calculation of the reliability index of a prediction. A list of residues can be provided (e.g. “234,324,57,2,436”). Ticking the *Guess* box, residues will be automatically selected that are within 16Å of the center selected for the docking.



To create a new model, pdb file of the protein must be submitted. All the files required for the calculation (gromacs coordinates and topology, docking template, etc) will be created automatically. However no refinement of the pdb is performed (e.g. missing atoms, side chains reorientations, removal of other ligands).

Coordinates for the docking center and radius for the docking sphere must be defined. In case of *Is it Cyt P450?* Is selected, it is possible to tick *Guess* for computing automatically the center of the docking (the radius of the docking sphere should always be defined).

A model using multiple protein conformations can be created. *This function of the interface is still to be tested.*

## RUNNING JOBS

The status of the jobs can be monitored from this page. Every screening show the model and version used, date of submission, the overall status and the type of the screening (*CAL*: calibration of a new model; *PRED*: prediction of query compounds).

Clicking on the *(+)* on left, details of the compounds submitted to screening are shown. Clicking on *Click Me*, chemical structure of the compound is shown in a new window of the browser. ΔGEXP is shown for calibration runs only, ΔGCALC at the end of the calculation. *CI* indicates the confidential interval of the prediction (0: high reliability, error within SDEP obtained during cross validation – 5: not reliable). Since the method has not been extensively tested, for the moment intermediate values should be carefully taken into account.

In case of local execution of the machinery the status of the job that will appear will be limited to “*SUBMITTED*, *DONE*, *FAILED*”. Intermediate values should not be visualized in this mode.

