# Collector v1.5.0 setup and quickstart manual

# Installation of collector

Collector is distributed as a tgz compressed file.

Uncompress it and follow the installation instructions in the README file.

# Starting Collector web application server

Open a terminal window.

Move to the collector installation path. For example:

[[collector@collector](mailto:collector@collector) ~]$ cd /opt/collector

Start the web application by typing:

[[collector@collector](mailto:collector@collector) ~]$ startCollector

Now you will see on the terminal the Collector starting messages.

# Stopping collector web application

Open a terminal window.

Move to the collector installation path. For example:

[[collector@collector](mailto:collector@collector) ~]$ cd /opt/collector

Stop the web application by typing:

[[collector@collector](mailto:collector@collector) ~]$ stopCollector

# Troubleshooting starting web application

If you find difficulties starting the application you can try to stop previous collector instances that were already running:

[[collector@collector](mailto:collector@collector) ~]$ stopCollector

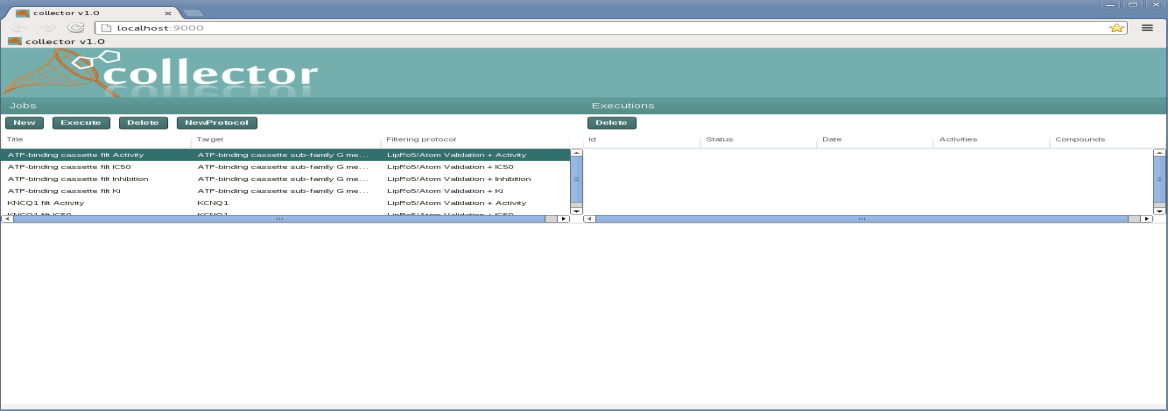
[[collector@collector](mailto:collector@collector) ~]$ startCollector

# Accessing Collector web application

Open a web browser and enter the web address:

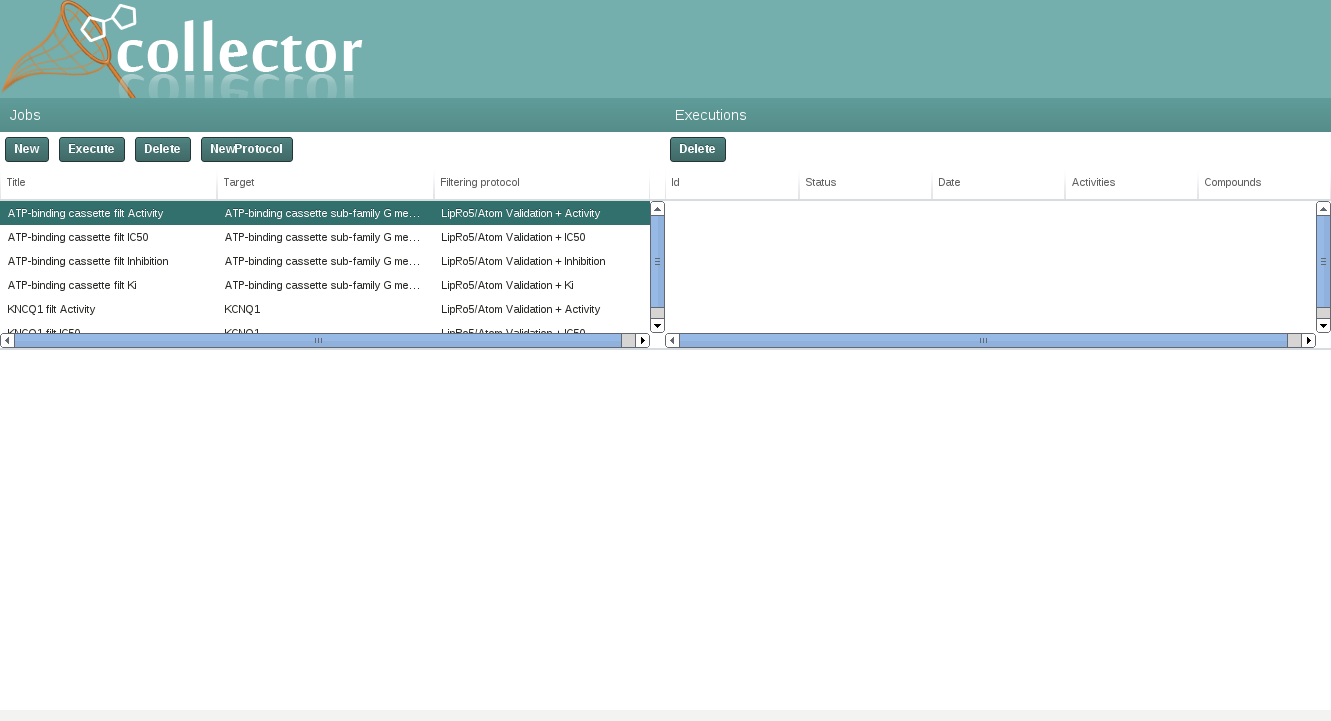
<http://localhost:9001>

After loading the page you should see something like:

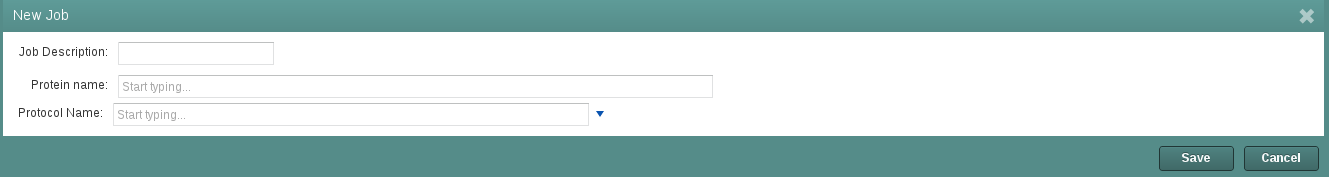


# Collector web user functions

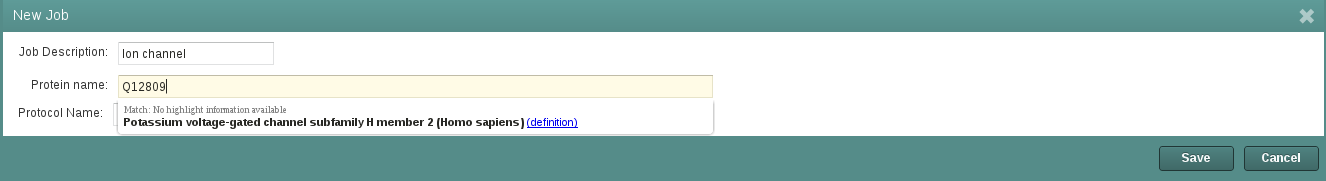
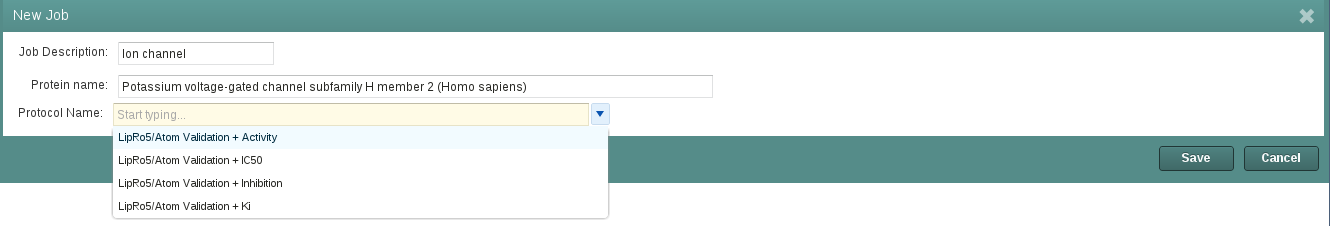
## Defining a new job



To define a new job push the button “New”:

The following form appears:

You have to complete:

* Job Description
* Protein Name: Enter the name of the target you we are interested in. You can use an UNIPROT accession number like Q12809 or a target name:
* The protocol to apply to the obtained data. You can select it from a list of predefined filtering protocols:
* LipRo5/Atom Validation + Activity
* LipRo5/Atom Validation + IC50
* LipRo5/Atom Validation + Inhibition
* LipRo5/Atom Validation + Ki

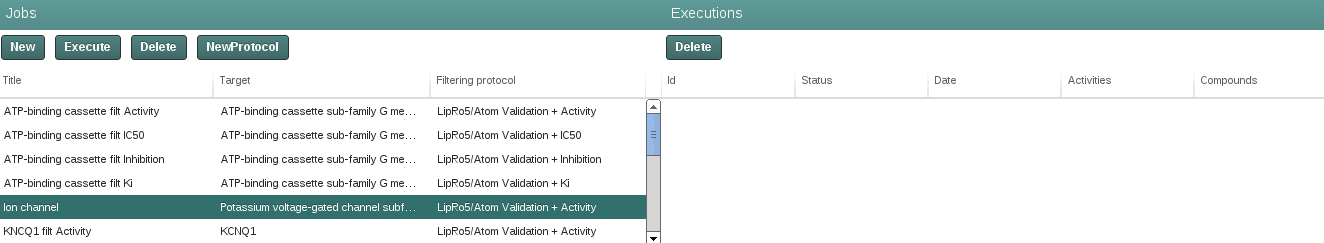
Each filtering protocol filters by a different activity type: Activity, IC50, Inhibition and Ki.

All these protocols apply two additional filters based on chemical properties:

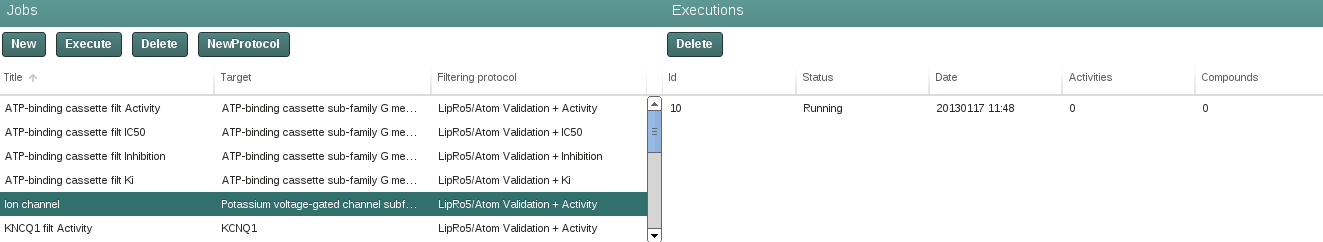
* Lipinski Rule of 5: compounds that do not meet “Lipinski Rule of 5” are filtered out
* Atom validation: compound containing any atom different from H, C, N, O, S, P, Cl, I, Br, F are filtered out

## Executing a job

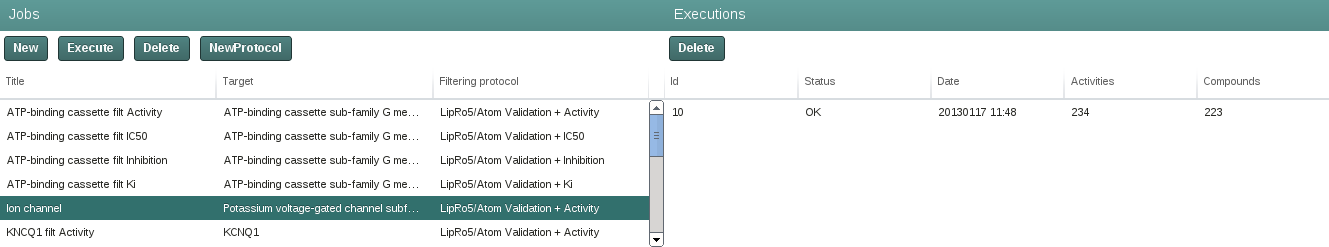
Once the job is defined you can execute it. You have to highlight the job in the “Jobs” panel and push the button “Execute”:



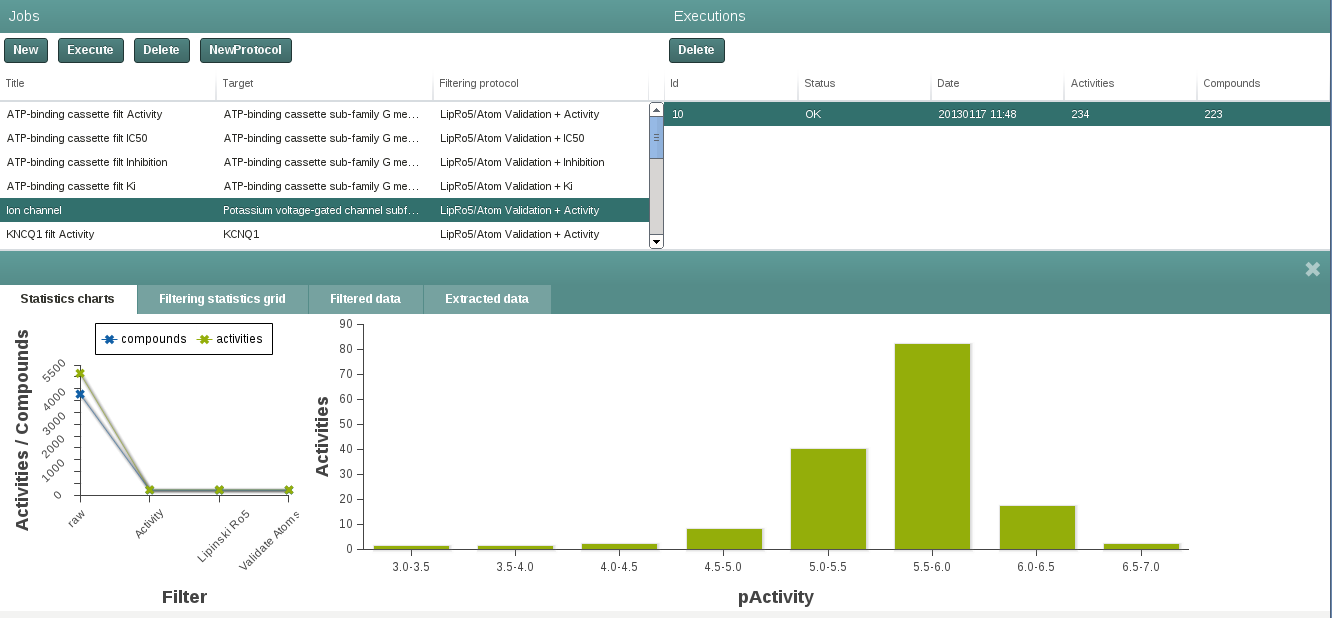
The job execution starts:



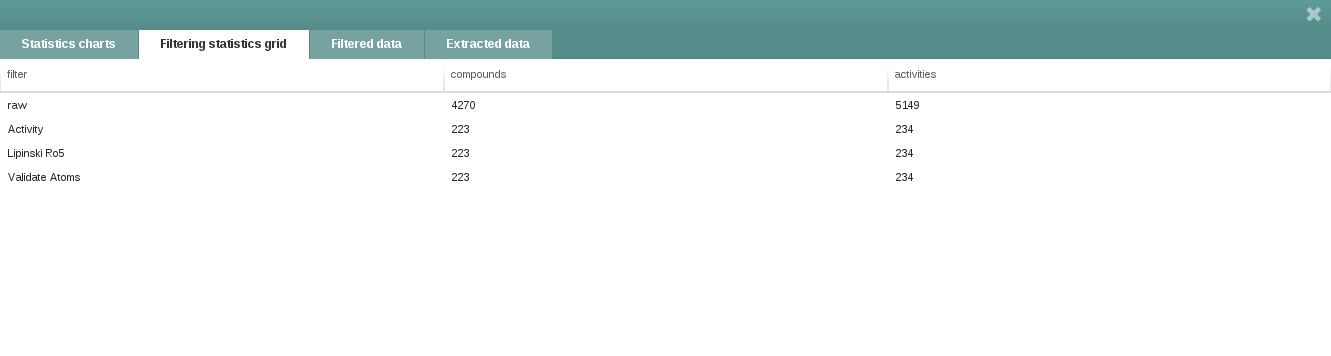
After some time the execution finishes showing “Status” OK:



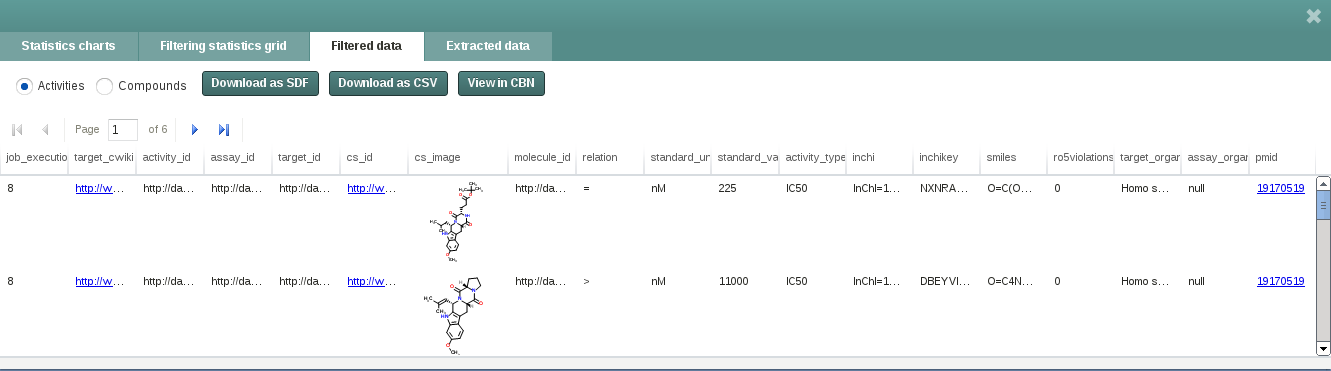
## Browse and export job execution data

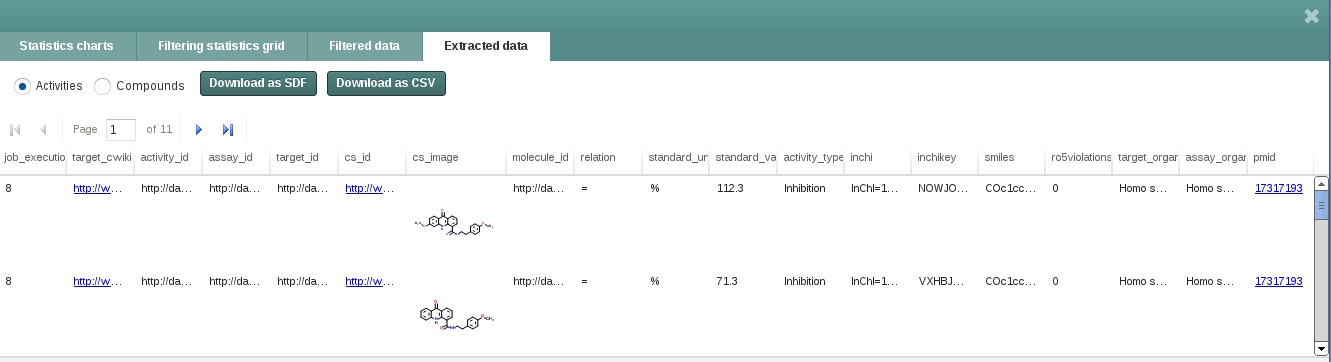
Now you can browse the extracted data clicking on the panel “Executions”:

In the first tab you can see, on the left hand side, a line chart showing the number of compounds that passed the different filters of the protocol. On the right hand side you can see a histogram of the distribution of the parameter of activity obtained.

The second tab shows a table representing the same data of the line chart in tabular form:

Please note the different “compounds” and “activities” counts; compound indicates the number of different compounds obtained and activities the number of activity annotations extracted. Very often both figures are different, since the most databases contain several activity annotations for the same compounds.

In the third and fourth tab we can browse the raw data extracted or the filtered data:



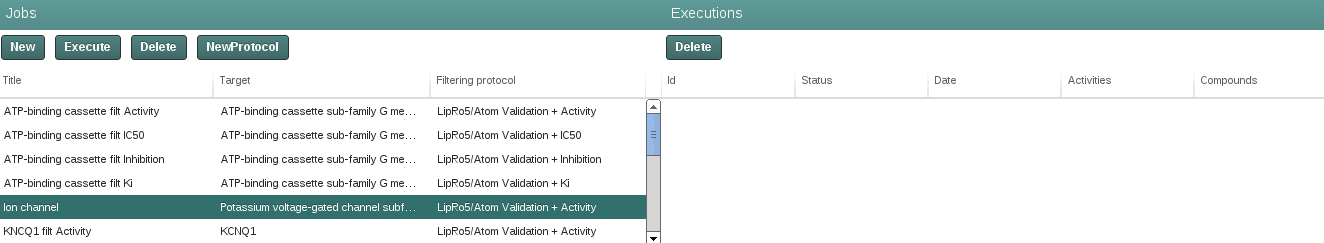
You can decide what to export by pressing Activities/Compounds radio button:

In the case of compounds, when we have different measurements for the same compound, Collector computes the median of the different values

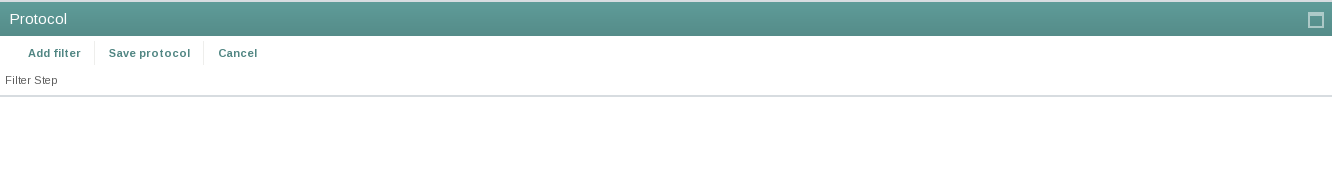
For exporting the results you can press the buttons:

* Download as CSV: if you want the data in plain text tab-separated format
* Download as SDF: in 2D SDF format

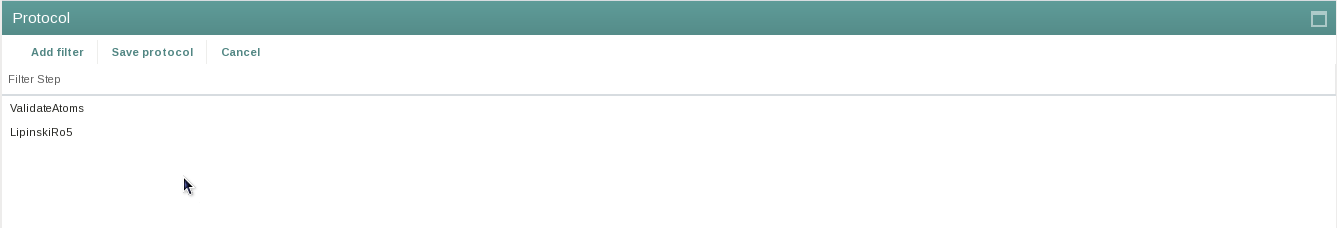
## Define new Protocols

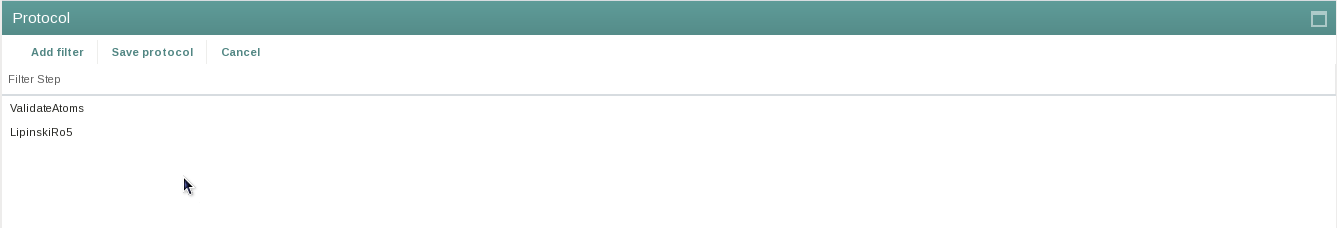


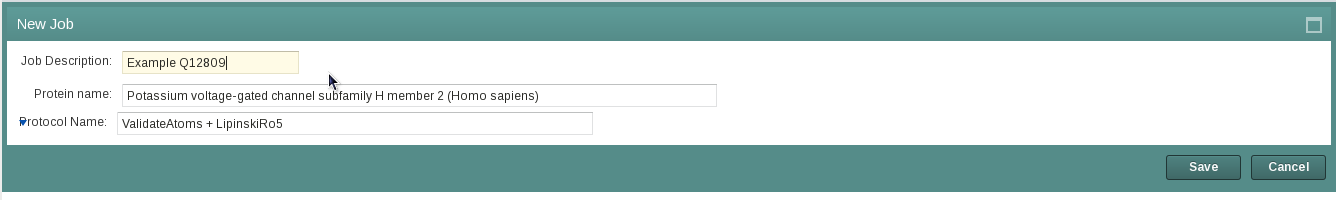
You can define a new protocol by pushing the NewProtocol button:

You can see a new form:

This form allows creating a new protocol by adding filters to the protocol. For example we can add a new filter:



When the user finishes adding the filters to the protocol, you can save it by pushing the “Save protocol” button:

Once the protocol is saved you can define a new job that uses the newly defined protocol:

# Collector Command line

Collector has a full command line interface.

To access it you have to open a terminal. Type “collector” to get a brief description of the commands available:

[[collector@collector](mailto:collector@collector) ~]$ collector

Collector command line commands:

collector listprotocols

* Lists the protocols available in the system. In the current implementation:

|  |  |  |  |
| --- | --- | --- | --- |
| job\_filtering\_id | job\_filtering\_description | filter\_description | curation\_order |
| 1 | No filtering | NoFiltering | 1 |
| 2 | LipRo5/Atom Validation + Activity | Activity | 1 |
| 2 | LipRo5/Atom Validation + Activity | LipinskiRo5 | 2 |
| 2 | LipRo5/Atom Validation + Activity | ValidateAtoms | 3 |
| 3 | LipRo5/Atom Validation + IC50 | IC50 | 1 |
| 3 | LipRo5/Atom Validation + IC50 | LipinskiRo5 | 2 |
| 3 | LipRo5/Atom Validation + IC50 | ValidateAtoms | 3 |
| 4 | LipRo5/Atom Validation + Inhibition | Inhibition | 1 |
| 4 | LipRo5/Atom Validation + Inhibition | LipinskiRo5 | 2 |
| 4 | LipRo5/Atom Validation + Inhibition | ValidateAtoms | 3 |
| 5 | LipRo5/Atom Validation + Ki | Ki | 1 |
| 5 | LipRo5/Atom Validation + Ki | LipinskiRo5 | 2 |
| 5 | LipRo5/Atom Validation + Ki | ValidateAtoms | 3 |

Every filtering protocol is identified by job\_filtering\_id. The curation\_order defines the order by which filters are applied. For example, the protocol 2 applies sequentially the filters

* Activity
* LipinskiRo5
* ValidateAtoms

The “No filtering” is protocol that applies no filtering; it only extracts the raw data from the source.

collector newjobuniprotid –-protocolid <protocol\_id> --uniprotid <uniprotaccession> --jobdescription <job\_description>

Defines a new job based on:

* <protocol\_id>: the protocol to apply (obtained in listprotocols call)
* the <uniprotaccession> of the target
* <job\_description>: a descriptive text of the job

collector executejob –jobdescription <job\_description>

Executes the job with <job\_description>

collector listjobexecutions --jobid <job\_id>

Lists all the executions of the given job\_id

collector export --raw **-–jobexecutionid <job\_execution\_id> --datatoexport activities|compounds –exportformat sdf|csv –filename <filename>**

Exports the data obtained in the job\_execution\_id depending on:

* activities: extracts the detailed activity data
* compounds: extracts the detailed activity data aggregated at compound level (median of all the activities reported for the same compound)
* sdf or csv: SDF 2D format or plain text tab-separated plain text file.
* <filename>: Filename to export.