EXAMPLES OF USE: LigRMSD

Link reference: https://ligrmsd.appsbio.utalca.cl/results/20191223220522/

LigRMSD is a free web server for automatic matching and RMSD calculations among identical or similar chemical compounds. For calculations, LigRMSD requires a set of molecules in MOL format (at least two). Once the user uploads the molecules, one of them (at least) must be selected as reference. Then, all molecules will be compared with the chosen references. It is essential to notice that the user must assure that MOL files submitted were correctly generated and the structures standardized. Special attention for the bond types and atom valences of molecules extracted from PDB structures; the PDB format does not include this type of information.

As it is mentioned in the manuscript, there are two assays available in LigRMSD: (i) matching and RMSD calculation between two compounds; and (ii) matching an RMSD calculation of one (or more) query against several reference compounds (data set). For any of these options, two different matching methods are configured. First, strict matching identifies equivalent atoms by considering identical atom types and bonds. Meanwhile, flexible matching contains the strict match and also may include different atom and bond types, only where their inclusion does not affect the chemical graph of the molecules.

I) Matching and RMSD calculation between two compounds:

In this first example, we are interested in comparing the reproduced protein-ligand docking of the structures of three thrombin-ligands complexes (81A, I50 and NLI) of the PDBids 1T4V, 2R2M and 3LDX (Figure 1).

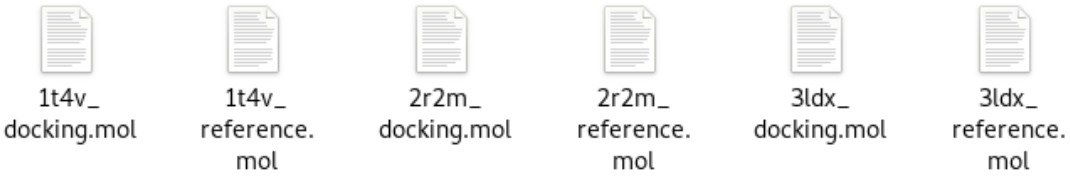


Figure 1: Set of molecules (thrombin-ligands docked and references) in MOL format

It is essential to mention that LigRMSD will only process molecules submitted in MOL format. Also, the server requires that all MOL files must be compressed as a ZIP file (Figure 2 and Figure 3).

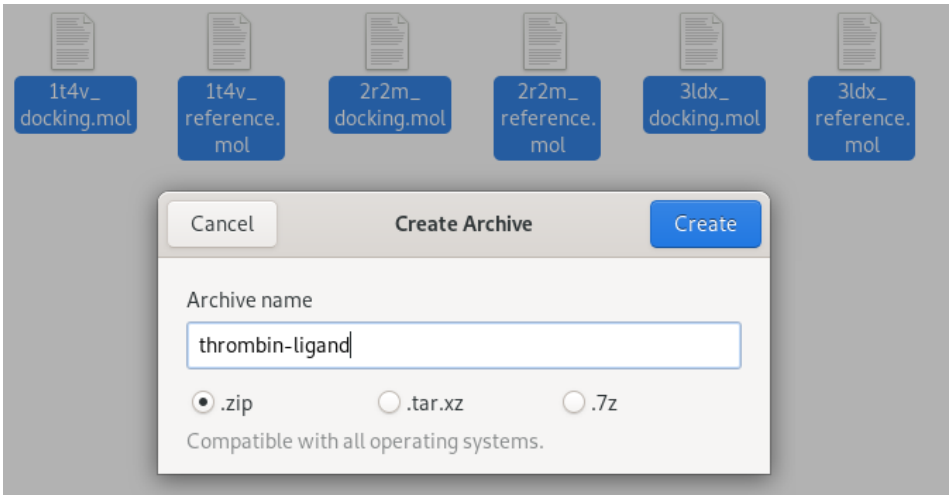


Figure 2: Creating the ZIP file

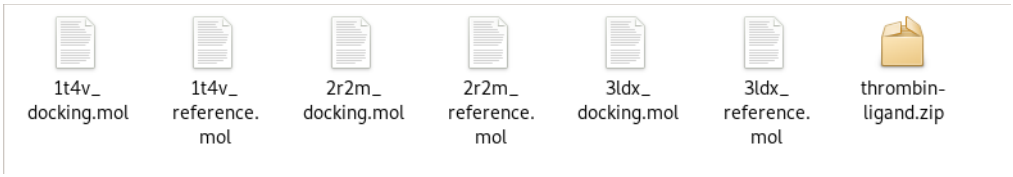


Figure 3: The ZIP file created

The next step is to upload the ZIP file to the server. Figure 4 shows the link for the submission of a new job.

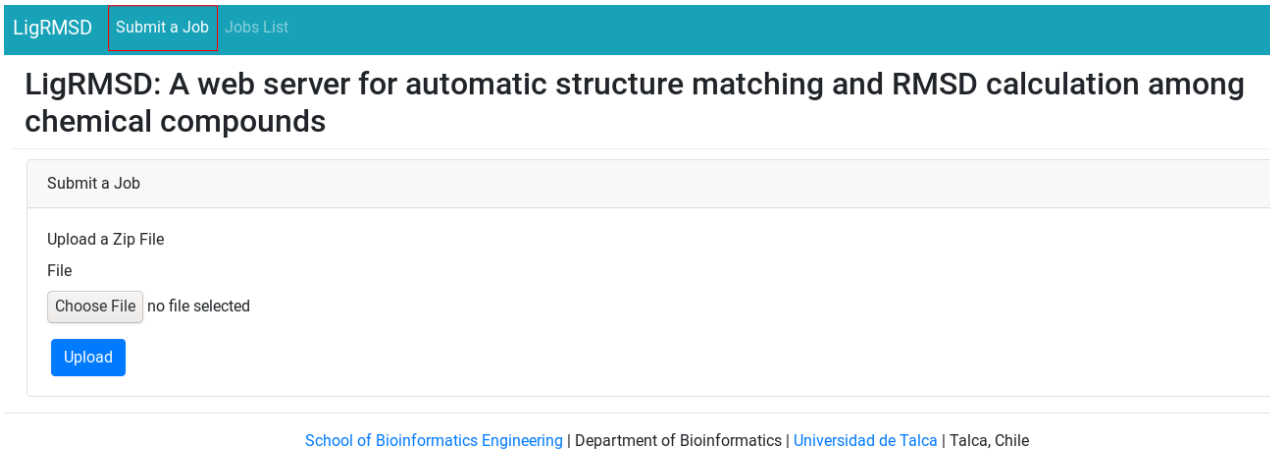


Figure 4: Submitting a new job

Now the user can press the button “Choose File” and select the previously created ZIP file (Figure 5 and Figure 6).

Figure 5: Selecting the ZIP file

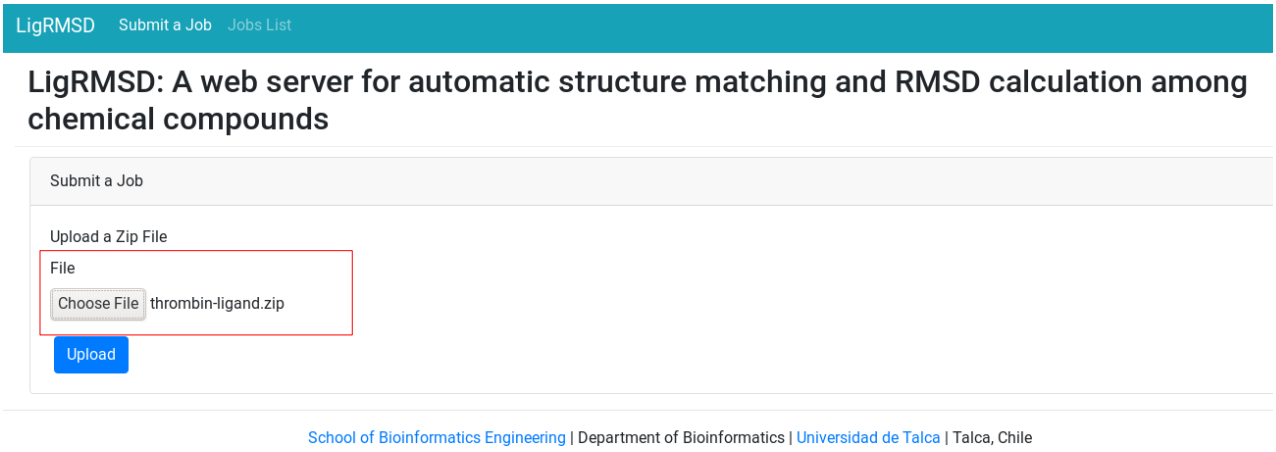


Figure 6: Uploading the ZIP file

Once the user presses the upload button, the ZIP file is submitted into a temporary directory and then is relocated into the server.

The next step to submit a job in LigRMSD is the selection of reference (s). For this, the web-server shows a 2D representation of each molecule included in the ZIP file submitted (Figure 7). Here, the user can select one or more references for the comparisons (Figure 8).

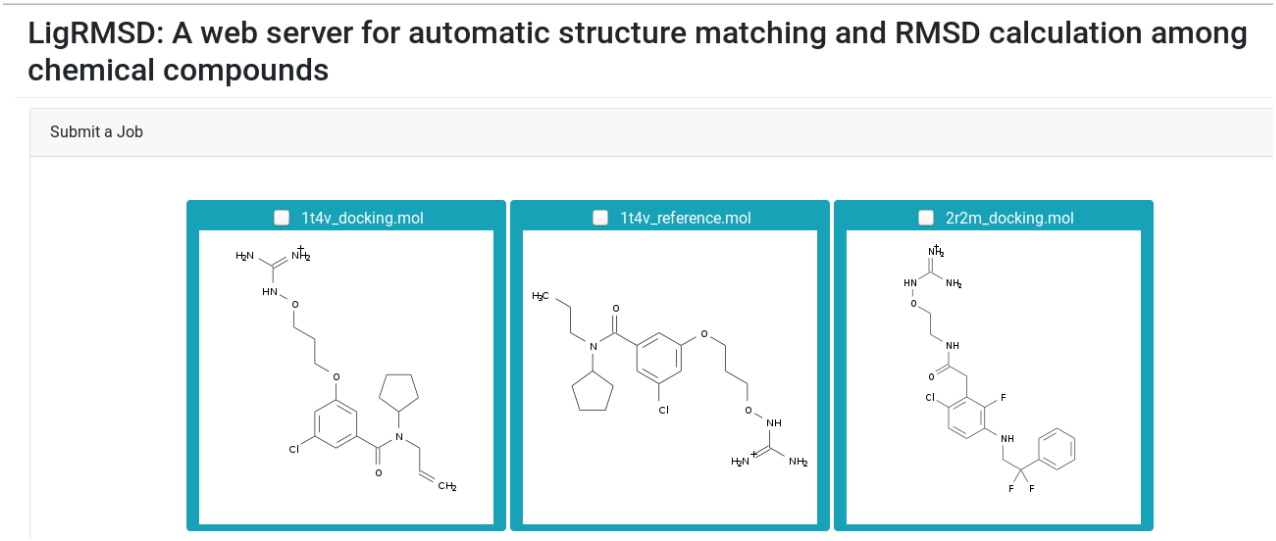


Figure 7: 2D representation of the MOL files submitted

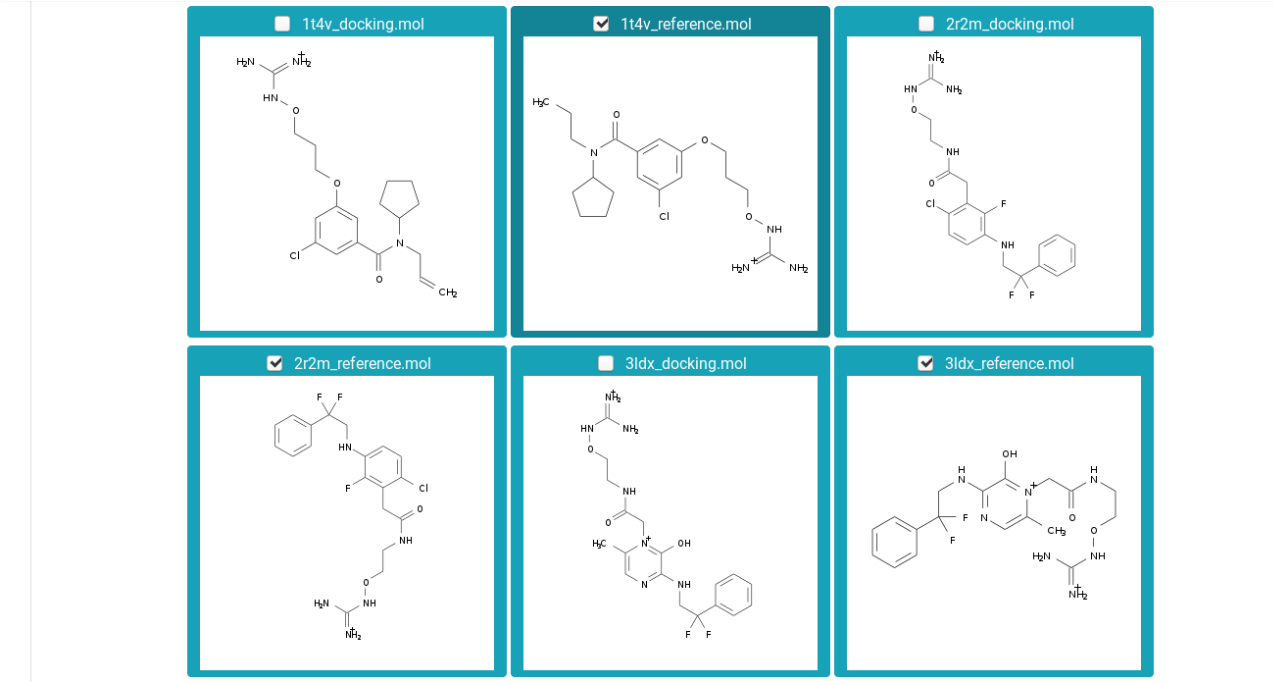


Figure 8: Selecting the references

Below the selection box of the references, the user can complete the form identification for the job (Figure 9). Even when these parameters are not mandatory, it is commendable to include representative job names. At the end of the measures, this information will be sent to the user e-mail. Also, the user must include a “Limit” value. This parameter represents the minimum structural match that must have a pair of molecules for calculation of the RMSD.

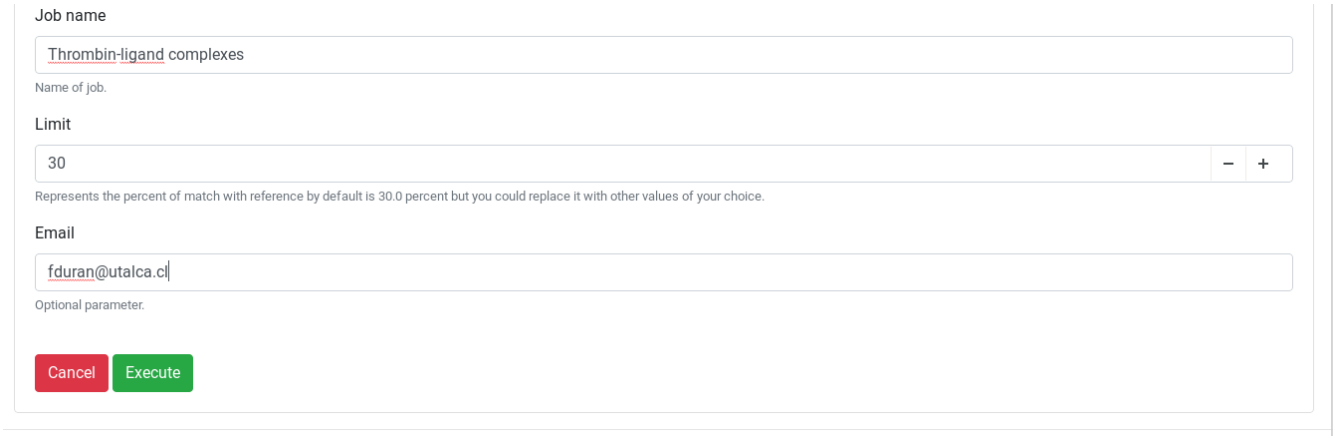


Figure 9: Job identification

Once completed the form and pressed the button "execute", LigRMSD redirects the browser to the Jobs list page (Figure 10). Thus, the user can visualize all jobs and their status (Waiting, Running, Finished and Error; Figure 10, Figure 11 and Figure 12).

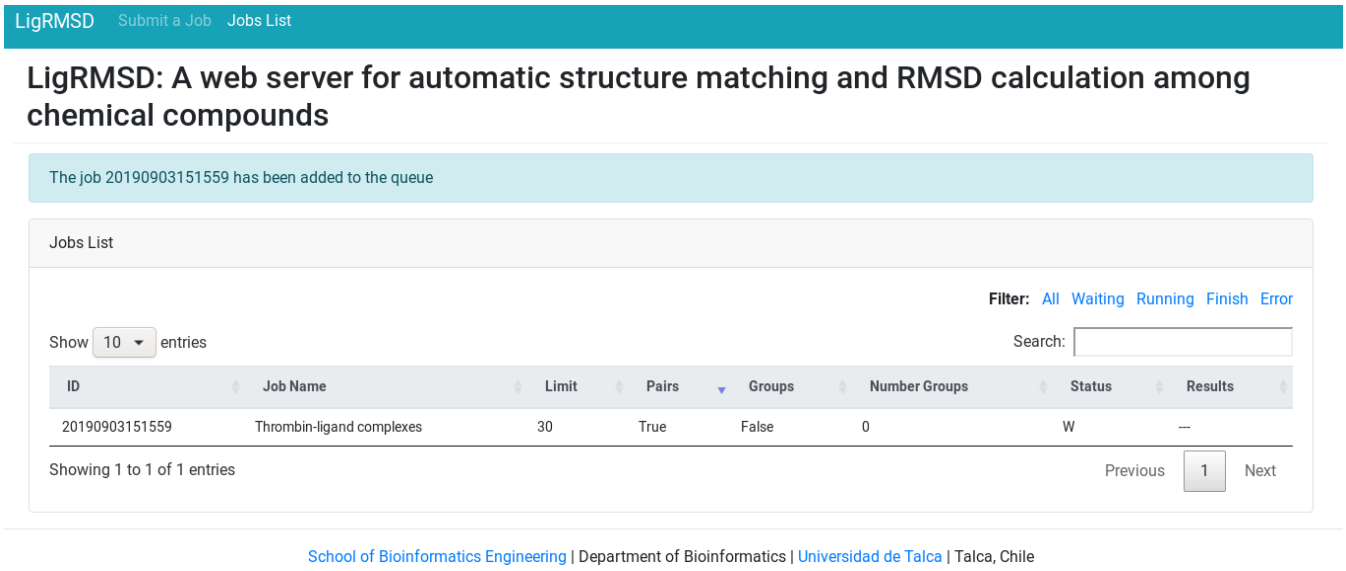


Figure 10: A new job is added to the Jobs list

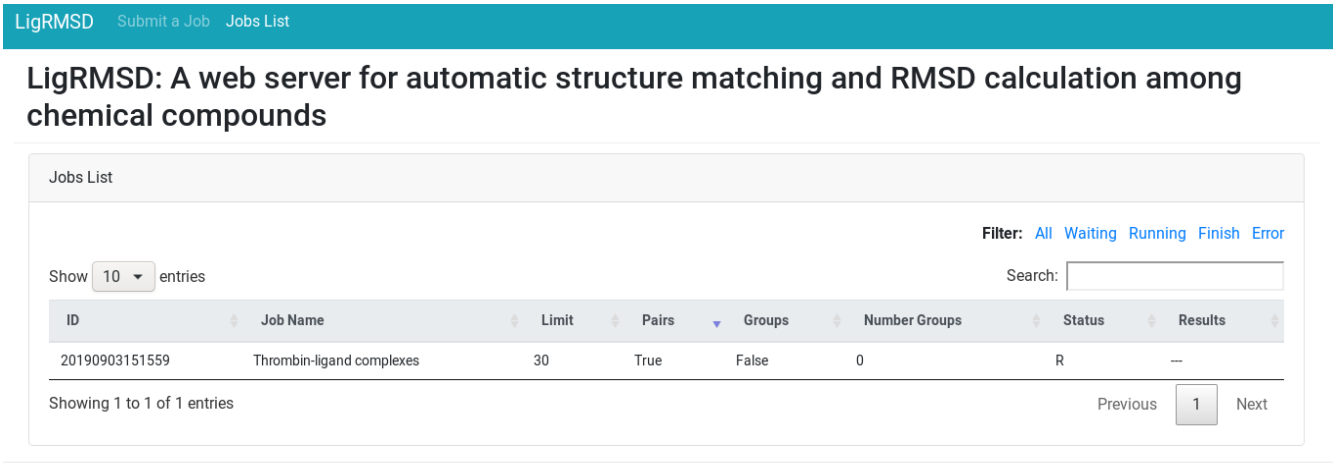


Figure 11: A Job in Running (R) status

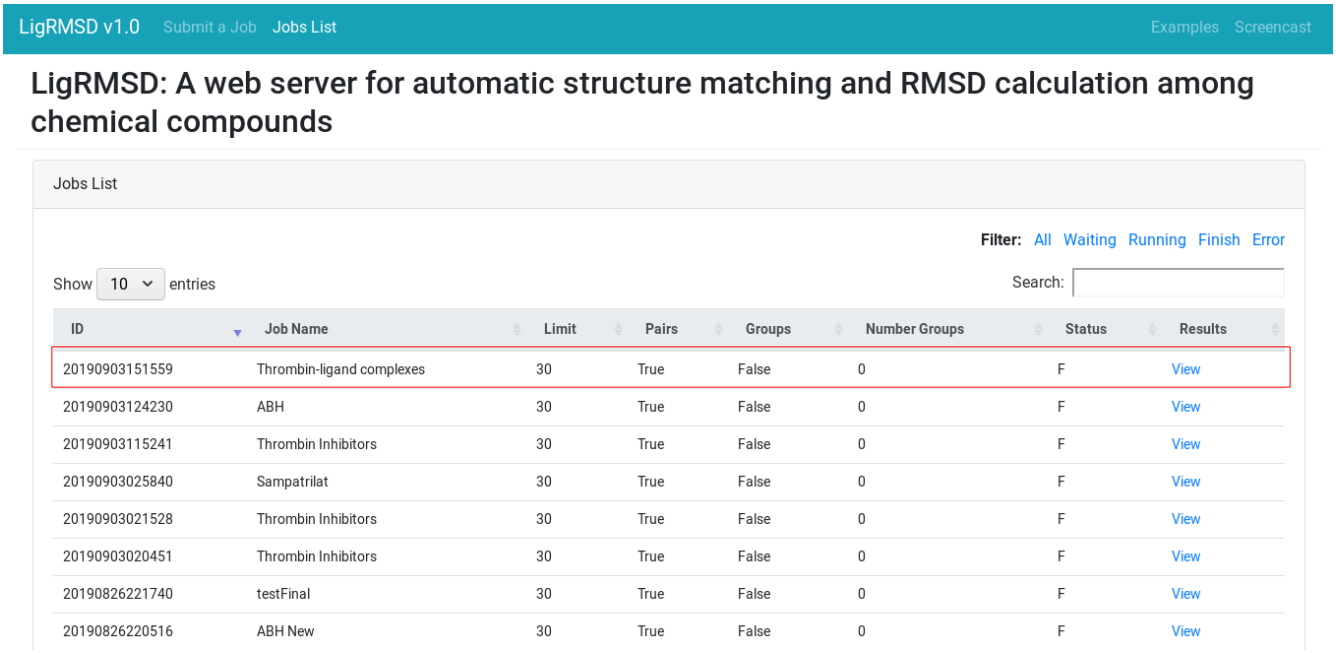


Figure 12: The Job denoted by the red box is in Finished (F) status

Visualizing the results

When the calculation has done (status F), the user can visualize the results by clicking on the "View" link (Figure 12). This link will forward the browser to another page with the results. The page of results has three sections. The first section (denoted by a red box in Figure 13) shows the reference molecules selected by the user. If an error is detected on any reference, this molecule would be mentioned in this section ("References with error”). For exploration of a particular reference result, the user must click on the name of the reference. The second section, named “Matching Method”, is denoted by a green box in Figure 13. Here the user can visualize all the pair of molecules evaluated using strict or flexible methods. The third section is the Database (right side of the green box in Figure 13). This section is used for comparing an extensive database of compounds against one or several query molecules (Figure 14). In this section, the user can view all the comparisons simultaneously.

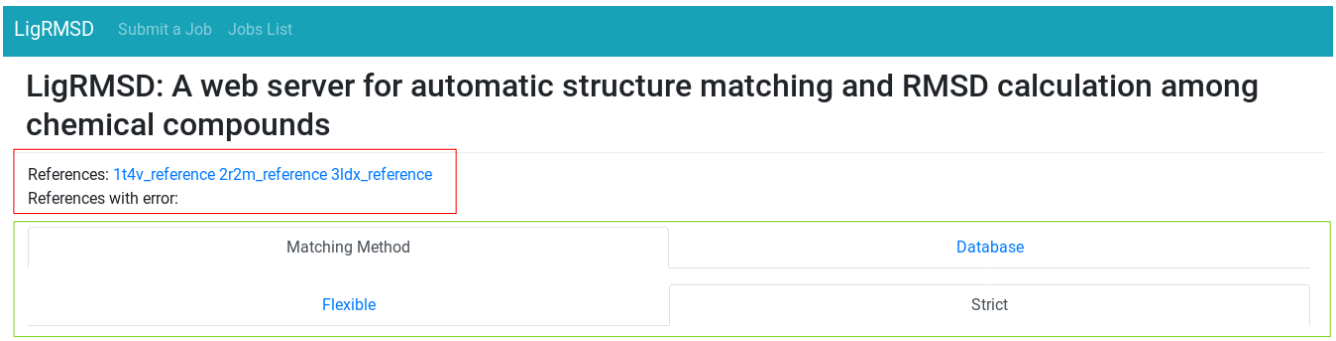


Figure 13: The three sections of results are denoted into the red and green boxes

Figure 14: Database results

If the user press on one of the references located in the first section of results (red box in Figure 12), then this reference will be shown as a 2D representation and in a 3D viewer (Figure 15).

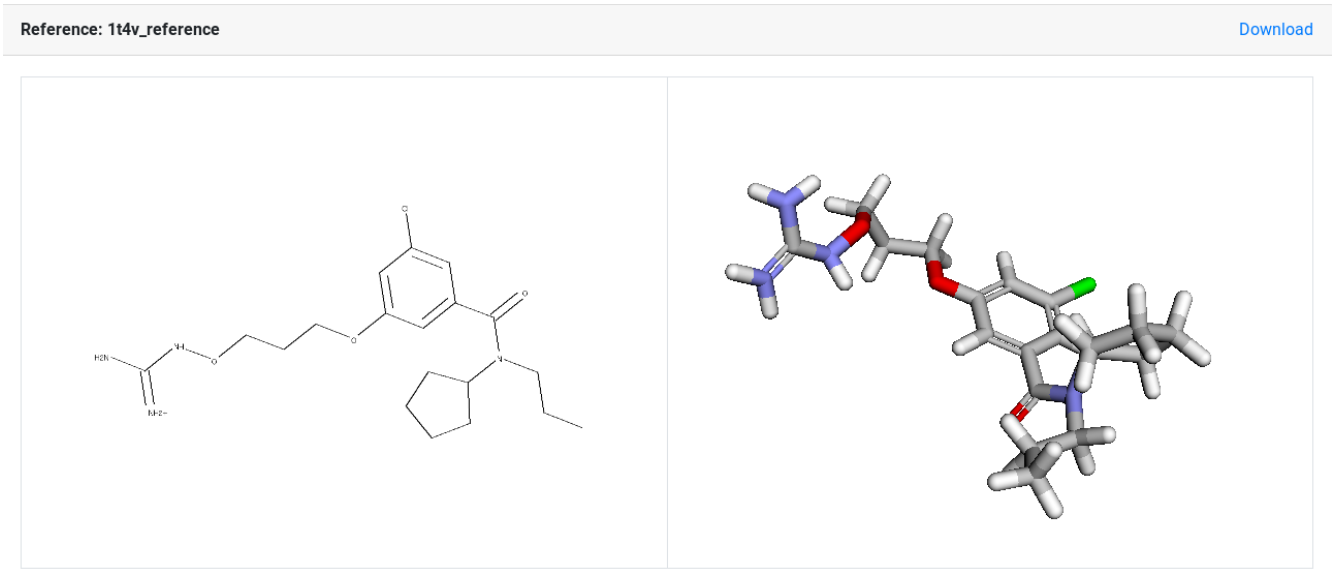


Figure 15: 2D representation and 3D viewer of a reference selected

Once selected a reference, the results can be filtered for both, the strict and the flexible method. As is shown in Figure 16 (flexible) and Figure 17 (strict), all the molecules are listed and represented as 2D structures. Those zones that were included in the maximum common substructure (MCS) between the molecule and the selected reference are remarked in red. The information about the percentage of the match and the RMSD values are delivered in the same list. 'Percent of Ref Match' refers to the percent of

common graphs between the docked and reference compounds concerning the total number of atoms of the reference compound. ‘Percent of Molecule Match’ refers to the percent of common graphs between the docked and reference compounds regarding the total number of atoms of the docked compound.

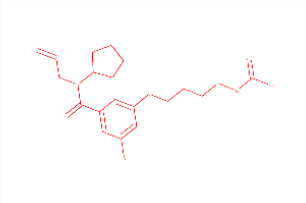
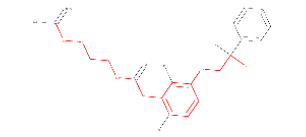
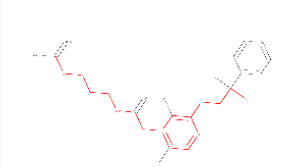
Molecule	Image	RMSD	Percent of ref match	Percent of molecule match	Details	Download
1t4v_docking		1.33	100.00	100.00	View	Download
2r2m_docking		1.56	66.67	60.00	View	Download
3ldx_docking		1.45	66.67	60.00	View	Download

Figure 16: The matching results for several molecules against reference (flexible method)

As we mentioned before, a “Limit” value of matching must be defined on each query. In these examples, we used an amount of 30%, denoting the minimum percent of structural matching allowed to realise the calculation of the RMSD. Because this value is related to the MCS, different results might be obtained between the flexible and the strict method. Thus and for comparisons purposes, LigRMSD will show the results for the two methods if at least in one of them, the “Limit” threshold is overcome. In this case, the molecules 2r2m_docking and 3ldx_docking, shows a 66,67 % of ‘Percent of Ref Match’ considering the flexible method (Figure 16). This value decreases to 25,93 % when the strict method is selected (Figure 17). It should be noted that low values of ‘Percent of Ref Match’ might bring unsatisfactory results. For example, as is shown in Figure 17, unlike the MCS detected between the reference and the 2r2m_docking, the MCS between the reference and the 3ldx_doking, comprise non-analogous groups yielding high values of RMSD (5,28).

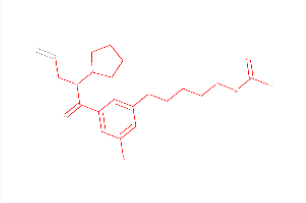
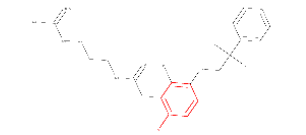
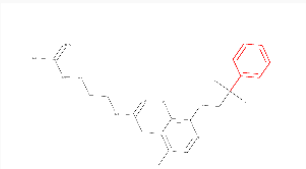
Molecule	Image	RMSD	Percent of ref match	Percent of molecule match	Details	Download
1t4v_docking		0.90	96.30	96.30	View	Download
2r2m_docking		1.12	25.93	23.33	View	Download
3ldx_docking		5.28	25.93	23.33	View	Download

Figure 17: The matching results for several molecules against reference (strict method)

Finally, the user can view a detailed comparison between a pair of compounds (Figure 18) or download the structures analyzed.

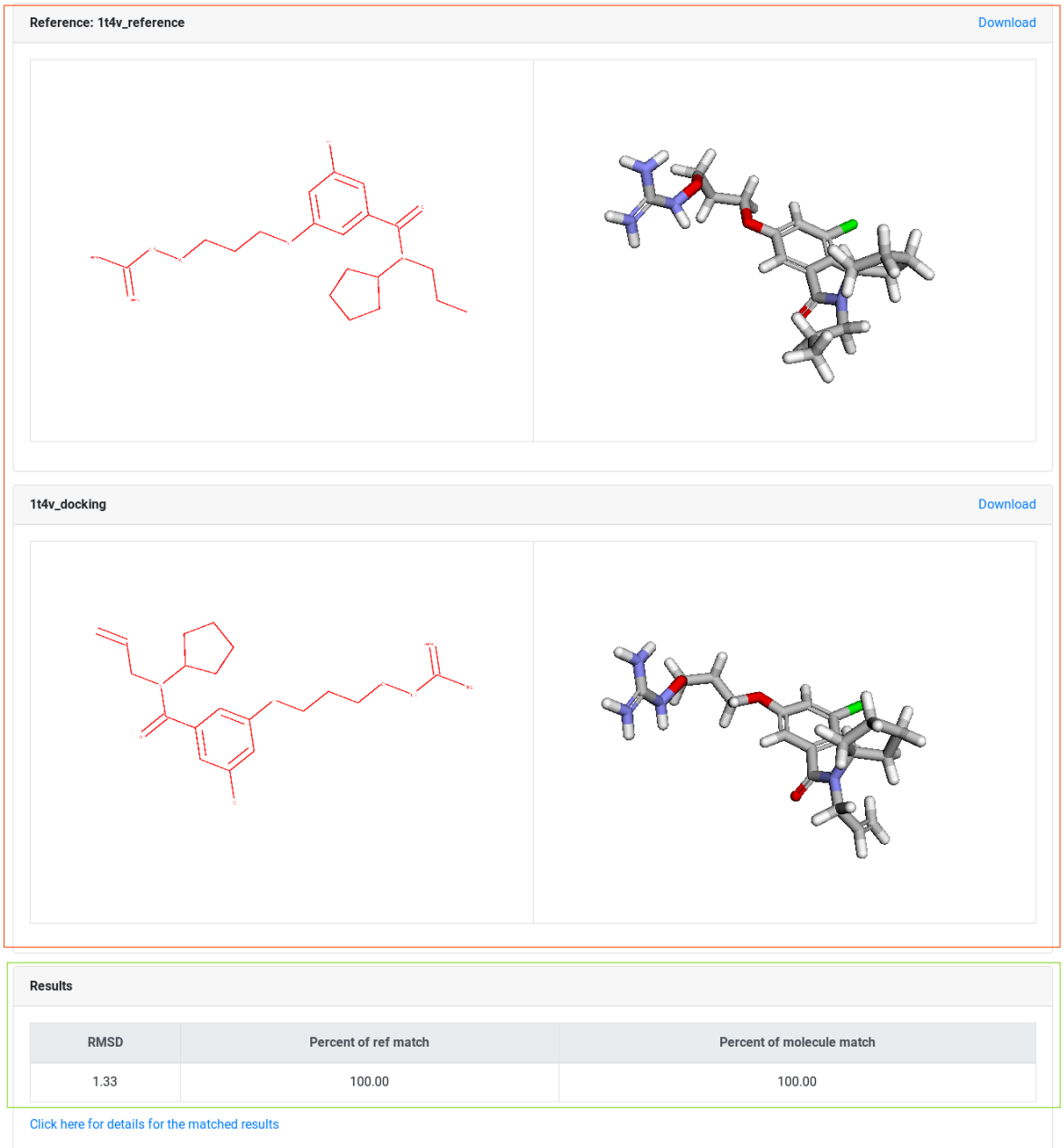


Figure 18: Detailed evaluation between two compounds. Structures of the compounds are presented in the red box and RMSD value and the percentage of ref and molecule match are available in the green box.