FORECASTER Suite 2018 Tutorial

FITTED, MATCHUP, PREPARE, PROCESS and SMART

Department of Chemistry,
McGill University
Montréal, Québec, Canada

Molecular Forecaster Inc.
Montréal, Québec, Canada

20/02/2018

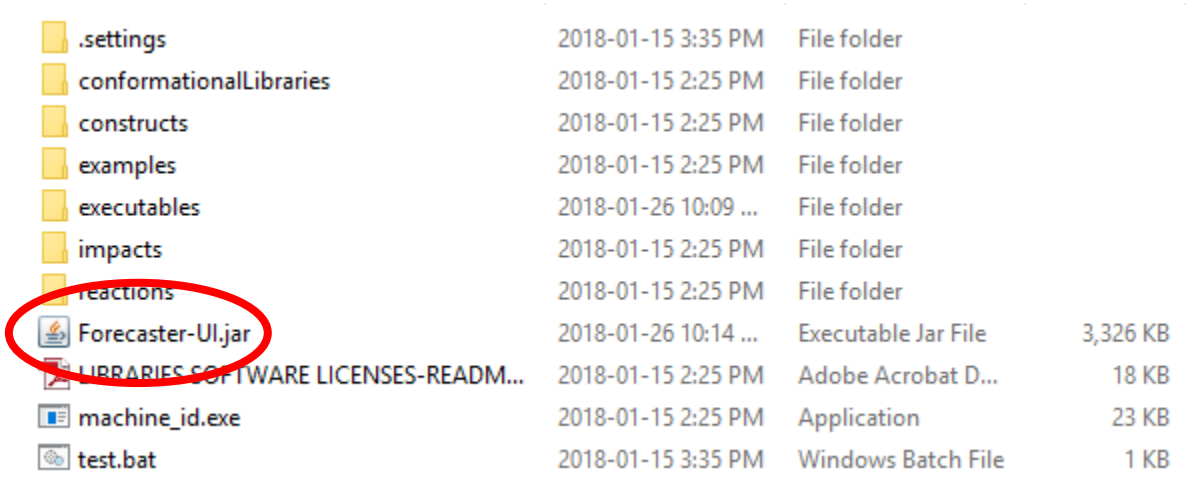
Table of contents

I. Running FORECASTER with the User Interface

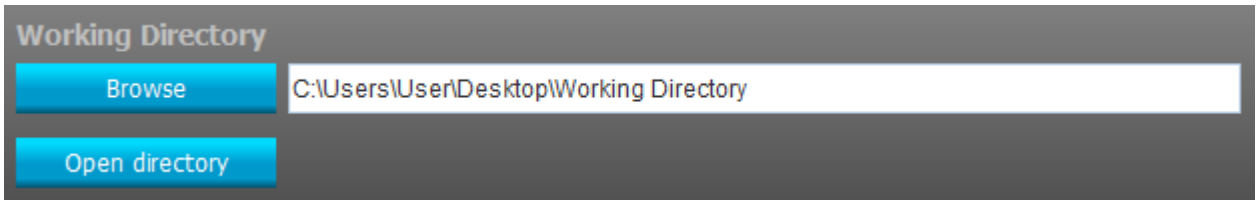
The user interface (UI) can be started by double clicking on the `Forecaster-UI.jar` file (`FrontEnd.jar` in earlier versions) in the `Forecaster` folder. This will open the main window.

Under Linux and Mac OSX, it is recommended to launch it from a terminal window by typing the command below. Make sure that you are located in the folder where this jar file is.

```
Forecaster@Linux/Forecaster:~$ java -jar Forecaster-UI.jar
```



The first step is to set the working directory. Click **Browse**, under **Working Directory**. You will be prompted to navigate to the folder where you will save various files while working with FORECASTER.



II. Rigid protein docking

Getting started.

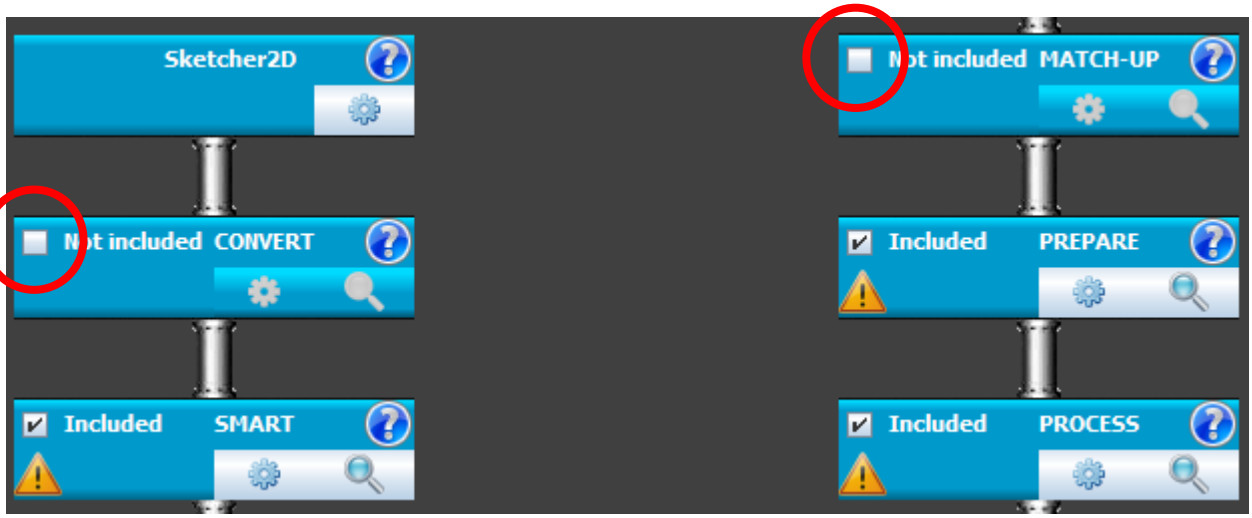
This example will use the thymidine kinase `1e2k.pdb` structure and perform a self-docking using either rigid protein docking or flexible protein docking modes in FITTED.

Click on **Start Forecasting** to expand the workflow. (If you did not already have the `1e2k.pdb` file, type the name of the file (without the extension) into the text box below download pdb file and press the button.).

The left branch will be used to prepare the ligand(s) of this workflow while the right branch will be used to prepare the protein structures and binding site files.

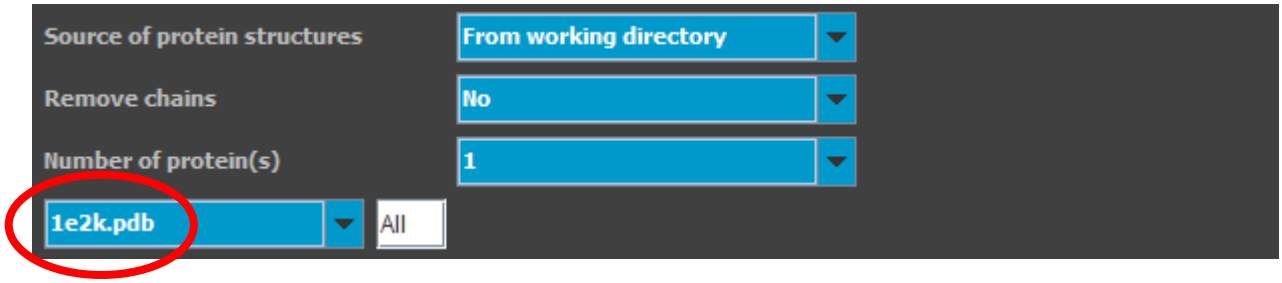
1. PREPARE: The preparation of the protein pdb file

Since we will be docking a small molecule to a rigid protein structure, only one protein structure file (e.g., from the protein databank, PDB) is required and the MATCH-UP step (used to superpose multiple protein structures) should be turned off by unchecking the box. We will use the ligand from the same pdb file, hence the CONVERT step (used to convert a 2D sketch into a 3D structure) should also be unchecked. It is important to set the parameters in the correct order (going down the workflow tree).



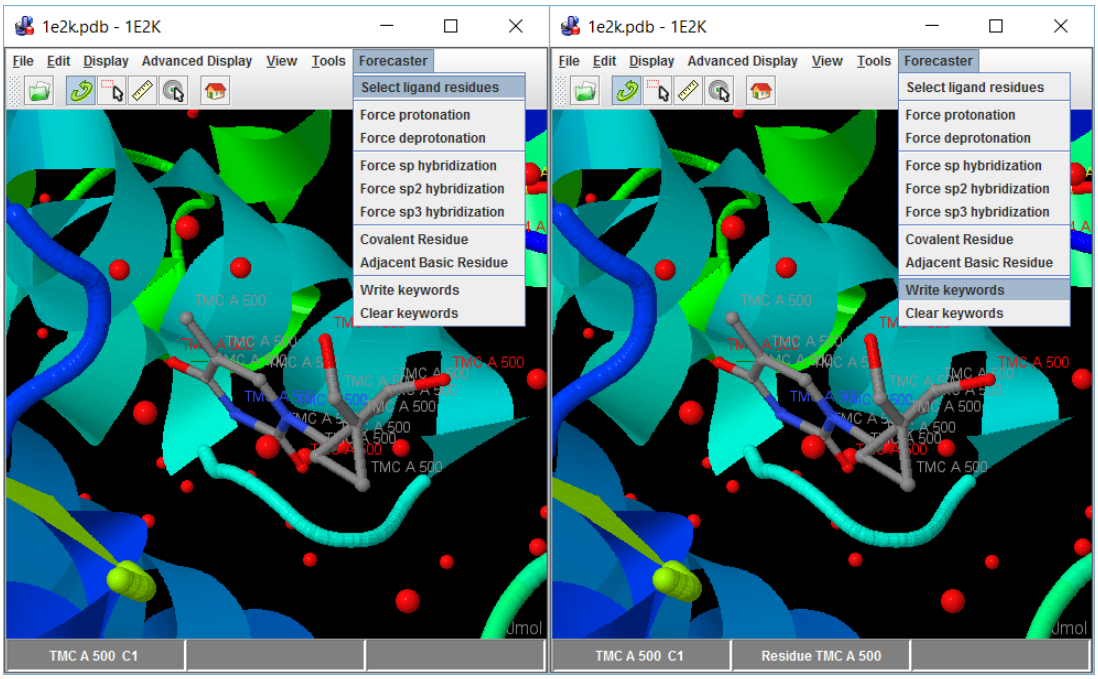
The first step in a rigid protein docking procedure is to take the protein/ligand complex from the pdb file and prepare the corresponding files in mol2 format by using PREPARE. This program adds hydrogens, generates the possible tautomers and optimizes the H-bond network by an iterative algorithm. Clicking on the gear icon will open the parameters section for PREPARE.

Since the `1e2k.pdb` file is already in the working directory, there is no need to download the file from the PDB site and it should be already listed in the protein menu.



The ligand residue(s) must be identified in order to define the active site and extract the ligands from the protein/ligand complexes. To automatically identify the ligand residues, click the **Load pdb structure** button and a 3D viewer will open with the protein loaded.

Within this 3D viewer, the ligand is selected by clicking one atom. The lower left box will then show the name of the selected ligand atom (TMC A 500 C1 in our case). If the correct ligand residue is selected, clicking **Select ligand residues** in the **Forecaster** menu will save the residue in the lower middle box (Residue TMC A 500).When the ligand is made of more than one residue, all of the residues must be listed. With some PDB files with poorly defined geometries, you may have to assign hybridization or protonation states to some atoms using a similar approach (e.g., **Force sp2 hybridization** in the **Forecaster** menu. Clicking on **Write keywords** will automatically save the information back to the PREPARE menu. The 3D viewer can then be closed to return to the parameters section.

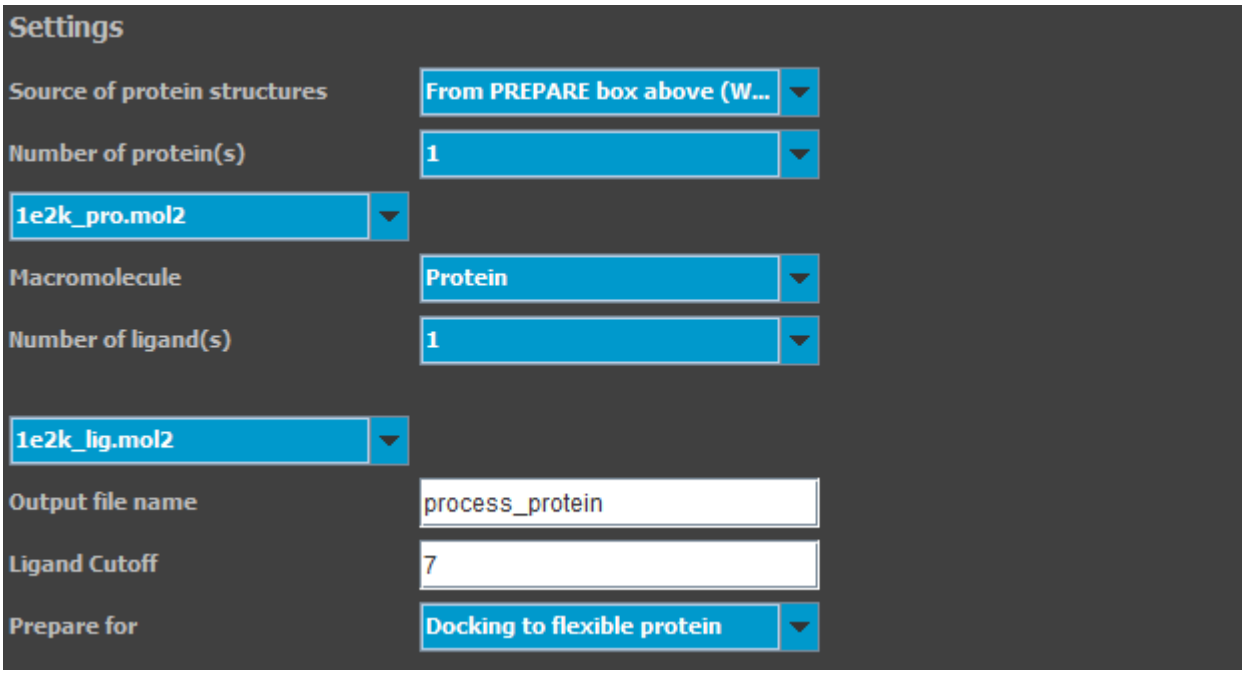


Once all the parameters are set, the keyword file needs to be written by clicking the **Write keyword file** button. Clicking the **Exit** button will close the PREPARE parameters section and return to the main workflow.

A green check should now appear in the PREPARE box indicating that this box is now configured.

2. PROCESS: The setup of the protein for docking

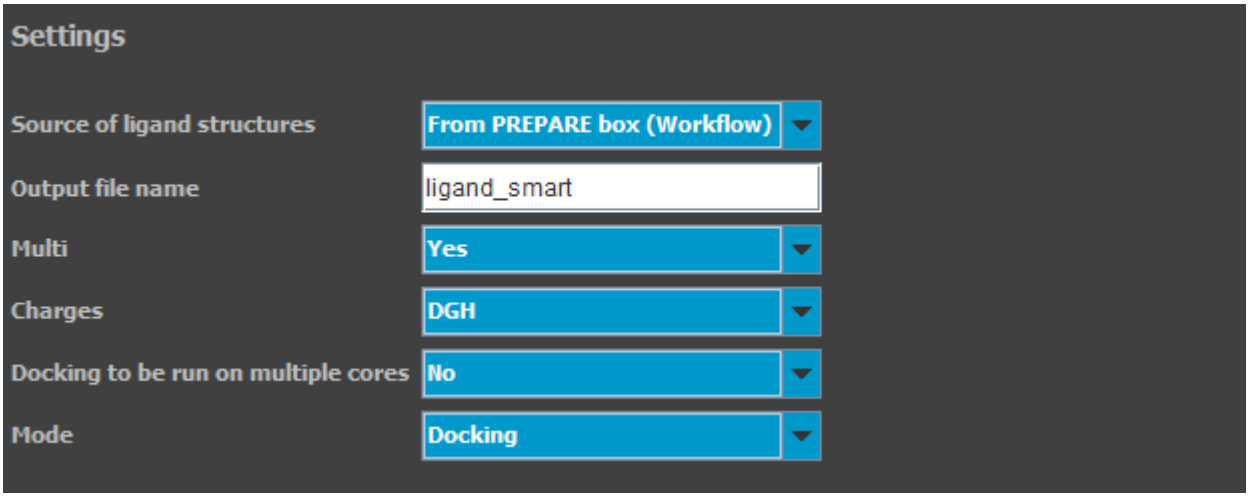
PROCESS is used to setup the protein for docking with FITTED. Since we are preparing a workflow, the files generated from the previous programs should be made available for the next step even if they are not physically present in the working directory. Thus, the **Source of protein structures** should be set to "From PREPARE box above". The **Number of protein(s)** should be set to "1" for rigid protein docking. The active site is determined by using the co-crystallized ligand. In this example, **Number of ligand(s)** is set to "1" and the ligand file set to `1e2k_lig.mol2`.



Once all the parameters are set, the keyword file needs to be written by clicking the **Write keyword file** button. Clicking the **Exit** button will close the PROCESS parameters section and return to the main workflow. A green check should now appear in the PROCESS box.

3. SMART: The setup of the ligand for docking

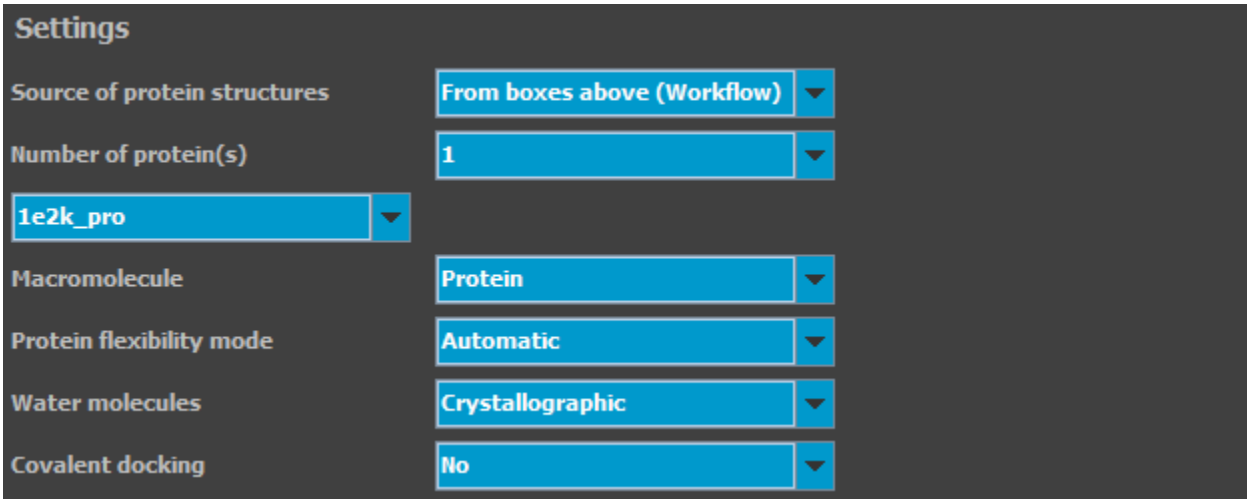
SMART will setup the ligand (3D molecules only) for docking with FITTED. For the same reason as above, the **Source of ligand structures** should be set to “From PREPARE box (Workflow)”. The **Output file name** can be filled with any desired name,“ligand_smart” is suggested. The atomic partial charges method can be selected under **Charges**. The methods available are DGH (recommended), none (no charge), and Input (will keep input file charges if any).



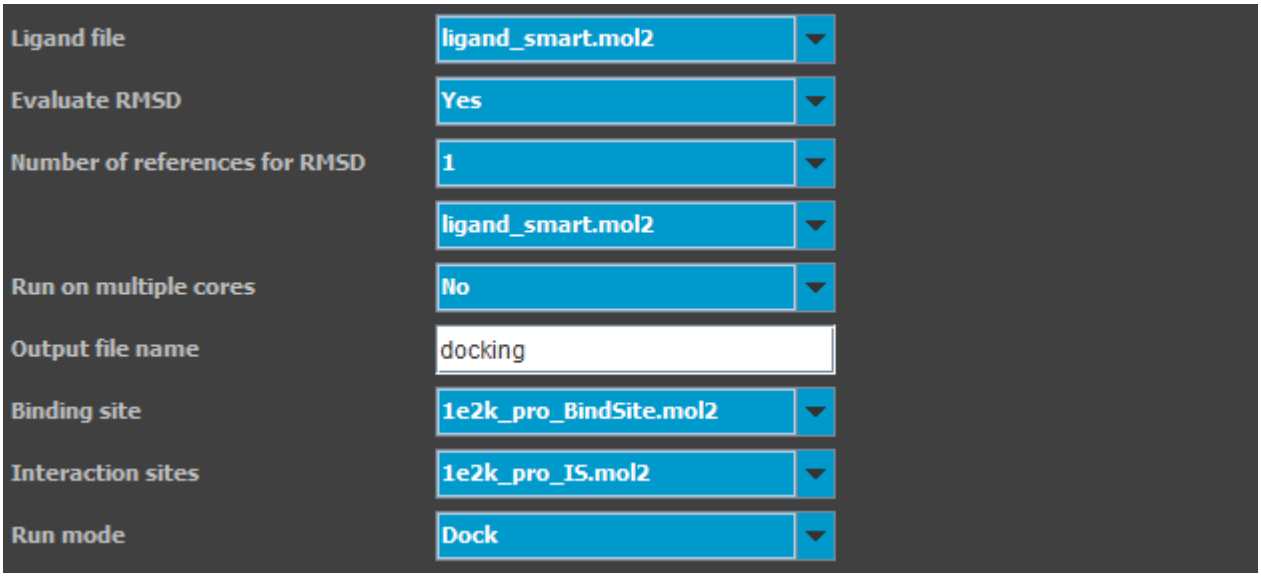
Once the settings are complete, the keyword file needs to be written by clicking the **Write keyword file** button. Clicking the **Exit** button will close the SMART parameters section and return to the main workflow. A green check should now appear in the SMART box.

4. FITTED: Rigid protein docking

The **Source of structures** should be set to “From boxes above (Workflow)”. The **Number of protein(s)** should be set to “1” for rigid protein docking. **Macromolecule** is used to differentiate between protein, metalloprotein, DNA and RNA.



The **Ligand file** is the ligand that will be prepared by SMART (`ligand_smart.mol2`). Since we are performing a self-docking, the RMSD value may be calculated by setting the option (**Evaluate RMSD**) to "Yes" and providing the reference ligand file `ligand_smart.mol2`. When multiple ligands are docked, the docking can be distributed over multiple cores. In this case, since there is only one ligand, the keyword **Run on multiple cores** should be set to "No".

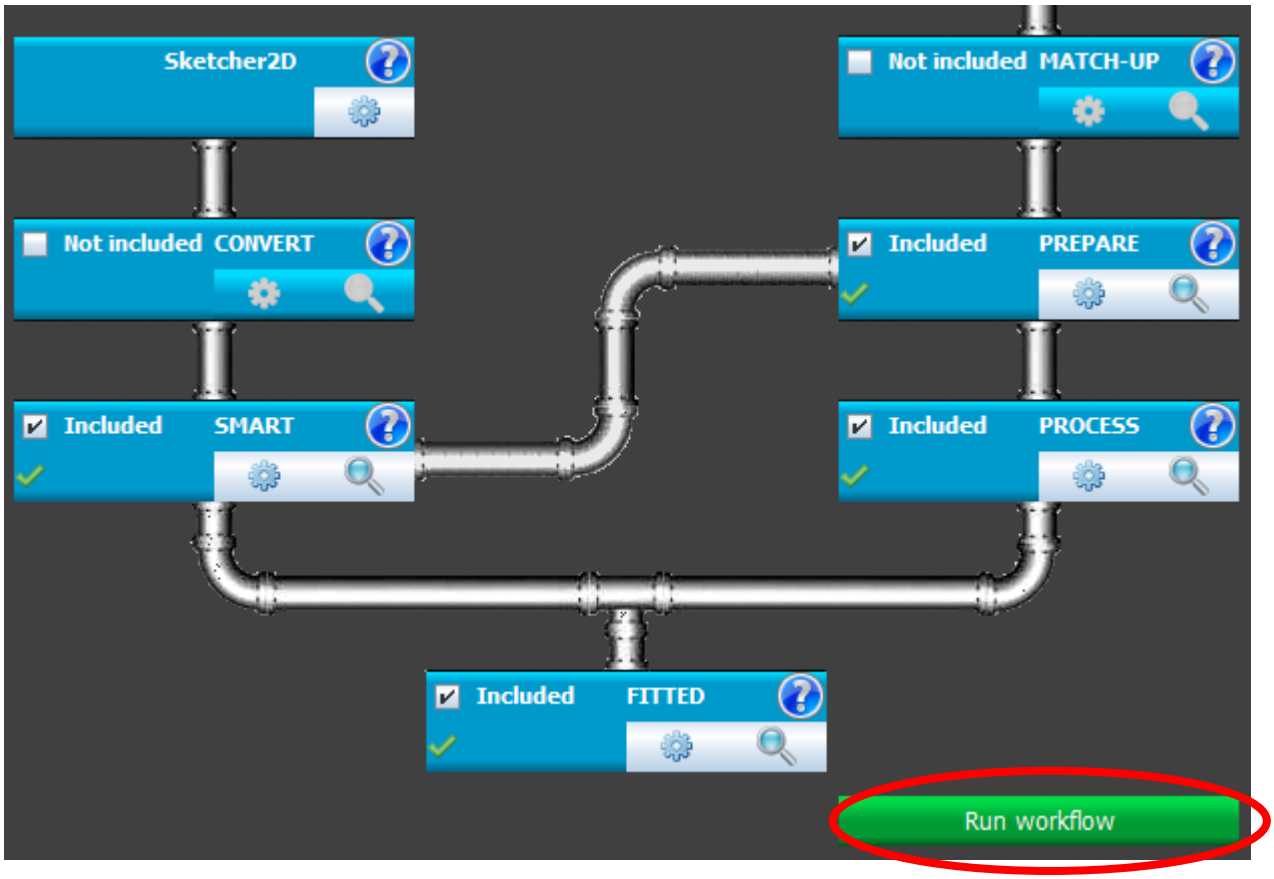


The keyword may then be written by clicking the **Write keyword file** button. Clicking the **Exit** button will close the FITTED parameters section and return to the main workflow. A green check should now appear in the FITTED box. The program will not run until the execution of the workflow is launched.

5. Workflow: Running the docking workflow

Once all the included steps are ready, the workflow can be executed by clicking the **Run workflow** button. The programs run in a terminal and once the complete workflow is complete, the terminal

window will close. Do not close the terminal window manually unless you want to stop the execution of the workflow (cannot be resumed).



Once the docking is complete, a new `output` folder will be automatically created and the docking results files will be placed in this folder.

6. Result Analysis

PREPARE takes a protein-ligand complex in pdb format and will create the `1e2k_pro.mol2` and `1e2k_lig.mol2` files, while the `prepare_protein.out` file contains the output information. Clicking the magnifier icon in the PREPARE box will allow the visualization of these files.

Several new files will be generated by PROCESS: `process.out` (output file), `1e2k_BindSite.mol2` (binding site definition file), `1e2k_pro_dock.mol2` (truncated protein), `1e2k_pro_score.mol2` (protein score file), `1e2k_pro_IS.mol2` (interaction sites file), and

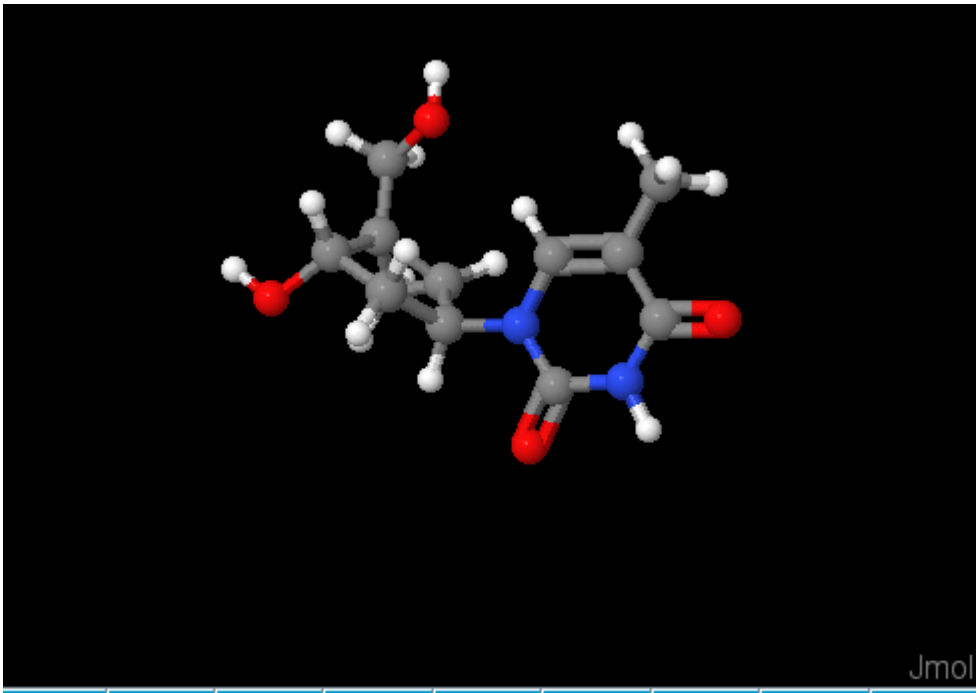
`1e2k_pro_site.txt` (flexible residues in binding site). Some files are automatically named according to the input protein mol2 file and others are named based on the filename provided in the **PROCESS**keyword file. Once again, the magnifier icon will allow you to visualize the files.

The files generated by SMART will be: `ligand_smart.mol2` and an output file named `ligand_smart.out`. Here as well, the files may be visualized using by clicking the magnifier icon.

FITTED will generate (within the `output` folder): `docking.log` (log file for potential errors), `docking.out` (output file), `docking_Docked_Poses.mol2` (the poses in mol2 format), `docking.sdf` (sdf file with the poses and scores/energy), and `docking-results.txt` (summary of the results in text format). If you open the `docking-results.txt` file, you can visualize the score, energy and rmsd value for each run.

The `docking.sdf` file contains the best pose of the docked ligand and it can be visualized together with the `1e2k_pro.mol2` file in your favorite 3D graphical program.

Alternatively, the poses (i.e., proposed binding modes) can be visualized by clicking the magnifier icon in the FITTED box. Clicking the **Open poses** button and selecting the file `docking.sdf` will open the results in a 3D viewer. Within this 3D viewer, a table of data with the corresponding scores will be available to easily visualize and analyze the docking results.



Entry	Name	Display	FITTED...	FITTED...	FITTED...	FITTED...	FITTED...	none
1	Protein	☐	n/a	n/a	n/a	n/a	n/a	n/a
2	Interact...	☐	n/a	n/a	n/a	n/a	n/a	n/a
3	docking	☑	3	-25.8540	-97.2494	0.47	41.92	n/a

In addition, a result file in text format is generated and contains all the scores and values for the docking poses. This file can be opened by clicking the **Read numerical results** button and selecting the file `docking-results.txt`.

III. Flexible protein docking

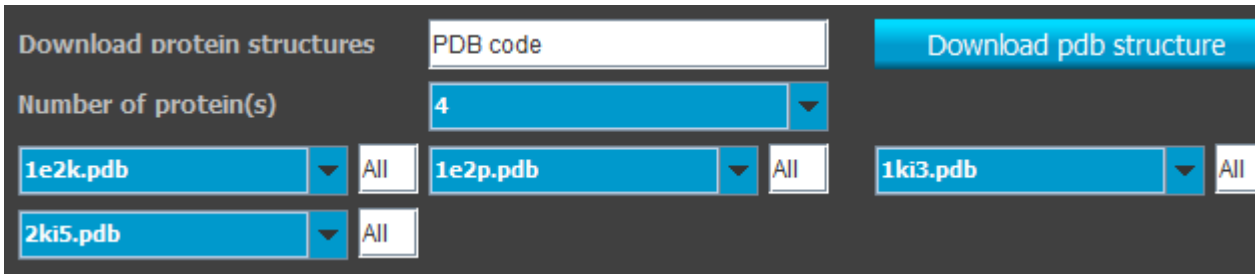
FITTED can dock ligands to a protein simulating the protein flexibility. This is achieved by providing a conformational ensemble of structures (more than one crystal structure for instance) which will be used by FITTED.

In this example, we will perform a docking of a putative ligand within a conformational ensemble of four protein files using the flexible protein docking mode. The first step in a flexible protein docking procedure is to take the protein/ligand complex from the pdb file and perform a sequence alignment and mutations to make all the pdb structures similar. In this workflow, all the steps will be included.

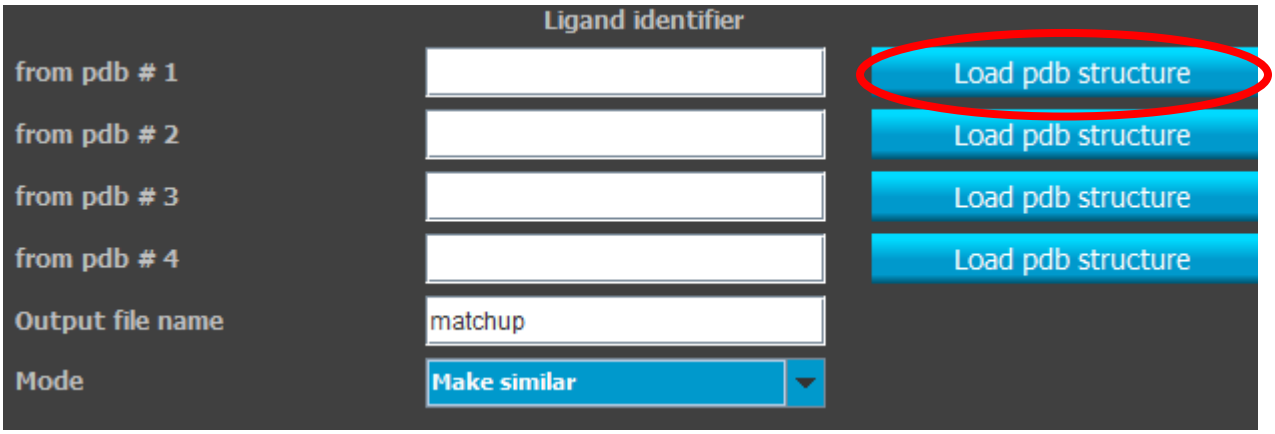
You will need to download 1e2k.pdb, 1e2p.pdb, 1ki3.pdb and 2ki5.pdb.

1. MATCH-UP: Superposition of the protein pdb files

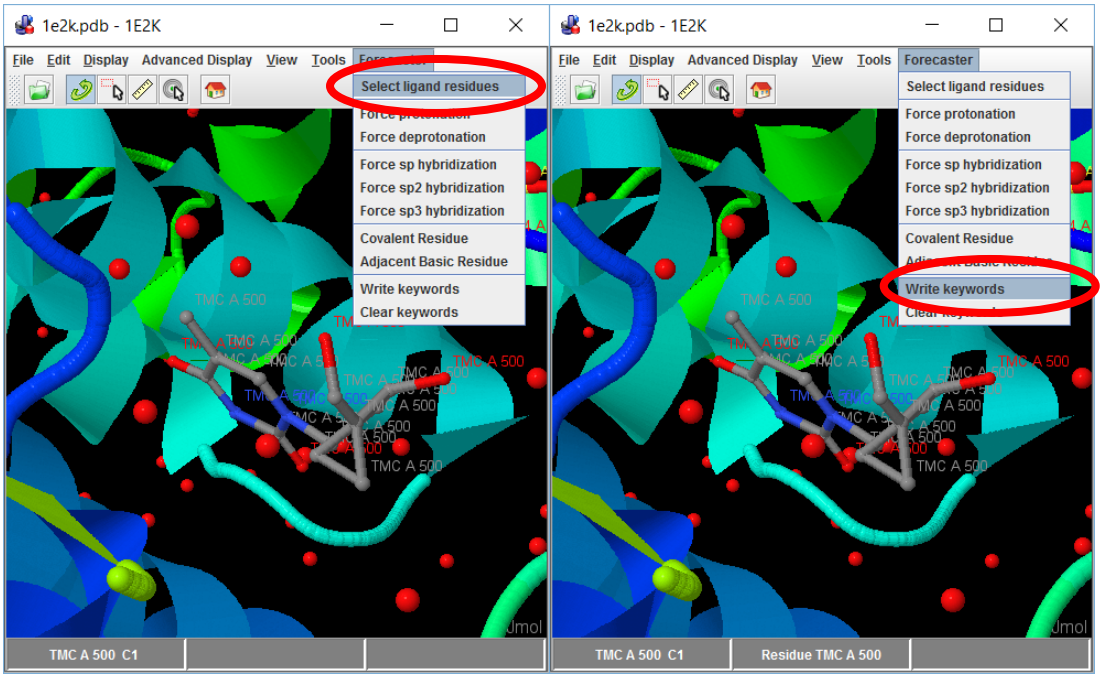
MATCH-UP takes two to fifteen pdb files and makes them similar (or only superposes them). Clicking on the gear icon will open the parameters section for MATCH-UP. After selecting the proper number of protein structures (4 in our example) to be used, the pdb files may be selected from the pull-down menus. The desired chain(s) (A, B, AB, AC, All, etc..) may be defined in the box next to the pdb file.

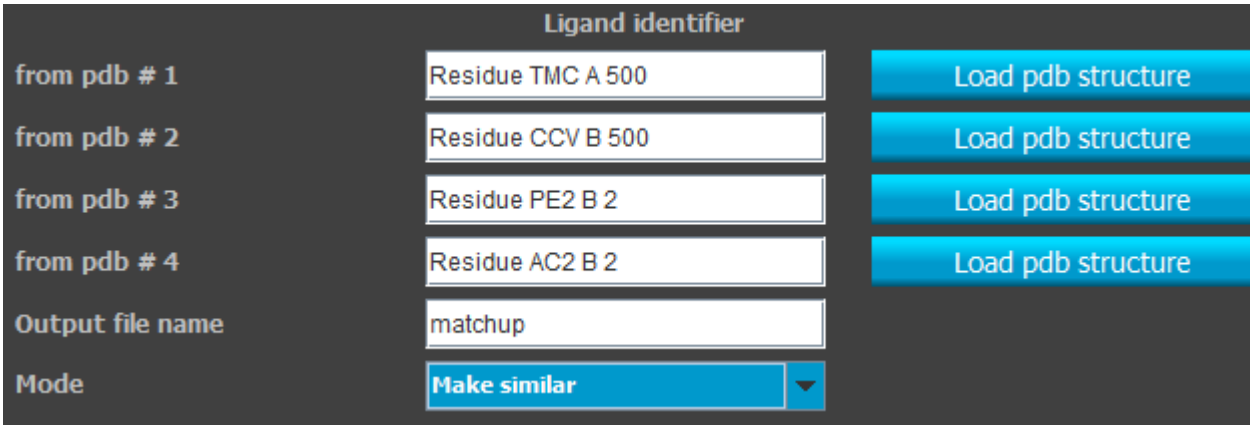


The ligand residues must be identified in order to define the active site and remove it from the binding site. The superposition of the proteins is performed with a focus around the active site. To automatically identify the ligand residues, click the **Load pdb structure** button and a 3D viewer will open with the protein loaded.



Within this 3D viewer, the ligand is selected by clicking one atom. The lower left box will then show the name of the selected ligand atom (TMC A 500 C1). If the correct ligand residue is selected, clicking **Select ligand residues** in the **Forecaster** menu will save the residue in the lower middle box (TMC A 500). When the ligand is made of more than one residue, all of the residues must be listed. Click on **Write keywords** to save the keyword. The 3D viewer can then be closed to return to the parameters section. This procedure needs to be done for every pdb file.





Once all the parameters are set, write the keyword file and exit MATCH-UP.

2. PREPARE: The preparation of the protein pdb files

The second step in a flexible protein docking procedure is to take the protein/ligand complexes from the aligned/mutated pdb files and prepare the corresponding files in mol2 format by using PREPARE.

This program adds hydrogens, generates the possible tautomers and optimizes the H-bond network by an iterative algorithm.

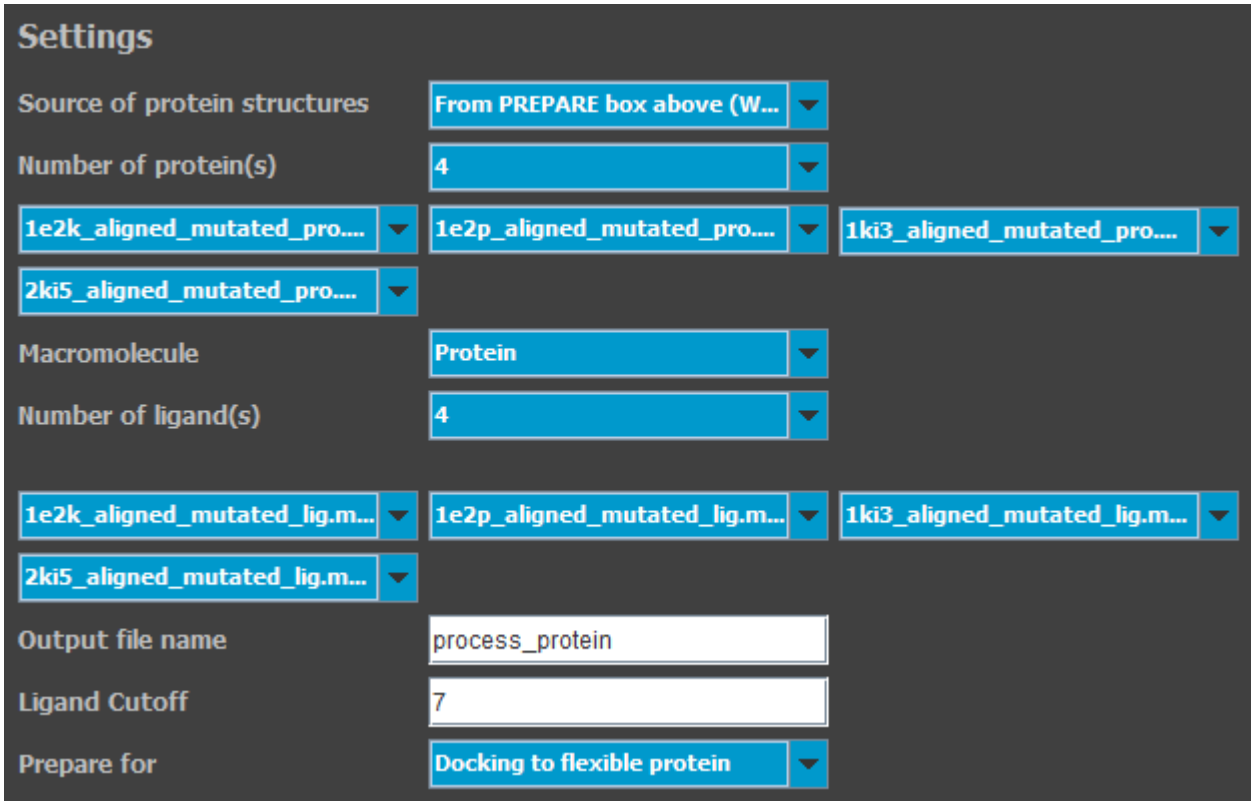
The **Source of protein** structures should be set to “From MATCH-UP box above”. The **Number of protein structure(s)** should already be set to “4” and all the protein filenames should already be selected. Since the ligand residues were already identified in the MATCH-UP step, they will be automatically updated. Other settings should keep their default values.

Once all the parameters are set, write the keyword file and exit PREPARE.

3. PROCESS: The setup of the protein for docking

PROCESS is used to setup the protein for docking with FITTED. The **Source of protein structures** should be set to “From PREPARE box above” which will automatically fills up most of the pull-down menus. The **Number of protein(s)** should be set to “4”. The active sites are determined by using

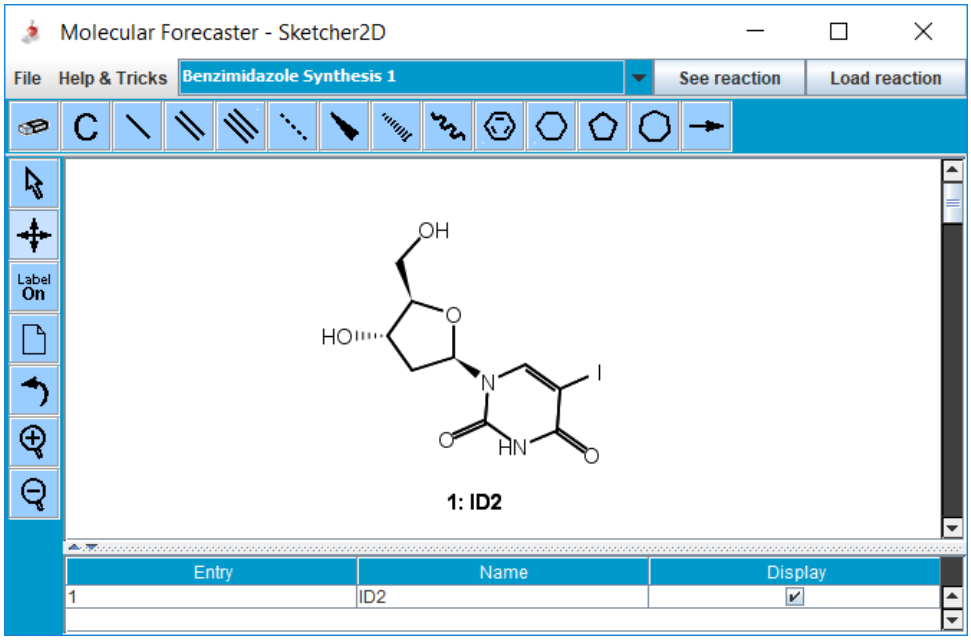
the co-crystallized ligands. In this example, **Number of ligand(s)** is set to "4" and the ligand files to `xxxx_aligned_mutated_lig.mol2`.



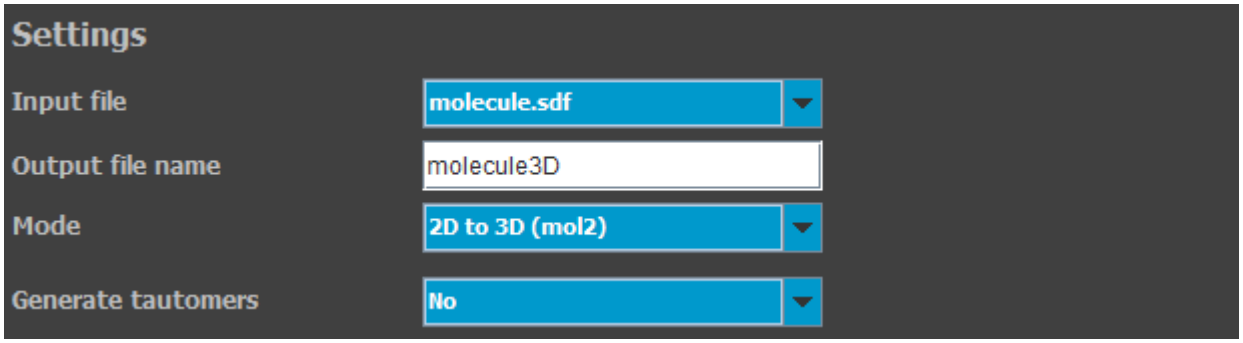
Once the parameters are configured, write the keyword and exit PROCESS.

4. CONVERT: The conversion of the ligand from 2D to 3D

CONVERT will convert a 2D molecule into an energy-minimized 3D structure. The 2D molecule can either be imported as a file (supported input formats are mol and sdf) or drawn directly using the 2D sketcher. In this example, we will draw the molecule with the sketcher. Click on the gear icon in the Sketcher2D box to open the sketcher window. Once the molecule is drawn, the file is saved by clicking on **File** and **Save structure (sdf)**. The sketcher window can then be closed.



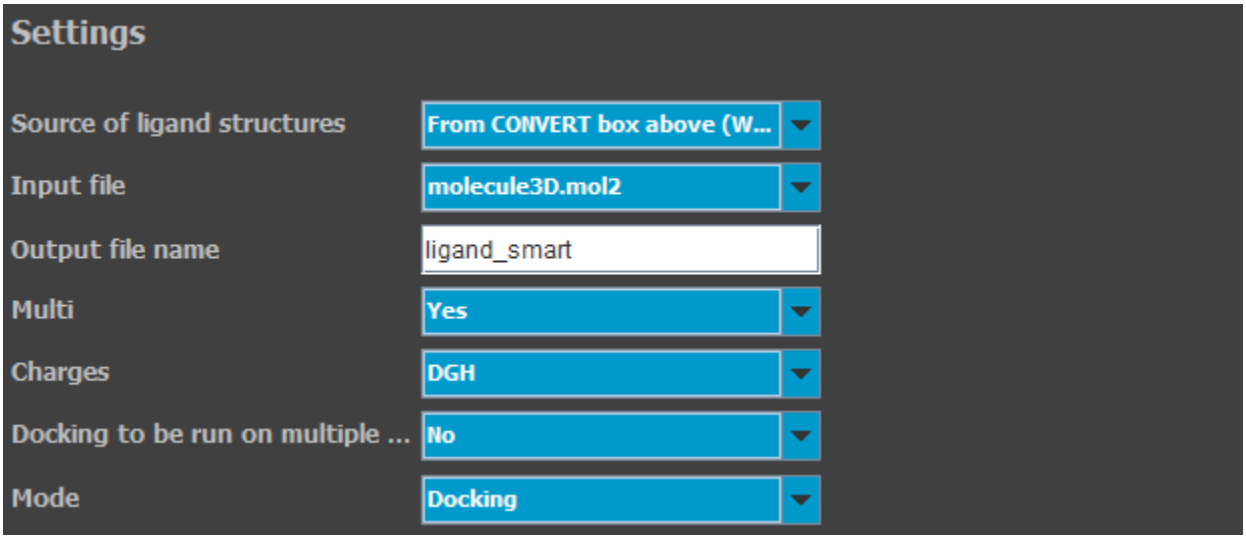
Clicking on the gear icon in the Convert box will open the parameters section. The newly created ligand file should be available in the **Input file** menu. Set a name for the **Ouput file** and select the **Mode** "2D to 3D (mol2)".



Once these parameters are set, write the keyword file and exit CONVERT.

5. SMART: The setup of the ligand for docking

SMART will setup the ligand (3D molecules only) for docking with FITTED. The **Source of ligand structures** should be set to "From CONVERT box (Workflow)". The **Input file** will automatically be selected. The **Output file name** text box can be filled with any desired name. The atomic partial charges method may be selected under **Charges**. The methods available are DGH (recommended), none (the molecule is not charged), and Input (will keep input charges in original input file).



Once the configuration is complete, write the keyword file and exit SMART.

6. FITTED: Flexible protein docking

The **Source of structures** should be set to "From boxes above (Workflow)". The **Number of protein(s)** should be set to "4". **Macromolecule** is used to differentiate between protein, metalloprotein, DNA and RNA. The **Protein flexibility mode** needs to be set to "Automatic". The **Ligand file** is the ligand that has been previously prepared by SMART. Since we are not performing a self-docking, the RMSD value should be set to "No". When multiple ligands are docked, the docking can be distributed over multiple cores. In this case, since there is only one ligand, the keyword **Run on multiple cores** should be set to "No." All other settings are set to their default values.

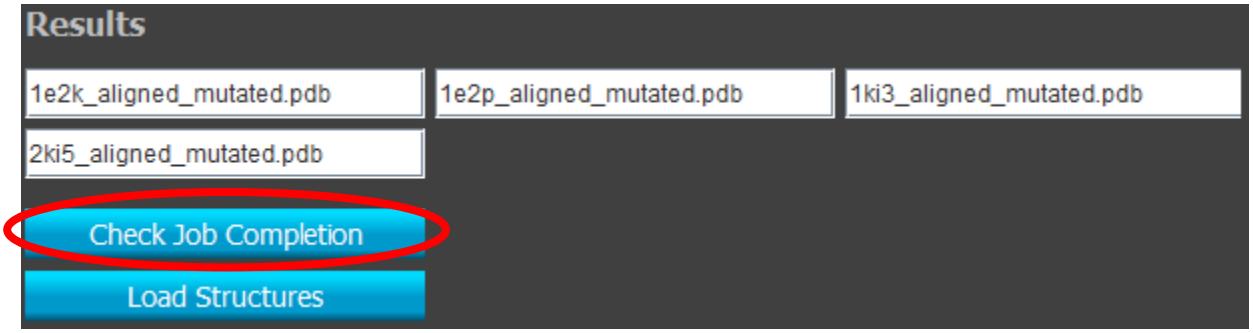
Write the keyword and exit FITTED.

7. Workflow: Running the docking workflow

You may now run the workflow. Once the forecasting is completed and the terminal window closes, you may proceed with the results analysis.

8. Result analysis

In this case, MATCH-UP will generate 4 new files. This step performs two distinct actions sequentially; the superposition and the mutation/deletion to make all the pdb files similar (xxxx_aligned_mutated.pdb). It is recommended to inspect the output files for any sequence alignment problem and/or deletion or mutation of residues close to the protein active site. Click the magnifier in the MATCH-UP box, followed by **Check Job Completion,** before loading the structures in order to verify them. If file names appear, this means that the job went to completion.



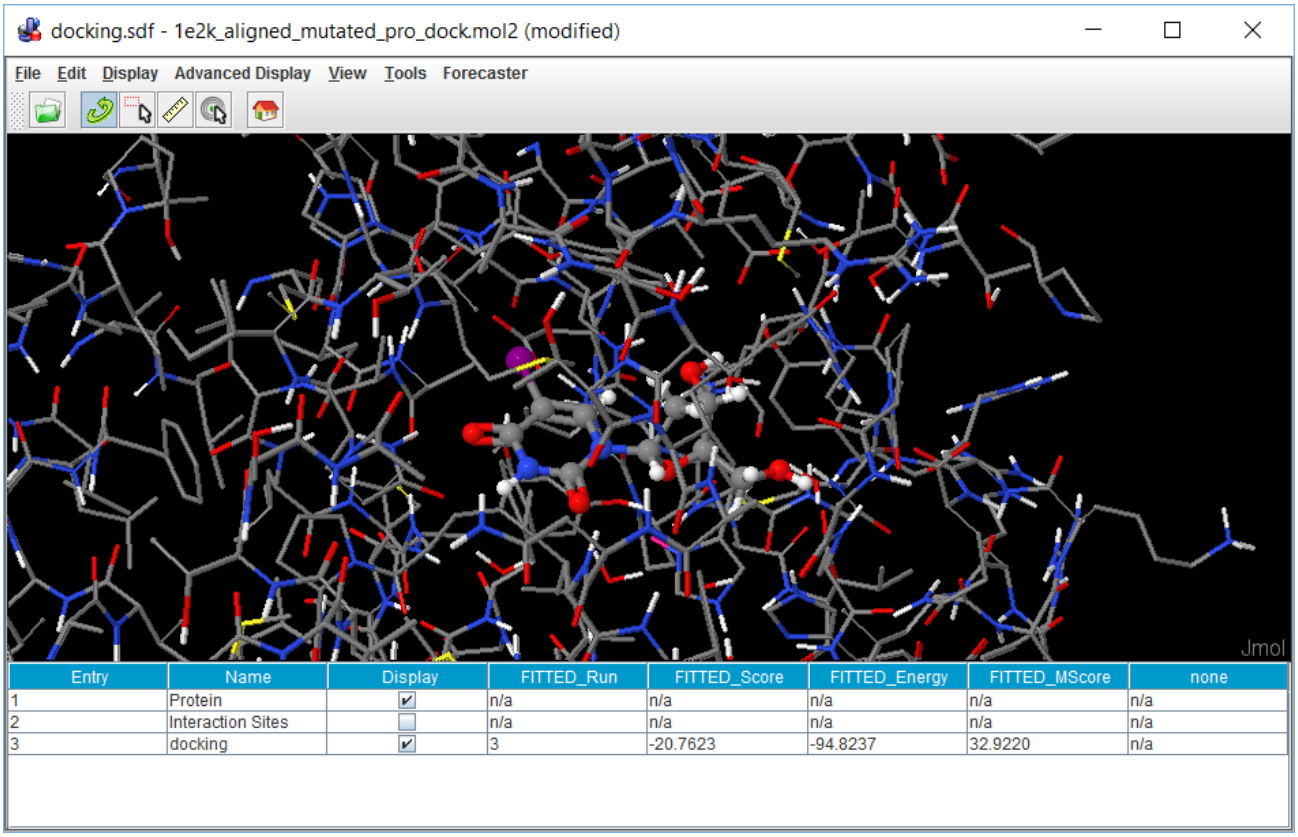
PREPARE takes the protein-ligand complexes in pdb format and creates the corresponding xxxx_aligned_mutated_pro.mol2 and xxxx_aligned_mutated_lig.mol2 files. Once again, we recommend inspecting the structures.

PROCESS will create several new files: the output file (process_protein.out), the global interaction sites (1e2k_aligned_mutated_pro_IS_flex.mol2), the binding site definition (1e2k_aligned_mutated_pro_BindSite_flex.mol2), the following files for each protein structure:xxxx_aligned_mutated_pro_dock.mol2, xxxx_aligned_mutated_score.mol2, xxxx_aligned_mutated_pro_IS.mol2, and xxxx_aligned_mutated_pro_site.txt. Some files are automatically named according to the input protein mol2 file and others are named based on the filename provided in the PROCESS keyword file. The structures may be inspected by opening them through PROCESS.

To make sure there was no problem during the 3D conversion and setup of the ligand, it is recommended to verify the output 3D file through CONVERT and the ligand output through SMART.

In the output folder, you will find the final results: docking.log (log file for potential errors), docking.out (output file), docking_Docked_Poses.mol2 (the poses in mol2 format), docking.sdf (sdf file with the poses and scores/energy), docking_Prot1_run1.mol2 (the protein corresponding to the pose in mol2), docking_Prot1_run1.pdb (the protein corresponding to the pose in pdb), and docking-results.txt (summary results file). If you open the docking-results.txt file, you can visualize the score and the energy value for each run. The docking.sdf file contains the best pose of the docked ligand and it can be visualized together with the 1e2k_pro.mol2 file in your favorite 3D graphical program.

Alternatively, the poses (i.e., proposed binding modes) can be visualized by clicking the magnifier icon of the FITTED box. Click **Open poses** and select docking.sdf. Within the 3D viewer, a table of data with the corresponding scores will be available to easily visualize and analyze the docking results. In addition, docking-results.txt contains all the scores and values for the docking poses. This file can be opened by clicking the **Read numerical results** button and selecting it.



Entry	Name	Display	FITTED_Run	FITTED_Score	FITTED_Energy	FITTED_MScore	none
1	Protein	☑	n/a	n/a	n/a	n/a	n/a
2	Interaction Sites	☐	n/a	n/a	n/a	n/a	n/a
3	docking	☑	3	-20.7623	-94.8237	32.9220	n/a