AutoDock

AutoDock
Virtual Screening and RC with Raccoon

Stefano Forli
Ruth Huey

The Scripps Research Institute
Molecular Graphics Laboratory
10550 N. Torrey Pines Rd.
La Jolla, California 92037-1000
USA

1 August 2009 v.0.1

Contents

A . Interface introduction

B. Virtual Screening generation
- -Ligands definition
- -Target definition
- -Grid box definition
- -Docking parameters definition
- -Jobs generation
- -Docking calculations and results analysis

C. RC virtual screening (HIV protease | MD targets)
- -Ligands definition
- -Target definition
- -Grid box definition
- -Docking parameters definition
- -Jobs generation
- -Docking calculations and results analysis

D. RC virtual screening (HIV protease | mutants)
- -Ligands definition
- -Target definition
- -Grid box definition
- -Docking parameters definition
- -Jobs generation
- -Docking calculations and results analysis

A. Introduction

This tutorial describe the procedure to follow for generating a virtual screening job for AutoDock using the tool Raccoon.
Raccoon is a graphical interface for processing ligand libraries in different formats (PDB, multistructureMOL2 and PDBQT), multiple receptor conformations (e.g. relaxed complex experiments) and flexible residues.

It automates some of the most common operations performed when preparing a virtual screening:

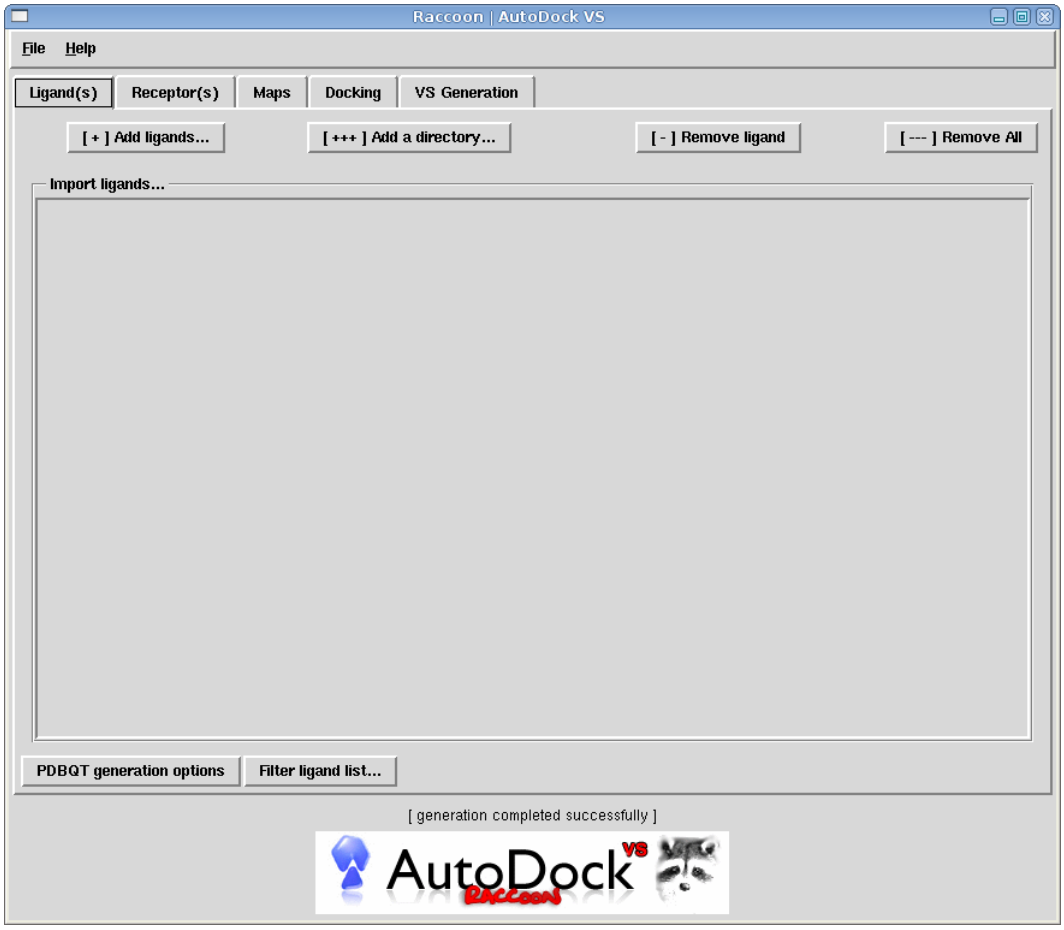
-splitting multistructure files (MOL2)
-generate PDBQT input files
-generate configuration files for both AutoGrid and AutoDock.
-generate scripts for running the virtual screening.

The interface is composed of five tabs for **Ligands**, **Receptors**, **Maps**, **Docking parameters** and **Virtual Screening Generation**.

Although few options require at least a ligand and/or a receptor to be activated (e.g. caching maps in advance), the workflow is rather flexible so there is no predefined order in filling required inputs for the VS generation.

The **Virtual Screening Generation** tab provide an overview of what's required in order to start the generation, and the inputs to be defined are colored in red. The button for starting the actual generation will be activated when all the requirements will be satisfied.

LIGANDS MENU



“ Add ligands”
It allows to load one or more ligands in any of the supported formats (PDB, single structure MOL2, multiMOL2, PDBQT).

“Add a directory”
It allows to load all the ligands in any of the supported formats contained in a directory.

Whenever a multiMOL2 is detected, a directory in which to split and convert the ligands will be prompted.

“Remove ligand” and “Remove All”
The buttons allows to remove a selected ligand or the entire set, respectively.

LIGAND MENU OPTIONS

PDBQT GENERATION OPTIONS

It contains all the available options for generating PDBQT ligands from PDB or MOL2.

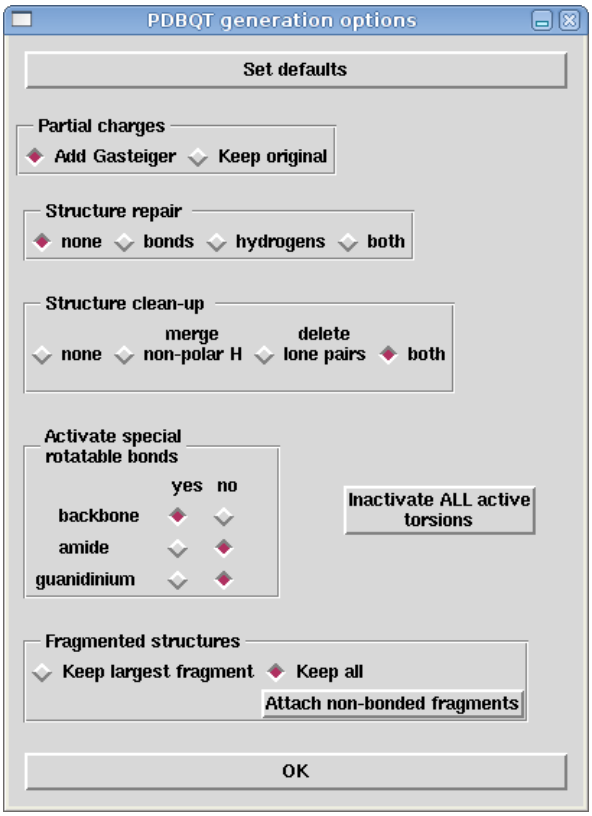
Partial charges
It allows to set the AutoDock Gasteiger charges or keep the original charges in the molecule.

Structure repair
Perform a structure repair restoring bonds and hydrogens.

Structure clean-up
Remove lone pairs and non-polar hydrogens.

Rotatable bonds
Allows to control the bonds that are going to be set as rotatable (or kept rigid)

Fragmented structures
Set the policy for managing problematic structures (e.g. containing multiple fragments, salts)



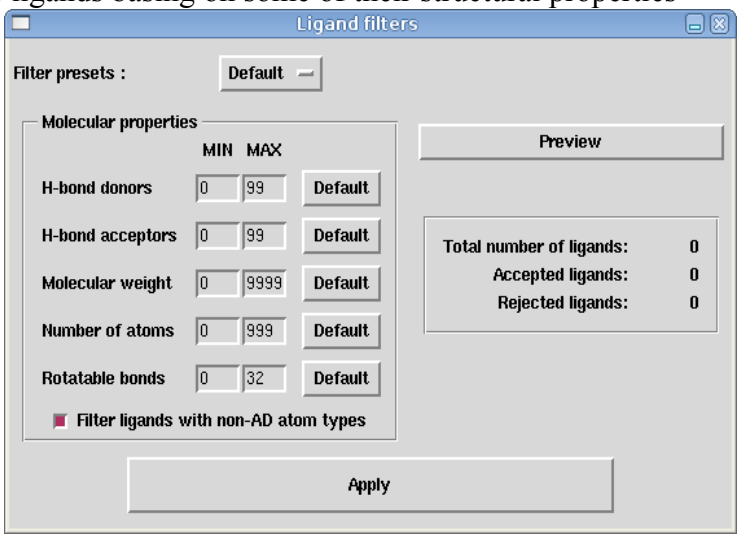
All the options will affect newly imported ligands only**.**

FILTER LIGAND LIST

It provides a simple set of parameters to filter ligands basing on some of their structural properties

Filter presets
A set of predefined values is accessible via the pull-down menu: Lipinski's rules (except for the C*log*P), DrugLike set, "DrugLike for fragments" set.

All the the sets can be previewed before being applied.
Once filters have been accepted they will be applied also to ligands loaded eventually.

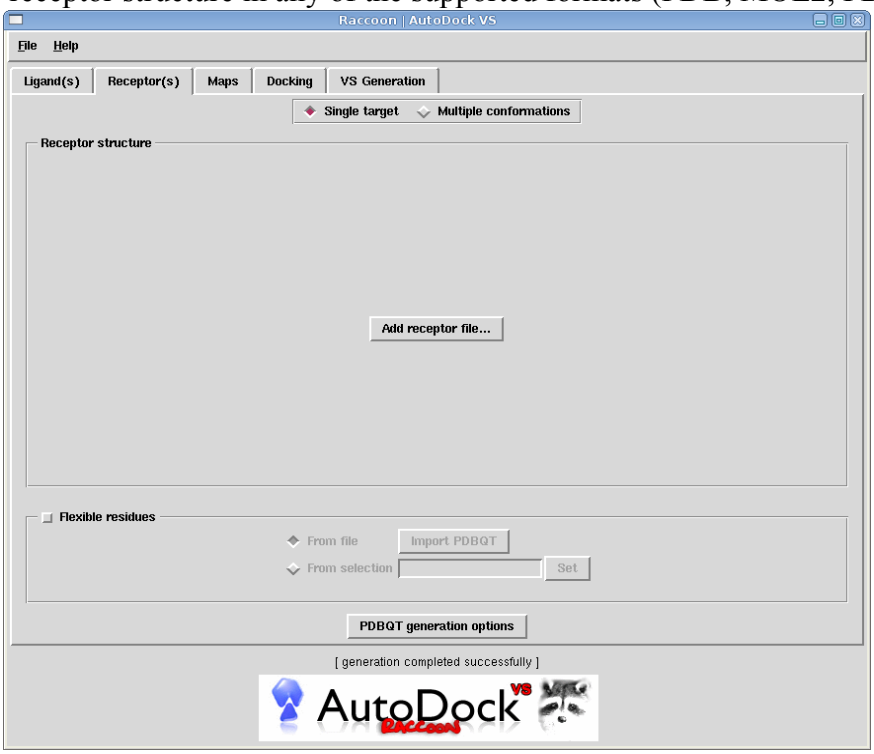


RECEPTORS MENU

SINGLE TARGET MODE

" Add Receptor File"

It allows to load the receptor structure in any of the supported formats (PDB, MOL2, PDBQT).



MULTI-TARGET MODE (Relaxed Complex)

" Add Receptor File"

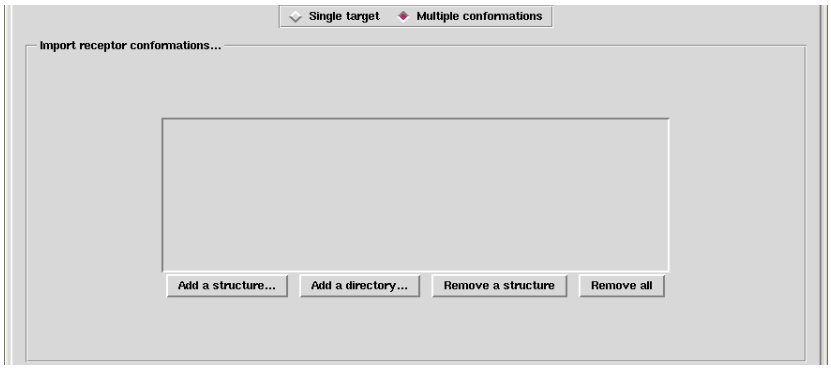
It allows to load the receptor structure in any of the supported formats (PDB, MOL2, PDBQT).

"Add a directory"

It allows to load all the receptors in any of the supported formats contained in a directory.

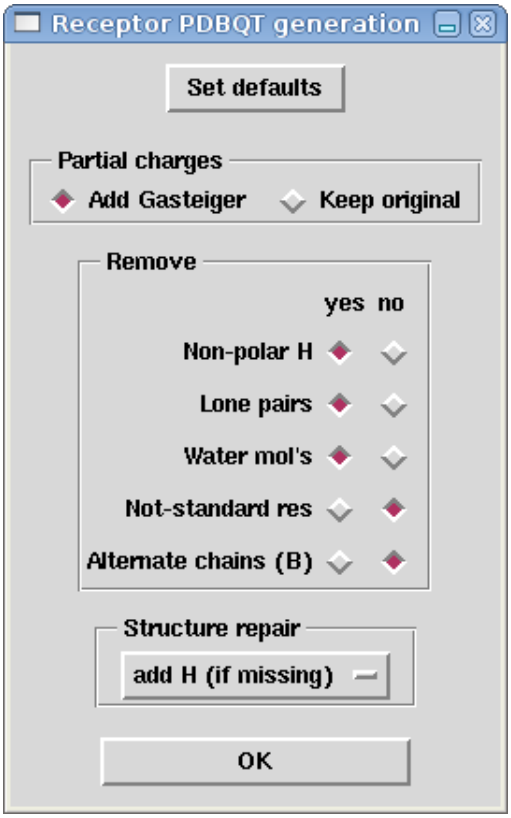
"Remove a structure" and "Remove All"

The buttons allows to remove a selected structure or the entire set, respectively.



RECEPTOR PDBQT GENERATION OPTIONS

It contains all the available options for generating the PDBQT receptor structure from PDB or MOL2.



Partial charges
It allows to set the AutoDock Gasteiger charges or keep the original charges in the molecule.

Remove
It allows to define the clean-up operations that will be performed on the receptor structure by selecting the features that will be filtered during the PDBQT generation.

Structure repair
Perform a structure repair restoring bonds and hydrogens.

All the options will affect the newly imported receptor only.

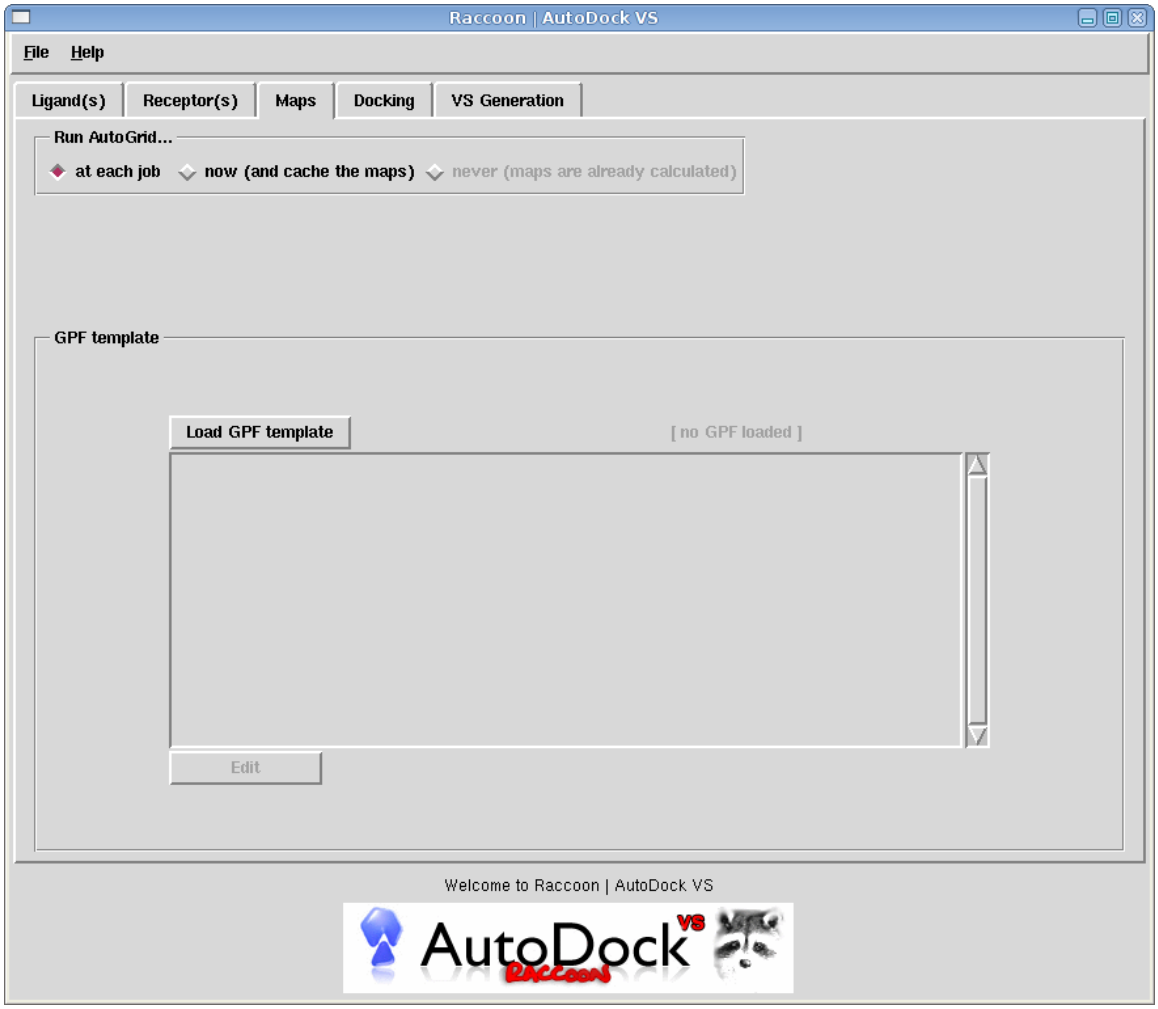
FLEXIBLE RESIDUES MENU

It contains all the available options for importing or generating the flexible residues PDBQT file.

From file *[available only in Single Target mode]*
It allows to import a flexible residue file generated with AutoDockTools

From selection
It allows to generate a flexible residues file by specifying one or more residue names. The presence of the selected residues will be checked either in the single structure or in the multiple set before being accepted.

MAPS MENU



AUTOGRID CALCULATION

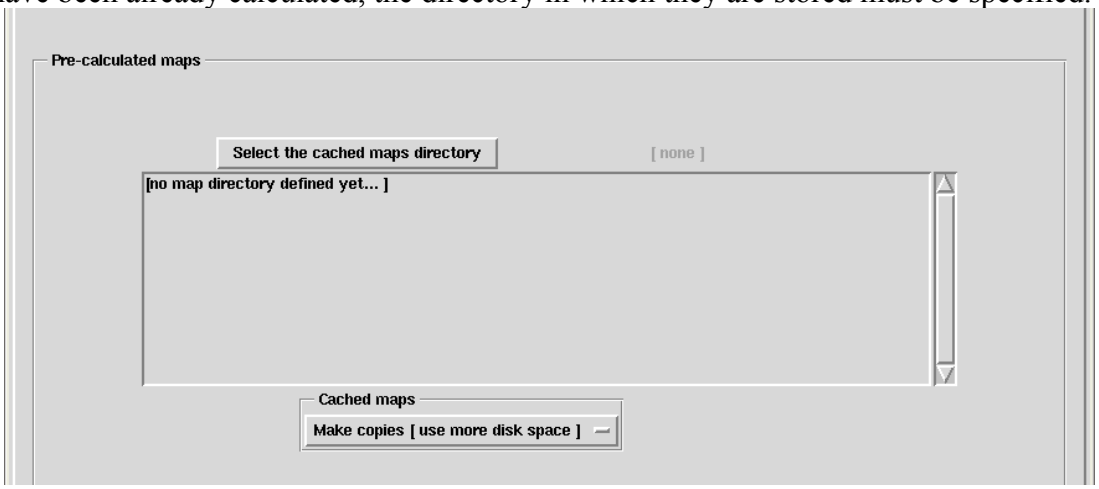
It is possible to choose if the affinity maps must be calculated in each ligand job, during the generation phase (automatically with Raccoon) or a set of precalculated maps is used in place.

For activating the cached maps options, at list a ligands and a receptor structure must be loaded.
If the maps must be calculated with Raccoon, the location of the AutoGrid binary must be specified.

Load GPF template

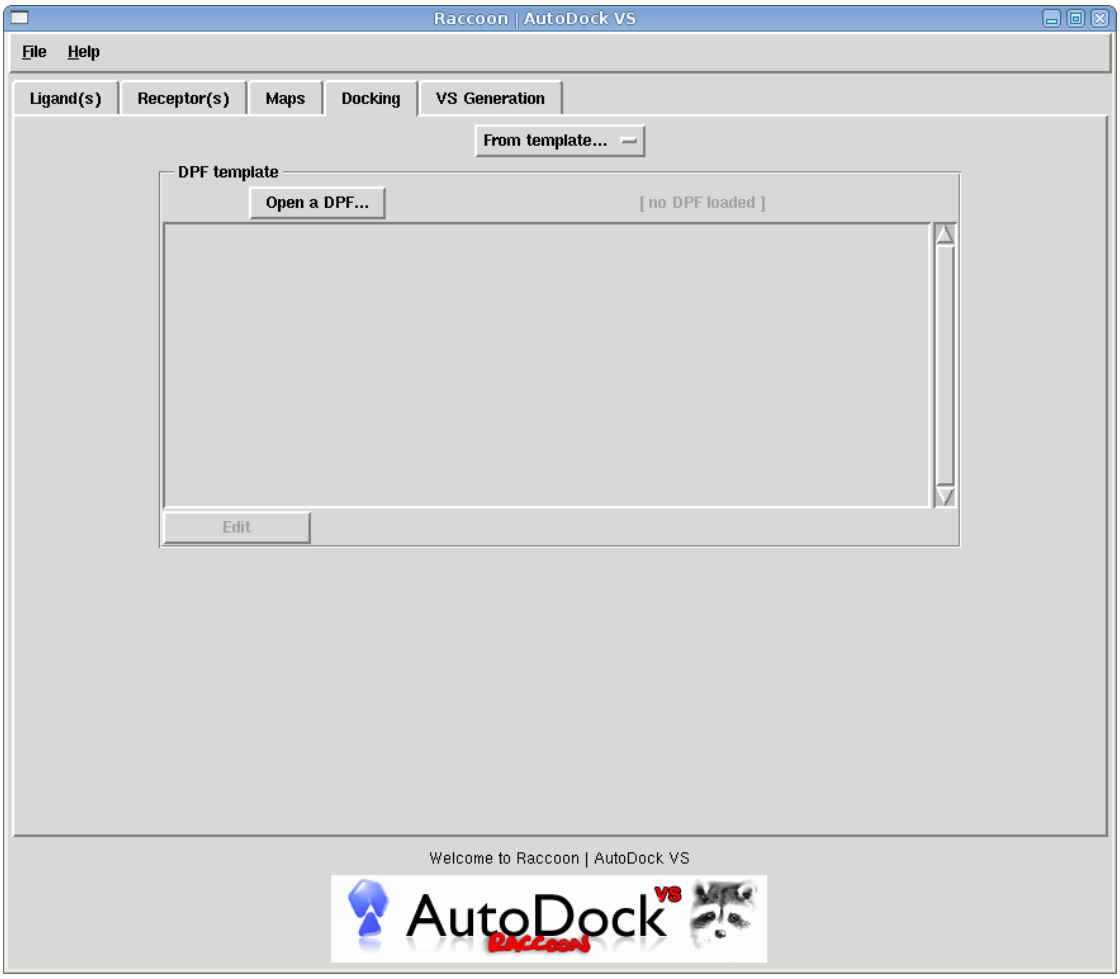
If maps need to be calculated during the VS or during the jobs preparation, a template GPF must be specified with this button. The GPF can be modified in place by pressing the “Edit” button. Recognized keywords will be represented with a bold font.

Select the cached maps directory *[available only in Single Target mode]*
If maps have been already calculated, the directory in which they are stored must be specified.



For pre-cached maps there is the choice of copying the maps in each ligand directory or create symbolic links (Linux/Macintosh only).

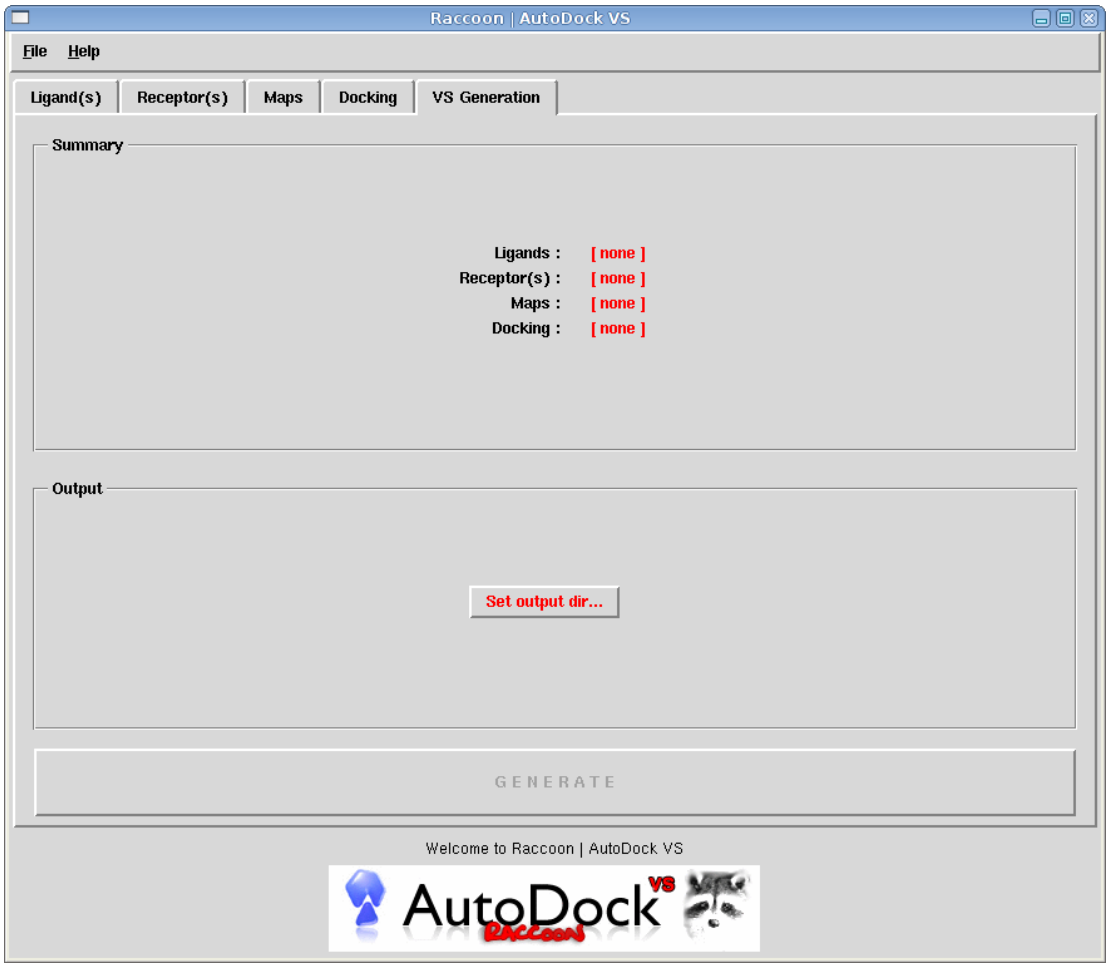
DOCKING MENU



Open a DPF...

Docking parameters are generated for each ligand by using a template DPF. The DPF can be modified in place by pressing the “Edit” button. Recognized keywords will be represented with a bold font.

VS GENERATION MENU



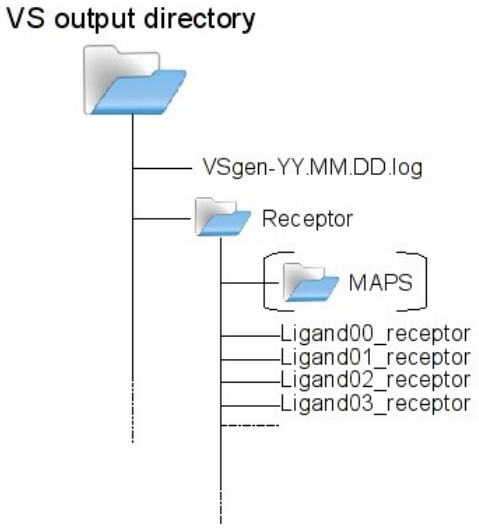
SUMMARY
A summary of the VS generation is provided in this panel. Required items that are missing are colored in red.

Set output dir...
It allows to set the VS output directory in which the jobs will be generated.

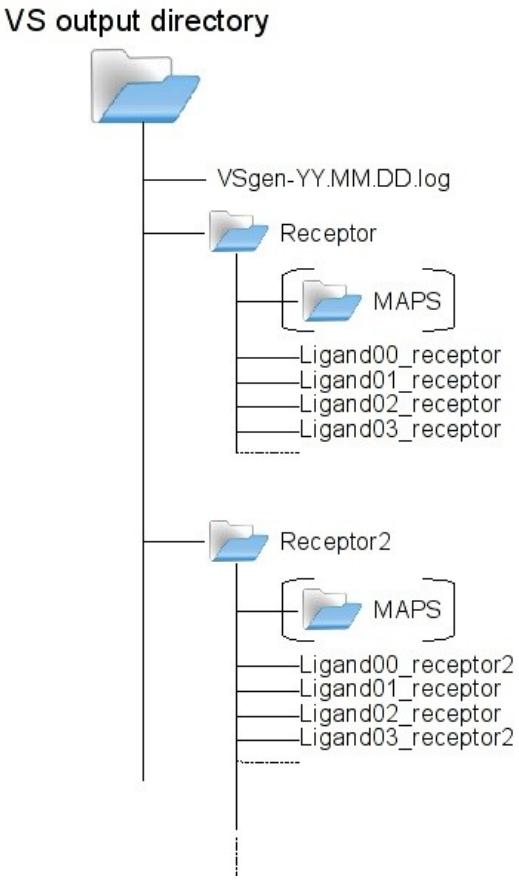
Once all the required items are defined, the "GENERATE" button will be activated.
A log file `Vsgen-YY.MM.DD.log` of the generation process will be created in the VS output directory. In each ligand directory a bash script file "run.sh" for running the single dockings will be created. In the directory containing the virtual screenings jobs is generated a master bash script to run the entire virtual screening.

VS FILES ARRANGEMENT

SINGLE TARGET



MULTIPLE TARGETS



GENERAL FILE STRUCTURE OF THE TUTORIAL

ligands : different ligand sets are available to be selected (12)

receptors : target protein structures are stored here

config : AutoGrid and AutoDock parameter files template (1 GPF, 1 DPF)

results : pre-calculated docking results to be used during the analysis step.

VS_summarize_results-XXXX-XX-XX.log

log file containing the summarized VS results.

1. VIRTUAL SCREENING GENERATION (C-Abl kinase)

WHAT: Docking of a small library of ligands against the C-Abl kinase from the AutoDock exercise. The known ligand (Imatinib) is used as a reference.
WHY: Perform a virtual screening in order to identify new molecules that could inhibit the protein.

LIGANDS DEFINITION ("Ligand(s)" tab)

1. A set of ligands will be selected to be screened on the target structure. A multistructure file downloaded from ZINC will be used as first input.

-Click on the **[+] Add ligands...** button

-Browse to the "**`ligands`**" directory and change the file type to MOL2:

-Open the file "**`small_lib.mol2`**"

The software identify the file as a multistructure file containing 10 ligands. Press "OK" to start the automatic split process. Select a directory name in which the files will be split and converted: "MySplittedMol2". Once the process is terminated the ligands will be imported in the ligand list.

2. The molecule present in the PDB structure, Imatinib, will imported to be used as a reference compound. Before doing that, check for the options that will be applied in the conversion to PDBQT by clicking on the button **PDBQT generation options**

The default options will be used for this conversion.

-Click on the **[+] Add ligands...** button

-Browse to the "**`ligands`**" directory and change the file type to PDB:

-Open the file "**`imatinib.pdbqt`**" (the reference ligand)

The PDB file will be converted and imported automatically.

3. The set of 11 ligands is ready to be used in the docking process. Ligands can be further selected by applying the filters accessible with the button **Filter ligand list...**

RECEPTOR DEFINITION ("Receptor(s)" tab)

4. If multiple conformations of the receptor have been sampled they can be used for generating a multi-target virtual screening. In this tutorial a single conformation will be used, then let the "**Single target**" option selected.

Imatinib doesn't exploit any water molecules to establish interactions with the protein, but in some ligand-protein complexes is possible that water plays a crucial role in the binding (i.e. HIV protease). By default, when generating the PDBQT file of the receptor, all the heteroatoms (non-aminoacid atoms) are removed. In order to keep the single water molecule, the PDBQT generation options must be changed by clicking on the button **PDBQT generation options** before the PDB structure is loaded.
In the "**Remove**" panel, set "**Water mol's**" to "**no**" and click on "**OK**".

5. Load the receptor file by clicking on the button **Add receptor file...** wse to the "`receptors`" directory. Change the filetype to PDB and load the receptor file "`1iep_receptor.pdbqt`"
The file will be converted in PDBQT and imported.

FLEXIBLE RESIDUES [OPTIONAL]

5.1. In this phase is possible to define residues that will be considered flexible during the docking process. As an exercise, activate the "**Flexible residues**" option and select the option "From selection". Type "**GLU286**" in the entry field and press "Set". Two residues should be selected, because the protease receptor is an dimer structure. Try "**A:GLU286**" to specify the A chain, so that a single residue should be selected.
In this tutorial the structure will be kept completely rigid, then disable the "**Flexible residues**" option.

MAPS DEFINITION ("Maps" tab)

6. AutoGrid affinity maps must be calculated before the actual docking is performed. As default option, the maps are calculated independently for each ligand by using a template GPF file as a reference. Especially when docking many ligands, though, the maps can be calculated in advance just once to save time. Eventually, the files can be copied or linked to the respective ligand directories.

Load the template by clicking on **Load GPF template** and browse to the "`config`" directory and select the file "`kinase_binding_site_template.gpf`" generated with AutoDockTools.
The grid parameter file will be shown in the editor panel. The text can be changed in place for "last minute" modifications by pressing the "Edit" button.

Select the option "**At each job**" to calculate the maps in each ligand job. This will generate a grid parameter file (GPF) for each ligand.

DOCKING PARAMETERS DEFINITION ("Docking" tab)

7. Generate a default DPF file by clicking on the **Generate DPF** button.
It's a good practice to reduce the output provided by AutoDock during a virtual screening. For this reason the "**outlev**" value in the DPF must be set to 0. Moreover, to sample more results, the number of conformations generated per ligand will be changed to 100 by setting the "**ga_run**" keyword.

Use the **Edit** to change the suggested values and click on **Apply** when finished.

VIRTUAL SCREENING GENERATION ("VS Generation" tab)

8. In the "Summary" panel all the input already defined are shown. Required inputs that are missing are colored in red. The last input to provide is the directory in which to generate the files for performing the virtual screening. Click on the **Set output dir...** button, and browse to the "virtualscreening" directory. Add "`VS_jobs`" to the "Selection" entry ["…/virtualscreening/**VS_jobs**"] and press "OK".

Since the directory doesn't exist, press "OK" to create it.

If all the previous operations succeed, the "**G E N E R A T E**" button should be now be activated. Press it and wait until it finish. The operations that will be performed are:

- creation of the directory containing the virtual screening (one per receptor);
- run AutoGrid for generating all the maps corresponding to the atom types in the ligand set using the GPF input template;
- create a sub-directory for each ligand and copy the ligand file;
- copy (or link) the required maps in each ligand directory;
- generate the docking parameter file using the DPF input template
- generate a script file for running the docking in each ligand directory ["`run.bat`"];
- generate a master script file for running the virtual screening ["`RunVS.bat`"];
- generate a log file containing all the information necessary to reproduce the virtual screening ["`Raccoon-YYYY.MM.DD.log`"].

RUNNING VIRTUAL SCREENING (terminal)

9. Open the Windows file browser and browse to the directory containing the the files generated by Raccoon. Inspect the log file using a text editor to check that all the virtual screening has been generated correctly:

```
VSgen-xxxx.xx.xx.log
```

10. Run the virtual screening by double-clicking on the `RunVS.bat` script file

ANALYZE THE RESULTS (terminal) [OPTIONAL]

11. On the terminal move to the directory containing the pre-calculated results:

```
cd ../../results/
```

12. Run the script for summarizing the DLG of the virtual screening results:

```
C:\Python25\python.exe [PATH]\Utilities24\summaryze_results4.py
```

VISUALLY INSPECT THE RESULTS IN ADT

13. Inspect the `VS_summarize_results-2009-12-10.log` file

14. Import and visually inspect the top results using ADT.

2. RC EXPERIMENT GENERATION
(HIV protease | MD targets)

WHAT: Docking of a small library of ligands against 5 frames from a 20 ns MD simulation of the HIV protease. Reference ligands (Indinavir, 5-cyclo-octapyranone) are used to compare the results from the virtual screening.

WHY: Sample multiple conformations of the same protein that are different from the crystallografic state.

LIGANDS DEFINITION ("Ligand(s)" tab)

1. A set of ligands will be selected to be screened on the target structure. A multistructure file downloaded from ZINC will be used as first input.

-Click on the **[+] Add ligands...** button

-Browse to the "**`ligands`**" directory and change the file type to MOL2:

-Open the file "**`small_lib.mol2`**"

The software identify the file as a multistructure file containing 10 ligands. Press "OK" to start the automatic split process. Select a directory name in which the files will be split and converted: "MySplittedMol2". Once the process is terminated the ligands will be imported in the ligand list.

2. The molecule present in the PDB structure, Indinavir, will imported to be used as a reference compound, together with the structure of another protease inhibitor, 5-cyclo-octalpyranone.
Before doing that, check for the options that will be applied in the conversion to PDBQT by clicking on the button **PDBQT generation options**

The default options will be used for this conversion.

-Click on the **[+] Add ligands...** button

-Browse to the "**`ligands`**" directory and change the file type to PDB:

-Open the files "**`indinavir.pdbqt`**" and "**`5-cyclo-octapyranone.pdbqt`**" (the reference ligands)

The PDB files will be converted and imported automatically.

3. The set of 11 ligands is ready to be used in the docking process. Ligands can be further selected by applying the filters accessible with the button **Filter ligand list...**

RECEPTOR DEFINITION ("Receptor(s)" tab)

4. Multiple conformations of the receptor will be used for generating a multi-target virtual screening. Switch to the "**Multiple conformations**" mode.

Indinavir exploits a water molecule to establish interactions with the protein, therefore that water plays a crucial role in the binding to the HIV protease. By default, when generating the PDBQT file of the receptor, all the heteroatoms (non-aminoacid atoms) are removed.
In order to keep the single water molecule, the PDBQT generation options should be changed by clicking on the button **PDBQT generation options** before the PDB structure is loaded then in the "**Remove**" panel, set "**Water mol's**" to "**no**" and click on "**OK**".
For this simulation, though, we're going to ignore any water molecule, therefore default parameters can be used as well.

5. Load the receptor file by clicking on the button **Add a structure**, browse to the "`targets`" directory and then in the "`20ns_molecular_dynamics`" sub-directory. Change the file type to PDBQT and load the receptor structures. Multiple structures can be loaded in the same session by using Ctrl and/or Shift while selecting the files.

MAPS DEFINITION ("Maps" tab)

6. AutoGrid affinity maps must be calculated before the actual docking is performed. As default option, the maps are calculated independently for each ligand by using a template GPF file as a reference. Especially when docking many ligands, though, the maps can be calculated in advance just once to save time. Eventually, the files can be copied or linked to the respective ligand directories.

Load the template by clicking on **Load GPF template** and browse to the "`config`" directory and select the file "`template_MD.gpf`" previously generated with AutoDockTools.
The grid parameter file will be shown in the editor panel. The text can be changed in place for "last minute" modifications by pressing the "Edit" button.

Select the option "**At each job**" to calculate the maps in each ligand job. This will generate a grid parameter file (GPF) for each ligand.

DOCKING PARAMETERS DEFINITION ("Docking" tab)

7. Generate a default DPF file by clicking on the **Generate DPF** button.
It's a good practice to reduce the output provided by AutoDock during a virtual screening. For this reason the "**outlev**" value in the DPF must be set to 0. Moreover, to sample more results, the number of conformations generated per ligand will be changed to 100 by setting the "**ga_run**" keyword.

Use the **Edit** to change the suggested values and click on **Apply** when finished.

VIRTUAL SCREENING GENERATION ("VS Generation" tab)

8. In the "Summary" panel all the input already defined are shown. Required inputs that are missing are colored in red. The last input to provide is the directory in which to generate the files for performing the virtual screening. Click on the **Set output dir...** button, and browse to the "virtualscreening" directory. Add "`VS_jobs`" to the "Selection" entry ["..../virtualscreening/**VS_jobs**"] and press "OK".

Since the directory doesn't exist, press "OK" to create it.

If all the previous operations succeed, the "**G E N E R A T E**" button should be now be activated. Press it and wait until it finish. The operations that will be performed are:

- creation of the directory containing the virtual screening (one per receptor);
- Generate the GPF to run AutoGrid by using the GPF template.
- create a sub-directory for each ligand and copy the ligand file;
- copy (or link) the required maps in each ligand directory;
- generate the docking parameter file using the DPF input template
- generate a script file for running the docking in each ligand directory ["`run.bat`"];
- generate a master script file for running the virtual screening ["`RunVS.bat`"];
- generate a log file containing all the information necessary to reproduce the virtual screening ["`Raccoon-YYYY.MM.DD.log`"].

RUNNING VIRTUAL SCREENING (terminal)

9. Open the Windows file browser and browse to the directory containing the the files generated by Raccoon. Inspect the log file using a text editor to check that all the virtual screening has been generated correctly:

```
VSgen-xxxx.xx.xx.log
```

10. Run the virtual screening by double-clicking on the `RunVS.bat` script file

ANALYZE THE RESULTS (terminal) [OPTIONAL]

11. On the terminal move to the directory containing the pre-calculated results:

```
cd ../../results/
```

12. Run the script for summarizing the DLG of the virtual screening results:

```
C:\Python25\python.exe [PATH]\Utilities24\summaryze_results4.py
```

VISUALLY INSPECT THE RESULTS IN ADT

13. Inspect the `VS_summarize_results-2009-12-10.log` file

14. Import and visually inspect the top results using ADT.

3. MULTIPLE TARGET EXPERIMENT GENERATION

(HIV protease | mutant targets)

WHAT: Docking of a small library of ligands against two mutants of the HIV protease. Reference ligands (Indinavir, 5-cyclo-octapyranone) are used to compare the results from the virtual screening.
WHY: Multiple mutants of the same protein can be used to screen the same library in order to find ligands able to overcome the mutations.

LIGANDS DEFINITION ("Ligand(s)" tab)

1. A set of ligands will be selected to be screened on the target structure. A multistructure file downloaded from ZINC will be used as first input.

-Click on the **[+] Add ligands...** button

-Browse to the "**`ligands`**" directory and change the file type to MOL2:

-Open the file "**`small_lib.mol2`**"
The software identify the file as a multistructure file containing 10 ligands. Press "OK" to start the automatic split process. Select a directory name in which the files will be split and converted: "MySplittedMol2". Once the process is terminated the ligands will be imported in the ligand list.

2. The molecule present in the PDB structure, Imatinib, will imported to be used as a reference compound. Before doing that, check for the options that will be applied in the conversion to PDBQT by clicking on the button **PDBQT generation options**

The default options will be used for this conversion.

-Click on the **[+] Add ligands...** button

-Browse to the "**`ligands`**" directory and change the file type to PDB:

-Open the files "**`indinavir.pdbqt`**" and "**`5-cyclo-octapyranone.pdbqt`**" (the reference ligands)

The PDB files will be converted and imported automatically.

3. The set of 11 ligands is ready to be used in the docking process. Ligands can be further selected by applying the filters accessible with the button **Filter ligand list...**

RECEPTOR DEFINITION ("Receptor(s)" tab)

4. In this experiment, different "flavors" of the same protein will be used for generating a multi-target virtual screening against two mutants. Switch to the "**Multiple conformations**" mode.

5. Load the receptor file by clicking on the button **Add a structure**, browse to the "`targets`" directory and then in the "`Xmutants`" sub-directory. Change the file type to PDB and load the receptor structures. Multiple structures can be loaded in the same session by using Ctrl and/or Shift while selecting the files. The files will be converted in PDBQT and imported.

MAPS DEFINITION ("Maps" tab)

6. AutoGrid affinity maps must be calculated before the actual docking is performed. As default option, the maps are calculated independently for each ligand by using a template GPF file as a reference. Especially when docking many ligands, though, the maps can be calculated in advance just once to save time. Eventually, the files can be copied or linked to the respective ligand directories.

Load the template by clicking on **Load GPF template** and browse to the "`config`" directory and select the file "`template_MUT.gpf`" generated with AutoDockTools.
The grid parameter file will be shown in the editor panel. The text can be changed in place for "last minute" modifications by pressing the "Edit" button.

Select the option "**Now (and cache the maps)**" to calculate the maps only once, when generating the VS. Select either "**Make copies**" or "**Make symbolic links**" in the "**Cached maps**" panel.

DOCKING PARAMETERS DEFINITION ("Docking" tab)

7. Generate a default DPF file by clicking on the **Generate DPF** button.
It's a good practice to reduce the output provided by AutoDock during a virtual screening. For this reason the "**outlev**" value in the DPF must be set to 0. Moreover, to sample more results, the number of conformations generated per ligand will be changed to 100 by setting the "**ga_run**" keyword.

Use the **Edit** to change the suggested values and click on **Apply** when finished.

VIRTUAL SCREENING GENERATION ("VS Generation" tab)

8. In the "Summary" panel all the input already defined are shown. Required inputs that are missing are colored in red. The last input to provide is the directory in which to generate the files for performing the virtual screening. Click on the **Set output dir...** button, and browse to the "virtualscreening" directory. Add "`VS_jobs`" to the "Selection" entry ["…/virtualscreening/**VS_jobs**"] and press "OK".

Since the directory doesn't exist, press “OK” to create it.

If all the previous operations succeed, the “**G E N E R A T E**” button should be now be activated. Press it and wait until it finish. The operations that will be performed are:

- creation of the directory containing the virtual screening (one per receptor);
- run AutoGrid for generating all the maps corresponding to the atom types in the ligand set using the GPF input template;
- create a sub-directory for each ligand and copy the ligand file;
- copy (or link) the required maps in each ligand directory;
- generate the docking parameter file using the DPF input template
- generate a script file for running the docking in each ligand directory [“`run.bat`”];
- generate a master script file for running the virtual screening [“`RunVS.bat`”];
- generate a log file containing all the information necessary to reproduce the virtual screening [“`Raccoon-YYYY.MM.DD.log`”].

RUNNING VIRTUAL SCREENING (terminal)

9. Open the Windows file browser and browse to the directory containing the the files generated by Raccoon. Inspect the log file using a text editor to check that all the virtual screening has been generated correctly:

```
VSgen-xxxx.xx.xx.log
```

10. Run the virtual screening by double-clicking on the `RunVS.bat` script file

ANALYZE THE RESULTS (terminal) [OPTIONAL]

11. On the terminal move to the directory containing the pre-calculated results:

```
cd ../../results/
```

12. Run the script for summarizing the DLG of the virtual screening results:

```
C:\Python25\python.exe [PATH]\Utilities24\summaryze_results4.py
```

VISUALLY INSPECT THE RESULTS IN ADT

13. Inspect the `VS_summarize_results-2009-12-10.log` file

14. Import and visually inspect the top results using ADT.