Portal

www.dockthor.lncc.br

User Guide

Version 1.0

Contributors

Diogo A. Marinho, Isabella A. Guedes, Eduardo Krempser, Camila S. de Magalhães, Hélio J. C. Barbosa and Laurent E. Dardenne

www.gmmsb.lncc.br

Laboratório Nacional de Computação Científica – LNCC/MCTI

14 July 2013

INPI Software Registration Number 13318-3

Portal DockThor

1. Introduction

The DockThor Portal, developed by the group GMMSB/LNCC, is a free receptor-ligand docking server idealized to facilitate and enable the use of the docking methodology by the academic community. The implemented DockThor® program is a flexible-ligand and rigid-receptor grid based method that employs a multiple solution genetic algorithm and the MMFF94S molecular force field scoring function. The major steps of the ligand and protein preparation are available on the DockThor Portal, being possible to change the protonation states of the amino acid residues and include cofactors (*e.g.* structural water molecules, metals, organic molecules) as rigid entities. The user can also customize the main parameters of the energy grid and the genetic algorithm.

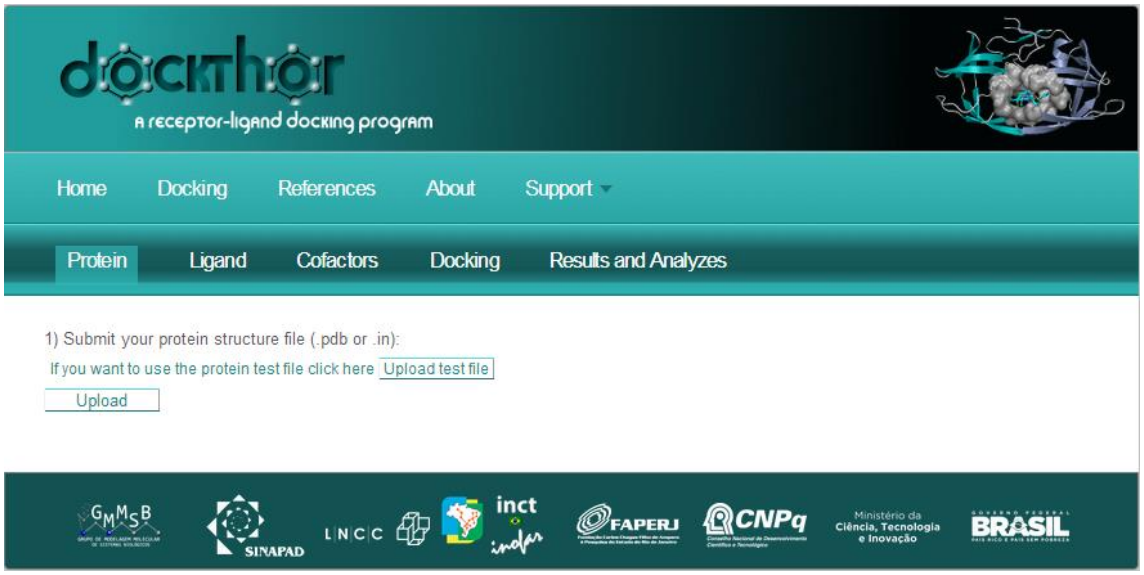
The results of the docking process are analyzed and ordered automatically. The parameters of the analysis can also be customized by the user. The DockThor Portal employs the computational facilities provided by the Brazilian SINAPAD (Sistema Nacional de Alto Desempenho) high performance platform.

2. Submiting a Docking Job

In the present version 1.0 of the DockThor Portal only protein receptors and only .pdb filetype are accepted (or pre-prepared DockThor input files *.in). Non-proteic receptors (*e.g.* DNA, RNA, another ligand) will be allowed in the next portal version.

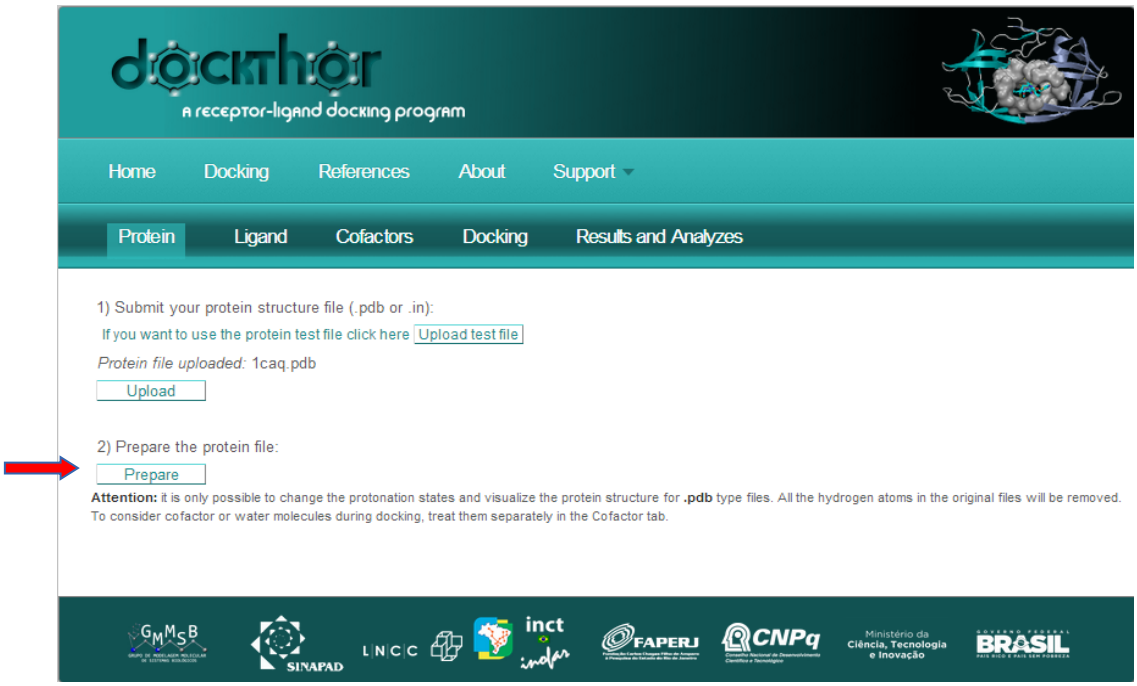
2.1 Protein Preparation

1. Click on the **Docking** tab. After, click on the **Protein** tab to open the protein preparation page.

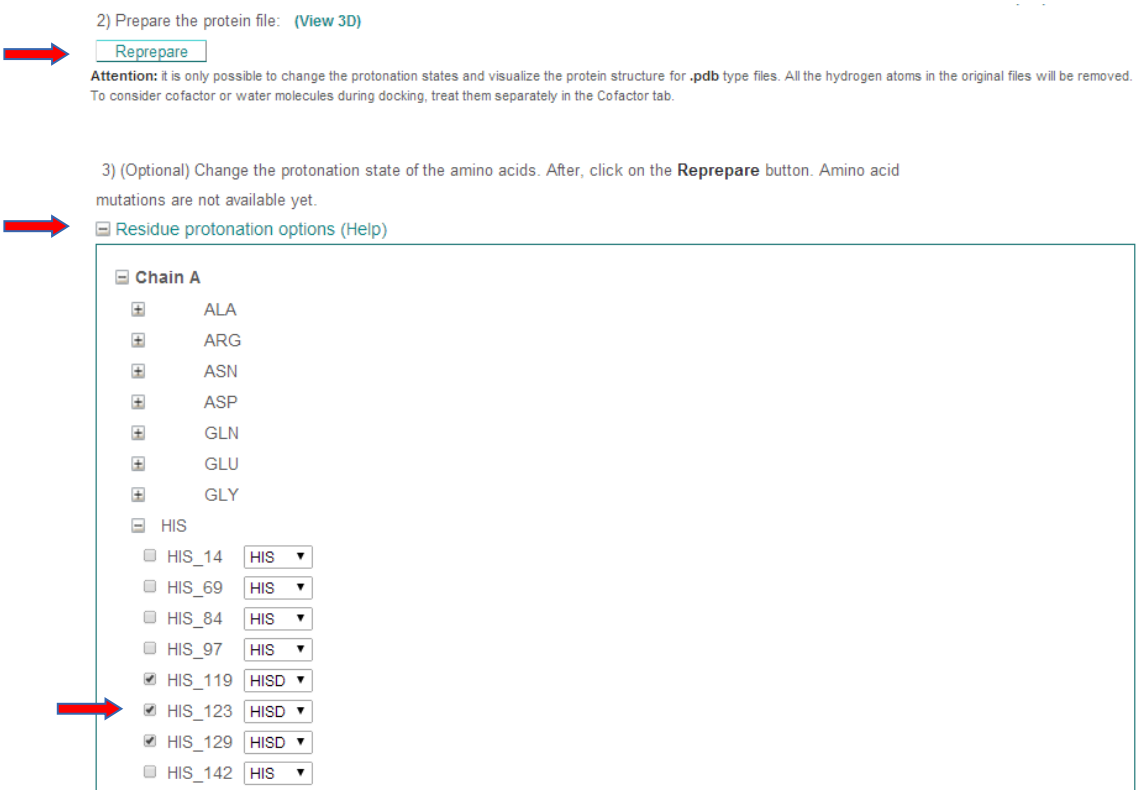


2. To submit your protein file click on the **Upload** button. So far, it is only possible to upload *Protein Data Bank* (.pdb) type or *DockThor input* (.in) protein files. The structure of the catalytic domain of human stromelysin complexed with a non-peptide inhibitor (PDB code 1CAQ) is available as a test case. To use this structure, click on **Upload test file**.
3. Prepare your protein file with the basic options clicking on **Prepare**. In this step, all the missing amino acid residue side chain atoms will be reconstructed. The protein atoms are recognized by the initial .pdb label 'ATOM', Atoms associated with the initial .pdb label 'HETATM' are ignored. All the atoms are also recognized by their .pdb atom label (*e.g.* CA, CB etc). If the atom label nomenclature is not correct the atom will be reconstructed. For this reason, all the hydrogen atom positions will probably be reconstructed[1]. If two side chain conformations are present for the same residue in the .pdb file, only the first one will be considered by the program.

[1] To maintain the original H's positions you have to use the same H atom .pdb nomenclature used by the DockThor program (download and examine the generated *.in file).



4. Set the protonation states of the residues (Asp, Cys, Glu and His are set to the default values - see appendix A). In the protein test file it is important to change the protonation states of some histidine residues, which interact with metals: His-123, His-119 and His-129 must be of type HISD.



5. Reprepare your protein file with the new protonation states clicking on **Reprepare**.
6. View the prepared protein file with JSmol clicking on (**View 3D)**. Useful information about the use of this plugin can be found in http://www.chem.uwec.edu/JmolTut/.
7. Click on **Download prepared files** to download the respective protein files.

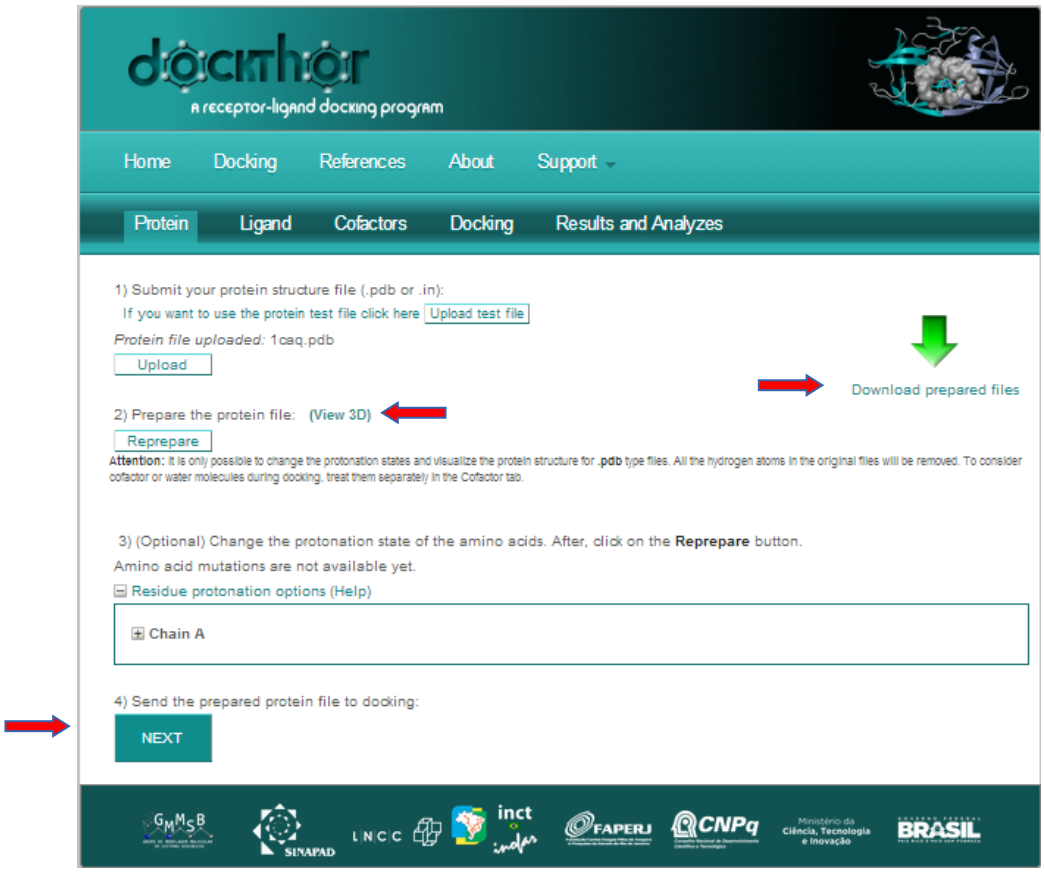
   protein.in – DockThor receptor input file (contains the MMFF94 atomic type number, the atomic coordinates, the bond connectivity, the atomic partial charges, the pdb type atom, residue and chain labels and the residue and chain numbers).

   resumo.out – contains the information of the protein preparation process.

   protein_prep.pdb – prepared protein file in the *Protein Data Bank* format (equivalent to the protein.in file for visualization).

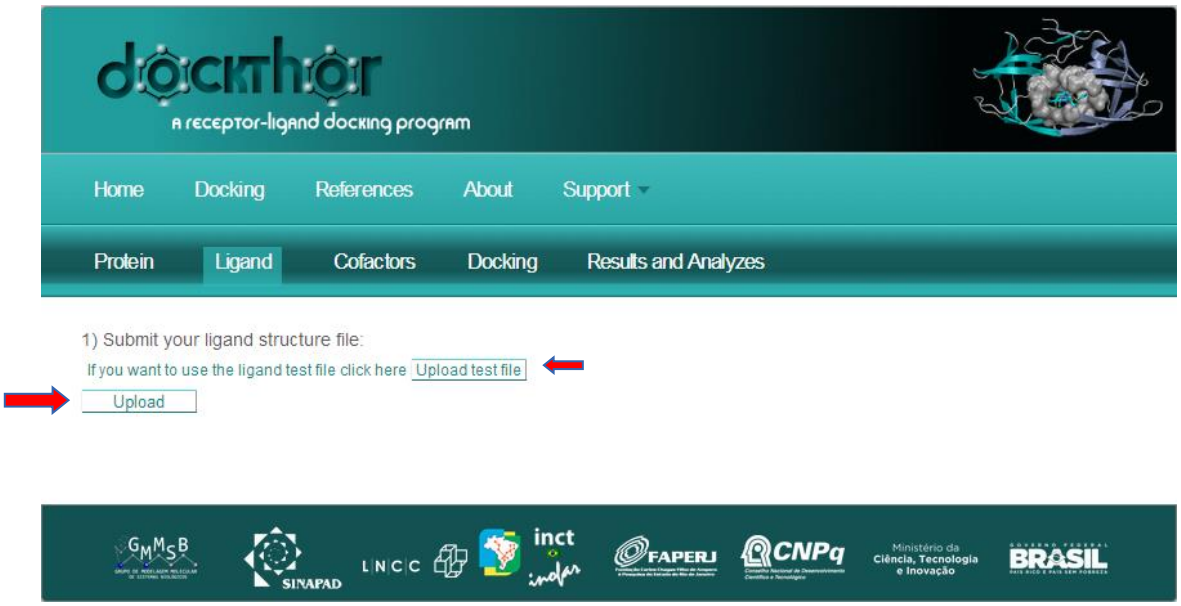
   protein.X - configuration file for each protein chain X. Contains the amino acid residue list and the respective protonation state label.

8. Click on **Next** to submit your prepared protein file to docking. It is important to notice that only the protein file in the format *.in* is necessary to docking.



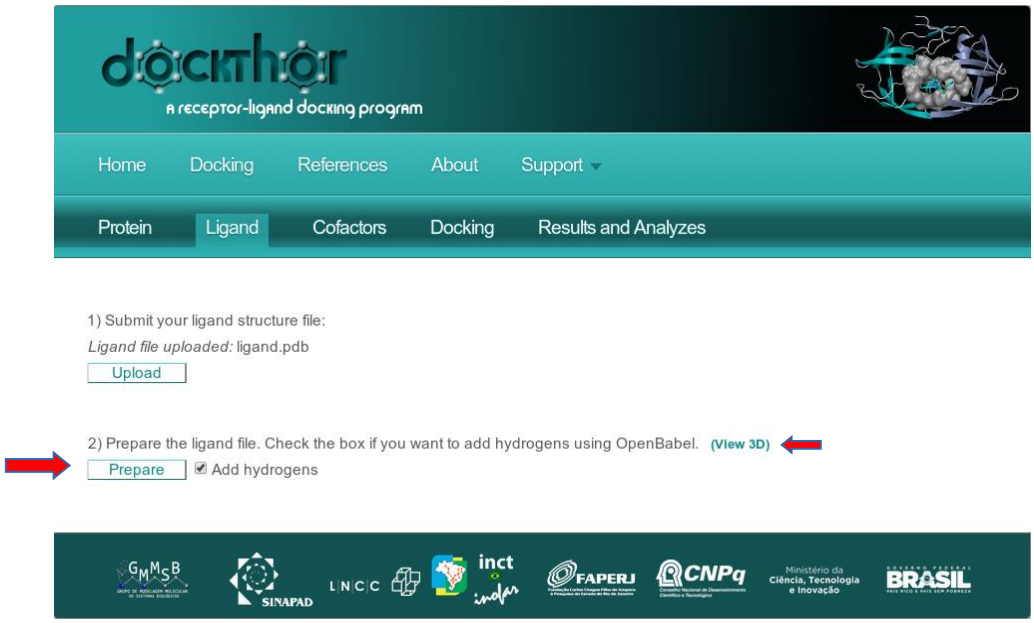
2.2 Ligand Preparation

1. To submit a single small molecule[2] to docking, click **Upload**. If you want to use the ligand test file (non-peptide inhibitor, PDB identifier DPS) click on **Upload test file**).



2. The recent uploaded ligand must be prepared to generate the respective topology file (.top) for docking. This step comprises the correct MMFF94S force field atom type assignment, the atomic partial charges calculation and the assignment of the rotatable chemical bonds. If you want to add hydrogen atoms automatically, just check the **Add hydrogens box** and **Reprepare** the molecule again. The hydrogen atoms will be added by OpenBabel tool using pH = 7.0.
3. To view the most recent structure of your ligand, click (View 3D).

[2] In the present portal version, it is possible to dock only one ligand at time. A virtual screening version of the DockThor Portal will be available in the future.



4. You can view the selected rotatable bonds just clicking on the "*Rotatable bonds*..." button. If you uncheck some flexible rotatable bonds click on **Reprepare**.

Attention: If you click on **Reprepare** twice without check/uncheck one or more rotatable bonds, all boxes will be automatically checked.

5. To download the prepared ligand files click on **Download prepared files**.

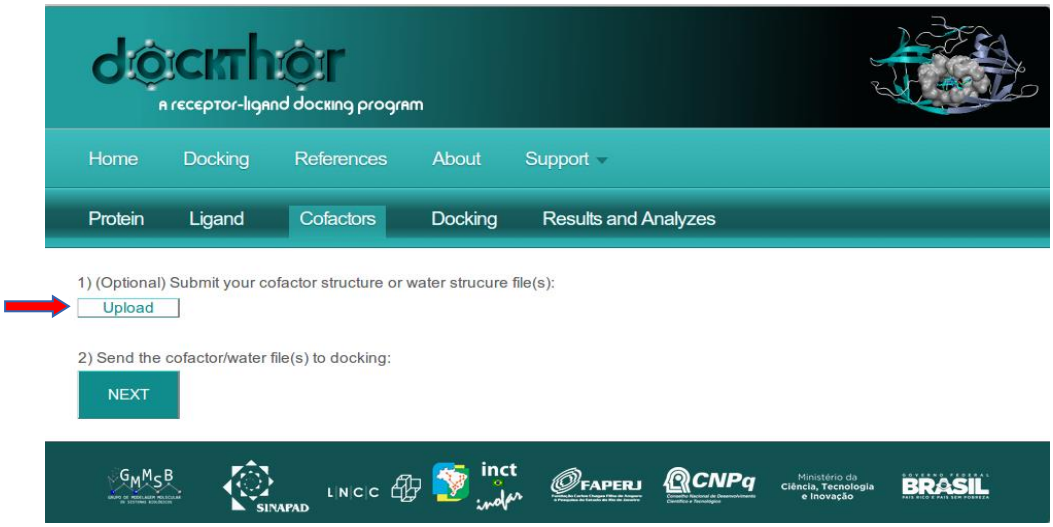
   ligand.top – DockThor ligand input file. It contains the atom name, the atom number, the MMFF94S atomic type number, the atomic partial charges, the atomic coordinates and the atom valence. It contains also: (i) the atom connectivity; (ii) force field torsional parameters; (iii) selected flexible bonds; (iv) non-bonded intramoleular atom interactions.

   new_ligand.pdb – prepared ligand in the Protein Data Bank format (for visualization). This file is generated only when the Add hydrogens option is chosen.

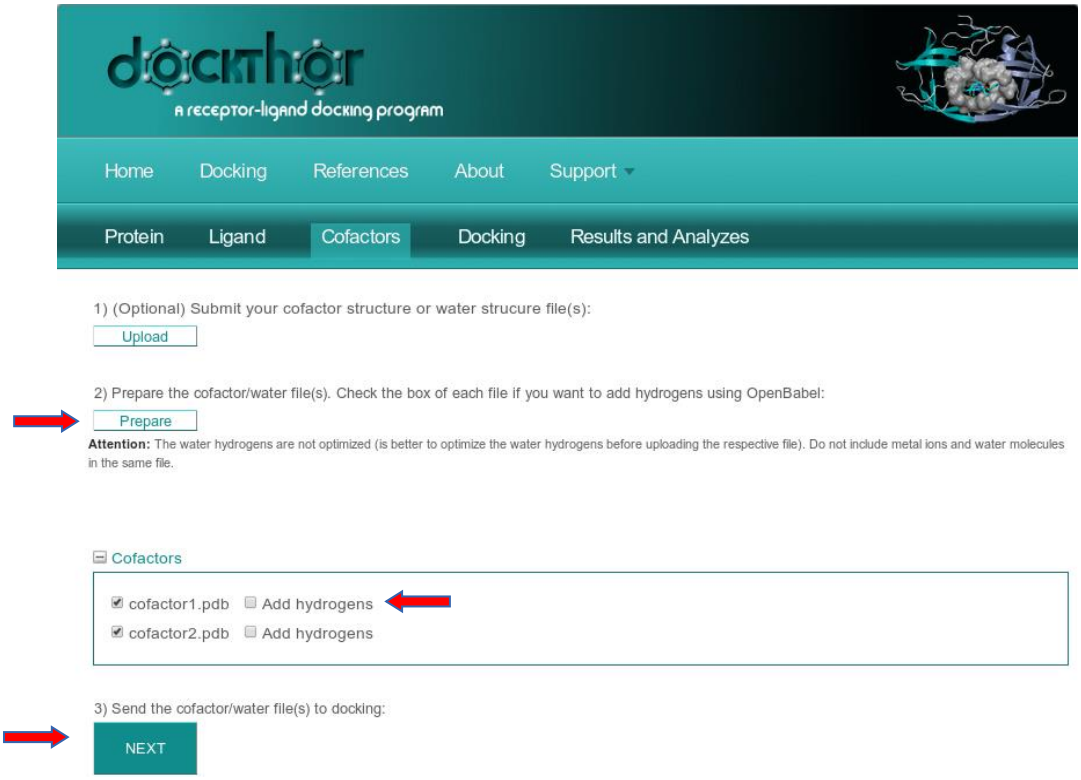
6. Click **Next** to send the most recent ligand file to docking.

2.3 Cofactor and Water Preparation

1. For some protein-ligand complexes it is important to consider cofactor (*e.g.* NAD, ATP, FAD, Mg, Zn etc.) and/or water molecules. The DockThor Portal allows one or more cofactor and water molecules to be included; they are kept fixed during the docking simulation. Upload each file at a time clicking on **Upload**.



2. As well as ligand, the cofactor and water files (one or more water molecules per file) need to be converted to the topology file. Check the respective boxes to add hydrogen atoms[3] and click **Prepare**.
3. It is also possible to download the prepared files correspondent to each cofactor/water clicking on **Download prepared files**. Each cofactor/water file will generate one link for download. In the case of our test file, all the five cofactors (two zinc and three calcium atoms) are into the same file.
4. Send the cofactor and/or water to docking clicking **Next.**



2.4 Docking Configuration

1. The next tab consists of the main options to run the docking job. It is possible to see the original files for the ligand and the protein in **Files uploaded**. Check if are correct.

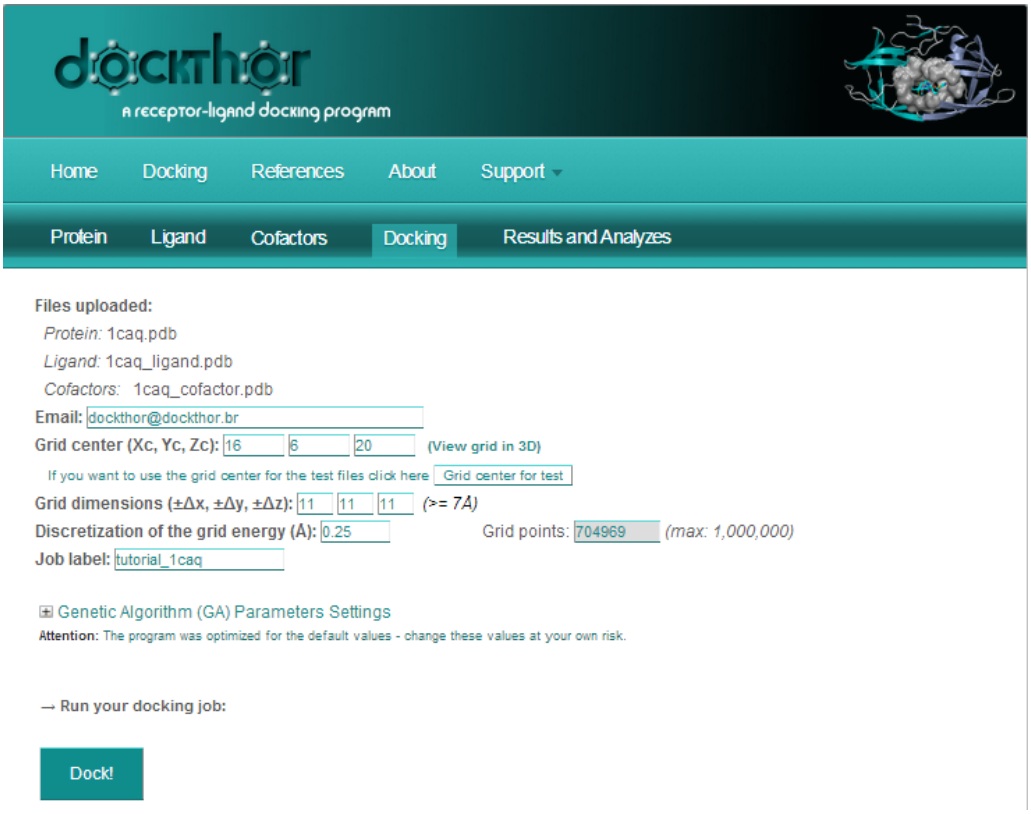
2. Fill the **Email** field to receipt the link for the results page when the docking job is finished. This field is mandatory.

[3] Since MMFFLigand do not optimize the hydrogen atoms position, it is recommended to do it previously and do not chose to include hydrogen atoms to these water files.

3. Insert the center coordinates of the grid energy in **Grid center** (Xc, Yc, Zc).

4. Insert the **Grid Dimensions** (±Δx, ±Δy, ±Δz). These values correspond to the half of the grid size in each dimension, *e.g.* (Xc-Δx) ≤ X Dimension ≤ (Xc+Δx)

5. Select the spatial discretization of the energy grid. This value corresponds to the spacing between the points of the grid (the default value is 0.25Å). Check if the number of grid points do not exceed the limit allowed (1,000,000) in the **Grid points** box.

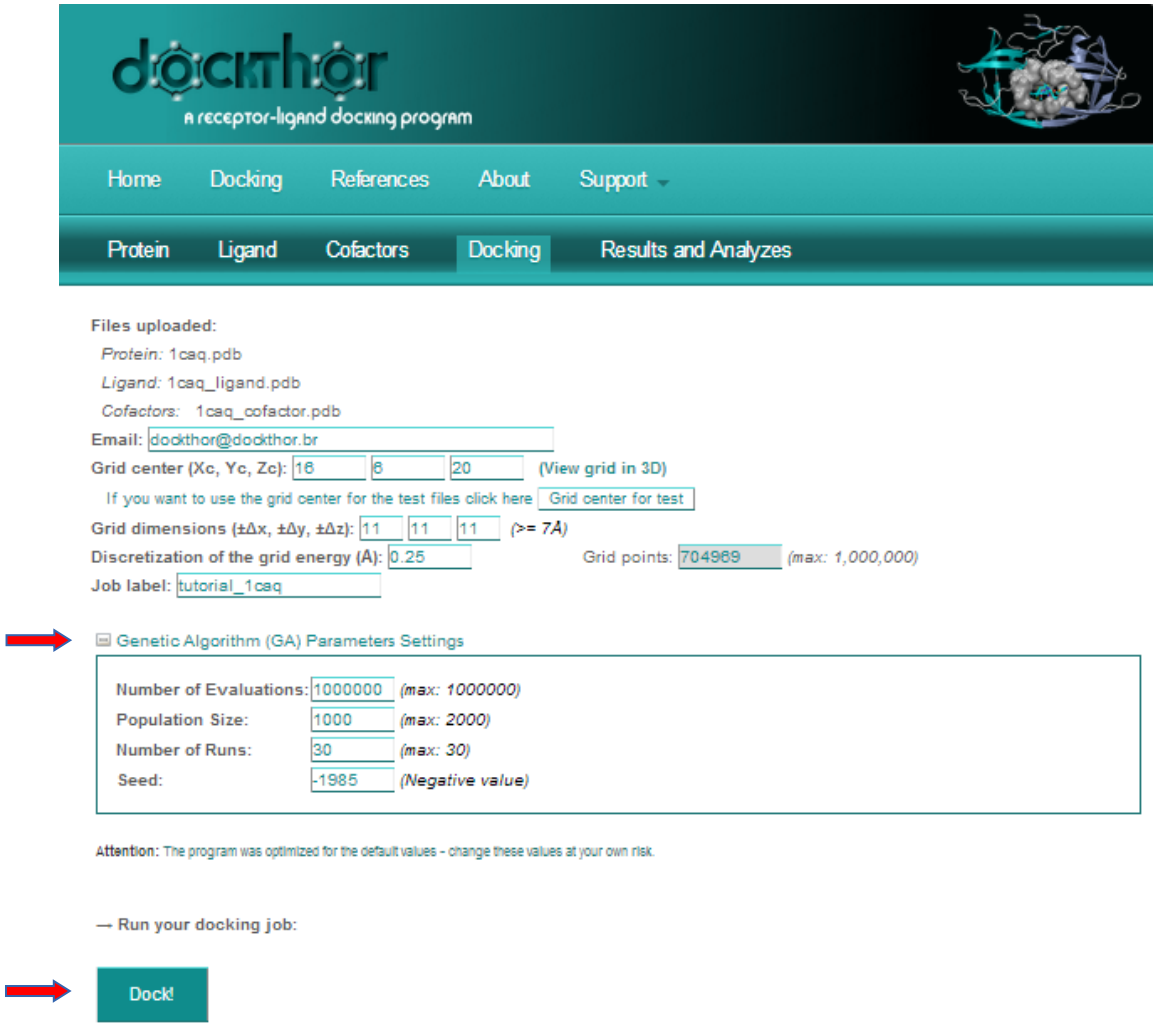


6. To view the grid dimensions and the receptor click (View grid in 3D).
7. Choose a label to facilitate the identification of your docking job.
8. Some genetic algorithm (GA) parameters can be modified.

- The GA multisolution algorithm was optimized to deal with highly flexible ligands and we do not recommend changing the standard number of evaluations per run (*i.e.* 1,000,000) or the GA population size (*i.e.* 1,000). Change these values at your own risk.

- Each submission job corresponds to 30 independent docking runs. A maximum of 50 docking runs per job is permitted. It is also possible to run more

independent docking runs submitting more than one job and changing the initial seed. The seed must to have a negative value and for each run its value is decreased by one. You should take this into account if you want to submit more than one job in order to obtain more than 50 independent docking runs.



9. Run your docking job clicking on **Dock!**
10. When your docking is finished an email will be sent with a link to the results page.

3. Analyzing the Docking Job

Once the docking job is finished, the user receives an email with the link to the Results page as below:

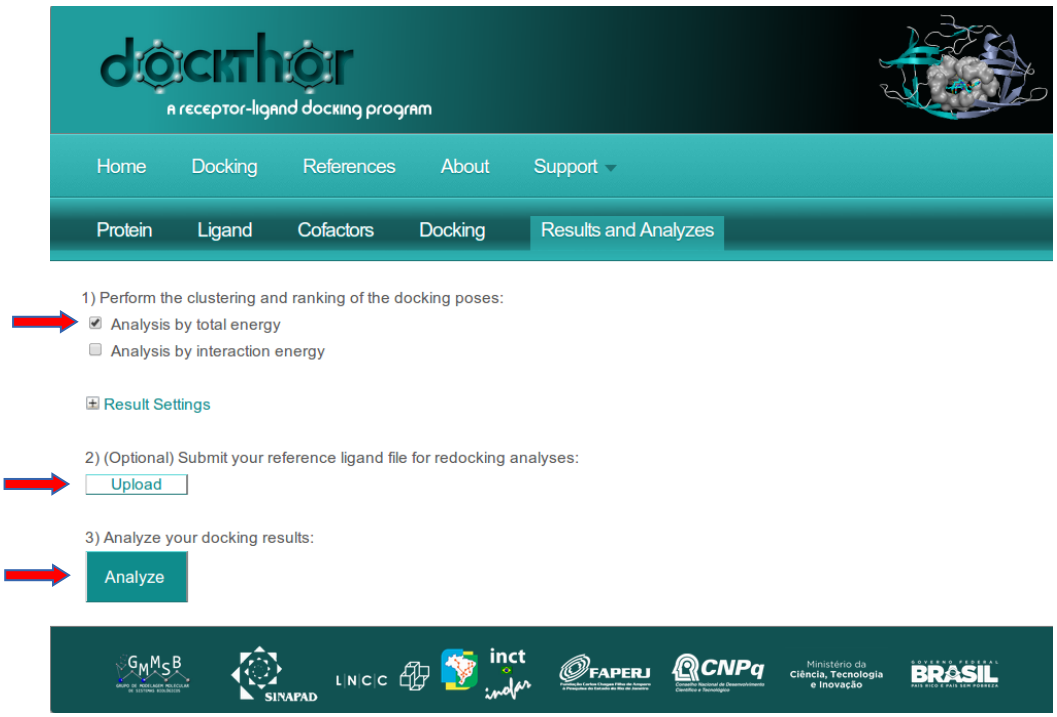
Your docking result is available at
http://www.dockthor.lncc.br/index.php?pg=submission&pgs=results&id=docking_gmmsb_1caq_ligand_9l84mb

Follow the link in the email to go to the Results and Analyses page.

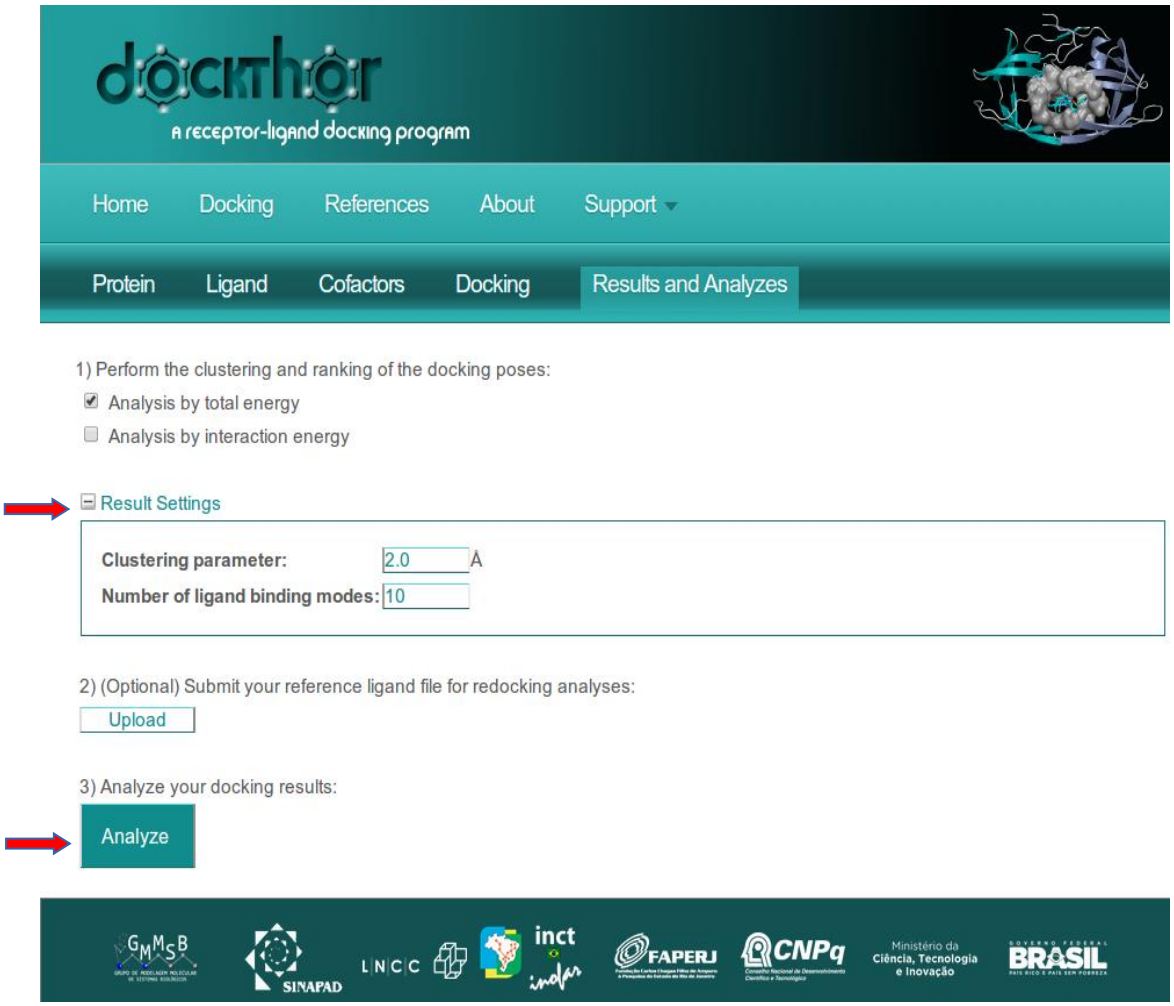
3.1 Results and Analyses

For each docking run the final population is clustered using the total energy and a RMSD = 1.0Å criteria. Only the leaders (top energy solution) of each cluster will be used to the final clustering analysis step:

1. It is possible to cluster and order the docking solutions according to two criterions: **Total Energy** (intermolecular ligand-receptor + intramolecular ligand energies) or **Interaction Energy** (only intermolecular ligand-receptor energy). The default is **Analysis by total energy**.

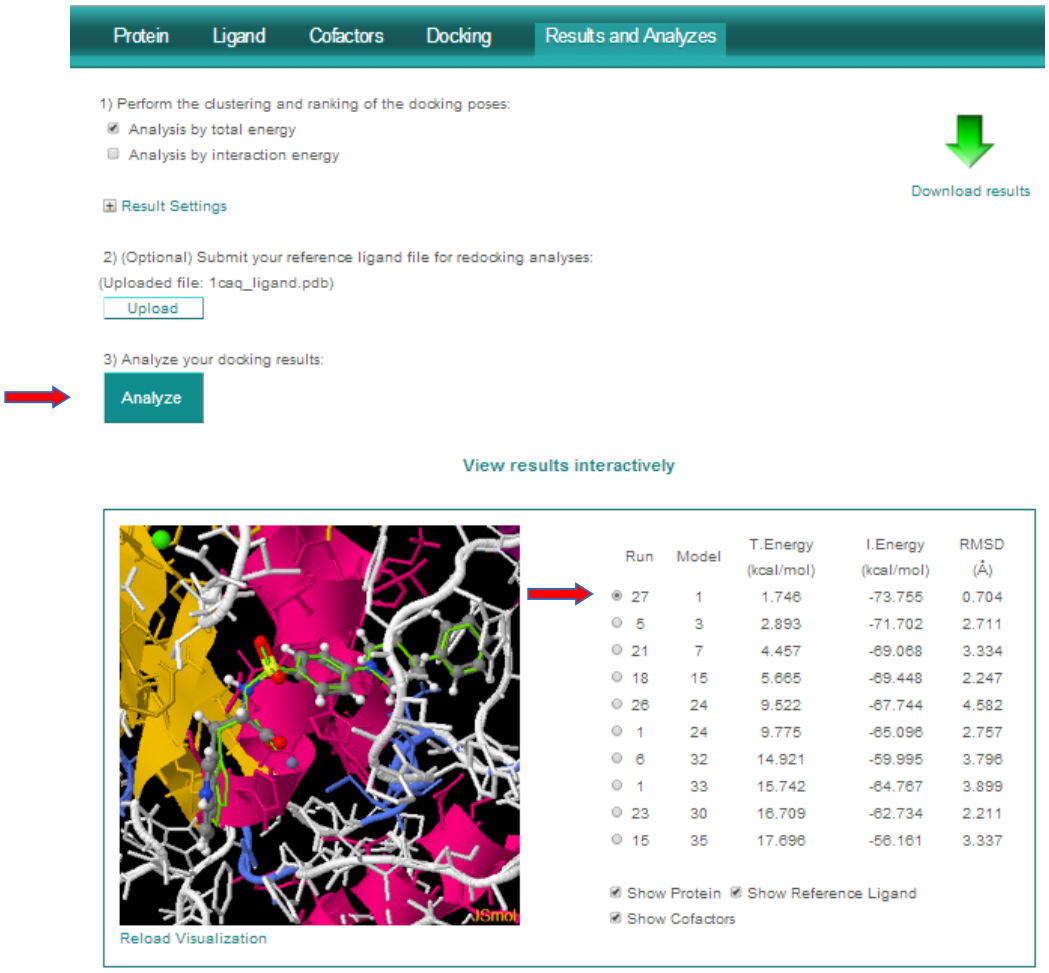


2. Click **Analyze** to perform the docking poses analyses.
3. You can change the clustering criterion (default = 2.0Å) and the number of the best ligand binding modes (default = 20) selected to analyses and visualization. Click on the ***Result Settings*** button.

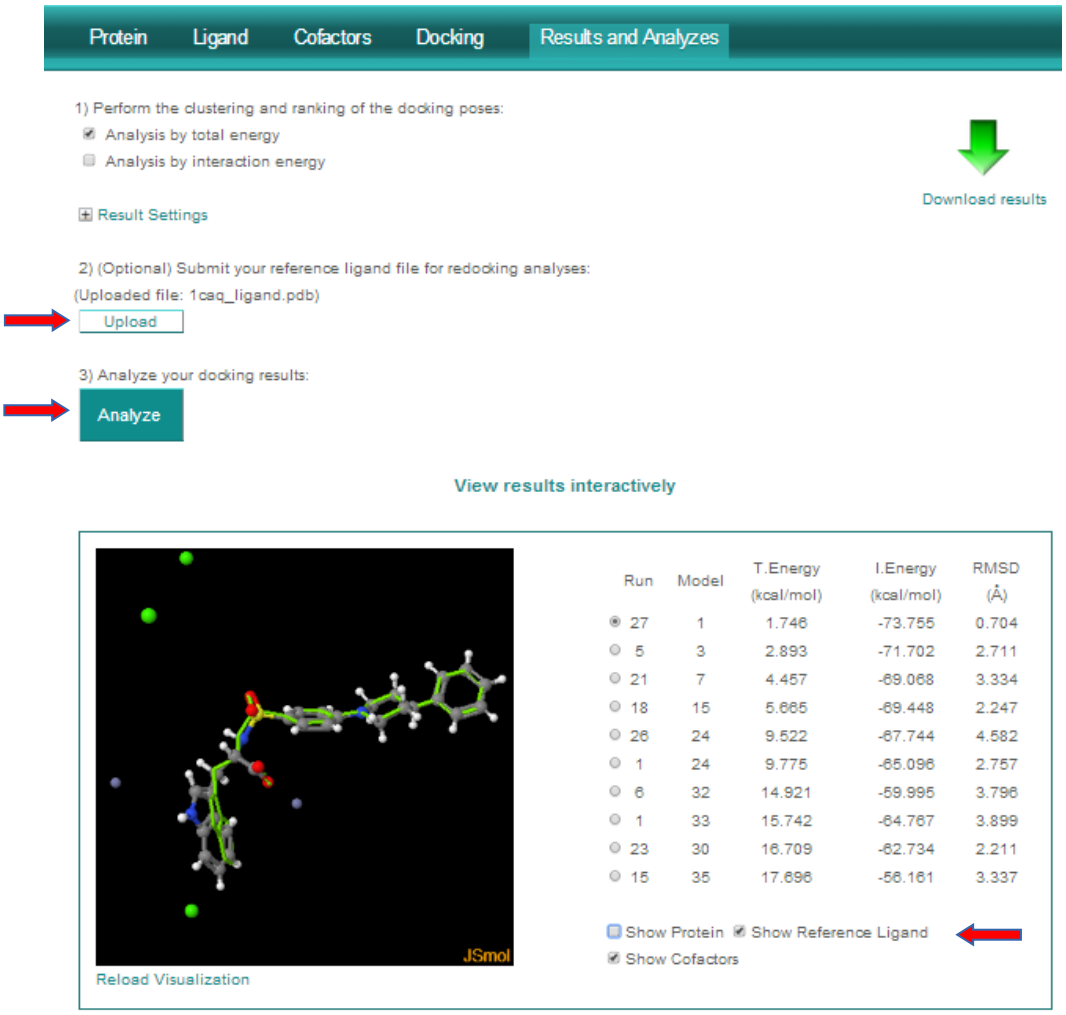


4. It is possible to view the results interactively in the website. Click on **View results interactively**.
5. To see each docking ligand binding mode (*i.e.* cluster leader) select the corresponding one on the results table. For each solution is shown the corresponding **Run** of the Genetic Algorithm, the **Model** (number of the cluster leader) in the corresponding GA run, the **Total Energy**, the **Intermolecular Energy** and the **RMSD** (root mean square deviation calculated using the non H atoms) relative to the top ranked pose. This value gives an idea of the conformational difference between the alternative ligand binding modes.

Attention: If you have submitted a reference file, the last column refers to the RMSD between each pose and the reference conformation.



6. It is also possible to upload a reference ligand conformation (.pdb type file) to perform redocking analyses or to help the investigation of the distinct binding modes. After this click **Analyze** to perform the docking poses analyses and click on **View results interactively**. The **RMSD** (in the last column) is now calculated relative to the uploaded reference ligand pose. It is possible to hide the protein to facilitate the visualization. If necessary, click on Reload Visualization to upload the JSMol scene.



7. To download the docking results, including the clustered conformations, click **Download Results**.

dockthor.out – general information about each DockThor run.
parameters.txt – general information about the docking parameters used.
results.out – total number of clusters obtained (after analyzing all runs).
out.log – summary of the clustering analysis for the best cluster leaders (ranking, energies, RMSD).
out.mol2 – contains the atomic coordinates of the best cluster leaders (multimodel .mol2 type file, ordered according to the results described in out.log).
ligand_run_X.log – contains the information of the cluster leaders obtained in run X (using a RMSD criterion of 1.0Å).
ligand_run_X.pdb – contains the atomic coordinates of the cluster leaders obtained in run X (multimodel .pdb type file).
protein.in – DockThor receptor input file.
ligand.top – DockThor ligand input file.

8. It is always possible to perform other analyses and download the respective files.

4. Softwares

The Portal DockThor uses the following programs:

***MMFFLigand*:** generates the topology file for the ligand and cofactor files through MMFF94S force field and OpenBabel tools;
***PdbThorBox*:** prepares the protein file (adds hydrogen atoms, changes amino acid protonation states, completes missing side chains) with the MMFF94 force field;
***DockThor*:** the docking program is a flexible ligand rigid receptor grid based method that employs a multiple solutions genetic algorithm as the search method and the MMFF94 force field as the scoring function.
Dtstatistic: clusters and ranks the docking poses according to total or interaction energy.

** All these programs were developed by the GMMSB/LNCC group.*

Appendix A

Residue Protonation States

ASP - Negatively charged aspartic acid **(default)**.
ASPN1 - Neutral aspartic acid with a H bonded to the Oδ1
ASPN2 - Neutral aspartic acid with a H bonded to the Oδ2

GLU - Negatively charged glutamic acid **(default)**.
GLUN1 - Neutral glutamic acid with a H bonded to the Oε1.
GLUN2 - Neutral glutamic acid with a H bonded to the Oε2.

CYSH - Neutral cysteine with a H bonded to S **(default)**.
CYS - Negatively charged cysteine.
CYSS - Neutral cysteine (disulfide bond).

HIS - Neutral histidine with a H bonded to Nτ **(default)**
HISD - Neutral histidine with a H bonded to Nπ.
HISP - Positively charged histidine.

ARG - Positively charged arginine **(default)**.
ARGN1 - Neutral arginine at $N\omega_1$.
ARGN2 - Neutral arginine at $N\omega_2$.

LYS - Positively charged lysine **(default)**.
LYSN - Neutral lysine.