Profa. Rafaela Salgado Ferreira
Departamento de Bioquímica e Imunologia - UFMG

Instruções para Tutorial DockThor

O docking (de um ligante da protease do HIV) será realizada através do webserver do programa DockThor:
http://dockthor.lncc.br/

Será realizado o procedimento de docking nativo, no qual emprega-se um ligante que já co-cristalizado com a proteína de interesse e considera-se a conformação proteica observada na presença deste ligante (obtida diretamente do PDB referente ao complexo em questão).

Preparação inicial de arquivos:

Utilizaremos o tutorial preparado pela própria equipe do DockThor. No entanto, antes de iniciar o tutorial é necessário preparar os arquivos da proteína e do ligante a ser avaliado. É necessário também definir o ponto central do Grid no qual será realizado o procedimento de docking.

1. Para preparar o arquivo da proteína:

- Faça o download do arquivo 1HBV.pdb diretamente no site do PDB:

http://www.pdb.org/pdb/explore/explore.do?structureId=1hbv

Estrutura 1HBV - A CHECK ON RATIONAL DRUG DESIGN. CRYSTAL STRUCTURE OF A COMPLEX OF HIV-1 PROTEASE WITH A NOVEL GAMMA-TURN MIMETIC

Não é necessária nenhuma preparação adicional. O receptor será preparado automaticamente no site do programa.

2. Para preparar arquivo do ligante:

Num editor de texto, abra o arquivo 1HBV.pdb e copie todas as linhas contendo informações sobre o ligante. Estas são identificadas pela coluna HETATM e pelo código do ligante GAN.

```
HETATM 1523  N1  GAN A 600      -4.428  16.771  18.898  1.00 49.56           N
HETATM 1524  C2  GAN A 600      -4.368  17.218  20.328  1.00 48.46           C
HETATM 1525  C3  GAN A 600      -3.362  18.372  20.475  1.00 51.15           C
                                   ...
HETATM 1564  C42 GAN A 600     -12.454  13.101  33.796  1.00 39.47           C
```

- Salve estas linhas em um novo arquivo: GAN.pdb.

3. Definição do ponto central do GRID:

Para definição do ponto central do GRID, selecione um átomo central no sítio ativo. Para visualizar o sítio ativo, utilize o programa Pymol e crie um objeto contendo o ligante GAN e todos os resíduos a uma distância de até 6 Å ao redor.

Considerando a posição central dos Asp catalíticos no sítio ativo, neste caso selecionei o oxigênio OD1 do Asp25 para definição do ponto central do grid. A partir da linha correspondente a este átomo no arquivo PDB, obtem-se as coordenadas para o GRID (destacadas em negrito abaixo).

ATOM 198 OD1 ASP A 25 **-11.544 21.448 28.812** 1.00 10.03 O

Estas coordenadas serão adicionadas no site do DockThor, no momento de submissão do docking.

Após a preparação destes arquivos, siga o tutorial a seguir, que foi retirado diretamente do site do DockThor.

Observações importantes:

- Na preparação da proteína, altere o estado de protonação de Asp25 para AspN1 (conforme indicado no próprio tutorial). O pka das cadeias laterais dos resíduos na estrutura 1HBV foram calculados utilizando o programa propka3.0. Conforme indicado abaixo, observa-se uma variação significativa no pKa de Asp25, quando comparado a Asp29:

RESIDUE	pKa	BURIED
ASP 25 A	5.91*	100 %
ASP 29 A	3.77	51 %

Essa variação é decorrente da proximidade espacial entre dois resíduos de aspartato.

- Nenhum cofator será utilizado.

Portal

www.dockthor.lncc.br

User Guide

Version 1.0

Contributors

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

GRUPO DE MODELAGEM MOLECULAR DE SISTEMAS BIOLÓGICOS

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 along the MMFF94S molecular force field scoring function. The main steps of the ligand and protein set up are available on the DockThor Portal, being possible to change the amino acid residues protonation states 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 can be analyzed and sorted automatically. The analysis parameters 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 pdb type files can be accepted (or pre-prepared DockThor input files *.in). Other types of receptors (*e.g.* DNA, RNA, another ligand) will be allowed in the next version.

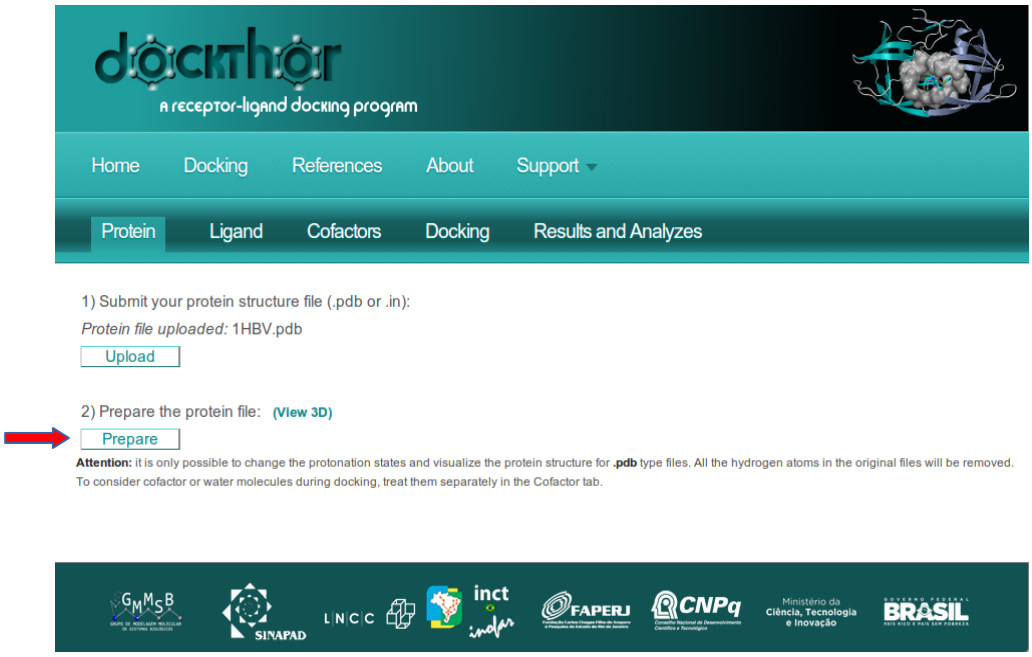
2.1 Protein Preparation

1. Click on the **Docking** tab. Then 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.

3. Prepare the protein file applying the basic options clicking on **Prepare**. At this step, all the missing atoms of the residue side chains will be reconstructed. The protein atoms are recognized by the initial .pdb label 'ATOM', Atoms associated with the initial .pdb label 'HETATM' will be ignored. All the atoms are also recognized by their .pdb atom label (*e.g.* CA, CB etc). If the atom label nomenclature is not right the atom will be reconstructed. For this reason it is very likely that all the hydrogen atom will be rebuilt[1]. If two side chain conformations are given 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 must use the same H atom .pdb nomenclature used by the DockThor program (please 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). Reprepare your protein file by applying the new protonation states clicking on **Reprepare**.

5. View the prepared protein file with JSmol clicking on **View 3D**.

6. To download the respective protein files click on **Download prepared files**.

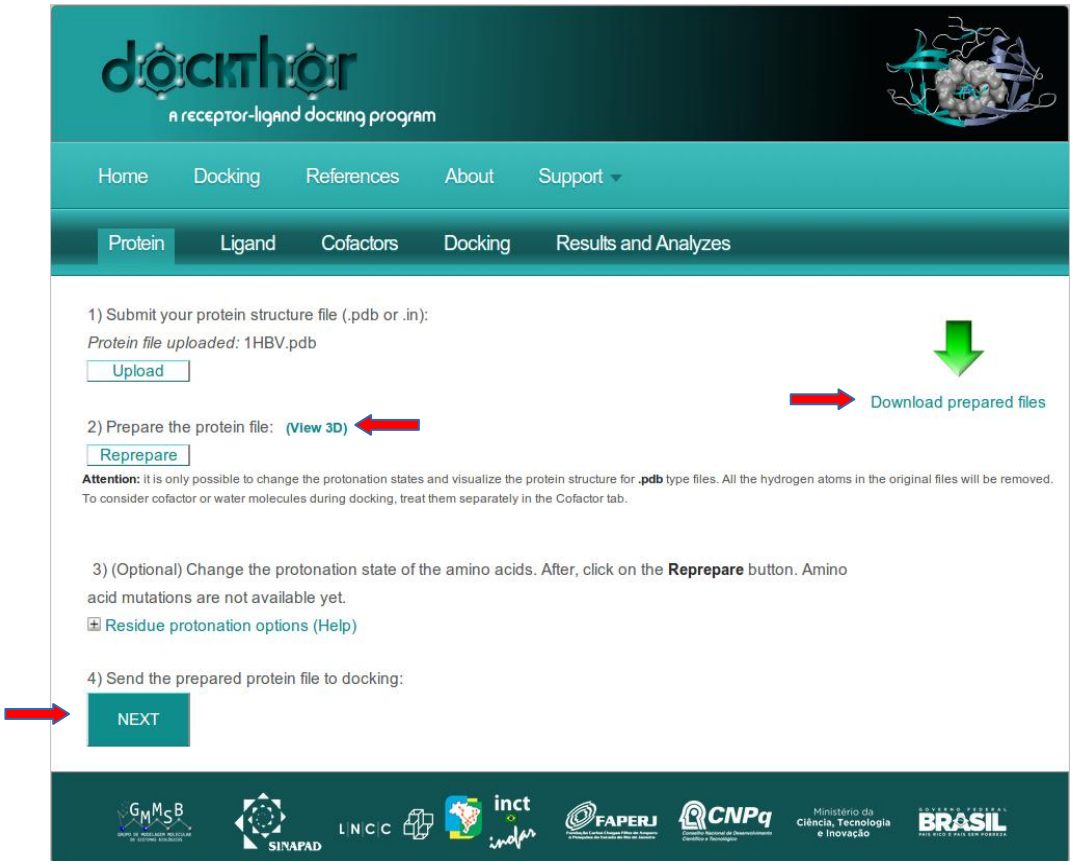
   protein.in – DockThor receptor input file (contains the MMFF94S atomic type number, the atomic coordinates, the bond connectivities, the atomic partial charges, the pdb type atom, residue and chain labels and the residue and chain numbers).

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

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

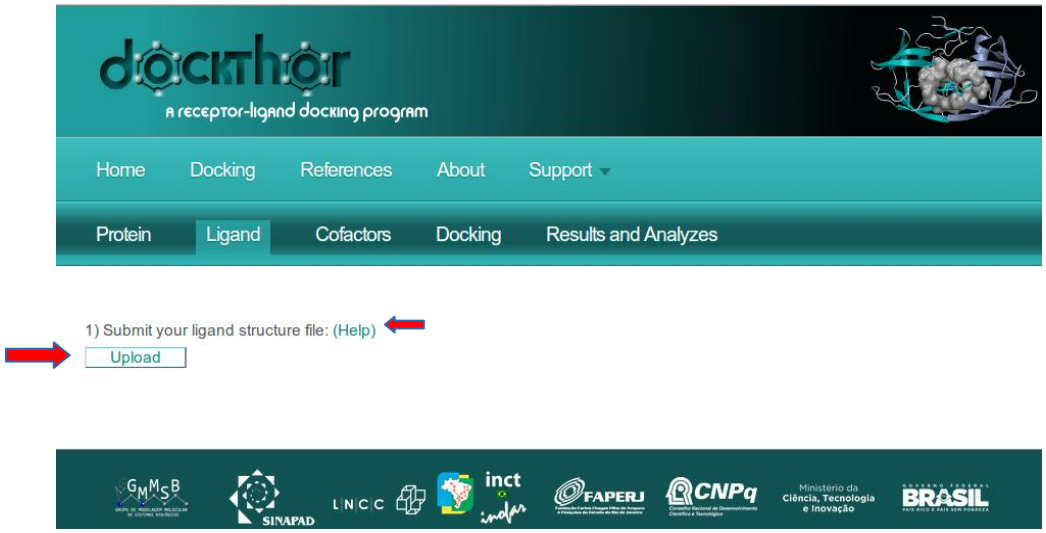
   protein.X - configuration file for each protein chain X. Contains the amino acid residue list and the respective protonation state label.

7. Click **Next** to submit your prepared protein file to docking. It is noteworthy that only the protein file in the *.in* format is necessary to docking.



2.2 Ligand Preparation

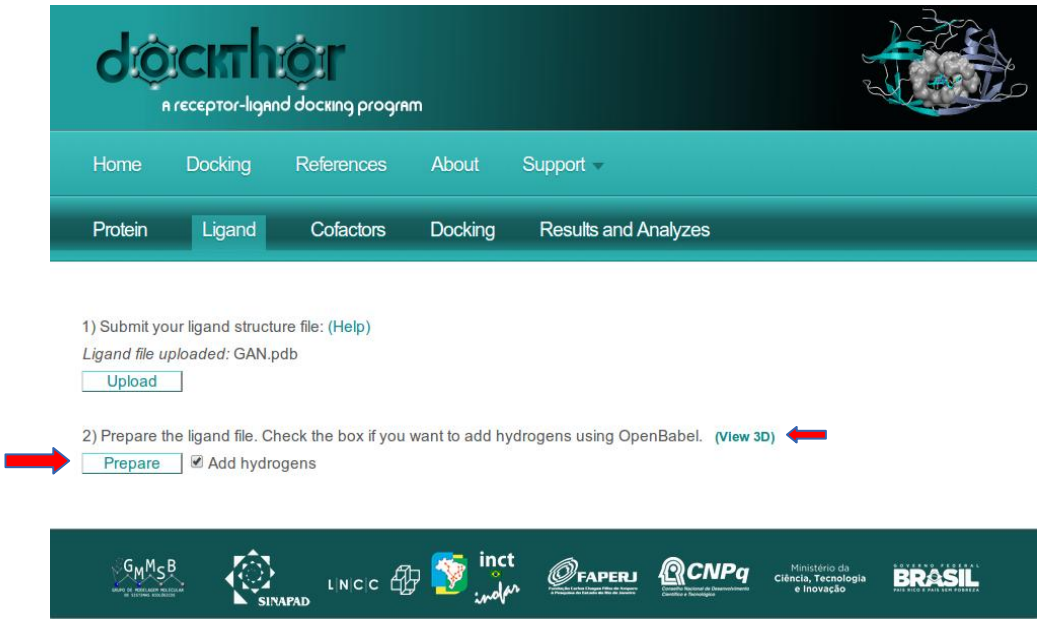
1. To submit a single small molecule[2] to docking, click **Upload**. The formats accepted are those recognized by OpenBabel. For a complete list of these files, click on the (Help) link.



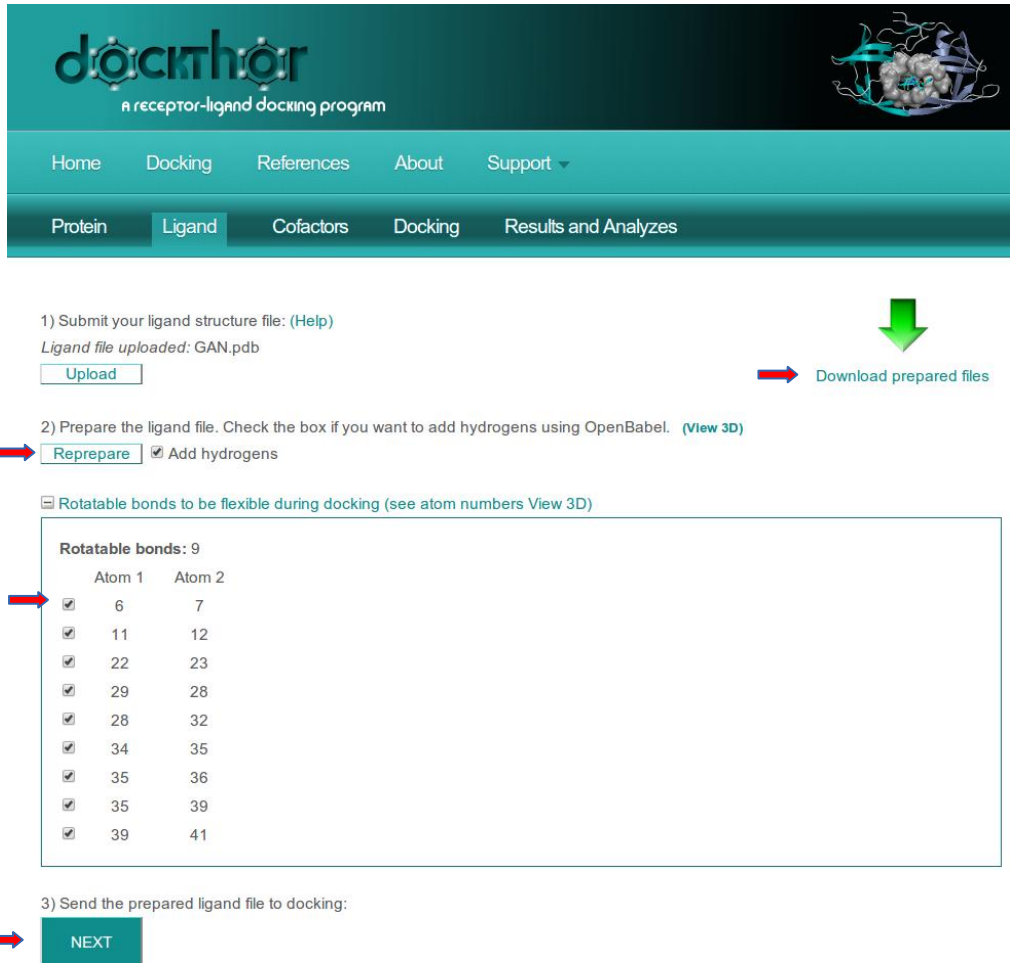
2. The uploaded ligand needs to be prepared to generate the respective topology file (.top) to docking. This step comprises the right MMFF94S force field atom type assignment, the atomic partial charges calculation and the assignment of the rotatable chemical bonds. If you want to add hydrogen atoms automatically, just check the **Add hydrogens box** and **Reprepare** the molecule again. The hydrogen atoms will be added through OpenBabel tool at pH = 7.0.

3. To view the current structure of your ligand, click **View 3D**.

[2] In the present portal version, it is only possible to dock 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 unselect some flexible rotatable bonds click on **Reprepare**.



5. If you click on **Reprepare** without unselecting any rotatable flexible bond, the original set of flexible bonds will be restored.

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

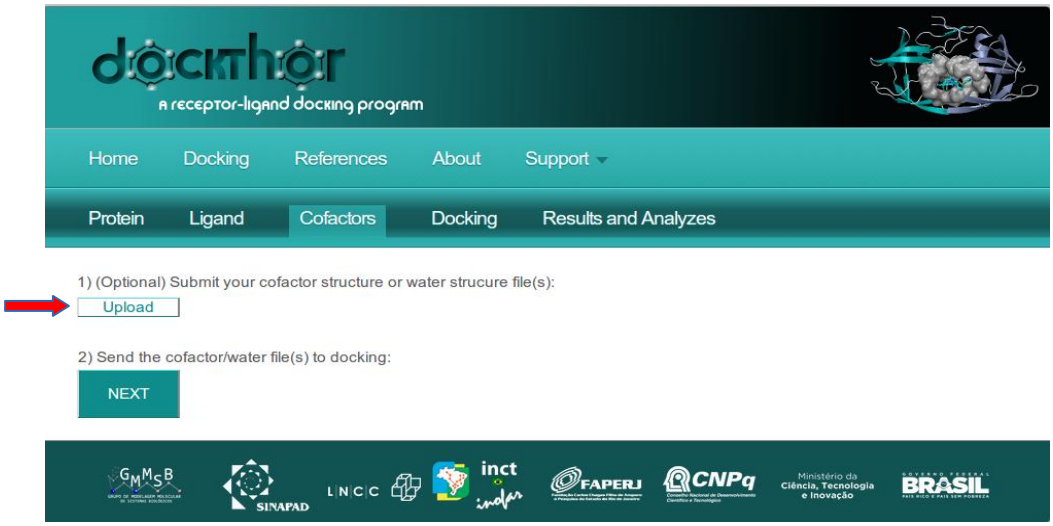
   ligand.top – DockThor ligand input file. It contains the atom name, the atom number, the MMFF94S atomic type number, the atomic partial charge, the atomic coordinates and the atom valence. It contains also: (i) the atom connectivity; (ii) force field torsional parameters; (iii) selected flexible bonds;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

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

7. Click **Next** to send the updated ligand file to docking.

2.3 Cofactor and Water Preparation

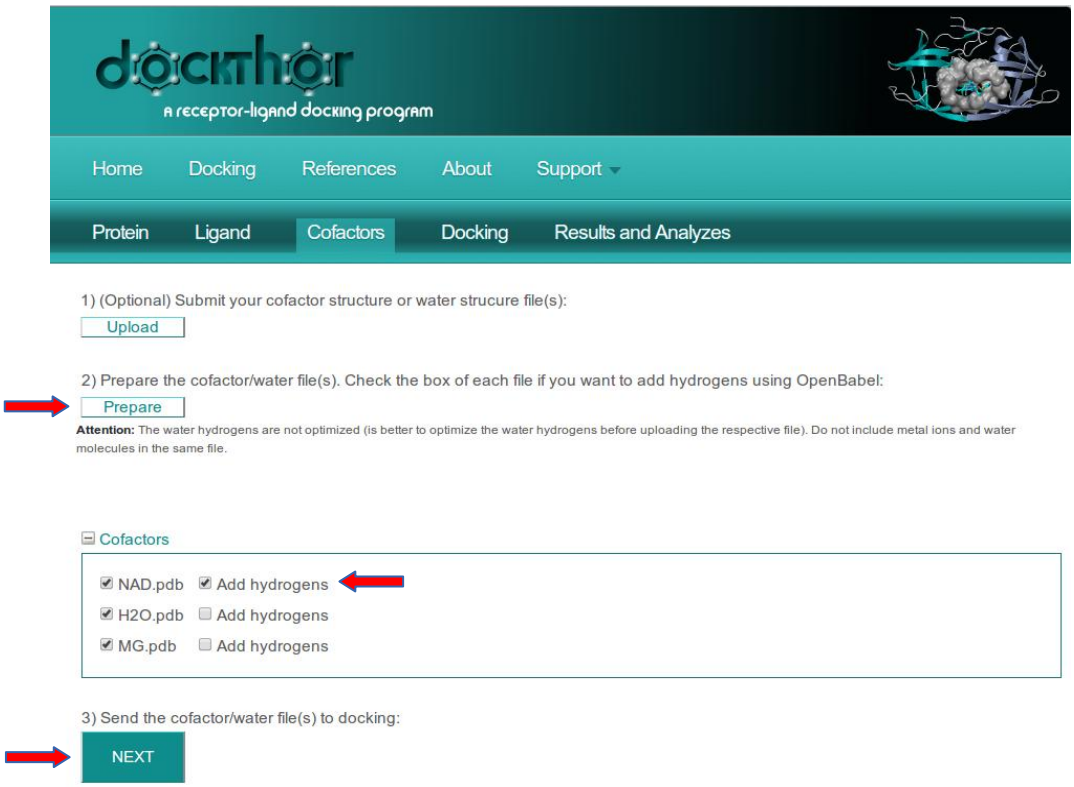
1. For some protein-ligand complexes it is important to consider cofactor (*e.g.* NAD, ATP, FAD, Mg, Zn etc.) and/or water/solvent molecules. The DockThor Portal allows the inclusion of one or more cofactor and solvent molecules; they are kept fixed during the docking simulation. Upload each file at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



2. As well as ligand file, the cofactor and water files (one or more water molecules per file) need to be converted to the topology file. To add hydrogen atoms[3], check the respective boxes and click **Prepare**.

3. It is also possible to download the prepared files corresponding to each cofactor/water clicking on **Download prepared files**. Each cofactor/water file will generate one link for download.

4. Send the cofactor and/or water to docking by clicking on **Next.**



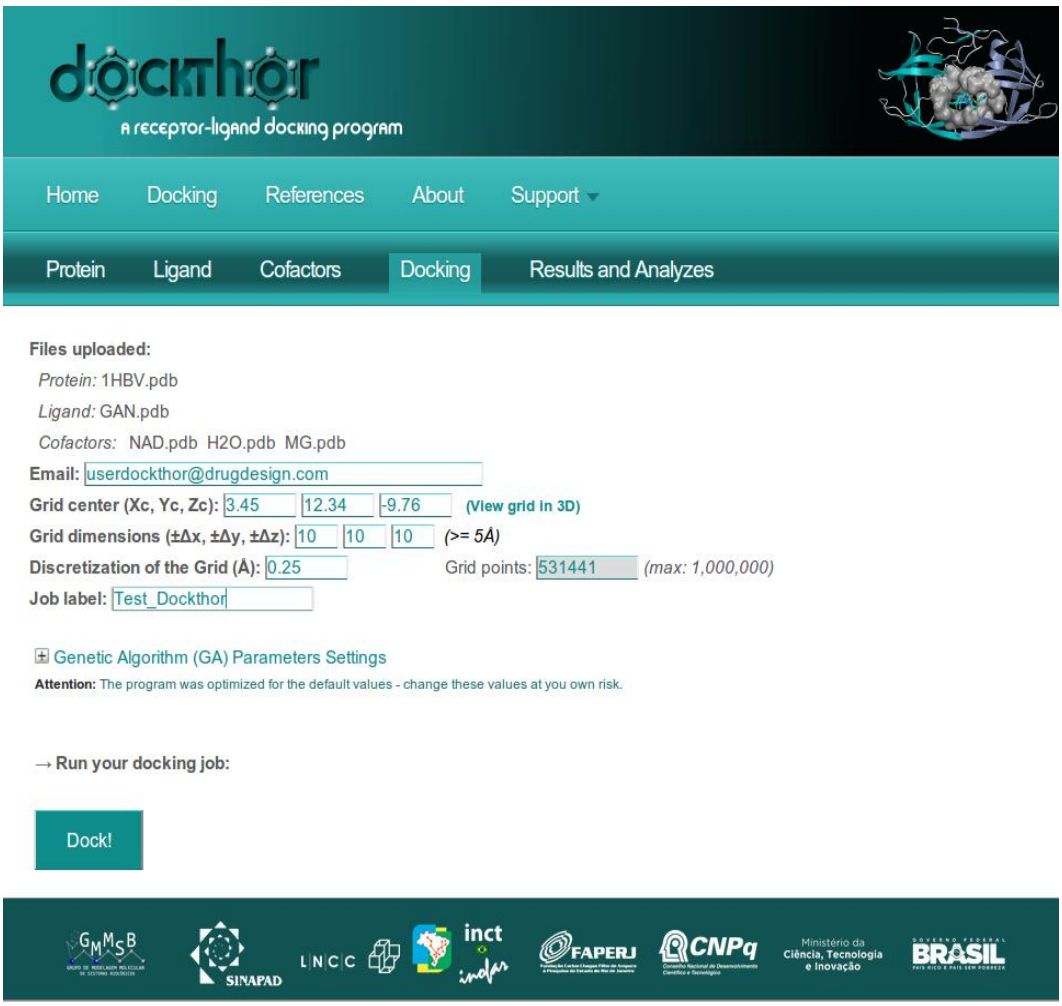
2.4 Docking Configuration

1. The next tab displays the main parameters to run the docking job. It is possible to see the original files for the ligand and protein in **Files uploaded**. Check if they are all right.

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

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

3. Fill the coordinates of the energy grid center in **Grid center** (Xc, Yc, Zc).
4. Fill the **Grid Dimensions** (±Δx, ±Δy, ±Δz). These values correspond to half of the grid size on 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 grid points (the default value is 0.25Å). Check if the number of grid points does 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 an identification label for your docking job.
8. Some genetic algorithm (GA) parameters can be modified.

- The GA multisolution algorithm was optimized to deal with highly flexible ligands. 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 possible to run more independent docking runs submitting more than one job and changing the initial seed. The seed must have a negative value, and for each successive run its value is diminished 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 by clicking on **Dock!**

10. When your docking process is finished an email will be sent at the previously stated e-mail, containing a link to the results page.

3. Analyzing the Docking Job

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

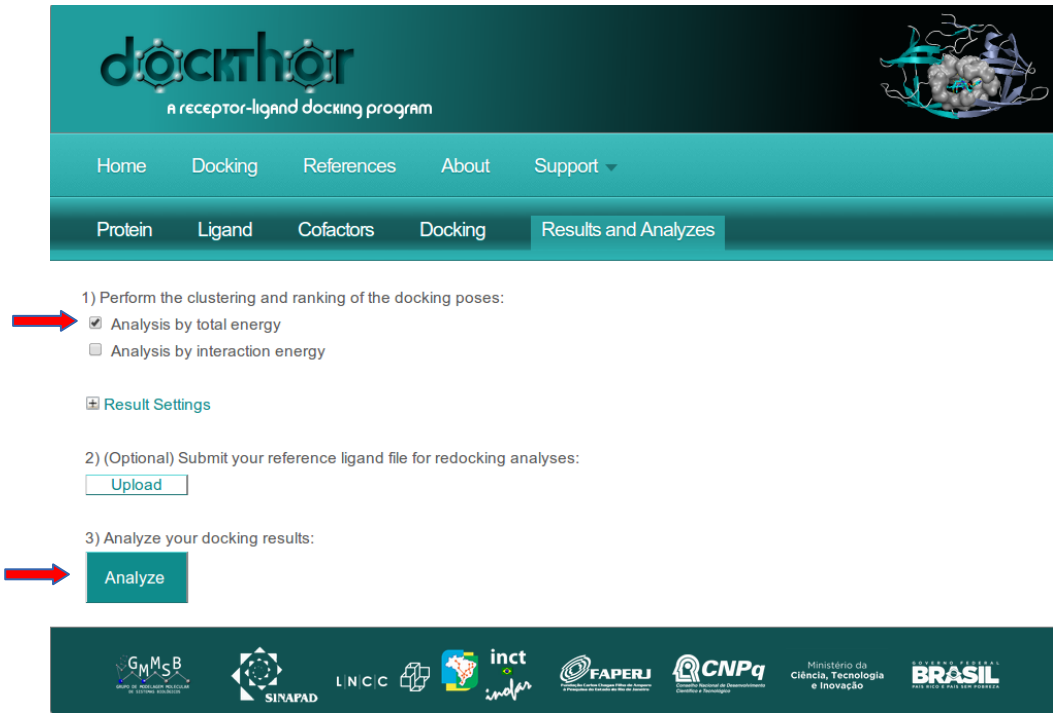
Your docking result is available at
http://www.dockthor.lncc.br/index.phppg=*submission&pgs=results&id=Test_Dockthor_r1rpaa*

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

3.1 Results and Analyses

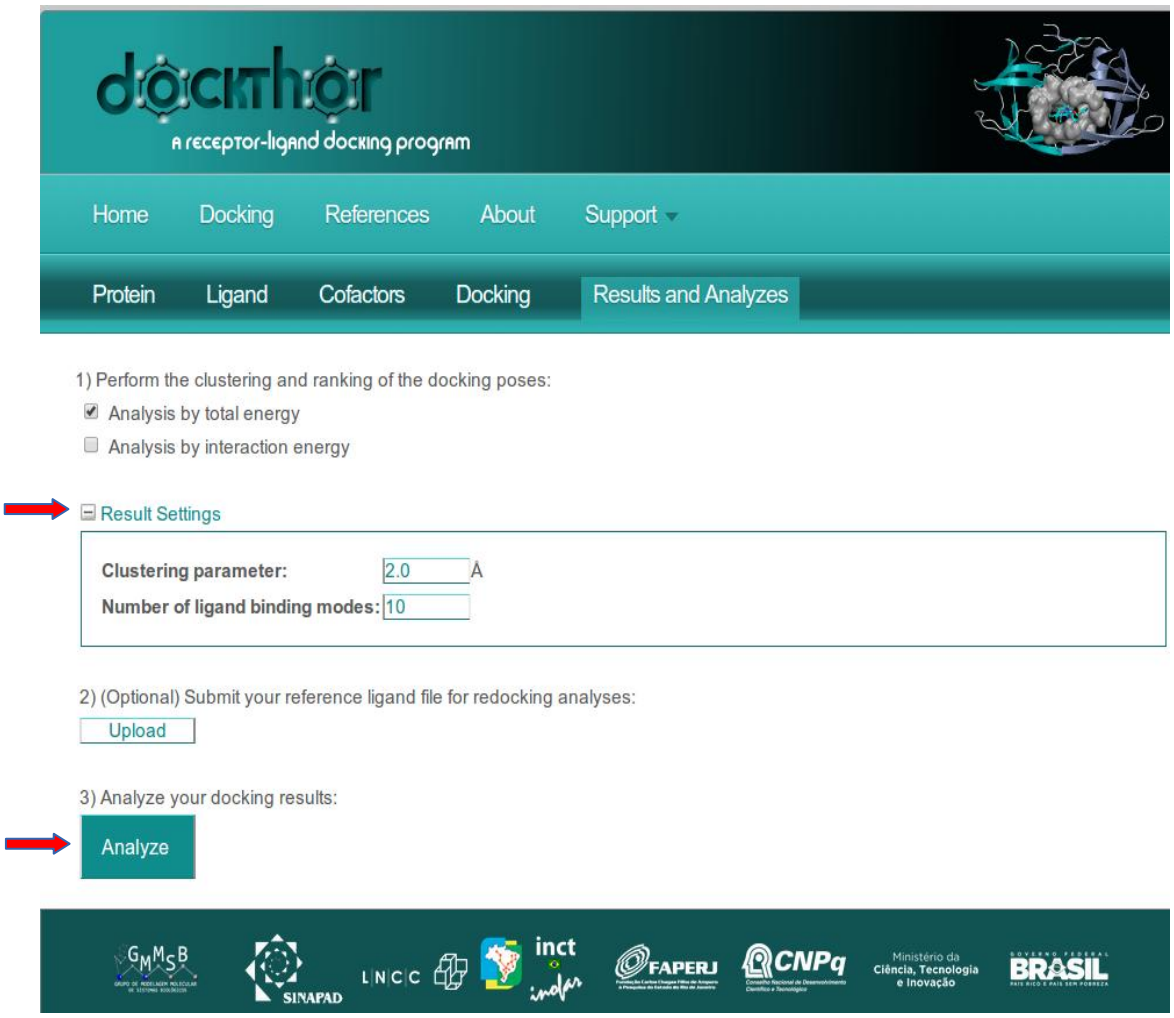
For each docking run the final population is clustered using the total energy and a RMSD = 1.0Å criterion. Only the leaders of each cluster will be used in the final clustering analysis step:

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



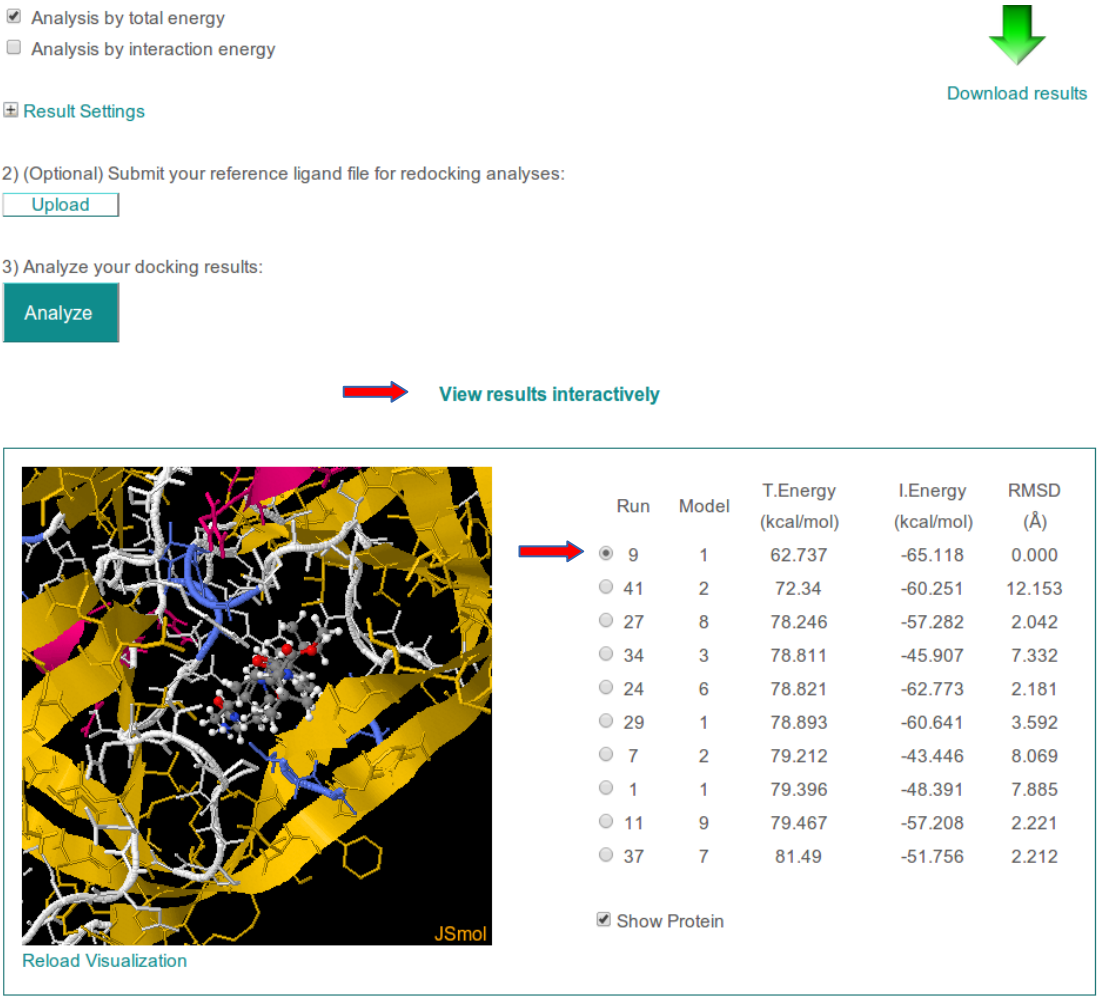
2. Click **Analyze** to perform the docking poses analyses.

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

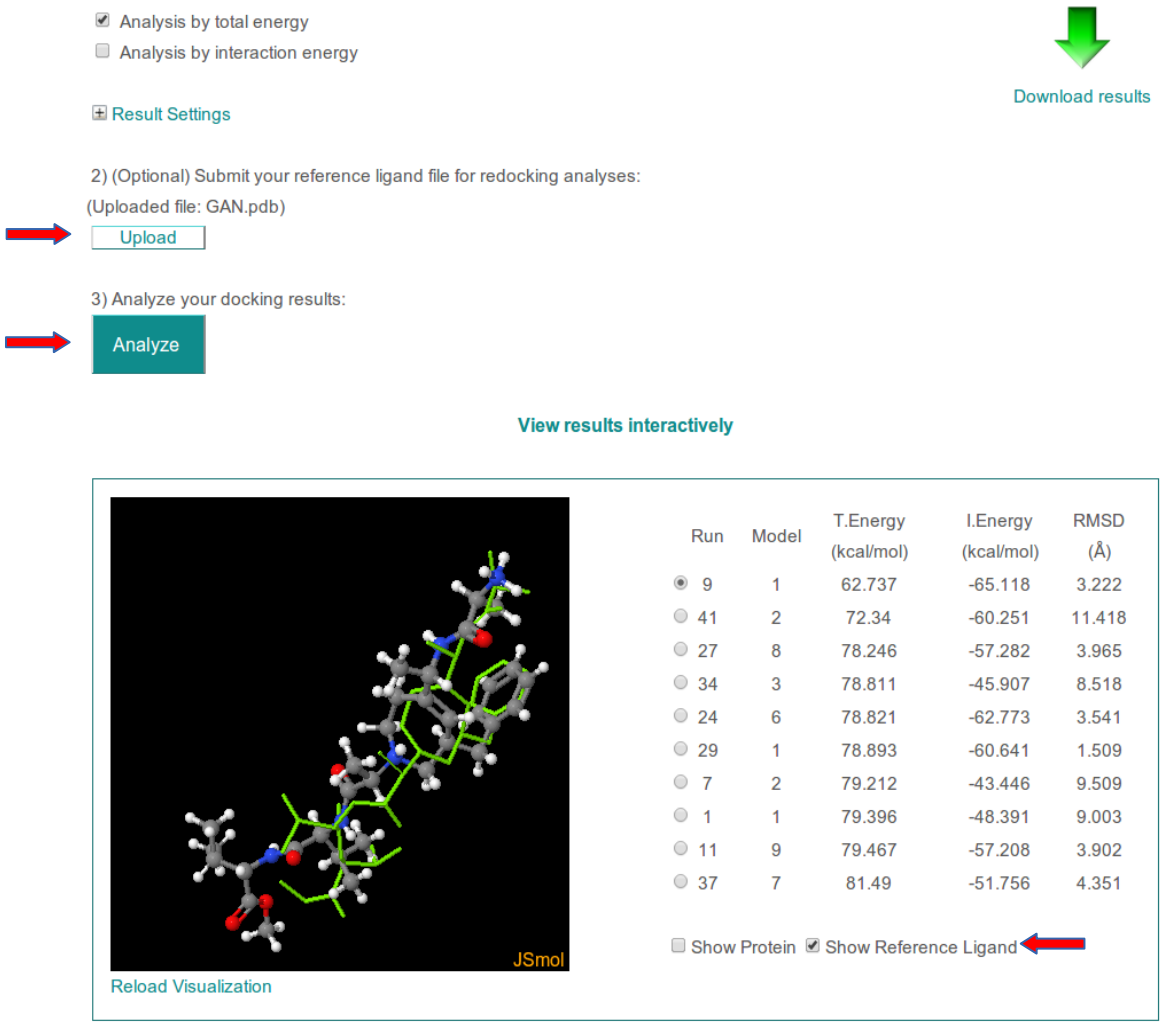


4. It is possible to view the results interactively on the website. Click on **View results interactively**.

5. To see each docking ligand binding mode (*i.e.* cluster leader) select the corresponding one in the results table. For each solution the system shows the corresponding **Run** of the Genetic Algorithm, the **Model** (number of the cluster leader) of 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 among the alternative ligand binding modes.



6. It is also possible to upload a reference ligand conformation (.pdb type file) to perform redocking analyses or help the investigation of the distinct binding modes. Then 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 structure to facilitate visualization.



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, sorted out 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 MMFF94S 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 using the MMFF94S force field as the scoring function.

Dtstatistic: clusters and ranks the docking poses according to total or interaction energies.

** All these programs were developed by the GMMSB/LNCC group.*
*** Ernesto R. Caffarena, Michel Loos and Isabelle Ortmans also contributed for the PDBTHORBOX development.*

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 $N\omega_2$.

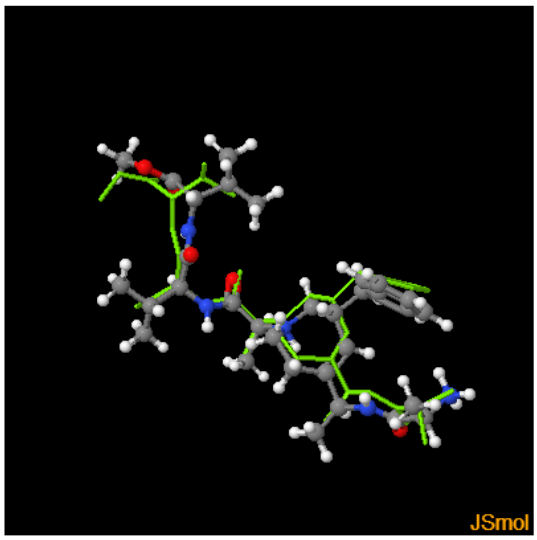
LYS - Positively charged lysine **(default)**.
LYSN - Neutral lysine.

Análise dos resultados:

No site do DockThor é possível visualizar os resultados, através de link enviado por e-mail. Resultados de uma submissão prévia podem ser acessados pelo link:

http://www.dockthor.lncc.br/index.php?pg=submission&pgs=results&id=docking_GAN_1HBV_171016_ligand_iwgvow

É possível comparar os resultados com o ligante de referência (desde que você acrescente o arquivo contendo a conformação cristalográfica antes de clicar em **Analyze**). Observe se o modo de ligação de menor energia (colorido por átomos) corresponde ao modo cristalográfico (em verde).



Reload Visualization

Run	Model	T.Energy (kcal/mol)	I.Energy (kcal/mol)	RMSD (Å)
5	1	92.892	-51.294	1.089
27	1	99.384	-49.169	3.207
19	3	99.52	-43.356	2.459
17	1	103.353	-47.051	1.809
15	5	104.908	-41.624	3.016
4	4	106.954	-35.01	3.372
12	3	108.358	-45.397	3.180
20	6	115.503	-26.997	4.963
12	9	118.478	-27.863	4.947
4	15	122.683	-20.367	1.794

☐ Show Protein ☑ Show Reference Ligand

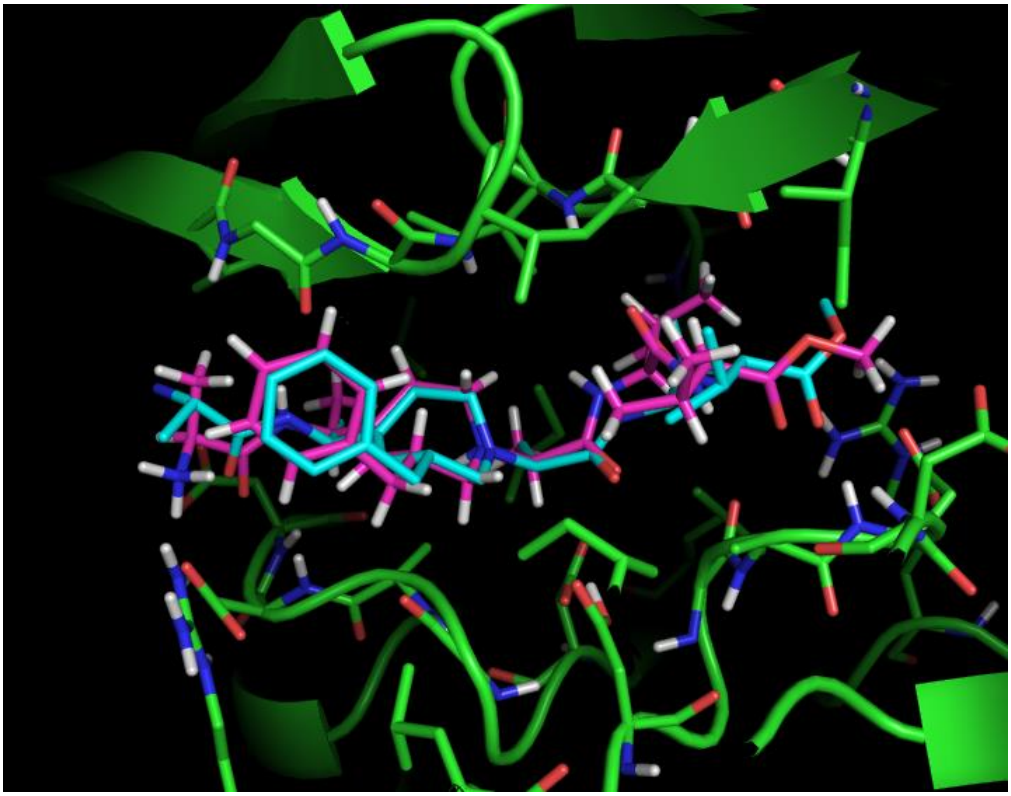
Para analisar cuidadosamente os resultados, abra no Pymol os arquivos:

- protein_******_prep.pdb (arquivo da proteína após preparação pelo DockThor)
- GAN.pdb (ligante com coordenadas cristalográficas)
- out.mol2 (conformações de menor energia)

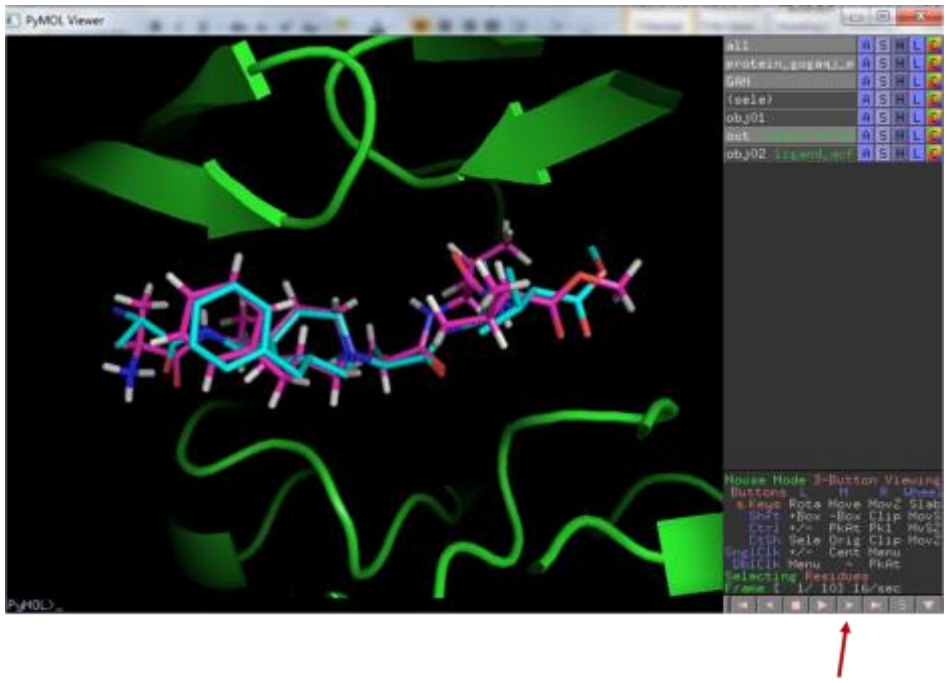
Altere as representações gráficas das proteínas e dos ligantes da maneira que achar mais conveniente para análise. Sugestão de etapas para a análise:

- crie um objeto contendo os resíduos presentes a uma distância de no máximo 6Å do ligante cristalográfico. Represente este objeto como **sticks**, colorindo por átomo.
- represente a proteína como **cartoon**
- represente o ligante GAN.pdb e os resultados de docking como **sticks**, colorindo ambos por átomo, porém colorindo de cores diferente os carbonos do ligante cristalográfico.
- no menu **Edit > Edit all** , altere o valor de **stick_radius** de 0.25 para 0.15 e dê enter.

Após todas estas etapas você deverá visualizar uma tela semelhante a mostrada abaixo:



A seguir, alterne entre as conformações de menor energia calculadas por docking, utilizando o botão indicado na figura abaixo:



Meça as distâncias das ligações de hidrogênio entre o ligante cristalográfico e a proteína. Observe se estas interações são preditas nos modos de ligação calculados por docking.