

Tutorial 1

*Redocking study of HIV-1 protease in
complex with amprenavir*

PDB code 1HPV [1]



www.dockthor.lncc.br

Laboratório Nacional de Computação Científica – LNCC/MCTI

April 2020

INPI Software Registration Number 13318-3

DockThor® registered mark

Introduction

The DockThor portal is a free protein-ligand docking server developed by the *Grupo de Modelagem Molecular em Sistemas Biológicos* at the National Laboratory for Scientific Computing (LNCC/MCTIC) idealized to facilitate and enable the use of the docking methodology by the scientific community [2]–[5]. The implemented DockThor® program is a flexible-ligand and rigid-receptor grid-based method that employs a multiple solution genetic algorithm[2] and the MMFF94S molecular force field[6] scoring function for pose prediction. Affinity prediction and ranking of protein-ligand complexes are performed with DockTScore, which is a linear empirical scoring function that takes into account important terms for protein-ligand binding, such as intermolecular interactions, ligand entropy and desolvation [7], [8]. 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 grid box and the genetic algorithm.

The results of docking experiments are clustered and ordered automatically. The DockThor Portal utilizes the computational facilities provided by the Brazilian SINAPAD (Sistema Nacional de Alto Desempenho) high-performance platform and the supercomputer SDumont, located at the LNCC/MCTIC, Brazil.

Submitting a Docking Job

The main purpose of this basic tutorial is to present the redocking process of a single ligand into the binding site of a protein to compare the predicted binding modes with its crystallographic structure. For this experiment, we will perform the docking of amprenavir (VX-478) into the binding site of HIV-1 protease (PDB code 1HPV [1]).

Currently, only protein receptors with .pdb or .in¹ files extensions are accepted. Non-protein receptors (e.g. DNA, RNA, another ligand) will be allowed on a further version of the DockThor Portal.

Attention: DockThor only works with a single tab at a time. If you open multiple tabs, only one of them will be able to submit docking experiments.

1.1 Protein Preparation

1. Click on the **Docking** tab. Then, click on the **Protein** tab to open the protein preparation page. Go to step ①.
2. Currently, it is only possible to upload *Protein Data Bank* (.pdb) or *DockThor input* (.in) protein files. The structure of the HIV-1 protease complexed with the drug amprenavir (PDB code 1HPV) is available as a test file. To use this structure, click on **Select Test File**. For other protein files, click on the **Add file** button to upload your protein file.
3. Click on the **Send** button. In this step, your file will be checked and you will be able to prepare your protein if your file is valid.
 - a. The *in-house* program PdbThorBox automatically parametrizes the protein atoms according to the MMFF94 atomic type and partial charges, completes missing side

¹ A prepared input file specific for the DockThor program.

chains and adds polar hydrogen atoms if necessary. Go to step ②.

4. In this step, all the missing atoms of the residue side chains will be reconstructed. The protein atoms are recognized in the .pdb format through the label "ATOM" located at the first column of the file. Atoms associated with the label "HETATM" are ignored.
 - a. All protein atoms are recognized by their .pdb atom label (e.g. CA, CB). If the nomenclature is not correct, the atom will be reconstructed.
 - b. 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.
 - c. The PdbThorBox program recognizes hydrogen atoms added from diverse software such as the Protein Preparation Wizard[9] from Schrödinger (Maestro, Schrödinger, LLC, New York, NY, 2020) and Protoss[10].
5. Check the protonation states of the amino acid residues. The Protonation States box provide the following information for each amino acid residue:
 - a. **Residue index:** position of the residue in the PDB file. It always starts at position 1. The residue index will replace the original residue number in the prepared .pdb file (*_prep.pdb).
 - b. **Residue number:** original residue number in the PDB file.
 - c. **Protonation state:** code of the protonation state. Default protonation states are set automatically if polar hydrogen atoms are absent or unrecognized - see Appendix A or click on ⓘ.
6. For the protein test file, it is important to change the protonation state of a catalytic aspartate, which interact with the inhibitor:

Asp25 of chain B (numbering according to the *residue index*²) must be of type ASPN1 (*i.e.*, protonated at the OD1 atom). Select **Chain B** → **Asp** and change the type of Asp-25 from ASP to ASPN1. An exclamation point appears for each residue with protonation state altered but not applied yet.

7. Prepare the protein with the new protonation states clicking on **Apply**. Check the modification log in the blue box. Go to Step ③.
8. View the prepared protein file with JSmol clicking on **View 3D**. Useful information about the use of this plugin can be found at <http://www.chem.uwec.edu/JmolTut/>.
9. Click on **Download** to download the compacted directory containing the prepared files in the OUTPUT/ folder. The files uploaded in DockThor Portal are stored with random names.
protein.in – DockThor receptor input file. It contains the MMFF94 atomic type numbers, the atomic coordinates, the bond connectivities, the atomic partial charges, the PDB atom types, residue and chain labels and the residue and chain numbers.
protein_prep.pdb – prepared protein file in the *Protein Data Bank* format for visualization purposes.
10. If you just want to download the prepared .pdb file, select **Prepared file (.pdb)**. In the same way, it is possible to download only the .in file for further docking studies (**Topology file (.in)**).
11. Click **Send to DockThor** → **Yes** to submit the protein file to docking.

Attention: it is mandatory to click this button to save the current protein state for docking.

1.2 Cofactor and Water Preparation

² Residue Index corresponds to the order of the residues after renumbering and Residue Number is the original number of the residue in the input .pdb file.

For some protein-ligand complexes, it is important to consider cofactors (e.g. NAD, ATP, FAD, Mg, Zn etc) and/or water molecules. The DockThor Portal allows the users to upload up to ten cofactor and/or water molecules; they are kept fixed during the docking simulation. If no cofactor is needed for your docking job, click on **Do Not Use Cofactors**.

1. Click **Select Test File** to upload the water molecule HOH201 from chain A of the 1HPV complex that mediates the protein-ligand interaction. To upload your own cofactor files, click on the **Add file** button.
2. Check **Add H** to add hydrogen atoms and click **Send**. In this step, the input cofactor files are converted to a topology file specific for DockThor (.top file type).
 - a. The *in-house* program MMFFLigand automatically parametrizes the cofactor atoms according to the MMFF94S atomic type and partial charges, and adds all hydrogen atoms if required by the user (**Add H** option enabled). Go to step ②.

Attention: Check carefully the hydrogen atoms added to the cofactor files. Hydrogens are added automatically to the cofactor structures without considering the protein-ligand environment. For waters and cofactors with hydrogen bond donor atoms, it is strongly recommended to perform this step on an external program to optimize the hydrogen bond network of the complex.

3. Visualize the prepared file with JSmol clicking on **View 3D**.
4. Click on **Download** to download the compacted directory containing the prepared files in the OUTPUT folder. The files uploaded in DockThor Portal are stored with random names, but you can easily identify them in the *mapfile.csv* table.

cofactor_log.txt – log of the MMFFLigand program.

cofactor_mapfile.csv – file that associates the random IDs with the original filenames.

invalid.txt – list of invalid structures.

cofactor_ID.top – DockThor 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 intramolecular atom interactions.

new_cofactor_ID.format – prepared ligand in the same format of the input file (SDF, MOL2 or PDB). This file is generated only when the **Add H** option is enabled.

5. If you just want to download the .top file for further docking studies select **Topology file(s) (.top)**.
6. Click **Send to DockThor** → **Yes** to save the cofactor files.

Attention: it is mandatory to click this button to save the current cofactors for docking.

1.3 Ligand Preparation

The DockThor Portal allows the users to upload up to a hundred structures as guest users, or a thousand structures if you have an approved project.

1. In this tutorial, we will perform a redocking experiment of the ligand originally complexed with the 1HPV structure (amprenavir, PDB identifier 478). Click **Select Test File** → **Single docking (1 molecule)** to upload the test ligand file.
 - a. If you wish to upload your own structure(s), click **Add File** to submit one or more files. It is possible to upload files containing a single compound or multiple molecules in the formats .pdb (single molecule), .sdf and .mol2.
2. Since X-ray structures do not generally contain hydrogen atoms, we need to add the hydrogens to the ligand to be a valid molecule for the force field. To add the hydrogen atoms automatically, check the **Add H** check box. Click **Send** to upload the file and prepare the amprenavir structure. The

hydrogen atoms will be added by the OpenBabel tool[11] using pH = 7.4.

Attention: The structures submitted in the Cofactor and Ligand tabs are validated if they can be recognized by the force field. For each input file, different symbols indicate if the molecules are valid ✓, invalid ✗ or contains some valid and invalid structures at the same time ⚠ - in this case, just the valid structures will be submitted to docking.

- a. Uploaded structures are checked according to some properties to remove invalid molecules. Molecules are considered as valid if:
 - i. Molecular weight \leq 1500 Da
 - ii. Rotatable bonds (including terminal hydroxyl groups) \leq 60
 - iii. Contains a 3D structure (*i.e.*, the column of the Z coordinate is not null)
 - iv. Atom types are parametrized in the MMFF94S force field³
 - b. In this step, the in-house program MMFFLigand automatically parametrizes the atoms according to the MMFF94S atom types and partial charges, and adds all the hydrogen atoms if required by the user (**Add H** option enabled).
 - c. Go to step ②.
3. In this tutorial we will keep the ligand as totally flexible, so keep all the rotatable bonds enabled.
- a. The option to disable/enable rotatable bonds is available when only one ligand is uploaded.
 - b. Go to the Step ③.
4. To view the prepared structure of the ligand(s), click **View 3D**.

³ Check the MMFF94S force field reference[6, p. 94] for a list of the allowed atom types.

5. Click on **Download** to get the compacted directory containing the prepared files (OUTPUT/ folder). Except for the test files, all files uploaded in DockThor Portal are stored with random names, but you can easily identify them in the *mapfile.csv* table.

ligand_log.txt – log of the MMFFLigand program.

ligand_mapfile.csv – contains the relationship between the random names and the original filenames. Also contains the ligand names present in the .mol2 and .sdf files according to the "@TRIPOS_MOLECULE" and "<name>" tags, respectively.

invalid.txt – list of invalid structures.

ligand_ID.top – DockThor input file for each compound. 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 intramolecular atom interactions.

new_ligand_ID.format – prepared ligand in the same format of the input file (SDF, MOL2 or PDB). This file is generated only when the **Add H** option is enabled.

6. Click **Send to DockThor** to send the prepared ligand file(s) to docking.

1.4 Docking Configuration

1. The next tab consists of the main options to run the docking job. It is possible to see the list of the input files. Check if it is correct.
2. In the Step ②, select the center and dimensions of the grid box that represents the binding site. For this tutorial, click **Test** to load the configuration for the 1HPV complex. It is possible to see a box representing the grid dimensions in the NGL Viewer plugin [12]–[14].

- a. The **grid center** corresponds to the center of the search space. This can be defined, for example, as the X, Y, Z coordinates of a reference atom in the 3D space. In the case of 1HPV, we selected as reference the atom C6 of amprenavir (X = 10.387, Y = 18.119, Z = 8.484).
- b. The **grid size** corresponds to the size of the grid box. The values to be inserted in the X, Y, Z fields for the grid size correspond to the **total value** to be used on each dimension. For example, the configuration for the test files (**Test** button) has the values of X = 20, Y = 20 and Z = 20 corresponding to a grid box with a total size of 20 Å on each dimension.

Attention: The **Blind Docking** option generates a grid box centered on the center of coordinates of the protein with a size covering the entire receptor. You can use this configuration as an initial definition to change interactively the grid box location and size. Use this option with caution if you intend to search for binding sites.

- c. The **discretization** value defines the distance between the grid points. The default value is 0.25 Å, but you can increase this value if you need to use large grid boxes. However, the increase of the discretization value will decrease the accuracy of the docking experiment and must be within the maximum number of points (950 000).

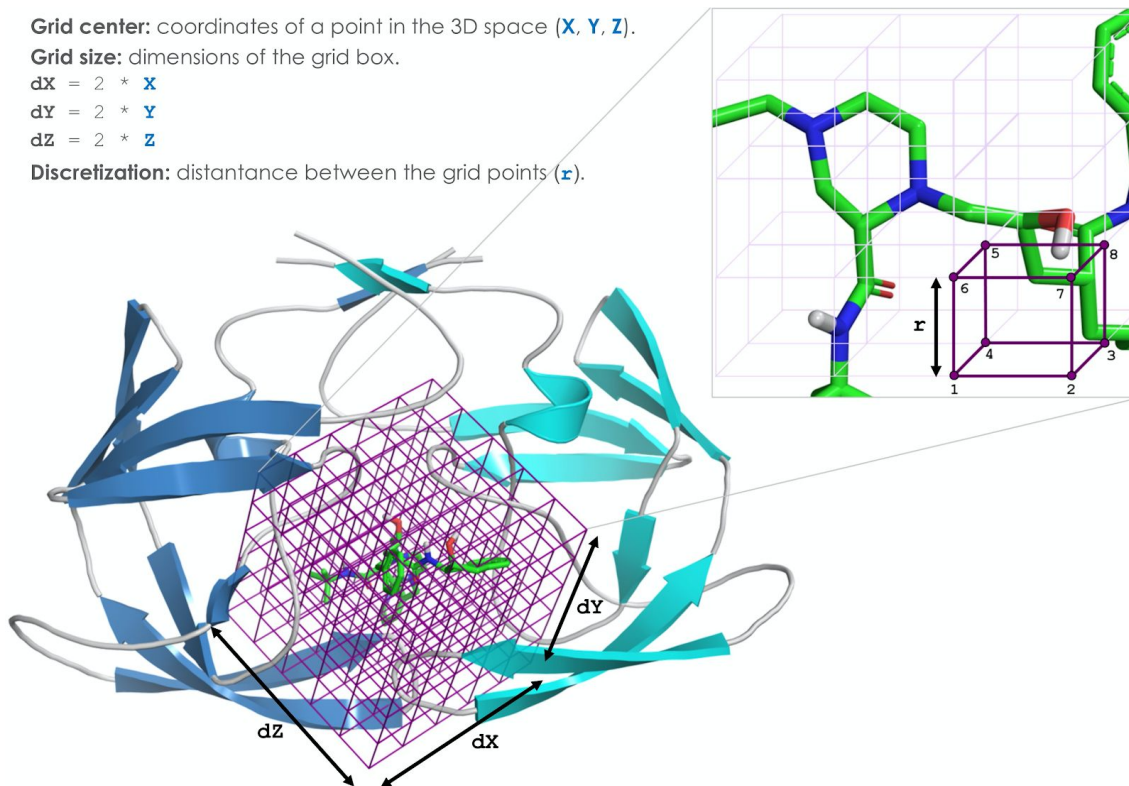


Figure 1. Configuration of the grid box representing the search space for the docking experiment.

3. Some genetic algorithm (GA) parameters can be modified in the DockThor Portal. In this simulation, we will keep the **Standard** configuration:
 - a. Number of evaluations = 1,000,000
 - b. Population size = 750
 - c. Initial seed = -1985
 - d. Number of runs = 24
 - e. Soft docking activated: softening the MMFF94S Buf-14-7 potential using a buffering constant of 0.35 instead of 0.07.
4. Choose a **Job name** to facilitate the identification of your docking job. It will be followed by a random identifier generated automatically.

5. Fill the **Email** field to receive the link for the results page when the docking job is submitted and finished. You can insert up to 5 addresses.
6. If you wish, check the option to receive DockThor e-Newsletters (e.g. information about new versions and improvements).
7. Accept the terms of use and submit the docking job clicking on **Dock!**
8. You will receive an email with the link for the results page when the job is submitted. In addition, an email will be sent as soon as your docking is finished.

2. Analyzing the Docking Job

Once the docking job is finished, the user receives an email with the link to the Results page. Click **View results** to go to the Analyses tab.

2.1 Results and Analyses

The docked poses obtained in all docking runs are clustered by the DTStatistic program and ranked according to the total energy (named T. Energy) from the MMFF94S force field, which is composed of the intermolecular and intramolecular interactions according to the electrostatic and vdW potentials and the torsional energy of the ligand. For each docking run, only the top-ranked leaders of each cluster will be used to the final clustering analysis step:

1. In the **Results** tab, you can change the clustering criterion (default = 2.0 Å), the number of different binding modes of the ligand (up to 10 poses for single docking and up to 3 poses for virtual screening) and upload a file with the reference ligand to calculate RMSD with docking poses when performing a redocking experiment. Change these parameters as you wish.

2. Get the reference ligand file of the amprenavir at the DockThor Portal (**Support** → **Help** → *1hpn_ligref.pdb*).
3. Click on **Analyse** to cluster and analyze docking results.
4. Explore the different binding modes obtained for the DPS ligand in the results table. For each top-ranked pose, it is shown the corresponding **docking run** of the Genetic Algorithm, the **model** (number of the cluster leader in the corresponding GA run), the predicted binding affinity by the DockTScore program (**Score**) and the **Total Energy**.
5. The **RMSD** column appears when you select a reference file to calculate the RMSD in redocking experiments. Instead, if you do not intend to compare with a reference ligand, the **Intermolecular Energy** is shown in the place of the RMSD column.

Attention: The predicted binding affinity (Score) and the energies from the MMFF94S force field (T. Energy and I. Energy) are given in kcal/mol units.

6. To download the docking results and the top-ranked docking poses, click **Download**.

parameters.txt – the configuration of the docking experiment.

COFACTOR/ – directory with the output files of the cofactors (if used).

LIGAND/ – directory with the output files of the ligands.

PROTEIN/ – directory with the output files of the protein.

RESULT/ – directory with the output files of the docking experiment containing the following files:

bestranking.mol2 – multimol2 file with the top-energy pose (*i.e.*, the pose with the lowest Total Energy) of all compounds, ranked according to the predicted binding affinity (Score).

bestranking.csv – sheet with the energies of the top-energy pose of all compounds, ranked according to the predicted binding affinity (Score). Also, the name of the original file(s) and the molecule name (if available) are provided.

result_*.log – summary of the clustering analysis for the top-energy docked poses (ranking, energies, RMSD and affinity prediction) ranked according to the Total Energy.

result_*.mol2 – contains the atomic coordinates of the top-ranked docked poses (multimodel .mol2 type file, ranked according to the respective .log file).

ligand*_run_X.log – contains the information of the cluster leaders obtained in run X (using a RMSD criterion of 1.0Å), where X is the number of the docking run.

ligand*_run_X.pdb – contains the atomic coordinates of the cluster leaders obtained in run X (multimodel .pdb type file), where X is the number of the docking run.

Attention: The Total Energy is used to rank different binding modes of the same compound. The affinity prediction is used to rank different ligands in virtual screening experiments considering only the top-energy pose of each compound.

7. Click **Remove Files** if you want to delete all files associated with the docking experiment from our server. This operation cannot be undone.

3. Software

The Portal DockThor uses the following programs developed by the GMMSB/LNCC research group:

- **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 the total energy from the MMFF94S force field.
- **DockTScore:** predicts the binding affinity of the protein-ligand complexes using a linear and empirical scoring function[8] consisting of physics-based terms related with the intermolecular interactions, ligand entropy and desolvation. Developed in collaboration with the *MTi - Molécules Thérapeutic in silico* (Université Paris Diderot - Inserm UMR-S 973).

References

- [1] E. E. Kim *et al.*, "Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme," *J. Am. Chem. Soc.*, vol. 117, no. 3, pp. 1181–1182, Jan. 1995, doi: 10.1021/ja00108a056.
- [2] C. S. de Magalhães, D. M. Almeida, H. J. C. Barbosa, and L. E. Dardenne, "A dynamic niching genetic algorithm strategy for docking highly flexible ligands," *Inf. Sci.*, vol. 289, pp. 206–224, Dec. 2014, doi: 10.1016/j.ins.2014.08.002.
- [3] C. S. de Magalhães, H. J. C. Barbosa, and L. E. Dardenne, "Selection-Insertion Schemes in Genetic Algorithms for the Flexible Ligand Docking Problem," in *Genetic and Evolutionary Computation – GECCO 2004*, vol. 3102, K. Deb, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 368–379.
- [4] K. B. dos Santos, I. A. Guedes, A. L. M. Karl, and L. Dardenne, "Highly Flexible Ligand Docking: Benchmarking of the DockThor Program on the LEADS-PEP Protein-peptide Dataset," *J. Chem. Inf. Model.*, p. acs.jcim.9b00905, Jan. 2020, doi: 10.1021/acs.jcim.9b00905.
- [5] D. M. Almeida, "DockThor: Implementação, Aprimoramento e Validação de um Programa de Atracamento Receptor-Ligante," Dissertação de Mestrado, Laboratório Nacional de Computação Científica – LNCC, 2011.
- [6] T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94," *J. Comput. Chem.*, vol. 17, no. 5–6, pp. 490–519, Apr. 1996, doi: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.
- [7] I. A. Guedes, A. M. S. Barreto, M. A. Miteva, and L. E. Dardenne, "Development of Empirical Scoring Functions for Predicting Protein-Ligand Binding Affinity," Laboratório Nacional de Computação Científica, Petrópolis - RJ, 2016.
- [8] I. A. Guedes *et al.*, "DockTScore: New Physics-Based and Target-Specific Empirical Scoring Functions (article under submission phase)," 2020.
- [9] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aided Mol. Des.*, vol. 27, no. 3, pp. 221–234, Mar. 2013, doi: 10.1007/s10822-013-9644-8.
- [10] S. Bietz, S. Urbaczek, B. Schulz, and M. Rarey, "Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes," *J. Cheminformatics*, vol. 6, no. 1, p. 12, 2014, doi: 10.1186/1758-2946-6-12.
- [11] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox," *J. Cheminformatics*, vol. 3, no. 1, p. 33, 2011, doi: 10.1186/1758-2946-3-33.
- [12] A. S. Rose, A. R. Bradley, Y. Valasatava, J. M. Duarte, A. Prlić, and P. W. Rose, "NGL viewer: web-based molecular graphics for large complexes," *Bioinformatics*, vol. 34, no. 21, pp. 3755–3758, Nov. 2018, doi: 10.1093/bioinformatics/bty419.
- [13] H. Nguyen, D. A. Case, and A. S. Rose, "NGLview—interactive molecular graphics for Jupyter notebooks," *Bioinformatics*, vol. 34, no. 7, pp. 1241–1242, Apr. 2018, doi: 10.1093/bioinformatics/btx789.
- [14] A. S. Rose and P. W. Hildebrand, "NGL Viewer: a web application for molecular visualization," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W576–W579, Jul. 2015, doi: 10.1093/nar/gkv402.

Appendix A

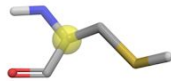



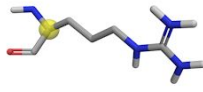
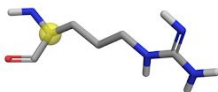
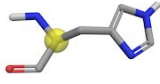
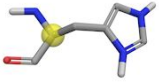
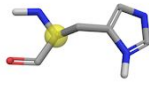
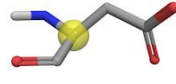
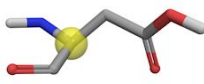
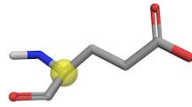
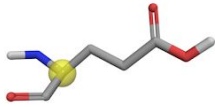
Protonation States of the Amino Acid Residues

Protonation States of Amino Acid Residues

Carbon, nitrogen, oxygen and sulfur atoms are colored grey, blue, red, and yellow, respectively.

Yellow spheres highlights the C α . * indicates the default protonation state.

The label **I** refers to the initial residue and **T** to the terminal residue.

Cysteine	Cys (C)	CYSH - neutral CYS* - negatively charged CYSS - disulfide bond	 CYSH	 CYS, CYSS	
Lysine	Lys (K)	LYS* - positively charged LYSN - neutral (NH2)	 LYS	 LYSN	
Arginine	Arg (R)	ARG* - positively charged ARGN1 - neutral at N ω ARGN2 - neutral at N ω'	 ARG	 ARGN1, ARGN2	
Histidine	His (H)	HISE* - neutral at NE τ HISD - neutral at N τ HISP - positively charged	 HISE	 HISP	 HISD
Aspartate	Asp (D)	ASP* - negatively charged ASP1 - neutral at OD1 ASP2 - neutral at OD2	 ASP	 ASP1, ASP2	
Glutamate	Glu (E)	GLU* - negatively charged GLUN1 - neutral at OE1 GLU2 - neutral at OE2	 GLU	 GLUN1, GLUN2	