Version 2

Aug 10, 2021

Analysis of protein structure using Molprobity V.2

In 1 collection

Chris Berndsen[1]

[1]James Madison University

1 Works for me Share

This protocol is published without a DOI.

Chris Berndsen
James Madison University

ABSTRACT

Molprobity is a valuable collection of structural analysis tools for the validation or "correctness" of protein structures and homology models. This protocol walks through the basics of running the webserver version of Molprobity and provides some information on how to interpret the results.

PROTOCOL CITATION

Chris Berndsen 2021. Analysis of protein structure using Molprobity. **protocols.io** https://protocols.io/view/analysis-of-protein-structure-using-molprobity-bw9tph6n Version created by Chris Berndsen

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Williams et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Science 27: 293-315.

COLLECTIONS

Biochemistry I methods

CREATED

Aug 10, 2021

LAST MODIFIED

Aug 10, 2021

PROTOCOL INTEGER ID

52243

PARENT PROTOCOLS

Part of collection
Biochemistry I methods

MATERIALS TEXT

computer with an internet connection
structural coordinates in PDB format
molecular viewing software (optional)

BEFORE STARTING

A .pdb file from experimental data such as X-ray crystallography or Cryo-EM or a model from one of many

softwares is required.

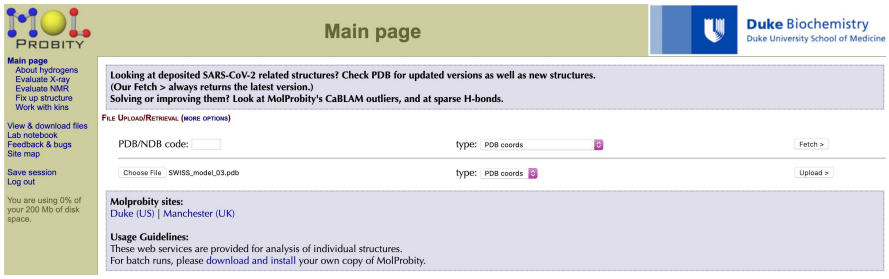
1 This protocol describes submitting a .pdb file to the Molprobity server hosted at Duke University.

Molprobity has now been incorporated into SWISS-Model results as well. The analysis aspects of either method are comparable and the analysis section can be applied to data from this latter instance. See the sections on the multi-criterion table and Ramachandran plot in this case.

Prepare file

2 Navigate to the Molprobity home page

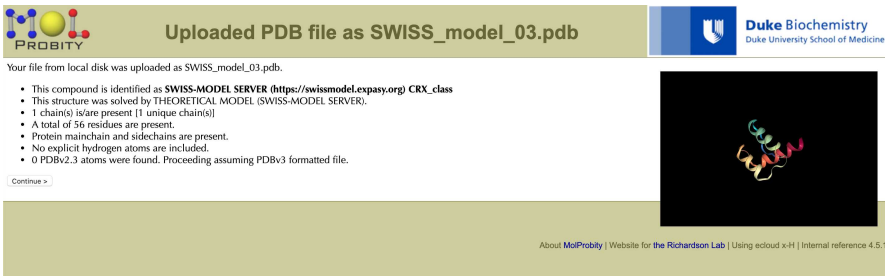
3 Upload the .pdb file and press the upload file button.



Upload screen for Molprobity

The server will then process the file. This can take a few seconds to a few minutes depending on size of the molecule, format, etc.

3.1 If successful, the a screen similar to the following will appear. Press Continue to move to the analysis phase.



Successful upload to the server

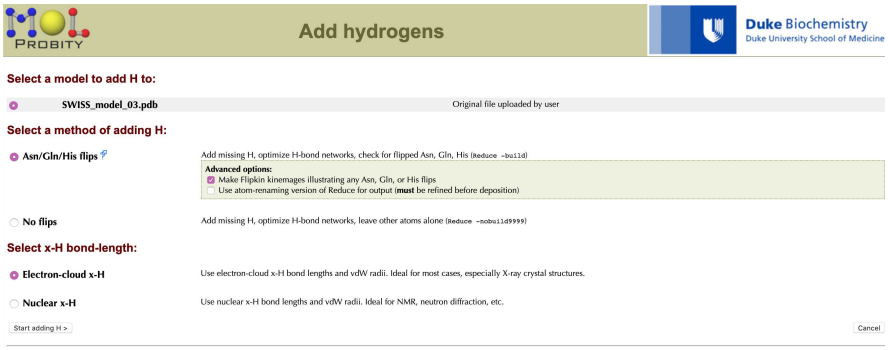
3.2 Record the name of the PDB file that you used and make sure that it is on OSF.

4 In the analysis menu, select Add Hydrogens.

Adding hydrogens is not required, but will improve the analysis reliability. Most homology modeling programs do not include hydrogens because hydrogens are not observed (for the most past) in X-ray crystallography experiments.

4.1 The defaults are fine for most analyses. Press Start adding H to begin adding hydrogens to the structure.

1. Flips optimizes hydrogen bonding, which can be good in models. If you have a crystal structure, then maybe run the analysis twice with and without flips and see the difference.
2. Electron cloud is appropriate for most structures.



4.2 Record any amino acids that were flipped in the table below. Add rows as needed.

A
Flipped amino acids

4.3 If hydrogens were added and geometry flipped. The model has been changed and the new model should be kept for further analysis.

Molprobity will ask if you want to download the file say that you do. Upload this to your project folder for this project and name the file:

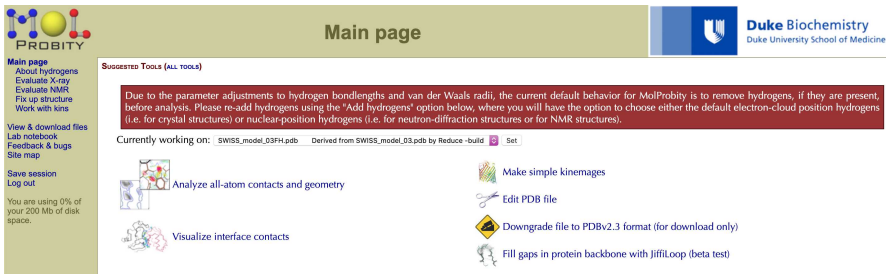
```
[date]_[sequencename]_[team_name]_Model_h_added.pdb
```

Replace **[Group_name]** with your name/group name without the brackets. Replace **[sequence_name]** with the name of the sequence.

4.4 Indicate your project file location as a link within a note on this step.

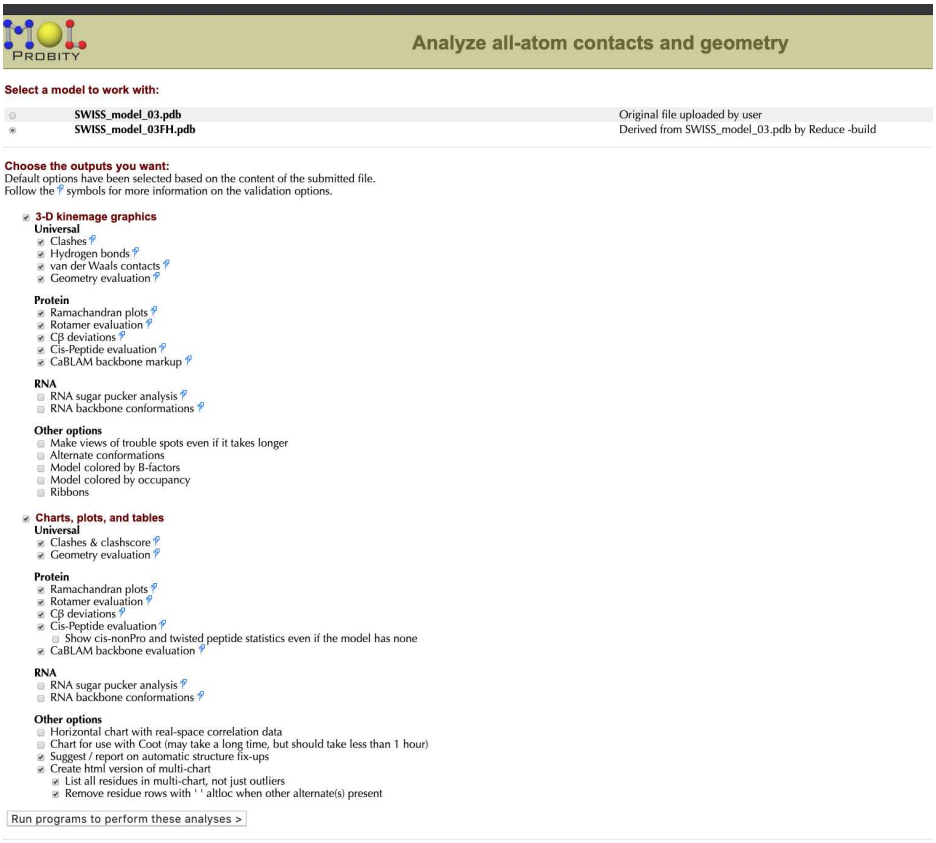
Analyze the structure

5 When the addition and optimization is complete. The screen with show the options below. Select Analyze all-atom contacts and geometry.



5.1 The menu on the next page allows for selection parameters to analyze. In general the selections shown below are appropriate for analysis of protein homology models.

Press Run programs to perform analyses and wait about 60 seconds.



6 The results screen brings up many tables and files.

6.1 The summary table is at the top of the screen and should appear similar as that shown below.

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	6.15		90th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	45	88.24%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	54	100.00%	Goal: >98%
	Rama distribution Z-score	-0.42 ± 0.89		Goal: abs(Z score) < 2
	MolProbity score^	1.34		98th percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 491	0.00%	Goal: 0%
	Bad angles:	2 / 659	0.30%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 2	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	0	0.0%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/70		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

In the two column results, the left column gives the raw count, right column gives the percentage.
* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.
^ MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.
Key to table colors and cutoffs here:

Summary table of a homology model

6.2 *Score descriptions:*

- **Clashscore, all atoms**: This score rates the sterics, lower number is better, while higher percentile is better. The percentile is the value that should be reported.
- **Rotamers**: Refers to the geometry of the amino acid side chains. The number indicates the number of amino acids in the poor or favored category and the percentage of amino acids that fall into those categories. If there is a high percentage of poor rotamers (>1.5%), this can be concerning.
- **Ramachandran**: Shows number of amino acids with poor or favored phi/psi angles. Phi/Psi angles are the dihedral angles in the protein backbone. Only certain angles are typically found in proteins.
- **Rama distribution Z-score**: Can show over fitting of angles to unrealistic levels. A value -4 < x < 2 is considered appropriate for normal structures. This pre-print provides more details: Link
- **Molprobity score**: Combines all the geometric scores into a single value to suggest quality. Lower values and higher percentiles are better.
- **Cβ, Bad bonds, band angles, cis-Prolines**: Indicate number of amino acids with poor geometry or chemical parameters.
- **CaBLAM outliers**: C-Alpha Based Low-resolution Annotation Method, looks at the backbone geometry.

Multi-criterion table

7 At the bottom of the page are several additional files that can be viewed for deeper analysis.

Multi-criterion visualizations

View in KiNG | View in NGL | Download (799 Kb)

View (73 Kb)

View (2.8 Kb)

Single-criterion visualizations

- **Clash list** (421 bytes): View
- **Ramachandran plot kinemage** (408 Kb): View in KiNG | View in NGL | Download
- **Ramachandran plot PDF** (1.7 Mb): View
- **Ramachandran distribution Z-score analysis** (4.2 Kb): View
- **Chiral volume report** (786 bytes): View
- **Cβ deviation scatter plot** (12 Kb): View in KiNG | View in NGL | Download

Continue >

7.1 The Multi-Criterion chart is the highest level of detail and resolution and shows the residue level problems in the protein.

- Boxes shown in pink indicate the bad parameter at that amino acid position.
- More than 5 consecutive amino acids with bad parameters may mean an issue with that part of the protein and should be inspected manually as well as reported.

#	Alt	Res	High B	Clash > 0.4Å	Ramachandran	Rotamer	Cβ deviation	CaBLAM	Bond lengths	Bond angles	Cis Peptides
			Avg: 0.62	Clashscore: 6.15	Outliers: 0 of 54	Poor rotamers: 0 of 51	Outliers: 0 of 56	Outliers: 0 of 52	Outliers: 0 of 56	Outliers: 2 of 56	Non-Trans: 0 of 55
A 38		LYS	0.23	-	-	Favored (97.3%) *mttt* chi angles: 289.7,181,179.2,177.7	0.03Å	-	-	-	-
A 39		GLN	0.39	-	Favored (38.65%) General / -56.9,143.2	Allowed (0.5%) *mm-40* chi angles: 267.6,257.9,39.9	0.03Å	-	-	-	-
A 40		ARG	0.28	-	Favored (10.57%) General / -102.1,166.5	Allowed (0.3%) *mtt-85* chi angles: 306.6,162.1,131.6,236	0.09Å	Favored (21.839%)	-	-	-
A 41		ARG	0.42	-	Favored (43.93%) General / -63.5,150.3	Favored (3.9%) *ptp-110* chi angles: 72.8,162.4,53.1,251.9	0.07Å	Favored (18.838%)	-	-	-
A 42		GLU	0.46	-	Favored (41.19%) General / -54.8,139.1	Favored (90.4%) *tt0* chi angles: 186.5,176.7,182.5	0.06Å	Favored (29.009%)	-	-	-
A 43		ARG	0.47	-	Favored (23.77%) General / -61.3,125.4	Favored (23.1%) *ttt90* chi angles: 191,176.4,169.8,110.2	0.02Å	Favored (41.789%)	-	-	-
A 44		THR	0.48	-	Favored (39.16%) General / -73.7,130.0	Favored (96.6%) *m* chi angles: 300.7	0.04Å	Favored (42.832%) beta sheet	-	-	-

7.2 Save the multi-criterion chart as a PDF. Upload this to your folder for this project and name the file:

--> Typically print to PDF works best in Chrome with a landscape orientation

```
[date]_[sequencename]_[groupname]_molprobity.pdf
```

Replace **[Group_name]** with your name/group name without the brackets. Replace **[sequence_name]** with the name of the sequence.

7.3 Indicate your file location as a link within a note on this step.

THIS IS ONE OF YOUR DATA FILE FOR THE ANALYSIS!

Ramachandran analysis

8 The Ramachandran plot is also important for analysis. Select the Ramachandran plot PDF.

For more information on the Ramachandran plot see:
http://proteopedia.org/wiki/index.php/Tutorial:Ramachandran_principle_and_phi_psi_angles

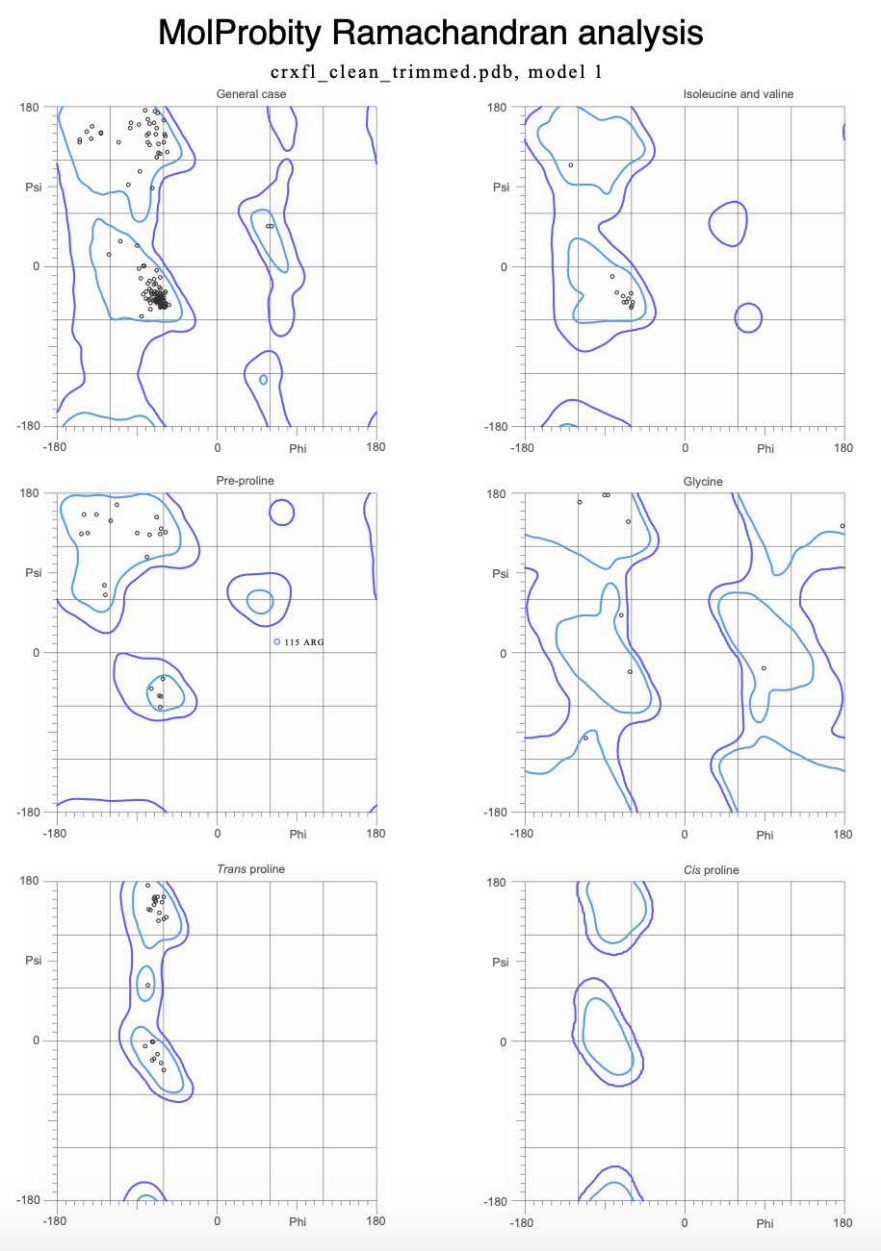
and

https://proteopedia.org/wiki/index.php/Ramachandran_Plots

8.1 The Ramachandran plot PDF should appear similar to those shown below.

- There are 6 plots, the main one to be concerned with early on is the general case shown at top left.
- Ideally the dots (which represent the angles for each amino acid) are all enclosed inside the blue lines.
- Outliers are marked with amino acid three letter code and position number.

- More than 5 consecutive amino acids with bad parameters may mean an issue with that part of the protein and should be inspected manually as well as reported.



8.2 Save the Ramachandran plot as a PDF. Upload this to your OSF folder for this project and name the file:

```
[date]_[sequencename]_[teamname]_ramachandran.pdf
```

Replace **[Group_name]** with your name/group name without the brackets. Replace **[sequence_name]** with the name of the sequence.

8.3 Indicate your file location as a link within a note on this step.

9 Save this protocol as a PDF and include in your project folder.