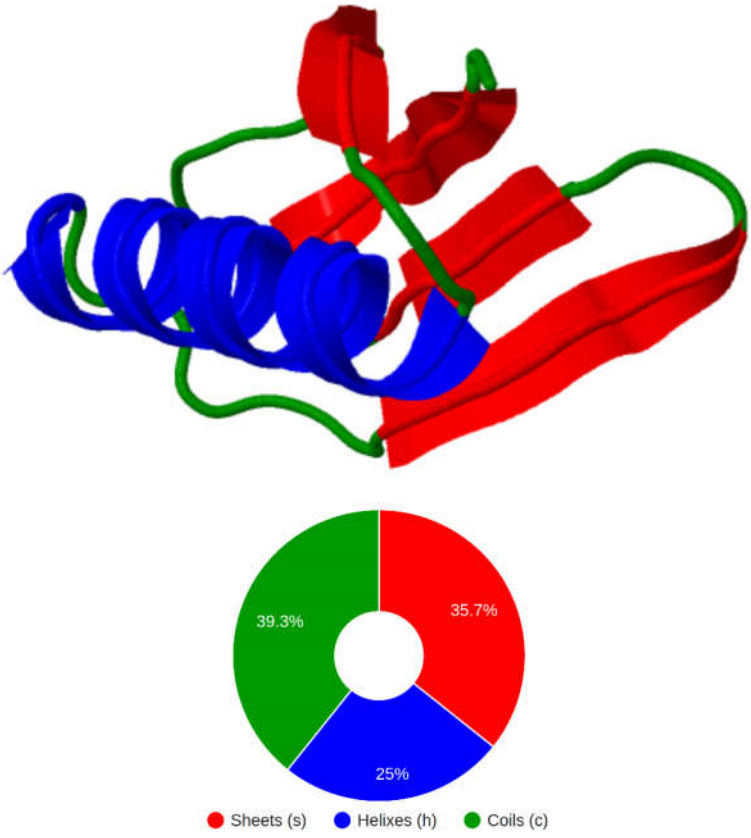
ProSSA: (Hydrogen-Bond-Independent) Proteins' Secondary Structure Assignment

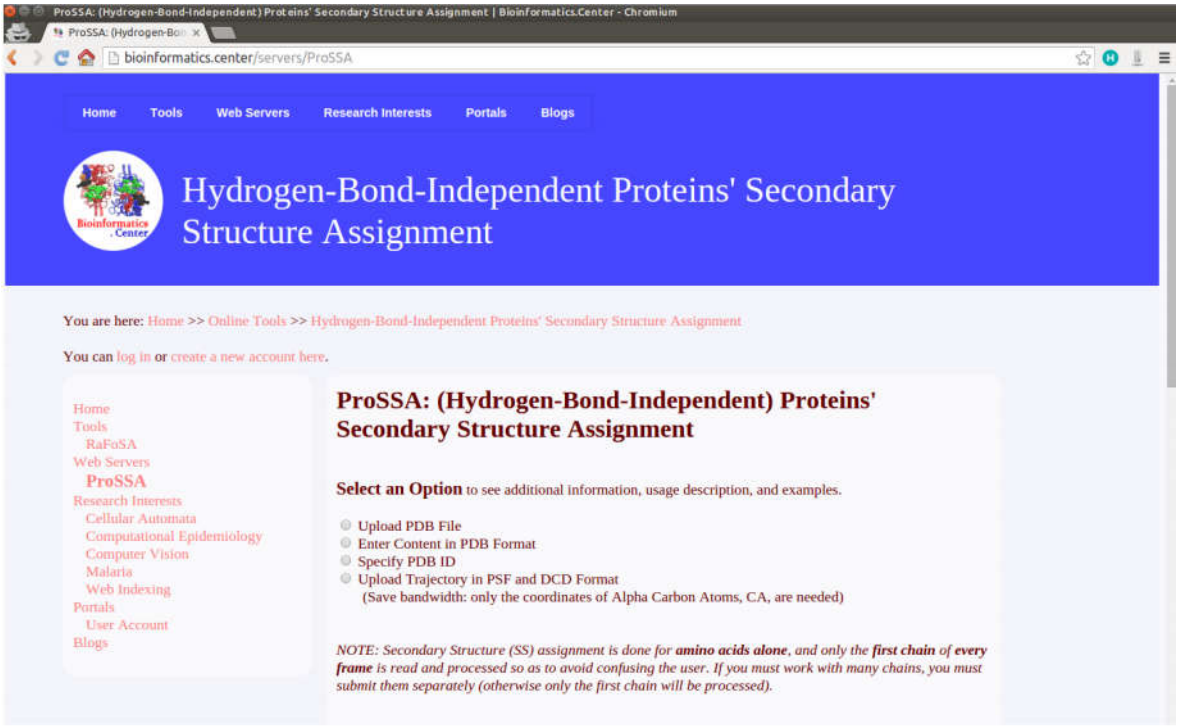
Concise User Guide



http://bioinformatics.center/servers/ProSSA

December, 2015

[1] To use ProSSA, visit http://bioinformatics.center/servers/ProSSA
You should see a page that looks like the screenshot below.



[2] Select the approach/format through which you intend to provide input for ProSSA.

There are four options available. You can either

[A] Upload PDB File

[B] Enter Content in PDB Format

[C] Specify PDB ID or

[D] Upload Trajectory in PSF and DCD Format

Regardless of the approach/format you select,

[I] Secondary structure assignment is done for protein residues alone (i.e. for amino acids alone). DNA, RNA, water, or ligand molecules are ignored.

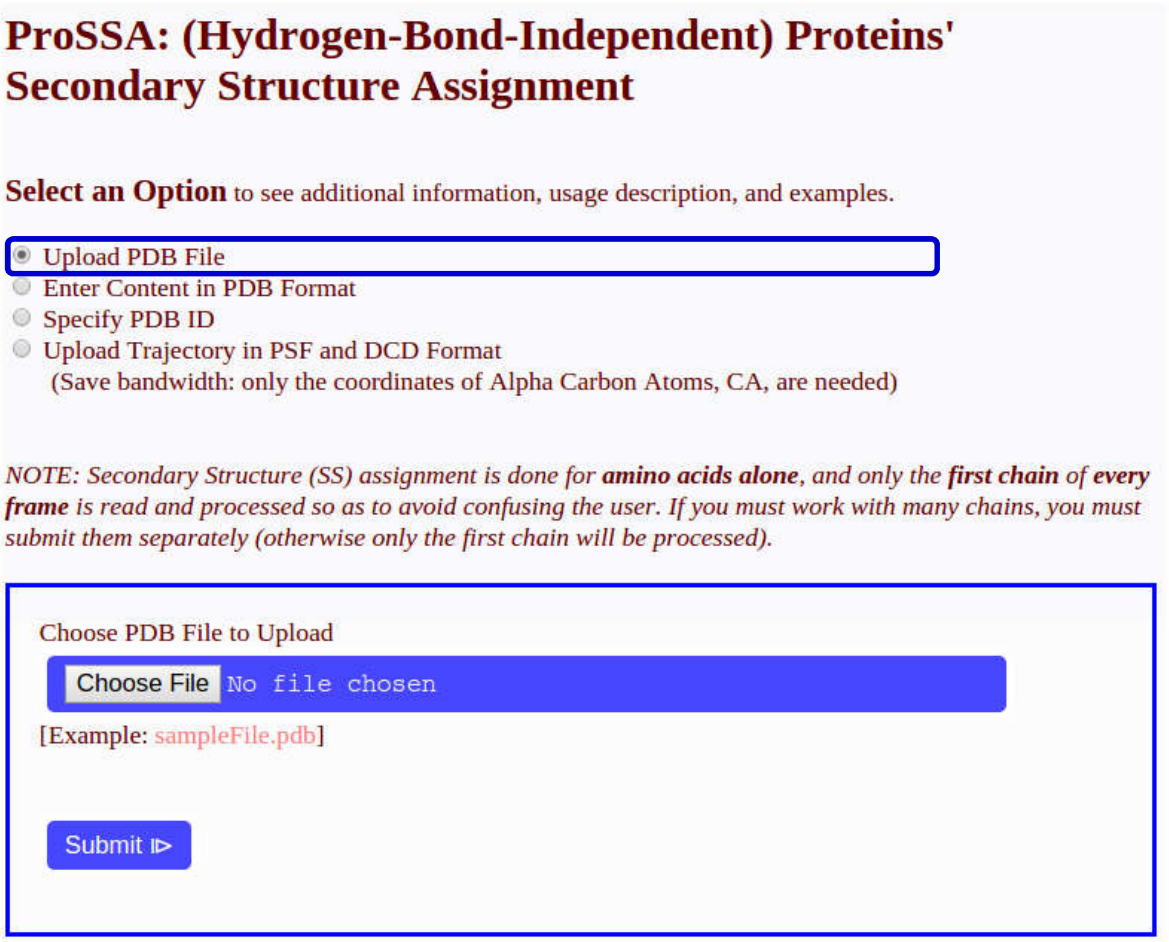
[II] Only the first protein chain in the data provided is used.

[III] All the frames/models are processed. In fact, we encourage submitting multiples frames whenever user has more than one frame/model for the same molecule of interest.

[IV] Only the information of Alpha Carbon atoms are needed and used regardless of whether the user submits all-atom information or not.

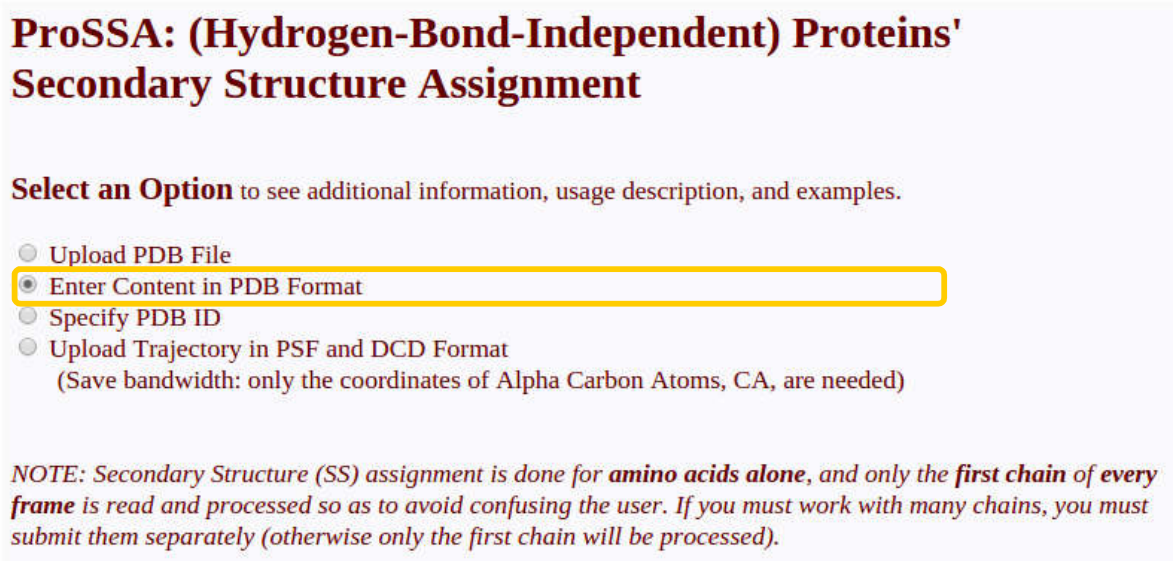
[V] We encourage users to submit information of Alpha Carbon atoms alone so as to save bandwidth, especially when submitting multiple jobs with multiple frames/models. We provide a tutorial on how to create PSF, DCD, or PDB file that contains only Alpha Carbon atoms here http://bioinformatics.center/blogs/1001/1001

[Option A] If the user chooses to submit a PDB File, a form for uploading PDB File is displayed such as in the screenshot below.



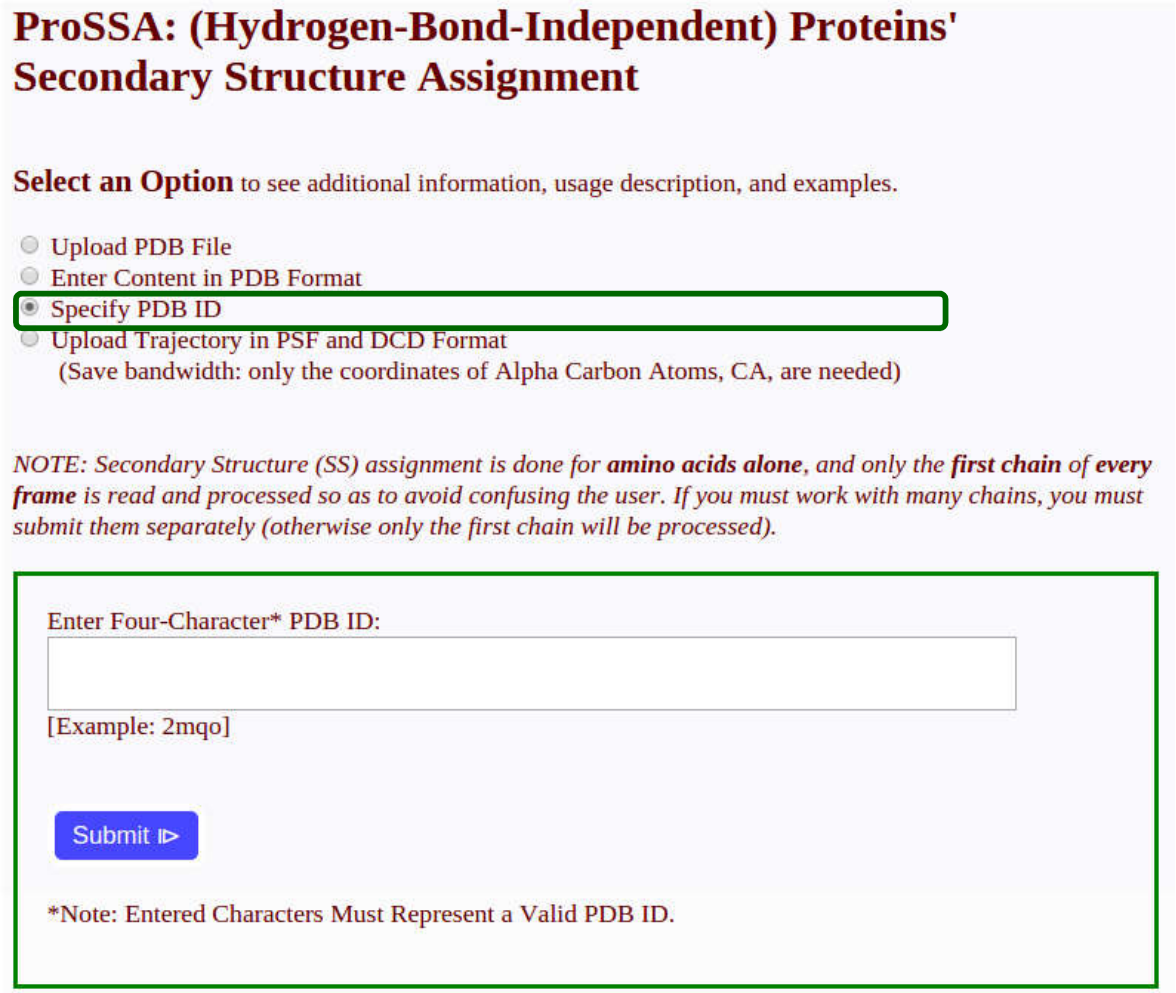
[Option B] If the user chooses to Enter Content in PDB Format, a form for entering the PDB Content is displayed such as in the screenshot below.

Users who do not have enough knowledge about PDB Format should generally avoid using this option. If necessary user may refer to this http://bioinformatics.center/blogs/1001/1002 short tutorial on PDB File Format.



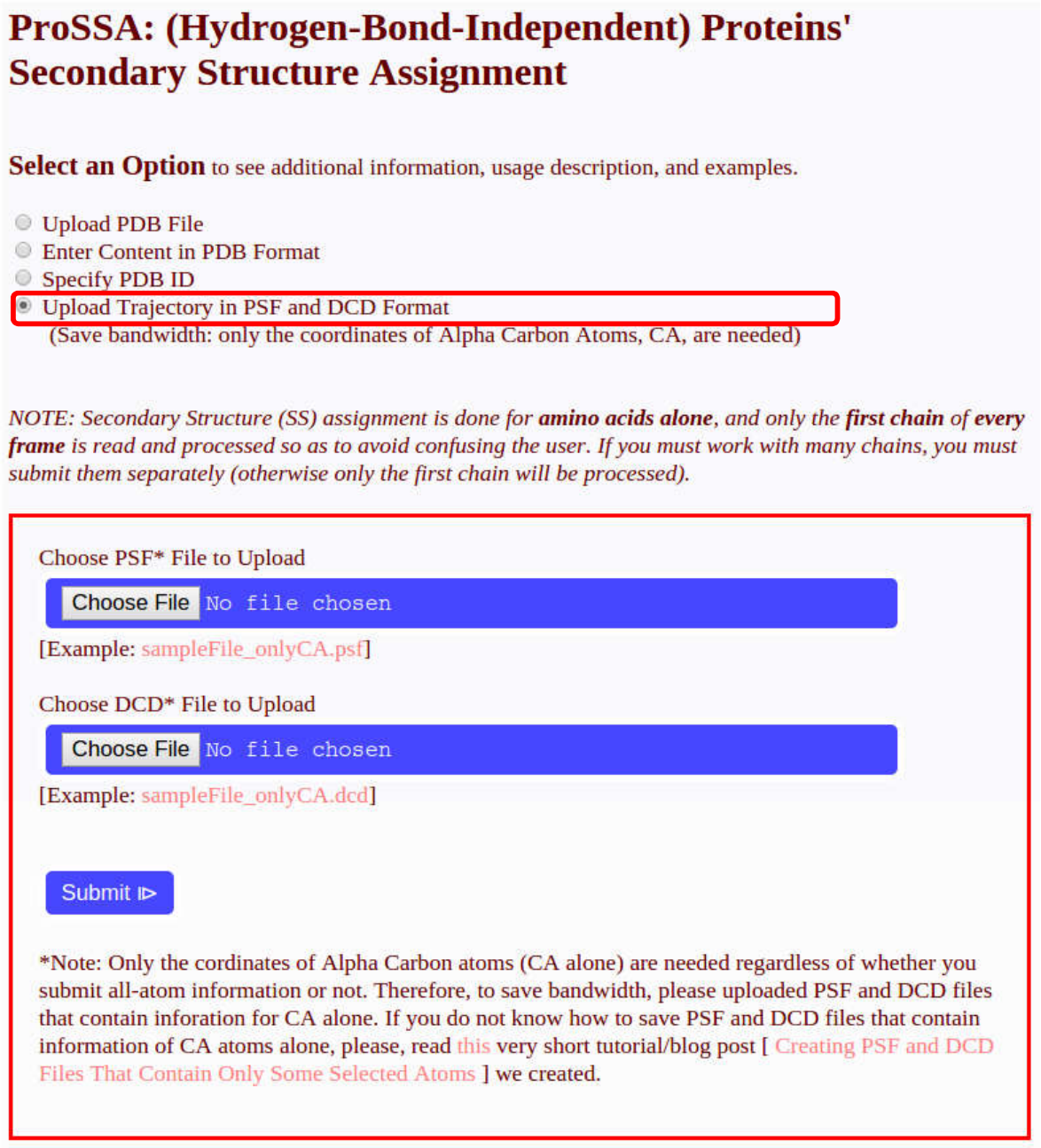
[Option C] If the user chooses to Specify PDB ID, a form for specifying the PDB ID is displayed such as in the screenshot below.

The PDB ID must be four characters in length and must be a valid identifier of a record in the Protein Data Bank. The record must contain protein. It does not matter whether the PDB record contains protein alone or protein alongside other biomolecules; but the record must contain protein.



[Option D] If the user chooses to provide input in PSF and DCD Format, a form for uploading the PSF and DCD Files is displayed such as in the screenshot below.

We encourage users to submit information of Alpha Carbon atoms alone so as to save bandwidth, especially when submitting multiple jobs with multiple frames/models. We provide a tutorial on how to create PSF, DCD, or PDB file that contains only Alpha Carbon atoms here http://bioinformatics.center/blogs/1001/1001



[3] Once you have selected an option in [2] above and clicked submit, you will be provide a link to the result page and the calculations will be start immediately. (The available computation resources are shared among ongoing computations. So, computations may take a moment.) Follow the link provided. If the computation is not ready, you will be told how much of additional time is needed. You will need to refresh the result page once the time has elapsed.

[4] Once the SS assignment is finished, you will see a result page similar to the one shown in the screenshots below (Part 1 to Part 3 of Results).

[Part 1 of Results] This part of the results is always available. It shows the job's ID, and the assigned secondary structure (in both 3-class and 7-class systems). If you submitted more than one frame/model, the results for all the frames are shown. The results for a frame span a line.

ProSSA: (Hydrogen-Bond-Independent) Proteins' Secondary Structure Assignment

Result Page

Job ID: 568a5a98g

Assigned Secondary Structure, 3-Class* System [Download]

Frames|Secondary Structures* Assigned to the Residues

```
1  ccsssssscccccssssscchhhhhhhhhhhhhhcccccsssssccssssc
2  cccsscccccccsssscccccchhhccchhhhcccccccccccccccccccc
3  ccccccccchhhccccssssscchhhhhhcccccccccccccccccsssccc
4  cccccccccccccccccssccchhhhhcccccccccccccccssccccccccc
5  ccccccccccccccccsssccccccccccccccccccccccccccccccccc
6  ccccccccchhhhcccccccccccchhhhcccccccccccccccccccccc
7  ccccccccchhhhhcccccccccchhhhhcccchhccccccccccccccc
8  ccccccccchhhhcccccccccccchhhcccccccccccccccccccccc
9  cccccccccccccccccccccccchhhhhhcccccccccccccccccccc
10 ccccccccchhhhcccccccccccchhhhhhcccccccccccccccccccc
11 cchhhccchhhcccccccccccchhhhhhhcccccccccccccccccccc
12 ccccccccchhhhcccccccccccchhhhhhhcccccccccccccccccc
```

* c = coil; h = helix; s = sheet

Assigned Secondary Structure, 7-Class** System [Download]

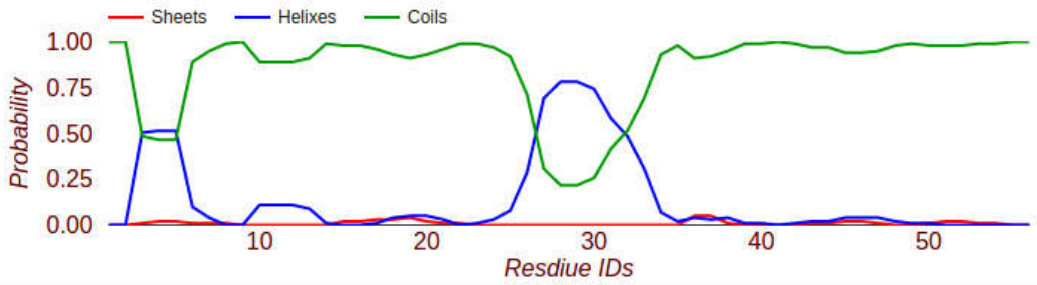
Frames|Secondary Structures** Assigned to the Residues

```
1  --EEEEEE-SS--SEEEEE-SSHHHHHHHHHHHHHHTT----EEEEETTTEEEE--
2  ---EET---TT-SSEEE---S-STTHHHTTSHHHHSSS-SS----TSTT-------
3  ----ST-SSHHHSSS--EEEEESSHHHHHHS--TTTTT--------STTEEE----
4  ----S--SSTT-SS----EESSSHHHHHTS--S-SSST-TS---EESTT-------
5  -----SS-STT--S--EEESTTSTTGGGT--TT-SSSS-TS--ST-TTTTSS-S--
6  --TTTS--SHHHHT----SSS-SSSTHHHH-S--SSSSS-S--S--STTSSSSS--
7  --STTT--SHHHHH-----S--SSSHHHHH--S-SHHHTT---S---TT-ST----
8  ---TTT--SHHHHS--------SST--HHHTTTTSS-TT-S--S--TTTTSS----
9  ----TT--TTTTTS--------SSS--HHHHHHTTS-TT-S--S-TSTT---SS--
10 --TTTTT--HHHHTS-------TTTTHHHHHHTTT--TTTT--TT--TS-SSS---
11 --HHHTT-SHHHTS--------SSSHHHHHHHTTT--TTTT--TT-TTS--SS---
12 --GGGTT--HHHHTS-----S-SSSTHHHHHHHT---TTTT--TT-STTSSTS---
```

** H = alpha helix; B = beta bridge; E = strand; G = helix-3;
I = helix-5; T = turn; S = bend; - = unknown

[Part 2 of Results] This part of the results is available only if more than one frame was submitted. This Part 2 of Results shows the probability that a residue at a position in the protein is part of a sheet, helix, or coil.

Distribution of the Assigned Secondary Structures, 3-Class System

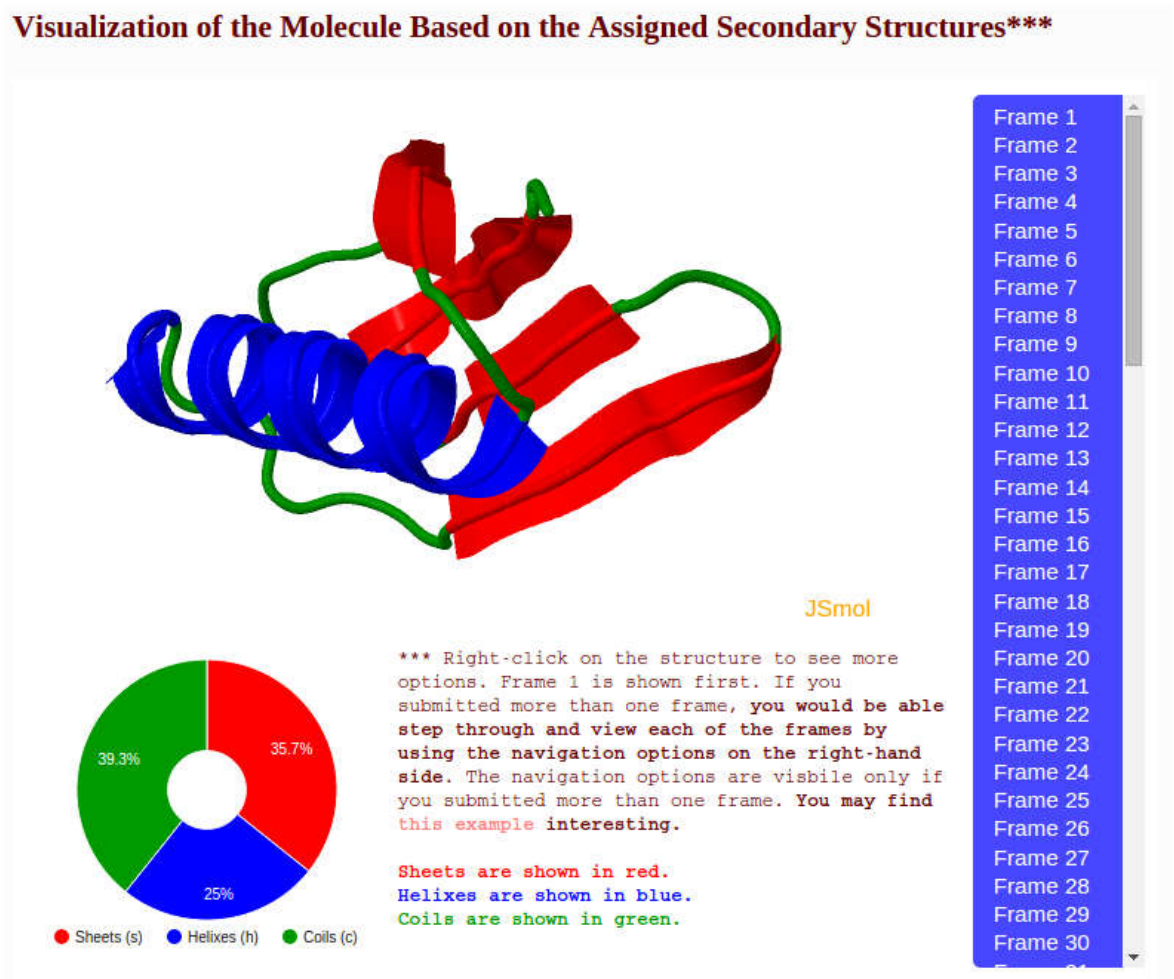


Position the mouse pointer on the chart to see each residue name, residue ID, frequency of sheets/helixes/coils [F(Sheet)/F(Helixes)/F(Coils)] for each residue, and probability of sheets/helixes/coils [P(Sheet)/P(Helixes)/P(Coils)] for each residue.

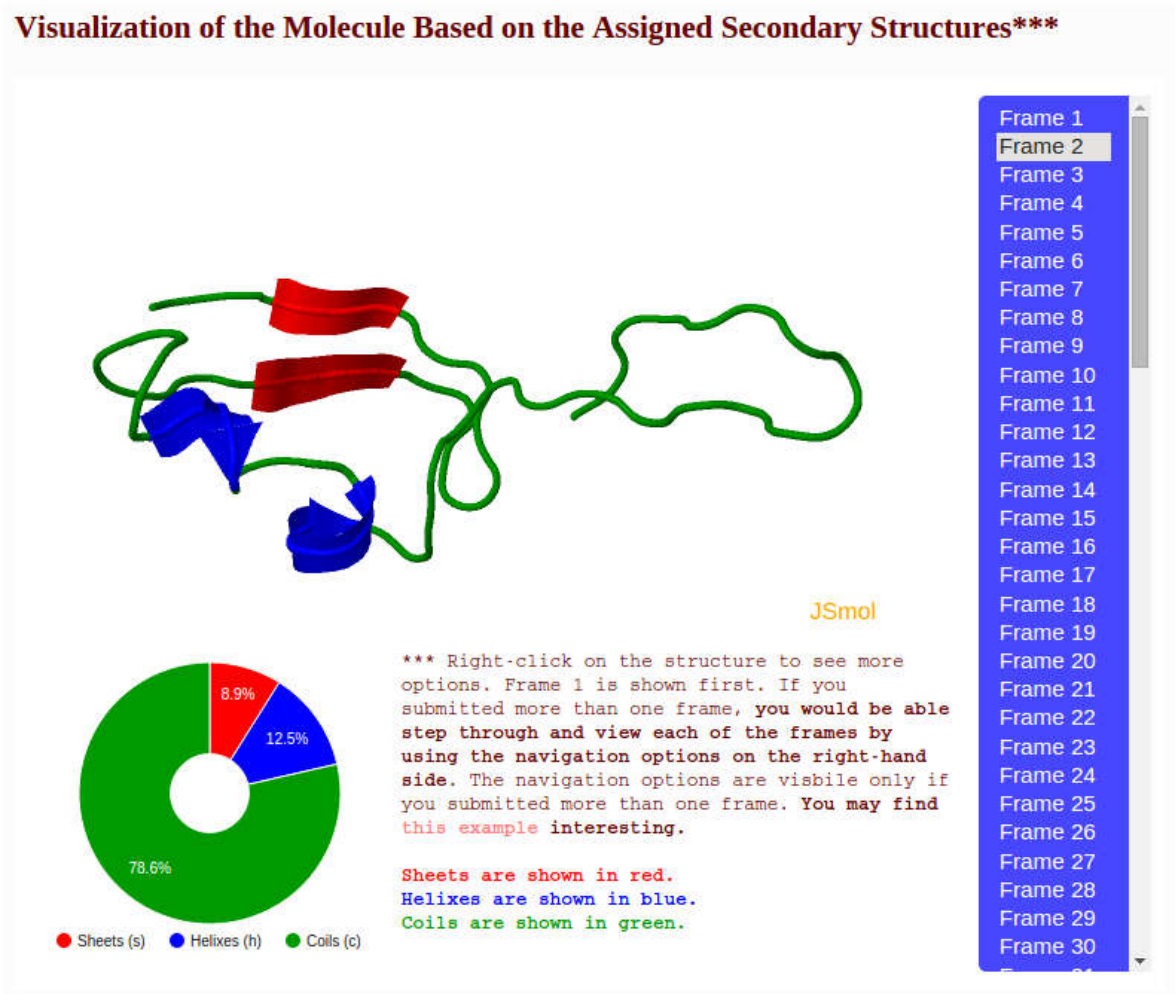
[Part 3 of Results] This part of the results is also always available, except the blue panel on the right-hand side which is available only if more than one frame/model was submitted. This part of the results shows the 3-D structure of the protein and the assigned secondary structure. Sheets are shown in red, helixes in blue, and coils in green. This part also shows the distribution of sheets, helixes, and coils in the protein based on the number of residues assigned to each of these three secondary structure types. This Part 3 of Results also allows the user to step through and visualized all the frames (if more than one frame was submitted). For example, we show the first three frames from job "568a5a98g" in the screenshots below.

[Frame 1 of Job 568a5a98g] For all jobs submitted to ProSSA, the results for Frame 1 is shown first. The screenshot below is from job of Job 568a5a98g available here http://bioinformatics.center/servers/ProSSA/results/568a5a98g#ssVisualization . The use can any frame he/she wants to show using the blue panel shown on the right-hand side.

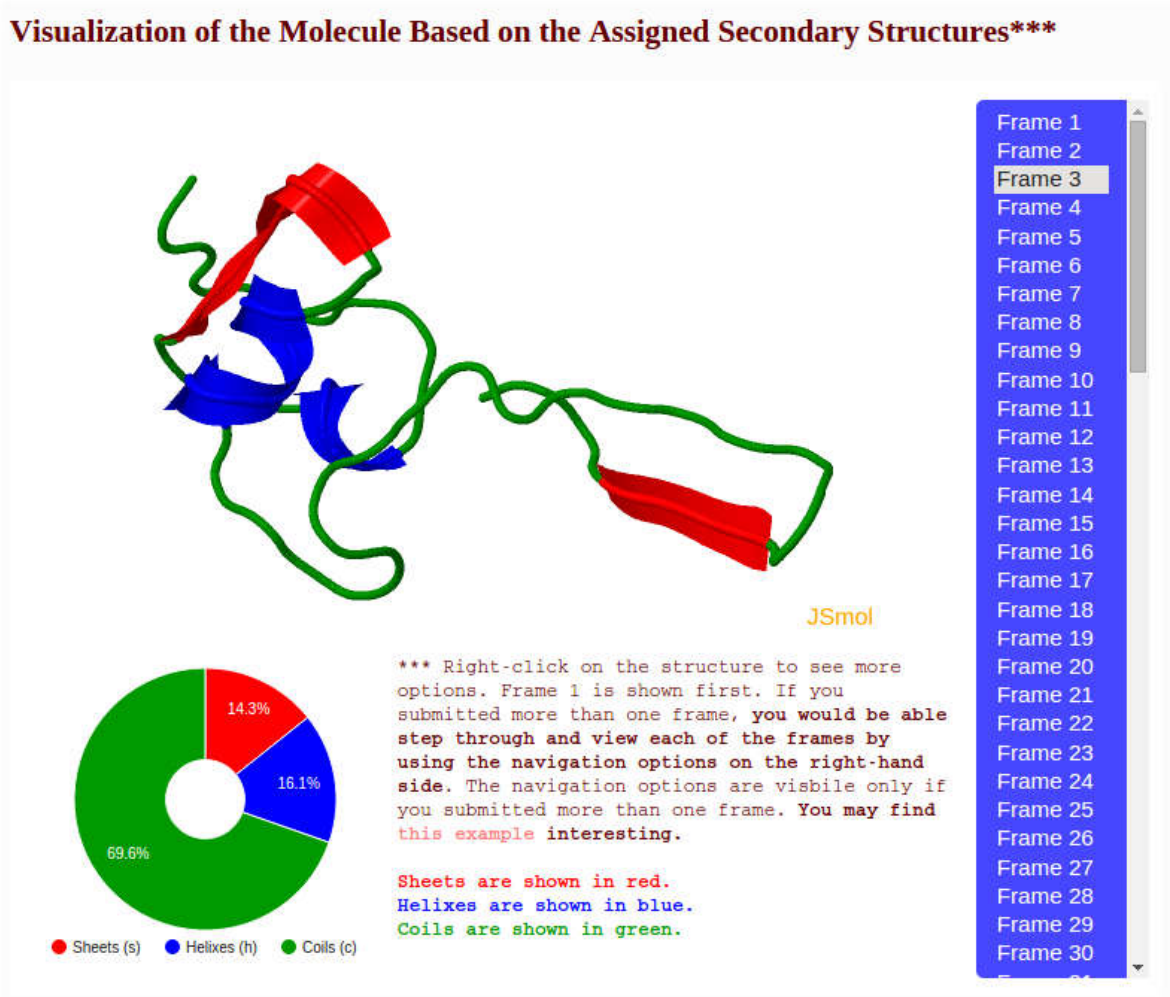
NOTE: The blue panel on the right-hand side is available only if more than one frame/model was submitted.



[Frame 2 of Job 568a5a98g] You can select "Frame 2" in the blue panel shown on the right to see the visualization of the assigned secondary structures for the submitted frame/model 2.



[Frame 3 of Job 568a5a98g] You can select "Frame 3" in the blue panel shown on the right to see the visualization of the assigned secondary structures for the submitted frame/model 3.



All the other frames can also be accessed.
Please, go here http://bioinformatics.center/servers/ProSSA/results/568a5a98g#ssVisualization to see this specific example.

If you need additional information or additional help, please do not hesitate to contact us.

Email: emmanuel@gapp.nthu.edu.tw, es13acp@herts.ac.uk
Tel: +18133888836

http://bioinformatics.center/servers/ProSSA