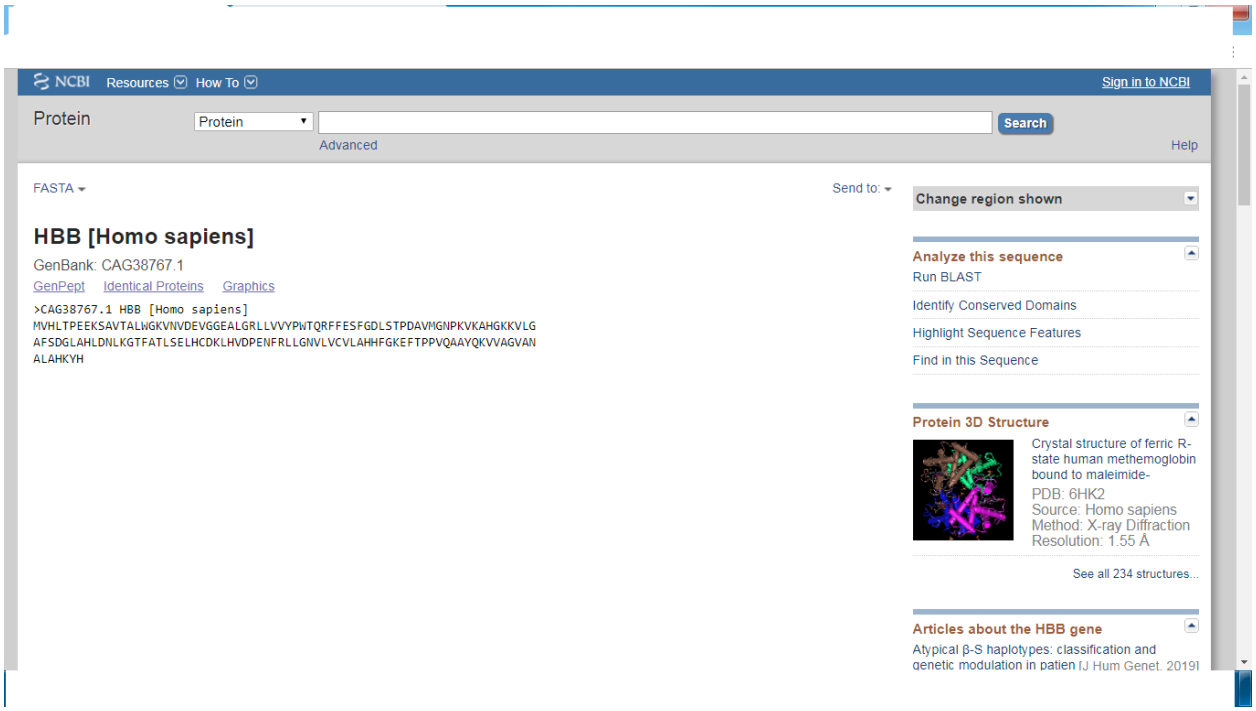
Bioinformatics-II (BIF501)

Assignment No.01

ZURAIZ BUTT

Question 1:

Select a protein of your choice and also write that why you selected that protein.



HBB [Homo sapiens]:

```
>CAG38767.1 HBB [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH
```

why you selected that protein:

The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta--3'.

Go to Protein databank server and download the structure of that protein:

I advice the students to copy whatever protein sequence they select and paste it on swiss model tool and download the PDB format:

Find any disease causing mutation in amino acid sequence of that particular protein. (You have to find it from literature):

β-S globin haplotype (β^S haplotype) characterization in sickle cell anemia (SCA) patients is important because it assists individualized treatment. However, the patient with atypical haplotypes do not present detailed studies such as clinical and laboratory data. To understand the phenotypic expression of atypical haplotype patients in relation to typical haplotype ones, it may be necessary to assess the main clinical and laboratorial parameters and investigate transcription factors, as possible genetic modulators that can contribute to the improvement of the SCA patients' clinical condition. The study group was composed of 600 SCA Brazilian patients of both genders ranging in age from 1 to 68 years. The atypical haplotypes were the third most frequent (5.7%) with 11 patterns numerically ranked according to occurrence. We verified that patients with atypical 1 haplotype in combination with Bantu haplotype presented milder clinical outcomes in relation to Bantu/Bantu and Benin/Benin patients, according to improved values of hemoglobin and hematocrit. In clinical severity, we did not observe significant statistical differences between typical and atypical haplotype patients, and this result can be explained with reference to the action of transcription factors in β-globin cluster. Thus, we presented the atypical haplotype relationship with SCA pathophysiology, reinforcing the hypothesis that individual genetic factors may be responsible for phenotypic diversity of the disease.

Manually add that mutation in the protein (amino acid) sequence by replacing the normal amino acid with mutant amino acidat that position and save a separate sequence file "Protein-name_mutation":

Mutant sequence crosspondind to Protein (Amino acid) S10A:

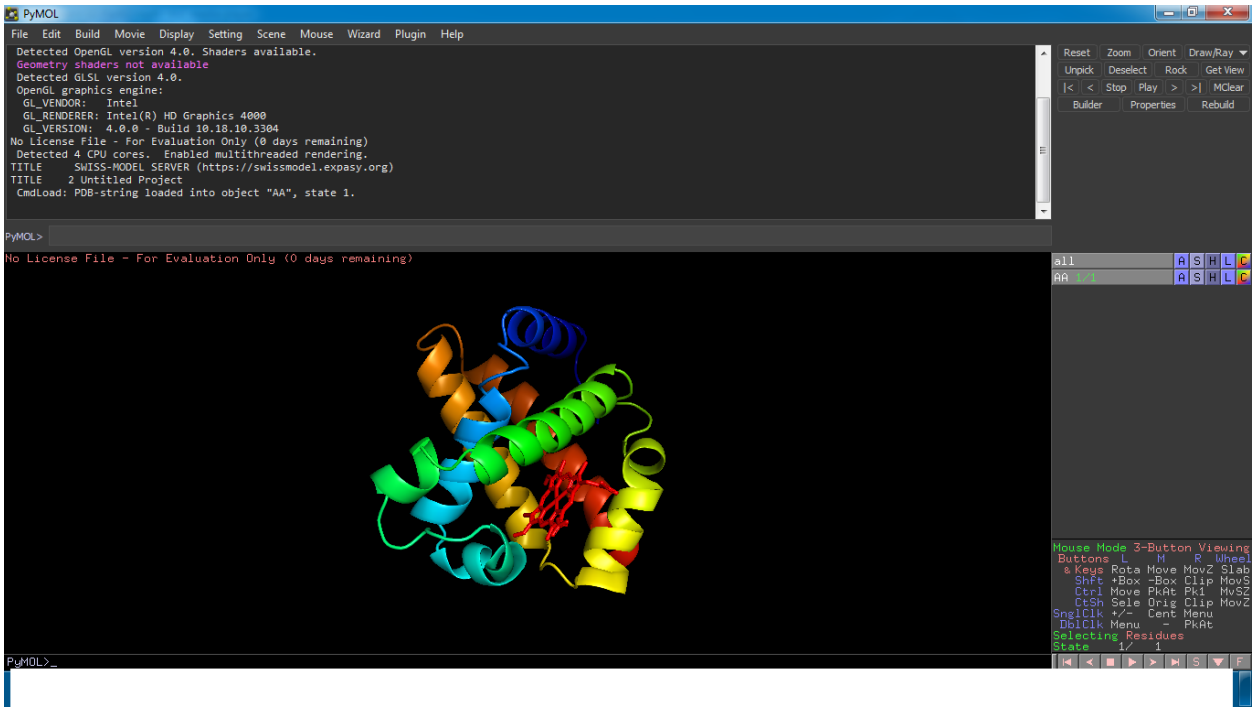
>CAG38767.1 HBB [Homo sapiens]
MVHLTPEEKAAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH

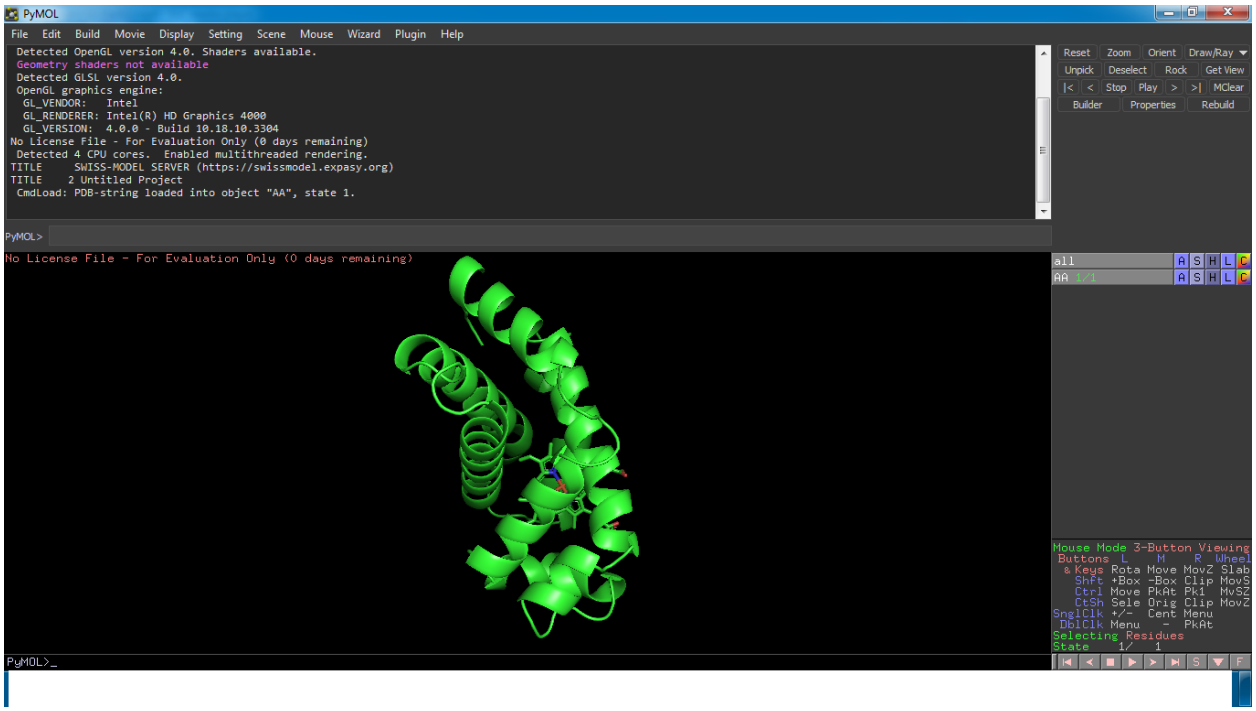
Go to FoldX server and build the mutant structure from mutant sequence. (FoldX is a Linux based tool so you have to install Linux OS on your system/PC/Laptop or you can ask your nearby VUP campuses to provide you the particular facility. Those who cannot use FoldX, are required to use Chimera/Pymol for this purpose:

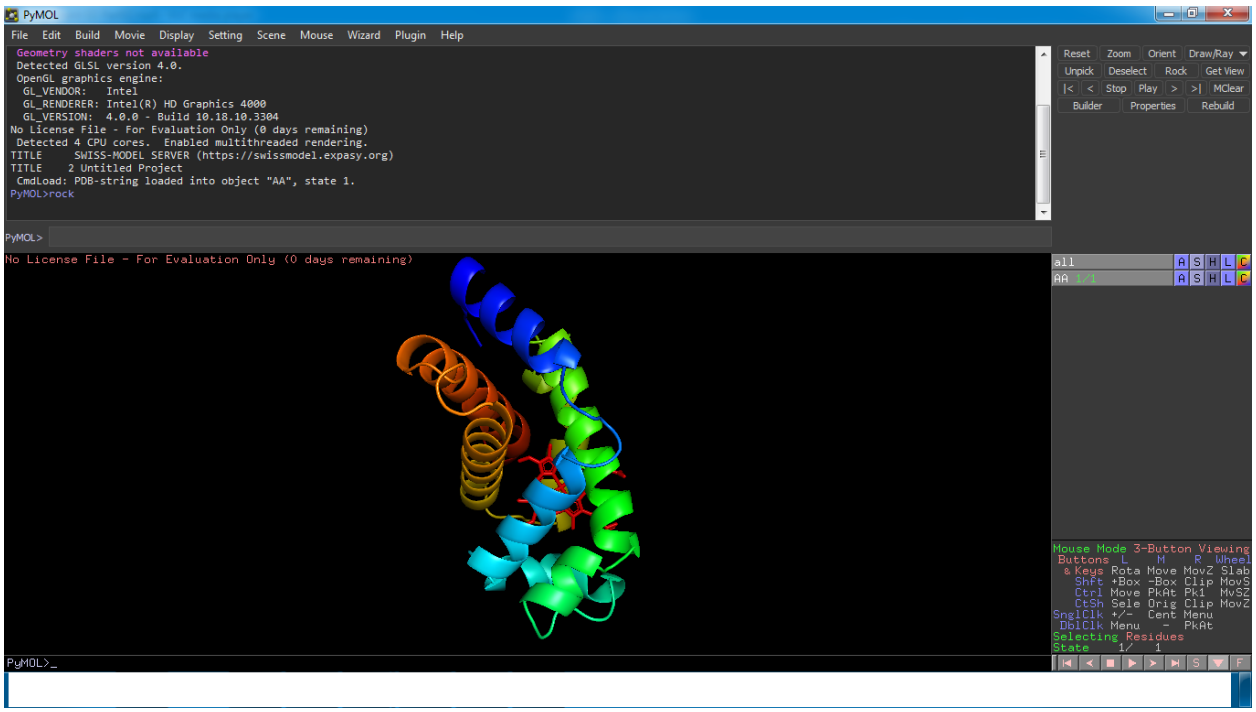
I use Pymol

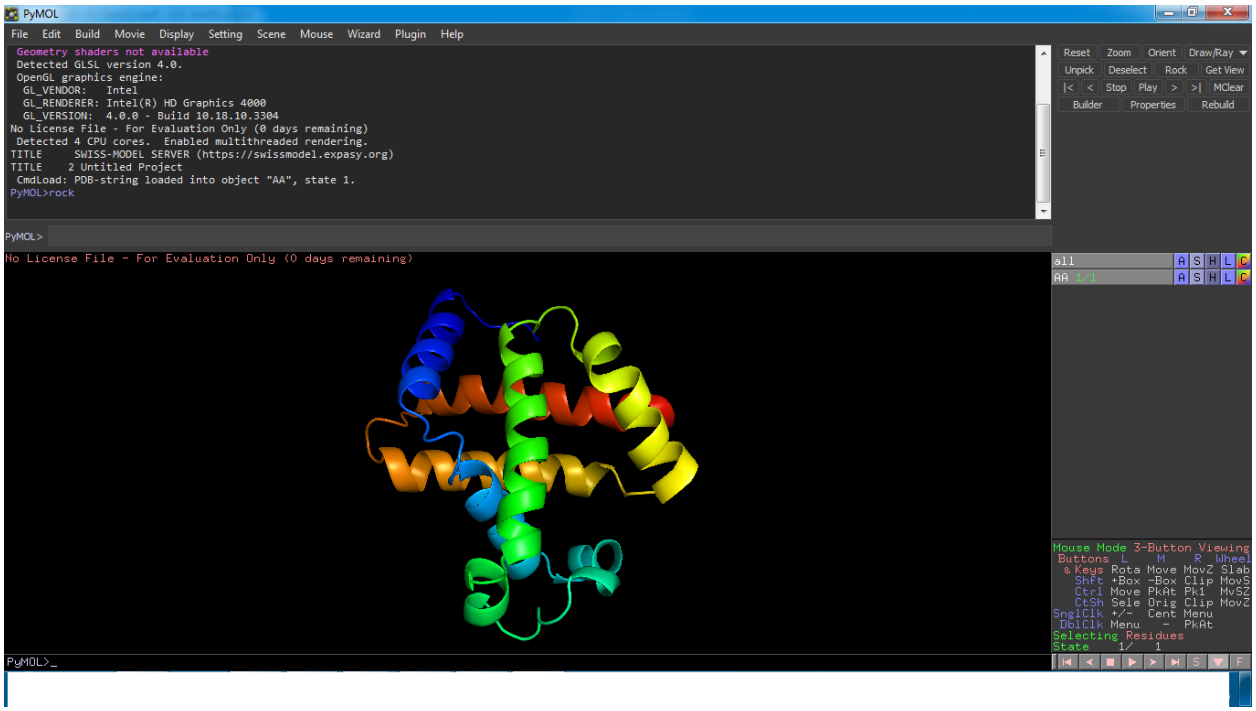
HBB [Homo sapiens]:

>CAG38767.1 HBB [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH




PyMOL
File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help
Detected OpenGL version 4.0. Shaders available.
Geometry shaders not available
Detected GLSL version 4.0.
OpenGL graphics engine:
GL_VENDOR: Intel
GL_RENDERER: Intel(R) HD Graphics 4000
GL_VERSION: 4.0.0 - Build 10.18.10.3304
No License File - For Evaluation Only (0 days remaining)
Detected 4 CPU cores. Enabled multithreaded rendering.
TITLE SWISS-MODEL SERVER (https://swissmodel.expasy.org)
TITLE 2 Untitled Project
CmdLoad: PDB-string loaded into object "AA", state 1.
Reset Zoom Orient Draw/Ray
Unpick Deselect Rock Get View
Stop Play MClear
Builder Properties Rebuild
PyMOL>
No License File - For Evaluation Only (0 days remaining)
all
AA 1/1
Mouse Mode 3-Button Viewing
Buttons L M R Wheel
Selecting Residues
State 1/ 1
PyMOL>_

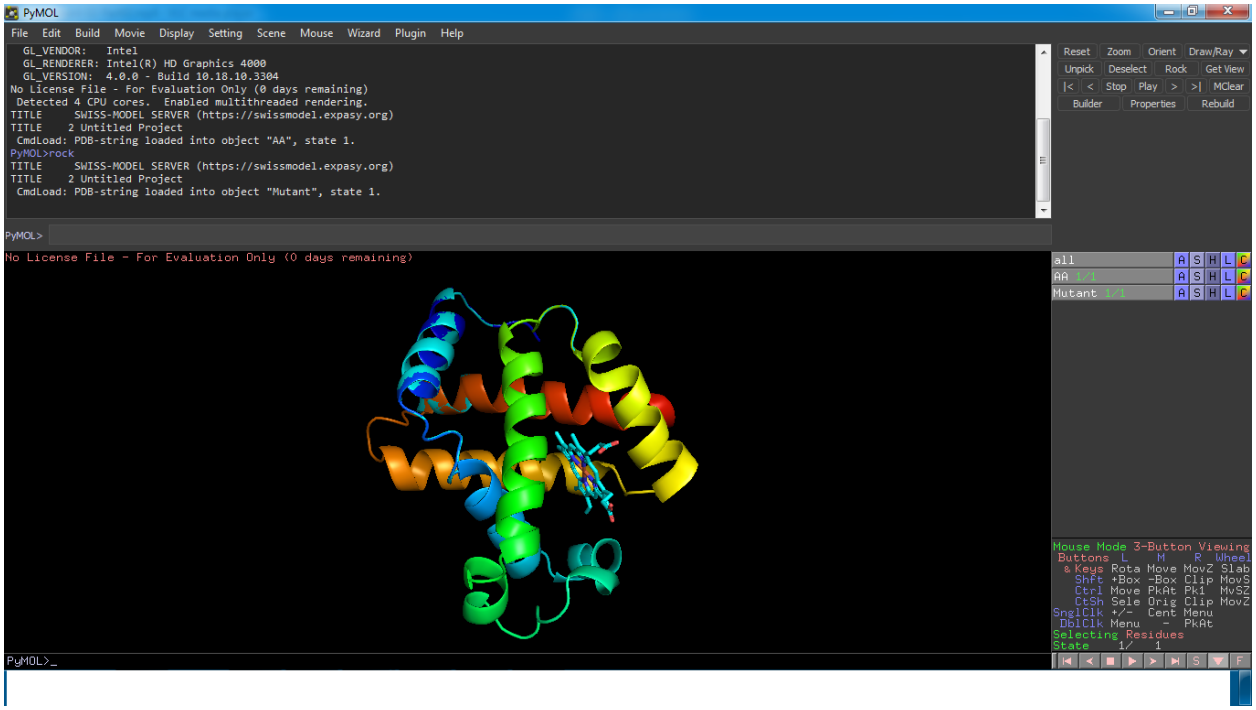

PyMOL
File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help
Geometry shaders not available
Detected GLSL version 4.0.
OpenGL graphics engine:
GL_VENDOR: Intel
GL_RENDERER: Intel(R) HD Graphics 4000
GL_VERSION: 4.0.0 - Build 10.18.10.3304
No License File - For Evaluation Only (0 days remaining)
Detected 4 CPU cores. Enabled multithreaded rendering.
TITLE SWISS-MODEL SERVER (https://swissmodel.expasy.org)
TITLE 2 Untitled Project
CmdLoad: PDB-string loaded into object "AA", state 1.
PyMOL>rock
Reset Zoom Orient Draw/Ray
Unpick Deselect Rock Get View
Stop Play MClear
Builder Properties Rebuild
PyMOL>
No License File - For Evaluation Only (0 days remaining)
all
AA 1/1
Mouse Mode 3-Button Viewing
Buttons L M R Wheel
Selecting Residues
State 1/ 1
PyMOL>_

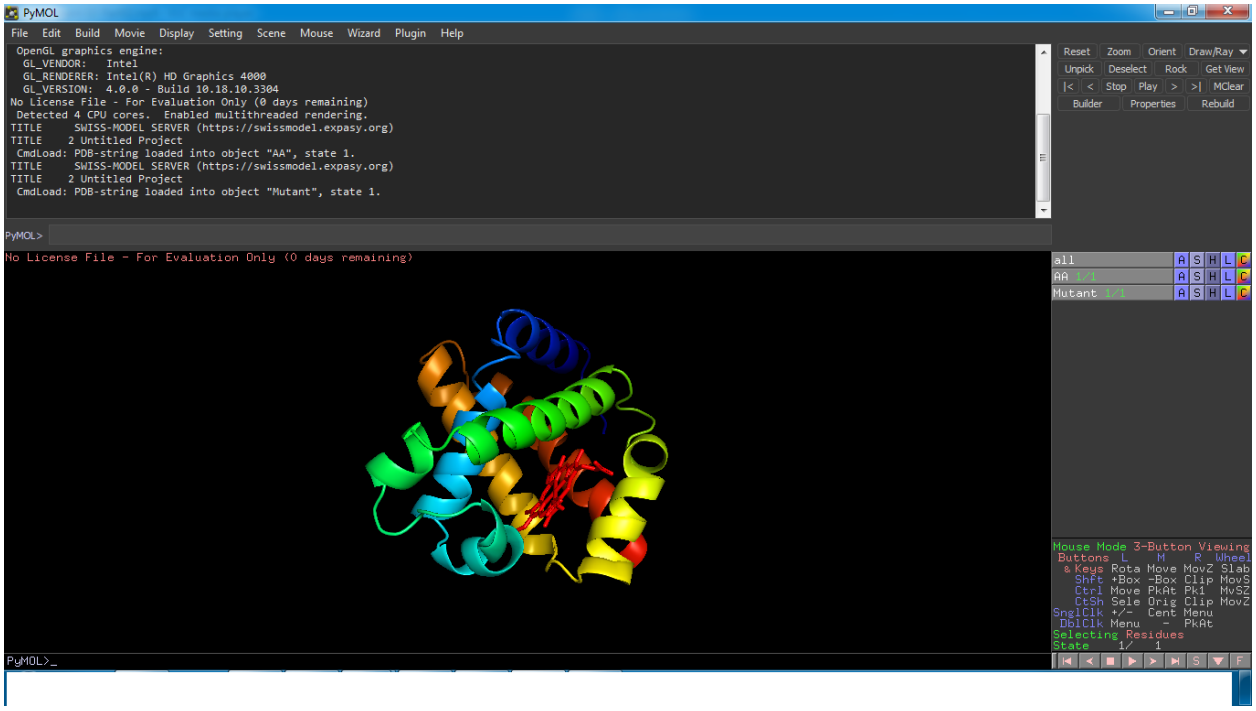

PyMOL
File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help
Geometry shaders not available
Detected GLSL version 4.0.
OpenGL graphics engine:
GL_VENDOR: Intel
GL_RENDERER: Intel(R) HD Graphics 4000
GL_VERSION: 4.0.0 - Build 10.18.10.3304
No License File - For Evaluation Only (0 days remaining)
Detected 4 CPU cores. Enabled multithreaded rendering.
TITLE SWISS-MODEL SERVER (https://swissmodel.expasy.org)
TITLE 2 Untitled Project
CmdLoad: PDB-string loaded into object "AA", state 1.
PyMOL>rock
Reset Zoom Orient Draw/Ray
Unpick Deselect Rock Get View
Stop Play MClear
Builder Properties Rebuild
PyMOL>
No License File - For Evaluation Only (0 days remaining)
all
AA 1/1
Mouse Mode 3-Button Viewing
Buttons L M R Wheel
Keys Rota Move MovZ Slab
ShFt +Box -Box Clip MovS
Ctrl Move PkAt Pk1 MvSZ
CtSh Sele Orig Clip MovZ
SnglClk +/- Cent Menu
DblClk Menu - PkAt
Selecting Residues
State 1/ 1
PyMOL>_

Mutant sequence crosspondind to Amino acid S10A

Highlight the structural change at that particular position where you pointed the mutation:

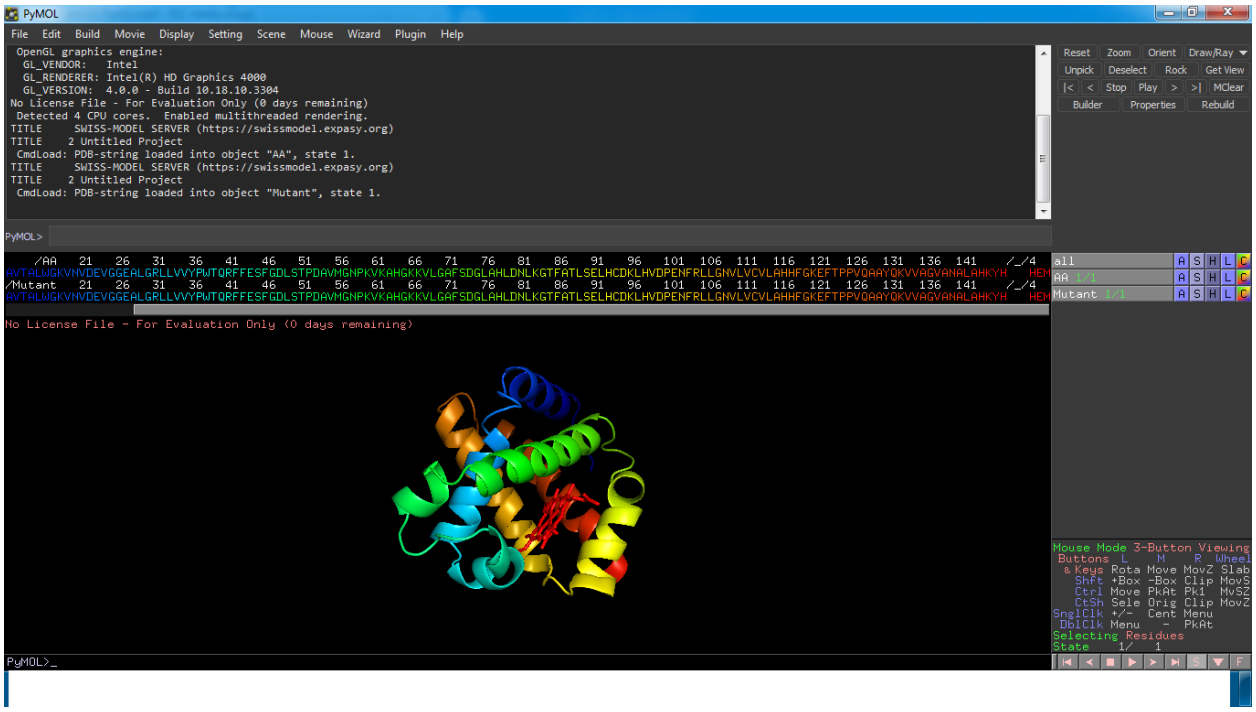
>CAG38767.1 HBB [Homo sapiens]
MVHLTPEEKAAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH

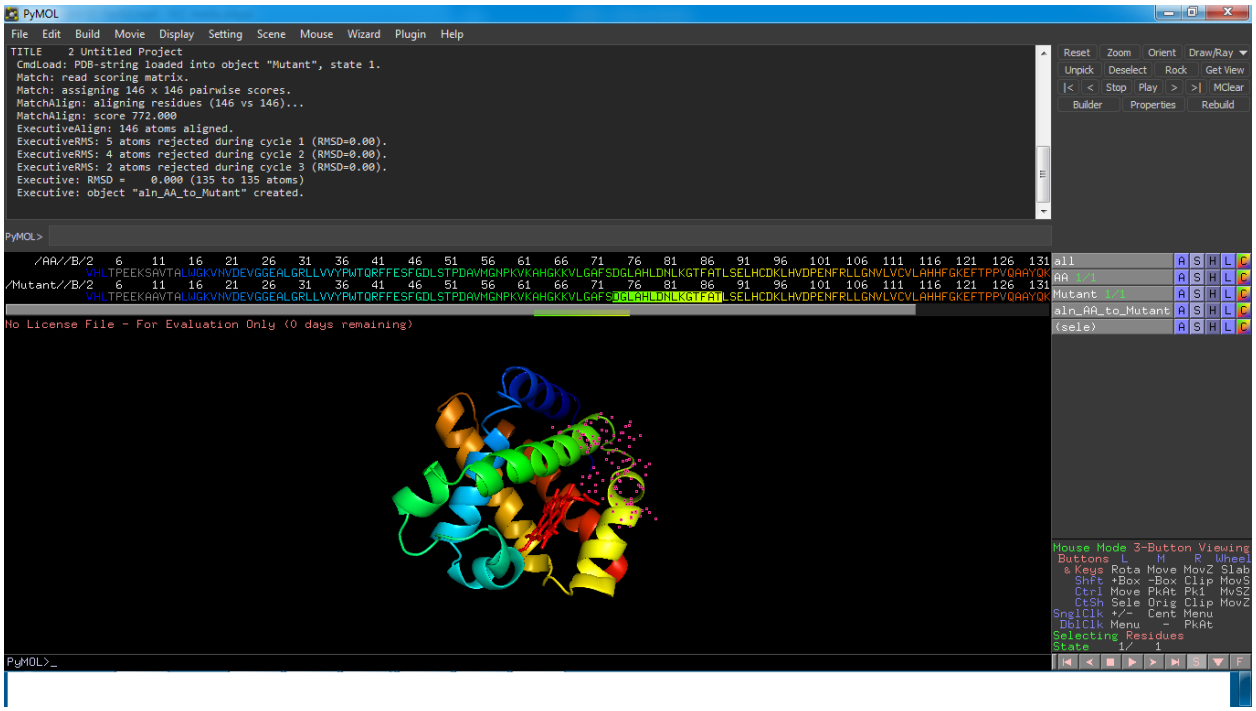




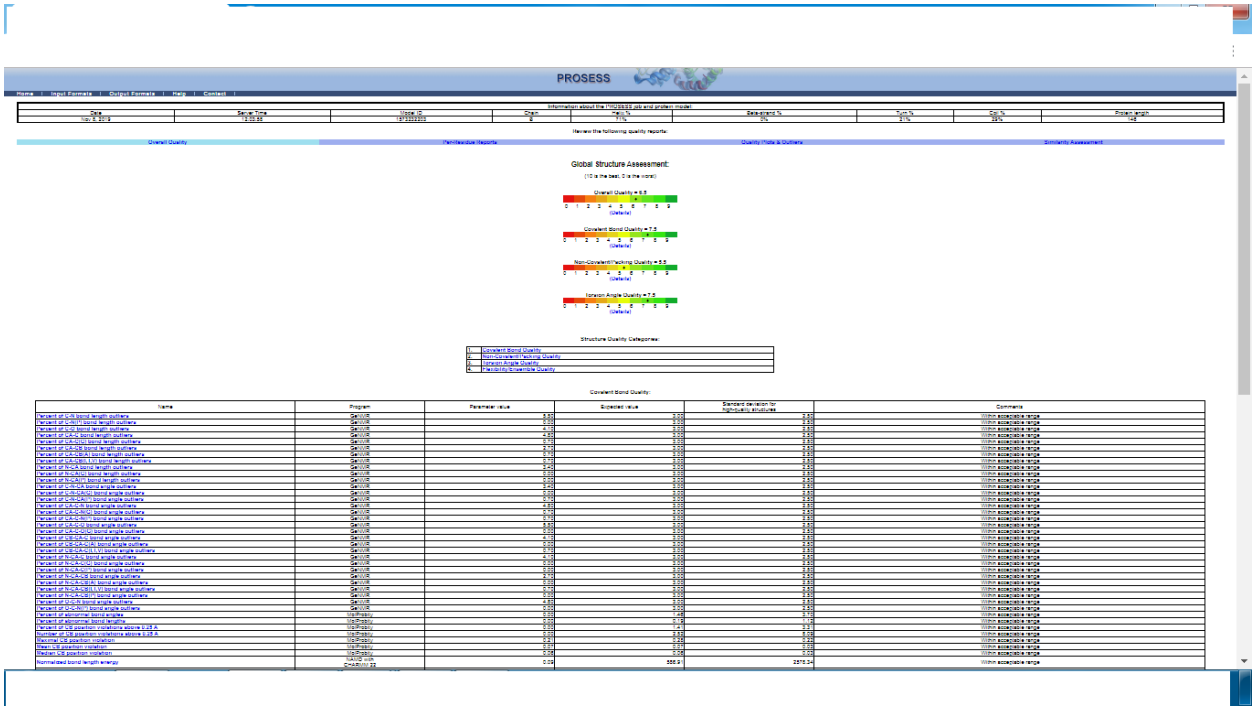
Refine both structures:

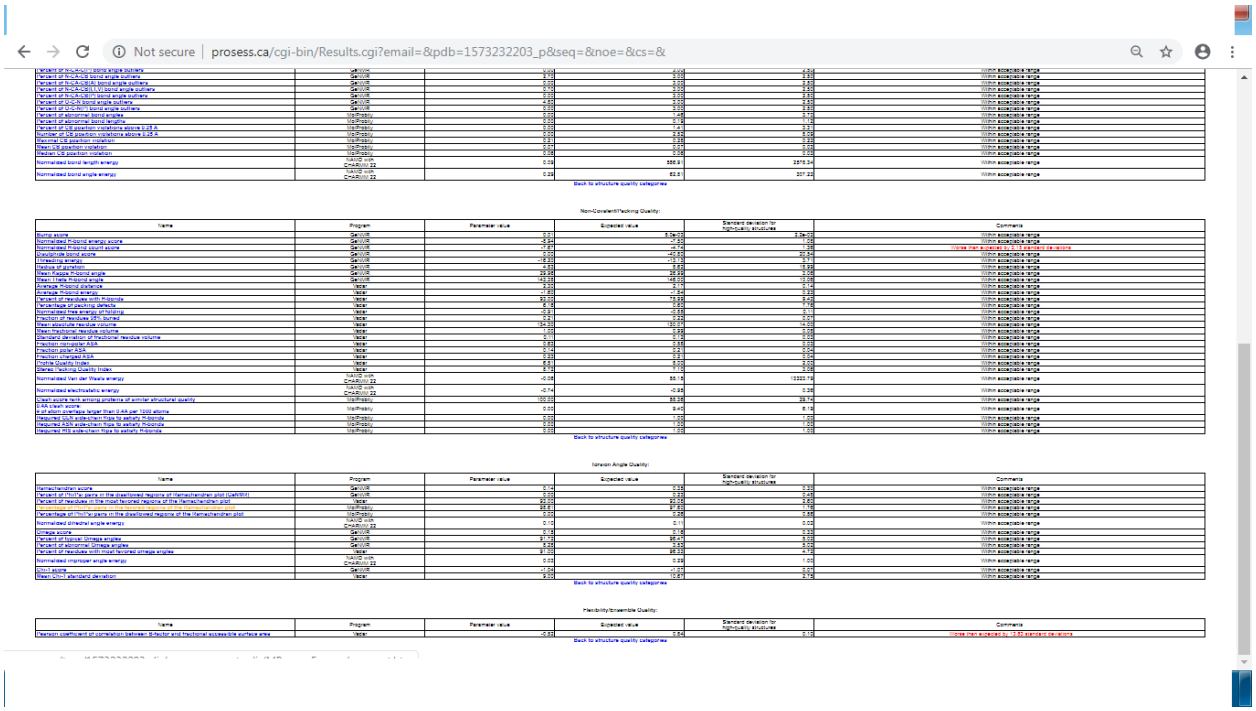
Highlight the structural change at that particular position where you pointed the mutation. This could be done using PyMol:





Go to PROSESS web-based tool (http://www.prosess.ca/) and evaluate the structure:





Global Structure Assessment:

(10 is the best, 0 is the worst)

Overall Quality = 6.5

(Details)

Covalent Bond Quality = 7.5

(Details)

Non-Covalent/Packing Quality = 5.5

(Details)

Torsion Angle Quality = 7.5

(Details)

Structure Quality Categories:

1.	Covalent Bond Quality
2.	Non-Covalent/Packing Quality
3.	Torsion Angle Quality
4.	Flexibility/Ensemble Quality

Covalent Bond Quality:

Name	Program	Parameter value	Expected value	Standard deviation for high-quality structures	Comments
Percent of C-N bond length outliers	GeNMR	5.50	3.00	2.50	Within acceptable range
Percent of C-N(P) bond length outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of C-O bond length outliers	GeNMR	4.10	3.00	2.50	Within acceptable range
Percent of CA-C bond length outliers	GeNMR	4.80	3.00	2.50	Within acceptable range
Percent of CA-C(G) bond length outliers	GeNMR	0.70	3.00	2.50	Within acceptable range
Percent of CA-CB bond length outliers	GeNMR	2.70	3.00	2.50	Within acceptable range
Percent of CA-CB(A) bond	GeNMR	0.70	3.00	2.50	Within acceptable range

length outliers					
Percent of CA-CB(I,T,V) bond length outliers	GeNMR	0.70	3.00	2.50	Within acceptable range
Percent of N-CA bond length outliers	GeNMR	3.40	3.00	2.50	Within acceptable range
Percent of N-CA(G) bond length outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of N-CA(P) bond length outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of C-N-CA bond angle outliers	GeNMR	3.40	3.00	2.50	Within acceptable range
Percent of C-N-CA(G) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of C-N-CA(P) bond angle outliers	GeNMR	0.70	3.00	2.50	Within acceptable range
Percent of CA-C-N bond angle outliers	GeNMR	4.80	3.00	2.50	Within acceptable range
Percent of CA-C-N(G) bond angle outliers	GeNMR	0.70	3.00	2.50	Within acceptable range
Percent of CA-C-N(P) bond angle outliers	GeNMR	0.70	3.00	2.50	Within acceptable range
Percent of CA-C-O bond angle outliers	GeNMR	5.50	3.00	2.50	Within acceptable range
Percent of CA-C-O(G) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of CB-CA-C bond angle outliers	GeNMR	4.10	3.00	2.50	Within acceptable range
Percent of CB-CA-C(A) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of CB-CA-C(I,T,V) bond angle outliers	GeNMR	0.70	3.00	2.50	Within acceptable range
Percent of N-CA-C bond angle outliers	GeNMR	4.10	3.00	2.50	Within acceptable range
Percent of N-CA-C(G) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of N-CA-C(P) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of N-CA-CB bond angle outliers	GeNMR	2.70	3.00	2.50	Within acceptable range
Percent of N-CA-CB(A) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of N-CA-CB(I,T,V) bond angle outliers	GeNMR	0.70	3.00	2.50	Within acceptable range

Percent of N-CA-CB(P) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of O-C-N bond angle outliers	GeNMR	4.80	3.00	2.50	Within acceptable range
Percent of O-C-N(P) bond angle outliers	GeNMR	0.00	3.00	2.50	Within acceptable range
Percent of abnormal bond angles	MolProbity	0.00	1.46	3.70	Within acceptable range
Percent of abnormal bond lengths	MolProbity	0.00	0.19	1.12	Within acceptable range
Percent of CB position violations above 0.25 A	MolProbity	0.00	1.41	3.31	Within acceptable range
Number of CB position violations above 0.25 A	MolProbity	0.00	2.52	8.09	Within acceptable range
Maximal CB position violation	MolProbity	0.21	0.28	0.22	Within acceptable range
Mean CB position violation	MolProbity	0.07	0.07	0.03	Within acceptable range
Median CB position violation	MolProbity	0.06	0.06	0.03	Within acceptable range
Normalized bond length energy	NAMD with CHARMM 22	0.09	586.91	2578.34	Within acceptable range
Normalized bond angle energy	NAMD with CHARMM 22	0.29	62.81	307.22	Within acceptable range

Back to structure quality categories

Non-Covalent/Packing Quality:

Name	Program	Parameter value	Expected value	Standard deviation for high-quality structures	Comments
Bump score	GeNMR	0.01	5.0e-03	2.2e-02	Within acceptable range
Normalized H-bond energy score	GeNMR	-8.94	-7.50	1.08	Within acceptable range
Normalized H-bond count score	GeNMR	-7.67	-4.74	1.36	Worse than expected by 2.15 standard deviations
Disulphide bond score	GeNMR	0.00	-40.80	20.54	Within acceptable range
Threading energy	GeNMR	-16.20	-13.13	3.71	Within acceptable range
Radius of gyration	GeNMR	4.83	8.62	18.99	Within acceptable range

Mean Kappa H-bond angle	GeNMR	29.96	26.99	3.06	Within acceptable range
Mean Theta H-bond angle	GeNMR	143.38	146.00	10.06	Within acceptable range
Average H-bond distance	Vadar	2.30	2.17	0.14	Within acceptable range
Average H-bond energy	Vadar	-1.60	-1.84	0.23	Within acceptable range
Percent of residues with H-bonds	Vadar	93.00	75.99	9.42	Within acceptable range
Percentage of packing defects	Vadar	6.16	0.60	7.76	Within acceptable range
Normalized free energy of folding	Vadar	-0.91	-0.85	0.11	Within acceptable range
Fraction of residues 95% buried	Vadar	0.21	0.22	0.07	Within acceptable range
Mean absolute residue volume	Vadar	134.30	130.07	14.00	Within acceptable range
Mean fractional residue volume	Vadar	1.00	0.99	0.08	Within acceptable range
Standard deviation of fractional residue volume	Vadar	0.11	0.12	0.03	Within acceptable range
Fraction non-polar ASA	Vadar	0.63	0.58	0.03	Within acceptable range
Fraction polar ASA	Vadar	0.14	0.21	0.04	Within acceptable range
Fraction charged ASA	Vadar	0.23	0.21	0.04	Within acceptable range
Profile Quality Index	Vadar	6.51	8.00	2.00	Within acceptable range
Stereo Packing Quality Index	Vadar	8.72	7.10	3.06	Within acceptable range
Normalized Van der Waals energy	NAMD with CHARMM 22	-0.06	55.15	13323.79	Within acceptable range
Normalized electrostatic energy	NAMD with CHARMM 22	-0.74	-0.95	0.36	Within acceptable range
Clash score rank among proteins of similar structural quality	MolProbity	100.00	58.36	29.74	Within acceptable range
0.4A clash score: # of atom overlaps larger than 0.4A per 1000 atoms	MolProbity	0.00	9.40	6.19	Within acceptable range
Required GLN side-chain flips to satisfy H-bonds	MolProbity	0.00	1.00	1.00	Within acceptable range
Required ASN side-chain flips to satisfy H-bonds	MolProbity	0.00	1.00	1.00	Within acceptable range
Required HIS side-chain flips to satisfy H-bonds	MolProbity	0.00	1.00	1.00	Within acceptable range

Back to structure quality categories

Torsion Angle Quality:

Name	Program	Parameter value	Expected value	Standard deviation for high-quality structures	Comments
Ramachandran score	GeNMR	0.14	0.35	0.30	Within acceptable range
Percent of Phi/Psi pairs in the disallowed regions of Ramachandran plot (GeNMR)	GeNMR	0.00	0.23	0.48	Within acceptable range
Percent of residues in the most favored regions of the Ramachandran plot	Vadar	93.00	93.08	2.60	Within acceptable range
Percentage of Phi/Psi pairs in the favored regions of the Ramachandran plot	MolProbity	98.61	97.60	1.76	Within acceptable range
Percentage of Phi/Psi pairs in the disallowed regions of the Ramachandran plot	MolProbity	0.00	0.26	0.56	Within acceptable range
Normalized dihedral angle energy	NAMD with CHARMM 22	0.10	0.11	0.02	Within acceptable range
Omega score	GeNMR	0.15	0.16	0.33	Within acceptable range
Percent of typical Omega angles	GeNMR	91.72	96.47	5.02	Within acceptable range
Percent of abnormal Omega angles	GeNMR	8.28	3.53	5.02	Within acceptable range
Percent of residues with most favored omega angles	Vadar	91.00	96.33	4.72	Within acceptable range
Normalized improper angle energy	NAMD with CHARMM 22	0.03	0.29	1.00	Within acceptable range
Chi-1 score	GeNMR	-1.04	-1.07	0.07	Within acceptable range
Mean Chi-1 standard deviation	Vadar	9.00	10.67	2.75	Within acceptable range

Back to structure quality categories

Flexibility/Ensemble Quality:

Name	Program	Parameter value	Expected value	Standard deviation for high-quality structures	Comments
Pearson coefficient of correlation between B-factor and fractional accessible surface area	Vadar	-0.52	0.84	0.10	Worse than expected by 13.63 standard deviations