wwPDB NMR Structure Validation Summary Report (i)

Jun 6, 2023 – 02:03 pm BST

PDB ID	:	5OGU
BMRB ID	:	34163
Title	:	Structure of DNA-binding HU protein from micoplasma Spiroplasma melliferum
Authors	:	Altukhov, D.A.; Talyzina, A.A.; Agapova, Y.K.; Vlaskina, A.V.; Korzhenevskiy, D.A.; Bocharov, E.V.; Rakitina, T.V.; Timofeev, V.I.
Deposited on	:	2017-07-13

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

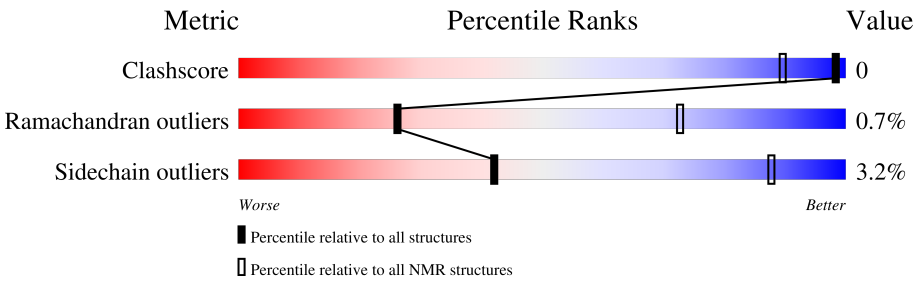
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 12%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	95	77% • 22%
1	B	95	91% 7% •

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:-1-A:55, A:77-A:93, B:-1-B:55, B:77-B:93 (148)	0.79	4
2	B:58-B:76 (19)	0.68	6

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7
2	8, 9, 10, 11, 12, 13, 14
Single-model clusters	15

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2948 atoms, of which 1500 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA-binding protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	95	1474	456	750	130	137	1	0
1	B	95	1474	456	750	130	137	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP A0A037USE5
A	0	HIS	-	expression tag	UNP A0A037USE5
B	-1	GLY	-	expression tag	UNP A0A037USE5
B	0	HIS	-	expression tag	UNP A0A037USE5

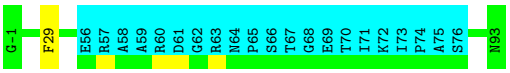
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

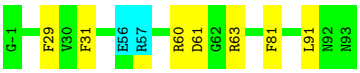
• Molecule 1: DNA-binding protein

Chain A: 77% • 22%



• Molecule 1: DNA-binding protein

Chain B: 91% 7% •

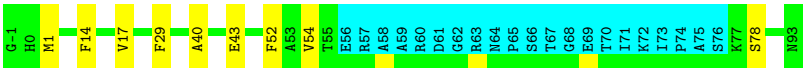


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

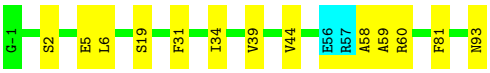
• Molecule 1: DNA-binding protein

Chain A: 68% 9% 22%



• Molecule 1: DNA-binding protein

Chain B: 84% 14% •



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 5000 calculated structures, 15 were deposited, based on the following criterion: *structures with the least restraint violations*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
GROMACS	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	320
Number of shifts mapped to atoms	320
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	12%

6 Model quality (i)

6.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.50±0.01	0±0/577 (0.0± 0.0%)	1.79±0.08	8±3/767 (1.1± 0.3%)
1	B	0.51±0.01	0±0/714 (0.0± 0.0%)	1.86±0.07	14±3/953 (1.4± 0.3%)
All	All	0.51	0/19365 (0.0%)	1.83	328/25800 (1.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	B	0.0±0.0	0.7±1.1
1	A	0.0±0.0	0.6±0.9
All	All	0	20

There are no bond-length outliers.

5 of 192 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	63	ARG	NE-CZ-NH2	15.75	128.17	120.30	6	4
1	B	63	ARG	NE-CZ-NH1	14.54	127.57	120.30	13	5
1	B	31	PHE	CB-CG-CD2	-13.42	111.41	120.80	8	5
1	B	60	ARG	NE-CZ-NH1	12.47	126.54	120.30	11	6
1	B	60	ARG	NE-CZ-NH2	-12.41	114.09	120.30	11	3

There are no chirality outliers.

5 of 16 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	14	PHE	Sidechain	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	B	63	ARG	Sidechain	2
1	A	81	PHE	Sidechain	2
1	B	81	PHE	Sidechain	1
1	B	67	THR	Mainchain	1

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	B	704	731	731	0±0
1	A	569	595	595	0±0
All	All	19095	19890	19890	6

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

5 of 6 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:34:ILE:HG23	1:B:44:VAL:HG11	0.56	1.78	4	1
1:B:88:LYS:HG3	1:B:89:THR:H	0.55	1.61	1	1
1:A:33:HIS:CE1	1:B:1:MET:CE	0.47	2.97	13	1
1:A:32:ASP:OD2	1:A:36:LYS:NZ	0.47	2.44	7	1
1:A:31:PHE:CD1	1:A:31:PHE:N	0.43	2.87	5	1

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	72/95 (76%)	66±2 (92±3%)	5±2 (7±3%)	0±1 (1±1%)	29 74

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	91/95 (96%)	82±3 (90±3%)	8±2 (9±2%)	1±1 (1±1%)	26	73
All	All	2445/2850 (86%)	2229 (91%)	200 (8%)	16 (1%)	26	73

5 of 12 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	47	ALA	4
1	B	85	LYS	2
1	A	20	LYS	1
1	B	83	ALA	1
1	A	29	PHE	1

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	59/75 (79%)	57±1 (97±2%)	2±1 (3±2%)	49	91
1	B	73/75 (97%)	70±1 (96±2%)	3±1 (4±2%)	37	85
All	All	1980/2250 (88%)	1917 (97%)	63 (3%)	42	88

5 of 33 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	29	PHE	7
1	B	61	ASP	6
1	B	91	LEU	5
1	A	16	ASP	3
1	A	29	PHE	3

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 12% for the well-defined parts and 13% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Spm_BMRB.tab*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	320
Number of shifts mapped to atoms	320
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	78	-0.05 ± 0.24	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	41	0.55 ± 0.20	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	78	0.40 ± 0.32	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 12%, i.e. 261 atoms were assigned a chemical shift out of a possible 2232. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	226/843 (27%)	104/344 (30%)	61/334 (18%)	61/165 (37%)
Sidechain	35/1227 (3%)	0/797 (0%)	35/386 (9%)	0/44 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/162 (0%)	0/84 (0%)	0/72 (0%)	0/6 (0%)
Overall	261/2232 (12%)	104/1225 (8%)	96/792 (12%)	61/215 (28%)

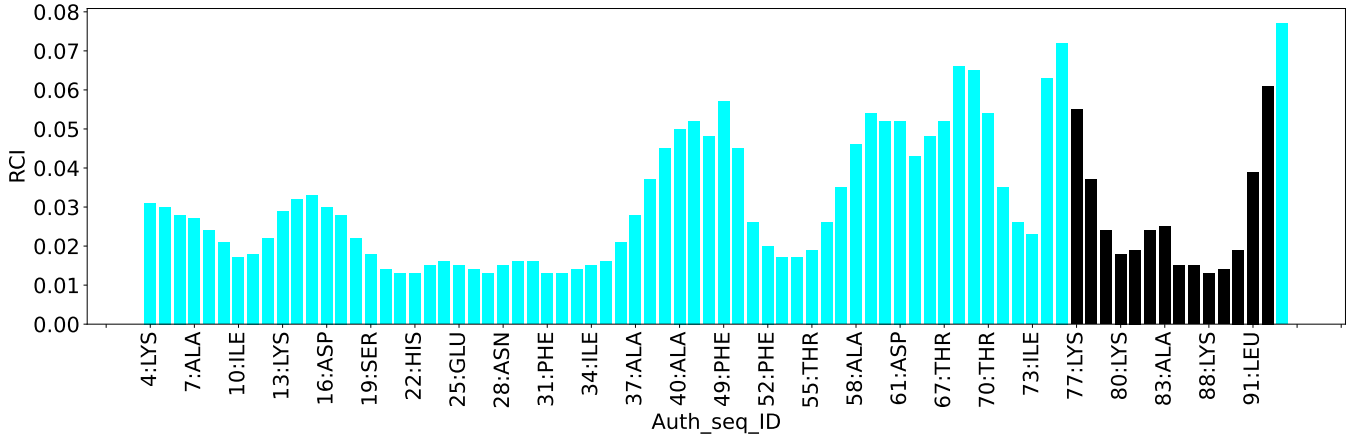
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



8 NMR restraints analysis (i)

8.1 Conformationally restricting restraints (i)

The following table provides the summary of experimentally observed NMR restraints in different categories. Restraints are classified into different categories based on the sequence separation of the atoms involved.

Description	Value
Total distance restraints	0
Intra-residue (|i-j|=0)	0
Sequential (|i-j|=1)	0
Medium range (|i-j|>1 and |i-j|<5)	0
Long range (|i-j|≥5)	0
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	255
Number of unmapped restraints	0
Number of restraints per residue	0
Number of long range restraints per residue[1]	0

[1]Long range hydrogen bonds and disulfide bonds are counted as long range restraints while calculating the number of long range restraints per residue

8.2 Residual restraint violations (i)

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

8.2.1 Average number of distance violations per model (i)

Distance violations less than 0.1 Å are not included in the calculation.There are no distance restraints

8.2.2 Average number of dihedral-angle violations per model (i)

Dihedral-angle violations less than 1° are not included in the calculation.

Bins (°)	Average number of violations per model	Max (°)
1.0-10.0 (Small)	54.9	10.0

Continued on next page...

Continued from previous page...

Bins (°)	Average number of violations per model	Max (°)
10.0-20.0 (Medium)	35.5	20.0
>20.0 (Large)	65.0	172.8

9 Distance violation analysis (i)

No distance restraints data found

10 Dihedral-angle violation analysis (i)

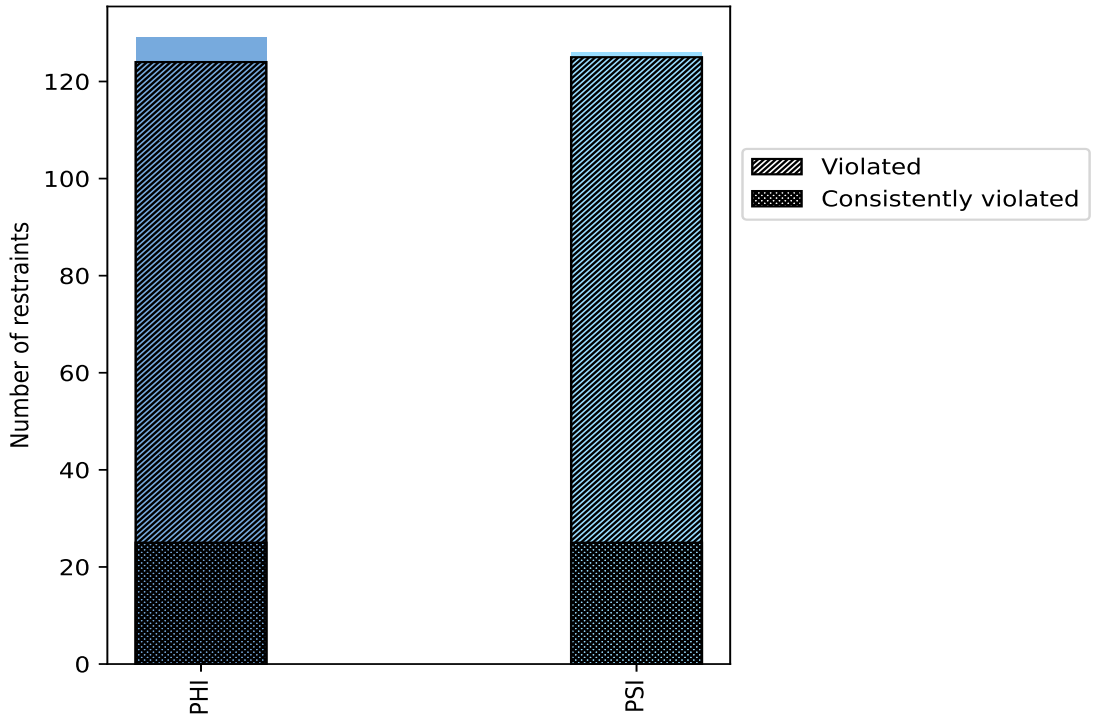
10.1 Summary of dihedral-angle violations (i)

The following table provides the summary of dihedral-angle violations in different dihedral-angle types. Violations less than 1° are not included in the calculation.

Angle type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
PHI	129	50.6	124	96.1	48.6	25	19.4	9.8
PSI	126	49.4	125	99.2	49.0	25	19.8	9.8
Total	255	100.0	249	97.6	97.6	50	19.6	19.6

[1] percentage calculated with respect to total number of dihedral-angle restraints, [2] percentage calculated with respect to number of restraints in a particular dihedral-angle type, [3] violated in at least one model, [4] violated in all the models

10.1.1 Bar chart : Distribution of dihedral-angles and violations (i)



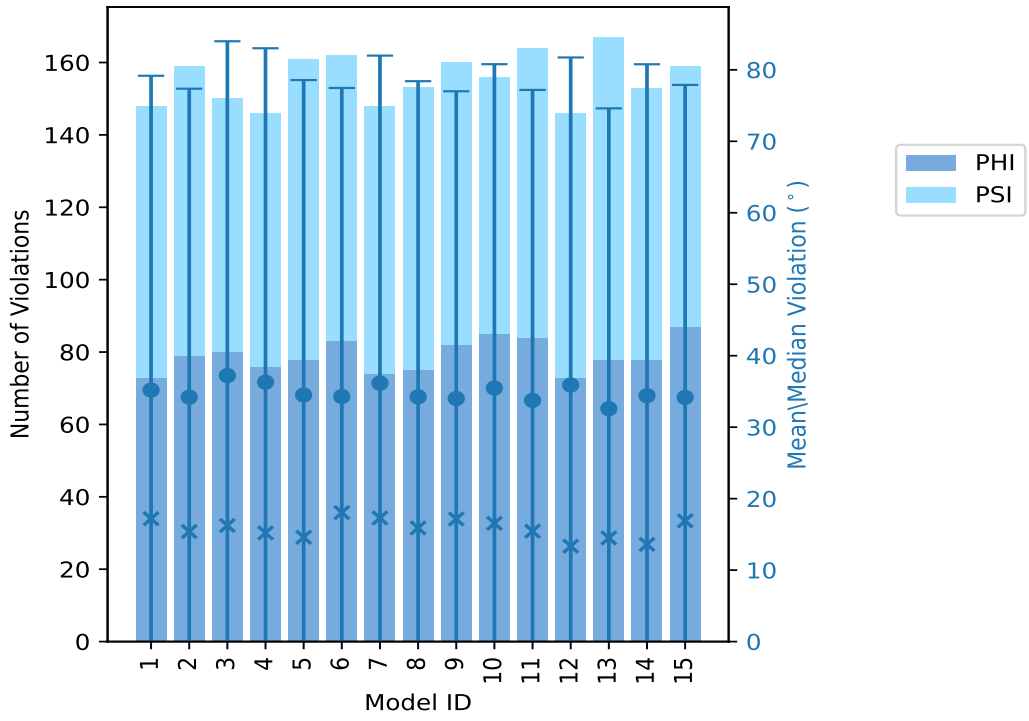
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

10.2 Dihedral-angle violation statistics for each model (i)

The following table provides the dihedral-angle violation statistics for each model in the ensemble. Violations less than 1° are not included in the statistics.

Model ID	Number of violations			Mean (°)	Max (°)	SD (°)	Median (°)
	PHI	PSI	Total				
1	73	75	148	35.19	171.1	43.98	17.2
2	79	80	159	34.2	171.2	43.16	15.4
3	80	70	150	37.22	163.2	46.78	16.25
4	76	70	146	36.3	171.7	46.72	15.2
5	78	83	161	34.5	169.7	44.07	14.6
6	83	79	162	34.29	171.4	43.17	18.05
7	74	74	148	36.16	171.1	45.83	17.3
8	75	78	153	34.24	172.8	44.17	15.9
9	82	78	160	34.01	172.4	42.99	17.15
10	85	71	156	35.47	170.3	45.32	16.55
11	84	80	164	33.74	170.9	43.45	15.45
12	73	73	146	35.9	169.7	45.83	13.35
13	78	89	167	32.59	169.0	42.02	14.5
14	78	75	153	34.41	170.2	46.37	13.6
15	87	72	159	34.17	166.3	43.72	16.9

10.2.1 Bar graph : Dihedral violation statistics for each model (i)



The mean(dot),median(x) and the standard deviation are shown in blue with respect to the y axis on the right

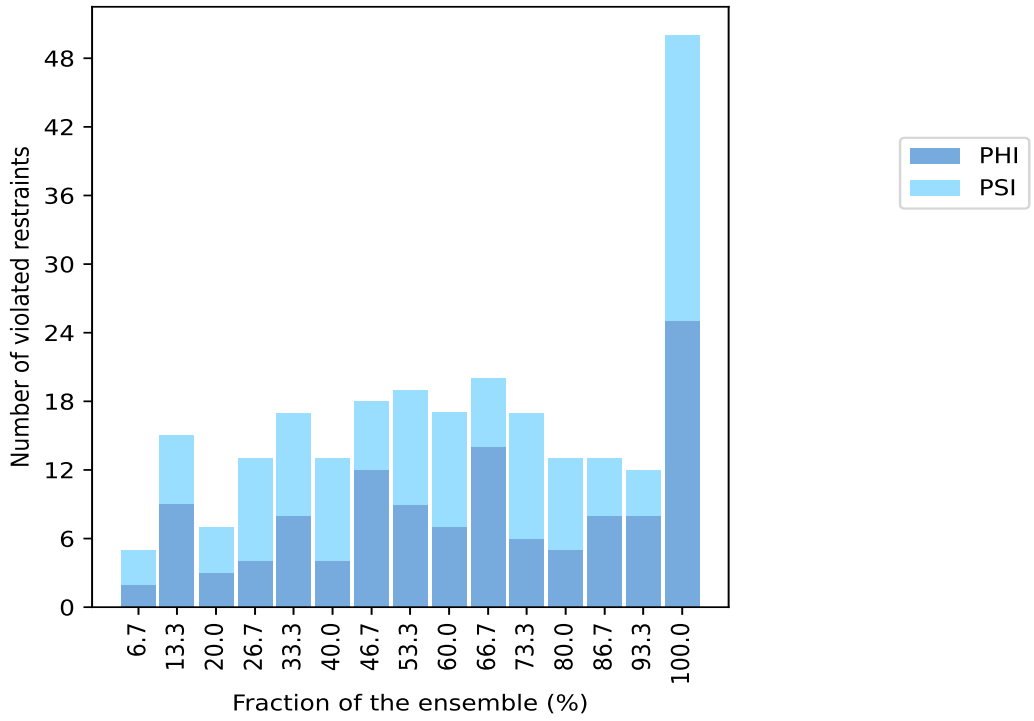
10.3 Dihedral-angle violation statistics for the ensemble (i)

Violation analysis may find that some restraints are violated in very few models and some are violated in most of models. The following table provides this information as number of violated restraints for a given fraction of ensemble.

Number of violated restraints			Fraction of the ensemble	
PHI	PSI	Total	Count[1]	%
2	3	5	1	6.7
9	6	15	2	13.3
3	4	7	3	20.0
4	9	13	4	26.7
8	9	17	5	33.3
4	9	13	6	40.0
12	6	18	7	46.7
9	10	19	8	53.3
7	10	17	9	60.0
14	6	20	10	66.7
6	11	17	11	73.3
5	8	13	12	80.0
8	5	13	13	86.7
8	4	12	14	93.3
25	25	50	15	100.0

[1] Number of models with violations

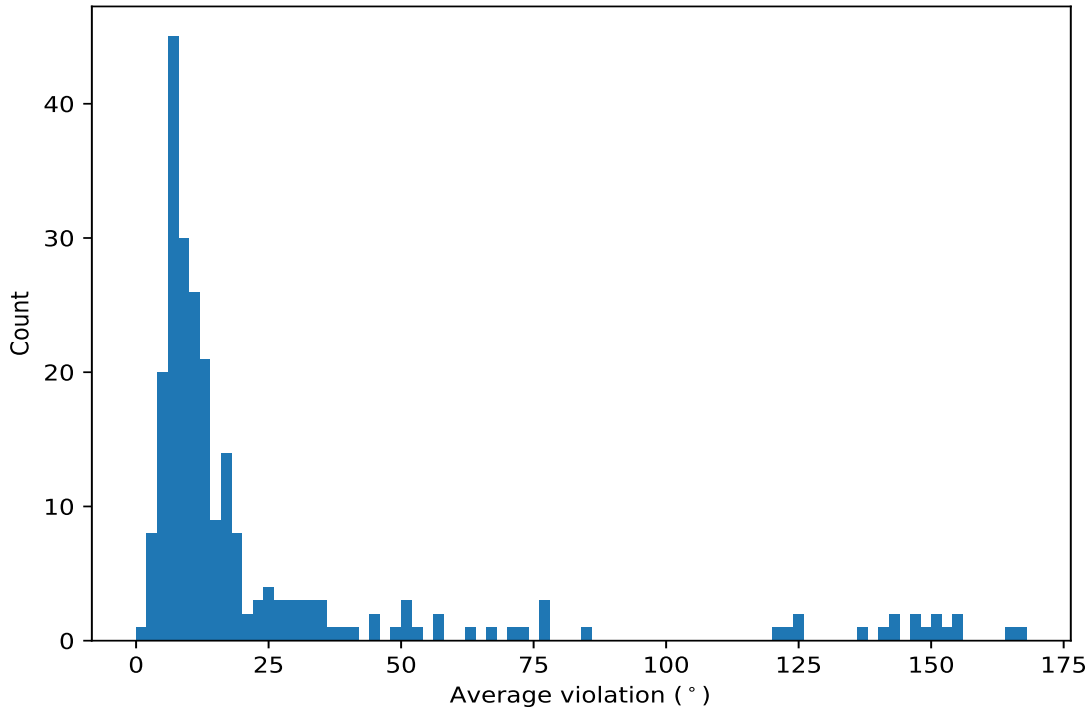
10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble (i)



10.4 Most violated dihedral-angle restraints in the ensemble (i)

10.4.1 Histogram : Distribution of mean dihedral-angle violations (i)

The following histogram shows the distribution of the average value of the violation. The average is calculated for each restraint that is violated in more than one model over all the violated models in the ensemble



10.4.2 Table: Most violated dihedral-angle restraints (i)

The following table provides the mean and the standard deviation of the violations for the 10 worst performing restraints, sorted by number of violated models and the mean violation value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint.

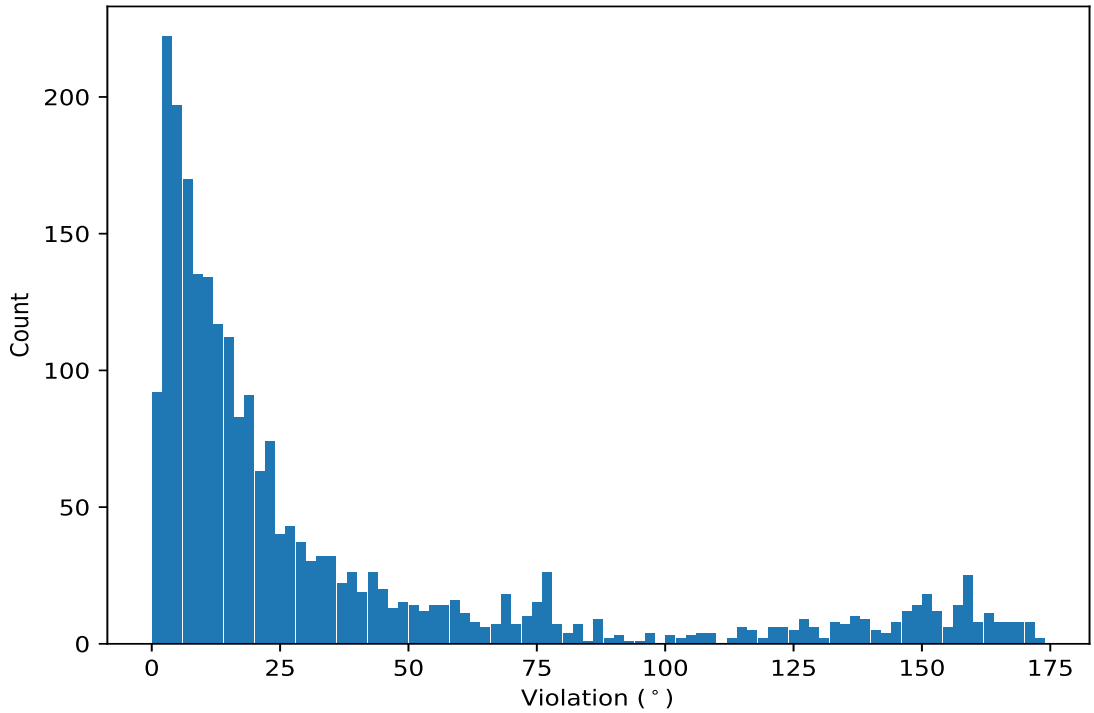
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	15	166.62	4.82	167.6
(1,133)	1:A:40:ALA:C	1:A:41:GLY:N	1:A:41:GLY:CA	1:A:41:GLY:C	15	165.7	4.59	167.5
(1,185)	1:A:67:THR:C	1:A:68:GLY:N	1:A:68:GLY:CA	1:A:68:GLY:C	15	155.83	9.96	161.3
(1,186)	1:B:67:THR:C	1:B:68:GLY:N	1:B:68:GLY:CA	1:B:68:GLY:C	15	155.59	7.42	158.7
(1,56)	1:A:21:THR:N	1:A:21:THR:CA	1:A:21:THR:C	1:A:22:HIS:N	15	153.03	7.77	152.7
(1,51)	1:A:20:LYS:N	1:A:20:LYS:CA	1:A:20:LYS:C	1:A:21:THR:N	15	150.82	7.88	150.7
(1,52)	1:B:20:LYS:N	1:B:20:LYS:CA	1:B:20:LYS:C	1:B:21:THR:N	15	150.27	6.33	151.9
(1,47)	1:A:19:SER:N	1:A:19:SER:CA	1:A:19:SER:C	1:A:20:LYS:N	15	149.75	5.45	149.5
(1,139)	1:B:42:LYS:N	1:B:42:LYS:CA	1:B:42:LYS:C	1:B:43:GLU:N	15	147.49	9.2	148.7
(1,55)	1:B:21:THR:N	1:B:21:THR:CA	1:B:21:THR:C	1:B:22:HIS:N	15	146.8	14.55	147.0

[1] Number of violated models, [2]Standard deviation, All angle values are in degree (°)

10.5 All violated dihedral-angle restraints (i)

10.5.1 Histogram : Distribution of violations (i)

The following histogram shows the distribution of the absolute value of the violation for all violated restraints in the ensemble.



10.5.2 Table: All violated dihedral-angle restraints (i)

The following table provides the list of violationsfor the 10 worst performing restraints, sorted by the violation value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint.

Key	Atom-1	Atom-2	Atom-3	Atom-4	Model ID	Violation (°)
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	8	172.8
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	9	172.4
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	4	171.7
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	6	171.4
(1,133)	1:A:40:ALA:C	1:A:41:GLY:N	1:A:41:GLY:CA	1:A:41:GLY:C	2	171.2
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	1	171.1
(1,133)	1:A:40:ALA:C	1:A:41:GLY:N	1:A:41:GLY:CA	1:A:41:GLY:C	7	171.1
(1,251)	1:B:92:ASN:N	1:B:92:ASN:CA	1:B:92:ASN:C	1:B:93:ASN:N	11	170.9
(1,133)	1:A:40:ALA:C	1:A:41:GLY:N	1:A:41:GLY:CA	1:A:41:GLY:C	10	170.3
(1,134)	1:B:40:ALA:C	1:B:41:GLY:N	1:B:41:GLY:CA	1:B:41:GLY:C	14	170.2