



wwPDB NMR Structure Validation Summary Report ⓘ

Aug 27, 2024 – 10:08 AM JST

PDB ID : 8KH1
BMRB ID : 28044
Title : Phage SPO1 protein Gp49 is a novel RNA binding protein that involves in host iron metabolism
Authors : Liu, B.
Deposited on : 2023-08-21

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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
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) : 2.38.2

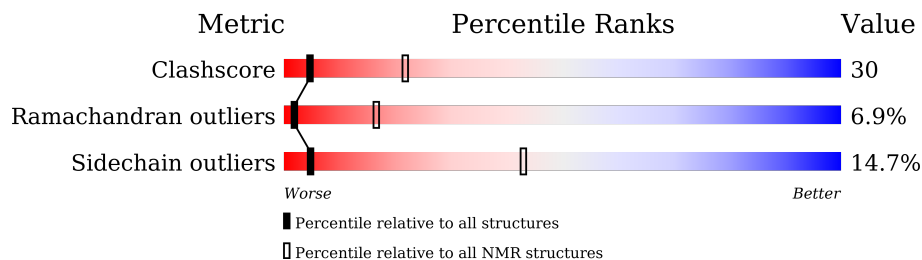
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 89%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	117	

2 Ensemble composition and analysis

This entry contains 10 models. Model 9 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:2-A:85, A:93-A:117 (109)	0.46	9

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 2 single-model clusters were found.

Cluster number	Models
1	3, 5, 6, 8, 9, 10
2	1, 4
Single-model clusters	2; 7

3 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 1875 atoms, of which 935 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative gene 49 protein.

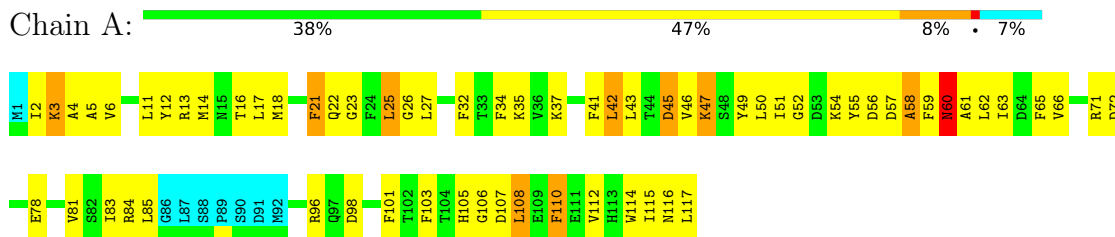
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	117	1875	606	935	151	179	4	0

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.

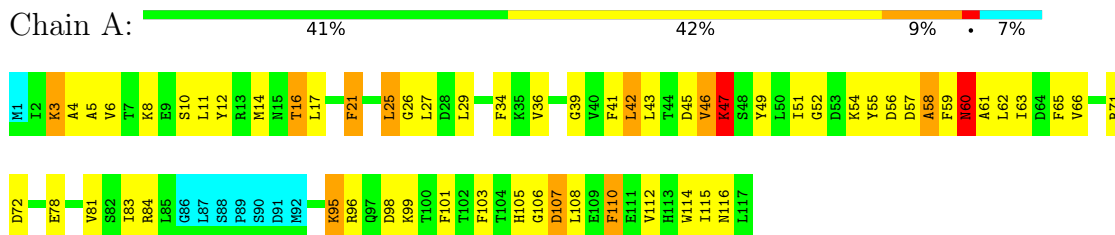
- Molecule 1: Putative gene 49 protein



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

The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: Putative gene 49 protein



5 Refinement protocol and experimental data overview

The models were refined using the following method: *distance geometry*.

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CNS	structure calculation	
ARIA	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	1418
Number of shifts mapped to atoms	1418
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

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	0.48±0.04	0±1/902 (0.0± 0.1%)	0.60±0.01	0±0/1216 (0.0± 0.0%)
All	All	0.48	4/9020 (0.0%)	0.60	0/12160 (0.0%)

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	55	TYR	CE2-CZ	-7.13	1.29	1.38	5	2
1	A	55	TYR	CE1-CZ	6.24	1.46	1.38	5	1
1	A	110	PHE	CE1-CZ	5.19	1.47	1.37	5	1

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	A	885	880	878	53±7
All	All	8850	8800	8780	527

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 30.

5 of 199 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:94:VAL:HB	1:A:96:ARG:NH2	0.79	1.91	2	1
1:A:27:LEU:HD21	1:A:45:ASP:HB2	0.78	1.55	1	8
1:A:5:ALA:HB1	1:A:115:ILE:HG13	0.77	1.57	6	8
1:A:46:VAL:HG11	1:A:59:PHE:CE2	0.76	2.15	8	10
1:A:6:VAL:HG21	1:A:66:VAL:HG21	0.76	1.57	5	7

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	108/117 (92%)	83±1 (77±1%)	17±1 (16±1%)	8±2 (7±1%)	2	16
All	All	1080/1170 (92%)	832 (77%)	173 (16%)	75 (7%)	2	16

5 of 16 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	16	THR	10
1	A	25	LEU	10
1	A	58	ALA	10
1	A	60	ASN	10
1	A	26	GLY	9

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	96/103 (93%)	82±1 (85±2%)	14±1 (15±2%)	5	43
All	All	960/1030 (93%)	819 (85%)	141 (15%)	5	43

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	A	3	LYS	10
1	A	47	LYS	10
1	A	110	PHE	10
1	A	42	LEU	9
1	A	60	ASN	9

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 oligosaccharides 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 89% for the well-defined parts and 87% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	1418
Number of shifts mapped to atoms	1418
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	116	0.03 ± 0.16	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	108	-0.04 ± 0.19	None needed (< 0.5 ppm)
$^{13}\text{C}'$	85	-0.07 ± 0.19	None needed (< 0.5 ppm)
^{15}N	112	-0.00 ± 0.39	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 89%, i.e. 1362 atoms were assigned a chemical shift out of a possible 1538. 0 out of 23 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	520/550 (95%)	223/224 (100%)	190/218 (87%)	107/108 (99%)
Sidechain	727/833 (87%)	496/541 (92%)	225/263 (86%)	6/29 (21%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	115/155 (74%)	58/76 (76%)	56/74 (76%)	1/5 (20%)
Overall	1362/1538 (89%)	777/841 (92%)	471/555 (85%)	114/142 (80%)

7.1.4 Statistically unusual chemical shifts [i](#)

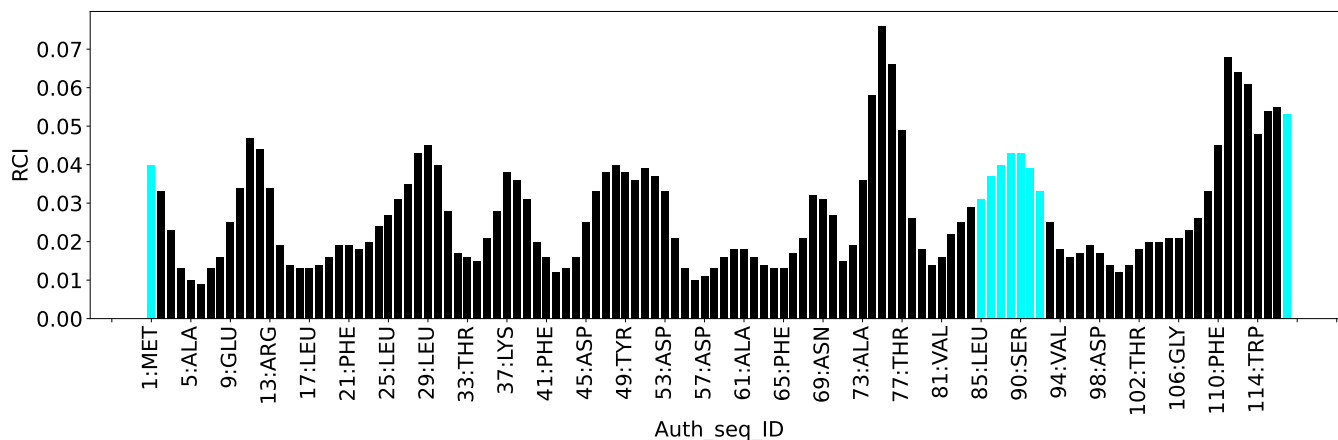
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	71	ARG	HG2	-0.32	0.26 – 2.87	-7.2
1	A	54	LYS	CD	21.20	23.50 – 34.42	-7.1
1	A	17	LEU	HB2	-0.17	-0.07 – 3.30	-5.3

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 \geq 5$)	0
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	158
Number of unmapped restraints	0
Number of restraints per residue	0
Number of long range restraints per residue ¹	0

¹Long 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)	24.6	9.98

Continued on next page...

Continued from previous page...

Bins (°)	Average number of violations per model	Max (°)
10.0-20.0 (Medium)	3.4	19.54
>20.0 (Large)	3.9	149.26

9 Distance violation analysis

No distance restraints data found

10 Dihedral-angle violation analysis [i](#)

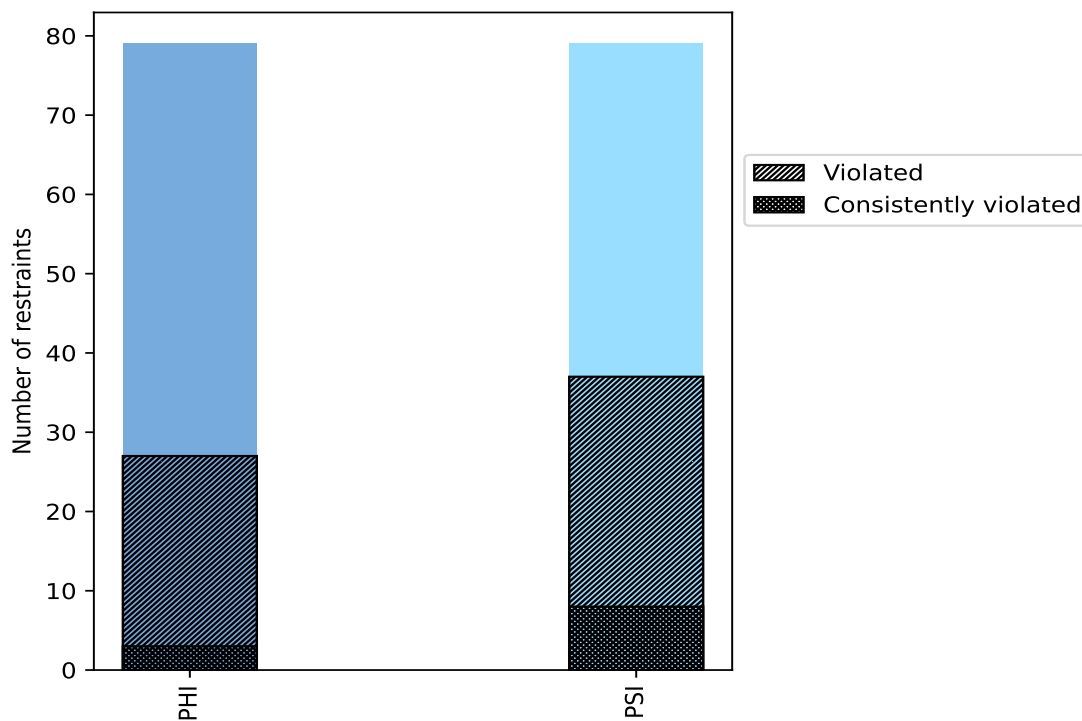
10.1 Summary of dihedral-angle violations [i](#)

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	79	50.0	27	34.2	17.1	3	3.8	1.9
PSI	79	50.0	37	46.8	23.4	8	10.1	5.1
Total	158	100.0	64	40.5	40.5	11	7.0	7.0

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](#)



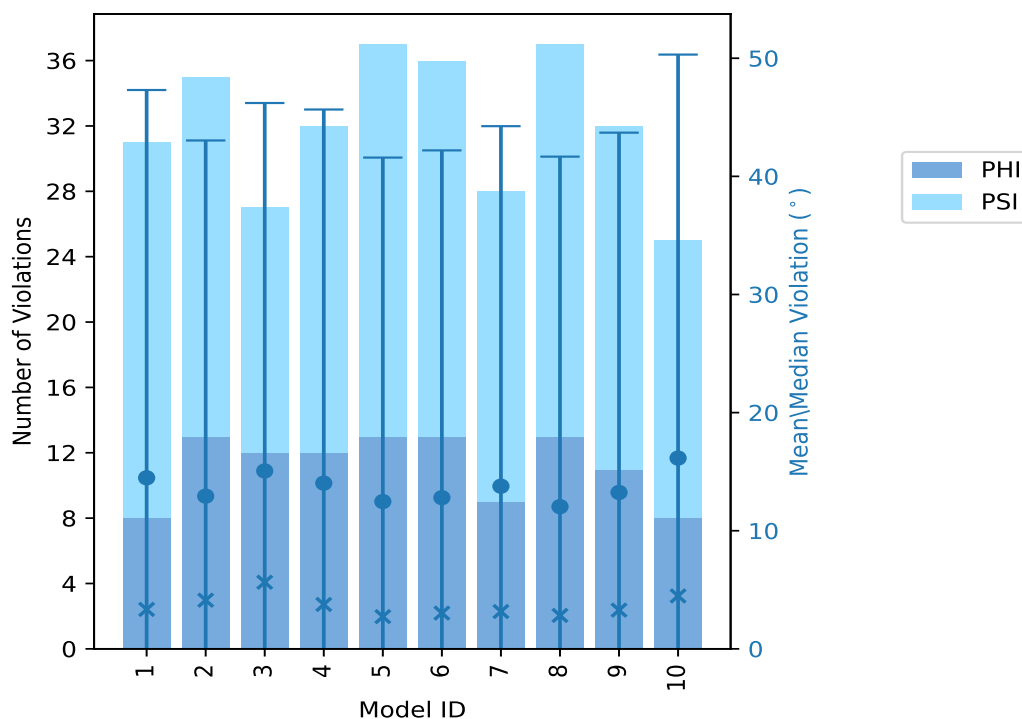
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

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	8	23	31	14.48	148.63	32.83	3.34
2	13	22	35	12.93	149.02	30.11	4.12
3	12	15	27	15.06	138.09	31.15	5.64
4	12	20	32	14.03	144.62	31.63	3.76
5	13	24	37	12.47	149.2	29.12	2.73
6	13	23	36	12.8	144.09	29.4	3.03
7	9	19	28	13.77	123.12	30.48	3.16
8	13	24	37	12.04	149.26	29.63	2.83
9	11	21	32	13.24	143.03	30.46	3.28
10	8	17	25	16.15	141.75	34.16	4.48

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

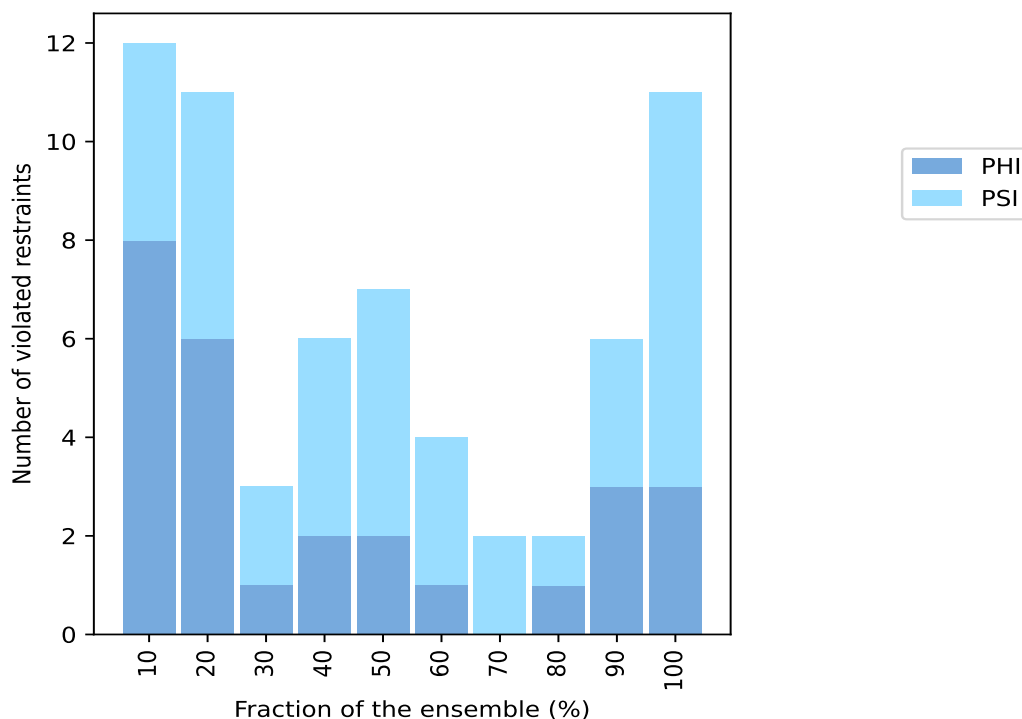
10.3 Dihedral-angle violation statistics for the ensemble [i](#)

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	%
8	4	12	1	10.0
6	5	11	2	20.0
1	2	3	3	30.0
2	4	6	4	40.0
2	5	7	5	50.0
1	3	4	6	60.0
0	2	2	7	70.0
1	1	2	8	80.0
3	3	6	9	90.0
3	8	11	10	100.0

¹ Number of models with violations

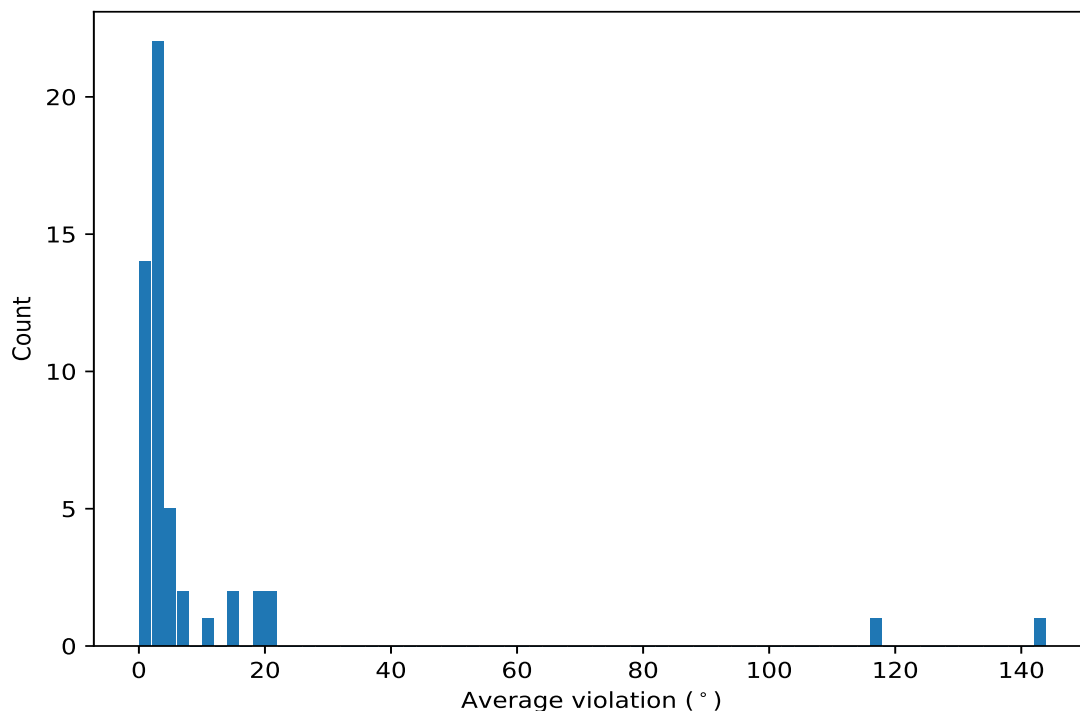
10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble [i](#)



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.

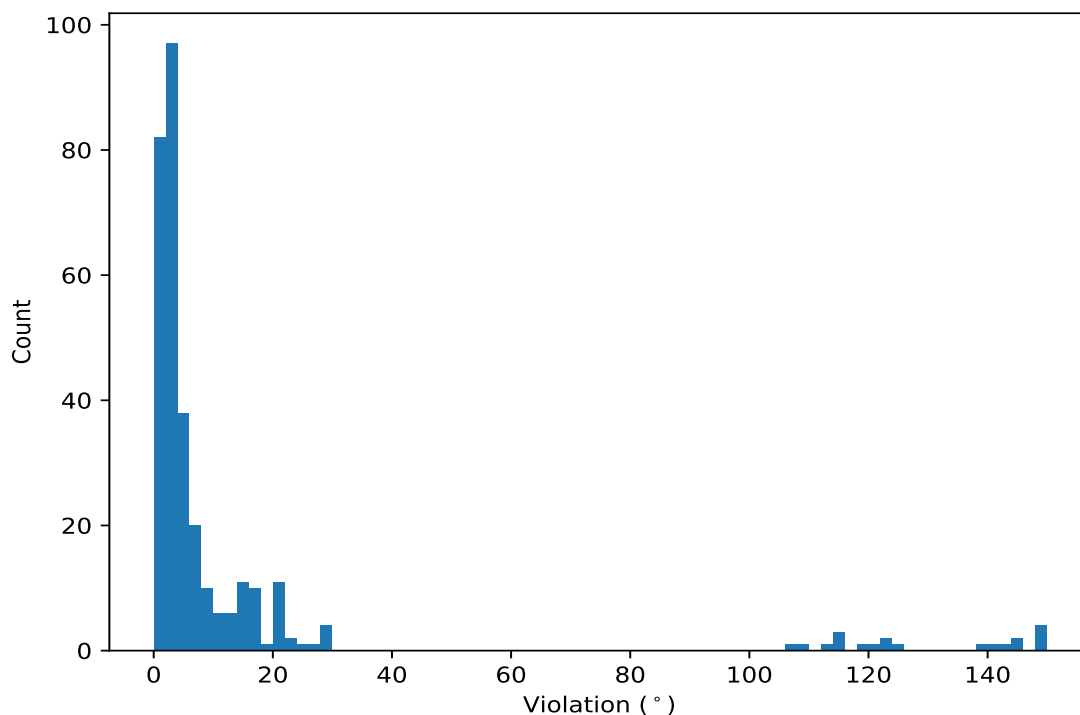
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models ¹	Mean	SD ²	Median
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	10	143.08	7.55	144.36
(1,13)	1:18:A:MET:C	1:19:A:GLU:N	1:19:A:GLU:CA	1:19:A:GLU:C	10	116.26	5.69	115.58
(1,36)	1:59:A:PHE:C	1:60:A:ASN:N	1:60:A:ASN:CA	1:60:A:ASN:C	10	20.27	2.6	20.88
(1,91)	1:18:A:MET:N	1:18:A:MET:CA	1:18:A:MET:C	1:19:A:GLU:N	10	19.39	5.96	17.11
(1,92)	1:19:A:GLU:N	1:19:A:GLU:CA	1:19:A:GLU:C	1:20:A:ALA:N	10	15.98	4.69	14.96
(1,114)	1:59:A:PHE:N	1:59:A:PHE:CA	1:59:A:PHE:C	1:60:A:ASN:N	10	14.95	1.52	15.52
(1,132)	1:87:A:LEU:N	1:87:A:LEU:CA	1:87:A:LEU:C	1:88:A:SER:N	10	7.04	3.07	7.82
(1,87)	1:14:A:MET:N	1:14:A:MET:CA	1:14:A:MET:C	1:15:A:ASN:N	10	5.95	1.01	6.38
(1,71)	1:105:A:HIS:C	1:106:A:GLY:N	1:106:A:GLY:CA	1:106:A:GLY:C	10	4.91	1.06	4.5
(1,112)	1:54:A:LYS:N	1:54:A:LYS:CA	1:54:A:LYS:C	1:55:A:TYR:N	10	3.82	1.39	3.48

¹ Number of violated models, ²Standard 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 violations for 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 (°)
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	8	149.26
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	5	149.2
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	2	149.02
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	1	148.63
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	4	144.62
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	6	144.09
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	9	143.03
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	10	141.75
(1,111)	1:52:A:GLY:N	1:52:A:GLY:CA	1:52:A:GLY:C	1:53:A:ASP:N	3	138.09
(1,13)	1:18:A:MET:C	1:19:A:GLU:N	1:19:A:GLU:CA	1:19:A:GLU:C	1	125.02