wwPDB NMR Structure Validation Summary Report (i)

Oct 8, 2024 – 10:21 AM EDT

PDB ID	:	8V0H
BMRB ID	:	50994
Title	:	Structural characterization of zika virus NS2B by NMR and RosettaMP
Authors	:	Penna, B.R.; Gomes-Neto, F.; Anobom, C.D.; Valente, A.P.
Deposited on	:	2023-11-17

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

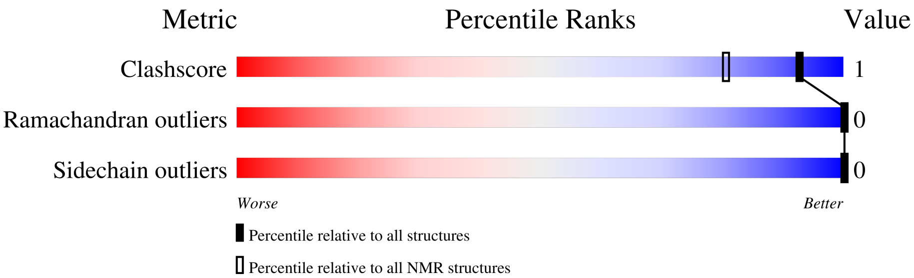
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 31%.

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	160	73% • 7% 19%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 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:75, A:86-A:130 (119)	0.76	19

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 6 single-model clusters were found.

Cluster number	Models
1	7, 10, 12, 13, 14, 18, 19, 20
2	2, 3, 6, 8, 9, 15
Single-model clusters	1; 4; 5; 11; 16; 17

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1939 atoms, of which 974 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Serine protease subunit NS2B.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	130	1939	621	974	153	183	8	0

There are 30 discrepancies between the modelled and reference sequences:

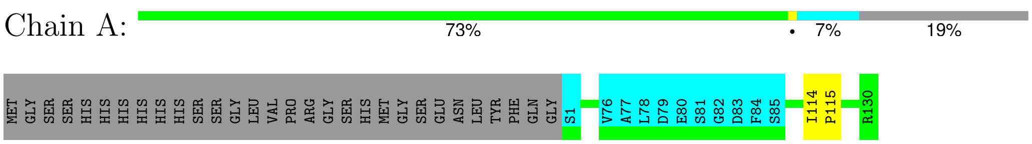
Chain	Residue	Modelled	Actual	Comment	Reference
A	-29	MET	-	initiating methionine	UNP Q32ZE1
A	-28	GLY	-	expression tag	UNP Q32ZE1
A	-27	SER	-	expression tag	UNP Q32ZE1
A	-26	SER	-	expression tag	UNP Q32ZE1
A	-25	HIS	-	expression tag	UNP Q32ZE1
A	-24	HIS	-	expression tag	UNP Q32ZE1
A	-23	HIS	-	expression tag	UNP Q32ZE1
A	-22	HIS	-	expression tag	UNP Q32ZE1
A	-21	HIS	-	expression tag	UNP Q32ZE1
A	-20	HIS	-	expression tag	UNP Q32ZE1
A	-19	SER	-	expression tag	UNP Q32ZE1
A	-18	SER	-	expression tag	UNP Q32ZE1
A	-17	GLY	-	expression tag	UNP Q32ZE1
A	-16	LEU	-	expression tag	UNP Q32ZE1
A	-15	VAL	-	expression tag	UNP Q32ZE1
A	-14	PRO	-	expression tag	UNP Q32ZE1
A	-13	ARG	-	expression tag	UNP Q32ZE1
A	-12	GLY	-	expression tag	UNP Q32ZE1
A	-11	SER	-	expression tag	UNP Q32ZE1
A	-10	HIS	-	expression tag	UNP Q32ZE1
A	-9	MET	-	expression tag	UNP Q32ZE1
A	-8	GLY	-	expression tag	UNP Q32ZE1
A	-7	SER	-	expression tag	UNP Q32ZE1
A	-6	GLU	-	expression tag	UNP Q32ZE1
A	-5	ASN	-	expression tag	UNP Q32ZE1
A	-4	LEU	-	expression tag	UNP Q32ZE1
A	-3	TYR	-	expression tag	UNP Q32ZE1
A	-2	PHE	-	expression tag	UNP Q32ZE1
A	-1	GLN	-	expression tag	UNP Q32ZE1
A	0	GLY	-	expression tag	UNP Q32ZE1

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.

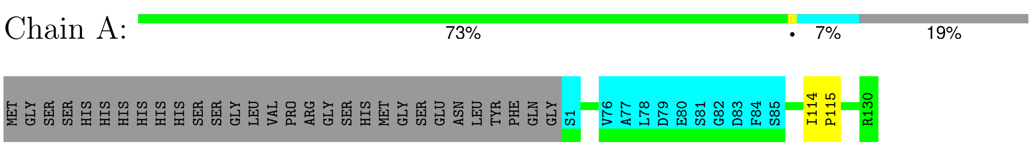
- Molecule 1: Serine protease subunit NS2B



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

The representative model is number 19. Colouring as in section 4.1 above.

- Molecule 1: Serine protease subunit NS2B



5 Refinement protocol and experimental data overview (i)

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

Of the 500000 calculated structures, 20 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
Rosetta	structure calculation	2020.08.61146
Rosetta	refinement	2020.08.61146
Rosetta	geometry optimization	2020.08.61146

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	640
Number of shifts mapped to atoms	539
Number of unparsed shifts	0
Number of shifts with mapping errors	101
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	31%

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

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	887	908	908	2±1
All	All	17740	18160	18160	45

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 1.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:130:ARG:OXT	1:A:130:ARG:HG3	0.60	1.96	11	1
1:A:74:LEU:C	1:A:74:LEU:HD23	0.55	2.22	16	2
1:A:74:LEU:HD23	1:A:74:LEU:O	0.50	2.07	16	2
1:A:114:ILE:N	1:A:115:PRO:HD2	0.49	2.22	7	20
1:A:114:ILE:HB	1:A:115:PRO:CD	0.47	2.40	8	20

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	118/160 (74%)	117±1 (99±1%)	1±1 (1±1%)	0±0 (0±0%)	100	100
All	All	2360/3200 (74%)	2343 (99%)	17 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	92/126 (73%)	92±0 (100±0%)	0±0 (0±0%)	100	100
All	All	1840/2520 (73%)	1840 (100%)	0 (0%)	100	100

There are no protein residues with a non-rotameric sidechain to report.

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 31% for the well-defined parts and 32% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_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	640
Number of shifts mapped to atoms	539
Number of unparsed shifts	0
Number of shifts with mapping errors	101
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 101) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	-29	MET	C	173.748	.	.
1	A	-29	MET	CA	53.553	.	.
1	A	-29	MET	CB	26.033	.	.
1	A	-28	GLY	H	7.54	.	.
1	A	-28	GLY	C	171.066	.	.
1	A	-28	GLY	CA	42.816	.	.
1	A	-28	GLY	N	108.752	.	.
1	A	-27	SER	H	7.709	.	.
1	A	-27	SER	C	170.69	.	.
1	A	-27	SER	CA	55.799	.	.
1	A	-27	SER	CB	61.501	.	.
1	A	-27	SER	N	115.003	.	.
1	A	-18	SER	H	8.215	.	.
1	A	-18	SER	C	172.46	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	-18	SER	CA	56.37	.	.
1	A	-18	SER	CB	61.396	.	.
1	A	-18	SER	N	116.222	.	.
1	A	-17	GLY	H	8.185	.	.
1	A	-17	GLY	C	171.354	.	.
1	A	-17	GLY	CA	42.951	.	.
1	A	-17	GLY	N	110.728	.	.
1	A	-16	LEU	H	7.897	.	.
1	A	-16	LEU	C	173.973	.	.
1	A	-16	LEU	CA	53.542	.	.
1	A	-16	LEU	CB	39.582	.	.
1	A	-16	LEU	N	121.502	.	.
1	A	-15	VAL	H	7.605	.	.
1	A	-15	VAL	C	171.684	.	.
1	A	-15	VAL	CA	57.289	.	.
1	A	-15	VAL	CB	29.596	.	.
1	A	-15	VAL	N	118.174	.	.
1	A	-14	PRO	C	174.048	.	.
1	A	-14	PRO	CA	60.247	.	.
1	A	-14	PRO	CB	31.606	.	.
1	A	-13	ARG	H	8.517	.	.
1	A	-13	ARG	C	174.475	.	.
1	A	-13	ARG	CA	54.159	.	.
1	A	-13	ARG	CB	27.485	.	.
1	A	-13	ARG	N	121.594	.	.
1	A	-12	GLY	H	8.221	.	.
1	A	-12	GLY	C	171.94	.	.
1	A	-12	GLY	CA	42.941	.	.
1	A	-12	GLY	N	109.523	.	.
1	A	-11	SER	H	7.977	.	.
1	A	-11	SER	C	172.038	.	.
1	A	-11	SER	CA	56.677	.	.
1	A	-11	SER	CB	61.293	.	.
1	A	-11	SER	N	115.225	.	.
1	A	-10	HIS	H	8.279	.	.
1	A	-10	HIS	C	174.137	.	.
1	A	-10	HIS	CA	54.102	.	.
1	A	-10	HIS	CB	26.649	.	.
1	A	-10	HIS	N	119.368	.	.
1	A	-9	MET	H	8.03	.	.
1	A	-9	MET	C	173.9	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	-9	MET	CA	59.966	.	.
1	A	-9	MET	CB	29.449	.	.
1	A	-9	MET	N	121.483	.	.
1	A	-8	GLY	H	8.231	.	.
1	A	-8	GLY	C	171.781	.	.
1	A	-8	GLY	CA	42.977	.	.
1	A	-8	GLY	N	109.74	.	.
1	A	-7	SER	H	7.991	.	.
1	A	-7	SER	C	172.19	.	.
1	A	-7	SER	CA	56.445	.	.
1	A	-7	SER	CB	61.399	.	.
1	A	-7	SER	N	115.488	.	.
1	A	-6	GLU	H	8.375	.	.
1	A	-6	GLU	C	173.416	.	.
1	A	-6	GLU	CA	54.464	.	.
1	A	-6	GLU	CB	26.67	.	.
1	A	-6	GLU	N	121.377	.	.
1	A	-5	ASN	H	8.028	.	.
1	A	-5	ASN	C	171.86	.	.
1	A	-5	ASN	CA	51.053	.	.
1	A	-5	ASN	CB	36.592	.	.
1	A	-5	ASN	N	118.062	.	.
1	A	-4	LEU	H	7.85	.	.
1	A	-4	LEU	C	174.135	.	.
1	A	-4	LEU	CA	52.809	.	.
1	A	-4	LEU	CB	38.908	.	.
1	A	-4	LEU	N	121.672	.	.
1	A	-3	TYR	H	7.874	.	.
1	A	-3	TYR	C	172.857	.	.
1	A	-3	TYR	CA	56.268	.	.
1	A	-3	TYR	CB	36.249	.	.
1	A	-3	TYR	N	119.841	.	.
1	A	-2	PHE	H	7.894	.	.
1	A	-2	PHE	C	173.014	.	.
1	A	-2	PHE	CA	55.291	.	.
1	A	-2	PHE	CB	36.369	.	.
1	A	-2	PHE	N	118.819	.	.
1	A	-1	GLN	H	8.05	.	.
1	A	-1	GLN	C	173.761	.	.
1	A	-1	GLN	CA	53.421	.	.
1	A	-1	GLN	CB	26.051	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	-1	GLN	N	120.716	.	.
1	A	0	GLY	H	7.474	.	.
1	A	0	GLY	C	171.042	.	.
1	A	0	GLY	CA	42.826	.	.
1	A	0	GLY	N	108.846	.	.

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	135	1.71 ± 0.22	Should be checked
$^{13}C_{\beta}$	116	3.62 ± 0.14	Should be checked
$^{13}C'$	137	1.99 ± 0.17	Should be applied
^{15}N	126	0.60 ± 0.34	None needed (imprecise)

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 31%, i.e. 485 atoms were assigned a chemical shift out of a possible 1568. 0 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	396/591 (67%)	95/242 (39%)	206/238 (87%)	95/111 (86%)
Sidechain	89/885 (10%)	0/589 (0%)	89/276 (32%)	0/20 (0%)
Aromatic	0/92 (0%)	0/44 (0%)	0/45 (0%)	0/3 (0%)
Overall	485/1568 (31%)	95/875 (11%)	295/559 (53%)	95/134 (71%)

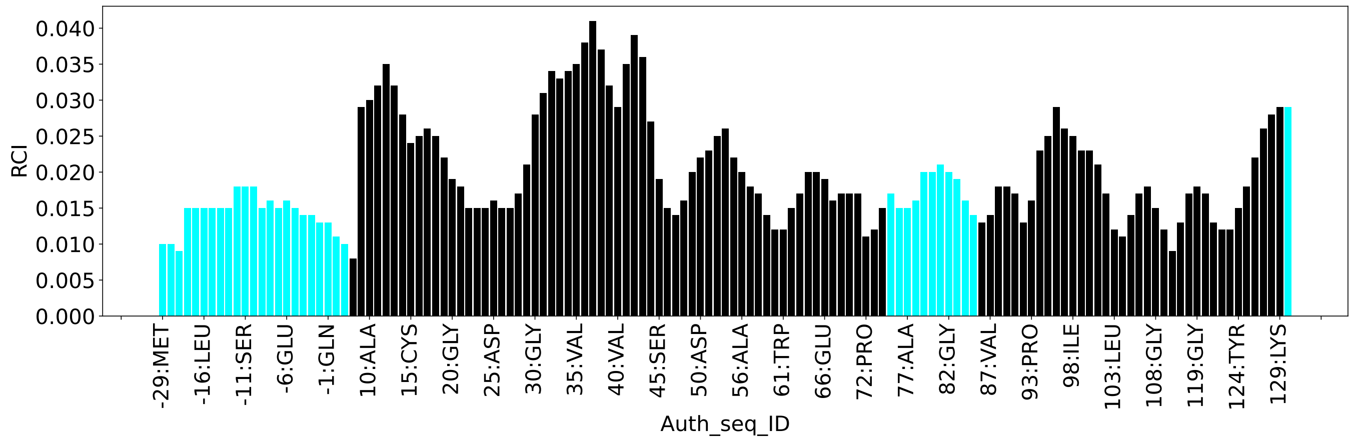
7.1.4 Statistically unusual chemical shifts (i)

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	89
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.There are no dihedral-angle violations

9 Distance violation analysis (i)

No distance restraints data found

10 Dihedral-angle violation analysis (i)

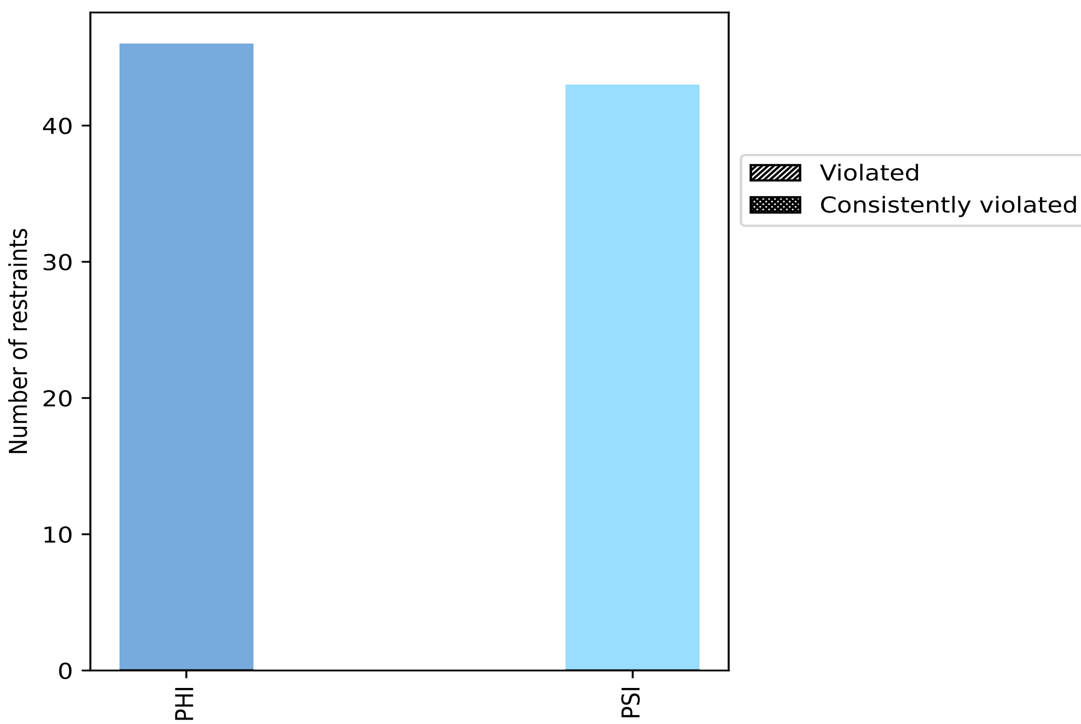
10.1 Summary of dihedral-angle violations (i)

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	46	51.7	0	0.0	0.0	0	0.0	0.0
PSI	43	48.3	0	0.0	0.0	0	0.0	0.0
Total	89	100.0	0	0.0	0.0	0	0.0	0.0

[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)

No violations found

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

No violations found

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

No violations found

10.5 All violated dihedral-angle restraints (i)

No violations found