wwPDB NMR Structure Validation Summary Report (i)

Dec 13, 2023 – 11:34 PM EST

PDB ID	:	2JOE
BMRB ID	:	15167
Title	:	NMR Structure of E. Coli YehR Protein. Northeast Structural Genomics Target ER538.
Authors	:	Ding, K.; Ramelot, T.A.; Cort, J.R.; Chen, C.X.; Jiang, M.; Cunningham, K.; Ma, L.; Xiao, R.; Liu, J.; Baran, M.C.; Swapna, G.V.T.; Acton, T.B.; Rost, B.; Montelione, G.T.; Kennedy, M.A.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2007-03-08

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:

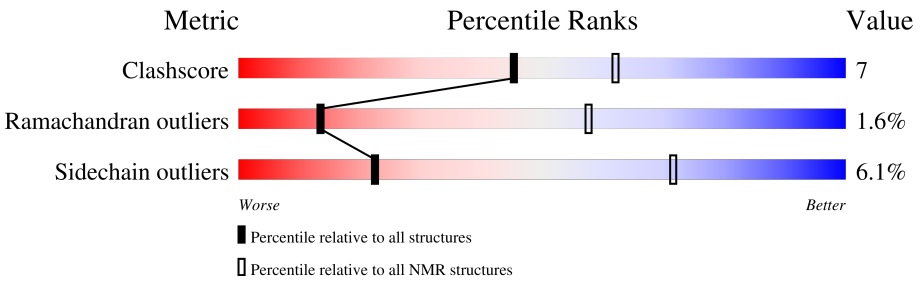
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)

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 97%.

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	139	73% 13% 14%

Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.36

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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 total 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:5-A:99, A:108-A:131 (119)	0.96	15

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2168 atoms, of which 1090 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hypothetical lipoprotein yehR.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	139	2168	677	1090	180	216	5	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	cloning artifact	UNP P33354
A	132	LEU	-	cloning artifact	UNP P33354
A	133	GLU	-	cloning artifact	UNP P33354
A	134	HIS	-	expression tag	UNP P33354
A	135	HIS	-	expression tag	UNP P33354
A	136	HIS	-	expression tag	UNP P33354
A	137	HIS	-	expression tag	UNP P33354
A	138	HIS	-	expression tag	UNP P33354
A	139	HIS	-	expression tag	UNP P33354

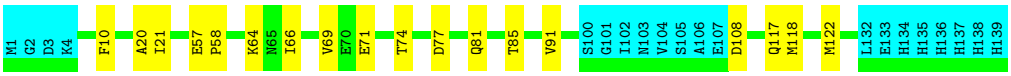
4 Residue-property plots (i)

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: Hypothetical lipoprotein yehR

Chain A: 73% 13% 14%

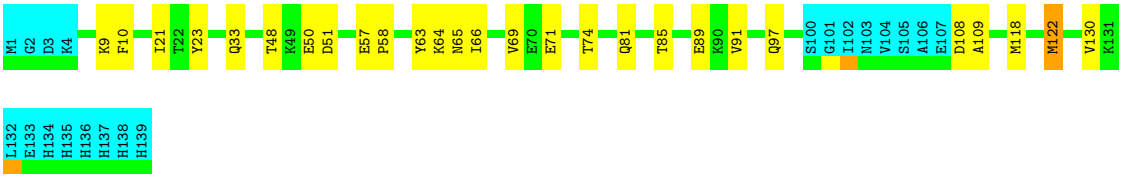


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 15. Colouring as in section 4.1 above.

• Molecule 1: Hypothetical lipoprotein yehR

Chain A: 66% 19% • 14%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 30 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest NOE and bond energy.*

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

Software name	**Classification**	**Version**
X-PLOR NIH	structure solution	2.15.0
CNS	refinement	1.1

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	1732
Number of shifts mapped to atoms	1732
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	97%

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	918	944	944	12±4
All	All	18360	18880	18880	250

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 7.

5 of 121 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:62:LYS:HB3	1:A:99:ILE:HD12	0.80	1.52	1	2
1:A:108:ASP:HA	1:A:111:LYS:HE2	0.75	1.59	6	2
1:A:8:LYS:HE2	1:A:129:GLU:HG2	0.71	1.59	2	1
1:A:69:VAL:HA	1:A:85:THR:O	0.68	1.89	15	12
1:A:10:PHE:CE2	1:A:21:ILE:HB	0.66	2.26	2	15

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	119/139 (86%)	108±3 (91±2%)	9±3 (7±2%)	2±1 (2±1%)	13 57
All	All	2380/2780 (86%)	2165 (91%)	178 (7%)	37 (2%)	13 57

5 of 11 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	77	ASP	12
1	A	64	LYS	9
1	A	131	LYS	2
1	A	91	VAL	2
1	A	66	ILE	2

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	98/115 (85%)	92±2 (94±2%)	6±2 (6±2%)	22 71
All	All	1960/2300 (85%)	1841 (94%)	119 (6%)	22 71

5 of 38 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	91	VAL	16
1	A	122	MET	13
1	A	71	GLU	11
1	A	118	MET	6
1	A	14	LEU	6

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 97% for the well-defined parts and 94% 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	1732
Number of shifts mapped to atoms	1732
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	138	-0.12 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	129	-0.13 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	130	0.02 ± 0.13	None needed (< 0.5 ppm)
^{15}N	137	-0.34 ± 0.19	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 97%, i.e. 1544 atoms were assigned a chemical shift out of a possible 1586. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	599/600 (100%)	244/244 (100%)	237/238 (100%)	118/118 (100%)
Sidechain	865/901 (96%)	584/584 (100%)	272/290 (94%)	9/27 (33%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	80/85 (94%)	40/40 (100%)	40/45 (89%)	0/0 (—%)
Overall	1544/1586 (97%)	868/868 (100%)	549/573 (96%)	127/145 (88%)

7.1.4 Statistically unusual chemical shifts (i)

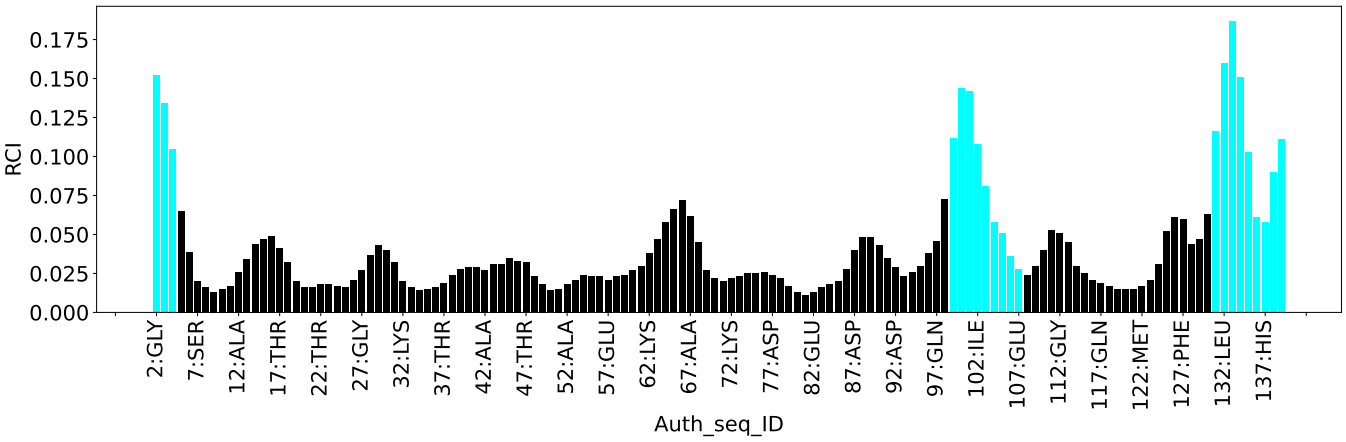
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	6	GLU	HB2	0.80	1.00 – 3.05	-6.0

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	1182
Intra-residue (|i-j|=0)	0
Sequential (|i-j|=1)	175
Medium range (|i-j|>1 and |i-j|<5)	314
Long range (|i-j|≥5)	621
Inter-chain	0
Hydrogen bond restraints	72
Disulfide bond restraints	0
Total dihedral-angle restraints	161
Number of unmapped restraints	0
Number of restraints per residue	9.7
Number of long range restraints per residue[1]	4.9

[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 violations

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)

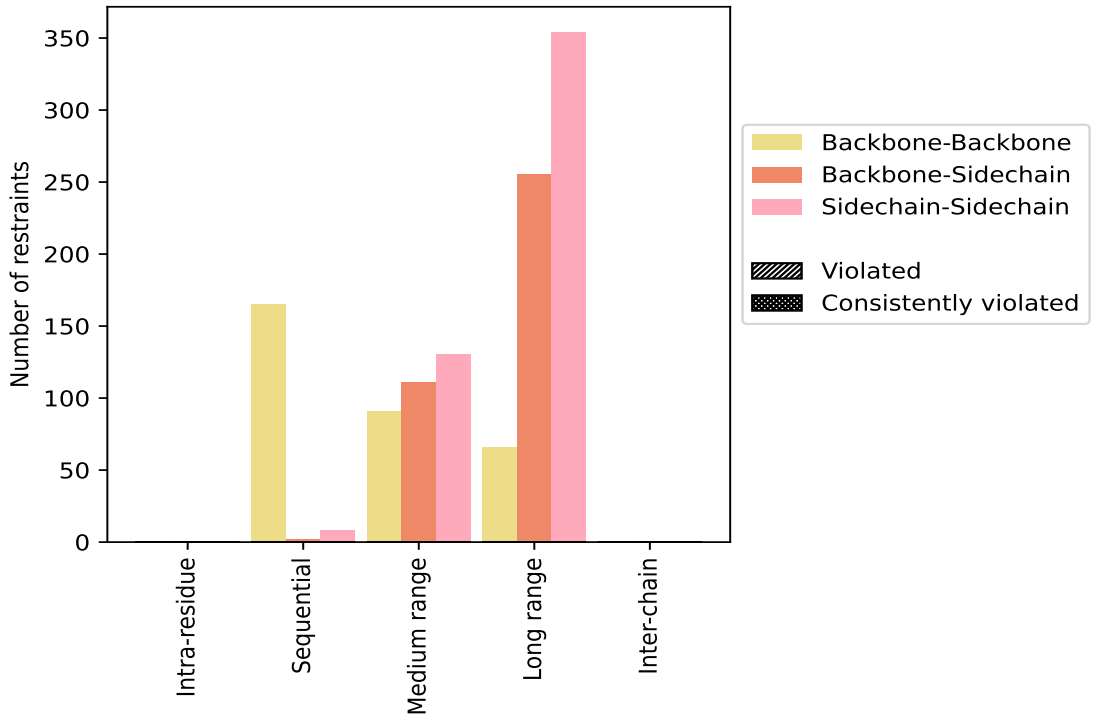
9.1 Summary of distance violations (i)

The following table shows the summary of distance violations in different restraint categories based on the sequence separation of the atoms involved. Each category is further sub-divided into three sub-categories based on the atoms involved. Violations less than 0.1 Å are not included in the statistics.

Restraints type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
Intra-residue (|i-j|=0)	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	175	14.8	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	165	14.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	2	0.2	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	8	0.7	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	314	26.6	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	91	7.7	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	93	7.9	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	130	11.0	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	621	52.5	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	66	5.6	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	201	17.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	354	29.9	0	0.0	0.0	0	0.0	0.0
Inter-chain	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	72	6.1	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	1182	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	322	27.2	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	368	31.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	492	41.6	0	0.0	0.0	0	0.0	0.0

[1] percentage calculated with respect to the total number of distance restraints, [2] percentage calculated with respect to the number of restraints in a particular restraint category, [3] violated in at least one model, [4] violated in all the models

9.1.1 Bar chart : Distribution of distance restraints and violations (i)



Violated and consistently violated restraints are shown using different hatch patterns in their respective categories. The hydrogen bonds and disulfied bonds are counted in their appropriate category on the x-axis

9.2 Distance violation statistics for each model (i)

No violations found

9.3 Distance violation statistics for the ensemble (i)

No violations found

9.4 Most violated distance restraints in the ensemble (i)

No violations found

9.5 All violated distance restraints (i)

No violations found

10 Dihedral-angle violation analysis (i)

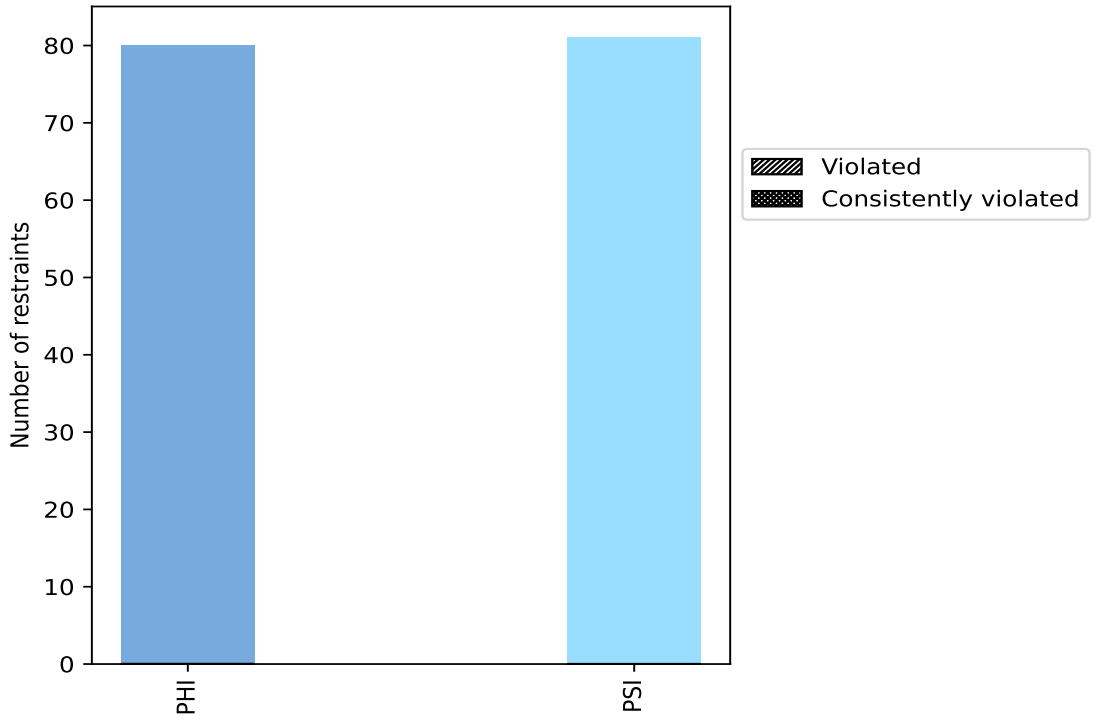
10.1 Summary of dihedral-angle violations (i)

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	80	49.7	0	0.0	0.0	0	0.0	0.0
PSI	81	50.3	0	0.0	0.0	0	0.0	0.0
Total	161	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