wwPDB NMR Structure Validation Summary Report

Jul 2, 2020 – 12:53 AM CDT

PDB ID	:	2JUW
Title	:	NMR solution structure of homodimer protein SO_2176 from Shewanella oneidensis. Northeast Structural Genomics Consortium target SoR77
Authors	:	Ramelot, T.A.; Cort, J.R.; Wang, D.; Nwosu, C.; Owens, 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	:	2007-09-03

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

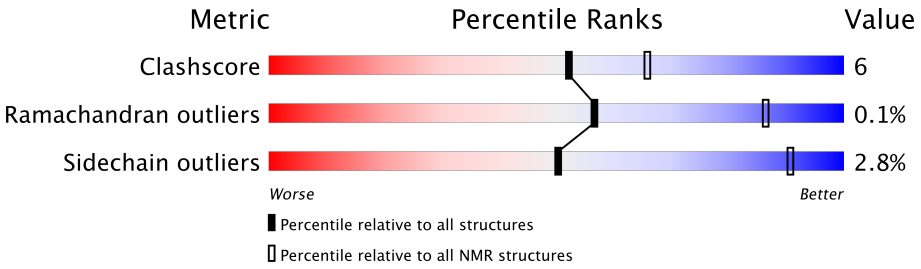
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.6.dev1
BMRB Restraints Analalysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.6.dev1

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 66%.

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	136327	12091
Ramachandran outliers	132723	10835
Sidechain outliers	132532	10811

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	80	69% 10% 21%
1	B	80	73% 8% 20%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:8-A:70, B:8-B:71 (127)	0.48	8

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 4 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2542 atoms, of which 1288 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called UPF0352 protein SO_2176.

Mol	Chain	Residues	Atoms						Trace
1	A	80	Total	C	H	N	O	S	0
			1271	391	644	116	117	3	
1	B	80	Total	C	H	N	O	S	0
			1271	391	644	116	117	3	

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	73	LEU	-	EXPRESSION TAG	UNP Q8EF26
A	74	GLU	-	EXPRESSION TAG	UNP Q8EF26
A	75	HIS	-	EXPRESSION TAG	UNP Q8EF26
A	76	HIS	-	EXPRESSION TAG	UNP Q8EF26
A	77	HIS	-	EXPRESSION TAG	UNP Q8EF26
A	78	HIS	-	EXPRESSION TAG	UNP Q8EF26
A	79	HIS	-	EXPRESSION TAG	UNP Q8EF26
A	80	HIS	-	EXPRESSION TAG	UNP Q8EF26
B	73	LEU	-	EXPRESSION TAG	UNP Q8EF26
B	74	GLU	-	EXPRESSION TAG	UNP Q8EF26
B	75	HIS	-	EXPRESSION TAG	UNP Q8EF26
B	76	HIS	-	EXPRESSION TAG	UNP Q8EF26
B	77	HIS	-	EXPRESSION TAG	UNP Q8EF26
B	78	HIS	-	EXPRESSION TAG	UNP Q8EF26
B	79	HIS	-	EXPRESSION TAG	UNP Q8EF26
B	80	HIS	-	EXPRESSION TAG	UNP Q8EF26

4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

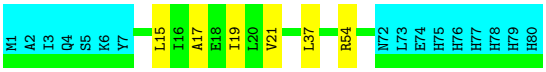
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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: UPF0352 protein SO_2176

Chain A:

• Molecule 1: UPF0352 protein SO_2176

Chain B:



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

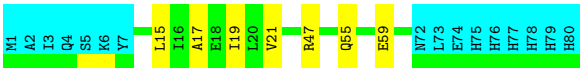
The representative model is number 8. Colouring as in section 4.1 above.

• Molecule 1: UPF0352 protein SO_2176

Chain A:

• Molecule 1: UPF0352 protein SO_2176

Chain B:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing.*

Of the 100 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
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)	2juw_nmr.cif
Number of chemical shift lists	1
Total number of shifts	1959
Number of shifts mapped to atoms	1217
Number of unparsed shifts	278
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	464
Assignment completeness (well-defined parts)	66%

6 Model quality

6.1 Standard geometry

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

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	478	506	504	8±2
1	B	484	511	509	7±2
All	All	19240	20340	20260	245

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 6.

5 of 107 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:62:ALA:HB1	1:B:62:ALA:HB1	0.73	1.60	7	3
1:B:48:LYS:HE2	1:B:48:LYS:HA	0.70	1.63	16	1
1:B:45:LEU:HB3	1:B:54:ARG:HG3	0.69	1.63	4	6
1:B:15:LEU:O	1:B:19:ILE:HG12	0.64	1.93	4	2
1:A:15:LEU:O	1:A:19:ILE:HG12	0.63	1.93	2	5

6.3 Torsion angles

6.3.1 Protein backbone

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	63/80 (79%)	62±1 (98±2%)	1±1 (1±2%)	0±0 (0±0%)	56 86
1	B	64/80 (80%)	63±1 (99±2%)	1±1 (1±2%)	0±0 (0±0%)	56 86
All	All	2540/3200 (79%)	2506 (99%)	32 (1%)	2 (0%)	56 86

All 2 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	8	SER	1
1	B	8	SER	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	54/70 (77%)	53±1 (98±2%)	1±1 (2±2%)	54 92
1	B	55/70 (79%)	53±1 (97±2%)	2±1 (3±2%)	48 89
All	All	2180/2800 (78%)	2120 (97%)	60 (3%)	50 91

5 of 35 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	43	HIS	5
1	B	47	ARG	4
1	A	24	GLU	3
1	B	63	LYS	3
1	B	52	GLU	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 carbohydrates 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

The completeness of assignment taking into account all chemical shift lists is 66% for the well-defined parts and 61% for the entire structure.

7.1 Chemical shift list 1

File name: 2juw_nmr.cif

Chemical shift list name: *nef_chemical_shift_list_2juw_rt.mr*

7.1.1 Bookkeeping

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	1959
Number of shifts mapped to atoms	1217
Number of unparsed shifts	278
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	464
Number of shift outliers (ShiftChecker)	0

The following errors were found when reading this chemical shift list.

- Chemical shift has been reported more than once. First 5 (of 278) occurences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
3	A	2	ALA	HB%	1.510	0.020	1
4	A	2	ALA	HB%	1.510	0.020	1
12	A	3	ILE	HD1%	0.870	0.020	1
13	A	3	ILE	HD1%	0.870	0.020	1
17	A	3	ILE	HG2%	0.920	0.020	1

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

- No matching atoms found in structure. First 5 (of 464) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	16	ILE	HG1y	1.78	0.02	2
B	31	ASP	HBy	2.88	0.02	2

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
B	59	GLU	HGx	2.32	0.02	2
B	39	ASN	HBy	2.85	0.02	2
B	55	GLN	HG%	2.32	0.02	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
$^{13}C_{\alpha}$	154	-0.56 ± 0.08	Should be applied
$^{13}C_{\beta}$	152	0.24 ± 0.06	None needed (< 0.5 ppm)
$^{13}C'$	148	-0.39 ± 0.10	None needed (< 0.5 ppm)
^{15}N	148	0.31 ± 0.22	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 66%, i.e. 1012 atoms were assigned a chemical shift out of a possible 1532. 32 out of 32 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	621/627 (99%)	248/250 (99%)	250/254 (98%)	123/123 (100%)
Sidechain	363/855 (42%)	46/496 (9%)	301/325 (93%)	16/34 (47%)
Aromatic	28/50 (56%)	12/26 (46%)	12/16 (75%)	4/8 (50%)
Overall	1012/1532 (66%)	306/772 (40%)	563/595 (95%)	143/165 (87%)

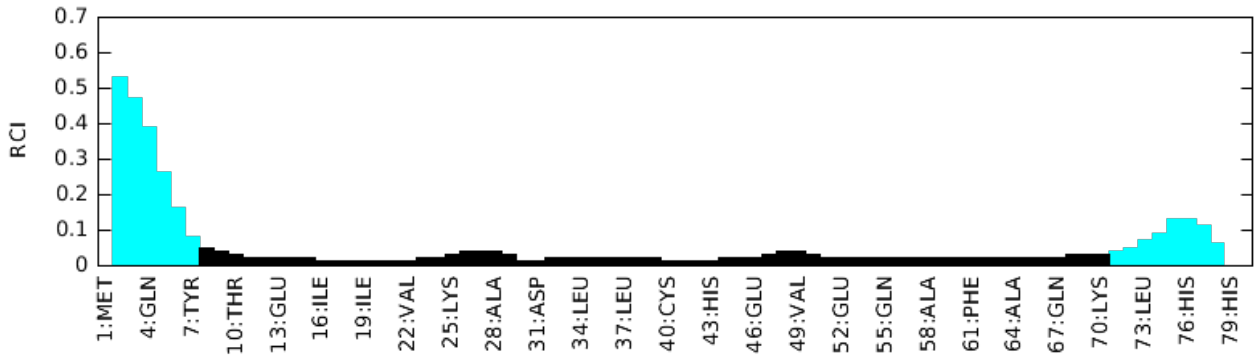
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

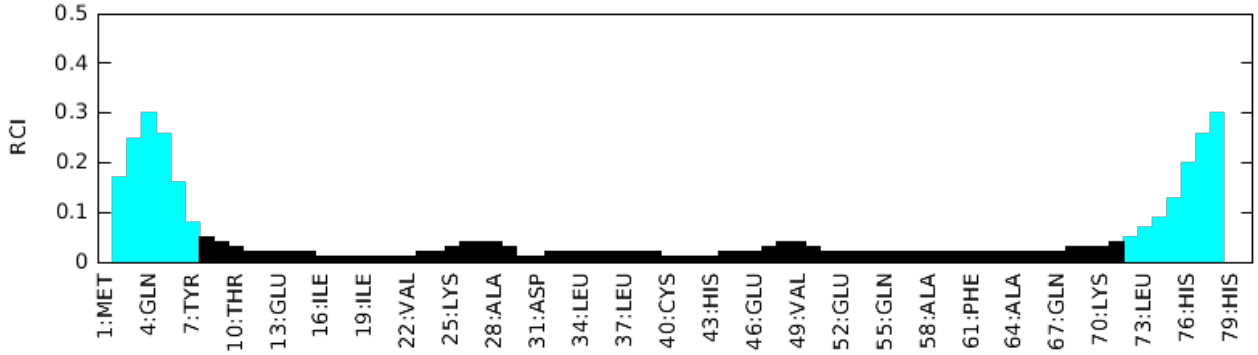
7.1.5 Random Coil Index (RCI) plots (i)

The images below report *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.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B:



8 NMR restraints analysis

8.1 Conformationally restricting restraints

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	1484
Intra-residue (|i-j|=0)	0
Sequential (|i-j|=1)	204
Medium range (|i-j|>1 and |i-j|<5)	856
Long range (|i-j|≥5)	116
Inter-chain	308
Total dihedral-anlge restraints	252
Total hydrogen bond restraints	244
Total disulfide bond restraints	0
Number of unmapped restraints	0
Number of restraints per residue	10.3
Number of long range restraints per residue	0.8

8.2 Residual restraint violations

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

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-anlge violations per model

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

Bins (°)	Average number of violations per model	Max (°)
1.0-10.0 (Small)	0.1	1.1
10.0-20.0 (Medium)	None	None
>20.0 (Large)	None	None

9 Distance violation analysis

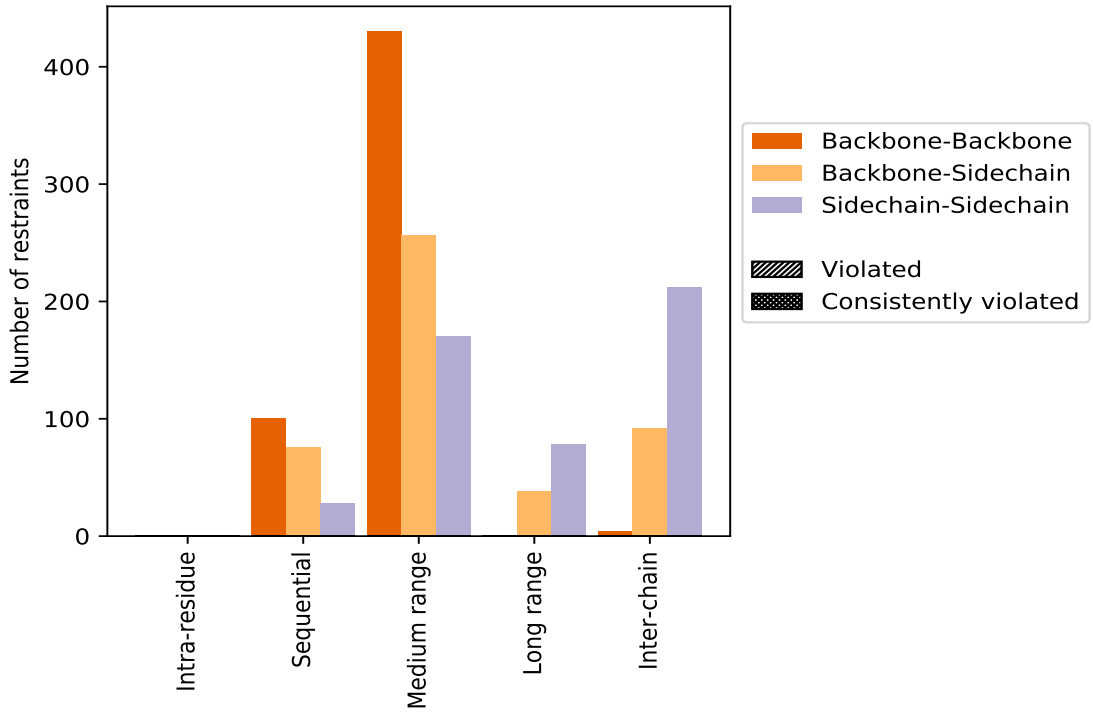
9.1 Summary of distance violations

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)	204	13.7	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	100	6.7	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	76	5.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	28	1.9	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	856	57.7	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	430	29.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	256	17.3	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	170	11.5	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	116	7.8	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	38	2.6	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	78	5.3	0	0.0	0.0	0	0.0	0.0
Inter-chain	308	20.8	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	4	0.3	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	92	6.2	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	212	14.3	0	0.0	0.0	0	0.0	0.0
Total	1484	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	534	36.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	462	31.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	488	32.9	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



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

9.2 Distance violation statistics for each model

No violations found

9.3 Distance violation statistics for the ensemble

No violations found

9.4 Most violated distance restraints in the ensemble

No violations found

9.5 All distance violations

No violations found

10 Dihedral-anlge violation analysis

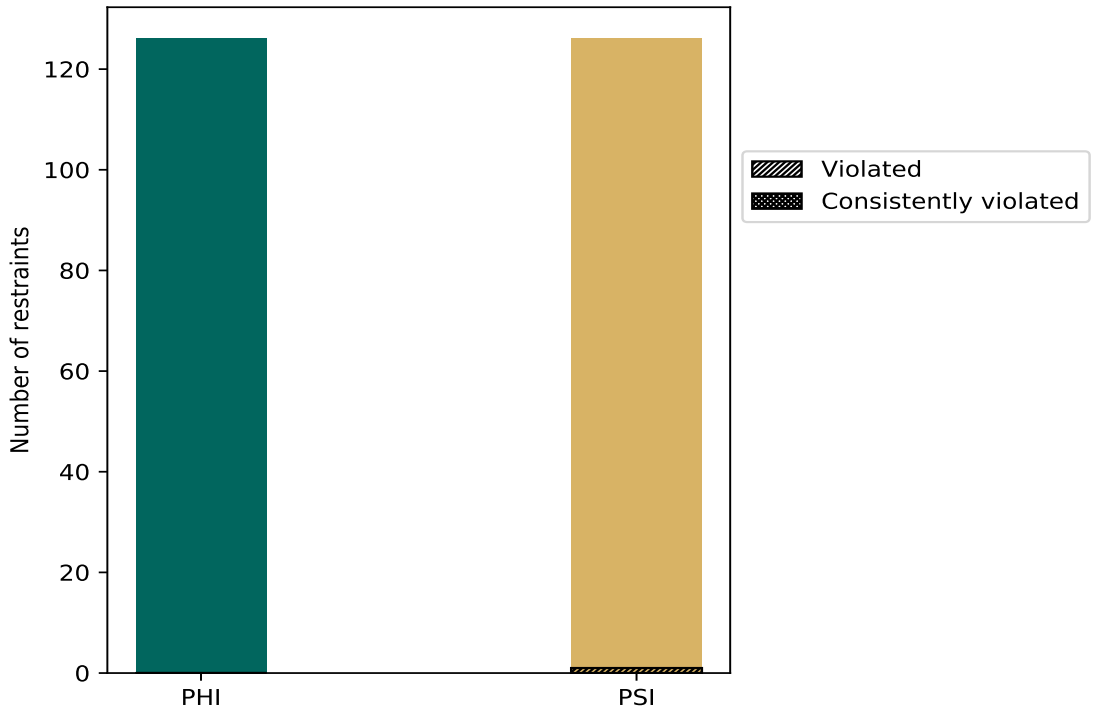
10.1 Summary of dihedral-anlge violations

The following table provides the summary of dihedral-anlge violations in different dihedral-anlge 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	126	50.0	0	0.0	0.0	0	0.0	0.0
PSI	126	50.0	1	0.8	0.4	0	0.0	0.0
Total	252	100.0	1	0.4	0.4	0	0.0	0.0

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

10.1.1 Bar chart : Distribution of dihedral-anlges and violations



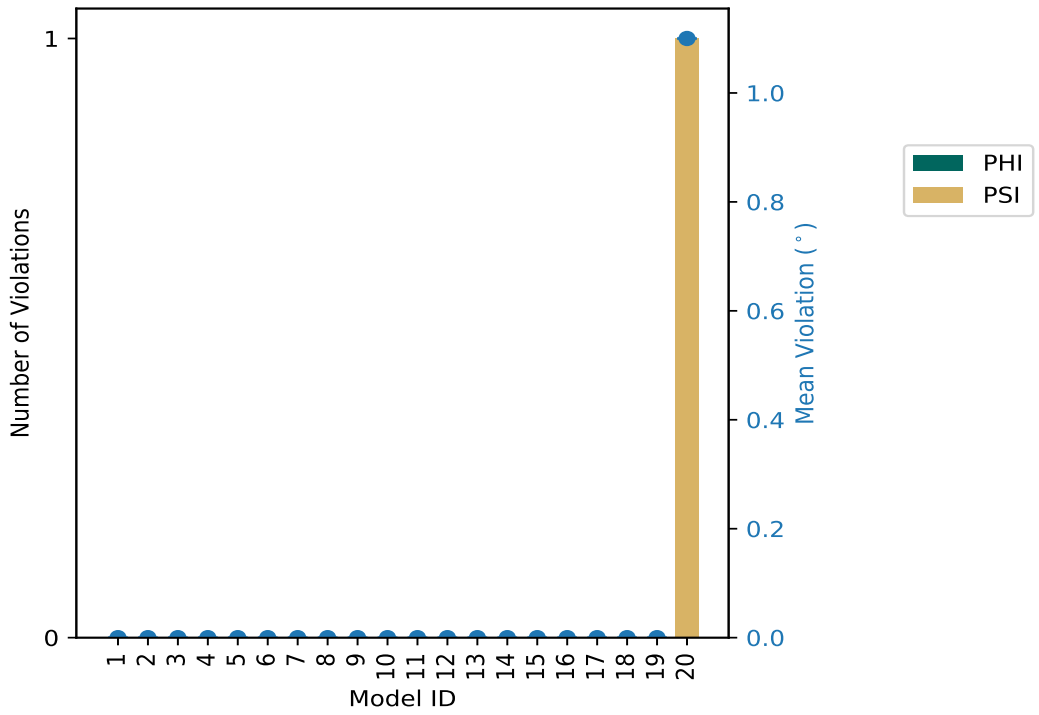
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

10.2 Dihedral-anlge violation statistics in each model

The following table provides the dihedral-anlge 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 (°)	Std. deviation (°)
	PHI	PSI	Total			
1	0	0	0	0.0	0.0	0.0
2	0	0	0	0.0	0.0	0.0
3	0	0	0	0.0	0.0	0.0
4	0	0	0	0.0	0.0	0.0
5	0	0	0	0.0	0.0	0.0
6	0	0	0	0.0	0.0	0.0
7	0	0	0	0.0	0.0	0.0
8	0	0	0	0.0	0.0	0.0
9	0	0	0	0.0	0.0	0.0
10	0	0	0	0.0	0.0	0.0
11	0	0	0	0.0	0.0	0.0
12	0	0	0	0.0	0.0	0.0
13	0	0	0	0.0	0.0	0.0
14	0	0	0	0.0	0.0	0.0
15	0	0	0	0.0	0.0	0.0
16	0	0	0	0.0	0.0	0.0
17	0	0	0	0.0	0.0	0.0
18	0	0	0	0.0	0.0	0.0
19	0	0	0	0.0	0.0	0.0
20	0	1	1	1.1	1.1	0.0

10.2.1 Bar graph : Dihedral violation statistics for each model



The mean and the standard deviation are shown in blue with respect to the y axis on the right

10.3 Violation statistics in the ensemble

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]	%
0	1	1	1	5.0
0	0	0	2	10.0
0	0	0	3	15.0
0	0	0	4	20.0
0	0	0	5	25.0
0	0	0	6	30.0
0	0	0	7	35.0
0	0	0	8	40.0
0	0	0	9	45.0
0	0	0	10	50.0
0	0	0	11	55.0
0	0	0	12	60.0

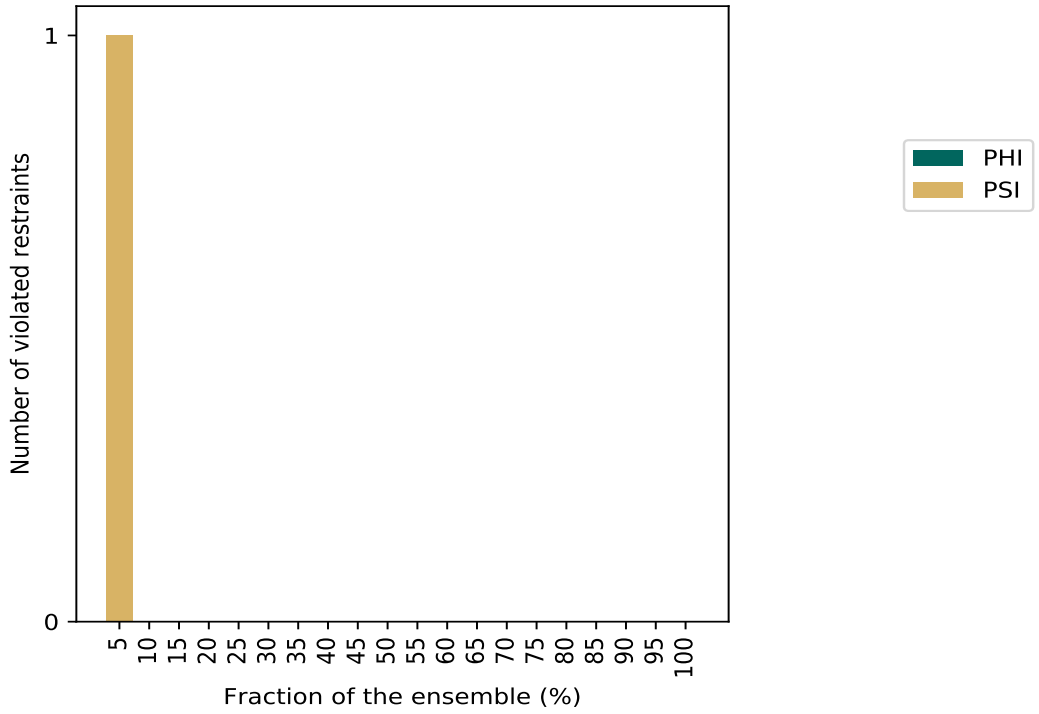
Continued on next page...

Continued from previous page...

Number of violated restraints			Fraction of the ensemble	
PHI	PSI	Total	Count[1]	%
0	0	0	13	65.0
0	0	0	14	70.0
0	0	0	15	75.0
0	0	0	16	80.0
0	0	0	17	85.0
0	0	0	18	90.0
0	0	0	19	95.0
0	0	0	20	100.0

[1] Number of models with violations

10.3.1 Bar graph : Dihedral-anlge Violation statistics for the ensemble



10.4 Most violated dihedral-anlge restraints

No violations found

10.5 All violated dihedral-anlge restraints

10.5.1 Histogram : Distribution of violations

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

10.5.2 Table: All violated dihedral-anlge restraints

The following table lists the absolute value of the violation for each restraint in the ensemble sorted by its 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 (°)
(1,72)	1:A:46:GLU:N	1:A:46:GLU:CA	1:A:46:GLU:C	1:A:47:ARG:N	20	1.1