wwPDB NMR Structure Validation Summary Report

Jul 2, 2020 – 12:12 AM CDT

PDB ID	:	2K2E
Title	:	Solution NMR structure of Bordetella pertussis protein BP2786, a Mth938-like domain. Northeast Structural Genomics Consortium target BeR31
Authors	:	Cort, J.R.; Ho, C.K.; Nwosu, C.; Maglaqui, M.; Xiao, R.; Liu, J.; Baran, M.C.; Swapna, G.; Acton, T.B.; Rost, B.; Montelione, G.T.; Kennedy, M.A.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2008-04-01

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:

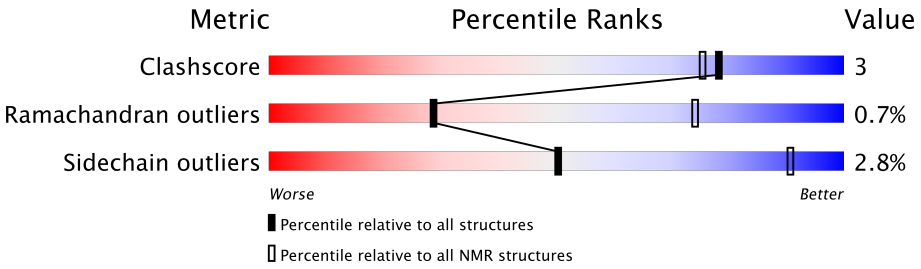
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.6.dev1
BMRB Restraints Analalysis	:	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 60%.

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	158	70% · 29%

2 Ensemble composition and analysis

This entry contains 20 models. Model 11 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:11-A:63, A:88-A:146 (112)	0.55	11

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2367 atoms, of which 1179 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Uncharacterized protein BP2786.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	158	2367	744	1179	219	221	4	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	151	LEU	-	EXPRESSION TAG	UNP Q7VV99
A	152	GLU	-	EXPRESSION TAG	UNP Q7VV99
A	153	HIS	-	EXPRESSION TAG	UNP Q7VV99
A	154	HIS	-	EXPRESSION TAG	UNP Q7VV99
A	155	HIS	-	EXPRESSION TAG	UNP Q7VV99
A	156	HIS	-	EXPRESSION TAG	UNP Q7VV99
A	157	HIS	-	EXPRESSION TAG	UNP Q7VV99
A	158	HIS	-	EXPRESSION TAG	UNP Q7VV99

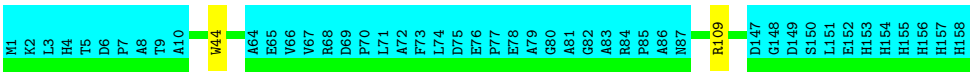
4 Residue-property plots

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: Uncharacterized protein BP2786

Chain A:

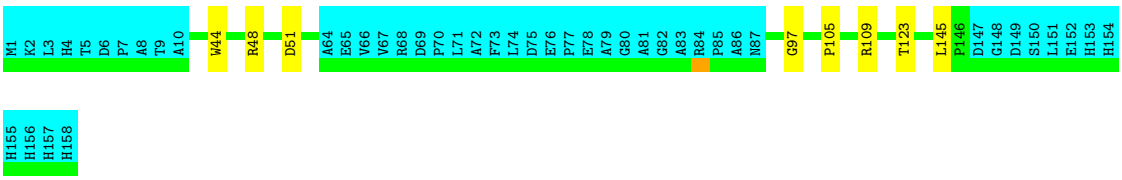


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 11. Colouring as in section 4.1 above.

- Molecule 1: Uncharacterized protein BP2786

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *all calculated structures submitted.*

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

Software name	Classification	Version
AutoStructure	refinement	
AutoStructure	structure solution	
PSVS	refinement	
CNS	refinement	
CNS	structure solution	
X-PLOR NIH	refinement	
X-PLOR NIH	structure solution	

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)	2k2e_nmr.cif
Number of chemical shift lists	1
Total number of shifts	1509
Number of shifts mapped to atoms	971
Number of unparsed shifts	194
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	344
Assignment completeness (well-defined parts)	60%

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	837	856	854	5±2
All	All	16740	17120	17080	104

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 3.

5 of 67 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:48:ARG:HB2	1:A:51:ASP:HB2	0.84	1.48	11	1
1:A:21:TYR:HE1	1:A:23:GLU:HB2	0.73	1.43	16	1
1:A:100:GLN:HG3	1:A:121:MET:HA	0.67	1.64	2	2
1:A:91:VAL:HG11	1:A:133:LEU:HD21	0.64	1.68	6	1
1:A:105:PRO:O	1:A:109:ARG:HB2	0.62	1.94	6	3

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	112/158 (71%)	104±2 (92±2%)	8±2 (7±2%)	1±1 (1±1%)	28 74
All	All	2240/3160 (71%)	2071 (92%)	153 (7%)	16 (1%)	28 74

5 of 6 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	25	ASN	6
1	A	146	PRO	5
1	A	18	GLY	2
1	A	98	ARG	1
1	A	26	GLN	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	85/120 (71%)	83±1 (97±2%)	2±1 (3±2%)	50 91
All	All	1700/2400 (71%)	1653 (97%)	47 (3%)	50 91

5 of 20 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	44	TRP	11
1	A	15	THR	6
1	A	23	GLU	5
1	A	96	THR	4
1	A	12	ASN	2

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 60% for the well-defined parts and 52% for the entire structure.

7.1 Chemical shift list 1

File name: 2k2e_nmr.cif

Chemical shift list name: *nef_chemical_shift_list_2k2e.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	1509
Number of shifts mapped to atoms	971
Number of unparsed shifts	194
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	344
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 194) occurences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
2	A	1	MET	HE%	2.070	0.020	1
3	A	1	MET	HE%	2.070	0.020	1
4	A	1	MET	HE%	2.070	0.020	1
5	A	1	MET	HE%	2.070	0.020	1
6	A	1	MET	HE%	2.070	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 344) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	47	GLN	HGx	2.28	0.02	2
A	139	ARG	HBx	1.91	0.02	2

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	54	ALA	HB%	1.53	0.02	1
A	144	LEU	HD1%	0.76	0.02	1
A	43	SER	HBy	3.84	0.02	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	145	-0.15 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	132	0.40 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	130	0.36 ± 0.13	None needed (< 0.5 ppm)
^{15}N	130	-0.04 ± 0.35	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 60%, i.e. 786 atoms were assigned a chemical shift out of a possible 1321. 24 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	520/544 (96%)	204/216 (94%)	214/224 (96%)	102/104 (98%)
Sidechain	236/707 (33%)	33/410 (8%)	202/263 (77%)	1/34 (3%)
Aromatic	30/70 (43%)	16/36 (44%)	13/29 (45%)	1/5 (20%)
Overall	786/1321 (60%)	253/662 (38%)	429/516 (83%)	104/143 (73%)

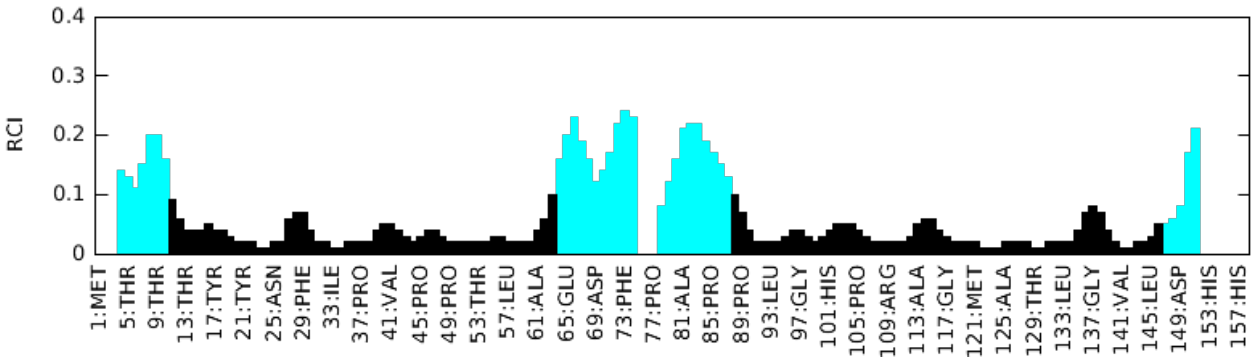
7.1.4 Statistically unusual chemical shifts (i)

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.

Random coil index (RCI) for chain A:



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	1125
Intra-residue (|i-j|=0)	164
Sequential (|i-j|=1)	237
Medium range (|i-j|>1 and |i-j|<5)	230
Long range (|i-j|≥5)	494
Inter-chain	0
Total dihedral-anlge restraints	111
Total hydrogen bond restraints	132
Total disulfide bond restraints	0
Number of unmapped restraints	0
Number of restraints per residue	7.5
Number of long range restraints per residue	3.3

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.

Bins (Å)	Average number of violations per model	Max (Å)
0.1-0.2 (Small)	0.1	0.13
0.2-0.5 (Medium)	None	None
>0.5 (Large)	None	None

8.2.2 Average number of dihedral-anlge violations per model

Dihedral-anlge violations less than 1° are not included in the calculation.There are no dihedral-anlge violations

9 Distance violation analysis

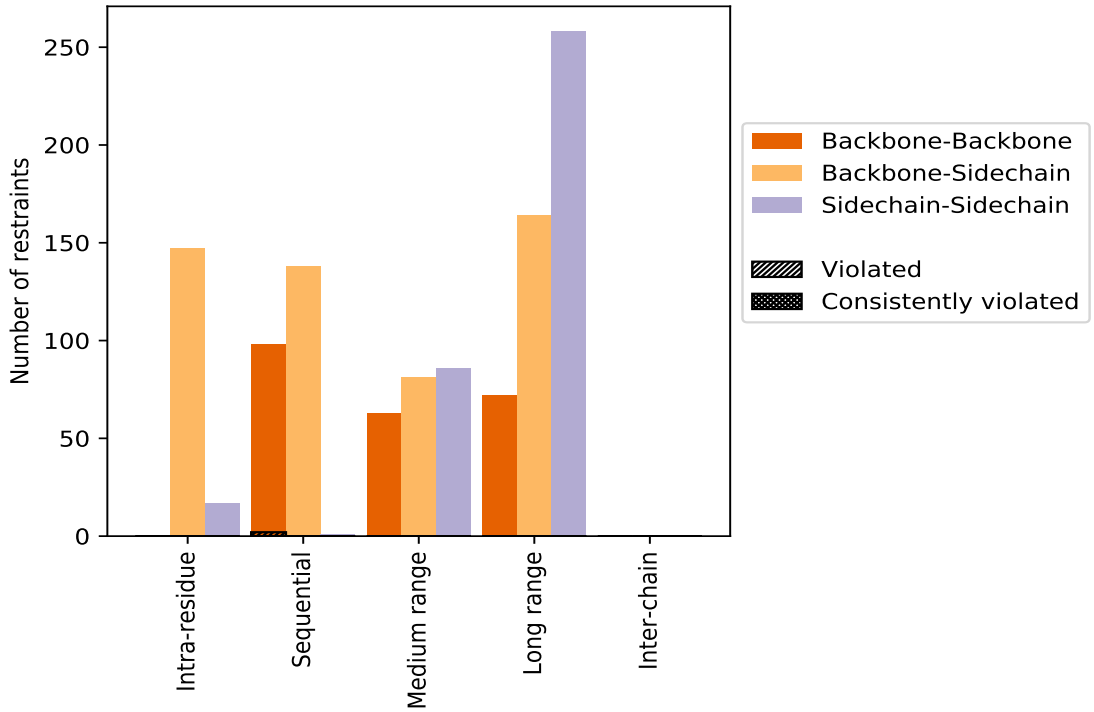
9.1 Summary of distance violations

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)	164	14.6	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	147	13.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	17	1.5	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	237	21.1	2	0.8	0.2	0	0.0	0.0
Backbone-Backbone	98	8.7	2	2.0	0.2	0	0.0	0.0
Backbone-Sidechain	138	12.3	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	1	0.1	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	230	20.4	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	63	5.6	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	81	7.2	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	86	7.6	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	494	43.9	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	72	6.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	164	14.6	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	258	22.9	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
Total	1125	100.0	2	0.2	0.2	0	0.0	0.0
Backbone-Backbone	233	20.7	2	0.9	0.2	0	0.0	0.0
Backbone-Sidechain	530	47.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	362	32.2	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

The following table provides the distance violation statistics for each model in the ensemble. Violations less than 0.1 Å are not included in the statistics.

Model ID	Number of violations						Mean (Å)	Max (Å)	SD[6] (Å)
	IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total			
1	0	0	0	0	0	0	0.0	0.0	0.0
2	0	0	0	0	0	0	0.0	0.0	0.0
3	0	0	0	0	0	0	0.0	0.0	0.0
4	0	0	0	0	0	0	0.0	0.0	0.0
5	0	0	0	0	0	0	0.0	0.0	0.0
6	0	0	0	0	0	0	0.0	0.0	0.0
7	0	0	0	0	0	0	0.0	0.0	0.0
8	0	0	0	0	0	0	0.0	0.0	0.0
9	0	0	0	0	0	0	0.0	0.0	0.0
10	0	0	0	0	0	0	0.0	0.0	0.0
11	0	0	0	0	0	0	0.0	0.0	0.0
12	0	0	0	0	0	0	0.0	0.0	0.0

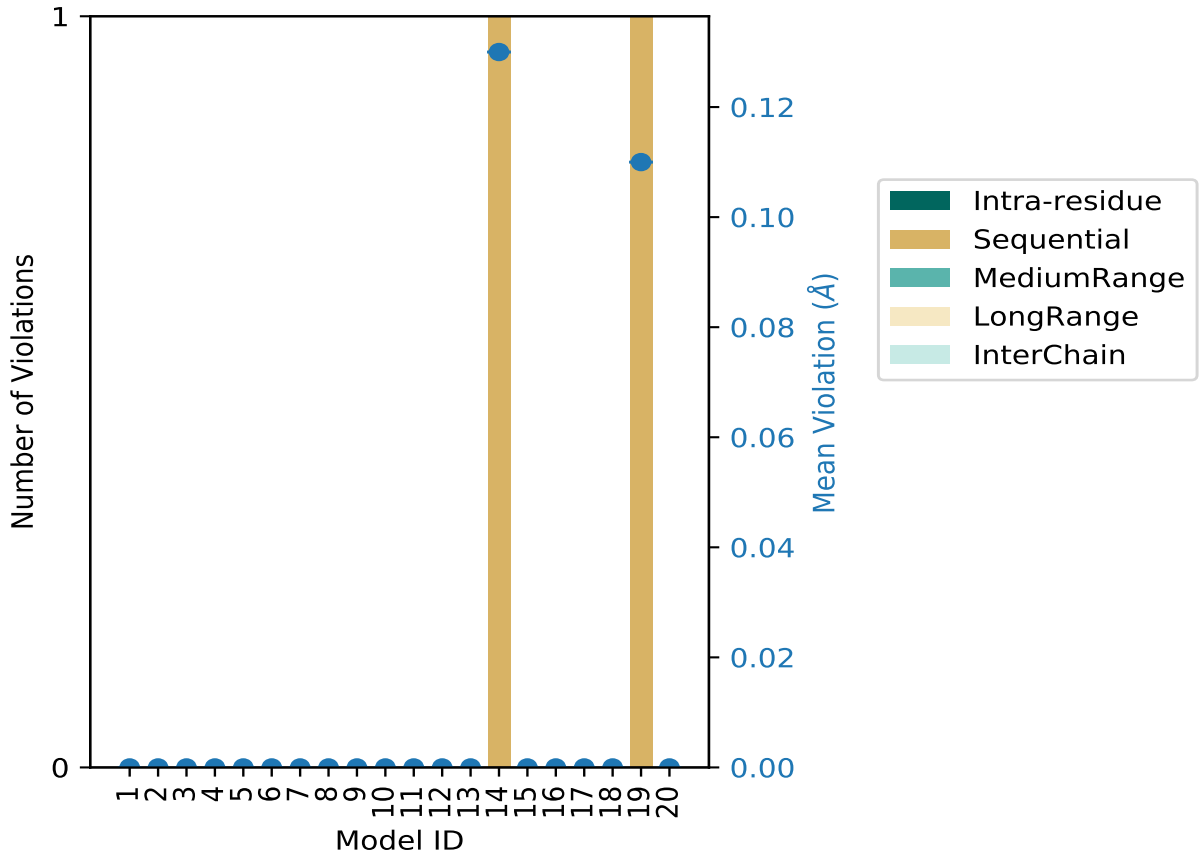
Continued on next page...

Continued from previous page...

Model ID	Number of violations						Mean (Å)	Max (Å)	SD[6] (Å)
	IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total			
13	0	0	0	0	0	0	0.0	0.0	0.0
14	0	1	0	0	0	1	0.13	0.13	0.0
15	0	0	0	0	0	0	0.0	0.0	0.0
16	0	0	0	0	0	0	0.0	0.0	0.0
17	0	0	0	0	0	0	0.0	0.0	0.0
18	0	0	0	0	0	0	0.0	0.0	0.0
19	0	1	0	0	0	1	0.11	0.11	0.0
20	0	0	0	0	0	0	0.0	0.0	0.0

[1]Intra-residue restraints, [2]Sequential restraints, [3]Medium range restraints, [4]Long range restraints, [5]Inter-chain restraints, [6]Standard deviation

9.2.1 Bar graph : Distance Violation statistics for each model



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

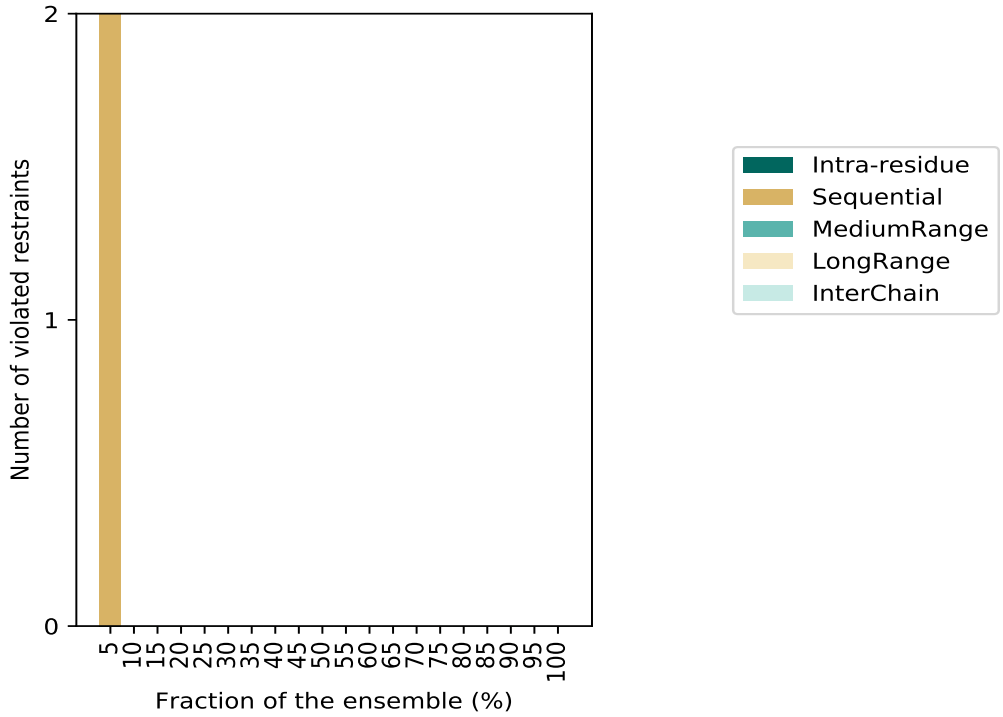
9.3 Distance violation statistics for the ensemble

Violation analysis may find that some restraints are violated in 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 the ensemble. In total, 1123(IR:164, SQ:235, MR:230, LR:494, IC:0) restraints are not violated in the ensemble.

Number of violated restraints						Fraction of the ensemble	
IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total	Count[6]	%
0	2	0	0	0	2	1	5.0
0	0	0	0	0	0	2	10.0
0	0	0	0	0	0	3	15.0
0	0	0	0	0	0	4	20.0
0	0	0	0	0	0	5	25.0
0	0	0	0	0	0	6	30.0
0	0	0	0	0	0	7	35.0
0	0	0	0	0	0	8	40.0
0	0	0	0	0	0	9	45.0
0	0	0	0	0	0	10	50.0
0	0	0	0	0	0	11	55.0
0	0	0	0	0	0	12	60.0
0	0	0	0	0	0	13	65.0
0	0	0	0	0	0	14	70.0
0	0	0	0	0	0	15	75.0
0	0	0	0	0	0	16	80.0
0	0	0	0	0	0	17	85.0
0	0	0	0	0	0	18	90.0
0	0	0	0	0	0	19	95.0
0	0	0	0	0	0	20	100.0

[1]Intra-residue restraints, [2]Sequential restraints, [3]Medium range restraints, [4]Long range restraints, [5]Inter-chain restraints, [6] Number of models with violations

9.3.1 Bar graph : Distance violation statistics for the ensemble



9.4 Most violated distance restraints in the ensemble

No violations found

9.5 All distance violations

9.5.1 Histogram : Distribution of distance 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

9.5.2 Table : All distance violations

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. Rows with same key represent combinatorial or ambiguous restraints and are counted as a single restraint.

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(1,69)	1:A:38:GLU:HA	1:A:39:GLY:H	14	0.13

Continued on next page...

Continued from previous page...

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(1,49)	1:A:137:GLY:H	1:A:138:ARG:H	19	0.11

10 Dihedral-anlge violation analysis

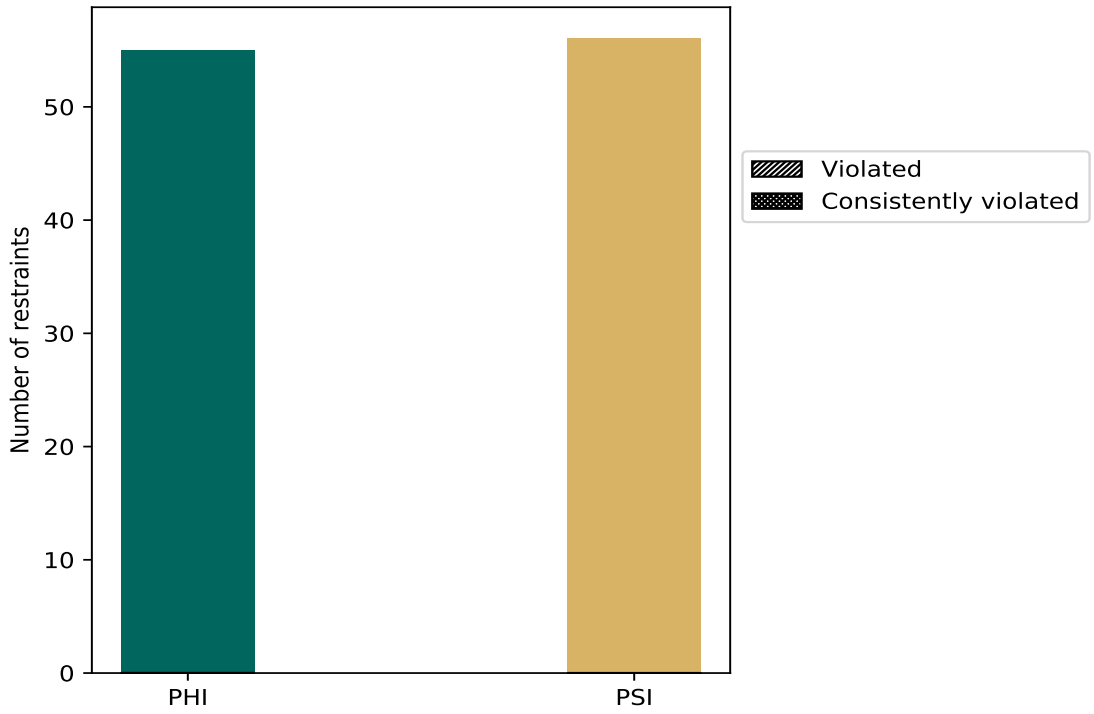
10.1 Summary of dihedral-anlge violations

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	55	49.5	0	0.0	0.0	0	0.0	0.0
PSI	56	50.5	0	0.0	0.0	0	0.0	0.0
Total	111	100.0	0	0.0	0.0	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

No violations found

10.3 Violation statistics in the ensemble

No violations found

10.4 Most violated dihedral-anlge restraints

No violations found

10.5 All violated dihedral-anlge restraints

No violations found