wwPDB NMR Structure Validation Summary Report (i)

Jun 6, 2023 – 06:35 pm BST

PDB ID	:	3ZFJ
BMRB ID	:	18943
Title	:	N-terminal domain of pneumococcal PhtD protein with bound Zn(II)
Authors	:	Bersch, B.; Bougault, C.; Favier, A.; Gabel, F.; Roux, L.; Vernet, T.; Durmort, C.
Deposited on	:	2012-12-11

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

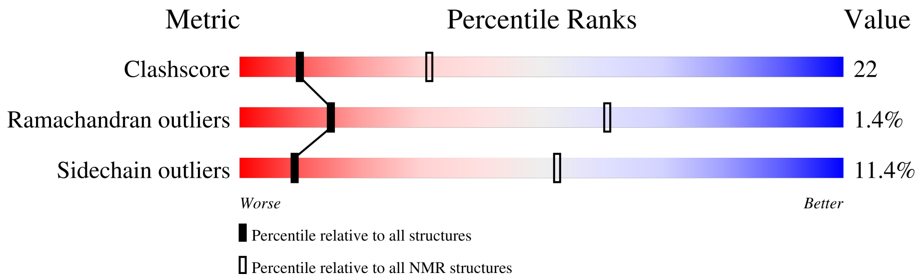
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 91%.

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	138	

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:51-A:106, A:111-A:158 (104)	0.58	3

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 3 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1897 atoms, of which 935 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PNEUMOCOCCAL HISTIDINE TRIAD PROTEIN D.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	120	1896	603	935	163	194	1	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	29	GLY	-	expression tag	UNP Q04KS8

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

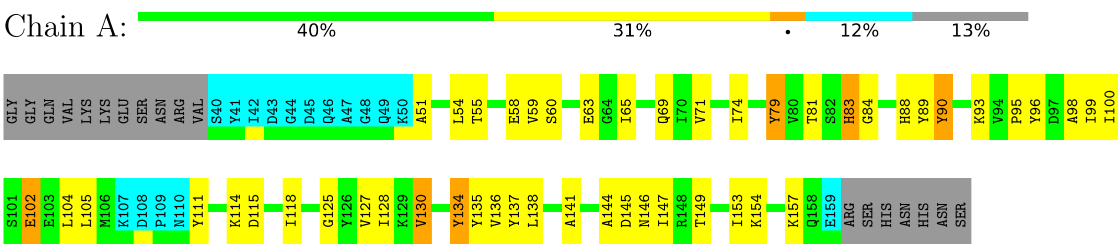
Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

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.

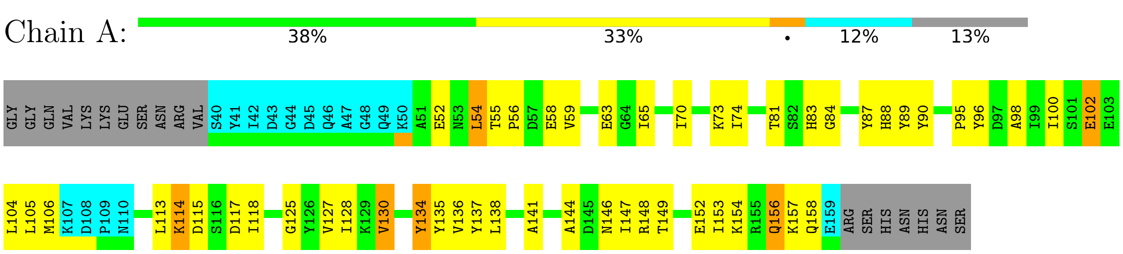
- Molecule 1: PNEUMOCOCCAL HISTIDINE TRIAD PROTEIN D



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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: PNEUMOCOCCAL HISTIDINE TRIAD PROTEIN D



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *UNIO10, ARIA*.

Of the 1000 calculated structures, 20 were deposited, based on the following criterion: *TOTAL ENERGY*.

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

Software name	Classification	Version
CNS	refinement	
CYANA	structure solution	
CNS	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1473
Number of shifts mapped to atoms	1472
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.58±0.20	1±2/862 (0.2± 0.2%)	0.55±0.06	0±1/1163 (0.0± 0.1%)
All	All	0.62	28/17240 (0.2%)	0.55	8/23260 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.1±0.3
All	All	0	2

5 of 10 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	137	TYR	CE1-CZ	19.73	1.64	1.38	6	1
1	A	134	TYR	CE2-CZ	18.70	1.62	1.38	15	1
1	A	134	TYR	CE1-CZ	-18.65	1.14	1.38	15	1
1	A	137	TYR	CE2-CZ	-18.45	1.14	1.38	6	1
1	A	79	TYR	CE2-CZ	-11.23	1.24	1.38	11	4

5 of 8 unique angle 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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	137	TYR	CD1-CE1-CZ	-7.45	113.10	119.80	6	1
1	A	137	TYR	CE1-CZ-OH	-7.19	100.68	120.10	6	1
1	A	134	TYR	CZ-CE2-CD2	-6.87	113.62	119.80	15	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	134	TYR	CD1-CE1-CZ	6.75	125.88	119.80	15	1
1	A	134	TYR	CE1-CZ-OH	6.47	137.58	120.10	15	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	111	TYR	Sidechain	1
1	A	137	TYR	Sidechain	1

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	846	831	830	36±5
All	All	16940	16620	16600	730

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 22.

5 of 217 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:65:ILE:HD12	1:A:84:GLY:HA3	0.90	1.41	13	17
1:A:60:SER:HB3	1:A:66:ASN:HA	0.82	1.48	12	4
1:A:51:ALA:H	1:A:89:TYR:HB3	0.81	1.33	12	2
1:A:138:LEU:HD23	1:A:141:ALA:HA	0.81	1.51	14	20
1:A:56:PRO:HG3	1:A:95:PRO:HG3	0.80	1.53	3	5

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	104/138 (75%)	92±1 (89±1%)	10±1 (10±1%)	2±1 (1±1%)	15	61
All	All	2080/2760 (75%)	1849 (89%)	201 (10%)	30 (1%)	15	61

All 5 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	144	ALA	18
1	A	51	ALA	8
1	A	146	ASN	2
1	A	66	ASN	1
1	A	158	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	91/120 (76%)	81±2 (89±2%)	10±2 (11±2%)	9	52
All	All	1820/2400 (76%)	1613 (89%)	207 (11%)	9	52

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	83	HIS	20
1	A	89	TYR	20
1	A	134	TYR	19
1	A	130	VAL	16
1	A	96	TYR	12

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)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 91% for the well-defined parts and 90% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *3zfj*

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	1473
Number of shifts mapped to atoms	1472
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

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

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	39	VAL	C	175.787	.	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
$^{13}C_{\alpha}$	119	0.48 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	110	0.34 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	117	0.12 ± 0.13	None needed (< 0.5 ppm)
^{15}N	116	0.31 ± 0.31	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 91%, i.e. 1315 atoms were assigned a chemical shift out of a possible 1444. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	521/523 (100%)	213/213 (100%)	206/208 (99%)	102/102 (100%)
Sidechain	696/799 (87%)	471/514 (92%)	221/255 (87%)	4/30 (13%)
Aromatic	98/122 (80%)	48/56 (86%)	42/58 (72%)	8/8 (100%)
Overall	1315/1444 (91%)	732/783 (93%)	469/521 (90%)	114/140 (81%)

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	129	LYS	HG2	0.02	0.13 – 2.61	-5.5

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	2676
Intra-residue (|i-j|=0)	1102
Sequential (|i-j|=1)	582
Medium range (|i-j|>1 and |i-j|<5)	296
Long range (|i-j|≥5)	696
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	186
Number of unmapped restraints	2
Number of restraints per residue	20.7
Number of long range restraints per residue[1]	5.0

[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.

Bins (Å)	Average number of violations per model	Max (Å)
0.1-0.2 (Small)	101.3	0.2
0.2-0.5 (Medium)	96.0	0.5
>0.5 (Large)	143.4	7.52

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)	16.6	9.8
10.0-20.0 (Medium)	0.3	14.8
>20.0 (Large)	0.9	82.2

9 Distance violation analysis (i)

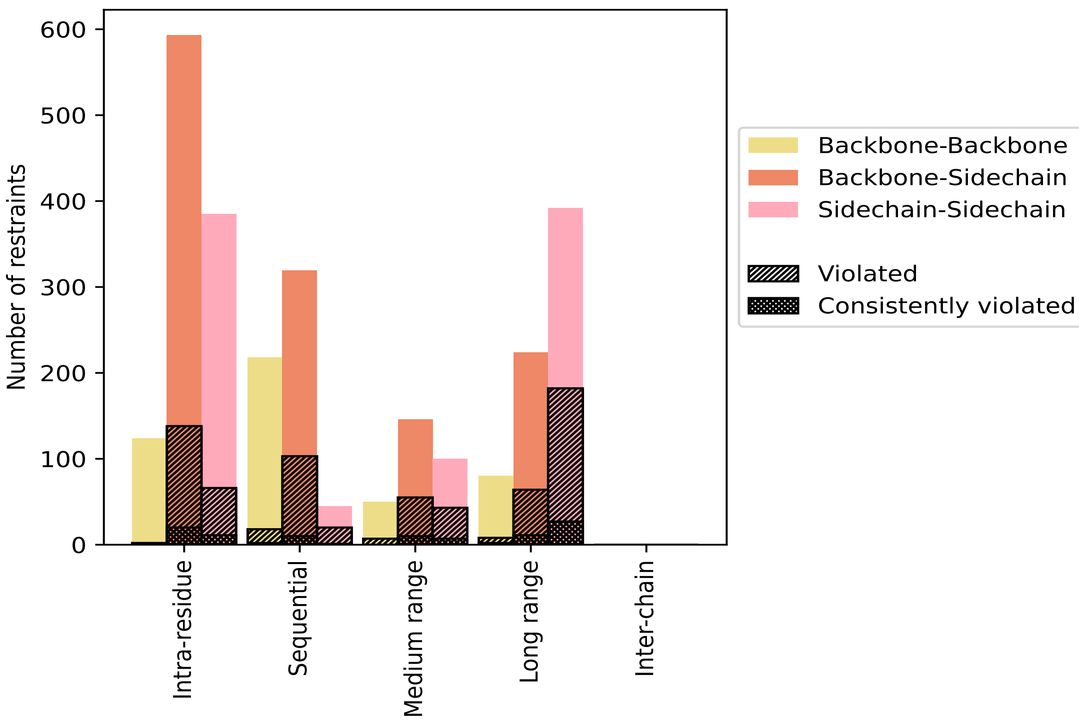
9.1 Summary of distance violations (i)

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)	1102	41.2	206	18.7	7.7	31	2.8	1.2
Backbone-Backbone	124	4.6	2	1.6	0.1	0	0.0	0.0
Backbone-Sidechain	593	22.2	138	23.3	5.2	20	3.4	0.7
Sidechain-Sidechain	385	14.4	66	17.1	2.5	11	2.9	0.4
Sequential (|i-j|=1)	582	21.7	141	24.2	5.3	13	2.2	0.5
Backbone-Backbone	218	8.1	18	8.3	0.7	2	0.9	0.1
Backbone-Sidechain	319	11.9	103	32.3	3.8	10	3.1	0.4
Sidechain-Sidechain	45	1.7	20	44.4	0.7	1	2.2	0.0
Medium range (|i-j|>1 & |i-j|<5)	296	11.1	105	35.5	3.9	17	5.7	0.6
Backbone-Backbone	50	1.9	7	14.0	0.3	0	0.0	0.0
Backbone-Sidechain	146	5.5	55	37.7	2.1	10	6.8	0.4
Sidechain-Sidechain	100	3.7	43	43.0	1.6	7	7.0	0.3
Long range (|i-j|≥5)	696	26.0	254	36.5	9.5	40	5.7	1.5
Backbone-Backbone	80	3.0	8	10.0	0.3	2	2.5	0.1
Backbone-Sidechain	224	8.4	64	28.6	2.4	11	4.9	0.4
Sidechain-Sidechain	392	14.6	182	46.4	6.8	27	6.9	1.0
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	0	0.0	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	2676	100.0	706	26.4	26.4	101	3.8	3.8
Backbone-Backbone	472	17.6	35	7.4	1.3	4	0.8	0.1
Backbone-Sidechain	1282	47.9	360	28.1	13.5	51	4.0	1.9
Sidechain-Sidechain	922	34.5	311	33.7	11.6	46	5.0	1.7

[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)

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)	Median (Å)
	IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total				
1	101	62	50	125	0	338	0.67	4.82	0.77	0.43
2	94	57	51	126	0	328	0.7	4.96	0.81	0.41
3	108	66	55	127	0	356	0.66	4.54	0.75	0.37
4	104	64	49	124	0	341	0.61	4.62	0.67	0.38
5	108	61	59	142	0	370	0.77	7.26	0.9	0.44
6	95	69	59	143	0	366	0.76	5.06	0.83	0.44
7	90	62	46	122	0	320	0.55	4.61	0.59	0.36
8	101	65	50	118	0	334	0.68	5.04	0.75	0.4
9	92	64	51	133	0	340	0.67	5.68	0.88	0.34
10	101	62	57	118	0	338	0.71	4.54	0.8	0.41
11	101	63	49	130	0	343	0.65	5.11	0.76	0.36

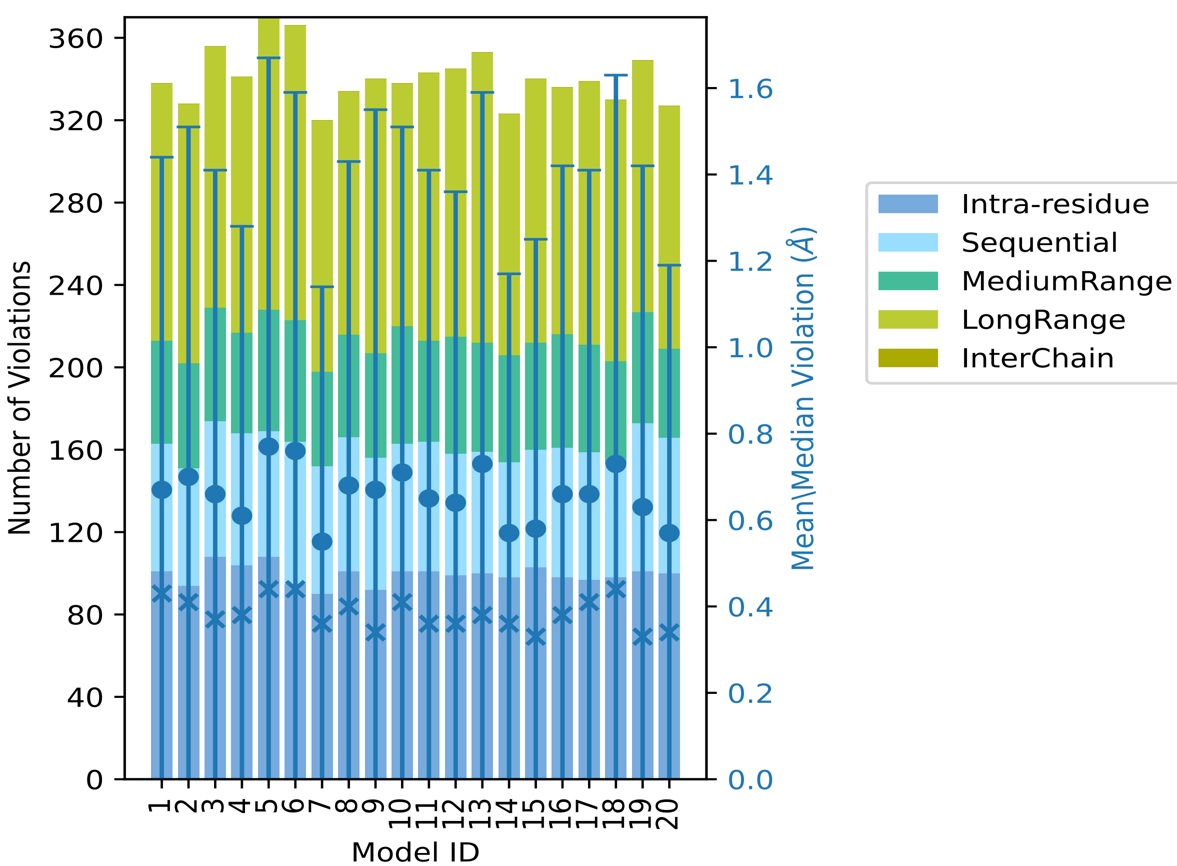
Continued on next page...

Continued from previous page...

Model ID	Number of violations						Mean (Å)	Max (Å)	SD[6] (Å)	Median (Å)
	IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total				
12	99	59	57	130	0	345	0.64	4.92	0.72	0.36
13	100	59	53	141	0	353	0.73	4.76	0.86	0.38
14	98	56	52	117	0	323	0.57	3.65	0.6	0.36
15	103	57	52	128	0	340	0.58	5.23	0.67	0.33
16	98	63	55	120	0	336	0.66	4.79	0.76	0.38
17	97	62	52	128	0	339	0.66	4.4	0.75	0.41
18	98	55	50	127	0	330	0.73	7.52	0.9	0.44
19	101	72	54	122	0	349	0.63	5.31	0.79	0.33
20	100	66	43	118	0	327	0.57	4.01	0.62	0.34

[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 (i)



The mean(dot),median(x) and the standard deviation are shown in blue with respect to the y axis on the right

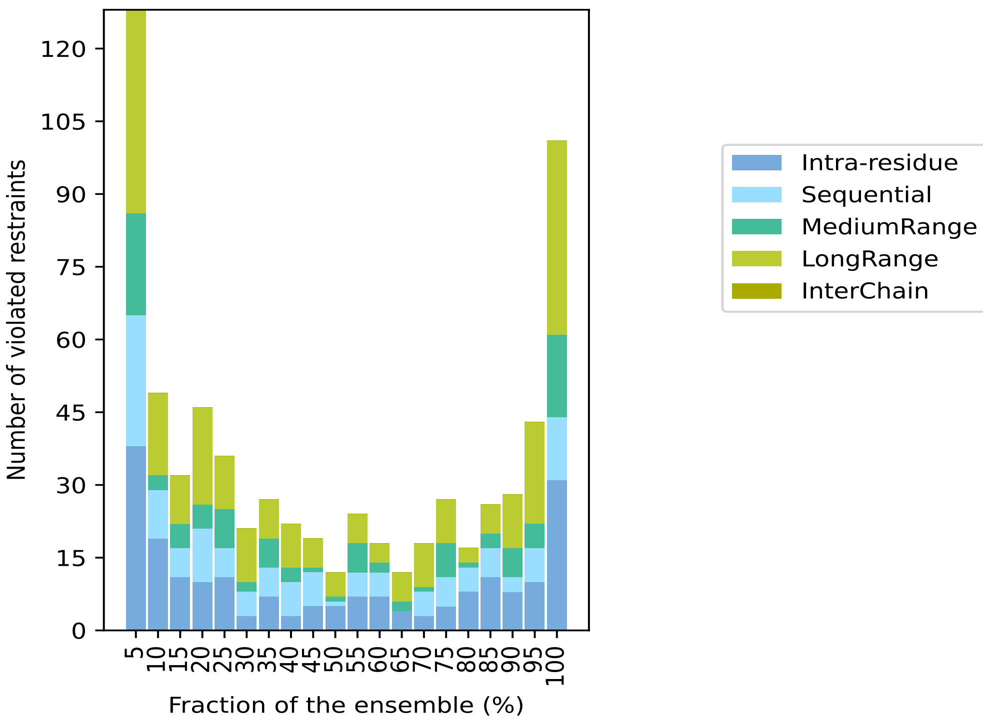
9.3 Distance violation statistics for the ensemble (i)

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, 1970(IR:896, SQ:441, MR:191, LR:442, 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]	%
38	27	21	42	0	128	1	5.0
19	10	3	17	0	49	2	10.0
11	6	5	10	0	32	3	15.0
10	11	5	20	0	46	4	20.0
11	6	8	11	0	36	5	25.0
3	5	2	11	0	21	6	30.0
7	6	6	8	0	27	7	35.0
3	7	3	9	0	22	8	40.0
5	7	1	6	0	19	9	45.0
5	1	1	5	0	12	10	50.0
7	5	6	6	0	24	11	55.0
7	5	2	4	0	18	12	60.0
4	0	2	6	0	12	13	65.0
3	5	1	9	0	18	14	70.0
5	6	7	9	0	27	15	75.0
8	5	1	3	0	17	16	80.0
11	6	3	6	0	26	17	85.0
8	3	6	11	0	28	18	90.0
10	7	5	21	0	43	19	95.0
31	13	17	40	0	101	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

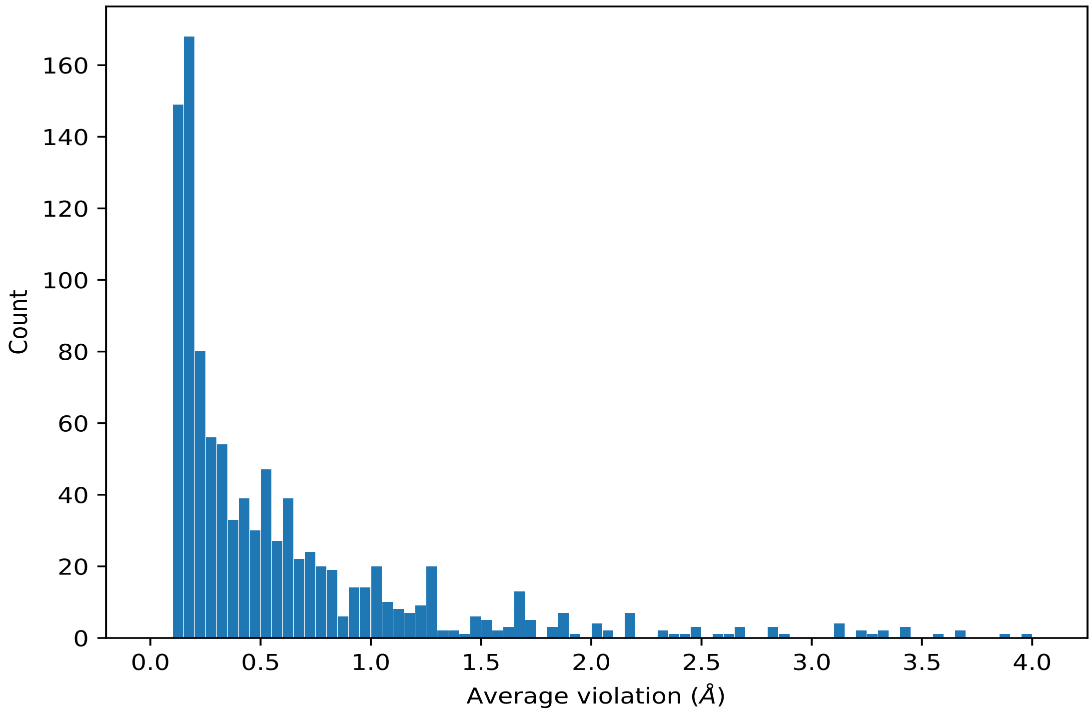
9.3.1 Bar graph : Distance violation statistics for the ensemble (i)



9.4 Most violated distance restraints in the ensemble (i)

9.4.1 Histogram : Distribution of mean distance 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



9.4.2 Table: Most violated distance 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. Rows with same key represent combinatorial or ambiguous restraints and are counted as a single restraint.

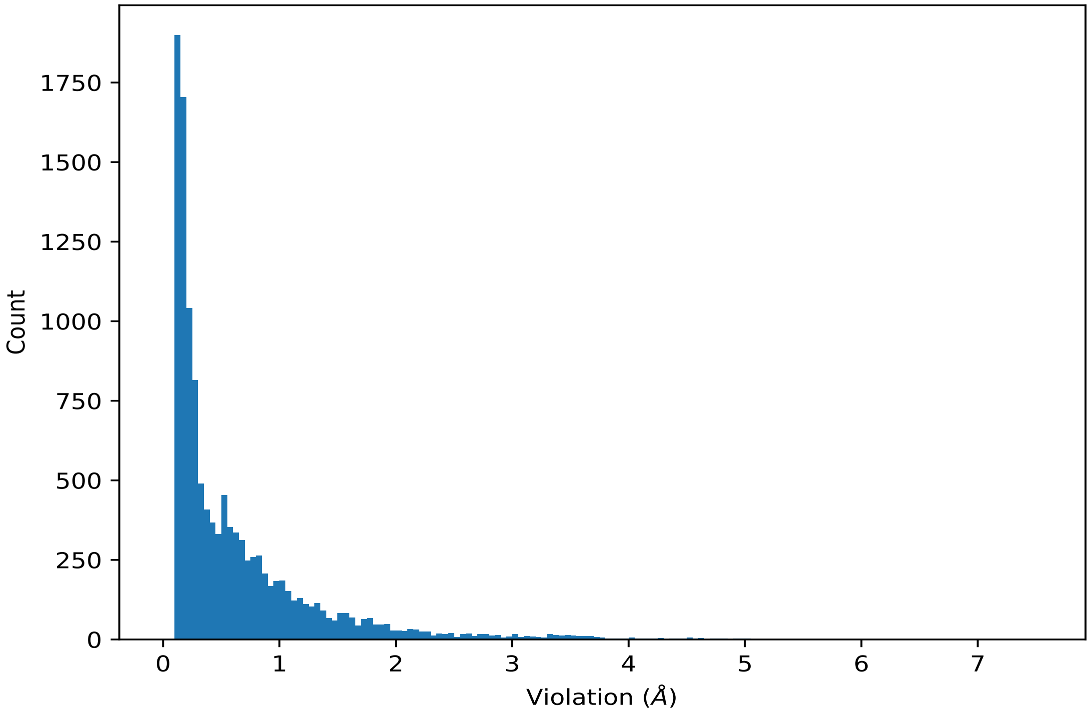
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	20	3.98	1.61	4.78
(1,211)	1:A:137:TYR:HE1	1:A:111:TYR:HE1	20	3.65	1.6	4.5
(1,1119)	1:A:88:HIS:HB2	1:A:90:TYR:HE2	20	3.59	1.73	3.04
(1,156)	1:A:75:THR:HA	1:A:134:TYR:HE1	20	3.34	0.75	3.5
(1,897)	1:A:113:LEU:HB3	1:A:111:TYR:HE1	20	3.22	1.35	3.84
(1,34)	1:A:137:TYR:HD1	1:A:111:TYR:HE1	20	3.11	1.67	4.0
(1,58)	1:A:134:TYR:HD1	1:A:75:THR:HA	20	2.55	0.56	2.68
(1,1425)	1:A:104:LEU:HD22	1:A:100:ILE:HA	20	2.01	0.13	2.02
(1,1425)	1:A:104:LEU:HD21	1:A:100:ILE:HA	20	2.01	0.13	2.02
(1,1425)	1:A:104:LEU:HD23	1:A:100:ILE:HA	20	2.01	0.13	2.02
(1,739)	1:A:95:PRO:HD2	1:A:79:TYR:HD1	20	1.94	1.21	1.08
(1,401)	1:A:94:VAL:HA	1:A:79:TYR:HD1	20	1.85	0.44	1.63

[1]Number of violated models, [2]Standard deviation

9.5 All violated distance restraints (i)

9.5.1 Histogram : Distribution of distance violations (i)

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



9.5.2 Table : All distance violations (i)

The following table provides 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. 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)
(1,1119)	1:A:88:HIS:HB2	1:A:90:TYR:HE2	18	7.52
(1,1119)	1:A:88:HIS:HB2	1:A:90:TYR:HE2	5	7.26
(1,1119)	1:A:88:HIS:HB2	1:A:90:TYR:HE2	9	5.68
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	19	5.31
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	15	5.23
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	11	5.11
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	6	5.06
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	8	5.04
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	9	5.03
(1,33)	1:A:137:TYR:HD1	1:A:111:TYR:HD1	2	4.96

10 Dihedral-angle violation analysis (i)

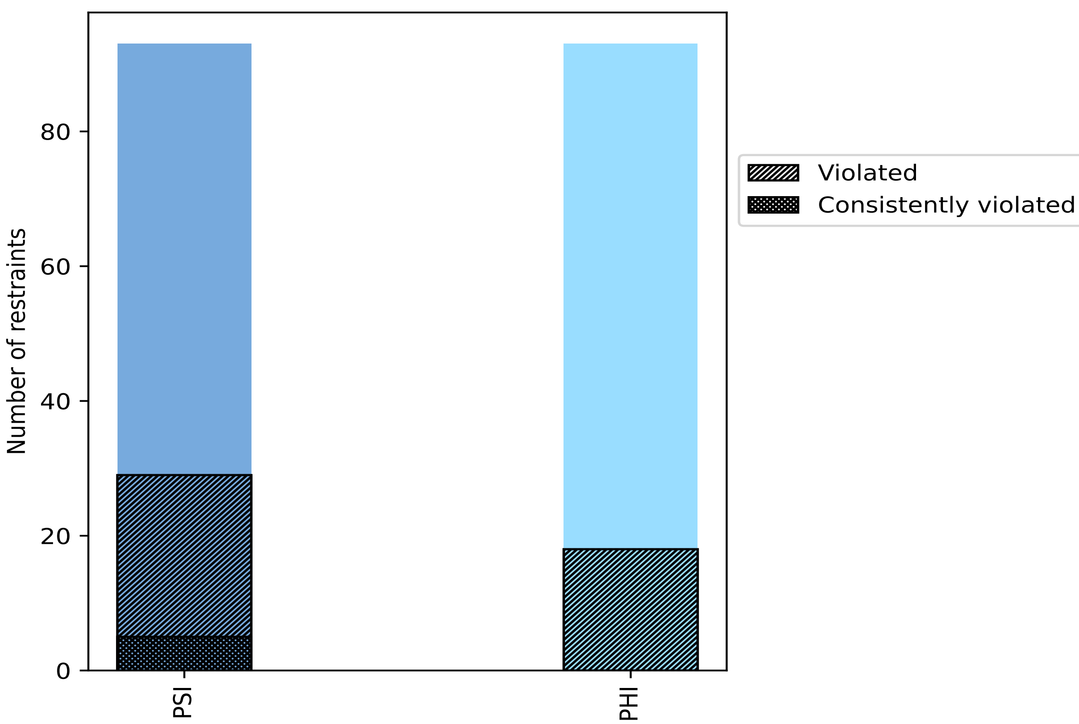
10.1 Summary of dihedral-angle violations (i)

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]
PSI	93	50.0	29	31.2	15.6	5	5.4	2.7
PHI	93	50.0	18	19.4	9.7	0	0.0	0.0
Total	186	100.0	47	25.3	25.3	5	2.7	2.7

[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)



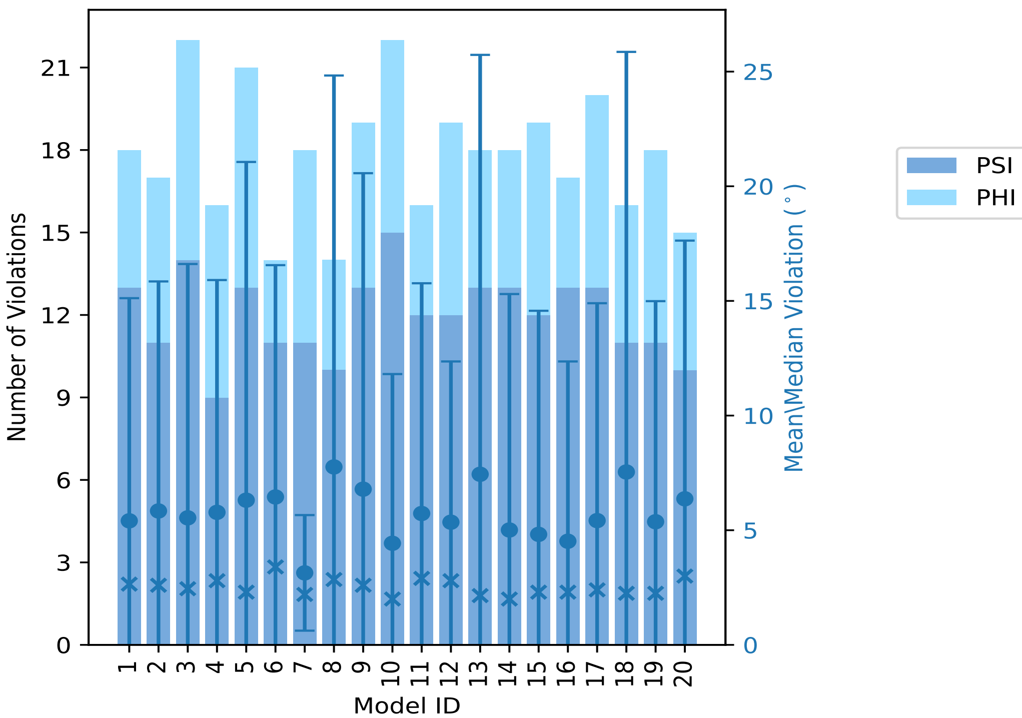
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

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 (°)
	PSI	PHI	Total				
1	13	5	18	5.41	44.6	9.71	2.65
2	11	6	17	5.84	45.3	10.01	2.6
3	14	8	22	5.54	55.0	11.07	2.45
4	9	7	16	5.78	44.3	10.13	2.8
5	13	8	21	6.31	71.2	14.75	2.3
6	11	3	14	6.45	42.3	10.11	3.4
7	11	7	18	3.14	10.4	2.52	2.2
8	10	4	14	7.76	69.0	17.07	2.85
9	13	6	19	6.79	64.5	13.78	2.6
10	15	7	22	4.43	36.7	7.38	2.0
11	12	4	16	5.73	43.9	10.04	2.9
12	12	7	19	5.35	31.8	7.01	2.8
13	13	5	18	7.44	82.2	18.29	2.15
14	13	5	18	5.01	46.7	10.29	2.0
15	12	7	19	4.82	45.8	9.75	2.3
16	13	4	17	4.52	35.6	7.84	2.3
17	13	7	20	5.42	45.4	9.48	2.4
18	11	5	16	7.54	78.0	18.32	2.25
19	11	7	18	5.37	44.2	9.62	2.25
20	10	5	15	6.37	47.7	11.26	3.0

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	
PSI	PHI	Total	Count[1]	%
6	8	14	1	5.0
3	1	4	2	10.0
2	0	2	3	15.0
1	1	2	4	20.0
4	0	4	5	25.0
2	2	4	6	30.0
0	0	0	7	35.0
0	1	1	8	40.0
0	0	0	9	45.0
1	0	1	10	50.0
1	0	1	11	55.0

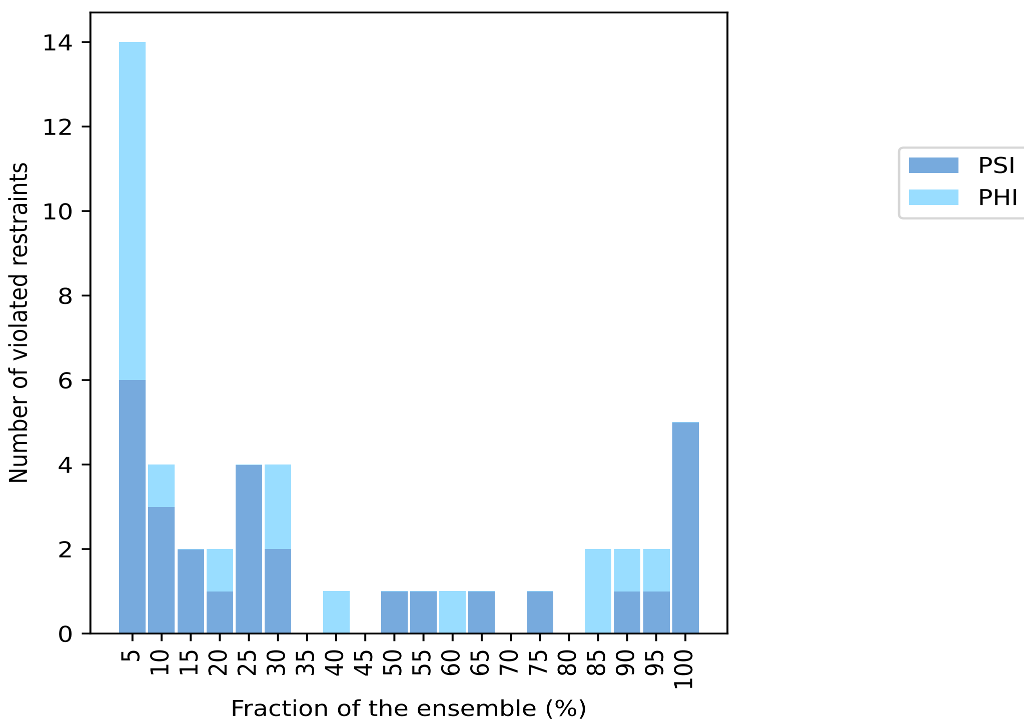
Continued on next page...

Continued from previous page...

Number of violated restraints			Fraction of the ensemble	
PSI	PHI	Total	Count[1]	%
0	1	1	12	60.0
1	0	1	13	65.0
0	0	0	14	70.0
1	0	1	15	75.0
0	0	0	16	80.0
0	2	2	17	85.0
1	1	2	18	90.0
1	1	2	19	95.0
5	0	5	20	100.0

[1] Number of models with violations

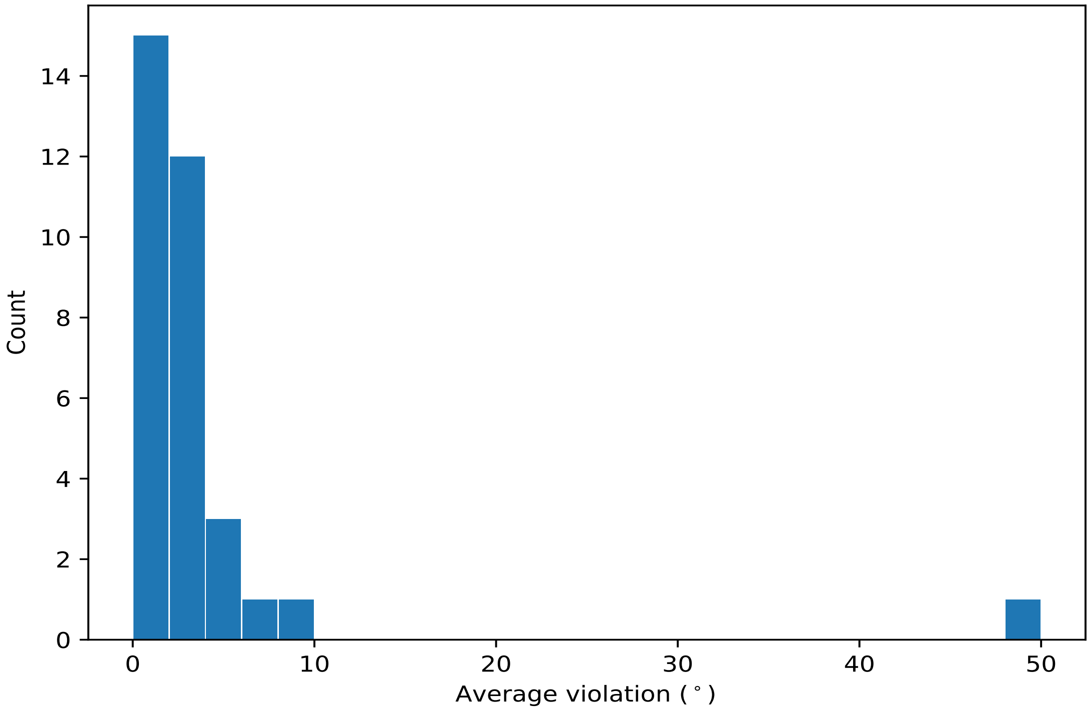
10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble (i)



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.

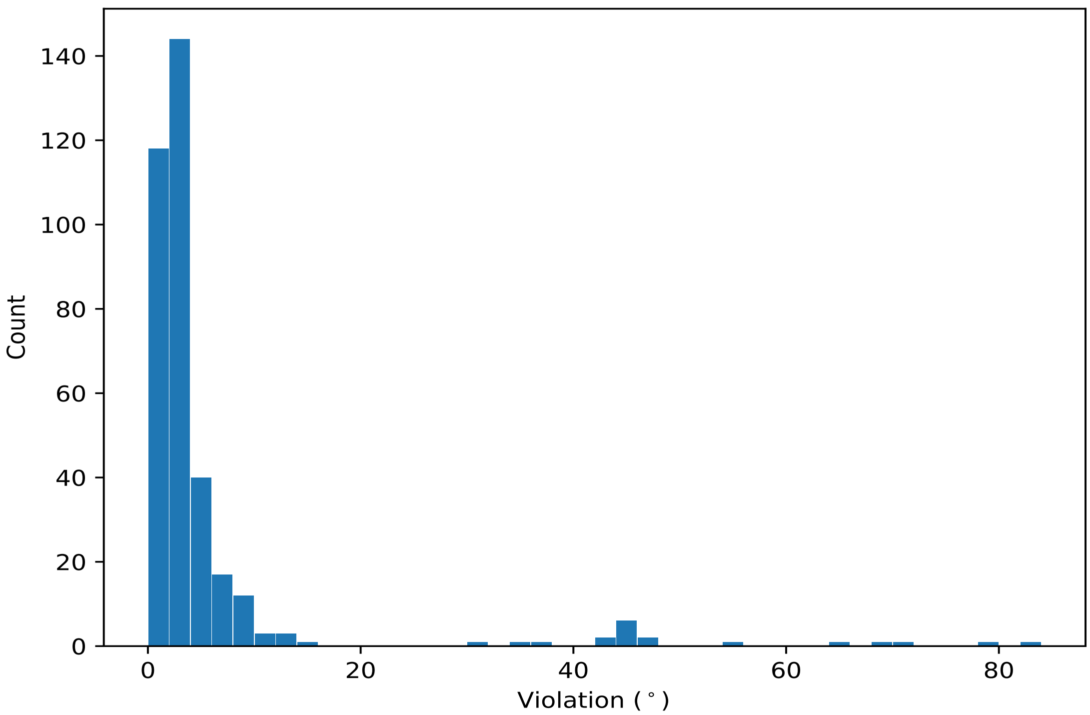
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	20	48.8	17.54	45.35
(1,86)	1:A:107:LYS:N	1:A:107:LYS:CA	1:A:107:LYS:C	1:A:108:ASP:N	20	9.34	2.56	8.65
(1,126)	1:A:127:VAL:N	1:A:127:VAL:CA	1:A:127:VAL:C	1:A:128:ILE:N	20	3.13	0.58	3.1
(1,130)	1:A:129:LYS:N	1:A:129:LYS:CA	1:A:129:LYS:C	1:A:130:VAL:N	20	2.74	0.55	2.65
(1,54)	1:A:87:TYR:N	1:A:87:TYR:CA	1:A:87:TYR:C	1:A:88:HIS:N	20	2.16	0.6	2.1
(1,87)	1:A:107:LYS:C	1:A:108:ASP:N	1:A:108:ASP:CA	1:A:108:ASP:C	19	6.34	1.01	6.1
(1,116)	1:A:122:ILE:N	1:A:122:ILE:CA	1:A:122:ILE:C	1:A:123:LYS:N	19	2.16	0.43	2.3
(1,160)	1:A:145:ASP:N	1:A:145:ASP:CA	1:A:145:ASP:C	1:A:146:ASN:N	18	4.74	1.2	4.7
(1,51)	1:A:84:GLY:C	1:A:85:ASP:N	1:A:85:ASP:CA	1:A:85:ASP:C	18	2.15	0.59	2.2
(1,161)	1:A:145:ASP:C	1:A:146:ASN:N	1:A:146:ASN:CA	1:A:146:ASN:C	17	4.21	1.21	3.9

[1] Number of violated models, [2]Standard 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 violationsfor 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 (°)
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	13	82.2
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	18	78.0
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	5	71.2
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	8	69.0
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	9	64.5
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	3	55.0
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	20	47.7
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	14	46.7
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	15	45.8
(1,186)	1:A:158:GLN:N	1:A:158:GLN:CA	1:A:158:GLN:C	1:A:159:GLU:N	17	45.4