wwPDB NMR Structure Validation Summary Report (i)

Jun 6, 2023 – 05:06 pm BST

PDB ID	:	6R5G
BMRB ID	:	34384
Title	:	C-SH2 domain of SHP-2 in complex with phospho-ITSM of PD-1
Authors	:	Marasco, M.
Deposited on	:	2019-03-25

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

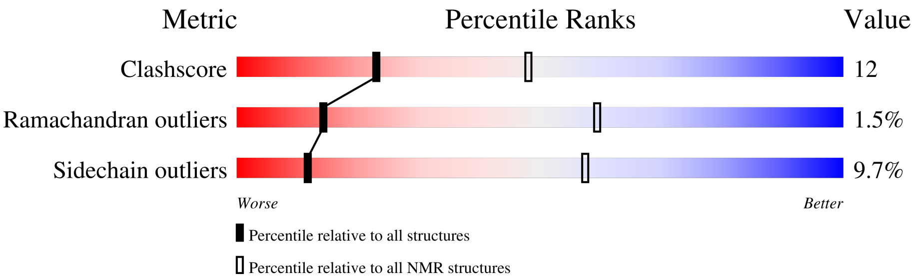
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	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	119	62% 18% 5% 14%
2	B	11	18% 9% 9% 64%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:107-A:154, A:166-A:219, B:1-B:4 (106)	0.44	4

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 3 single-model clusters were found.

Cluster number	**Models**
1	1, 2, 4, 8, 9
2	6, 7
Single-model clusters	3; 5; 10

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2039 atoms, of which 1010 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 11.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	119	1858	578	925	168	183	4	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	GLY	-	expression tag	UNP Q06124
A	103	PRO	-	expression tag	UNP Q06124
A	104	MET	-	expression tag	UNP Q06124

- Molecule 2 is a protein called ITSM.

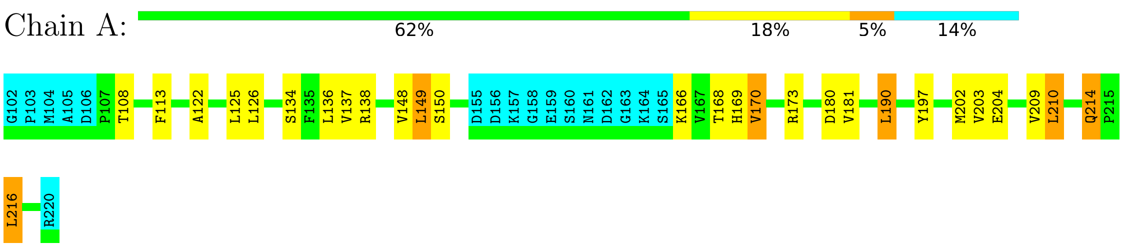
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	11	181	60	85	12	23	1	0

4 Residue-property plots (i)

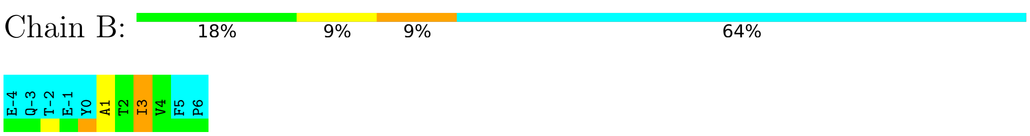
4.1 Average score per residue in the NMR ensemble

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: Tyrosine-protein phosphatase non-receptor type 11



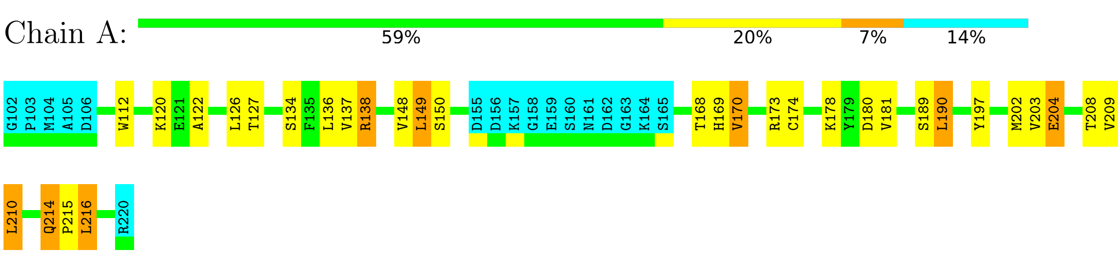
- Molecule 2: ITSM



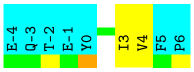
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Tyrosine-protein phosphatase non-receptor type 11



- Molecule 2: ITSM


E-4
Q-3
T-2
E-1
Y0
I3
V4
F5
P6

5 Refinement protocol and experimental data overview (i)

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

Of the 150 calculated structures, 10 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**
ARIA	structure calculation	
CNS	structure calculation	

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	1544
Number of shifts mapped to atoms	1544
Number of unparsed shifts	0
Number of shifts with mapping errors	0
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: PTR

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	810	816	810	21±3
2	B	27	32	32	2±1
All	All	8370	8480	8420	209

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 12.

5 of 65 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:149:LEU:HB3	1:A:170:VAL:HG22	0.79	1.53	4	10
1:A:202:MET:SD	1:A:210:LEU:HD23	0.67	2.30	10	4
1:A:170:VAL:HG11	1:A:202:MET:SD	0.65	2.30	2	4
1:A:168:THR:HG21	1:A:210:LEU:HD11	0.64	1.67	4	10
1:A:108:THR:HB	1:A:113:PHE:CG	0.63	2.28	9	6

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	102/119 (86%)	99±1 (97±1%)	2±1 (2±1%)	0±0 (0±0%)	32	76
2	B	4/11 (36%)	3±0 (65±12%)	0±0 (8±11%)	1±0 (28±8%)	0	1
All	All	1060/1300 (82%)	1020 (96%)	24 (2%)	16 (2%)	14	59

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	3	ILE	10
1	A	115	GLY	5
2	B	1	ALA	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	91/104 (88%)	82±1 (90±1%)	9±1 (10±1%)	11	57
2	B	3/9 (33%)	3±0 (93±13%)	0±0 (7±13%)	20	68
All	All	940/1130 (83%)	849 (90%)	91 (10%)	12	57

5 of 21 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	210	LEU	10
1	A	216	LEU	10
1	A	125	LEU	9
1	A	149	LEU	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	170	VAL	7

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	PTR	B	0	2	15,16,17	0.83±0.03	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	PTR	B	0	2	19,22,24	0.96±0.05	1±0 (5±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PTR	B	0	2	-	0±0,10,11,13	0±0,1,1,1

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
2	B	0	PTR	O2P-P-OH	3.11	114.97	105.24	8	10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 91% for the well-defined parts and 89% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list*

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	1544
Number of shifts mapped to atoms	1544
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	118	-0.34 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	106	0.05 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	112	-0.08 ± 0.09	None needed (< 0.5 ppm)
^{15}N	112	-0.62 ± 0.38	None needed (imprecise)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 91%, i.e. 1331 atoms were assigned a chemical shift out of a possible 1459. 0 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	514/532 (97%)	217/218 (100%)	200/212 (94%)	97/102 (95%)
Sidechain	736/809 (91%)	505/525 (96%)	220/254 (87%)	11/30 (37%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	81/118 (69%)	43/58 (74%)	37/47 (79%)	1/13 (8%)
Overall	1331/1459 (91%)	765/801 (96%)	457/513 (89%)	109/145 (75%)

7.1.4 Statistically unusual chemical shifts (i)

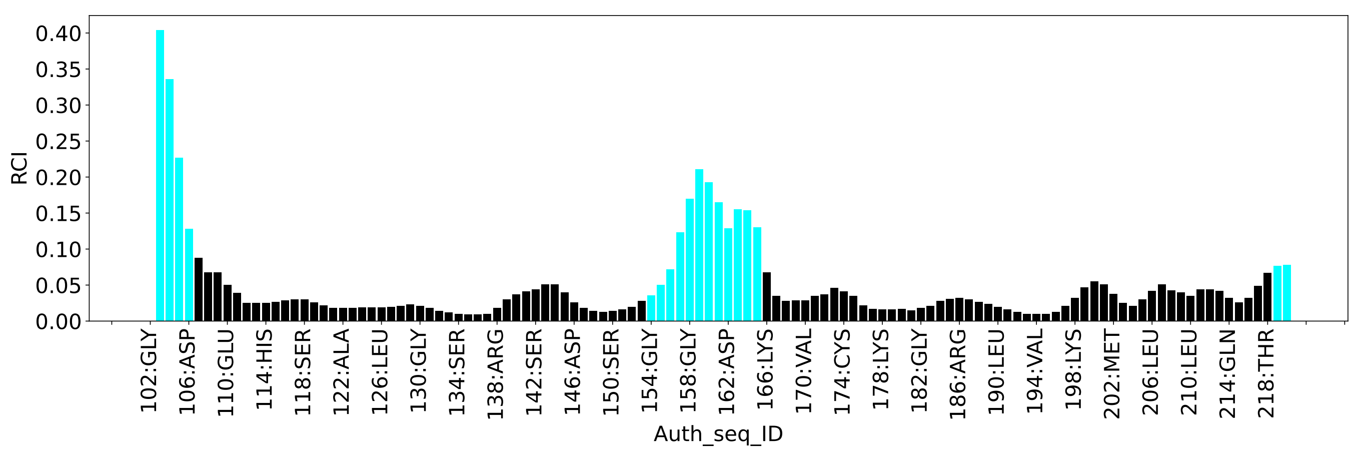
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	-2	THR	HG1	8.19	0.08 – 2.19	33.4
1	A	127	THR	HG1	5.21	0.08 – 2.19	19.3
1	A	110	GLU	HB2	0.48	1.00 – 3.05	-7.5
1	A	110	GLU	HB3	0.48	0.95 – 3.05	-7.2

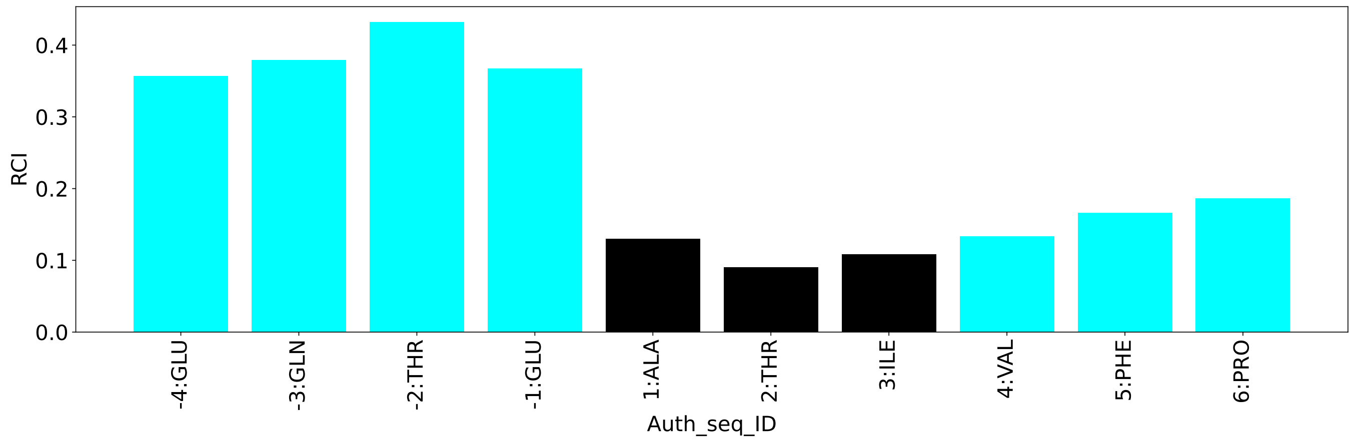
7.1.5 Random Coil Index (RCI) plots (i)

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:



Random coil index (RCI) for chain B:


0.4
0.3
0.2
0.1
0.0
RCI
-4:GLU
-3:GLN
-2:THR
-1:GLU
1:ALA
2:THR
3:ILE
4:VAL
5:PHE
6:PRO
Auth_seq_ID

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	6108
Intra-residue (|i-j|=0)	1644
Sequential (|i-j|=1)	1387
Medium range (|i-j|>1 and |i-j|<5)	821
Long range (|i-j|≥5)	2040
Inter-chain	216
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	190
Number of unmapped restraints	5
Number of restraints per residue	48.4
Number of long range restraints per residue[1]	15.7

[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)	53.9	0.2
0.2-0.5 (Medium)	55.7	0.5
>0.5 (Large)	59.0	2.92

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)	21.8	5.7
10.0-20.0 (Medium)	None	None
>20.0 (Large)	None	None

9 Distance violation analysis (i)

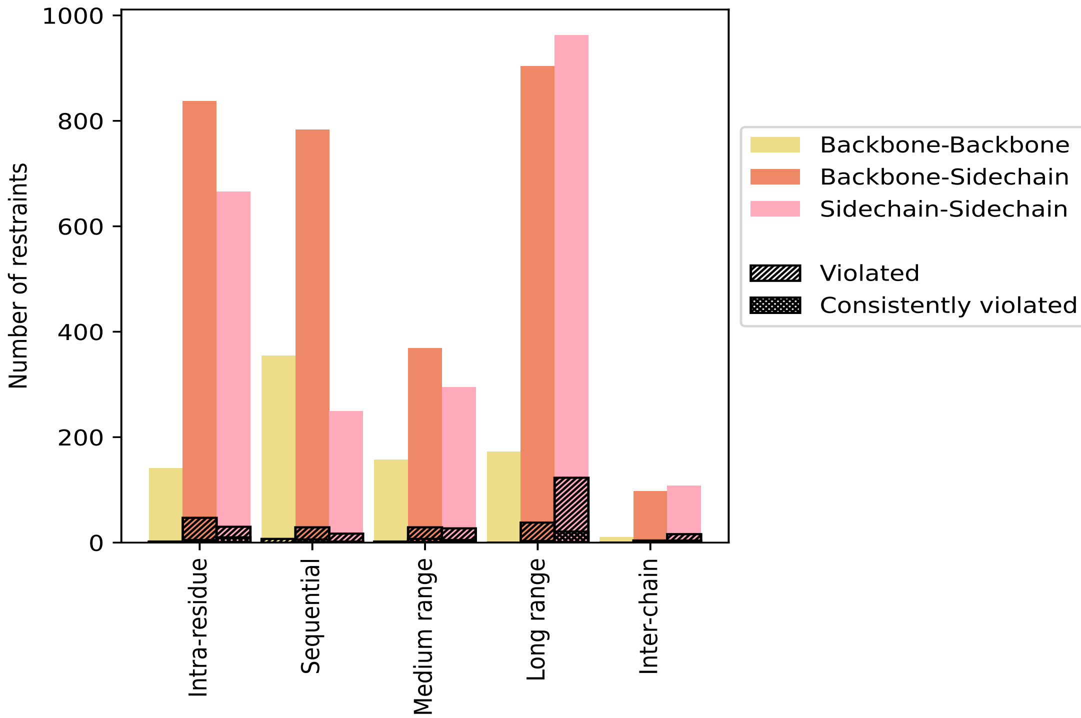
9.1 Summary of distance violations (i)

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)	1644	26.9	79	4.8	1.3	16	1.0	0.3
Backbone-Backbone	141	2.3	2	1.4	0.0	1	0.7	0.0
Backbone-Sidechain	837	13.7	47	5.6	0.8	5	0.6	0.1
Sidechain-Sidechain	666	10.9	30	4.5	0.5	10	1.5	0.2
Sequential (|i-j|=1)	1387	22.7	53	3.8	0.9	8	0.6	0.1
Backbone-Backbone	355	5.8	7	2.0	0.1	0	0.0	0.0
Backbone-Sidechain	783	12.8	29	3.7	0.5	6	0.8	0.1
Sidechain-Sidechain	249	4.1	17	6.8	0.3	2	0.8	0.0
Medium range (|i-j|>1 & |i-j|<5)	821	13.4	58	7.1	0.9	13	1.6	0.2
Backbone-Backbone	157	2.6	2	1.3	0.0	1	0.6	0.0
Backbone-Sidechain	369	6.0	29	7.9	0.5	7	1.9	0.1
Sidechain-Sidechain	295	4.8	27	9.2	0.4	5	1.7	0.1
Long range (|i-j|≥5)	2040	33.4	161	7.9	2.6	24	1.2	0.4
Backbone-Backbone	173	2.8	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	904	14.8	38	4.2	0.6	3	0.3	0.0
Sidechain-Sidechain	963	15.8	123	12.8	2.0	21	2.2	0.3
Inter-chain	216	3.5	20	9.3	0.3	4	1.9	0.1
Backbone-Backbone	10	0.2	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	98	1.6	4	4.1	0.1	0	0.0	0.0
Sidechain-Sidechain	108	1.8	16	14.8	0.3	4	3.7	0.1
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	6108	100.0	371	6.1	6.1	65	1.1	1.1
Backbone-Backbone	836	13.7	11	1.3	0.2	2	0.2	0.0
Backbone-Sidechain	2991	49.0	147	4.9	2.4	21	0.7	0.3
Sidechain-Sidechain	2281	37.3	213	9.3	3.5	42	1.8	0.7

[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)

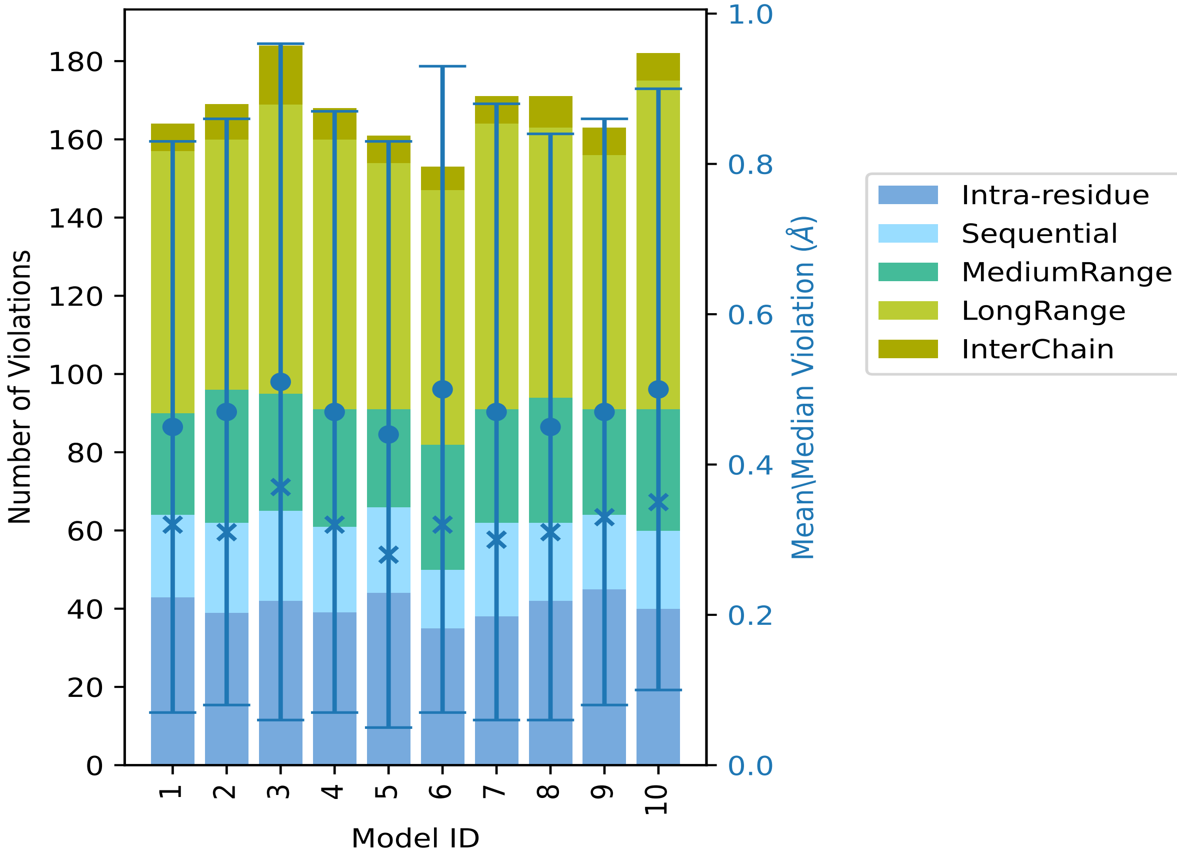
The following table provides the distance violation statistics for each model in the ensemble. Violations less than 0.1 Å are not included in the statistics.

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	43	21	26	67	7	164	0.45	2.31	0.38	0.32
2	39	23	34	64	9	169	0.47	2.26	0.39	0.31
3	42	23	30	74	15	184	0.51	2.92	0.45	0.37
4	39	22	30	69	8	168	0.47	2.21	0.4	0.32
5	44	22	25	63	7	161	0.44	2.24	0.39	0.28
6	35	15	32	65	6	153	0.5	2.12	0.43	0.32
7	38	24	29	73	7	171	0.47	2.28	0.41	0.3
8	42	20	32	69	8	171	0.45	2.22	0.39	0.31
9	45	19	27	65	7	163	0.47	2.1	0.39	0.33
10	40	20	31	84	7	182	0.5	2.05	0.4	0.35

[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, 5737(IR:1565, SQ:1334, MR:763, LR:1879, IC:196) 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]	%
18	17	18	61	10	124	1	10.0
8	12	6	17	2	45	2	20.0
4	4	2	13	0	23	3	30.0
8	3	3	4	0	18	4	40.0

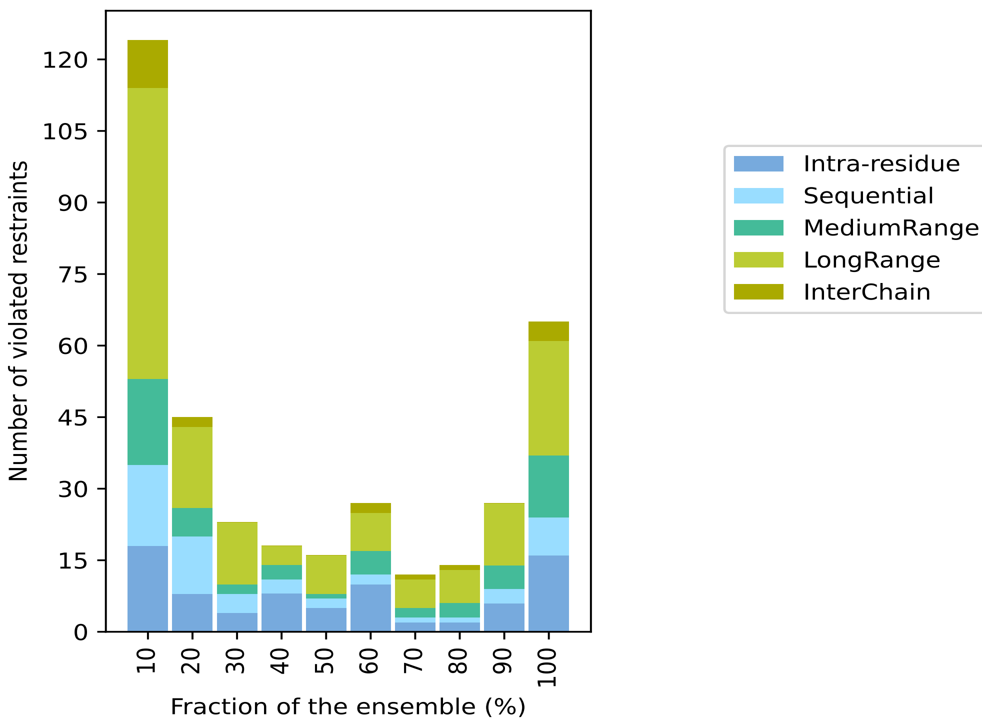
Continued on next page...

Continued from previous page...

Number of violated restraints						Fraction of the ensemble	
IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total	Count[6]	%
5	2	1	8	0	16	5	50.0
10	2	5	8	2	27	6	60.0
2	1	2	6	1	12	7	70.0
2	1	3	7	1	14	8	80.0
6	3	5	13	0	27	9	90.0
16	8	13	24	4	65	10	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

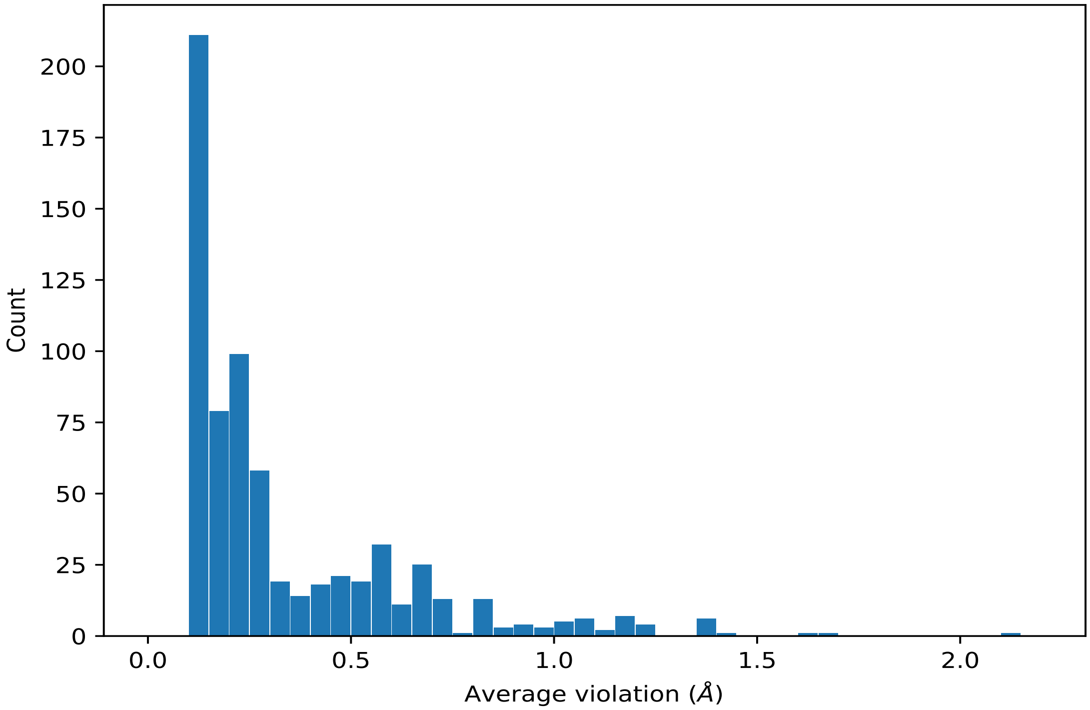
9.3.1 Bar graph : Distance violation statistics for the ensemble (i)



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)
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	10	2.14	0.13	2.16
(2,2518)	1:A:149:LEU:HB2	1:A:170:VAL:HB	10	1.67	0.07	1.66
(2,3865)	1:A:203:VAL:HG22	1:A:207:GLY:HA3	10	1.37	0.03	1.36
(2,3865)	1:A:203:VAL:HG21	1:A:207:GLY:HA3	10	1.37	0.03	1.36
(2,3865)	1:A:203:VAL:HG23	1:A:207:GLY:HA3	10	1.37	0.03	1.36
(2,5114)	1:A:107:PRO:HD3	1:A:113:PHE:HZ	10	1.21	0.03	1.21
(2,4240)	1:A:129:LYS:HB3	1:A:216:LEU:HD12	10	1.17	0.07	1.15
(2,4240)	1:A:129:LYS:HB3	1:A:216:LEU:HD11	10	1.17	0.07	1.15
(2,4240)	1:A:129:LYS:HB3	1:A:216:LEU:HD13	10	1.17	0.07	1.15
(2,2626)	1:A:133:GLY:HA3	1:A:151:VAL:HG12	10	1.17	0.06	1.18
(2,2626)	1:A:133:GLY:HA3	1:A:151:VAL:HG11	10	1.17	0.06	1.18
(2,2626)	1:A:133:GLY:HA3	1:A:151:VAL:HG13	10	1.17	0.06	1.18
(2,1850)	1:A:123:GLU:HA	1:A:126:LEU:HB3	10	1.12	0.19	1.08
(2,1889)	1:A:125:LEU:HA	1:A:128:GLU:HB3	10	1.1	0.24	1.17
(2,2065)	1:A:129:LYS:HD3	1:A:125:LEU:HD22	10	1.07	0.29	0.96
(2,2065)	1:A:129:LYS:HD3	1:A:125:LEU:HD21	10	1.07	0.29	0.96

Continued on next page...

Continued from previous page...

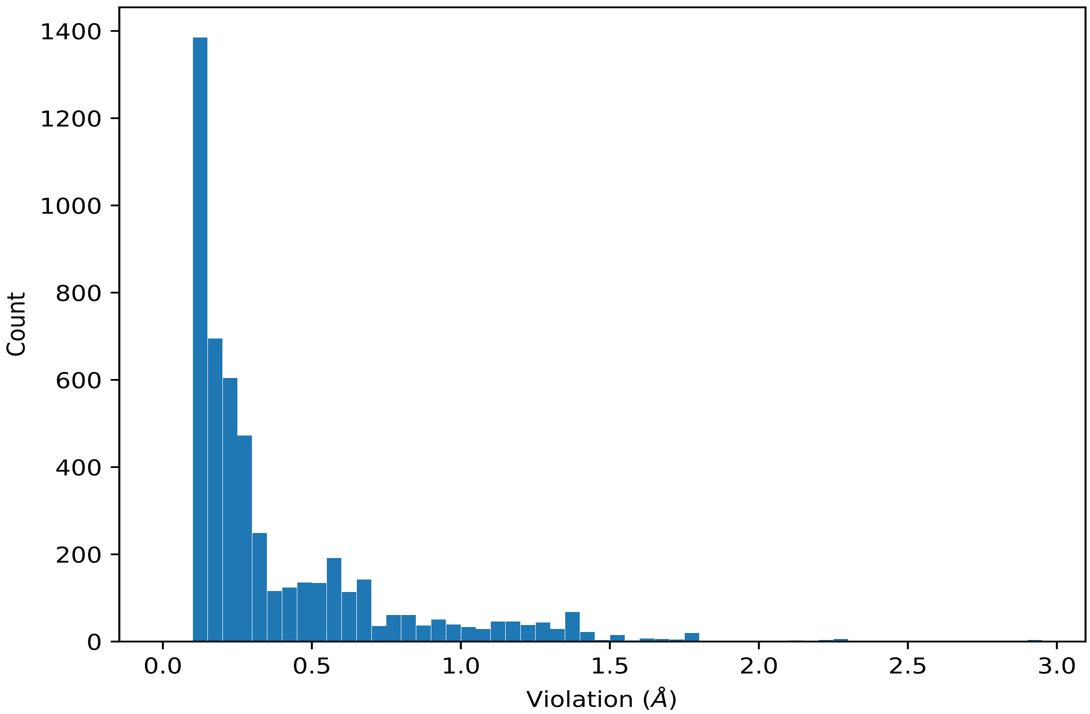
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(2,2065)	1:A:129:LYS:HD3	1:A:125:LEU:HD23	10	1.07	0.29	0.96
(2,4375)	1:A:103:PRO:HB3	1:A:104:MET:HA	10	1.06	0.23	1.03

[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)
(2,3263)	1:A:181:VAL:HG23	1:A:172:ILE:HG13	3	2.92
(2,3263)	1:A:181:VAL:HG21	1:A:172:ILE:HG13	3	2.92
(2,3263)	1:A:181:VAL:HG22	1:A:172:ILE:HG13	3	2.92

Continued on next page...

Continued from previous page...

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	1	2.31
(2,3295)	1:A:181:VAL:HG11	2:B:5:PHE:HZ	3	2.29
(2,3295)	1:A:181:VAL:HG13	2:B:5:PHE:HZ	3	2.29
(2,3295)	1:A:181:VAL:HG12	2:B:5:PHE:HZ	3	2.29
(2,3433)	1:A:190:LEU:HB3	1:A:107:PRO:HB3	7	2.28
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	2	2.26
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	5	2.24
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	8	2.22
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	4	2.21
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	6	2.12
(2,2184)	1:A:136:LEU:HB3	1:A:138:ARG:HB2	9	2.1

10 Dihedral-angle violation analysis (i)

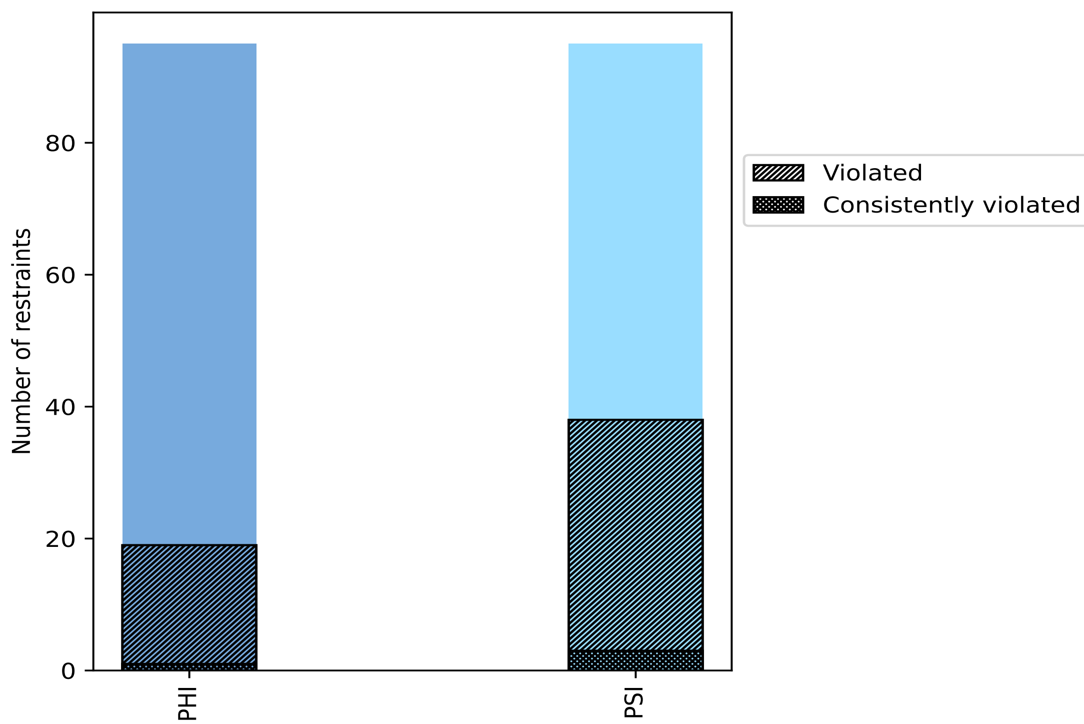
10.1 Summary of dihedral-angle violations (i)

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	95	50.0	19	20.0	10.0	1	1.1	0.5
PSI	95	50.0	38	40.0	20.0	3	3.2	1.6
Total	190	100.0	57	30.0	30.0	4	2.1	2.1

[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)



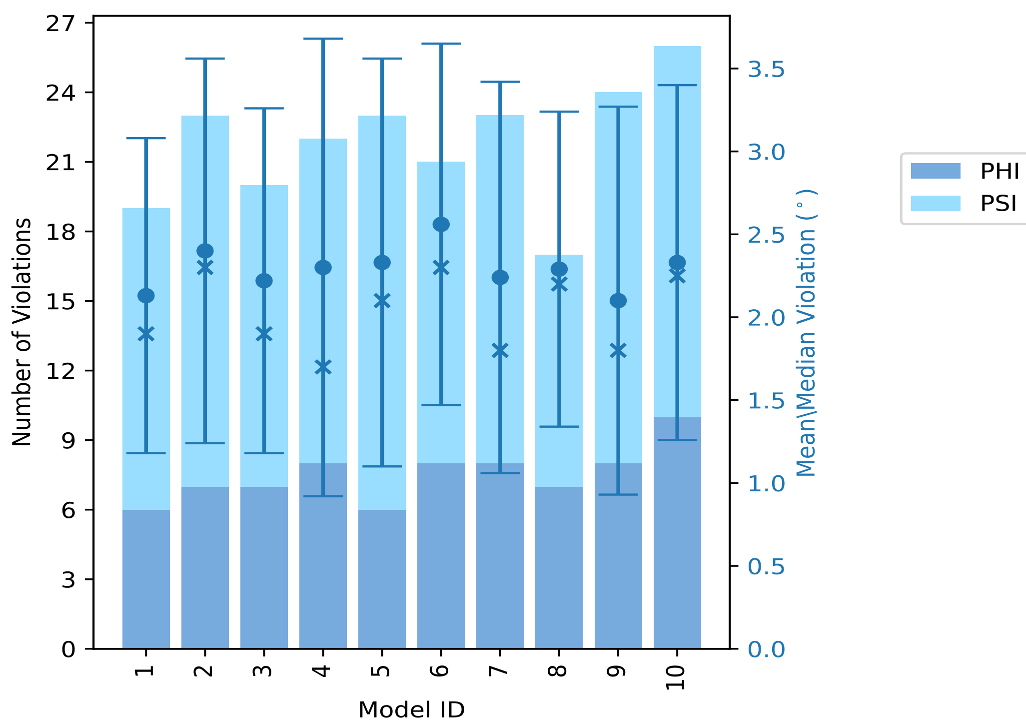
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

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 (°)
	PHI	PSI	Total				
1	6	13	19	2.13	4.7	0.95	1.9
2	7	16	23	2.4	5.3	1.16	2.3
3	7	13	20	2.22	5.3	1.04	1.9
4	8	14	22	2.3	5.4	1.38	1.7
5	6	17	23	2.33	5.7	1.23	2.1
6	8	13	21	2.56	5.0	1.09	2.3
7	8	15	23	2.24	5.3	1.18	1.8
8	7	10	17	2.29	4.4	0.95	2.2
9	8	16	24	2.1	5.1	1.17	1.8
10	10	16	26	2.33	4.9	1.07	2.25

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

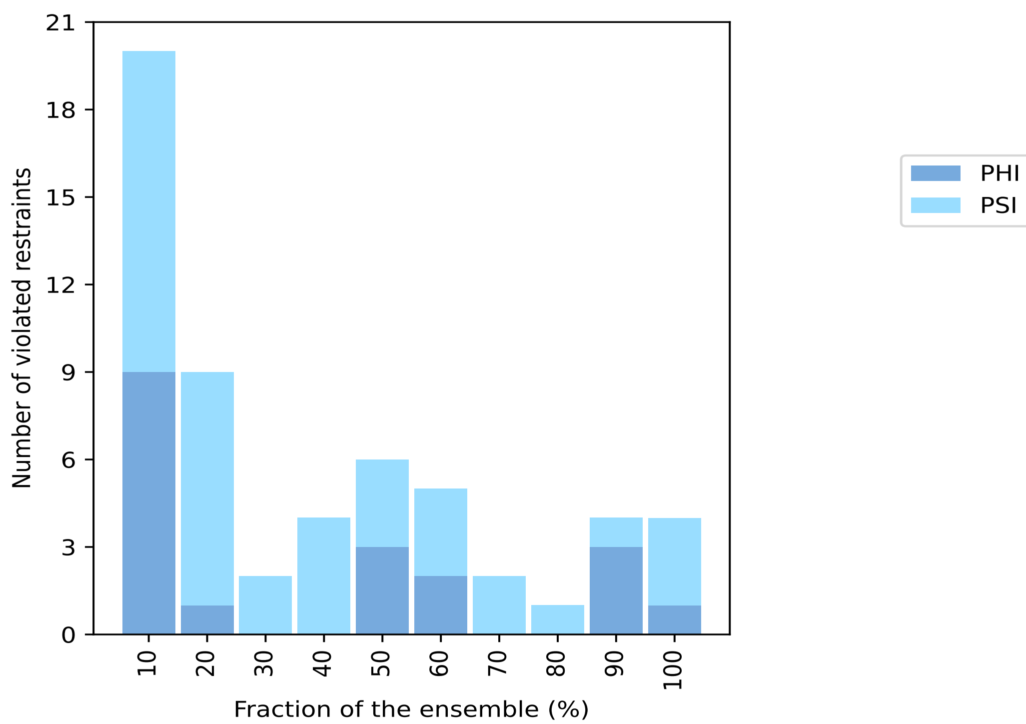
10.3 Dihedral-angle violation statistics for the ensemble (i)

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	
PHI	PSI	Total	Count[1]	%
9	11	20	1	10.0
1	8	9	2	20.0
0	2	2	3	30.0
0	4	4	4	40.0
3	3	6	5	50.0
2	3	5	6	60.0
0	2	2	7	70.0
0	1	1	8	80.0
3	1	4	9	90.0
1	3	4	10	100.0

[1] Number of models with violations

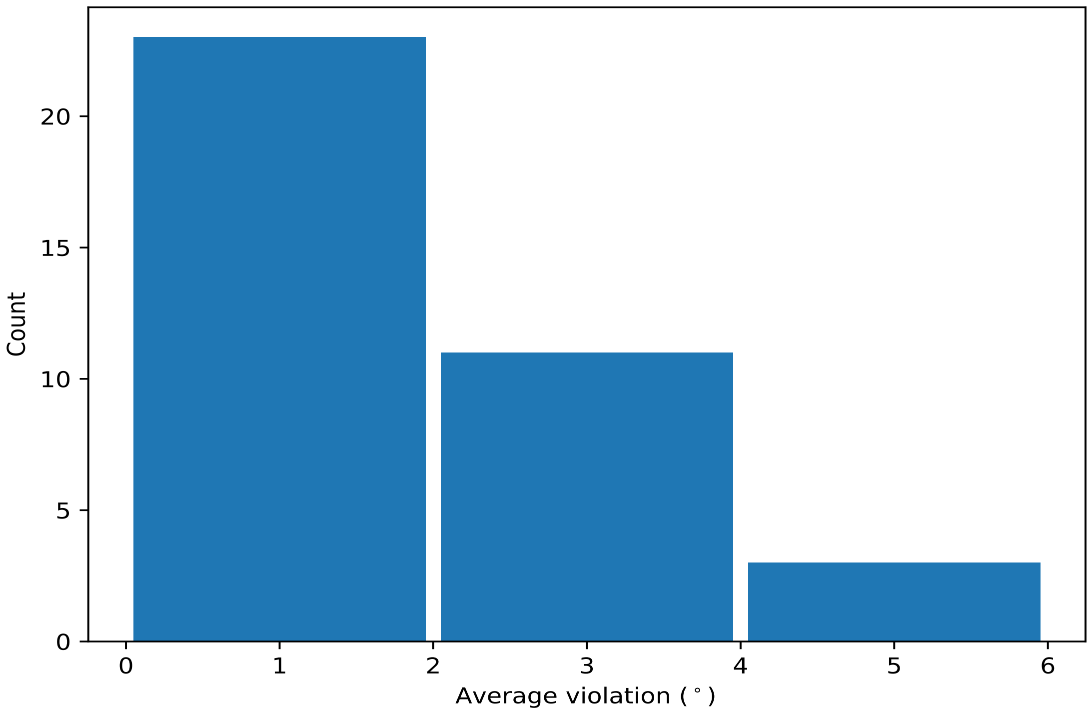
10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble (i)



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.

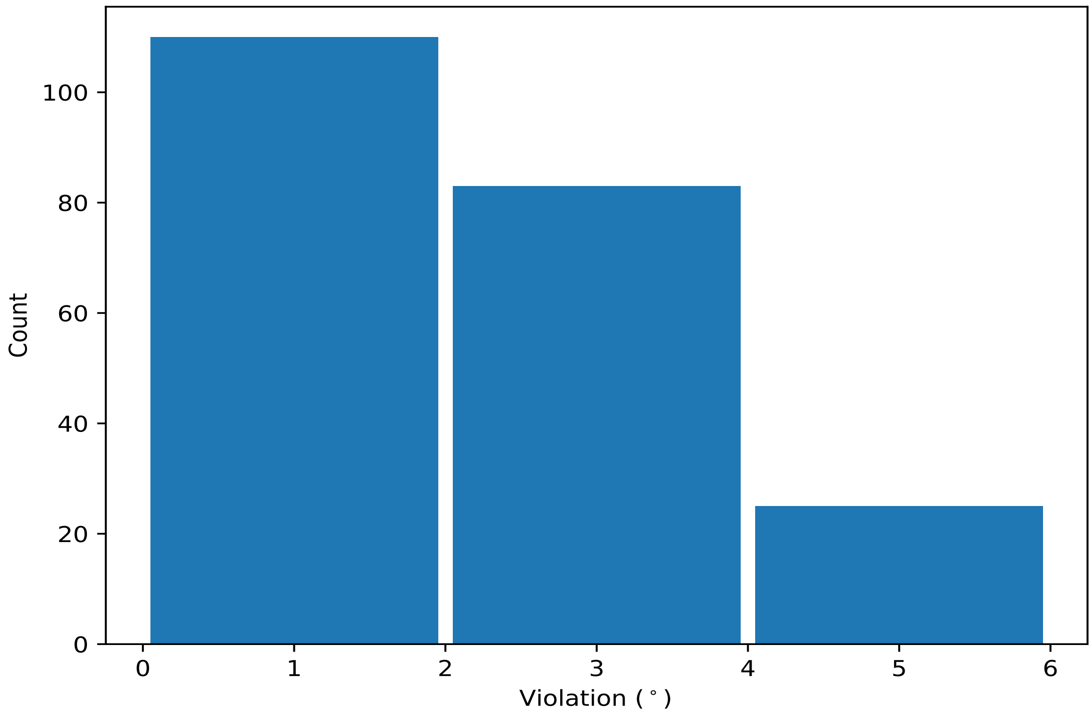
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	10	4.88	0.44	4.95
(1,34)	1:A:125:LEU:N	1:A:125:LEU:CA	1:A:125:LEU:C	1:A:126:LEU:N	10	4.29	0.7	4.35
(1,170)	1:A:208:THR:N	1:A:208:THR:CA	1:A:208:THR:C	1:A:209:VAL:N	10	4.2	0.98	4.4
(1,142)	1:A:193:LEU:N	1:A:193:LEU:CA	1:A:193:LEU:C	1:A:194:VAL:N	10	2.09	0.29	2.1
(1,123)	1:A:182:GLY:C	1:A:183:GLY:N	1:A:183:GLY:CA	1:A:183:GLY:C	9	3.41	0.73	3.3
(1,99)	1:A:169:HIS:C	1:A:170:VAL:N	1:A:170:VAL:CA	1:A:170:VAL:C	9	3.1	0.42	3.3
(1,33)	1:A:124:LYS:C	1:A:125:LEU:N	1:A:125:LEU:CA	1:A:125:LEU:C	9	2.68	0.58	2.7
(1,80)	1:A:149:LEU:N	1:A:149:LEU:CA	1:A:149:LEU:C	1:A:150:SER:N	9	2.12	0.37	2.0
(1,86)	1:A:152:ARG:N	1:A:152:ARG:CA	1:A:152:ARG:C	1:A:153:THR:N	8	1.9	0.74	1.7
(1,32)	1:A:124:LYS:N	1:A:124:LYS:CA	1:A:124:LYS:C	1:A:125:LEU:N	7	2.1	0.51	1.9

[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 (°)
(1,170)	1:A:208:THR:N	1:A:208:THR:CA	1:A:208:THR:C	1:A:209:VAL:N	5	5.7
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	4	5.4
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	3	5.3
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	7	5.3
(1,34)	1:A:125:LEU:N	1:A:125:LEU:CA	1:A:125:LEU:C	1:A:126:LEU:N	2	5.3
(1,170)	1:A:208:THR:N	1:A:208:THR:CA	1:A:208:THR:C	1:A:209:VAL:N	4	5.3
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	9	5.1
(1,34)	1:A:125:LEU:N	1:A:125:LEU:CA	1:A:125:LEU:C	1:A:126:LEU:N	4	5.1
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	6	5.0
(1,35)	1:A:125:LEU:C	1:A:126:LEU:N	1:A:126:LEU:CA	1:A:126:LEU:C	10	4.9