wwPDB NMR Structure Validation Summary Report (i)

Jun 3, 2023 – 11:44 PM EDT

PDB ID	:	2KM8
BMRB ID	:	16425
Title	:	Interdomain RRM packing contributes to RNA recognition in the rna15, hrp1, anchor RNA 3' processing ternary complex
Authors	:	Leeper, T.C.; Varani, G.
Deposited on	:	2009-07-24

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 (1)) were used in the production of this report:

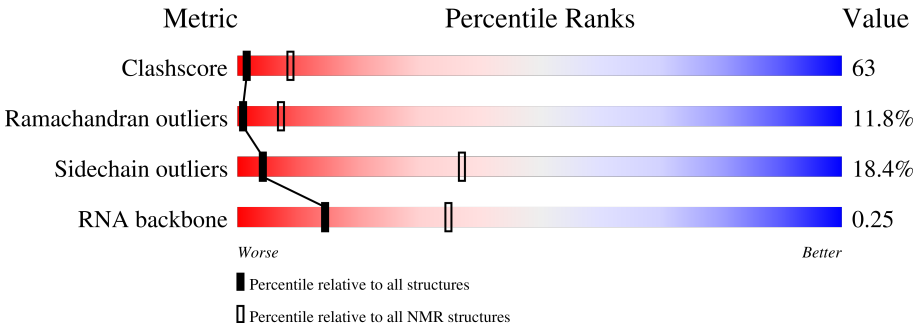
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal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

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
RNA backbone	4643	676

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	13	77% 23%
2	B	84	21% 58% 18% •
3	C	167	31% 54% 12% •

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

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	B:22-B:103, C:158-C:319 (244)	2.27	1

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	1, 3, 4, 6
2	7, 8
3	2, 5
4	9, 10

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4375 atoms, of which 2102 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called 5'-R(P*UP*AP*UP*AP*UP*AP*UP*AP*AP*UP*AP*AP*U)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	13	414	124	139	47	91	13	0

- Molecule 2 is a protein called mRNA 3'-end-processing protein RNA15.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	84	1296	411	640	112	129	4	0

- Molecule 3 is a protein called Nuclear polyadenylated RNA-binding protein 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	C	167	2665	849	1323	232	256	5	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

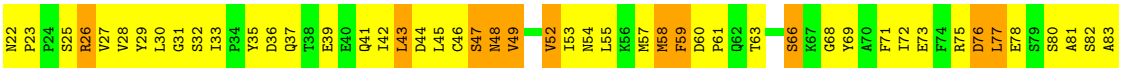
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for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

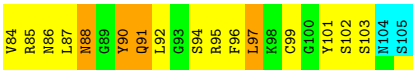
• Molecule 1: 5'-R(P*UP*AP*UP*AP*UP*AP*UP*AP*AP*UP*AP*AP*U)-3'

Chain A: 77% 23%

• Molecule 2: mRNA 3'-end-processing protein RNA15

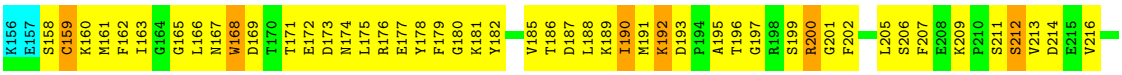
Chain B: 21% 58% 18% •

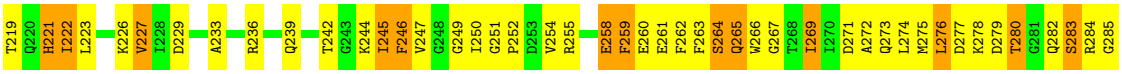


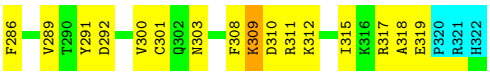


• Molecule 3: Nuclear polyadenylated RNA-binding protein 4

Chain C: 31% 54% 12% •







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

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: 5'-R(P*UP*AP*UP*AP*UP*AP*UP*AP*AP*UP*AP*AP*U)-3'

Chain A: 15% 54% 31%

U1 A2 U3 A4 U5 A6 U7 A8 A9 U10 A11 A12 U13

- Molecule 2: mRNA 3'-end-processing protein RNA15

Chain B: 26% 49% 21%

N22 S25 R26 V27 V28 Y29 L30 G31 S32 I33 P34 Y35 D36 Q37 T38 E39 E40 Q41 I42 L43 D44 L45 C46 S47 N48 V49 V52 I53 N54 L55 K56 M57 M58 F59 D60 P61 R65 S66 K67 G68 Y69 A70 F71 I72 E73 F74 R75 D76 L77 E78 A81 S82 A83 V84 R85 N86

L87 N88 G89 Y90 Q91 L92 G93 S94 R95 F96 L97 G100 Y101 S102 S103 N104 S105

- Molecule 3: Nuclear polyadenylated RNA-binding protein 4

Chain C: 37% 46% 12%

K156 E157 S158 C159 K160 M161 F162 G165 L166 N167 W168 D169 T170 T171 N174 L175 R176 E177 Y178 F179 G180 K181 Y182 D187 L188 K189 I190 M191 K192 D193 T196 S199 R200 G201 F202 L205 S206 K209 P210 S211 S212 V213 D214 E215 V216 V217 H221 I222 L223 D224 G225 K226

V227 I228 D229 R232 A233 D237 E238 Q239 T242 G243 K244 I245 F246 V247 G248 G249 V254 E258 F259 E260 E261 F262 F263 S264 Q265 W266 G267 T268 I269 I270 D271 A272 Q273 L274 M275 L276 D277 K278 D279 T280 G281 Q282 S283 R284 G285 F286 G287 F288 V289 T290 Y291 D292 R299 V300

C301 Q302 N303 K304 F308 K309 D310 R311 E314 I315 K316 R317 A318 E319 P320 R321 H322

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 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
X-PLOR NIH	refinement	
CYANA	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	922
Number of shifts mapped to atoms	922
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	25%

6 Model quality (i)

6.1 Standard geometry (i)

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	1.12±0.10	0±1/307 (0.1± 0.4%)	1.65±0.05	11±2/473 (2.2± 0.4%)
2	B	1.05±0.02	0±0/655 (0.0± 0.0%)	0.90±0.02	0±0/886 (0.0± 0.0%)
3	C	1.05±0.00	0±0/1322 (0.0± 0.0%)	0.89±0.00	0±0/1776 (0.0± 0.0%)
All	All	1.06	5/22840 (0.0%)	1.04	106/31350 (0.3%)

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	1

All 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	11	A	O3'-P	8.23	1.71	1.61	7	1
1	A	12	A	P-O5'	6.47	1.66	1.59	7	1
1	A	11	A	N9-C4	-6.31	1.34	1.37	7	1
1	A	11	A	C3'-O3'	5.52	1.49	1.42	7	1
2	B	68	GLY	N-CA	5.20	1.53	1.46	7	1

5 of 24 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	9	A	O4'-C1'-N9	10.28	116.43	108.20	5	7
1	A	12	A	O4'-C1'-N9	7.37	114.10	108.20	7	1
1	A	9	A	C1'-O4'-C4'	-6.79	104.46	109.90	5	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	8	A	C5'-C4'-C3'	-6.77	105.17	116.00	10	6
1	A	3	U	O4'-C1'-N1	6.53	113.42	108.20	2	1

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	4	A	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	275	139	138	98±23
2	B	641	629	626	99±25
3	C	1295	1274	1273	134±14
All	All	22110	20420	20370	2660

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 63.

5 of 1438 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:A:H3'	2:B:69:TYR:N	1.31	1.37	7	1
1:A:6:A:H4'	1:A:7:U:H5'	1.13	1.17	9	7
1:A:12:A:H5'	2:B:67:LYS:O	1.09	1.47	7	1
1:A:12:A:C5'	2:B:67:LYS:O	1.07	2.00	7	1
1:A:10:U:H5'	2:B:71:PHE:CE1	1.03	1.86	5	4

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
2	B	81/84 (96%)	56±2 (69±3%)	13±2 (16±3%)	12±2 (15±2%)	0 4
3	C	162/167 (97%)	119±3 (74±2%)	26±3 (16±2%)	17±2 (10±1%)	1 9
All	All	2430/2510 (97%)	1752 (72%)	392 (16%)	286 (12%)	1 7

5 of 83 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	26	ARG	10
2	B	47	SER	10
2	B	76	ASP	10
2	B	77	LEU	10
2	B	91	GLN	10

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
2	B	72/74 (97%)	58±2 (80±3%)	14±2 (20±3%)	4 34
3	C	140/145 (97%)	115±3 (82±2%)	25±3 (18±2%)	4 39
All	All	2120/2190 (97%)	1730 (82%)	390 (18%)	4 37

5 of 129 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)
2	B	90	TYR	10
2	B	37	GLN	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	97	LEU	9
3	C	171	THR	9
3	C	222	ILE	9

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	12/13 (92%)	6±1 (46±11%)	1±1 (8±8%)	0.25±0.07
All	All	120/130 (92%)	55 (46%)	9 (8%)	0.25

The overall RNA backbone suiteness is 0.25.

5 of 11 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	3	U	8
1	A	13	U	7
1	A	4	A	6
1	A	10	U	6
1	A	12	A	6

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	5	U	3
1	A	6	A	3
1	A	10	U	3

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 25% for the well-defined parts and 25% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	922
Number of shifts mapped to atoms	922
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	80	-0.14 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	65	0.11 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	77	-0.38 ± 0.12	None needed (< 0.5 ppm)
^{15}N	78	0.64 ± 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 25%, i.e. 902 atoms were assigned a chemical shift out of a possible 3549. 0 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	392/1221 (32%)	163/500 (33%)	153/488 (31%)	76/233 (33%)
Sidechain	477/1819 (26%)	347/1169 (30%)	121/568 (21%)	9/82 (11%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	33/274 (12%)	33/133 (25%)	0/137 (0%)	0/4 (0%)
Sugar	0/143 (0%)	0/78 (0%)	0/65 (0%)	0/0 (—%)
Base	0/92 (0%)	0/53 (0%)	0/26 (0%)	0/13 (0%)
Overall	902/3549 (25%)	543/1933 (28%)	274/1284 (21%)	85/332 (26%)

7.1.4 Statistically unusual chemical shifts (i)

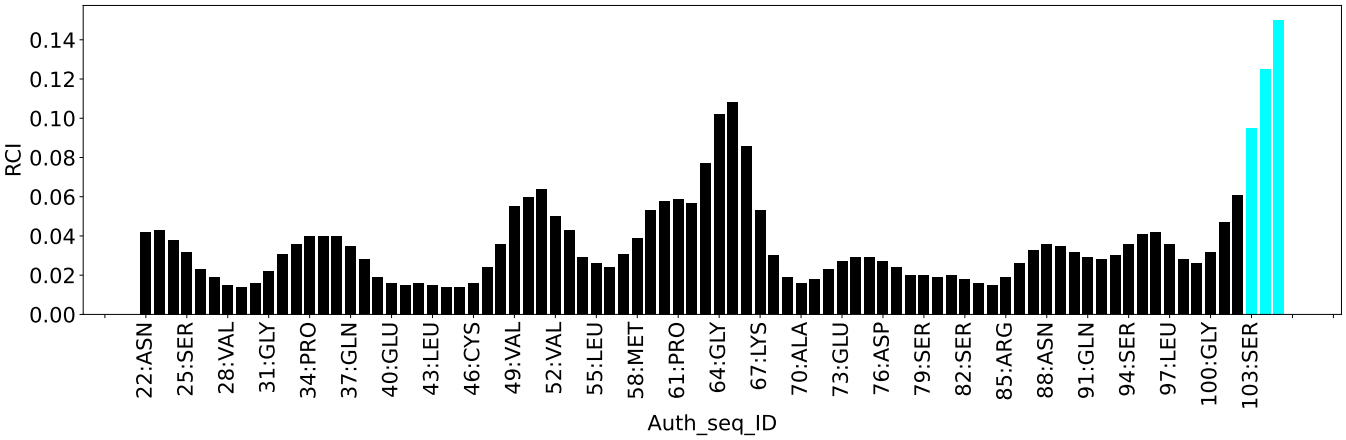
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	23	PRO	HA	1.69	2.78 – 6.00	-8.4

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 B:



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	4935
Intra-residue (|i-j|=0)	1296
Sequential (|i-j|=1)	1018
Medium range (|i-j|>1 and |i-j|<5)	672
Long range (|i-j|≥5)	1524
Inter-chain	293
Hydrogen bond restraints	132
Disulfide bond restraints	0
Total dihedral-angle restraints	202
Number of unmapped restraints	0
Number of restraints per residue	19.5
Number of long range restraints per residue[1]	6.1

[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)	54.1	0.2
0.2-0.5 (Medium)	12.1	0.5
>0.5 (Large)	9.0	2.88

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)	17.0	6.8
10.0-20.0 (Medium)	None	None
>20.0 (Large)	None	None

9 Distance violation analysis (i)

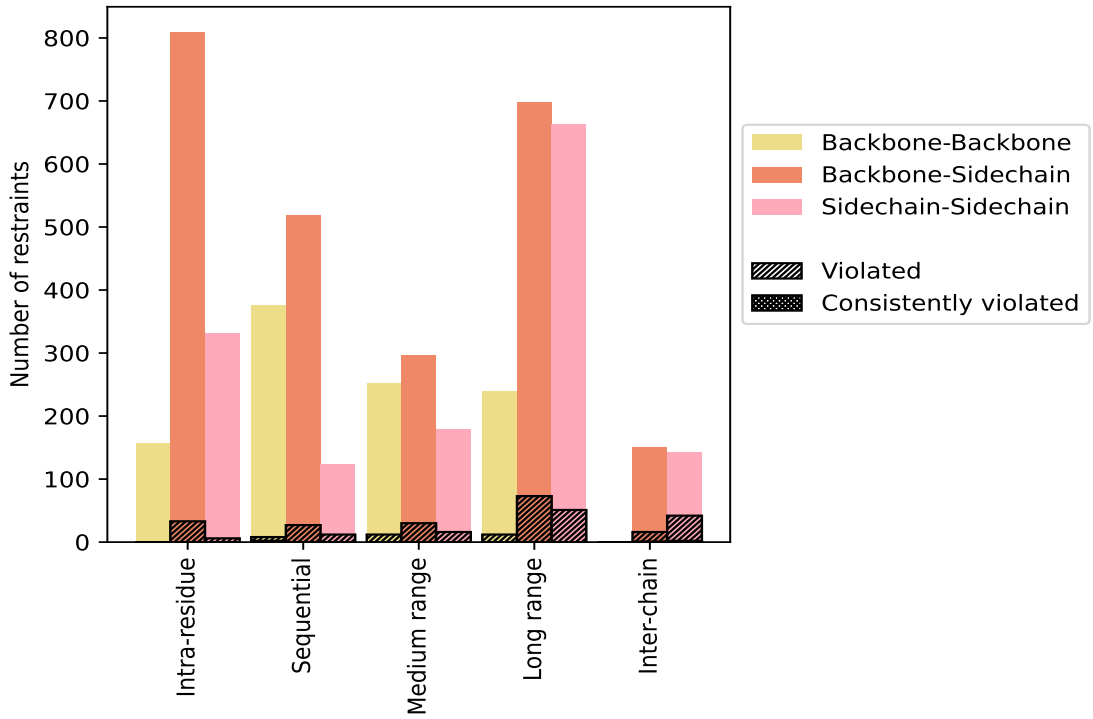
9.1 Summary of distance violations (i)

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)	1296	26.3	39	3.0	0.8	0	0.0	0.0
Backbone-Backbone	156	3.2	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	809	16.4	33	4.1	0.7	0	0.0	0.0
Sidechain-Sidechain	331	6.7	6	1.8	0.1	0	0.0	0.0
Sequential (|i-j|=1)	1018	20.6	47	4.6	1.0	2	0.2	0.0
Backbone-Backbone	375	7.6	8	2.1	0.2	2	0.5	0.0
Backbone-Sidechain	519	10.5	27	5.2	0.5	0	0.0	0.0
Sidechain-Sidechain	124	2.5	12	9.7	0.2	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	672	13.6	58	8.6	1.2	0	0.0	0.0
Backbone-Backbone	196	4.0	12	6.1	0.2	0	0.0	0.0
Backbone-Sidechain	297	6.0	30	10.1	0.6	0	0.0	0.0
Sidechain-Sidechain	179	3.6	16	8.9	0.3	0	0.0	0.0
Long range (|i-j|≥5)	1524	30.9	133	8.7	2.7	0	0.0	0.0
Backbone-Backbone	163	3.3	9	5.5	0.2	0	0.0	0.0
Backbone-Sidechain	698	14.1	73	10.5	1.5	0	0.0	0.0
Sidechain-Sidechain	663	13.4	51	7.7	1.0	0	0.0	0.0
Inter-chain	293	5.9	58	19.8	1.2	3	1.0	0.1
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	150	3.0	16	10.7	0.3	1	0.7	0.0
Sidechain-Sidechain	143	2.9	42	29.4	0.9	2	1.4	0.0
Hydrogen bond	132	2.7	3	2.3	0.1	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	4935	100.0	338	6.8	6.8	5	0.1	0.1
Backbone-Backbone	1022	20.7	32	3.1	0.6	2	0.2	0.0
Backbone-Sidechain	2473	50.1	179	7.2	3.6	1	0.0	0.0
Sidechain-Sidechain	1440	29.2	127	8.8	2.6	2	0.1	0.0

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

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



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

9.2 Distance violation statistics for each model (i)

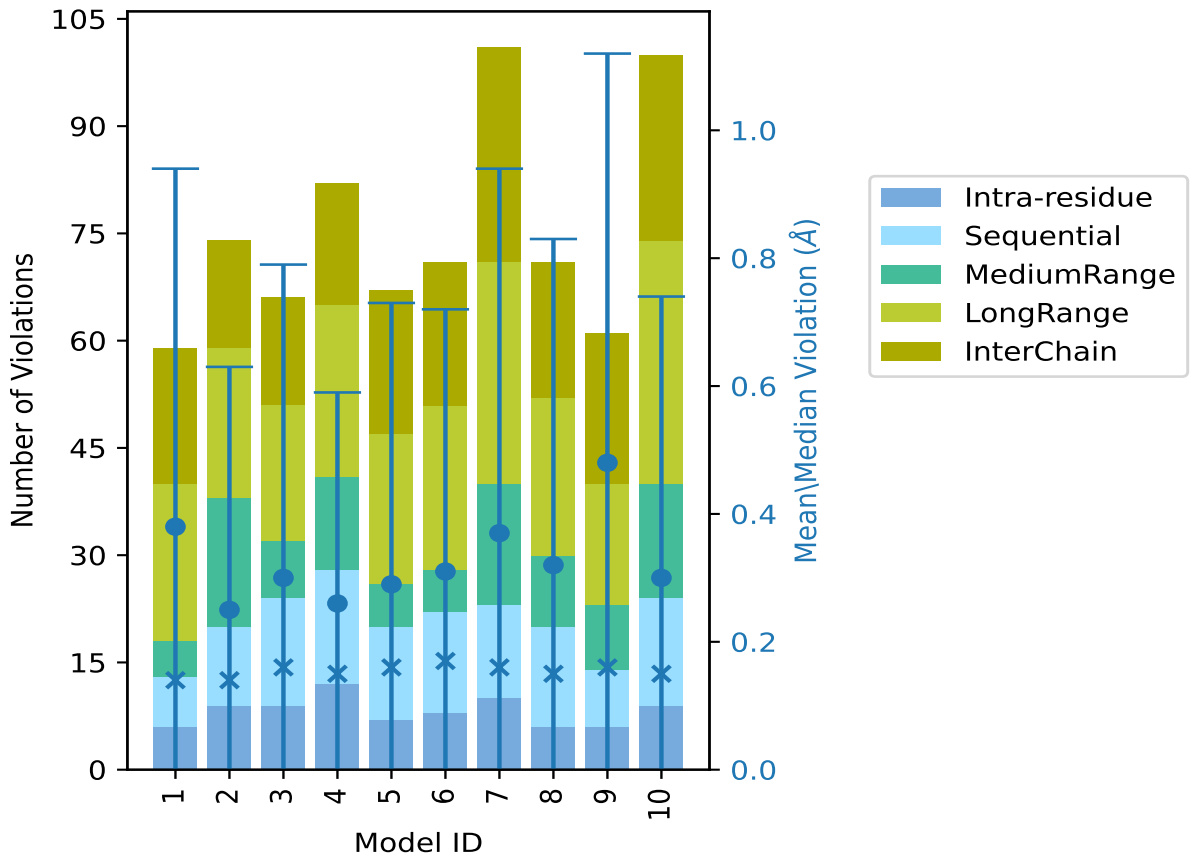
The following table provides the distance violation statistics for each model in the ensemble. Violations less than 0.1 Å are not included in the statistics.

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	6	7	5	22	19	59	0.38	2.49	0.56	0.14
2	9	11	18	21	15	74	0.25	2.37	0.38	0.14
3	9	15	8	19	15	66	0.3	2.55	0.49	0.16
4	12	16	13	24	17	82	0.26	1.82	0.33	0.15
5	7	13	6	21	20	67	0.29	2.47	0.44	0.16
6	8	14	6	23	20	71	0.31	2.31	0.41	0.17
7	10	13	17	31	30	101	0.37	2.88	0.57	0.16
8	6	14	10	22	19	71	0.32	2.7	0.51	0.15
9	6	8	9	17	21	61	0.48	2.73	0.64	0.16
10	9	15	16	34	26	100	0.3	2.57	0.44	0.15

[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, 4468(IR:1257, SQ:971, MR:614, LR:1391, IC:235) 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]	%
23	23	37	84	20	187	1	10.0
5	11	9	32	9	66	2	20.0
5	3	6	6	8	28	3	30.0
2	2	1	1	4	10	4	40.0

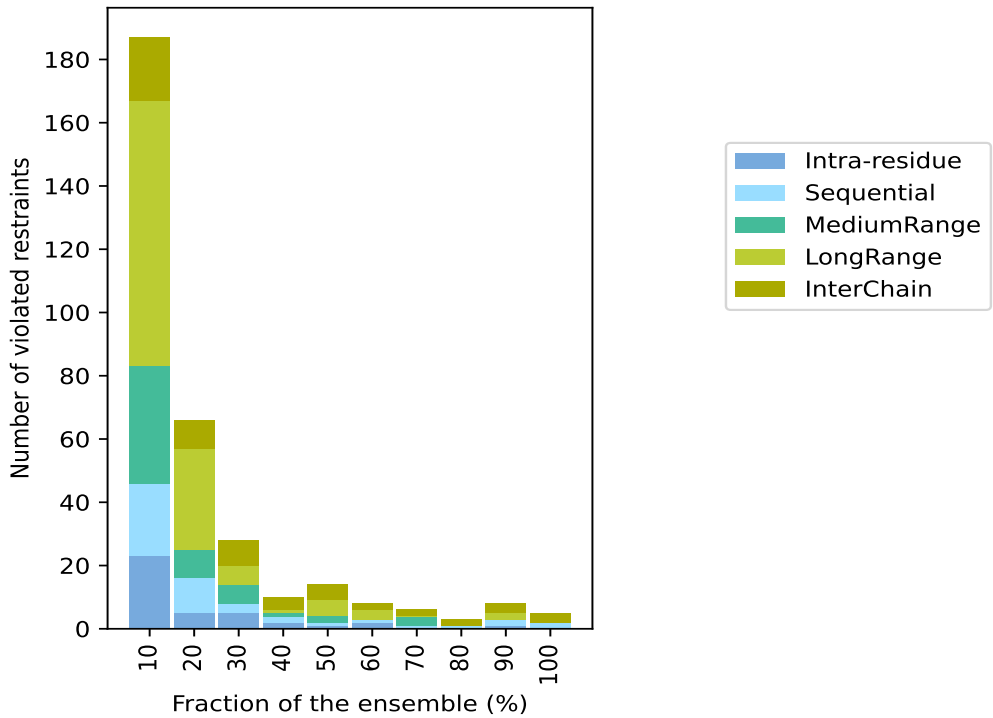
Continued on next page...

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]	%
1	1	2	5	5	14	5	50.0
2	1	0	3	2	8	6	60.0
0	1	3	0	2	6	7	70.0
0	1	0	0	2	3	8	80.0
1	2	0	2	3	8	9	90.0
0	2	0	0	3	5	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

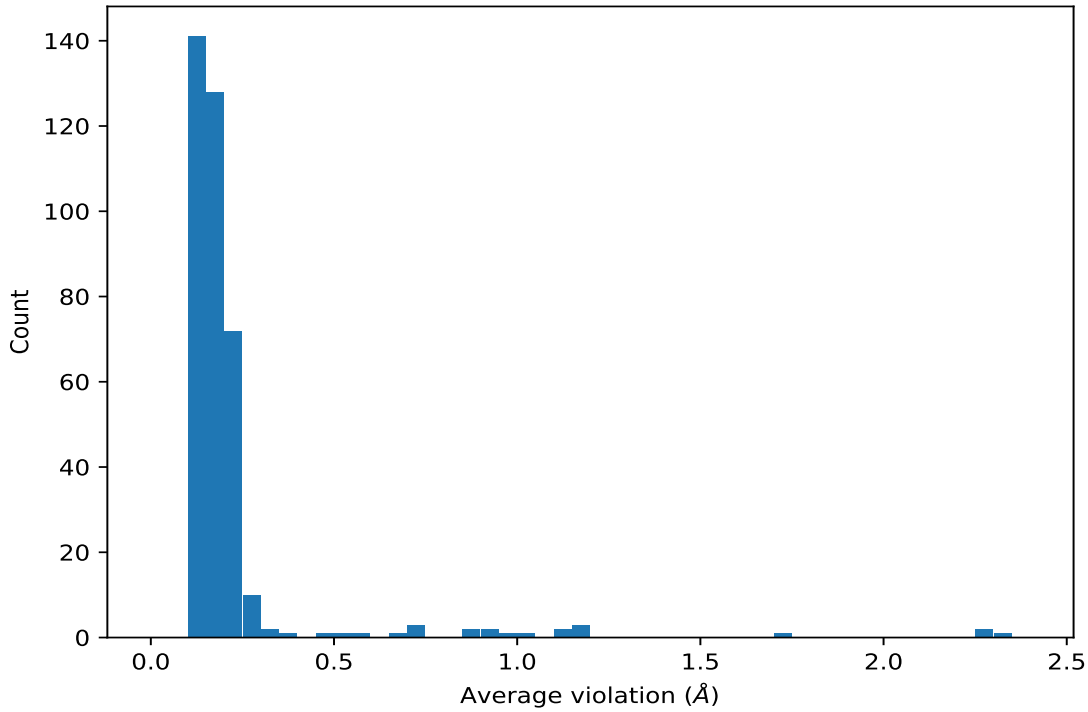
9.3.1 Bar graph : Distance violation statistics for the ensemble (i)



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.

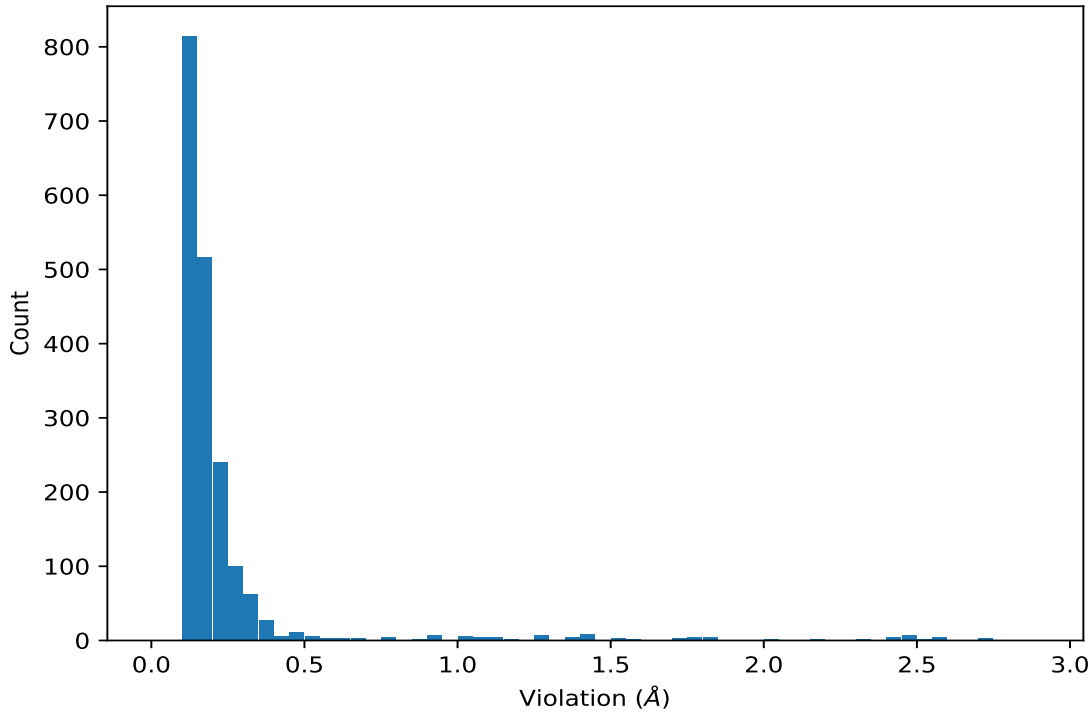
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(3,133)	2:B:94:SER:H	1:A:5:U:H2'	10	2.32	0.43	2.45
(3,130)	2:B:91:GLN:HE21	1:A:5:U:H2'	10	2.25	0.37	2.44
(3,130)	2:B:91:GLN:HE22	1:A:5:U:H2'	10	2.25	0.37	2.44
(3,4)	2:B:86:ASN:HD21	1:A:3:U:H2'	10	1.14	0.51	1.27
(3,4)	2:B:86:ASN:HD22	1:A:3:U:H2'	10	1.14	0.51	1.27
(1,846)	2:B:95:ARG:H	2:B:96:PHE:H	10	0.19	0.02	0.2
(1,1022)	2:B:74:PHE:HA	2:B:75:ARG:HA	10	0.13	0.01	0.12
(1,4652)	3:C:286:PHE:HZ	1:A:2:A:H2'	9	1.72	0.66	1.82
(3,127)	2:B:89:GLY:H	1:A:5:U:H2'	9	1.18	0.55	1.34
(3,70)	2:B:33:ILE:H	1:A:5:U:H2'	9	1.17	0.72	0.87
(1,3574)	3:C:262:PHE:HA	3:C:308:PHE:HE1	9	0.19	0.04	0.18
(1,3574)	3:C:262:PHE:HA	3:C:308:PHE:HE2	9	0.19	0.04	0.18
(1,734)	2:B:76:ASP:H	2:B:76:ASP:HB3	9	0.16	0.03	0.17

[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)
(3,133)	2:B:94:SER:H	1:A:5:U:H2'	7	2.88
(3,133)	2:B:94:SER:H	1:A:5:U:H2'	9	2.73
(3,70)	2:B:33:ILE:H	1:A:5:U:H2'	7	2.7
(1,4652)	3:C:286:PHE:HZ	1:A:2:A:H2'	8	2.7
(3,133)	2:B:94:SER:H	1:A:5:U:H2'	10	2.57
(3,133)	2:B:94:SER:H	1:A:5:U:H2'	3	2.55
(3,130)	2:B:91:GLN:HE21	1:A:5:U:H2'	7	2.55
(3,130)	2:B:91:GLN:HE22	1:A:5:U:H2'	7	2.55
(3,130)	2:B:91:GLN:HE21	1:A:5:U:H2'	8	2.52
(3,130)	2:B:91:GLN:HE22	1:A:5:U:H2'	8	2.52
(3,130)	2:B:91:GLN:HE21	1:A:5:U:H2'	1	2.49

Continued on next page...

Continued from previous page...

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(3,130)	2:B:91:GLN:HE22	1:A:5:U:H2'	1	2.49
(3,133)	2:B:94:SER:H	1:A:5:U:H2'	5	2.47

10 Dihedral-angle violation analysis (i)

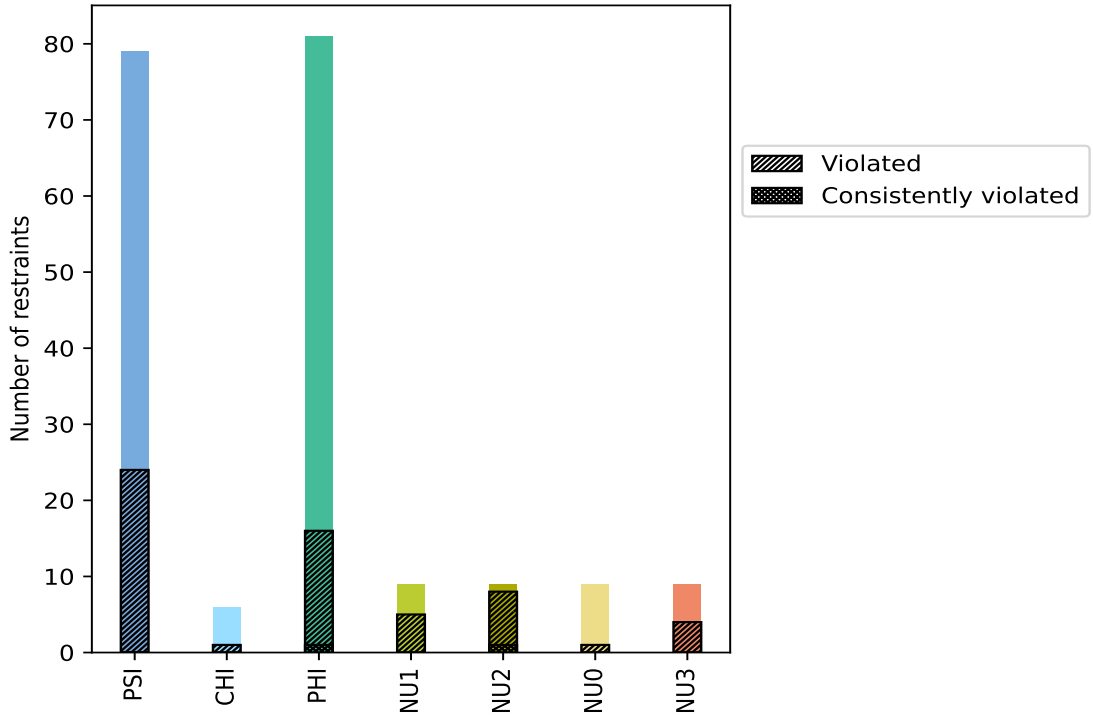
10.1 Summary of dihedral-angle violations (i)

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	79	39.1	24	30.4	11.9	0	0.0	0.0
CHI	6	3.0	1	16.7	0.5	0	0.0	0.0
PHI	81	40.1	16	19.8	7.9	1	1.2	0.5
NU1	9	4.5	5	55.6	2.5	0	0.0	0.0
NU2	9	4.5	8	88.9	4.0	1	11.1	0.5
NU0	9	4.5	1	11.1	0.5	0	0.0	0.0
NU3	9	4.5	4	44.4	2.0	0	0.0	0.0
Total	202	100.0	59	29.2	29.2	2	1.0	1.0

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

10.1.1 Bar chart : Distribution of dihedral-angles and violations (i)



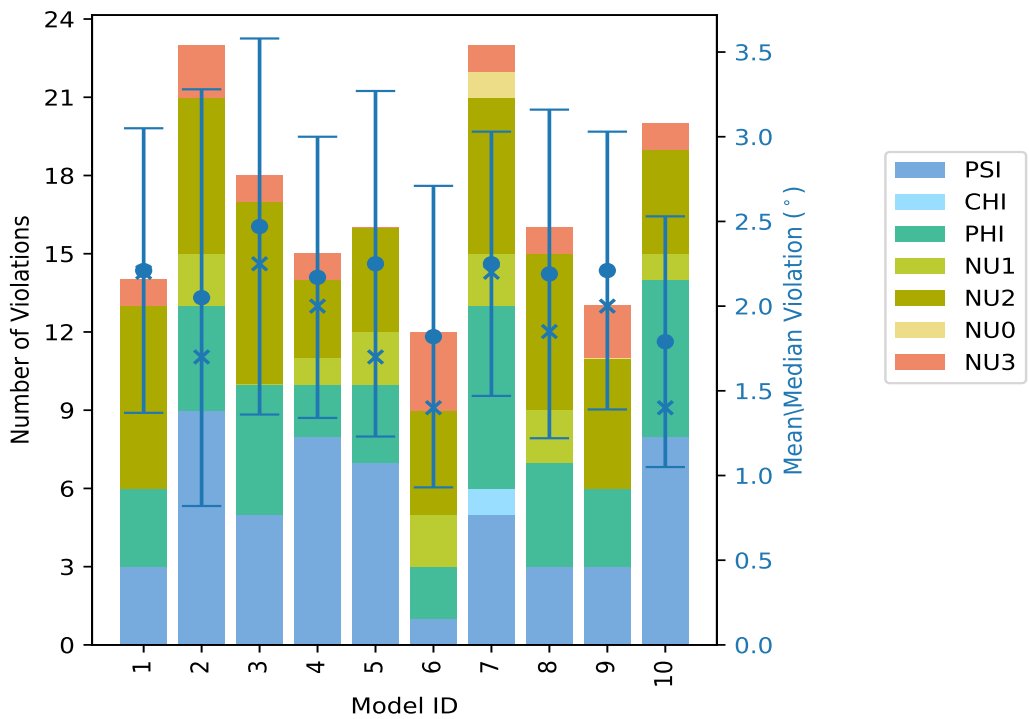
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

10.2 Dihedral-angle violation statistics for each model (i)

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	CHI	PHI	NU1	NU2	NU0	NU3	Total				
1	3	0	3	0	7	0	1	14	2.21	4.0	0.84	2.2
2	9	0	4	2	6	0	2	23	2.05	6.8	1.23	1.7
3	5	0	5	0	7	0	1	18	2.47	5.5	1.11	2.25
4	8	0	2	1	3	0	1	15	2.17	3.7	0.83	2.0
5	7	0	3	2	4	0	0	16	2.25	4.2	1.02	1.7
6	1	0	2	2	4	0	3	12	1.82	4.2	0.89	1.4
7	5	1	7	2	6	1	1	23	2.25	3.9	0.78	2.2
8	3	0	4	2	6	0	1	16	2.19	4.2	0.97	1.85
9	3	0	3	0	5	0	2	13	2.21	3.9	0.82	2.0
10	8	0	6	1	4	0	1	20	1.79	4.1	0.74	1.4

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)

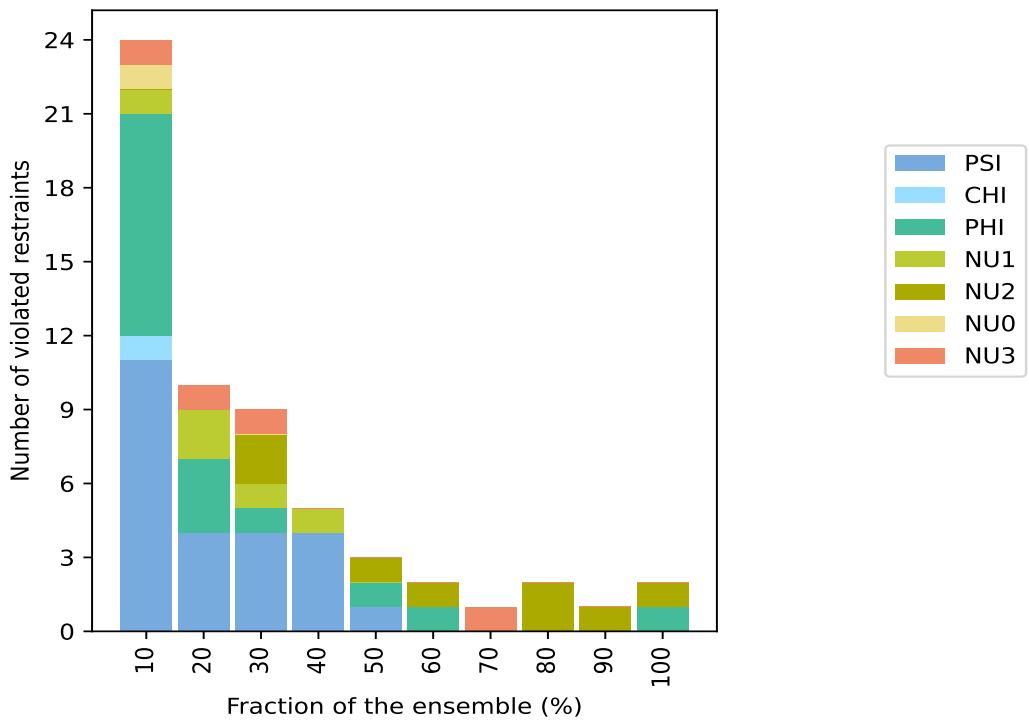
Violation analysis may find that some restraints are violated in very few models and some are violated in most of models. The following table provides this information as number of violated

restraints for a given fraction of ensemble.

Number of violated restraints								Fraction of the ensemble	
PSI	CHI	PHI	NU1	NU2	NU0	NU3	Total	Count[1]	%
11	1	9	1	0	1	1	24	1	10.0
4	0	3	2	0	0	1	10	2	20.0
4	0	1	1	2	0	1	9	3	30.0
4	0	0	1	0	0	0	5	4	40.0
1	0	1	0	1	0	0	3	5	50.0
0	0	1	0	1	0	0	2	6	60.0
0	0	0	0	0	0	1	1	7	70.0
0	0	0	0	2	0	0	2	8	80.0
0	0	0	0	1	0	0	1	9	90.0
0	0	1	0	1	0	0	2	10	100.0

[1] Number of models with violations

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

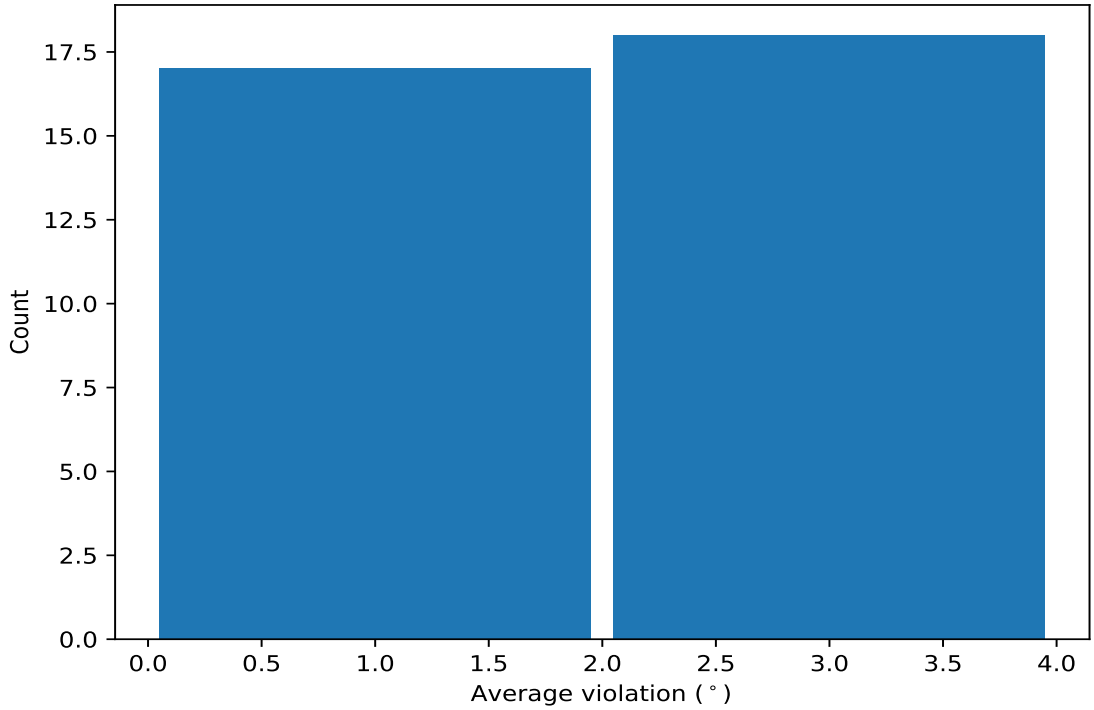


10.4 Most violated dihedral-angle restraints in the ensemble (i)

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.

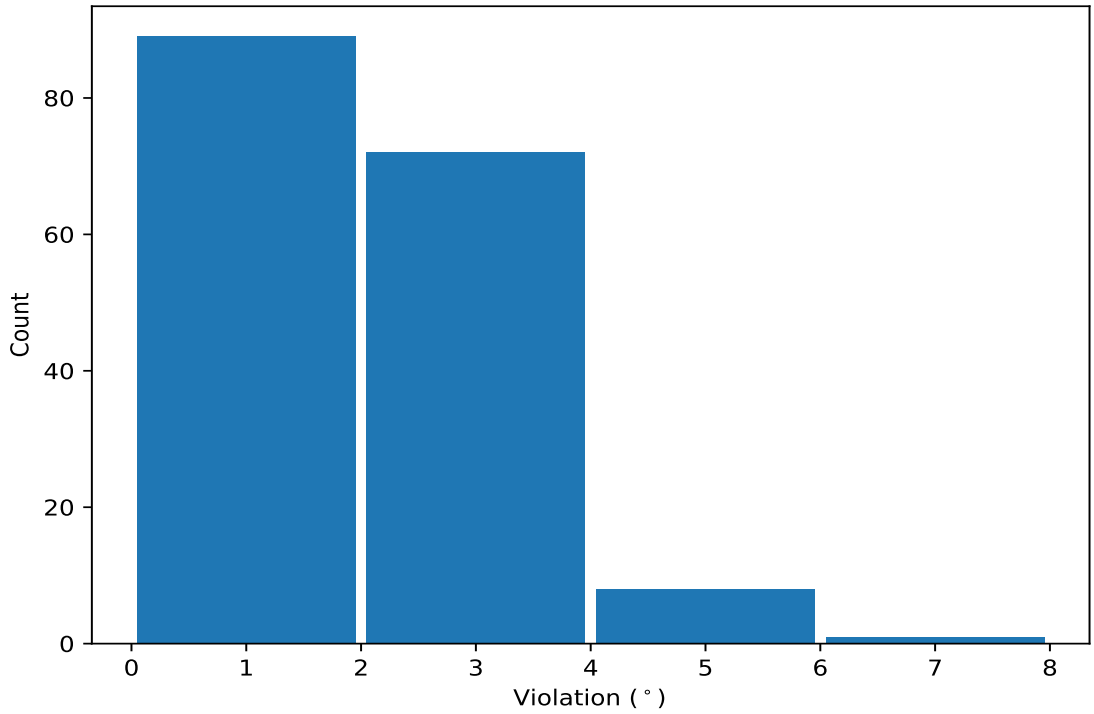
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,200)	3:C:302:GLN:C	3:C:303:ASN:N	3:C:303:ASN:CA	3:C:303:ASN:C	10	3.81	0.76	3.8
(1,93)	1:A:1:U:C1'	1:A:1:U:C2'	1:A:1:U:C3'	1:A:1:U:C4'	10	1.98	0.27	2.0
(1,131)	1:A:13:U:C1'	1:A:13:U:C2'	1:A:13:U:C3'	1:A:13:U:C4'	9	1.33	0.05	1.3
(1,97)	1:A:2:A:C1'	1:A:2:A:C2'	1:A:2:A:C3'	1:A:2:A:C4'	8	2.28	0.34	2.3
(1,105)	1:A:4:A:C1'	1:A:4:A:C2'	1:A:4:A:C3'	1:A:4:A:C4'	8	2.16	0.67	2.25
(1,110)	1:A:6:A:C2'	1:A:6:A:C3'	1:A:6:A:C4'	1:A:6:A:O4'	7	1.69	0.3	1.6
(1,83)	2:B:95:ARG:C	2:B:96:PHE:N	2:B:96:PHE:CA	2:B:96:PHE:C	6	2.83	0.91	3.25
(1,123)	1:A:11:A:C1'	1:A:11:A:C2'	1:A:11:A:C3'	1:A:11:A:C4'	6	1.68	0.39	1.85
(1,13)	2:B:35:TYR:N	2:B:35:TYR:CA	2:B:35:TYR:C	2:B:36:ASP:N	5	2.52	0.91	2.6
(1,137)	3:C:166:LEU:C	3:C:167:ASN:N	3:C:167:ASN:CA	3:C:167:ASN:C	5	1.56	0.39	1.4

[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 (°)
(1,202)	3:C:308:PHE:N	3:C:308:PHE:CA	3:C:308:PHE:C	3:C:309:LYS:N	2	6.8
(1,200)	3:C:302:GLN:C	3:C:303:ASN:N	3:C:303:ASN:CA	3:C:303:ASN:C	3	5.5
(1,89)	2:B:99:CYS:N	2:B:99:CYS:CA	2:B:99:CYS:C	2:B:100:GLY:N	3	4.3
(1,202)	3:C:308:PHE:N	3:C:308:PHE:CA	3:C:308:PHE:C	3:C:309:LYS:N	8	4.2
(1,200)	3:C:302:GLN:C	3:C:303:ASN:N	3:C:303:ASN:CA	3:C:303:ASN:C	5	4.2
(1,200)	3:C:302:GLN:C	3:C:303:ASN:N	3:C:303:ASN:CA	3:C:303:ASN:C	6	4.2
(1,144)	3:C:173:ASP:C	3:C:174:ASN:N	3:C:174:ASN:CA	3:C:174:ASN:C	3	4.2
(1,200)	3:C:302:GLN:C	3:C:303:ASN:N	3:C:303:ASN:CA	3:C:303:ASN:C	10	4.1
(1,157)	3:C:201:GLY:N	3:C:201:GLY:CA	3:C:201:GLY:C	3:C:202:PHE:N	1	4.0
(1,82)	2:B:95:ARG:N	2:B:95:ARG:CA	2:B:95:ARG:C	2:B:96:PHE:N	5	3.9