wwPDB NMR Structure Validation Summary Report

Jun 5, 2023 – 06:30 PM JST

PDB ID	:	7D3V
BMRB ID	:	36387
Title	:	Non-specific and specific interactions work cooperatively to promote cytidine deamination catalyzed by APOBEC3A
Authors	:	Cao, C.Y.; Liu, Y.P.; Lan, W.X.
Deposited on	:	2020-09-21

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:

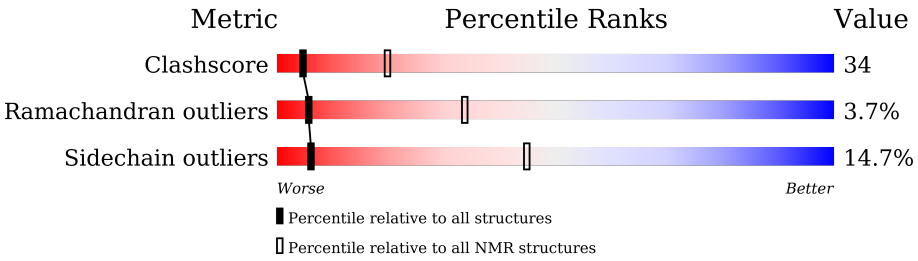
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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 77%.

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	199	46% 41% 11%
1	B	199	59% 29% 6% 6%
2	C	10	40% 40% 20%
2	D	10	70% 30%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 5 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:6-A:199, B:204-B:246, B:250-B:300, B:305-B:398 (382)	0.95	5

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. No single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6978 atoms, of which 3326 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA dC->dU-editing enzyme APOBEC-3A.

Mol	Chain	Residues	Atoms						Trace
1	A	199	Total	C	H	N	O	S	0
			3172	1029	1546	295	293	9	
1	B	199	Total	C	H	N	O	S	0
			3172	1029	1546	295	293	9	

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	63	ASN	LEU	engineered mutation	UNP P31941
A	64	SER	CYS	engineered mutation	UNP P31941
A	72	GLN	GLU	engineered mutation	UNP P31941
A	171	GLN	CYS	engineered mutation	UNP P31941
B	262	ASN	LEU	engineered mutation	UNP P31941
B	263	SER	CYS	engineered mutation	UNP P31941
B	271	GLN	GLU	engineered mutation	UNP P31941
B	370	GLN	CYS	engineered mutation	UNP P31941

- Molecule 2 is a DNA chain called DNA (5'-D(*AP*TP*TP*TP*TP*CP*AP*AP*TP*T)-3').

Mol	Chain	Residues	Atoms						Trace
2	C	10	Total	C	H	N	O	P	0
			316	99	117	30	61	9	
2	D	10	Total	C	H	N	O	P	0
			316	99	117	30	61	9	

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	
3	A	1	Total	Zn
			1	1
3	B	1	Total	Zn
			1	1

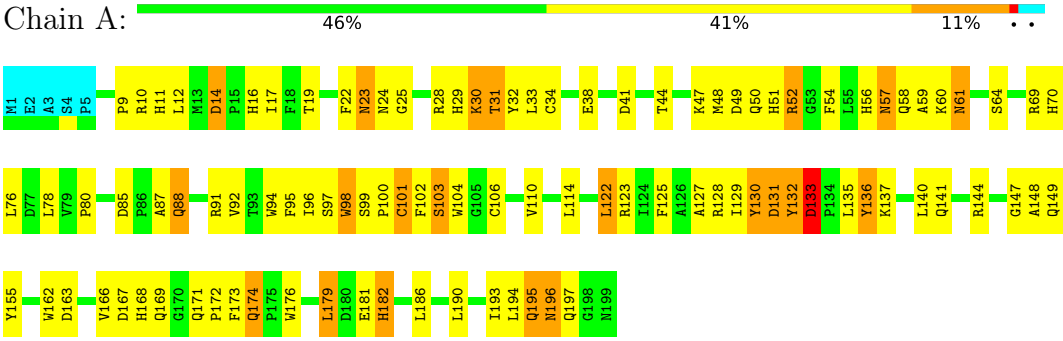
4 Residue-property plots

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.

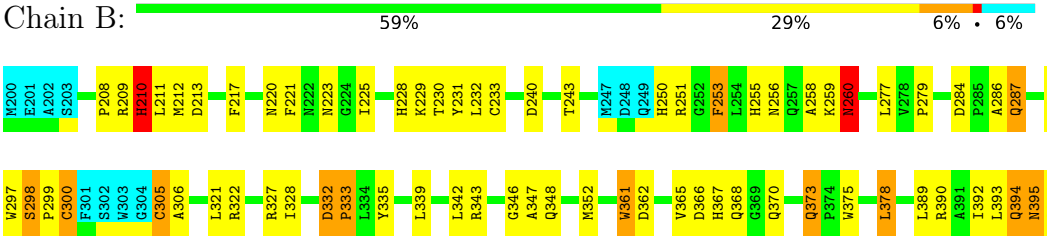
• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3A

Chain A:



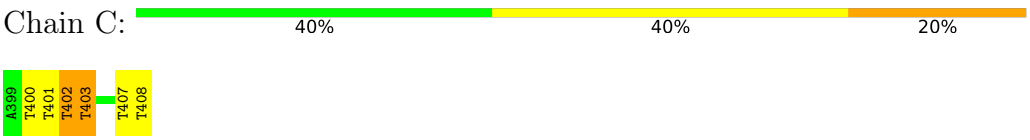
• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3A

Chain B:



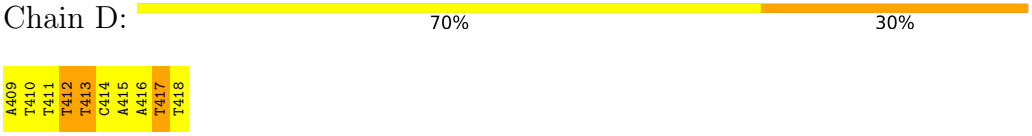
• Molecule 2: DNA (5'-D(*AP*TP*TP*TP*TP*CP*AP*AP*TP*T)-3')

Chain C:



• Molecule 2: DNA (5'-D(*AP*TP*TP*TP*TP*CP*AP*AP*TP*T)-3')

Chain D:

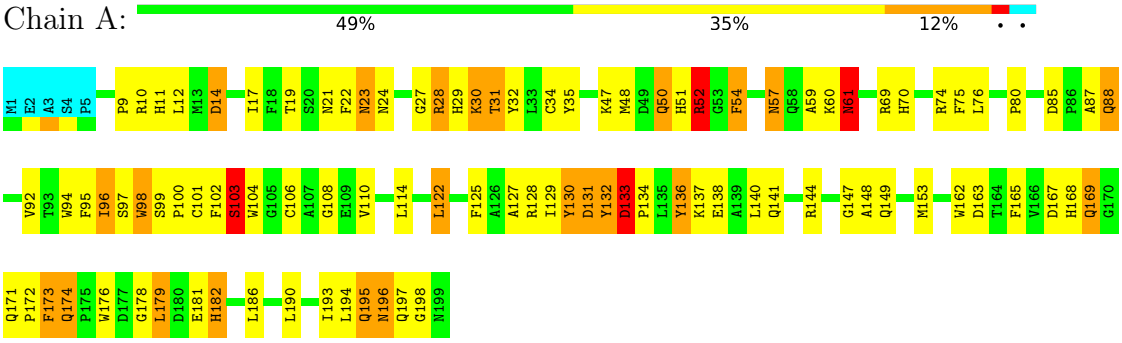


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 5. Colouring as in section 4.1 above.

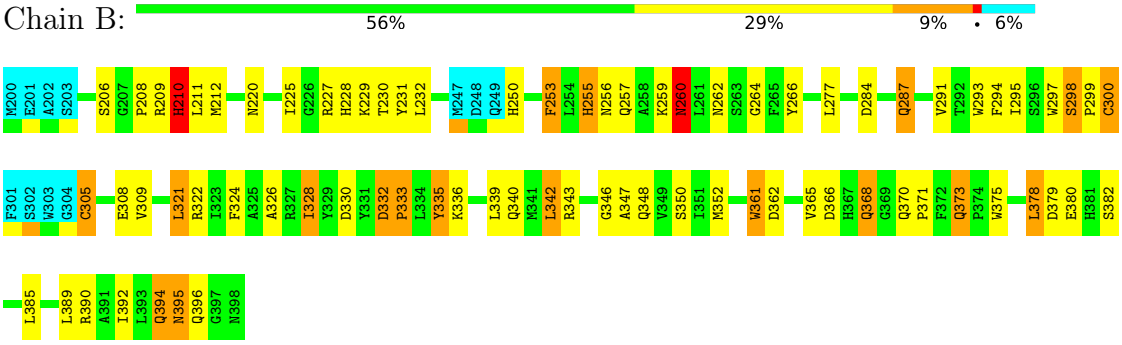
• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3A

Chain A:



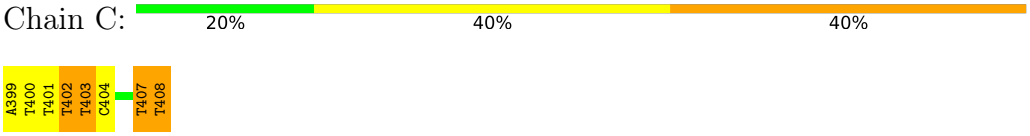
• Molecule 1: DNA dC->dU-editing enzyme APOBEC-3A

Chain B:



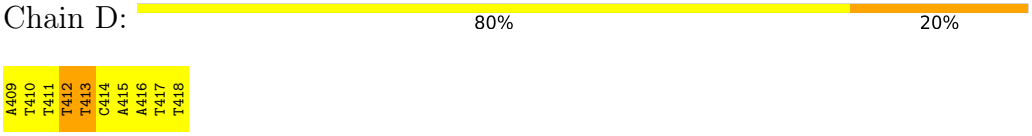
• Molecule 2: DNA (5'-D(*AP*TP*TP*TP*TP*CP*AP*AP*TP*T)-3')

Chain C:



• Molecule 2: DNA (5'-D(*AP*TP*TP*TP*TP*CP*AP*AP*TP*T)-3')

Chain D:



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
X-PLOR NIH	structure calculation	
CNS	refinement	

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	4564
Number of shifts mapped to atoms	4564
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	77%

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	1.00±0.00	0±0/1638 (0.0± 0.0%)	0.92±0.00	1±0/2218 (0.0± 0.0%)
1	B	1.00±0.00	0±0/1583 (0.0± 0.0%)	0.90±0.00	0±0/2144 (0.0± 0.0%)
2	C	1.38±0.04	1±1/221 (0.5± 0.6%)	1.74±0.01	11±1/339 (3.3± 0.4%)
2	D	1.28±0.01	1±1/221 (0.4± 0.3%)	1.71±0.01	12±1/339 (3.4± 0.2%)
All	All	1.05	37/73260 (0.1%)	1.06	464/100800 (0.5%)

5 of 11 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
2	C	402	DT	C5-C7	6.17	1.53	1.50	8	3
2	C	401	DT	C5-C7	5.70	1.53	1.50	8	8
2	C	407	DT	C5-C7	5.25	1.53	1.50	8	1
2	D	412	DT	C5-C7	5.23	1.53	1.50	11	4
2	D	418	DT	C5-C7	5.16	1.53	1.50	20	4

5 of 27 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
2	C	402	DT	C6-C5-C7	-7.81	118.22	122.90	17	19
2	C	401	DT	C6-C5-C7	-6.89	118.77	122.90	16	19
2	C	403	DT	C6-C5-C7	-6.88	118.77	122.90	16	20
2	D	417	DT	C6-C5-C7	-6.88	118.77	122.90	20	20
2	D	410	DT	C6-C5-C7	-6.51	119.00	122.90	4	20

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	1591	1512	1503	135±10
1	B	1538	1471	1462	100±11
2	C	199	117	118	5±2
2	D	199	117	118	22±3
All	All	70580	64340	64020	4584

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 34.

5 of 1402 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:378:LEU:H	1:B:378:LEU:HD22	0.94	1.23	10	12
1:A:179:LEU:HD22	1:A:179:LEU:H	0.93	1.22	3	15
1:B:250:HIS:NE2	1:B:281:LEU:HD21	0.90	1.81	10	1
1:B:361:TRP:CE3	1:B:365:VAL:HG21	0.88	2.03	9	17
1:B:343:ARG:HH11	1:B:393:LEU:HD11	0.88	1.28	17	1

6.3 Torsion angles

6.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	193/199 (97%)	172±2 (89±1%)	12±2 (6±1%)	9±1 (5±1%)	4	27
1	B	187/199 (94%)	168±2 (90±1%)	14±2 (7±1%)	5±1 (3±1%)	8	42
All	All	7600/7960 (95%)	6804 (90%)	514 (7%)	282 (4%)	6	34

5 of 30 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	30	LYS	20
1	A	31	THR	20
1	A	61	ASN	20
1	A	132	TYR	20
1	A	148	ALA	20

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	167/171 (98%)	140±3 (84±2%)	27±3 (16±2%)	5	42
1	B	162/171 (95%)	141±2 (87±1%)	21±2 (13±1%)	7	48
All	All	6580/6840 (96%)	5615 (85%)	965 (15%)	6	45

5 of 108 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	88	GLN	20
1	A	98	TRP	20
1	A	99	SER	20
1	A	101	CYS	20
1	A	122	LEU	20

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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 77% for the well-defined parts and 76% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *A3A_shifts-2plus2_20200918.txt*

7.1.1 Bookkeeping

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

Total number of shifts	4564
Number of shifts mapped to atoms	4564
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	21

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	384	0.04 ± 0.09	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	350	0.68 ± 0.09	Should be checked
$^{13}C'$	345	-0.02 ± 0.08	None needed (< 0.5 ppm)
^{15}N	362	-0.31 ± 0.23	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 77%, i.e. 4403 atoms were assigned a chemical shift out of a possible 5725. 0 out of 54 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	1811/1903 (95%)	749/774 (97%)	709/764 (93%)	353/365 (97%)
Sidechain	2151/2806 (77%)	1539/1813 (85%)	592/852 (69%)	20/141 (14%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	275/640 (43%)	232/313 (74%)	38/280 (14%)	5/47 (11%)
Sugar	140/240 (58%)	140/140 (100%)	0/100 (0%)	0/0 (—%)
Base	26/136 (19%)	26/76 (34%)	0/40 (0%)	0/20 (0%)
Overall	4403/5725 (77%)	2686/3116 (86%)	1339/2036 (66%)	378/573 (66%)

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.

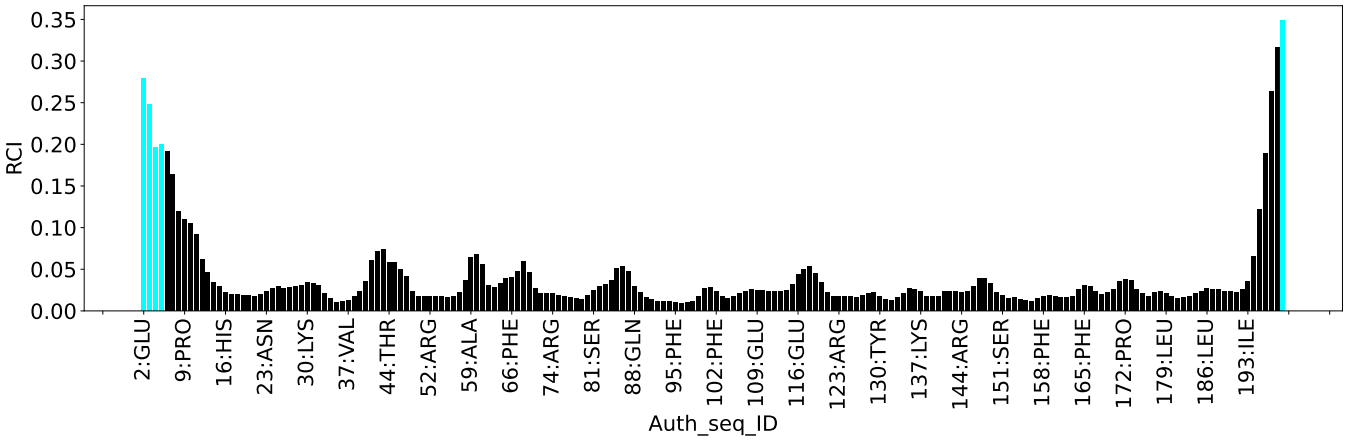
List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	44	THR	HG1	5.07	0.08 – 2.19	18.6
1	B	243	THR	HG1	5.07	0.08 – 2.19	18.6
1	A	70	HIS	CE1	118.42	126.08 – 149.12	-8.3
1	B	269	HIS	CE1	118.42	126.08 – 149.12	-8.3
1	A	94	TRP	HE1	5.58	6.88 – 13.28	-7.0
1	B	293	TRP	HE1	5.58	6.88 – 13.28	-7.0
1	A	66	PHE	CD2	124.86	125.53 – 137.61	-5.6
1	B	265	PHE	CD2	124.86	125.53 – 137.61	-5.6
1	A	72	GLN	CB	38.88	20.34 – 37.98	5.5
1	B	271	GLN	CB	38.88	20.34 – 37.98	5.5
1	A	131	ASP	CB	32.21	32.98 – 48.76	-5.5
1	B	330	ASP	CB	32.21	32.98 – 48.76	-5.5
1	A	66	PHE	CD1	124.86	125.33 – 137.83	-5.4
1	B	265	PHE	CD1	124.86	125.33 – 137.83	-5.4
1	B	369	GLY	H	11.62	5.23 – 11.42	5.3
1	A	170	GLY	H	11.61	5.23 – 11.42	5.3
1	B	378	LEU	HG	-0.22	-0.13 – 3.16	-5.2
1	A	179	LEU	HG	-0.21	-0.13 – 3.16	-5.2
1	A	97	SER	HB3	2.44	2.49 – 5.20	-5.2
1	B	296	SER	HB3	2.44	2.49 – 5.20	-5.2
1	B	371	PRO	CG	32.94	21.69 – 32.72	5.2

7.1.5 Random Coil Index (RCI) plots (i)

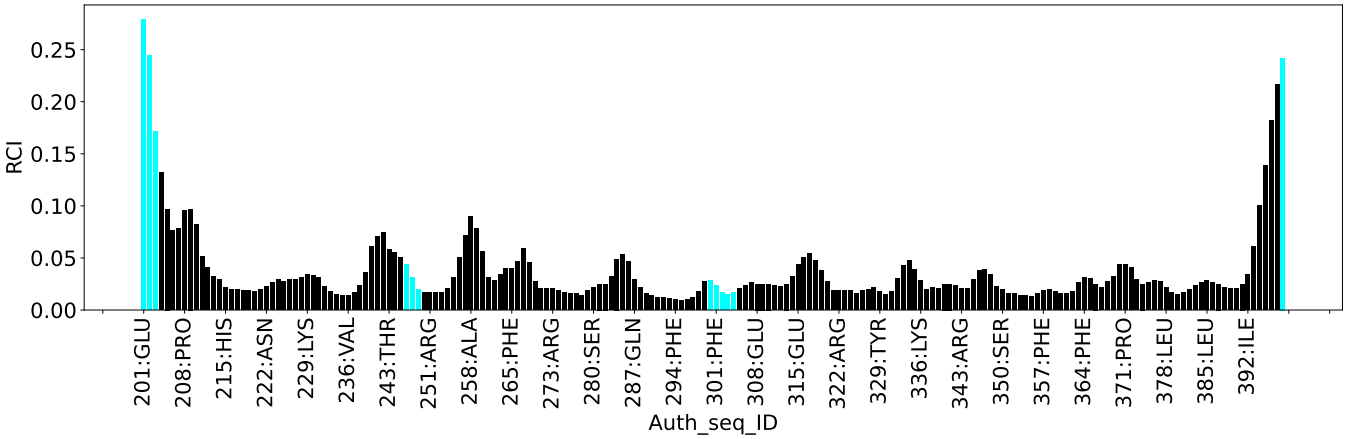
The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted 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:



8 NMR restraints analysis

8.1 Conformationally restricting restraints

The following table provides the summary of experimentally observed NMR restraints in different categories. Restraints are classified into different categories based on the sequence separation of the atoms involved.

Description	Value
Total distance restraints	7003
Intra-residue (|i-j|=0)	3464
Sequential (|i-j|=1)	1379
Medium range (|i-j|>1 and |i-j|<5)	741
Long range (|i-j|≥5)	1316
Inter-chain	103
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	16.8
Number of long range restraints per residue[1]	3.1

[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

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

8.2.1 Average number of distance violations per model

Distance violations less than 0.1 Å are not included in the calculation.

Bins (Å)	Average number of violations per model	Max (Å)
0.1-0.2 (Small)	118.0	0.2
0.2-0.5 (Medium)	38.5	0.5
>0.5 (Large)	107.7	59.27

8.2.2 Average number of dihedral-angle violations per model (i)

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

9 Distance violation analysis (i)

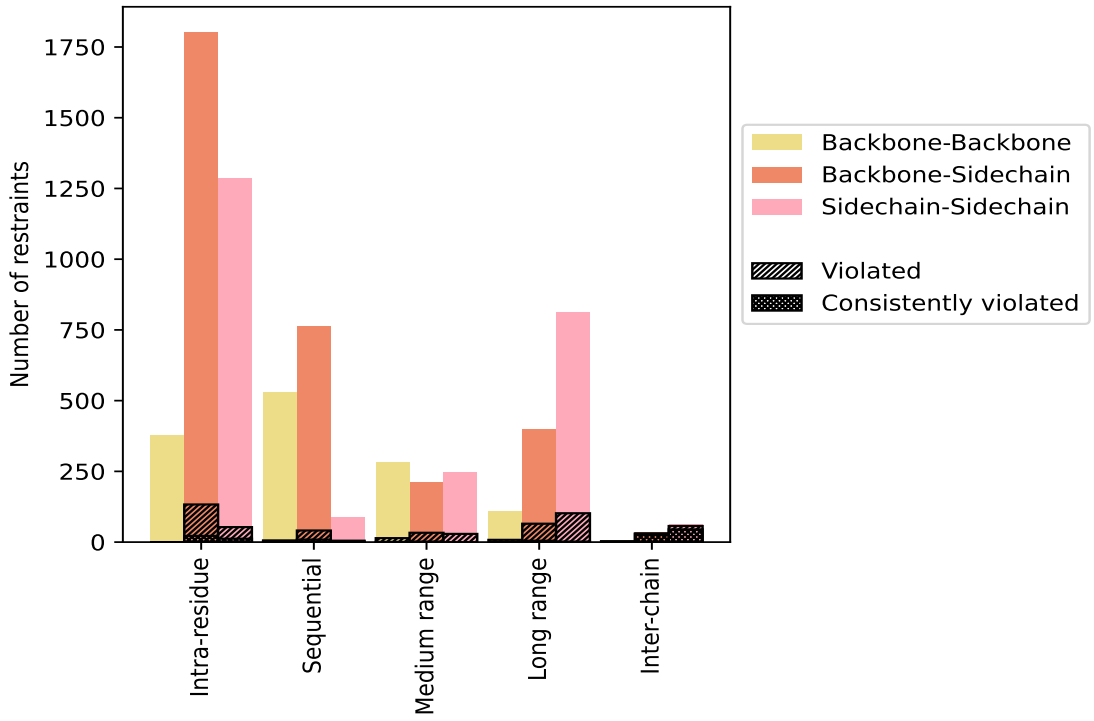
9.1 Summary of distance violations (i)

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)	3464	49.5	186	5.4	2.7	34	1.0	0.5
Backbone-Backbone	376	5.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	1802	25.7	133	7.4	1.9	22	1.2	0.3
Sidechain-Sidechain	1286	18.4	53	4.1	0.8	12	0.9	0.2
Sequential (|i-j|=1)	1379	19.7	52	3.8	0.7	12	0.9	0.2
Backbone-Backbone	530	7.6	6	1.1	0.1	2	0.4	0.0
Backbone-Sidechain	761	10.9	41	5.4	0.6	9	1.2	0.1
Sidechain-Sidechain	88	1.3	5	5.7	0.1	1	1.1	0.0
Medium range (|i-j|>1 & |i-j|<5)	741	10.6	76	10.3	1.1	3	0.4	0.0
Backbone-Backbone	282	4.0	14	5.0	0.2	1	0.4	0.0
Backbone-Sidechain	211	3.0	33	15.6	0.5	2	0.9	0.0
Sidechain-Sidechain	248	3.5	29	11.7	0.4	0	0.0	0.0
Long range (|i-j|≥5)	1316	18.8	175	13.3	2.5	7	0.5	0.1
Backbone-Backbone	108	1.5	8	7.4	0.1	0	0.0	0.0
Backbone-Sidechain	397	5.7	65	16.4	0.9	5	1.3	0.1
Sidechain-Sidechain	811	11.6	102	12.6	1.5	2	0.2	0.0
Inter-chain	103	1.5	91	88.3	1.3	75	72.8	1.1
Backbone-Backbone	6	0.1	3	50.0	0.0	2	33.3	0.0
Backbone-Sidechain	35	0.5	31	88.6	0.4	28	80.0	0.4
Sidechain-Sidechain	62	0.9	57	91.9	0.8	45	72.6	0.6
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	7003	100.0	580	8.3	8.3	131	1.9	1.9
Backbone-Backbone	1302	18.6	31	2.4	0.4	5	0.4	0.1
Backbone-Sidechain	3206	45.8	303	9.5	4.3	66	2.1	0.9
Sidechain-Sidechain	2495	35.6	246	9.9	3.5	60	2.4	0.9

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

9.1.1 Bar chart : Distribution of distance restraints and violations



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

9.2 Distance violation statistics for each model

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

Model ID	Number of violations						Mean (Å)	Max (Å)	SD[6] (Å)	Median (Å)
	IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total				
1	90	25	25	51	85	276	6.12	58.17	13.02	0.31
2	78	16	25	55	85	259	6.48	58.88	13.3	0.31
3	78	21	21	55	86	261	6.38	58.94	13.27	0.27
4	88	18	22	51	85	264	6.41	58.95	13.25	0.26
5	83	17	25	39	84	248	6.77	57.57	13.53	0.34
6	89	23	24	47	85	268	6.22	58.96	13.15	0.24
7	97	22	27	56	86	288	5.87	58.8	12.74	0.22
8	89	22	29	53	85	278	6.16	59.27	13.11	0.22
9	81	17	31	41	82	252	6.57	58.77	13.38	0.34
10	80	21	29	38	84	252	6.71	58.65	13.57	0.28
11	83	22	29	41	84	259	6.41	58.47	13.37	0.23

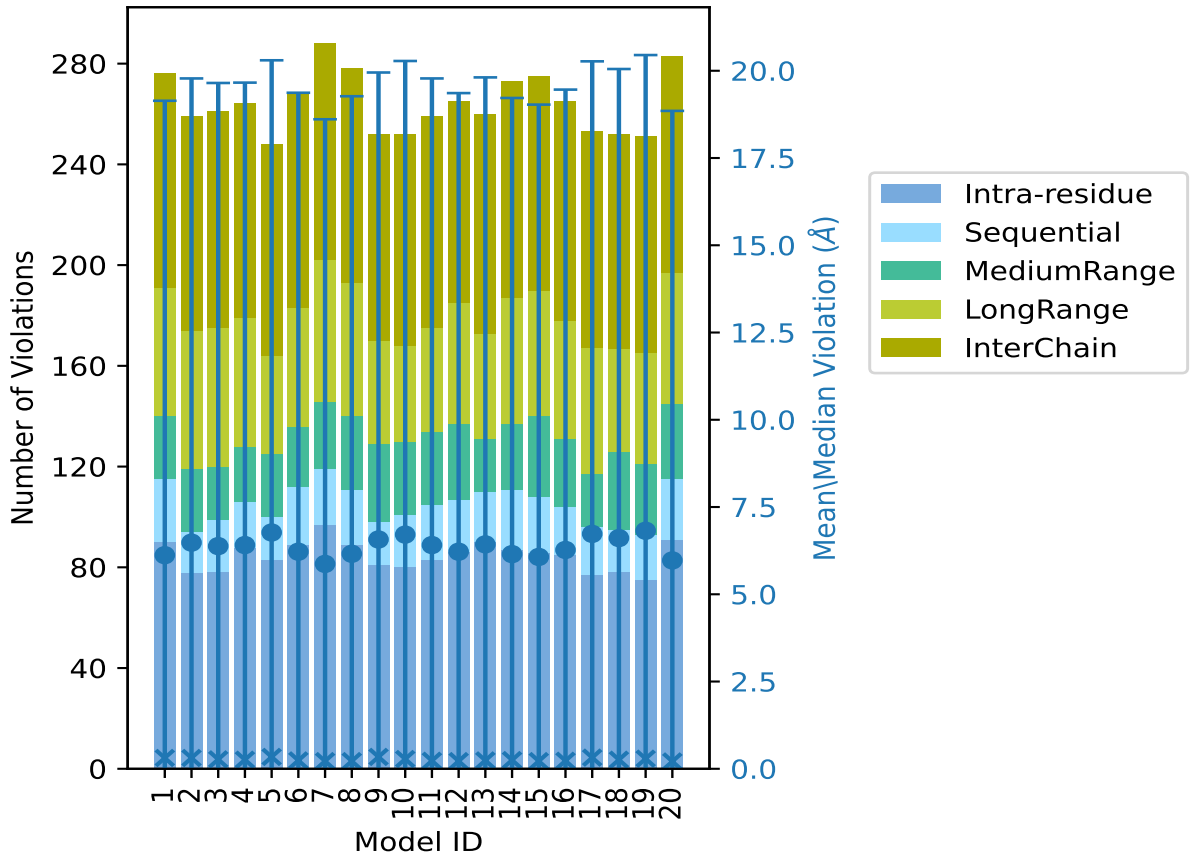
Continued on next page...

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	86	21	30	48	80	265	6.22	58.32	13.14	0.22
13	90	20	21	42	87	260	6.43	58.93	13.38	0.24
14	87	24	26	50	86	273	6.15	58.61	13.07	0.25
15	86	22	32	50	85	275	6.07	57.58	12.96	0.22
16	85	19	27	47	87	265	6.27	58.88	13.19	0.23
17	77	19	21	50	86	253	6.73	58.88	13.54	0.32
18	78	17	31	41	85	252	6.61	57.55	13.44	0.26
19	75	18	28	44	86	251	6.82	59.23	13.63	0.29
20	91	24	30	52	86	283	5.97	57.67	12.88	0.21

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

9.2.1 Bar graph : Distance Violation statistics for each model



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

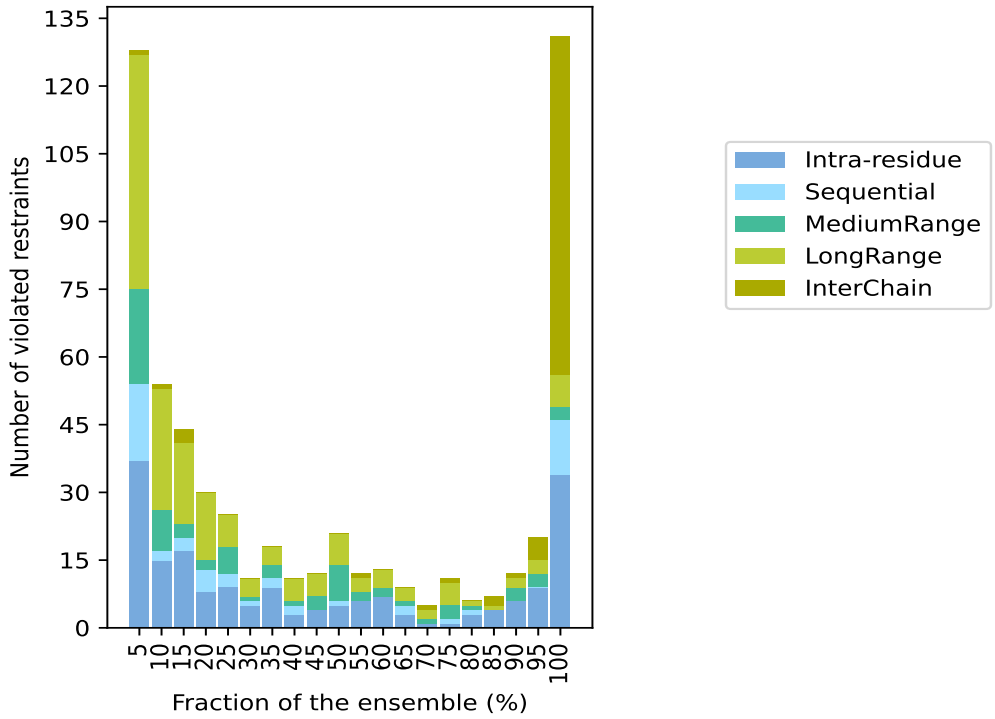
9.3 Distance violation statistics for the ensemble (i)

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, 6423(IR:3278, SQ:1327, MR:665, LR:1141, IC:12) 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]	%
37	17	21	52	1	128	1	5.0
15	2	9	27	1	54	2	10.0
17	3	3	18	3	44	3	15.0
8	5	2	15	0	30	4	20.0
9	3	6	7	0	25	5	25.0
5	1	1	4	0	11	6	30.0
9	2	3	4	0	18	7	35.0
3	2	1	5	0	11	8	40.0
4	0	3	5	0	12	9	45.0
5	1	8	7	0	21	10	50.0
6	0	2	3	1	12	11	55.0
7	0	2	4	0	13	12	60.0
3	2	1	3	0	9	13	65.0
1	0	1	2	1	5	14	70.0
1	1	3	5	1	11	15	75.0
3	1	1	1	0	6	16	80.0
4	0	0	1	2	7	17	85.0
6	0	3	2	1	12	18	90.0
9	0	3	3	5	20	19	95.0
34	12	3	7	75	131	20	100.0

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

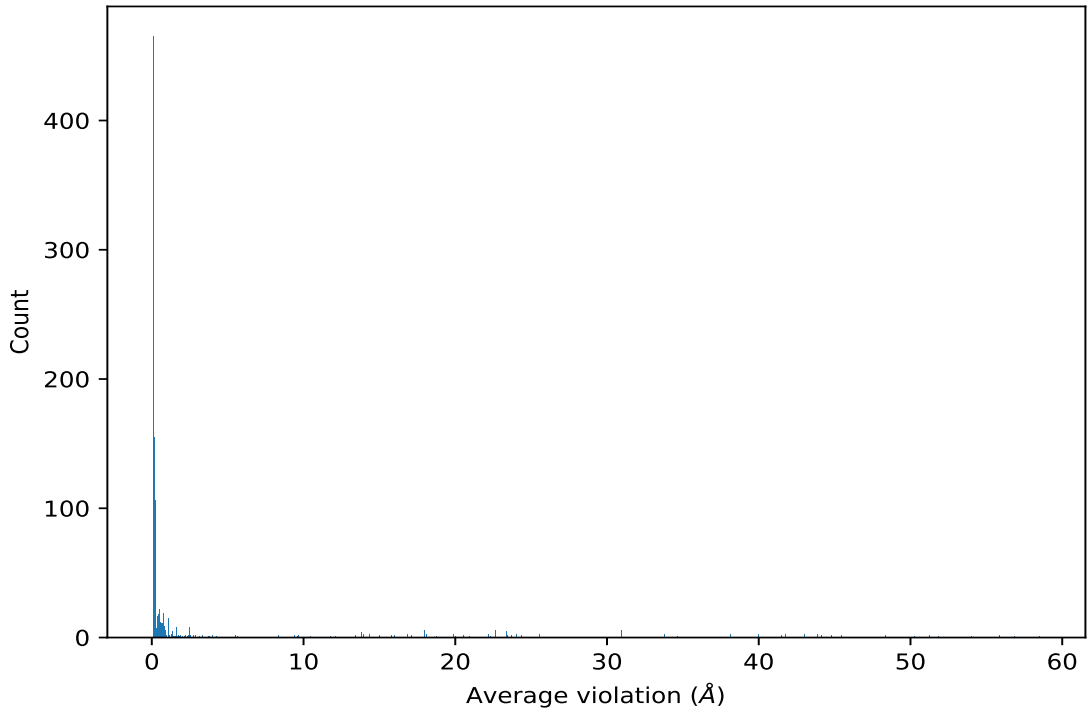
9.3.1 Bar graph : Distance violation statistics for the ensemble



9.4 Most violated distance restraints in the ensemble

9.4.1 Histogram : Distribution of mean distance violations

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

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.

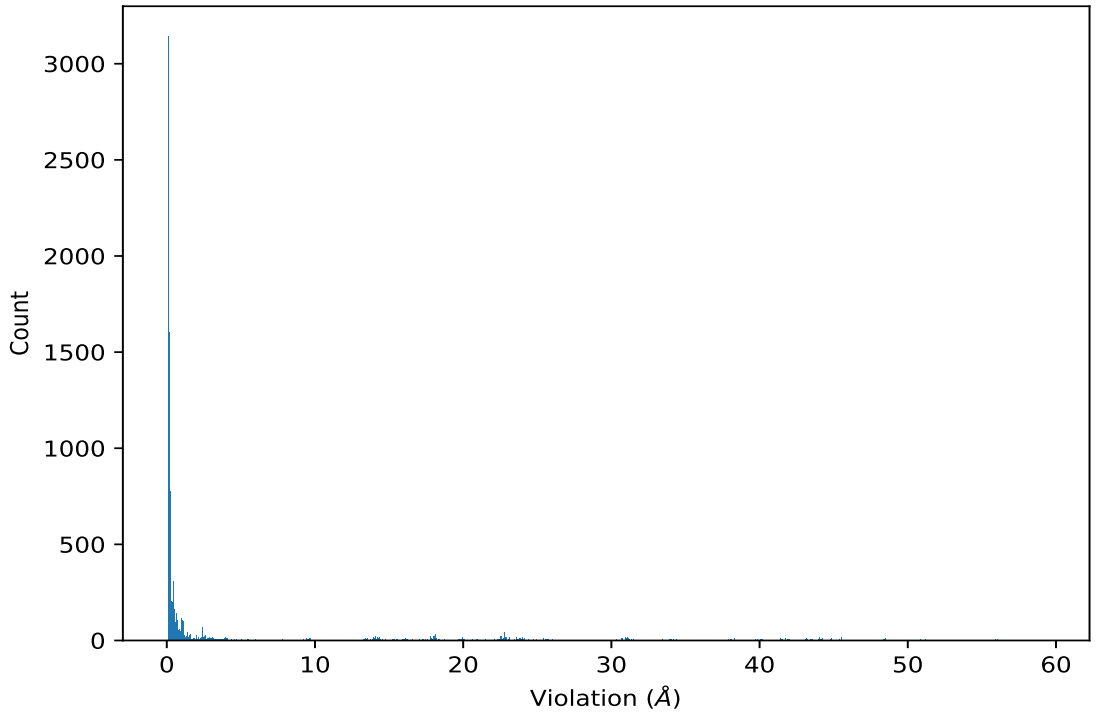
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	20	58.53	0.59	58.78
(1,6716)	2:C:399:DA:H1'	1:A:184:GLN:HG3	20	56.82	0.59	56.6
(1,6722)	2:C:399:DA:H8	1:A:181:GLU:HA	20	55.87	0.17	55.88
(1,6721)	2:C:399:DA:H4'	1:A:185:ALA:HA	20	55.85	0.34	55.96
(1,6718)	2:C:399:DA:H5"	1:A:185:ALA:HA	20	54.01	0.35	54.14
(1,6726)	2:C:400:DT:H5'	1:A:185:ALA:HA	20	51.82	0.33	51.94
(1,6724)	2:C:399:DA:H8	1:A:184:GLN:HG2	20	51.28	0.27	51.24
(1,6724)	2:C:399:DA:H8	1:A:184:GLN:HG3	20	51.28	0.27	51.24
(1,6728)	2:C:400:DT:H5"	1:A:185:ALA:HA	20	50.3	0.34	50.43
(1,6723)	2:C:399:DA:H8	1:A:181:GLU:HB2	20	48.39	0.17	48.41
(1,6723)	2:C:399:DA:H8	1:A:181:GLU:HB3	20	48.39	0.17	48.41
(1,6731)	2:C:400:DT:H5'	1:A:181:GLU:HB3	20	45.46	0.19	45.5

[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)
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	8	59.27
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	19	59.23
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	6	58.96
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	4	58.95
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	3	58.94
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	13	58.93
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	2	58.88
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	16	58.88
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	17	58.88
(1,6717)	2:C:399:DA:H1'	1:A:184:GLN:HG2	7	58.8

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found