wwPDB NMR Structure Validation Summary Report (i)

Jun 5, 2023 – 02:42 PM JST

PDB ID	:	7D5F
BMRB ID	:	36390
Title	:	Left-handed G-quadruplex containing 3 bulges
Authors	:	Winnerdy, F.R.; Das, P.; Ngo, K.H.; Maity, A.; Bakalar, B.; Mechulam, Y.; Schmitt, E.; Phan, A.T.
Deposited on	:	2020-09-25

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:

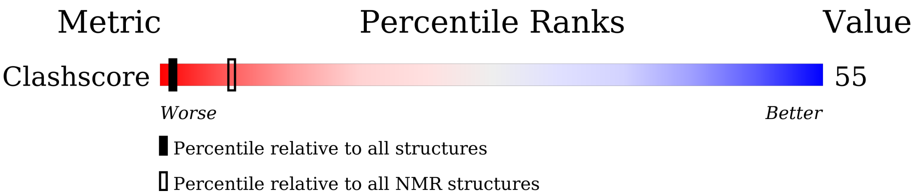
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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 30%.

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

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	27	59% 41%

2 Ensemble composition and analysis (i)

This entry contains 10 models. This entry does not contain polypeptide chains, therefore identification of well-defined residues and clustering analysis are not possible. All residues are included in the validation scores.

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 879 atoms, of which 310 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called 3xBulge-LHG4motif.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	27	879	270	310	102	171	26	0

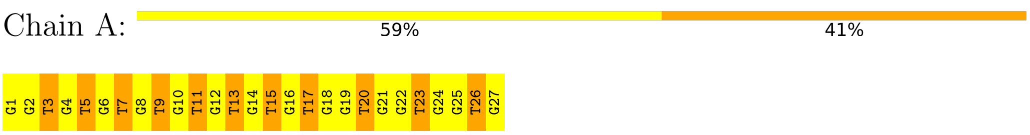
4 Residue-property plots (i)

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: 3xBulge-LHG4motif

Chain A:

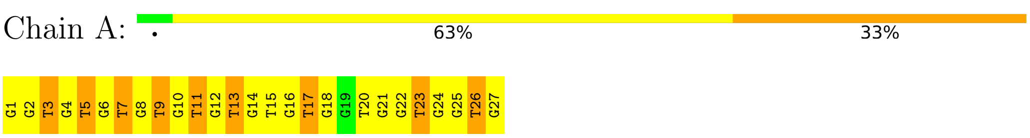


4.2 Residue scores for the representative (author defined) model from the NMR ensemble

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

- Molecule 1: 3xBulge-LHG4motif

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing, molecular 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	structure calculation	
X-PLOR NIH	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	181
Number of shifts mapped to atoms	181
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	30%

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	1.06±0.01	1±1/638 (0.2± 0.2%)	1.43±0.00	22±0/990 (2.2± 0.0%)
All	All	1.06	10/6380 (0.2%)	1.43	217/9900 (2.2%)

5 of 9 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	3	DT	C5-C7	5.09	1.53	1.50	6	2
1	A	20	DT	C5-C7	5.08	1.53	1.50	4	1
1	A	26	DT	C5-C7	5.05	1.53	1.50	10	1
1	A	13	DT	C5-C7	5.04	1.53	1.50	2	1
1	A	23	DT	C5-C7	5.04	1.53	1.50	10	1

5 of 22 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	20	DT	C6-C5-C7	-5.86	119.38	122.90	3	10
1	A	7	DT	C6-C5-C7	-5.86	119.39	122.90	10	10
1	A	11	DT	C6-C5-C7	-5.85	119.39	122.90	5	10
1	A	3	DT	C6-C5-C7	-5.84	119.40	122.90	8	10
1	A	5	DT	C6-C5-C7	-5.83	119.40	122.90	9	10

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	569	310	310	48±12
All	All	5690	3100	3100	484

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 55.

5 of 278 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:6:DG:H3'	1:A:6:DG:OP2	0.97	1.59	1	1
1:A:21:DG:H3'	1:A:21:DG:OP2	0.96	1.60	2	5
1:A:24:DG:H3'	1:A:24:DG:OP2	0.95	1.60	7	6
1:A:27:DG:H3'	1:A:27:DG:OP2	0.95	1.61	2	3
1:A:4:DG:H3'	1:A:4:DG:OP2	0.91	1.66	6	5

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

There are no protein molecules in this entry.

6.3.2 Protein sidechains (i)

There are no protein molecules in this entry.

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

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 30% for the well-defined parts and 30% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

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	181
Number of shifts mapped to atoms	181
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

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 30%, i.e. 160 atoms were assigned a chemical shift out of a possible 534. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Sugar	121/324 (37%)	121/189 (64%)	0/135 (0%)	0/0 (—%)
Base	39/210 (19%)	39/129 (30%)	0/38 (0%)	0/43 (0%)
Overall	160/534 (30%)	160/318 (50%)	0/173 (0%)	0/43 (0%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

No *random coil index*(RCI) plot could be generated from the current chemical shift list. RCI is only applicable to proteins

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	654
Intra-residue (|i-j|=0)	365
Sequential (|i-j|=1)	105
Medium range (|i-j|>1 and |i-j|<5)	125
Long range (|i-j|≥5)	59
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	16
Number of unmapped restraints	0
Number of restraints per residue	24.8
Number of long range restraints per residue[1]	2.2

[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)	9.6	0.19
0.2-0.5 (Medium)	None	None
>0.5 (Large)	None	None

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)

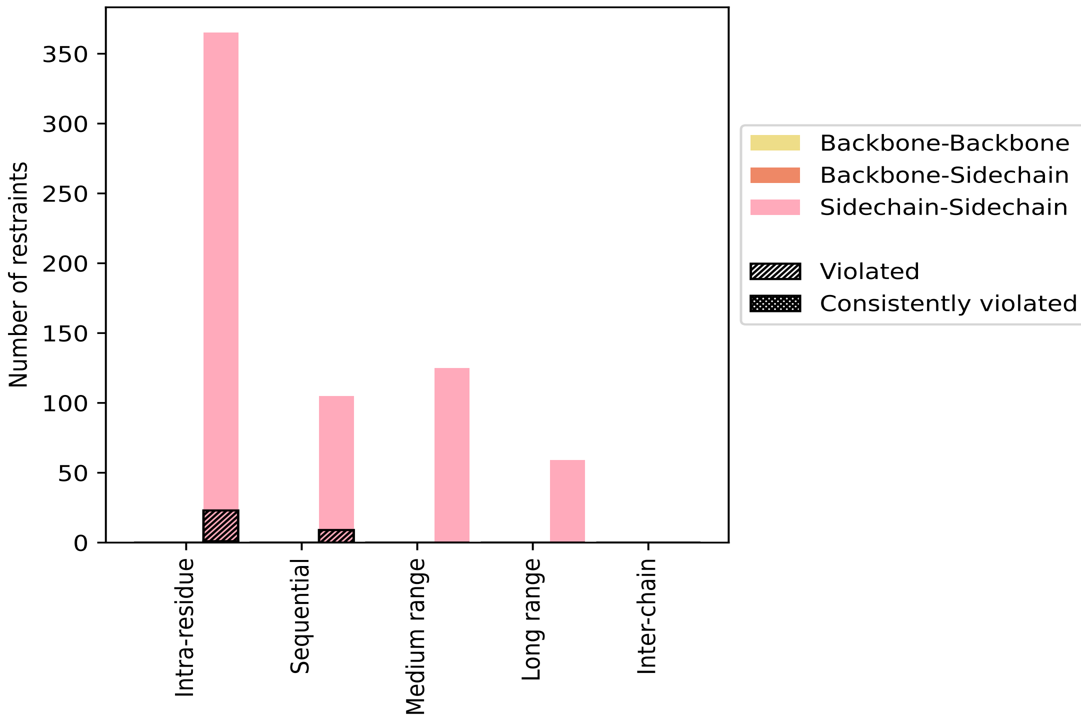
9.1 Summary of distance violations (i)

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)	365	55.8	23	6.3	3.5	1	0.3	0.2
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	365	55.8	23	6.3	3.5	1	0.3	0.2
Sequential (|i-j|=1)	105	16.1	9	8.6	1.4	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	105	16.1	9	8.6	1.4	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	125	19.1	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	125	19.1	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	59	9.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	59	9.0	0	0.0	0.0	0	0.0	0.0
Inter-chain	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
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	654	100.0	32	4.9	4.9	1	0.2	0.2
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	654	100.0	32	4.9	4.9	1	0.2	0.2

[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)

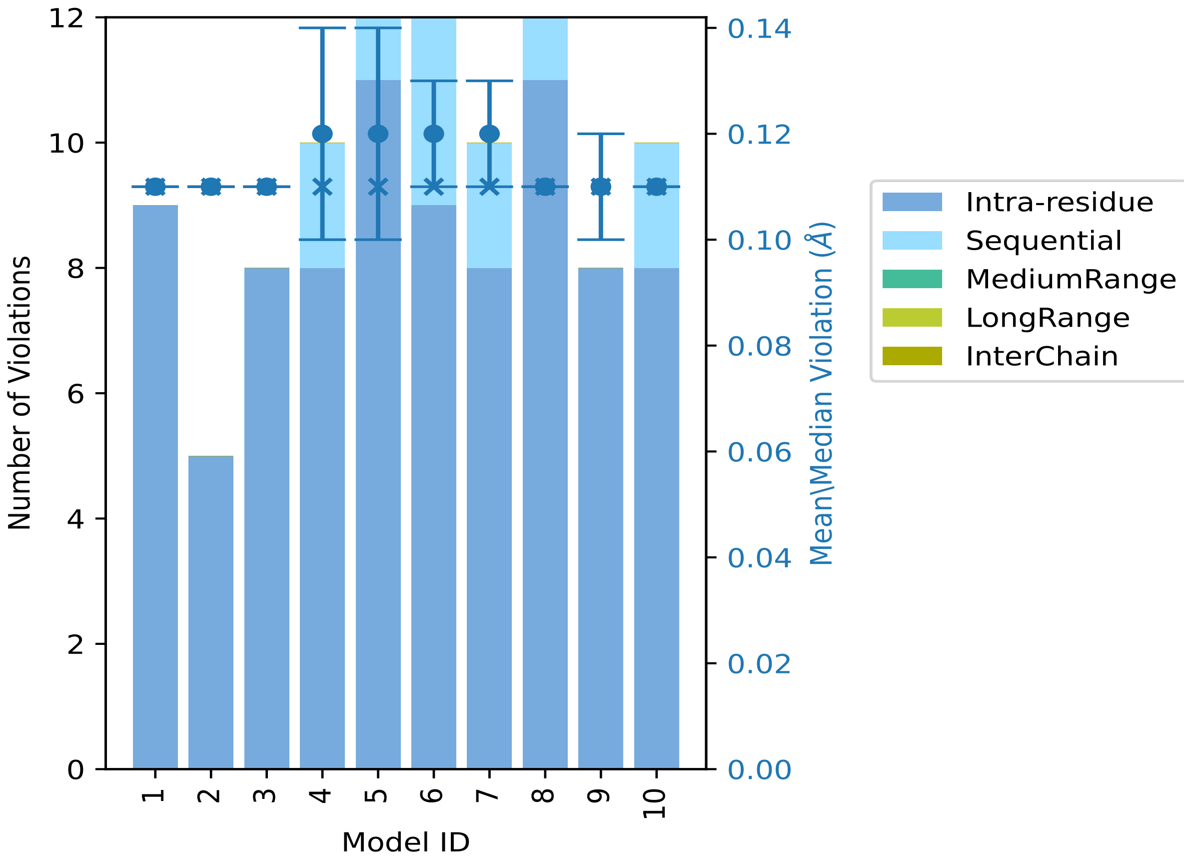
The following table provides the distance violation statistics for each model in the ensemble. Violations less than 0.1 Å are not included in the statistics.

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	9	0	0	0	0	9	0.11	0.12	0.0	0.11
2	5	0	0	0	0	5	0.11	0.12	0.0	0.11
3	8	0	0	0	0	8	0.11	0.12	0.0	0.11
4	8	2	0	0	0	10	0.12	0.18	0.02	0.11
5	11	1	0	0	0	12	0.12	0.19	0.02	0.11
6	9	3	0	0	0	12	0.12	0.16	0.01	0.11
7	8	2	0	0	0	10	0.12	0.16	0.01	0.11
8	11	1	0	0	0	12	0.11	0.12	0.0	0.11
9	8	0	0	0	0	8	0.11	0.13	0.01	0.11
10	8	2	0	0	0	10	0.11	0.12	0.0	0.11

[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, 622(IR:342, SQ:96, MR:125, LR:59, IC:0) restraints are not violated in the ensemble.

Number of violated restraints						Fraction of the ensemble	
IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total	Count[6]	%
4	7	0	0	0	11	1	10.0
3	2	0	0	0	5	2	20.0
6	0	0	0	0	6	3	30.0
3	0	0	0	0	3	4	40.0

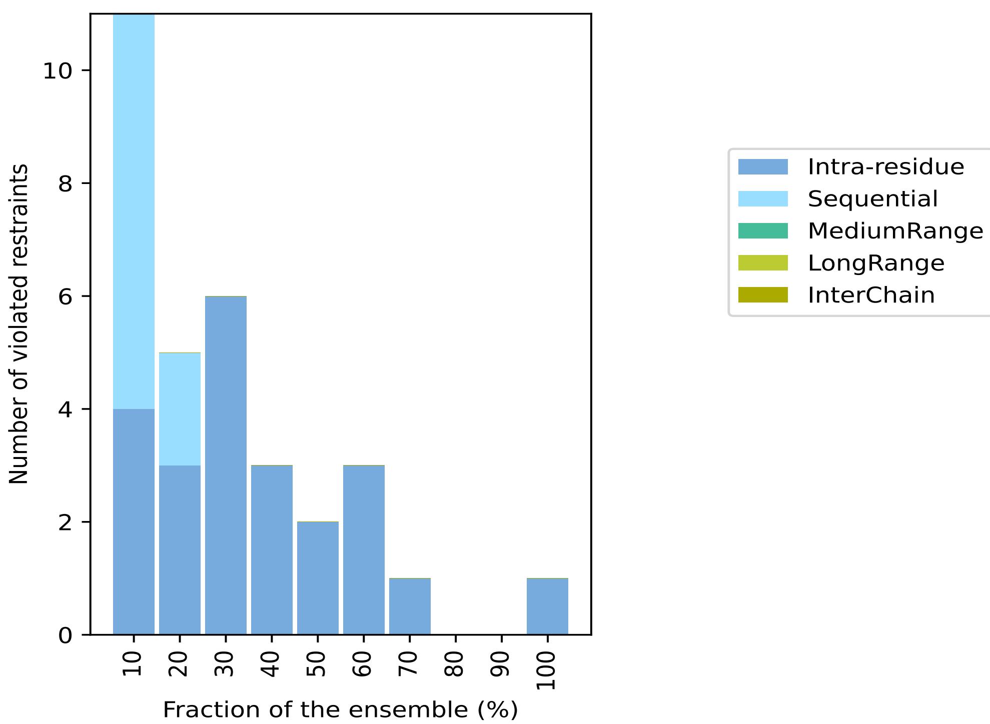
Continued on next page...

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

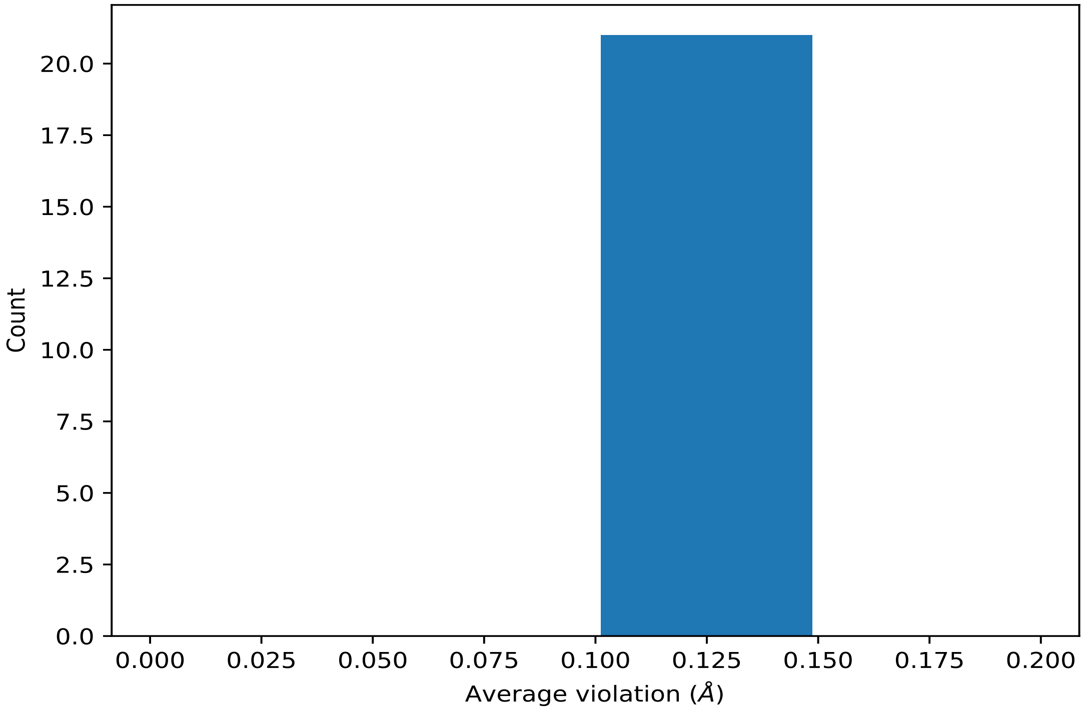
9.3.1 Bar graph : Distance violation statistics for the ensemble (i)



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.

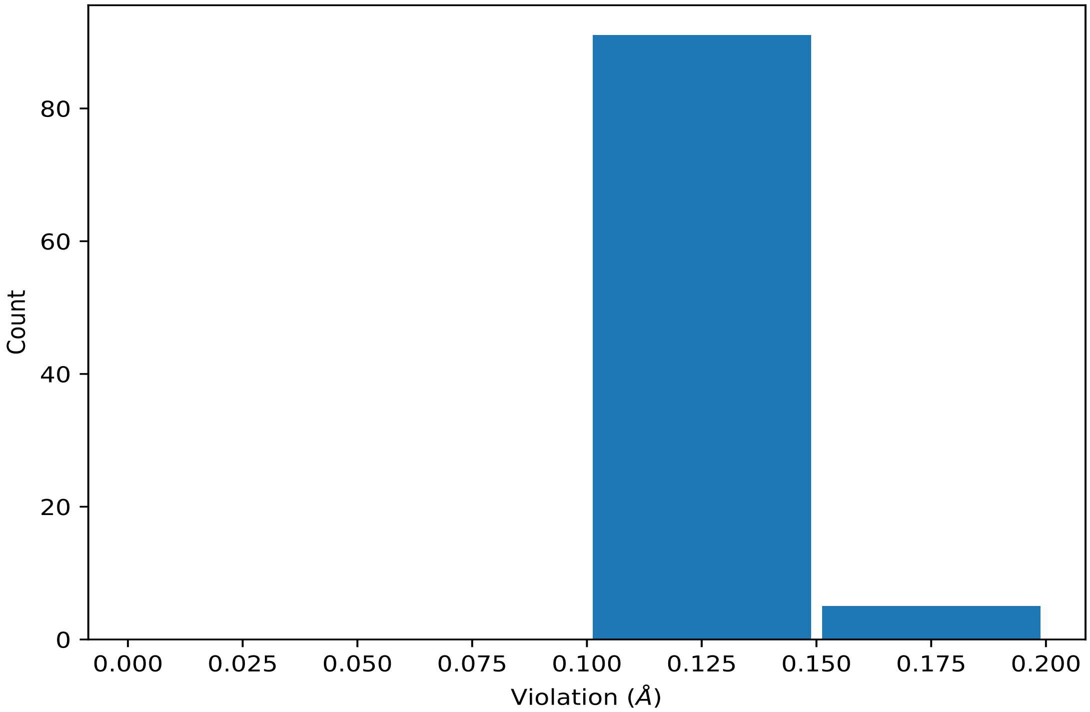
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(2,27)	1:A:1:DG:H5"	1:A:1:DG:H5'	10	0.12	0.0	0.12
(2,322)	1:A:25:DG:H2'	1:A:25:DG:H2"	7	0.11	0.0	0.11
(2,47)	1:A:3:DT:H2"	1:A:3:DT:H2'	6	0.11	0.0	0.11
(2,81)	1:A:6:DG:H2'	1:A:6:DG:H2"	6	0.11	0.0	0.11
(2,97)	1:A:7:DT:H2'	1:A:7:DT:H2"	6	0.11	0.0	0.11
(2,123)	1:A:10:DG:H2'	1:A:10:DG:H2"	5	0.11	0.0	0.11
(2,246)	1:A:20:DT:H2"	1:A:20:DT:H2'	5	0.11	0.0	0.11
(2,31)	1:A:2:DG:H2'	1:A:2:DG:H2"	4	0.11	0.0	0.11
(2,173)	1:A:14:DG:H2'	1:A:14:DG:H2"	4	0.11	0.0	0.11
(2,184)	1:A:16:DG:H2'	1:A:16:DG:H2"	4	0.11	0.0	0.11

[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)
(2,210)	1:A:17:DT:H4’	1:A:17:DT:H1’	5	0.19
(1,17)	1:A:7:DT:H1’	1:A:6:DG:H1	4	0.18
(2,490)	1:A:22:DG:H3’	1:A:21:DG:H1’	6	0.16
(2,279)	1:A:22:DG:H2’	1:A:22:DG:H8	7	0.16
(1,65)	1:A:17:DT:H1’	1:A:16:DG:H1	5	0.16
(2,279)	1:A:22:DG:H2’	1:A:22:DG:H8	9	0.13
(2,501)	1:A:24:DG:H1’	1:A:25:DG:H8	4	0.12
(2,482)	1:A:21:DG:H1’	1:A:22:DG:H8	6	0.12
(2,482)	1:A:21:DG:H1’	1:A:22:DG:H8	8	0.12
(2,462)	1:A:18:DG:H1’	1:A:19:DG:H8	10	0.12

10 Dihedral-angle violation analysis (i)

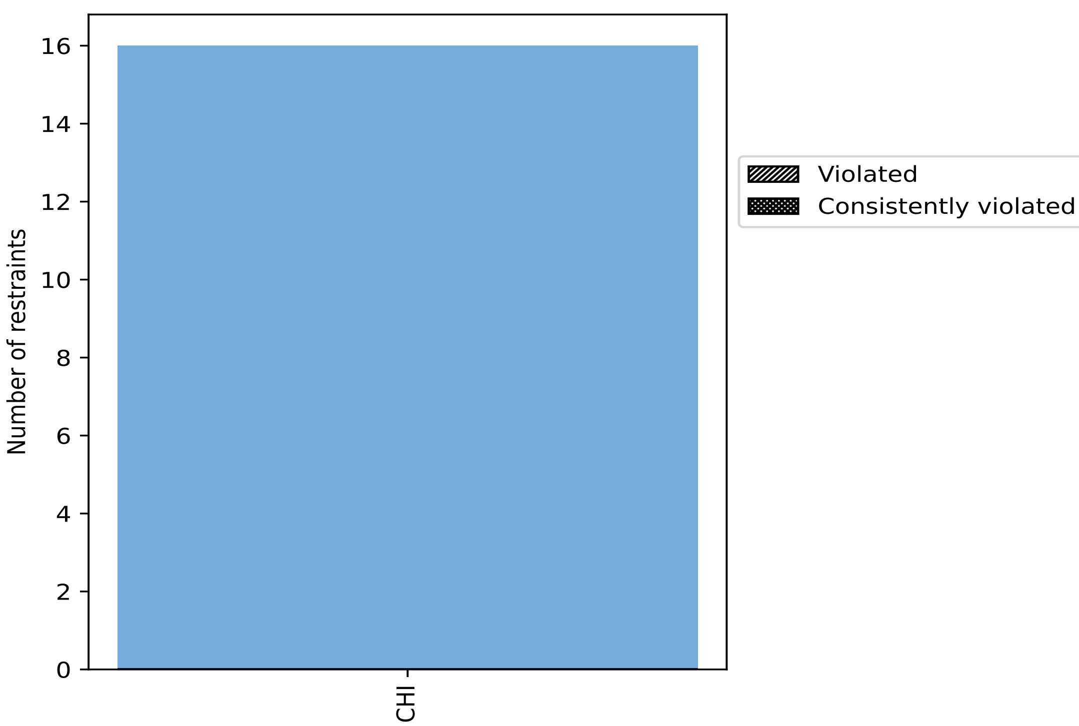
10.1 Summary of dihedral-angle violations (i)

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]
CHI	16	100.0	0	0.0	0.0	0	0.0	0.0
Total	16	100.0	0	0.0	0.0	0	0.0	0.0

[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)

No violations found

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

No violations found

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

No violations found

10.5 All violated dihedral-angle restraints (i)

No violations found