wwPDB NMR Structure Validation Summary Report

Jun 5, 2023 – 09:27 AM EDT

PDB ID	:	2M39
BMRB ID	:	18950
Title	:	The solution structure of 3',5'-LINKED 2'-O-(2-METHOXYETHYL)-RNA DUPLEX
Authors	:	Plevnik, M.; Cevec, M.; Plavec, J.
Deposited on	:	2013-01-15

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	:	FAILED
Mogul	:	1.8.5 (274361), 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 7%.

There are no overall percentile quality scores available for this entry.

The sequence quality summary graphics cannot be shown.

2 Ensemble composition and analysis (i)

This entry contains 9 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 1024 atoms, of which 418 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called RNA (5'-R(*(C5L)P*(G48)P*(C5L)P*(G48)P*(A44)P*(A44)P*(T39)P*(T39)P*(C5L)P*(G48)P*(C5L)P*DG)-3').

Mol	Chain	Residues	Atoms							Trace
1	A	12	Total	C	H	Ho	N	O	P	0
			512	153	209	1	46	92	11	
1	B	12	Total	C	H	Ho	N	O	P	0
			512	153	209	1	46	92	11	

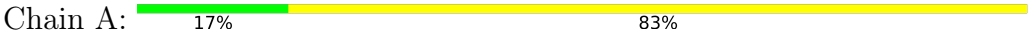
4 Residue-property plots (i)

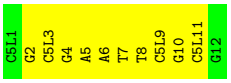
4.1 Average score per residue in the NMR ensemble

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: RNA (5'-R(*(C5L)P*(G48)P*(C5L)P*(G48)P*(A44)P*(A44)P*(T39)P*(T39)P*(C5L)P*(G48)P*(C5L)P*DG)-3')

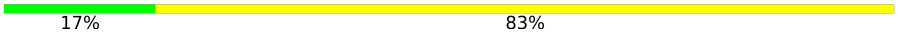
Chain A:

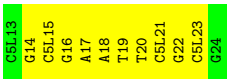




• Molecule 1: RNA (5'-R(*(C5L)P*(G48)P*(C5L)P*(G48)P*(A44)P*(A44)P*(T39)P*(T39)P*(C5L)P*(G48)P*(C5L)P*DG)-3')

Chain B:



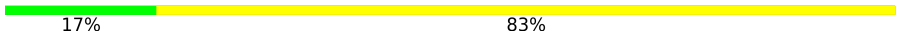


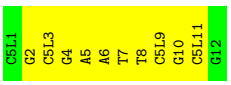
4.2 Residue scores for the representative (author defined) model from the NMR ensemble

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

• Molecule 1: RNA (5'-R(*(C5L)P*(G48)P*(C5L)P*(G48)P*(A44)P*(A44)P*(T39)P*(T39)P*(C5L)P*(G48)P*(C5L)P*DG)-3')

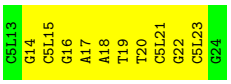
Chain A:





• Molecule 1: RNA (5'-R(*(C5L)P*(G48)P*(C5L)P*(G48)P*(A44)P*(A44)P*(T39)P*(T39)P*(C5L)P*(G48)P*(C5L)P*DG)-3')

Chain B:


C5L13
G14
C5L15
G16
A17
A18
T19
T20
C5L21
G22
C5L23
G24

5 Refinement protocol and experimental data overview

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

Of the 50 calculated structures, 9 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
Amber	structure solution	7
Amber	refinement	7

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	247
Number of shifts mapped to atoms	247
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	7%

6 Model quality

6.1 Standard geometry

MolProbity failed to run properly - this section will have to be empty.

6.2 Too-close contacts

MolProbity failed to run properly - this section will have to be empty.

6.3 Torsion angles

6.3.1 Protein backbone

MolProbity failed to run properly - this section will have to be empty.

6.3.2 Protein sidechains

MolProbity failed to run properly - this section will have to be empty.

6.3.3 RNA

MolProbity failed to run properly - this section will have to be empty.

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

MolProbity failed to run properly - this section will have to be empty.

6.5 Carbohydrates

MolProbity failed to run properly - this section will have to be empty.

6.6 Ligand geometry

MolProbity failed to run properly - this section will have to be empty.

6.7 Other polymers

MolProbity failed to run properly - this section will have to be empty.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 7% for the well-defined parts and 7% 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

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	247
Number of shifts mapped to atoms	247
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

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

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 7%, i.e. 3 atoms were assigned a chemical shift out of a possible 42. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Sugar	2/24 (8%)	2/14 (14%)	0/10 (0%)	0/0 (—%)
Base	1/18 (6%)	1/12 (8%)	0/2 (0%)	0/4 (0%)
Overall	3/42 (7%)	3/26 (12%)	0/12 (0%)	0/4 (0%)

7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots

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

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	239
Intra-residue (|i-j|=0)	58
Sequential (|i-j|=1)	107
Medium range (|i-j|>1 and |i-j|<5)	19
Long range (|i-j|≥5)	55
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	10.0
Number of long range restraints per residue[1]	2.3

[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)	9.9	0.2
0.2-0.5 (Medium)	11.2	0.49
>0.5 (Large)	78.6	30.21

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)

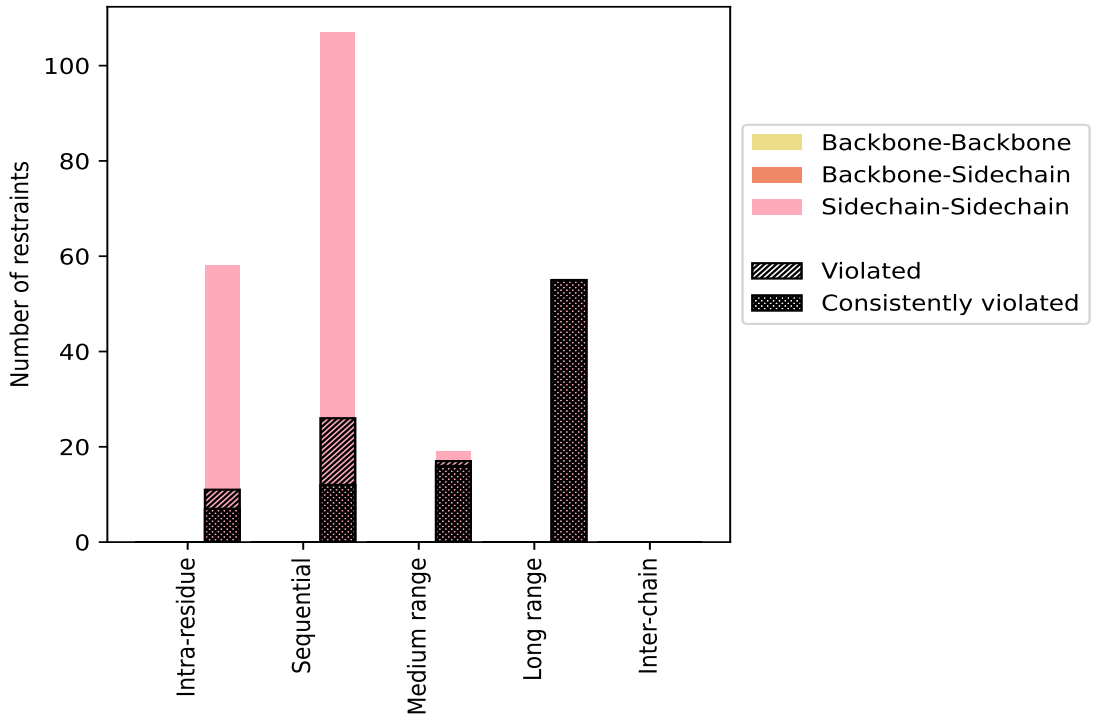
9.1 Summary of distance violations (i)

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)	58	24.3	11	19.0	4.6	7	12.1	2.9
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	58	24.3	11	19.0	4.6	7	12.1	2.9
Sequential (|i-j|=1)	107	44.8	26	24.3	10.9	12	11.2	5.0
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	107	44.8	26	24.3	10.9	12	11.2	5.0
Medium range (|i-j|>1 & |i-j|<5)	19	7.9	17	89.5	7.1	16	84.2	6.7
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	19	7.9	17	89.5	7.1	16	84.2	6.7
Long range (|i-j|≥5)	55	23.0	55	100.0	23.0	55	100.0	23.0
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	55	23.0	55	100.0	23.0	55	100.0	23.0
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	239	100.0	109	45.6	45.6	90	37.7	37.7
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	239	100.0	109	45.6	45.6	90	37.7	37.7

[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



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

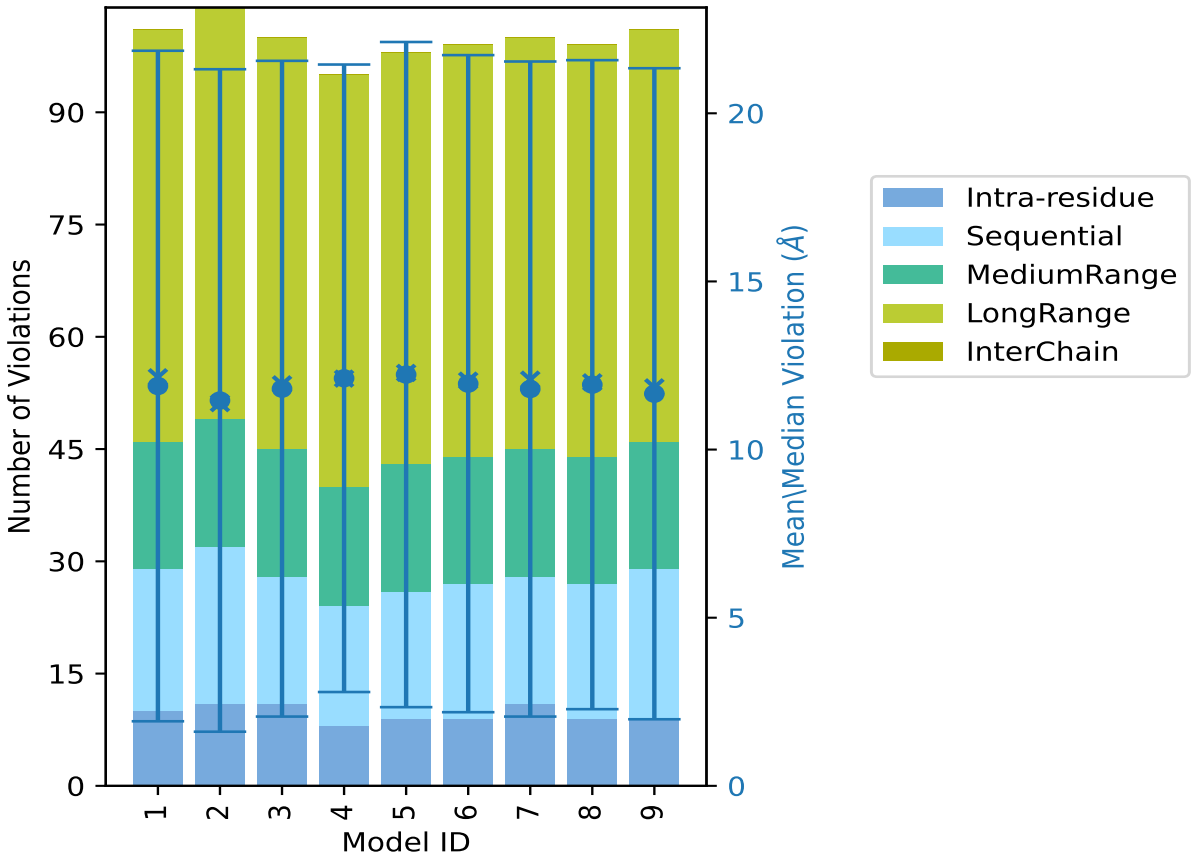
9.2 Distance violation statistics for each model

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	10	19	17	55	0	101	11.89	30.11	9.97	12.15
2	11	21	17	55	0	104	11.46	29.33	9.85	11.38
3	11	17	17	55	0	100	11.81	29.68	9.75	11.94
4	8	16	16	55	0	95	12.12	28.35	9.33	12.11
5	9	17	17	55	0	98	12.23	30.21	9.89	12.28
6	9	18	17	55	0	99	11.96	29.82	9.77	12.05
7	11	17	17	55	0	100	11.8	29.09	9.74	12.08
8	9	18	17	55	0	99	11.93	28.88	9.65	12.0
9	9	20	17	55	0	101	11.66	28.69	9.68	11.85

[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, 130(IR:47, SQ:81, MR:2, LR:0, 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]	%
0	4	0	0	0	4	1	11.1
0	0	0	0	0	0	2	22.2
0	0	0	0	0	0	3	33.3
1	6	0	0	0	7	4	44.4
1	1	0	0	0	2	5	55.6
0	1	0	0	0	1	6	66.7

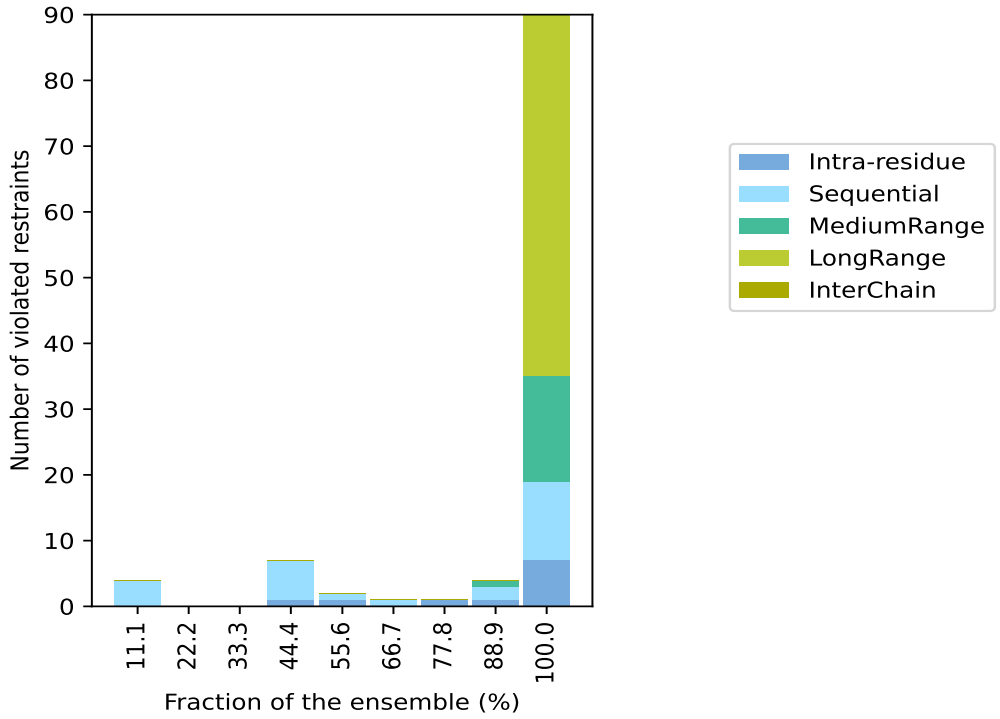
Continued on next page...

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	0	0	0	0	1	7	77.8
1	2	1	0	0	4	8	88.9
7	12	16	55	0	90	9	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

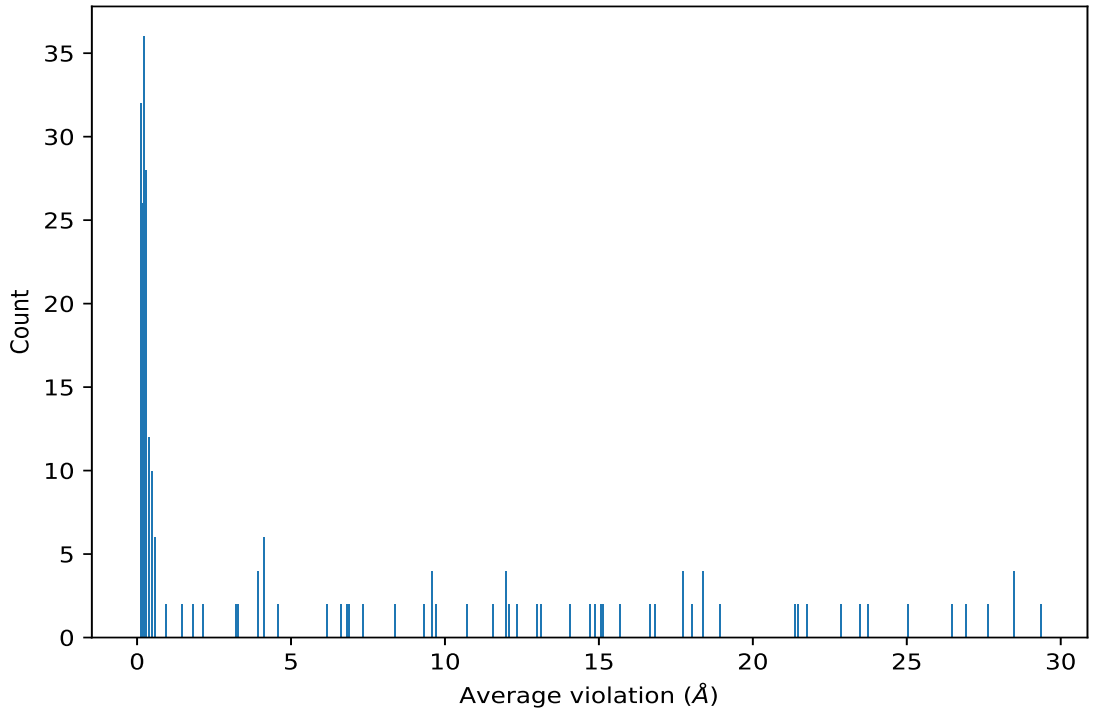
9.3.1 Bar graph : Distance violation statistics for the ensemble



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 (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.

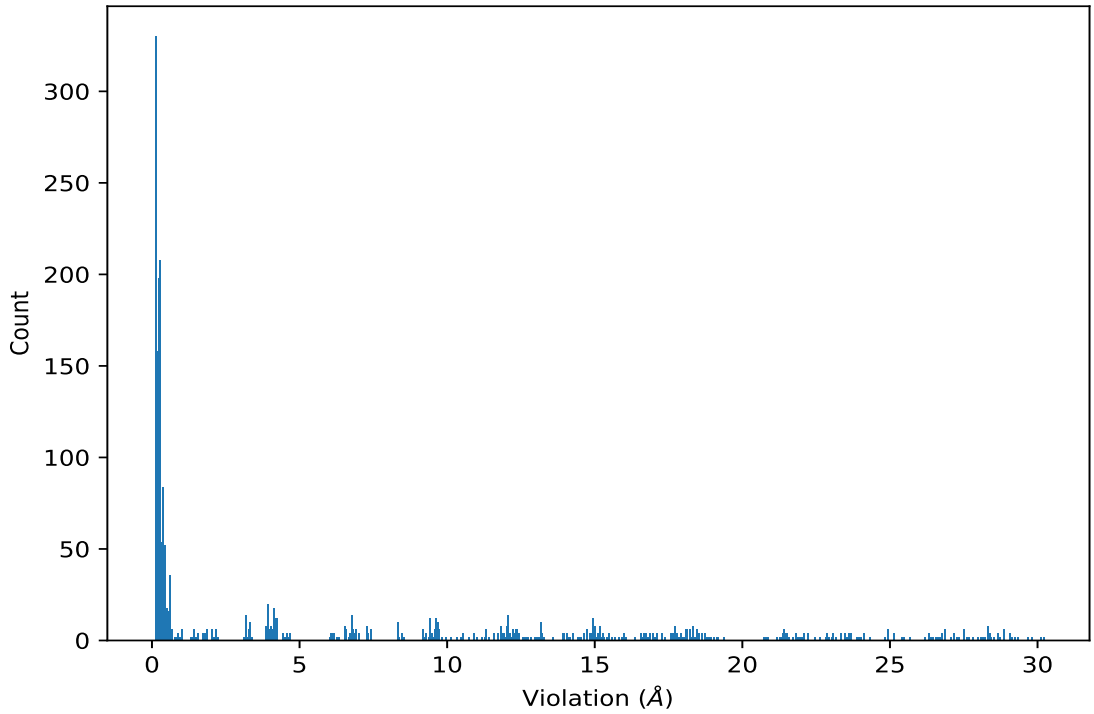
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(2,47)	1:B:13:C5L:O2	1:B:24:DG:H21	9	29.35	0.61	29.33
(2,52)	1:B:13:C5L:O2	1:B:24:DG:H21	9	29.35	0.61	29.33
(2,44)	1:B:13:C5L:N3	1:B:24:DG:H1	9	28.49	0.5	28.38
(2,49)	1:B:13:C5L:N3	1:B:24:DG:H1	9	28.49	0.5	28.38
(2,43)	1:B:13:C5L:H41	1:B:24:DG:O6	9	28.46	0.39	28.41
(2,48)	1:B:13:C5L:H41	1:B:24:DG:O6	9	28.46	0.39	28.41
(2,45)	1:B:13:C5L:N3	1:B:24:DG:N1	9	27.61	0.49	27.53
(2,50)	1:B:13:C5L:N3	1:B:24:DG:N1	9	27.61	0.49	27.53
(2,46)	1:B:13:C5L:N4	1:B:24:DG:O6	9	26.93	0.38	26.89
(2,51)	1:B:13:C5L:N4	1:B:24:DG:O6	9	26.93	0.38	26.89

[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)
(2,52)	1:B:13:C5L:O2	1:B:24:DG:H21	5	30.21
(2,47)	1:B:13:C5L:O2	1:B:24:DG:H21	5	30.21
(2,52)	1:B:13:C5L:O2	1:B:24:DG:H21	1	30.11
(2,47)	1:B:13:C5L:O2	1:B:24:DG:H21	1	30.11
(2,52)	1:B:13:C5L:O2	1:B:24:DG:H21	6	29.82
(2,47)	1:B:13:C5L:O2	1:B:24:DG:H21	6	29.82
(2,52)	1:B:13:C5L:O2	1:B:24:DG:H21	3	29.68
(2,47)	1:B:13:C5L:O2	1:B:24:DG:H21	3	29.68
(2,52)	1:B:13:C5L:O2	1:B:24:DG:H21	2	29.33
(2,47)	1:B:13:C5L:O2	1:B:24:DG:H21	2	29.33

10 Dihedral-angle violation analysis (i)

Dihedral angle analysis failed due to data error in the dihedral angle restraints, possibly missing target value