wwPDB NMR Structure Validation Summary Report (i)

Jun 6, 2023 – 08:36 pm BST

PDB ID	:	1QKG
BMRB ID	:	4409
Title	:	DNA DECAMER DUPLEX CONTAINING T-T DEWAR PHOTOPRODUCT
Authors	:	Lee, J.-H.; Bae, S.-H.; Choi, Y.-J.; Choi, B.-S.
Deposited on	:	1999-07-20

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

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

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

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

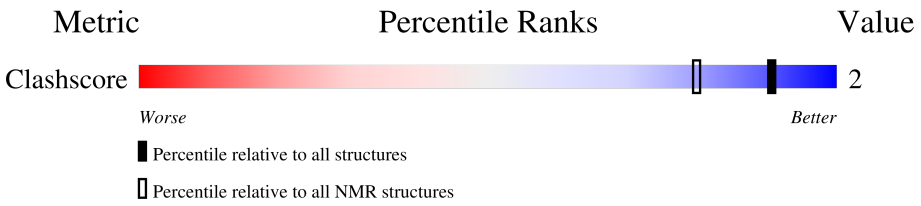
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 38%.

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	10	100%
2	B	10	80% 20%

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 633 atoms, of which 228 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*CP*GP*CP*AP*(HYD)TP*+TP*AP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	10	313	96	114	36	58	9	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*CP*GP*TP*GP*AP*TP*GP*CP*G)-3').

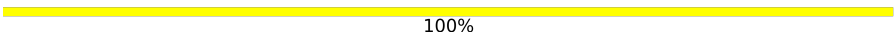
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	10	320	98	114	40	59	9	0

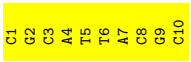
4 Residue-property plots (i)

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 (5'-D(*CP*GP*CP*AP*(HYD)TP*+TP*AP*CP*GP*C)-3')

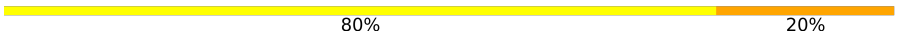
Chain A:





- Molecule 2: DNA (5'-D(*GP*CP*GP*TP*GP*AP*TP*GP*CP*G)-3')

Chain B:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING AND FULL RELAXATION MATRIX REFINEMENT*.

Of the 8 calculated structures, 1 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	**Classification**	**Version**
X-PLOR	refinement	3.1
X-PLOR	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	2
Total number of shifts	163
Number of shifts mapped to atoms	162
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	38%

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: TA3, 64T

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	1.64	2/177 (1.1%)	2.82	15/268 (5.6%)
2	B	1.72	2/231 (0.9%)	3.01	31/356 (8.7%)
All	All	1.69	4/408 (1.0%)	2.93	46/624 (7.4%)

All bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	17	DT	N1-C2	6.64	1.43	1.38
1	A	4	DA	N9-C4	6.57	1.41	1.37
2	B	14	DT	N1-C2	5.60	1.42	1.38
1	A	7	DA	N9-C4	5.18	1.41	1.37

5 of 46 angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	3	DC	P-O3'-C3'	17.91	141.19	119.70
2	B	17	DT	P-O3'-C3'	15.36	138.14	119.70
2	B	18	DG	P-O3'-C3'	14.69	137.33	119.70
2	B	11	DG	P-O3'-C3'	14.17	136.71	119.70
1	A	9	DG	P-O3'-C3'	13.75	136.21	119.70

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
2	B	206	114	114	1
All	All	405	228	228	1

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 2.

All clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
2:B:14:DT:H2"	2:B:15:DG:C8	0.48	2.43

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)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	64T	A	5	1	17,22,23	2.41	5 (29%)

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
1	TA3	A	6	1	18,21,22	3.66	7 (38%)

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

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
1	64T	A	5	1	24,33,36	1.49	2 (8%)
1	TA3	A	6	1	11,32,35	1.39	2 (18%)

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

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	64T	A	5	1	-	0,7,40,41	0,2,2,2
1	TA3	A	6	1	-	0,3,45,46	0,3,3,3

5 of 12 bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	6	TA3	C4-C5	12.52	1.37	1.54
1	A	5	64T	C2-N1	6.47	1.44	1.35
1	A	6	TA3	C6-N3	5.29	1.56	1.48
1	A	5	64T	C1'-N1	4.23	1.51	1.45
1	A	6	TA3	C1'-N1	4.03	1.50	1.45

All angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	5	64T	C1'-N1-C2	4.87	112.02	118.50
1	A	6	TA3	C2'-C1'-N1	3.46	110.91	115.59
1	A	5	64T	O5-C5-C5M	2.91	104.21	109.39
1	A	6	TA3	O4'-C1'-N1	2.70	111.85	108.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 38% for the well-defined parts and 38% 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	83
Number of shifts mapped to atoms	82
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	5	64T	H62	4.39	.	.

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 18%, i.e. 64 atoms were assigned a chemical shift out of a possible 363. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Sugar	40/216 (19%)	40/126 (32%)	0/90 (0%)	0/0 (—%)
Base	24/147 (16%)	24/93 (26%)	0/29 (0%)	0/25 (0%)
Overall	64/363 (18%)	64/219 (29%)	0/119 (0%)	0/25 (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

7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

7.2.1 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	80
Number of shifts mapped to atoms	80
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

7.2.2 Chemical shift referencing (i)

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

7.2.3 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 20%, i.e. 74 atoms were assigned a chemical shift out of a possible 363. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Sugar	50/216 (23%)	50/126 (40%)	0/90 (0%)	0/0 (—%)
Base	24/147 (16%)	24/93 (26%)	0/29 (0%)	0/25 (0%)
Overall	74/363 (20%)	74/219 (34%)	0/119 (0%)	0/25 (0%)

7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 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	375
Intra-residue (|i-j|=0)	212
Sequential (|i-j|=1)	75
Medium range (|i-j|>1 and |i-j|<5)	0
Long range (|i-j|≥5)	0
Inter-chain	88
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	18.8
Number of long range restraints per residue[1]	0.0

[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)	17.0	0.2
0.2-0.5 (Medium)	34.0	0.49
>0.5 (Large)	37.0	3.22

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)

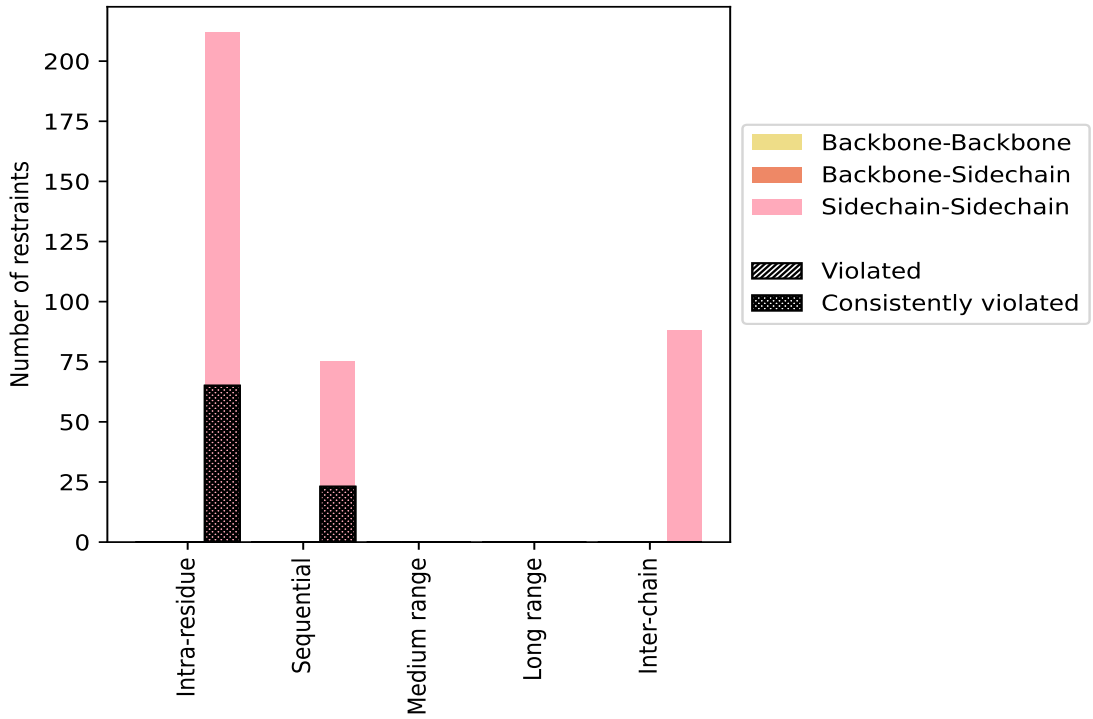
9.1 Summary of distance violations (i)

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)	212	56.5	65	30.7	17.3	65	30.7	17.3
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	212	56.5	65	30.7	17.3	65	30.7	17.3
Sequential (|i-j|=1)	75	20.0	23	30.7	6.1	23	30.7	6.1
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	75	20.0	23	30.7	6.1	23	30.7	6.1
Medium range (|i-j|>1 & |i-j|<5)	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
Long range (|i-j|≥5)	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
Inter-chain	88	23.5	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	88	23.5	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	375	100.0	88	23.5	23.5	88	23.5	23.5
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	375	100.0	88	23.5	23.5	88	23.5	23.5

[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

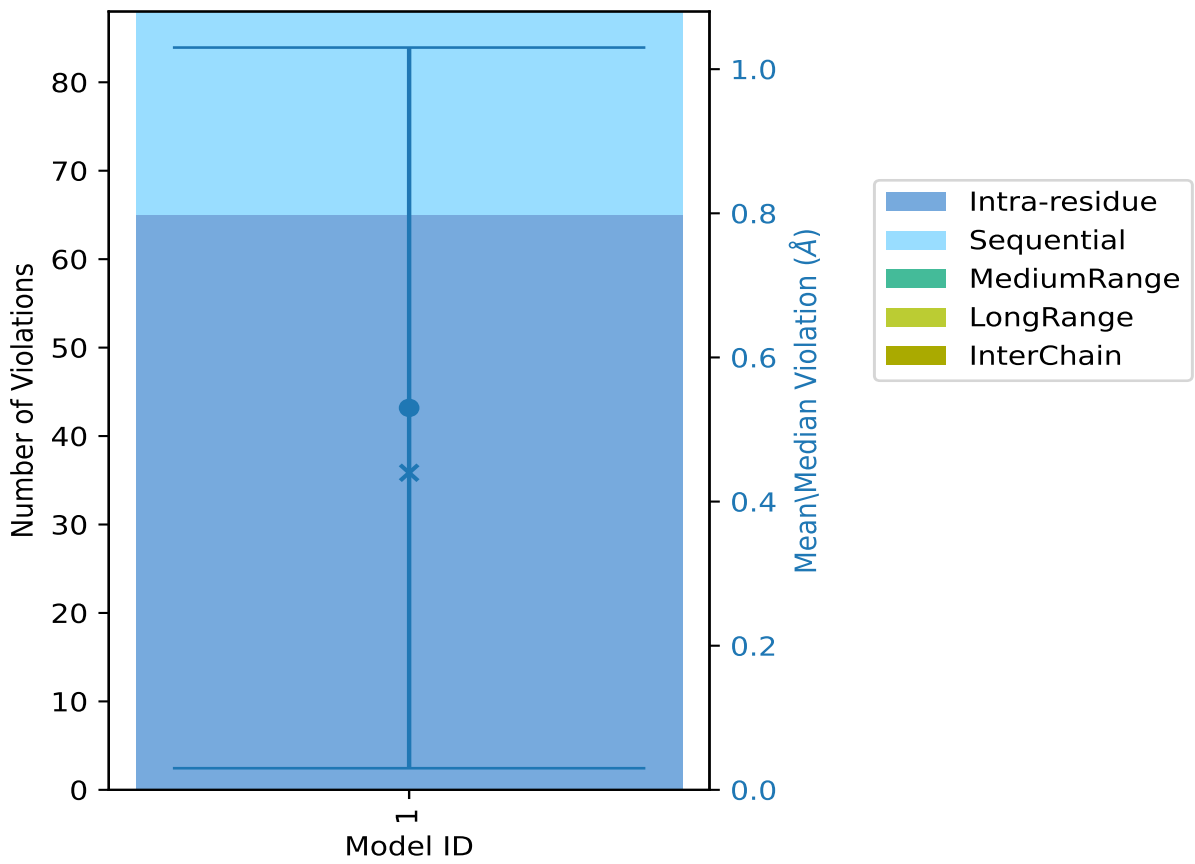
9.2 Distance violation statistics for each model (i)

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	65	23	0	0	0	88	0.53	3.22	0.5	0.44

[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

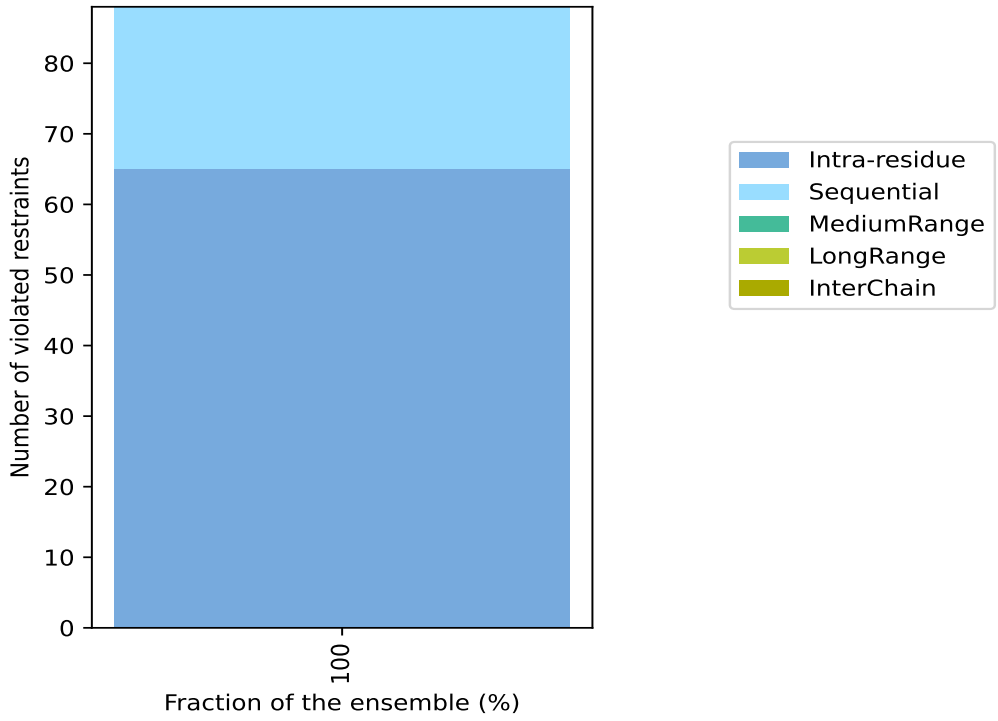
9.3 Distance violation statistics for the ensemble (i)

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, 287(IR:147, SQ:52, MR:0, LR:0, IC:88) 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]	%
65	23	0	0	0	88	1	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)



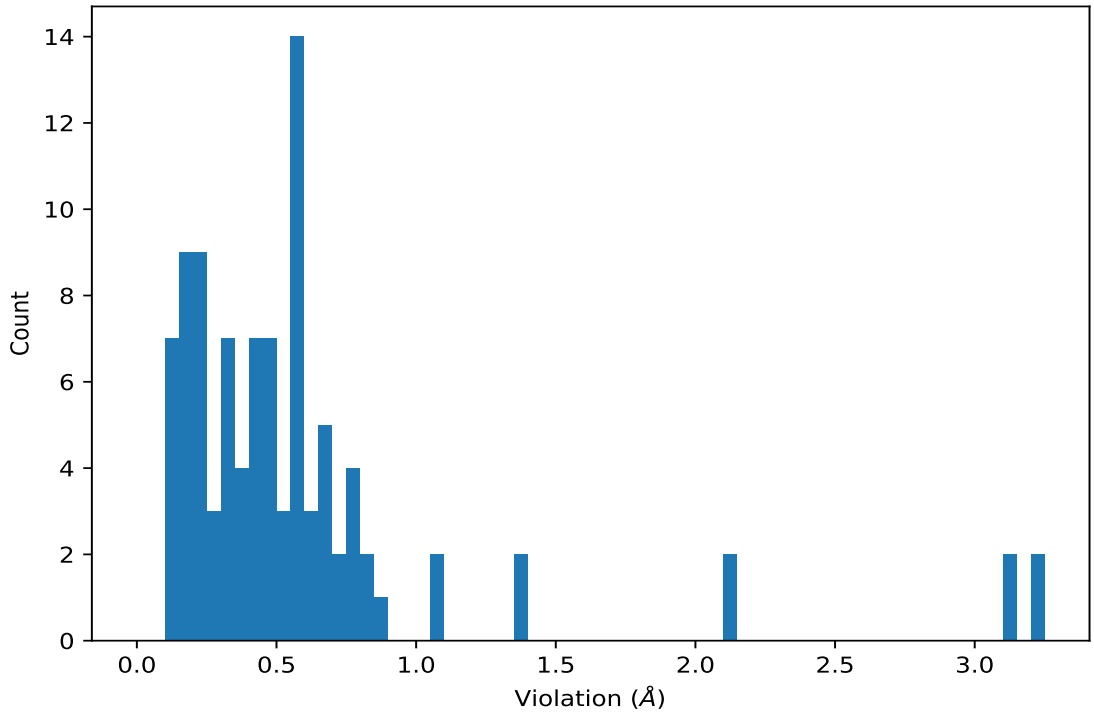
9.4 Most violated distance restraints in the ensemble (i)

No violations found

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)
(1,154)	2:B:13:DG:H8	2:B:14:DT:H5'	1	3.22
(1,154)	2:B:13:DG:H8	2:B:14:DT:H5"	1	3.22
(1,155)	2:B:16:DA:H8	2:B:17:DT:H5'	1	3.14
(1,155)	2:B:16:DA:H8	2:B:17:DT:H5"	1	3.14
(1,153)	1:A:4:DA:H8	1:A:5:64T:H5'	1	2.13
(1,153)	1:A:4:DA:H8	1:A:5:64T:H5"	1	2.13
(1,150)	1:A:6:TA3:H5'	1:A:5:64T:H61	1	1.35
(1,150)	1:A:6:TA3:H5"	1:A:5:64T:H61	1	1.35
(1,79)	2:B:16:DA:H2'	2:B:17:DT:H6	1	1.08
(1,113)	1:A:4:DA:H8	1:A:4:DA:H4'	1	1.06
(1,31)	1:A:3:DC:H1'	1:A:4:DA:H8	1	0.89
(1,72)	1:A:7:DA:H2'	1:A:8:DC:H6	1	0.83
(1,94)	1:A:7:DA:H8	1:A:7:DA:H3'	1	0.81
(1,95)	1:A:8:DC:H6	1:A:8:DC:H3'	1	0.79

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found