



wwPDB NMR Structure Validation Summary Report ⓘ

Jan 13, 2025 – 12:08 PM JST

PDB ID : 9KAD
BMRB ID : 36700
Title : Solution NMR structure of an RNA duplex containing pure CAG repeats
Authors : Guo, P.; Wan, L.
Deposited on : 2024-10-28

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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
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) : 2.40

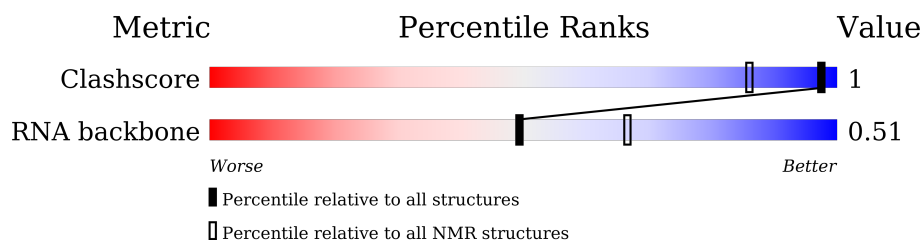
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

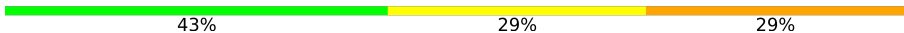

The overall completeness of chemical shifts assignment is 9%.

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	210492	14027
RNA backbone	6643	756

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	14	 43% 29% 29%
1	B	14	 43% 36% 21%

2 Ensemble composition and analysis ⓘ

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 912 atoms, of which 312 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called RNA (5'-R(*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*C)-3').

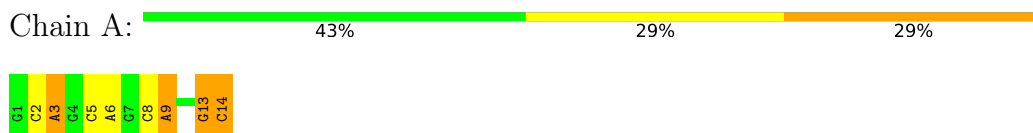
Mol	Chain	Residues	Atoms						Trace
1	A	14	Total	C	H	N	O	P	0
			456	135	156	60	92	13	
1	B	14	Total	C	H	N	O	P	0
			456	135	156	60	92	13	

4 Residue-property plots [i](#)

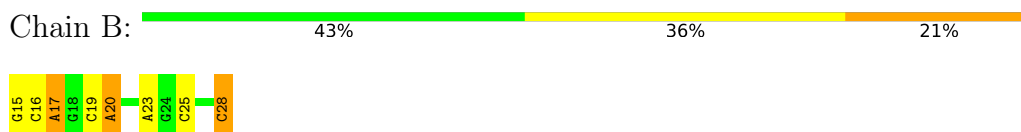
4.1 Average score per residue in the NMR ensemble

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(*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*C)-3')



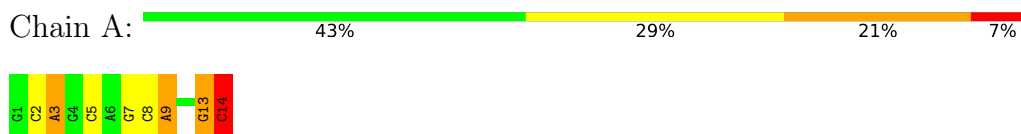
- Molecule 1: RNA (5'-R(*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*C)-3')



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(*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*C)-3')



- Molecule 1: RNA (5'-R(*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*CP*AP*GP*C)-3')



5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
Amber	structure calculation	

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	49
Number of shifts mapped to atoms	49
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	9%

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	1.21±0.02	2±0/336 (0.5± 0.1%)	1.90±0.05	8±3/523 (1.5± 0.5%)
1	B	1.19±0.02	1±0/336 (0.4± 0.1%)	1.95±0.04	9±2/523 (1.8± 0.4%)
All	All	1.20	28/6720 (0.4%)	1.93	174/10460 (1.7%)

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	1.8±0.6
1	B	0.0±0.0	2.6±1.0
All	All	0	44

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	3	A	N3-C4	6.36	1.38	1.34	8	10
1	A	6	A	N3-C4	5.80	1.38	1.34	10	5
1	B	17	A	N3-C4	5.79	1.38	1.34	4	10
1	A	9	A	N3-C4	5.48	1.38	1.34	6	1
1	B	23	A	N3-C4	5.05	1.37	1.34	4	2

5 of 44 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(°)	Ideal(°)	Models	
								Worst	Total
1	B	23	A	O4'-C1'-N9	9.59	115.87	108.20	4	4
1	B	28	C	O4'-C1'-N1	8.08	114.66	108.20	8	10
1	A	9	A	O4'-C1'-N9	7.99	114.59	108.20	2	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	14	C	O4'-C1'-N1	7.97	114.58	108.20	4	10
1	B	20	A	O4'-C1'-N9	7.93	114.55	108.20	6	6

There are no chirality outliers.

5 of 13 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	B	17	A	Sidechain	8
1	B	15	G	Sidechain	8
1	A	3	A	Sidechain	6
1	A	13	G	Sidechain	4
1	A	7	G	Sidechain	3

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	300	156	156	0±1
All	All	6000	3120	3120	5

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:13:G:C4	1:A:14:C:C5	0.42	3.06	2	2
1:A:8:C:H2'	1:A:9:A:C8	0.42	2.49	4	3

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](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	13/14 (93%)	3±1 (21±8%)	2±1 (12±8%)	0.49±0.07
1	B	13/14 (93%)	2±1 (16±4%)	1±0 (8±3%)	0.54±0.04
All	All	260/280 (93%)	48 (18%)	26 (10%)	0.51

The overall RNA backbone suiteness is 0.51.

5 of 11 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	28	C	10
1	A	14	C	9
1	B	20	A	7
1	A	6	A	7
1	A	13	G	5

5 of 6 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	20	A	7
1	A	9	A	6
1	A	13	G	5
1	A	6	A	5
1	B	23	A	2

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 oligosaccharides 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 9% for the well-defined parts and 9% 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	49
Number of shifts mapped to atoms	49
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 9%, i.e. 49 atoms were assigned a chemical shift out of a possible 542. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Sugar	14/308 (5%)	14/168 (8%)	0/140 (0%)	0/0 (—%)
Base	35/234 (15%)	35/150 (23%)	0/46 (0%)	0/38 (0%)
Overall	49/542 (9%)	49/318 (15%)	0/186 (0%)	0/38 (0%)

7.1.4 Statistically unusual chemical shifts [i](#)

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 [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	190
Intra-residue ($ i-j =0$)	40
Sequential ($ i-j =1$)	100
Medium range ($ i-j >1$ and $ i-j <5$)	0
Long range ($ i-j \geq 5$)	0
Inter-chain	26
Hydrogen bond restraints	24
Disulfide bond restraints	0
Total dihedral-angle restraints	28
Number of unmapped restraints	0
Number of restraints per residue	7.8
Number of long range restraints per residue ¹	0.0

¹Long 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. There are no distance violations

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](#)

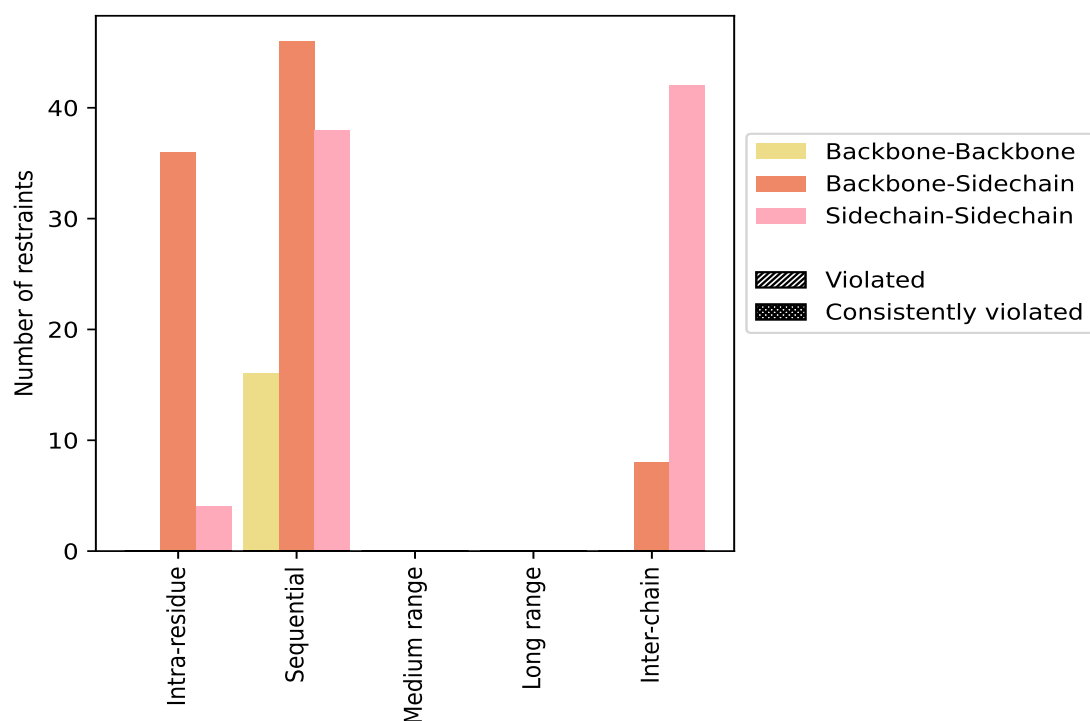
9.1 Summary of distance violations [i](#)

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.

Restrains type	Count	% ¹	Violated ³			Consistently Violated ⁴		
			Count	% ²	% ¹	Count	% ²	% ¹
Intra-residue (i-j =0)	40	21.1	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	36	18.9	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	4	2.1	0	0.0	0.0	0	0.0	0.0
Sequential (i-j =1)	100	52.6	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	16	8.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	46	24.2	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	38	20.0	0	0.0	0.0	0	0.0	0.0
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	26	13.7	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	8	4.2	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	18	9.5	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	24	12.6	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	190	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	16	8.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	90	47.4	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	84	44.2	0	0.0	0.0	0	0.0	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](#)

No violations found

9.3 Distance violation statistics for the ensemble [i](#)

No violations found

9.4 Most violated distance restraints in the ensemble [i](#)

No violations found

9.5 All violated distance restraints [i](#)

No violations found

10 Dihedral-angle violation analysis [i](#)

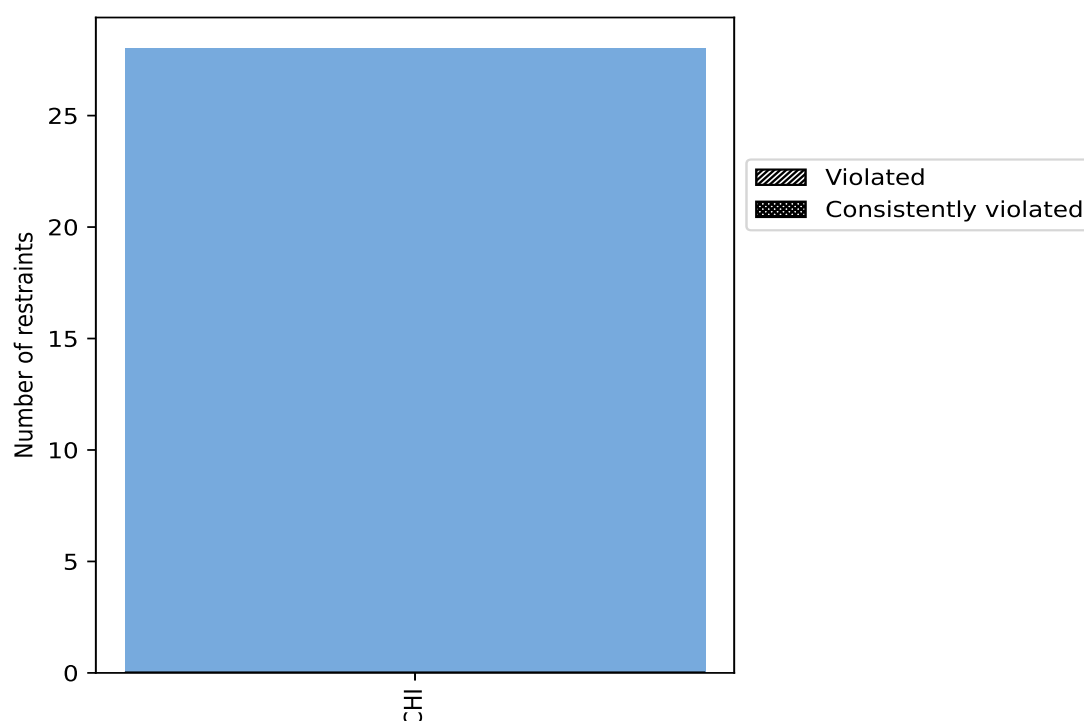
10.1 Summary of dihedral-angle violations [i](#)

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	28	100.0	0	0.0	0.0	0	0.0	0.0
Total	28	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