WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

Jul 29, 2024 – 10:42 AM EDT

PDB ID	:	9CPI
BMRB ID	:	31190
Title	:	Structures of small molecules bound to RNA repeat expansions that cause Huntington's disease-like 2 and myotonic dystrophy type 1
Authors	:	Chen, J.L.; Taghavi, A.; Disney, M.D.; Fountain, M.A.; Childs-Disney, J.L.
Deposited on	:	2024-07-18

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

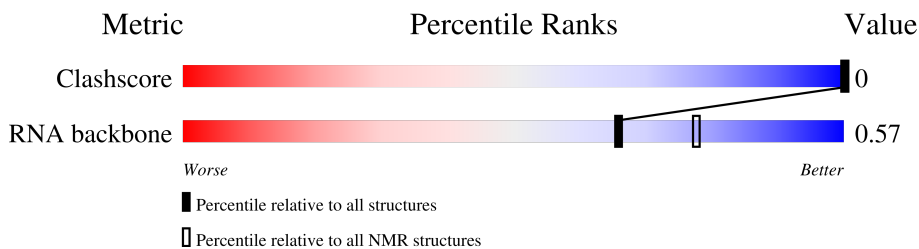|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 14%.

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	Percentile Ranks	Value
Clashscore		0
RNA backbone		0.57

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
RNA backbone	4643	676

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	13	62% 38%
1	B	13	54% 46%

2 Ensemble composition and analysis ⓘ

This entry contains 20 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 ⓘ

There are 2 unique types of molecules in this entry. The entry contains 885 atoms, of which 311 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	A	13	Total	C	H	N	O	P	0
			415	123	142	48	90	12	
1	B	13	Total	C	H	N	O	P	0
			415	123	142	48	90	12	

- Molecule 2 is 4-[1-(2-aminoethyl)-4,5-dihydro-1H-imidazol-2-yl]-N-[4-(4,5-dihydro-1H-imidazol-2-yl)phenyl]benzamide (three-letter code: A1AZL) (formula: $C_{21}H_{24}N_6O$) (labeled as "Ligand of Interest" by depositor).
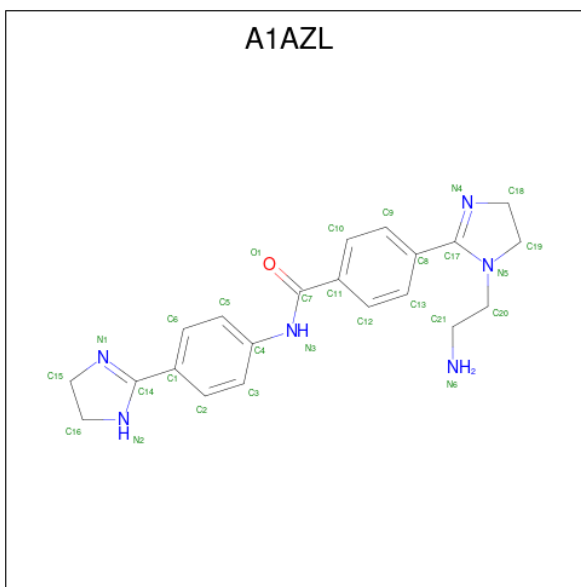
Mol	Chain	Residues	Atoms				
2	B	1	Total	C	H	N	O
			55	21	27	6	1

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

Chain A:

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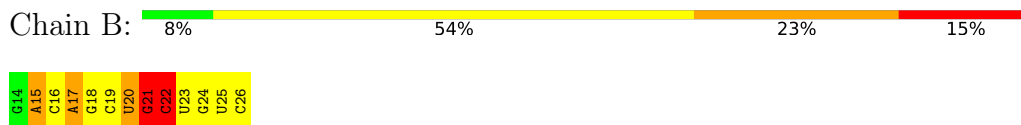
Chain B:

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

Chain A:

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

Chain B:

5 Refinement protocol and experimental data overview (i)

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 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	78
Number of shifts mapped to atoms	78
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	14%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: A1AZL

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	1.52±0.04	0±0/304 (0.0± 0.1%)	2.27±0.07	22±3/472 (4.6± 0.6%)
1	B	1.53±0.03	0±0/304 (0.1± 0.1%)	2.30±0.06	23±2/472 (4.8± 0.4%)
All	All	1.52	7/12160 (0.1%)	2.28	889/18880 (4.7%)

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	5.0±0.9
1	B	0.0±0.0	5.7±1.9
All	All	0	212

5 of 6 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	B	15	A	N3-C4	5.99	1.38	1.34	12	2
1	B	26	C	C4-N4	-5.81	1.28	1.33	11	1
1	B	15	A	N7-C5	5.14	1.42	1.39	14	1
1	A	13	C	C4-N4	-5.14	1.29	1.33	8	1
1	B	24	G	C2-N2	-5.12	1.29	1.34	20	1

5 of 146 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	15	A	N1-C6-N6	-11.35	111.79	118.60	18	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models Worst	Models Total
1	A	4	A	N1-C6-N6	-10.95	112.03	118.60	19	20
1	B	21	G	O4'-C1'-N9	10.12	116.30	108.20	13	14
1	B	17	A	N1-C6-N6	-9.93	112.64	118.60	5	18
1	B	16	C	N3-C2-O2	-9.63	115.16	121.90	1	20

There are no chirality outliers.

5 of 22 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	A	5	G	Sidechain	19
1	A	9	C	Sidechain	19
1	A	8	G	Sidechain	17
1	B	17	A	Sidechain	15
1	B	22	C	Sidechain	15

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
All	All	11480	6220	5674	-

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 -.

There are no clashes.

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

There are no protein molecules in this entry.

6.3.2 Protein sidechains ⓘ

There are no protein molecules in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	13/13 (100%)	3±1 (21±8%)	1±0 (7±3%)	0.53±0.06
1	B	12/13 (92%)	2±1 (12±7%)	0±1 (3±5%)	0.61±0.04
All	All	495/520 (95%)	85 (17%)	26 (5%)	0.57

The overall RNA backbone suiteness is 0.57.

5 of 9 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	21	G	17
1	A	2	A	15
1	A	11	G	15
1	A	8	G	9
1	A	9	C	9

5 of 7 unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	1	G	15
1	B	20	U	4
1	A	6	C	2
1	B	19	C	2
1	A	7	U	1

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

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

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

1 ligand is 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
2	A1AZL	B	101	-	31,31,31	1.11±0.05	1±0 (3±1%)

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
2	A1AZL	B	101	-	39,42,42	1.10±0.13	4±1 (10±3%)

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
2	A1AZL	B	101	-	-	0±0,19,36,36	0±0,4,4,4

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
2	B	101	A1AZL	C17-N4	3.43	1.39	1.30	11	20
2	B	101	A1AZL	C7-N3	2.11	1.41	1.35	5	1
2	B	101	A1AZL	C2-C1	2.02	1.42	1.39	3	1
2	B	101	A1AZL	C10-C11	2.02	1.42	1.39	18	1

5 of 10 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	B	101	A1AZL	C20-N5-C19	4.09	114.02	121.80	10	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models Worst	Models Total
2	B	101	A1AZL	C18-C19-N5	4.02	105.35	101.05	11	17
2	B	101	A1AZL	C8-C17-N4	3.73	115.57	123.18	3	18
2	B	101	A1AZL	C15-C16-N2	3.55	105.81	101.69	13	15
2	B	101	A1AZL	C18-N4-C17	3.13	112.54	106.50	4	12

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand A1AZL B 101

Bond lengths

Bond angles

Torsions

Rings
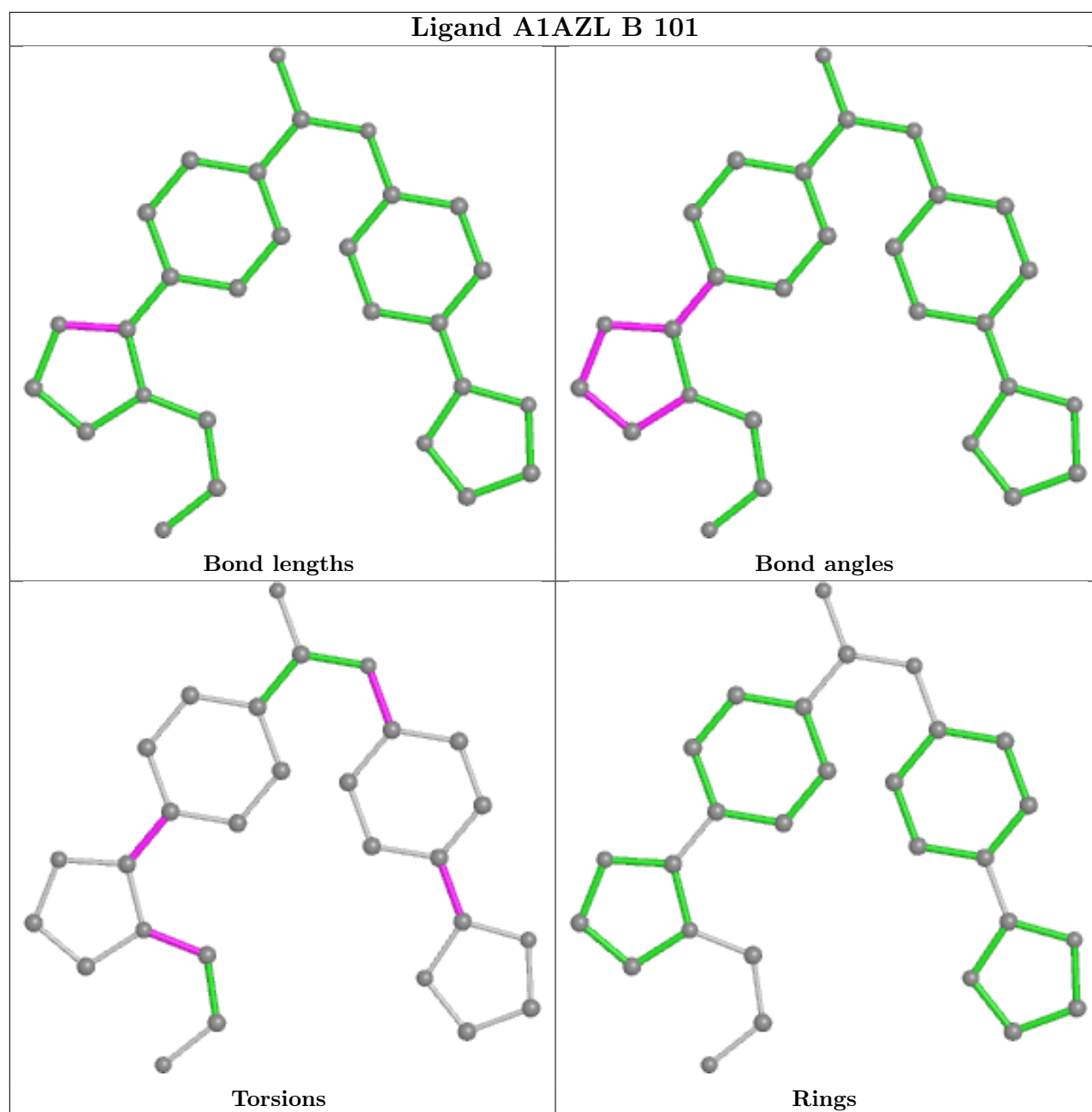
6.7 Other polymers ⓘ

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 14% for the well-defined parts and 14% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *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	78
Number of shifts mapped to atoms	78
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 14%, i.e. 67 atoms were assigned a chemical shift out of a possible 490. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Sugar	36/286 (13%)	36/156 (23%)	0/130 (0%)	0/0 (—%)
Base	31/204 (15%)	31/126 (25%)	0/44 (0%)	0/34 (0%)
Overall	67/490 (14%)	67/282 (24%)	0/174 (0%)	0/34 (0%)

7.1.4 Statistically unusual chemical shifts ⓘ

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 ⓘ

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	289
Intra-residue ($|i-j|$=0)	122
Sequential ($|i-j|$=1)	104
Medium range ($|i-j|$>1 and $|i-j|$<5)	0
Long range ($|i-j|$≥5)	0
Inter-chain	37
Hydrogen bond restraints	26
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	10.7
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 ⓘ

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.There are no distance violations

8.2.2 Average number of dihedral-angle violations per model ⓘ

Dihedral-angle violations less than 1° are not included in the calculation.There are no dihedral-angle violations
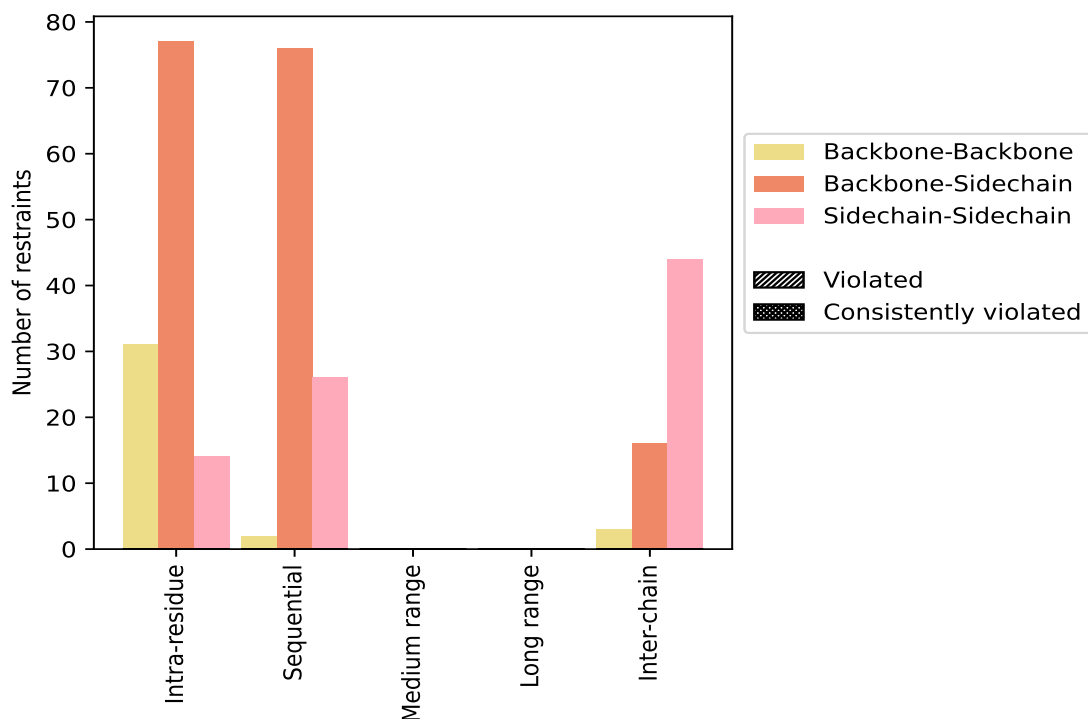
9   Distance violation analysis ⓘ

9.1   Summary of distance violations ⓘ

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)	122	42.2	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	31	10.7	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	77	26.6	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	14	4.8	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	104	36.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	2	0.7	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	76	26.3	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	26	9.0	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	37	12.8	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	3	1.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	16	5.5	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	18	6.2	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	26	9.0	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	289	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	36	12.5	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	169	58.5	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	84	29.1	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 ⓘ

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 ⓘ

No violations found

9.3 Distance violation statistics for the ensemble ⓘ

No violations found

9.4 Most violated distance restraints in the ensemble ⓘ

No violations found

9.5 All violated distance restraints ⓘ

No violations found

10 Dihedral-angle violation analysis ⓘ

No dihedral-angle restraints found