W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB NMR Structure Validation Summary Report (i)

Jun 3, 2023 – 06:01 AM EDT

PDB ID	:	5UE2
Title	:	proMMP-7 with heparin octasaccharide bridging between domains
Authors	:	Fulcher, Y.G.; Prior, S.H.; Linhardt, R.J.; Van Doren, S.R.
Deposited on	:	2016-12-29

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	:	1.8.5 (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 ⓘ

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

The overall completeness of chemical shifts assignment is 86%.

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.
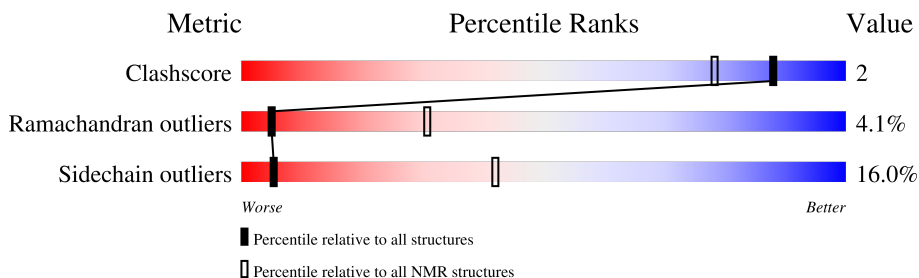
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	247	63% 15% • 21%
2	B	8	100%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	B	IDS	6	1	-

2 Ensemble composition and analysis ⓘ

This entry contains 16 models. Model 1 is the overall representative, medoid model (most similar to other models).

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:15-A:22, A:34-A:71, A:83-A:216, A:224-A:239 (196)	0.71	1

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 3 single-model clusters were found.

Cluster number	Models
1	1, 3, 5, 8, 12, 13
2	2, 4, 7, 10, 15
3	14, 16
Single-model clusters	6; 9; 11

W O R L D W I D E
PDB
PROTEIN DATA BANK

3 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 4032 atoms, of which 1945 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Matrilysin.

Mol	Chain	Residues	Atoms						Trace
1	A	247	Total	C	H	N	O	S	0
			3828	1234	1885	338	362	9	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	195	ALA	GLU	engineered mutation	UNP P09237

- Molecule 2 is an oligosaccharide called 2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose.
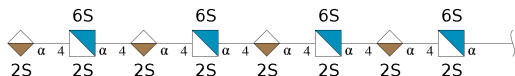
Mol	Chain	Residues	Atoms						Trace
2	B	8	Total	C	H	N	O	S	0
			200	48	60	4	76	12	

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
3	A	2	Total	Ca
			2	2

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

| Mol	Chain	Residues	Atoms	
4	A	2	Total	Zn
			2	2

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: Matrilysin
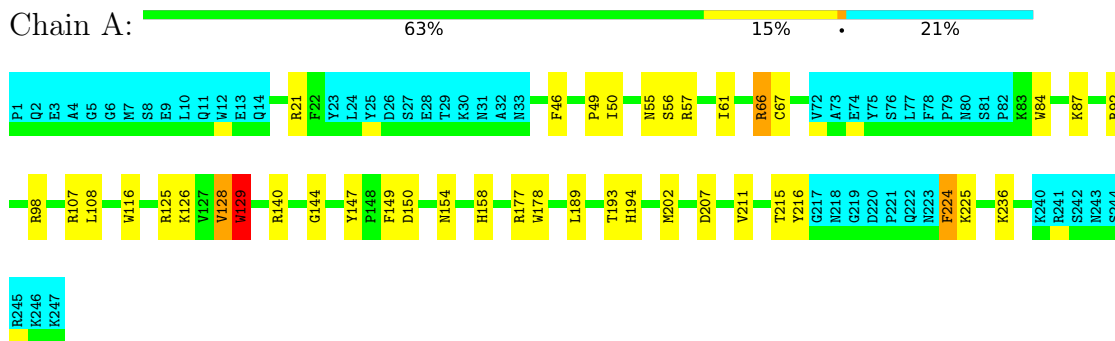
• Molecule 2: 2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose
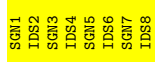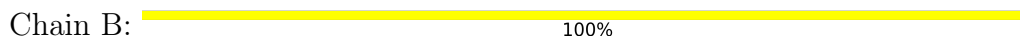
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

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

• Molecule 1: Matrilysin

P1 Q2 E3 A4 G5 G6 M7 S8 E9 L10 Q11 W12 E13 Q14 R21 F22 Y23 L24 Y25 D26 S27 E28 T29 K30 N31 A32 N33 L39 Q43 F46 I50 N55 S56 R57 K64 P65 R66 C67 V72 A73 E74 Y75 S76 L77 F78 P79 N80 S81 P82 K83 W84

K87 R92 R98 R107 L108 W116 E119 R125 K126 V127 V128 W129 R140 Y147 P148 F149 D150 N154 T164 R177 W178 T179 D180 L184 L189 H194 H198 V211 M212 Y213 P214 T215 Y216 G217 N218 G219 D220 P221 Q222 N223 F224 K225 D229

K236 K240 R241 S242 N243 S244 R245 K246 K247

● Molecule 2: 2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-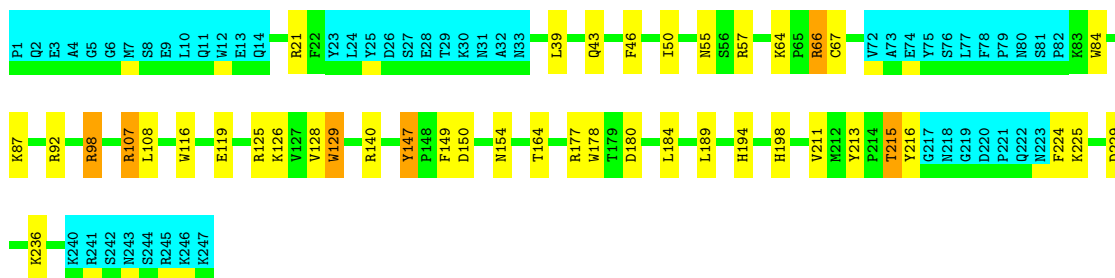deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose

Chain B: 100%

SGN1 IDS2 SGN3 IDS4 SGN5 IDS6 SGN7 IDS8

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *torsion angle dynamics.*

Of the 120 calculated structures, 16 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
CYANA	refinement	2.1
SYBYL-X	refinement	2.1.1
CYANA	structure calculation	2.1

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	2746
Number of shifts mapped to atoms	2746
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CA, ZN, SGN, IDS

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	0.99±0.01	0±0/1582 (0.0± 0.0%)	1.17±0.02	13±1/2141 (0.6± 0.1%)
All	All	0.99	0/25312 (0.0%)	1.17	204/34256 (0.6%)

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	1.2±1.2
All	All	0	19

There are no bond-length outliers.

5 of 23 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	107	ARG	NE-CZ-NH1	8.23	124.41	120.30	6	13
1	A	140	ARG	NE-CZ-NH1	7.97	124.28	120.30	14	15
1	A	21	ARG	NE-CZ-NH1	7.69	124.14	120.30	14	14
1	A	125	ARG	NE-CZ-NH1	7.60	124.10	120.30	9	15
1	A	66	ARG	NE-CZ-NH1	7.22	123.91	120.30	10	10

There are no chirality outliers.

5 of 11 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	216	TYR	Sidechain	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	46	PHE	Sidechain	3
1	A	224	PHE	Sidechain	2
1	A	147	TYR	Sidechain	2
1	A	190	TYR	Sidechain	1

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	1538	1512	1511	6±2
2	B	140	60	45	0±0
All	All	26912	25152	24897	96

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.

5 of 53 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:46:PHE:CD1	1:A:61:ILE:HD12	0.62	2.29	14	3
1:A:158:HIS:NE2	1:A:171:HIS:ND1	0.58	2.46	16	8
1:A:46:PHE:CD2	1:A:61:ILE:HD12	0.57	2.34	11	2
1:A:46:PHE:CE1	1:A:61:ILE:HD12	0.56	2.36	15	2
1:A:194:HIS:CD2	1:A:198:HIS:NE2	0.55	2.73	14	5

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	196/247 (79%)	170±3 (87±2%)	18±3 (9±1%)	8±2 (4±1%)	5	31

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	3136/3952 (79%)	2716 (87%)	291 (9%)	129 (4%)	5 31

5 of 23 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	55	ASN	16
1	A	154	ASN	16
1	A	147	TYR	15
1	A	144	GLY	11
1	A	66	ARG	10

6.3.2   Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	161/205 (79%)	135±5 (84±3%)	26±5 (16±3%)	5 42
All	All	2576/3280 (79%)	2165 (84%)	411 (16%)	5 42

5 of 86 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	108	LEU	16
1	A	128	VAL	16
1	A	129	TRP	16
1	A	236	LYS	16
1	A	126	LYS	15

6.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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 ⓘ

8 monosaccharides 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
2	SGN	B	1	2	18,19,20	3.74±0.05	3±0 (18±2%)
2	IDS	B	2	2	16,16,17	1.25±0.02	1±0 (6±1%)
2	SGN	B	3	2	18,19,20	3.77±0.05	5±0 (26±2%)
2	IDS	B	4	2	16,16,17	1.26±0.02	1±0 (6±0%)
2	SGN	B	5	2	18,19,20	3.77±0.07	5±0 (26±2%)
2	IDS	B	6	2	16,16,17	1.27±0.01	1±0 (6±0%)
2	SGN	B	7	2	18,19,20	3.76±0.05	4±0 (25±2%)
2	IDS	B	8	2	16,16,17	1.20±0.02	1±0 (6±0%)

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	SGN	B	1	2	22,29,31	1.42±0.15	3±1 (13±4%)
2	IDS	B	2	2	17,24,26	1.12±0.14	1±1 (5±4%)
2	SGN	B	3	2	22,29,31	1.63±0.20	4±2 (20±7%)
2	IDS	B	4	2	17,24,26	1.08±0.19	0±1 (2±4%)
2	SGN	B	5	2	22,29,31	1.60±0.17	4±1 (20±6%)
2	IDS	B	6	2	17,24,26	1.04±0.13	1±1 (3±3%)
2	SGN	B	7	2	22,29,31	1.55±0.21	4±1 (16±5%)
2	IDS	B	8	2	17,24,26	1.06±0.19	1±1 (4±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	SGN	B	1	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	2	2	-	0±0,9,26,29	0±0,1,1,1
2	SGN	B	3	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	4	2	-	0±0,9,26,29	0±0,1,1,1
2	SGN	B	5	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	6	2	-	0±0,9,26,29	0±0,1,1,1
2	SGN	B	7	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	8	2	-	0±0,9,26,29	0±0,1,1,1

5 of 25 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	Models Total
2	B	5	SGN	S1-N2	15.30	1.80	1.59	5	16
2	B	3	SGN	S1-N2	15.22	1.80	1.59	5	16
2	B	7	SGN	S1-N2	15.13	1.79	1.59	2	16
2	B	1	SGN	S1-N2	15.12	1.79	1.59	2	16
2	B	7	SGN	O2S-S1	3.15	1.45	1.42	11	16

5 of 62 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	Models Total
2	B	3	SGN	O6-C6-C5	4.92	116.79	107.62	11	4
2	B	8	IDS	O2-C2-C3	4.46	113.19	106.95	13	7
2	B	5	SGN	O3-C3-C2	4.27	100.62	109.47	11	3
2	B	7	SGN	O1S-S1-O2S	4.14	110.39	120.16	14	9
2	B	4	IDS	O2-C2-C3	4.13	112.72	106.95	11	1

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	B	6	IDS	C1	1

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 oligosaccharide.

Oligosaccharide Chain B

Bond lengths

Bond angles

Torsions

Rings
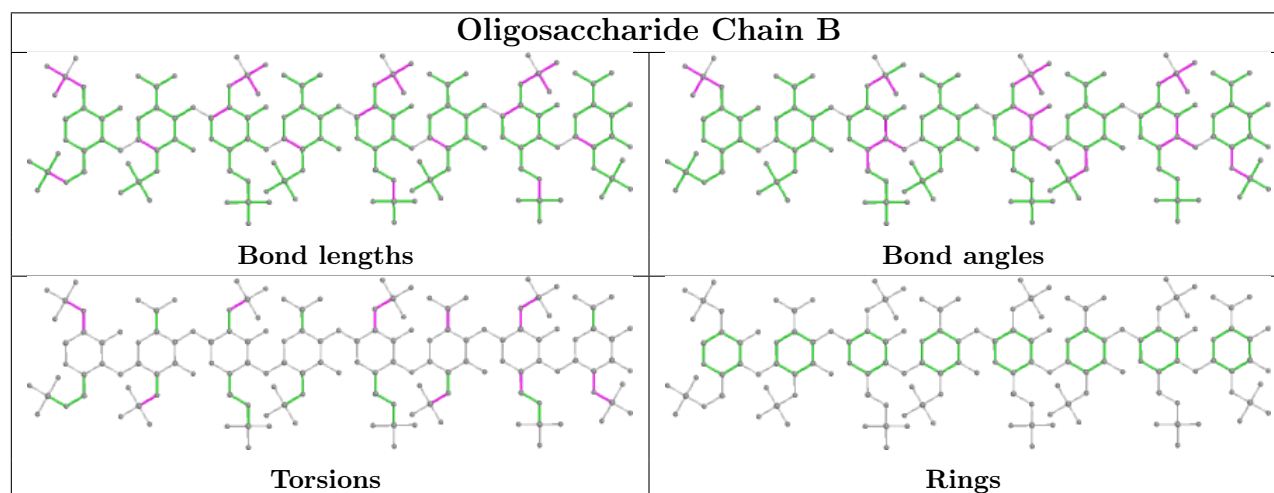
6.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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 86% for the well-defined parts and 83% 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	2746
Number of shifts mapped to atoms	2746
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	15

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	236	-0.28 ± 0.20	None needed (< 0.5 ppm)
$^{13}C_\beta$	212	0.28 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	232	0.03 ± 0.15	None needed (< 0.5 ppm)
^{15}N	220	-0.11 ± 0.26	None needed (< 0.5 ppm)

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 86%, i.e. 2286 atoms were assigned a chemical shift out of a possible 2646. 0 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	952/981 (97%)	388/403 (96%)	383/392 (98%)	181/186 (97%)
Sidechain	1166/1389 (84%)	789/908 (87%)	367/431 (85%)	10/50 (20%)

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

|	Total	^{1}H	^{13}C	^{15}N
Aromatic	168/276 (61%)	87/138 (63%)	77/126 (61%)	4/12 (33%)
Overall	2286/2646 (86%)	1264/1449 (87%)	827/949 (87%)	195/248 (79%)

7.1.4   Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	107	ARG	NE	116.85	76.53 − 92.65	20.0
1	A	57	ARG	NE	115.24	76.53 − 92.65	19.0
1	A	92	ARG	NE	114.04	76.53 − 92.65	18.3
1	A	66	ARG	NE	112.31	76.53 − 92.65	17.2
1	A	162	PRO	HB2	-0.32	0.37 − 3.78	-7.0
1	A	60	GLU	HB2	0.60	1.00 − 3.05	-6.9
1	A	247	LYS	CE	36.10	37.57 − 46.21	-6.7
1	A	60	GLU	HB3	0.64	0.95 − 3.05	-6.5
1	A	60	GLU	CG	43.68	30.20 − 42.01	6.4
1	A	65	PRO	HA	2.44	2.78 − 6.00	-6.1
1	A	3	GLU	CG	43.11	30.20 − 42.01	5.9
1	A	62	MET	HG2	0.38	0.65 − 4.19	-5.8
1	A	242	SER	N	135.65	99.14 − 133.45	5.6
1	A	211	VAL	H	11.59	4.98 − 11.56	5.0
1	A	153	GLY	H	11.43	5.23 − 11.42	5.0

7.1.5   Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:
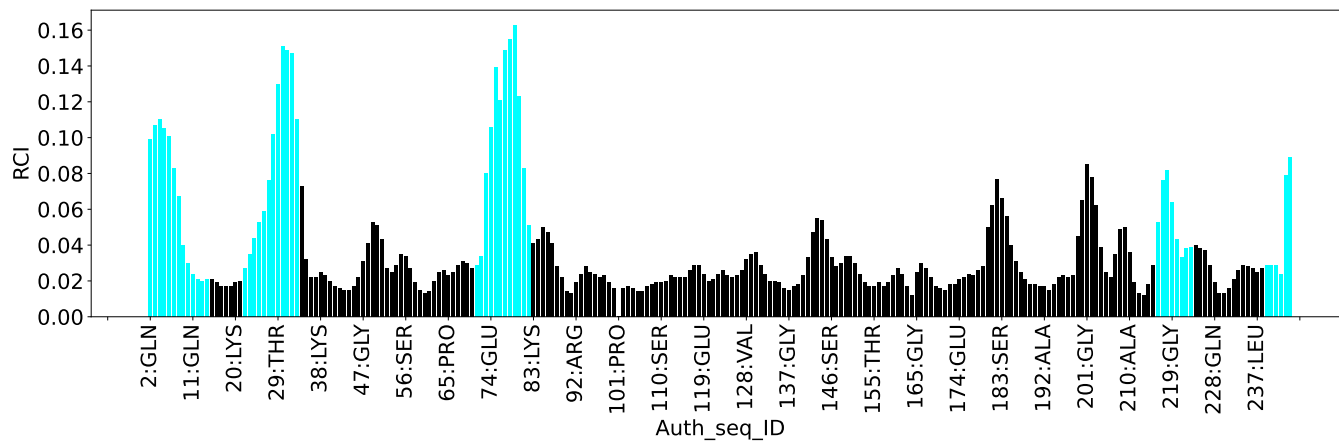
8   NMR restraints analysis ⓘ

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	2224
Intra-residue ($\|i\text{-}j\|$=0)	427
Sequential ($\|i\text{-}j\|$=1)	543
Medium range ($\|i\text{-}j\|$>1 and $\|i\text{-}j\|$<5)	329
Long range ($\|i\text{-}j\|$≥5)	691
Inter-chain	220
Hydrogen bond restraints	13
Disulfide bond restraints	1
Total dihedral-angle restraints	0
Number of unmapped restraints	220
Number of restraints per residue	9.0
Number of long range restraints per residue[1]	2.9

[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)	57.3	0.2
0.2-0.5 (Medium)	102.5	0.5
>0.5 (Large)	52.4	2.82

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

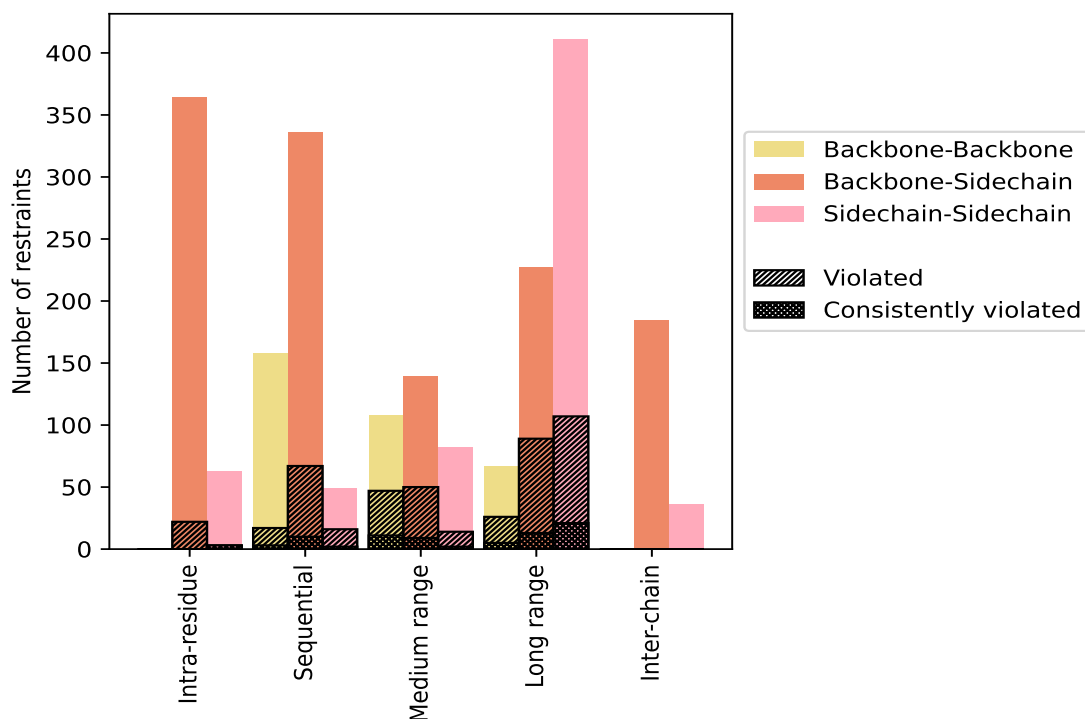# 9 Distance violation analysis ⓘ

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)	427	19.2	25	5.9	1.1	3	0.7	0.1
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	364	16.4	22	6.0	1.0	0	0.0	0.0
Sidechain-Sidechain	63	2.8	3	4.8	0.1	3	4.8	0.1
Sequential (|i-j|=1)	543	24.4	100	18.4	4.5	15	2.8	0.7
Backbone-Backbone	158	7.1	17	10.8	0.8	3	1.9	0.1
Backbone-Sidechain	336	15.1	67	19.9	3.0	10	3.0	0.4
Sidechain-Sidechain	49	2.2	16	32.7	0.7	2	4.1	0.1
Medium range (|i-j|>1 & |i-j|<5)	329	14.8	111	33.7	5.0	22	6.7	1.0
Backbone-Backbone	108	4.9	47	43.5	2.1	11	10.2	0.5
Backbone-Sidechain	139	6.2	50	36.0	2.2	9	6.5	0.4
Sidechain-Sidechain	82	3.7	14	17.1	0.6	2	2.4	0.1
Long range (|i-j|≥5)	691	31.1	218	31.5	9.8	38	5.5	1.7
Backbone-Backbone	55	2.5	23	41.8	1.0	4	7.3	0.2
Backbone-Sidechain	226	10.2	88	38.9	4.0	13	5.8	0.6
Sidechain-Sidechain	410	18.4	107	26.1	4.8	21	5.1	0.9
Inter-chain	220	9.9	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	184	8.3	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	36	1.6	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	13	0.6	4	30.8	0.2	1	7.7	0.0
Disulfide bond	1	0.0	0	0.0	0.0	0	0.0	0.0
Total	2224	100.0	458	20.6	20.6	79	3.6	3.6
Backbone-Backbone	333	15.0	90	27.0	4.0	19	5.7	0.9
Backbone-Sidechain	1250	56.2	228	18.2	10.3	32	2.6	1.4
Sidechain-Sidechain	641	28.8	140	21.8	6.3	28	4.4	1.3

[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	7	54	50	98	0	209	0.37	1.3	0.22	0.31
2	10	46	56	93	0	205	0.38	1.33	0.23	0.33
3	9	53	59	99	0	220	0.38	1.17	0.23	0.34
4	11	41	51	93	0	196	0.38	1.36	0.24	0.33
5	8	45	58	98	0	209	0.38	1.24	0.21	0.34
6	12	50	52	114	0	228	0.44	2.82	0.33	0.34
7	10	47	56	101	0	214	0.38	1.28	0.23	0.34
8	10	47	48	110	0	215	0.39	1.26	0.23	0.35
9	8	46	56	109	0	219	0.38	1.21	0.24	0.3
10	10	43	49	106	0	208	0.39	1.42	0.24	0.33
11	10	44	58	100	0	212	0.4	1.39	0.24	0.35

Continued on next page...

Continued from previous page...

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				
12	11	49	56	110	0	226	0.36	1.31	0.23	0.3
13	11	49	55	103	0	218	0.36	1.24	0.23	0.3
14	9	44	62	97	0	212	0.39	1.26	0.23	0.35
15	11	47	52	97	0	207	0.39	1.25	0.24	0.34
16	8	43	50	97	0	198	0.37	1.44	0.23	0.34

[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 ⓘ
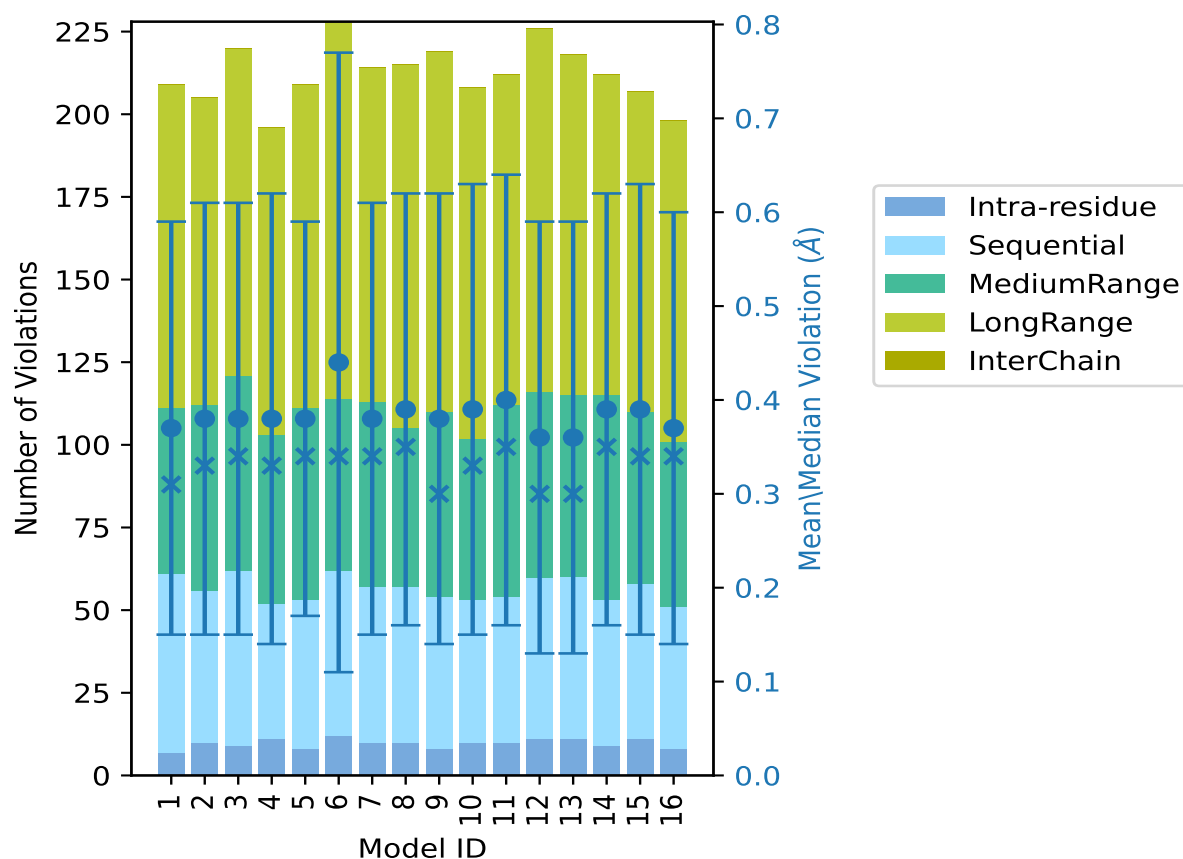
The mean(dot),median(x) and the standard deviation are shown in blue with respect to the y axis on the right

9.3   Distance violation statistics for the ensemble ⓘ

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, 1756(IR:402, SQ:443, MR:218, LR:473, IC:220) 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]	%
3	16	21	47	0	87	1	6.2
3	11	12	23	0	49	2	12.5
3	8	8	14	0	33	3	18.8
4	4	3	11	0	22	4	25.0
2	6	5	10	0	23	5	31.2
1	5	4	7	0	17	6	37.5
1	3	3	7	0	14	7	43.8
1	7	6	11	0	25	8	50.0
0	3	4	10	0	17	9	56.2
3	4	3	5	0	15	10	62.5
0	5	7	9	0	21	11	68.8
1	6	4	8	0	19	12	75.0
0	2	4	6	0	12	13	81.2
0	2	1	4	0	7	14	87.5
0	3	4	8	0	15	15	93.8
3	15	22	38	0	78	16	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 ⓘ
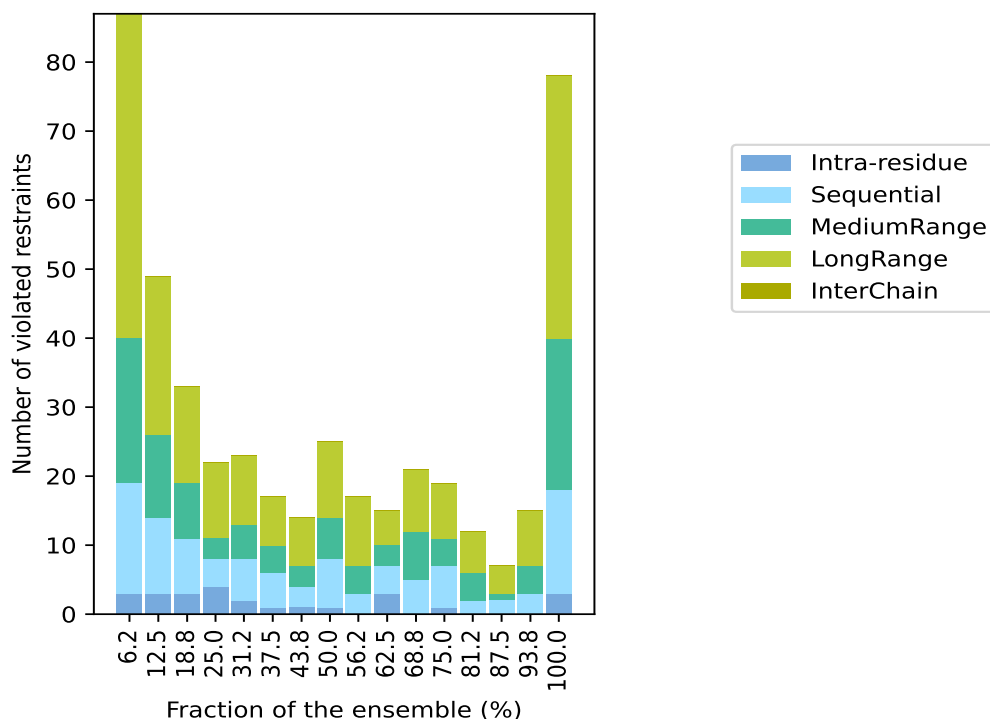
9.4 Most violated distance restraints in the ensemble ⓘ

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
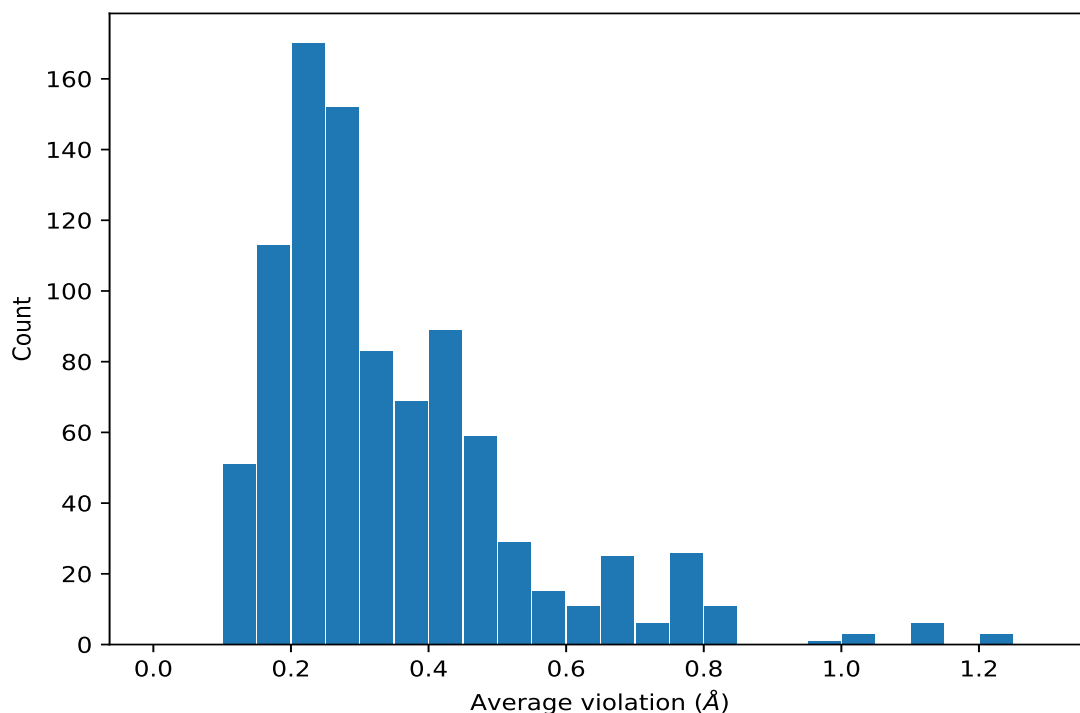
9.4.2 Table: Most violated distance restraints ⓘ

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)
(7,401)	1:A:156:LEU:HD21	1:A:186:ILE:H	16	1.22	0.12	1.25
(7,401)	1:A:156:LEU:HD22	1:A:186:ILE:H	16	1.22	0.12	1.25
(7,401)	1:A:156:LEU:HD23	1:A:186:ILE:H	16	1.22	0.12	1.25
(7,261)	1:A:108:LEU:HD11	1:A:111:LYS:HE2	16	1.15	0.12	1.15
(7,261)	1:A:108:LEU:HD11	1:A:111:LYS:HE3	16	1.15	0.12	1.15
(7,261)	1:A:108:LEU:HD12	1:A:111:LYS:HE2	16	1.15	0.12	1.15
(7,261)	1:A:108:LEU:HD12	1:A:111:LYS:HE3	16	1.15	0.12	1.15
(7,261)	1:A:108:LEU:HD13	1:A:111:LYS:HE2	16	1.15	0.12	1.15
(7,261)	1:A:108:LEU:HD13	1:A:111:LYS:HE3	16	1.15	0.12	1.15
(7,307)	1:A:113:LEU:HB3	1:A:196:LEU:HD11	16	1.04	0.16	1.03
(7,307)	1:A:113:LEU:HB3	1:A:196:LEU:HD12	16	1.04	0.16	1.03
(7,307)	1:A:113:LEU:HB3	1:A:196:LEU:HD13	16	1.04	0.16	1.03
(1,1066)	1:A:156:LEU:HG	1:A:186:ILE:HG13	16	0.96	0.14	0.94
(1,712)	1:A:108:LEU:H	1:A:188:PHE:HD1	16	0.84	0.14	0.84
(1,712)	1:A:108:LEU:H	1:A:188:PHE:HD2	16	0.84	0.14	0.84
(7,206)	1:A:94:VAL:HG21	1:A:135:MET:HG2	16	0.82	0.07	0.82

*Continued on next page...*

Continued from previous page...

Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(7,206)	1:A:94:VAL:HG21	1:A:135:MET:HG3	16	0.82	0.07	0.82
(7,206)	1:A:94:VAL:HG22	1:A:135:MET:HG2	16	0.82	0.07	0.82
(7,206)	1:A:94:VAL:HG22	1:A:135:MET:HG3	16	0.82	0.07	0.82
(7,206)	1:A:94:VAL:HG23	1:A:135:MET:HG2	16	0.82	0.07	0.82
(7,206)	1:A:94:VAL:HG23	1:A:135:MET:HG3	16	0.82	0.07	0.82
(1,1288)	1:A:198:HIS:H	1:A:200:LEU:HG	16	0.79	0.07	0.8
(7,260)	1:A:108:LEU:HD11	1:A:188:PHE:HD1	16	0.79	0.21	0.8
(7,260)	1:A:108:LEU:HD11	1:A:188:PHE:HD2	16	0.79	0.21	0.8
(7,260)	1:A:108:LEU:HD12	1:A:188:PHE:HD1	16	0.79	0.21	0.8
(7,260)	1:A:108:LEU:HD12	1:A:188:PHE:HD2	16	0.79	0.21	0.8
(7,260)	1:A:108:LEU:HD13	1:A:188:PHE:HD1	16	0.79	0.21	0.8
(7,260)	1:A:108:LEU:HD13	1:A:188:PHE:HD2	16	0.79	0.21	0.8
(1,336)	1:A:60:GLU:H	1:A:62:MET:HG2	16	0.78	0.26	0.81
(1,336)	1:A:60:GLU:H	1:A:62:MET:HG3	16	0.78	0.26	0.81
(1,909)	1:A:127:VAL:HA	1:A:129:TRP:HD1	16	0.77	0.2	0.79

[1]Number of violated models, [2]Standard deviation

9.5   All violated distance restraints ⓘ

9.5.1   Histogram : Distribution of distance violations ⓘ

The following histogram shows the distribution of the absolute value of the violation for all violated restraints in the ensemble.
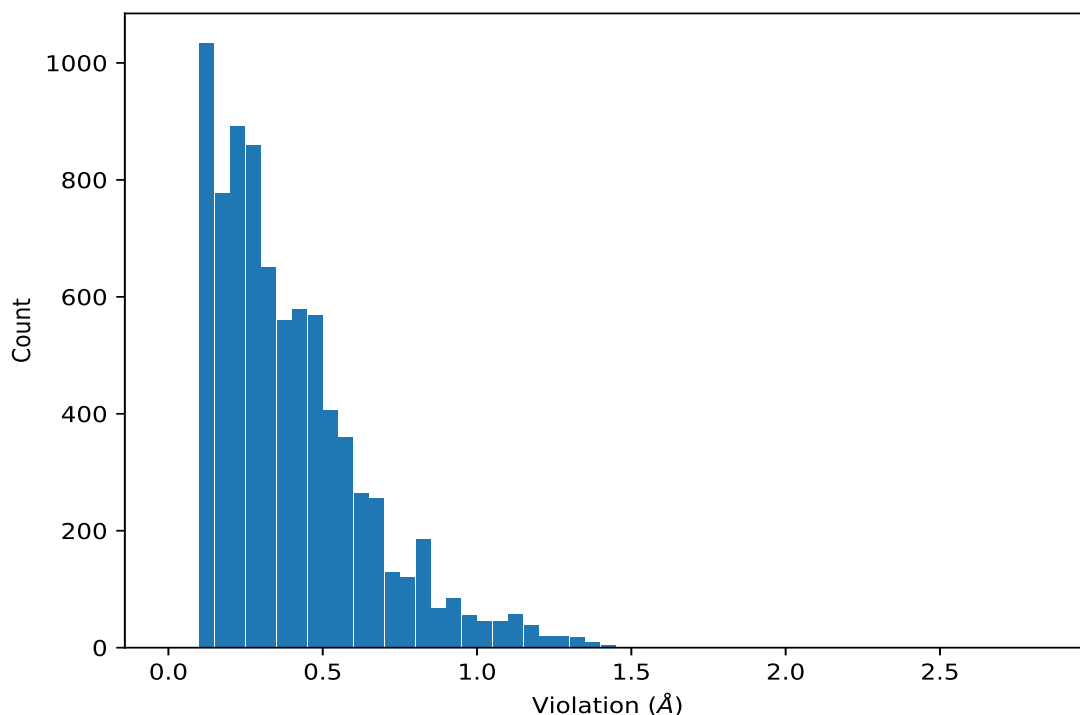
9.5.2　Table : All distance violations ⓘ

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)
(4,23)	4:A:304:ZN:ZN	1:A:194:HIS:NE2	6	2.82
(4,24)	4:A:304:ZN:ZN	1:A:194:HIS:CG	6	2.39
(5,2)	1:A:66:ARG:NH2	1:A:71:ASP:OD1	16	1.44
(7,401)	1:A:156:LEU:HD21	1:A:186:ILE:H	10	1.42
(7,401)	1:A:156:LEU:HD22	1:A:186:ILE:H	10	1.42
(7,401)	1:A:156:LEU:HD23	1:A:186:ILE:H	10	1.42
(7,401)	1:A:156:LEU:HD21	1:A:186:ILE:H	11	1.39
(7,401)	1:A:156:LEU:HD22	1:A:186:ILE:H	11	1.39
(7,401)	1:A:156:LEU:HD23	1:A:186:ILE:H	11	1.39
(7,261)	1:A:108:LEU:HD11	1:A:111:LYS:HE2	4	1.36
(7,261)	1:A:108:LEU:HD11	1:A:111:LYS:HE3	4	1.36
(7,261)	1:A:108:LEU:HD12	1:A:111:LYS:HE2	4	1.36
(7,261)	1:A:108:LEU:HD12	1:A:111:LYS:HE3	4	1.36
(7,261)	1:A:108:LEU:HD13	1:A:111:LYS:HE2	4	1.36
(7,261)	1:A:108:LEU:HD13	1:A:111:LYS:HE3	4	1.36
(4,12)	4:A:303:ZN:ZN	1:A:158:HIS:NE2	6	1.36
(7,401)	1:A:156:LEU:HD21	1:A:186:ILE:H	2	1.33

Continued on next page...

Continued from previous page...

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(7,401)	1:A:156:LEU:HD22	1:A:186:ILE:H	2	1.33
(7,401)	1:A:156:LEU:HD23	1:A:186:ILE:H	2	1.33
(7,307)	1:A:113:LEU:HB3	1:A:196:LEU:HD11	4	1.33
(7,307)	1:A:113:LEU:HB3	1:A:196:LEU:HD12	4	1.33
(7,307)	1:A:113:LEU:HB3	1:A:196:LEU:HD13	4	1.33
(7,401)	1:A:156:LEU:HD21	1:A:186:ILE:H	12	1.31

10 Dihedral-angle violation analysis ⓘ

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