WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

Apr 9, 2024 – 01:06 pm BST

PDB ID	:	8S03
BMRB ID	:	34899
Title	:	NMR solution structure of the CysD2 domain of MUC2
Authors	:	Recktenwald, C.; Karlsson, B.G.; Garcia-Bonnete, M.-J.; Katona, G.; Jensen, M.; Lymer, R.; Baeckstroem, M.; Johansson, M.E.V.; Hansson, G.C.; Trillo-Muyo, S.
Deposited on	:	2024-02-13

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

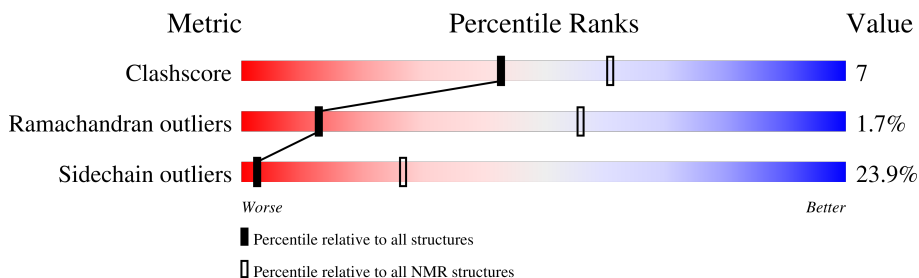|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 77%.

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		7
Ramachandran outliers		1.7%
Sidechain outliers		23.9%

Worse   *Better*

▮ Percentile relative to all structures

◲ Percentile relative to all NMR structures

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	142	35% 14% • 20% 30%

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy.*

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:1785-A:1793, A:1804-A:1852, A:1864-A:1876 (71)	0.94	2

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 4, 5, 6, 7, 8, 9, 10
2	1, 3

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 1440 atoms, of which 701 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mucin-2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	100	1439	464	701	122	140	12	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1751	ASP	-	expression tag	UNP Q02817
A	1752	ALA	-	expression tag	UNP Q02817
A	1753	ALA	-	expression tag	UNP Q02817
A	1754	GLN	-	expression tag	UNP Q02817
A	1755	PRO	-	expression tag	UNP Q02817
A	1756	ALA	-	expression tag	UNP Q02817
A	1757	ARG	-	expression tag	UNP Q02817
A	1758	ARG	-	expression tag	UNP Q02817
A	1759	ALA	-	expression tag	UNP Q02817
A	1760	VAL	-	expression tag	UNP Q02817
A	1761	ARG	-	expression tag	UNP Q02817
A	1762	SER	-	expression tag	UNP Q02817
A	1763	SER	-	expression tag	UNP Q02817
A	1764	SER	-	expression tag	UNP Q02817
A	1765	GLU	-	expression tag	UNP Q02817
A	1766	LEU	-	expression tag	UNP Q02817
A	1767	THR	-	expression tag	UNP Q02817
A	1768	SER	-	expression tag	UNP Q02817
A	1769	GLU	-	expression tag	UNP Q02817
A	1770	GLN	-	expression tag	UNP Q02817
A	1771	LYS	-	expression tag	UNP Q02817
A	1772	LEU	-	expression tag	UNP Q02817
A	1773	ILE	-	expression tag	UNP Q02817
A	1774	SER	-	expression tag	UNP Q02817
A	1775	GLU	-	expression tag	UNP Q02817
A	1776	GLU	-	expression tag	UNP Q02817
A	1777	ASP	-	expression tag	UNP Q02817
A	1778	LEU	-	expression tag	UNP Q02817
A	1779	SER	-	expression tag	UNP Q02817
A	1780	SER	-	expression tag	UNP Q02817

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms
2	A	1	Total Ca 1 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.
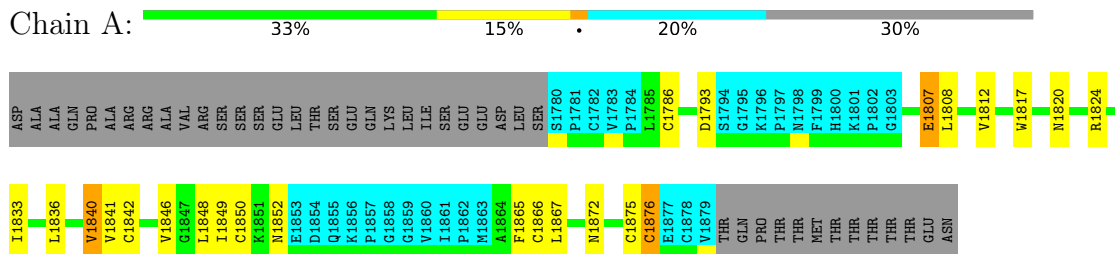
• Molecule 1: Mucin-2
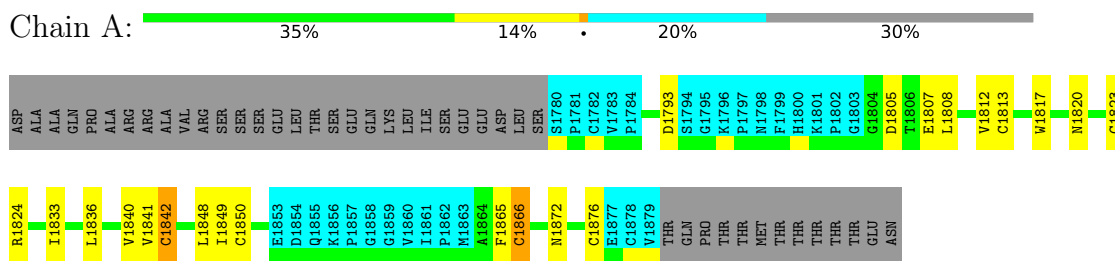
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

• Molecule 1: Mucin-2

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *ENERGY MINIMIZATION*.

Of the 200 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
YASARA	refinement	2.5
CYANA	structure calculation	3.0

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	812
Number of shifts mapped to atoms	812
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	77%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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

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	1.28±0.17	4±3/540 (0.8± 0.5%)	1.23±0.11	3±3/741 (0.4± 0.3%)
All	All	1.29	43/5400 (0.8%)	1.24	26/7410 (0.4%)

5 of 23 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	1817	TRP	CB-CG	-8.16	1.35	1.50	7	6
1	A	1828	TYR	CD1-CE1	8.09	1.51	1.39	6	1
1	A	1840	VAL	CB-CG2	7.35	1.68	1.52	9	1
1	A	1817	TRP	CG-CD2	7.32	1.56	1.43	6	6
1	A	1791	TRP	CD2-CE2	7.29	1.50	1.41	8	4

5 of 18 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	A	1840	VAL	CA-CB-CG2	8.35	123.42	110.90	6	1
1	A	1824	ARG	NE-CZ-NH1	6.66	123.63	120.30	5	5
1	A	1836	LEU	CB-CG-CD1	6.43	121.93	111.00	1	1
1	A	1842	CYS	CA-CB-SG	6.37	125.47	114.00	6	1
1	A	1869	TYR	CB-CG-CD2	-6.36	117.18	121.00	9	2

There are no chirality outliers.

There are no planarity outliers.

WORLDWIDE wwPDB PROTEIN DATA BANK

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
1	A	528	497	505	7±2
2	A	1	0	0	1±1
All	All	5290	4970	5050	73

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 7.

5 of 38 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:1833:ILE:HD12	1:A:1840:VAL:HG11	0.69	1.64	5	1
1:A:1850:CYS:CB	1:A:1866:CYS:HG	0.69	2.00	6	3
1:A:1840:VAL:HG12	1:A:1850:CYS:SG	0.67	2.29	9	1
1:A:1850:CYS:HG	1:A:1866:CYS:CB	0.65	2.05	7	6
1:A:1793:ASP:OD1	2:A:1901:CA:CA	0.63	1.76	9	3

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	71/142 (50%)	64±1 (90±1%)	6±1 (8±2%)	1±1 (2±1%)	13	56
All	All	710/1420 (50%)	640 (90%)	58 (8%)	12 (2%)	13	56

All 3 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	1866	CYS	6
1	A	1814	GLY	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	1865	PHE	3

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	59/122 (48%)	45±2 (76±3%)	14±2 (24±3%)	2	27
All	All	590/1220 (48%)	449 (76%)	141 (24%)	2	27

5 of 42 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	1836	LEU	9
1	A	1808	LEU	8
1	A	1833	ILE	8
1	A	1876	CYS	8
1	A	1812	VAL	7

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 ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 77% for the well-defined parts and 65% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *nef_chemical_shift_list_ShiftList_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	812
Number of shifts mapped to atoms	812
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	83	0.05 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_\beta$	72	-0.40 ± 0.31	None needed (< 0.5 ppm)
$^{13}C'$	2	—	None (insufficient data)
^{15}N	71	-0.96 ± 0.34	Should be applied

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 77%, i.e. 684 atoms were assigned a chemical shift out of a possible 889. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	268/357 (75%)	138/147 (94%)	70/142 (49%)	60/68 (88%)
Sidechain	374/468 (80%)	257/308 (83%)	117/148 (79%)	0/12 (0%)

Continued on next page...

Continued from previous page...

|	Total	^{1}H	^{13}C	^{15}N
Aromatic	42/64 (66%)	30/31 (97%)	9/30 (30%)	3/3 (100%)
Overall	684/889 (77%)	425/486 (87%)	196/320 (61%)	63/83 (76%)

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	1789	THR	HG1	6.14	0.08 − 2.19	23.7
1	A	1874	GLN	HB2	-1.31	0.80 − 3.29	-13.5
1	A	1872	ASN	HB2	-0.03	1.27 − 4.34	-9.2
1	A	1874	GLN	HG2	0.48	1.01 − 3.62	-7.0
1	A	1794	SER	H	11.14	5.45 − 11.10	5.1

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:
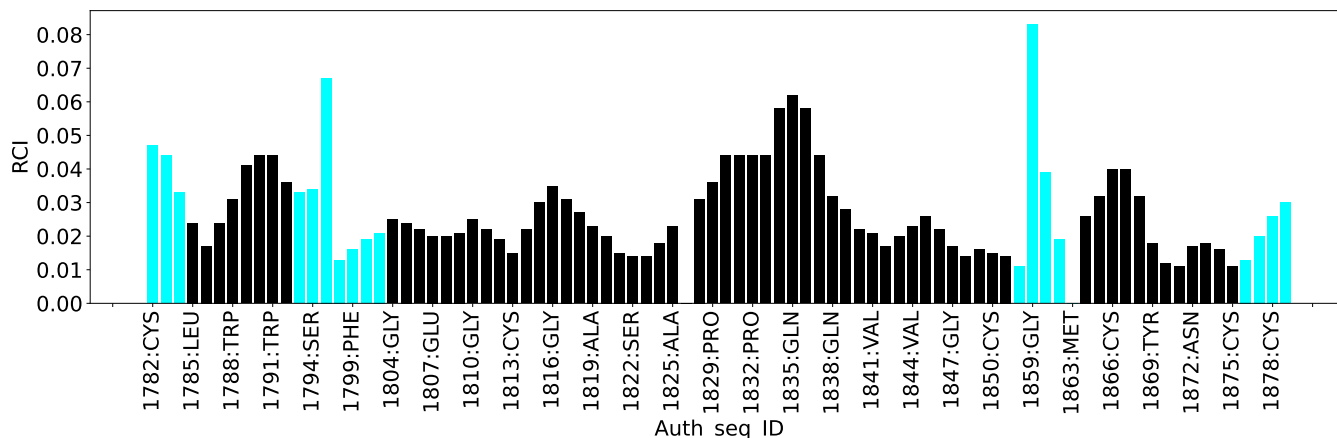
8 NMR restraints analysis ⓘ

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	732
Intra-residue ($|i\text{-}j|$=0)	340
Sequential ($|i\text{-}j|$=1)	203
Medium range ($|i\text{-}j|$>1 and $|i\text{-}j|$<5)	39
Long range ($|i\text{-}j|$≥5)	150
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	151
Number of unmapped restraints	0
Number of restraints per residue	6.2
Number of long range restraints per residue[1]	1.0

[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.

Bins (Å)	Average number of violations per model	Max (Å)
0.1-0.2 (Small)	49.9	0.2
0.2-0.5 (Medium)	103.2	0.5
>0.5 (Large)	90.0	10.47

8.2.2 Average number of dihedral-angle violations per model ⓘ

Dihedral-angle violations less than 1° are not included in the calculation.

Bins (°)	Average number of violations per model	Max (°)
1.0-10.0 (Small)	1.7	7.1
10.0-20.0 (Medium)	1.9	19.46
>20.0 (Large)	2.2	130.56

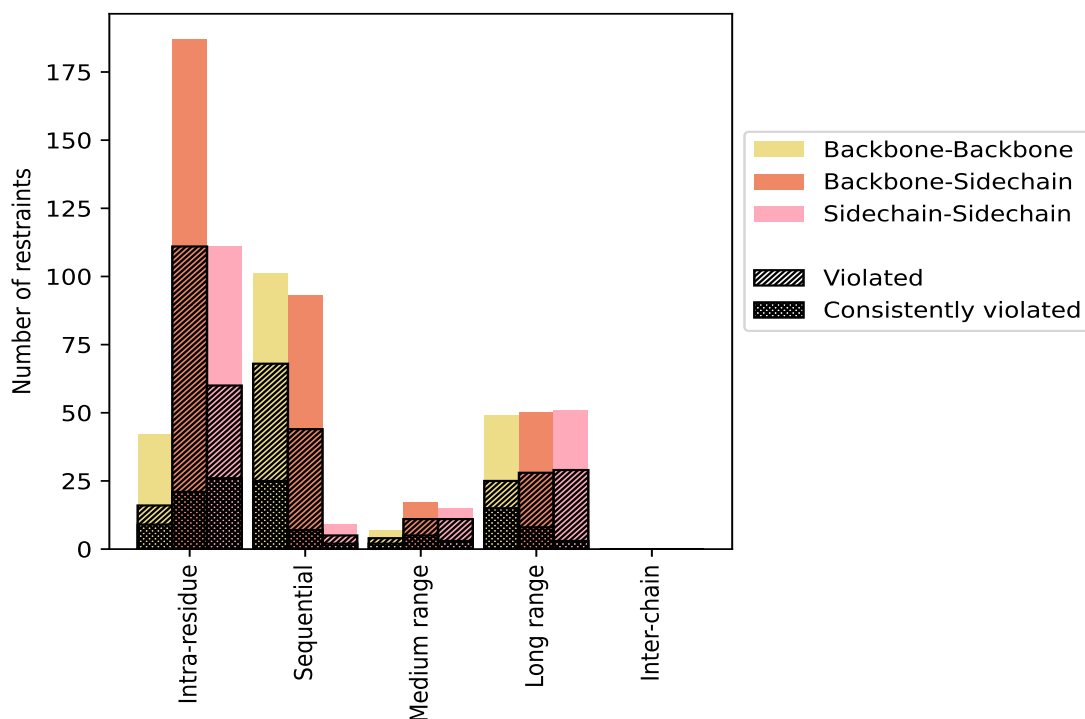# 9 Distance violation analysis ⓘ

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)	340	46.4	187	55.0	25.5	56	16.5	7.7
Backbone-Backbone	42	5.7	16	38.1	2.2	9	21.4	1.2
Backbone-Sidechain	187	25.5	111	59.4	15.2	21	11.2	2.9
Sidechain-Sidechain	111	15.2	60	54.1	8.2	26	23.4	3.6
Sequential (|i-j|=1)	203	27.7	117	57.6	16.0	34	16.7	4.6
Backbone-Backbone	101	13.8	68	67.3	9.3	25	24.8	3.4
Backbone-Sidechain	93	12.7	44	47.3	6.0	7	7.5	1.0
Sidechain-Sidechain	9	1.2	5	55.6	0.7	2	22.2	0.3
Medium range (|i-j|>1 & |i-j|<5)	39	5.3	26	66.7	3.6	10	25.6	1.4
Backbone-Backbone	7	1.0	4	57.1	0.5	2	28.6	0.3
Backbone-Sidechain	17	2.3	11	64.7	1.5	5	29.4	0.7
Sidechain-Sidechain	15	2.0	11	73.3	1.5	3	20.0	0.4
Long range (|i-j|≥5)	150	20.5	82	54.7	11.2	26	17.3	3.6
Backbone-Backbone	49	6.7	25	51.0	3.4	15	30.6	2.0
Backbone-Sidechain	50	6.8	28	56.0	3.8	8	16.0	1.1
Sidechain-Sidechain	51	7.0	29	56.9	4.0	3	5.9	0.4
Inter-chain	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	732	100.0	412	56.3	56.3	126	17.2	17.2
Backbone-Backbone	199	27.2	113	56.8	15.4	51	25.6	7.0
Backbone-Sidechain	347	47.4	194	55.9	26.5	41	11.8	5.6
Sidechain-Sidechain	186	25.4	105	56.5	14.3	34	18.3	4.6

[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
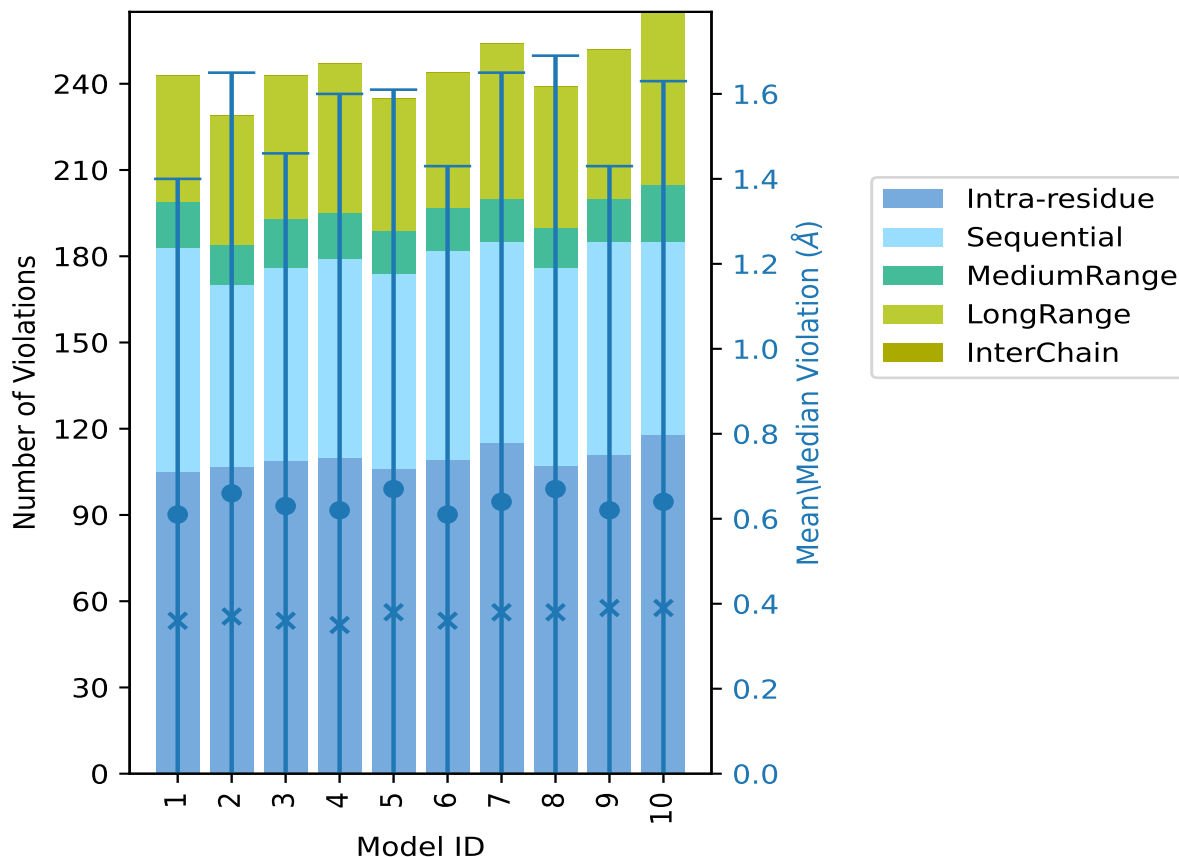
9.2 Distance violation statistics for each model ⓘ

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	105	78	16	44	0	243	0.61	5.89	0.79	0.36
2	107	63	14	45	0	229	0.66	8.99	0.99	0.37
3	109	67	17	50	0	243	0.63	7.33	0.83	0.36
4	110	69	16	52	0	247	0.62	8.63	0.98	0.35
5	106	68	15	46	0	235	0.67	7.17	0.94	0.38
6	109	73	15	47	0	244	0.61	6.35	0.82	0.36
7	115	70	15	54	0	254	0.64	10.47	1.01	0.38
8	107	69	14	49	0	239	0.67	8.88	1.02	0.38
9	111	74	15	52	0	252	0.62	6.43	0.81	0.39
10	118	67	20	60	0	265	0.64	10.44	0.99	0.39

[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, 320(IR:153, SQ:86, MR:13, LR:68, IC:0) restraints are not violated in the ensemble.

Number of violated restraints						Fraction of the ensemble	
IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total	Count[6]	%
28	11	5	7	0	51	1	10.0
22	13	1	9	0	45	2	20.0
16	11	2	8	0	37	3	30.0
13	11	1	9	0	34	4	40.0

Continued on next page...

Continued from previous page...

Number of violated restraints						Fraction of the ensemble	
IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total	Count[6]	%
11	12	5	4	0	32	5	50.0
9	7	0	9	0	25	6	60.0
9	8	1	2	0	20	7	70.0
14	4	1	6	0	25	8	80.0
9	6	0	2	0	17	9	90.0
56	34	10	26	0	126	10	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 ⓘ
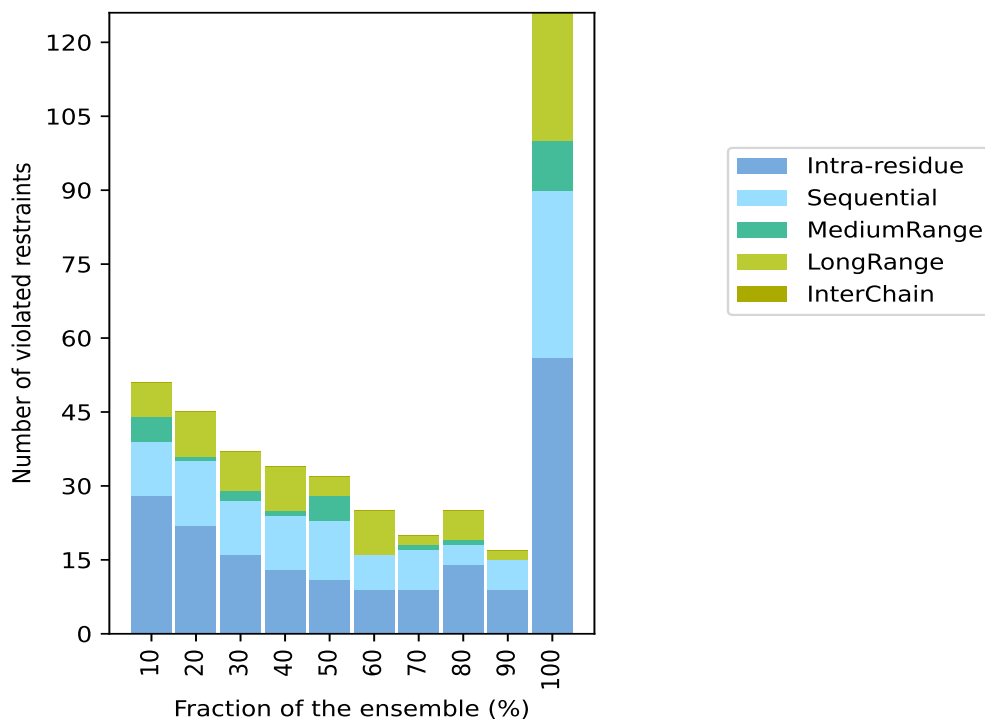
9.4   Most violated distance restraints in the ensemble ⓘ

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
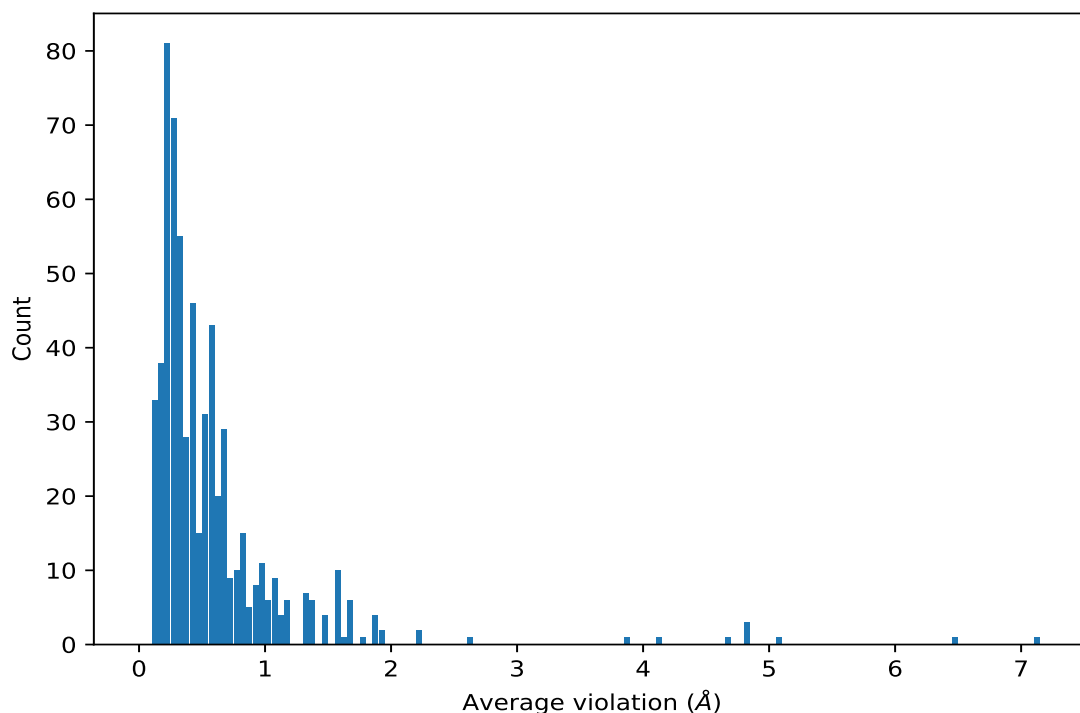
9.4.2   Table: Most violated distance restraints ⓘ

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)
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	10	7.14	2.61	7.9
(1,335)	1:1805:A:ASP:H	1:1870:A:GLU:HA	10	6.45	0.41	6.39
(1,267)	1:1823:A:CYS:HA	1:1847:A:GLY:HA3	10	5.08	0.27	5.0
(1,222)	1:1818:A:ALA:H	1:1844:A:VAL:HG21	10	4.81	0.9	4.72
(1,222)	1:1818:A:ALA:H	1:1844:A:VAL:HG22	10	4.81	0.9	4.72
(1,222)	1:1818:A:ALA:H	1:1844:A:VAL:HG23	10	4.81	0.9	4.72
(1,266)	1:1823:A:CYS:HA	1:1847:A:GLY:HA3	10	4.68	0.27	4.6
(1,218)	1:1818:A:ALA:H	1:1874:A:GLN:H	10	4.12	0.41	4.26
(1,228)	1:1818:A:ALA:H	1:1874:A:GLN:H	10	3.89	0.42	4.03
(1,422)	1:1787:A:ASN:HA	1:1820:A:ASN:H	10	2.61	0.36	2.52
(1,689)	1:1828:A:TYR:HE1	1:1829:A:PRO:HD3	10	2.24	0.06	2.23
(1,689)	1:1828:A:TYR:HE2	1:1829:A:PRO:HD3	10	2.24	0.06	2.23
(1,685)	1:1828:A:TYR:HE1	1:1829:A:PRO:HD2	10	1.93	0.2	1.88

[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.
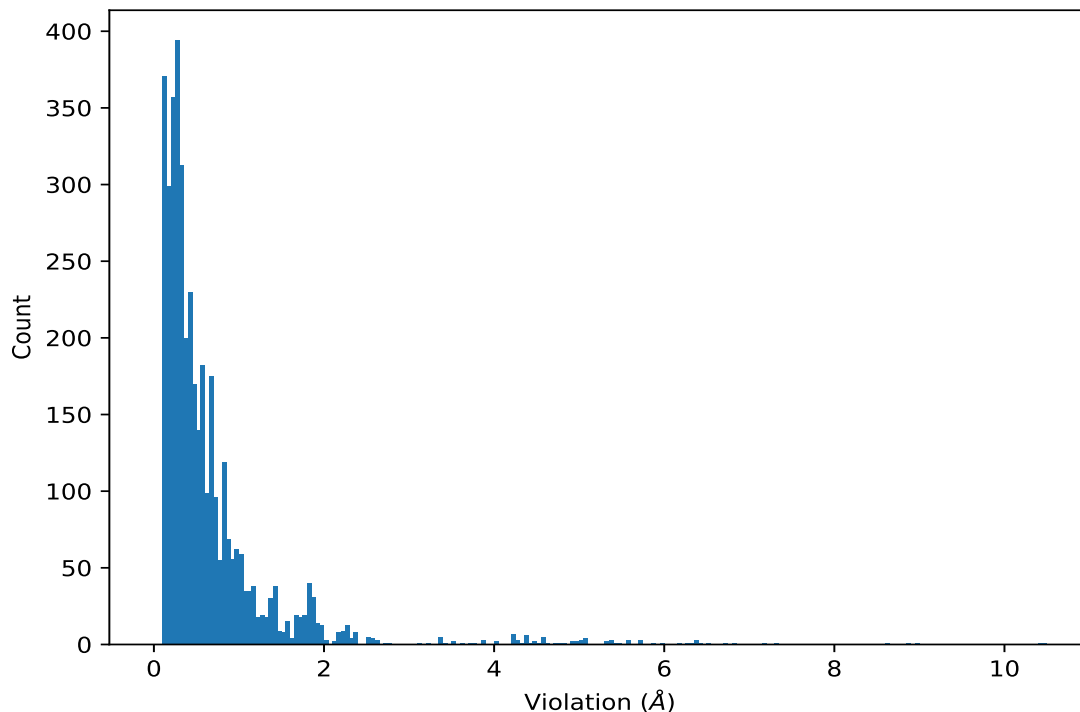
9.5.2 Table : All distance violations ⓘ

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)
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	7	10.47
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	10	10.44
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	2	8.99
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	8	8.88
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	4	8.63
(1,335)	1:1805:A:ASP:H	1:1870:A:GLU:HA	3	7.33
(1,293)	1:1795:A:GLY:H	1:1799:A:PHE:HA	5	7.17
(1,335)	1:1805:A:ASP:H	1:1870:A:GLU:HA	8	6.84
(1,335)	1:1805:A:ASP:H	1:1870:A:GLU:HA	4	6.74
(1,335)	1:1805:A:ASP:H	1:1870:A:GLU:HA	5	6.51

10 Dihedral-angle violation analysis ⓘ
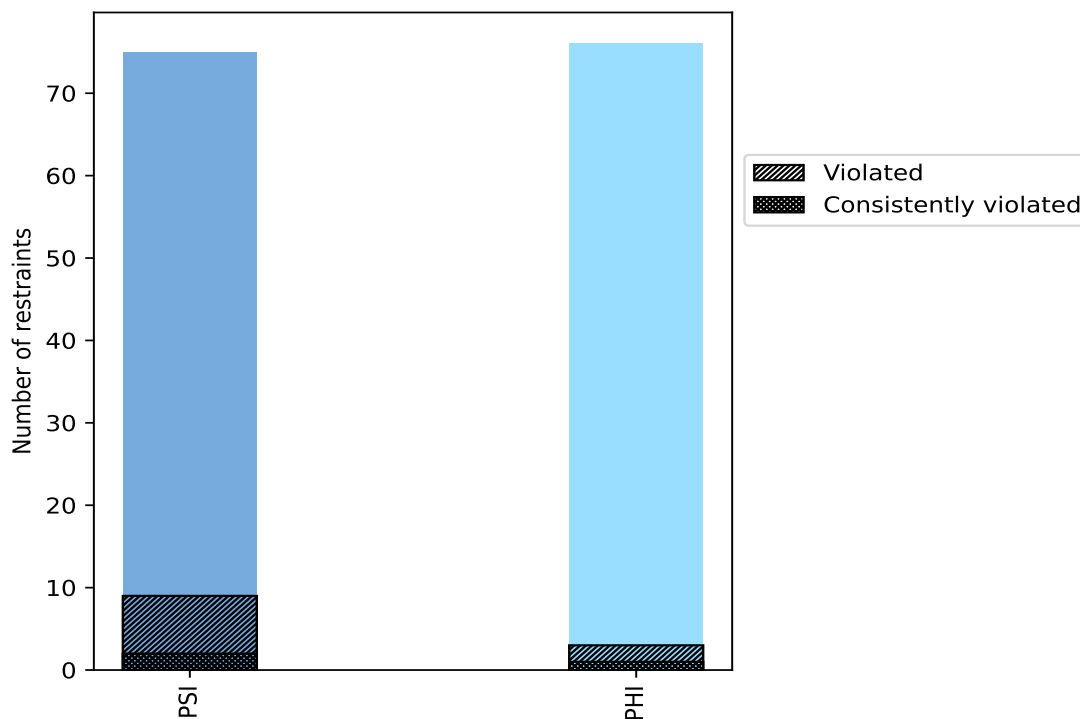
10.1 Summary of dihedral-angle violations ⓘ

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	$\%^2$	$\%^1$
PSI	75	49.7	9	12.0	6.0	2	2.7	1.3
PHI	76	50.3	3	3.9	2.0	1	1.3	0.7
Total	151	100.0	12	7.9	7.9	3	2.0	2.0

[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 ⓘ


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)

The following table provides the dihedral-angle violation statistics for each model in the ensemble. Violations less than 1° are not included in the statistics.

Model ID	Number of violations			Mean (°)	Max (°)	SD (°)	Median (°)
	PSI	PHI	Total				
1	3	1	4	59.96	124.76	47.99	52.22
2	4	1	5	53.59	118.33	43.46	36.02
3	6	1	7	36.46	129.22	45.84	10.72
4	4	1	5	47.96	127.17	50.52	12.48
5	3	1	4	54.73	110.03	46.49	50.9
6	3	3	6	39.62	121.77	49.87	7.84
7	5	2	7	35.78	130.56	47.8	10.59
8	5	1	6	45.54	129.7	46.32	17.86
9	5	2	7	33.07	112.66	42.71	10.39
10	5	2	7	34.49	113.74	45.24	6.2

10.2.1 Bar graph : Dihedral violation statistics for each model (i)
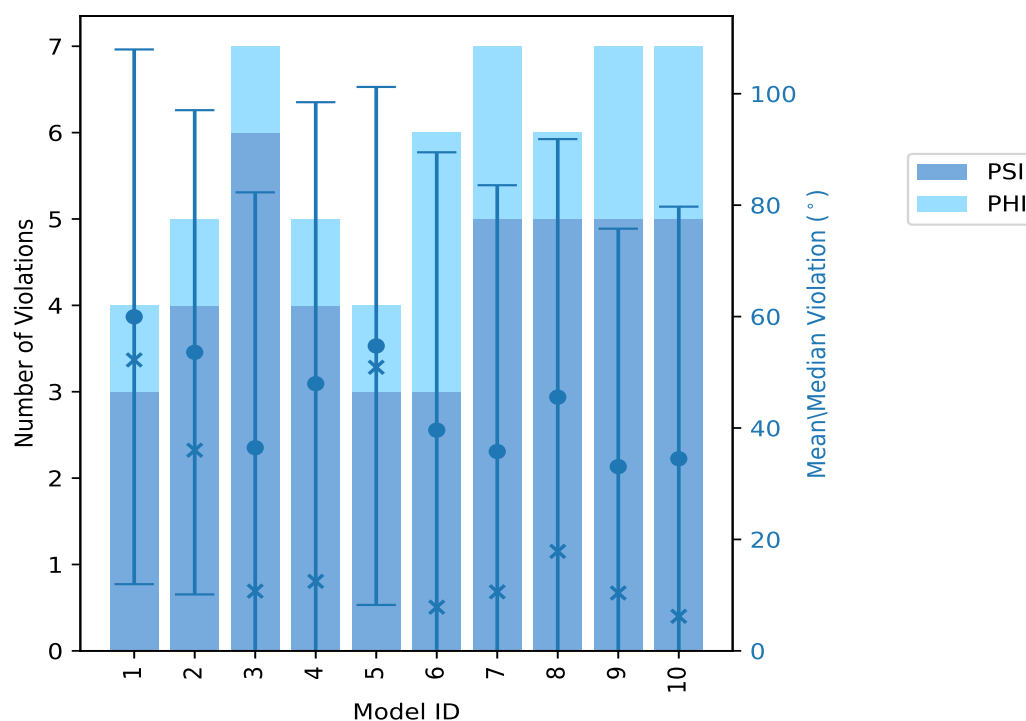
The mean(dot),median(x) and the standard deviation are shown in blue with respect to the y axis on the right

## 10.3 Dihedral-angle violation statistics for the ensemble ⓘ

Violation analysis may find that some restraints are violated in very 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 ensemble.

Number of violated restraints			Fraction of the ensemble	
PSI	PHI	Total	Count[1]	%
3	1	4	1	10.0
0	0	0	2	20.0
1	0	1	3	30.0
2	1	3	4	40.0
0	0	0	5	50.0
0	0	0	6	60.0
0	0	0	7	70.0
0	0	0	8	80.0
1	0	1	9	90.0
2	1	3	10	100.0

[1] Number of models with violations

10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble ⓘ
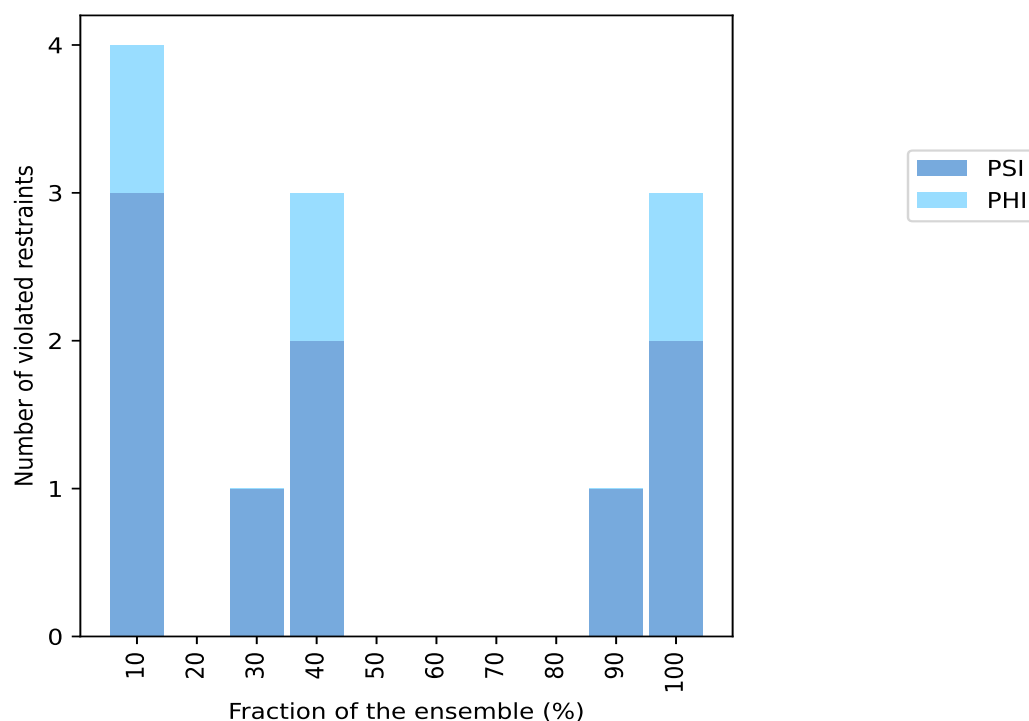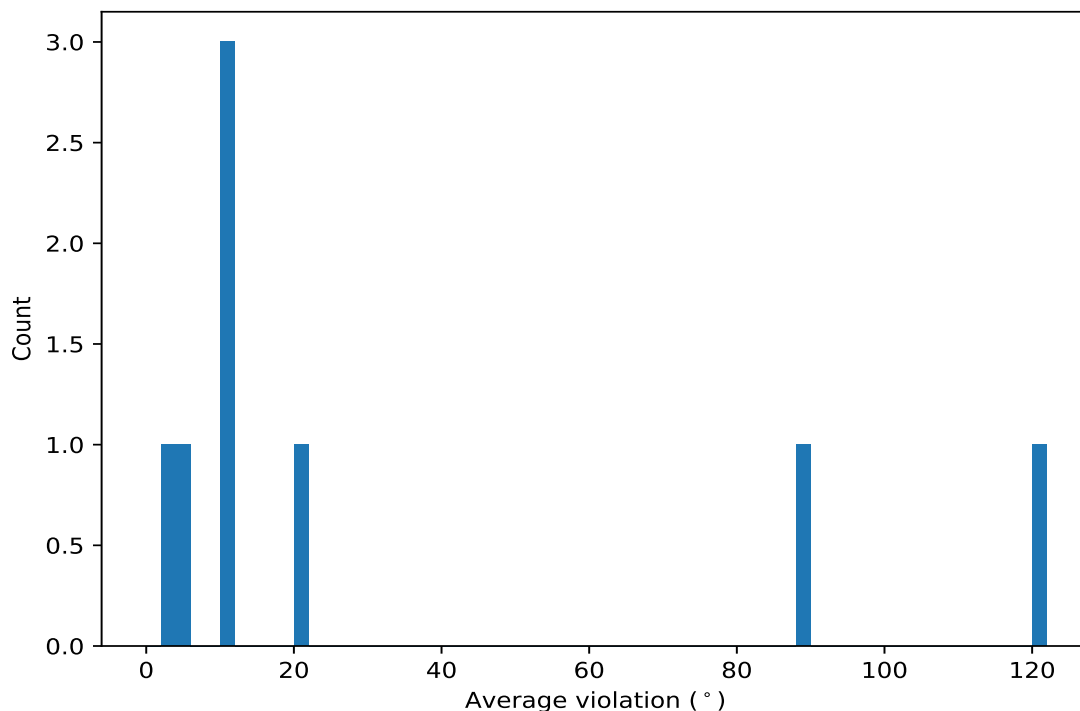
10.4   Most violated dihedral-angle restraints in the ensemble ⓘ

10.4.1   Histogram : Distribution of mean dihedral-angle 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

10.4.2   Table: Most violated dihedral-angle restraints ⓘ

The following table provides the mean and the standard deviation of the violation for each restraint sorted by number of violated models and the mean value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint.
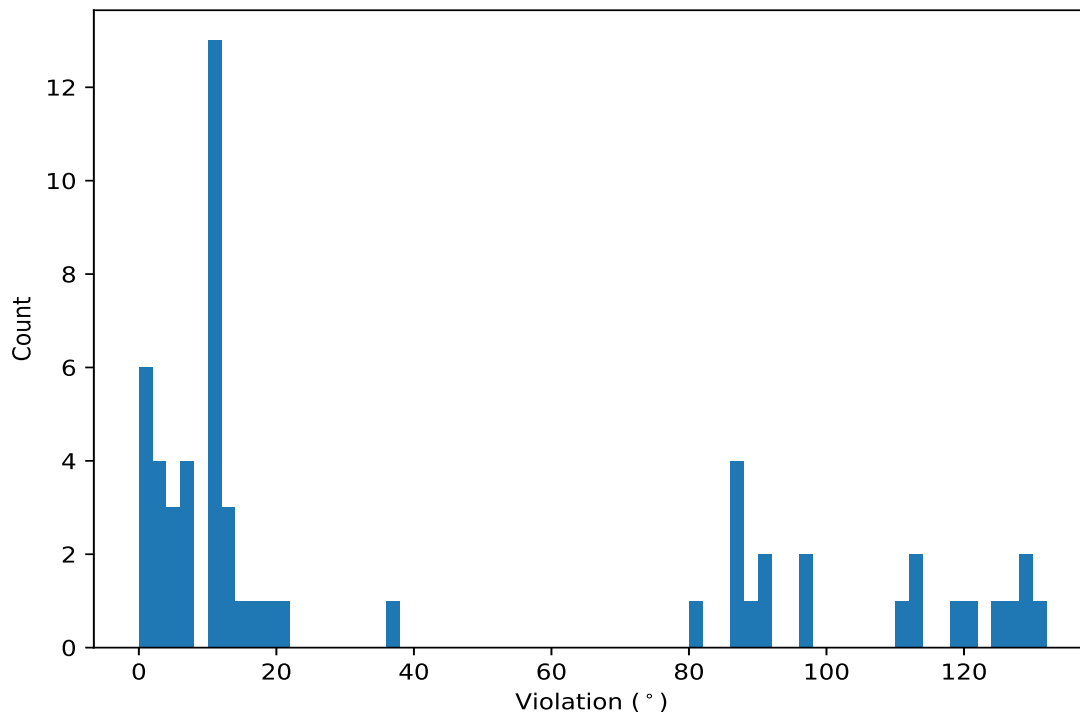
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Med
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	10	121.79	7.29	123
(1,70)	1:1827:A:MET:N	1:1827:A:MET:CA	1:1827:A:MET:C	1:1828:A:TYR:N	10	89.49	4.55	87
(1,67)	1:1825:A:ALA:C	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	10	11.45	4.7	12.
(1,126)	1:1864:A:ALA:N	1:1864:A:ALA:CA	1:1864:A:ALA:C	1:1865:A:PHE:N	9	10.41	0.2	10.
(1,4)	1:1782:A:CYS:N	1:1782:A:CYS:CA	1:1782:A:CYS:C	1:1783:A:VAL:N	4	10.41	0.13	10.
(1,99)	1:1842:A:CYS:C	1:1843:A:ASP:N	1:1843:A:ASP:CA	1:1843:A:ASP:C	4	4.0	2.14	4.
(1,94)	1:1840:A:VAL:N	1:1840:A:VAL:CA	1:1840:A:VAL:C	1:1841:A:VAL:N	4	2.42	0.75	2.5
(1,2)	1:1781:A:PRO:N	1:1781:A:PRO:CA	1:1781:A:PRO:C	1:1782:A:CYS:N	3	20.74	12.64	21.

[1] Number of violated models, [2]Standard deviation, All angle values are in degree (°)

10.5 All violated dihedral-angle restraints ⓘ

10.5.1 Histogram : Distribution of violations ⓘ

The following histogram shows the distribution of the absolute value of the violation for all violated restraints in the ensemble.

10.5.2 Table: All violated dihedral-angle restraints ⓘ

The following table provides the list of violationsfor 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.

Key	Atom-1	Atom-2	Atom-3	Atom-4	Model ID	Violation (°)
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	7	130.56
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	8	129.7
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	3	129.22
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	4	127.17
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	1	124.76
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	6	121.77
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	2	118.33
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	10	113.74
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	9	112.66
(1,68)	1:1826:A:THR:N	1:1826:A:THR:CA	1:1826:A:THR:C	1:1827:A:MET:N	5	110.03