wwPDB NMR Structure Validation Summary Report

Sep 2, 2024 – 05:18 pm BST

PDB ID	:	8R1X
BMRB ID	:	34874
Title	:	Solution structure and chemical shift assignments for HMG-D Y12F mutant complexed to a 14:12 dA2 bulge DNA
Authors	:	Yang, J.C.; Hill, G.R.; Neuhaus, D.
Deposited on	:	2023-11-02

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 (1)) were used in the production of this report:

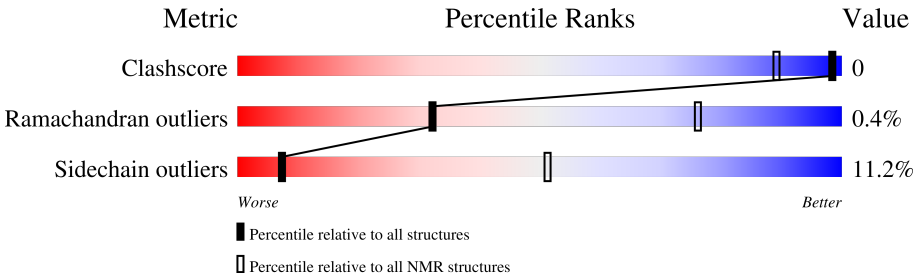
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

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 59%.

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
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	112	57% 5% · 35%
2	B	14	14% 79% 7%
3	C	12	17% 83%

2 Ensemble composition and analysis

This entry contains 20 models. Model 5 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:5-A:74 (70)	0.59	5

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 and 2 single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 7, 11, 12, 13, 14, 15, 16, 18, 19
2	1, 6, 8, 9, 10
Single-model clusters	17; 20

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2000 atoms, of which 886 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called High mobility group protein D.

Mol	Chain	Residues	Atoms						Trace
1	A	73	Total	C	H	N	O	S	0
			1176	367	590	108	109	2	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	PHE	TYR	engineered mutation	UNP Q05783

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

Mol	Chain	Residues	Atoms						Trace
2	B	14	Total	C	H	N	O	P	0
			444	137	159	55	80	13	

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

Mol	Chain	Residues	Atoms						Trace
3	C	12	Total	C	H	N	O	P	0
			380	117	137	45	70	11	

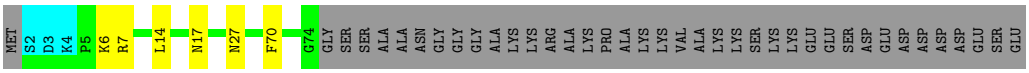
4 Residue-property plots

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.

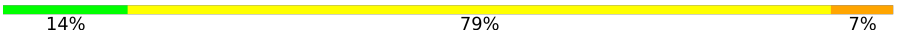
• Molecule 1: High mobility group protein D

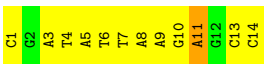
Chain A:



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

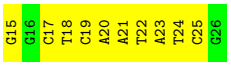
Chain B:





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

Chain C:



4.2 Residue scores for the representative (medoid) model from the NMR ensemble

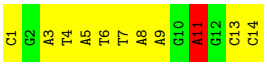
The representative model is number 5. Colouring as in section 4.1 above.

• Molecule 1: High mobility group protein D

Chain A:

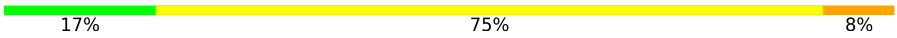
• Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B:



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

Chain C:



5 Refinement protocol and experimental data overview

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

Of the 60 calculated structures, 20 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
X-PLOR NIH	structure calculation	2.28
Amber	refinement	11

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	1096
Number of shifts mapped to atoms	930
Number of unparsed shifts	0
Number of shifts with mapping errors	166
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	59%

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	0.72±0.01	0±0/574 (0.0± 0.0%)	1.06±0.03	1±1/766 (0.2± 0.1%)
2	B	1.61±0.02	0±0/320 (0.0± 0.0%)	2.41±0.02	29±1/492 (5.8± 0.2%)
3	C	1.63±0.01	0±0/272 (0.0± 0.0%)	2.35±0.02	21±1/418 (5.1± 0.3%)
All	All	1.26	0/23320 (0.0%)	1.90	1023/33520 (3.1%)

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	0.2±0.4
2	B	0.0±0.0	1.5±0.6
3	C	0.0±0.0	0.8±0.9
All	All	0	52

There are no bond-length outliers.

5 of 71 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	9	DA	N1-C6-N6	-10.49	112.31	118.60	15	20
2	B	8	DA	N1-C6-N6	-9.59	112.85	118.60	14	20
3	C	20	DA	N1-C6-N6	-9.18	113.09	118.60	3	20
2	B	5	DA	N1-C6-N6	-8.85	113.29	118.60	7	20
3	C	23	DA	N1-C6-N6	-8.74	113.36	118.60	5	20

There are no chirality outliers.

5 of 12 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	11	DA	Sidechain	20
3	C	19	DC	Sidechain	7
2	B	10	DG	Sidechain	6
1	A	63	TYR	Sidechain	5
3	C	20	DA	Sidechain	5

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	563	568	568	0±1
2	B	285	159	159	0±0
All	All	21820	17280	17280	7

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:43:TRP:CZ2	1:A:49:LYS:HE2	0.46	2.46	13	1
1:A:8:PRO:HD3	1:A:63:TYR:CG	0.44	2.47	3	3
1:A:60:LYS:HE2	2:B:11:DA:O4'	0.43	2.13	3	3

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	69/112 (62%)	66±1 (95±2%)	3±1 (5±2%)	0±1 (0±1%)	32	76
All	All	1380/2240 (62%)	1310 (95%)	64 (5%)	6 (0%)	32	76

All 2 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	32	VAL	4
1	A	7	ARG	2

6.3.2 Protein sidechains (i)

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	56/88 (64%)	50±2 (89±3%)	6±2 (11±3%)	7 51
All	All	1120/1760 (64%)	995 (89%)	125 (11%)	7 51

5 of 22 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	27	ASN	18
1	A	14	LEU	14
1	A	17	ASN	12
1	A	6	LYS	11
1	A	49	LYS	9

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

There are no ligands in this entry.

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 59% for the well-defined parts and 60% 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

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	1096
Number of shifts mapped to atoms	930
Number of unparsed shifts	0
Number of shifts with mapping errors	166
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

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

- No matching atom found in the structure. First 5 (of 166) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	MET	HE1	1.375	0.02	1
1	A	1	MET	HE2	1.375	0.02	1
1	A	1	MET	HE3	1.375	0.02	1
1	A	1	MET	CE	15.618	0.10	1
1	A	75	GLY	H	8.011	0.02	1
1	A	75	GLY	HA2	3.399	0.02	2
1	A	75	GLY	HA3	3.254	0.02	2
1	A	75	GLY	CA	45.741	0.10	1
1	A	75	GLY	N	108.622	0.10	1
1	A	76	SER	H	8.296	0.02	1
1	A	76	SER	HA	4.555	0.02	1
1	A	76	SER	HB2	3.937	0.02	2
1	A	76	SER	HB3	3.937	0.02	2
1	A	76	SER	CA	58.839	0.10	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	76	SER	CB	64.309	0.10	1
1	A	76	SER	N	116.173	0.10	1
1	A	77	SER	H	8.497	0.02	1
1	A	77	SER	HA	4.507	0.02	1
1	A	77	SER	HB2	3.928	0.02	2
1	A	77	SER	HB3	3.928	0.02	2
1	A	77	SER	CA	59.057	0.10	1
1	A	77	SER	CB	64.254	0.10	1
1	A	77	SER	N	118.285	0.10	1
1	A	78	ALA	H	8.243	0.02	1
1	A	78	ALA	HA	4.327	0.02	1
1	A	78	ALA	HB1	1.396	0.02	1
1	A	78	ALA	HB2	1.396	0.02	1
1	A	78	ALA	HB3	1.396	0.02	1
1	A	78	ALA	CA	53.087	0.10	1
1	A	78	ALA	CB	19.652	0.10	1
1	A	78	ALA	N	125.47	0.10	1
1	A	79	ALA	H	8.184	0.02	1
1	A	79	ALA	HA	4.293	0.02	1
1	A	79	ALA	HB1	1.401	0.02	1
1	A	79	ALA	HB2	1.401	0.02	1
1	A	79	ALA	HB3	1.401	0.02	1
1	A	79	ALA	CA	53.069	0.10	1
1	A	79	ALA	CB	19.47	0.10	1
1	A	79	ALA	N	122.089	0.10	1
1	A	80	ASN	H	8.306	0.02	1
1	A	80	ASN	HA	4.724	0.02	1
1	A	80	ASN	HB2	2.837	0.02	2
1	A	80	ASN	HB3	2.837	0.02	2
1	A	80	ASN	HD21	7.637	0.02	1
1	A	80	ASN	HD22	6.948	0.02	1
1	A	80	ASN	CA	53.642	0.10	1
1	A	80	ASN	CB	39.338	0.10	1
1	A	80	ASN	N	116.879	0.10	1
1	A	80	ASN	ND2	112.16	0.10	1
1	A	87	ARG	HA	4.307	0.02	1
1	A	87	ARG	HB2	1.787	0.02	2
1	A	87	ARG	HB3	1.787	0.02	2
1	A	87	ARG	CA	56.542	0.10	1
1	A	87	ARG	CB	31.426	0.10	1
1	A	88	ALA	H	8.374	0.02	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	88	ALA	HA	4.301	0.02	1
1	A	88	ALA	HB1	1.37	0.02	1
1	A	88	ALA	HB2	1.37	0.02	1
1	A	88	ALA	HB3	1.37	0.02	1
1	A	88	ALA	CA	52.824	0.10	1
1	A	88	ALA	CB	19.846	0.10	1
1	A	88	ALA	N	125.461	0.10	1
1	A	90	PRO	HA	4.427	0.02	1
1	A	90	PRO	HB2	1.909	0.02	2
1	A	90	PRO	HB3	2.318	0.02	2
1	A	90	PRO	HD2	3.853	0.02	2
1	A	90	PRO	HD3	3.655	0.02	2
1	A	90	PRO	HG2	2.046	0.02	2
1	A	90	PRO	HG3	2.046	0.02	2
1	A	90	PRO	CA	63.367	0.10	1
1	A	90	PRO	CB	32.52	0.10	1
1	A	90	PRO	CD	51.07	0.10	1
1	A	90	PRO	CG	27.9	0.10	1
1	A	91	ALA	H	8.443	0.02	1
1	A	91	ALA	HA	4.307	0.02	1
1	A	91	ALA	HB1	1.397	0.02	1
1	A	91	ALA	HB2	1.397	0.02	1
1	A	91	ALA	HB3	1.397	0.02	1
1	A	91	ALA	CB	19.935	0.10	1
1	A	91	ALA	N	124.573	0.10	1
1	A	94	VAL	HA	4.102	0.02	1
1	A	94	VAL	HB	2.048	0.02	1
1	A	94	VAL	HG11	0.934	0.02	2
1	A	94	VAL	HG12	0.934	0.02	2
1	A	94	VAL	HG13	0.934	0.02	2
1	A	94	VAL	HG21	0.937	0.02	2
1	A	94	VAL	HG22	0.937	0.02	2
1	A	94	VAL	HG23	0.937	0.02	2
1	A	94	VAL	CA	62.347	0.10	1
1	A	94	VAL	CB	33.158	0.10	1
1	A	94	VAL	CG1	21.637	0.10	2
1	A	94	VAL	CG2	21.0	0.10	2
1	A	95	ALA	H	8.398	0.02	1
1	A	95	ALA	HA	4.307	0.02	1
1	A	95	ALA	HB1	1.364	0.02	1
1	A	95	ALA	HB2	1.364	0.02	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	95	ALA	HB3	1.364	0.02	1
1	A	95	ALA	CB	19.669	0.10	1
1	A	95	ALA	N	127.81	0.10	1
1	A	98	SER	HA	4.445	0.02	1
1	A	98	SER	HB2	3.866	0.02	2
1	A	98	SER	HB3	3.866	0.02	2
1	A	98	SER	CA	58.826	0.10	1
1	A	98	SER	CB	64.372	0.10	1
1	A	99	LYS	H	8.37	0.02	1
1	A	99	LYS	HA	4.329	0.02	1
1	A	99	LYS	HB2	1.845	0.02	2
1	A	99	LYS	HB3	1.742	0.02	2
1	A	99	LYS	HD2	1.678	0.02	2
1	A	99	LYS	HD3	1.678	0.02	2
1	A	99	LYS	HE2	3.189	0.02	2
1	A	99	LYS	HE3	3.189	0.02	2
1	A	99	LYS	HG2	1.477	0.02	2
1	A	99	LYS	HG3	1.477	0.02	2
1	A	99	LYS	CA	56.93	0.10	1
1	A	99	LYS	CB	33.561	0.10	1
1	A	99	LYS	CD	27.718	0.10	1
1	A	99	LYS	CE	43.946	0.10	1
1	A	99	LYS	N	122.773	0.10	1
1	A	102	GLU	CA	57.472	0.10	1
1	A	102	GLU	CB	30.286	0.10	1
1	A	103	SER	H	8.089	0.02	1
1	A	103	SER	CA	58.483	0.10	1
1	A	103	SER	CB	64.5	0.10	1
1	A	103	SER	N	114.51	0.10	1
1	A	104	ASP	H	8.329	0.02	1
1	A	104	ASP	N	122.285	0.10	1
1	A	109	ASP	CA	54.82	0.10	1
1	A	109	ASP	CB	41.692	0.10	1
1	A	110	GLU	H	8.39	0.02	1
1	A	110	GLU	CA	56.815	0.10	1
1	A	110	GLU	CB	30.504	0.10	1
1	A	110	GLU	N	121.073	0.10	1
1	A	111	SER	H	8.316	0.02	1
1	A	111	SER	CA	59.058	0.10	1
1	A	111	SER	CB	64.474	0.10	1
1	A	111	SER	N	117.273	0.10	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	112	GLU	H	8.056	0.02	1
1	A	112	GLU	N	127.478	0.10	1
1	A	1	MET	H1'	5.757	0.02	1
1	A	1	MET	H2'	1.911	0.02	2
1	A	1	MET	H3'	4.715	0.02	1
1	A	1	MET	H4'	4.07	0.02	1
1	A	1	MET	H41	7.006	0.02	2
1	A	1	MET	H42	8.105	0.02	2
1	A	1	MET	H5	5.903	0.02	1
1	A	1	MET	H5'	3.997	0.02	2
1	A	1	MET	H6	7.625	0.02	1
1	A	14	LEU	H1'	6.252	0.02	1
1	A	14	LEU	H2'	2.265	0.02	2
1	A	14	LEU	H3'	4.556	0.02	1
1	A	14	LEU	H4'	4.047	0.02	1
1	A	14	LEU	H5	5.74	0.02	1
1	A	14	LEU	H5'	4.145	0.02	2
1	A	14	LEU	H6	7.648	0.02	1
1	A	15	TRP	H1'	5.72	0.02	1
1	A	15	TRP	H2'	2.528	0.02	2
1	A	15	TRP	H3'	4.826	0.02	1
1	A	15	TRP	H4'	4.189	0.02	1
1	A	15	TRP	H8	7.863	0.02	1
1	A	26	GLU	H1'	6.158	0.02	1
1	A	26	GLU	H2'1	2.358	0.02	2
1	A	26	GLU	H2'2	2.605	0.02	2
1	A	26	GLU	H3'1	4.682	0.02	1
1	A	26	GLU	H4'	4.184	0.02	1
1	A	26	GLU	H8	7.94	0.02	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	88	-0.54 ± 0.12	Should be checked
$^{13}C_{\beta}$	83	-0.22 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	83	-0.56 ± 0.33	None needed (imprecise)

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 59%, i.e. 885 atoms were assigned a chemical shift out of a possible 1489. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	275/348 (79%)	141/141 (100%)	68/140 (49%)	66/67 (99%)
Sidechain	395/562 (70%)	280/359 (78%)	112/172 (65%)	3/31 (10%)
Aromatic	64/65 (98%)	32/32 (100%)	29/30 (97%)	3/3 (100%)
Sugar	105/312 (34%)	105/182 (58%)	0/130 (0%)	0/0 (—%)
Base	46/202 (23%)	46/124 (37%)	0/46 (0%)	0/32 (0%)
Overall	885/1489 (59%)	604/838 (72%)	209/518 (40%)	72/133 (54%)

7.1.4 Statistically unusual chemical shifts

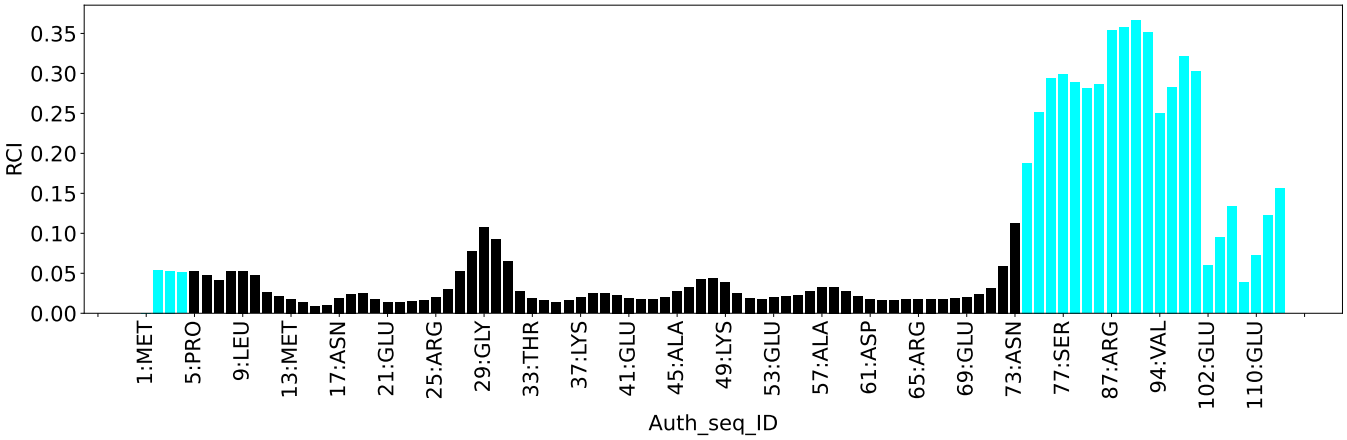
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	49	LYS	HE2	1.54	1.95 – 3.88	-7.1
1	A	49	LYS	HE3	1.83	1.92 – 3.89	-5.5
1	A	51	GLU	HG3	1.11	1.20 – 3.30	-5.4

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:


RCI
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1:MET
5:PRO
9:LEU
13:MET
17:ASN
21:GLU
25:ARG
29:GLY
33:THR
37:LYS
41:GLU
45:ALA
49:LYS
53:GLU
57:ALA
61:ASP
65:ARG
69:GLU
73:ASN
77:SER
87:ARG
94:VAL
102:GLU
110:GLU
Auth_seq_ID

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	1414
Intra-residue (|i-j|=0)	524
Sequential (|i-j|=1)	345
Medium range (|i-j|>1 and |i-j|<5)	235
Long range (|i-j|≥5)	136
Inter-chain	86
Hydrogen bond restraints	88
Disulfide bond restraints	0
Total dihedral-angle restraints	617
Number of unmapped restraints	80
Number of restraints per residue	14.7
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)	12.9	0.2
0.2-0.5 (Medium)	1.6	0.48
>0.5 (Large)	1.1	21.65

8.2.2 Average number of dihedral-angle violations per model (i)

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)	43.3	10.0
10.0-20.0 (Medium)	4.7	19.97
>20.0 (Large)	1.2	28.91

9 Distance violation analysis (i)

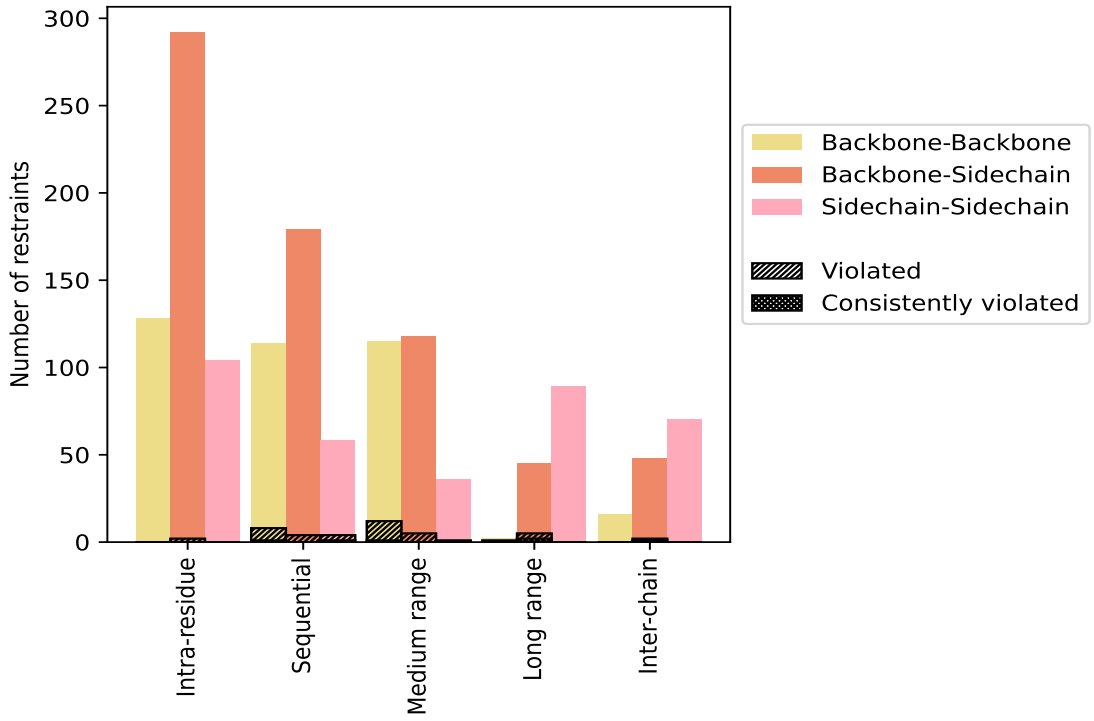
9.1 Summary of distance violations (i)

The following table shows the summary of distance violations in different restraint categories based on the sequence separation of the atoms involved. Each category is further sub-divided into three sub-categories based on the atoms involved. Violations less than 0.1 Å are not included in the statistics.

Restraints type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
Intra-residue (|i-j|=0)	524	37.1	2	0.4	0.1	0	0.0	0.0
Backbone-Backbone	128	9.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	292	20.7	2	0.7	0.1	0	0.0	0.0
Sidechain-Sidechain	104	7.4	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	345	24.4	16	4.6	1.1	2	0.6	0.1
Backbone-Backbone	114	8.1	8	7.0	0.6	1	0.9	0.1
Backbone-Sidechain	179	12.7	4	2.2	0.3	0	0.0	0.0
Sidechain-Sidechain	52	3.7	4	7.7	0.3	1	1.9	0.1
Medium range (|i-j|>1 & |i-j|<5)	235	16.6	14	6.0	1.0	1	0.4	0.1
Backbone-Backbone	115	8.1	12	10.4	0.8	1	0.9	0.1
Backbone-Sidechain	84	5.9	1	1.2	0.1	0	0.0	0.0
Sidechain-Sidechain	36	2.5	1	2.8	0.1	0	0.0	0.0
Long range (|i-j|≥5)	136	9.6	6	4.4	0.4	2	1.5	0.1
Backbone-Backbone	2	0.1	1	50.0	0.1	0	0.0	0.0
Backbone-Sidechain	45	3.2	5	11.1	0.4	2	4.4	0.1
Sidechain-Sidechain	89	6.3	0	0.0	0.0	0	0.0	0.0
Inter-chain	86	6.1	2	2.3	0.1	1	1.2	0.1
Backbone-Backbone	16	1.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	48	3.4	2	4.2	0.1	1	2.1	0.1
Sidechain-Sidechain	22	1.6	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	88	6.2	4	4.5	0.3	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	1414	100.0	44	3.1	3.1	6	0.4	0.4
Backbone-Backbone	375	26.5	21	5.6	1.5	2	0.5	0.1
Backbone-Sidechain	682	48.2	18	2.6	1.3	3	0.4	0.2
Sidechain-Sidechain	357	25.2	5	1.4	0.4	1	0.3	0.1

[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	0	5	6	2	1	14	1.61	20.56	5.26	0.14
2	1	6	9	2	1	19	1.16	19.53	4.33	0.14
3	1	5	7	3	2	18	1.23	19.6	4.46	0.12
4	1	8	8	3	1	21	1.1	20.42	4.32	0.12
5	1	6	9	3	1	20	1.22	21.65	4.69	0.12
6	1	7	7	3	1	19	1.19	19.94	4.42	0.13
7	1	6	9	3	1	20	1.18	20.64	4.47	0.12
8	1	7	6	2	1	17	1.3	19.91	4.65	0.13
9	0	8	6	3	2	19	1.17	19.81	4.39	0.11
10	0	7	6	3	1	17	1.25	18.94	4.42	0.14

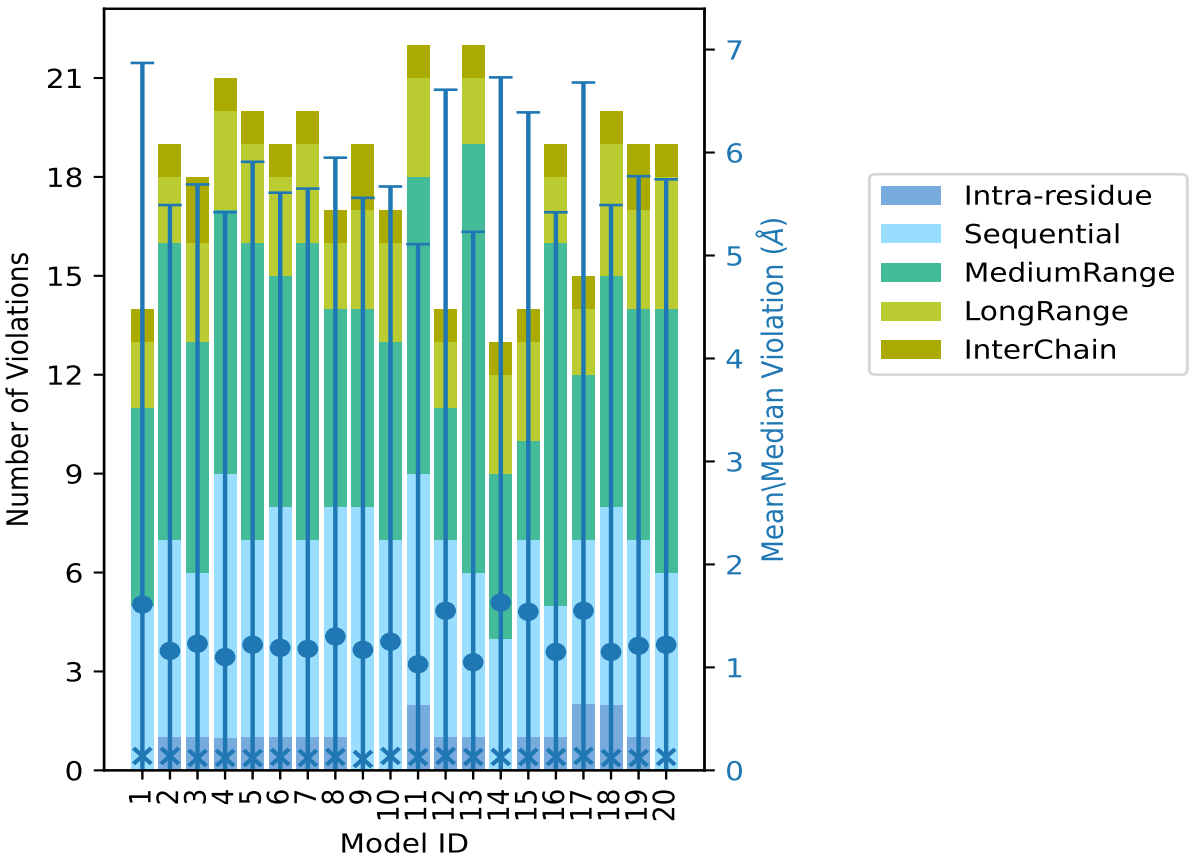
Continued on next page...

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				
11	2	7	9	3	1	22	1.03	19.73	4.08	0.13
12	1	6	4	2	1	14	1.55	19.78	5.06	0.14
13	1	5	13	2	1	22	1.05	20.23	4.18	0.13
14	0	4	5	3	1	13	1.63	19.28	5.1	0.13
15	1	6	3	3	1	14	1.54	19.02	4.85	0.13
16	1	4	11	2	1	19	1.15	19.25	4.27	0.13
17	2	5	5	2	1	15	1.55	20.75	5.13	0.14
18	2	6	7	4	1	20	1.15	20.05	4.34	0.12
19	1	6	7	3	2	19	1.21	20.57	4.56	0.12
20	0	6	8	4	1	19	1.22	20.39	4.52	0.13

[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

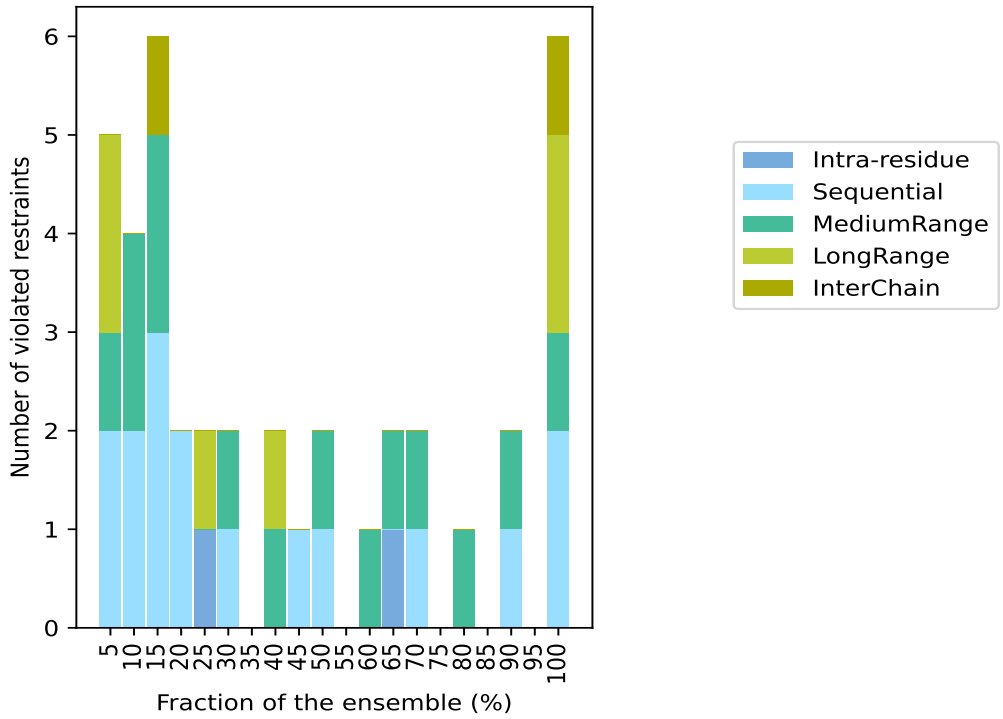
9.3 Distance violation statistics for the ensemble (i)

Violation analysis may find that some restraints are violated in few models and some are violated in most of models. The following table provides this information as number of violated restraints for a given fraction of the ensemble. In total, 1286(IR:522, SQ:329, MR:221, LR:130, IC:84) 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]	%
0	2	1	2	0	5	1	5.0
0	2	2	0	0	4	2	10.0
0	3	2	0	1	6	3	15.0
0	2	0	0	0	2	4	20.0
1	0	0	1	0	2	5	25.0
0	1	1	0	0	2	6	30.0
0	0	0	0	0	0	7	35.0
0	0	1	1	0	2	8	40.0
0	1	0	0	0	1	9	45.0
0	1	1	0	0	2	10	50.0
0	0	0	0	0	0	11	55.0
0	0	1	0	0	1	12	60.0
1	0	1	0	0	2	13	65.0
0	1	1	0	0	2	14	70.0
0	0	0	0	0	0	15	75.0
0	0	1	0	0	1	16	80.0
0	0	0	0	0	0	17	85.0
0	1	1	0	0	2	18	90.0
0	0	0	0	0	0	19	95.0
0	2	1	2	1	6	20	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

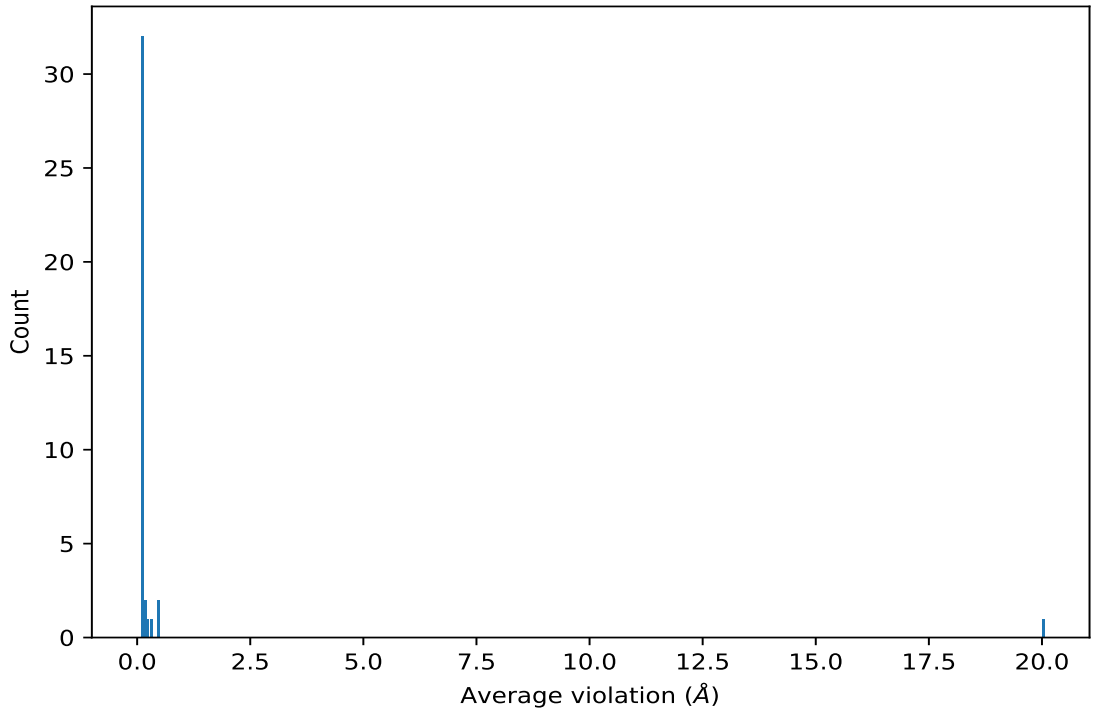
9.3.1 Bar graph : Distance violation statistics for the ensemble



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.

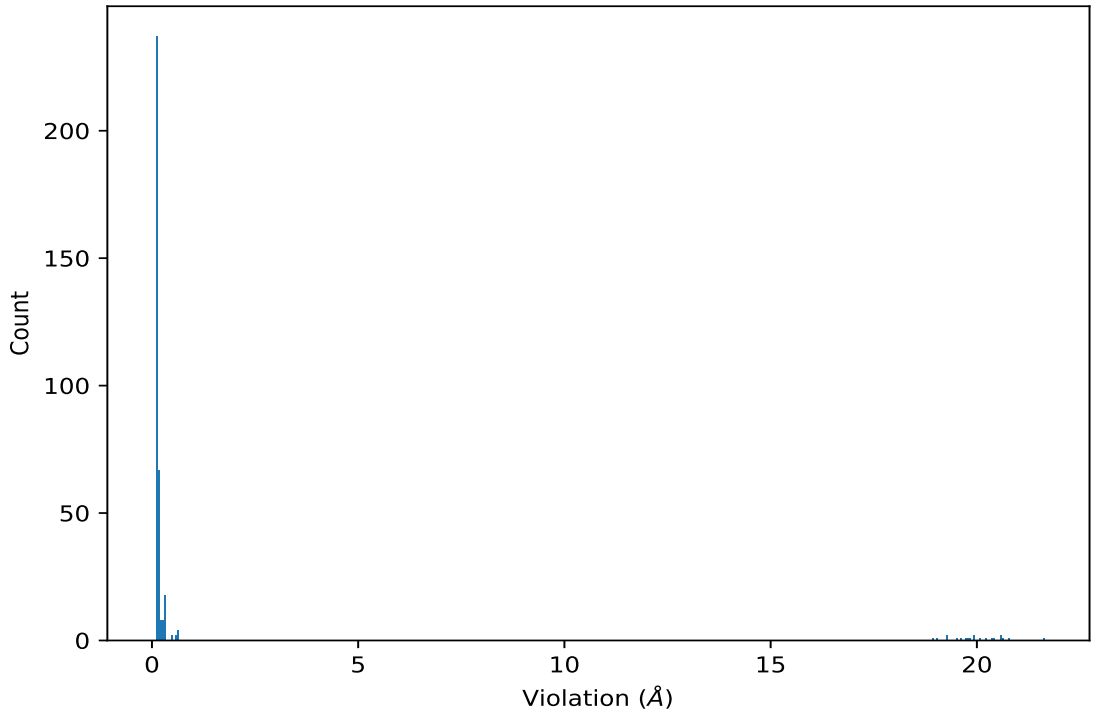
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	20	20.0	0.65	19.92
(1,987)	2:11:B:DA:H2	2:12:B:DG:H8	20	0.31	0.02	0.31
(1,525)	1:46:A:MET:H	1:47:A:LYS:H	20	0.2	0.06	0.2
(1,117)	1:12:A:PHE:HZ	1:43:A:TRP:HA	20	0.15	0.02	0.15
(1,583)	1:51:A:GLU:HA	1:54:A:ALA:H	20	0.15	0.02	0.15
(1,90)	1:11:A:ALA:HA	1:43:A:TRP:HH2	20	0.14	0.02	0.15
(1,834)	1:73:A:ASN:HA	1:74:A:GLY:H	18	0.16	0.03	0.15
(1,109)	1:12:A:PHE:HA	1:15:A:TRP:H	18	0.12	0.01	0.12
(1,458)	1:41:A:GLU:HA	1:44:A:ARG:H	16	0.12	0.02	0.12
(1,938)	2:5:B:DA:H8	2:6:B:DT:H6	14	0.12	0.01	0.12

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

9.5 All violated distance restraints (i)

9.5.1 Histogram : Distribution of distance violations (i)

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



9.5.2 Table : All distance violations (i)

The following table provides the 10 worst performing restraints, sorted by the violation value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint. Rows with same key represent combinatorial or ambiguous restraints and are counted as a single restraint.

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	5	21.65
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	17	20.75
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	7	20.64
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	19	20.57
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	1	20.56
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	4	20.42
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	20	20.39
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	13	20.23
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	18	20.05
(2,150)	3:25:C:DC:H6	1:26:A:GLU:H	6	19.94

10 Dihedral-angle violation analysis

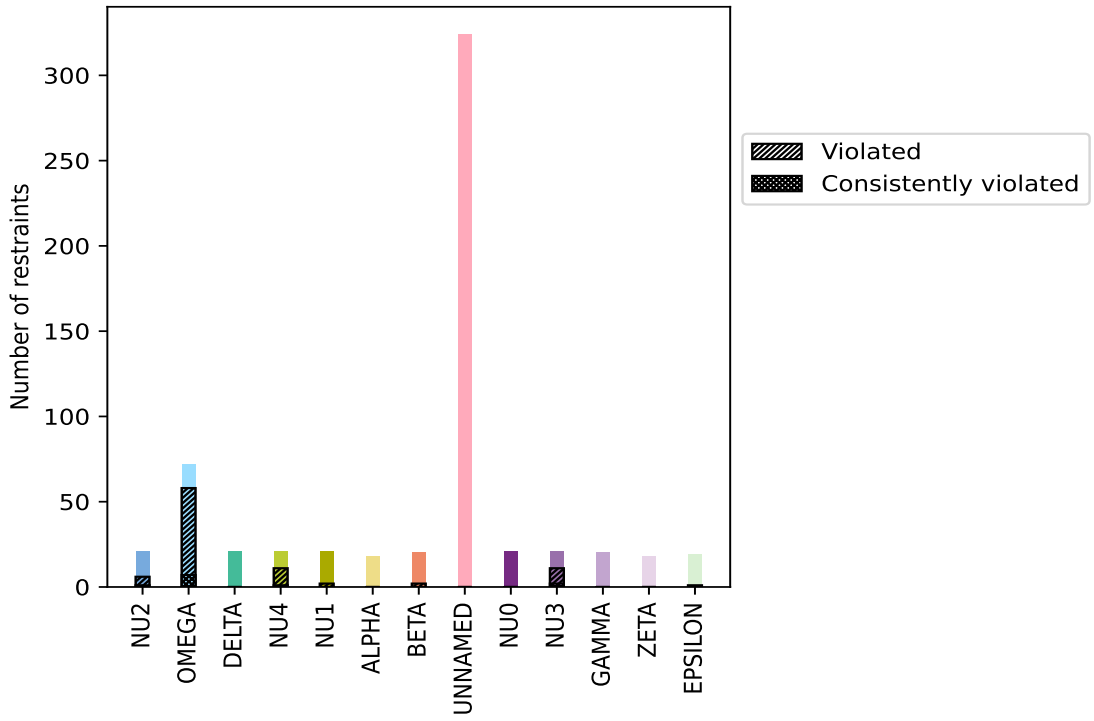
10.1 Summary of dihedral-angle violations

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]
NU2	21	3.4	6	28.6	1.0	1	4.8	0.2
OMEGA	72	11.7	58	80.6	9.4	7	9.7	1.1
DELTA	21	3.4	0	0.0	0.0	0	0.0	0.0
NU4	21	3.4	11	52.4	1.8	1	4.8	0.2
NU1	21	3.4	2	9.5	0.3	0	0.0	0.0
ALPHA	18	2.9	0	0.0	0.0	0	0.0	0.0
BETA	20	3.2	2	10.0	0.3	0	0.0	0.0
UNNAMED	324	52.5	0	0.0	0.0	0	0.0	0.0
NU0	21	3.4	0	0.0	0.0	0	0.0	0.0
NU3	21	3.4	11	52.4	1.8	2	9.5	0.3
GAMMA	20	3.2	0	0.0	0.0	0	0.0	0.0
ZETA	18	2.9	0	0.0	0.0	0	0.0	0.0
EPSILON	19	3.1	1	5.3	0.2	0	0.0	0.0
Total	617	100.0	91	14.7	14.7	11	1.8	1.8

[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)



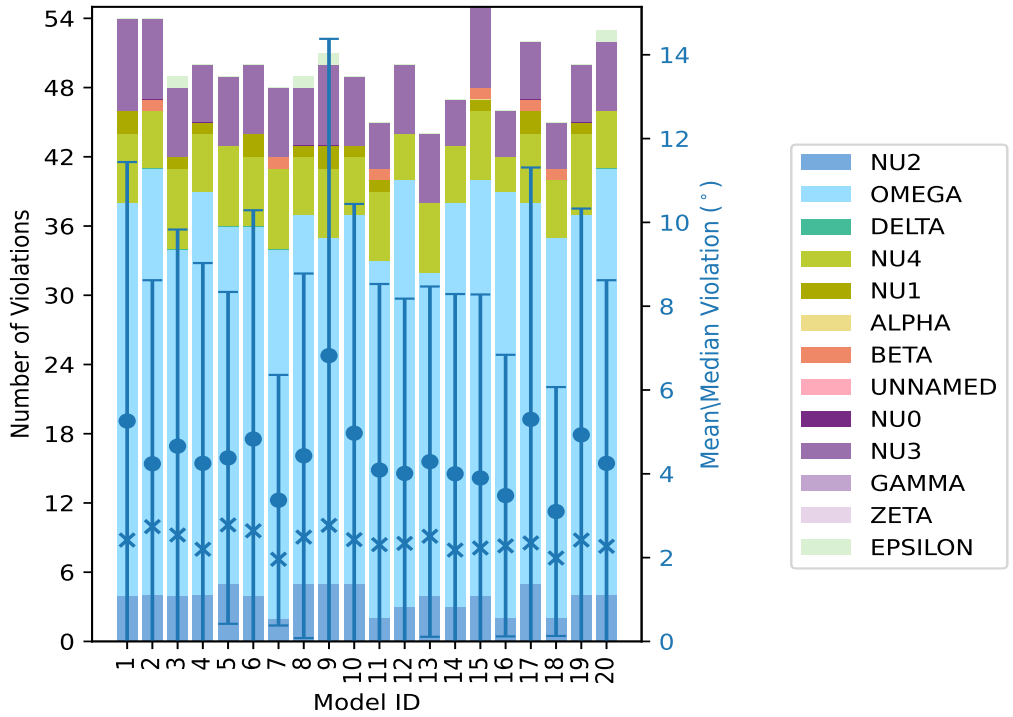
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

10.2 Dihedral-angle violation statistics for each model (i)

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
	NU2	OMEGA	DELTA	NU4	NU1	ALPHA	BETA	UNNAMED	NU0	NU3	GAMMA	ZETA	EPSILON	Total	
1	4	34	0	6	2	0	0	0	0	8	0	0	0	54	5.26
2	4	37	0	5	0	0	1	0	0	7	0	0	0	54	4.24
3	4	30	0	7	1	0	0	0	0	6	0	0	1	49	4.66
4	4	35	0	5	1	0	0	0	0	5	0	0	0	50	4.25
5	5	31	0	7	0	0	0	0	0	6	0	0	0	49	4.38
6	4	32	0	6	2	0	0	0	0	6	0	0	0	50	4.83
7	2	32	0	7	0	0	1	0	0	6	0	0	0	48	3.37
8	5	32	0	5	1	0	0	0	0	5	0	0	1	49	4.43
9	5	30	0	6	2	0	0	0	0	7	0	0	1	51	6.82
10	5	32	0	5	1	0	0	0	0	6	0	0	0	49	4.97
11	2	31	0	6	1	0	1	0	0	4	0	0	0	45	4.09
12	3	37	0	4	0	0	0	0	0	6	0	0	0	50	4.01
13	4	28	0	6	0	0	0	0	0	6	0	0	0	44	4.29
14	3	35	0	5	0	0	0	0	0	4	0	0	0	47	4.0
15	4	36	0	6	1	0	1	0	0	7	0	0	0	55	3.9
16	2	37	0	3	0	0	0	0	0	4	0	0	0	46	3.48
17	5	33	0	6	2	0	1	0	0	5	0	0	0	52	5.3
18	2	33	0	5	0	0	1	0	0	4	0	0	0	45	3.1
19	4	33	0	7	1	0	0	0	0	5	0	0	0	50	4.93
20	4	37	0	5	0	0	0	0	0	6	0	0	1	53	4.25

10.2.1 Bar graph : Dihedral 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

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.

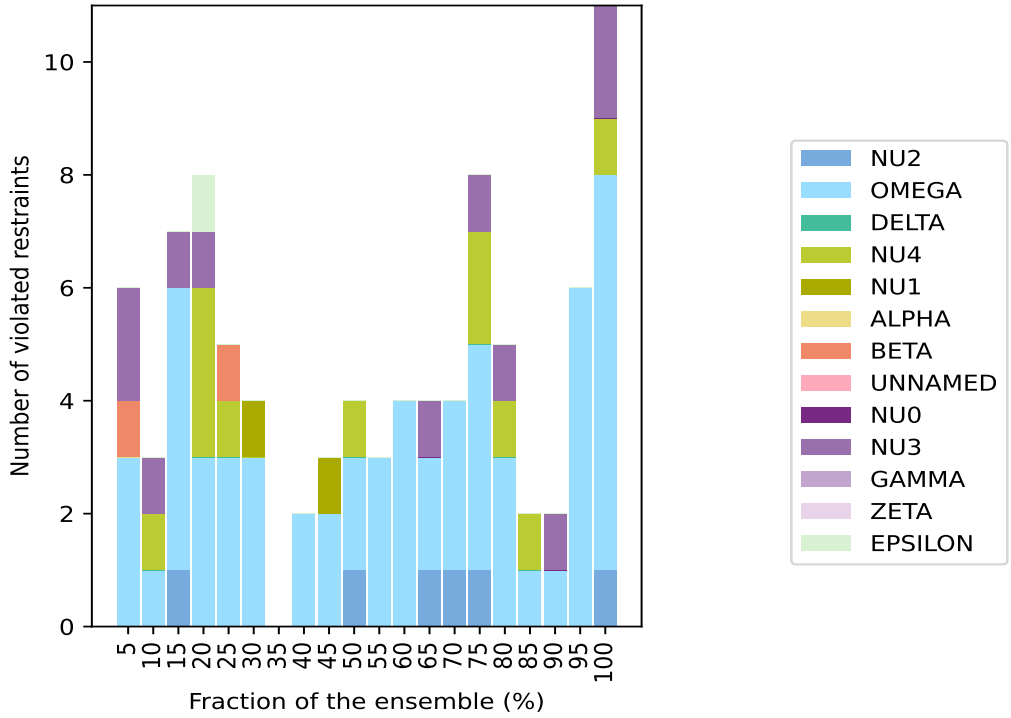
Number of violated restraints												
NU2	OMEGA	DELTA	NU4	NU1	ALPHA	BETA	UNNAMED	NU0	NU3	GAMMA	ZETA	EPSILON
0	3	0	0	0	0	1	0	0	2	0	0	0
0	1	0	1	0	0	0	0	0	1	0	0	0
1	5	0	0	0	0	0	0	0	1	0	0	0
0	3	0	3	0	0	0	0	0	1	0	0	1
0	3	0	1	0	0	1	0	0	0	0	0	0
0	3	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	2	0	0	0	0	0	0	0	0	0	0	0
0	2	0	0	1	0	0	0	0	0	0	0	0
1	2	0	1	0	0	0	0	0	0	0	0	0
0	3	0	0	0	0	0	0	0	0	0	0	0
0	4	0	0	0	0	0	0	0	0	0	0	0
1	2	0	0	0	0	0	0	0	1	0	0	0
1	3	0	0	0	0	0	0	0	0	0	0	0
1	4	0	2	0	0	0	0	0	1	0	0	0

Continued from previous page...

Number of violated restraints												
NU2	OMEGA	DELTA	NU4	NU1	ALPHA	BETA	UNNAMED	NU0	NU3	GAMMA	ZETA	EPSILON
0	3	0	1	0	0	0	0	0	1	0	0	0
0	1	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	1	0	0	0
0	6	0	0	0	0	0	0	0	0	0	0	0
1	7	0	1	0	0	0	0	0	2	0	0	0

[1] Number of models with violations

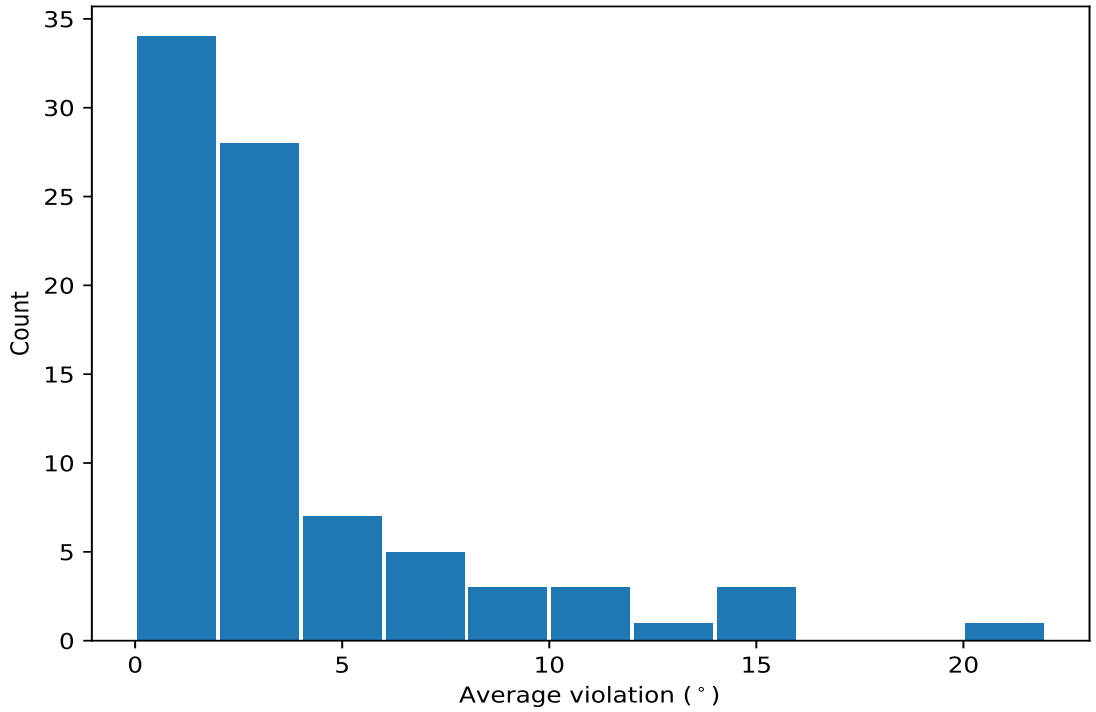
10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble



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 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.

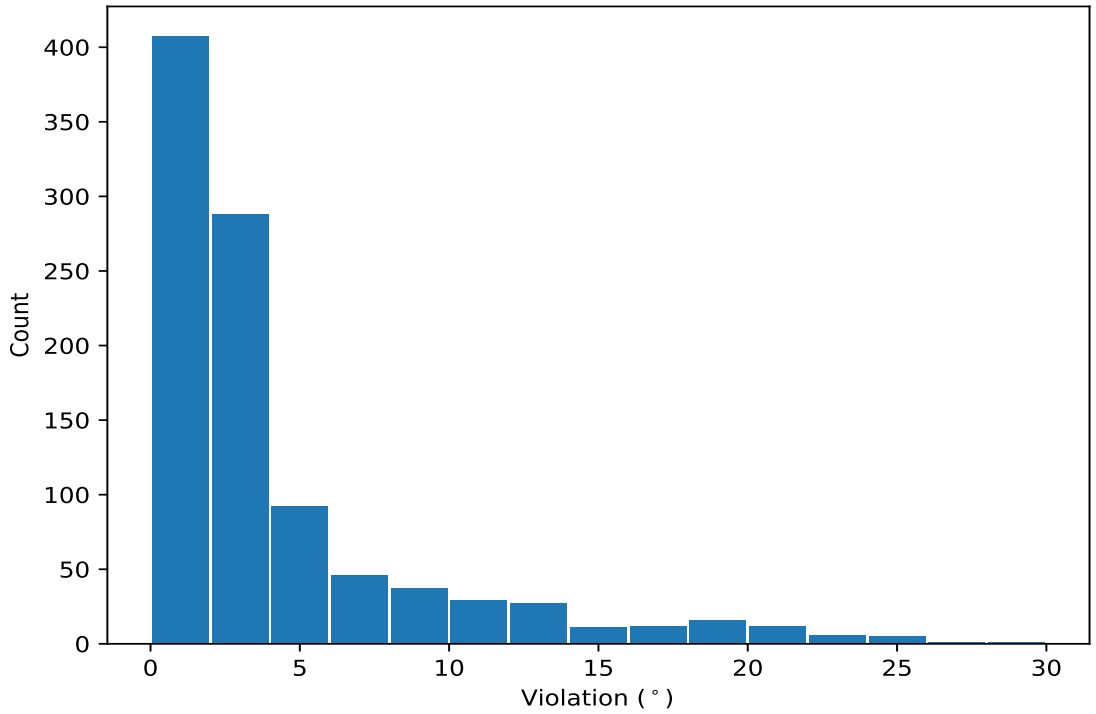
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	20	20.01	4.48	20.17
(1,140)	2:6:B:DT:C1'	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	20	15.95	5.19	16.26
(1,186)	3:18:C:DT:C2'	3:18:C:DT:C3'	3:18:C:DT:C4'	3:18:C:DT:O4'	20	9.83	4.64	9.33
(1,142)	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	2:6:B:DT:C1'	20	7.86	2.06	8.06
(1,607)	1:63:A:TYR:CA	1:63:A:TYR:N	1:62:A:ASP:C	1:62:A:ASP:CA	20	3.44	0.28	3.47
(1,579)	1:35:A:VAL:CA	1:35:A:VAL:N	1:34:A:GLU:C	1:34:A:GLU:CA	20	2.82	0.47	2.76
(1,578)	1:34:A:GLU:CA	1:34:A:GLU:N	1:33:A:THR:C	1:33:A:THR:CA	20	2.36	0.6	2.29
(1,595)	1:51:A:GLU:CA	1:51:A:GLU:N	1:50:A:SER:C	1:50:A:SER:CA	20	2.31	0.44	2.25
(1,594)	1:50:A:SER:CA	1:50:A:SER:N	1:49:A:LYS:C	1:49:A:LYS:CA	20	2.28	0.77	2.3
(1,563)	1:19:A:ALA:CA	1:19:A:ALA:N	1:18:A:SER:C	1:18:A:SER:CA	20	1.85	0.33	1.82

[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 violations for 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 (°)
(1,190)	3:19:C:DC:C1'	3:19:C:DC:C2'	3:19:C:DC:C3'	3:19:C:DC:C4'	9	28.91
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	9	26.75
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	17	25.92
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	1	25.49
(1,191)	3:19:C:DC:C2'	3:19:C:DC:C3'	3:19:C:DC:C4'	3:19:C:DC:O4'	9	24.52
(1,140)	2:6:B:DT:C1'	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	9	24.4
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	3	24.21
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	4	23.85
(1,141)	2:6:B:DT:C2'	2:6:B:DT:C3'	2:6:B:DT:C4'	2:6:B:DT:O4'	19	23.81
(1,190)	3:19:C:DC:C1'	3:19:C:DC:C2'	3:19:C:DC:C3'	3:19:C:DC:C4'	1	23.59