wwPDB NMR Structure Validation Summary Report

Jun 5, 2023 – 08:28 AM EDT

PDB ID	:	2M45
BMRB ID	:	18986
Title	:	NMR solution structure of the C-terminus of the minichromosome maintenance protein MCM from Sulfolobus solfataricus
Authors	:	Wiedemann, C.; Ohlenschlager, O.; Medagli, B.; Onesti, S.; Gorlach, M.
Deposited on	:	2013-01-29

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

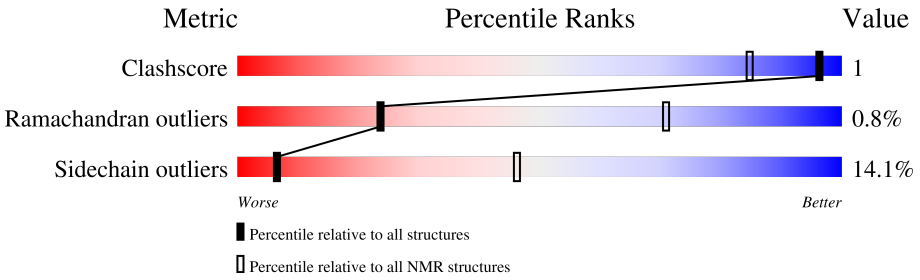
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 94%.

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	87	79% 7% 14%

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:610-A:618 (9)	0.22	4
2	A:621-A:686 (66)	0.65	6

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1384 atoms, of which 714 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Minichromosome maintenance protein MCM.

Mol	Chain	Residues	Atoms						Trace
1	A	87	Total	C	H	N	O	S	0
			1384	419	714	112	132	7	

There are 5 discrepancies between the modelled and reference sequences:

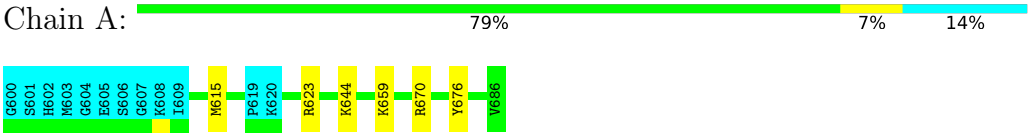
Chain	Residue	Modelled	Actual	Comment	Reference
A	600	GLY	-	expression tag	UNP Q9UXG1
A	601	SER	-	expression tag	UNP Q9UXG1
A	602	HIS	-	expression tag	UNP Q9UXG1
A	603	MET	-	expression tag	UNP Q9UXG1
A	604	GLY	-	expression tag	UNP Q9UXG1

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.

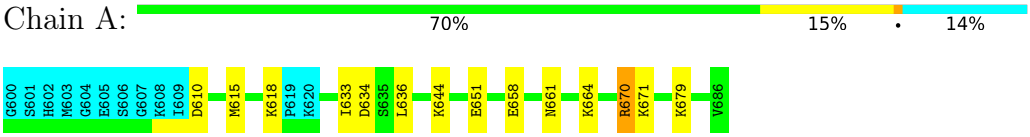
• Molecule 1: Minichromosome maintenance protein MCM



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

The representative model is number 6. Colouring as in section 4.1 above.

• Molecule 1: Minichromosome maintenance protein MCM



5 Refinement protocol and experimental data overview

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
CYANA	structure solution	3
OPAL	refinement	
CYANA	refinement	

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	1118
Number of shifts mapped to atoms	1118
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	94%

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	0.53±0.01	0±0/588 (0.0± 0.0%)	0.99±0.05	1±1/781 (0.1± 0.1%)
All	All	0.53	0/11760 (0.0%)	0.99	20/15620 (0.1%)

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	1.6±1.1
All	All	0	32

There are no bond-length outliers.

5 of 13 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	623	ARG	NE-CZ-NH2	-8.71	115.94	120.30	15	2
1	A	676	TYR	CB-CG-CD2	-7.64	116.42	121.00	5	2
1	A	623	ARG	NE-CZ-NH1	6.73	123.67	120.30	7	3
1	A	676	TYR	CA-CB-CG	5.87	124.55	113.40	5	1
1	A	670	ARG	NE-CZ-NH2	-5.86	117.37	120.30	20	2

There are no chirality outliers.

All 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	670	ARG	Sidechain	12
1	A	623	ARG	Sidechain	7
1	A	676	TYR	Sidechain	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	683	TYR	Sidechain	6

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	586	627	627	1±1
All	All	11720	12540	12540	19

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 1.

5 of 18 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:626:MET:CE	1:A:665:LEU:HD11	0.60	2.25	14	1
1:A:626:MET:HA	1:A:629:ILE:HG13	0.54	1.78	9	1
1:A:632:ILE:HG21	1:A:651:GLU:OE1	0.53	2.03	14	1
1:A:625:LYS:O	1:A:629:ILE:HD12	0.52	2.05	13	1
1:A:632:ILE:HD13	1:A:651:GLU:OE2	0.49	2.08	14	1

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	74/87 (85%)	67±2 (91±3%)	6±2 (8±3%)	1±1 (1±1%)	24	71
All	All	1480/1740 (85%)	1349 (91%)	119 (8%)	12 (1%)	24	71

5 of 8 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	681	GLU	3
1	A	678	ALA	2
1	A	659	LYS	2
1	A	640	SER	1
1	A	617	GLY	1

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	67/76 (88%)	58±2 (86±3%)	9±2 (14±3%)	6	46
All	All	1340/1520 (88%)	1151 (86%)	189 (14%)	6	46

5 of 45 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	670	ARG	13
1	A	644	LYS	12
1	A	615	MET	11
1	A	610	ASP	9
1	A	613	THR	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 monosaccharides 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 94% for the well-defined parts and 94% 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	1118
Number of shifts mapped to atoms	1118
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following errors were found when reading this chemical shift list.

- Chemical shift has been reported more than once. First 5 (of 0) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	636	LEU	HD12	0.736	0.020	2
1	A	636	LEU	HD13	0.736	0.020	2
1	A	649	LEU	HD12	0.827	0.020	2
1	A	649	LEU	HD13	0.827	0.020	2
1	A	665	LEU	HD12	0.716	0.020	2
1	A	665	LEU	HD13	0.716	0.020	2
1	A	666	LEU	HD12	0.253	0.020	2
1	A	666	LEU	HD13	0.253	0.020	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	87	-0.40 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	81	0.07 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	83	-0.29 ± 0.10	None needed (< 0.5 ppm)
^{15}N	82	0.21 ± 0.22	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 94%, i.e. 982 atoms were assigned a chemical shift out of a possible 1041. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	374/376 (99%)	152/152 (100%)	148/150 (99%)	74/74 (100%)
Sidechain	592/647 (91%)	404/420 (96%)	188/206 (91%)	0/21 (0%)
Aromatic	16/18 (89%)	8/8 (100%)	8/10 (80%)	0/0 (—%)
Overall	982/1041 (94%)	564/580 (97%)	344/366 (94%)	74/95 (78%)

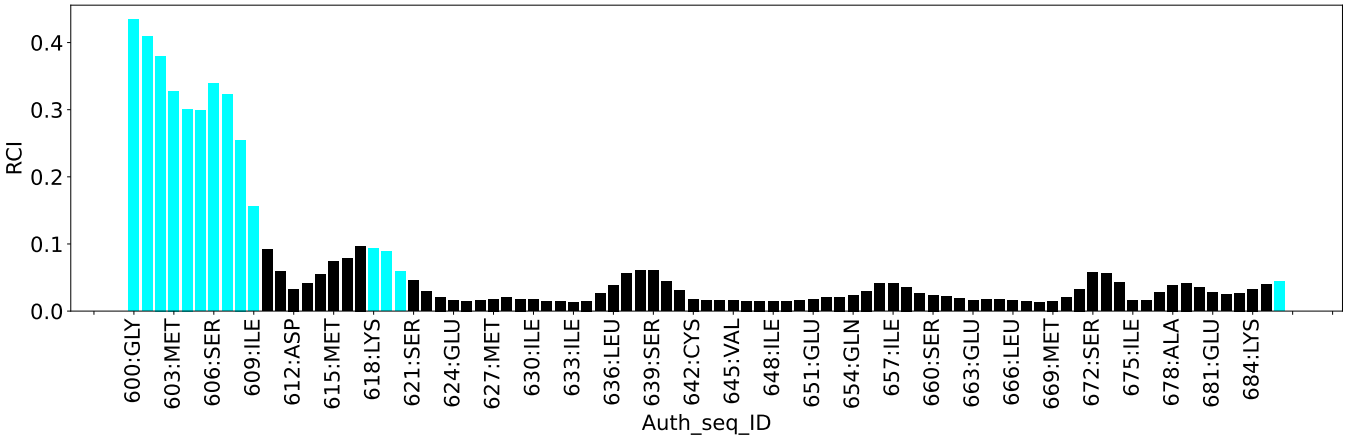
7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

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
0.4
0.3
0.2
0.1
0.0
600:GLY
603:MET
606:SER
609:ILE
612:ASP
615:MET
618:LYS
621:SER
624:GLU
627:MET
630:ILE
633:ILE
636:LEU
639:SER
642:CYS
645:VAL
648:ILE
651:GLU
654:GLN
657:ILE
660:SER
663:GLU
666:LEU
669:MET
672:SER
675:ILE
678:ALA
681:GLU
684:LYS
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	1755
Intra-residue (|i-j|=0)	234
Sequential (|i-j|=1)	476
Medium range (|i-j|>1 and |i-j|<5)	472
Long range (|i-j|≥5)	493
Inter-chain	0
Hydrogen bond restraints	80
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	20.2
Number of long range restraints per residue[1]	5.8

[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)	0.1	0.11
0.2-0.5 (Medium)	None	None
>0.5 (Large)	None	None

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

Dihedral-angle violations less than 1° are not included in the calculation.There are no dihedral-angle violations

9 Distance violation analysis (i)

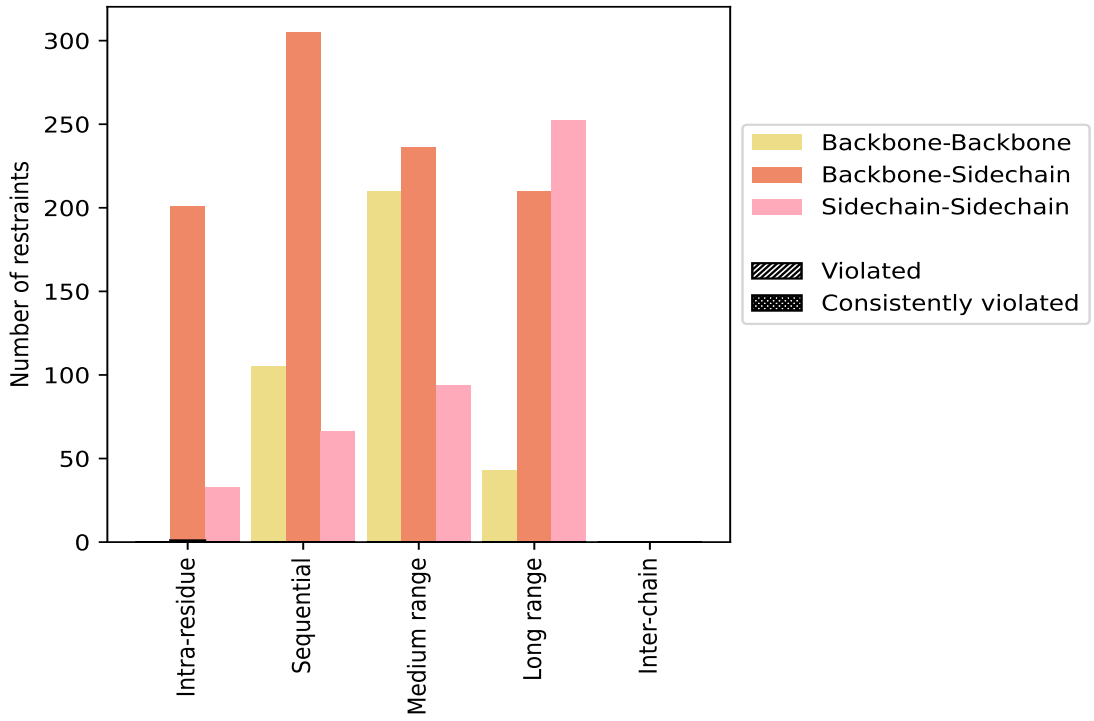
9.1 Summary of distance violations (i)

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

Restraints type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
Intra-residue (|i-j|=0)	234	13.3	1	0.4	0.1	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	201	11.5	1	0.5	0.1	0	0.0	0.0
Sidechain-Sidechain	33	1.9	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	476	27.1	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	105	6.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	305	17.4	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	66	3.8	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	472	26.9	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	142	8.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	236	13.4	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	94	5.4	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	493	28.1	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	31	1.8	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	210	12.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	252	14.4	0	0.0	0.0	0	0.0	0.0
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	80	4.6	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	1755	100.0	1	0.1	0.1	0	0.0	0.0
Backbone-Backbone	358	20.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	952	54.2	1	0.1	0.1	0	0.0	0.0
Sidechain-Sidechain	445	25.4	0	0.0	0.0	0	0.0	0.0

[1] percentage calculated with respect to the total number of distance restraints, [2] percentage calculated with respect to the number of restraints in a particular restraint category, [3] violated in at least one model, [4] violated in all the models

9.1.1 Bar chart : Distribution of distance restraints and violations



Violated and consistently violated restraints are shown using different hatch patterns in their respective categories. The hydrogen bonds and disulfied bonds are counted in their appropriate category on the x-axis

9.2 Distance violation statistics for each model

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	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	0.0
3	0	0	0	0	0	0	0.0	0.0	0.0	0.0
4	0	0	0	0	0	0	0.0	0.0	0.0	0.0
5	0	0	0	0	0	0	0.0	0.0	0.0	0.0
6	0	0	0	0	0	0	0.0	0.0	0.0	0.0
7	0	0	0	0	0	0	0.0	0.0	0.0	0.0
8	0	0	0	0	0	0	0.0	0.0	0.0	0.0
9	0	0	0	0	0	0	0.0	0.0	0.0	0.0
10	0	0	0	0	0	0	0.0	0.0	0.0	0.0
11	0	0	0	0	0	0	0.0	0.0	0.0	0.0

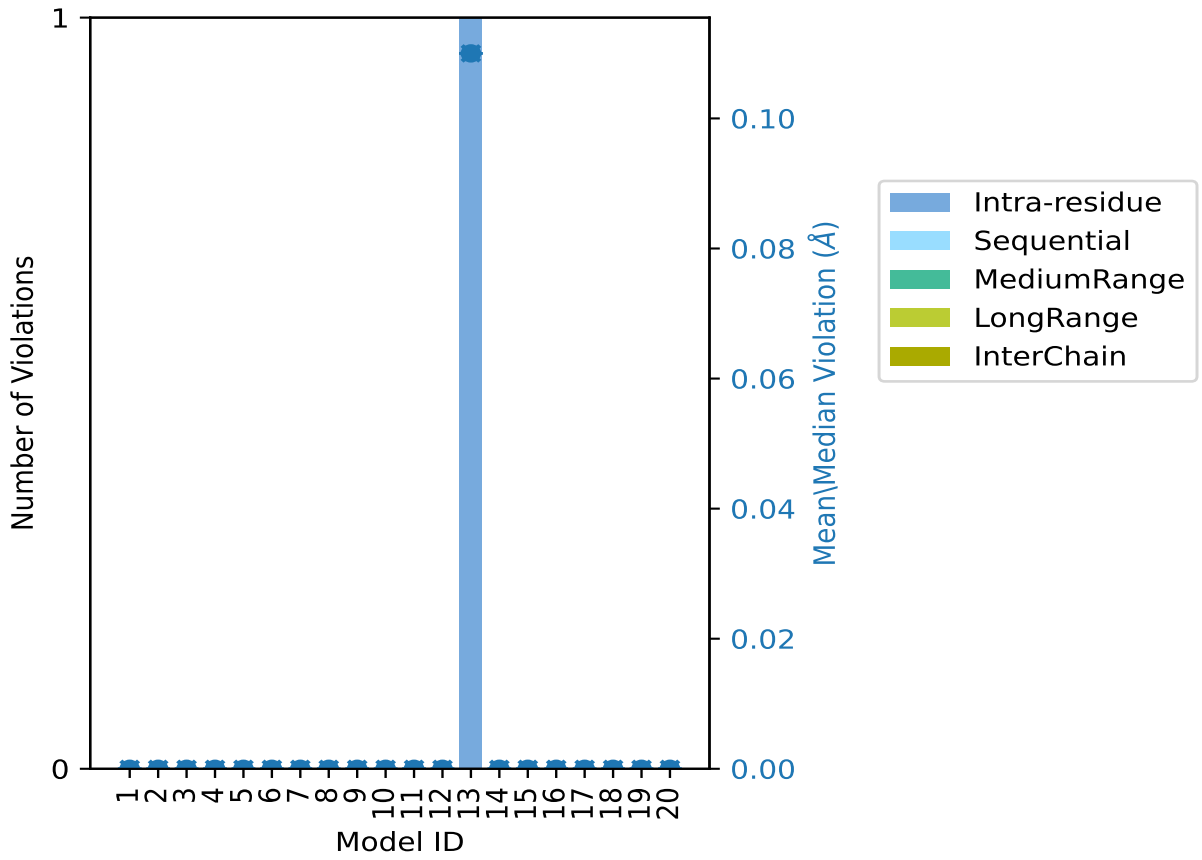
Continued on next page...

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	0	0	0	0	0	0	0.0	0.0	0.0	0.0
13	1	0	0	0	0	1	0.11	0.11	0.0	0.11
14	0	0	0	0	0	0	0.0	0.0	0.0	0.0
15	0	0	0	0	0	0	0.0	0.0	0.0	0.0
16	0	0	0	0	0	0	0.0	0.0	0.0	0.0
17	0	0	0	0	0	0	0.0	0.0	0.0	0.0
18	0	0	0	0	0	0	0.0	0.0	0.0	0.0
19	0	0	0	0	0	0	0.0	0.0	0.0	0.0
20	0	0	0	0	0	0	0.0	0.0	0.0	0.0

[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

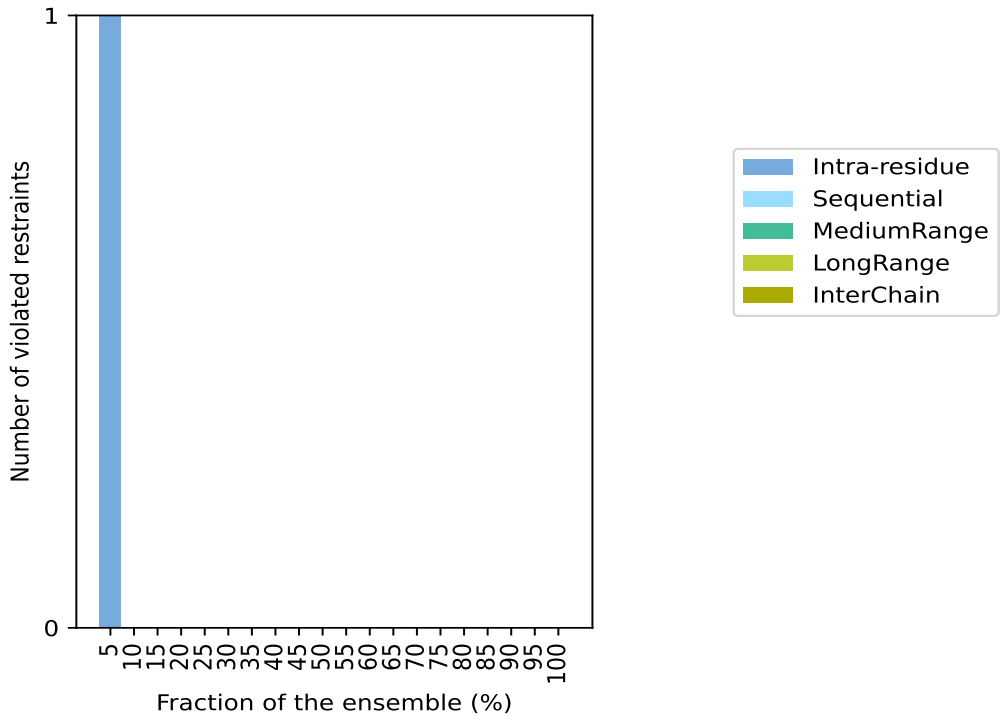
9.3 Distance violation statistics for the ensemble (i)

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, 1674(IR:233, SQ:476, MR:472, LR:493, 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]	%
1	0	0	0	0	1	1	5.0
0	0	0	0	0	0	2	10.0
0	0	0	0	0	0	3	15.0
0	0	0	0	0	0	4	20.0
0	0	0	0	0	0	5	25.0
0	0	0	0	0	0	6	30.0
0	0	0	0	0	0	7	35.0
0	0	0	0	0	0	8	40.0
0	0	0	0	0	0	9	45.0
0	0	0	0	0	0	10	50.0
0	0	0	0	0	0	11	55.0
0	0	0	0	0	0	12	60.0
0	0	0	0	0	0	13	65.0
0	0	0	0	0	0	14	70.0
0	0	0	0	0	0	15	75.0
0	0	0	0	0	0	16	80.0
0	0	0	0	0	0	17	85.0
0	0	0	0	0	0	18	90.0
0	0	0	0	0	0	19	95.0
0	0	0	0	0	0	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

No violations found

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.
Data insufficient to plot histogram

9.5.2 Table : All distance violations

The following table lists the absolute value of the violation for each restraint in the ensemble sorted by its 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)
(1,14)	1:A:609:ILE:H	1:A:609:ILE:HB	13	0.11

10 Dihedral-angle violation analysis (i)

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