wwPDB NMR Structure Validation Summary Report

Jun 6, 2023 – 05:02 pm BST

PDB ID	:	7AEP
BMRB ID	:	34560
Title	:	Solution structure of U1-A RRM2 (190-282)
Authors	:	Campagne, S.; Allain, F.H.
Deposited on	:	2020-09-18

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

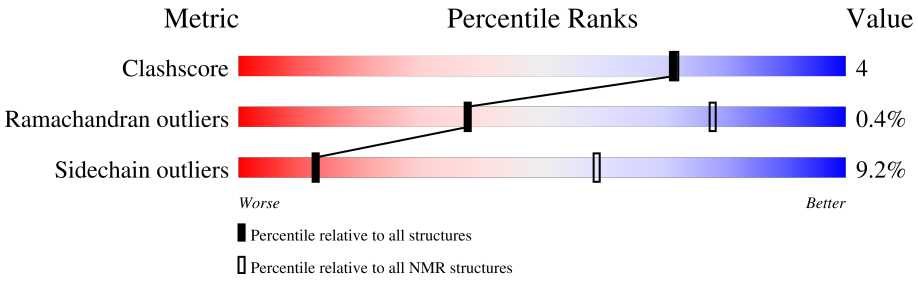
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal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

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 91%.

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	126	48% 6% • 18% 26%

2 Ensemble composition and analysis (i)

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:208-A:240, A:244-A:280 (70)	0.14	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 1 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1463 atoms, of which 729 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called U1 small nuclear ribonucleoprotein A.

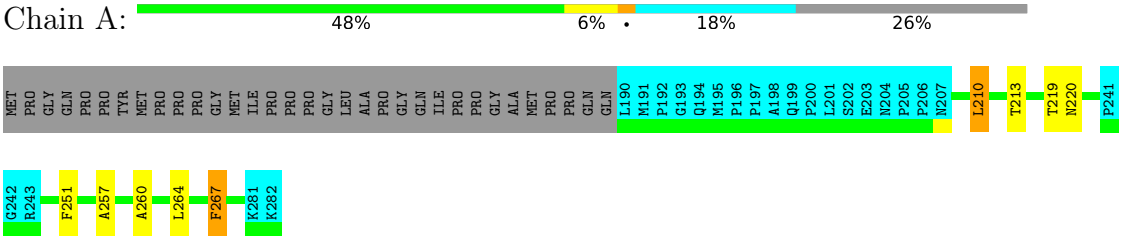
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	93	1463	470	729	125	134	5	0

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.

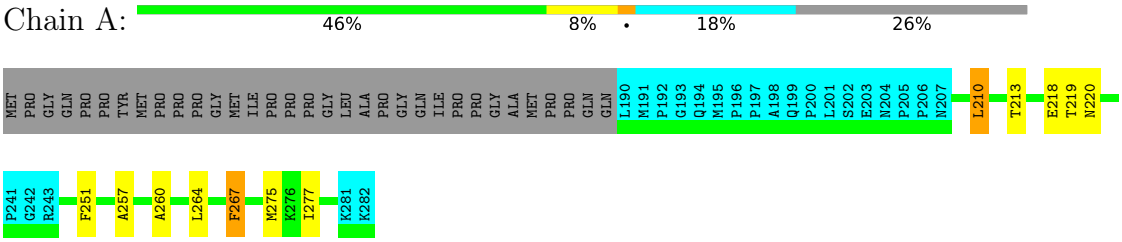
• Molecule 1: U1 small nuclear ribonucleoprotein A



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: U1 small nuclear ribonucleoprotein A



5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing.*

Of the 50 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
Amber	refinement	20
CYANA	structure calculation	3.94

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	1322
Number of shifts mapped to atoms	1090
Number of unparsed shifts	0
Number of shifts with mapping errors	232
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

6 Model quality

6.1 Standard geometry

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	561	551	551	4±1
All	All	11220	11020	11020	78

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 4.

5 of 17 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:228:PHE:CE1	1:A:264:LEU:HD12	0.56	2.35	12	1
1:A:251:PHE:CZ	1:A:260:ALA:HB2	0.55	2.36	12	1
1:A:251:PHE:CE1	1:A:260:ALA:HB2	0.55	2.37	11	6
1:A:251:PHE:CE2	1:A:260:ALA:HB2	0.54	2.38	13	13
1:A:215:LEU:HD21	1:A:275:MET:SD	0.53	2.43	13	1

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	70/126 (56%)	62±1 (88±1%)	8±1 (11±1%)	0±1 (0±1%)	38	78
All	All	1400/2520 (56%)	1237 (88%)	157 (11%)	6 (0%)	38	78

All 4 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	275	MET	2
1	A	232	PRO	2
1	A	208	HIS	1
1	A	265	GLN	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	60/106 (57%)	54±1 (91±2%)	6±1 (9±2%)	13	59
All	All	1200/2120 (57%)	1090 (91%)	110 (9%)	13	59

5 of 9 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	210	LEU	20
1	A	220	ASN	20
1	A	219	THR	19
1	A	267	PHE	19
1	A	213	THR	19

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 (i)

There are no ligands in this entry.

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 91% for the well-defined parts and 85% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *U1A_RRM2_link.str*

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	1322
Number of shifts mapped to atoms	1090
Number of unparsed shifts	0
Number of shifts with mapping errors	232
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

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 232) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	158	PRO	C	176.906	0.3	1
1	A	158	PRO	CA	56.564	0.3	1
1	A	158	PRO	CB	30.865	0.3	1
1	A	158	PRO	CG	27.161	0.3	1
1	A	158	PRO	CD	49.922	0.3	1
1	A	158	PRO	HA	4.335	0.020	1
1	A	159	GLY	C	173.397	0.3	1
1	A	159	GLY	CA	45.361	0.3	1
1	A	159	GLY	H	8.453	0.020	1
1	A	159	GLY	HA2	3.99	0.020	1
1	A	159	GLY	HA3	3.99	0.020	1
1	A	159	GLY	N	109.963	0.3	1
1	A	160	GLN	CA	53.732	0.3	1
1	A	160	GLN	CB	29.648	0.3	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	160	GLN	CG	34.014	0.3	1
1	A	160	GLN	H	8.344	0.020	1
1	A	160	GLN	HA	4.363	0.020	1
1	A	160	GLN	HB2	2.028	0.020	1
1	A	160	GLN	HB3	2.028	0.020	1
1	A	160	GLN	HG2	2.389	0.020	1
1	A	160	GLN	HG3	2.389	0.020	1
1	A	160	GLN	N	121.113	0.3	1
1	A	162	PRO	C	176.239	0.3	1
1	A	162	PRO	CA	63.039	0.3	1
1	A	162	PRO	CB	31.675	0.3	1
1	A	162	PRO	CG	27.014	0.3	1
1	A	162	PRO	CD	50.17	0.3	1
1	A	162	PRO	HA	4.47	0.020	1
1	A	163	TYR	C	174.905	0.3	1
1	A	163	TYR	CA	57.885	0.3	1
1	A	163	TYR	CB	38.789	0.3	1
1	A	163	TYR	CD1	133.272	0.3	1
1	A	163	TYR	CE1	118.347	0.3	1
1	A	163	TYR	H	8.001	0.020	1
1	A	163	TYR	HA	4.559	0.020	1
1	A	163	TYR	HB2	3.123	0.020	2
1	A	163	TYR	HB3	2.927	0.020	2
1	A	163	TYR	HD1	7.13	0.020	1
1	A	163	TYR	HD2	7.13	0.020	1
1	A	163	TYR	HE1	6.848	0.020	1
1	A	163	TYR	HE2	6.848	0.020	1
1	A	163	TYR	N	119.843	0.3	1
1	A	164	MET	CA	52.121	0.3	1
1	A	164	MET	CB	33.155	0.3	1
1	A	164	MET	CG	33.129	0.3	1
1	A	164	MET	H	7.854	0.020	1
1	A	164	MET	HA	4.788	0.020	1
1	A	164	MET	HB2	1.863	0.020	1
1	A	164	MET	HB3	1.863	0.020	1
1	A	164	MET	HG2	1.995	0.020	1
1	A	164	MET	HG3	1.995	0.020	1
1	A	164	MET	N	125.342	0.3	1
1	A	165	PRO	CA	62.988	0.3	1
1	A	165	PRO	CB	32.644	0.3	1
1	A	165	PRO	CD	52.145	0.3	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	167	PRO	C	177.796	0.3	1
1	A	167	PRO	CA	63.377	0.3	1
1	A	167	PRO	CB	31.933	0.3	1
1	A	167	PRO	HA	4.449	0.020	1
1	A	168	GLY	C	176.239	0.3	1
1	A	168	GLY	CA	45.361	0.3	1
1	A	168	GLY	H	8.418	0.020	1
1	A	168	GLY	HA2	3.997	0.020	1
1	A	168	GLY	HA3	3.997	0.020	1
1	A	168	GLY	N	109.083	0.3	1
1	A	169	MET	C	174.435	0.3	1
1	A	169	MET	CA	55.532	0.3	1
1	A	169	MET	CB	31.918	0.3	1
1	A	169	MET	CG	32.0	0.3	1
1	A	169	MET	CE	17.0	0.3	1
1	A	169	MET	H	8.018	0.020	1
1	A	169	MET	HA	4.469	0.020	1
1	A	169	MET	HB2	2.075	0.020	1
1	A	169	MET	HB3	2.075	0.020	1
1	A	169	MET	N	119.624	0.3	1
1	A	170	ILE	CA	57.591	0.3	1
1	A	170	ILE	CB	40.732	0.3	1
1	A	170	ILE	CG2	17.037	0.3	1
1	A	170	ILE	CD1	23.667	0.3	1
1	A	170	ILE	H	7.835	0.020	1
1	A	170	ILE	HA	4.284	0.020	1
1	A	170	ILE	HB	1.834	0.020	1
1	A	170	ILE	HG21	0.926	0.020	1
1	A	170	ILE	HG22	0.926	0.020	1
1	A	170	ILE	HG23	0.926	0.020	1
1	A	170	ILE	HD11	0.83	0.020	1
1	A	170	ILE	HD12	0.83	0.020	1
1	A	170	ILE	HD13	0.83	0.020	1
1	A	170	ILE	N	121.338	0.3	1
1	A	173	PRO	C	177.573	0.3	1
1	A	173	PRO	CA	63.31	0.3	1
1	A	173	PRO	CB	31.675	0.3	1
1	A	173	PRO	CG	27.132	0.3	1
1	A	173	PRO	CD	50.164	0.3	1
1	A	173	PRO	HA	4.455	0.020	1
1	A	173	PRO	HB2	2.333	0.020	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	173	PRO	HB3	2.333	0.020	1
1	A	174	GLY	C	173.891	0.3	1
1	A	174	GLY	CA	45.295	0.3	1
1	A	174	GLY	H	8.361	0.020	1
1	A	174	GLY	HA2	3.979	0.020	1
1	A	174	GLY	HA3	3.979	0.020	1
1	A	174	GLY	N	108.595	0.3	1
1	A	175	LEU	C	176.56	0.3	1
1	A	175	LEU	CA	54.937	0.3	1
1	A	175	LEU	CB	42.785	0.3	1
1	A	175	LEU	CG	26.821	0.3	1
1	A	175	LEU	CD1	23.7	0.3	1
1	A	175	LEU	CD2	23.805	0.3	1
1	A	175	LEU	H	7.905	0.020	1
1	A	175	LEU	HA	4.443	0.020	1
1	A	175	LEU	HB2	1.638	0.020	1
1	A	175	LEU	HB3	1.638	0.020	1
1	A	175	LEU	HD11	0.909	0.020	2
1	A	175	LEU	HD12	0.909	0.020	2
1	A	175	LEU	HD13	0.909	0.020	2
1	A	175	LEU	HD21	0.969	0.020	2
1	A	175	LEU	HD22	0.969	0.020	2
1	A	175	LEU	HD23	0.969	0.020	2
1	A	175	LEU	N	121.429	0.3	1
1	A	176	ALA	CA	50.415	0.3	1
1	A	176	ALA	CB	18.207	0.3	1
1	A	176	ALA	H	8.265	0.020	1
1	A	176	ALA	HA	4.657	0.020	1
1	A	176	ALA	HB1	1.408	0.020	1
1	A	176	ALA	HB2	1.408	0.020	1
1	A	176	ALA	HB3	1.408	0.020	1
1	A	176	ALA	N	126.406	0.3	1
1	A	177	PRO	C	177.771	0.3	1
1	A	177	PRO	CA	63.794	0.3	1
1	A	177	PRO	CB	31.816	0.3	1
1	A	177	PRO	CG	27.132	0.3	1
1	A	177	PRO	CD	50.282	0.3	1
1	A	177	PRO	HA	4.417	0.020	1
1	A	177	PRO	HB2	2.342	0.020	1
1	A	177	PRO	HB3	2.342	0.020	1
1	A	178	GLY	C	174.114	0.3	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	178	GLY	CA	45.402	0.3	1
1	A	178	GLY	H	8.518	0.020	1
1	A	178	GLY	HA2	3.941	0.020	2
1	A	178	GLY	HA3	4.037	0.020	2
1	A	178	GLY	N	109.56	0.3	1
1	A	179	GLN	C	175.596	0.3	1
1	A	179	GLN	CA	55.829	0.3	1
1	A	179	GLN	CB	29.75	0.3	1
1	A	179	GLN	CG	33.819	0.3	1
1	A	179	GLN	H	8.056	0.020	1
1	A	179	GLN	HA	4.396	0.020	1
1	A	179	GLN	HB2	2.155	0.020	2
1	A	179	GLN	HB3	2.066	0.020	2
1	A	179	GLN	N	119.634	0.3	1
1	A	180	ILE	CA	58.448	0.3	1
1	A	180	ILE	CB	38.739	0.3	1
1	A	180	ILE	CG1	27.024	0.3	1
1	A	180	ILE	CG2	17.214	0.3	1
1	A	180	ILE	CD1	12.679	0.3	1
1	A	180	ILE	H	8.156	0.020	1
1	A	180	ILE	HA	4.519	0.020	1
1	A	180	ILE	HB	1.918	0.020	1
1	A	180	ILE	HG12	1.232	0.020	2
1	A	180	ILE	HG13	1.579	0.020	2
1	A	180	ILE	HG21	1.005	0.020	1
1	A	180	ILE	HG22	1.005	0.020	1
1	A	180	ILE	HG23	1.005	0.020	1
1	A	180	ILE	HD11	0.911	0.020	1
1	A	180	ILE	HD12	0.911	0.020	1
1	A	180	ILE	HD13	0.911	0.020	1
1	A	180	ILE	N	123.603	0.3	1
1	A	182	PRO	C	177.746	0.3	1
1	A	182	PRO	CA	63.377	0.3	1
1	A	182	PRO	CB	31.933	0.3	1
1	A	182	PRO	CG	27.432	0.3	1
1	A	182	PRO	CD	50.282	0.3	1
1	A	182	PRO	HA	4.455	0.020	1
1	A	182	PRO	HB2	2.333	0.020	1
1	A	182	PRO	HB3	2.333	0.020	1
1	A	183	GLY	C	173.817	0.3	1
1	A	183	GLY	CA	45.295	0.3	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	183	GLY	H	8.418	0.020	1
1	A	183	GLY	HA2	4.079	0.020	2
1	A	183	GLY	HA3	3.984	0.020	2
1	A	183	GLY	N	108.789	0.3	1
1	A	184	ALA	C	177.277	0.3	1
1	A	184	ALA	CA	52.181	0.3	1
1	A	184	ALA	CB	19.417	0.3	1
1	A	184	ALA	H	7.984	0.020	1
1	A	184	ALA	HA	4.415	0.020	1
1	A	184	ALA	HB1	1.418	0.020	1
1	A	184	ALA	HB2	1.418	0.020	1
1	A	184	ALA	HB3	1.418	0.020	1
1	A	184	ALA	N	123.157	0.3	1
1	A	185	MET	CA	53.244	0.3	1
1	A	185	MET	CB	32.785	0.3	1
1	A	185	MET	CG	32.208	0.3	1
1	A	185	MET	CE	17.18	0.3	1
1	A	185	MET	H	8.193	0.020	1
1	A	185	MET	HA	4.859	0.020	1
1	A	185	MET	HB2	2.014	0.020	2
1	A	185	MET	HB3	2.111	0.020	2
1	A	185	MET	HG2	2.622	0.020	2
1	A	185	MET	HG3	2.707	0.020	2
1	A	185	MET	N	120.42	0.3	1
1	A	187	PRO	C	177.079	0.3	1
1	A	187	PRO	CA	63.177	0.3	1
1	A	187	PRO	CB	31.866	0.3	1
1	A	187	PRO	CG	27.203	0.3	1
1	A	187	PRO	CD	50.246	0.3	1
1	A	187	PRO	HA	4.462	0.020	1
1	A	187	PRO	HB2	2.347	0.020	1
1	A	187	PRO	HB3	2.347	0.020	1
1	A	188	GLN	C	175.893	0.3	1
1	A	188	GLN	CA	56.131	0.3	1
1	A	188	GLN	CB	29.528	0.3	1
1	A	188	GLN	CG	35.115	0.3	1
1	A	188	GLN	H	8.413	0.020	1
1	A	188	GLN	HA	4.332	0.020	1
1	A	188	GLN	HB2	2.137	0.020	1
1	A	188	GLN	HB3	2.137	0.020	1
1	A	188	GLN	HG2	2.286	0.020	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	188	GLN	HG3	2.236	0.020	2
1	A	188	GLN	N	119.859	0.3	1
1	A	189	GLN	C	175.522	0.3	1
1	A	189	GLN	CA	56.171	0.3	1
1	A	189	GLN	CB	29.634	0.3	1
1	A	189	GLN	CG	33.852	0.3	1
1	A	189	GLN	H	8.339	0.020	1
1	A	189	GLN	HA	4.334	0.020	1
1	A	189	GLN	HB2	2.135	0.020	1
1	A	189	GLN	HB3	2.135	0.020	1
1	A	189	GLN	HG2	2.436	0.020	1
1	A	189	GLN	HG3	2.436	0.020	1
1	A	189	GLN	N	120.982	0.3	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	117	-0.04 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	107	-0.19 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	100	0.22 ± 0.14	None needed (< 0.5 ppm)
^{15}N	100	0.21 ± 0.58	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 91%, i.e. 890 atoms were assigned a chemical shift out of a possible 980. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	342/349 (98%)	139/141 (99%)	137/140 (98%)	66/68 (97%)
Sidechain	484/537 (90%)	328/350 (94%)	146/168 (87%)	10/19 (53%)
Aromatic	64/94 (68%)	36/48 (75%)	28/44 (64%)	0/2 (0%)
Overall	890/980 (91%)	503/539 (93%)	311/352 (88%)	76/89 (85%)

7.1.4 Statistically unusual chemical shifts (i)

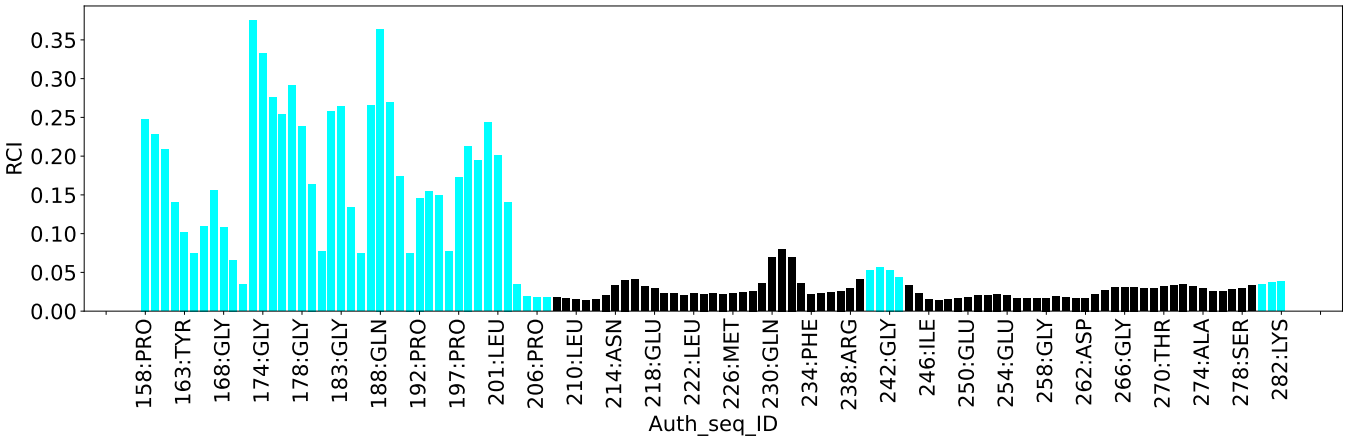
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	170	ILE	CD1	23.67	5.18 – 21.60	6.3
1	A	204	ASN	CB	30.34	30.50 – 46.89	-5.1

7.1.5 Random Coil Index (RCI) plots (i)

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	1717
Intra-residue (|i-j|=0)	287
Sequential (|i-j|=1)	413
Medium range (|i-j|>1 and |i-j|<5)	258
Long range (|i-j|≥5)	639
Inter-chain	0
Hydrogen bond restraints	120
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	13.6
Number of long range restraints per residue[1]	5.4

[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)	59.8	0.2
0.2-0.5 (Medium)	17.5	0.41
>0.5 (Large)	0.1	0.56

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)

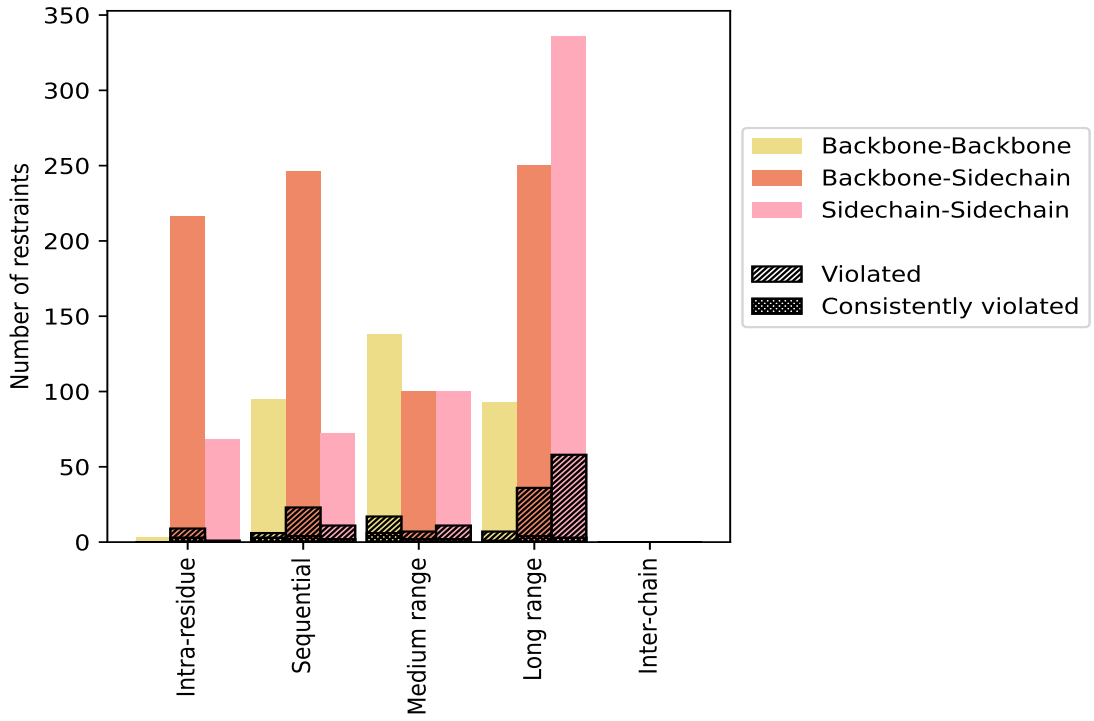
9.1 Summary of distance violations (i)

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)	287	16.7	10	3.5	0.6	3	1.0	0.2
Backbone-Backbone	3	0.2	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	216	12.6	9	4.2	0.5	3	1.4	0.2
Sidechain-Sidechain	68	4.0	1	1.5	0.1	0	0.0	0.0
Sequential (|i-j|=1)	413	24.1	40	9.7	2.3	9	2.2	0.5
Backbone-Backbone	95	5.5	6	6.3	0.3	3	3.2	0.2
Backbone-Sidechain	246	14.3	23	9.3	1.3	4	1.6	0.2
Sidechain-Sidechain	72	4.2	11	15.3	0.6	2	2.8	0.1
Medium range (|i-j|>1 & |i-j|<5)	258	15.0	23	8.9	1.3	4	1.6	0.2
Backbone-Backbone	58	3.4	5	8.6	0.3	0	0.0	0.0
Backbone-Sidechain	100	5.8	7	7.0	0.4	2	2.0	0.1
Sidechain-Sidechain	100	5.8	11	11.0	0.6	2	2.0	0.1
Long range (|i-j|≥5)	639	37.2	98	15.3	5.7	8	1.3	0.5
Backbone-Backbone	53	3.1	4	7.5	0.2	1	1.9	0.1
Backbone-Sidechain	250	14.6	36	14.4	2.1	4	1.6	0.2
Sidechain-Sidechain	336	19.6	58	17.3	3.4	3	0.9	0.2
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	120	7.0	15	12.5	0.9	6	5.0	0.3
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	1717	100.0	186	10.8	10.8	30	1.7	1.7
Backbone-Backbone	329	19.2	30	9.1	1.7	10	3.0	0.6
Backbone-Sidechain	812	47.3	75	9.2	4.4	13	1.6	0.8
Sidechain-Sidechain	576	33.5	81	14.1	4.7	7	1.2	0.4

[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	5	19	19	36	0	79	0.17	0.4	0.05	0.15
2	5	22	21	30	0	78	0.17	0.56	0.07	0.16
3	6	19	20	32	0	77	0.17	0.38	0.06	0.16
4	5	19	20	31	0	75	0.17	0.37	0.06	0.16
5	5	17	19	35	0	76	0.17	0.36	0.05	0.16
6	5	23	20	30	0	78	0.18	0.39	0.06	0.16
7	5	16	21	29	0	71	0.17	0.36	0.06	0.17
8	5	18	18	29	0	70	0.19	0.52	0.07	0.17
9	5	16	20	32	0	73	0.17	0.36	0.06	0.16
10	5	19	19	38	0	81	0.17	0.35	0.05	0.16
11	4	16	22	35	0	77	0.17	0.37	0.06	0.16

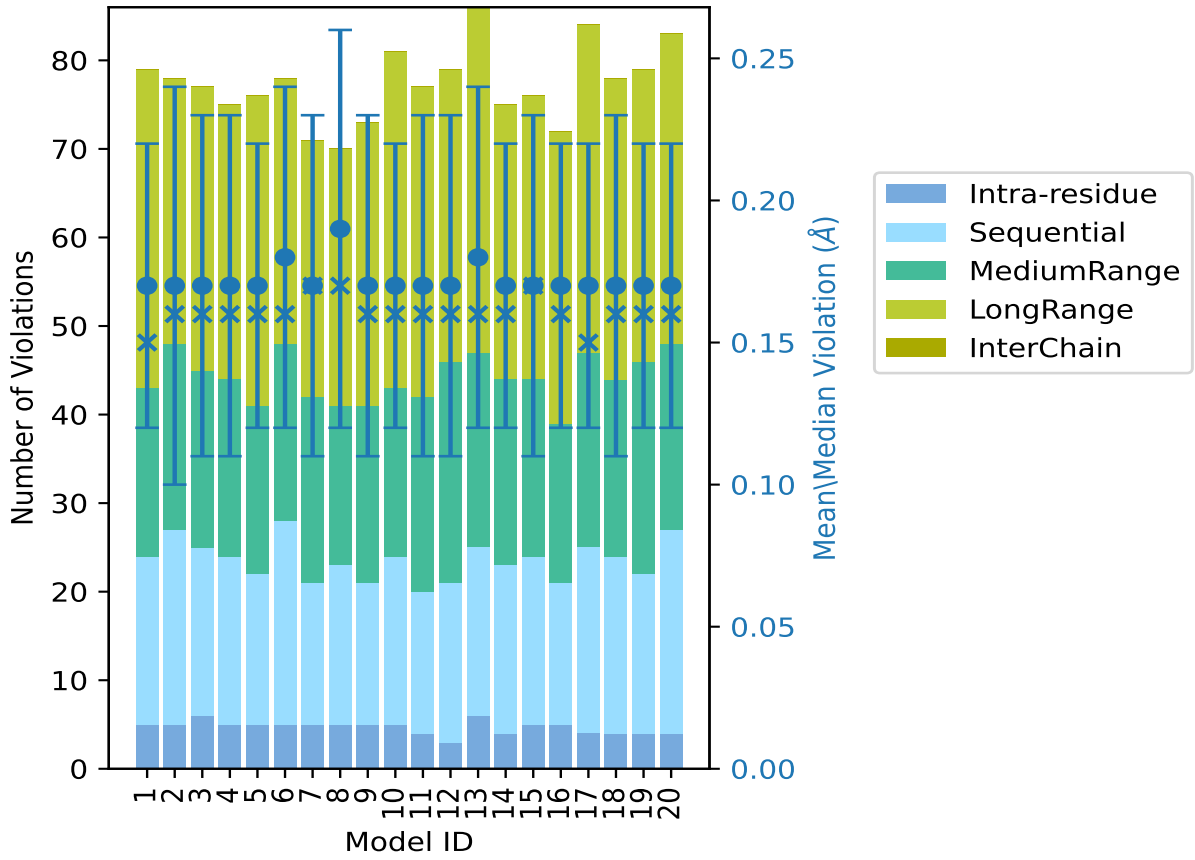
Continued on next page...

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	3	18	25	33	0	79	0.17	0.37	0.06	0.16
13	6	19	22	39	0	86	0.18	0.35	0.06	0.16
14	4	19	21	31	0	75	0.17	0.37	0.05	0.16
15	5	19	20	32	0	76	0.17	0.39	0.06	0.17
16	5	16	18	33	0	72	0.17	0.37	0.05	0.16
17	4	21	22	37	0	84	0.17	0.34	0.05	0.15
18	4	20	20	34	0	78	0.17	0.41	0.06	0.16
19	4	18	24	33	0	79	0.17	0.39	0.05	0.16
20	4	23	21	35	0	83	0.17	0.35	0.05	0.16

[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

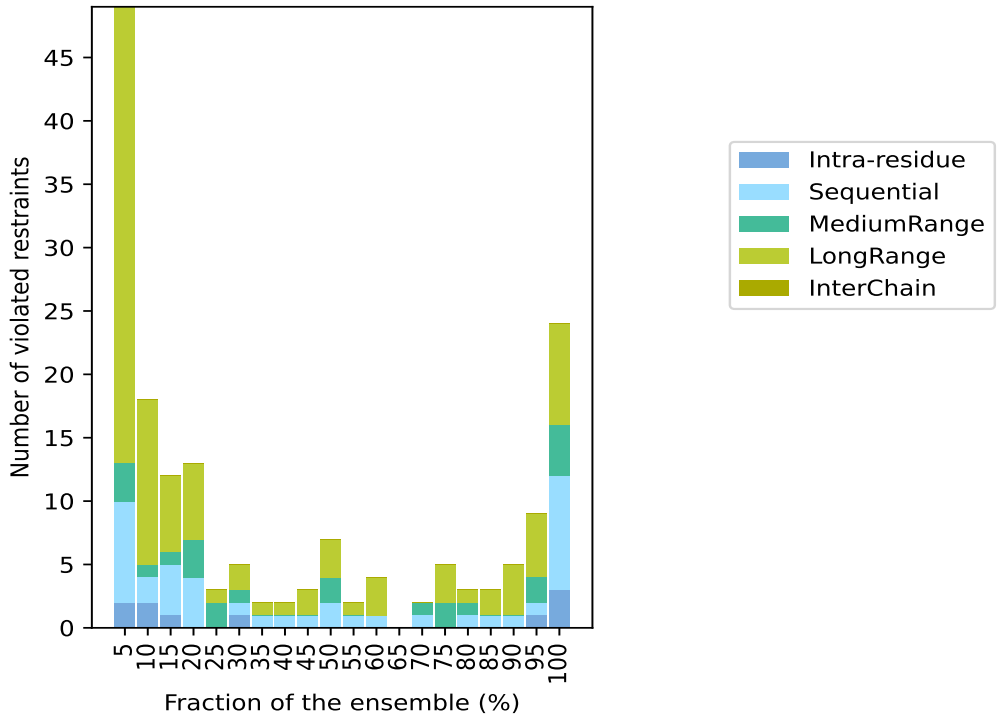
9.3 Distance violation statistics for the ensemble (i)

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, 1426(IR:277, SQ:373, MR:235, LR:541, 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]	%
2	8	3	36	0	49	1	5.0
2	2	1	13	0	18	2	10.0
1	4	1	6	0	12	3	15.0
0	4	3	6	0	13	4	20.0
0	0	2	1	0	3	5	25.0
1	1	1	2	0	5	6	30.0
0	1	0	1	0	2	7	35.0
0	1	0	1	0	2	8	40.0
0	1	0	2	0	3	9	45.0
0	2	2	3	0	7	10	50.0
0	1	0	1	0	2	11	55.0
0	1	0	3	0	4	12	60.0
0	0	0	0	0	0	13	65.0
0	1	1	0	0	2	14	70.0
0	0	2	3	0	5	15	75.0
0	1	1	1	0	3	16	80.0
0	1	0	2	0	3	17	85.0
0	1	0	4	0	5	18	90.0
1	1	2	5	0	9	19	95.0
3	9	4	8	0	24	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

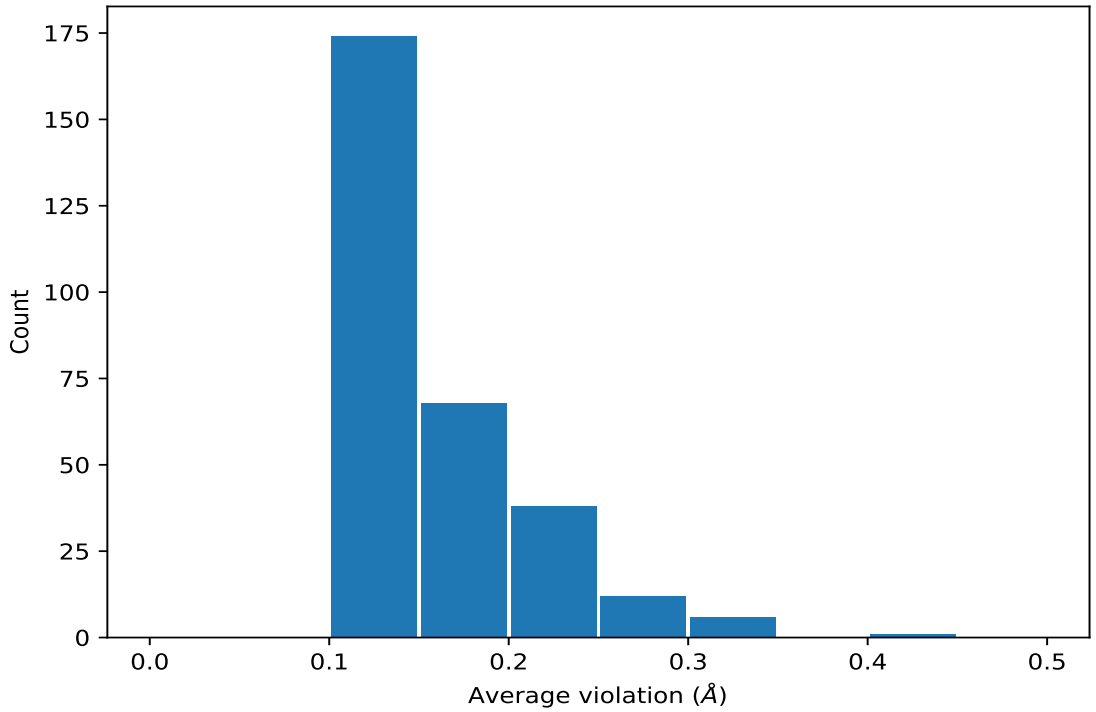
9.3.1 Bar graph : Distance violation statistics for the ensemble



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 (i)

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)
(1,1144)	1:A:250:GLU:HA	1:A:257:ALA:H	20	0.34	0.05	0.35
(1,316)	1:A:212:LEU:HA	1:A:276:LYS:HB3	20	0.31	0.07	0.33
(1,214)	1:A:210:LEU:HB3	1:A:257:ALA:HB1	20	0.29	0.01	0.29
(1,214)	1:A:210:LEU:HB3	1:A:257:ALA:HB2	20	0.29	0.01	0.29
(1,214)	1:A:210:LEU:HB3	1:A:257:ALA:HB3	20	0.29	0.01	0.29
(1,1344)	1:A:263:ALA:HB1	1:A:265:GLN:HB2	20	0.24	0.02	0.24
(1,1344)	1:A:263:ALA:HB2	1:A:265:GLN:HB2	20	0.24	0.02	0.24
(1,1344)	1:A:263:ALA:HB3	1:A:265:GLN:HB2	20	0.24	0.02	0.24
(2,44)	1:A:265:GLN:H	1:A:261:ARG:O	20	0.22	0.03	0.23
(2,52)	1:A:263:ALA:H	1:A:259:ALA:O	20	0.22	0.02	0.22
(1,741)	1:A:226:MET:HG2	1:A:227:LEU:HD21	20	0.22	0.02	0.23
(1,741)	1:A:226:MET:HG2	1:A:227:LEU:HD22	20	0.22	0.02	0.23
(1,741)	1:A:226:MET:HG2	1:A:227:LEU:HD23	20	0.22	0.02	0.23
(1,741)	1:A:226:MET:HG3	1:A:227:LEU:HD21	20	0.22	0.02	0.23
(1,741)	1:A:226:MET:HG3	1:A:227:LEU:HD22	20	0.22	0.02	0.23
(1,741)	1:A:226:MET:HG3	1:A:227:LEU:HD23	20	0.22	0.02	0.23

Continued on next page...

Continued from previous page...

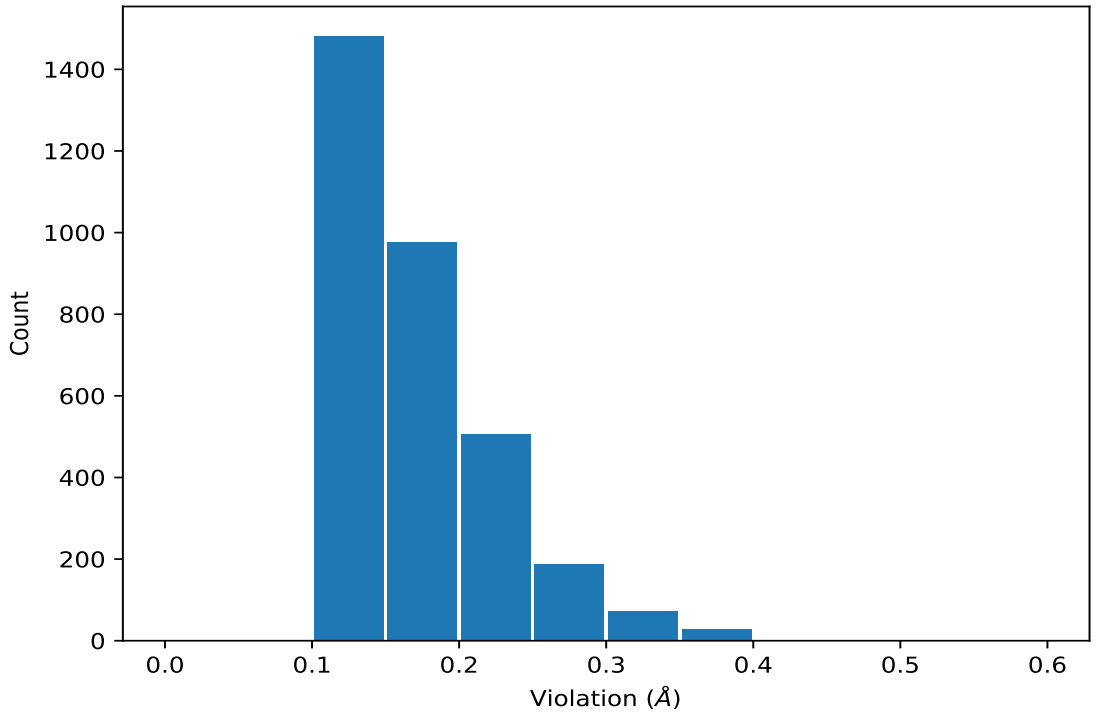
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(1,115)	1:A:207:ASN:HD22	1:A:248:PHE:HD1	20	0.22	0.05	0.21
(1,115)	1:A:207:ASN:HD22	1:A:248:PHE:HD2	20	0.22	0.05	0.21
(1,580)	1:A:220:ASN:HD21	1:A:221:GLU:H	20	0.21	0.02	0.21
(1,346)	1:A:212:LEU:HD11	1:A:277:ILE:H	20	0.2	0.06	0.18

[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)
(1,30)	1:A:197:PRO:HA	1:A:198:ALA:H	2	0.56

Continued on next page...

Continued from previous page...

Key	Atom-1	Atom-2	Model ID	Violation (Å)
(1,30)	1:A:197:PRO:HA	1:A:198:ALA:H	8	0.52
(1,1144)	1:A:250:GLU:HA	1:A:257:ALA:H	18	0.41
(1,1144)	1:A:250:GLU:HA	1:A:257:ALA:H	1	0.4
(1,958)	1:A:236:GLU:HB3	1:A:237:VAL:H	8	0.39
(1,534)	1:A:218:GLU:H	1:A:218:GLU:HG2	15	0.39
(1,534)	1:A:218:GLU:H	1:A:218:GLU:HG3	15	0.39
(1,316)	1:A:212:LEU:HA	1:A:276:LYS:HB3	19	0.39
(1,115)	1:A:207:ASN:HD22	1:A:248:PHE:HD1	6	0.39
(1,115)	1:A:207:ASN:HD22	1:A:248:PHE:HD2	6	0.39
(1,1144)	1:A:250:GLU:HA	1:A:257:ALA:H	3	0.38
(1,316)	1:A:212:LEU:HA	1:A:276:LYS:HB3	2	0.37

10 Dihedral-angle violation analysis (i)

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