wwPDB NMR Structure Validation Summary Report

Dec 24, 2024 – 03:26 PM EST

PDB ID	:	2LEG
BMRB ID	:	17710
Title	:	Membrane protein complex DsbB-DsbA structure by joint calculations with solid-state NMR and X-ray experimental data
Authors	:	Tang, M.; Sperling, L.J.; Berthold, D.A.; Schwieters, C.D.; Nesbitt, A.E.; Nieuwkoop, A.J.; Gennis, R.B.; Rienstra, C.M.
Deposited on	:	2011-06-15

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

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

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

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

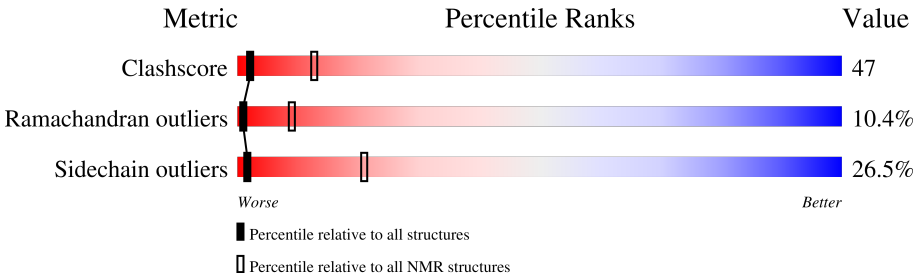
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 33%.

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	189	32% 58% 8%
2	B	176	15% 41% 16% 24%

2 Ensemble composition and analysis

This entry contains 10 models. Model 9 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:2-A:188, B:14-B:36, B:40-B:115, B:120-B:126, B:142-B:162 (314)	1.35	9

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 5 single-model clusters were found.

Cluster number	Models
1	1, 5, 9
2	3, 8
Single-model clusters	2; 4; 6; 7; 10

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5154 atoms, of which 2579 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Thiol:disulfide interchange protein DsbA.

Mol	Chain	Residues	Atoms						Trace
1	A	188	Total	C	H	N	O	S	0
			2930	946	1453	242	282	7	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	ALA	CYS	engineered mutation	UNP P0AEG4

- Molecule 2 is a protein called Disulfide bond formation protein B.

Mol	Chain	Residues	Atoms						Trace
2	B	134	Total	C	H	N	O	S	0
			2205	741	1126	164	165	9	

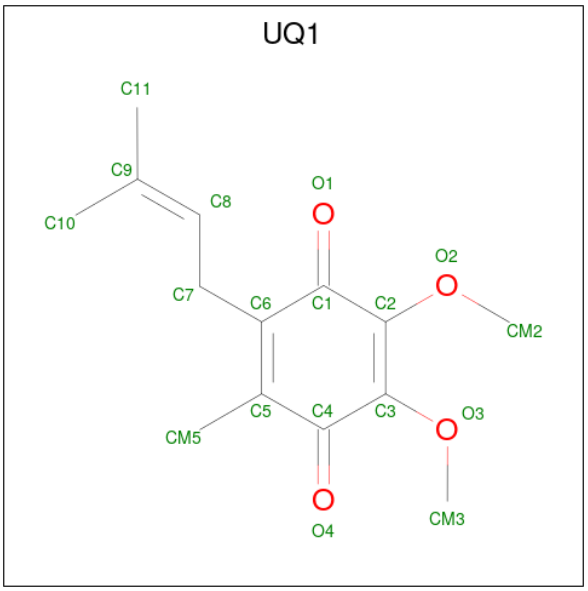
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	8	ALA	CYS	engineered mutation	UNP P0A6M2
B	49	VAL	CYS	engineered mutation	UNP P0A6M2
B	130	SER	CYS	engineered mutation	UNP P0A6M2

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
3	A	1	Total	Zn
			1	1

- Molecule 4 is UBIQUINONE-1 (three-letter code: UQ1) (formula: $C_{14}H_{18}O_4$).



Mol	Chain	Residues	Atoms		
			Total	C	O
4	B	1	18	14	4

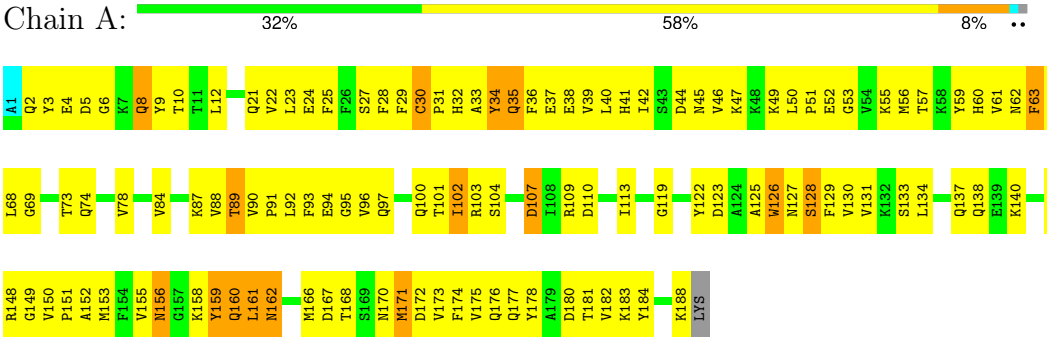
4 Residue-property plots

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.

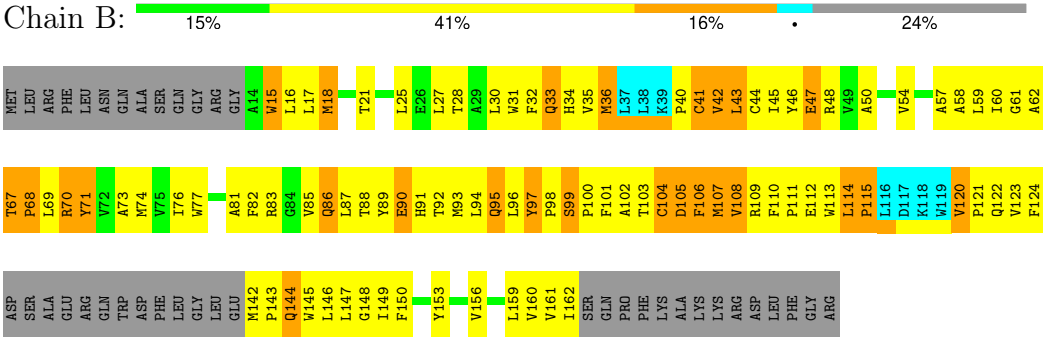
• Molecule 1: Thiol:disulfide interchange protein DsbA

Chain A:



• Molecule 2: Disulfide bond formation protein B

Chain B:

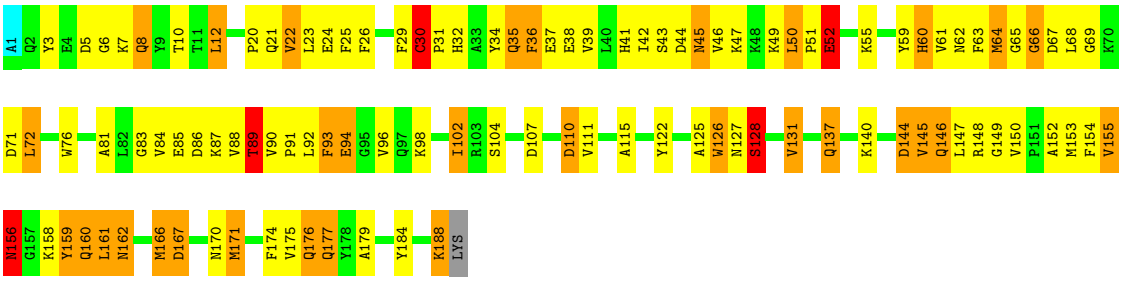


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

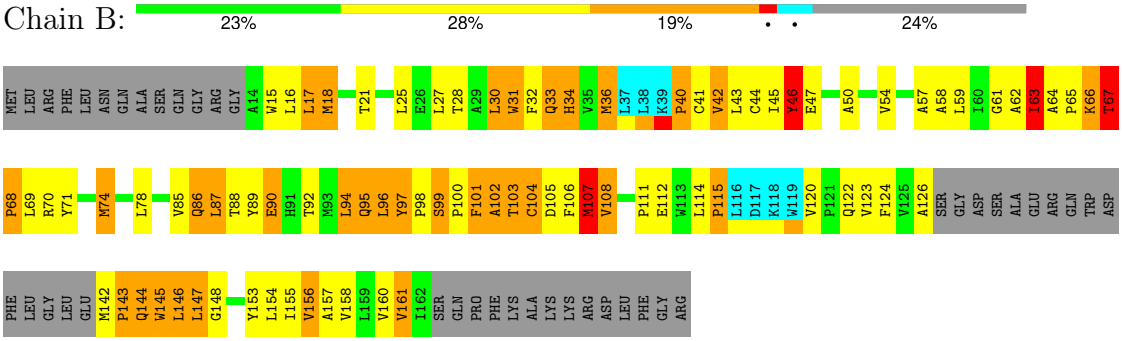
The representative model is number 9. Colouring as in section 4.1 above.

• Molecule 1: Thiol:disulfide interchange protein DsbA

Chain A:



• Molecule 2: Disulfide bond formation protein B



5 Refinement protocol and experimental data overview

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

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

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

Software name	Classification	Version
X-PLOR NIH	refinement	
X-PLOR NIH	structure solution	

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	1555
Number of shifts mapped to atoms	1486
Number of unparsed shifts	0
Number of shifts with mapping errors	69
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	33%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, UQ1

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	0.75±0.02	0±0/1504 (0.0± 0.0%)	0.77±0.01	1±0/2034 (0.0± 0.0%)
2	B	0.75±0.01	0±0/1049 (0.0± 0.0%)	0.83±0.02	0±0/1439 (0.0± 0.0%)
All	All	0.75	0/25530 (0.0%)	0.80	8/34730 (0.0%)

There are no bond-length outliers.

All 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	59	TYR	CB-CG-CD2	-6.15	117.31	121.00	9	6
2	B	89	TYR	CB-CG-CD2	-6.13	117.32	121.00	6	1
2	B	46	TYR	CB-CG-CD2	-5.16	117.90	121.00	9	1

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	1472	1446	1443	121±11
2	B	1015	1053	1049	121±18
4	B	18	0	18	9±4
All	All	25060	24990	25100	2340

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 47.

5 of 1662 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:41:CYS:SG	4:B:501:UQ1:HM52	1.05	1.92	2	1
2:B:147:LEU:HD13	2:B:148:GLY:N	0.95	1.76	7	3
2:B:87:LEU:HD13	2:B:87:LEU:H	0.94	1.22	4	1
1:A:102:ILE:HD13	1:A:102:ILE:H	0.93	1.22	9	1
2:B:41:CYS:O	2:B:43:LEU:N	0.90	2.04	9	2

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	186/189 (98%)	150±4 (81±2%)	23±5 (13±3%)	13±2 (7±1%)	2	17
2	B	123/176 (70%)	85±3 (69±2%)	18±3 (15±3%)	20±2 (16±2%)	0	4
All	All	3090/3650 (85%)	2353 (76%)	415 (13%)	322 (10%)	1	9

5 of 102 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	120	VAL	10
2	B	108	VAL	9
1	A	30	CYS	8
1	A	89	THR	8
2	B	115	PRO	8

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	157/158 (99%)	118±7 (75±4%)	39±7 (25±4%)	2 24
2	B	106/147 (72%)	75±2 (71±2%)	31±2 (29±2%)	1 17
All	All	2630/3050 (86%)	1933 (73%)	697 (27%)	1 21

5 of 215 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	159	TYR	9
2	B	71	TYR	9
1	A	93	PHE	8
2	B	86	GLN	8
2	B	107	MET	8

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 (i)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
4	UQ1	B	501	-	18,18,18	2.22±0.01	2±0 (11±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
4	UQ1	B	501	-	24,25,25	1.44±0.00	4±0 (16±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	UQ1	B	501	-	-	0±0,9,33,33	0±0,1,1,1

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
4	B	501	UQ1	C6-C5	8.16	1.49	1.35	7	10
4	B	501	UQ1	C3-C2	3.84	1.50	1.36	1	10

All 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
4	B	501	UQ1	CM5-C5-C6	3.31	119.01	124.45	7	10
4	B	501	UQ1	C7-C6-C5	3.13	119.52	124.89	7	10
4	B	501	UQ1	C7-C6-C1	2.69	121.64	118.52	7	10
4	B	501	UQ1	C11-C9-C10	2.19	119.63	114.59	10	10

There are no chirality outliers.

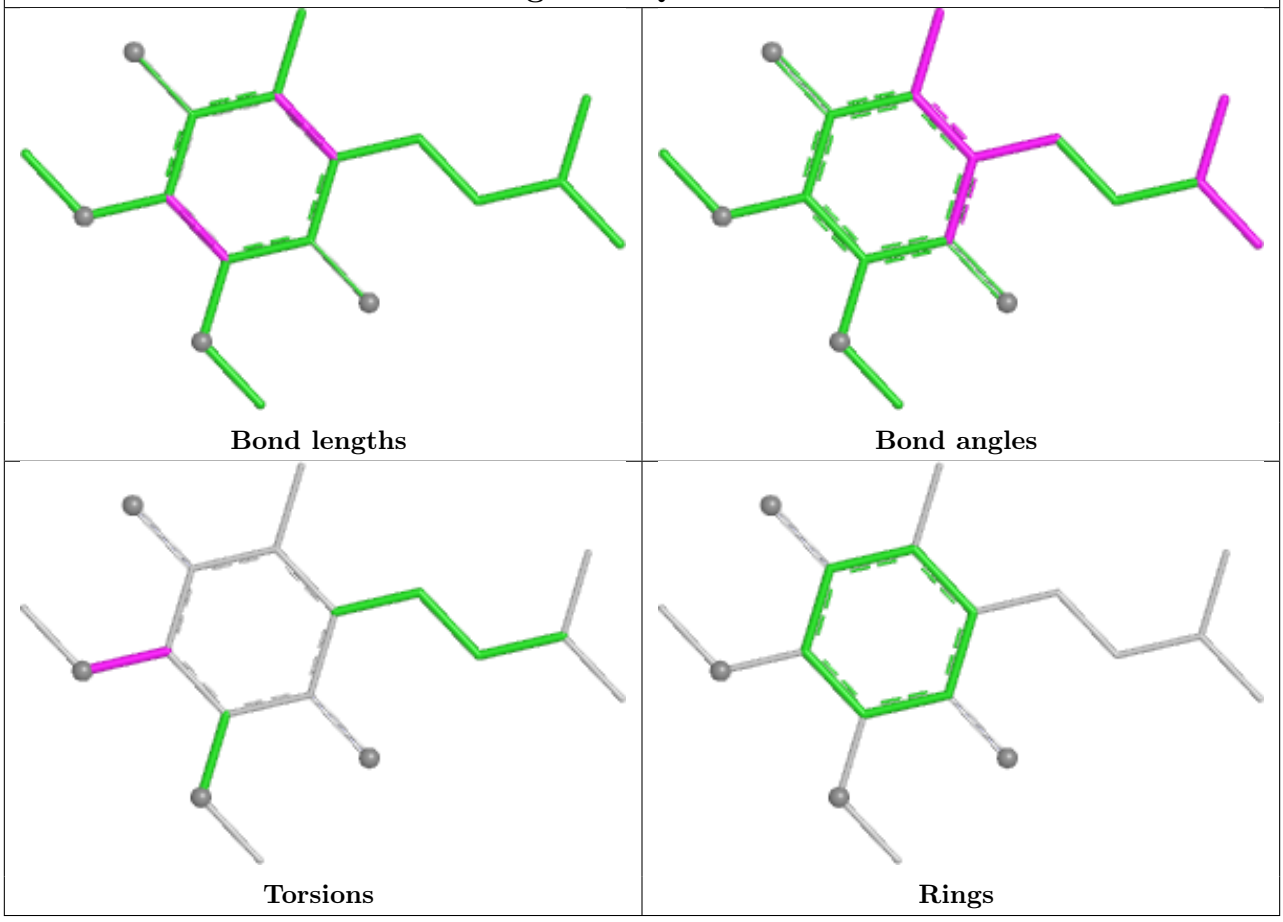
There are no torsion outliers.

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In

addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand UQ1 B 501



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 33% for the well-defined parts and 32% 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	1555
Number of shifts mapped to atoms	1486
Number of unparsed shifts	0
Number of shifts with mapping errors	69
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

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 69) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	6	ASN	C	179.258	0.145	1
1	B	6	ASN	CA	57.564	0.000	1
1	B	6	ASN	CB	38.167	0.000	1
1	B	7	GLN	H	9.338	0.026	1
1	B	7	GLN	C	180.404	0.092	1
1	B	7	GLN	CA	57.611	0.012	1
1	B	7	GLN	CB	26.649	0.210	1
1	B	7	GLN	CG	31.228	0.000	1
1	B	7	GLN	N	120.525	0.189	1
1	B	8	ALA	H	9.436	0.052	1
1	B	8	ALA	C	179.313	0.000	1
1	B	8	ALA	CA	54.645	0.000	1
1	B	8	ALA	CB	17.967	0.000	1
1	B	8	ALA	N	123.894	0.176	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	9	SER	C	173.894	0.126	1
1	B	9	SER	CA	60.565	0.137	1
1	B	9	SER	CB	63.79	0.070	1
1	B	10	GLN	H	7.835	0.000	1
1	B	10	GLN	C	174.766	0.000	1
1	B	10	GLN	CB	29.955	0.000	1
1	B	10	GLN	N	119.198	0.200	1
1	B	11	GLY	H	8.433	0.005	1
1	B	11	GLY	N	108.456	0.067	1
1	B	12	ARG	C	178.053	0.000	1
1	B	12	ARG	CA	57.504	0.000	1
1	B	12	ARG	CB	30.522	0.000	1
1	B	12	ARG	CD	43.193	0.107	1
1	B	12	ARG	CZ	159.227	0.000	1
1	B	13	GLY	C	175.091	0.000	1
1	B	13	GLY	CA	45.326	0.000	1
1	B	13	GLY	N	108.265	0.000	1
1	B	127	SER	H	8.145	0.026	1
1	B	127	SER	CA	58.101	0.000	1
1	B	127	SER	CB	64.539	0.000	1
1	B	127	SER	N	112.716	0.005	1
1	B	128	GLY	H	8.595	0.058	1
1	B	128	GLY	C	177.047	0.000	1
1	B	128	GLY	CA	46.296	0.123	1
1	B	128	GLY	N	106.487	0.017	1
1	B	129	ASP	C	177.916	0.000	1
1	B	129	ASP	CB	43.975	0.134	1
1	B	130	SER	H	9.948	0.003	1
1	B	130	SER	C	173.312	0.000	1
1	B	130	SER	CA	64.581	0.074	1
1	B	130	SER	CB	42.879	0.043	1
1	B	130	SER	N	122.984	0.016	1
1	B	131	ALA	H	9.869	0.029	1
1	B	131	ALA	CA	52.64	0.000	1
1	B	131	ALA	CB	18.497	0.000	1
1	B	131	ALA	N	121.534	0.128	1
1	B	136	ASP	C	172.916	0.000	1
1	B	137	PHE	H	8.959	0.005	1
1	B	137	PHE	CA	57.31	0.000	1
1	B	137	PHE	CB	42.927	0.000	1
1	B	137	PHE	N	117.677	0.047	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	138	LEU	CB	43.547	0.000	1
1	B	139	GLY	C	176.262	0.000	1
1	B	139	GLY	CA	46.123	0.000	1
1	B	140	LEU	CB	40.503	0.012	1
1	B	140	LEU	N	124.207	0.000	1
1	B	163	SER	H	9.427	0.022	1
1	B	163	SER	CA	60.441	0.222	1
1	B	163	SER	CB	63.788	0.000	1
1	B	163	SER	N	112.832	0.000	1
1	B	165	PRO	CA	62.249	0.011	1
1	B	165	PRO	CD	51.723	0.128	1
1	B	165	PRO	N	134.035	0.022	1
1	B	168	ALA	CA	51.195	0.007	1
1	B	168	ALA	CB	18.891	0.003	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	306	-0.68 ± 0.07	Should be checked
$^{13}C_{\beta}$	267	0.35 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	277	-0.38 ± 0.10	None needed (< 0.5 ppm)
^{15}N	286	0.67 ± 0.26	Should be applied

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 33%, i.e. 1421 atoms were assigned a chemical shift out of a possible 4367. 0 out of 68 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	862/1556 (55%)	54/630 (9%)	550/628 (88%)	258/298 (87%)
Sidechain	528/2361 (22%)	0/1561 (0%)	528/736 (72%)	0/64 (0%)
Aromatic	31/450 (7%)	0/218 (0%)	29/215 (13%)	2/17 (12%)
Overall	1421/4367 (33%)	54/2409 (2%)	1107/1579 (70%)	260/379 (69%)

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

7.1.4 Statistically unusual chemical shifts (i)

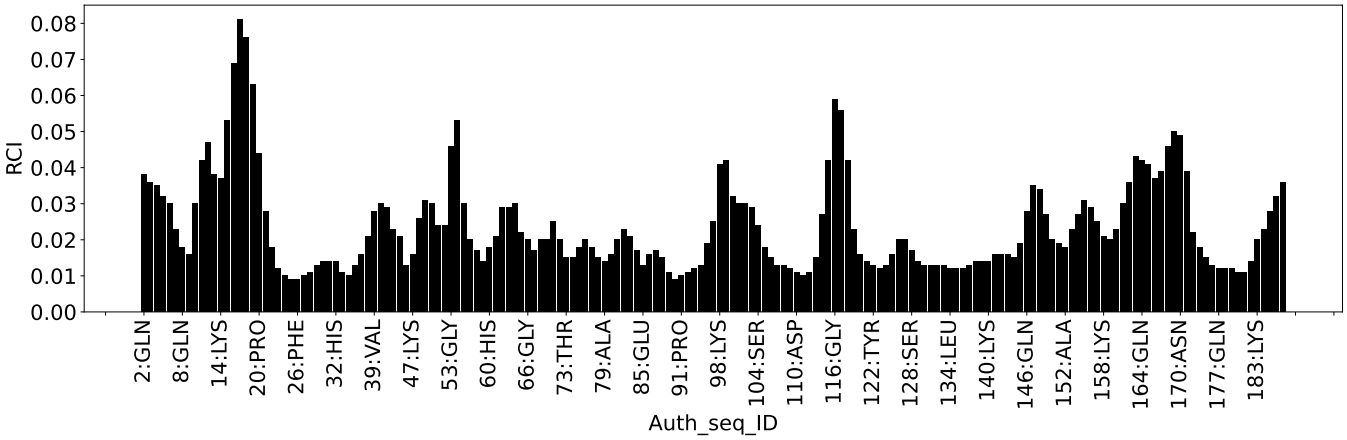
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	130	SER	CB	42.88	56.28 – 71.32	-13.9
1	A	24	GLU	CB	40.08	21.56 – 38.37	6.0
1	B	92	THR	CG2	27.48	16.06 – 27.03	5.4
1	A	5	ASP	CB	32.43	32.98 – 48.76	-5.3

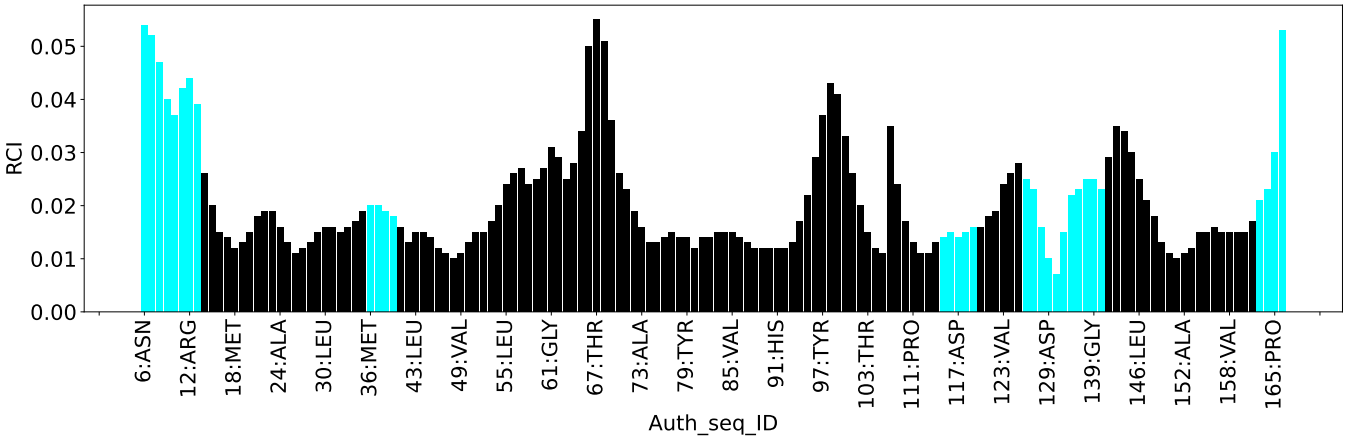
7.1.5 Random Coil Index (RCI) plots (i)

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:



Random coil index (RCI) for chain B:


RCI
0.00
0.01
0.02
0.03
0.04
0.05
6:ASN
12:ARG
18:MET
24:ALA
30:LEU
36:MET
43:LEU
49:VAL
55:LEU
61:GLY
67:THR
73:ALA
79:TYR
85:VAL
91:HIS
97:TYR
103:THR
111:PRO
117:ASP
123:VAL
129:ASP
139:GLY
146:LEU
152:ALA
158:VAL
165:PRO
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	811
Intra-residue (|i-j|=0)	190
Sequential (|i-j|=1)	329
Medium range (|i-j|>1 and |i-j|<5)	267
Long range (|i-j|≥5)	25
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	399
Number of unmapped restraints	1
Number of restraints per residue	3.3
Number of long range restraints per residue[1]	0.1

[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)	4.0	0.2
0.2-0.5 (Medium)	55.4	0.5
>0.5 (Large)	2.0	0.7

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)	8.6	2.64
10.0-20.0 (Medium)	None	None
>20.0 (Large)	None	None

9 Distance violation analysis (i)

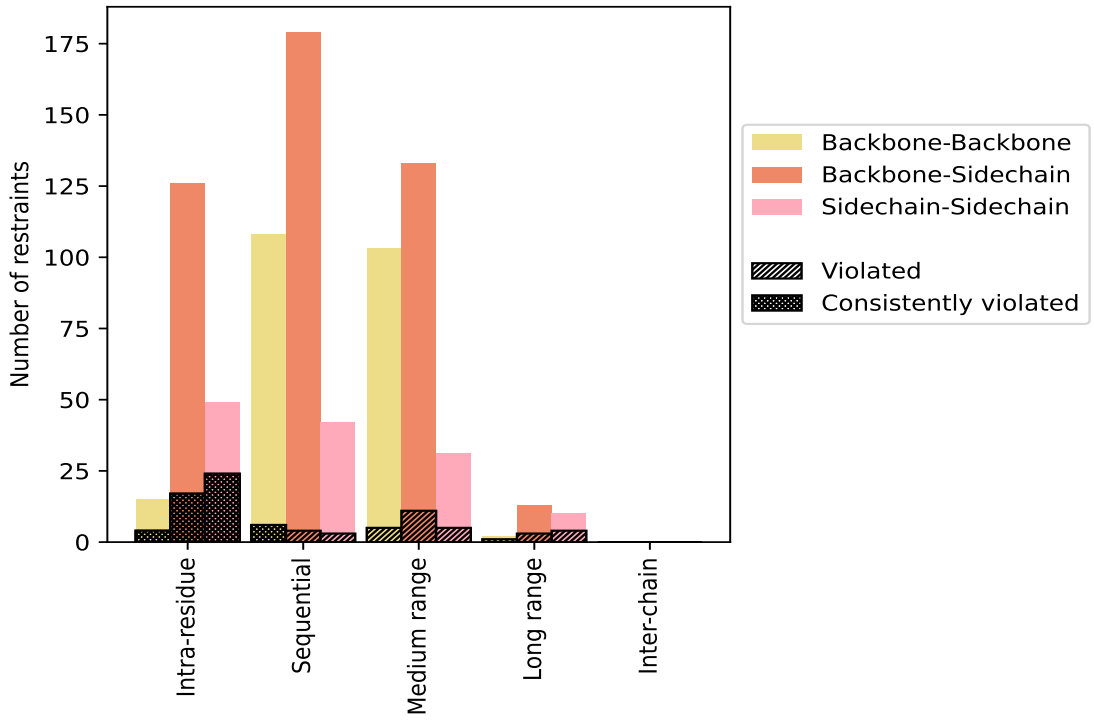
9.1 Summary of distance violations (i)

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)	190	23.4	45	23.7	5.5	45	23.7	5.5
Backbone-Backbone	15	1.8	4	26.7	0.5	4	26.7	0.5
Backbone-Sidechain	126	15.5	17	13.5	2.1	17	13.5	2.1
Sidechain-Sidechain	49	6.0	24	49.0	3.0	24	49.0	3.0
Sequential (|i-j|=1)	329	40.6	13	4.0	1.6	6	1.8	0.7
Backbone-Backbone	108	13.3	6	5.6	0.7	6	5.6	0.7
Backbone-Sidechain	179	22.1	4	2.2	0.5	0	0.0	0.0
Sidechain-Sidechain	42	5.2	3	7.1	0.4	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	267	32.9	21	7.9	2.6	0	0.0	0.0
Backbone-Backbone	103	12.7	5	4.9	0.6	0	0.0	0.0
Backbone-Sidechain	133	16.4	11	8.3	1.4	0	0.0	0.0
Sidechain-Sidechain	31	3.8	5	16.1	0.6	0	0.0	0.0
Long range (|i-j|≥5)	25	3.1	8	32.0	1.0	0	0.0	0.0
Backbone-Backbone	2	0.2	1	50.0	0.1	0	0.0	0.0
Backbone-Sidechain	13	1.6	3	23.1	0.4	0	0.0	0.0
Sidechain-Sidechain	10	1.2	4	40.0	0.5	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	0	0.0	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	811	100.0	87	10.7	10.7	51	6.3	6.3
Backbone-Backbone	228	28.1	16	7.0	2.0	10	4.4	1.2
Backbone-Sidechain	451	55.6	35	7.8	4.3	17	3.8	2.1
Sidechain-Sidechain	132	16.3	36	27.3	4.4	24	18.2	3.0

[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

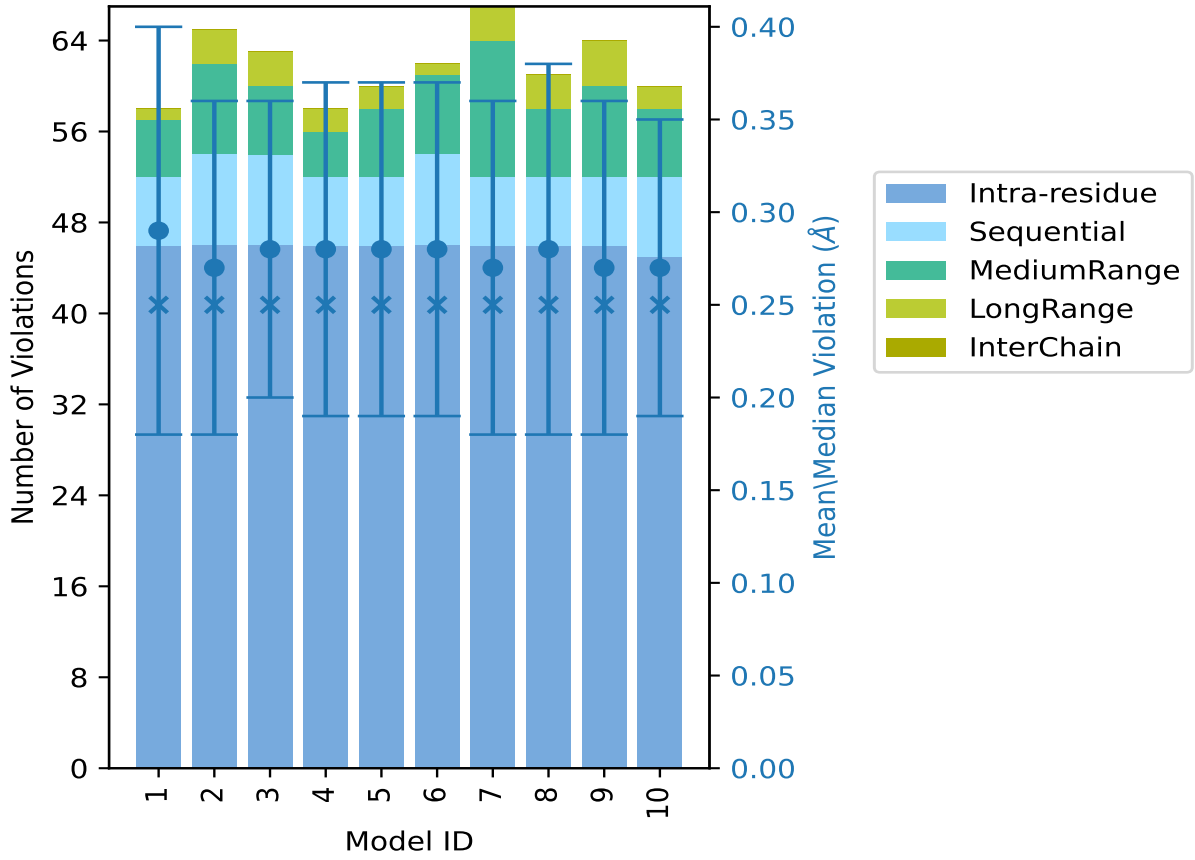
9.2 Distance violation statistics for each model

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	46	6	5	1	0	58	0.29	0.7	0.11	0.25
2	46	8	8	3	0	65	0.27	0.52	0.09	0.25
3	46	8	6	3	0	63	0.28	0.51	0.08	0.25
4	46	6	4	2	0	58	0.28	0.5	0.09	0.25
5	46	6	6	2	0	60	0.28	0.69	0.09	0.25
6	46	8	7	1	0	62	0.28	0.51	0.09	0.25
7	46	6	12	3	0	67	0.27	0.51	0.09	0.25
8	46	6	6	3	0	61	0.28	0.52	0.1	0.25
9	46	6	8	4	0	64	0.27	0.52	0.09	0.25
10	45	7	6	2	0	60	0.27	0.53	0.08	0.25

[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 (i)



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, 724(IR:145, SQ:316, MR:246, LR:17, 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]	%
0	7	8	3	0	18	1	10.0
0	0	1	1	0	2	2	20.0
0	0	1	1	0	2	3	30.0

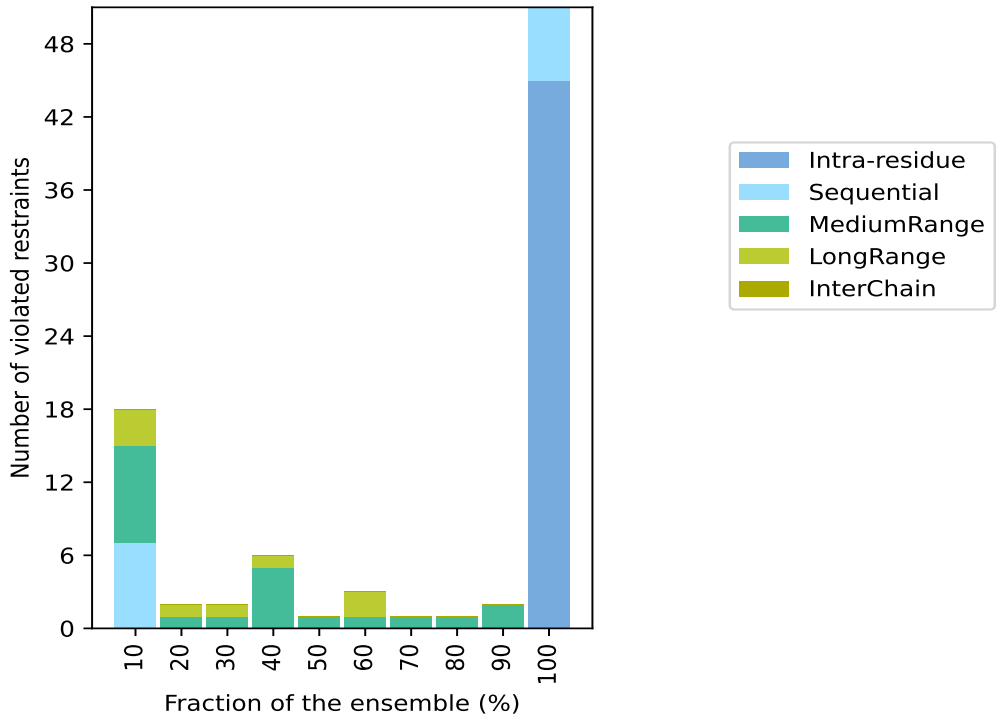
Continued on next page...

Continued from previous page...

Number of violated restraints						Fraction of the ensemble	
IR[1]	SQ[2]	MR[3]	LR[4]	IC[5]	Total	Count[6]	%
0	0	5	1	0	6	4	40.0
0	0	1	0	0	1	5	50.0
0	0	1	2	0	3	6	60.0
0	0	1	0	0	1	7	70.0
0	0	1	0	0	1	8	80.0
0	0	2	0	0	2	9	90.0
45	6	0	0	0	51	10	100.0

[1]Intra-residue restraints, [2]Sequential restraints, [3]Medium range restraints, [4]Long range restraints, [5]Inter-chain restraints, [6] Number of models with violations

9.3.1 Bar graph : Distance violation statistics for the ensemble

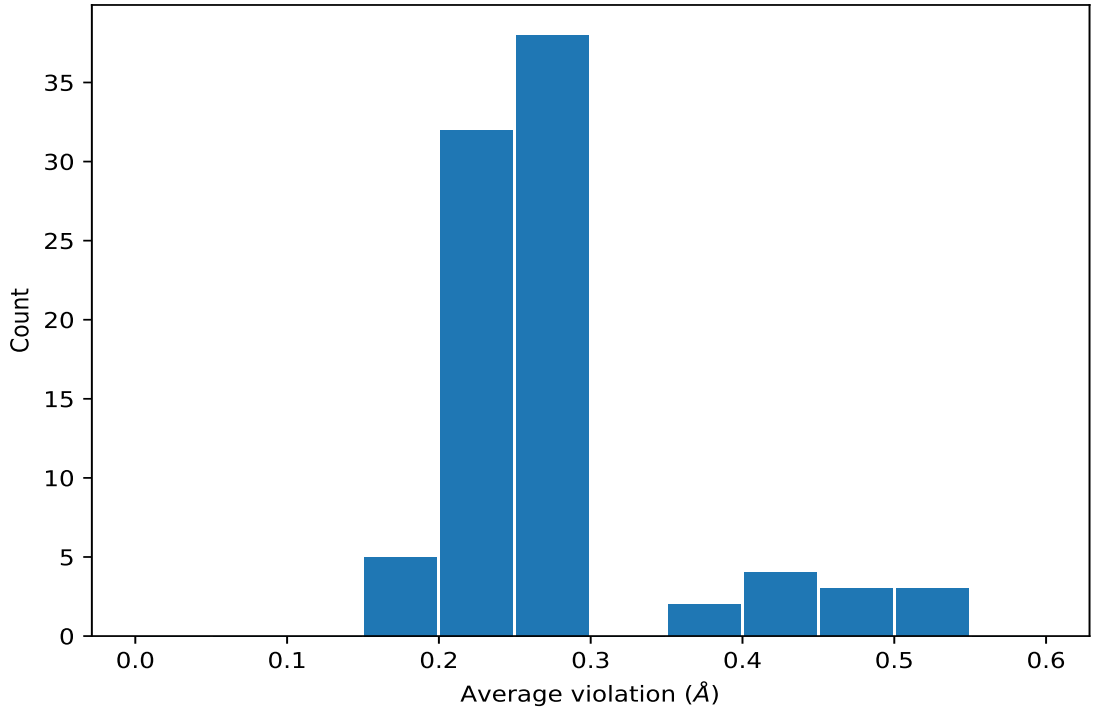


9.4 Most violated distance restraints in the ensemble

9.4.1 Histogram : Distribution of mean distance violations

The following histogram shows the distribution of the average value of the violation. The average is calculated for each restraint that is violated in more than one model over all the violated models

in the ensemble



9.4.2 Table: Most violated distance restraints

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

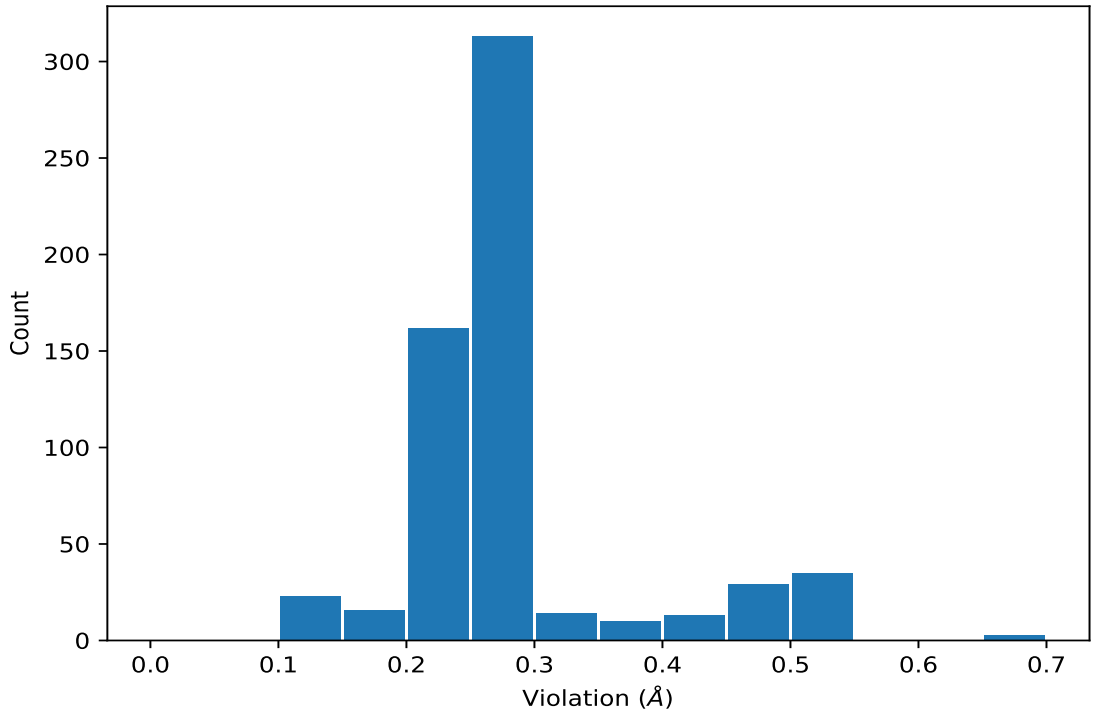
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(1,102)	2:25:B:LEU:CA	2:24:B:ALA:C	10	0.51	0.01	0.51
(1,38)	2:100:B:PRO:N	2:99:B:SER:CA	10	0.5	0.02	0.5
(1,39)	2:103:B:THR:N	2:102:B:ALA:CA	10	0.5	0.01	0.49
(1,37)	2:98:B:PRO:N	2:97:B:TYR:CA	10	0.49	0.01	0.5
(1,27)	2:65:B:PRO:N	2:64:B:ALA:CA	10	0.48	0.03	0.48
(1,31)	2:68:B:PRO:N	2:67:B:THR:CA	10	0.45	0.01	0.45
(1,118)	2:76:B:ILE:CG2	2:76:B:ILE:CG1	10	0.42	0.02	0.42
(1,176)	2:68:B:PRO:CG	2:68:B:PRO:CB	10	0.29	0.01	0.29
(1,177)	2:68:B:PRO:CG	2:68:B:PRO:CB	10	0.29	0.01	0.29
(1,174)	2:115:B:PRO:CG	2:115:B:PRO:CB	10	0.28	0.01	0.29

[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)
(1,229)	2:68:B:PRO:CB	2:70:B:ARG:C	1	0.7
(1,714)	2:62:B:ALA:CA	2:59:B:LEU:C	5	0.69
(1,230)	2:155:B:ILE:CG1	2:151:B:ILE:C	1	0.69
(1,102)	2:25:B:LEU:CA	2:24:B:ALA:C	10	0.53
(1,600)	2:63:B:ILE:CG1	2:60:B:ILE:C	9	0.52
(1,102)	2:25:B:LEU:CA	2:24:B:ALA:C	8	0.52
(1,39)	2:103:B:THR:N	2:102:B:ALA:CA	1	0.52
(1,38)	2:100:B:PRO:N	2:99:B:SER:CA	2	0.52
(1,102)	2:25:B:LEU:CA	2:24:B:ALA:C	3	0.51
(1,102)	2:25:B:LEU:CA	2:24:B:ALA:C	5	0.51

10 Dihedral-angle violation analysis

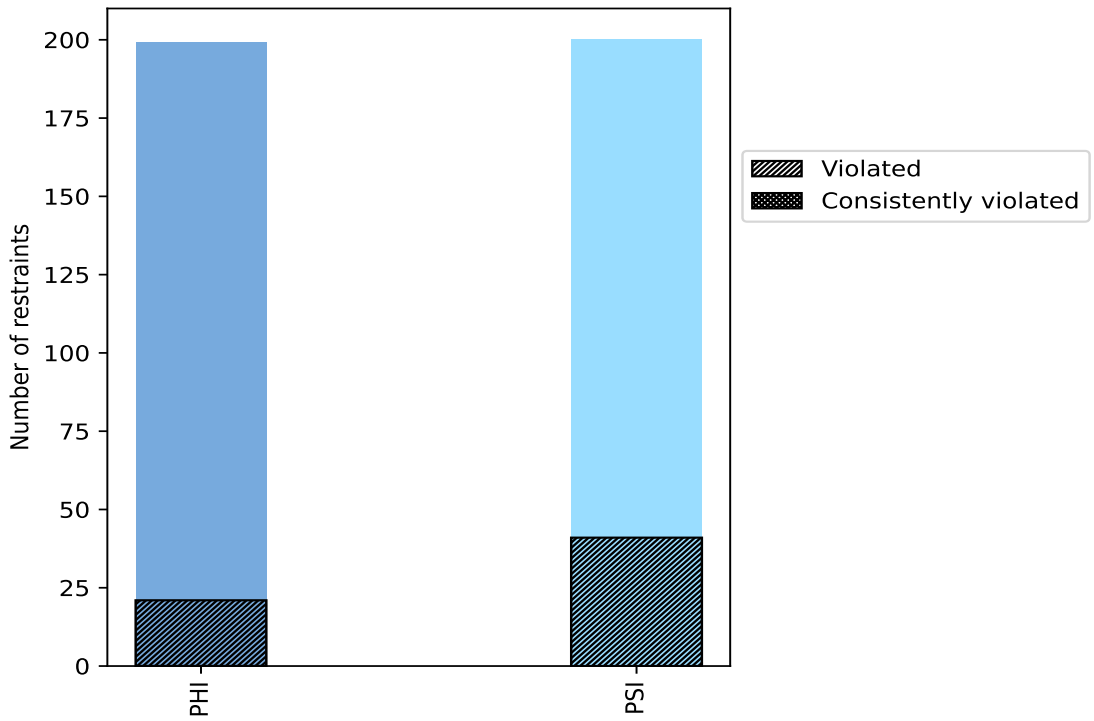
10.1 Summary of dihedral-angle violations

The following table provides the summary of dihedral-angle violations in different dihedral-angle types. Violations less than 1° are not included in the calculation.

Angle type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
PHI	199	49.9	21	10.6	5.3	0	0.0	0.0
PSI	200	50.1	41	20.5	10.3	0	0.0	0.0
Total	399	100.0	62	15.5	15.5	0	0.0	0.0

[1] percentage calculated with respect to total number of dihedral-angle restraints, [2] percentage calculated with respect to number of restraints in a particular dihedral-angle type, [3] violated in at least one model, [4] violated in all the models

10.1.1 Bar chart : Distribution of dihedral-angles and violations



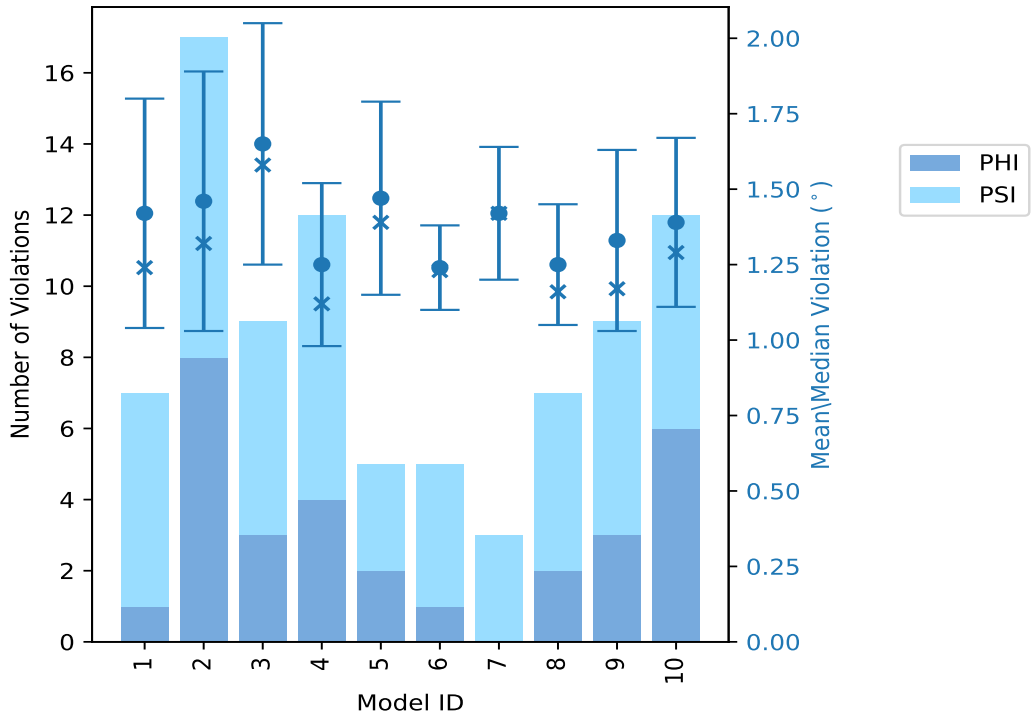
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

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

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

Model ID	Number of violations			Mean (°)	Max (°)	SD (°)	Median (°)
	PHI	PSI	Total				
1	1	6	7	1.42	2.2	0.38	1.24
2	8	9	17	1.46	2.4	0.43	1.32
3	3	6	9	1.65	2.64	0.4	1.58
4	4	8	12	1.25	1.83	0.27	1.12
5	2	3	5	1.47	1.94	0.32	1.39
6	1	4	5	1.24	1.4	0.14	1.23
7	0	3	3	1.42	1.69	0.22	1.42
8	2	5	7	1.25	1.57	0.2	1.16
9	3	6	9	1.33	1.94	0.3	1.17
10	6	6	12	1.39	2.1	0.28	1.29

10.2.1 Bar graph : Dihedral violation statistics for each model (i)



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

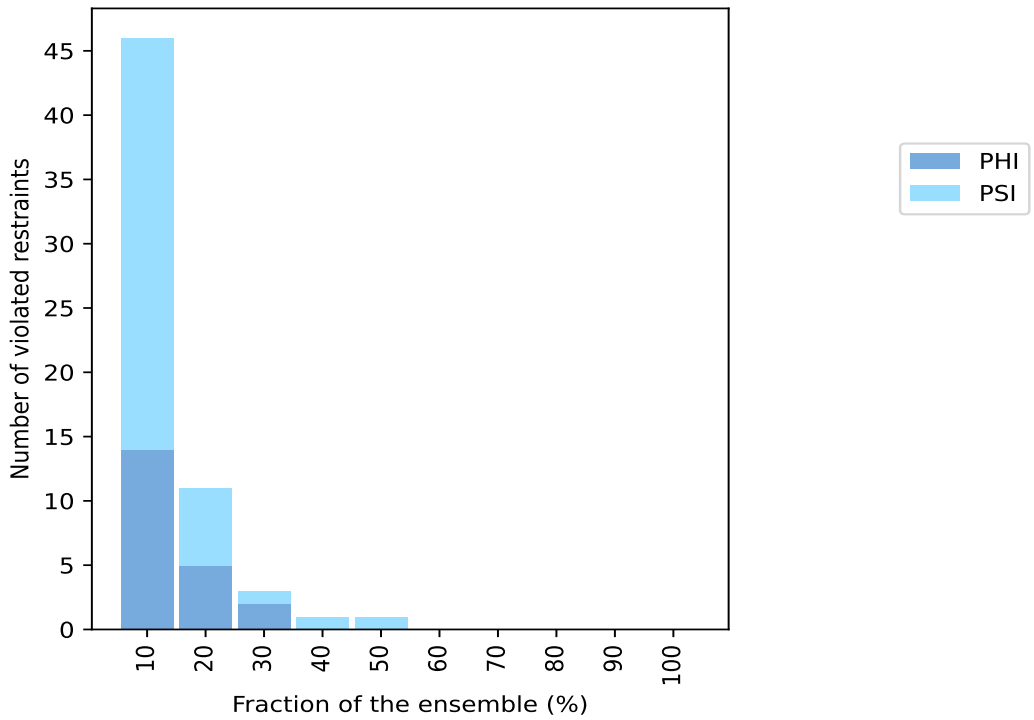
10.3 Dihedral-angle violation statistics for the ensemble (i)

Violation analysis may find that some restraints are violated in very few models and some are violated in most of models. The following table provides this information as number of violated restraints for a given fraction of ensemble.

Number of violated restraints			Fraction of the ensemble	
PHI	PSI	Total	Count[1]	%
14	32	46	1	10.0
5	6	11	2	20.0
2	1	3	3	30.0
0	1	1	4	40.0
0	1	1	5	50.0
0	0	0	6	60.0
0	0	0	7	70.0
0	0	0	8	80.0
0	0	0	9	90.0
0	0	0	10	100.0

[1] Number of models with violations

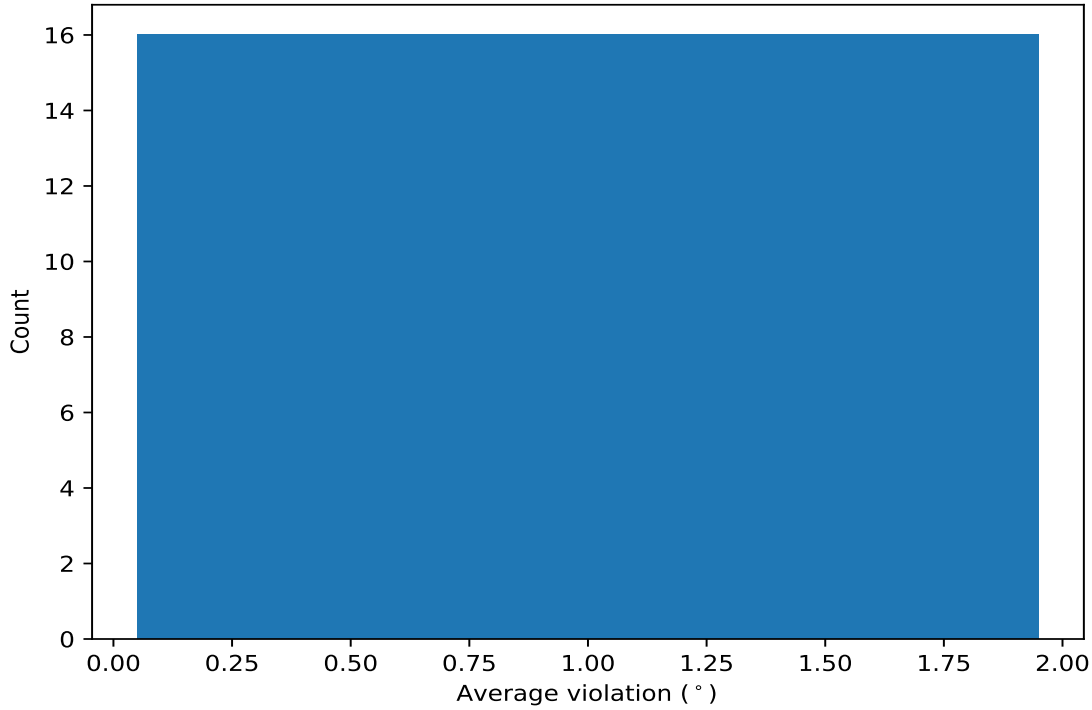
10.3.1 Bar graph : Dihedral-angle Violation statistics for the ensemble (i)



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.

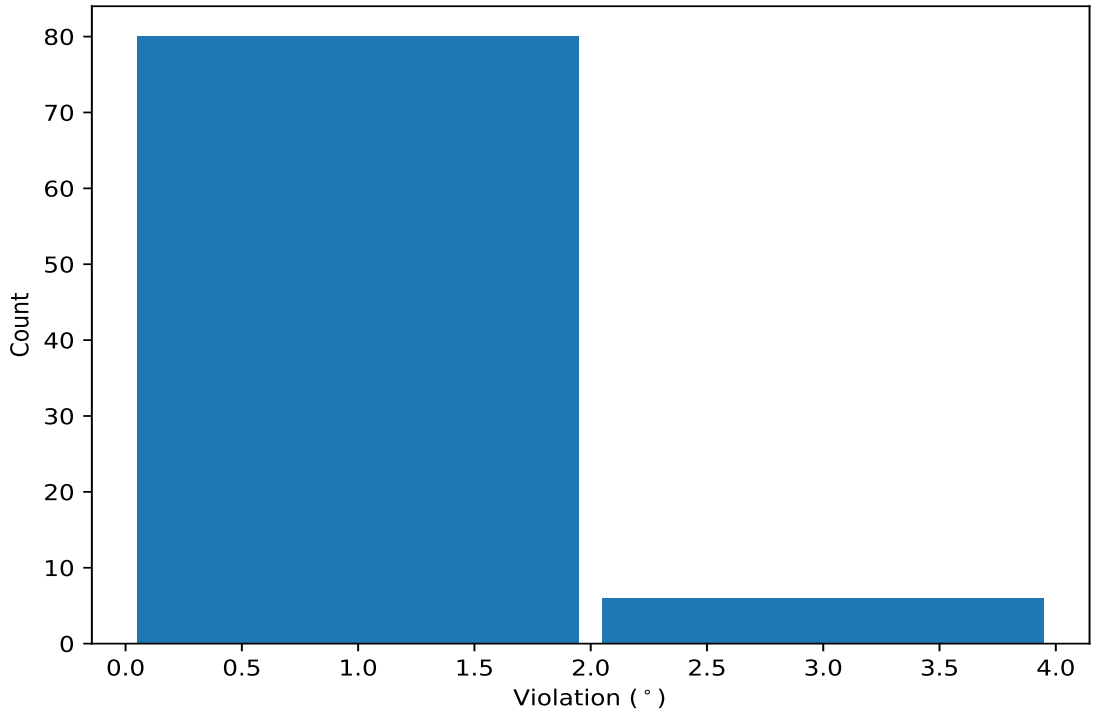
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,125)	1:95:A:GLY:N	1:95:A:GLY:CA	1:95:A:GLY:C	1:96:A:VAL:N	5	1.61	0.41	1.43
(1,43)	1:36:A:PHE:N	1:36:A:PHE:CA	1:36:A:PHE:C	1:37:A:GLU:N	4	1.34	0.24	1.3
(1,220)	1:158:A:LYS:C	1:159:A:TYR:N	1:159:A:TYR:CA	1:159:A:TYR:C	3	1.92	0.4	1.94
(1,214)	1:153:A:MET:C	1:154:A:PHE:N	1:154:A:PHE:CA	1:154:A:PHE:C	3	1.52	0.25	1.65
(1,175)	1:125:A:ALA:N	1:125:A:ALA:CA	1:125:A:ALA:C	1:126:A:TRP:N	3	1.19	0.12	1.13
(1,24)	1:21:A:GLN:N	1:21:A:GLN:CA	1:21:A:GLN:C	1:22:A:VAL:N	2	1.88	0.77	1.88
(1,128)	1:97:A:GLN:C	1:98:A:LYS:N	1:98:A:LYS:CA	1:98:A:LYS:C	2	1.74	0.36	1.74
(1,221)	1:159:A:TYR:N	1:159:A:TYR:CA	1:159:A:TYR:C	1:160:A:GLN:N	2	1.73	0.2	1.73
(1,33)	1:31:A:PRO:N	1:31:A:PRO:CA	1:31:A:PRO:C	1:32:A:HIS:N	2	1.7	0.22	1.7
(1,222)	1:159:A:TYR:C	1:160:A:GLN:N	1:160:A:GLN:CA	1:160:A:GLN:C	2	1.64	0.08	1.64

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

10.5 All violated dihedral-angle restraints (i)

10.5.1 Histogram : Distribution of violations (i)

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 (i)

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 (°)
(1,24)	1:21:A:GLN:N	1:21:A:GLN:CA	1:21:A:GLN:C	1:22:A:VAL:N	3	2.64
(1,220)	1:158:A:LYS:C	1:159:A:TYR:N	1:159:A:TYR:CA	1:159:A:TYR:C	2	2.4
(1,366)	2:103:B:THR:C	2:104:B:CYS:N	2:104:B:CYS:CA	2:104:B:CYS:C	2	2.23
(1,125)	1:95:A:GLY:N	1:95:A:GLY:CA	1:95:A:GLY:C	1:96:A:VAL:N	2	2.23
(1,37)	1:33:A:ALA:N	1:33:A:ALA:CA	1:33:A:ALA:C	1:34:A:TYR:N	1	2.2
(1,128)	1:97:A:GLN:C	1:98:A:LYS:N	1:98:A:LYS:CA	1:98:A:LYS:C	10	2.1
(1,220)	1:158:A:LYS:C	1:159:A:TYR:N	1:159:A:TYR:CA	1:159:A:TYR:C	9	1.94
(1,125)	1:95:A:GLY:N	1:95:A:GLY:CA	1:95:A:GLY:C	1:96:A:VAL:N	5	1.94
(1,221)	1:159:A:TYR:N	1:159:A:TYR:CA	1:159:A:TYR:C	1:160:A:GLN:N	2	1.93
(1,33)	1:31:A:PRO:N	1:31:A:PRO:CA	1:31:A:PRO:C	1:32:A:HIS:N	3	1.93