wwPDB NMR Structure Validation Summary Report

Jun 4, 2023 – 11:45 PM EDT

PDB ID	:	2M1E
BMRB ID	:	18859
Title	:	Biosynthetic engineered B28K-B29P human insulin monomer structure in in water solutions.
Authors	:	Bocian, W.; Kozerski, L.
Deposited on	:	2012-11-26

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:

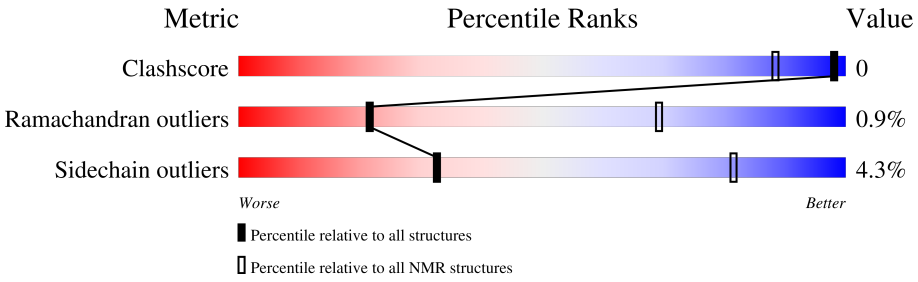
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal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

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 48%.

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	21	100%
2	B	30	80% 7% 13%

2 Ensemble composition and analysis

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:1-A:21, B:1-B:26 (47)	0.61	4

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 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 786 atoms, of which 381 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Insulin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	21	312	99	149	25	35	4	0

- Molecule 2 is a protein called Insulin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	30	474	158	232	40	42	2	0

There are 2 discrepancies between the modelled and reference sequences:

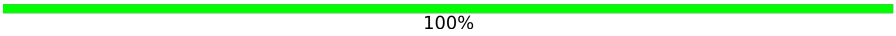
Chain	Residue	Modelled	Actual	Comment	Reference
B	28	LYS	PRO	engineered mutation	UNP P01308
B	29	PRO	LYS	engineered mutation	UNP P01308

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.

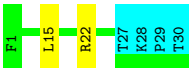
• Molecule 1: Insulin

Chain A: 

There are no outlier residues in this chain.

• Molecule 2: Insulin

Chain B: 

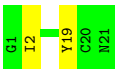


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

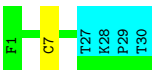
• Molecule 1: Insulin

Chain A: 



• Molecule 2: Insulin

Chain B: 



5 Refinement protocol and experimental data overview

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

Of the 100 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	11
Amber	structure solution	11
CYANA	refinement	2.1

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	2
Total number of shifts	323
Number of shifts mapped to atoms	323
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	48%

6 Model quality

6.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.72±0.01	0±0/164 (0.0± 0.0%)	0.87±0.03	0±0/220 (0.0± 0.0%)
2	B	0.82±0.01	0±0/218 (0.0± 0.0%)	0.96±0.05	1±1/295 (0.3± 0.3%)
All	All	0.78	0/7640 (0.0%)	0.93	16/10300 (0.2%)

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
2	B	0.0±0.0	0.1±0.3
All	All	0	2

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
2	B	22	ARG	NE-CZ-NH1	6.59	123.59	120.30	16	11
2	B	22	ARG	NE-CZ-NH2	5.51	123.06	120.30	6	3
2	B	22	ARG	NH1-CZ-NH2	-5.11	113.78	119.40	12	2

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	22	ARG	Sidechain	2

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	163	149	149	0±0
All	All	7480	6940	6940	4

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:2:ILE:HD13	1:A:19:TYR:CE1	0.48	2.42	1	2
1:A:2:ILE:HD12	1:A:19:TYR:CD1	0.44	2.47	4	2

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	19/21 (90%)	18±1 (93±5%)	1±1 (6±5%)	0±0 (1±2%)	18	66
2	B	25/30 (83%)	23±1 (91±4%)	2±1 (9±3%)	0±1 (1±2%)	24	71
All	All	880/1020 (86%)	805 (91%)	67 (8%)	8 (1%)	21	69

All 5 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	10	ILE	4
2	B	26	TYR	1
2	B	23	GLY	1
2	B	24	PHE	1
2	B	5	HIS	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	20/20 (100%)	19±1 (97±4%)	1±1 (3±4%)	46 90
2	B	22/26 (85%)	21±1 (94±4%)	1±1 (6±4%)	24 73
All	All	840/920 (91%)	804 (96%)	36 (4%)	33 81

5 of 8 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)
2	B	15	LEU	11
2	B	7	CYS	10
1	A	21	ASN	6
1	A	6	CYS	3
2	B	3	ASN	3

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 48% for the well-defined parts and 48% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

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	129
Number of shifts mapped to atoms	129
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

7.1.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

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 21%, i.e. 129 atoms were assigned a chemical shift out of a possible 624. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	40/239 (17%)	40/98 (41%)	0/94 (0%)	0/47 (0%)
Sidechain	81/305 (27%)	81/200 (40%)	0/96 (0%)	0/9 (0%)
Aromatic	8/80 (10%)	8/39 (21%)	0/39 (0%)	0/2 (0%)
Overall	129/624 (21%)	129/337 (38%)	0/229 (0%)	0/58 (0%)

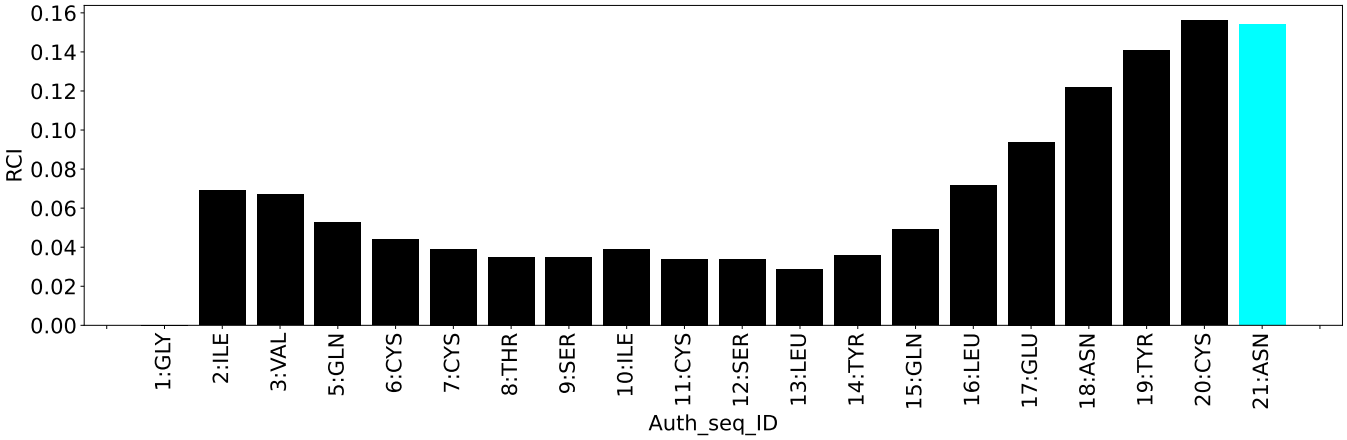
7.1.4 Statistically unusual chemical shifts

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:



7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2_dup*

7.2.1 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	194
Number of shifts mapped to atoms	194
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

7.2.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

7.2.3 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 27%, i.e. 168 atoms were assigned a chemical shift out of a possible 624. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	50/239 (21%)	50/98 (51%)	0/94 (0%)	0/47 (0%)
Sidechain	97/305 (32%)	97/200 (48%)	0/96 (0%)	0/9 (0%)
Aromatic	21/80 (26%)	21/39 (54%)	0/39 (0%)	0/2 (0%)
Overall	168/624 (27%)	168/337 (50%)	0/229 (0%)	0/58 (0%)

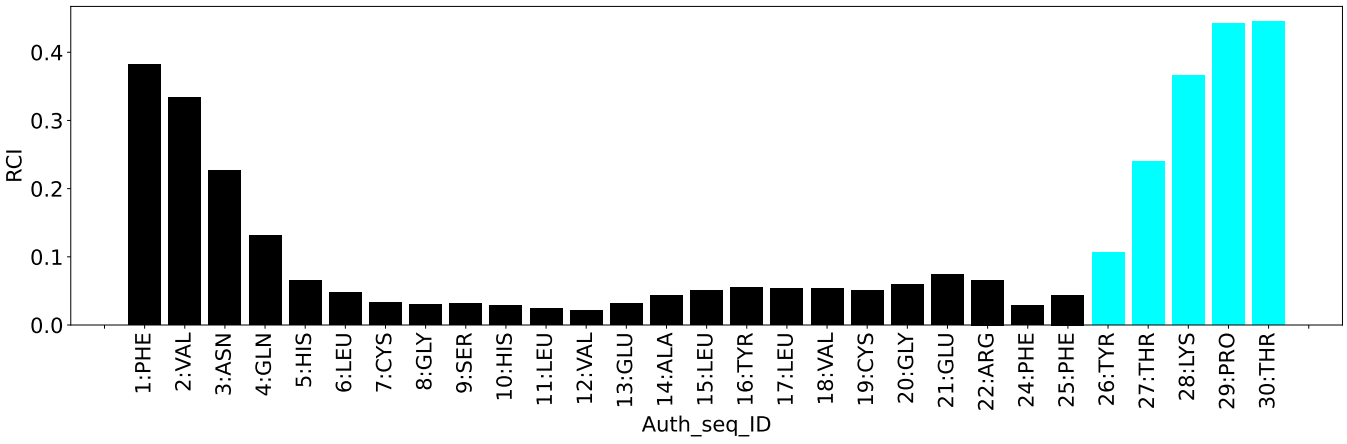
7.2.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.2.5 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 B:



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	782
Intra-residue (|i-j|=0)	249
Sequential (|i-j|=1)	215
Medium range (|i-j|>1 and |i-j|<5)	162
Long range (|i-j|≥5)	51
Inter-chain	105
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	152
Number of unmapped restraints	0
Number of restraints per residue	18.3
Number of long range restraints per residue[1]	1.0

[1]Long range hydrogen bonds and disulfide bonds are counted as long range restraints while calculating the number of long range restraints per residue

8.2 Residual restraint violations

This section provides the overview of the restraint violations analysis. The violations are binned as small, medium and large violations based on its absolute value. Average number of violations per model is calculated by dividing the total number of violations in each bin by the size of the ensemble.

8.2.1 Average number of distance violations per model

Distance violations less than 0.1 Å are not included in the calculation.

Bins (Å)	Average number of violations per model	Max (Å)
0.1-0.2 (Small)	1.5	0.2
0.2-0.5 (Medium)	0.2	0.5
>0.5 (Large)	0.9	0.73

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)	6.1	10.0
10.0-20.0 (Medium)	1.4	19.7
>20.0 (Large)	25.6	105.0

9 Distance violation analysis (i)

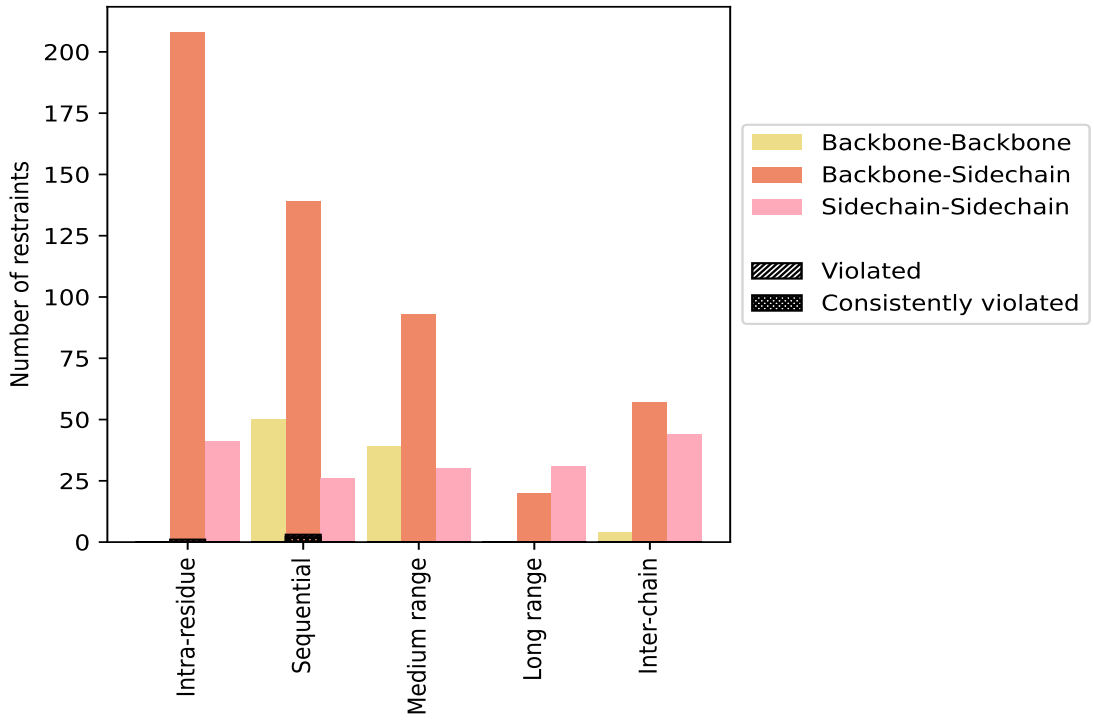
9.1 Summary of distance violations (i)

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)	249	31.8	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	208	26.6	1	0.5	0.1	0	0.0	0.0
Sidechain-Sidechain	41	5.2	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	215	27.5	3	1.4	0.4	2	0.9	0.3
Backbone-Backbone	50	6.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	139	17.8	3	2.2	0.4	2	1.4	0.3
Sidechain-Sidechain	26	3.3	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	162	20.7	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	39	5.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	93	11.9	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	30	3.8	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	51	6.5	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	20	2.6	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	31	4.0	0	0.0	0.0	0	0.0	0.0
Inter-chain	105	13.4	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	4	0.5	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	57	7.3	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	44	5.6	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	782	100.0	4	0.5	0.5	2	0.3	0.3
Backbone-Backbone	93	11.9	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	517	66.1	4	0.8	0.5	2	0.4	0.3
Sidechain-Sidechain	172	22.0	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	1	2	0	0	0	3	0.31	0.6	0.21	0.19
2	1	2	0	0	0	3	0.35	0.72	0.27	0.19
3	1	2	0	0	0	3	0.33	0.67	0.24	0.19
4	1	2	0	0	0	3	0.31	0.63	0.23	0.19
5	0	2	0	0	0	2	0.39	0.58	0.19	0.39
6	1	2	0	0	0	3	0.33	0.66	0.24	0.2
7	0	3	0	0	0	3	0.32	0.6	0.2	0.19
8	0	3	0	0	0	3	0.28	0.5	0.16	0.21
9	0	2	0	0	0	2	0.39	0.57	0.18	0.39
10	0	3	0	0	0	3	0.31	0.61	0.21	0.19
11	0	2	0	0	0	2	0.43	0.66	0.24	0.43

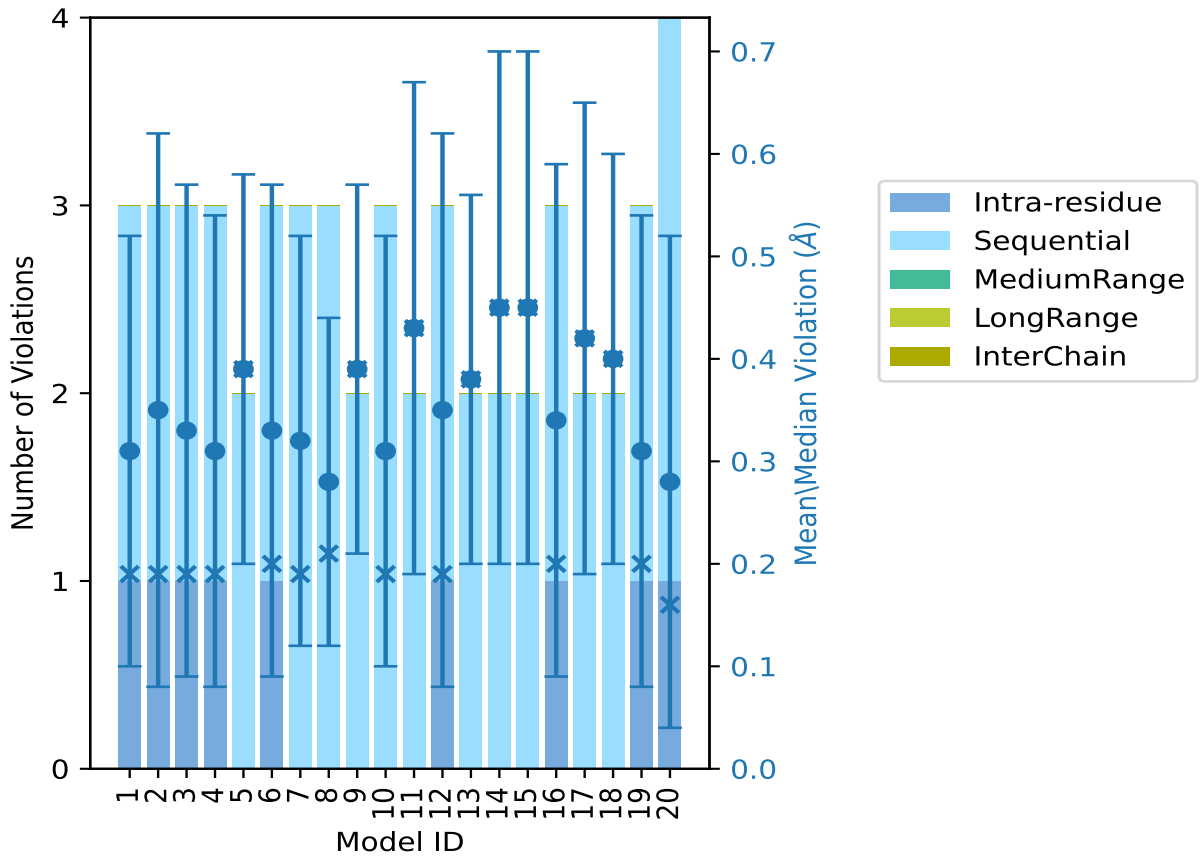
Continued on next page...

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	1	2	0	0	0	3	0.35	0.73	0.27	0.19
13	0	2	0	0	0	2	0.38	0.56	0.18	0.38
14	0	2	0	0	0	2	0.45	0.71	0.25	0.45
15	0	2	0	0	0	2	0.45	0.7	0.25	0.45
16	1	2	0	0	0	3	0.34	0.69	0.25	0.2
17	0	2	0	0	0	2	0.42	0.65	0.23	0.42
18	0	2	0	0	0	2	0.4	0.61	0.2	0.4
19	1	2	0	0	0	3	0.31	0.63	0.23	0.2
20	1	3	0	0	0	4	0.28	0.69	0.24	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

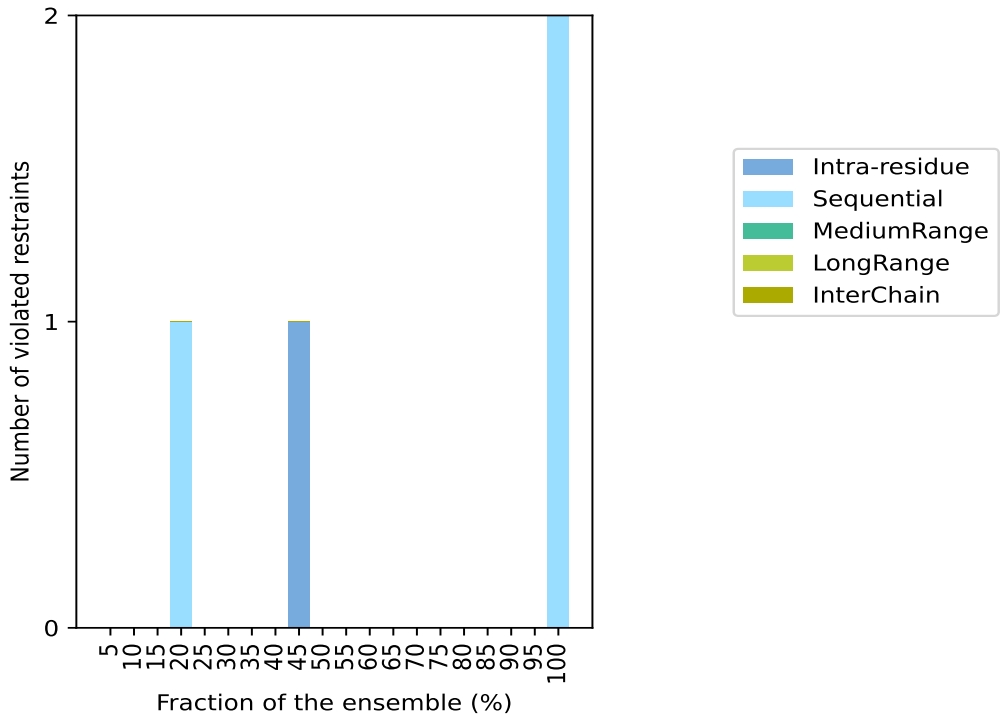
9.3 Distance violation statistics for the ensemble (i)

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, 778(IR:248, SQ:212, MR:162, LR:51, IC:105) 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	0	0	0	0	0	1	5.0
0	0	0	0	0	0	2	10.0
0	0	0	0	0	0	3	15.0
0	1	0	0	0	1	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
1	0	0	0	0	1	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	2	0	0	0	2	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

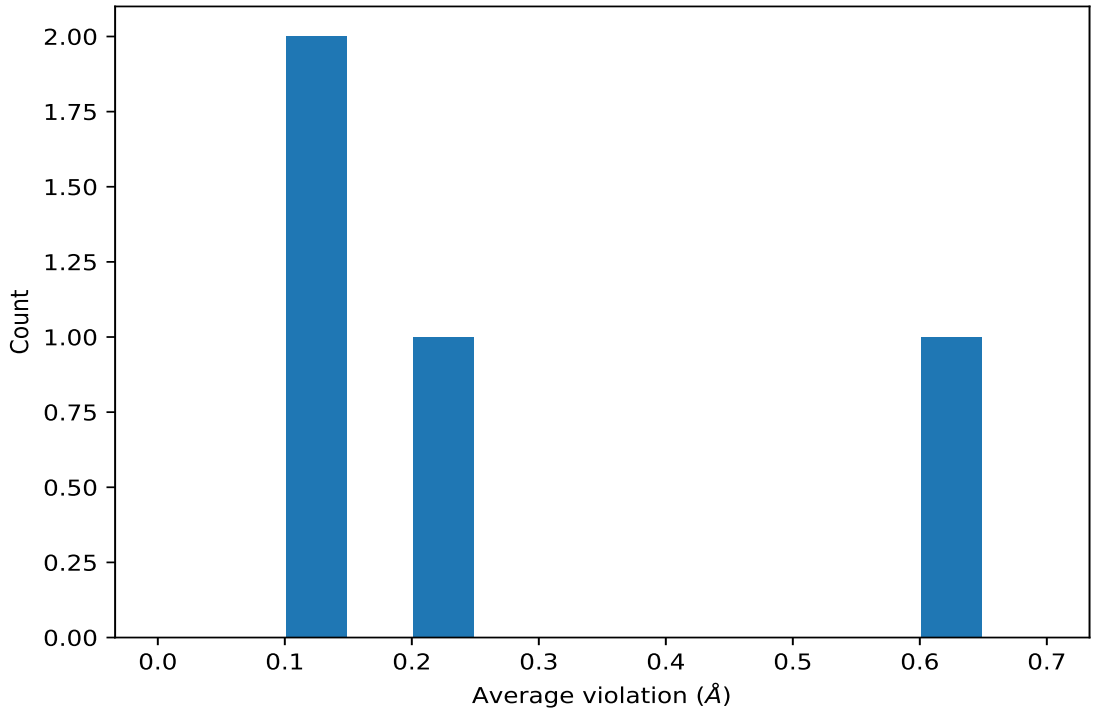
9.3.1 Bar graph : Distance violation statistics for the ensemble



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 violation for each restraint sorted by number of violated models and the mean value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint. Rows with same key represent combinatorial or ambiguous restraints and are counted as a single restraint.

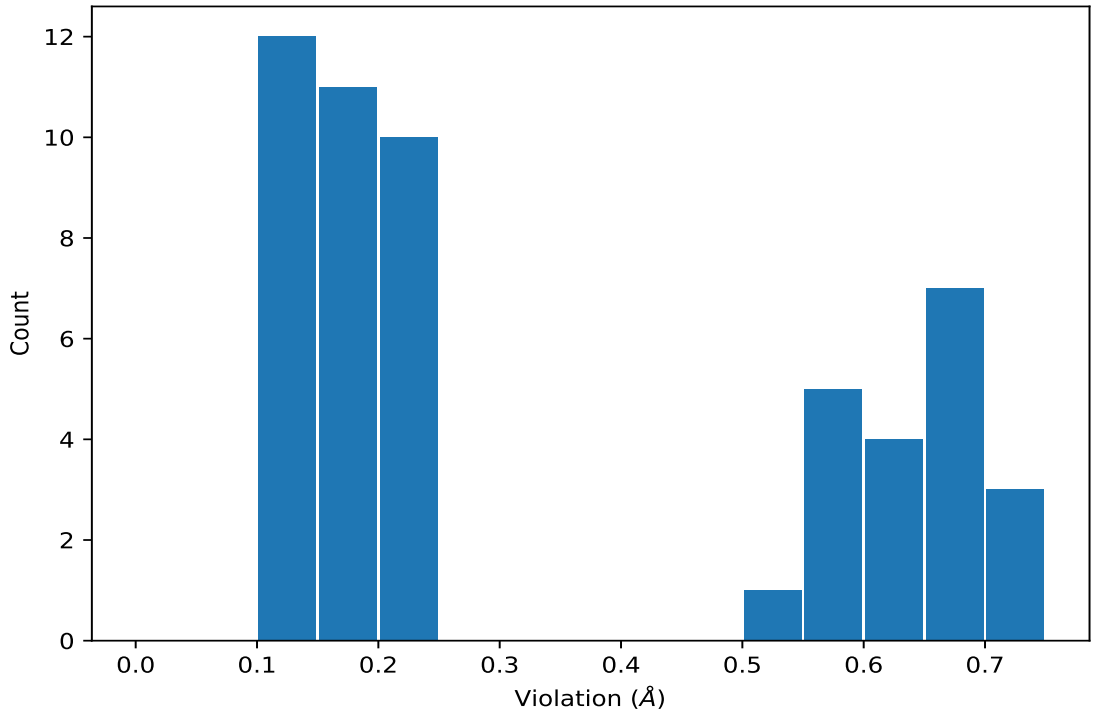
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	20	0.64	0.06	0.64
(1,255)	1:A:14:TYR:HB3	1:A:15:GLN:H	20	0.2	0.01	0.2
(1,263)	1:A:14:TYR:H	1:A:14:TYR:HB3	9	0.12	0.01	0.12
(1,424)	1:A:9:SER:HB3	1:A:10:ILE:H	4	0.14	0.01	0.14

[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)
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	12	0.73
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	2	0.72
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	14	0.71
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	15	0.7
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	16	0.69
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	20	0.69
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	3	0.67
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	6	0.66
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	11	0.66
(1,280)	1:A:13:LEU:H	1:A:14:TYR:HB3	17	0.65

10 Dihedral-angle violation analysis

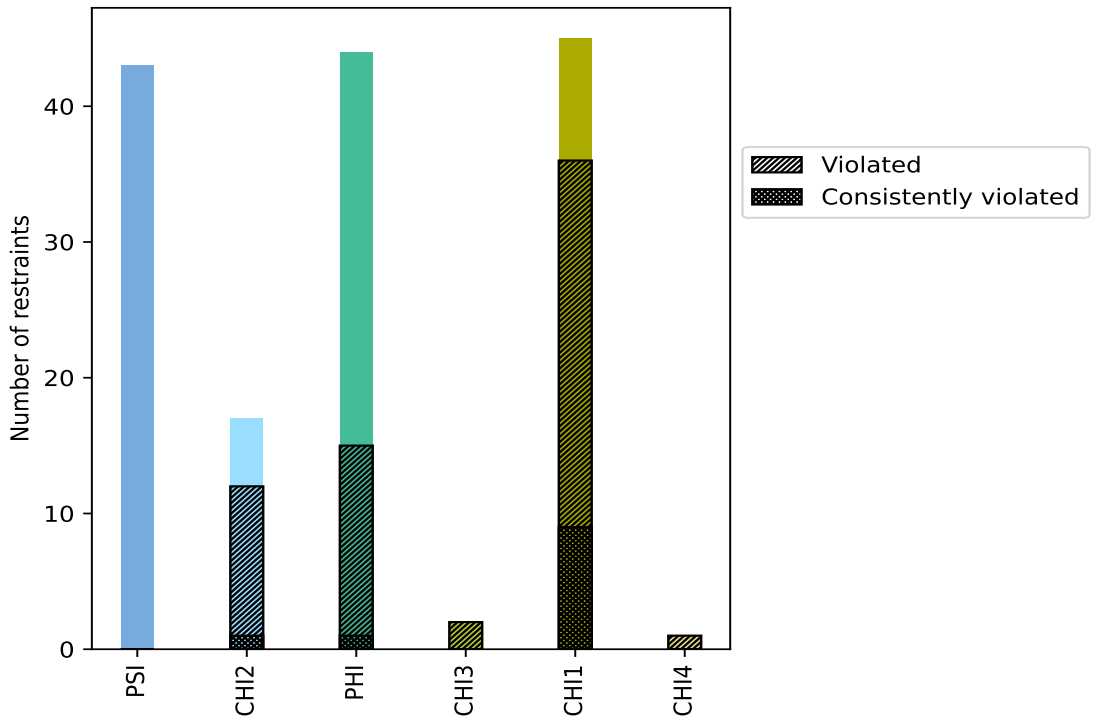
10.1 Summary of dihedral-angle violations

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

Angle type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
PSI	43	28.3	0	0.0	0.0	0	0.0	0.0
CHI2	17	11.2	12	70.6	7.9	1	5.9	0.7
PHI	44	28.9	15	34.1	9.9	1	2.3	0.7
CHI3	2	1.3	2	100.0	1.3	0	0.0	0.0
CHI1	45	29.6	36	80.0	23.7	9	20.0	5.9
CHI4	1	0.7	1	100.0	0.7	0	0.0	0.0
Total	152	100.0	66	43.4	43.4	11	7.2	7.2

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

10.1.1 Bar chart : Distribution of dihedral-angles and violations



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

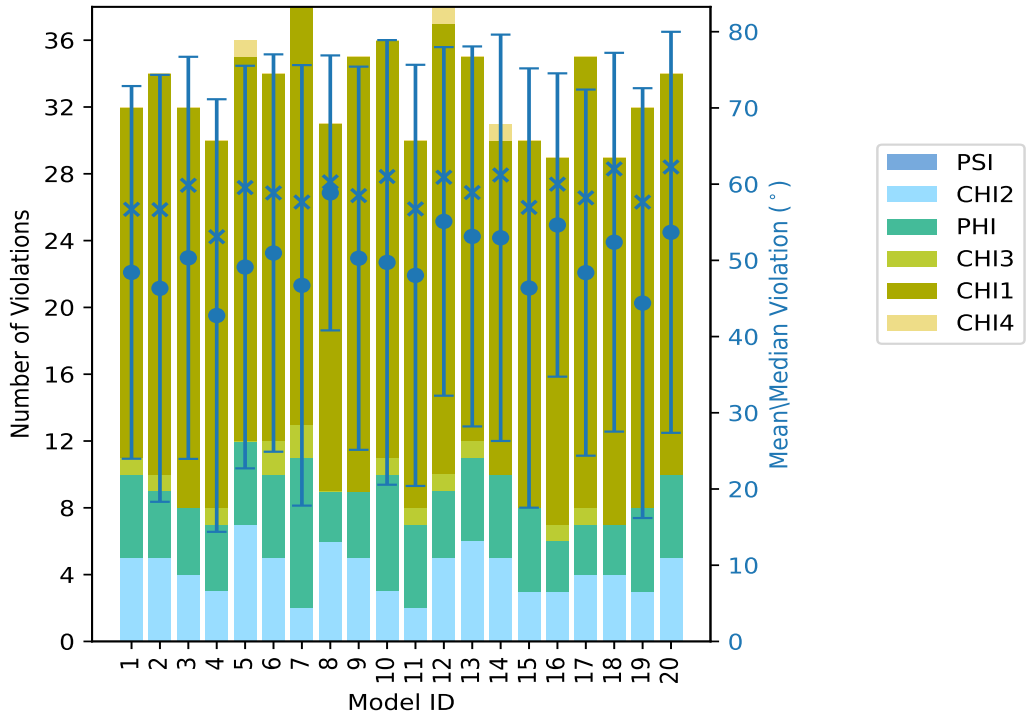
respective categories

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

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

Model ID	Number of violations							Mean (°)	Max (°)	SD (°)	Median
	PSI	CHI2	PHI	CHI3	CHI1	CHI4	Total				
1	0	5	5	1	21	0	32	48.42	95.7	24.45	56.7
2	0	5	4	1	24	0	34	46.33	98.0	28.01	56.65
3	0	4	4	0	24	0	32	50.33	101.2	26.38	59.85
4	0	3	4	1	22	0	30	42.76	103.0	28.38	53.1
5	0	7	5	0	23	1	36	49.12	102.6	26.41	59.55
6	0	5	5	2	22	0	34	50.96	103.4	26.07	58.85
7	0	2	9	2	25	0	38	46.72	101.2	28.9	57.65
8	0	6	3	0	22	0	31	58.85	97.6	18.04	60.3
9	0	5	4	0	26	0	35	50.28	99.9	25.14	58.5
10	0	3	7	1	25	0	36	49.73	105.0	29.17	61.0
11	0	2	5	1	22	0	30	48.03	95.9	27.63	56.75
12	0	5	4	1	27	1	38	55.11	103.7	22.87	60.9
13	0	6	5	1	23	0	35	53.14	96.7	24.93	58.9
14	0	5	5	0	20	1	31	52.96	102.1	26.66	61.2
15	0	3	5	0	22	0	30	46.37	103.7	28.82	56.95
16	0	3	3	1	22	0	29	54.64	101.2	19.9	60.0
17	0	4	3	1	27	0	35	48.39	82.8	24.02	58.2
18	0	4	3	0	22	0	29	52.38	102.7	24.85	62.0
19	0	3	5	0	24	0	32	44.39	102.8	28.2	57.65
20	0	5	5	0	24	0	34	53.68	104.9	26.31	62.25

10.2.1 Bar graph : Dihedral violation statistics for each model



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

10.3 Dihedral-angle violation statistics for the ensemble

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

Number of violated restraints							Fraction of the ensemble	
PSI	CHI2	PHI	CHI3	CHI1	CHI4	Total	Count[1]	%
0	2	5	0	3	0	10	1	5.0
0	1	0	0	1	0	2	2	10.0
0	2	2	0	0	1	5	3	15.0
0	1	2	1	1	0	5	4	20.0
0	1	1	0	1	0	3	5	25.0
0	0	1	0	3	0	4	6	30.0
0	0	1	0	0	0	1	7	35.0
0	1	0	0	2	0	3	8	40.0
0	0	0	1	1	0	2	9	45.0
0	1	0	0	1	0	2	10	50.0
0	0	0	0	0	0	0	11	55.0

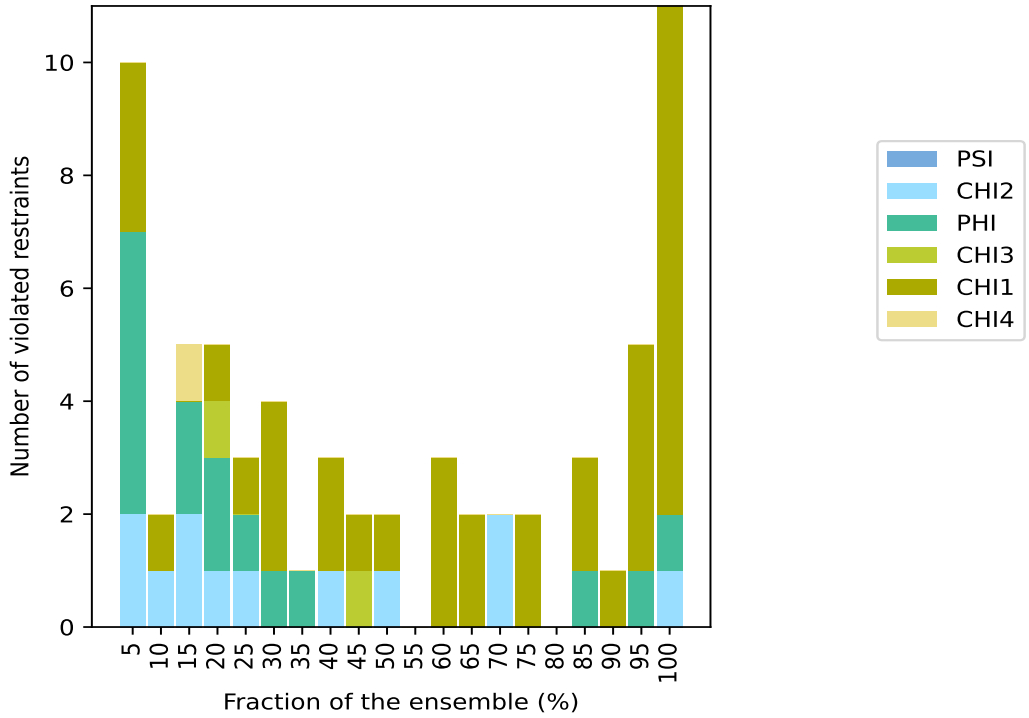
Continued on next page...

Continued from previous page...

Number of violated restraints							Fraction of the ensemble	
PSI	CHI2	PHI	CHI3	CHI1	CHI4	Total	Count[1]	%
0	0	0	0	3	0	3	12	60.0
0	0	0	0	2	0	2	13	65.0
0	2	0	0	0	0	2	14	70.0
0	0	0	0	2	0	2	15	75.0
0	0	0	0	0	0	0	16	80.0
0	0	1	0	2	0	3	17	85.0
0	0	0	0	1	0	1	18	90.0
0	0	1	0	4	0	5	19	95.0
0	1	1	0	9	0	11	20	100.0

[1] Number of models with violations

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

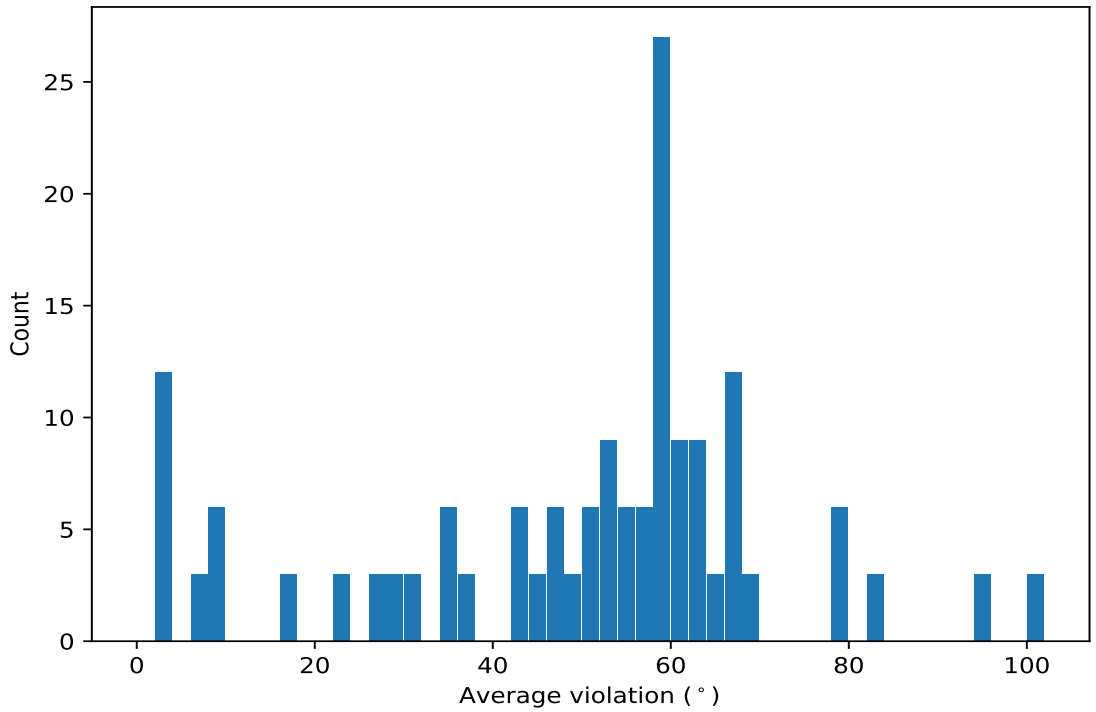


10.4 Most violated dihedral-angle restraints in the ensemble

10.4.1 Histogram : Distribution of mean dihedral-angle violations

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

in the ensemble



10.4.2 Table: Most violated dihedral-angle restraints

The following table provides the mean and the standard deviation of the 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
(1,99)	1:A:10:ILE:N	1:A:10:ILE:CA	1:A:10:ILE:CB	1:A:10:ILE:CG1	20	68.91	5.27	69.8
(1,99)	1:A:10:ILE:N	1:A:10:ILE:CA	1:A:10:ILE:CB	1:A:10:ILE:CG1	20	68.91	5.27	69.8
(1,99)	1:A:10:ILE:N	1:A:10:ILE:CA	1:A:10:ILE:CB	1:A:10:ILE:CG1	20	68.91	5.27	69.8
(1,105)	1:A:14:TYR:N	1:A:14:TYR:CA	1:A:14:TYR:CB	1:A:14:TYR:CG	20	66.35	1.82	66.45
(1,105)	1:A:14:TYR:N	1:A:14:TYR:CA	1:A:14:TYR:CB	1:A:14:TYR:CG	20	66.35	1.82	66.45
(1,105)	1:A:14:TYR:N	1:A:14:TYR:CA	1:A:14:TYR:CB	1:A:14:TYR:CG	20	66.35	1.82	66.45
(1,126)	2:B:10:HIS:N	2:B:10:HIS:CA	2:B:10:HIS:CB	2:B:10:HIS:CG	20	66.1	4.36	67.75
(1,126)	2:B:10:HIS:N	2:B:10:HIS:CA	2:B:10:HIS:CB	2:B:10:HIS:CG	20	66.1	4.36	67.75
(1,126)	2:B:10:HIS:N	2:B:10:HIS:CA	2:B:10:HIS:CB	2:B:10:HIS:CG	20	66.1	4.36	67.75
(1,122)	2:B:6:LEU:N	2:B:6:LEU:CA	2:B:6:LEU:CB	2:B:6:LEU:CG	20	63.43	5.13	61.75
(1,122)	2:B:6:LEU:N	2:B:6:LEU:CA	2:B:6:LEU:CB	2:B:6:LEU:CG	20	63.43	5.13	61.75
(1,122)	2:B:6:LEU:N	2:B:6:LEU:CA	2:B:6:LEU:CB	2:B:6:LEU:CG	20	63.43	5.13	61.75
(1,108)	1:A:16:LEU:N	1:A:16:LEU:CA	1:A:16:LEU:CB	1:A:16:LEU:CG	20	63.0	2.8	63.5
(1,108)	1:A:16:LEU:N	1:A:16:LEU:CA	1:A:16:LEU:CB	1:A:16:LEU:CG	20	63.0	2.8	63.5
(1,108)	1:A:16:LEU:N	1:A:16:LEU:CA	1:A:16:LEU:CB	1:A:16:LEU:CG	20	63.0	2.8	63.5
(1,113)	1:A:19:TYR:N	1:A:19:TYR:CA	1:A:19:TYR:CB	1:A:19:TYR:CG	20	59.83	2.86	59.3
(1,113)	1:A:19:TYR:N	1:A:19:TYR:CA	1:A:19:TYR:CB	1:A:19:TYR:CG	20	59.83	2.86	59.3
(1,113)	1:A:19:TYR:N	1:A:19:TYR:CA	1:A:19:TYR:CB	1:A:19:TYR:CG	20	59.83	2.86	59.3
(1,125)	2:B:9:SER:N	2:B:9:SER:CA	2:B:9:SER:CB	2:B:9:SER:OG	20	57.81	18.42	65.5
(1,125)	2:B:9:SER:N	2:B:9:SER:CA	2:B:9:SER:CB	2:B:9:SER:OG	20	57.81	18.42	65.5

Continued on next page...

Continued from previous page...

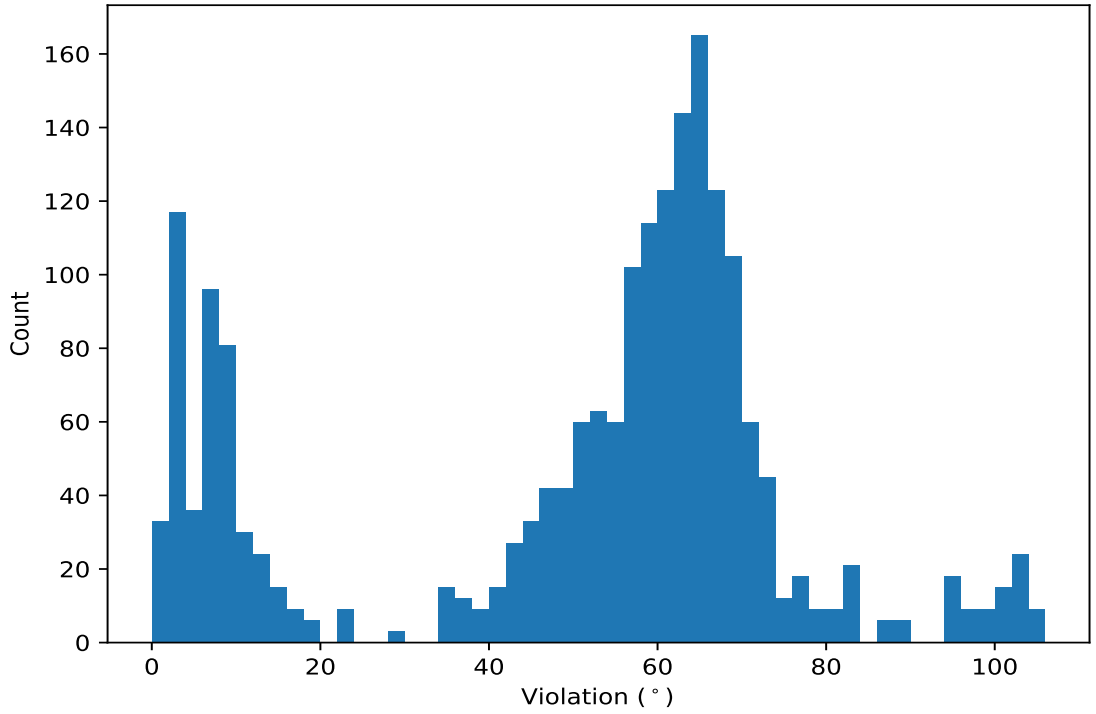
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,125)	2:B:9:SER:N	2:B:9:SER:CA	2:B:9:SER:CB	2:B:9:SER:OG	20	57.81	18.42	65.5
(1,110)	1:A:17:GLU:N	1:A:17:GLU:CA	1:A:17:GLU:CB	1:A:17:GLU:CG	20	43.46	5.05	43.05
(1,110)	1:A:17:GLU:N	1:A:17:GLU:CA	1:A:17:GLU:CB	1:A:17:GLU:CG	20	43.46	5.05	43.05
(1,110)	1:A:17:GLU:N	1:A:17:GLU:CA	1:A:17:GLU:CB	1:A:17:GLU:CG	20	43.46	5.05	43.05
(1,97)	1:A:8:THR:N	1:A:8:THR:CA	1:A:8:THR:CB	1:A:8:THR:OG1	20	34.39	30.16	10.2
(1,97)	1:A:8:THR:N	1:A:8:THR:CA	1:A:8:THR:CB	1:A:8:THR:OG1	20	34.39	30.16	10.2
(1,97)	1:A:8:THR:N	1:A:8:THR:CA	1:A:8:THR:CB	1:A:8:THR:OG1	20	34.39	30.16	10.2
(1,128)	2:B:11:LEU:CA	2:B:11:LEU:CB	2:B:11:LEU:CG	2:B:11:LEU:CD1	20	27.11	17.22	16.7

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

10.5 All violated dihedral-angle restraints

10.5.1 Histogram : Distribution of violations

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



10.5.2 Table: All violated dihedral-angle restraints

The following table provides the list of violationsfor the 10 worst performing restraints, sorted by the violation value. The Key (restraint list ID, restraint ID) is the unique identifier for a given restraint.

Key	Atom-1	Atom-2	Atom-3	Atom-4	Model ID	Violation (°)
(1,83)	2:B:25:PHE:C	2:B:26:TYR:N	2:B:26:TYR:CA	2:B:26:TYR:C	10	105.0
(1,83)	2:B:25:PHE:C	2:B:26:TYR:N	2:B:26:TYR:CA	2:B:26:TYR:C	10	105.0
(1,83)	2:B:25:PHE:C	2:B:26:TYR:N	2:B:26:TYR:CA	2:B:26:TYR:C	10	105.0
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	20	104.9
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	20	104.9
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	20	104.9
(1,39)	1:A:20:CYS:C	1:A:21:ASN:N	1:A:21:ASN:CA	1:A:21:ASN:C	20	104.3
(1,39)	1:A:20:CYS:C	1:A:21:ASN:N	1:A:21:ASN:CA	1:A:21:ASN:C	20	104.3
(1,39)	1:A:20:CYS:C	1:A:21:ASN:N	1:A:21:ASN:CA	1:A:21:ASN:C	20	104.3
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	12	103.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	12	103.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	12	103.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	15	103.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	15	103.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	15	103.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	6	103.4
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	6	103.4
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	6	103.4
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	4	103.0
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	4	103.0
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	4	103.0
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	19	102.8
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	19	102.8
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	19	102.8
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	18	102.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	18	102.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	18	102.7
(1,47)	2:B:4:GLN:C	2:B:5:HIS:N	2:B:5:HIS:CA	2:B:5:HIS:C	5	102.6