wwPDB NMR Structure Validation Summary Report

Nov 7, 2023 – 03:37 PM EST

PDB ID	:	2K03
BMRB ID	:	15635
Title	:	Structure of SDF1 in complex with the CXCR4 N-terminus containing a sulfotyrosine at postition 21
Authors	:	Volkman, B.F.; Veldkamp, C.T.; Peterson, F.C.
Deposited on	:	2008-01-24

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:

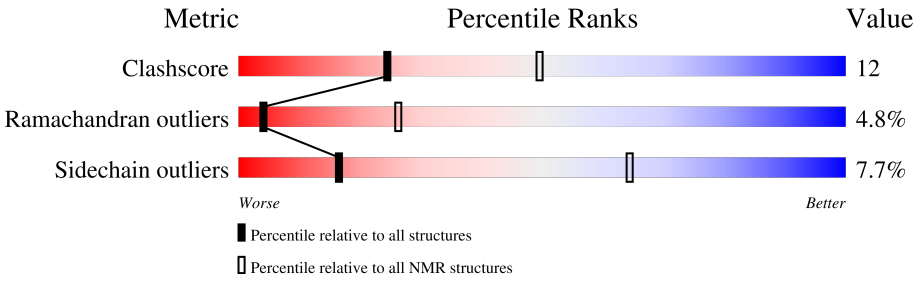
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
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)	:	2.36

1 Overall quality at a glance

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

The overall completeness of chemical shifts assignment is 88%.

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	70	56% 24% • 16% •
1	C	70	53% 29% 16% •
2	B	40	22% 10% 62% 5%
2	D	40	25% 12% 58% 5%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:8-A:43, A:47-A:67, B:114-B:120, B:122-B:127, C:208-C:243, C:247-C:267, D:312-D:320, D:322-D:327 (142)	1.02	11

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 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3417 atoms, of which 1686 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Stromal cell-derived factor 1.

Mol	Chain	Residues	Atoms						Trace
1	A	68	Total	C	H	N	O	S	0
			1136	353	579	106	92	6	
1	C	68	Total	C	H	N	O	S	0
			1136	353	579	106	92	6	

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P48061
A	0	MET	-	expression tag	UNP P48061
A	36	CYS	LEU	engineered mutation	UNP P48061
A	65	CYS	ALA	engineered mutation	UNP P48061
C	199	GLY	-	expression tag	UNP P48061
C	200	MET	-	expression tag	UNP P48061
C	236	CYS	LEU	engineered mutation	UNP P48061
C	265	CYS	ALA	engineered mutation	UNP P48061

- Molecule 2 is a protein called C-X-C chemokine receptor type 4.

Mol	Chain	Residues	Atoms						Trace
2	B	38	Total	C	H	N	O	S	0
			572	185	264	47	72	4	
2	D	38	Total	C	H	N	O	S	0
			573	185	264	47	73	4	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	99	GLY	-	expression tag	UNP P61073
B	100	SER	-	expression tag	UNP P61073
B	128	ALA	CYS	engineered mutation	UNP P61073
D	299	GLY	-	expression tag	UNP P61073
D	300	SER	-	expression tag	UNP P61073
D	328	ALA	CYS	engineered mutation	UNP P61073

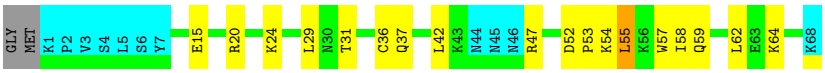
4 Residue-property plots

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.

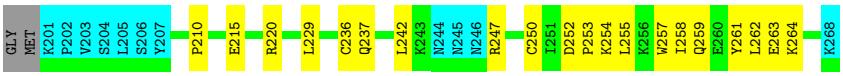
• Molecule 1: Stromal cell-derived factor 1

Chain A:



• Molecule 1: Stromal cell-derived factor 1

Chain C:

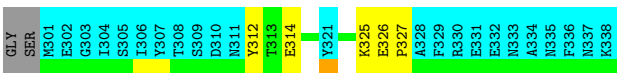


• Molecule 2: C-X-C chemokine receptor type 4

Chain B:

• Molecule 2: C-X-C chemokine receptor type 4

Chain D:

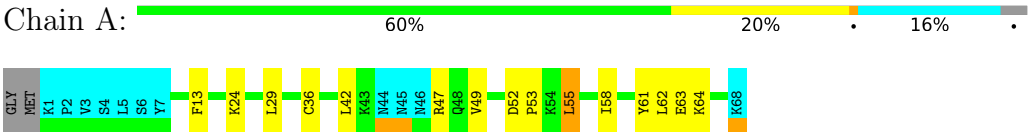


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 11. Colouring as in section 4.1 above.

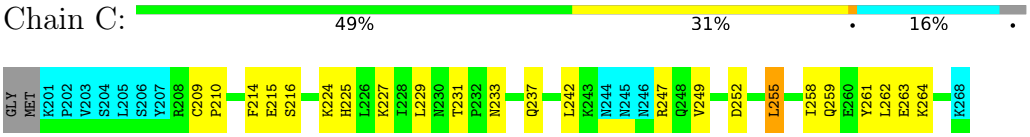
• Molecule 1: Stromal cell-derived factor 1

Chain A:



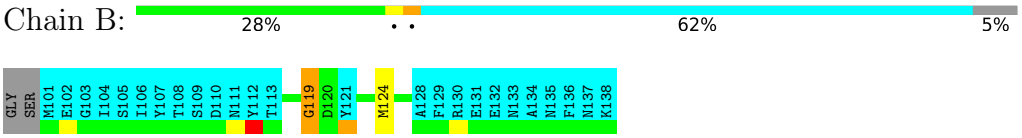
• Molecule 1: Stromal cell-derived factor 1

Chain C:



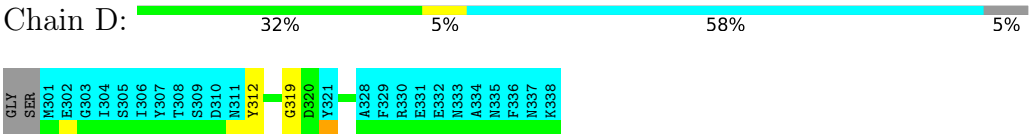
• Molecule 2: C-X-C chemokine receptor type 4

Chain B:



• Molecule 2: C-X-C chemokine receptor type 4

Chain D:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *AUTOMATED METHODS WERE USED FOR BACKBONE CHEMICAL SHIFT ASSIGNMENT AND ITERATIVE NOE REFINEMENT. FINAL STRUCTURES WERE OBTAINED BY MOLECULAR DYNAMICS IN EXPLICIT SOLVENT.*.

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

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

Software name	Classification	Version
Xplor-NIH	refinement	2.9.3

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	2478
Number of shifts mapped to atoms	2478
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	88%

6 Model quality

6.1 Standard geometry

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

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	0.85±0.03	0±0/479 (0.0± 0.0%)	0.75±0.03	0±0/646 (0.0± 0.0%)
1	C	0.86±0.03	0±0/479 (0.0± 0.0%)	0.74±0.03	0±0/646 (0.0± 0.0%)
2	B	0.89±0.10	0±0/95 (0.0± 0.0%)	0.88±0.07	0±0/124 (0.0± 0.0%)
2	D	0.93±0.07	0±0/115 (0.0± 0.0%)	0.92±0.10	0±0/152 (0.0± 0.1%)
All	All	0.86	0/23360 (0.0%)	0.77	1/31360 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.1±0.4
1	C	0.0±0.0	0.1±0.2
All	All	0	4

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	D	312	TYR	CB-CG-CD1	-5.57	117.66	121.00	2	1

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	47	ARG	Sidechain	2
1	A	20	ARG	Sidechain	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	C	220	ARG	Sidechain	1

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	469	489	487	12±3
1	C	469	489	487	14±3
2	B	95	80	80	2±1
2	D	114	96	96	4±2
All	All	22940	23080	23000	543

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 12.

5 of 230 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:C:254:LYS:O	2:D:327:PRO:HA	0.75	1.81	14	9
1:C:229:LEU:HG	2:D:312:TYR:CE2	0.72	2.19	16	7
1:C:220:ARG:HG2	1:C:257:TRP:CE3	0.65	2.25	6	3
1:C:215:GLU:HG2	1:C:250:CYS:O	0.65	1.91	3	4
1:A:10:PRO:HA	2:B:114:GLU:O	0.64	1.92	8	1

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	57/70 (81%)	51±1 (90±2%)	4±1 (7±2%)	1±1 (2±2%)	9	45
1	C	57/70 (81%)	52±2 (90±3%)	4±2 (7±3%)	1±1 (2±1%)	9	45

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	B	13/40 (32%)	7±2 (55±15%)	4±2 (31±12%)	2±1 (14±7%)	1 5
2	D	15/40 (38%)	8±2 (55±10%)	5±1 (31±8%)	2±1 (15±7%)	0 4
All	All	2840/4400 (65%)	2367 (83%)	337 (12%)	136 (5%)	4 26

5 of 36 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	35	ALA	8
2	B	119	GLY	8
1	C	235	ALA	8
2	B	126	GLU	8
2	D	326	GLU	7

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	53/65 (82%)	48±1 (91±3%)	5±1 (9±3%)	13 60
1	C	53/65 (82%)	49±2 (92±3%)	4±2 (8±3%)	16 64
2	B	11/33 (33%)	10±1 (95±5%)	0±1 (5±5%)	31 80
2	D	13/33 (39%)	12±1 (95±5%)	1±1 (5±5%)	28 77
All	All	2600/3920 (66%)	2401 (92%)	199 (8%)	16 64

5 of 49 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	47	ARG	16
1	C	247	ARG	14
1	A	55	LEU	11
1	C	260	GLU	10
1	A	14	PHE	9

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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
2	TYS	B	121	2	15,16,17	2.69±0.21	2±0 (10±3%)
2	TYS	D	321	2	15,16,17	2.69±0.19	2±1 (11±3%)

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
2	TYS	B	121	2	18,22,24	1.04±0.03	2±1 (9±3%)
2	TYS	D	321	2	18,22,24	1.03±0.04	1±1 (7±3%)

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
2	TYS	B	121	2	-	0±0,10,11,13	0±0,1,1,1
2	TYS	D	321	2	-	0±0,10,11,13	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
2	D	321	TYS	OH-S	11.03	1.41	1.58	4	20
2	B	121	TYS	OH-S	10.93	1.41	1.58	8	20
2	D	321	TYS	OH-CZ	2.81	1.38	1.42	4	12
2	B	121	TYS	OH-CZ	2.74	1.38	1.42	20	11
2	D	321	TYS	CB-CG	2.00	1.46	1.51	1	1

5 of 6 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	D	321	TYS	OH-S-O2	3.08	116.68	107.71	14	19
2	B	121	TYS	OH-S-O2	2.87	116.07	107.71	4	20
2	B	121	TYS	O3-S-O2	2.21	100.82	108.49	4	11
2	D	321	TYS	O3-S-O2	2.12	101.12	108.49	10	8
2	D	321	TYS	OH-S-O1	2.06	113.72	107.71	6	1

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 88% for the well-defined parts and 84% 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	2478
Number of shifts mapped to atoms	2478
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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	200	-0.48 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	192	-0.42 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	186	-0.27 ± 0.07	None needed (< 0.5 ppm)
^{15}N	192	-0.49 ± 0.43	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 88%, i.e. 1759 atoms were assigned a chemical shift out of a possible 2002. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	676/698 (97%)	276/280 (99%)	266/284 (94%)	134/134 (100%)
Sidechain	1001/1181 (85%)	684/760 (90%)	303/363 (83%)	14/58 (24%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	82/123 (67%)	48/60 (80%)	32/53 (60%)	2/10 (20%)
Overall	1759/2002 (88%)	1008/1100 (92%)	601/700 (86%)	150/202 (74%)

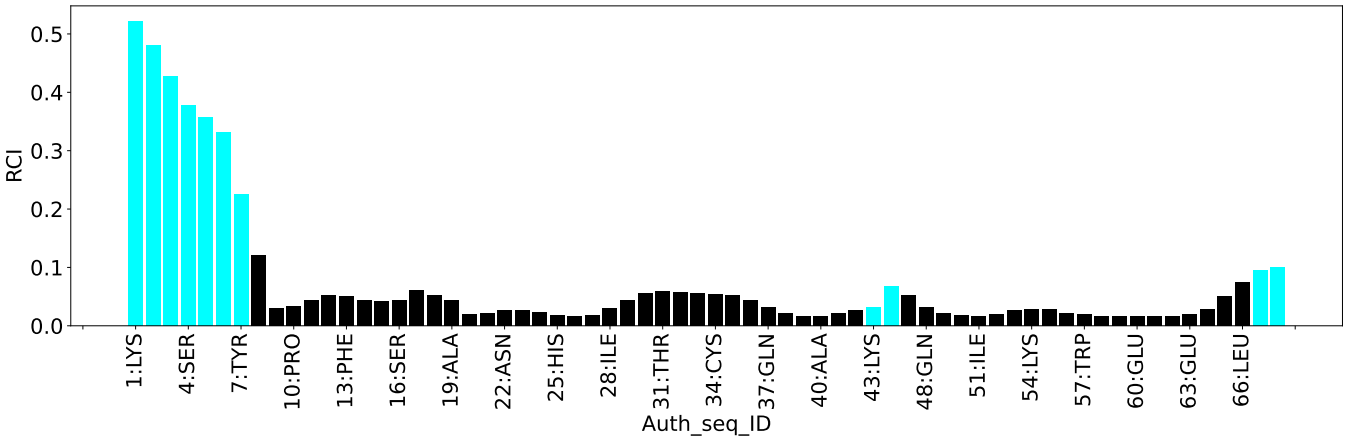
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

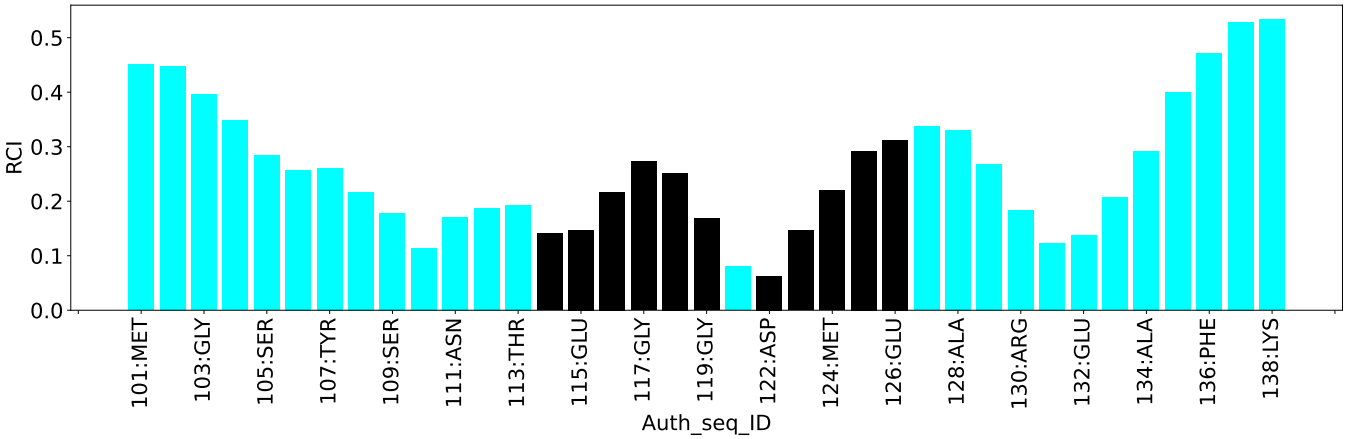
7.1.5 Random Coil Index (RCI) plots (i)

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.

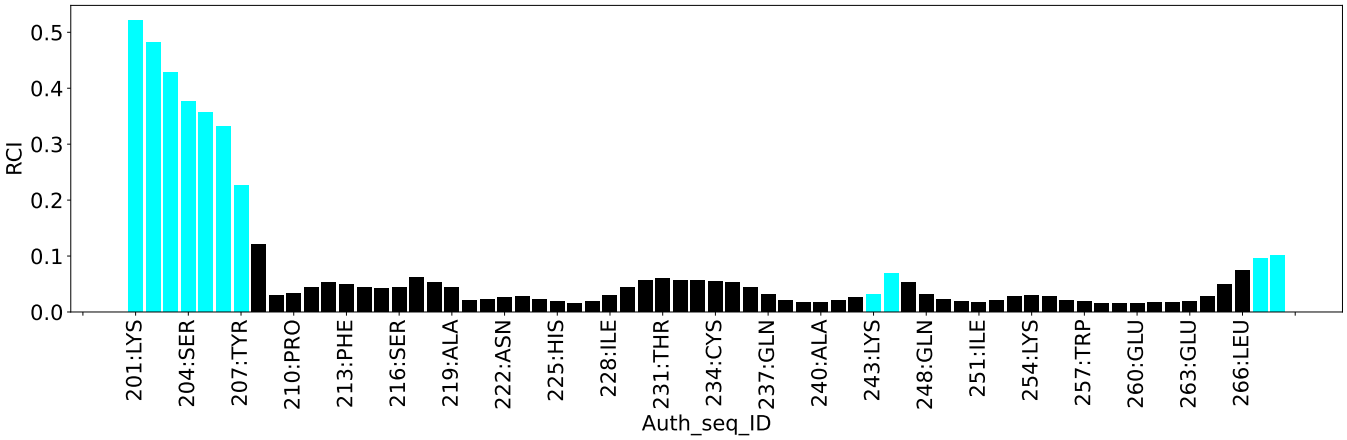
Random coil index (RCI) for chain A:



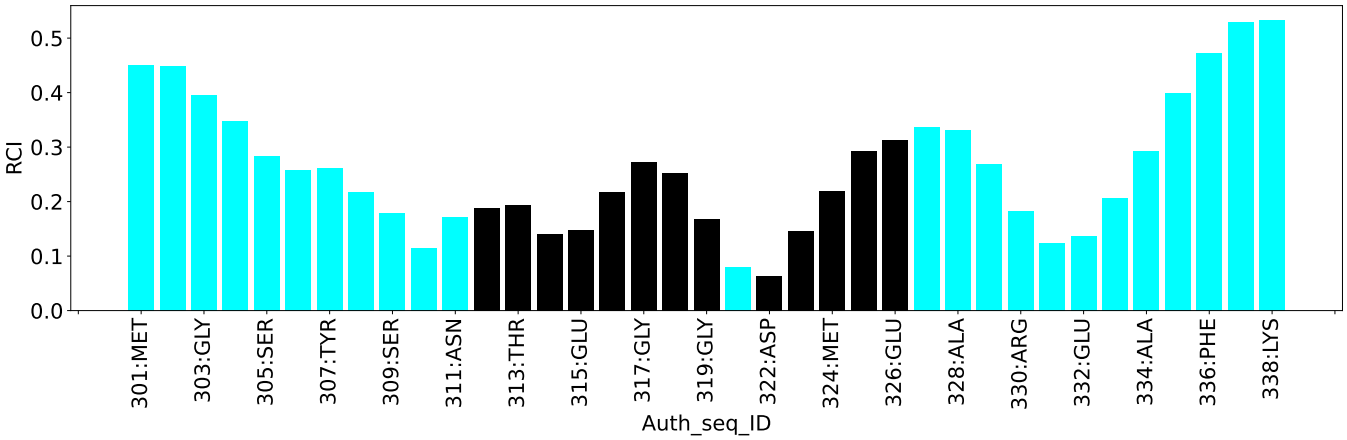
Random coil index (RCI) for chain B:



Random coil index (RCI) for chain C:



Random coil index (RCI) for chain D:



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	1694
Intra-residue (|i-j|=0)	528
Sequential (|i-j|=1)	408
Medium range (|i-j|>1 and |i-j|<5)	197
Long range (|i-j|≥5)	376
Inter-chain	185
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	128
Number of unmapped restraints	0
Number of restraints per residue	8.3
Number of long range restraints per residue[1]	1.7

[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)	42.4	0.2
0.2-0.5 (Medium)	25.8	0.5
>0.5 (Large)	0.2	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)	20.2	5.04
10.0-20.0 (Medium)	None	None
>20.0 (Large)	None	None

9 Distance violation analysis (i)

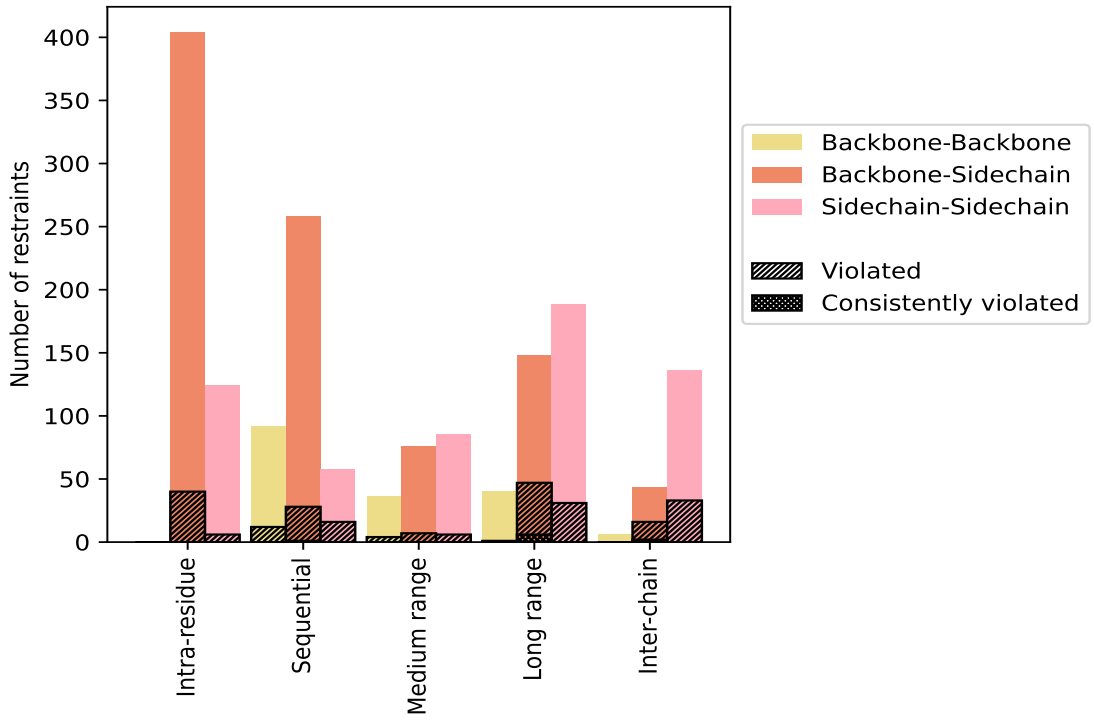
9.1 Summary of distance violations (i)

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)	528	31.2	46	8.7	2.7	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	404	23.8	40	9.9	2.4	0	0.0	0.0
Sidechain-Sidechain	124	7.3	6	4.8	0.4	0	0.0	0.0
Sequential (|i-j|=1)	408	24.1	56	13.7	3.3	1	0.2	0.1
Backbone-Backbone	92	5.4	12	13.0	0.7	0	0.0	0.0
Backbone-Sidechain	258	15.2	28	10.9	1.7	1	0.4	0.1
Sidechain-Sidechain	58	3.4	16	27.6	0.9	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	197	11.6	17	8.6	1.0	0	0.0	0.0
Backbone-Backbone	36	2.1	4	11.1	0.2	0	0.0	0.0
Backbone-Sidechain	76	4.5	7	9.2	0.4	0	0.0	0.0
Sidechain-Sidechain	85	5.0	6	7.1	0.4	0	0.0	0.0
Long range (|i-j|≥5)	376	22.2	79	21.0	4.7	6	1.6	0.4
Backbone-Backbone	40	2.4	1	2.5	0.1	0	0.0	0.0
Backbone-Sidechain	148	8.7	47	31.8	2.8	6	4.1	0.4
Sidechain-Sidechain	188	11.1	31	16.5	1.8	0	0.0	0.0
Inter-chain	185	10.9	49	26.5	2.9	2	1.1	0.1
Backbone-Backbone	6	0.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	43	2.5	16	37.2	0.9	2	4.7	0.1
Sidechain-Sidechain	136	8.0	33	24.3	1.9	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	1694	100.0	247	14.6	14.6	9	0.5	0.5
Backbone-Backbone	174	10.3	17	9.8	1.0	0	0.0	0.0
Backbone-Sidechain	929	54.8	138	14.9	8.1	9	1.0	0.5
Sidechain-Sidechain	591	34.9	92	15.6	5.4	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	9	10	1	35	13	68	0.21	0.42	0.1	0.19
2	9	12	2	35	17	75	0.19	0.42	0.09	0.15
3	14	11	6	31	20	82	0.18	0.59	0.08	0.16
4	13	12	2	26	16	69	0.19	0.7	0.1	0.16
5	16	18	1	29	14	78	0.2	0.43	0.1	0.16
6	12	11	6	32	15	76	0.21	0.44	0.1	0.16
7	10	16	4	28	16	74	0.19	0.42	0.08	0.16
8	10	14	4	31	15	74	0.21	0.44	0.09	0.16
9	12	15	1	34	11	73	0.2	0.45	0.09	0.17
10	13	14	1	28	15	71	0.2	0.41	0.08	0.17
11	7	15	3	25	12	62	0.2	0.44	0.09	0.18

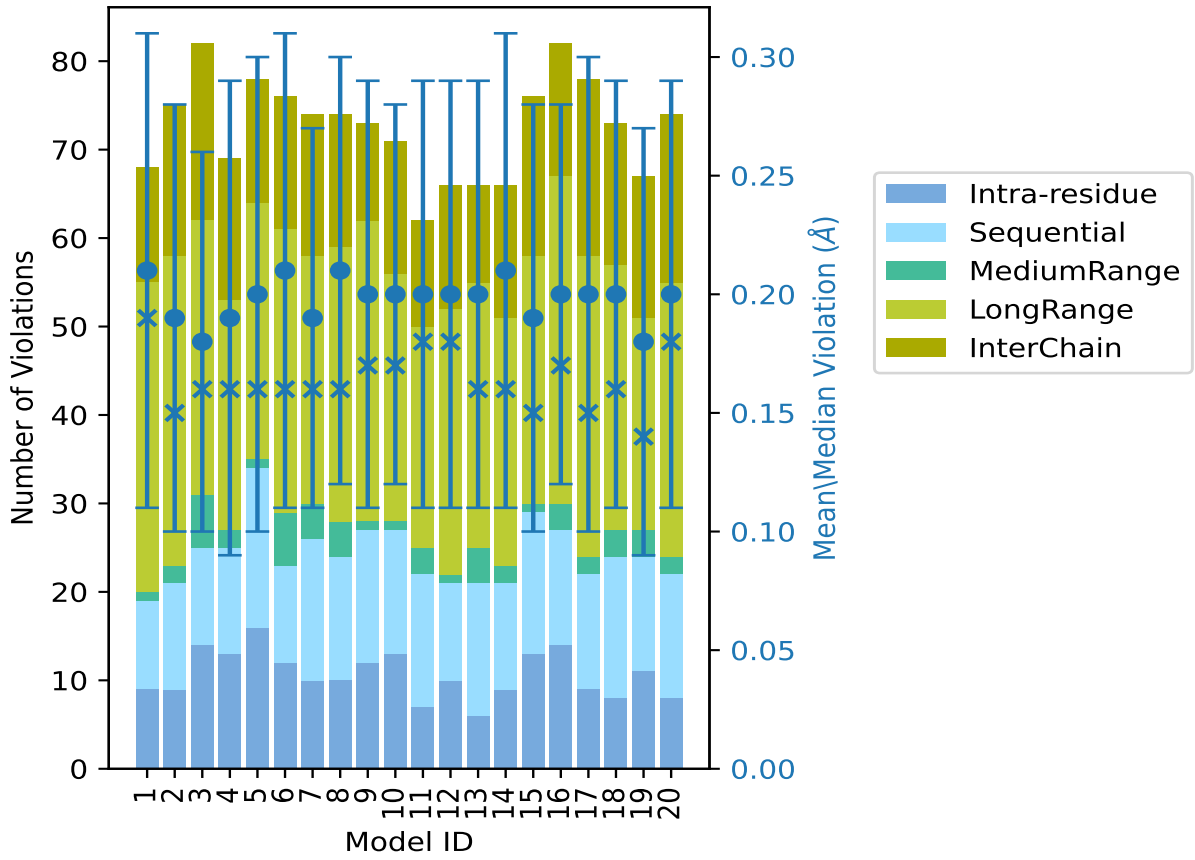
Continued on next page...

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	10	11	1	30	14	66	0.2	0.55	0.09	0.18
13	6	15	4	30	11	66	0.2	0.41	0.09	0.16
14	9	12	2	28	15	66	0.21	0.5	0.1	0.16
15	13	16	1	28	18	76	0.19	0.45	0.09	0.15
16	14	13	3	37	15	82	0.2	0.5	0.08	0.17
17	9	13	2	34	20	78	0.2	0.44	0.1	0.15
18	8	16	3	30	16	73	0.2	0.54	0.09	0.16
19	11	13	3	24	16	67	0.18	0.64	0.09	0.14
20	8	14	2	31	19	74	0.2	0.44	0.09	0.18

[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

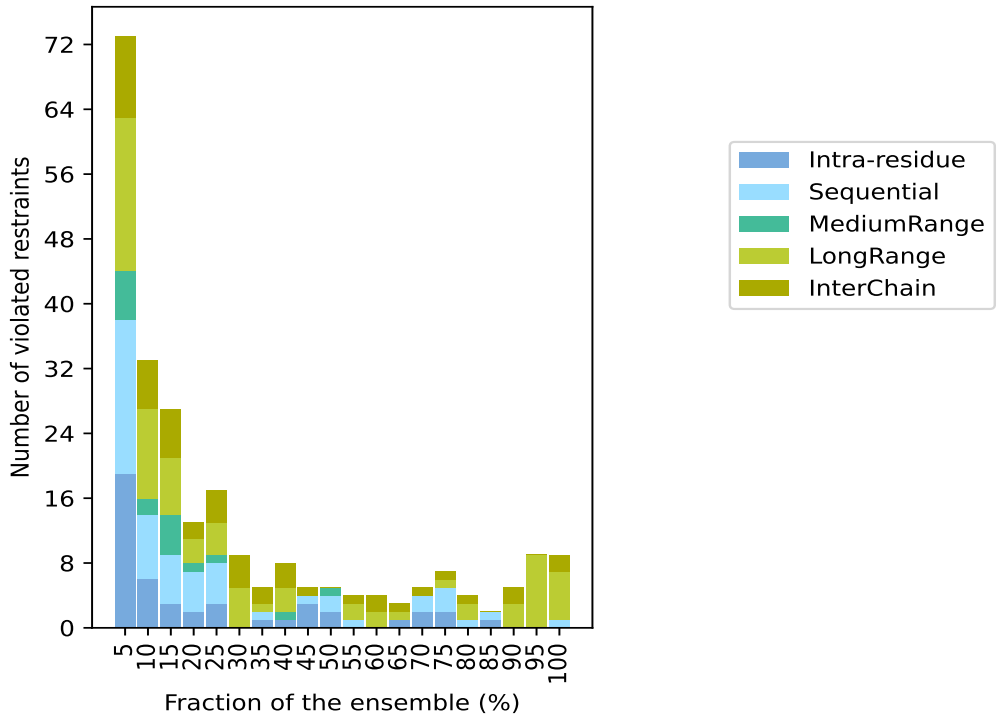
9.3 Distance violation statistics for the ensemble (i)

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, 1447(IR:482, SQ:352, MR:180, LR:297, IC:136) 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]	%
19	19	6	19	10	73	1	5.0
6	8	2	11	6	33	2	10.0
3	6	5	7	6	27	3	15.0
2	5	1	3	2	13	4	20.0
3	5	1	4	4	17	5	25.0
0	0	0	5	4	9	6	30.0
1	1	0	1	2	5	7	35.0
1	0	1	3	3	8	8	40.0
3	1	0	0	1	5	9	45.0
2	2	1	0	0	5	10	50.0
0	1	0	2	1	4	11	55.0
0	0	0	2	2	4	12	60.0
1	0	0	1	1	3	13	65.0
2	2	0	0	1	5	14	70.0
2	3	0	1	1	7	15	75.0
0	1	0	2	1	4	16	80.0
1	1	0	0	0	2	17	85.0
0	0	0	3	2	5	18	90.0
0	0	0	9	0	9	19	95.0
0	1	0	6	2	9	20	100.0

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

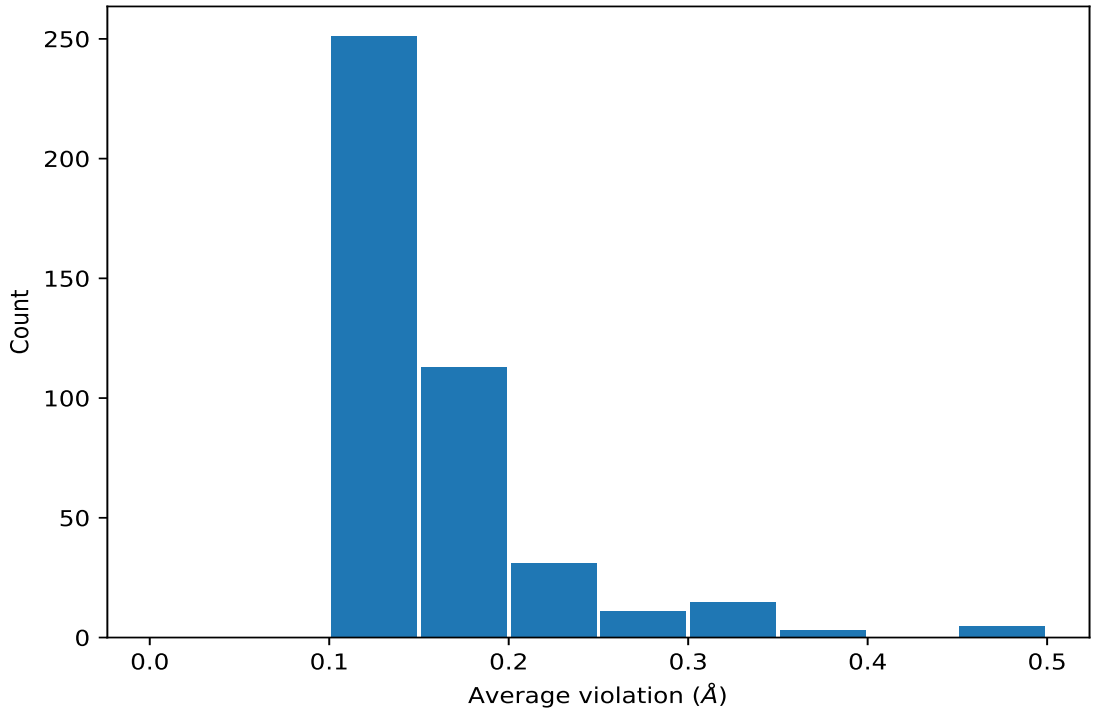
9.3.1 Bar graph : Distance violation statistics for the ensemble



9.4 Most violated distance restraints in the ensemble

9.4.1 Histogram : Distribution of mean distance violations

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



9.4.2 Table: Most violated distance restraints

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

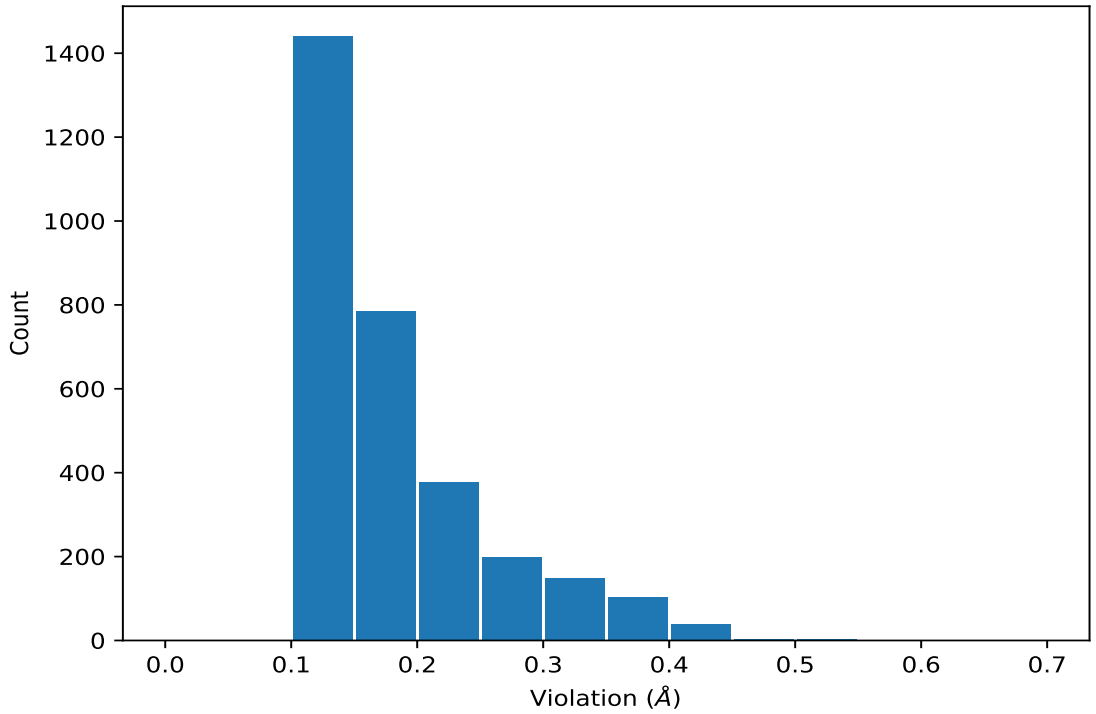
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(3,19)	1:38:A:ILE:O	1:51:A:ILE:N	20	0.36	0.03	0.36
(3,34)	1:238:C:ILE:O	1:251:C:ILE:N	20	0.35	0.05	0.36
(3,28)	1:224:C:LYS:N	1:241:C:ARG:O	20	0.31	0.07	0.32
(3,18)	1:38:A:ILE:O	1:51:A:ILE:H	20	0.3	0.05	0.31
(3,33)	1:238:C:ILE:O	1:251:C:ILE:H	20	0.3	0.06	0.32
(3,14)	1:28:A:ILE:N	1:226:C:LEU:O	20	0.29	0.04	0.3
(3,4)	1:24:A:LYS:N	1:41:A:ARG:O	20	0.28	0.06	0.32
(3,8)	1:26:A:LEU:O	1:228:C:ILE:N	20	0.28	0.06	0.29
(1,631)	1:62:A:LEU:HG	1:63:A:GLU:H	20	0.12	0.01	0.12
(2,16)	1:211:C:CYS:CB	1:250:C:CYS:CB	19	0.35	0.09	0.36

[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)
(1,86)	1:18:A:VAL:HG11	2:121:B:TYS:HD2	4	0.7
(1,86)	1:18:A:VAL:HG23	2:121:B:TYS:HE2	19	0.64
(1,86)	1:18:A:VAL:HG23	2:121:B:TYS:HE1	3	0.59
(1,971)	1:218:C:VAL:HG21	2:321:D:TYS:HE1	12	0.55
(1,86)	1:18:A:VAL:HG23	2:121:B:TYS:HE1	18	0.54
(2,10)	1:65:A:CYS:CB	1:236:C:CYS:CB	16	0.5
(2,4)	1:11:A:CYS:CB	1:50:A:CYS:CB	14	0.5
(2,16)	1:211:C:CYS:CB	1:250:C:CYS:CB	14	0.46
(2,16)	1:211:C:CYS:CB	1:250:C:CYS:CB	9	0.45
(2,16)	1:211:C:CYS:CB	1:250:C:CYS:CB	15	0.45

10 Dihedral-angle violation analysis

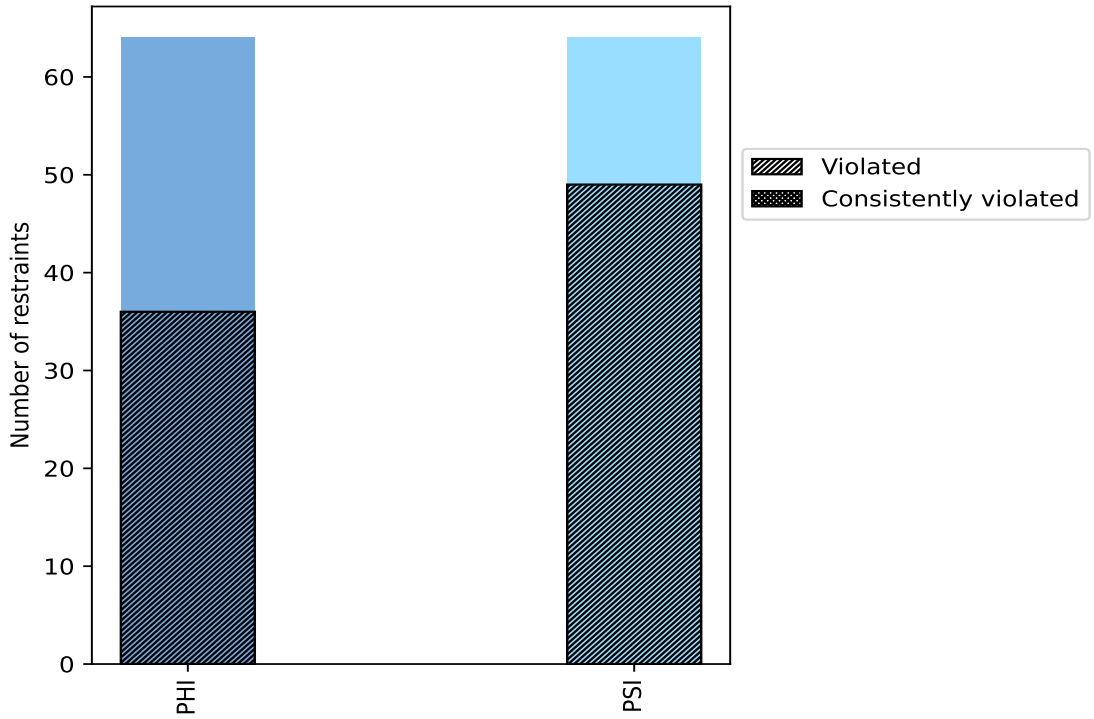
10.1 Summary of dihedral-angle violations

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	64	50.0	36	56.2	28.1	0	0.0	0.0
PSI	64	50.0	49	76.6	38.3	0	0.0	0.0
Total	128	100.0	85	66.4	66.4	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



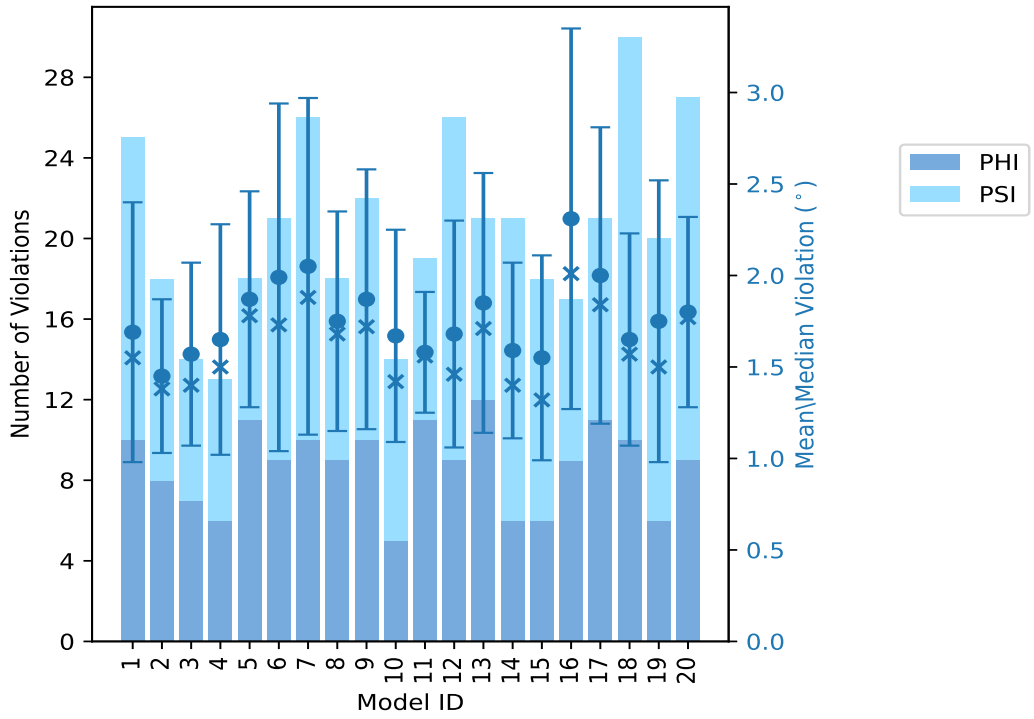
Violated and consistently violated restraints are shown using different hatch patterns in their respective categories

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	10	15	25	1.69	4.56	0.71	1.55
2	8	10	18	1.45	2.9	0.42	1.38
3	7	7	14	1.57	2.54	0.5	1.4
4	6	7	13	1.65	3.3	0.63	1.5
5	11	7	18	1.87	3.41	0.59	1.78
6	9	12	21	1.99	5.04	0.95	1.73
7	10	16	26	2.05	4.8	0.92	1.88
8	9	9	18	1.75	3.84	0.6	1.68
9	10	12	22	1.87	3.84	0.71	1.72
10	5	9	14	1.67	2.82	0.58	1.42
11	11	8	19	1.58	2.32	0.33	1.56
12	9	17	26	1.68	3.06	0.62	1.46
13	12	9	21	1.85	3.23	0.71	1.71
14	6	15	21	1.59	2.67	0.48	1.4
15	6	12	18	1.55	2.78	0.56	1.32
16	9	8	17	2.31	4.89	1.04	2.01
17	11	10	21	2.0	4.33	0.81	1.84
18	10	20	30	1.65	3.16	0.58	1.57
19	6	14	20	1.75	4.54	0.77	1.5
20	9	18	27	1.8	3.18	0.52	1.77

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	
PHI	PSI	Total	Count[1]	%
11	9	20	1	5.0
6	9	15	2	10.0
3	7	10	3	15.0
2	4	6	4	20.0
2	5	7	5	25.0
1	0	1	6	30.0
1	3	4	7	35.0
1	2	3	8	40.0
4	1	5	9	45.0
0	3	3	10	50.0
1	2	3	11	55.0

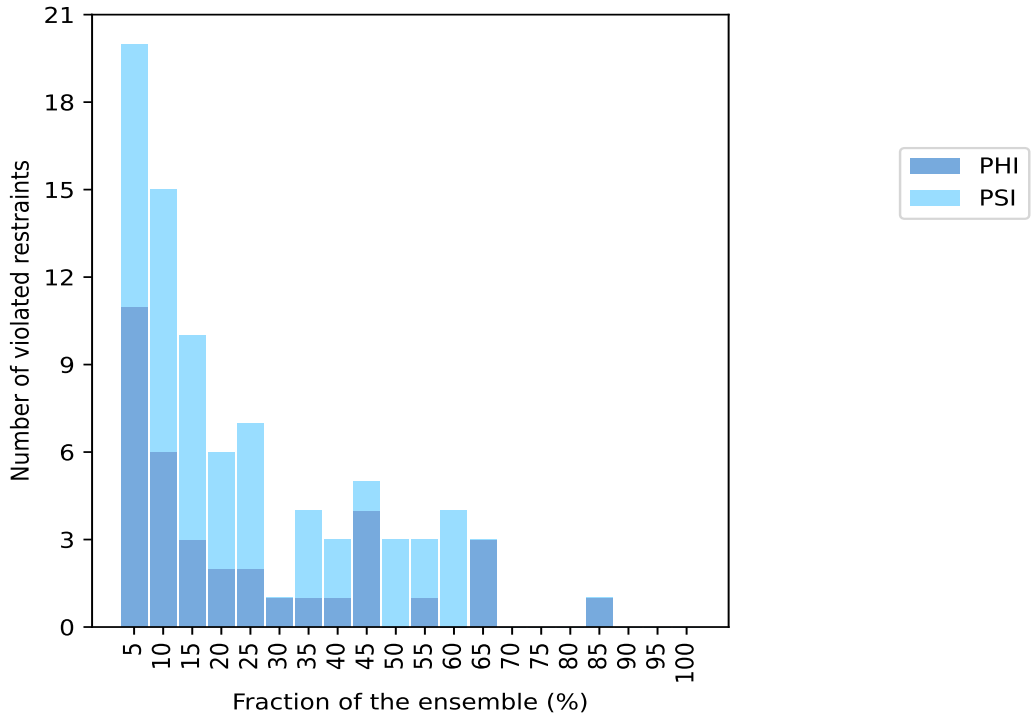
Continued on next page...

Continued from previous page...

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

[1] Number of models with violations

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

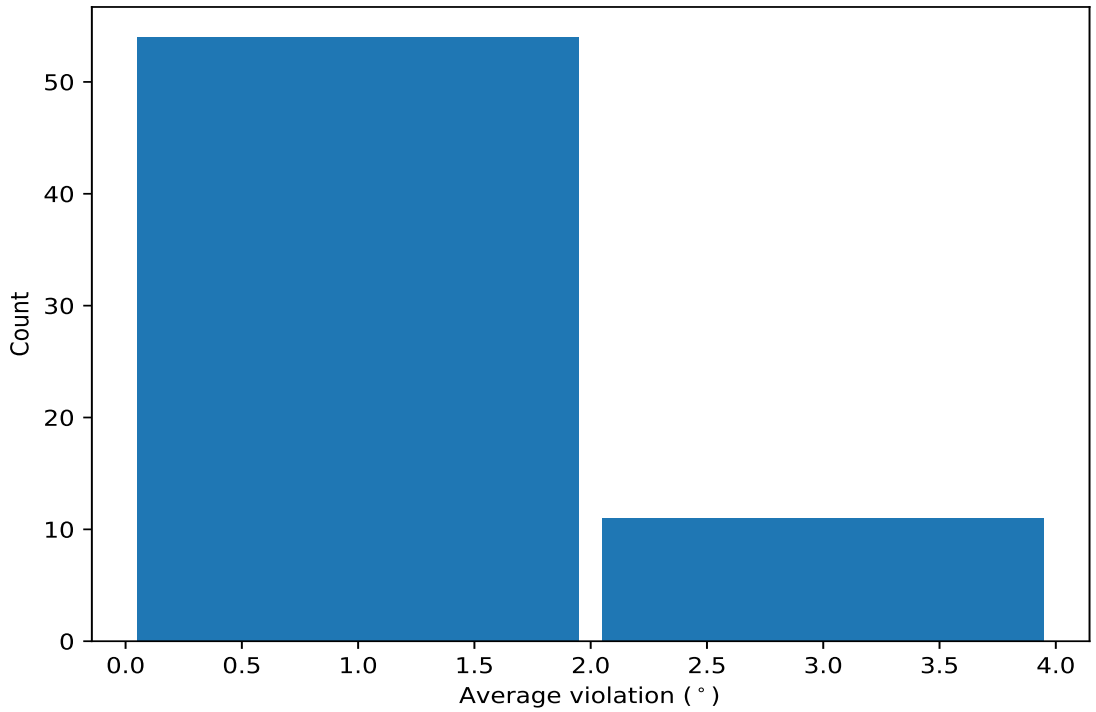


10.4 Most violated dihedral-angle restraints in the ensemble

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.

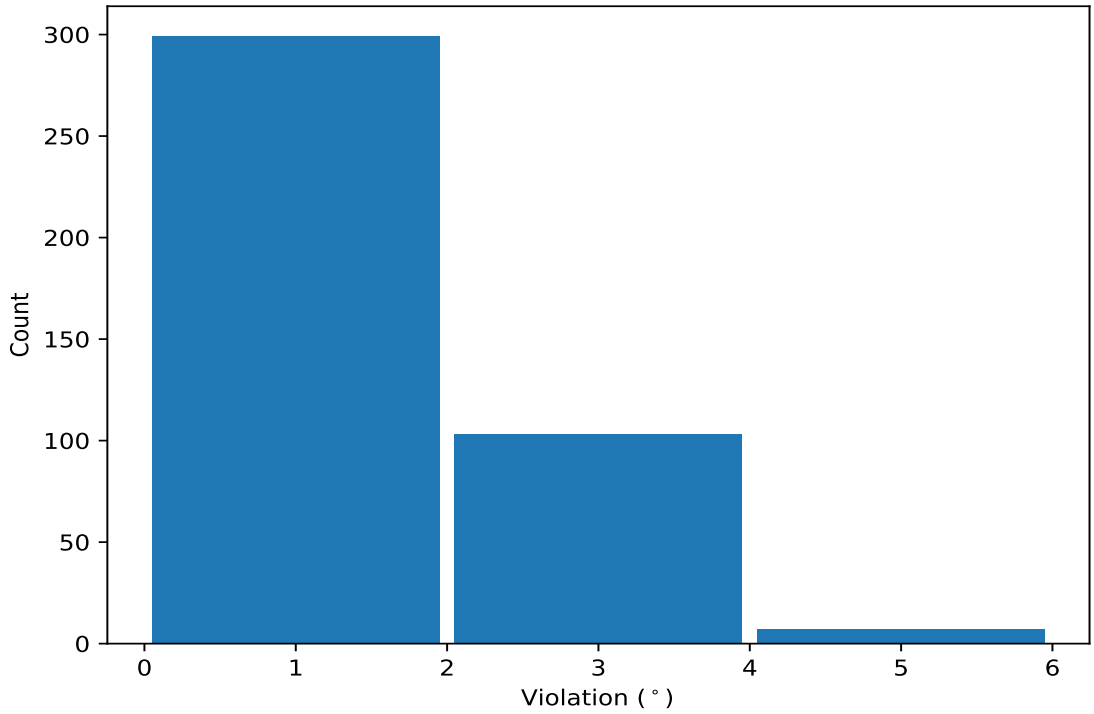
Key	Atom-1	Atom-2	Atom-3	Atom-4	Models[1]	Mean	SD[2]	Median
(1,86)	1:225:C:HIS:C	1:226:C:LEU:N	1:226:C:LEU:CA	1:226:C:LEU:C	17	1.86	0.54	1.77
(1,111)	1:250:C:CYS:C	1:251:C:ILE:N	1:251:C:ILE:CA	1:251:C:ILE:C	13	2.07	0.88	1.83
(1,47)	1:50:A:CYS:C	1:51:A:ILE:N	1:51:A:ILE:CA	1:51:A:ILE:C	13	2.04	0.76	1.77
(1,22)	1:25:A:HIS:C	1:26:A:LEU:N	1:26:A:LEU:CA	1:26:A:LEU:C	13	1.91	0.5	1.89
(1,2)	1:8:A:ARG:N	1:8:A:ARG:CA	1:8:A:ARG:C	1:9:A:CYS:N	12	2.6	0.97	2.74
(1,50)	1:57:A:TRP:N	1:57:A:TRP:CA	1:57:A:TRP:C	1:58:A:ILE:N	12	1.99	0.48	1.94
(1,97)	1:234:C:CYS:N	1:234:C:CYS:CA	1:234:C:CYS:C	1:235:C:ALA:N	12	1.84	0.7	1.58
(1,114)	1:257:C:TRP:N	1:257:C:TRP:CA	1:257:C:TRP:C	1:258:C:ILE:N	12	1.72	0.42	1.73
(1,66)	1:208:C:ARG:N	1:208:C:ARG:CA	1:208:C:ARG:C	1:209:C:CYS:N	11	2.81	1.19	2.13
(1,40)	1:40:A:ALA:C	1:41:A:ARG:N	1:41:A:ARG:CA	1:41:A:ARG:C	11	1.86	0.59	1.77

[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 (°)
(1,66)	1:208:C:ARG:N	1:208:C:ARG:CA	1:208:C:ARG:C	1:209:C:CYS:N	6	5.04
(1,67)	1:208:C:ARG:C	1:209:C:CYS:N	1:209:C:CYS:CA	1:209:C:CYS:C	16	4.89
(1,66)	1:208:C:ARG:N	1:208:C:ARG:CA	1:208:C:ARG:C	1:209:C:CYS:N	7	4.8
(1,3)	1:8:A:ARG:C	1:9:A:CYS:N	1:9:A:CYS:CA	1:9:A:CYS:C	1	4.56
(1,2)	1:8:A:ARG:N	1:8:A:ARG:CA	1:8:A:ARG:C	1:9:A:CYS:N	19	4.54
(1,111)	1:250:C:CYS:C	1:251:C:ILE:N	1:251:C:ILE:CA	1:251:C:ILE:C	17	4.33
(1,66)	1:208:C:ARG:N	1:208:C:ARG:CA	1:208:C:ARG:C	1:209:C:CYS:N	16	4.07
(1,67)	1:208:C:ARG:C	1:209:C:CYS:N	1:209:C:CYS:CA	1:209:C:CYS:C	6	3.97
(1,2)	1:8:A:ARG:N	1:8:A:ARG:CA	1:8:A:ARG:C	1:9:A:CYS:N	7	3.92
(1,104)	1:240:C:ALA:C	1:241:C:ARG:N	1:241:C:ARG:CA	1:241:C:ARG:C	9	3.84