wwPDB NMR Structure Validation Summary Report (i)

Jun 5, 2023 – 02:55 PM JST

PDB ID	:	7EN4
BMRB ID	:	36420
Title	:	Multi-state structure determination and dynamics analysis elucidate a new u biquitin-recognition mechanism of yeast ubiquitin C-terminal hydrolase.
Authors	:	Okada, M.; Tateishi, Y.; Nojiri, E.; Mikawa, T.; Rajesh, S.; Ogasawara, H.; Ueda, T.; Yagi, H.; Kohno, T.; Kigawa, T.; Shimada, I.; Guentert, P.; Yutaka, I.; Ikeya, T.
Deposited on	:	2021-04-15

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:

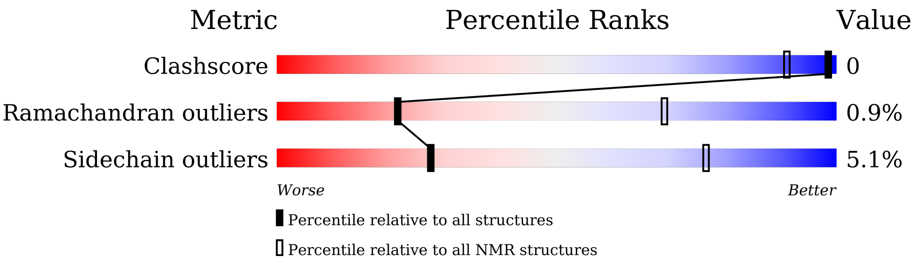
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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 87%.

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	236	84% 16%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 14 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 the center of the largest cluster*.

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:14-A:64, A:77-A:151, A:164-A:236 (199)	1.07	14

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 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3654 atoms, of which 1793 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ubiquitin carboxyl-terminal hydrolase YUH1.

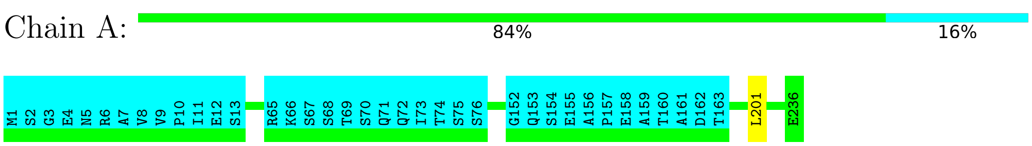
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	236	3654	1176	1793	304	377	4	0

4 Residue-property plots (i)

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.

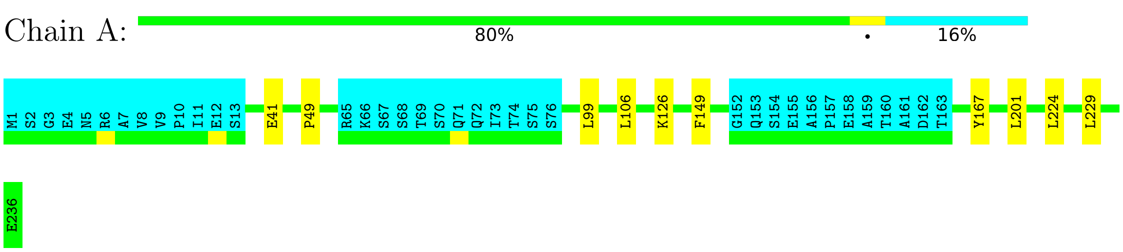
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase YUH1



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

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: Ubiquitin carboxyl-terminal hydrolase YUH1



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *energy minimization*.

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *structures randomly selected from each of five clusters*.

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

Software name	Classification	Version
CYANA	structure calculation	3.99.0
OPAL	refinement	1.4

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	2734
Number of shifts mapped to atoms	2734
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

6 Model quality (i)

6.1 Standard geometry (i)

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts (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	1595	1539	1539	0±1
All	All	31900	30780	30780	7

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.

5 of 6 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:185:LEU:HD13	1:A:185:LEU:H	0.48	1.67	2	2
1:A:122:THR:HG22	1:A:123:SER:H	0.48	1.69	1	1
1:A:97:HIS:CD2	1:A:169:THR:HG23	0.46	2.45	7	1
1:A:170:TYR:CD2	1:A:170:TYR:N	0.46	2.83	7	1
1:A:14:ASN:N	1:A:15:PRO:CD	0.40	2.83	12	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	198/236 (84%)	183±4 (92±2%)	13±3 (7±2%)	2±1 (1±1%)	21	69
All	All	3960/4720 (84%)	3661 (92%)	265 (7%)	34 (1%)	21	69

5 of 14 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	201	LEU	8
1	A	49	PRO	6
1	A	15	PRO	3
1	A	128	ARG	2
1	A	129	PHE	2

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	179/210 (85%)	170±2 (95±1%)	9±2 (5±1%)	27	77
All	All	3580/4200 (85%)	3396 (95%)	184 (5%)	27	77

5 of 75 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	112	LEU	9
1	A	93	TYR	8
1	A	208	ARG	8
1	A	126	LYS	7
1	A	99	LEU	7

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 (i)

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 87% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

7.1.1 Bookkeeping (i)

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	2734
Number of shifts mapped to atoms	2734
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 (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	231	0.08 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	221	0.66 ± 0.06	Should be checked
$^{13}C'$	211	0.17 ± 0.12	None needed (< 0.5 ppm)
^{15}N	222	-1.59 ± 0.16	Should be applied

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 87%, i.e. 2350 atoms were assigned a chemical shift out of a possible 2715. 0 out of 40 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	953/987 (97%)	389/399 (97%)	375/398 (94%)	189/190 (99%)
Sidechain	1280/1479 (87%)	866/960 (90%)	393/471 (83%)	21/48 (44%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	117/249 (47%)	83/122 (68%)	31/121 (26%)	3/6 (50%)
Overall	2350/2715 (87%)	1338/1481 (90%)	799/990 (81%)	213/244 (87%)

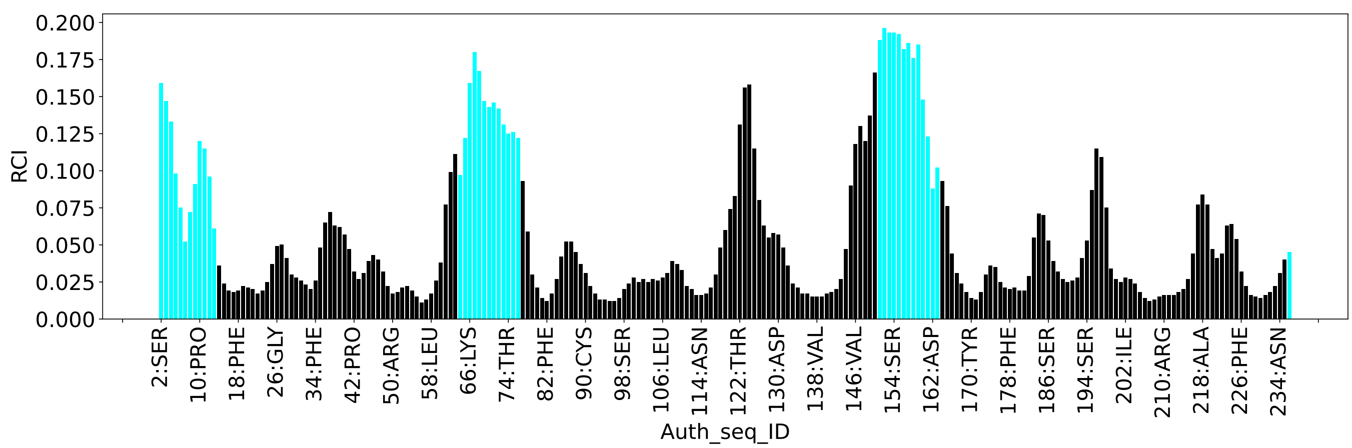
7.1.4 Statistically unusual chemical shifts (i)

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:



8 NMR restraints analysis (i)

8.1 Conformationally restricting restraints (i)

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	6446
Intra-residue (|i-j|=0)	322
Sequential (|i-j|=1)	1090
Medium range (|i-j|>1 and |i-j|<5)	1257
Long range (|i-j|≥5)	3752
Inter-chain	0
Hydrogen bond restraints	25
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	0
Number of restraints per residue	27.3
Number of long range restraints per residue[1]	16.0

[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 (i)

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 (i)

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)	92.5	0.2
0.2-0.5 (Medium)	243.7	0.5
>0.5 (Large)	1230.0	35.79

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

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

9 Distance violation analysis (i)

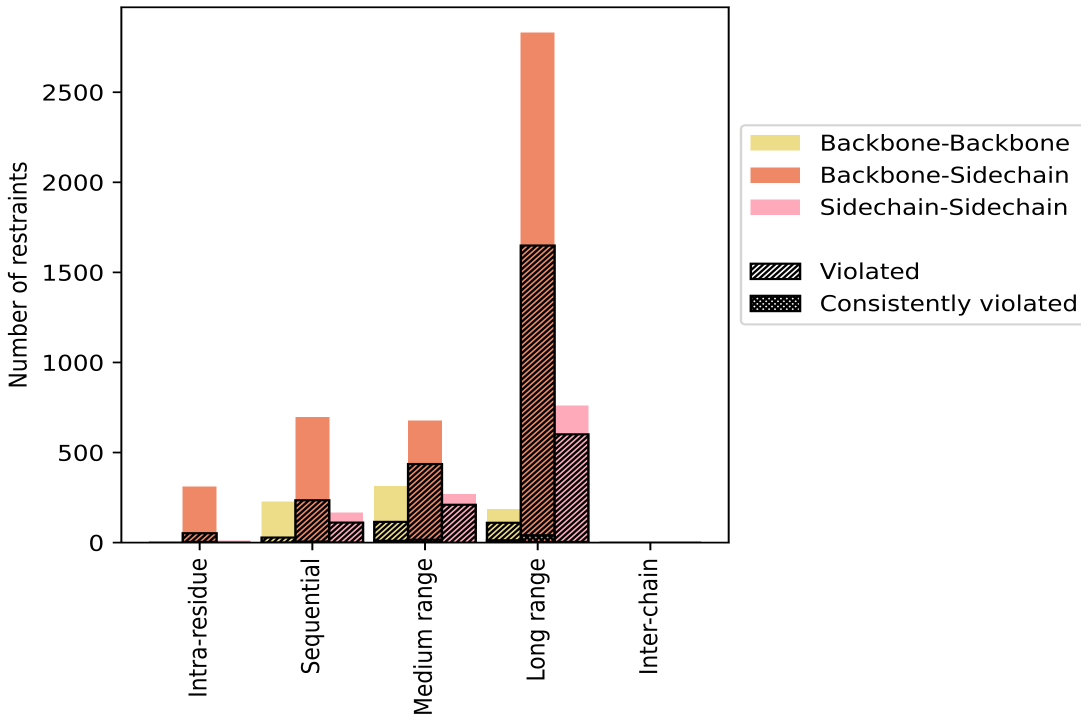
9.1 Summary of distance violations (i)

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

Restraints type	Count	%[1]	Violated[3]			Consistently Violated[4]		
			Count	%[2]	%[1]	Count	%[2]	%[1]
Intra-residue (|i-j|=0)	322	5.0	52	16.1	0.8	4	1.2	0.1
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	311	4.8	52	16.7	0.8	4	1.3	0.1
Sidechain-Sidechain	11	0.2	0	0.0	0.0	0	0.0	0.0
Sequential (|i-j|=1)	1090	16.9	374	34.3	5.8	8	0.7	0.1
Backbone-Backbone	228	3.5	28	12.3	0.4	1	0.4	0.0
Backbone-Sidechain	696	10.8	235	33.8	3.6	6	0.9	0.1
Sidechain-Sidechain	166	2.6	111	66.9	1.7	1	0.6	0.0
Medium range (|i-j|>1 & |i-j|<5)	1257	19.5	759	60.4	11.8	26	2.1	0.4
Backbone-Backbone	310	4.8	113	36.5	1.8	10	3.2	0.2
Backbone-Sidechain	678	10.5	436	64.3	6.8	16	2.4	0.2
Sidechain-Sidechain	269	4.2	210	78.1	3.3	0	0.0	0.0
Long range (|i-j|≥5)	3752	58.2	2337	62.3	36.3	58	1.5	0.9
Backbone-Backbone	162	2.5	87	53.7	1.3	13	8.0	0.2
Backbone-Sidechain	2830	43.9	1649	58.3	25.6	41	1.4	0.6
Sidechain-Sidechain	760	11.8	601	79.1	9.3	4	0.5	0.1
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	25	0.4	25	100.0	0.4	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	6446	100.0	3547	55.0	55.0	96	1.5	1.5
Backbone-Backbone	725	11.2	253	34.9	3.9	24	3.3	0.4
Backbone-Sidechain	4515	70.0	2372	52.5	36.8	67	1.5	1.0
Sidechain-Sidechain	1206	18.7	922	76.5	14.3	5	0.4	0.1

[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 (i)



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 (i)

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	24	154	372	978	0	1528	2.46	21.88	3.32	1.23
2	16	155	375	1123	0	1669	3.86	30.58	5.31	1.48
3	24	158	367	905	0	1454	2.28	20.82	3.05	1.2
4	18	157	373	1099	0	1647	4.06	33.68	5.68	1.39
5	19	161	356	987	0	1523	2.67	23.25	3.4	1.27
6	16	150	357	974	0	1497	2.73	19.86	3.66	1.2
7	15	148	340	1126	0	1629	4.39	35.25	6.46	1.4
8	19	148	346	975	0	1488	2.53	24.63	3.69	1.17
9	17	150	328	1063	0	1558	4.02	35.49	5.65	1.37
10	22	168	351	1152	0	1693	4.04	31.03	5.54	1.49
11	21	163	356	997	0	1537	2.07	18.08	2.51	1.19

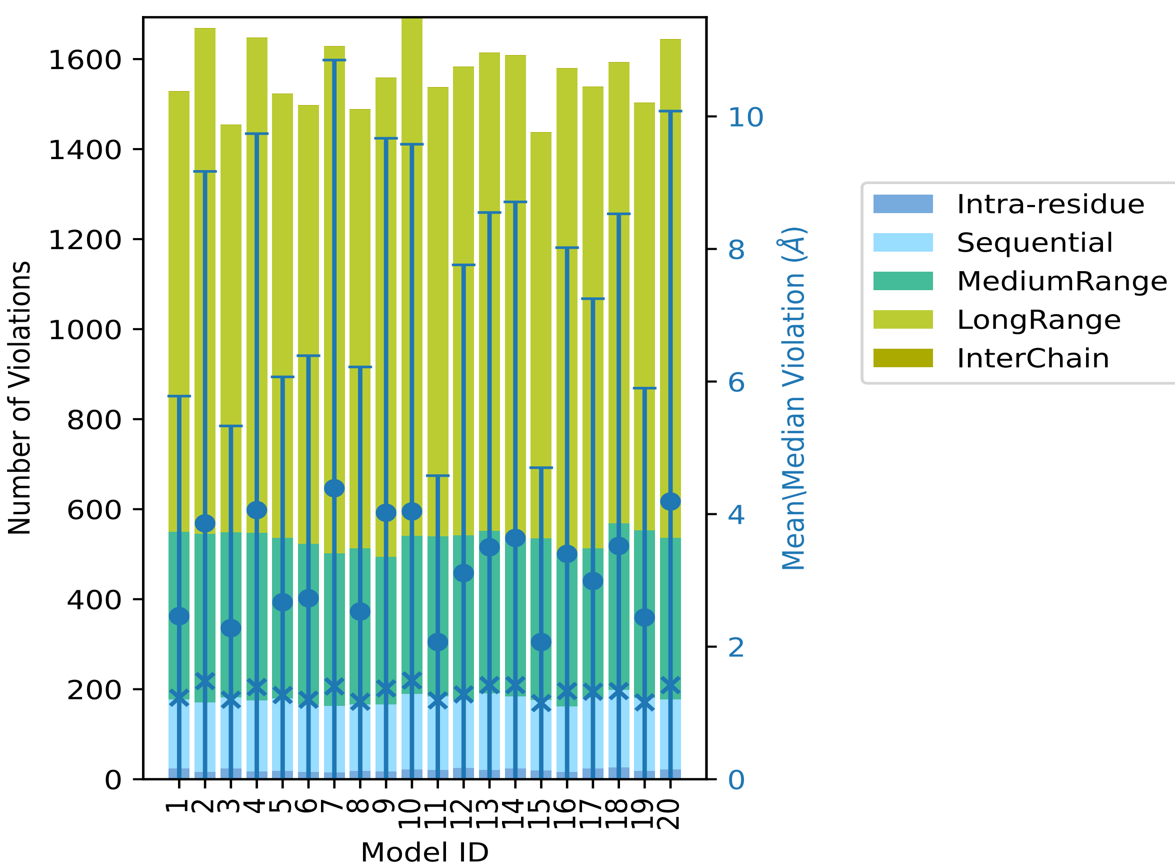
Continued on next page...

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	25	152	365	1041	0	1583	3.11	35.4	4.65	1.28
13	21	170	361	1062	0	1614	3.5	31.14	5.05	1.42
14	24	160	349	1075	0	1608	3.64	28.71	5.07	1.42
15	20	156	360	901	0	1437	2.07	18.1	2.63	1.15
16	16	146	342	1076	0	1580	3.4	26.05	4.62	1.33
17	24	158	332	1024	0	1538	2.99	30.27	4.26	1.32
18	26	173	370	1024	0	1593	3.52	35.79	5.01	1.33
19	19	162	372	950	0	1503	2.44	22.78	3.46	1.16
20	22	156	358	1108	0	1644	4.19	30.87	5.89	1.42

[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

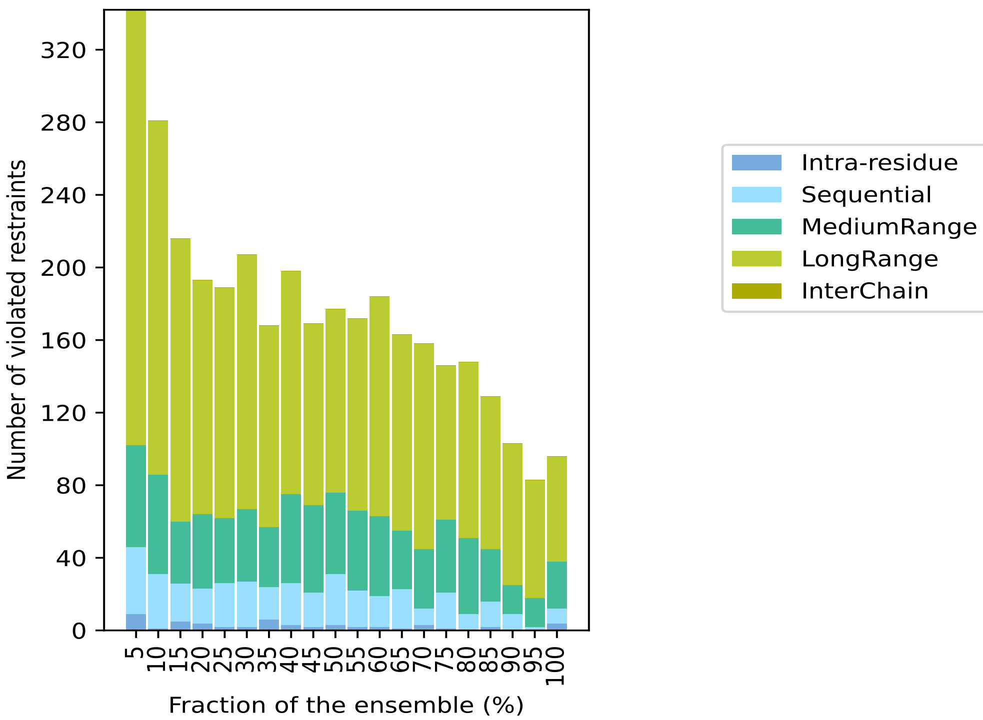
9.3 Distance violation statistics for the ensemble (i)

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, 2899(IR:270, SQ:716, MR:498, LR:1415, 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]	%
9	37	56	240	0	342	1	5.0
1	30	55	195	0	281	2	10.0
5	21	34	156	0	216	3	15.0
4	19	41	129	0	193	4	20.0
2	24	36	127	0	189	5	25.0
2	25	40	140	0	207	6	30.0
6	18	33	111	0	168	7	35.0
3	23	49	123	0	198	8	40.0
2	19	48	100	0	169	9	45.0
3	28	45	101	0	177	10	50.0
2	20	44	106	0	172	11	55.0
2	17	44	121	0	184	12	60.0
1	22	32	108	0	163	13	65.0
3	9	33	113	0	158	14	70.0
1	20	40	85	0	146	15	75.0
0	9	42	97	0	148	16	80.0
2	14	29	84	0	129	17	85.0
0	9	16	78	0	103	18	90.0
0	2	16	65	0	83	19	95.0
4	8	26	58	0	96	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

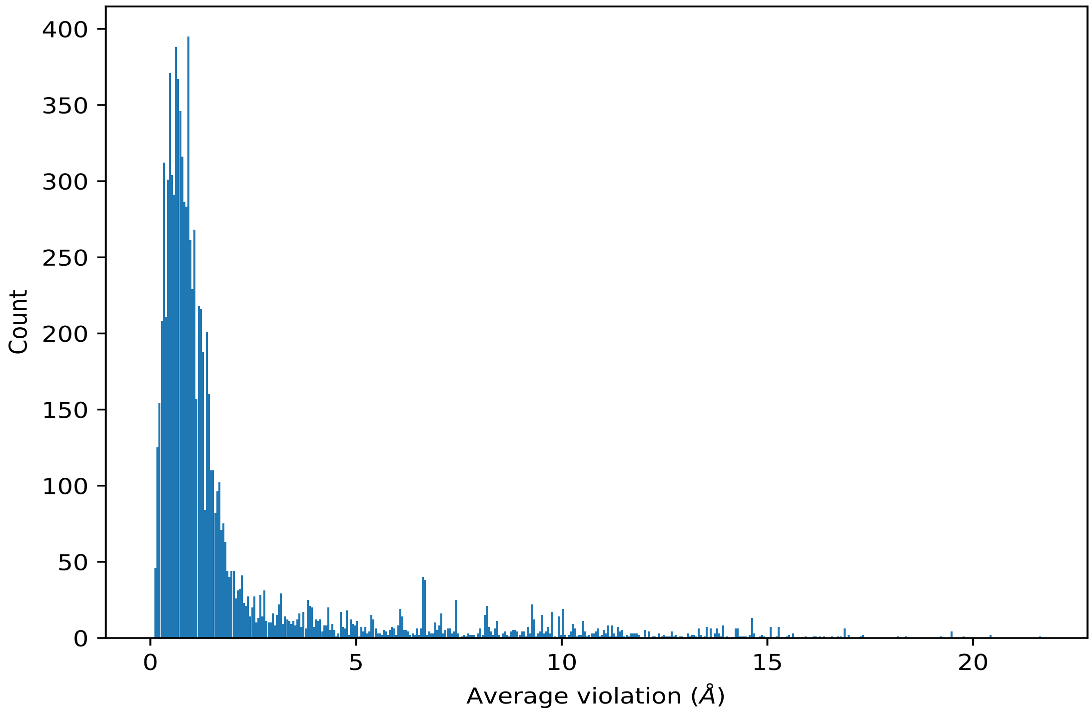
9.3.1 Bar graph : Distance violation statistics for the ensemble (i)



9.4 Most violated distance restraints in the ensemble (i)

9.4.1 Histogram : Distribution of mean distance violations (i)

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



9.4.2 Table: Most violated distance restraints (i)

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

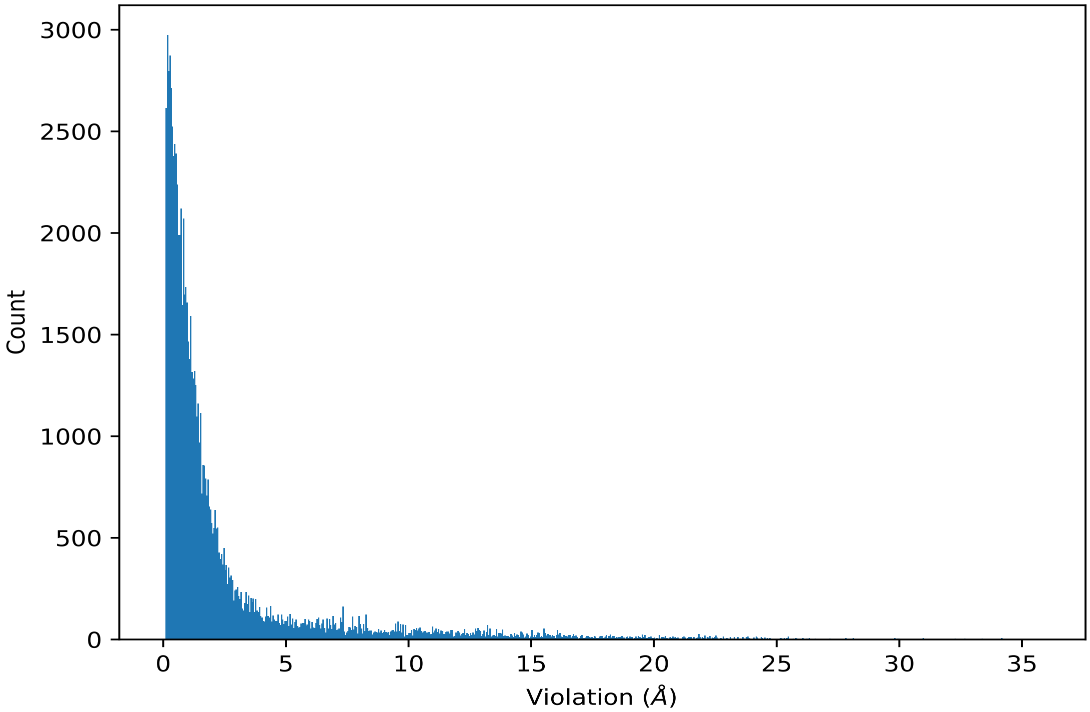
Key	Atom-1	Atom-2	Models[1]	Mean (Å)	SD[1] (Å)	Median (Å)
(4,2982)	1:A:4:GLU:H	1:A:147:GLN:H	20	19.79	11.91	17.78
(4,2234)	1:A:4:GLU:H	1:A:146:VAL:H	20	19.23	11.39	17.26
(5,373)	1:A:5:ASN:ND2	1:A:77:TYR:H	20	16.98	7.21	15.38
(5,95)	1:A:39:LEU:H	1:A:140:ASN:ND2	20	13.53	1.02	13.65
(1,545)	1:A:5:ASN:ND2	1:A:6:ARG:H	20	13.09	1.19	12.61
(4,1692)	1:A:9:VAL:H	1:A:159:ALA:H	20	11.3	8.03	12.36
(1,161)	1:A:59:PHE:H	1:A:225:ASN:ND2	20	11.25	1.3	10.97
(1,501)	1:A:4:GLU:H	1:A:5:ASN:ND2	20	11.07	1.05	10.8
(5,479)	1:A:5:ASN:ND2	1:A:78:ASP:H	20	10.28	6.75	9.14
(1,608)	1:A:5:ASN:ND2	1:A:7:ALA:H	20	10.23	1.68	10.18

[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)
(4,2982)	1:A:4:GLU:H	1:A:147:GLN:H	18	35.79
(4,2982)	1:A:4:GLU:H	1:A:147:GLN:H	9	35.49
(4,473)	1:A:4:GLU:H	1:A:143:LYS:H	9	35.43
(4,2982)	1:A:4:GLU:H	1:A:147:GLN:H	12	35.4
(4,473)	1:A:4:GLU:H	1:A:143:LYS:H	7	35.25
(4,473)	1:A:4:GLU:H	1:A:143:LYS:H	12	35.0
(4,945)	1:A:4:GLU:H	1:A:143:LYS:HD2	9	34.73
(4,945)	1:A:4:GLU:H	1:A:143:LYS:HD3	9	34.73
(4,2982)	1:A:4:GLU:H	1:A:147:GLN:H	7	34.6
(4,2234)	1:A:4:GLU:H	1:A:146:VAL:H	18	34.43
(4,2234)	1:A:4:GLU:H	1:A:146:VAL:H	9	34.3

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found