wwPDB NMR Structure Validation Summary Report (i)

Jun 3, 2023 – 08:55 PM EDT

PDB ID	:	2N1R
BMRB ID	:	25573
Title	:	NMR Structure of the Myristylated Feline Immunodeficiency Virus Matrix Protein
Authors	:	Brown, L.A.; Cox, C.; Button, R.J.; Baptiste, J.; Bahlow, K.; Spurrier, V.; Luttge, B.G.; Kuo, L.; Freed, E.O.; Summers, M.F.; Kyser, J.; Summers, H.R.
Deposited on	:	2015-04-15

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:

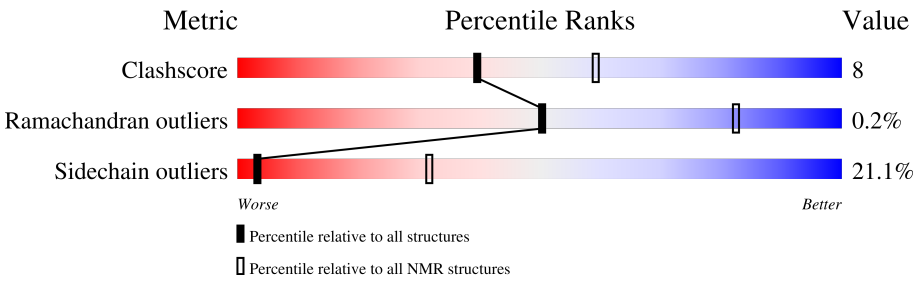
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 61%.

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	135	53% 17% 30%

2 Ensemble composition and analysis (i)

This entry contains 20 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 12 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:7-A:23, A:29-A:46, A:56-A:114 (94)	0.75	12

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2071 atoms, of which 1045 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Matrix protein p15.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	135	2071	640	1045	183	194	9	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MYR	-	modified residue	UNP P16087
A	5	ALA	GLN	engineered mutation	UNP P16087
A	6	SER	GLY	engineered mutation	UNP P16087

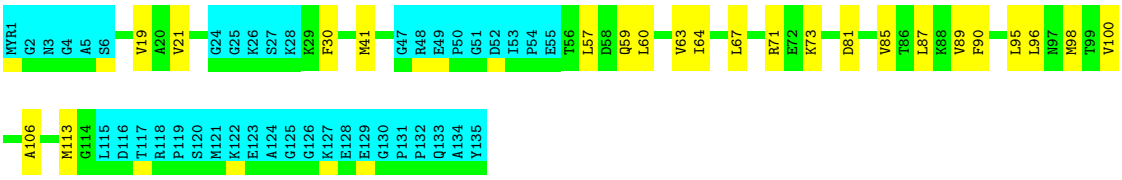
4 Residue-property plots (i)

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: Matrix protein p15

Chain A: 53% 17% 30%

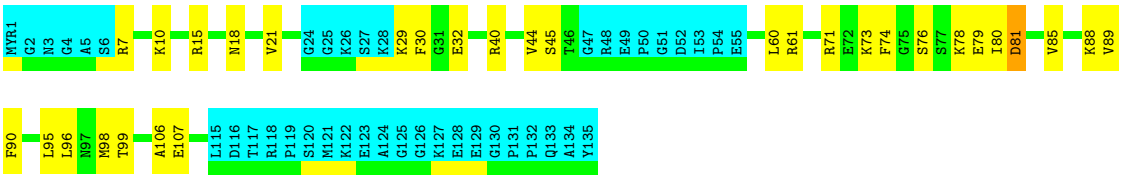


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 12. Colouring as in section 4.1 above.

- Molecule 1: Matrix protein p15

Chain A: 47% 22% • 30%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *CYANA*.

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *20 structures for lowest energy*.

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

Software name	Classification	Version
CYANA	structure calculation	
CYANA	refinement	

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	1005
Number of shifts mapped to atoms	1005
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	61%

6 Model quality (i)

6.1 Standard geometry (i)

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

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	728	749	749	11±4
All	All	14560	14979	14980	228

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 8.

5 of 91 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:100:VAL:HG21	1:A:106:ALA:HB2	1.03	1.29	16	7
1:A:64:ILE:HD11	1:A:84:ILE:HG23	0.91	1.38	5	4
1:A:57:LEU:HD21	1:A:95:LEU:HD13	0.78	1.54	18	1
1:A:42:ALA:HB2	1:A:67:LEU:HD22	0.73	1.57	2	1
1:A:21:VAL:HG22	1:A:30:PHE:CZ	0.68	2.23	12	11

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	94/135 (70%)	87±2 (93±2%)	7±2 (7±2%)	0±1 (0±1%)	50 82
All	All	1880/2700 (70%)	1743 (93%)	133 (7%)	4 (0%)	50 82

All 2 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	100	VAL	3
1	A	99	THR	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	77/105 (73%)	61±3 (79±3%)	16±3 (21±3%)	3 32
All	All	1540/2100 (73%)	1215 (79%)	325 (21%)	3 32

5 of 52 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	96	LEU	16
1	A	98	MET	15
1	A	73	LYS	12
1	A	41	MET	11
1	A	113	MET	11

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 61% for the well-defined parts and 57% 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 (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	1005
Number of shifts mapped to atoms	1005
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	130	1.94 ± 0.21	Should be applied
$^{13}C_{\beta}$	106	2.65 ± 0.30	Should be applied
$^{13}C'$	1	—	None (insufficient data)
^{15}N	117	0.84 ± 0.33	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 61%, i.e. 786 atoms were assigned a chemical shift out of a possible 1294. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	357/476 (75%)	179/194 (92%)	92/188 (49%)	86/94 (91%)
Sidechain	396/745 (53%)	258/487 (53%)	137/223 (61%)	1/35 (3%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	33/73 (45%)	16/36 (44%)	15/35 (43%)	2/2 (100%)
Overall	786/1294 (61%)	453/717 (63%)	244/446 (55%)	89/131 (68%)

7.1.4 Statistically unusual chemical shifts (i)

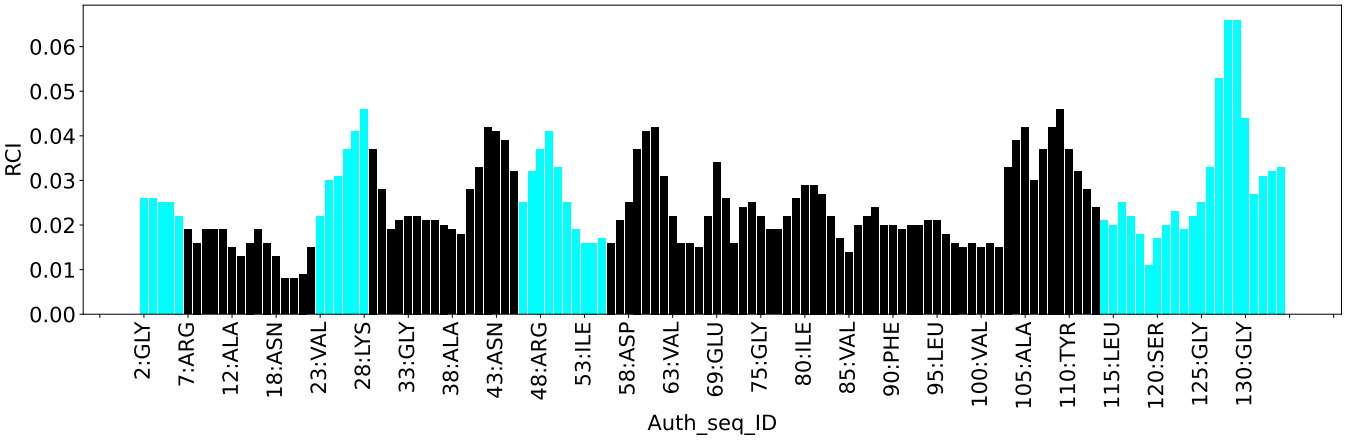
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	8	ASP	CB	26.50	32.98 – 48.76	-9.1
1	A	60	LEU	HB2	4.13	-0.07 – 3.30	7.5
1	A	10	LYS	CE	35.46	37.57 – 46.21	-7.4
1	A	119	PRO	HA	2.08	2.78 – 6.00	-7.2
1	A	87	LEU	CG	19.55	21.37 – 32.19	-6.7
1	A	87	LEU	CB	31.82	33.11 – 51.34	-5.7
1	A	60	LEU	CG	20.96	21.37 – 32.19	-5.4
1	A	67	LEU	CB	32.75	33.11 – 51.34	-5.2
1	A	62	LEU	CG	21.32	21.37 – 32.19	-5.0

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:


RCI
0.00
0.01
0.02
0.03
0.04
0.05
0.06
2:GLY
7:ARG
12:ALA
18:ASN
23:VAL
28:LYS
33:GLY
38:ALA
43:ASN
48:ARG
53:ILE
58:ASP
63:VAL
69:GLU
75:GLY
80:ILE
85:VAL
90:PHE
95:LEU
100:VAL
105:ALA
110:TYR
115:LEU
120:SER
125:GLY
130:GLY
Auth_seq_ID

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	372
Intra-residue (|i-j|=0)	2
Sequential (|i-j|=1)	105
Medium range (|i-j|>1 and |i-j|<5)	173
Long range (|i-j|≥5)	92
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	3
Number of restraints per residue	2.8
Number of long range restraints per residue[1]	0.7

[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 (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.There are no distance violations

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)

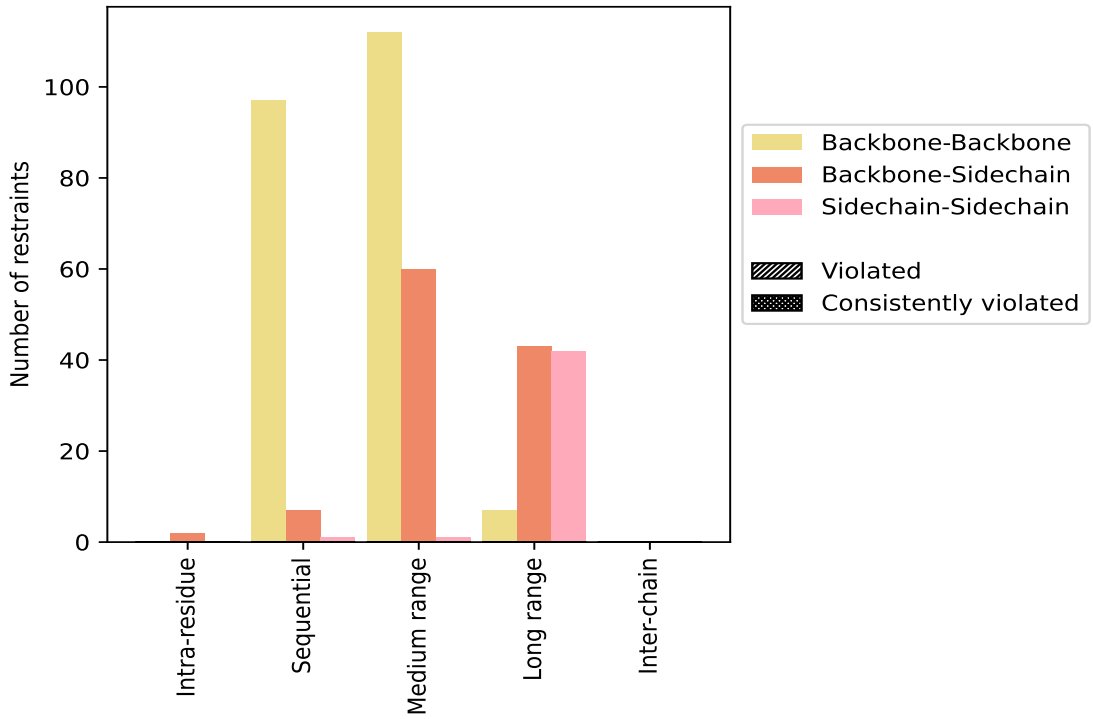
9.1 Summary of distance violations (i)

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)	2	0.5	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	2	0.5	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
Sequential (|i-j|=1)	105	28.2	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	97	26.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	7	1.9	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	1	0.3	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	173	46.5	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	112	30.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	60	16.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	1	0.3	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	92	24.7	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	7	1.9	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	43	11.6	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	42	11.3	0	0.0	0.0	0	0.0	0.0
Inter-chain	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	0	0.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	0	0.0	0	0.0	0.0	0	0.0	0.0
Hydrogen bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Disulfide bond	0	0.0	0	0.0	0.0	0	0.0	0.0
Total	372	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	216	58.1	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	112	30.1	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	44	11.8	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 (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)

No violations found

9.3 Distance violation statistics for the ensemble (i)

No violations found

9.4 Most violated distance restraints in the ensemble (i)

No violations found

9.5 All violated distance restraints (i)

No violations found

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found