Full wwPDB NMR Structure Validation Report (i)

Jun 5, 2023 – 02:05 PM JST

PDB ID	:	5XM4
BMRB ID	:	36083
Title	:	Solution NMR structure of a new lasso peptide subterisin
Authors	:	Hemmi, H.; Kodani, S.; Kuroha, M.; Ohnishi-Kameyama, M.
Deposited on	:	2017-05-12

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

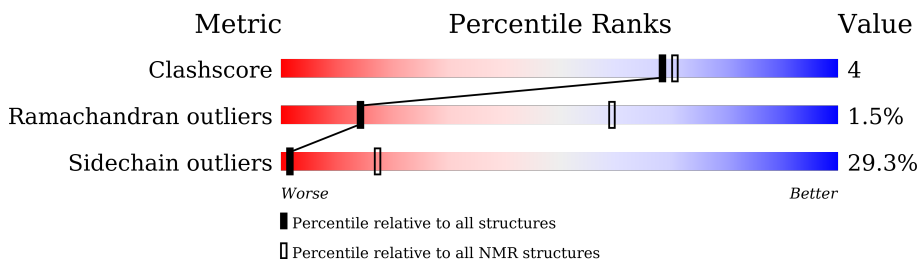|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)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

1 Overall quality at a glance ⓘ

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

The overall completeness of chemical shifts assignment is 79%.

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	Percentile Ranks	Value
Clashscore		4
Ramachandran outliers		1.5%
Sidechain outliers		29.3%

Worse   *Better*

▌ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	17	59% 18% 24%

2 Ensemble composition and analysis ⓘ

This entry contains 15 models. Model 15 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:2-A:14 (13)	0.23	15

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. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 14, 15
2	7, 11

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 239 atoms, of which 120 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Uncharacterized protein.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	17	239	78	120	21	20	0

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: Uncharacterized protein

Chain A:

59% 18% 24%

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Uncharacterized protein

Chain A:

59% 12% 6% 24%

4.2.2 Score per residue for model 2

• Molecule 1: Uncharacterized protein
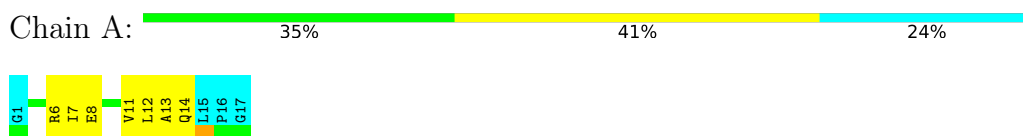
Chain A:

35% 41% 24%

4.2.3 Score per residue for model 3

• Molecule 1: Uncharacterized protein
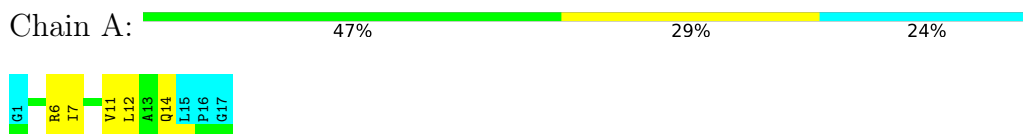
Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Uncharacterized protein
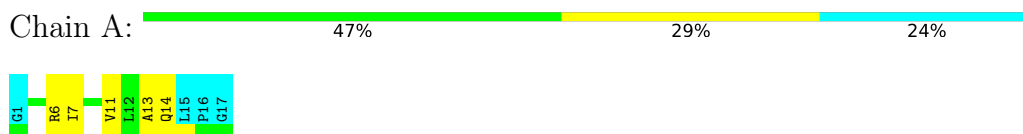
Chain A:

4.2.5 Score per residue for model 5

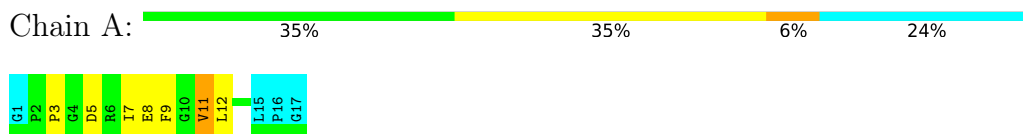• Molecule 1: Uncharacterized protein

Chain A:

4.2.6 Score per residue for model 6

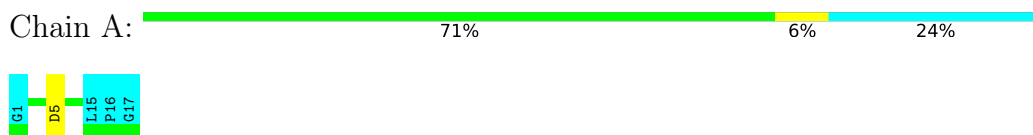• Molecule 1: Uncharacterized protein

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: Uncharacterized protein

Chain A:

4.2.8 Score per residue for model 8

• Molecule 1: Uncharacterized protein

Chain A:

4.2.9 Score per residue for model 9

• Molecule 1: Uncharacterized protein

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Uncharacterized protein

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Uncharacterized protein

Chain A:

4.2.12 Score per residue for model 12

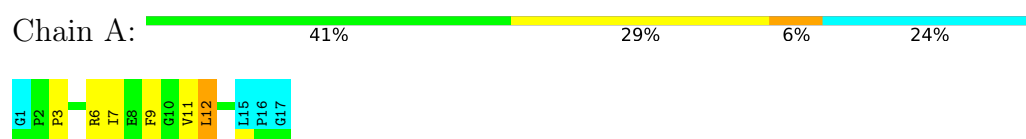• Molecule 1: Uncharacterized protein

Chain A:

4.2.13 Score per residue for model 13

• Molecule 1: Uncharacterized protein

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: Uncharacterized protein

Chain A:

4.2.15 Score per residue for model 15 (medoid)

• Molecule 1: Uncharacterized protein

Chain A:

5 Refinement protocol and experimental data overview ⓘ

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

Of the 300 calculated structures, 15 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
CNS	refinement	
CNS	structure calculation	

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	173
Number of shifts mapped to atoms	173
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	79%

6   Model quality ⓘ

6.1   Standard geometry ⓘ

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 ⓘ

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	97	96	96	1±1
All	All	1455	1440	1440	11

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 4.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:6:ARG:HG3	1:A:13:ALA:HB3	0.61	1.72	6	3
1:A:9:PHE:O	1:A:9:PHE:CD1	0.47	2.68	11	2
1:A:3:PRO:HB3	1:A:12:LEU:HD13	0.47	1.86	14	1
1:A:3:PRO:CB	1:A:12:LEU:HD13	0.46	2.40	14	1
1:A:12:LEU:HD12	1:A:12:LEU:O	0.45	2.11	14	1
1:A:8:GLU:HB2	1:A:13:ALA:HB2	0.45	1.89	2	1
1:A:9:PHE:O	1:A:9:PHE:CD2	0.45	2.69	4	1
1:A:8:GLU:N	1:A:11:VAL:O	0.43	2.51	7	1

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	13/17 (76%)	8±1 (64±5%)	4±1 (34±6%)	0±0 (2±3%)	14	59
All	All	195/255 (76%)	125 (64%)	67 (34%)	3 (2%)	14	59

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	3	PRO	3

6.3.2   Protein sidechains ⓘ

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	10/12 (83%)	7±1 (71±14%)	3±1 (29±14%)	1	17
All	All	150/180 (83%)	106 (71%)	44 (29%)	1	17

All 7 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	11	VAL	13
1	A	6	ARG	7
1	A	7	ILE	7
1	A	12	LEU	7
1	A	14	GLN	6
1	A	5	ASP	3
1	A	9	PHE	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 79% for the well-defined parts and 80% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *subterisin-chemical-shifts.tbl*

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	173
Number of shifts mapped to atoms	173
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 79%, i.e. 137 atoms were assigned a chemical shift out of a possible 173. 0 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	39/63 (62%)	26/26 (100%)	13/26 (50%)	0/11 (0%)
Sidechain	88/100 (88%)	61/65 (94%)	27/31 (87%)	0/4 (0%)
Aromatic	10/10 (100%)	5/5 (100%)	5/5 (100%)	0/0 (—%)
Overall	137/173 (79%)	92/96 (96%)	45/62 (73%)	0/15 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 171 atoms were assigned a chemical shift out of a possible 215. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

|	Total	^{1}H	^{13}C	^{15}N
Backbone	51/83 (61%)	34/35 (97%)	17/34 (50%)	0/14 (0%)
Sidechain	110/122 (90%)	76/80 (95%)	34/38 (89%)	0/4 (0%)
Aromatic	10/10 (100%)	5/5 (100%)	5/5 (100%)	0/0 (—%)
Overall	171/215 (80%)	115/120 (96%)	56/77 (73%)	0/18 (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:
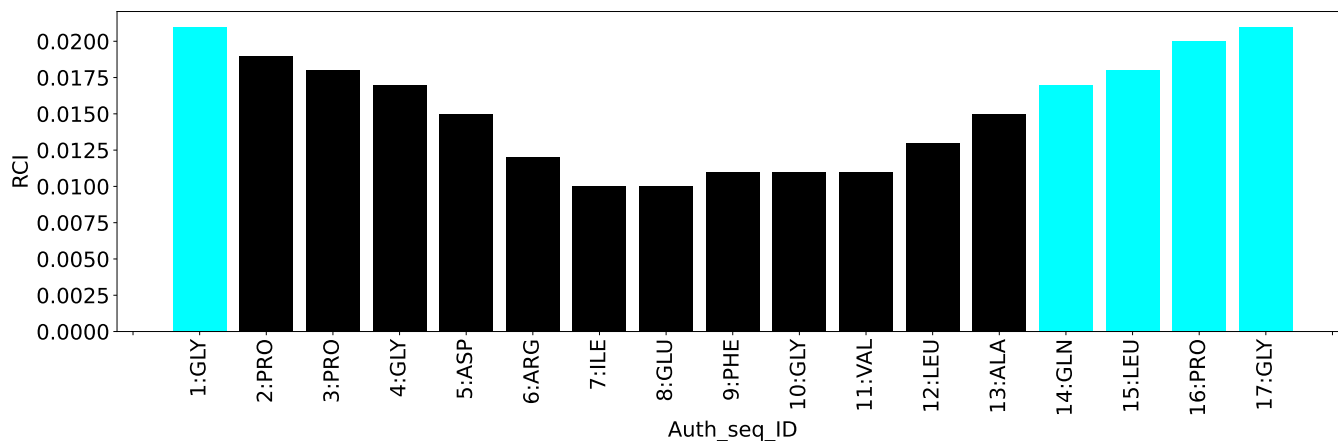
8 NMR restraints analysis ⓘ

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	183
Intra-residue ($|i\text{-}j|$=0)	23
Sequential ($|i\text{-}j|$=1)	80
Medium range ($|i\text{-}j|$>1 and $|i\text{-}j|$<5)	19
Long range ($|i\text{-}j|$≥5)	61
Inter-chain	0
Hydrogen bond restraints	0
Disulfide bond restraints	0
Total dihedral-angle restraints	10
Number of unmapped restraints	8
Number of restraints per residue	11.4
Number of long range restraints per residue[1]	3.6

[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.There are no distance violations

8.2.2 Average number of dihedral-angle violations per model ⓘ

Dihedral-angle violations less than 1° are not included in the calculation.There are no dihedral-angle violations
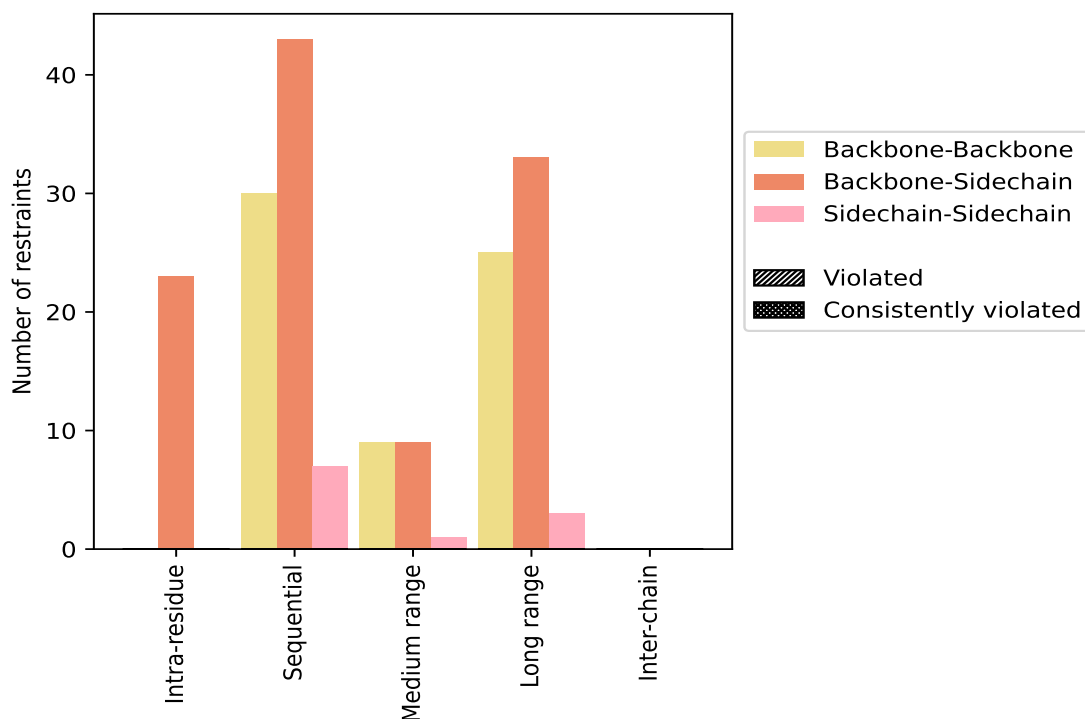
9 Distance violation analysis ⓘ

9.1 Summary of distance violations ⓘ

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)	23	12.6	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	23	12.6	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)	80	43.7	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	30	16.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	43	23.5	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	7	3.8	0	0.0	0.0	0	0.0	0.0
Medium range (|i-j|>1 & |i-j|<5)	19	10.4	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	9	4.9	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	9	4.9	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	1	0.5	0	0.0	0.0	0	0.0	0.0
Long range (|i-j|≥5)	61	33.3	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	25	13.7	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	33	18.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	3	1.6	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	183	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	64	35.0	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	108	59.0	0	0.0	0.0	0	0.0	0.0
Sidechain-Sidechain	11	6.0	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 (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)

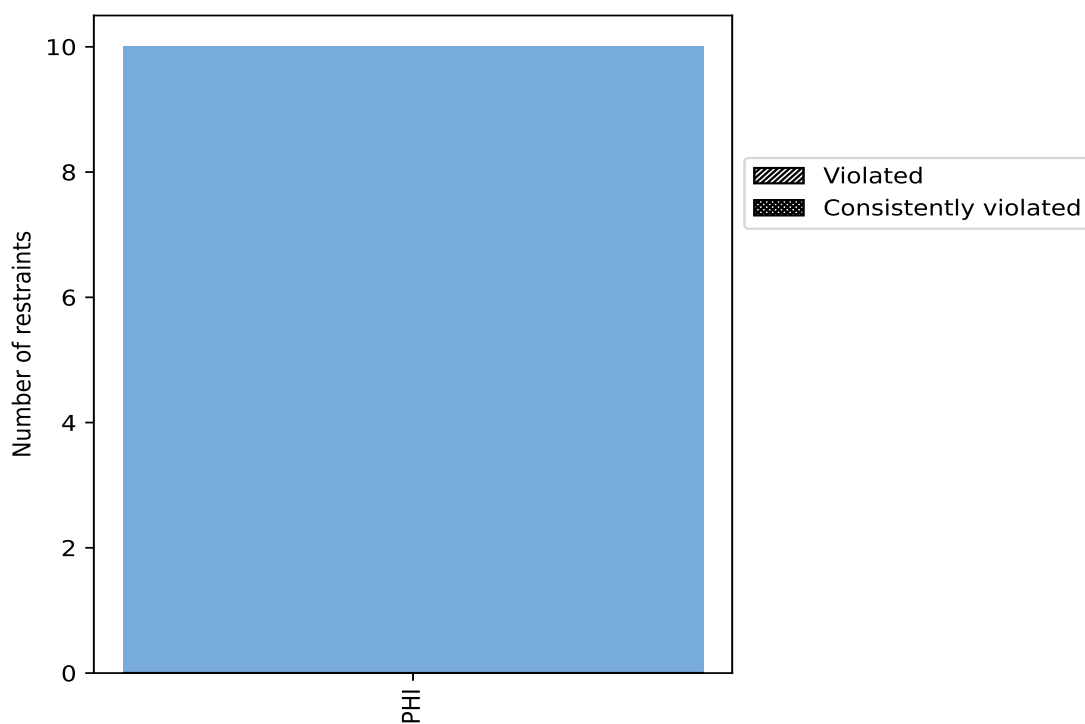No violations found

10 Dihedral-angle violation analysis (i)

10.1 Summary of dihedral-angle violations (i)

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	10	100.0	0	0.0	0.0	0	0.0	0.0
Total	10	100.0	0	0.0	0.0	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 (i)

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)

No violations found

10.3 Dihedral-angle violation statistics for the ensemble ⓘ

No violations found

10.4 Most violated dihedral-angle restraints in the ensemble ⓘ

No violations found

10.5 All violated dihedral-angle restraints ⓘ

No violations found