Full wwPDB NMR Structure Validation Report (i)

Jun 3, 2023 – 11:30 PM EDT

PDB ID	:	2KHK
BMRB ID	:	16247
Title	:	NMR solution structure of the b30-82 domain of subunit b of Escherichia coli F1FO ATP synthase
Authors	:	Priya, R.; Biukovic, G.; Gayen, S.; Vivekanandan, S.; Gruber, G.
Deposited on	:	2009-04-08

This is a Full wwPDB NMR Structure Validation 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:

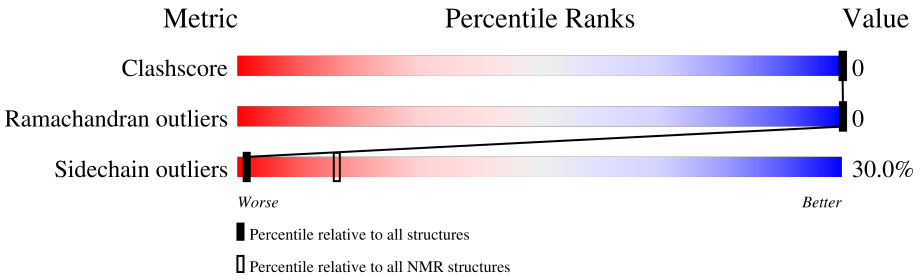
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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 89%.

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	53	40% · 58%

2 Ensemble composition and analysis

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:12-A:33 (22)	0.21	1

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, 4, 5, 7, 8, 9, 10
2	3, 6
Single-model clusters	2

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 828 atoms, of which 425 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ATP synthase subunit b.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	53	828	246	425	77	79	1	0

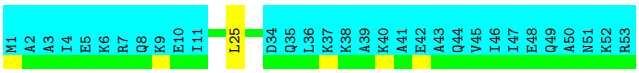
4 Residue-property plots

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: ATP synthase subunit b

Chain A:



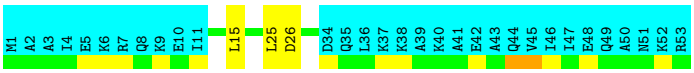
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: ATP synthase subunit b

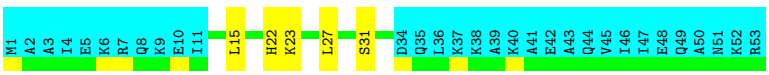
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: ATP synthase subunit b

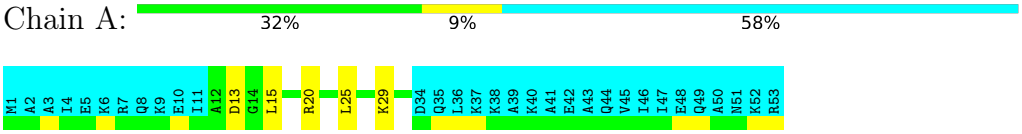
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: ATP synthase subunit b

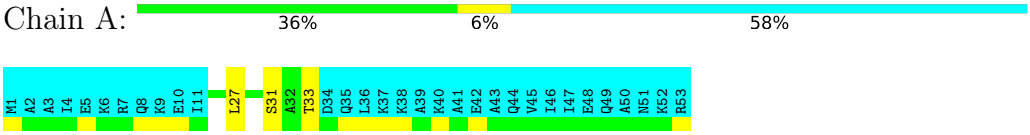
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: ATP synthase subunit b

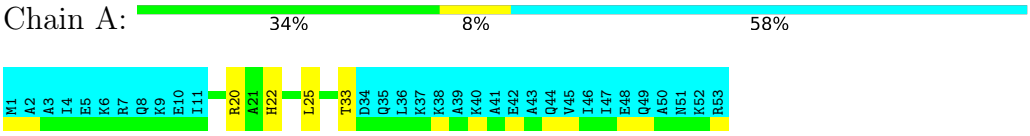
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: ATP synthase subunit b

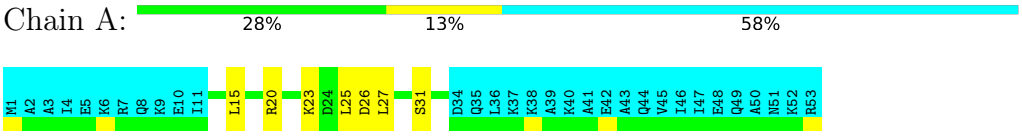
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: ATP synthase subunit b

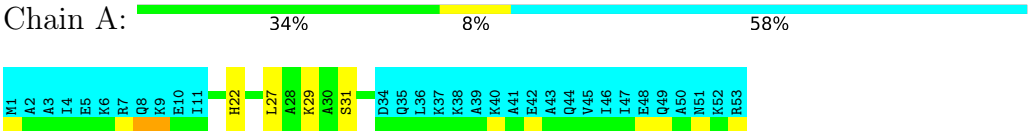
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: ATP synthase subunit b

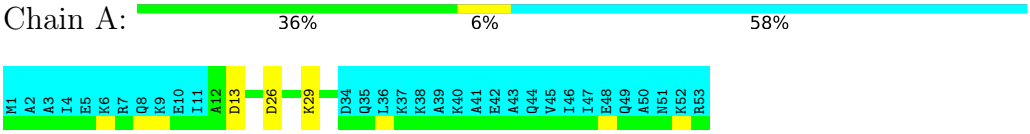
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: ATP synthase subunit b

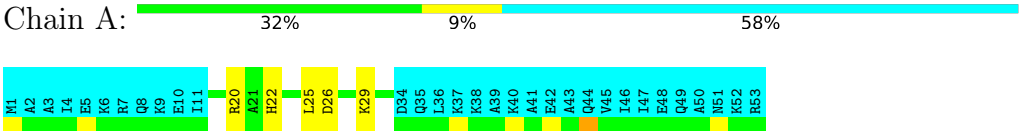
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: ATP synthase subunit b

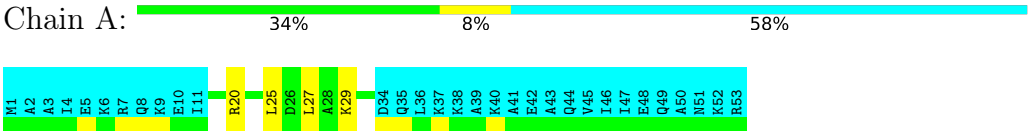
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: ATP synthase subunit b

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics*.

Of the 10 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CYANA	refinement	2.1
MOLMOL	structure solution	

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	587
Number of shifts mapped to atoms	586
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

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	154	152	152	0±0
All	All	1540	1520	1520	-

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 -.

There are no clashes.

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	22/53 (42%)	22±0 (100±1%)	0±0 (0±1%)	0±0 (0±0%)	100	100
All	All	220/530 (42%)	219 (100%)	1 (0%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	14/39 (36%)	10±1 (70±9%)	4±1 (30±9%)	1 16
All	All	140/390 (36%)	98 (70%)	42 (30%)	1 16

All 11 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	25	LEU	6
1	A	27	LEU	5
1	A	20	ARG	5
1	A	29	LYS	5
1	A	26	ASP	4
1	A	22	HIS	4
1	A	31	SER	4
1	A	15	LEU	3
1	A	23	LYS	2
1	A	13	ASP	2
1	A	33	THR	2

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

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 89% for the well-defined parts and 81% 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	587
Number of shifts mapped to atoms	586
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	10	GLU	HE2	2.727	0.020	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	52	-0.92 ± 0.12	Should be checked
$^{13}C_{\beta}$	51	0.39 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	48	-1.05 ± 0.14	Should be applied
^{15}N	51	-1.06 ± 0.14	Should be applied

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 89%, i.e. 235 atoms were assigned a chemical shift out of a possible 263. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	110/111 (99%)	45/45 (100%)	43/44 (98%)	22/22 (100%)
Sidechain	125/145 (86%)	86/95 (91%)	39/45 (87%)	0/5 (0%)
Aromatic	0/7 (0%)	0/4 (0%)	0/2 (0%)	0/1 (0%)
Overall	235/263 (89%)	131/144 (91%)	82/91 (90%)	22/28 (79%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 81%, i.e. 585 atoms were assigned a chemical shift out of a possible 720. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	253/266 (95%)	102/107 (95%)	100/106 (94%)	51/53 (96%)
Sidechain	332/447 (74%)	222/288 (77%)	110/137 (80%)	0/22 (0%)
Aromatic	0/7 (0%)	0/4 (0%)	0/2 (0%)	0/1 (0%)
Overall	585/720 (81%)	324/399 (81%)	210/245 (86%)	51/76 (67%)

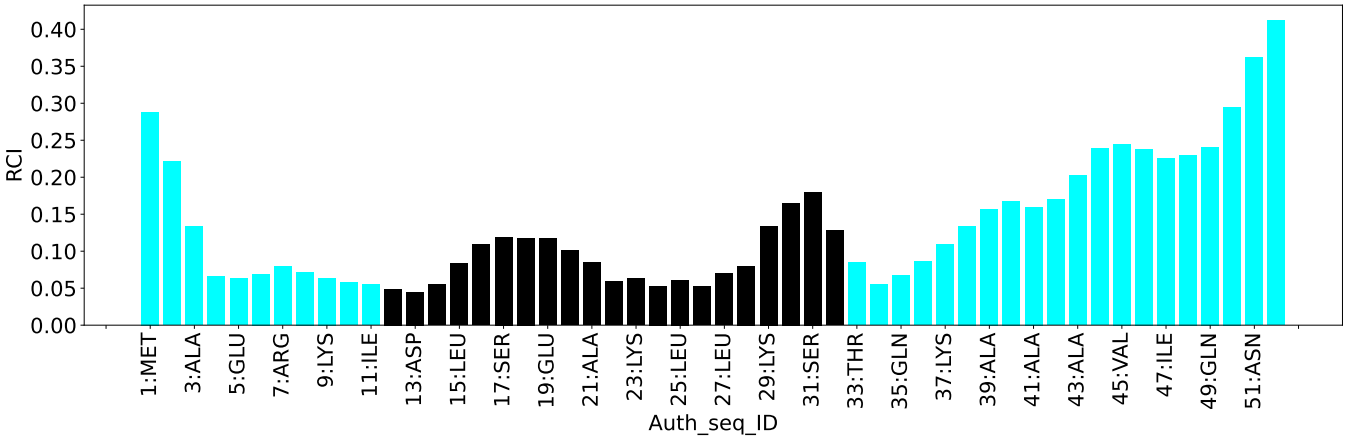
7.1.4 Statistically unusual chemical shifts

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:


RCI
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1:MET
3:ALA
5:GLU
7:ARG
9:LYS
11:ILE
13:ASP
15:LEU
17:SER
19:GLU
21:ALA
23:LYS
25:LEU
27:LEU
29:LYS
31:SER
33:THR
35:GLN
37:LYS
39:ALA
41:ALA
43:ALA
45:VAL
47:ILE
49:GLN
51:ASN
Auth_seq_ID

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	158
Intra-residue (|i-j|=0)	73
Sequential (|i-j|=1)	68
Medium range (|i-j|>1 and |i-j|<5)	6
Long range (|i-j|≥5)	0
Inter-chain	0
Hydrogen bond restraints	11
Disulfide bond restraints	0
Total dihedral-angle restraints	0
Number of unmapped restraints	4
Number of restraints per residue	3.0
Number of long range restraints per residue[1]	0.0

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

8.2 Residual restraint violations

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

8.2.1 Average number of distance violations per model

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

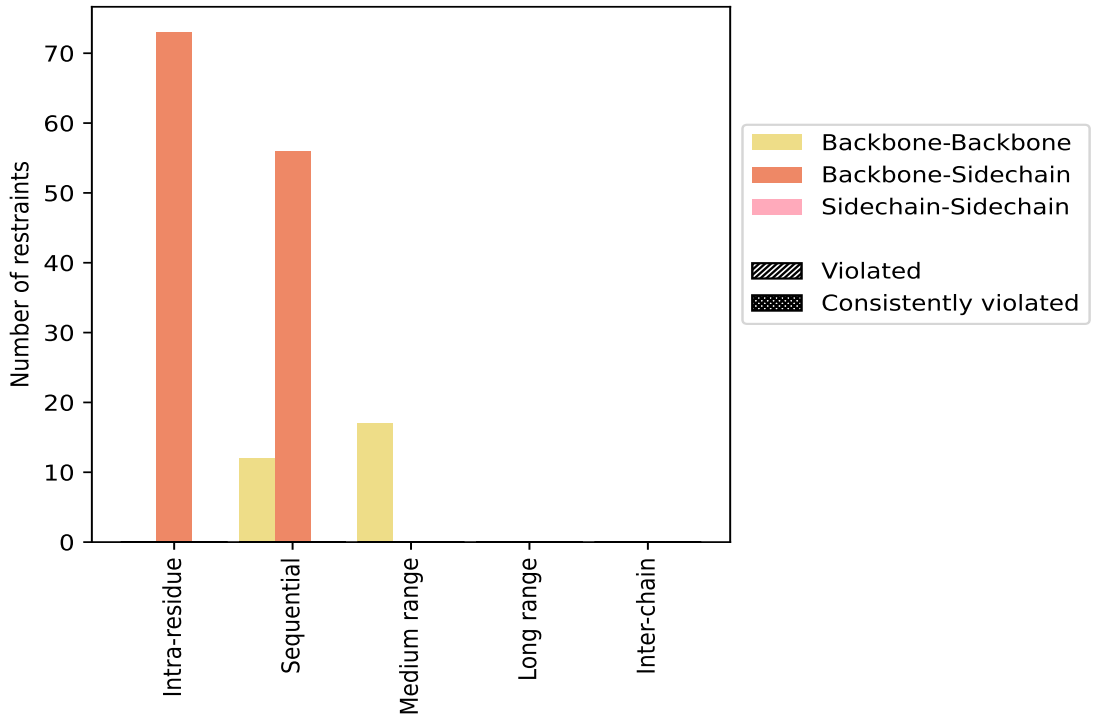
9.1 Summary of distance violations (i)

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)	73	46.2	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	73	46.2	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)	68	43.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	12	7.6	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	56	35.4	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
Medium range (|i-j|>1 & |i-j|<5)	6	3.8	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	6	3.8	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
Long range (|i-j|≥5)	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
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	11	7.0	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	158	100.0	0	0.0	0.0	0	0.0	0.0
Backbone-Backbone	29	18.4	0	0.0	0.0	0	0.0	0.0
Backbone-Sidechain	129	81.6	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

[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

No violations found

9.3 Distance violation statistics for the ensemble

No violations found

9.4 Most violated distance restraints in the ensemble

No violations found

9.5 All violated distance restraints

No violations found

10 Dihedral-angle violation analysis (i)

Dihedral angle analysis failed due to data error in the dihedral angle restraints, possibly missing target value