WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Apr 21, 2024 – 05:44 PM EDT

PDB ID	:	2M5H
BMRB ID	:	18548
Title	:	NMR structure note: solution structure of monomeric human FAM96A
Authors	:	Ouyang, B.; Xia, B.
Deposited on	:	2013-02-25

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

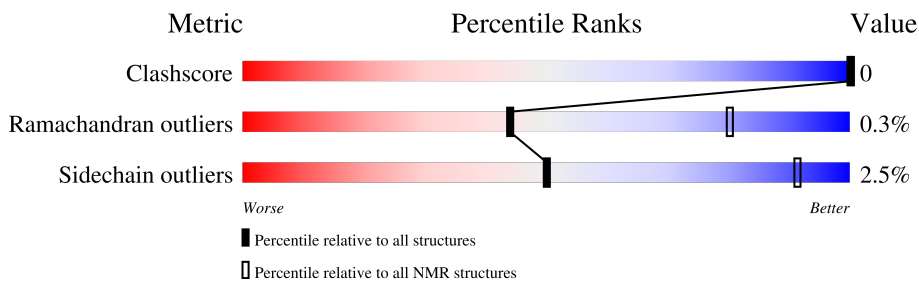|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		0.3%
Sidechain outliers		2.5%

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	139	86% · 13%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. 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: *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:33-A:119, A:125-A:158 (121)	0.84	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 5 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2276 atoms, of which 1141 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called MIP18 family protein FAM96A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	139	2276	709	1141	195	223	8	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	27	MET	-	expression tag	UNP Q9H5X1
A	161	GLN	-	expression tag	UNP Q9H5X1
A	162	LEU	-	expression tag	UNP Q9H5X1
A	163	GLU	-	expression tag	UNP Q9H5X1
A	164	HIS	-	expression tag	UNP Q9H5X1
A	165	HIS	-	expression tag	UNP Q9H5X1

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.
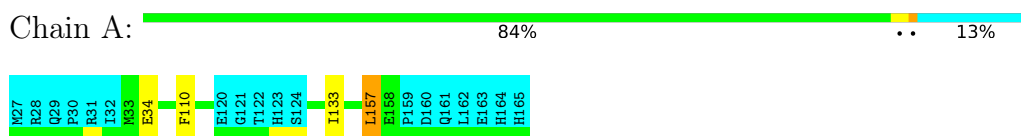
• Molecule 1: MIP18 family protein FAM96A

Chain A:
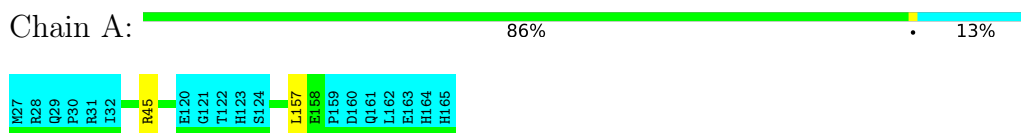
4.2 Scores per residue for each member of the ensemble
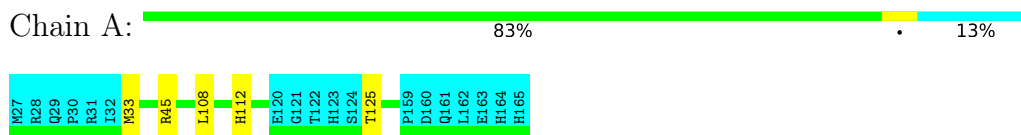
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
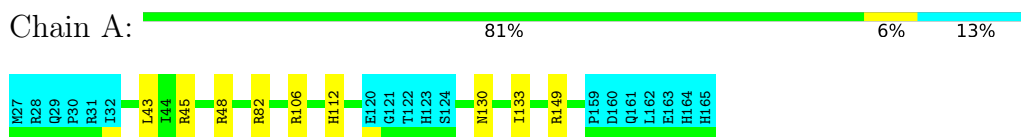
• Molecule 1: MIP18 family protein FAM96A

Chain A:

4.2.2 Score per residue for model 2
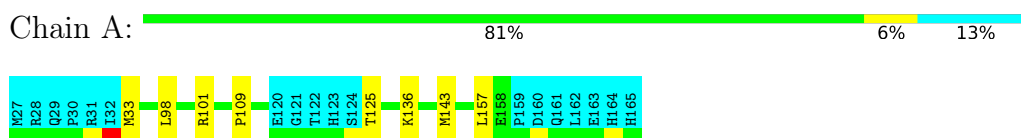
• Molecule 1: MIP18 family protein FAM96A

Chain A:

4.2.3 Score per residue for model 3
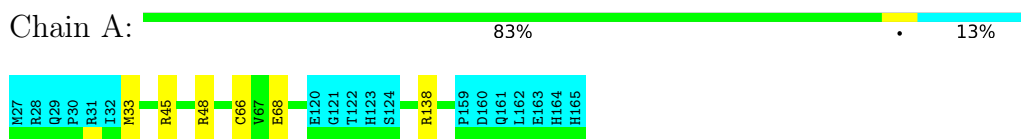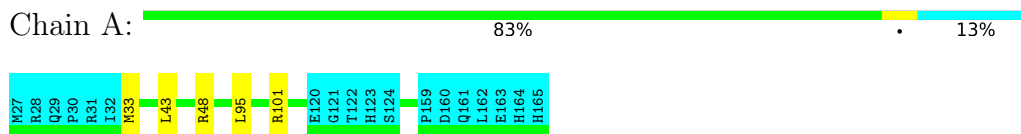
• Molecule 1: MIP18 family protein FAM96A

Chain A:

81%   6%   13%

M27 R28 Q29 P30 R31 I32 L43 I44 R45 R48 R82 R106 H112 E120 G121 T122 H123 S124 N130 I133 R149 P159 D160 Q161 L162 E163 H164 H165

4.2.4 Score per residue for model 4

• Molecule 1: MIP18 family protein FAM96A

Chain A:

81%   6%   13%

M27 R28 Q29 P30 R31 I32 M33 L98 R101 P109 E120 G121 T122 H123 S124 T125 K136 M143 L157 E158 P159 D160 Q161 L162 E163 H164 H165

4.2.5 Score per residue for model 5

• Molecule 1: MIP18 family protein FAM96A

Chain A:

83%   •   13%

M27 R28 Q29 P30 R31 I32 M33 R45 R48 C66 V67 E68 E120 G121 T122 H123 S124 R138 P159 D160 Q161 L162 E163 H164 H165

4.2.6 Score per residue for model 6

• Molecule 1: MIP18 family protein FAM96A

Chain A:

83%   •   13%

M27 R28 Q29 P30 R31 I32 M33 L43 R48 L95 R101 E120 G121 T122 H123 S124 P159 D160 Q161 L162 E163 H164 H165

4.2.7 Score per residue for model 7

• Molecule 1: MIP18 family protein FAM96A

Chain A:

83%   •   13%

M27 R28 Q29 P30 R31 I32 M33 R45 E68 E120 G121 T122 H123 S124 R149 L157 E158 P159 D160 Q161 L162 E163 H164 H165
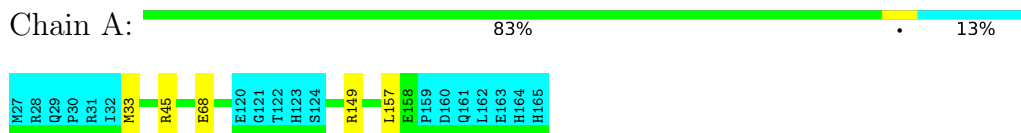
4.2.8 Score per residue for model 8

• Molecule 1: MIP18 family protein FAM96A

Chain A:

84% 13%

M27 R28 Q29 P30 R31 I32 L43 R82 E120 G121 T122 H123 S124 R138 L157 E158 P159 D160 Q161 L162 E163 H164 H165
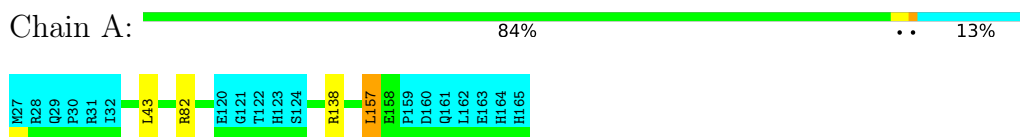
4.2.9 Score per residue for model 9

• Molecule 1: MIP18 family protein FAM96A

Chain A:

85% 13%

M27 R28 Q29 P30 R31 I32 R45 E120 G121 T122 H123 S124 T125 R149 P159 D160 Q161 L162 E163 H164 H165
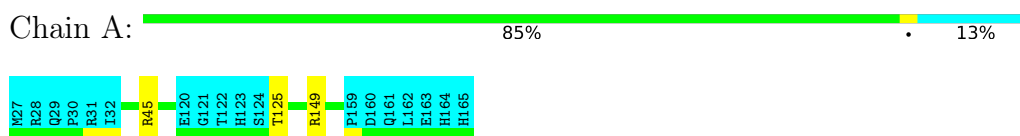
4.2.10 Score per residue for model 10

• Molecule 1: MIP18 family protein FAM96A

Chain A:

80% 6% 13%

M27 R28 Q29 P30 R31 I32 M33 L43 I44 R45 N73 E74 R82 R106 E120 G121 T122 H123 S124 R138 R149 L157 E158 P159 D160 Q161 L162 E163 H164 H165
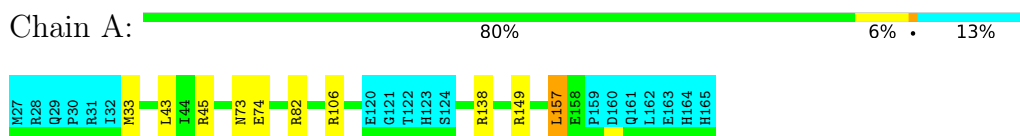
4.2.11 Score per residue for model 11

• Molecule 1: MIP18 family protein FAM96A

Chain A:

84% 13%

M27 R28 Q29 P30 R31 I32 L43 I44 R45 N73 R106 E120 G121 T122 H123 S124 P159 D160 Q161 L162 E163 H164 H165

4.2.12 Score per residue for model 12 (medoid)

• Molecule 1: MIP18 family protein FAM96A

Chain A:

84% 13%

M27 R28 Q29 P30 R31 I32 R45 H112 I116 E120 G121 T122 H123 S124 L157 E158 P159 D160 Q161 L162 E163 H164 H165

4.2.13 Score per residue for model 13

• Molecule 1: MIP18 family protein FAM96A

Chain A:

83% • 13%

M27 R28 Q29 P30 R31 I32 R45 C66 V67 E68 N73 C90 E120 G121 T122 H123 S124 R149 P159 D160 Q161 L162 E163 H164 H165

4.2.14 Score per residue for model 14

• Molecule 1: MIP18 family protein FAM96A

Chain A:

83% • 13%

M27 R28 Q29 P30 R31 I32 R45 R82 R106 E120 G121 T122 H123 S124 T125 E126 R138 P159 D160 Q161 L162 E163 H164 H165

4.2.15 Score per residue for model 15

• Molecule 1: MIP18 family protein FAM96A

Chain A:

84% • 13%

M27 R28 Q29 P30 R31 I32 R45 T86 E120 G121 T122 H123 S124 R138 V156 P159 D160 Q161 L162 E163 H164 H165

4.2.16 Score per residue for model 16

• Molecule 1: MIP18 family protein FAM96A

Chain A:

81% 5% • 13%

M27 R28 Q29 P30 R31 I32 L43 I44 R45 R48 E64 R82 R106 I118 S119 E120 G121 T122 H123 S124 L157 E158 P159 D160 Q161 L162 E163 H164 H165

4.2.17 Score per residue for model 17

• Molecule 1: MIP18 family protein FAM96A

Chain A:

83% • 13%

M27 R28 Q29 P30 R31 I32 M33 L43 I44 R45 R82 K111 E120 G121 T122 H123 S124 P159 D160 Q161 L162 E163 H164 H165

4.2.18 Score per residue for model 18

• Molecule 1: MIP18 family protein FAM96A

Chain A:

4.2.19 Score per residue for model 19
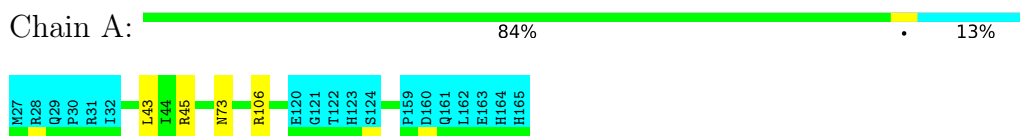
• Molecule 1: MIP18 family protein FAM96A

Chain A:

4.2.20 Score per residue for model 20
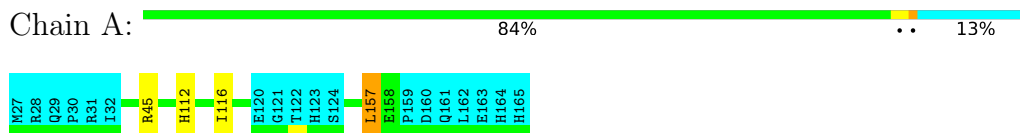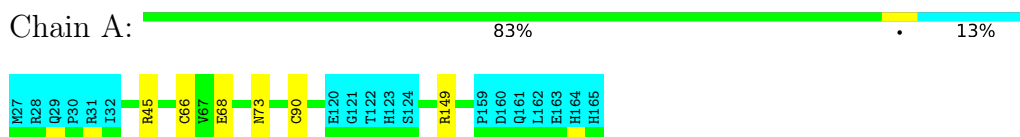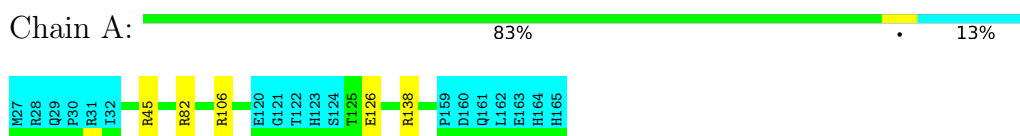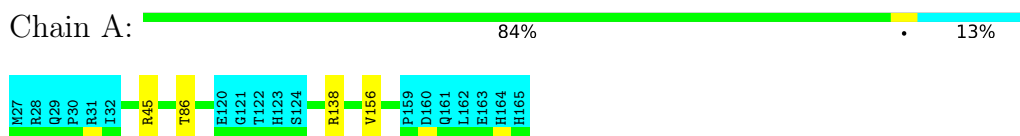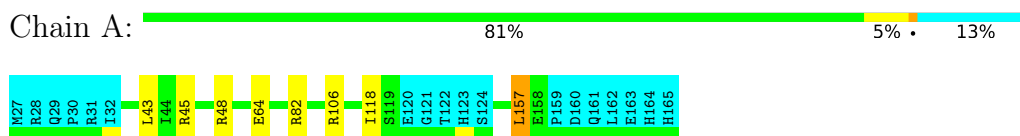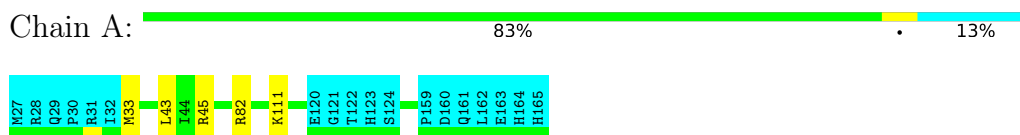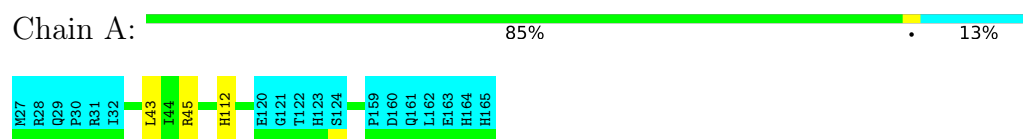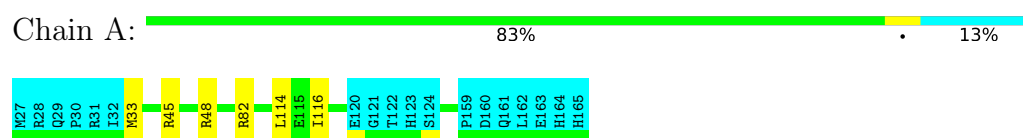
• Molecule 1: MIP18 family protein FAM96A

Chain A:
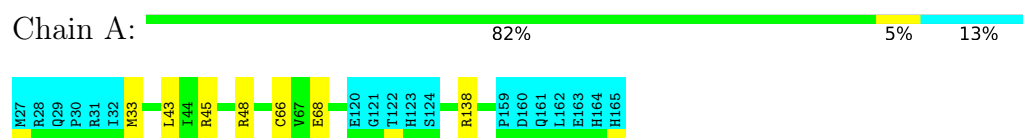
5 Refinement protocol and experimental data overview ⓘ

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

Of the 100 calculated structures, 20 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
Amber	refinement	9.0
DYANA	refinement	2.1
CYANA	structure solution	2.1

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	1735
Number of shifts mapped to atoms	1735
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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	0.63±0.00	0±0/996 (0.0± 0.0%)	0.94±0.03	3±2/1351 (0.3± 0.1%)
All	All	0.63	0/19920 (0.0%)	0.94	68/27020 (0.3%)

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±0.2
All	All	0	1

There are no bond-length outliers.

All 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
1	A	82	ARG	NE-CZ-NH1	7.64	124.12	120.30	16	7
1	A	149	ARG	NE-CZ-NH2	-7.42	116.59	120.30	13	5
1	A	45	ARG	NE-CZ-NH2	-6.64	116.98	120.30	2	14
1	A	45	ARG	NE-CZ-NH1	6.45	123.53	120.30	2	16
1	A	149	ARG	NE-CZ-NH1	6.38	123.49	120.30	13	5
1	A	48	ARG	NE-CZ-NH1	6.37	123.48	120.30	6	4
1	A	101	ARG	NE-CZ-NH1	6.12	123.36	120.30	4	2
1	A	106	ARG	NE-CZ-NH1	6.00	123.30	120.30	3	5
1	A	138	ARG	NE-CZ-NH1	5.96	123.28	120.30	15	4
1	A	157	LEU	CB-CA-C	5.25	120.17	110.20	16	5
1	A	48	ARG	NE-CZ-NH2	-5.15	117.72	120.30	3	1

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	48	ARG	Sidechain	1

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
All	All	19660	20000	20000	-

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	121/139 (87%)	118±2 (97±2%)	3±2 (2±1%)	0±1 (0±1%)	44	80
All	All	2420/2780 (87%)	2359 (97%)	53 (2%)	8 (0%)	44	80

All 6 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	73	ASN	3
1	A	110	PHE	1
1	A	109	PRO	1
1	A	74	GLU	1
1	A	90	CYS	1
1	A	111	LYS	1

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	115/132 (87%)	112±1 (97±1%)	3±1 (3±1%)	50	91
All	All	2300/2640 (87%)	2242 (97%)	58 (3%)	50	91

All 23 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	33	MET	9
1	A	43	LEU	9
1	A	157	LEU	7
1	A	112	HIS	4
1	A	68	GLU	4
1	A	125	THR	3
1	A	66	CYS	3
1	A	133	ILE	2
1	A	138	ARG	2
1	A	116	ILE	2
1	A	34	GLU	1
1	A	108	LEU	1
1	A	130	ASN	1
1	A	98	LEU	1
1	A	136	LYS	1
1	A	143	MET	1
1	A	95	LEU	1
1	A	126	GLU	1
1	A	86	THR	1
1	A	156	VAL	1
1	A	64	GLU	1
1	A	118	ILE	1
1	A	114	LEU	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 89% 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: *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	1735
Number of shifts mapped to atoms	1735
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

7.1.2　Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	138	-0.88 ± 0.10	Should be checked
$^{13}C_\beta$	136	0.04 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	126	-0.68 ± 0.38	None needed (imprecise)
^{15}N	121	0.91 ± 0.33	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. 1539 atoms were assigned a chemical shift out of a possible 1730. 0 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	574/594 (97%)	232/237 (98%)	233/242 (96%)	109/115 (95%)
Sidechain	953/1075 (89%)	647/696 (93%)	293/341 (86%)	13/38 (34%)

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	12/61 (20%)	12/30 (40%)	0/29 (0%)	0/2 (0%)
Overall	1539/1730 (89%)	891/963 (93%)	526/612 (86%)	122/155 (79%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 87%, i.e. 1735 atoms were assigned a chemical shift out of a possible 1984. 0 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	648/681 (95%)	263/272 (97%)	264/278 (95%)	121/131 (92%)
Sidechain	1074/1221 (88%)	728/789 (92%)	331/386 (86%)	15/46 (33%)
Aromatic	13/82 (16%)	13/42 (31%)	0/35 (0%)	0/5 (0%)
Overall	1735/1984 (87%)	1004/1103 (91%)	595/699 (85%)	136/182 (75%)

7.1.4 Statistically unusual chemical shifts ⓘ

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	36	LYS	HD3	0.42	0.54 − 2.65	-5.6
1	A	29	GLN	CD	173.04	173.59 − 185.85	-5.5
1	A	36	LYS	HD2	0.50	0.58 − 2.64	-5.4
1	A	36	LYS	HE3	1.86	1.92 − 3.89	-5.3
1	A	105	GLN	CD	173.26	173.59 − 185.85	-5.3
1	A	154	GLN	CD	173.26	173.59 − 185.85	-5.3
1	A	36	LYS	HE2	1.92	1.95 − 3.88	-5.1

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: