Full wwPDB NMR Structure Validation Report (i)

Apr 21, 2024 – 07:53 AM EDT

PDB ID	:	2L6N
BMRB ID	:	17317
Title	:	NMR solution structure of the protein YP_001092504.1
Authors	:	Mohanty, B.; Serrano, P.; Geralt, M.; Horst, R.; Wuthrich, K.; Joint Center for Structural Genomics (JCSG)
Deposited on	:	2010-11-23

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:

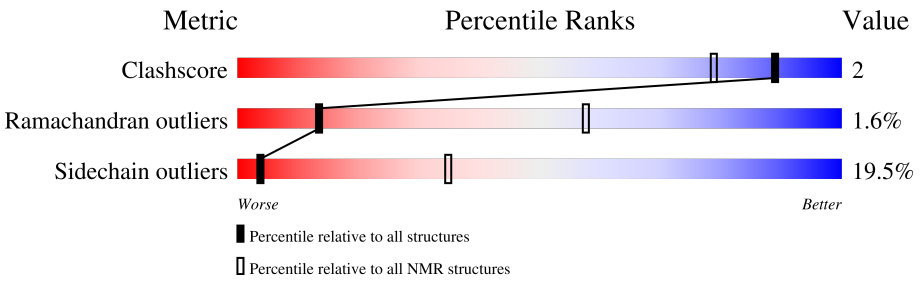
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment was not calculated.

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	132	67% 14% 18%

2 Ensemble composition and analysis (i)

This entry contains 20 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:6-A:46, A:56-A:122 (108)	0.81	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 1 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1929 atoms, of which 885 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called uncharacterized protein YP_001092504.1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	132	1929	662	885	185	194	3	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP A3Q9U7

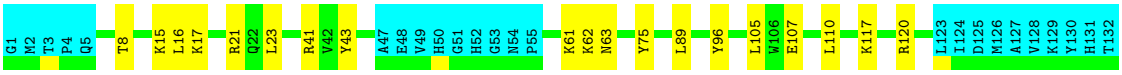
4 Residue-property plots (i)

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: uncharacterized protein YP_001092504.1

Chain A: 67% 14% 18%



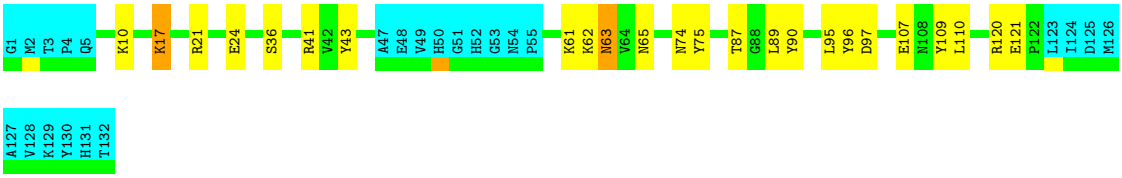
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: uncharacterized protein YP_001092504.1

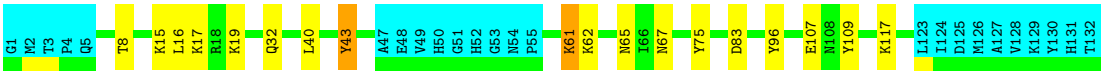
Chain A: 64% 17% • 18%



4.2.2 Score per residue for model 2

• Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 68% 12% • 18%



4.2.3 Score per residue for model 3

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 61% 20% • 18%

G1 M2 T3 P4 Q5 S6 D7 T8 L16 K17 R21 Q22 L23 S26 N29 Q32 E38 R41 S44 A47 E48 V49 H50 G51 H52 G53 N54 P55 V56 L57 K61 K62 N63 G88 L89 K93 Y96 D97 N108 Y109 L110 R114 K117

R120 L123 I124 D125 M126 A127 V128 K129 Y130 H131 T132

4.2.4 Score per residue for model 4

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 61% 17% • • 18%

G1 M2 T3 P4 Q5 T8 L14 K15 L16 K17 R18 R21 Q22 L23 E24 D28 N29 R41 V42 Y43 A47 E48 V49 H50 G51 H52 G53 N54 P55 T59 H60 K61 K62 N65 D82 D83 G84 H85 L89 Y96 N108 Y109 L113 R120 L123

I124 D125 M126 A127 V128 K129 Y130 H131 T132

4.2.5 Score per residue for model 5

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 64% 17% • 18%

G1 M2 T3 P4 Q5 K10 L16 K17 R18 R21 N29 L40 Y43 A47 E48 V49 H50 G51 H52 G53 N54 P55 K62 I66 N67 Y75 A76 V77 K78 D83 D86 T87 G88 L89 Y90 K93 Y96 E107 N108 L123 I124 D125 M126 A127

V128 K129 Y130 H131 T132

4.2.6 Score per residue for model 6

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 66% 14% • 18%

G1 M2 T3 P4 Q5 S6 D7 T8 K15 L16 K17 L23 R41 V42 Y43 A47 E48 V49 H50 G51 H52 G53 N54 P55 K61 K62 N63 G73 N74 Y75 H85 G88 L89 L95 Y96 D97 S100 E107 L123 I124 D125 M126 A127 V128 K129 Y130 H131 T132

4.2.7 Score per residue for model 7

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 63% 17% • 18%

G1 M2 T3 P4 Q5 T8 P9 K10 K15 L16 K17 R21 Y43 A47 E48 V49 H50 G51 H52 G53 N54 P55 K62 N63 I69 Y75 A76 V77 K78 H85 D86 Y90 Y96 S100 E107 L110 A111 R112 L113 R114 K117 A118 S119 R120

L123 I124 D125 M126 A127 V128 K129 Y130 H131 T132

4.2.8 Score per residue for model 8

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 62% 18% • 18%

G1 M2 T3 P4 Q5 S6 D7 T8 P9 K10 K15 L16 K17 L23 R41 V42 Y43 A47 E48 V49 H50 G51 H52 G53 N54 P55 V56 L57 K62 N63 I66 I69 T70 N74 V77 H85 Y96 E107 N108 R112 L113 R114 S119 R120 L123

I124 D125 M126 A127 V128 K129 Y130 H131 T132

4.2.9 Score per residue for model 9

- Molecule 1: uncharacterized protein YP_001092504.1

Chain A: 64% 14% • • 18%

G1 M2 T3 P4 Q5 S6 D7 T8 P9 K10 K15 L16 K17 R18 R21 Q22 L23 T34 L40 R41 S44 A47 E48 V49 H50 G51 H52 G53 N54 P55 K62 N63 I69 Y75 A76 V77 Y96 D97 S100 L110 K117 R120 L123 I124 D125

M126 A127 V128 K129 Y130 H131 T132

4.2.10 Score per residue for model 10

- Molecule 1: uncharacterized protein YP_001092504.1

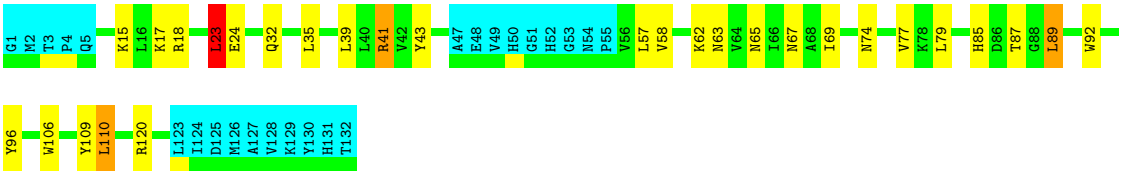
Chain A: 65% 14% • 18%

G1 M2 T3 P4 Q5 S6 D7 K10 L16 K17 S20 L23 S26 Q32 L40 R41 V42 Y43 A47 E48 V49 H50 G51 H52 G53 N54 P55 H60 Y75 D83 G84 H85 L89 Y96 N108 K117 R120 E121 P122 L123 I124 D125 M126 A127

4.2.11 Score per residue for model 11

• Molecule 1: uncharacterized protein YP_001092504.1

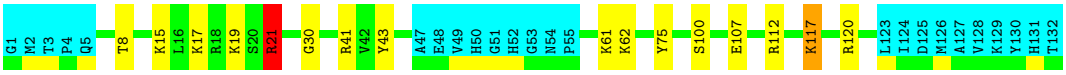
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: uncharacterized protein YP_001092504.1

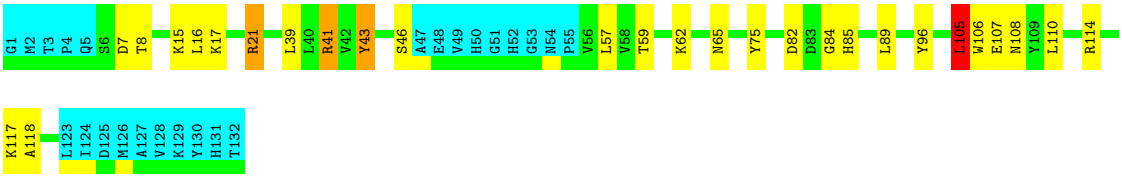
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: uncharacterized protein YP_001092504.1

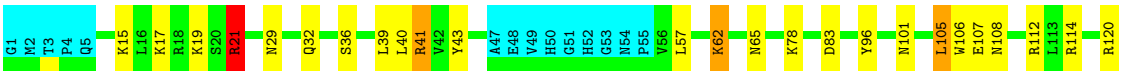
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: uncharacterized protein YP_001092504.1

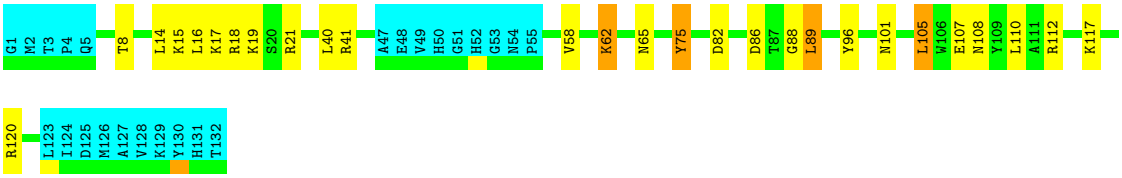
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: uncharacterized protein YP_001092504.1

Chain A:



4.2.16 Score per residue for model 16

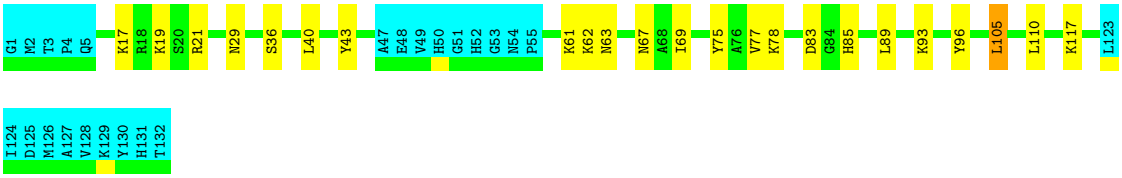
• Molecule 1: uncharacterized protein YP_001092504.1

Chain A:

4.2.17 Score per residue for model 17

• Molecule 1: uncharacterized protein YP_001092504.1

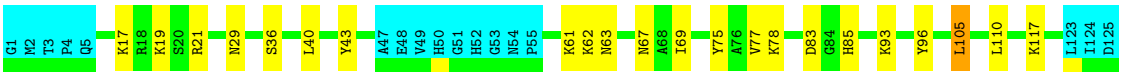
Chain A:

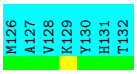


4.2.18 Score per residue for model 18

• Molecule 1: uncharacterized protein YP_001092504.1

Chain A:

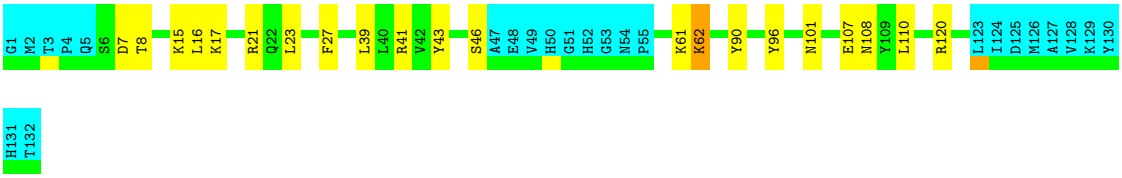




4.2.19 Score per residue for model 19

• Molecule 1: uncharacterized protein YP_001092504.1

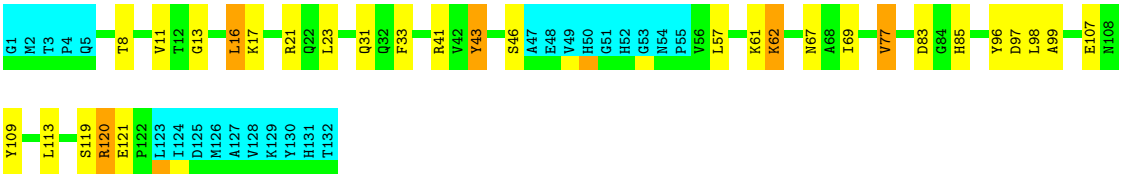
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: uncharacterized protein YP_001092504.1

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 80 calculated structures, 20 were deposited, based on the following criterion: *CYANA target function*.

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

Software name	Classification	Version
CYANA	structure calculation	3.0
UNIO	structure solution	2.0.1
UNIO	refinement	2.0.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	0.61±0.01	0±0/881 (0.0± 0.0%)	1.09±0.04	1±1/1196 (0.1± 0.1%)
All	All	0.61	0/17620 (0.0%)	1.10	26/23920 (0.1%)

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	1.6±0.9
All	All	0	33

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	41	ARG	NE-CZ-NH2	-6.67	116.96	120.30	14	1
1	A	105	LEU	CB-CG-CD2	6.45	121.97	111.00	17	2
1	A	110	LEU	CB-CG-CD1	6.35	121.79	111.00	9	5
1	A	96	TYR	CB-CG-CD2	-6.04	117.38	121.00	15	2
1	A	121	GLU	CB-CA-C	6.00	122.40	110.40	14	1
1	A	120	ARG	NE-CZ-NH2	-5.73	117.43	120.30	20	1
1	A	21	ARG	NE-CZ-NH2	-5.66	117.47	120.30	12	2
1	A	109	TYR	CB-CG-CD2	-5.56	117.67	121.00	4	1
1	A	59	THR	CA-CB-CG2	-5.54	104.65	112.40	4	1
1	A	23	LEU	CB-CG-CD2	5.47	120.31	111.00	3	3
1	A	77	VAL	CA-CB-CG1	5.39	118.98	110.90	20	1
1	A	75	TYR	CB-CG-CD2	-5.28	117.83	121.00	9	2
1	A	41	ARG	NE-CZ-NH1	5.19	122.90	120.30	13	1
1	A	43	TYR	CB-CG-CD1	-5.19	117.89	121.00	20	1
1	A	105	LEU	CB-CA-C	5.17	120.03	110.20	13	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	121	GLU	CB-CG-CD	-5.12	100.39	114.20	14	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	21	ARG	Sidechain	6
1	A	109	TYR	Sidechain	5
1	A	43	TYR	Sidechain	5
1	A	90	TYR	Sidechain	3
1	A	18	ARG	Sidechain	2
1	A	114	ARG	Sidechain	2
1	A	120	ARG	Sidechain	2
1	A	101	ASN	Peptide	2
1	A	60	HIS	Sidechain,Peptide	1
1	A	41	ARG	Sidechain	1
1	A	85	HIS	Peptide	1
1	A	112	ARG	Sidechain	1
1	A	27	PHE	Sidechain	1

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	863	739	874	3±2
All	All	17260	14780	17480	62

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:59:THR:HG22	1:A:118:ALA:HB1	0.69	1.62	13	1
1:A:39:LEU:HD11	1:A:105:LEU:HD11	0.65	1.66	13	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:110:LEU:H	1:A:110:LEU:HD22	0.59	1.58	1	3
1:A:16:LEU:HD12	1:A:23:LEU:HD23	0.56	1.77	9	2
1:A:33:PHE:CE2	1:A:99:ALA:HB2	0.55	2.37	20	1
1:A:69:ILE:HG23	1:A:77:VAL:HG11	0.54	1.77	7	3
1:A:62:LYS:H	1:A:62:LYS:HE2	0.54	1.61	9	1
1:A:14:LEU:HD23	1:A:23:LEU:HD21	0.53	1.81	16	1
1:A:16:LEU:HD11	1:A:23:LEU:HD12	0.53	1.80	10	2
1:A:39:LEU:HD11	1:A:105:LEU:CD1	0.53	2.34	16	2
1:A:113:LEU:HD12	1:A:118:ALA:HB3	0.50	1.82	7	1
1:A:105:LEU:HD12	1:A:106:TRP:N	0.50	2.22	14	3
1:A:69:ILE:CG2	1:A:77:VAL:HG21	0.49	2.37	9	3
1:A:113:LEU:CD1	1:A:118:ALA:HB3	0.49	2.37	7	1
1:A:105:LEU:HD12	1:A:105:LEU:C	0.49	2.28	13	2
1:A:106:TRP:CZ2	1:A:110:LEU:HD21	0.48	2.43	13	1
1:A:23:LEU:HD22	1:A:24:GLU:N	0.47	2.24	11	3
1:A:35:LEU:HD22	1:A:39:LEU:HD23	0.46	1.88	11	1
1:A:60:HIS:CG	1:A:61:LYS:N	0.46	2.83	4	1
1:A:69:ILE:HG23	1:A:77:VAL:CG1	0.45	2.40	7	2
1:A:110:LEU:HD12	1:A:110:LEU:H	0.45	1.71	19	2
1:A:17:LYS:HE3	1:A:24:GLU:CD	0.45	2.32	1	1
1:A:16:LEU:HD22	1:A:23:LEU:HD23	0.45	1.88	4	1
1:A:62:LYS:H	1:A:62:LYS:CE	0.45	2.24	9	2
1:A:95:LEU:HD12	1:A:95:LEU:H	0.44	1.72	1	2
1:A:113:LEU:HD21	1:A:119:SER:O	0.44	2.12	20	1
1:A:90:TYR:HB3	1:A:95:LEU:HD11	0.43	1.90	1	1
1:A:106:TRP:CE2	1:A:110:LEU:HD11	0.43	2.48	11	1
1:A:110:LEU:HD12	1:A:110:LEU:N	0.43	2.28	19	2
1:A:106:TRP:CE2	1:A:110:LEU:HD21	0.43	2.49	16	1
1:A:21:ARG:NE	1:A:65:ASN:HD22	0.43	2.11	14	1
1:A:77:VAL:HG13	1:A:79:LEU:HD13	0.43	1.91	11	1
1:A:110:LEU:H	1:A:110:LEU:CD2	0.42	2.26	1	1
1:A:101:ASN:HB3	1:A:105:LEU:HD23	0.42	1.92	15	1
1:A:17:LYS:HE2	1:A:34:THR:HG21	0.42	1.91	9	1
1:A:43:TYR:CD1	1:A:98:LEU:HD21	0.42	2.49	20	1
1:A:16:LEU:HD21	1:A:66:ILE:HB	0.42	1.91	8	1
1:A:110:LEU:HD22	1:A:110:LEU:N	0.41	2.28	1	1
1:A:77:VAL:HB	1:A:92:TRP:CH2	0.41	2.51	11	1
1:A:113:LEU:C	1:A:113:LEU:HD13	0.40	2.36	8	1
1:A:106:TRP:O	1:A:110:LEU:HD13	0.40	2.17	11	1
1:A:61:LYS:HA	1:A:61:LYS:HE3	0.40	1.93	2	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	108/132 (82%)	95±5 (88±4%)	12±4 (11±4%)	2±1 (2±1%)	13 57
All	All	2160/2640 (82%)	1892 (88%)	233 (11%)	35 (2%)	13 57

All 19 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	62	LYS	5
1	A	88	GLY	4
1	A	89	LEU	4
1	A	117	LYS	3
1	A	63	ASN	2
1	A	14	LEU	2
1	A	74	ASN	2
1	A	83	ASP	2
1	A	44	SER	1
1	A	66	ILE	1
1	A	73	GLY	1
1	A	60	HIS	1
1	A	87	THR	1
1	A	30	GLY	1
1	A	84	GLY	1
1	A	90	TYR	1
1	A	11	VAL	1
1	A	13	GLY	1
1	A	85	HIS	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	96/115 (83%)	77±2 (81±2%)	19±2 (19±2%)	4 35
All	All	1920/2300 (83%)	1546 (81%)	374 (19%)	4 35

All 59 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	17	LYS	20
1	A	62	LYS	19
1	A	96	TYR	18
1	A	43	TYR	16
1	A	41	ARG	14
1	A	107	GLU	13
1	A	15	LYS	13
1	A	75	TYR	12
1	A	120	ARG	12
1	A	8	THR	12
1	A	21	ARG	12
1	A	63	ASN	10
1	A	89	LEU	10
1	A	16	LEU	10
1	A	108	ASN	10
1	A	61	LYS	9
1	A	117	LYS	9
1	A	23	LEU	9
1	A	40	LEU	8
1	A	85	HIS	8
1	A	7	ASP	7
1	A	29	ASN	7
1	A	57	LEU	7
1	A	10	LYS	6
1	A	65	ASN	6
1	A	19	LYS	6
1	A	67	ASN	6
1	A	83	ASP	6
1	A	105	LEU	6
1	A	97	ASP	5
1	A	32	GLN	5
1	A	78	LYS	5
1	A	36	SER	4
1	A	121	GLU	4
1	A	93	LYS	4
1	A	114	ARG	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	100	SER	4
1	A	112	ARG	4
1	A	18	ARG	3
1	A	82	ASP	3
1	A	86	ASP	3
1	A	46	SER	3
1	A	74	ASN	2
1	A	87	THR	2
1	A	26	SER	2
1	A	28	ASP	2
1	A	58	VAL	2
1	A	38	GLU	1
1	A	59	THR	1
1	A	113	LEU	1
1	A	77	VAL	1
1	A	70	THR	1
1	A	119	SER	1
1	A	44	SER	1
1	A	6	SER	1
1	A	20	SER	1
1	A	14	LEU	1
1	A	39	LEU	1
1	A	31	GLN	1

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)

No chemical shift data were provided