Full wwPDB NMR Structure Validation Report (i)

Mar 1, 2022 – 10:11 AM EST

PDB ID	:	2G1E
Title	:	Solution Structure of TA0895
Authors	:	Yeo, I.Y.; Hong, E.; Jung, J.; Lee, W.; Yee, A.; Arrowsmith, C.H.
Deposited on	:	2006-02-14

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

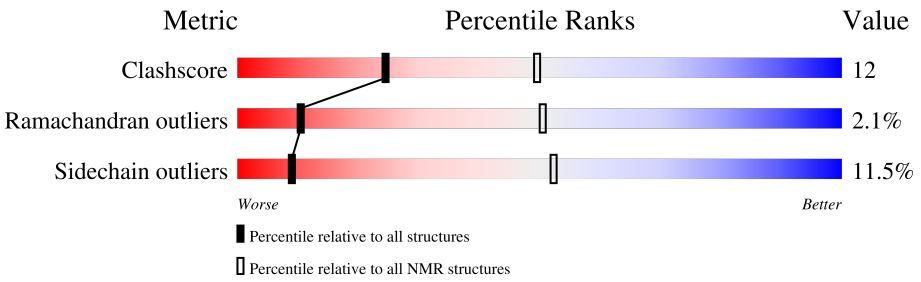
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	90	

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:1-A:86 (86)	0.24	8

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1452 atoms, of which 735 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called hypothetical protein Ta0895.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	90	1452	458	735	117	139	3	0

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: hypothetical protein Ta0895

Chain A: 66% 29%

M1 V2 A8 T9 L10 R11 P12 I13 T14 T20 I24 L30 L31 E32 R33 V36 E37 Y38 L52 F53 V56 I57 I58 L59 V60 I65 M68 K69 T73 E74 I75 D79 K80 I81 D82 L83 V87 A88 G89 G90

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: hypothetical protein Ta0895

Chain A: 62% 31%

M1 Y6 Y7 A8 T9 L10 R11 P12 T20 F21 N22 G23 I24 I27 S28 E29 L30 L31 E32 K35 V36 E37 Y38 F53 K54 N55 V56 I57 I58 L59 V60 I65 M68 K69 G70 L71 D72 T73 E74 I75 I81 D82 L83 V87 A88 G89 G90

4.2.2 Score per residue for model 2

• Molecule 1: hypothetical protein Ta0895

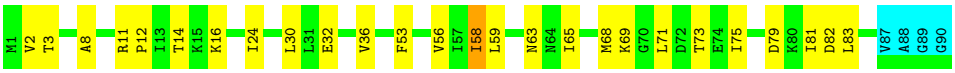
Chain A: 70% 23%

M1 A8 T9 L10 R11 P12 I13 T14 K15 K16 T20 I24 L30 L31 E32 R33 V36 E37 Y38 F53 V56 I57 I58 M68 K69 D78 D79 K80 I81 D82 L83 F84 P85 P86 V87 A88 G89 G90

4.2.3 Score per residue for model 3

- Molecule 1: hypothetical protein Ta0895

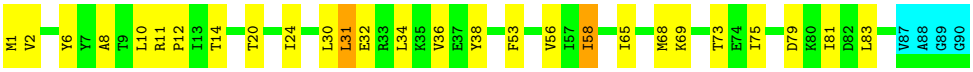
Chain A: 67% 28%



4.2.4 Score per residue for model 4

- Molecule 1: hypothetical protein Ta0895

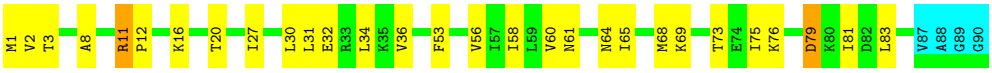
Chain A: 66% 28%



4.2.5 Score per residue for model 5

- Molecule 1: hypothetical protein Ta0895

Chain A: 63% 30%



4.2.6 Score per residue for model 6

- Molecule 1: hypothetical protein Ta0895

Chain A: 63% 29%

4.2.7 Score per residue for model 7

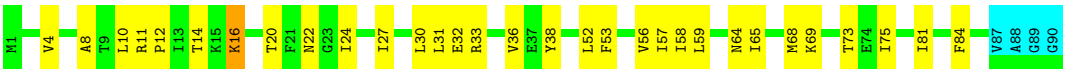
- Molecule 1: hypothetical protein Ta0895

Chain A: 69% 24%

4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: hypothetical protein Ta0895

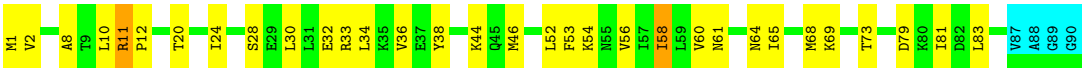
Chain A: 61% 33%



4.2.9 Score per residue for model 9

- Molecule 1: hypothetical protein Ta0895

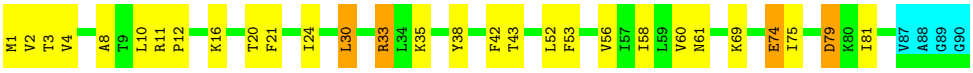
Chain A: 60% 33%



4.2.10 Score per residue for model 10

- Molecule 1: hypothetical protein Ta0895

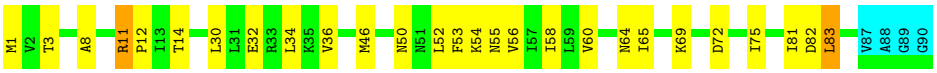
Chain A: 63% 28%



4.2.11 Score per residue for model 11

- Molecule 1: hypothetical protein Ta0895

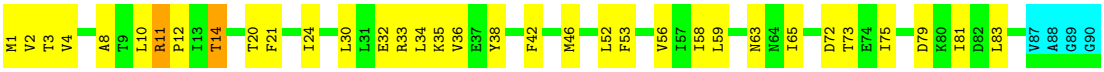
Chain A: 66% 28%



4.2.12 Score per residue for model 12

- Molecule 1: hypothetical protein Ta0895

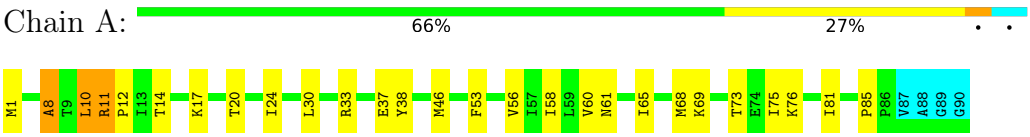
Chain A: 58% 36%



4.2.13 Score per residue for model 13

- Molecule 1: hypothetical protein Ta0895

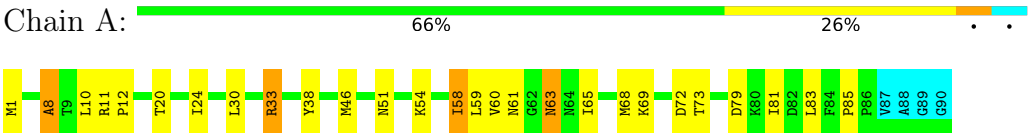
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: hypothetical protein Ta0895

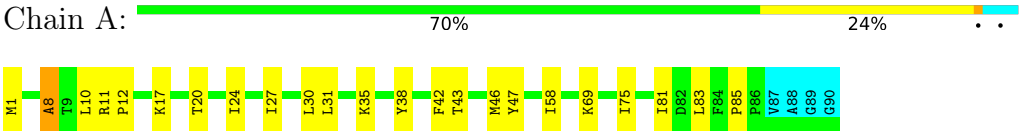
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: hypothetical protein Ta0895

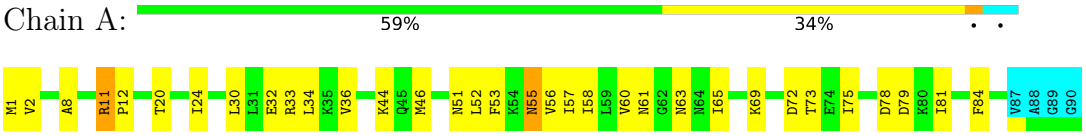
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: hypothetical protein Ta0895

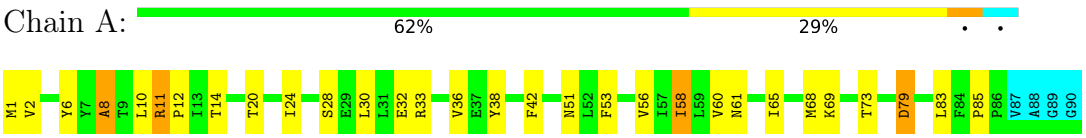
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: hypothetical protein Ta0895

Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: hypothetical protein Ta0895

Chain A: 66% 30%

M1 A8 T9 L10 R11 P12 I13 T14 K15 K16 T20 I24 S28 E29 L30 L31 E32 R33 L34 K35 V36 E37 Y38 L52 F53 V56 I57 I58 L59 V60 N64 I65 K69 I75 I81 V87 A88 G89 G90

4.2.19 Score per residue for model 19

- Molecule 1: hypothetical protein Ta0895

Chain A: 57% 37%

M1 V2 T3 V4 A8 T9 L10 R11 P12 I13 T14 K15 K16 E19 T20 F21 I24 I27 L30 L31 E37 Y38 D48 F53 V56 I57 I58 L59 V60 N63 N64 I65 K69 G70 L71 I75 D78 D79 K80 I81 D82 L83 F84 P85 P86 V87 A88

G89 G90

4.2.20 Score per residue for model 20

- Molecule 1: hypothetical protein Ta0895

Chain A: 61% 32%

M1 V2 T3 V4 R5 A8 T9 L10 R11 P12 I24 S25 K26 L30 L31 E32 R33 V36 E37 Y38 M46 L52 F53 I57 I58 L59 V60 N61 G62 N63 K69 G70 L71 I75 D79 K80 I81 F84 V87 A88 G89 G90

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 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
CYANA	structure solution	2.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	697	715	717	17±4
All	All	13940	14300	14340	333

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 12.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:65:ILE:HG22	1:A:73:THR:HG21	0.78	1.56	8	12
1:A:10:LEU:HD13	1:A:83:LEU:HD13	0.75	1.57	19	2
1:A:10:LEU:HD11	1:A:38:TYR:CE2	0.71	2.21	19	12
1:A:10:LEU:HD21	1:A:38:TYR:CZ	0.69	2.23	18	9
1:A:24:ILE:HD13	1:A:30:LEU:HD22	0.68	1.64	7	7
1:A:27:ILE:HD11	1:A:60:VAL:HG22	0.66	1.67	1	1
1:A:34:LEU:HD21	1:A:83:LEU:HD11	0.65	1.67	12	1
1:A:53:PHE:O	1:A:56:VAL:HG12	0.65	1.92	4	17
1:A:65:ILE:CG2	1:A:73:THR:HG21	0.64	2.23	13	12
1:A:24:ILE:HD13	1:A:30:LEU:CD2	0.64	2.23	6	11
1:A:10:LEU:HD11	1:A:38:TYR:CD2	0.63	2.28	4	2
1:A:75:ILE:CD1	1:A:81:ILE:HD11	0.63	2.24	5	15
1:A:2:VAL:HG23	1:A:79:ASP:C	0.62	2.15	4	2
1:A:10:LEU:HD12	1:A:83:LEU:HD22	0.61	1.73	1	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:27:ILE:HD11	1:A:60:VAL:HG12	0.60	1.72	19	1
1:A:10:LEU:CD1	1:A:83:LEU:HD13	0.60	2.26	15	2
1:A:30:LEU:CD1	1:A:81:ILE:HG21	0.60	2.26	12	1
1:A:27:ILE:HD11	1:A:60:VAL:CG1	0.60	2.27	19	1
1:A:24:ILE:HD11	1:A:33:ARG:CD	0.59	2.28	14	2
1:A:4:VAL:HG11	1:A:21:PHE:CE2	0.59	2.32	12	3
1:A:24:ILE:HD11	1:A:33:ARG:HD3	0.59	1.75	14	2
1:A:2:VAL:HG12	1:A:79:ASP:HB3	0.58	1.76	17	10
1:A:6:TYR:CD1	1:A:10:LEU:HD23	0.58	2.33	4	2
1:A:30:LEU:HG	1:A:81:ILE:HD13	0.58	1.74	4	17
1:A:60:VAL:CG1	1:A:65:ILE:HG22	0.58	2.29	11	2
1:A:24:ILE:HD11	1:A:33:ARG:HG2	0.57	1.75	10	2
1:A:57:ILE:HG23	1:A:84:PHE:HB2	0.57	1.74	20	3
1:A:4:VAL:HG22	1:A:81:ILE:HB	0.57	1.75	20	1
1:A:60:VAL:HG12	1:A:65:ILE:HG22	0.56	1.78	11	2
1:A:30:LEU:HD11	1:A:81:ILE:HG21	0.56	1.77	12	1
1:A:31:LEU:HD12	1:A:53:PHE:CE1	0.56	2.35	19	1
1:A:24:ILE:HD13	1:A:30:LEU:HD23	0.55	1.77	15	9
1:A:60:VAL:HG22	1:A:61:ASN:ND2	0.55	2.16	14	3
1:A:10:LEU:O	1:A:10:LEU:HD23	0.55	2.01	15	4
1:A:60:VAL:HG22	1:A:61:ASN:OD1	0.54	2.02	16	6
1:A:30:LEU:HD12	1:A:81:ILE:HD13	0.54	1.78	12	1
1:A:14:THR:HG22	1:A:16:LYS:HE2	0.54	1.80	8	2
1:A:32:GLU:O	1:A:36:VAL:HG23	0.54	2.03	1	14
1:A:75:ILE:HD11	1:A:81:ILE:HD11	0.53	1.78	1	9
1:A:10:LEU:HD23	1:A:10:LEU:O	0.53	2.03	6	4
1:A:24:ILE:HG21	1:A:30:LEU:HD23	0.53	1.81	6	6
1:A:34:LEU:CD2	1:A:83:LEU:HD11	0.53	2.34	4	2
1:A:27:ILE:HD11	1:A:60:VAL:CG2	0.52	2.34	1	1
1:A:14:THR:HG21	1:A:38:TYR:CE1	0.52	2.39	12	2
1:A:57:ILE:HG23	1:A:84:PHE:CB	0.51	2.35	20	1
1:A:14:THR:HG22	1:A:16:LYS:CE	0.51	2.35	8	1
1:A:8:ALA:HB2	1:A:85:PRO:N	0.51	2.20	19	6
1:A:11:ARG:CB	1:A:12:PRO:CD	0.51	2.89	12	20
1:A:10:LEU:CD2	1:A:83:LEU:HD13	0.51	2.35	2	1
1:A:60:VAL:HG22	1:A:61:ASN:HD22	0.50	1.67	14	3
1:A:30:LEU:CD1	1:A:81:ILE:HD13	0.50	2.37	12	1
1:A:58:ILE:HG22	1:A:83:LEU:HD13	0.50	1.83	3	5
1:A:10:LEU:HD21	1:A:38:TYR:CE2	0.49	2.42	4	4
1:A:46:MET:HB3	1:A:52:LEU:HD22	0.49	1.83	6	1
1:A:24:ILE:HD11	1:A:33:ARG:NE	0.49	2.23	13	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:75:ILE:HD13	1:A:81:ILE:HD11	0.48	1.83	6	3
1:A:24:ILE:HD11	1:A:33:ARG:HE	0.48	1.69	13	1
1:A:4:VAL:HG21	1:A:21:PHE:CE2	0.48	2.44	10	1
1:A:30:LEU:CD2	1:A:34:LEU:HD22	0.48	2.38	12	1
1:A:10:LEU:HD11	1:A:38:TYR:HE2	0.47	1.70	20	4
1:A:46:MET:O	1:A:52:LEU:HD13	0.46	2.10	6	1
1:A:34:LEU:HD11	1:A:83:LEU:HD11	0.46	1.87	11	1
1:A:59:LEU:HD23	1:A:63:ASN:C	0.45	2.32	12	4
1:A:24:ILE:HG23	1:A:29:GLU:HG3	0.45	1.86	1	1
1:A:27:ILE:O	1:A:31:LEU:HD23	0.45	2.12	8	2
1:A:52:LEU:HD13	1:A:53:PHE:N	0.45	2.26	20	1
1:A:27:ILE:HG22	1:A:31:LEU:HD11	0.44	1.90	5	1
1:A:6:TYR:HD1	1:A:10:LEU:HD23	0.44	1.70	4	1
1:A:52:LEU:C	1:A:52:LEU:HD13	0.44	2.33	12	7
1:A:34:LEU:CD1	1:A:83:LEU:HD11	0.44	2.43	11	2
1:A:24:ILE:HD11	1:A:33:ARG:CG	0.44	2.41	10	1
1:A:4:VAL:HG12	1:A:81:ILE:HB	0.44	1.90	8	1
1:A:35:LYS:HE2	1:A:43:THR:HG21	0.44	1.89	10	2
1:A:4:VAL:HG12	1:A:19:GLU:O	0.43	2.13	19	1
1:A:59:LEU:HD23	1:A:64:ASN:N	0.43	2.28	8	3
1:A:60:VAL:HG13	1:A:65:ILE:CG2	0.43	2.44	19	1
1:A:30:LEU:HD13	1:A:30:LEU:C	0.43	2.34	1	6
1:A:30:LEU:CD1	1:A:34:LEU:HD22	0.43	2.44	4	2
1:A:30:LEU:C	1:A:30:LEU:HD13	0.42	2.35	8	4
1:A:10:LEU:HD12	1:A:42:PHE:CZ	0.42	2.49	12	1
1:A:6:TYR:CD1	1:A:10:LEU:HD13	0.42	2.50	1	1
1:A:10:LEU:HD13	1:A:42:PHE:CE1	0.42	2.49	17	1
1:A:30:LEU:HD13	1:A:34:LEU:HG	0.42	1.92	9	2
1:A:31:LEU:HD11	1:A:58:ILE:HD12	0.41	1.90	4	1
1:A:46:MET:O	1:A:52:LEU:HD23	0.41	2.14	9	1
1:A:31:LEU:HD21	1:A:53:PHE:CE1	0.41	2.50	18	1
1:A:10:LEU:HD12	1:A:42:PHE:CE1	0.41	2.49	12	3
1:A:30:LEU:C	1:A:30:LEU:HD23	0.41	2.36	12	1
1:A:30:LEU:HD21	1:A:34:LEU:HD22	0.41	1.93	12	1
1:A:60:VAL:HG11	1:A:74:GLU:O	0.41	2.16	1	1
1:A:60:VAL:HG21	1:A:74:GLU:O	0.40	2.16	10	1
1:A:31:LEU:HD13	1:A:53:PHE:HE2	0.40	1.76	4	1
1:A:11:ARG:N	1:A:12:PRO:HD2	0.40	2.32	14	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	85/90 (94%)	74±1 (88±1%)	9±2 (10±2%)	2±1 (2±1%)	10 50
All	All	1700/1800 (94%)	1489 (88%)	176 (10%)	35 (2%)	10 50

All 5 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	8	ALA	20
1	A	68	MET	12
1	A	50	ASN	1
1	A	79	ASP	1
1	A	47	TYR	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	80/81 (99%)	71±2 (88±3%)	9±2 (11±3%)	9 52
All	All	1600/1620 (99%)	1416 (88%)	184 (12%)	9 52

All 39 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	58	ILE	20
1	A	69	LYS	18
1	A	20	THR	17
1	A	1	MET	15
1	A	11	ARG	11

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	14	THR	9
1	A	33	ARG	8
1	A	3	THR	7
1	A	46	MET	7
1	A	16	LYS	6
1	A	28	SER	5
1	A	54	LYS	4
1	A	71	LEU	4
1	A	79	ASP	4
1	A	72	ASP	4
1	A	83	LEU	3
1	A	78	ASP	3
1	A	82	ASP	3
1	A	64	ASN	3
1	A	37	GLU	3
1	A	51	ASN	3
1	A	63	ASN	3
1	A	22	ASN	2
1	A	35	LYS	2
1	A	80	LYS	2
1	A	76	LYS	2
1	A	44	LYS	2
1	A	55	ASN	2
1	A	17	LYS	2
1	A	68	MET	1
1	A	31	LEU	1
1	A	52	LEU	1
1	A	67	SER	1
1	A	30	LEU	1
1	A	74	GLU	1
1	A	10	LEU	1
1	A	48	ASP	1
1	A	5	ARG	1
1	A	26	LYS	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