Full wwPDB NMR Structure Validation Report (i)

Feb 23, 2022 – 02:09 PM EST

PDB ID	:	1YGO
Title	:	Solution Structure of the pelle Death Domain
Authors	:	Moncrieffe, M.C.; Stott, K.M.; Gay, N.J.
Deposited on	:	2005-01-05

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:

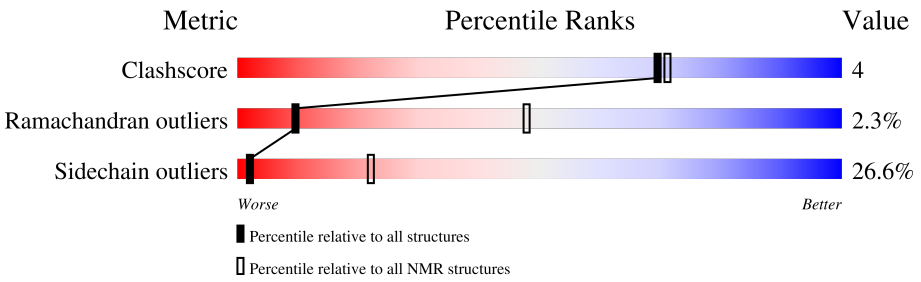
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	110	63% 14% • 21%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:11-A:56, A:60-A:100 (87)	0.40	11

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 3 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1805 atoms, of which 911 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Probable serine/threonine-protein kinase pelle.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	110	1805	569	911	163	158	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP Q05652
A	2	SER	-	cloning artifact	UNP Q05652
A	3	HIS	-	cloning artifact	UNP Q05652
A	4	MET	-	cloning artifact	UNP Q05652

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: Probable serine/threonine-protein kinase pelle

Chain A: 63% 14% • 21%

G1 S2 H3 M4 S5 H6 L7 D8 N9 T10 I13 R14 L15 L16 P17 L18 P19 V20 R21 A22 Q23 L24 Q36 K42 Q56 R57 G58 R59 H75 L95 I96 K97 D98 Y99 V100 S101 E102 D103 L104 H105 K106 Y107 I108 P109 R110

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: Probable serine/threonine-protein kinase pelle

Chain A: 54% 23% • 21%

G1 S2 H3 M4 S5 H6 L7 D8 N9 T10 R14 L15 L16 P17 L18 P19 Q23 L31 W34 Q35 Q36 L37 A40 V41 K42 Q47 V48 E49 Q54 R57 G58 R59 S62 N63 E64 G71 N74 H75 Q78 M93 R94 L95 I96 K97 D98 S101

E102 D103 L104 H105 K106 Y107 I108 P109 R110

4.2.2 Score per residue for model 2

• Molecule 1: Probable serine/threonine-protein kinase pelle

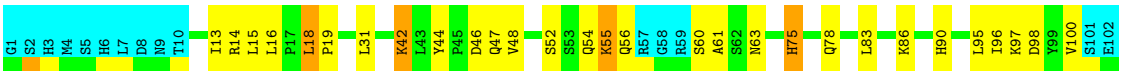
Chain A: 58% 19% • 21%

G1 S2 H3 M4 S5 H6 L7 D8 N9 T10 I13 R14 L15 L16 P17 L18 P19 V20 R21 A22 Q23 L24 L28 Q36 K42 D46 Q50 I51 S52 S53 R57 G58 R59 S60 N63 G71 K85 L95 I96 K97 D98 Y99 V100 S101 E102 D103 L104 H105 K106

4.2.3 Score per residue for model 3

• Molecule 1: Probable serine/threonine-protein kinase pelle

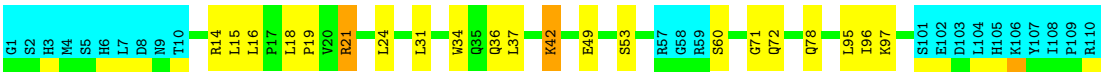
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Probable serine/threonine-protein kinase pelle

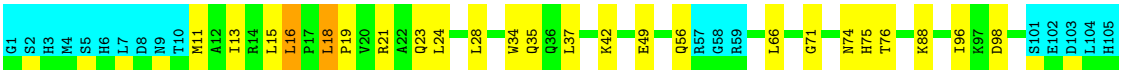
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Probable serine/threonine-protein kinase pelle

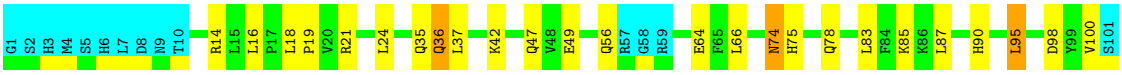
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Probable serine/threonine-protein kinase pelle

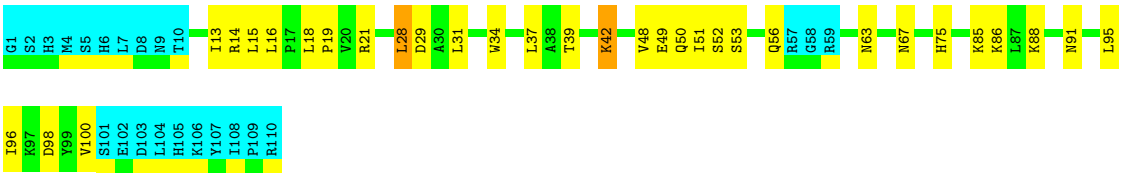
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Probable serine/threonine-protein kinase pelle

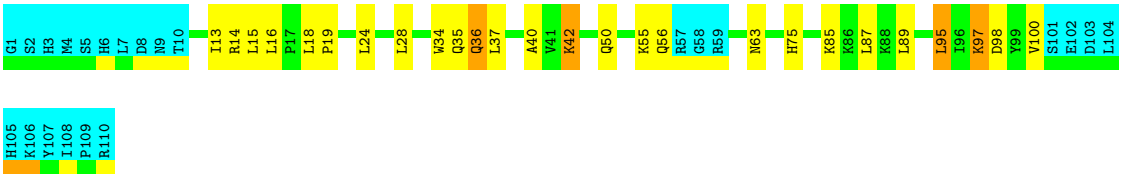
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Probable serine/threonine-protein kinase pelle

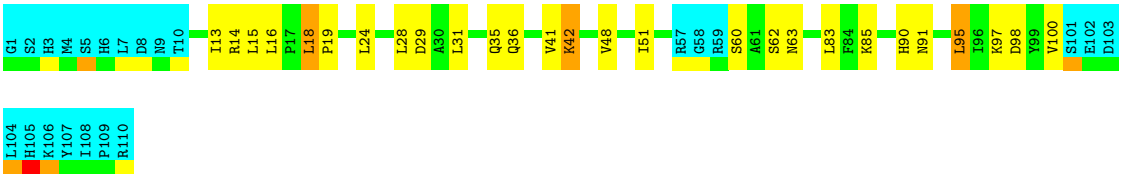
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Probable serine/threonine-protein kinase pelle

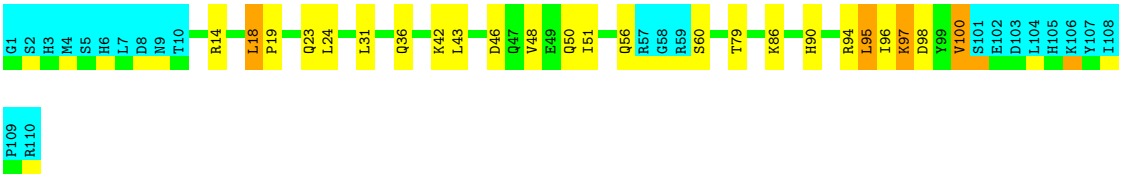
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Probable serine/threonine-protein kinase pelle

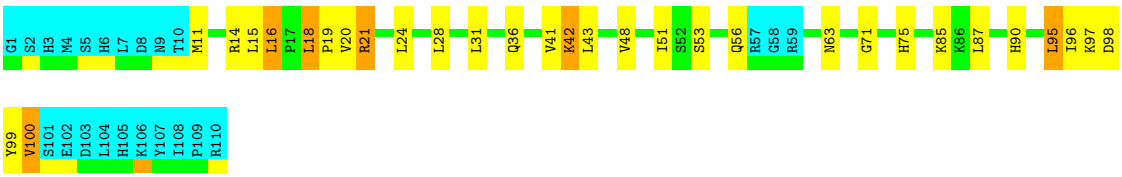
Chain A:



4.2.11 Score per residue for model 11 (medoid)

- Molecule 1: Probable serine/threonine-protein kinase pelle

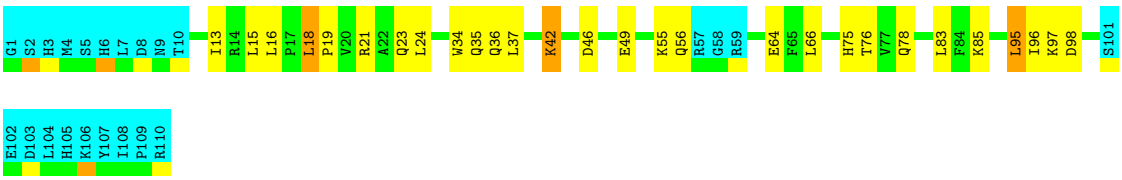
Chain A: 51% 23% 5% 21%



4.2.12 Score per residue for model 12

- Molecule 1: Probable serine/threonine-protein kinase pelle

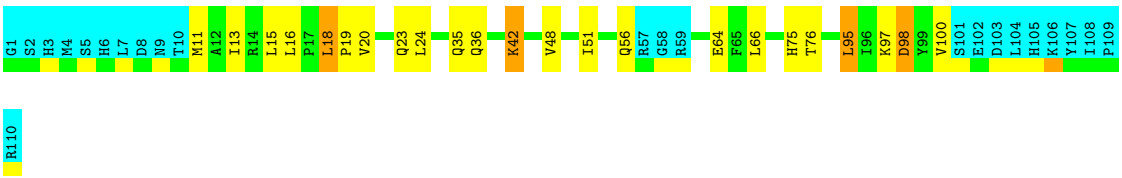
Chain A: 54% 23% • 21%



4.2.13 Score per residue for model 13

- Molecule 1: Probable serine/threonine-protein kinase pelle

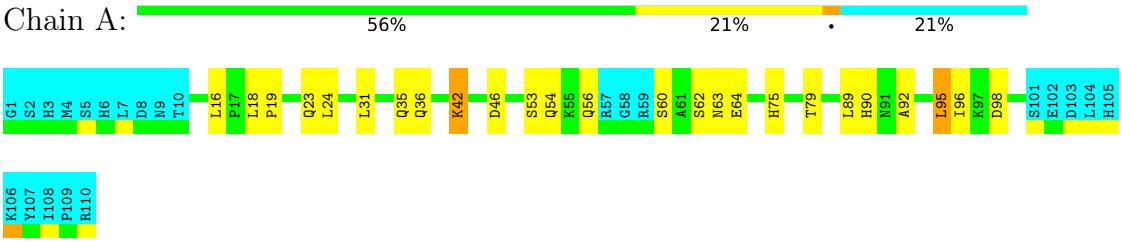
Chain A: 58% 17% • 21%



4.2.14 Score per residue for model 14

- Molecule 1: Probable serine/threonine-protein kinase pelle

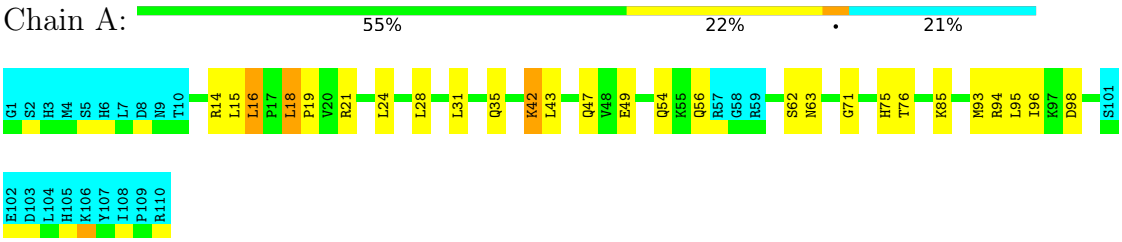
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Probable serine/threonine-protein kinase pelle

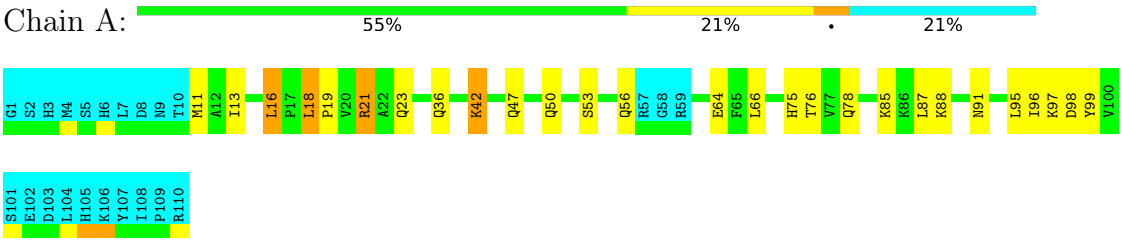
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Probable serine/threonine-protein kinase pelle

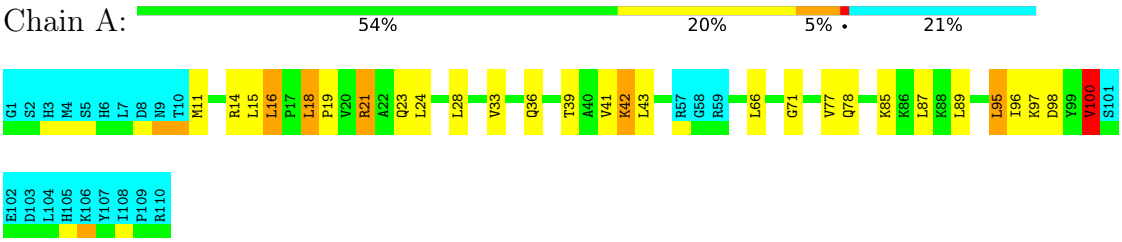
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Probable serine/threonine-protein kinase pelle

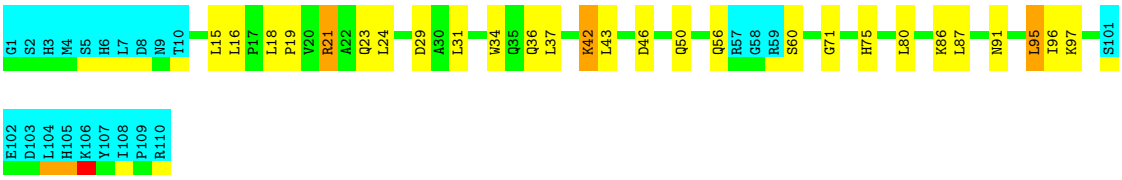
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Probable serine/threonine-protein kinase pelle

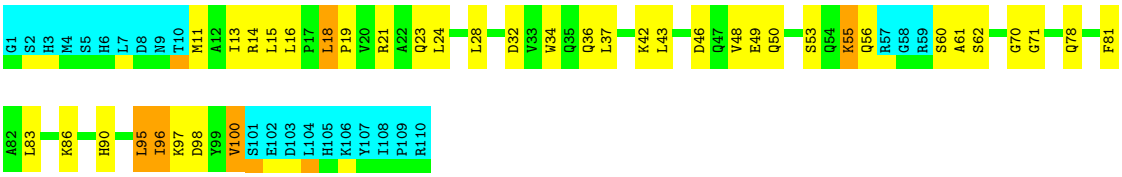
Chain A: 55% 22% • 21%



4.2.19 Score per residue for model 19

- Molecule 1: Probable serine/threonine-protein kinase pelle

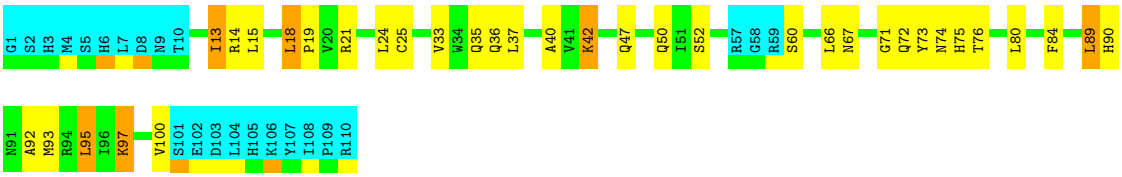
Chain A: 44% 31% 5% 21%



4.2.20 Score per residue for model 20

- Molecule 1: Probable serine/threonine-protein kinase pelle

Chain A: 47% 26% 5% 21%



5 Refinement protocol and experimental data overview (i)

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

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
CNS	refinement	1.1

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	704	727	724	5±3
All	All	14080	14540	14480	105

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:24:LEU:HD23	1:A:95:LEU:HD13	0.72	1.60	8	5
1:A:84:PHE:HB3	1:A:89:LEU:HD11	0.63	1.68	20	1
1:A:28:LEU:HD11	1:A:89:LEU:HD13	0.63	1.70	17	1
1:A:16:LEU:HD12	1:A:21:ARG:HG2	0.61	1.73	16	5
1:A:74:ASN:O	1:A:76:THR:HG23	0.61	1.96	20	1
1:A:34:TRP:HA	1:A:37:LEU:HD12	0.59	1.73	8	8
1:A:18:LEU:N	1:A:19:PRO:HD2	0.57	2.14	15	13
1:A:24:LEU:HB2	1:A:95:LEU:HD22	0.56	1.78	18	3
1:A:24:LEU:HD13	1:A:95:LEU:HD22	0.56	1.77	20	3
1:A:89:LEU:HD22	1:A:92:ALA:HB3	0.55	1.77	20	1
1:A:24:LEU:HD12	1:A:95:LEU:HB3	0.53	1.80	10	3
1:A:24:LEU:O	1:A:24:LEU:HD22	0.51	2.06	19	1
1:A:37:LEU:HD23	1:A:83:LEU:HD21	0.51	1.83	6	1
1:A:48:VAL:HA	1:A:51:ILE:HD12	0.51	1.82	11	5

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:28:LEU:HD22	1:A:37:LEU:HD11	0.50	1.83	7	1
1:A:13:ILE:HD12	1:A:75:HIS:HB3	0.48	1.84	3	1
1:A:36:GLN:O	1:A:40:ALA:HB2	0.48	2.09	8	3
1:A:44:TYR:O	1:A:48:VAL:HG23	0.47	2.09	3	1
1:A:24:LEU:CD1	1:A:28:LEU:HD12	0.47	2.40	19	1
1:A:55:LYS:HG3	1:A:61:ALA:HB2	0.47	1.87	19	2
1:A:77:VAL:HG12	1:A:100:VAL:HG13	0.47	1.86	17	1
1:A:89:LEU:N	1:A:89:LEU:HD12	0.47	2.25	14	2
1:A:33:VAL:HB	1:A:36:GLN:HG2	0.47	1.87	20	1
1:A:21:ARG:HH22	1:A:96:ILE:HG21	0.46	1.71	19	1
1:A:81:PHE:HB2	1:A:100:VAL:HG21	0.46	1.86	19	1
1:A:21:ARG:HH22	1:A:24:LEU:HD23	0.45	1.71	4	1
1:A:80:LEU:HD11	1:A:84:PHE:CZ	0.45	2.46	20	1
1:A:33:VAL:HG12	1:A:36:GLN:NE2	0.45	2.25	17	1
1:A:13:ILE:HG21	1:A:70:GLY:HA3	0.45	1.87	19	1
1:A:24:LEU:O	1:A:28:LEU:HD23	0.45	2.12	5	3
1:A:97:LYS:O	1:A:100:VAL:HG23	0.45	2.12	10	2
1:A:18:LEU:N	1:A:19:PRO:CD	0.45	2.80	18	7
1:A:20:VAL:HG11	1:A:99:TYR:OH	0.45	2.12	11	1
1:A:24:LEU:CD1	1:A:95:LEU:HD22	0.44	2.41	13	1
1:A:24:LEU:HD13	1:A:95:LEU:HB2	0.44	1.88	14	1
1:A:24:LEU:CD2	1:A:28:LEU:HD23	0.44	2.43	8	2
1:A:73:TYR:O	1:A:74:ASN:HB2	0.44	2.13	20	1
1:A:97:LYS:O	1:A:100:VAL:HG22	0.43	2.12	8	1
1:A:20:VAL:O	1:A:23:GLN:HG2	0.43	2.13	13	1
1:A:21:ARG:HH22	1:A:24:LEU:HD12	0.43	1.74	17	1
1:A:24:LEU:CD1	1:A:28:LEU:HD23	0.42	2.44	2	2
1:A:36:GLN:NE2	1:A:37:LEU:HG	0.42	2.30	6	1
1:A:24:LEU:HD11	1:A:28:LEU:CD1	0.42	2.44	19	1
1:A:84:PHE:CB	1:A:89:LEU:HD21	0.42	2.45	20	1
1:A:24:LEU:HD11	1:A:92:ALA:HB1	0.42	1.91	14	1
1:A:13:ILE:HB	1:A:75:HIS:HB2	0.41	1.92	20	1
1:A:41:VAL:HG13	1:A:83:LEU:HD13	0.41	1.91	9	1
1:A:16:LEU:HD21	1:A:99:TYR:HB3	0.41	1.91	16	1
1:A:24:LEU:HD13	1:A:95:LEU:CB	0.41	2.45	14	1
1:A:36:GLN:HG3	1:A:37:LEU:N	0.41	2.29	20	1
1:A:21:ARG:HH21	1:A:96:ILE:HG21	0.41	1.76	18	1
1:A:21:ARG:HH21	1:A:80:LEU:HD21	0.41	1.75	18	1
1:A:84:PHE:HA	1:A:89:LEU:HD21	0.41	1.93	20	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	87/110 (79%)	76±2 (88±2%)	9±2 (10±2%)	2±1 (2±1%)	9 48
All	All	1740/2200 (79%)	1527 (88%)	173 (10%)	40 (2%)	9 48

All 7 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	42	LYS	19
1	A	71	GLY	10
1	A	98	ASP	4
1	A	74	ASN	3
1	A	96	ILE	2
1	A	60	SER	1
1	A	100	VAL	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	76/97 (78%)	56±3 (73±4%)	20±3 (27±4%)	2 22
All	All	1520/1940 (78%)	1115 (73%)	405 (27%)	2 22

All 56 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	95	LEU	19
1	A	16	LEU	18
1	A	42	LYS	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	15	LEU	15
1	A	36	GLN	14
1	A	97	LYS	14
1	A	98	ASP	14
1	A	56	GLN	14
1	A	14	ARG	13
1	A	18	LEU	13
1	A	75	HIS	13
1	A	96	ILE	11
1	A	23	GLN	10
1	A	31	LEU	10
1	A	35	GLN	10
1	A	21	ARG	10
1	A	85	LYS	10
1	A	63	ASN	9
1	A	13	ILE	9
1	A	100	VAL	9
1	A	49	GLU	8
1	A	78	GLN	8
1	A	50	GLN	8
1	A	60	SER	8
1	A	90	HIS	8
1	A	46	ASP	7
1	A	53	SER	7
1	A	66	LEU	7
1	A	47	GLN	6
1	A	64	GLU	6
1	A	11	MET	6
1	A	87	LEU	6
1	A	43	LEU	6
1	A	62	SER	5
1	A	86	LYS	5
1	A	76	THR	5
1	A	54	GLN	4
1	A	52	SER	4
1	A	55	LYS	4
1	A	91	ASN	4
1	A	93	MET	3
1	A	94	ARG	3
1	A	83	LEU	3
1	A	88	LYS	3
1	A	29	ASP	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	72	GLN	2
1	A	39	THR	2
1	A	67	ASN	2
1	A	79	THR	2
1	A	41	VAL	2
1	A	74	ASN	1
1	A	28	LEU	1
1	A	32	ASP	1
1	A	48	VAL	1
1	A	25	CYS	1
1	A	89	LEU	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