Full wwPDB NMR Structure Validation Report (i)

Feb 22, 2022 – 08:53 AM EST

PDB ID	:	1WIL
Title	:	Solution Structure of the RING Finger Domain of the Human KIAA1045 Protein
Authors	:	Miyamoto, K.; Tochio, N.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2004-05-28

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:

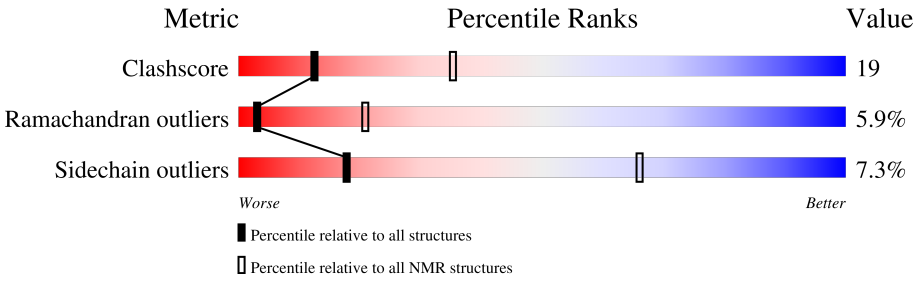
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	89	21% 17% • 61%

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: *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:16-A:50 (35)	0.38	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 3 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1276 atoms, of which 604 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called KIAA1045 protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	89	1274	405	604	116	139	10	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP Q9UPV7
A	2	SER	-	cloning artifact	UNP Q9UPV7
A	3	SER	-	cloning artifact	UNP Q9UPV7
A	4	GLY	-	cloning artifact	UNP Q9UPV7
A	5	SER	-	cloning artifact	UNP Q9UPV7
A	6	SER	-	cloning artifact	UNP Q9UPV7
A	7	GLY	-	cloning artifact	UNP Q9UPV7
A	84	SER	-	cloning artifact	UNP Q9UPV7
A	85	GLY	-	cloning artifact	UNP Q9UPV7
A	86	PRO	-	cloning artifact	UNP Q9UPV7
A	87	SER	-	cloning artifact	UNP Q9UPV7
A	88	SER	-	cloning artifact	UNP Q9UPV7
A	89	GLY	-	cloning artifact	UNP Q9UPV7

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

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: KIAA1045 protein

Chain A: 21% 17% • 61%

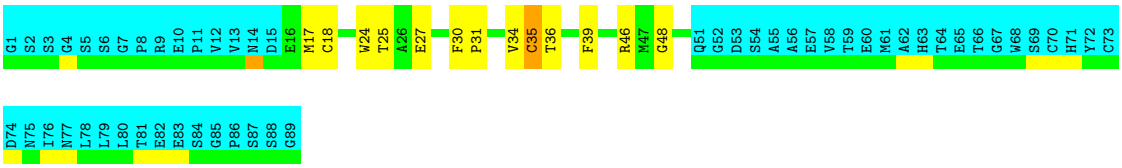
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: KIAA1045 protein

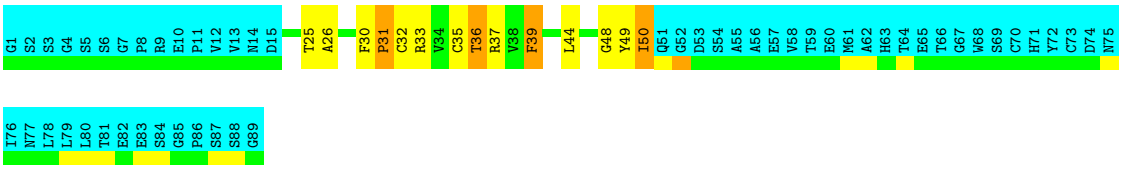
Chain A: 25% 13% • 61%



4.2.2 Score per residue for model 2

• Molecule 1: KIAA1045 protein

Chain A: 24% 11% • 61%



4.2.3 Score per residue for model 3

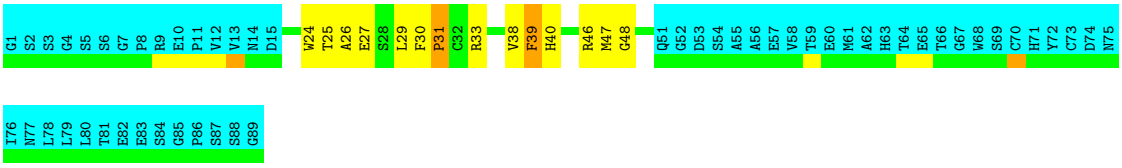
- Molecule 1: KIAA1045 protein

Chain A: 22% 13% • 61%

4.2.4 Score per residue for model 4

- Molecule 1: KIAA1045 protein

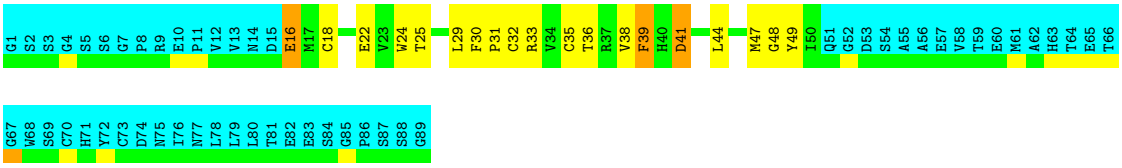
Chain A: 24% 13% • 61%



4.2.5 Score per residue for model 5

- Molecule 1: KIAA1045 protein

Chain A: 18% 18% • 61%



4.2.6 Score per residue for model 6

- Molecule 1: KIAA1045 protein

Chain A: 24% 15% • 61%

G1 S2 S3 G4 S5 S6 G7 P8 R9 E10 P11 V12 V13 N14 D15 E16 M17 W24 T25 L29 F30 P31 C32 R33 V34 C35 V38 F39 H40 L44 R45 Q51 G52 D53 S54 A55 A56 E57 V58 T59 E60 M61 A62 H63 T64 E65 T66 G67 W68 S69 C70 H71 Y72 C73 D74

N75 I76 N77 L78 L79 L80 T81 E82 E83 S84 G85 P86 S87 S88 G89

4.2.7 Score per residue for model 7

- Molecule 1: KIAA1045 protein

Chain A: 16% 22% • 61%

G1 S2 S3 G4 S5 S6 G7 P8 R9 E10 P11 V12 V13 N14 D15 C18 D19 E22 Y23 W24 T25 A26 L29 F30 P31 C32 R33 V34 C35 T36 R37 V38 F39 H40 L44 R45 R46 Q51 G52 D53 S54 A55 A56 E57 V58 T59 E60 M61 A62 H63 T64 E65 T66 G67 W68

S69 C70 H71 Y72 C73 D74 N75 I76 N77 L78 L79 L80 T81 E82 E83 S84 G85 P86 S87 S88 G89

4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: KIAA1045 protein

Chain A: 27% 10% • 61%

G1 S2 S3 G4 S5 S6 G7 P8 R9 E10 P11 V12 V13 N14 D15 V20 W24 T25 F30 P31 C35 T36 R45 R46 M47 G48 Q51 G52 D53 S54 A55 A56 E57 V58 T59 E60 M61 A62 H63 T64 E65 T66 G67 W68 S69 C70 H71 Y72 C73 D74 N75 I76 N77 L78

L79 L80 T81 E82 E83 S84 G85 P86 S87 S88 G89

4.2.9 Score per residue for model 9

- Molecule 1: KIAA1045 protein

Chain A: 30% 9% 61%

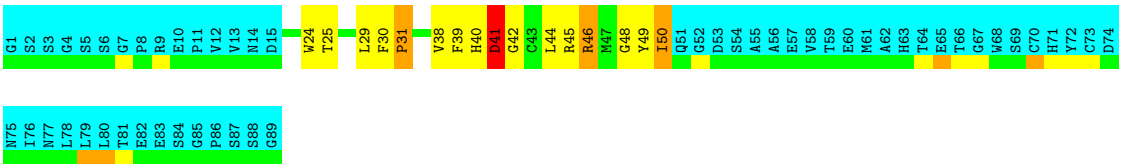
G1 S2 S3 G4 S5 S6 G7 P8 R9 E10 P11 V12 V13 N14 D15 E16 W24 T25 L29 F30 P31 V38 F39 Q51 G52 D53 S54 A55 A56 E57 V58 T59 E60 M61 A62 H63 T64 E65 T66 G67 W68 S69 C70 H71 Y72 C73 D74 N75 I76 N77 L78 L79 L80 T81 E82 E83

S84 G85 P86 S87 S88 G89

4.2.10 Score per residue for model 10

- Molecule 1: KIAA1045 protein

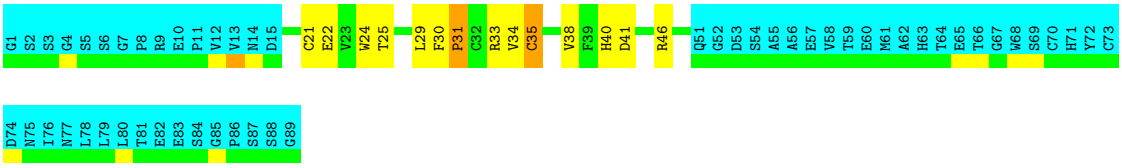
Chain A: 21% 13% • • 61%



4.2.11 Score per residue for model 11

- Molecule 1: KIAA1045 protein

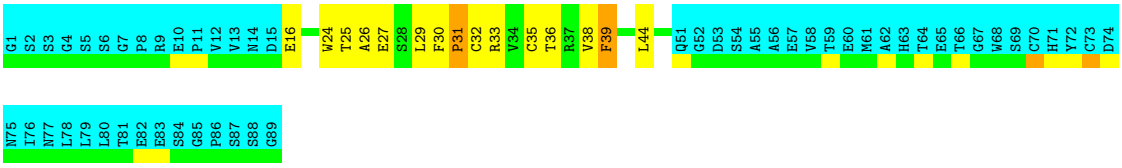
Chain A: 24% 13% • 61%



4.2.12 Score per residue for model 12

- Molecule 1: KIAA1045 protein

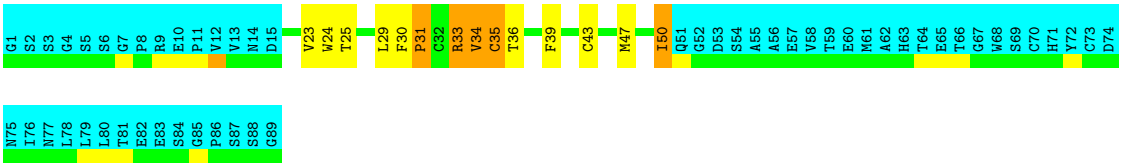
Chain A: 22% 15% • 61%



4.2.13 Score per residue for model 13

- Molecule 1: KIAA1045 protein

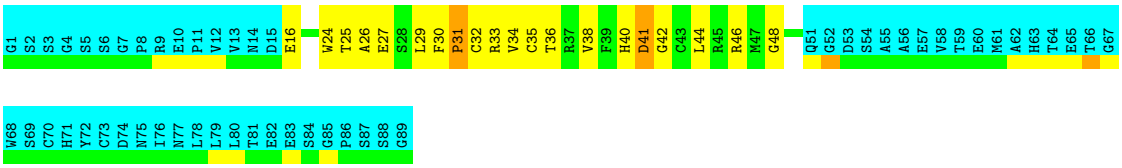
Chain A: 24% 10% 6% 61%



4.2.14 Score per residue for model 14

- Molecule 1: KIAA1045 protein

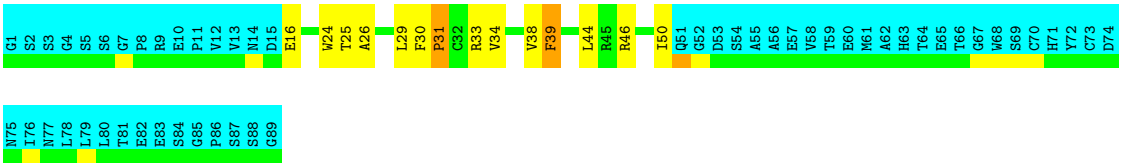
Chain A: 17% 20% • 61%



4.2.15 Score per residue for model 15

- Molecule 1: KIAA1045 protein

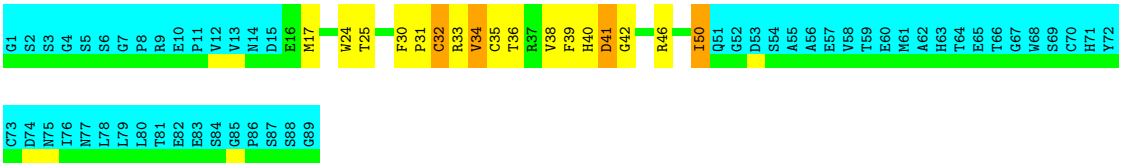
Chain A: 24% 13% • 61%



4.2.16 Score per residue for model 16

- Molecule 1: KIAA1045 protein

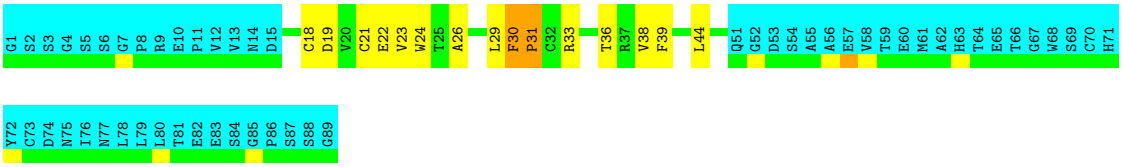
Chain A: 20% 15% • 61%



4.2.17 Score per residue for model 17

- Molecule 1: KIAA1045 protein

Chain A: 22% 15% • 61%



4.2.18 Score per residue for model 18

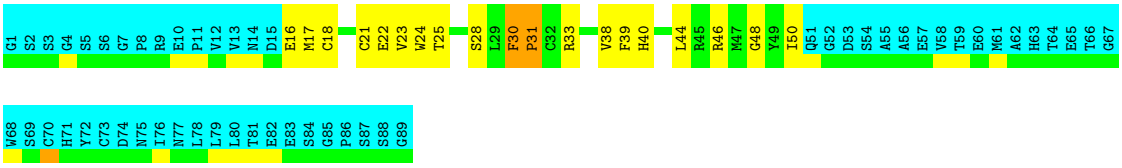
- Molecule 1: KIAA1045 protein

Chain A: 20% 18% • 61%

4.2.19 Score per residue for model 19

- Molecule 1: KIAA1045 protein

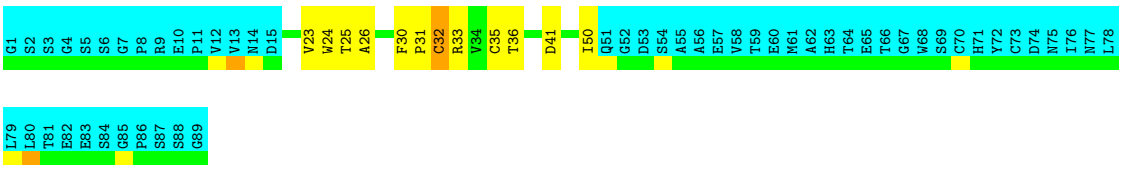
Chain A: 18% 19% • 61%



4.2.20 Score per residue for model 20

- Molecule 1: KIAA1045 protein

Chain A: 26% 12% • 61%



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, structures with the lowest energy, target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	283	266	266	10±3
All	All	5700	5320	5320	208

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 19.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:21:CYS:O	1:A:23:VAL:HG23	0.78	1.78	19	2
1:A:39:PHE:CD1	1:A:44:LEU:HD11	0.70	2.20	10	5
1:A:16:GLU:HB2	1:A:38:VAL:HG21	0.68	1.65	19	1
1:A:29:LEU:HD22	1:A:38:VAL:HB	0.61	1.71	11	8
1:A:24:TRP:CE3	1:A:25:THR:O	0.59	2.55	15	16
1:A:29:LEU:HD11	1:A:40:HIS:NE2	0.58	2.12	7	1
1:A:39:PHE:CD2	1:A:44:LEU:HD21	0.57	2.33	15	1
1:A:44:LEU:HB2	1:A:50:ILE:HD11	0.57	1.76	19	1
1:A:29:LEU:HD23	1:A:38:VAL:HB	0.56	1.76	7	3
1:A:29:LEU:HD23	1:A:30:PHE:N	0.56	2.16	5	10
1:A:41:ASP:HA	1:A:44:LEU:HD12	0.55	1.77	5	3
1:A:16:GLU:N	1:A:16:GLU:CD	0.54	2.61	5	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:36:THR:O	1:A:38:VAL:HG13	0.54	2.01	14	1
1:A:50:ILE:H	1:A:50:ILE:HD13	0.54	1.63	16	1
1:A:23:VAL:HG22	1:A:24:TRP:N	0.53	2.18	20	2
1:A:29:LEU:HD11	1:A:40:HIS:CD2	0.53	2.38	4	1
1:A:16:GLU:HB3	1:A:38:VAL:HG11	0.53	1.80	5	2
1:A:33:ARG:O	1:A:34:VAL:C	0.52	2.48	16	5
1:A:24:TRP:CE3	1:A:25:THR:C	0.52	2.83	11	9
1:A:24:TRP:CZ3	1:A:25:THR:O	0.51	2.63	15	6
1:A:34:VAL:O	1:A:35:CYS:C	0.51	2.49	11	2
1:A:29:LEU:HD21	1:A:40:HIS:CD2	0.50	2.41	3	2
1:A:39:PHE:HD1	1:A:44:LEU:HD11	0.50	1.65	2	1
1:A:16:GLU:OE1	1:A:33:ARG:NE	0.50	2.45	15	1
1:A:36:THR:O	1:A:36:THR:HG22	0.50	2.07	2	2
1:A:19:ASP:OD2	1:A:37:ARG:CZ	0.50	2.59	7	1
1:A:40:HIS:CD2	1:A:40:HIS:N	0.50	2.80	11	1
1:A:16:GLU:OE1	1:A:33:ARG:CD	0.49	2.60	15	1
1:A:24:TRP:CZ3	1:A:26:ALA:HA	0.49	2.42	3	7
1:A:16:GLU:HB3	1:A:38:VAL:HG21	0.49	1.83	15	1
1:A:16:GLU:O	1:A:17:MET:CE	0.49	2.61	19	1
1:A:31:PRO:HG3	1:A:39:PHE:CE2	0.49	2.43	15	2
1:A:33:ARG:O	1:A:35:CYS:N	0.49	2.46	16	1
1:A:31:PRO:HD3	1:A:39:PHE:CE1	0.48	2.43	16	1
1:A:40:HIS:O	1:A:42:GLY:N	0.48	2.45	16	3
1:A:34:VAL:HG13	1:A:35:CYS:N	0.48	2.24	3	2
1:A:40:HIS:O	1:A:41:ASP:C	0.48	2.52	14	4
1:A:16:GLU:CG	1:A:38:VAL:HG11	0.48	2.38	12	1
1:A:35:CYS:O	1:A:36:THR:C	0.47	2.53	8	11
1:A:31:PRO:HG3	1:A:39:PHE:CE1	0.47	2.44	19	2
1:A:31:PRO:HD3	1:A:39:PHE:CD1	0.47	2.43	13	3
1:A:40:HIS:N	1:A:40:HIS:CD2	0.47	2.82	14	2
1:A:49:TYR:O	1:A:50:ILE:C	0.47	2.52	10	2
1:A:24:TRP:CZ3	1:A:25:THR:C	0.47	2.87	11	4
1:A:30:PHE:HA	1:A:31:PRO:C	0.46	2.31	19	20
1:A:25:THR:O	1:A:26:ALA:C	0.46	2.53	2	2
1:A:29:LEU:HD13	1:A:38:VAL:HG21	0.46	1.87	14	1
1:A:31:PRO:O	1:A:32:CYS:C	0.46	2.52	20	1
1:A:18:CYS:O	1:A:22:GLU:N	0.46	2.48	18	6
1:A:20:VAL:HG21	1:A:47:MET:CE	0.45	2.41	8	1
1:A:21:CYS:O	1:A:22:GLU:CG	0.45	2.65	19	2
1:A:23:VAL:HG22	1:A:24:TRP:H	0.45	1.71	20	1
1:A:16:GLU:HG2	1:A:38:VAL:HG11	0.45	1.89	12	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:39:PHE:HD2	1:A:44:LEU:HD11	0.44	1.71	15	2
1:A:29:LEU:HD23	1:A:38:VAL:CB	0.44	2.42	4	2
1:A:35:CYS:O	1:A:37:ARG:N	0.44	2.50	2	1
1:A:17:MET:O	1:A:18:CYS:C	0.44	2.55	1	1
1:A:17:MET:C	1:A:38:VAL:HG23	0.43	2.34	6	1
1:A:16:GLU:C	1:A:38:VAL:HG21	0.43	2.33	9	1
1:A:46:ARG:O	1:A:47:MET:C	0.43	2.57	4	1
1:A:39:PHE:CD1	1:A:39:PHE:N	0.43	2.86	12	5
1:A:31:PRO:HD3	1:A:39:PHE:CD2	0.42	2.50	15	1
1:A:30:PHE:CD1	1:A:30:PHE:N	0.42	2.85	19	2
1:A:24:TRP:CE3	1:A:25:THR:N	0.42	2.88	11	1
1:A:29:LEU:CD2	1:A:38:VAL:HB	0.42	2.45	4	2
1:A:31:PRO:O	1:A:32:CYS:O	0.42	2.37	16	1
1:A:32:CYS:HB2	1:A:34:VAL:HG22	0.41	1.92	7	1
1:A:45:ARG:HG3	1:A:46:ARG:N	0.41	2.31	7	3
1:A:32:CYS:O	1:A:35:CYS:SG	0.41	2.79	16	1
1:A:42:GLY:O	1:A:45:ARG:N	0.41	2.53	18	1
1:A:24:TRP:O	1:A:25:THR:HG23	0.41	2.15	19	1
1:A:47:MET:CE	1:A:49:TYR:CE2	0.41	3.04	5	1
1:A:33:ARG:HD2	1:A:34:VAL:N	0.41	2.31	13	1
1:A:43:CYS:O	1:A:47:MET:CE	0.41	2.68	13	1
1:A:34:VAL:CG1	1:A:35:CYS:N	0.40	2.84	3	1
1:A:34:VAL:O	1:A:36:THR:N	0.40	2.54	1	1
1:A:29:LEU:CD2	1:A:30:PHE:O	0.40	2.69	15	1
1:A:17:MET:N	1:A:38:VAL:CG2	0.40	2.85	16	1
1:A:16:GLU:O	1:A:17:MET:HE3	0.40	2.17	19	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	35/89 (39%)	26±2 (74±6%)	7±2 (20±5%)	2±1 (6±3%)	3 21
All	All	700/1780 (39%)	519 (74%)	140 (20%)	41 (6%)	3 21

All 9 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	31	PRO	13
1	A	48	GLY	9
1	A	35	CYS	4
1	A	34	VAL	4
1	A	27	GLU	3
1	A	41	ASP	3
1	A	50	ILE	2
1	A	32	CYS	2
1	A	36	THR	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	32/76 (42%)	30±1 (93±3%)	2±1 (7±3%)	18	66
All	All	640/1520 (42%)	593 (93%)	47 (7%)	18	66

All 13 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	ARG	11
1	A	46	ARG	8
1	A	32	CYS	8
1	A	39	PHE	7
1	A	41	ASP	3
1	A	50	ILE	2
1	A	30	PHE	2
1	A	29	LEU	1
1	A	27	GLU	1
1	A	16	GLU	1
1	A	45	ARG	1
1	A	19	ASP	1
1	A	28	SER	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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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