wwPDB NMR Structure Validation Summary Report (i)

Feb 22, 2022 – 10:56 AM EST

PDB ID	:	1V9V
Title	:	Solution structure of putative domain of human KIAA0561 protein
Authors	:	Tochio, N.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2004-02-03

This is a wwPDB NMR Structure Validation Summary 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:

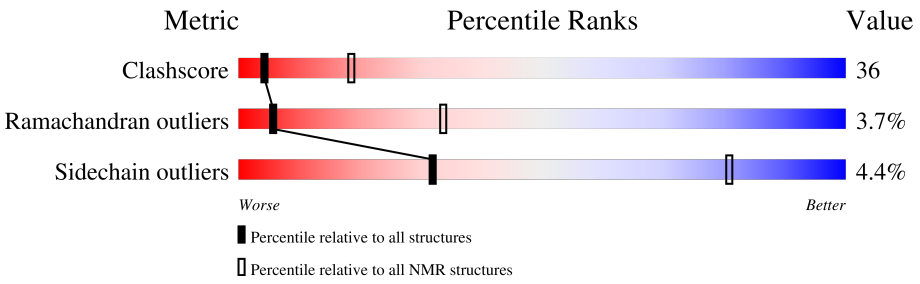
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	114	27% 41% • 28%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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:14-A:27, A:37-A:104 (82)	0.13	15

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1755 atoms, of which 884 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called KIAA0561 protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	114	1755	543	884	158	167	3	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP O60307
A	2	SER	-	cloning artifact	UNP O60307
A	3	SER	-	cloning artifact	UNP O60307
A	4	GLY	-	cloning artifact	UNP O60307
A	5	SER	-	cloning artifact	UNP O60307
A	6	SER	-	cloning artifact	UNP O60307
A	7	GLY	-	cloning artifact	UNP O60307
A	109	SER	-	cloning artifact	UNP O60307
A	110	GLY	-	cloning artifact	UNP O60307
A	111	PRO	-	cloning artifact	UNP O60307
A	112	SER	-	cloning artifact	UNP O60307
A	113	SER	-	cloning artifact	UNP O60307
A	114	GLY	-	cloning artifact	UNP O60307

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: KIAA0561 protein

Chain A: 27% 41% • 28%

G1 S2 S3 G4 S5 S6 G7 P8 K9 A10 T11 A12 Q13 M14 E15 G16 R17 L18 L22 Y25 G28 A29 R30 L31 A32 L33 A34 D35 G36 V37 L38 G39 F40 I41 Q44 I45 V46 E47 L48 A49 L53 A54 K55 V61 T62 S63 R64 Y65 F66 L67 E68 M69 L73

E74 R75 L76 L77 Q78 D79 A80 H81 E82 R83 S84 D85 S86 E87 E88 V89 S90 F91 I92 V93 Q94 L95 V96 R97 K98 L99 L100 I101 I102 I103 S104 R105 P106 A107 R108 S109 G110 P111 S112 S113 G114

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: KIAA0561 protein

Chain A: 29% 40% • 28%

G1 S2 S3 G4 S5 S6 G7 P8 K9 A10 T11 A12 Q13 M14 L22 Y25 A26 P27 G28 A29 R30 L31 A32 L33 A34 D35 G36 V37 L38 G39 F40 I41 Q44 I45 V46 E47 L48 A49 R50 D51 K55 V61 T62 S63 R64 Y65 F66 L67 E68 M69 L73 E74 R75 L76

L77 Q78 D79 A80 H81 E82 R83 S84 D85 S86 E87 E88 V89 I92 V93 Q94 L95 V96 R97 K98 L99 L100 I101 I102 I103 S104 R105 P106 A107 R108 S109 G110 P111 S112 S113 G114

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, 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)

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	668	684	681	48±5
All	All	13360	13680	13620	965

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 36.

5 of 158 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:73:LEU:HD22	1:A:99:LEU:HD23	0.94	1.40	14	20
1:A:73:LEU:HD11	1:A:100:LEU:HD21	0.86	1.46	4	20
1:A:89:VAL:HG12	1:A:93:VAL:HG23	0.78	1.54	14	11
1:A:38:LEU:C	1:A:38:LEU:HD13	0.76	2.01	4	19
1:A:61:VAL:HG12	1:A:65:TYR:CE2	0.75	2.16	16	15

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	82/114 (72%)	55±3 (67±4%)	24±3 (30±4%)	3±1 (4±1%)	6 34
All	All	1640/2280 (72%)	1092 (67%)	488 (30%)	60 (4%)	6 34

5 of 11 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	86	SER	19
1	A	84	SER	14
1	A	37	VAL	10
1	A	77	LEU	4
1	A	40	PHE	4

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	74/93 (80%)	71±1 (96±2%)	3±1 (4±2%)	32 81
All	All	1480/1860 (80%)	1415 (96%)	65 (4%)	32 81

5 of 14 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	83	ARG	20
1	A	14	MET	8
1	A	69	MET	8
1	A	22	LEU	6
1	A	51	ASP	5

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