Full wwPDB NMR Structure Validation Report (i)

Feb 19, 2022 – 09:12 PM EST

PDB ID	:	1UEP
Title	:	Solution Structure of The Third PDZ Domain of Human Atrophin-1 Interacting Protein 1 (KIAA0705 Protein)
Authors	:	Miyamoto, K.; Kigawa, T.; Koshiba, S.; Inoue, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2003-05-20

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:

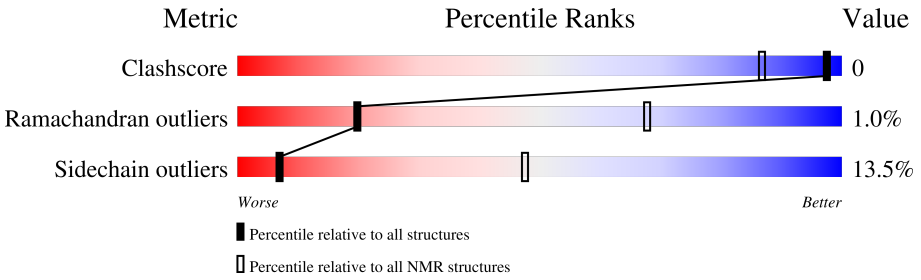
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	103	80% 6% 15%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 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:8-A:95 (88)	0.32	16

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 4 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1532 atoms, of which 764 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	103	1532	475	764	147	143	3	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP Q86UL8
A	2	SER	-	cloning artifact	UNP Q86UL8
A	3	SER	-	cloning artifact	UNP Q86UL8
A	4	GLY	-	cloning artifact	UNP Q86UL8
A	5	SER	-	cloning artifact	UNP Q86UL8
A	6	SER	-	cloning artifact	UNP Q86UL8
A	7	GLY	-	cloning artifact	UNP Q86UL8
A	98	SER	-	cloning artifact	UNP Q86UL8
A	99	GLY	-	cloning artifact	UNP Q86UL8
A	100	PRO	-	cloning artifact	UNP Q86UL8
A	101	SER	-	cloning artifact	UNP Q86UL8
A	102	SER	-	cloning artifact	UNP Q86UL8
A	103	GLY	-	cloning artifact	UNP Q86UL8

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: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 80% 6% 15%

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: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 71% 14% • 15%

4.2.2 Score per residue for model 2

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 72% 14% 15%

4.2.3 Score per residue for model 3

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 69% 16% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 K9 R16 F22 R25 I26 A40 V41 R49 R52 L53 D77 L78 H81 Q87 V88 N89 R93 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.4 Score per residue for model 4

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 70% 15% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 D12 F22 G23 F24 R25 I26 D50 G51 R52 Y61 K70 R73 Y74 L78 L90 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.5 Score per residue for model 5

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 71% 15% 15%

G1 S2 S3 G4 S5 S6 G7 L11 F24 R25 I26 G33 A40 D50 G51 R52 L53 H54 Y74 D77 L78 L90 R93 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.6 Score per residue for model 6

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 74% 10% • 15%

G1 S2 S3 G4 S5 S6 G7 F22 R25 I26 G33 R49 R52 E58 D63 R73 D77 H80 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.7 Score per residue for model 7

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 73% 12% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 K9 E10 L11 M18 E19 F22 I26 R49 H54 D77 H81 R93 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.8 Score per residue for model 8

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 68% 16% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 K9 D12 R17 F22 G23 F24 R25 I26 L27 S46 R49 E58 L78 M79 H80 R84 M89 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.9 Score per residue for model 9

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 71% 14% • 15%

G1 S2 S3 G4 S5 S6 G7 L11 R16 E19 F22 I26 L27 R52 L53 H54 D57 D77 L78 R84 V92 R93 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.10 Score per residue for model 10

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 70% 13% • 15%

G1 S2 S3 G4 S5 S6 G7 R17 M18 E19 F22 I26 L27 D30 D50 G51 R52 E58 D77 L78 M79 H80 R93 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.11 Score per residue for model 11

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 75% 11% 15%

G1 S2 S3 G4 S5 S6 G7 L11 H14 E19 F22 I26 A40 D50 G51 R52 E58 H81 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.12 Score per residue for model 12

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 74% 11% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 K9 E19 F22 I26 R52 E58 Y74 D77 L78 M79 H80 H81 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.13 Score per residue for model 13

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 73% 11% • 15%

4.2.14 Score per residue for model 14

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 74% 11% • 15%

4.2.15 Score per residue for model 15

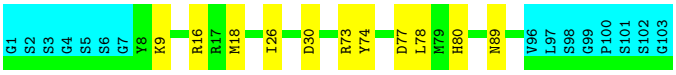
• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 74% 12% 15%

4.2.16 Score per residue for model 16 (medoid)

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

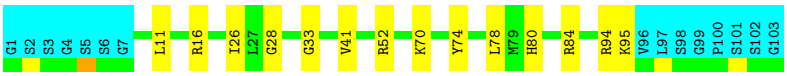
Chain A: 75% 11% 15%



4.2.17 Score per residue for model 17

• Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 72% 14% 15%



4.2.18 Score per residue for model 18

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 69% 16% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 K9 E10 L11 R16 F22 G23 F24 R25 I26 L27 G28 G33 V41 R49 G64 M79 H80 H81 L90 R94 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.19 Score per residue for model 19

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 78% 8% 15%

G1 S2 S3 G4 S5 S6 G7 Y8 L11 F22 I26 Y74 L78 H81 K95 V96 L97 S98 G99 P100 S101 S102 G103

4.2.20 Score per residue for model 20

- Molecule 1: Membrane Associated Guanylate Kinase Inverted-2 (MAGI-2)

Chain A: 69% 15% • 15%

G1 S2 S3 G4 S5 S6 G7 Y8 R17 S20 G21 F22 G23 F24 R25 I26 V41 D57 E58 L59 G64 I65 K70 R73 Y74 L78 V92 K95 V96 L97 S98 G99 P100 S101 S102 G103

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics, restrainted molecular 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	1.0.7
OPALp	refinement	

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	0.56±0.01	0±0/698 (0.0± 0.0%)	1.05±0.04	2±1/941 (0.2± 0.1%)
All	All	0.56	0/13960 (0.0%)	1.06	30/18820 (0.2%)

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	2.1±1.1
All	All	0	42

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	57	ASP	CB-CG-OD2	-7.66	111.41	118.30	9	1
1	A	8	TYR	CB-CG-CD1	-7.24	116.66	121.00	12	3
1	A	57	ASP	CB-CG-OD1	6.08	123.77	118.30	9	1
1	A	52	ARG	NE-CZ-NH2	-5.91	117.35	120.30	4	3
1	A	8	TYR	CB-CG-CD2	-5.85	117.49	121.00	14	1
1	A	49	ARG	NE-CZ-NH2	-5.72	117.44	120.30	8	2
1	A	16	ARG	NE-CZ-NH2	-5.70	117.45	120.30	2	2
1	A	94	ARG	NE-CZ-NH2	-5.67	117.47	120.30	14	3
1	A	16	ARG	NE-CZ-NH1	5.63	123.12	120.30	9	2
1	A	94	ARG	CD-NE-CZ	5.61	131.46	123.60	14	2
1	A	41	VAL	CA-CB-CG2	5.57	119.26	110.90	20	1
1	A	25	ARG	NE-CZ-NH2	-5.39	117.60	120.30	8	1
1	A	17	ARG	NE-CZ-NH2	-5.36	117.62	120.30	13	2
1	A	93	ARG	NE-CZ-NH1	5.28	122.94	120.30	2	2
1	A	73	ARG	NE-CZ-NH2	-5.19	117.70	120.30	4	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	24	PHE	CB-CG-CD1	-5.11	117.22	120.80	20	1
1	A	59	LEU	CB-CG-CD2	5.05	119.59	111.00	20	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	52	ARG	Sidechain	7
1	A	74	TYR	Sidechain	6
1	A	25	ARG	Sidechain	4
1	A	84	ARG	Sidechain	3
1	A	61	TYR	Sidechain	3
1	A	93	ARG	Sidechain	3
1	A	17	ARG	Sidechain	3
1	A	73	ARG	Sidechain	2
1	A	19	GLU	Peptide	2
1	A	94	ARG	Sidechain	2
1	A	49	ARG	Sidechain	2
1	A	28	GLY	Peptide	2
1	A	18	MET	Peptide	1
1	A	16	ARG	Sidechain	1
1	A	8	TYR	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	684	687	687	1±1
All	All	13680	13740	13740	13

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:70:LYS:HE3	1:A:74:TYR:CE2	0.53	2.38	4	1
1:A:17:ARG:HB2	1:A:22:PHE:CE1	0.50	2.42	10	2
1:A:70:LYS:HE3	1:A:74:TYR:CZ	0.48	2.43	4	1
1:A:24:PHE:CE2	1:A:90:LEU:HD11	0.48	2.44	5	2
1:A:70:LYS:HE3	1:A:74:TYR:CE1	0.48	2.44	17	1
1:A:59:LEU:HD23	1:A:92:VAL:HG12	0.47	1.87	20	1
1:A:27:LEU:HD12	1:A:27:LEU:C	0.43	2.33	8	1
1:A:57:ASP:OD2	1:A:92:VAL:HG11	0.42	2.14	9	1
1:A:24:PHE:CE1	1:A:90:LEU:HD11	0.42	2.49	18	1
1:A:8:TYR:CD1	1:A:8:TYR:C	0.40	2.94	12	1
1:A:70:LYS:CB	1:A:74:TYR:CD2	0.40	3.04	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	88/103 (85%)	78±2 (88±2%)	9±2 (11±2%)	1±1 (1±1%)	20	68
All	All	1760/2060 (85%)	1554 (88%)	189 (11%)	17 (1%)	20	68

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	33	GLY	4
1	A	40	ALA	3
1	A	19	GLU	3
1	A	22	PHE	2
1	A	23	GLY	2
1	A	64	GLY	2
1	A	8	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	70/80 (88%)	61±2 (86±3%)	9±2 (13±3%)	7 47
All	All	1400/1600 (88%)	1211 (86%)	189 (14%)	7 47

All 40 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	26	ILE	20
1	A	95	LYS	16
1	A	22	PHE	14
1	A	78	LEU	12
1	A	77	ASP	10
1	A	81	HIS	10
1	A	94	ARG	9
1	A	11	LEU	9
1	A	80	HIS	9
1	A	58	GLU	7
1	A	9	LYS	6
1	A	50	ASP	5
1	A	49	ARG	4
1	A	41	VAL	4
1	A	12	ASP	4
1	A	54	HIS	4
1	A	79	MET	4
1	A	19	GLU	3
1	A	20	SER	3
1	A	46	SER	3
1	A	16	ARG	3
1	A	89	ASN	3
1	A	30	ASP	3
1	A	87	GLN	2
1	A	53	LEU	2
1	A	73	ARG	2
1	A	18	MET	2
1	A	27	LEU	2
1	A	57	ASP	2
1	A	48	ASP	2
1	A	31	GLU	1
1	A	25	ARG	1
1	A	63	ASP	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	17	ARG	1
1	A	84	ARG	1
1	A	93	ARG	1
1	A	14	HIS	1
1	A	52	ARG	1
1	A	15	LEU	1
1	A	65	ILE	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