wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 07:53 PM BST

PDB ID	:	2E63
Title	:	Solution structure of the NEUZ domain in KIAA1787 protein
Authors	:	He, F.; Muto, Y.; Inoue, M.; Kigawa, T.; Shirouzu, M.; Terada, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2006-12-25

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
http://wwpdb.org/validation/2016/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:

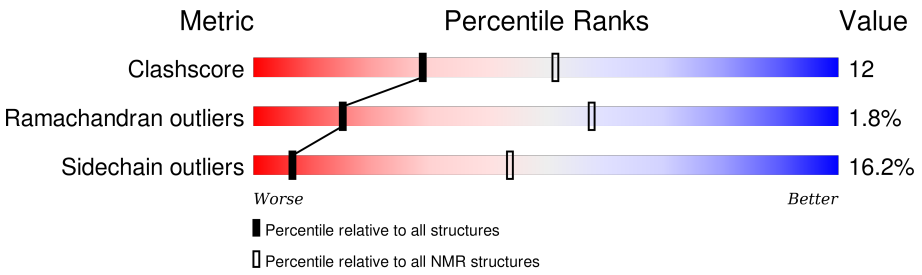
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 is 89%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	170	55% 31% • 14%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 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:9-A:18, A:25-A:28, A:38-A:170 (147)	0.15	2

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2546 atoms, of which 1270 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called KIAA1787 protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	170	2546	786	1270	241	245	4	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP Q96JN8
A	2	SER	-	CLONING ARTIFACT	UNP Q96JN8
A	3	SER	-	CLONING ARTIFACT	UNP Q96JN8
A	4	GLY	-	CLONING ARTIFACT	UNP Q96JN8
A	5	SER	-	CLONING ARTIFACT	UNP Q96JN8
A	6	SER	-	CLONING ARTIFACT	UNP Q96JN8
A	7	GLY	-	CLONING ARTIFACT	UNP Q96JN8

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: KIAA1787 protein

Chain A: 55% 31% • 14%

G1 S2 S3 G4 S5 S6 G7 E8 L9 T13 V17 S18 L19 S20 A21 C22 G23 R24 T25 A26 R27 R28 Q29 Q30 P31 G32 Q33 E34 F35 N36 H37 V40 L41 S42 R43 L46 R50 V51 F52 I56 D57 R58 K59 I67 E68 I69 G70 V71 T72 D75 P76 S77

D80 F81 P82 S83 S84 A85 V94 C98 S99 V100 L101 R102 V107 L108 E109 Q113 D114 L115 L118 G119 E120 G121 D122 R123 V124 E127 R128 E133 L134 A146 A156 L160 K163 I167 L170

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

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: KIAA1787 protein

Chain A: 49% 38% 14%

G1 S2 S3 G4 S5 S6 G7 E8 L9 T13 G14 R15 L16 V17 S18 L19 S20 A21 C22 G23 R24 T25 A26 Q29 Q30 P31 G32 Q33 E34 F35 N36 H37 V40 L41 S42 R43 L46 R47 R50 V51 F52 I56 D57 R58 K59 S62 I67 E68 I69 G70 V71 T72

D75 P76 S77 D80 F81 P82 A85 S92 W93 V94 C98 S99 V100 L101 R102 V107 L108 E109 Q113 D114 L115 D116 Q117 L118 G119 E120 G121 D122 R123 V124 G125 V126 E127 R128 L134 R135 L136 N139 V145 A146 A156 V157 V158 D159 L160 K163 C164

L170

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	2.1
CYANA	refinement	2.1

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	BMRB entry 11348
Number of chemical shift lists	1
Total number of shifts	2005
Number of shifts mapped to atoms	2005
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	89%

No validations of the models with respect to experimental NMR restraints is performed at this time.

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	1115	1131	1131	28±3
All	All	22300	22620	22620	553

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.

5 of 104 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:41:LEU:HD21	1:A:72:THR:HG21	0.77	1.56	13	20
1:A:46:LEU:HD11	1:A:156:ALA:HB2	0.75	1.58	15	14
1:A:41:LEU:HD11	1:A:157:VAL:HG22	0.74	1.57	17	8
1:A:17:VAL:HG11	1:A:40:VAL:HG13	0.71	1.62	3	18
1:A:9:LEU:HD11	1:A:167:ILE:HD11	0.69	1.65	16	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	146/170 (86%)	132±2 (91±1%)	11±2 (8±1%)	3±1 (2±0%)	15 58
All	All	2920/3400 (86%)	2647 (91%)	219 (8%)	54 (2%)	15 58

All 4 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	119	GLY	20
1	A	120	GLU	19
1	A	97	GLY	10
1	A	14	GLY	5

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	121/138 (88%)	101±3 (84±3%)	20±3 (16±3%)	7 44
All	All	2420/2760 (88%)	2027 (84%)	393 (16%)	7 44

5 of 47 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	160	LEU	20
1	A	98	CYS	20
1	A	101	LEU	20
1	A	77	SER	20
1	A	128	ARG	19

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 carbohydrates 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)

The completeness of assignment taking into account all chemical shift lists is 89% for the well-defined parts and 88% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 11348

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	2005
Number of shifts mapped to atoms	2005
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	166	-0.11 ± 0.07	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	144	0.08 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	164	0.19 ± 0.14	None needed (< 0.5 ppm)
^{15}N	154	0.36 ± 0.28	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 89%, i.e. 1569 atoms were assigned a chemical shift out of a possible 1757. 0 out of 37 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	721/723 (100%)	288/288 (100%)	293/294 (100%)	140/141 (99%)
Sidechain	763/945 (81%)	462/548 (84%)	294/341 (86%)	7/56 (12%)

Continued on next page...

Continued from previous page...

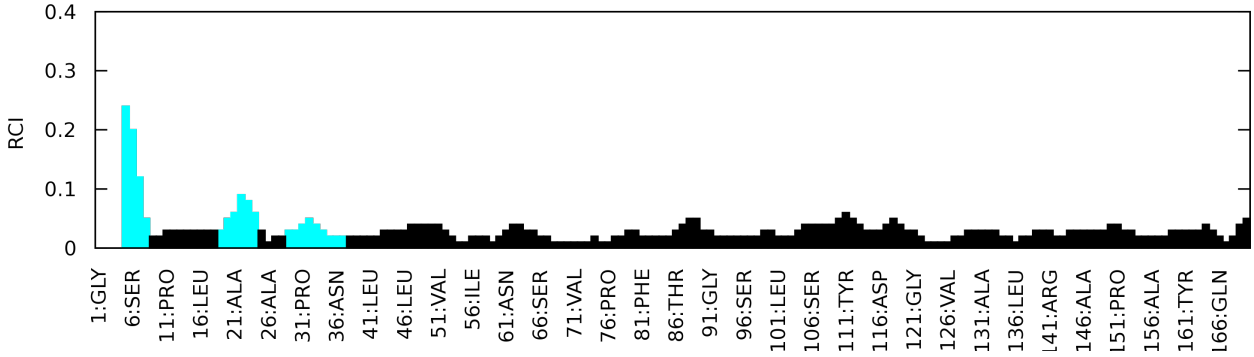
	Total	^{1}H	^{13}C	^{15}N
Aromatic	85/89 (96%)	43/46 (93%)	38/38 (100%)	4/5 (80%)
Overall	1569/1757 (89%)	793/882 (90%)	625/673 (93%)	151/202 (75%)

7.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	35	PHE	HB3	-0.19	4.85 – 1.05	-8.3
1	A	75	ASP	HA	2.37	6.15 – 3.05	-7.2
1	A	142	ASP	HB2	1.06	4.07 – 1.37	-6.1
1	A	125	GLY	HA3	5.94	5.80 – 2.00	5.4
1	A	102	ARG	HG3	0.01	3.00 – 0.10	-5.3
1	A	142	ASP	HA	2.99	6.15 – 3.05	-5.2

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: