wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 07:28 PM BST

PDB ID	:	2DGO
Title	:	Solution structure of the RNA binding domain in cytotoxic granule-associated RNA binding protein 1
Authors	:	Abe, C.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2006-03-15

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:

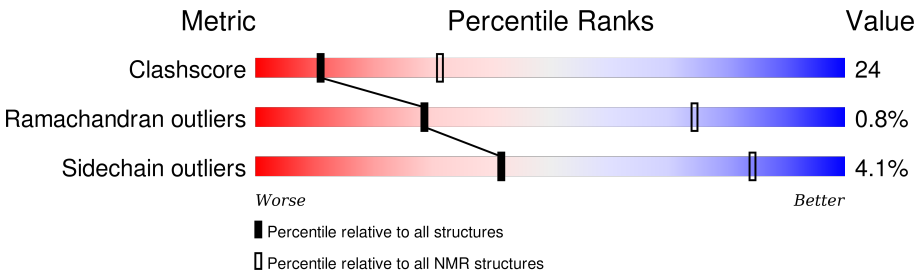
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	115	43% 23% • 34%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 13 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:78-A:153 (76)	0.15	13

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1717 atoms, of which 841 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cytotoxic granule-associated RNA binding protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	115	1717	546	841	158	170	2	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	63	GLY	-	CLONING ARTIFACT	UNP P52912
A	64	SER	-	CLONING ARTIFACT	UNP P52912
A	65	SER	-	CLONING ARTIFACT	UNP P52912
A	66	GLY	-	CLONING ARTIFACT	UNP P52912
A	67	SER	-	CLONING ARTIFACT	UNP P52912
A	68	SER	-	CLONING ARTIFACT	UNP P52912
A	69	GLY	-	CLONING ARTIFACT	UNP P52912
A	172	SER	-	CLONING ARTIFACT	UNP P52912
A	173	GLY	-	CLONING ARTIFACT	UNP P52912
A	174	PRO	-	CLONING ARTIFACT	UNP P52912
A	175	SER	-	CLONING ARTIFACT	UNP P52912
A	176	SER	-	CLONING ARTIFACT	UNP P52912
A	177	GLY	-	CLONING ARTIFACT	UNP P52912

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: Cytotoxic granule-associated RNA binding protein 1

Chain A: 43% 23% • 34%

G63 S64 S65 G66 S67 S68 G69 Q70 K71 K72 D73 T74 S75 N76 H77 F78 H79 V82 L85 I89 T90 T91 E92 D93 I94 K95 A99 P100 V109 V110 K111 D112 G116 Y121 V124 S125 F126 K129 W130 D131 A135 M139 G140 L144 G145 R150 T151

A154 T155 R156 K157 P158 P159 A160 P161 K162 S163 T164 Y165 E166 S167 N168 T169 K170 Q171 S172 G173 P174 S175 S176 G177

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

The representative model is number 13. Colouring as in section 4.1 above.

- Molecule 1: Cytotoxic granule-associated RNA binding protein 1

Chain A: 39% 26% • 34%

G63 S64 S65 G66 S67 S68 G69 Q70 K71 K72 D73 T74 S75 N76 H77 F78 H79 V82 L85 I89 T90 T91 E92 D93 I94 K95 A99 P100 F101 I104 S105 D106 A107 R108 V109 V110 K111 D112 G116 Y121 V124 S125 F126 K129 W130 D131 A132 A135 M139

G145 R150 A154 T155 R156 K157 P158 P159 A160 P161 K162 S163 T164 Y165 E166 S167 N168 T169 K170 Q171 S172 G173 P174 S175 S176 G177

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: *target function,structures with the lowest energy,structures with the least restraint violations.*

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

Software name	Classification	Version
CYANA	structure solution	2.0
CYANA	refinement	2.0

No chemical shift data was provided. 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	601	578	578	28±3
All	All	12020	11560	11560	561

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 24.

5 of 107 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:110:VAL:HG11	1:A:121:TYR:CE2	0.82	2.08	1	11
1:A:89:ILE:HD11	1:A:145:GLY:HA3	0.81	1.52	8	20
1:A:82:VAL:HG12	1:A:85:LEU:HD21	0.81	1.51	12	17
1:A:110:VAL:HG12	1:A:119:LYS:HG3	0.78	1.53	17	8
1:A:110:VAL:HG11	1:A:121:TYR:CZ	0.75	2.16	12	8

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	76/115 (66%)	63±2 (83±3%)	12±2 (16±3%)	1±1 (1±1%)	29 74
All	All	1520/2300 (66%)	1267 (83%)	241 (16%)	12 (1%)	29 74

All 3 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	142	GLN	6
1	A	140	GLY	5
1	A	136	ILE	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	60/92 (65%)	58±1 (96±2%)	2±1 (4±2%)	42 84
All	All	1200/1840 (65%)	1151 (96%)	49 (4%)	42 84

5 of 11 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	130	TRP	16
1	A	79	HIS	12
1	A	86	SER	6
1	A	93	ASP	4
1	A	108	ARG	3

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)

No chemical shift data were provided