wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:30 AM BST

PDB ID	:	2ROK
Title	:	Solution structure of the cap-binding domain of PARN complexed with the cap analog
Authors	:	Nagata, T.; Suzuki, S.; Endo, R.; Shirouzu, M.; Terada, T.; Inoue, M.; Kigawa, T.; Guntert, P.; Hayashizaki, Y.; Muto, Y.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2008-03-28

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:

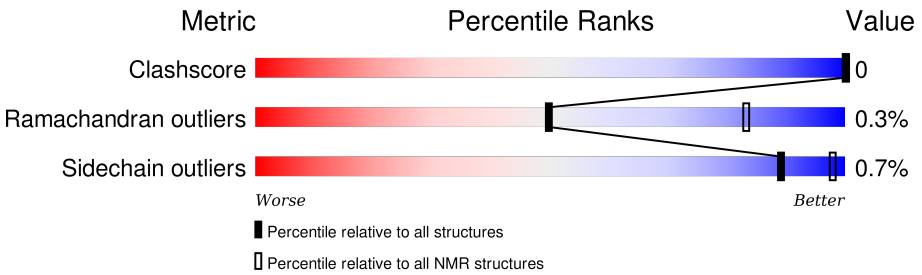
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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	100	69% 31%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 4 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:437-A:505 (69)	0.23	4

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 2 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1616 atoms, of which 785 are hydrogens and 0 are deuteriums.

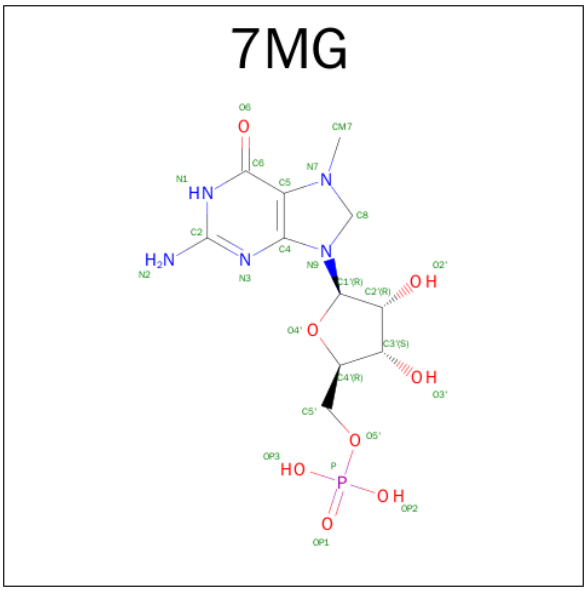
- Molecule 1 is a protein called poly(A)-specific ribonuclease.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	100	1537	492	758	132	155	0

There are 13 discrepancies between the modelled and reference sequences:

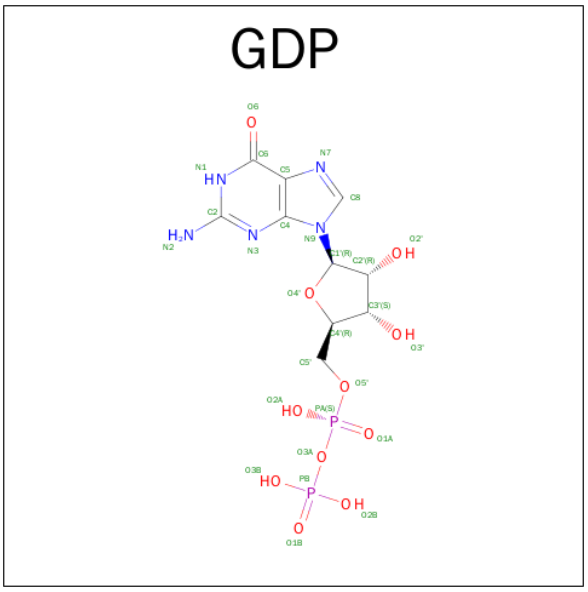
Chain	Residue	Modelled	Actual	Comment	Reference
A	423	GLY	-	EXPRESSION TAG	UNP Q3TUQ8
A	424	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	425	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	426	GLY	-	EXPRESSION TAG	UNP Q3TUQ8
A	427	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	428	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	429	GLY	-	EXPRESSION TAG	UNP Q3TUQ8
A	517	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	518	GLY	-	EXPRESSION TAG	UNP Q3TUQ8
A	519	PRO	-	EXPRESSION TAG	UNP Q3TUQ8
A	520	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	521	SER	-	EXPRESSION TAG	UNP Q3TUQ8
A	522	GLY	-	EXPRESSION TAG	UNP Q3TUQ8

- Molecule 2 is 7N-METHYL-8-HYDROGUANOSINE-5'-MONOPHOSPHATE (three-letter code: 7MG) (formula: $C_{11}H_{18}N_5O_8P$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
2	A	1	39	11	15	5	7	1

- Molecule 3 is GUANOSINE-5'-DIPHOSPHATE (three-letter code: GDP) (formula: $C_{10}H_{15}N_5O_{11}P_2$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
3	A	1	40	10	12	5	11	2

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: poly(A)-specific ribonuclease

Chain A:

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

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: poly(A)-specific ribonuclease

Chain A:

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics, simulated annealing.*

Of the 40 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
CYANA	structure solution	2.1
AMBER	refinement	9

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)

Bond lengths and bond angles in the following residue types are not validated in this section: GDP, 7MG

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	0.77±0.00	0±0/589 (0.0±0.0%)	0.92±0.01	1±1/804 (0.1±0.1%)
All	All	0.77	0/11780 (0.0%)	0.92	10/16080 (0.1%)

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	437	ARG	NE-CZ-NH1	6.36	123.48	120.30	3	6
1	A	498	ARG	NE-CZ-NH1	5.74	123.17	120.30	5	4

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
All	All	12480	11360	11380	-

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 -.

There are no clashes.

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	69/100 (69%)	66±1 (96±2%)	2±1 (4±2%)	0±0 (0±1%)	50 83
All	All	1380/2000 (69%)	1327 (96%)	49 (4%)	4 (0%)	50 83

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	495	GLU	2
1	A	439	HIS	1
1	A	496	SER	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	63/86 (73%)	63±1 (99±1%)	0±1 (1±1%)	89 98
All	All	1260/1720 (73%)	1251 (99%)	9 (1%)	89 98

5 of 9 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	444	THR	1
1	A	465	GLN	1
1	A	447	LYS	1
1	A	455	TYR	1
1	A	437	ARG	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	7MG	A	523	3	20,26,27	1.73±0.02	0±0 (0±1%)
3	GDP	A	524	2	24,30,30	0.92±0.02	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	7MG	A	523	3	22,39,42	1.37±0.04	0±0 (0±0%)
3	GDP	A	524	2	23,47,47	1.18±0.08	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	7MG	A	523	3	-	0±0,7,37,38	0±0,3,3,3
3	GDP	A	524	2	-	0±0,12,32,32	0±0,3,3,3

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	A	523	7MG	C8-N9	5.04	1.38	1.45	7	1

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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