wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 10:21 pm BST

PDB ID	:	2K3R
Title	:	Pfu Rpp21 structure and assignments
Authors	:	Amero, C.D.; Foster, M.P.; Boomershine, W.P.; Xu, Y.
Deposited on	:	2008-05-15

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
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 (1)) were used in the production of this report:

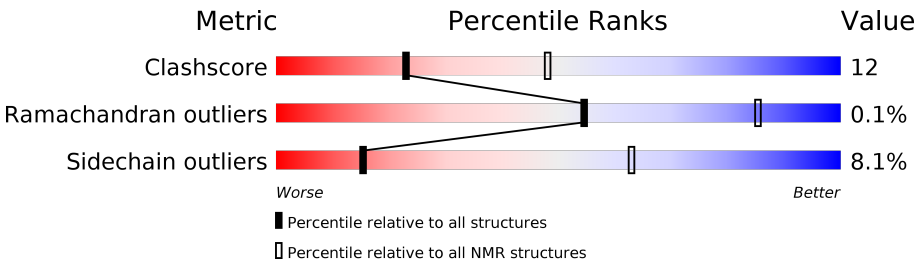
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 70%.

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	123	45% 22% 5% • 26%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:16-A:103 (88)	0.36	7

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 3 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1610 atoms, of which 839 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ribonuclease P protein component 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	91	1609	503	839	147	114	6	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	118	GLY	-	expression tag	UNP Q8U0H6
A	119	GLY	-	expression tag	UNP Q8U0H6
A	120	LEU	-	expression tag	UNP Q8U0H6
A	121	VAL	-	expression tag	UNP Q8U0H6
A	122	PRO	-	expression tag	UNP Q8U0H6
A	123	ARG	-	expression tag	UNP Q8U0H6

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

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: Ribonuclease P protein component 4

Chain A: 45% 22% 5% • 26%

MET ALA LYS TYR ASN GLU LYS LYS GLU LYS LYS ARG ILE ALA K15 F29 P30 Y31 S32 L35 A36 V40 E41 I54 W58 Y62 C63 C66 H67 A68 F69 L70 G73 I74 N75 A76 L80 K83 R84 M85 P86 H87 I88 V89 V90 K91 C92 L93 E94 C95

I98 Y101 I104 K105 GLU ILE LYS LYS ARG ARG LYS GLU LYS MET GLU TYR GLY GLY LEU VAL PRO ARG

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

The representative model is number 7. Colouring as in section 4.1 above.

- Molecule 1: Ribonuclease P protein component 4

Chain A: 43% 21% 7% • 26%

MET ALA LYS TYR ASN GLU LYS LYS GLU LYS LYS ARG ILE ALA K15 F22 E26 R27 Y28 F29 P30 L35 A36 E41 Q47 K53 I54 W58 K59 R60 R61 Y62 C66 H67 A68 F69 L70 I74 N75 A76 R77 L80 R81 D82 K83 R84 M85 P86 H87

V90 K91 C92 C95 I98 Y101 I104 K105 GLU ILE LYS LYS ARG ARG LYS GLU LYS MET GLU TYR GLY GLY LEU VAL PRO ARG

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 100 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
X-PLOR NIH	refinement	
X-PLOR NIH	structure solution	
X-PLOR NIH	geometry optimization	
NMRView	structure solution	
CARA	structure solution	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1003
Number of shifts mapped to atoms	1003
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	70%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	744	802	799	19±3
All	All	14900	16040	15980	383

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:40:VAL:HG11	1:A:67:HIS:O	0.83	1.72	14	4
1:A:66:CYS:SG	1:A:68:ALA:HB3	0.79	2.17	16	1
1:A:54:ILE:HG21	1:A:58:TRP:H	0.78	1.39	17	7
1:A:66:CYS:SG	1:A:95:CYS:N	0.77	2.56	2	10
1:A:65:LYS:HB3	1:A:95:CYS:SG	0.72	2.24	13	10

5.2 Torsion angles (i)

5.2.1 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	88/123 (72%)	85±1 (96±1%)	3±1 (4±1%)	0±0 (0±0%)	54	85
All	All	1760/2460 (72%)	1696 (96%)	62 (4%)	2 (0%)	54	85

All 2 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	82	GLN	1
1	A	55	PRO	1

5.2.2 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	80/111 (72%)	74±2 (92±2%)	6±2 (8±2%)	15	63
All	All	1600/2220 (72%)	1471 (92%)	129 (8%)	15	63

5 of 16 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	91	LYS	20
1	A	101	TYR	17
1	A	92	CYS	17
1	A	70	LEU	13
1	A	54	ILE	11

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 70% for the well-defined parts and 67% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 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	1003
Number of shifts mapped to atoms	1003
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	80	-0.69 ± 0.18	Should be applied
$^{13}C_{\beta}$	78	-0.21 ± 0.20	None needed (< 0.5 ppm)
$^{13}C'$	62	-0.10 ± 0.14	None needed (< 0.5 ppm)
^{15}N	71	0.25 ± 0.48	None needed (< 0.5 ppm)

6.1.3 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 70%, i.e. 878 atoms were assigned a chemical shift out of a possible 1254. 14 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	357/428 (83%)	148/170 (87%)	139/176 (79%)	70/82 (85%)
Sidechain	452/723 (63%)	261/432 (60%)	188/244 (77%)	3/47 (6%)

Continued on next page...

Continued from previous page...

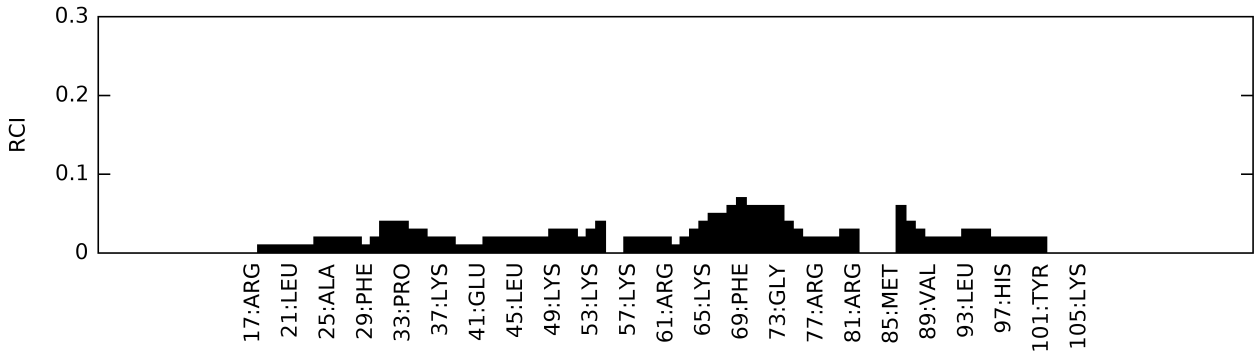
	Total	^{1}H	^{13}C	^{15}N
Aromatic	69/103 (67%)	35/53 (66%)	33/43 (77%)	1/7 (14%)
Overall	878/1254 (70%)	444/655 (68%)	360/463 (78%)	74/136 (54%)

6.1.4 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	22	PHE	HB2	0.44	4.85 – 1.15	-6.9

6.1.5 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: