



wwPDB NMR Structure Validation Summary Report ⓘ

May 28, 2020 – 10:36 pm BST

PDB ID : 2KM8
Title : Interdomain RRM packing contributes to RNA recognition in the rna15, hrp1, anchor RNA 3' processing ternary complex
Authors : Leeper, T.C.; Varani, G.
Deposited on : 2009-07-24

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

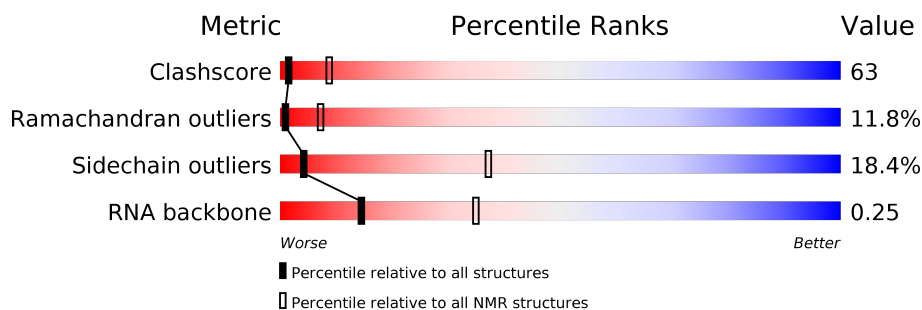
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 25%.

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
RNA backbone	4643	676

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	13	 77% 23%
2	B	84	 14% 65% 17% ••
3	C	167	 23% 62% 12% •

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 1 is the overall representative, medoid model (most similar to other models).

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	B:22-B:103, C:158-C:319 (244)	1.31	1

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. No single-model clusters were found.

Cluster number	Models
1	1, 3, 4, 6
2	7, 8
3	2, 5
4	9, 10

3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4375 atoms, of which 2102 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called 5'-R(P*UP*AP*UP*AP*UP*AP*UP*AP*AP*UP*AP*AP*U)-3'.


Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	13	414	124	139	47	91	13	0

- Molecule 2 is a protein called mRNA 3'-end-processing protein RNA15.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	84	1296	411	640	112	129	4	0

- Molecule 3 is a protein called Nuclear polyadenylated RNA-binding protein 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	C	167	2665	849	1323	232	256	5	0

Chain A: 

U1
A2
U3
A4
U5
A6
U7
A8
A9
U10
A11
A12
U13

- Molecule 2: mRNA 3'-end-processing protein RNA15

Chain B: 

R22
S25
R26
V27
Y29
L30
G31
S32
I33
P34
Y35
D36
Q37
T38
E39
E40
Q41
L42
L43
D44
L45
C46
S47
R48
V49
V52
I53
N54
L55
R56
M57
M58
F59
D60
P61
R65
S66
R67
G68
Y69
A70
F71
I72
E73
F74
R75
D76
L77
E78
A81
S82
A83
V84
R85
N86

L87
M88
G89
Y90
Q91
L92
G93
S94
R95
F96
L97
G100
Y101
S102
S103
N104
S105

- Molecule 3: Nuclear polyadenylated RNA-binding protein 4

Chain C: 

K156
E157
S158
C159
K160
M161
F162
G165
L166
M167
M168
D169
T170
T171
M174
L175
R176
E177
Y178
F179
G180
K181
Y182
D187
L188
K189
I190
M191
K192
D193
T196
S199
R200
G201
F202
L205
S206
K209
P210
S211
Q212
V213
D214
E215
V216
V217
H221
I222
L223
D224
G225
K226

Y227
I228
D229
R232
A233
D237
E238
Q239
T242
G243
K244
T245
F246
V247
G248
G249
V254
E258
F259
E260
E261
F262
F263
S264
Q265
W266
Q267
T268
I269
D271
A272
Q273
L274
M275
L276
D277
K278
D279
T280
G281
Q282
S283
R284
G285
F286
G287
F288
V289
T290
Y291
D292
R299
V300

C301
Q302
N303
K304
F308
K309
D310
R311
E314
I315
K316
R317
A318
E319
P320
R321
R322

5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 10 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	
CYANA	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	922
Number of shifts mapped to atoms	922
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	25%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 Too-close contacts

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	275	139	138	98±23
2	B	641	629	626	99±25
3	C	1295	1274	1273	134±14
All	All	22110	20420	20370	2660

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 63.

5 of 1467 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:A:H3'	2:B:69:TYR:N	1.31	1.37	7	1
1:A:6:A:H4'	1:A:7:U:H5'	1.13	1.17	9	7
1:A:12:A:H5'	2:B:67:LYS:O	1.09	1.47	7	1
1:A:12:A:C5'	2:B:67:LYS:O	1.07	2.00	7	1
1:A:10:U:H5'	2:B:71:PHE:CE1	1.03	1.86	5	4

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	
2	B	81/84 (96%)	56±2 (69±3%)	13±2 (16±3%)	12±2 (15±2%)	0	4
3	C	162/167 (97%)	119±3 (74±2%)	26±3 (16±2%)	17±2 (10±1%)	1	9
All	All	2430/2510 (97%)	1752 (72%)	392 (16%)	286 (12%)	1	7

5 of 83 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	77	LEU	10
2	B	26	ARG	10
2	B	91	GLN	10
2	B	76	ASP	10
2	B	47	SER	10

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	
2	B	72/74 (97%)	58±2 (80±3%)	14±2 (20±3%)	4	34
3	C	140/145 (97%)	115±3 (82±2%)	25±3 (18±2%)	4	39
All	All	2120/2190 (97%)	1730 (82%)	390 (18%)	4	37

5 of 129 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)
2	B	90	TYR	10
3	C	276	LEU	9
3	C	171	THR	9
2	B	37	GLN	9
3	C	222	ILE	9

5.2.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	A	12/13 (92%)	6±1 (46±11%)	1±1 (8±8%)	0.25±0.07
All	All	120/130 (92%)	55 (46%)	9 (8%)	0.25

The overall RNA backbone suiteness is 0.25.

5 of 11 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	3	U	8
1	A	13	U	7
1	A	4	A	6
1	A	10	U	6
1	A	12	A	6

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	A	5	U	3
1	A	6	A	3
1	A	10	U	3

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](#)

There are no ligands in this entry.

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 25% for the well-defined parts and 25% 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	922
Number of shifts mapped to atoms	922
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 \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	80	-0.13 ± 0.14	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	65	0.10 ± 0.15	None needed (< 0.5 ppm)
$^{13}\text{C}'$	77	-0.38 ± 0.22	None needed (< 0.5 ppm)
^{15}N	78	0.60 ± 0.76	None needed (imprecise)

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 25%, i.e. 812 atoms were assigned a chemical shift out of a possible 3265. 9 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	387/1198 (32%)	158/477 (33%)	153/488 (31%)	76/233 (33%)
Sidechain	392/1583 (25%)	264/933 (28%)	119/568 (21%)	9/82 (11%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	33/249 (13%)	33/133 (25%)	0/112 (0%)	0/4 (0%)
Overall	812/3265 (25%)	455/1674 (27%)	272/1259 (22%)	85/332 (26%)

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
2	B	23	PRO	HA	1.69	6.05 – 2.75	-8.2

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 B:

