wwPDB NMR Structure Validation Summary Report (i)

May 29, 2020 – 09:17 am BST

PDB ID	:	5ZUX
Title	:	Solution Structure of the DNA complex of the C-terminal Domain of Rok
Authors	:	Xia, B.; Duan, B.
Deposited on	:	2018-05-08

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:

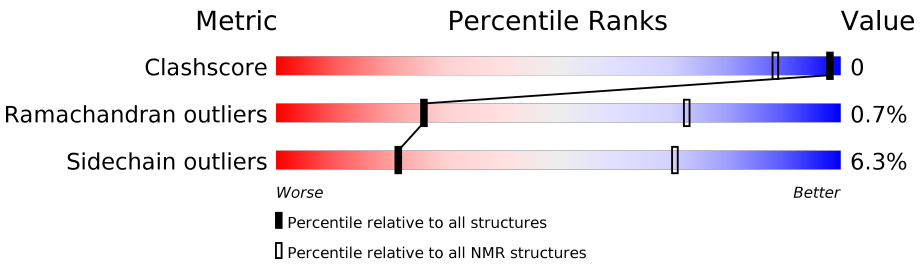
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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 68%.

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	92	75% • • 18%
2	B	18	100%
2	C	18	6% 83% 11%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 18 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:109-A:183 (75)	0.39	18

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 and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2663 atoms, of which 1185 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Rok.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	92	1517	463	771	145	135	3	0

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	18	573	178	207	65	106	17	0
2	C	18	573	178	207	65	106	17	0

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: Rok

Chain A: 75% • • 18%

G102 T103 T104 R105 R106 R107 R108 R112 R121 R130 T156 Y173 R174 E184 S185 L186 E187 H188 H189 H190 H191 H192 H193

• Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain B: 100%

C1 T2 A3 A4 T5 A6 A7 C8 T9 A10 G11 T12 T13 A14 T15 T16 A17 G18

• Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP*AP*G)-3')

Chain C: 6% 83% 11%

C19 T20 A21 A22 T23 A24 A25 C26 T27 A28 G29 T30 T31 A32 T33 T34 A35 G36

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

The representative model is number 18. Colouring as in section 4.1 above.

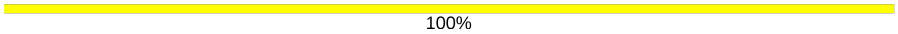
• Molecule 1: Rok

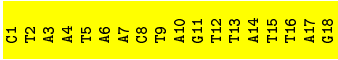
Chain A: 71% 10% • 18%

G102 T103 T104 R105 R106 R107 R108 R112 R121 E122 I125 R130 E143 L150 T156 Y173 R174 E184 S185 L186 E187 H188 H189 H190 H191 H192 H193

• Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP *AP*G)-3')

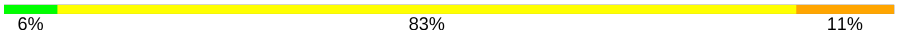
Chain B:





• Molecule 2: DNA (5'-D(*CP*TP*AP*AP*TP*AP*AP*CP*TP*AP*GP*TP*TP*AP*TP*TP *AP*G)-3')

Chain C:



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
AMBER	refinement	
CYANA	structure calculation	

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	1409
Number of shifts mapped to atoms	1409
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	68%

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	591	626	625	1±0
All	All	26460	20800	20780	14

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:173:TYR:CG	1:A:174:ARG:N	0.47	2.83	5	14

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	75/92 (82%)	71±1 (95±1%)	3±1 (4±1%)	1±0 (1±1%)	26	73
All	All	1500/1840 (82%)	1423 (95%)	66 (4%)	11 (1%)	26	73

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	174	ARG	9
1	A	153	PRO	2

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	63/79 (80%)	59±1 (94±2%)	4±1 (6±2%)	21	70
All	All	1260/1580 (80%)	1181 (94%)	79 (6%)	21	70

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	156	THR	20
1	A	112	ARG	17
1	A	121	ARG	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	143	GLU	5
1	A	122	GLU	5

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)

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 68% for the well-defined parts and 66% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *starch_output*

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	1409
Number of shifts mapped to atoms	1409
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	88	-0.41 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	81	-0.03 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	62	-0.22 ± 0.14	None needed (< 0.5 ppm)
^{15}N	78	0.08 ± 0.55	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 68%, i.e. 1130 atoms were assigned a chemical shift out of a possible 1656. 7 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	342/367 (93%)	146/146 (100%)	129/150 (86%)	67/71 (94%)
Sidechain	477/552 (86%)	298/328 (91%)	174/197 (88%)	5/27 (19%)

Continued on next page...

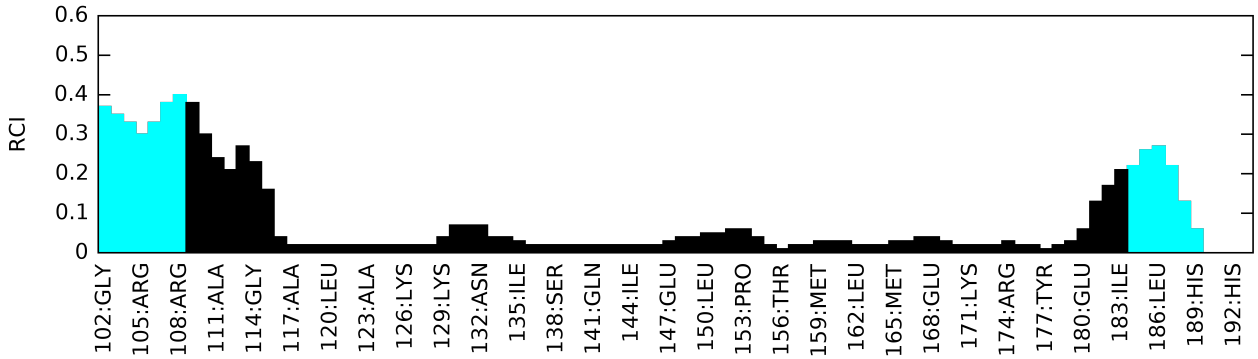
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	11/41 (27%)	11/21 (52%)	0/18 (0%)	0/2 (0%)
Overall	1130/1656 (68%)	755/903 (84%)	303/613 (49%)	72/140 (51%)

6.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: