W O R L D W I D E

wwPDB
PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report ⓘ

May 29, 2020 – 12:08 am BST

PDB ID	:	2N0Y
Title	:	NMR structure of the complex between the C-terminal domain of the Rift Valley fever virus protein NSs and the PH domain of the Tfb1 subunit of TFIIH
Authors	:	Cyr, N.; de la Fuente, C.; Lecoq, L.; Guendel, I.; Chabot, P.R.; Kehn-Hall, K.; Omichinski, J.G.
Deposited on	:	2015-03-18

The following versions of software and data (see references ⓘ) were used in the production of this report:

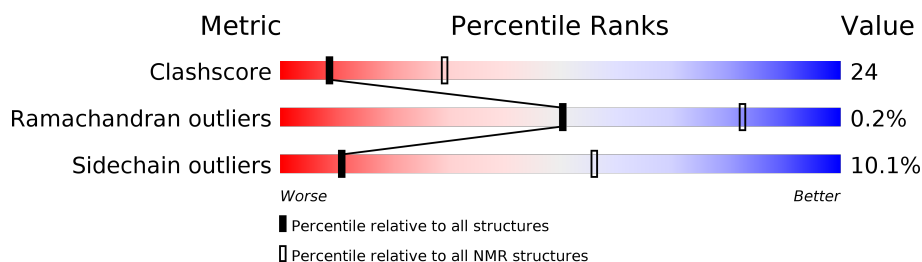|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 91%.

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	Percentile Ranks	Value
Clashscore		24
Ramachandran outliers		0.2%
Sidechain outliers		10.1%

Worse   *Better*

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	115	44% 32% • 21%
2	B	24	13% 88%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 12 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:3-A:66,　　　A:85-A:111, B:261-B:263 (94)	0.22	12

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2137 atoms, of which 1056 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called RNA polymerase II transcription factor B subunit 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	115	1821	559	919	161	177	5	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	PRO	MET	ENGINEERED MUTATION	UNP P32776

- Molecule 2 is a protein called Non-structural protein NS-S.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	24	316	104	137	24	50	1	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	242	GLY	-	EXPRESSION TAG	UNP P21698
B	243	GLY	-	EXPRESSION TAG	UNP P21698
B	244	GLY	-	EXPRESSION TAG	UNP P21698
B	245	GLY	-	EXPRESSION TAG	UNP P21698
B	246	TYR	-	EXPRESSION TAG	UNP P21698

4 Residue-property plots ⓘ

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: RNA polymerase II transcription factor B subunit 1
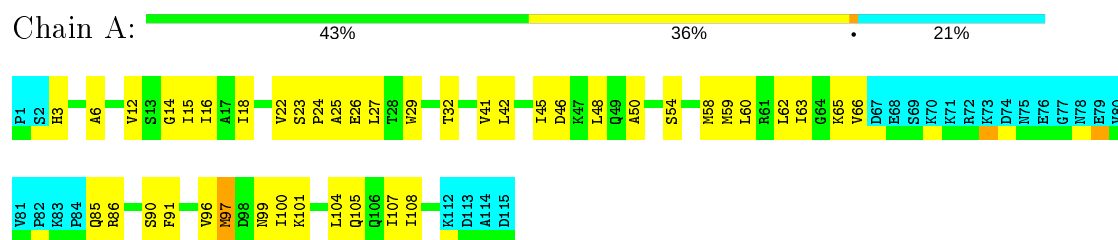

- Molecule 2: Non-structural protein NS-S

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

The representative model is number 12. Colouring as in section 4.1 above.

- Molecule 1: RNA polymerase II transcription factor B subunit 1

- Molecule 2: Non-structural protein NS-S

G242 G243 G244 G245 Y246 D247 V248 E249 M250 E251 S252 E253 E254 E255 S256 D257 D258 D259 G260 F261 V264 D265

5   Refinement protocol and experimental data overview (i)

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

Of the 260 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
CNS	refinement	

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	1693
Number of shifts mapped to atoms	1693
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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	714	736	733	35±4
2	B	27	24	24	1±1
All	All	14820	15200	15140	705

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:48:LEU:HD23	1:A:105:GLN:HG3	0.98	1.36	12	16
1:A:48:LEU:HD22	1:A:104:LEU:HB2	0.84	1.50	3	20
1:A:60:LEU:HB2	1:A:91:PHE:CE1	0.81	2.10	4	10
1:A:22:VAL:HG11	1:A:26:GLU:HB2	0.78	1.54	14	10
1:A:48:LEU:HD23	1:A:105:GLN:HG2	0.77	1.52	16	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	
1	A	91/115 (79%)	84±1 (93±1%)	7±1 (7±1%)	0±0 (0±0%)	50	82
2	B	3/24 (12%)	3±0 (100±0%)	0±0 (0±0%)	0±0 (0±0%)	100	100
All	All	1880/2780 (68%)	1745 (93%)	131 (7%)	4 (0%)	50	82

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	25	ALA	4

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	81/103 (79%)	73±1 (90±2%)	8±1 (10±2%)	11	56
2	B	3/19 (16%)	3±0 (90±15%)	0±0 (10±15%)	11	56
All	All	1680/2440 (69%)	1511 (90%)	169 (10%)	11	56

5 of 26 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	16	ILE	20
1	A	97	MET	20
1	A	12	VAL	19
1	A	58	MET	19
1	A	85	GLN	15

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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 91% for the well-defined parts and 91% 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 ⓘ

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	1693
Number of shifts mapped to atoms	1693
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

6.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	139	-0.12 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_\beta$	129	-0.10 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	133	0.13 ± 0.09	None needed (< 0.5 ppm)
^{15}N	132	-0.25 ± 0.28	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 91%, i.e. 1072 atoms were assigned a chemical shift out of a possible 1172. 14 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	462/466 (99%)	185/186 (99%)	186/188 (99%)	91/92 (99%)
Sidechain	542/626 (87%)	332/365 (91%)	202/232 (87%)	8/29 (28%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	68/80 (85%)	36/42 (86%)	31/31 (100%)	1/7 (14%)
Overall	1072/1172 (91%)	553/593 (93%)	419/451 (93%)	100/128 (78%)

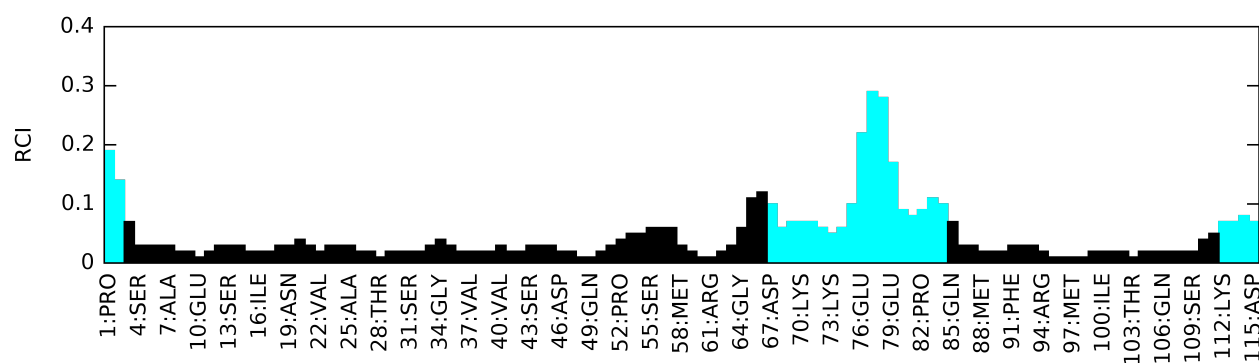### 6.1.4　Statistically unusual chemical shifts (i)

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	10	GLU	HG2	0.62	3.33 − 1.23	-7.9
1	A	10	GLU	HG3	1.05	3.31 − 1.21	-5.7
1	A	9	PHE	HB2	1.12	4.85 − 1.15	-5.1

6.1.5　Random Coil Index (RCI) plots (i)

The images below report *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:

Random coil index (RCI) for chain B: