wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 04:30 pm BST

PDB ID	:	5Y6J
Title	:	Structure of Tomato spotted wilt virus nucleocapsid protein with alternative oligomerization state
Authors	:	Guo, Y.; Dong, S.; Lou, Z.
Deposited on	:	2017-08-12
Resolution	:	2.81 Å(reported)

This is a wwPDB X-ray 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/XrayValidationReportHelp
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:

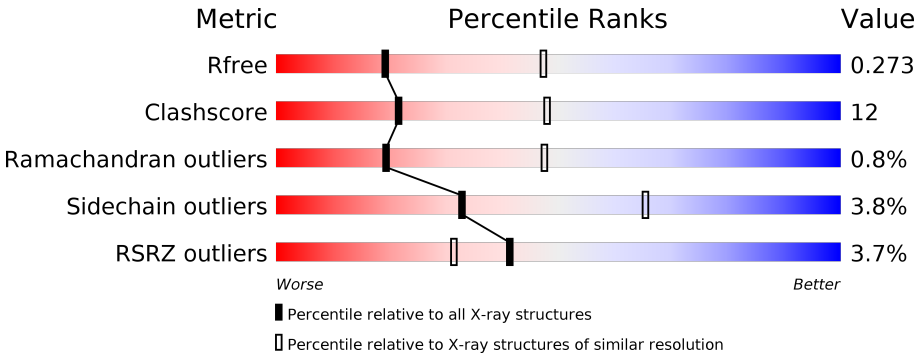
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.81 Å.

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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3617 (2.84-2.80)
Clashscore	141614	4060 (2.84-2.80)
Ramachandran outliers	138981	3978 (2.84-2.80)
Sidechain outliers	138945	3980 (2.84-2.80)
RSRZ outliers	127900	3552 (2.84-2.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	258	2% 71% 21% 6%
1	B	258	4% 69% 23% 5%
1	C	258	5% 74% 20%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5762 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	243	Total	C	N	O	S	0	0	0
			1896	1213	314	358	11			
1	B	246	Total	C	N	O	S	0	0	0
			1927	1229	317	370	11			
1	C	247	Total	C	N	O	S	0	0	0
			1916	1222	318	366	10			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	ALA	ARG	engineered mutation	UNP H6UMW9
A	95	ALA	ARG	engineered mutation	UNP H6UMW9
B	94	ALA	ARG	engineered mutation	UNP H6UMW9
B	95	ALA	ARG	engineered mutation	UNP H6UMW9
C	94	ALA	ARG	engineered mutation	UNP H6UMW9
C	95	ALA	ARG	engineered mutation	UNP H6UMW9

- Molecule 2 is water.

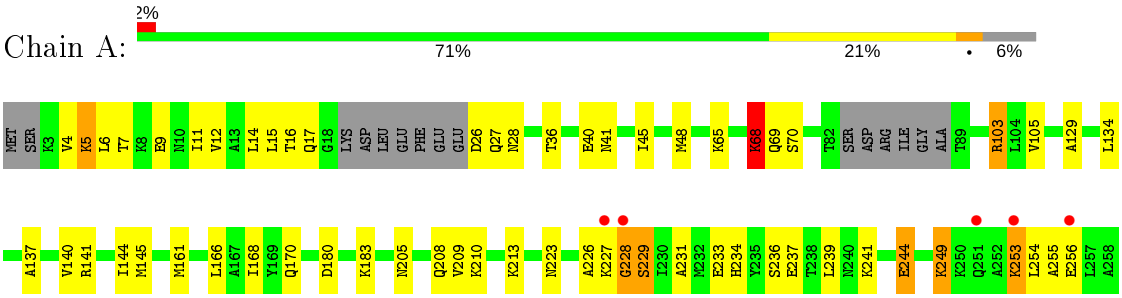
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	11	Total	O	0	0
			11	11		
2	B	7	Total	O	0	0
			7	7		
2	C	5	Total	O	0	0
			5	5		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

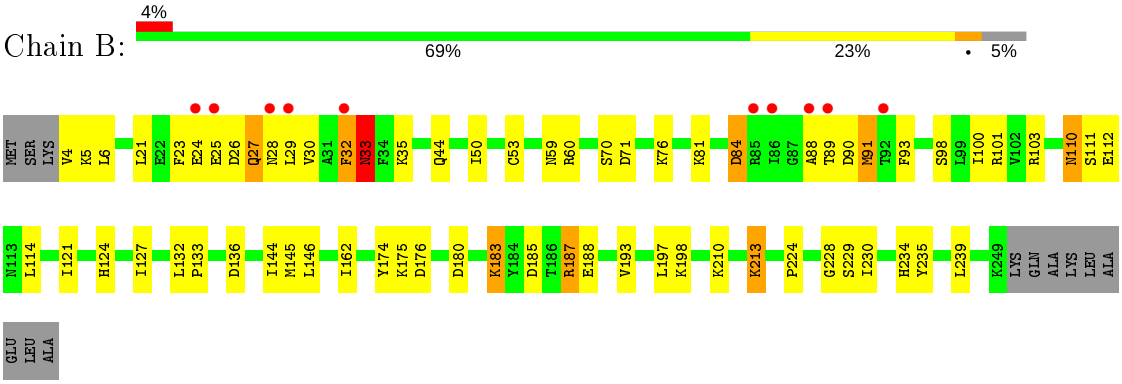
- Molecule 1: Nucleoprotein

Chain A:



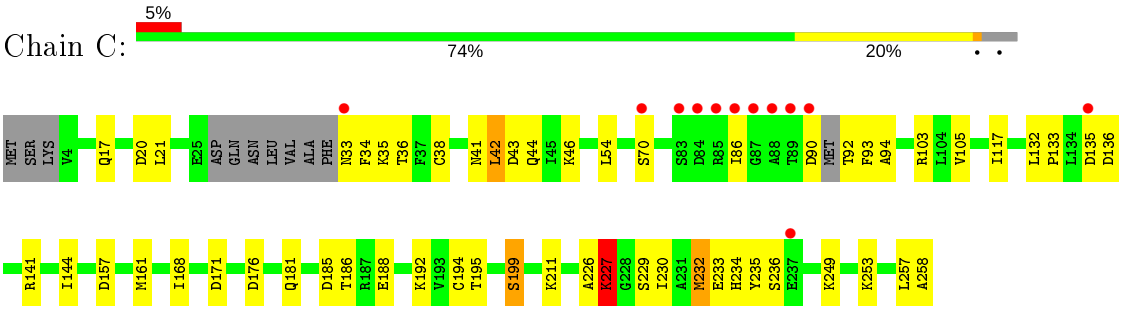
- Molecule 1: Nucleoprotein

Chain B:



- Molecule 1: Nucleoprotein

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.36Å 76.51Å 98.44Å 90.00° 93.55° 90.00°	Depositor
Resolution (Å)	49.49 – 2.81 49.49 – 2.81	Depositor EDS
% Data completeness (in resolution range)	98.3 (49.49-2.81) 98.3 (49.49-2.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.45 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.232 , 0.275 0.231 , 0.273	Depositor DCC
R_{free} test set	1006 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	42.2	Xtriage
Anisotropy	0.094	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 44.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	5762	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.26% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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 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	0.28	0/1917	0.58	4/2567 (0.2%)
1	B	0.26	0/1951	0.52	3/2618 (0.1%)
1	C	0.46	0/1937	0.55	3/2596 (0.1%)
All	All	0.35	0/5805	0.55	10/7781 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	134	LEU	CB-CA-C	-8.69	93.69	110.20
1	B	89	THR	N-CA-C	-6.89	92.39	111.00
1	B	91	MET	CB-CA-C	6.55	123.50	110.40
1	A	228	GLY	N-CA-C	-6.50	96.86	113.10
1	A	134	LEU	N-CA-C	6.35	128.15	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	229	SER	Peptide
1	A	253	LYS	Peptide

5.2 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 the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1896	0	1986	53	0
1	B	1927	0	1997	49	0
1	C	1916	0	1985	45	0
2	A	11	0	0	0	0
2	B	7	0	0	0	0
2	C	5	0	0	0	0
All	All	5762	0	5968	139	0

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.

The worst 5 of 139 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:253:LYS:CE	1:A:255:ALA:HB2	1.59	1.25
1:A:253:LYS:HE3	1:A:255:ALA:CB	1.64	1.22
1:C:227:LYS:H	1:C:230:ILE:HD13	1.11	1.09
1:A:253:LYS:HG3	1:A:255:ALA:H	1.24	1.00
1:C:227:LYS:N	1:C:230:ILE:HD13	1.78	0.98

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

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	237/258 (92%)	215 (91%)	19 (8%)	3 (1%)	12 34

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	244/258 (95%)	218 (89%)	24 (10%)	2 (1%)	19 47
1	C	241/258 (93%)	226 (94%)	14 (6%)	1 (0%)	34 64
All	All	722/774 (93%)	659 (91%)	57 (8%)	6 (1%)	19 47

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	68	LYS
1	C	227	LYS
1	B	27	GLN
1	B	33	ASN
1	A	27	GLN

5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	211/225 (94%)	205 (97%)	6 (3%)	43 76
1	B	216/225 (96%)	206 (95%)	10 (5%)	27 58
1	C	211/225 (94%)	203 (96%)	8 (4%)	33 65
All	All	638/675 (94%)	614 (96%)	24 (4%)	33 65

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	103	ARG
1	B	183	LYS
1	C	249	LYS
1	B	110	ASN
1	B	136	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	44	GLN
1	C	234	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	243/258 (94%)	-0.15	5 (2%)	63	54	15, 33, 71, 112	0
1	B	246/258 (95%)	-0.08	10 (4%)	37	27	19, 37, 90, 150	0
1	C	247/258 (95%)	0.10	12 (4%)	29	20	22, 41, 83, 169	0
All	All	736/774 (95%)	-0.04	27 (3%)	41	31	15, 37, 83, 169	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	87	GLY	11.6
1	C	90	ASP	8.6
1	C	88	ALA	6.0
1	B	92	THR	4.2
1	C	83	SER	3.9

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.