wwPDB X-ray Structure Validation Summary Report (i)

Aug 23, 2016 – 02:29 PM EDT

PDB ID	:	4YPI
Title	:	Structure of Ebola virus nucleoprotein N-terminal fragment bound to a peptide derived from Ebola VP35
Authors	:	Leung, D.W.; Borek, D.M.; Binning, J.M.; Otwinowski, Z.; Amarasinghe, G.K.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on	:	2015-03-13
Resolution	:	3.71 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
http://wwpdb.org/validation/2016/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:

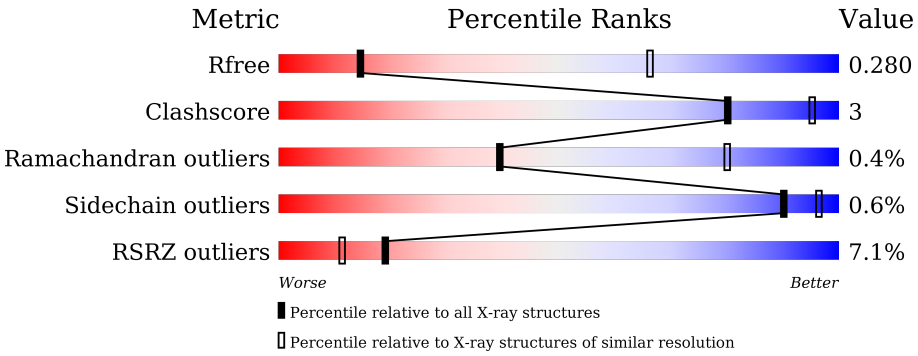
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 3.71 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}	91344	1129 (3.94-3.50)
Clashscore	102246	1252 (3.94-3.50)
Ramachandran outliers	100387	1199 (3.94-3.50)
Sidechain outliers	100360	1197 (3.94-3.50)
RSRZ outliers	91569	1136 (3.94-3.50)

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. 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	348	2%, 91%, 8%
1	B	348	9%, 92%, 7%
1	C	348	93%, 6%
1	D	348	20%, 92%, 6%
2	E	28	75%, 21%
2	F	28	75%, 25%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	28	
2	H	28	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11780 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	C	348	Total	C	N	O	S	0	0	0
			2739	1741	486	501	11			
1	A	346	Total	C	N	O	S	0	0	0
			2721	1731	483	496	11			
1	B	347	Total	C	N	O	S	0	0	0
			2730	1736	484	499	11			
1	D	347	Total	C	N	O	S	0	0	0
			2730	1736	484	499	11			

- Molecule 2 is a protein called Polymerase cofactor VP35.

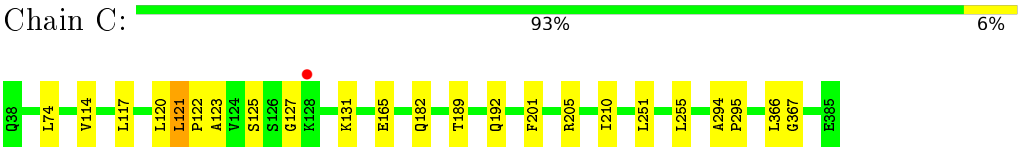
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	28	Total	C	N	O	S	0	0	0
			215	137	33	42	3			
2	E	28	Total	C	N	O	S	0	0	0
			215	137	33	42	3			
2	F	28	Total	C	N	O	S	0	0	0
			215	137	33	42	3			
2	H	28	Total	C	N	O	S	0	0	0
			215	137	33	42	3			

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 errors 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.

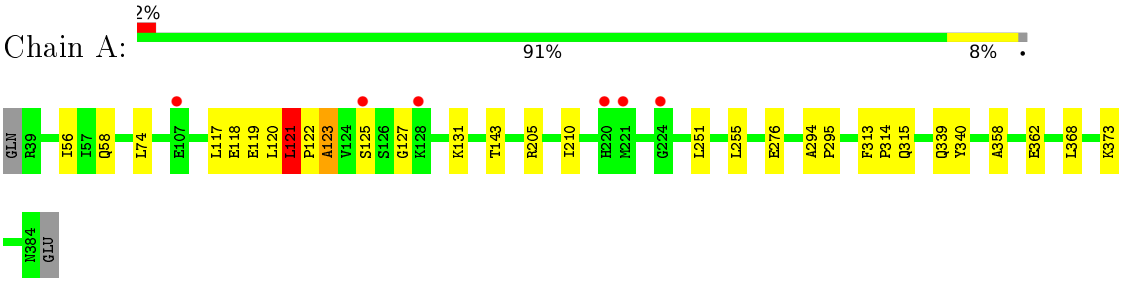
- Molecule 1: Nucleoprotein

Chain C:



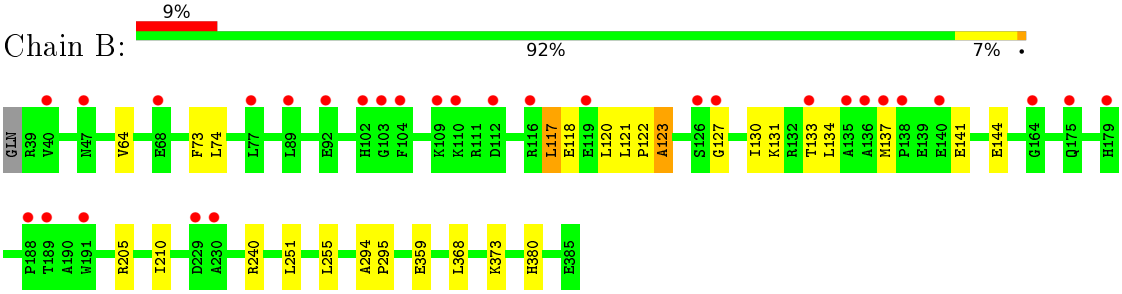
- Molecule 1: Nucleoprotein

Chain A:



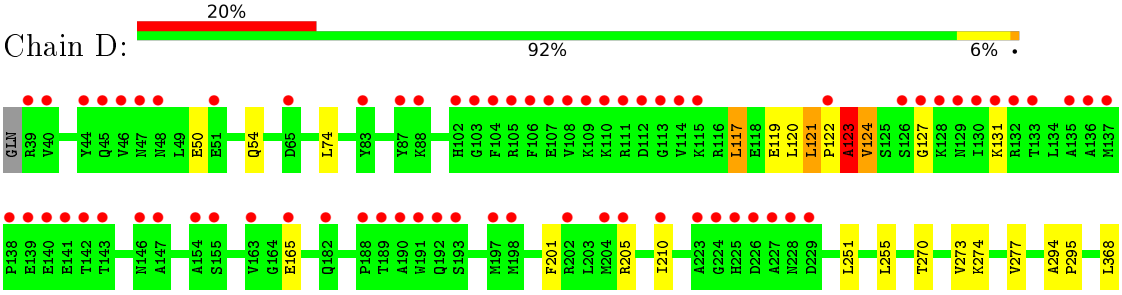
- Molecule 1: Nucleoprotein

Chain B:



- Molecule 1: Nucleoprotein

Chain D:

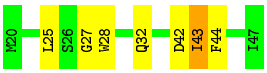


• Molecule 2: Polymerase cofactor VP35

Chain G:

• Molecule 2: Polymerase cofactor VP35

Chain E:

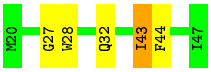


• Molecule 2: Polymerase cofactor VP35

Chain F:

• Molecule 2: Polymerase cofactor VP35

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	F 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	149.75Å 194.82Å 347.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 3.71 24.96 – 3.71	Depositor EDS
% Data completeness (in resolution range)	84.8 (25.00-3.71) 85.3 (24.96-3.71)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	3.80 (at 3.74Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.253 , 0.285 0.251 , 0.280	Depositor DCC
R_{free} test set	1163 reflections (5.31%)	DCC
Wilson B-factor (Å^2)	96.0	Xtriage
Anisotropy	0.165	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 85.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	11780	wwPDB-VP
Average B, all atoms (Å^2)	154.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% 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	0.34	0/2770	0.65	3/3734 (0.1%)
1	B	0.33	0/2779	0.52	1/3746 (0.0%)
1	C	0.34	0/2788	0.54	2/3758 (0.1%)
1	D	0.34	0/2779	0.54	2/3746 (0.1%)
2	E	0.33	0/220	0.54	0/297
2	F	0.36	0/220	0.60	0/297
2	G	0.35	0/220	0.56	0/297
2	H	0.36	0/220	0.55	0/297
All	All	0.34	0/11996	0.56	8/16172 (0.0%)

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	1
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	4

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	205	ARG	NE-CZ-NH2	15.40	128.00	120.30
1	A	205	ARG	NE-CZ-NH1	-14.63	112.98	120.30
1	A	121	LEU	CA-CB-CG	7.04	131.48	115.30
1	D	124	VAL	N-CA-C	6.06	127.36	111.00
1	C	205	ARG	NE-CZ-NH2	-5.69	117.45	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	123	ALA	Peptide
1	B	123	ALA	Peptide
1	C	123	ALA	Peptide
1	D	123	ALA	Peptide

5.2 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 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	2721	0	2752	21	0
1	B	2730	0	2758	23	1
1	C	2739	0	2766	11	0
1	D	2730	0	2758	13	6
2	E	215	0	208	4	0
2	F	215	0	208	5	0
2	G	215	0	209	2	1
2	H	215	0	209	3	0
All	All	11780	0	11868	80	8

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 3.

The worst 5 of 80 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:D:127:GLY:O	1:D:131:LYS:HG2	1.76	0.85
1:A:276:GLU:OE1	1:A:315:GLN:HB2	1.77	0.84
1:A:118:GLU:O	1:A:131:LYS:HE3	1.87	0.74
1:B:359:GLU:OE1	1:B:380:HIS:CG	2.42	0.73
1:A:276:GLU:OE2	1:A:313:PHE:HA	1.95	0.66

The worst 5 of 8 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:123:ALA:O	1:D:123:ALA:O[8_544]	1.45	0.75
1:D:123:ALA:C	1:D:123:ALA:O[8_544]	1.59	0.61
1:D:123:ALA:N	1:D:123:ALA:O[8_544]	1.87	0.33
1:B:141:GLU:OE1	1:B:141:GLU:OE1[4_554]	1.88	0.32
1:D:124:VAL:CG1	1:D:127:GLY:CA[8_544]	1.98	0.22

5.3 Torsion angles (i)

5.3.1 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 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	344/348 (99%)	314 (91%)	28 (8%)	2 (1%)	30 75
1	B	345/348 (99%)	313 (91%)	32 (9%)	0	100 100
1	C	346/348 (99%)	313 (90%)	31 (9%)	2 (1%)	30 75
1	D	345/348 (99%)	315 (91%)	28 (8%)	2 (1%)	30 75
2	E	26/28 (93%)	25 (96%)	1 (4%)	0	100 100
2	F	26/28 (93%)	25 (96%)	1 (4%)	0	100 100
2	G	26/28 (93%)	25 (96%)	1 (4%)	0	100 100
2	H	26/28 (93%)	24 (92%)	2 (8%)	0	100 100
All	All	1484/1504 (99%)	1354 (91%)	124 (8%)	6 (0%)	39 80

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	123	ALA
1	C	125	SER
1	A	125	SER
1	C	121	LEU
1	A	121	LEU

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	288/290 (99%)	287 (100%)	1 (0%)	94 98
1	B	289/290 (100%)	288 (100%)	1 (0%)	94 98
1	C	290/290 (100%)	288 (99%)	2 (1%)	88 95
1	D	289/290 (100%)	288 (100%)	1 (0%)	94 98
2	E	25/25 (100%)	23 (92%)	2 (8%)	15 55
2	F	25/25 (100%)	25 (100%)	0	100 100
2	G	25/25 (100%)	25 (100%)	0	100 100
2	H	25/25 (100%)	24 (96%)	1 (4%)	38 75
All	All	1256/1260 (100%)	1248 (99%)	8 (1%)	90 96

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	25	LEU
2	H	43	ILE
1	B	117	LEU
1	A	121	LEU
2	E	43	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	327	HIS
1	D	327	HIS
1	B	342	GLN
1	A	129	ASN
2	F	32	GLN

5.3.3 RNA (i)

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	346/348 (99%)	-0.29	6 (1%) 73 59	39, 116, 203, 231	0
1	B	347/348 (99%)	0.45	30 (8%) 13 9	53, 211, 279, 299	0
1	C	348/348 (100%)	-0.47	1 (0%) 94 90	30, 80, 157, 194	0
1	D	347/348 (99%)	0.74	70 (20%) 1 1	80, 255, 359, 554	0
2	E	28/28 (100%)	-0.52	0 100 100	62, 87, 123, 138	0
2	F	28/28 (100%)	-0.54	0 100 100	73, 98, 132, 133	0
2	G	28/28 (100%)	-0.63	0 100 100	43, 61, 107, 113	0
2	H	28/28 (100%)	-0.38	0 100 100	94, 131, 148, 151	0
All	All	1500/1504 (99%)	0.06	107 (7%) 19 11	30, 129, 299, 554	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	136	ALA	7.2
1	D	113	GLY	6.9
1	D	112	ASP	6.8
1	D	138	PRO	6.7
1	D	140	GLU	6.1

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

There are no such residues in this entry.