



wwPDB X-ray Structure Validation Summary Report

Jan 13, 2024 – 11:49 pm GMT

PDB ID : 6Y9I
Title : Structure of apo Chimpanzee Polyomavirus VP1
Authors : Stroh, L.J.; Rustmeier, N.H.; Stehle, T.
Deposited on : 2020-03-09
Resolution : 1.90 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  symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

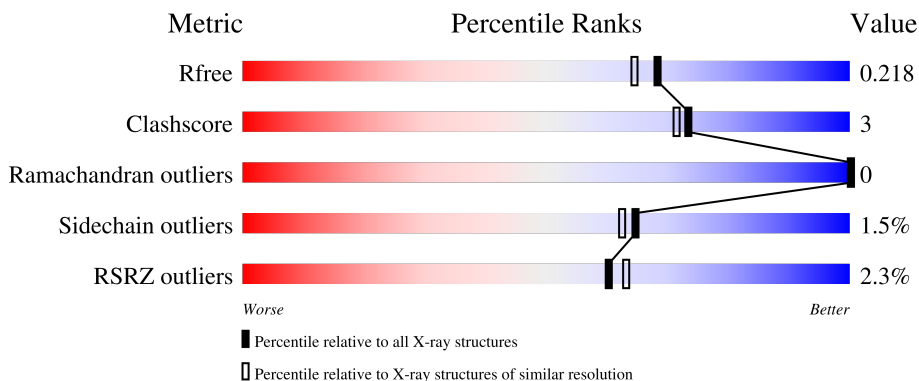
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 of 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 $\leq 5\%$. 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	AAA	292	 2% 89% 8% .
1	BBB	292	 2% 89% 7% .
1	CCC	292	 2% 88% 8% . .
1	DDD	292	 3% 87% 10% .
1	EEE	292	 % 88% 8% . .

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 11575 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 Major Capsid Protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	282	2130	1357	364	397	12	0	2	0
1	BBB	281	2134	1356	365	400	13	0	3	0
1	CCC	280	2109	1341	359	396	13	0	1	0
1	DDD	281	2123	1349	363	399	12	0	2	0
1	EEE	280	2116	1347	362	395	12	0	1	0

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	1	Total 1	Mg 1	0	0
2	BBB	2	Total 2	Mg 2	0	0
2	CCC	2	Total 2	Mg 2	0	0
2	DDD	2	Total 2	Mg 2	0	0
2	EEE	3	Total 3	Mg 3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	209	Total 209	O 209	0	0
3	BBB	206	Total 209	O 209	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	CCC	168	Total 168	O 168	0	0
3	DDD	176	Total 178	O 178	0	2
3	EEE	188	Total 189	O 189	0	1

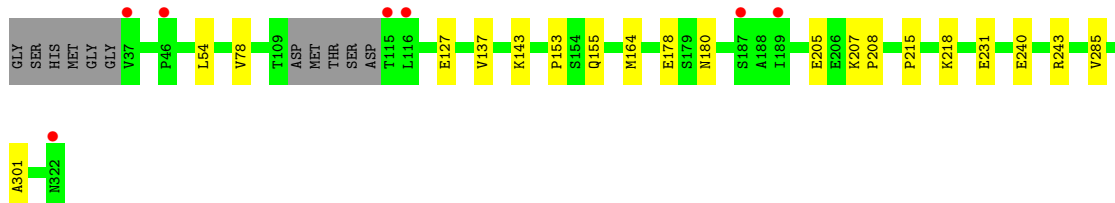
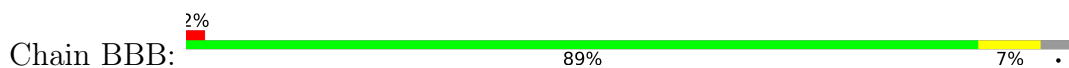
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

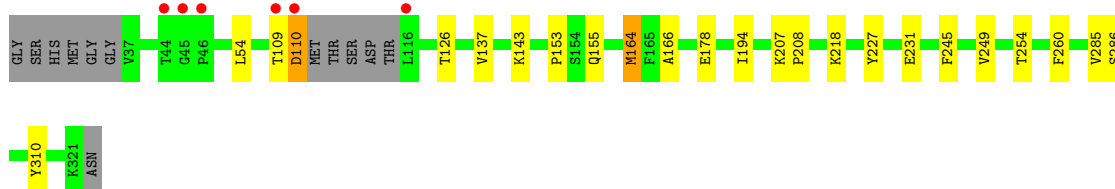
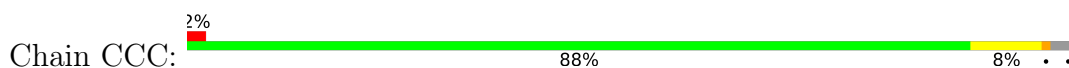
- Molecule 1: Major Capsid Protein VP1



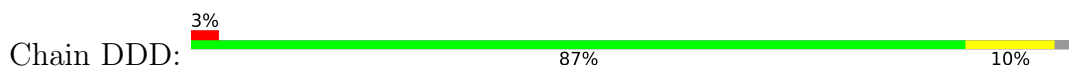
- Molecule 1: Major Capsid Protein VP1

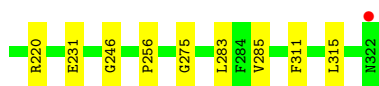


- Molecule 1: Major Capsid Protein VP1

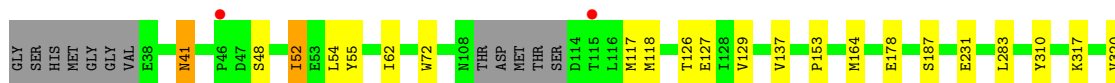
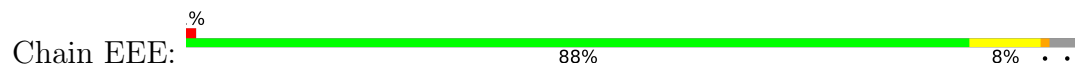


- Molecule 1: Major Capsid Protein VP1





- Molecule 1: Major Capsid Protein VP1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	63.41Å 82.19Å 82.66Å 68.61° 77.36° 77.49°	Depositor
Resolution (Å)	46.46 – 1.90 46.46 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.6 (46.46-1.90) 99.6 (46.46-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.16 (at 1.90Å)	Xtrriage
Refinement program	REFMAC 5.8.0253	Depositor
R, R_{free}	0.171 , 0.213 0.178 , 0.218	Depositor DCC
R_{free} test set	5886 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	25.0	Xtrriage
Anisotropy	0.072	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 52.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.018 for -h,-l,-k	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	11575	wwPDB-VP
Average B, all atoms (Å ²)	31.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.28% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	AAA	0.69	0/2177	0.84	0/2961
1	BBB	0.72	1/2181 (0.0%)	0.85	2/2966 (0.1%)
1	CCC	0.70	0/2156	0.85	0/2931
1	DDD	0.69	0/2170	0.86	0/2952
1	EEE	0.70	0/2163	0.85	0/2940
All	All	0.70	1/10847 (0.0%)	0.85	2/14750 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	BBB	240	GLU	CD-OE1	5.22	1.31	1.25

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	BBB	243	ARG	NE-CZ-NH1	5.37	122.98	120.30
1	BBB	243	ARG	NE-CZ-NH2	-5.26	117.67	120.30

There are no chirality outliers.

There are no planarity outliers.

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	AAA	2130	0	2078	15	0
1	BBB	2134	0	2080	11	0
1	CCC	2109	0	2037	13	0
1	DDD	2123	0	2047	15	0
1	EEE	2116	0	2064	17	0
2	AAA	1	0	0	0	0
2	BBB	2	0	0	0	0
2	CCC	2	0	0	0	0
2	DDD	2	0	0	0	0
2	EEE	3	0	0	0	0
3	AAA	209	0	0	2	0
3	BBB	209	0	0	2	0
3	CCC	168	0	0	0	0
3	DDD	178	0	0	1	0
3	EEE	189	0	0	1	0
All	All	11575	0	10306	68	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 3.

The worst 5 of 68 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:EEE:52:ILE:HD11	1:EEE:283:LEU:CD2	1.95	0.95
1:EEE:52:ILE:HD11	1:EEE:283:LEU:HD22	1.55	0.85
1:EEE:52:ILE:HD11	1:EEE:283:LEU:HD23	1.65	0.78
1:AAA:140[B]:LEU:HD23	1:AAA:141:GLU:HG3	1.69	0.75
1:CCC:137:VAL:HB	1:CCC:153:PRO:HB2	1.75	0.68

There are no symmetry-related clashes.

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	AAA	280/292 (96%)	268 (96%)	12 (4%)	0	100	100
1	BBB	280/292 (96%)	271 (97%)	9 (3%)	0	100	100
1	CCC	277/292 (95%)	266 (96%)	11 (4%)	0	100	100
1	DDD	279/292 (96%)	268 (96%)	11 (4%)	0	100	100
1	EEE	277/292 (95%)	267 (96%)	10 (4%)	0	100	100
All	All	1393/1460 (95%)	1340 (96%)	53 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	219/244 (90%)	215 (98%)	4 (2%)	59	55
1	BBB	222/244 (91%)	220 (99%)	2 (1%)	78	79
1	CCC	215/244 (88%)	213 (99%)	2 (1%)	78	79
1	DDD	216/244 (88%)	214 (99%)	2 (1%)	78	79
1	EEE	218/244 (89%)	212 (97%)	6 (3%)	43	36
All	All	1090/1220 (89%)	1074 (98%)	16 (2%)	65	62

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	EEE	164	MET
1	EEE	127	GLU
1	DDD	64	SER
1	EEE	54	LEU
1	CCC	164	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 10 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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, 95th 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	AAA	282/292 (96%)	-0.22	6 (2%) 63 66	20, 27, 46, 76	0
1	BBB	281/292 (96%)	-0.22	7 (2%) 57 60	19, 27, 45, 69	0
1	CCC	280/292 (95%)	-0.23	6 (2%) 63 66	20, 28, 47, 73	0
1	DDD	281/292 (96%)	-0.01	10 (3%) 42 45	20, 30, 54, 78	0
1	EEE	280/292 (95%)	-0.13	3 (1%) 80 82	21, 29, 47, 76	0
All	All	1404/1460 (96%)	-0.16	32 (2%) 60 63	19, 28, 48, 78	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	DDD	322	ASN	6.5
1	BBB	322	ASN	5.1
1	AAA	46	PRO	5.0
1	CCC	46	PRO	4.5
1	AAA	109	THR	4.5

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	MG	EEE	401	1/1	0.88	0.19	56,56,56,56	0
2	MG	DDD	402	1/1	0.96	0.27	44,44,44,44	0
2	MG	CCC	402	1/1	0.96	0.11	43,43,43,43	0
2	MG	BBB	401	1/1	0.97	0.25	50,50,50,50	0
2	MG	DDD	401	1/1	0.97	0.17	48,48,48,48	0
2	MG	EEE	402	1/1	0.97	0.15	53,53,53,53	0
2	MG	CCC	401	1/1	0.99	0.15	44,44,44,44	0
2	MG	AAA	401	1/1	0.99	0.27	39,39,39,39	0
2	MG	BBB	402	1/1	0.99	0.33	45,45,45,45	0
2	MG	EEE	403	1/1	0.99	0.20	39,39,39,39	0

6.5 Other polymers [i](#)

There are no such residues in this entry.