wwPDB X-ray Structure Validation Summary Report (i)

Nov 21, 2023 – 08:06 AM JST

PDB ID	:	7F8I
Title	:	Crystal structure of HPV6 L1 pentamer
Authors	:	Wang, Z.P.; Wang, D.N.; Gu, Y.; Li, S.W.
Deposited on	:	2021-07-02
Resolution	:	3.37 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

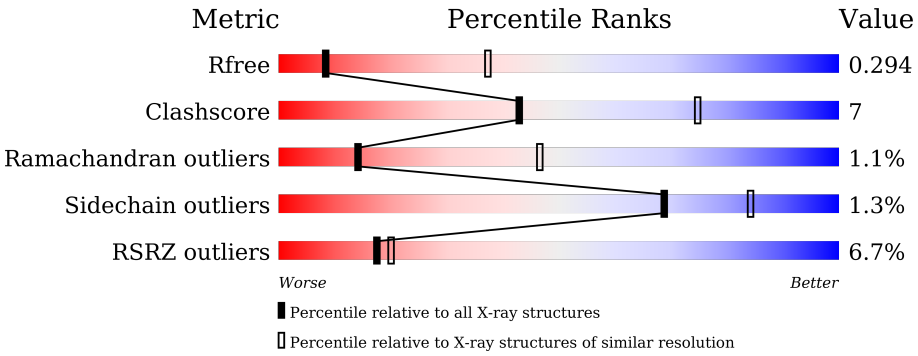
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.37 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	1558 (3.42-3.30)
Clashscore	141614	1627 (3.42-3.30)
Ramachandran outliers	138981	1599 (3.42-3.30)
Sidechain outliers	138945	1598 (3.42-3.30)
RSRZ outliers	127900	1507 (3.42-3.30)

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 <=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	496	3% 68% 15% • 16%
1	B	496	3% 70% 12% • 17%
1	C	496	6% 68% 15% • 16%
1	D	496	3% 69% 14% 17%
1	E	496	4% 67% 16% • 16%
1	F	496	5% 67% 14% • 18%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	496	7% 70% 14% 16%
1	H	496	7% 68% 15% 16%
1	I	496	9% 66% 15% 18%
1	J	496	7% 68% 15% 16%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 32368 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 L1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	B	410	Total	C	N	O	S	0	0	0
			3213	2042	540	612	19			
1	C	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	D	410	Total	C	N	O	S	0	0	0
			3216	2044	541	613	18			
1	E	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	F	407	Total	C	N	O	S	0	0	0
			3194	2031	536	609	18			
1	G	418	Total	C	N	O	S	0	0	0
			3274	2080	550	625	19			
1	H	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	I	405	Total	C	N	O	S	0	0	0
			3181	2025	534	604	18			
1	J	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	initiating methionine	UNP Q9W9C6
A	376	VAL	MET	conflict	UNP Q9W9C6
B	-2	MET	-	initiating methionine	UNP Q9W9C6
B	376	VAL	MET	conflict	UNP Q9W9C6
C	-2	MET	-	initiating methionine	UNP Q9W9C6
C	376	VAL	MET	conflict	UNP Q9W9C6
D	-2	MET	-	initiating methionine	UNP Q9W9C6
D	376	VAL	MET	conflict	UNP Q9W9C6
E	-2	MET	-	initiating methionine	UNP Q9W9C6

Continued on next page...

Continued from previous page...

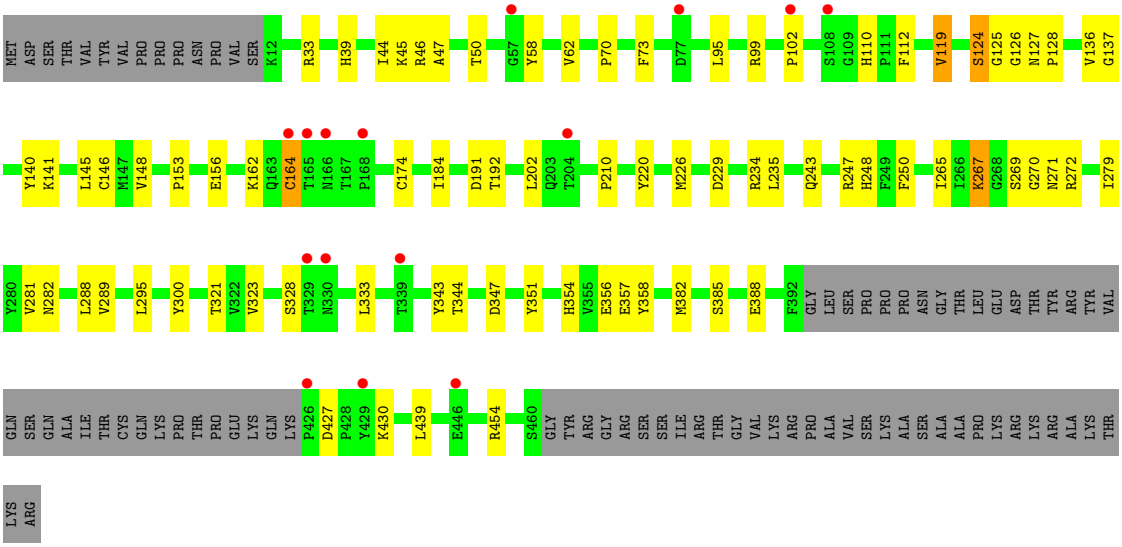
Chain	Residue	Modelled	Actual	Comment	Reference
E	376	VAL	MET	conflict	UNP Q9W9C6
F	-2	MET	-	initiating methionine	UNP Q9W9C6
F	376	VAL	MET	conflict	UNP Q9W9C6
G	-2	MET	-	initiating methionine	UNP Q9W9C6
G	376	VAL	MET	conflict	UNP Q9W9C6
H	-2	MET	-	initiating methionine	UNP Q9W9C6
H	376	VAL	MET	conflict	UNP Q9W9C6
I	-2	MET	-	initiating methionine	UNP Q9W9C6
I	376	VAL	MET	conflict	UNP Q9W9C6
J	-2	MET	-	initiating methionine	UNP Q9W9C6
J	376	VAL	MET	conflict	UNP Q9W9C6

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.

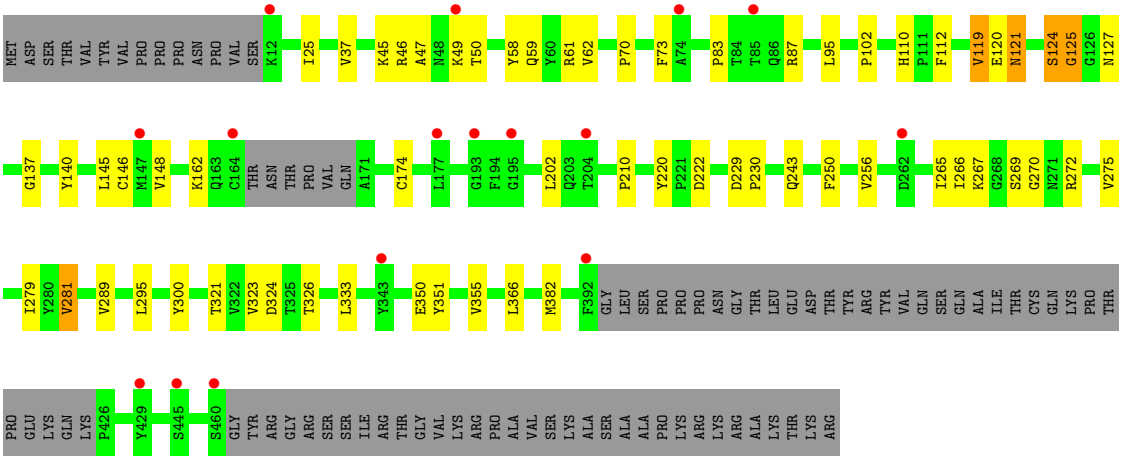
• Molecule 1: Major capsid protein L1

Chain A:



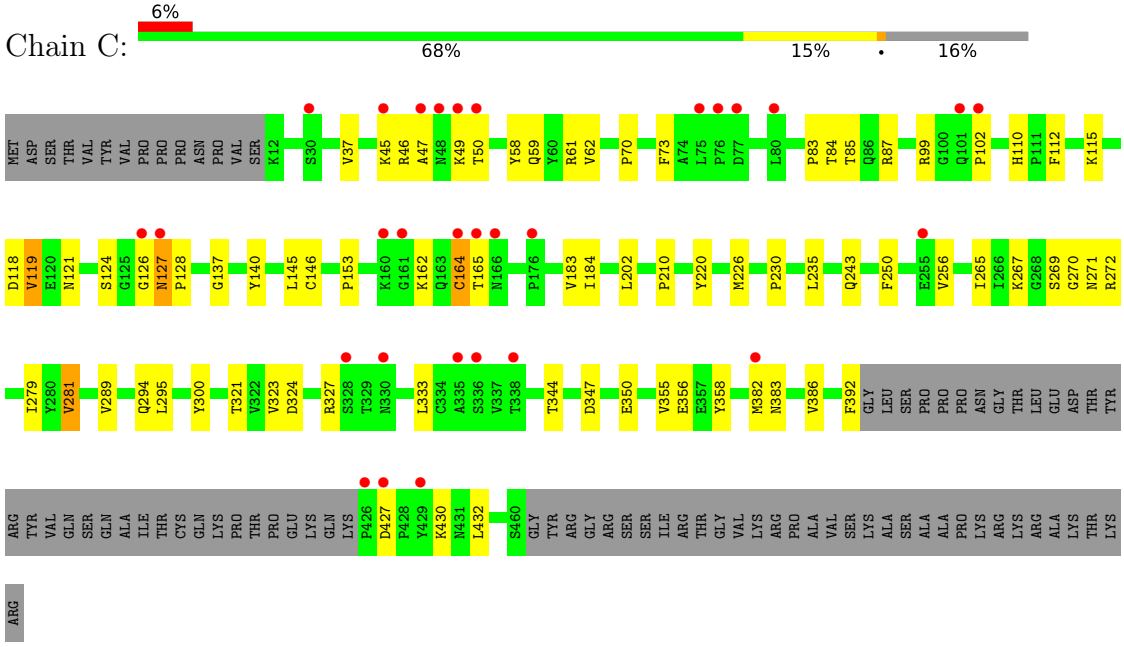
• Molecule 1: Major capsid protein L1

Chain B:



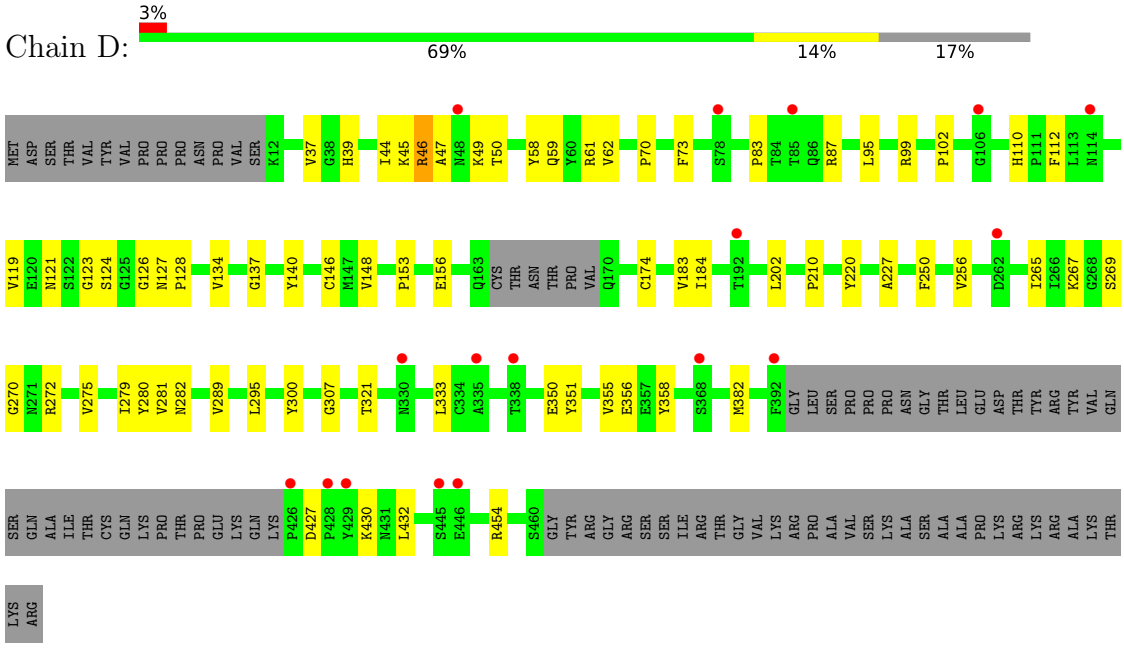
• Molecule 1: Major capsid protein L1

Chain C:



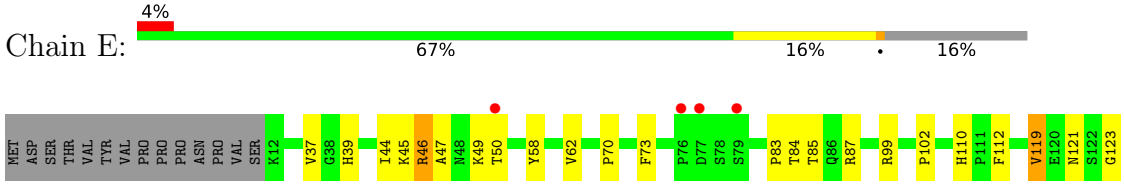
• Molecule 1: Major capsid protein L1

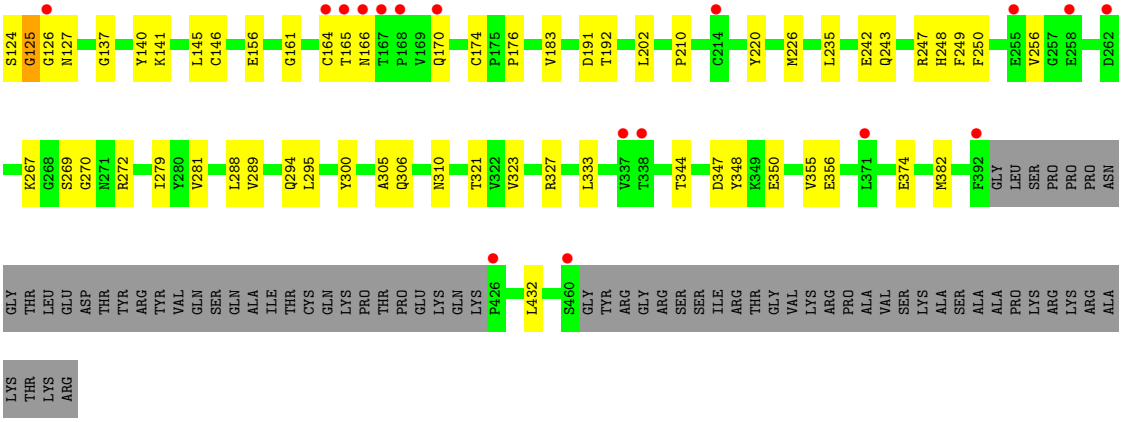
Chain D:



• Molecule 1: Major capsid protein L1

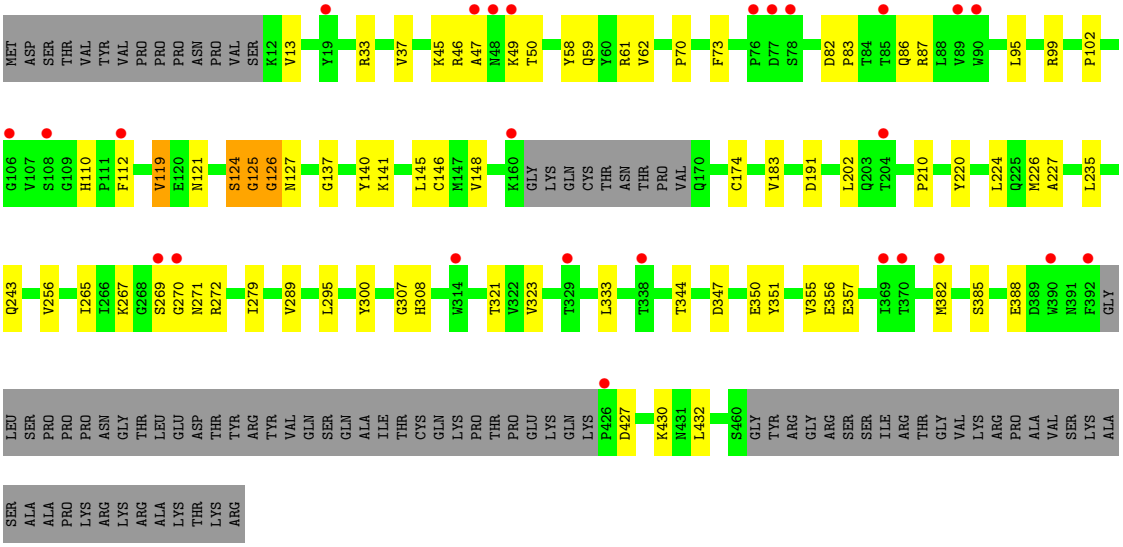
Chain E:





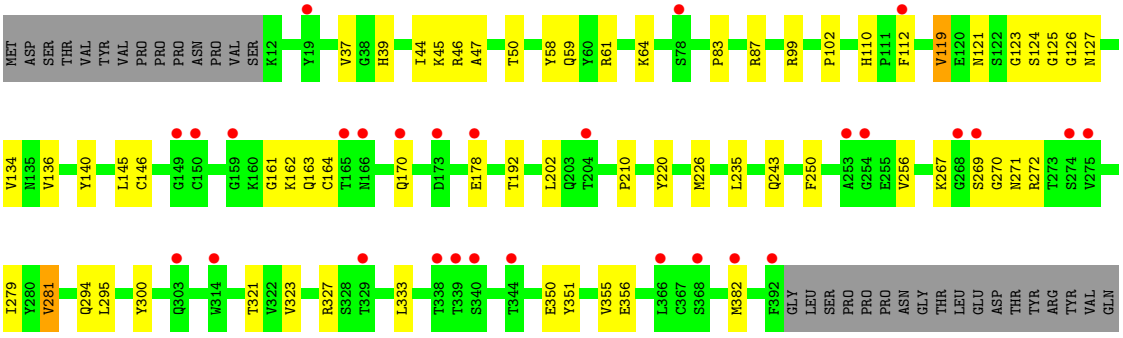
• Molecule 1: Major capsid protein L1

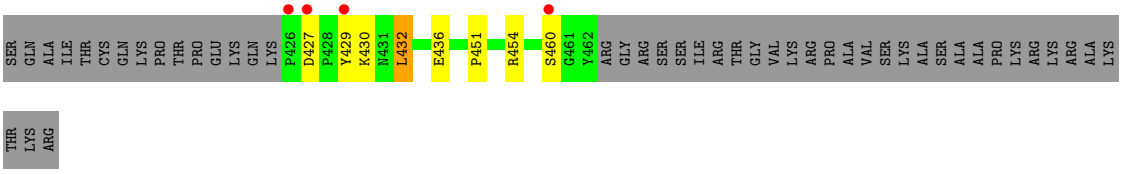
Chain F:



• Molecule 1: Major capsid protein L1

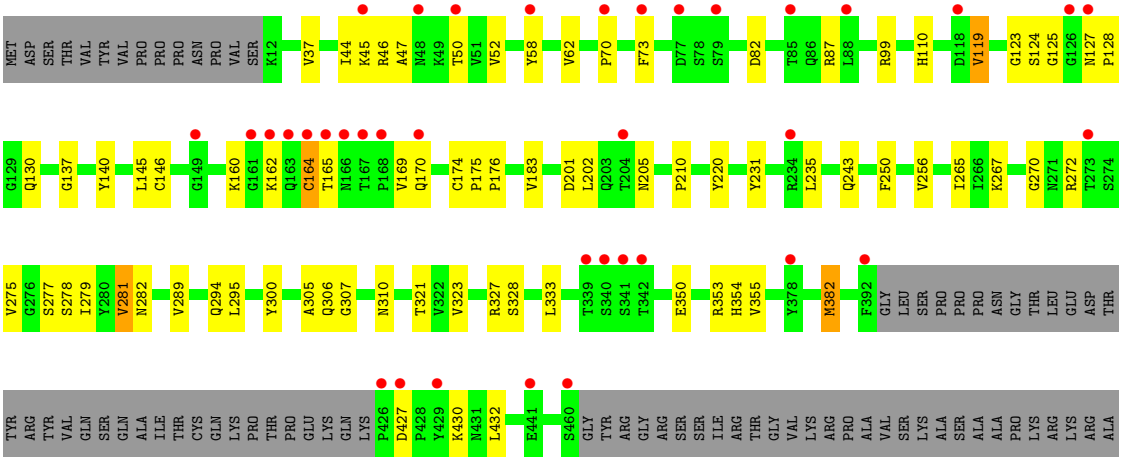
Chain G:





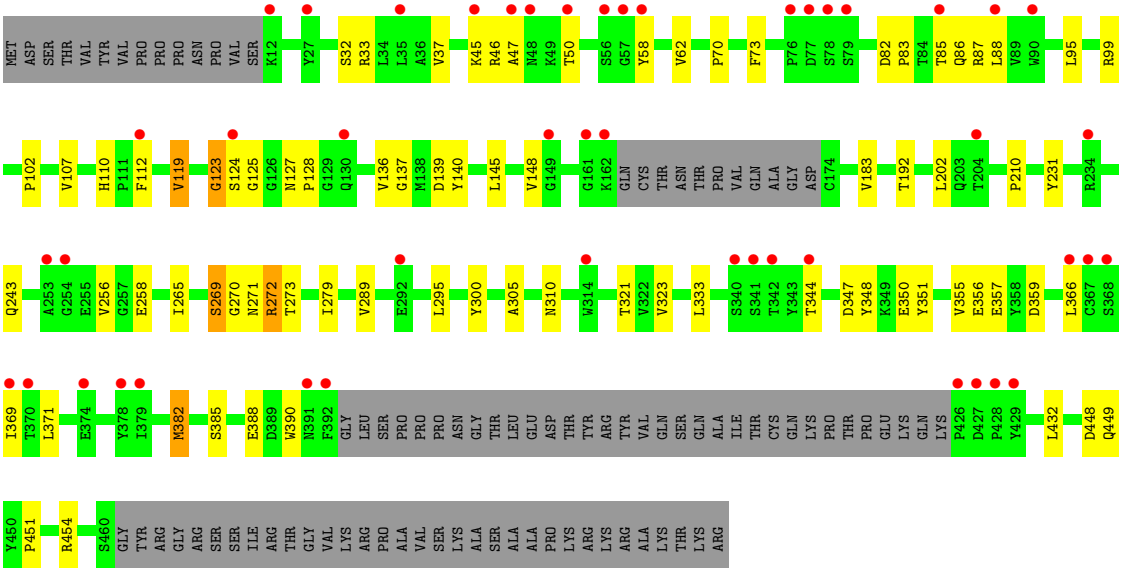
• Molecule 1: Major capsid protein L1

Chain H:



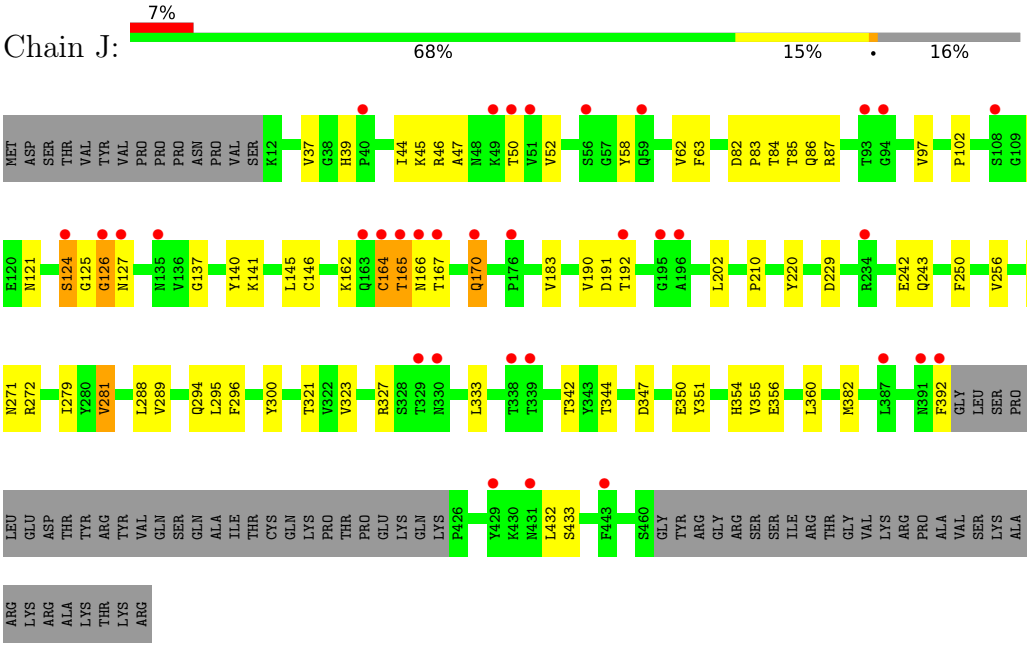
• Molecule 1: Major capsid protein L1

Chain I:



• Molecule 1: Major capsid protein L1

Chain J:


7%
68%
15%
16%
MET ASP SER THR VAL TYR VAL PRO PRO PRO ASN PRO VAL SER K12 V37 G38 H39 P40 I44 K45 R46 A47 N48 K49 T50 V51 V52 S56 G57 Y58 Q59 V62 F63 D82 P83 T84 T85 Q86 R87 T93 G94 V97 P102 S108 G109 H110 P111 F112 V119
E120 N121 S124 G125 G126 N127 N135 V136 G137 Y140 K141 L145 C146 K162 Q163 C164 T165 N166 T167 Q170 P176 V183 V190 D191 T192 G195 A196 L202 P210 Y220 D229 R234 E242 Q243 F250 V256 I265 I266 S269 G270
N271 R272 I279 Y280 V281 L288 V289 Q294 L295 F296 Y300 T321 V322 V323 R327 S328 T329 N330 L333 T338 T339 T342 Y343 T344 D347 E350 Y351 H354 V355 E356 L360 M382 L387 N391 F392 GLY LEU SER PRO PRO PRO ASN GLY THR
LEU GLU ASP THR TYR ARG TYR VAL GLN SER GLN ALA ILE THR CYS GLN LYS PRO THR PRO GLU LYS GLN LYS P426 Y429 K430 M431 L432 S433 F443 S460 GLY TYR ARG GLY ARG SER SER ILE ARG THR GLY VAL LYS ARG PRO ALA VAL SER LYS ALA SER ALA ALA PRO LYS
ARG LYS ARG ALA LYS THR LYS ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	308.96Å 106.66Å 198.71Å 90.00° 125.96° 90.00°	Depositor
Resolution (Å)	33.88 – 3.37 33.88 – 3.37	Depositor EDS
% Data completeness (in resolution range)	94.4 (33.88-3.37) 94.4 (33.88-3.37)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.35 (at 3.39Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.256 , 0.294 0.256 , 0.294	Depositor DCC
R_{free} test set	3592 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	95.6	Xtriage
Anisotropy	0.203	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 22.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	32368	wwPDB-VP
Average B, all atoms (Å^2)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.01% 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	0.26	0/3342	0.48	0/4544
1	B	0.26	0/3295	0.47	0/4476
1	C	0.25	0/3342	0.46	0/4544
1	D	0.26	0/3298	0.46	0/4480
1	E	0.25	0/3342	0.46	0/4544
1	F	0.25	0/3276	0.46	0/4452
1	G	0.25	0/3359	0.46	0/4567
1	H	0.26	0/3342	0.47	0/4544
1	I	0.26	0/3263	0.47	0/4433
1	J	0.25	0/3342	0.46	0/4544
All	All	0.25	0/33201	0.47	0/45128

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	F	0	1
1	H	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	125	GLY	Peptide
1	H	123	GLY	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	3258	0	3150	61	0
1	B	3213	0	3105	49	0
1	C	3258	0	3150	54	0
1	D	3216	0	3108	57	0
1	E	3258	0	3150	58	0
1	F	3194	0	3084	57	0
1	G	3274	0	3162	51	0
1	H	3258	0	3150	49	0
1	I	3181	0	3080	54	0
1	J	3258	0	3150	52	0
All	All	32368	0	31289	430	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 7.

The worst 5 of 430 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:119:VAL:HG11	1:D:280:TYR:HE1	1.43	0.83
1:D:350:GLU:HB2	1:E:279:ILE:HD13	1.59	0.83
1:D:279:ILE:HD13	1:G:350:GLU:HB2	1.65	0.78
1:H:350:GLU:HB2	1:I:279:ILE:HD13	1.66	0.78
1:I:83:PRO:O	1:I:87:ARG:NH1	2.15	0.78

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	412/496 (83%)	389 (94%)	19 (5%)	4 (1%)	15	49
1	B	404/496 (82%)	385 (95%)	15 (4%)	4 (1%)	15	49
1	C	412/496 (83%)	391 (95%)	15 (4%)	6 (2%)	10	39
1	D	404/496 (82%)	387 (96%)	15 (4%)	2 (0%)	29	63
1	E	412/496 (83%)	390 (95%)	18 (4%)	4 (1%)	15	49
1	F	401/496 (81%)	382 (95%)	15 (4%)	4 (1%)	15	49
1	G	414/496 (84%)	391 (94%)	19 (5%)	4 (1%)	15	49
1	H	412/496 (83%)	393 (95%)	15 (4%)	4 (1%)	15	49
1	I	399/496 (80%)	382 (96%)	12 (3%)	5 (1%)	12	42
1	J	412/496 (83%)	388 (94%)	17 (4%)	7 (2%)	9	36
All	All	4082/4960 (82%)	3878 (95%)	160 (4%)	44 (1%)	14	46

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	382	MET
1	F	126	GLY
1	I	269	SER
1	I	271	ASN
1	J	126	GLY

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	363/432 (84%)	360 (99%)	3 (1%)	81	91
1	B	357/432 (83%)	350 (98%)	7 (2%)	55	78
1	C	363/432 (84%)	359 (99%)	4 (1%)	73	86
1	D	357/432 (83%)	354 (99%)	3 (1%)	81	91
1	E	363/432 (84%)	358 (99%)	5 (1%)	67	83

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	F	355/432 (82%)	353 (99%)	2 (1%)	86 93
1	G	364/432 (84%)	358 (98%)	6 (2%)	62 81
1	H	363/432 (84%)	356 (98%)	7 (2%)	57 79
1	I	354/432 (82%)	351 (99%)	3 (1%)	81 91
1	J	363/432 (84%)	355 (98%)	8 (2%)	52 76
All	All	3602/4320 (83%)	3554 (99%)	48 (1%)	69 84

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	281	VAL
1	H	281	VAL
1	G	432	LEU
1	H	170	GLN
1	I	32	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	H	130	GLN

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)

There are no ligands in this entry.

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

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	416/496 (83%)	0.03	15 (3%) 42 45	47, 73, 124, 205	0
1	B	410/496 (82%)	0.08	16 (3%) 39 41	56, 87, 150, 194	0
1	C	416/496 (83%)	0.31	30 (7%) 15 18	67, 99, 168, 259	0
1	D	410/496 (82%)	0.19	17 (4%) 37 39	64, 88, 139, 190	0
1	E	416/496 (83%)	0.19	21 (5%) 28 31	70, 93, 152, 222	0
1	F	407/496 (82%)	0.42	26 (6%) 19 21	67, 117, 175, 226	0
1	G	418/496 (84%)	0.36	33 (7%) 12 14	68, 104, 163, 232	0
1	H	416/496 (83%)	0.40	37 (8%) 9 11	62, 103, 161, 236	0
1	I	405/496 (81%)	0.64	47 (11%) 4 5	73, 126, 190, 225	0
1	J	416/496 (83%)	0.43	34 (8%) 11 13	75, 105, 162, 213	0
All	All	4130/4960 (83%)	0.30	276 (6%) 17 20	47, 99, 164, 259	0

The worst 5 of 276 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	426	PRO	12.0
1	H	165	THR	8.8
1	F	392	PHE	8.5
1	C	165	THR	8.2
1	I	392	PHE	7.8

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.