



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 02:25 pm BST

PDB ID : 6IGD
Title : Crystal structure of HPV58/33 chimeric L1 pentamer
Authors : Li, Z.H.; Song, S.; He, M.Z.; Gu, Y.; Li, S.W.
Deposited on : 2018-09-25
Resolution : 2.50 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 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	:	2.11
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

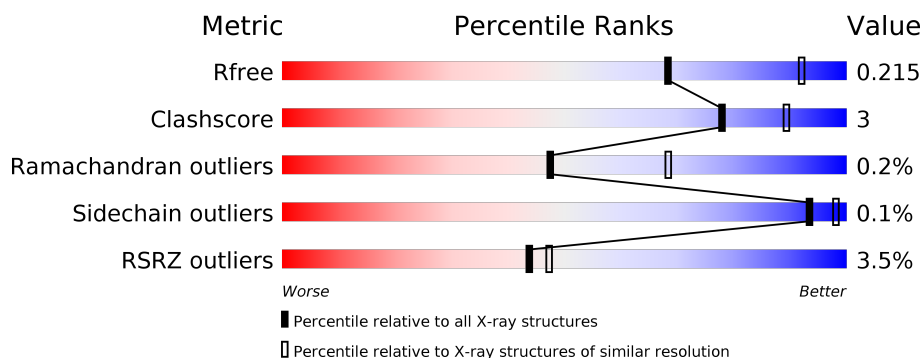
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 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	A	524	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>5%</div> <div>21%</div> </div> </div>
1	B	524	<div> <div>2%</div> <div> <div></div> <div>75%</div> <div>•</div> <div>21%</div> </div> </div>
1	C	524	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>•</div> <div>21%</div> </div> </div>
1	D	524	<div> <div>2%</div> <div> <div></div> <div>74%</div> <div>5%</div> <div>21%</div> </div> </div>
1	E	524	<div> <div>5%</div> <div> <div></div> <div>73%</div> <div>7%</div> <div>20%</div> </div> </div>
1	F	524	<div> <div>2%</div> <div> <div></div> <div>73%</div> <div>7%</div> <div>21%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	524	<div><div><div></div><div></div><div></div></div><div><div>2%</div><div>74%</div><div>5%</div><div>21%</div></div></div>
1	H	524	<div><div><div></div><div></div><div></div></div><div><div>2%</div><div>75%</div><div>•</div><div>21%</div></div></div>
1	I	524	<div><div><div></div><div></div><div></div></div><div><div>4%</div><div>74%</div><div>5%</div><div>21%</div></div></div>
1	J	524	<div><div><div></div><div></div><div></div></div><div><div>4%</div><div>74%</div><div>6%</div><div>20%</div></div></div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 66727 atoms, of which 32076 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	H	N	O	S	0	0	0
			6527	2120	3206	550	632	19			
1	B	414	Total	C	H	N	O	S	0	0	0
			6505	2114	3195	548	629	19			
1	C	415	Total	C	H	N	O	S	0	0	0
			6516	2117	3200	549	631	19			
1	D	416	Total	C	H	N	O	S	0	0	0
			6537	2123	3213	551	631	19			
1	E	420	Total	C	H	N	O	S	0	0	0
			6574	2133	3228	557	637	19			
1	F	416	Total	C	H	N	O	S	0	0	0
			6526	2120	3205	550	632	19			
1	G	414	Total	C	H	N	O	S	0	0	0
			6505	2114	3195	548	629	19			
1	H	414	Total	C	H	N	O	S	0	0	0
			6501	2112	3194	548	628	19			
1	I	416	Total	C	H	N	O	S	0	0	0
			6527	2120	3206	550	632	19			
1	J	421	Total	C	H	N	O	S	0	0	0
			6589	2138	3234	558	640	19			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	54	ASN	SER	engineered mutation	UNP P26535
A	56	THR	ASN	engineered mutation	UNP P26535
A	58	ALA	ASN	engineered mutation	UNP P26535
A	61	LEU	VAL	engineered mutation	UNP P26535
A	176	SER	CYS	engineered mutation	UNP P26535
A	352	SER	GLY	engineered mutation	UNP P26535
B	54	ASN	SER	engineered mutation	UNP P26535
B	56	THR	ASN	engineered mutation	UNP P26535
B	58	ALA	ASN	engineered mutation	UNP P26535

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	61	LEU	VAL	engineered mutation	UNP P26535
B	176	SER	CYS	engineered mutation	UNP P26535
B	352	SER	GLY	engineered mutation	UNP P26535
C	54	ASN	SER	engineered mutation	UNP P26535
C	56	THR	ASN	engineered mutation	UNP P26535
C	58	ALA	ASN	engineered mutation	UNP P26535
C	61	LEU	VAL	engineered mutation	UNP P26535
C	176	SER	CYS	engineered mutation	UNP P26535
C	352	SER	GLY	engineered mutation	UNP P26535
D	54	ASN	SER	engineered mutation	UNP P26535
D	56	THR	ASN	engineered mutation	UNP P26535
D	58	ALA	ASN	engineered mutation	UNP P26535
D	61	LEU	VAL	engineered mutation	UNP P26535
D	176	SER	CYS	engineered mutation	UNP P26535
D	352	SER	GLY	engineered mutation	UNP P26535
E	54	ASN	SER	engineered mutation	UNP P26535
E	56	THR	ASN	engineered mutation	UNP P26535
E	58	ALA	ASN	engineered mutation	UNP P26535
E	61	LEU	VAL	engineered mutation	UNP P26535
E	176	SER	CYS	engineered mutation	UNP P26535
E	352	SER	GLY	engineered mutation	UNP P26535
F	54	ASN	SER	engineered mutation	UNP P26535
F	56	THR	ASN	engineered mutation	UNP P26535
F	58	ALA	ASN	engineered mutation	UNP P26535
F	61	LEU	VAL	engineered mutation	UNP P26535
F	176	SER	CYS	engineered mutation	UNP P26535
F	352	SER	GLY	engineered mutation	UNP P26535
G	54	ASN	SER	engineered mutation	UNP P26535
G	56	THR	ASN	engineered mutation	UNP P26535
G	58	ALA	ASN	engineered mutation	UNP P26535
G	61	LEU	VAL	engineered mutation	UNP P26535
G	176	SER	CYS	engineered mutation	UNP P26535
G	352	SER	GLY	engineered mutation	UNP P26535
H	54	ASN	SER	engineered mutation	UNP P26535
H	56	THR	ASN	engineered mutation	UNP P26535
H	58	ALA	ASN	engineered mutation	UNP P26535
H	61	LEU	VAL	engineered mutation	UNP P26535
H	176	SER	CYS	engineered mutation	UNP P26535
H	352	SER	GLY	engineered mutation	UNP P26535
I	54	ASN	SER	engineered mutation	UNP P26535
I	56	THR	ASN	engineered mutation	UNP P26535
I	58	ALA	ASN	engineered mutation	UNP P26535

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	61	LEU	VAL	engineered mutation	UNP P26535
I	176	SER	CYS	engineered mutation	UNP P26535
I	352	SER	GLY	engineered mutation	UNP P26535
J	54	ASN	SER	engineered mutation	UNP P26535
J	56	THR	ASN	engineered mutation	UNP P26535
J	58	ALA	ASN	engineered mutation	UNP P26535
J	61	LEU	VAL	engineered mutation	UNP P26535
J	176	SER	CYS	engineered mutation	UNP P26535
J	352	SER	GLY	engineered mutation	UNP P26535

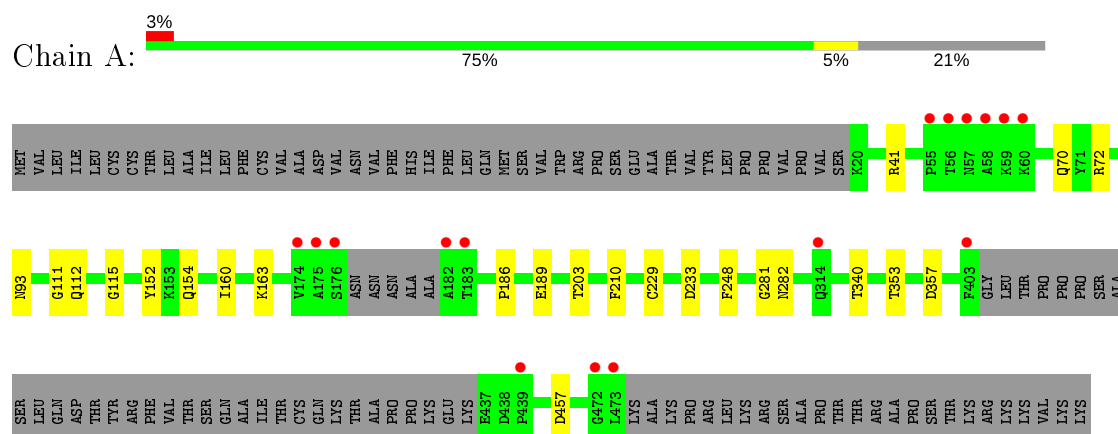
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	146	Total O 146 146	0	0
2	B	123	Total O 123 123	0	0
2	C	154	Total O 154 154	0	0
2	D	172	Total O 172 172	0	0
2	E	133	Total O 133 133	0	0
2	F	144	Total O 144 144	0	0
2	G	154	Total O 154 154	0	0
2	H	139	Total O 139 139	0	0
2	I	132	Total O 132 132	0	0
2	J	123	Total O 123 123	0	0

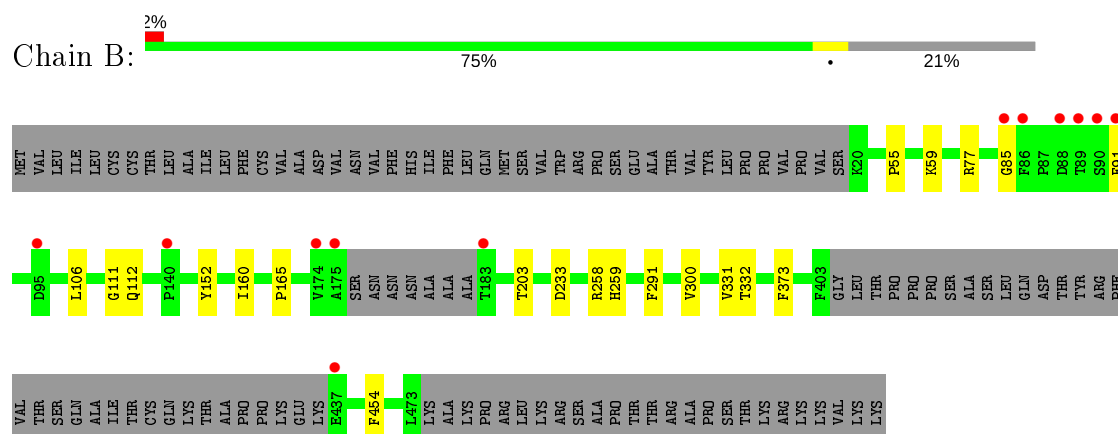
3 Residue-property plots [i](#)

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.

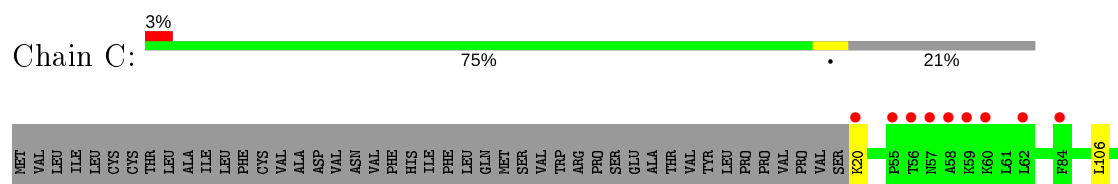
- Molecule 1: Major capsid protein L1

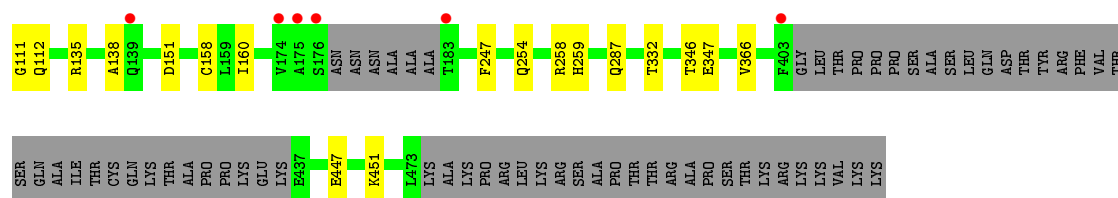


- Molecule 1: Major capsid protein L1

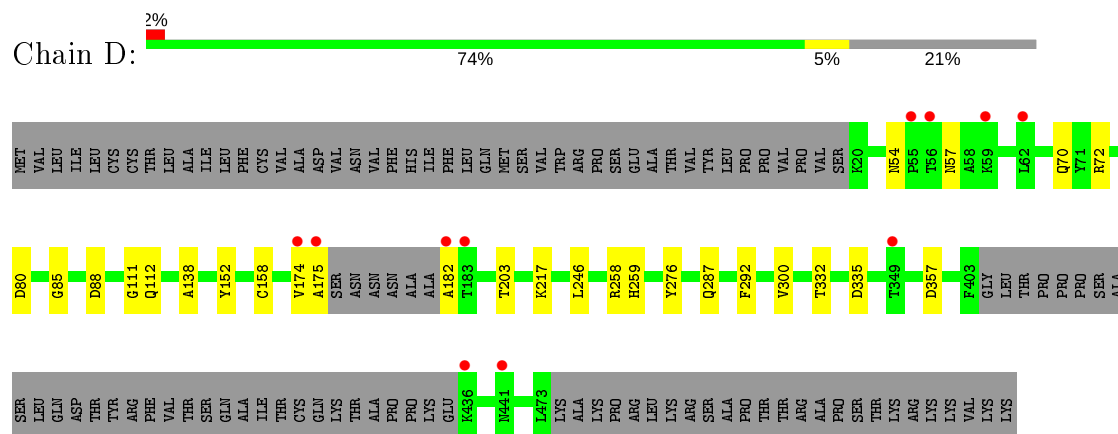


- Molecule 1: Major capsid protein L1

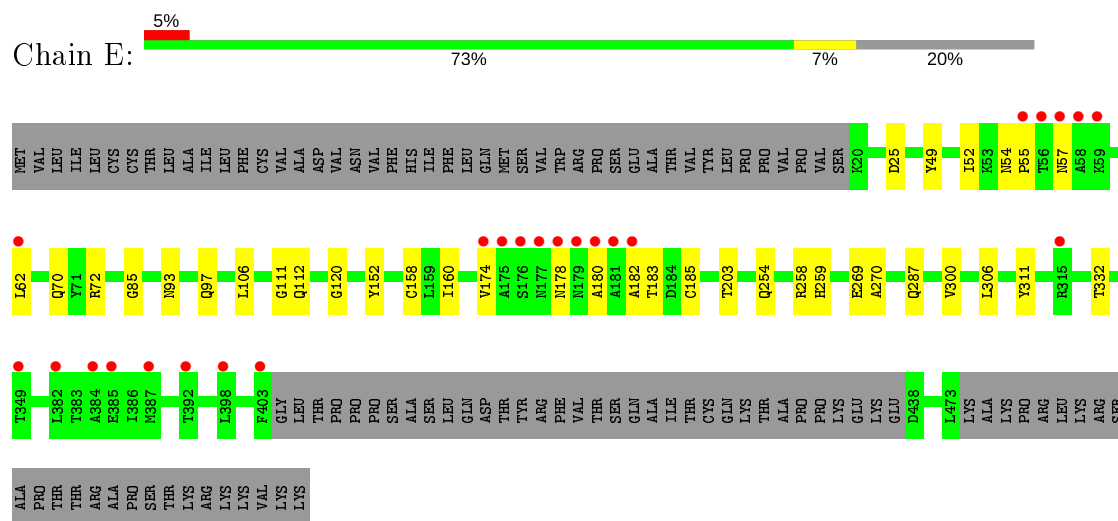




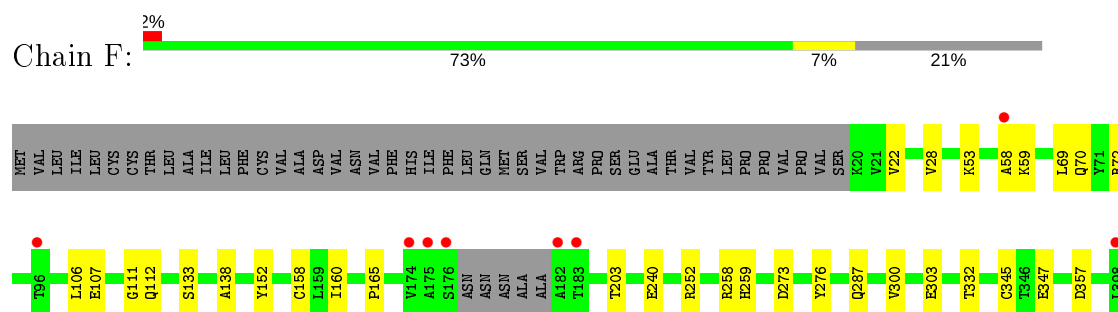
• Molecule 1: Major capsid protein L1

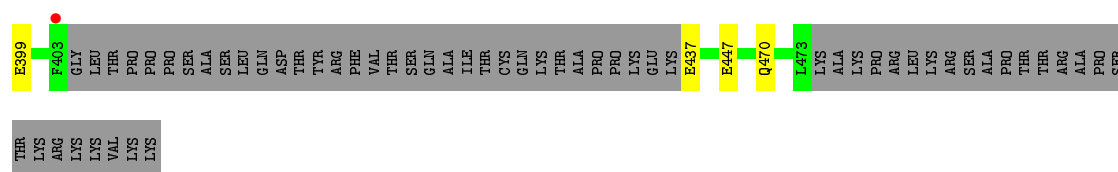


• Molecule 1: Major capsid protein L1

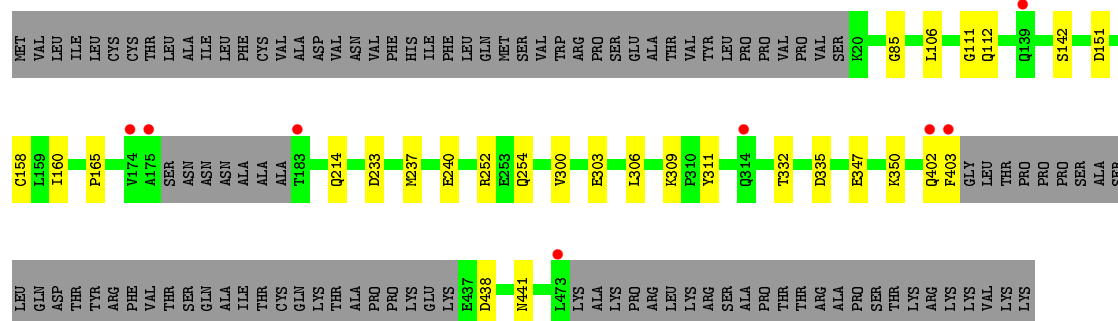
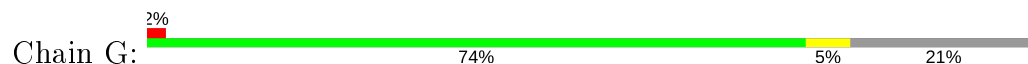


• Molecule 1: Major capsid protein L1

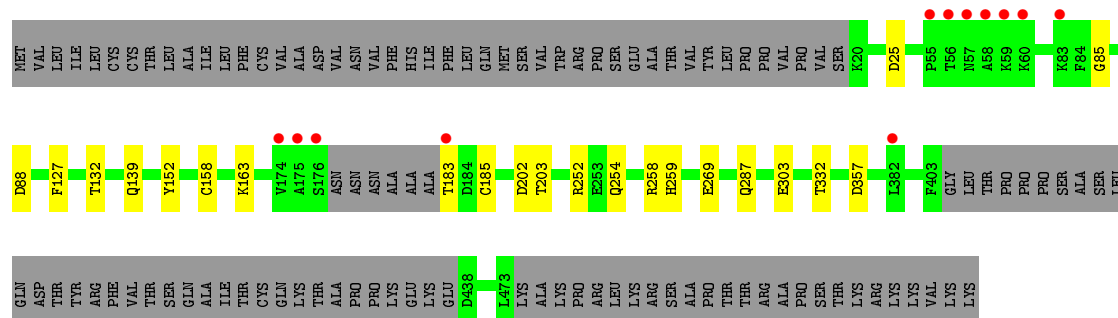
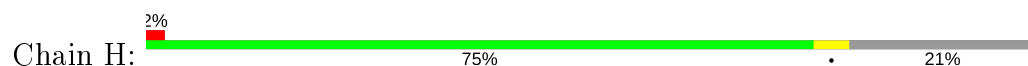




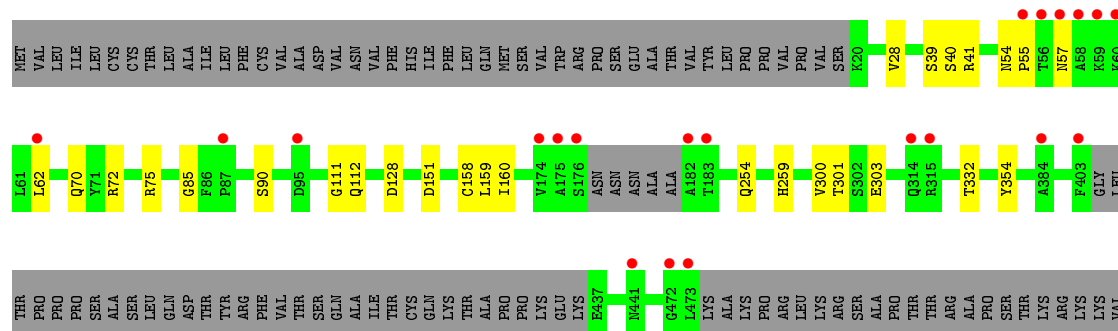
• Molecule 1: Major capsid protein L1



• Molecule 1: Major capsid protein L1

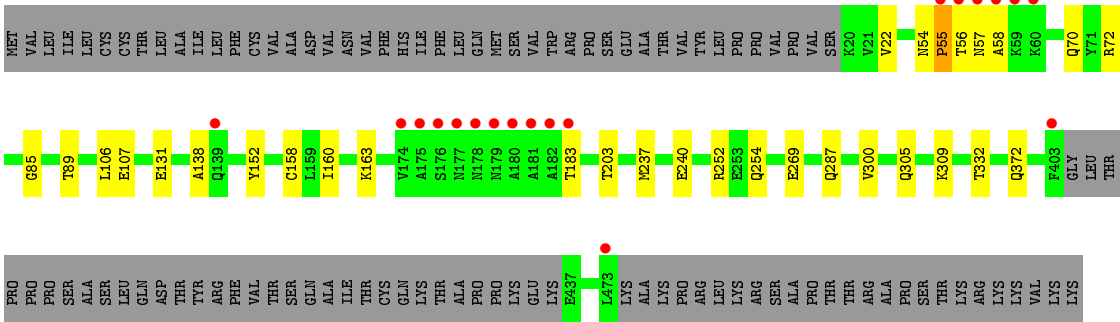


• Molecule 1: Major capsid protein L1



LYS
LYS

● Molecule 1: Major capsid protein L1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	153.69Å 105.83Å 154.71Å 90.00° 99.55° 90.00°	Depositor
Resolution (Å)	40.80 – 2.50 40.78 – 2.49	Depositor EDS
% Data completeness (in resolution range)	99.5 (40.80-2.50) 99.5 (40.78-2.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.13	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.51 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.182 , 0.216 0.182 , 0.215	Depositor DCC
R_{free} test set	8264 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å ²)	32.3	Xtriage
Anisotropy	0.729	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 43.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.013 for l,-k,h	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	66727	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.54% 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 ⓘ

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	0.30	0/3407	0.51	0/4621
1	B	0.29	0/3396	0.50	0/4606
1	C	0.29	0/3402	0.49	0/4614
1	D	0.29	0/3410	0.50	0/4624
1	E	0.29	0/3433	0.51	0/4659
1	F	0.29	0/3407	0.51	0/4621
1	G	0.30	0/3396	0.51	0/4606
1	H	0.29	0/3393	0.49	0/4602
1	I	0.29	0/3407	0.50	0/4621
1	J	0.29	0/3442	0.51	0/4671
All	All	0.29	0/34093	0.50	0/46245

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	3321	3206	3204	19	1
1	B	3310	3195	3194	17	0
1	C	3316	3200	3199	16	0
1	D	3324	3213	3212	18	0
1	E	3346	3228	3227	24	1
1	F	3321	3205	3204	25	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	3310	3195	3194	21	1
1	H	3307	3194	3193	17	0
1	I	3321	3206	3204	22	0
1	J	3355	3234	3233	29	0
2	A	146	0	0	7	1
2	B	123	0	0	5	0
2	C	154	0	0	8	0
2	D	172	0	0	8	0
2	E	133	0	0	7	0
2	F	144	0	0	11	1
2	G	154	0	0	6	0
2	H	139	0	0	5	0
2	I	132	0	0	8	0
2	J	123	0	0	9	0
All	All	34651	32076	32064	190	3

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 190 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:B:59:LYS:NZ	2:B:1401:HOH:O	1.96	0.97
1:G:403:PHE:O	2:G:501:HOH:O	1.83	0.95
1:E:306:LEU:O	1:E:311:TYR:OH	1.84	0.93
1:G:306:LEU:O	1:G:311:TYR:OH	1.87	0.92
1:G:142:SER:OG	2:G:502:HOH:O	1.90	0.90

All (3) 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)
2:A:633:HOH:O	2:F:502:HOH:O[1_554]	1.82	0.38
1:A:457:ASP:OD1	1:E:178:ASN:ND2[2_454]	2.10	0.10
1:F:59:LYS:NZ	1:G:402:GLN:O[2_555]	2.11	0.09

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	410/524 (78%)	397 (97%)	13 (3%)	0	100	100
1	B	408/524 (78%)	395 (97%)	12 (3%)	1 (0%)	47	68
1	C	409/524 (78%)	398 (97%)	11 (3%)	0	100	100
1	D	410/524 (78%)	398 (97%)	11 (3%)	1 (0%)	47	68
1	E	416/524 (79%)	404 (97%)	11 (3%)	1 (0%)	47	68
1	F	410/524 (78%)	397 (97%)	13 (3%)	0	100	100
1	G	408/524 (78%)	395 (97%)	12 (3%)	1 (0%)	47	68
1	H	408/524 (78%)	396 (97%)	11 (3%)	1 (0%)	47	68
1	I	410/524 (78%)	398 (97%)	11 (3%)	1 (0%)	47	68
1	J	417/524 (80%)	404 (97%)	11 (3%)	2 (0%)	29	48
All	All	4106/5240 (78%)	3982 (97%)	116 (3%)	8 (0%)	47	68

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	85	GLY
1	G	85	GLY
1	H	85	GLY
1	I	85	GLY
1	J	85	GLY

5.3.2 Protein sidechains

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	370/466 (79%)	370 (100%)	0	100	100
1	B	369/466 (79%)	369 (100%)	0	100	100
1	C	370/466 (79%)	370 (100%)	0	100	100
1	D	370/466 (79%)	370 (100%)	0	100	100
1	E	372/466 (80%)	371 (100%)	1 (0%)	92	97
1	F	370/466 (79%)	369 (100%)	1 (0%)	92	97
1	G	369/466 (79%)	369 (100%)	0	100	100
1	H	369/466 (79%)	369 (100%)	0	100	100
1	I	370/466 (79%)	370 (100%)	0	100	100
1	J	373/466 (80%)	372 (100%)	1 (0%)	92	97
All	All	3702/4660 (79%)	3699 (100%)	3 (0%)	93	98

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	174	VAL
1	F	69	LEU
1	J	163	LYS

Some 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 carbohydrates 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 [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	A	416/524 (79%)	-0.02	16 (3%) 40 43	23, 36, 72, 123	0
1	B	414/524 (79%)	0.03	12 (2%) 51 55	23, 38, 68, 149	0
1	C	415/524 (79%)	-0.07	15 (3%) 42 46	25, 35, 63, 124	0
1	D	416/524 (79%)	-0.07	11 (2%) 56 59	26, 35, 68, 128	0
1	E	420/524 (80%)	0.14	24 (5%) 23 25	21, 37, 72, 154	0
1	F	416/524 (79%)	-0.07	9 (2%) 62 65	25, 35, 62, 133	0
1	G	414/524 (79%)	-0.14	8 (1%) 66 69	24, 36, 55, 103	0
1	H	414/524 (79%)	-0.05	12 (2%) 51 55	27, 37, 64, 129	0
1	I	416/524 (79%)	0.05	21 (5%) 28 30	26, 40, 80, 115	0
1	J	421/524 (80%)	0.18	19 (4%) 33 36	24, 37, 77, 167	0
All	All	4162/5240 (79%)	-0.00	147 (3%) 44 47	21, 37, 68, 167	0

The worst 5 of 147 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	180	ALA	18.2
1	F	182	ALA	12.0
1	B	175	ALA	10.9
1	I	182	ALA	9.5
1	E	178	ASN	9.1

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 carbohydrates 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.