wwPDB X-ray Structure Validation Summary Report (i)

Jun 6, 2016 – 06:57 PM EDT

PDB ID	:	5J7U
Title	:	Faustovirus major capsid protein
Authors	:	Klose, T.; Rossmann, M.G.
Deposited on	:	2016-04-06
Resolution	:	2.44 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:

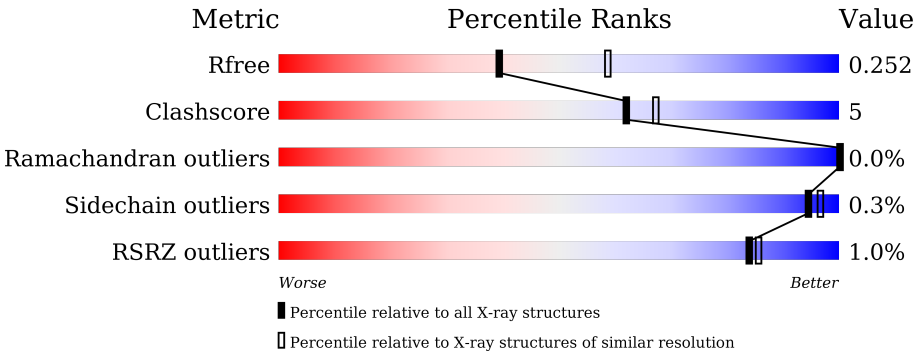
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027674
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027674

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 2.44 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	1003 (2.46-2.42)
Clashscore	102246	1071 (2.46-2.42)
Ramachandran outliers	100387	1065 (2.46-2.42)
Sidechain outliers	100360	1065 (2.46-2.42)
RSRZ outliers	91569	1005 (2.46-2.42)

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	645	83% 11% 6%
1	B	645	83% 13% .
1	C	645	% 82% 14% .
1	D	645	% 82% 12% 6%
1	E	645	2% 82% 13% .
1	F	645	% 83% 11% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	645	81% 13% 5%
1	H	645	85% 10% 5%
1	I	645	82% 13%
1	J	645	84% 11% 5%
1	K	645	82% 13% 5%
1	L	645	82% 12% 5%
2	M	21	100%
2	N	21	95% 5%
2	O	21	100%
2	P	21	95% 5%
2	Q	21	76% 24%
2	R	21	71% 19% 10%
2	S	21	86% 14%
2	T	21	100%
2	U	21	95% 5%
2	V	21	86% 14%
2	W	21	90% 10%
2	X	21	67% 24% 10%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 61035 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	609	Total	C	N	O	S	0	0	0
			4785	3060	803	908	14			
1	B	616	Total	C	N	O	S	0	0	0
			4837	3092	812	919	14			
1	C	616	Total	C	N	O	S	0	0	0
			4842	3098	812	918	14			
1	D	609	Total	C	N	O	S	0	0	0
			4785	3060	803	908	14			
1	E	616	Total	C	N	O	S	0	0	0
			4841	3097	812	918	14			
1	F	610	Total	C	N	O	S	0	0	0
			4793	3064	804	911	14			
1	G	611	Total	C	N	O	S	0	0	0
			4799	3067	805	913	14			
1	H	615	Total	C	N	O	S	0	0	0
			4831	3089	811	917	14			
1	I	616	Total	C	N	O	S	0	0	0
			4842	3098	812	918	14			
1	J	610	Total	C	N	O	S	0	0	0
			4793	3064	804	911	14			
1	K	614	Total	C	N	O	S	0	0	0
			4823	3083	810	916	14			
1	L	610	Total	C	N	O	S	0	0	0
			4793	3064	804	911	14			

- Molecule 2 is a protein called unknown.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	M	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	N	21	Total	C	N	O	0	0	0
			105	63	21	21			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	O	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	P	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	Q	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	R	19	Total	C	N	O	0	0	0
			95	57	19	19			
2	S	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	T	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	U	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	V	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	W	21	Total	C	N	O	0	0	0
			105	63	21	21			
2	X	19	Total	C	N	O	0	0	0
			95	57	19	19			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	173	Total	O	0	0
			173	173		
3	B	170	Total	O	0	0
			170	170		
3	C	164	Total	O	0	0
			164	164		
3	D	191	Total	O	0	0
			191	191		
3	E	168	Total	O	0	0
			168	168		
3	F	172	Total	O	0	0
			172	172		
3	N	2	Total	O	0	0
			2	2		
3	O	2	Total	O	0	0
			2	2		
3	P	5	Total	O	0	0
			5	5		

Continued on next page...

Continued from previous page...

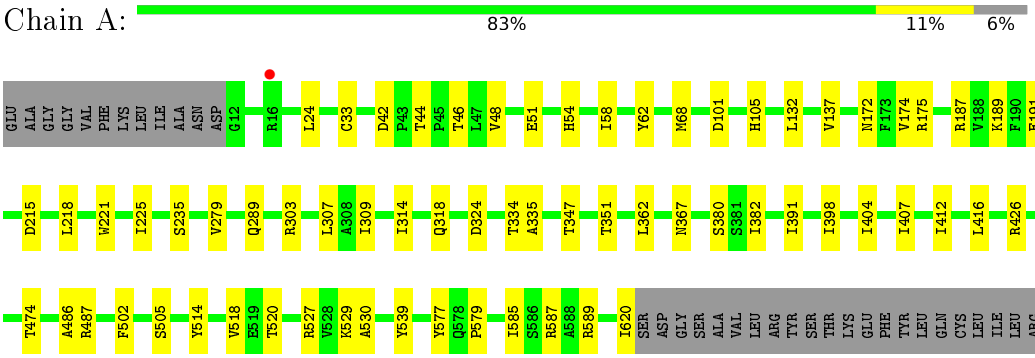
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	Q	2	Total O 2 2	0	0
3	R	2	Total O 2 2	0	0
3	G	176	Total O 176 176	0	0
3	H	165	Total O 165 165	0	0
3	I	133	Total O 133 133	0	0
3	J	163	Total O 163 163	0	0
3	K	156	Total O 156 156	0	0
3	L	171	Total O 171 171	0	0
3	S	1	Total O 1 1	0	0
3	U	7	Total O 7 7	0	0
3	V	3	Total O 3 3	0	0
3	W	2	Total O 2 2	0	0
3	X	3	Total O 3 3	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 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.

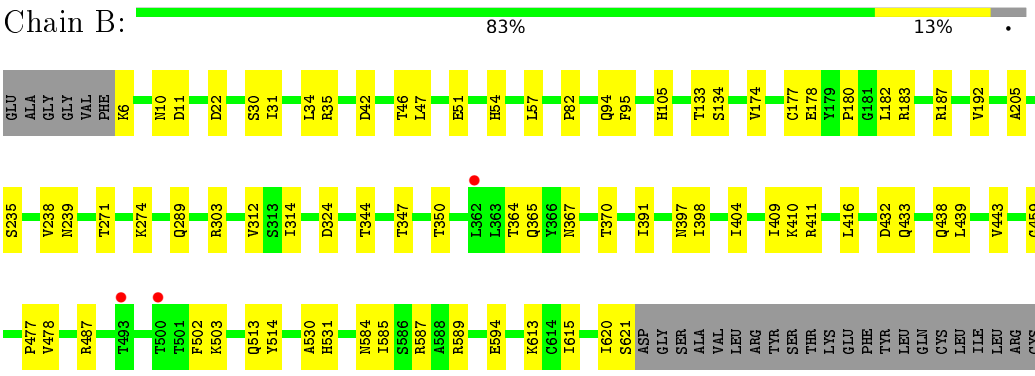
- Molecule 1: major capsid protein

Chain A:



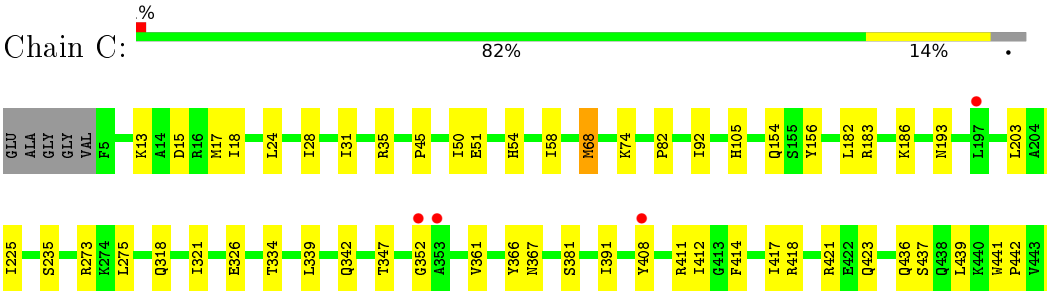
- Molecule 1: major capsid protein

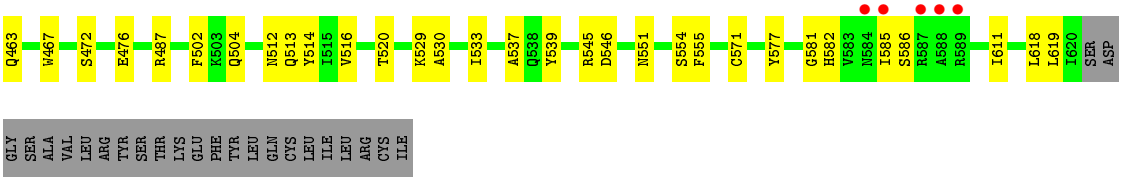
Chain B:



- Molecule 1: major capsid protein

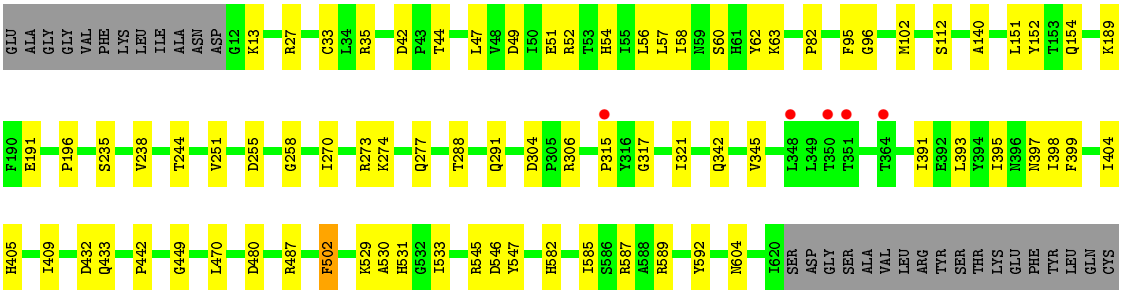
Chain C:





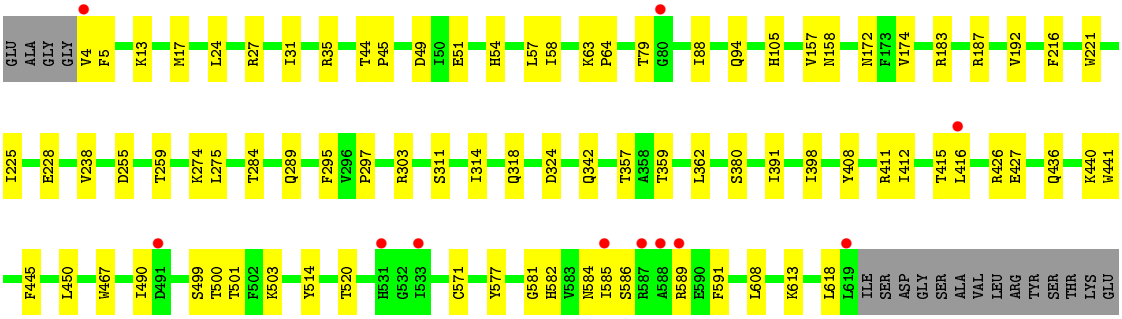
• Molecule 1: major capsid protein

Chain D:



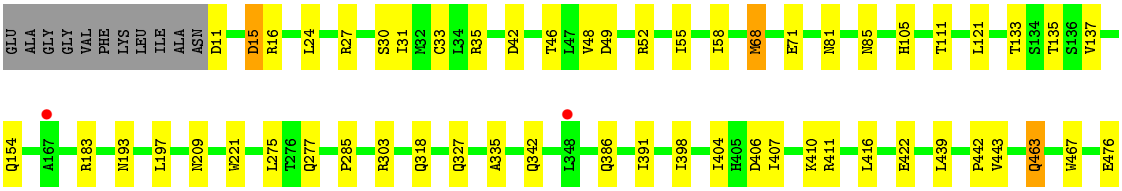
• Molecule 1: major capsid protein

Chain E:



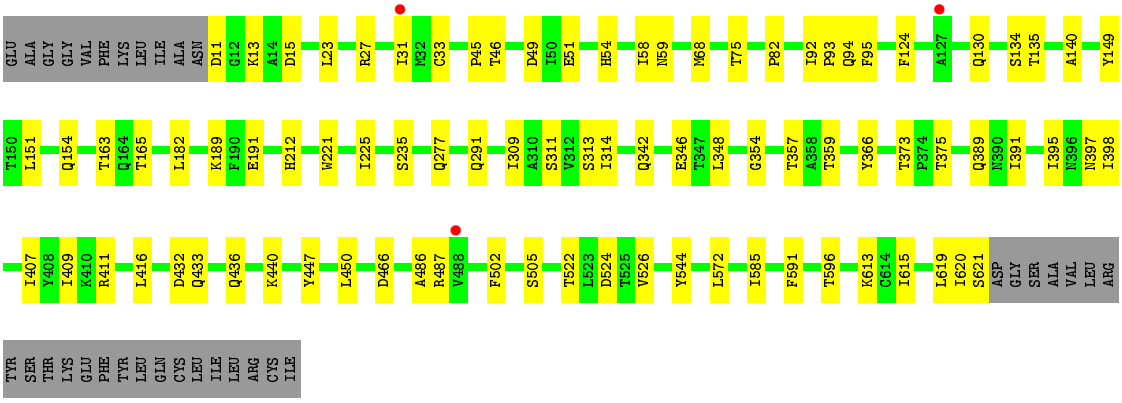
• Molecule 1: major capsid protein

Chain F:



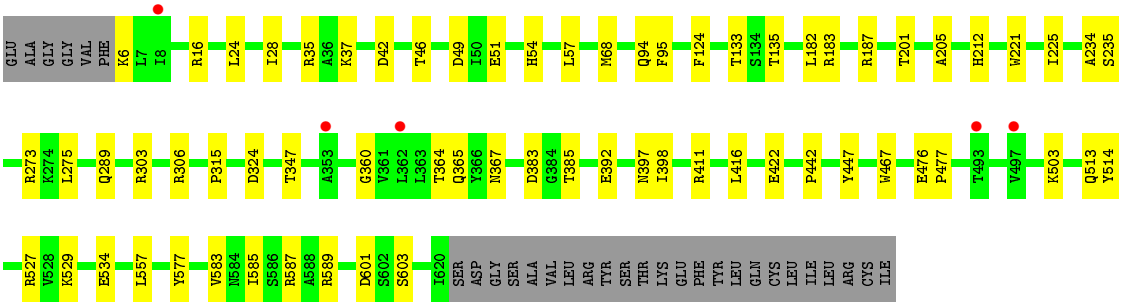
• Molecule 1: major capsid protein

Chain G:



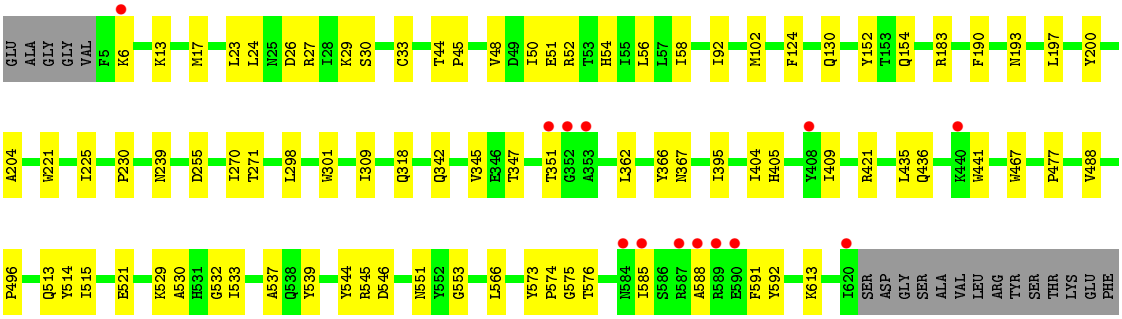
• Molecule 1: major capsid protein

Chain H:



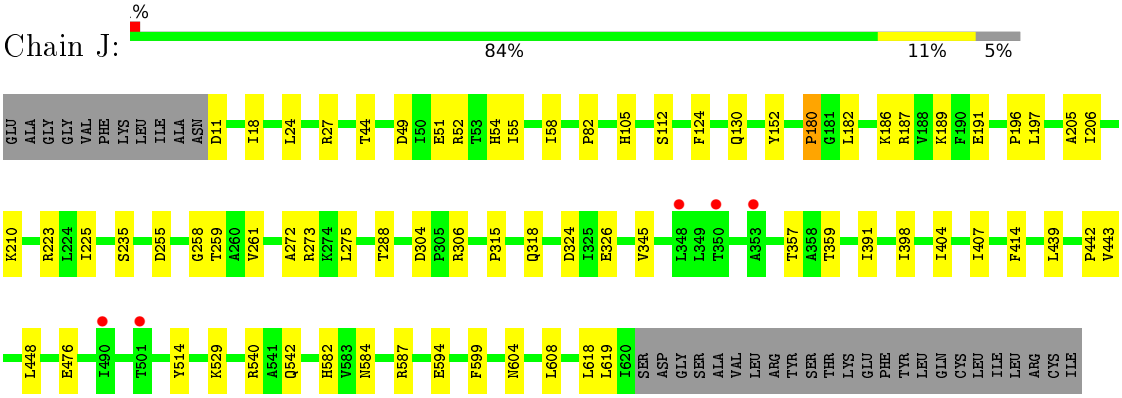
• Molecule 1: major capsid protein

Chain I:



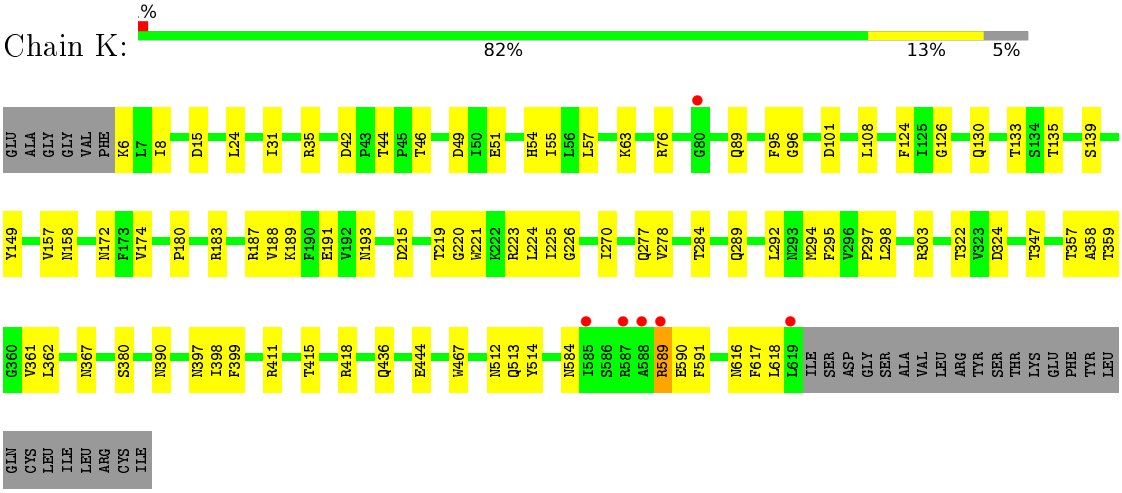
- Molecule 1: major capsid protein

Chain J:



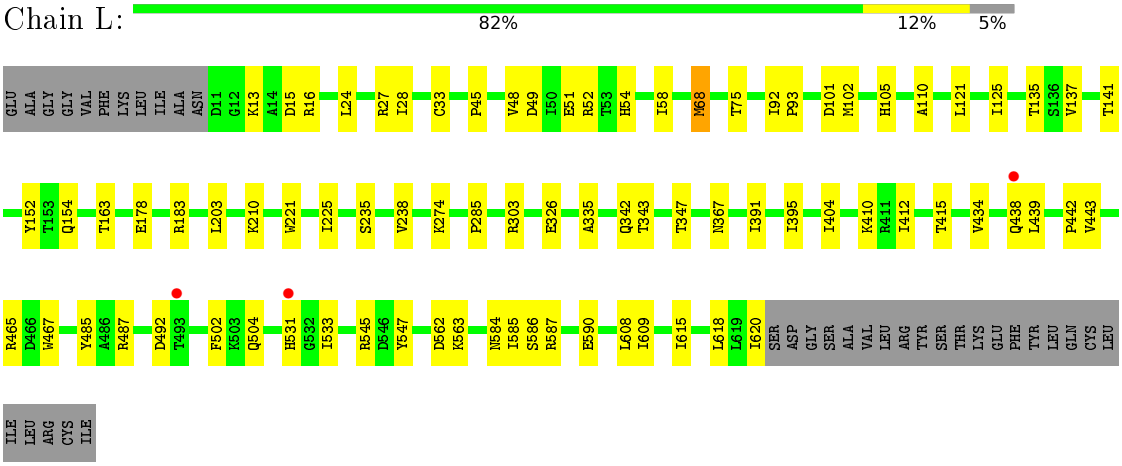
- Molecule 1: major capsid protein

Chain K:



- Molecule 1: major capsid protein

Chain L:



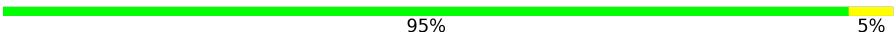
- Molecule 2: unknown

Chain M:

There are no outlier residues recorded for this chain.

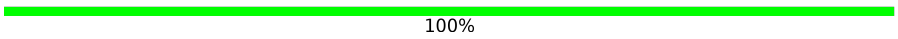
- Molecule 2: unknown

Chain N:



- Molecule 2: unknown

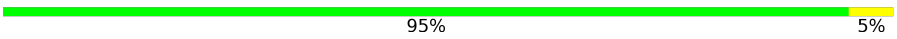
Chain O:



There are no outlier residues recorded for this chain.

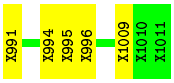
- Molecule 2: unknown

Chain P:



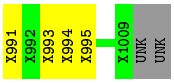
- Molecule 2: unknown

Chain Q:



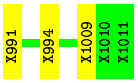
- Molecule 2: unknown

Chain R:



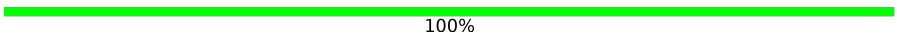
- Molecule 2: unknown

Chain S:



- Molecule 2: unknown

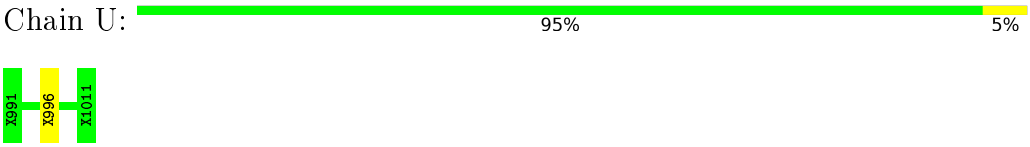
Chain T:



There are no outlier residues recorded for this chain.

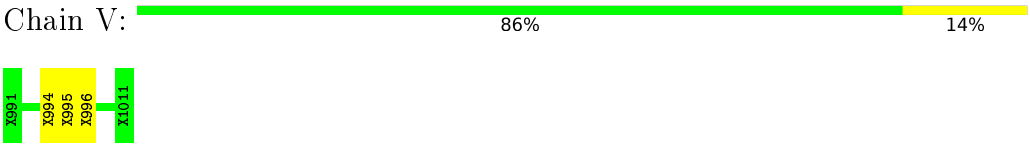
- Molecule 2: unknown

Chain U:



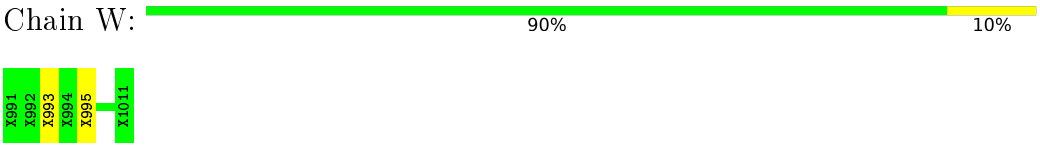
- Molecule 2: unknown

Chain V:



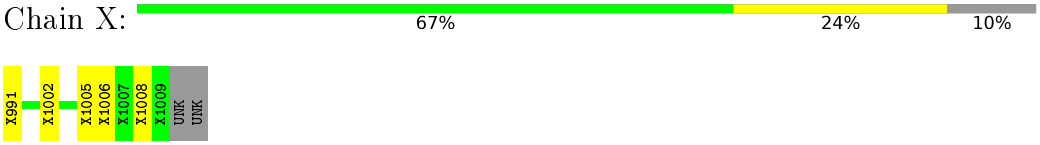
- Molecule 2: unknown

Chain W:



- Molecule 2: unknown

Chain X:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	86.38Å 390.83Å 130.14Å 90.00° 91.66° 90.00°	Depositor
Resolution (Å)	72.92 – 2.44 72.92 – 2.44	Depositor EDS
% Data completeness (in resolution range)	99.6 (72.92-2.44) 99.8 (72.92-2.44)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 2.45Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.205 , 0.252 0.201 , 0.252	Depositor DCC
R_{free} test set	15717 reflections (4.96%)	DCC
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 42.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.039 for h,-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	61035	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 46.29 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.1691e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0/4901	0.48	0/6697
1	B	0.26	0/4953	0.48	0/6767
1	C	0.27	0/4959	0.49	0/6775
1	D	0.26	0/4901	0.47	0/6697
1	E	0.26	0/4958	0.48	0/6774
1	F	0.26	0/4909	0.48	0/6708
1	G	0.27	0/4915	0.48	0/6716
1	H	0.26	0/4947	0.47	0/6759
1	I	0.26	0/4959	0.47	0/6775
1	J	0.26	0/4909	0.47	0/6708
1	K	0.27	0/4939	0.48	0/6748
1	L	0.26	0/4909	0.48	0/6708
All	All	0.26	0/59159	0.48	0/80832

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	4785	0	4711	72	0
1	B	4837	0	4766	61	0
1	C	4842	0	4770	63	0
1	D	4785	0	4711	65	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	4841	0	4768	72	0
1	F	4793	0	4715	57	0
1	G	4799	0	4720	56	0
1	H	4831	0	4761	47	0
1	I	4842	0	4770	59	0
1	J	4793	0	4715	55	0
1	K	4823	0	4750	61	0
1	L	4793	0	4715	57	0
2	M	105	0	24	0	0
2	N	105	0	24	1	0
2	O	105	0	24	0	0
2	P	105	0	24	1	0
2	Q	105	0	24	4	0
2	R	95	0	21	4	0
2	S	105	0	24	3	0
2	T	105	0	24	0	0
2	U	105	0	25	1	0
2	V	105	0	25	3	0
2	W	105	0	24	1	0
2	X	95	0	22	5	0
3	A	173	0	0	3	0
3	B	170	0	0	5	0
3	C	164	0	0	2	0
3	D	191	0	0	4	0
3	E	168	0	0	3	0
3	F	172	0	0	3	0
3	G	176	0	0	5	0
3	H	165	0	0	7	0
3	I	133	0	0	3	0
3	J	163	0	0	4	0
3	K	156	0	0	4	0
3	L	171	0	0	5	0
3	N	2	0	0	1	0
3	O	2	0	0	0	0
3	P	5	0	0	0	0
3	Q	2	0	0	1	0
3	R	2	0	0	2	0
3	S	1	0	0	1	0
3	U	7	0	0	0	0
3	V	3	0	0	0	0
3	W	2	0	0	1	0
3	X	3	0	0	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	61035	0	57157	628	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 5.

The worst 5 of 628 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:527:ARG:HH12	1:A:529:LYS:CE	1.25	1.44
1:D:62:TYR:CE1	1:E:416:LEU:CD1	2.02	1.43
1:D:62:TYR:CE1	1:E:416:LEU:HD13	1.53	1.40
1:A:527:ARG:NH1	1:A:529:LYS:HE3	1.36	1.36
1:A:351:THR:OG1	1:A:362:LEU:HD22	1.14	1.30

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	A	607/645 (94%)	593 (98%)	14 (2%)	0	100	100
1	B	614/645 (95%)	597 (97%)	17 (3%)	0	100	100
1	C	614/645 (95%)	598 (97%)	16 (3%)	0	100	100
1	D	607/645 (94%)	591 (97%)	16 (3%)	0	100	100
1	E	614/645 (95%)	597 (97%)	17 (3%)	0	100	100
1	F	608/645 (94%)	583 (96%)	25 (4%)	0	100	100
1	G	609/645 (94%)	592 (97%)	17 (3%)	0	100	100
1	H	613/645 (95%)	593 (97%)	20 (3%)	0	100	100
1	I	614/645 (95%)	596 (97%)	18 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	J	608/645 (94%)	592 (97%)	16 (3%)	0	100 100
1	K	612/645 (95%)	590 (96%)	21 (3%)	1 (0%)	52 64
1	L	608/645 (94%)	588 (97%)	20 (3%)	0	100 100
All	All	7328/7740 (95%)	7110 (97%)	217 (3%)	1 (0%)	100 100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	K	590	GLU

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	524/554 (95%)	522 (100%)	2 (0%)	93 96
1	B	530/554 (96%)	530 (100%)	0	100 100
1	C	530/554 (96%)	529 (100%)	1 (0%)	95 97
1	D	524/554 (95%)	523 (100%)	1 (0%)	95 97
1	E	530/554 (96%)	528 (100%)	2 (0%)	93 96
1	F	525/554 (95%)	522 (99%)	3 (1%)	90 94
1	G	526/554 (95%)	524 (100%)	2 (0%)	93 96
1	H	529/554 (96%)	528 (100%)	1 (0%)	95 97
1	I	530/554 (96%)	530 (100%)	0	100 100
1	J	525/554 (95%)	524 (100%)	1 (0%)	95 97
1	K	528/554 (95%)	525 (99%)	3 (1%)	90 94
1	L	525/554 (95%)	522 (99%)	3 (1%)	90 94
All	All	6326/6648 (95%)	6307 (100%)	19 (0%)	94 97

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	463	GLN
1	G	619	LEU
1	K	589	ARG
1	F	68	MET
1	L	68	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	81	ASN
1	L	318	GLN
1	H	513	GLN
1	C	193	ASN
1	F	318	GLN

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	609/645 (94%)	-0.18	1 (0%)	95	96	24, 39, 59, 72	0
1	B	616/645 (95%)	-0.18	3 (0%)	91	92	24, 39, 62, 93	0
1	C	616/645 (95%)	-0.16	9 (1%)	76	77	26, 40, 67, 112	0
1	D	609/645 (94%)	-0.14	5 (0%)	87	88	27, 40, 61, 84	0
1	E	616/645 (95%)	-0.12	11 (1%)	71	70	29, 42, 72, 120	0
1	F	610/645 (94%)	-0.07	7 (1%)	82	84	24, 39, 66, 127	0
1	G	611/645 (94%)	-0.10	3 (0%)	91	92	26, 41, 63, 93	0
1	H	615/645 (95%)	-0.16	5 (0%)	87	88	23, 40, 65, 98	0
1	I	616/645 (95%)	-0.07	13 (2%)	67	67	29, 46, 70, 123	0
1	J	610/645 (94%)	-0.17	5 (0%)	87	88	25, 39, 65, 85	0
1	K	614/645 (95%)	-0.17	6 (0%)	84	85	29, 43, 72, 106	0
1	L	610/645 (94%)	-0.10	3 (0%)	91	92	26, 41, 65, 117	0
2	M	0/21	-	-			-	-
2	N	0/21	-	-			-	-
2	O	0/21	-	-			-	-
2	P	0/21	-	-			-	-
2	Q	0/21	-	-			-	-
2	R	0/21	-	-			-	-
2	S	0/21	-	-			-	-
2	T	0/21	-	-			-	-
2	U	0/21	-	-			-	-
2	V	0/21	-	-			-	-
2	W	0/21	-	-			-	-
2	X	0/21	-	-			-	-
All	All	7352/7992 (91%)	-0.13	71 (0%)	84	85	23, 41, 65, 127	0

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	588	ALA	8.0
1	I	587	ARG	6.1
1	C	587	ARG	6.1
1	H	362	LEU	4.6
1	C	588	ALA	4.3

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.