wwPDB X-ray Structure Validation Summary Report (i)

Jan 30, 2023 – 01:27 pm GMT

PDB ID	:	7PA8
Title	:	JC polyomavirus VP1 in complex with Fab 27C2
Authors	:	Stroeh, L.J.; Harprecht, C.; Freytag, J.; Stehle, T.
Deposited on	:	2021-07-29
Resolution	:	3.15 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 (i)) were used in the production of this report:

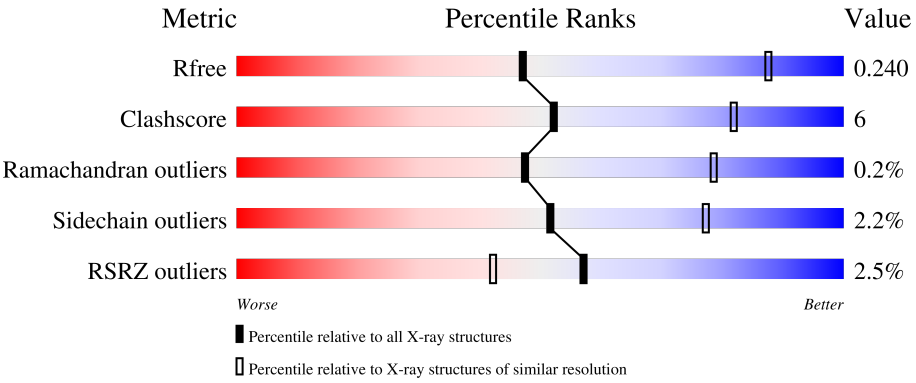
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.31.3
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 3.15 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.



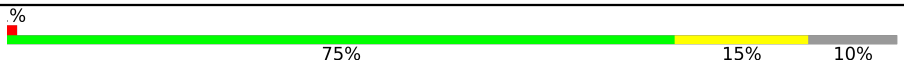
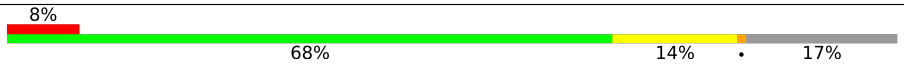
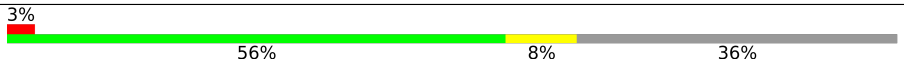
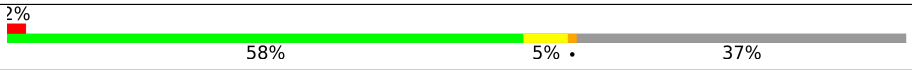
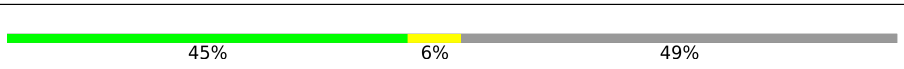
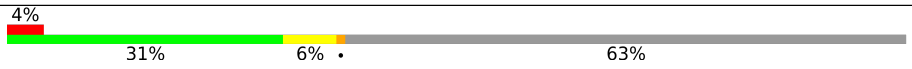
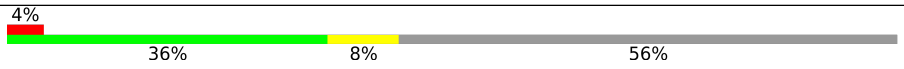
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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	AAA	272	81% 14% .
1	BBB	272	85% 10% 5%
1	CCC	272	85% 9% 6%
1	DDD	272	85% 11% .
1	EEE	272	82% 14% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	FFF	212	
2	GGG	212	
2	III	212	
2	KKK	212	
2	MMM	212	
3	HHH	398	
3	JJJ	398	
3	LLL	398	
3	NNN	398	
3	YYY	398	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 21266 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
1	AAA	260	Total	C	N	O	S	0	0	0
			1997	1263	343	380	11			
1	BBB	258	Total	C	N	O	S	0	0	0
			2004	1263	344	386	11			
1	CCC	257	Total	C	N	O	S	0	0	0
			1989	1255	340	383	11			
1	DDD	261	Total	C	N	O	S	0	0	0
			2010	1268	344	387	11			
1	EEE	262	Total	C	N	O	S	0	0	0
			2014	1268	346	389	11			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	18	GLY	-	expression tag	UNP P03089
AAA	19	SER	-	expression tag	UNP P03089
AAA	20	HIS	-	expression tag	UNP P03089
AAA	21	MET	-	expression tag	UNP P03089
BBB	18	GLY	-	expression tag	UNP P03089
BBB	19	SER	-	expression tag	UNP P03089
BBB	20	HIS	-	expression tag	UNP P03089
BBB	21	MET	-	expression tag	UNP P03089
CCC	18	GLY	-	expression tag	UNP P03089
CCC	19	SER	-	expression tag	UNP P03089
CCC	20	HIS	-	expression tag	UNP P03089
CCC	21	MET	-	expression tag	UNP P03089
DDD	18	GLY	-	expression tag	UNP P03089
DDD	19	SER	-	expression tag	UNP P03089
DDD	20	HIS	-	expression tag	UNP P03089
DDD	21	MET	-	expression tag	UNP P03089
EEE	18	GLY	-	expression tag	UNP P03089
EEE	19	SER	-	expression tag	UNP P03089
EEE	20	HIS	-	expression tag	UNP P03089

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
EEE	21	MET	-	expression tag	UNP P03089

- Molecule 2 is a protein called Fab 27C2 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	FFF	191	Total	C	N	O	S	0	0	0
			1329	827	223	275	4			
2	GGG	176	Total	C	N	O	S	0	0	0
			1213	755	208	246	4			
2	III	136	Total	C	N	O	S	0	0	0
			952	594	163	192	3			
2	KKK	134	Total	C	N	O	S	0	0	0
			952	597	162	191	2			
2	MMM	108	Total	C	N	O	S	0	0	0
			779	485	133	159	2			

- Molecule 3 is a protein called Fab 27C2 heavy chain.

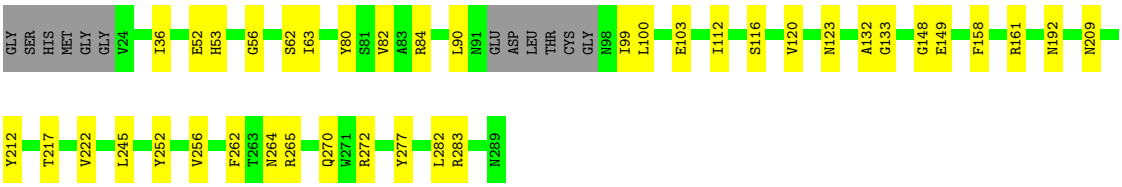
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	HHH	149	Total	C	N	O	S	0	0	0
			1125	714	186	220	5			
3	JJJ	177	Total	C	N	O	S	0	0	0
			1310	827	219	257	7			
3	LLL	170	Total	C	N	O	S	0	0	0
			1267	796	214	251	6			
3	NNN	118	Total	C	N	O	S	0	0	0
			906	571	155	175	5			
3	YYY	189	Total	C	N	O	S	0	0	0
			1419	902	239	271	7			

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.

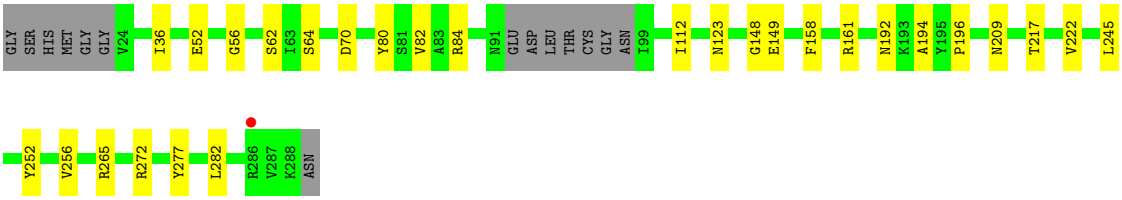
• Molecule 1: Major capsid protein VP1

Chain AAA:



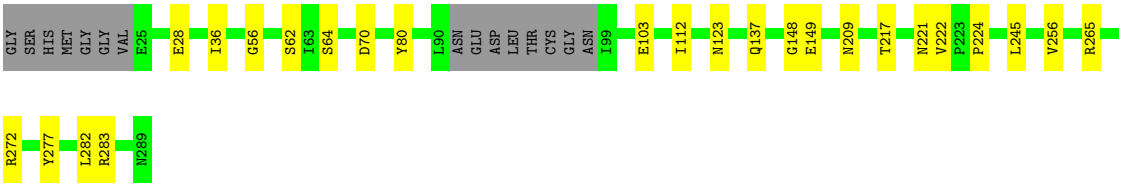
• Molecule 1: Major capsid protein VP1

Chain BBB:



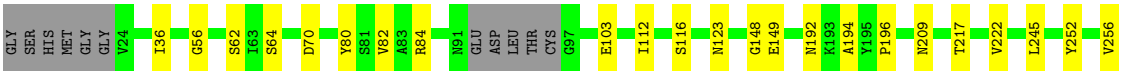
• Molecule 1: Major capsid protein VP1

Chain CCC:



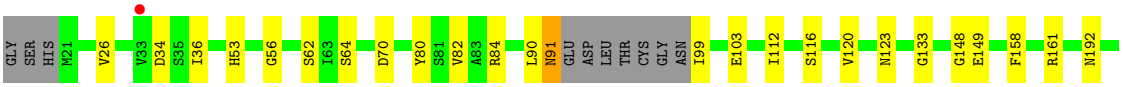
• Molecule 1: Major capsid protein VP1

Chain DDD:



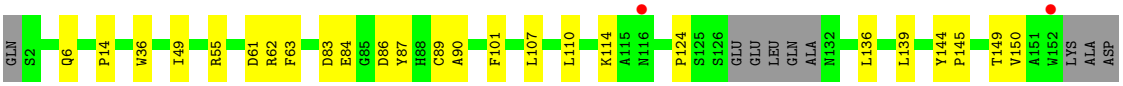
• Molecule 1: Major capsid protein VP1

Chain EEE:



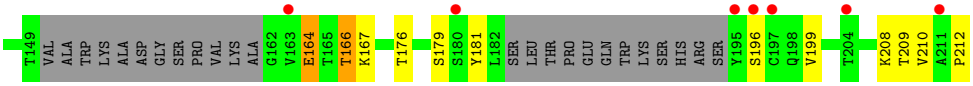
• Molecule 2: Fab 27C2 light chain

Chain FFF:



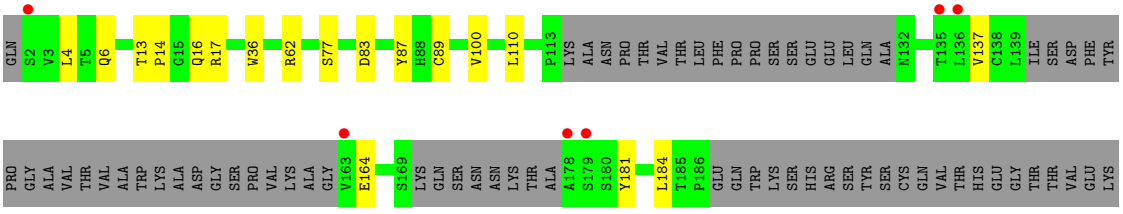
• Molecule 2: Fab 27C2 light chain

Chain GGG:



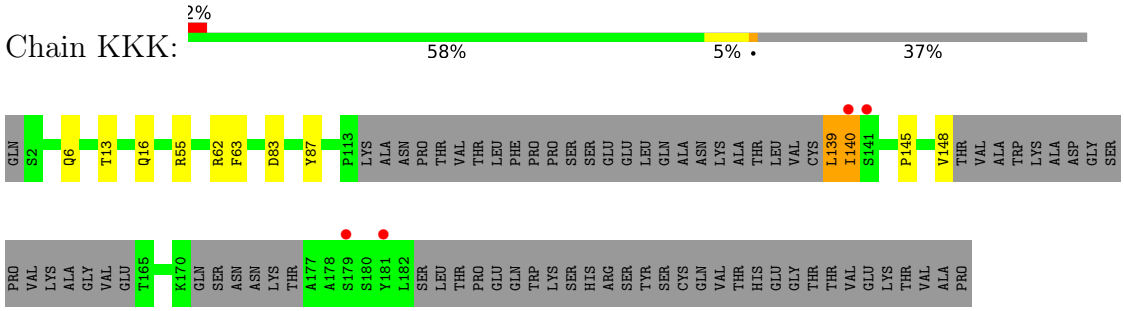
• Molecule 2: Fab 27C2 light chain

Chain III:



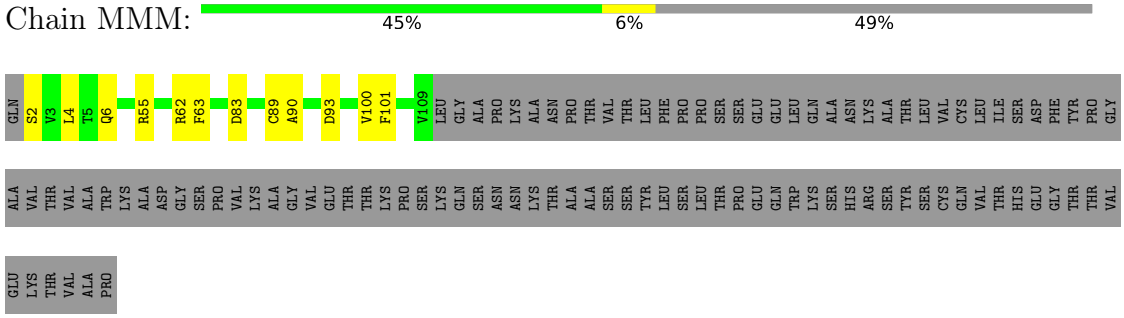
• Molecule 2: Fab 27C2 light chain

Chain KKK:



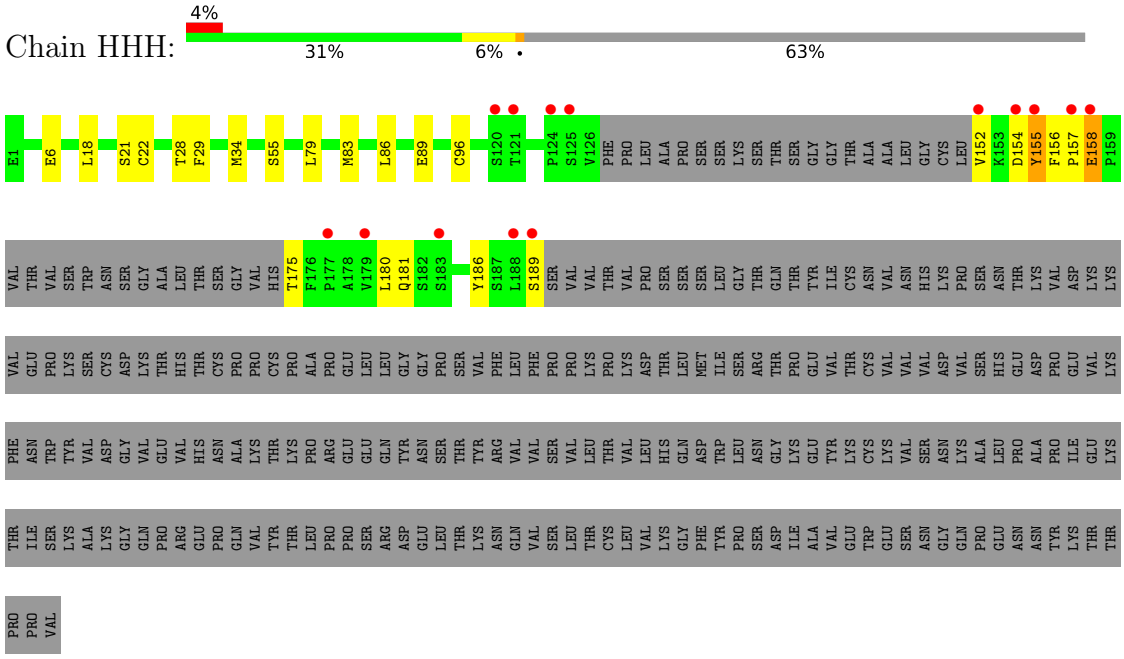
• Molecule 2: Fab 27C2 light chain

Chain MMM:



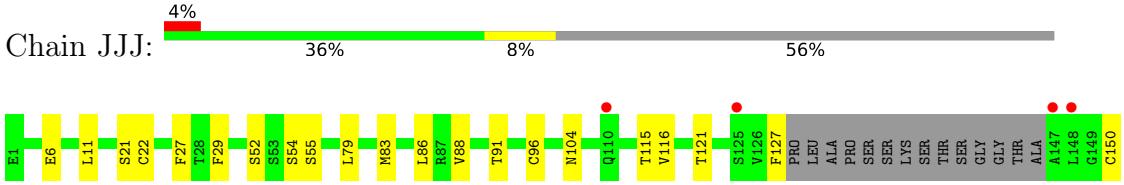
• Molecule 3: Fab 27C2 heavy chain

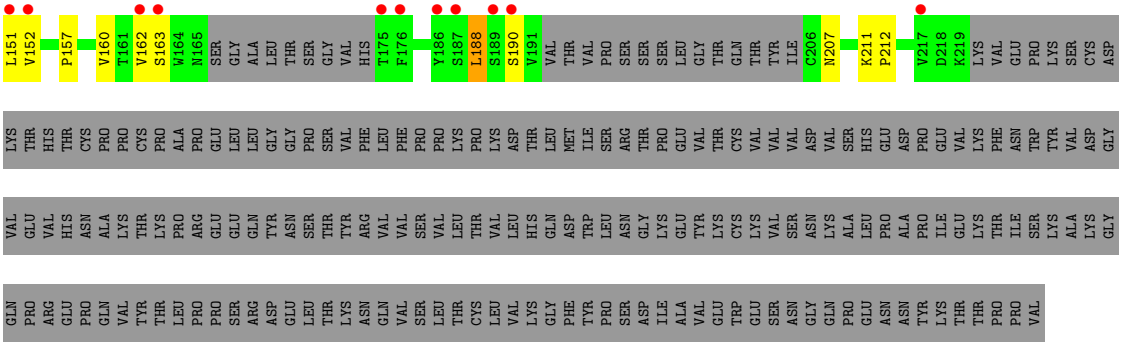
Chain HHH:



• Molecule 3: Fab 27C2 heavy chain

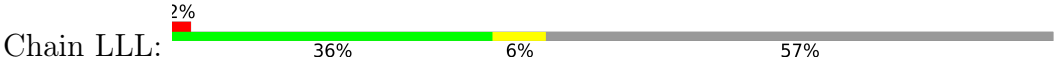
Chain JJJ:

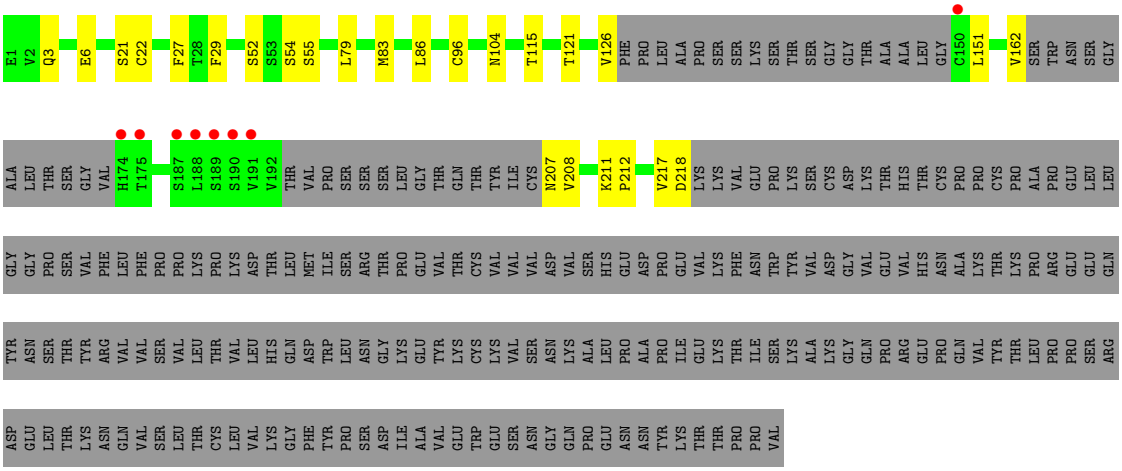




• Molecule 3: Fab 27C2 heavy chain

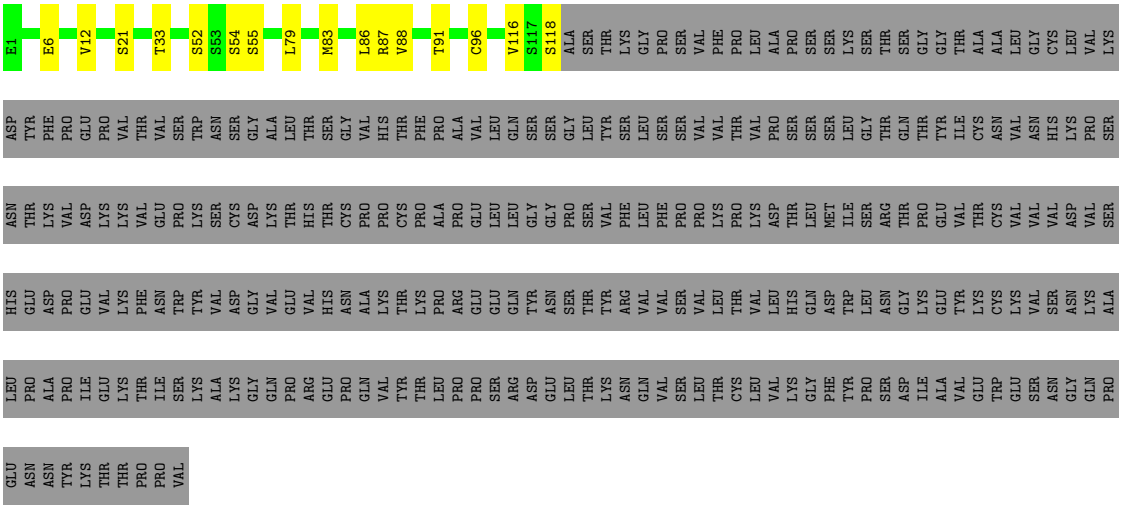
Chain LLL:





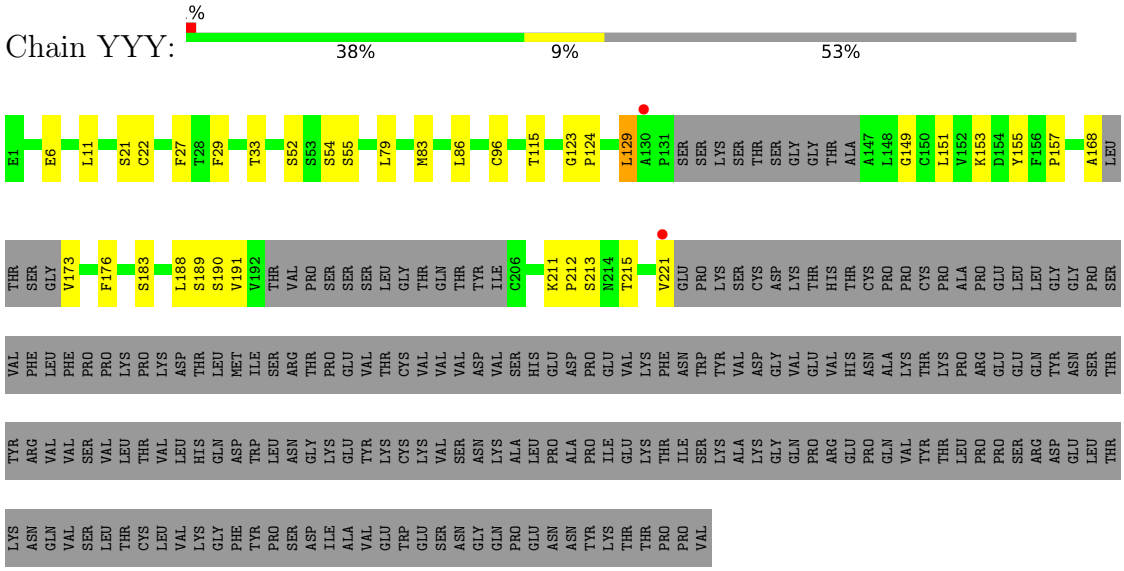
• Molecule 3: Fab 27C2 heavy chain

Chain NNN:



• Molecule 3: Fab 27C2 heavy chain

Chain YYY:


%
38%
9%
53%
E1 E6 L11 S21 C22 F27 T28 F29 T33 S52 S53 S54 S55 L79 M83 L86 C96 T115 G123 P124 L129 A130 P131 SER SER LYS SER THR SER GLY GLY THR ALA A147 L148 G149 C150 L151 V152 K153 D154 Y155 F156 P157 A168 LEU
THR SER GLY V173 F176 S183 L188 S189 S190 V191 V192 THR VAL PRO SER SER SER LEU GLY THR GLN THR TYR ILE C206 K211 P212 S213 N214 T215 V221 GLU PRO LYS SER CYS ASP LYS THR HIS THR CYS PRO PRO CYS PRO ALA PRO GLU LEU LEU GLY GLY PRO SER
VAL PHE LEU PHE PRO PRO LYS PRO LYS ASP THR LEU MET ILE SER ARG THR PRO GLU VAL THR CYS VAL VAL VAL ASP VAL SER HIS GLU ASP PRO GLU VAL LYS PHE ASN TRP TYR VAL ASP GLY VAL GLU VAL HIS ASN ALA LYS THR LYS PRO ARG GLU GLU GLN TYR ASN SER THR
TYR ARG VAL VAL SER VAL LEU THR VAL LEU HIS GLN ASP TRP LEU ASN GLY LYS GLU TYR LYS CYS LYS VAL SER ASN LYS ALA LEU PRO ALA PRO ILE GLU LYS THR ILE SER LYS ALA LYS GLY GLN PRO ARG GLU PRO GLN VAL TYR THR LEU PRO PRO SER ARG ASP GLU LEU THR
LYS ASN GLN VAL SER LEU THR CYS LEU VAL LYS GLY PHE TYR PRO SER ASP ILE ALA VAL GLU TRP GLU SER ASN GLY GLN PRO GLU ASN ASN TYR LYS THR THR PRO PRO VAL

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	137.07Å 110.44Å 158.21Å 90.00° 93.11° 90.00°	Depositor
Resolution (Å)	49.01 – 3.15 49.01 – 3.15	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.01-3.15) 99.7 (49.01-3.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.29 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.212 , 0.241 0.213 , 0.240	Depositor DCC
R_{free} test set	4079 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	47.5	Xtriage
Anisotropy	0.886	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 36.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	21266	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.35% 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	AAA	0.69	0/2042	0.87	0/2777
1	BBB	0.70	0/2049	0.86	0/2784
1	CCC	0.69	0/2034	0.88	0/2765
1	DDD	0.70	0/2055	0.89	0/2794
1	EEE	0.69	0/2059	0.88	0/2799
2	FFF	0.75	0/1359	0.89	1/1867 (0.1%)
2	GGG	0.77	0/1240	0.89	0/1701
2	III	0.73	0/972	0.85	0/1325
2	KKK	0.71	0/974	0.85	0/1327
2	MMM	0.75	0/799	0.89	0/1089
3	HHH	0.72	0/1153	0.89	0/1568
3	JJJ	0.74	0/1342	0.85	0/1832
3	LLL	0.73	0/1296	0.87	0/1768
3	NNN	0.71	0/928	0.90	0/1259
3	YYY	0.73	0/1454	0.91	0/1980
All	All	0.72	0/21756	0.88	1/29635 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	FFF	86	ASP	CB-CG-OD2	5.20	122.98	118.30

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	AAA	1997	0	1935	37	0
1	BBB	2004	0	1945	19	0
1	CCC	1989	0	1928	13	0
1	DDD	2010	0	1944	16	0
1	EEE	2014	0	1942	29	0
2	FFF	1329	0	1234	18	0
2	GGG	1213	0	1119	26	0
2	III	952	0	894	12	0
2	KKK	952	0	897	6	0
2	MMM	779	0	734	7	0
3	HHH	1125	0	1048	11	0
3	JJJ	1310	0	1186	17	0
3	LLL	1267	0	1167	11	0
3	NNN	906	0	850	9	0
3	YYY	1419	0	1346	22	0
All	All	21266	0	20169	229	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 6.

The worst 5 of 229 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:AAA:161:ARG:HH22	1:EEE:53:HIS:HB3	1.16	1.02
1:AAA:53:HIS:HB3	1:BBB:161:ARG:HH22	1.29	0.95
3:NNN:87:ARG:O	3:NNN:116:VAL:HG11	1.66	0.95
1:CCC:36:ILE:HD12	1:CCC:282:LEU:O	1.65	0.93
2:III:6:GLN:OE1	2:III:89:CYS:SG	2.27	0.92

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	256/272 (94%)	247 (96%)	9 (4%)	0	100	100
1	BBB	254/272 (93%)	244 (96%)	10 (4%)	0	100	100
1	CCC	253/272 (93%)	239 (94%)	14 (6%)	0	100	100
1	DDD	257/272 (94%)	244 (95%)	13 (5%)	0	100	100
1	EEE	258/272 (95%)	244 (95%)	14 (5%)	0	100	100
2	FFF	183/212 (86%)	166 (91%)	16 (9%)	1 (0%)	29	65
2	GGG	168/212 (79%)	155 (92%)	12 (7%)	1 (1%)	25	62
2	III	128/212 (60%)	122 (95%)	6 (5%)	0	100	100
2	KKK	126/212 (59%)	116 (92%)	10 (8%)	0	100	100
2	MMM	106/212 (50%)	102 (96%)	4 (4%)	0	100	100
3	HHH	143/398 (36%)	133 (93%)	8 (6%)	2 (1%)	11	43
3	JJJ	169/398 (42%)	157 (93%)	11 (6%)	1 (1%)	25	62
3	LLL	162/398 (41%)	149 (92%)	12 (7%)	1 (1%)	25	62
3	NNN	116/398 (29%)	110 (95%)	6 (5%)	0	100	100
3	YYY	181/398 (46%)	167 (92%)	14 (8%)	0	100	100
All	All	2760/4410 (63%)	2595 (94%)	159 (6%)	6 (0%)	47	78

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	HHH	157	PRO
2	GGG	142	ASP
3	HHH	158	GLU
3	JJJ	104	ASN
3	LLL	104	ASN

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	219/237 (92%)	218 (100%)	1 (0%)	88	95

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	BBB	223/237 (94%)	222 (100%)	1 (0%)	91 96
1	CCC	221/237 (93%)	219 (99%)	2 (1%)	78 91
1	DDD	222/237 (94%)	220 (99%)	2 (1%)	78 91
1	EEE	222/237 (94%)	217 (98%)	5 (2%)	50 76
2	FFF	144/175 (82%)	138 (96%)	6 (4%)	30 62
2	GGG	127/175 (73%)	122 (96%)	5 (4%)	32 64
2	III	102/175 (58%)	102 (100%)	0	100 100
2	KKK	102/175 (58%)	98 (96%)	4 (4%)	32 64
2	MMM	86/175 (49%)	86 (100%)	0	100 100
3	HHH	121/352 (34%)	114 (94%)	7 (6%)	20 52
3	JJJ	139/352 (40%)	135 (97%)	4 (3%)	42 72
3	LLL	138/352 (39%)	133 (96%)	5 (4%)	35 67
3	NNN	97/352 (28%)	94 (97%)	3 (3%)	40 70
3	YYY	155/352 (44%)	149 (96%)	6 (4%)	32 64
All	All	2318/3820 (61%)	2267 (98%)	51 (2%)	52 77

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	HHH	175	THR
2	KKK	145	PRO
3	YYY	129	LEU
3	JJJ	54	SER
3	JJJ	188	LEU

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)

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	AAA	260/272 (95%)	-0.43	0 100 100	21, 34, 70, 115	0
1	BBB	258/272 (94%)	-0.38	1 (0%) 92 89	20, 36, 74, 101	0
1	CCC	257/272 (94%)	-0.49	0 100 100	18, 33, 54, 71	0
1	DDD	261/272 (95%)	-0.47	0 100 100	18, 31, 57, 116	0
1	EEE	262/272 (96%)	-0.42	1 (0%) 92 89	20, 33, 77, 131	0
2	FFF	191/212 (90%)	-0.05	2 (1%) 82 73	24, 60, 101, 121	0
2	GGG	176/212 (83%)	0.43	18 (10%) 6 3	30, 74, 142, 165	0
2	III	136/212 (64%)	0.07	6 (4%) 34 20	35, 69, 110, 124	0
2	KKK	134/212 (63%)	0.00	4 (2%) 50 33	29, 52, 103, 123	0
2	MMM	108/212 (50%)	-0.14	0 100 100	28, 55, 87, 97	0
3	HHH	149/398 (37%)	0.05	14 (9%) 8 4	26, 45, 110, 140	0
3	JJJ	177/398 (44%)	0.56	15 (8%) 10 5	29, 80, 127, 146	0
3	LLL	170/398 (42%)	-0.11	8 (4%) 31 17	23, 49, 103, 121	0
3	NNN	118/398 (29%)	-0.36	0 100 100	24, 43, 60, 81	0
3	YYY	189/398 (47%)	-0.31	2 (1%) 80 70	22, 40, 93, 104	0
All	All	2846/4410 (64%)	-0.18	71 (2%) 57 42	18, 42, 105, 165	0

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	GGG	123	PRO	4.5
3	JJJ	175	THR	4.3
3	HHH	154	ASP	4.0
3	HHH	125	SER	3.6
3	JJJ	189	SER	3.5

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.