



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:41 PM GMT

PDB ID : 4PLK
Title : Hepatitis E Virus E2s domain (Genotype I) in complex with a neutralizing antibody 8G12
Authors : Tang, X.H.; Li, S.W.; Sivaraman, J.
Deposited on : 2014-05-18
Resolution : 4.00 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

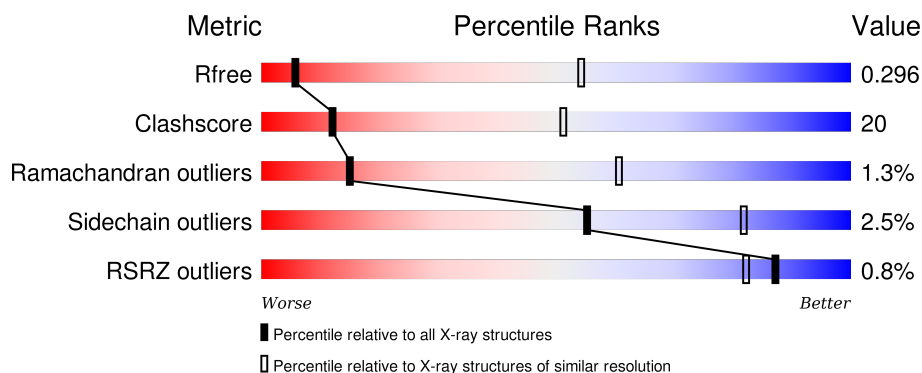
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.00 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	1010 (4.42-3.56)
Clashscore	102246	1052 (4.40-3.60)
Ramachandran outliers	100387	1005 (4.40-3.60)
Sidechain outliers	100360	1013 (4.42-3.58)
RSRZ outliers	91569	1013 (4.42-3.56)

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 $\leq 5\%$. 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	147	<div> <div>57%</div> <div>42%</div> <div>.</div> </div>
1	B	147	<div> <div>67%</div> <div>32%</div> <div>.</div> </div>
1	E	147	<div> <div>%</div> <div>64%</div> <div>35%</div> <div>.</div> </div>
1	F	147	<div> <div>55%</div> <div>44%</div> <div>.</div> </div>
2	C	212	<div> <div>46%</div> <div>42%</div> <div>.</div> <div>7%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	212	<div><div></div><div>58%32%7%</div></div>
2	K	212	<div><div></div><div>62%33%</div></div>
2	L	212	<div><div>2%</div><div></div><div>61%34%</div></div>
3	D	229	<div><div>3%</div><div></div><div>65%29%5%</div></div>
3	H	229	<div><div></div><div>63%29%6%</div></div>
3	I	229	<div><div>%</div><div></div><div>64%26%5%</div></div>
3	J	229	<div><div></div><div>66%27%6%</div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17362 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 Capsid protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	147	Total	C	N	O	0	0	0
			1104	702	186	216			
1	B	147	Total	C	N	O	0	0	0
			1104	702	186	216			
1	E	147	Total	C	N	O	0	0	0
			1104	702	186	216			
1	F	147	Total	C	N	O	0	0	0
			1104	702	186	216			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	604	PRO	-	expression tag	UNP L0L7P5
A	605	PRO	-	expression tag	UNP L0L7P5
B	604	PRO	-	expression tag	UNP L0L7P5
B	605	PRO	-	expression tag	UNP L0L7P5
E	604	PRO	-	expression tag	UNP L0L7P5
E	605	PRO	-	expression tag	UNP L0L7P5
F	604	PRO	-	expression tag	UNP L0L7P5
F	605	PRO	-	expression tag	UNP L0L7P5

- Molecule 2 is a protein called 8G12 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	L	208	Total	C	N	O	S	0	0	0
			1618	1010	271	331	6			
2	C	198	Total	C	N	O	S	0	0	0
			1533	959	257	312	5			
2	K	208	Total	C	N	O	S	0	0	0
			1618	1010	271	331	6			
2	G	198	Total	C	N	O	S	0	0	0
			1531	957	257	312	5			

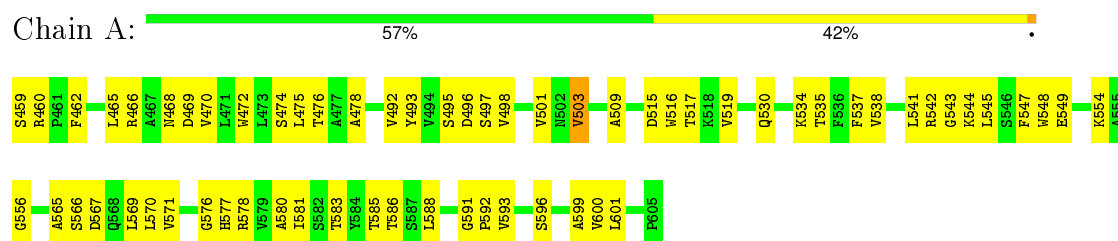
- Molecule 3 is a protein called 8G12 heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	216	Total	C	N	O	S	0	0	0
			1654	1056	271	320	7			
3	D	218	Total	C	N	O	S	0	0	0
			1669	1064	274	324	7			
3	J	216	Total	C	N	O	S	0	0	0
			1654	1056	271	320	7			
3	I	218	Total	C	N	O	S	0	0	0
			1669	1064	274	324	7			

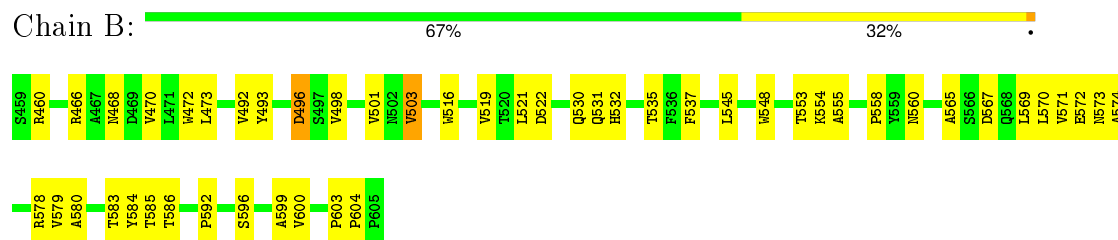
3 Residue-property plots [i](#)

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.

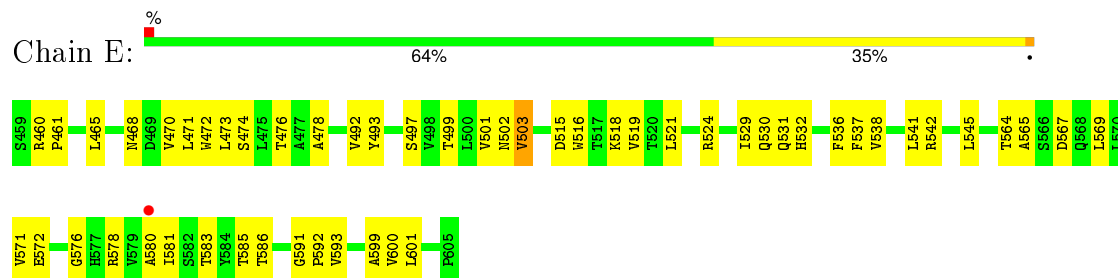
• Molecule 1: Capsid protein



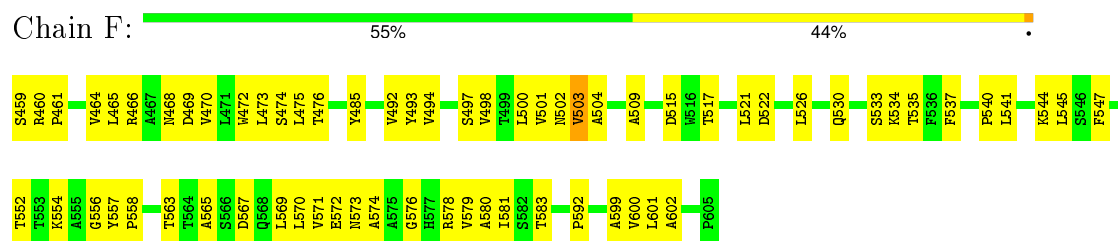
• Molecule 1: Capsid protein



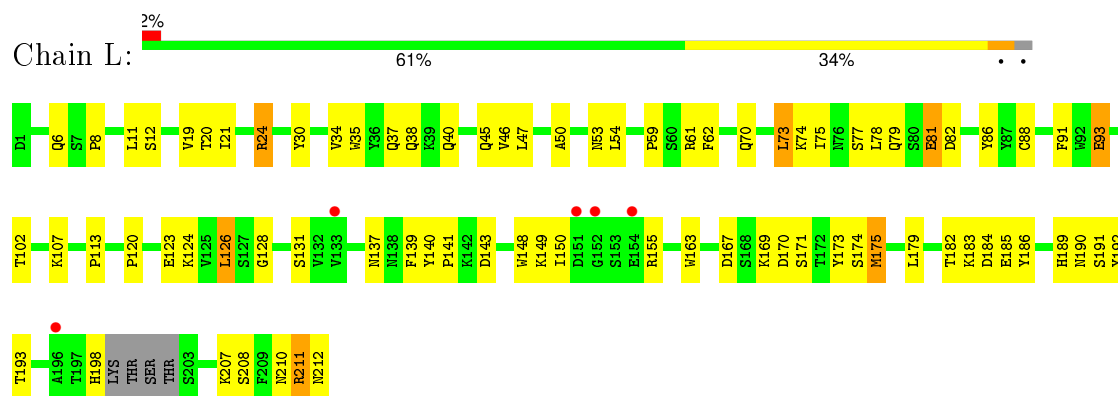
• Molecule 1: Capsid protein



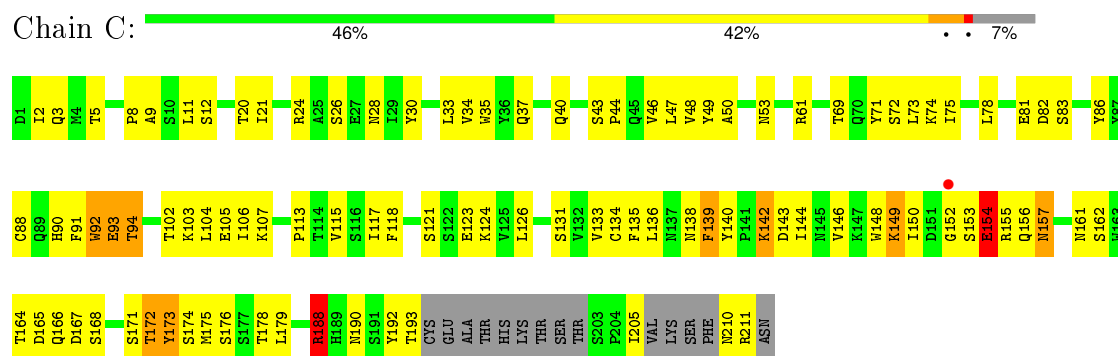
• Molecule 1: Capsid protein



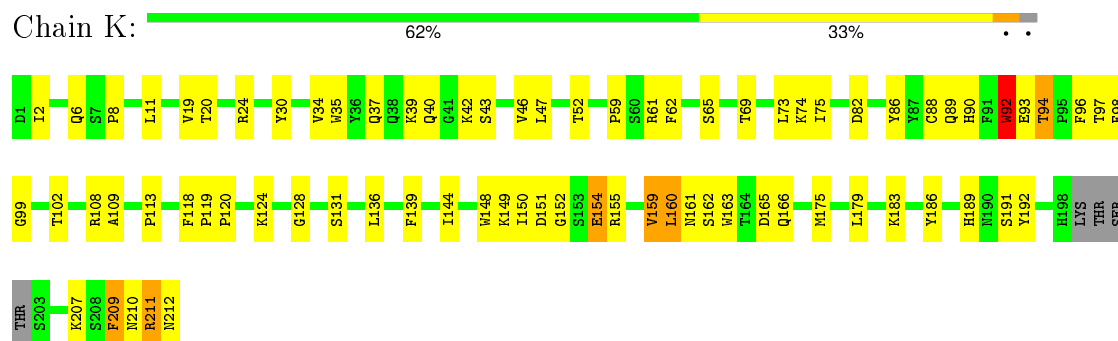
- Molecule 2: 8G12 light chain



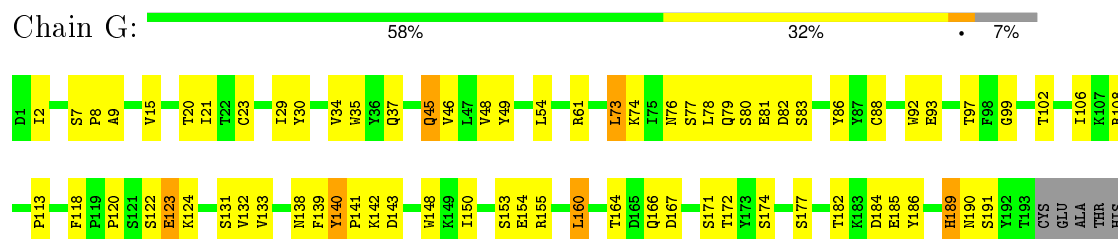
- Molecule 2: 8G12 light chain



- Molecule 2: 8G12 light chain

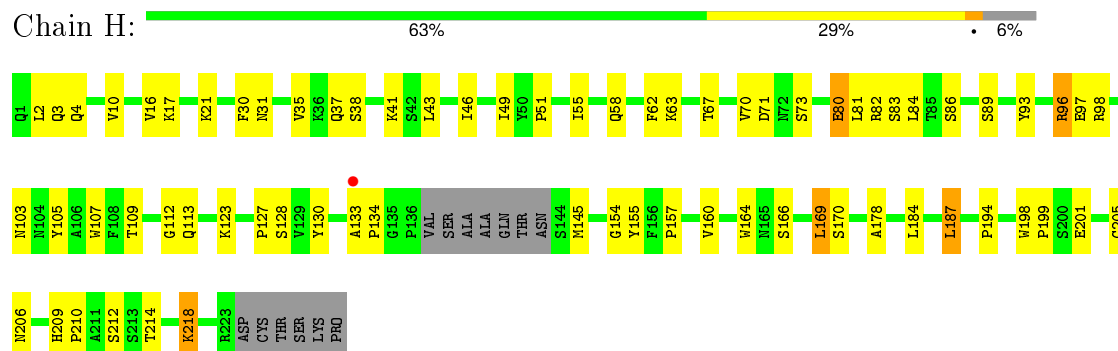


- Molecule 2: 8G12 light chain

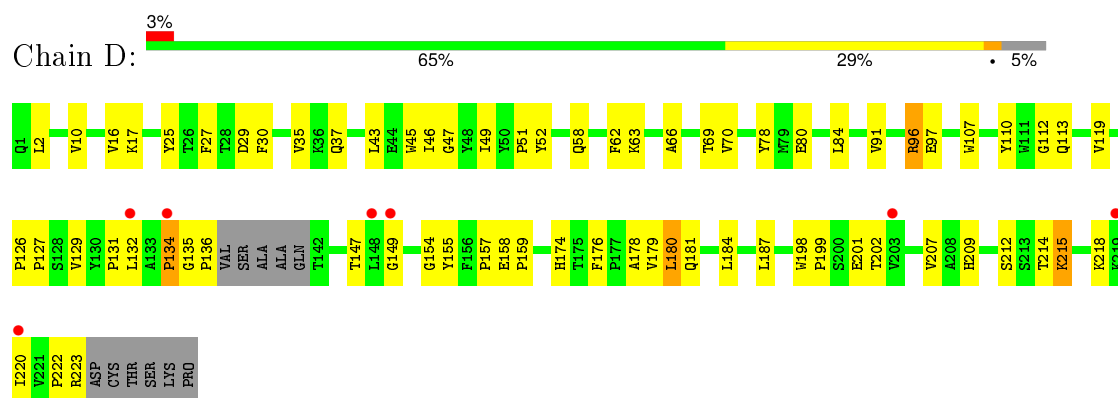


LYS	THR	SER	THR	S203	P204	T205	VAL	LYS	SER	PHE	R210	R211	ASN
-----	-----	-----	-----	------	------	------	-----	-----	-----	-----	------	------	-----

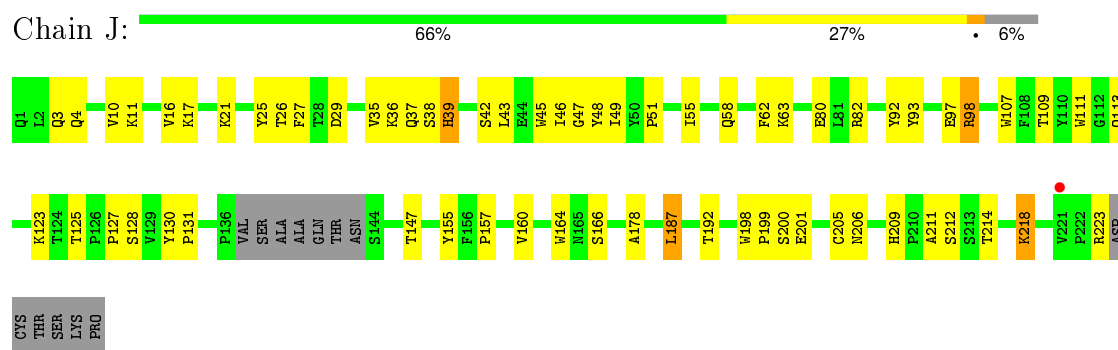
• Molecule 3: 8G12 heavy chain



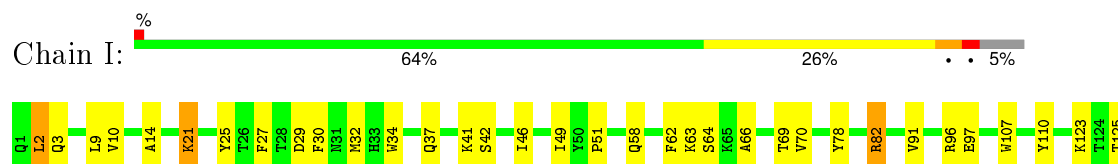
• Molecule 3: 8G12 heavy chain

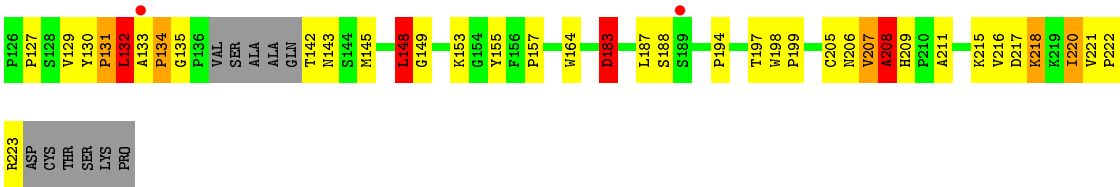


• Molecule 3: 8G12 heavy chain



• Molecule 3: 8G12 heavy chain





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	54.20 Å 191.99 Å 192.58 Å 90.00° 90.03° 90.00°	Depositor
Resolution (Å)	33.99 – 4.00 33.99 – 3.90	Depositor EDS
% Data completeness (in resolution range)	91.2 (33.99-4.00) 79.8 (33.99-3.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.65 (at 3.87 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.264 , 0.296 0.270 , 0.296	Depositor DCC
R_{free} test set	1701 reflections (6.17%)	DCC
Wilson B-factor (Å ²)	81.1	Xtriage
Anisotropy	0.567	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 42.8	EDS
Estimated twinning fraction	0.358 for -h,l,k 0.360 for -h,-l,-k 0.380 for h,-k,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Outliers	0 of 31796 reflections	Xtriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	17362	wwPDB-VP
Average B, all atoms (Å ²)	141.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% 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, 0.375 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	0.29	0/1131	0.62	0/1552
1	B	0.30	0/1131	0.61	0/1552
1	E	0.32	0/1131	0.66	1/1552 (0.1%)
1	F	0.31	0/1131	0.65	0/1552
2	C	0.36	0/1567	0.84	3/2126 (0.1%)
2	G	0.37	0/1565	0.87	2/2122 (0.1%)
2	K	0.34	0/1655	0.81	4/2246 (0.2%)
2	L	0.33	0/1655	0.76	3/2246 (0.1%)
3	D	0.34	0/1717	0.73	2/2346 (0.1%)
3	H	0.28	0/1702	0.63	2/2325 (0.1%)
3	I	0.33	0/1717	0.86	9/2346 (0.4%)
3	J	0.29	0/1702	0.67	1/2325 (0.0%)
All	All	0.32	0/17804	0.74	27/24290 (0.1%)

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
2	C	0	1
2	G	0	1
2	K	0	2
3	I	0	2
All	All	0	6

There are no bond length outliers.

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	K	92	TRP	CB-CA-C	-9.47	91.46	110.40
3	I	132	LEU	N-CA-C	9.33	136.19	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	I	131	PRO	C-N-CA	8.04	141.80	121.70
3	I	131	PRO	CA-C-N	-7.84	99.96	117.20
3	I	131	PRO	N-CA-C	7.21	130.85	112.10

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	172	THR	Peptide
2	G	141	PRO	Peptide
3	I	207	VAL	Peptide
2	K	159	VAL	Peptide
2	K	209	PHE	Peptide

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	A	1104	0	1092	65	0
1	B	1104	0	1090	53	0
1	E	1104	0	1092	40	0
1	F	1104	0	1092	56	0
2	C	1533	0	1464	103	0
2	G	1531	0	1456	62	0
2	K	1618	0	1544	75	0
2	L	1618	0	1542	95	0
3	D	1669	0	1624	64	0
3	H	1654	0	1611	51	0
3	I	1669	0	1624	64	0
3	J	1654	0	1611	52	0
All	All	17362	0	16842	680	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 20.

The worst 5 of 680 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:554:LYS:CD	2:C:92:TRP:CZ2	2.15	1.27
1:B:554:LYS:HD2	2:C:92:TRP:CE2	1.70	1.24
1:B:554:LYS:CD	2:C:92:TRP:CE2	2.19	1.23
1:B:554:LYS:HD2	2:C:92:TRP:CZ2	1.79	1.16
1:A:554:LYS:HE2	2:L:93:GLU:HG3	1.23	1.15

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	145/147 (99%)	139 (96%)	4 (3%)	2 (1%)	14	59
1	B	145/147 (99%)	140 (97%)	4 (3%)	1 (1%)	26	71
1	E	145/147 (99%)	138 (95%)	6 (4%)	1 (1%)	26	71
1	F	145/147 (99%)	136 (94%)	6 (4%)	3 (2%)	9	53
2	C	192/212 (91%)	176 (92%)	11 (6%)	5 (3%)	7	48
2	G	192/212 (91%)	176 (92%)	13 (7%)	3 (2%)	12	57
2	K	204/212 (96%)	193 (95%)	8 (4%)	3 (2%)	13	58
2	L	204/212 (96%)	192 (94%)	9 (4%)	3 (2%)	13	58
3	D	214/229 (93%)	202 (94%)	10 (5%)	2 (1%)	21	66
3	H	212/229 (93%)	208 (98%)	4 (2%)	0	100	100
3	I	214/229 (93%)	205 (96%)	5 (2%)	4 (2%)	10	54
3	J	212/229 (93%)	207 (98%)	3 (1%)	2 (1%)	21	66
All	All	2224/2352 (95%)	2112 (95%)	83 (4%)	29 (1%)	15	60

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	577	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	503	VAL
2	L	169	LYS
2	C	154	GLU
2	C	157	ASN

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	120/120 (100%)	119 (99%)	1 (1%)	86	93
1	B	120/120 (100%)	118 (98%)	2 (2%)	68	88
1	E	120/120 (100%)	118 (98%)	2 (2%)	68	88
1	F	120/120 (100%)	119 (99%)	1 (1%)	86	93
2	C	175/190 (92%)	167 (95%)	8 (5%)	33	70
2	G	175/190 (92%)	170 (97%)	5 (3%)	50	79
2	K	186/190 (98%)	184 (99%)	2 (1%)	80	91
2	L	186/190 (98%)	182 (98%)	4 (2%)	60	84
3	D	190/199 (96%)	187 (98%)	3 (2%)	70	88
3	H	188/199 (94%)	181 (96%)	7 (4%)	41	75
3	I	190/199 (96%)	180 (95%)	10 (5%)	28	67
3	J	188/199 (94%)	185 (98%)	3 (2%)	70	88
All	All	1958/2036 (96%)	1910 (98%)	48 (2%)	55	82

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

Mol	Chain	Res	Type
2	C	188	ARG
1	E	530	GLN
3	I	183	ASP
3	D	58	GLN
3	D	215	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	138	ASN
1	E	577	HIS
3	I	3	GLN
3	D	174	HIS
1	F	508	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 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 ⓘ

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, 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	147/147 (100%)	-0.25	0 100 100	73, 103, 146, 165	0
1	B	147/147 (100%)	-0.19	0 100 100	79, 104, 153, 193	0
1	E	147/147 (100%)	-0.17	1 (0%) 89 84	81, 108, 159, 180	0
1	F	147/147 (100%)	-0.28	0 100 100	68, 107, 158, 175	0
2	C	198/212 (93%)	-0.23	1 (0%) 91 88	55, 169, 218, 283	0
2	G	198/212 (93%)	-0.19	0 100 100	51, 167, 208, 237	0
2	K	208/212 (98%)	-0.23	0 100 100	63, 167, 209, 234	0
2	L	208/212 (98%)	-0.09	5 (2%) 62 51	49, 163, 234, 272	0
3	D	218/229 (95%)	-0.15	7 (3%) 51 38	96, 135, 210, 233	0
3	H	216/229 (94%)	-0.24	1 (0%) 91 88	85, 128, 202, 253	0
3	I	218/229 (95%)	-0.25	2 (0%) 85 80	98, 135, 211, 235	0
3	J	216/229 (94%)	-0.29	1 (0%) 91 88	95, 134, 192, 219	0
All	All	2268/2352 (96%)	-0.21	18 (0%) 87 82	49, 135, 210, 283	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	152	GLY	4.8
2	L	154	GLU	4.3
2	L	152	GLY	4.0
3	H	133	ALA	3.8
3	D	149	GLY	3.6

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 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.