W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Sep 26, 2023 – 02:49 PM EDT

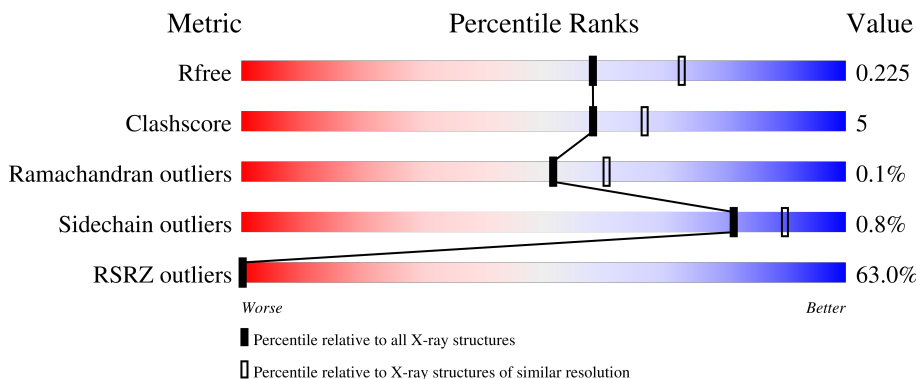|		
PDB ID	:	6BZW
Title	:	Structure of the Hepatitis C virus envelope glycoprotein E2 antigenic region 412-423 bound to the GL precursor of the broadly neutralizing antibody AP33
Authors	:	Tzarum, N.; Aleman, F.; Wilson, I.A.; Law, M.
Deposited on	:	2017-12-26
Resolution	:	2.20 Å(reported)

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 2.20 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.
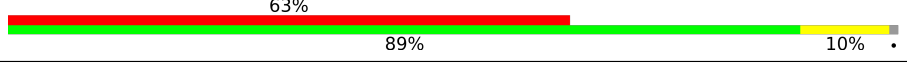
Metric	Percentile Ranks	Value
Rfree		0.225
Clashscore		5
Ramachandran outliers		0.1%
Sidechain outliers		0.8%
RSRZ outliers		63.0%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
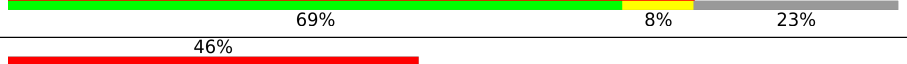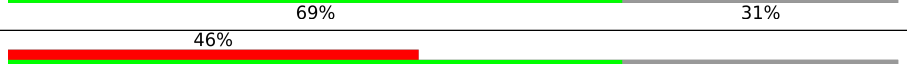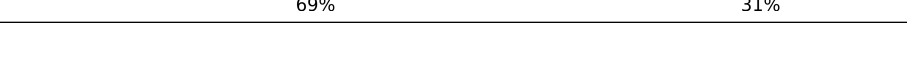
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	223	58% / 88% / 10% / ·
1	C	223	61% / 86% / 12% / ·
1	E	223	59% / 84% / 14% / ·
1	G	223	65% / 86% / 11% / ·
2	B	219	60% / 88% / 11% / ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	219	62% / 85% / 14%
2	F	219	63% / 89% / 10%
2	H	219	66% / 84% / 16%
3	I	13	62% / 62% / 8% / 31%
3	J	13	77% / 69% / 8% / 23%
3	K	13	46% / 69% / 31%
3	L	13	46% / 69% / 31%

2   Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 14240 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 AP33 GL Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	218	Total	C	N	O	S	0	0	0
			1655	1049	267	333	6			
1	A	219	Total	C	N	O	S	0	0	0
			1661	1052	268	335	6			
1	E	219	Total	C	N	O	S	0	0	0
			1661	1052	268	335	6			
1	G	219	Total	C	N	O	S	0	0	0
			1661	1052	268	335	6			

- Molecule 2 is a protein called AP33 GL Light Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	216	Total	C	N	O	S	0	0	0
			1652	1028	279	339	6			
2	B	217	Total	C	N	O	S	0	0	0
			1661	1033	280	342	6			
2	F	217	Total	C	N	O	S	0	0	0
			1661	1033	280	343	5			
2	H	218	Total	C	N	O	S	0	0	0
			1667	1036	281	344	6			

- Molecule 3 is a protein called E2 AS412 peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	I	9	Total	C	N	O	0	0	0
			73	46	14	13			
3	J	10	Total	C	N	O	0	0	0
			81	52	15	14			
3	K	9	Total	C	N	O	0	0	0
			73	46	14	13			
3	L	9	Total	C	N	O	0	0	0
			73	46	14	13			

• Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	89	Total 89	O 89	0	0
4	D	80	Total 80	O 80	0	0
4	A	81	Total 81	O 81	0	0
4	B	79	Total 79	O 79	0	0
4	E	69	Total 69	O 69	0	0
4	F	95	Total 95	O 95	0	0
4	G	62	Total 62	O 62	0	0
4	H	91	Total 91	O 91	0	0
4	I	5	Total 5	O 5	0	0
4	J	3	Total 3	O 3	0	0
4	K	3	Total 3	O 3	0	0
4	L	4	Total 4	O 4	0	0

3 Residue-property plots ⓘ

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: AP33 GL Heavy Chain

• Molecule 1: AP33 GL Heavy Chain

• Molecule 1: AP33 GL Heavy Chain

Chain E:

• Molecule 1: AP33 GL Heavy Chain

Chain G:

• Molecule 2: AP33 GL Light Chain

Chain D:

L201 SER S203 P204 V205 T206 K207 S208 F209 N210 R211 G212 E213 C214

• Molecule 2: AP33 GL Light Chain

60%

Chain B:

88%   11%

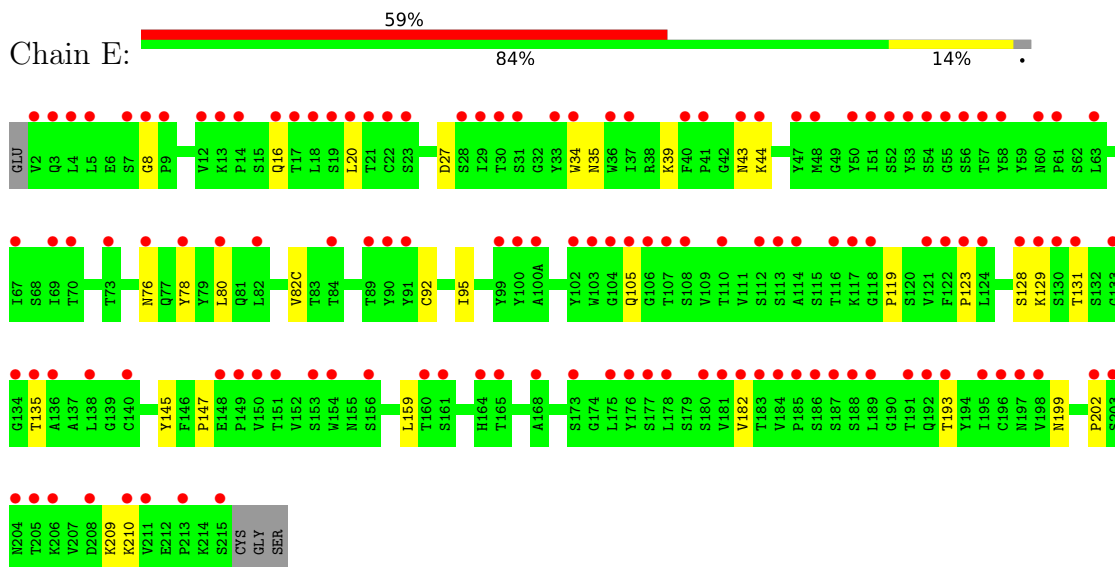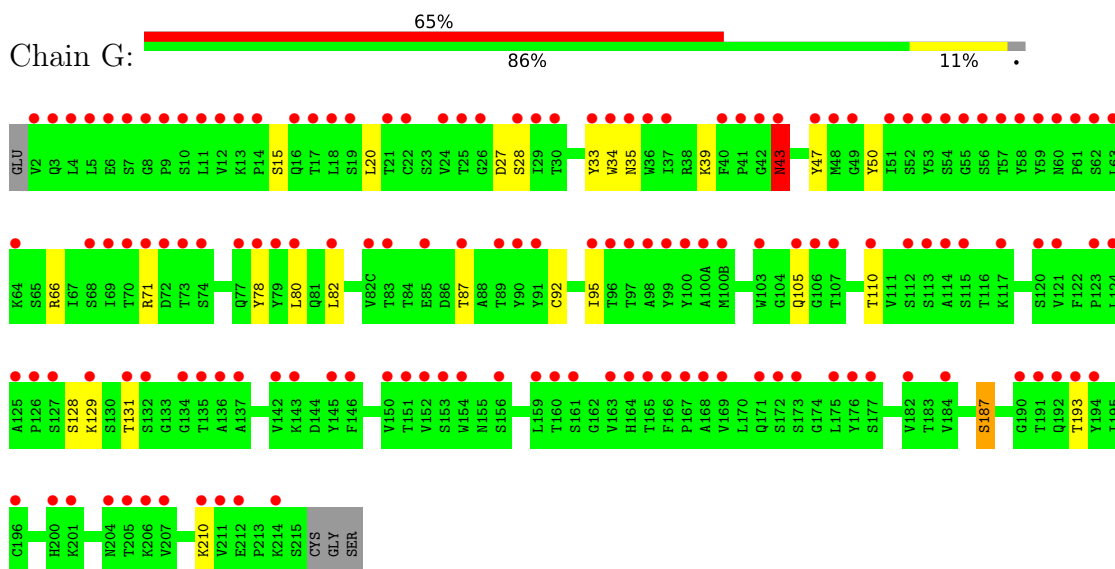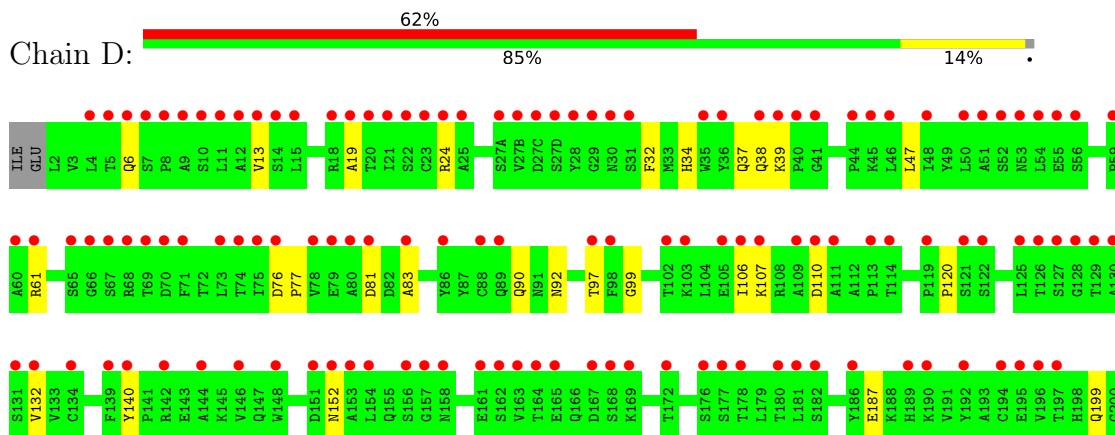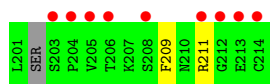ILE E1 L2 V3 L4 T5 Q6 S7 P8 A9 S10 L11 A12 V13 S14 L15 G16 Q17 R18 I21 S22 C23 R24 A25 S27A V27B D27C S27D Y28 G29 N30 S31 F32 M33 H34 W35 Y36 Q37 Q38 K39 P40 G41 Q42 P43 P44 K45 L46 L47 I48 Y49 L50 A51 S52 N53 L54 E55 S56 G57

V58 P59 A60 R61 F62 S65 G66 S67 R68 T69 D70 F71 T72 L73 T74 I75 D76 P77 V78 D81 D82 Y87 C88 Q89 Q90 N91 N92 P95 W96 T97 F98 G99 T102 K103 L104 E105 I106 K107 R108 A109 D110 A111 A112 P113 T114 I117 F118 P119 P120 S121 S122 L125 T126

S127 G128 T129 A130 S131 V132 V133 C134 L135 L136 N137 Y140 P141 R142 E143 A144 K145 V146 Q147 N152 A153 L154 Q155 S156 E161 S162 V163 T164 E165 Q166 D167 S168 K169 D170 S171 T172 Y173 S177 T178 L179 T180 L181 S182 K183 Y186 E187 K188 H189 K190 V191 V192 A193 C194 E195 V196

T197 H198 Q199 G200 L201 SER S203 P204 V205 T206 N210 R211 G212 E213 C214

• Molecule 2: AP33 GL Light Chain

63%

Chain F:

89%   10%

ILE E1 L2 V3 Q6 S7 P8 A9 S10 L11 A12 V13 S14 A19 T20 I21 S22 C23 R24 A25 S26 E27 S27A V27B D27C S27D Y28 G29 N30 S31 F32 M33 H34 W35 Y36 Q37 Q38 G41 Q42 P43 P44 K45 L46 L47 I48 Y49 L50 A51 S52 N53 L54 E55 S56 G57 V58 P59 A60

R61 F62 S63 G64 S65 G66 S67 R68 T69 D70 F71 T72 L73 T74 I75 D76 P77 V78 D81 T85 Y86 Y87 C88 Q89 Q90 N91 N92 E93 W96 T97 F98 G99 K103 L104 E105 I106 A109 D110 A111 A112 P113 T114 V115 S116 I117 F118 P119 P120 S121 S122 L125 T126 T129

A130 S131 V132 V133 C134 L135 L136 N137 N138 F139 Y140 P141 R142 E143 V146 Q147 W148 K149 V150 L154 S159 Q160 E161 S162 V163 T164 E165 S168 K169 D170 S171 T172 Y173 S174 L175 S176 S177 T178 L179 T180 L181 S182 K183 A184 D185 Y186 E187 K188 H189 K190 V191 Y192 A193 C194 E195 V196
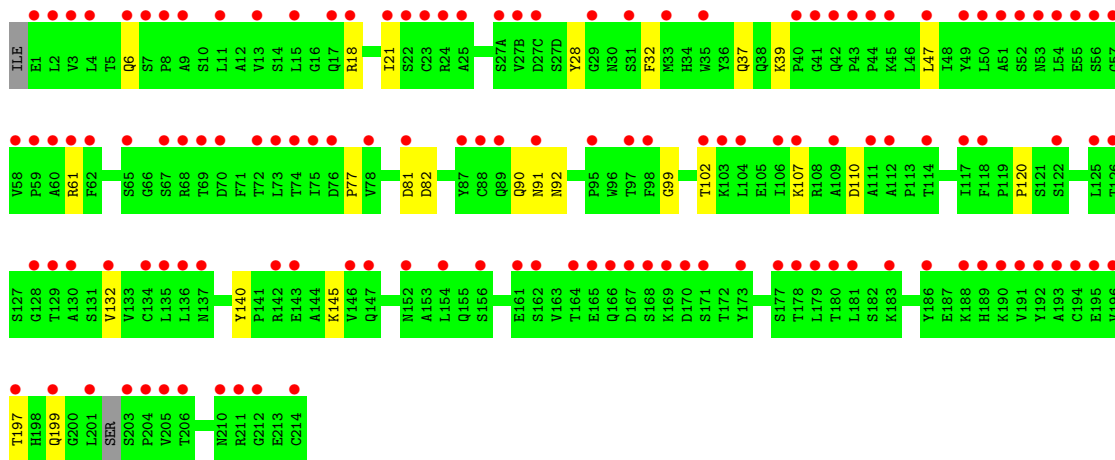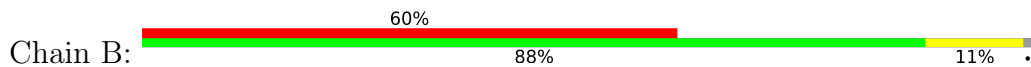
T197 H198 Q199 G200 L201 S202 S203 P204 V205 F209 N210 R211 G212 E213 CYS

• Molecule 2: AP33 GL Light Chain

66%

Chain H:

84%   16%

ILE E1 L2 V3 L4 T5 Q6 S7 P8 L11 A12 V13 S14 L15 A19 T20 I21 R24 A25 S26 E27 S27A V27B D27C S27D Y28 G29 N30 S31 F32 M33 H34 W35 Y36 Q37 Q38 K39 P40 G41 Q42 P43 L47 I48 Y49 L50 A51 S52 N53 L54 E55 S56 G57 A60 G64

S65 G66 S67 R68 T69 D70 F71 T72 L73 T74 I75 D76 P77 V78 E79 D82 A83 A84 T85 Y86 Y87 C88 Q89 Q90 N91 N92 E93 D94 P95 W96 T97 F98 G99 T102 K103 L104 E105 I106 K107 R108 A109 D110 A111 A112 P113 T114 V115 S116 P119 P120 S121 S122 Q123 Q124 L125 T126 S127

G128 T129 A130 S131 V132 V133 C134 L135 L136 N137 N138 F139 Y140 P141 R142 V146 Q147 W148 K149 V150 D151 N152 A153 L154 Q155 E161 S162 V163 Q166 D167 S168 K169 D170 S171 T172 Y173 S174 L175 S176 S177 T178 L179 T180 L181 S182 K183 A184 D185 Y186 E187 K188 H189 K190 V191 Y192 A193 C194

E195 V196 T197 H198 Q199 G200 L201 S202 S203 P204 F209 N210 R211 G212 E213 C214

• Molecule 3: E2 AS412 peptide

Chain I:

62%

62%   8%   31%

ARG GLN L413 I414 N415 T416 N417 G418 S419 W420 H421 ILE ASN

• Molecule 3: E2 AS412 peptide

Chain J:

77%

69%   8%   23%

ARG GLN L413 I414 N415 T416 N417 G418 S419 W420 H421 I422 ASN

• Molecule 3: E2 AS412 peptide

Chain K:

46%

69%   31%

ARG GLN L413 I414 N415 T416 N417 G418 S419 W420 H421 ILE ASN

• Molecule 3: E2 AS412 peptide

Chain L:

46%

69%   31%

ARG GLN L413 I414 N415 T416 N417 G418 S419 W420 H421 ILE ASN

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	60.36Å 84.71Å 192.92Å 90.00° 89.96° 90.00°	Depositor
Resolution (Å)	49.16 – 2.20 49.15 – 2.20	Depositor EDS
% Data completeness (in resolution range)	94.9 (49.16-2.20) 94.9 (49.15-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	113.35 (at 2.20Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, R_{free}	0.192 , 0.231 0.190 , 0.225	Depositor DCC
R_{free} test set	4557 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	20.8	Xtriage
Anisotropy	0.740	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 20.1	EDS
L-test for twinning[2]	$< |L| > = 0.42, < L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.419 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	14240	wwPDB-VP
Average B, all atoms (Å^2)	21.0	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 26.05 % 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 2.8314e-03. 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, 0.333 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	0.25	0/1703	0.49	0/2327
1	C	0.25	0/1697	0.50	0/2319
1	E	0.25	0/1703	0.49	0/2327
1	G	0.25	0/1703	0.51	0/2327
2	B	0.25	0/1696	0.46	0/2306
2	D	0.25	0/1687	0.46	0/2294
2	F	0.25	0/1697	0.47	0/2309
2	H	0.25	0/1703	0.47	0/2317
3	I	0.21	0/75	0.42	0/102
3	J	0.20	0/83	0.42	0/113
3	K	0.21	0/75	0.42	0/102
3	L	0.20	0/75	0.43	0/102
All	All	0.25	0/13897	0.48	0/18945

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	1661	0	1627	12	1
1	C	1655	0	1622	18	1
1	E	1661	0	1627	20	0
1	G	1661	0	1627	18	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1661	0	1591	16	0
2	D	1652	0	1582	19	0
2	F	1661	0	1592	12	0
2	H	1667	0	1597	21	0
3	I	73	0	65	1	0
3	J	81	0	76	1	0
3	K	73	0	65	0	0
3	L	73	0	65	0	0
4	A	81	0	0	3	0
4	B	79	0	0	2	0
4	C	89	0	0	2	0
4	D	80	0	0	2	0
4	E	69	0	0	0	0
4	F	95	0	0	1	0
4	G	62	0	0	2	0
4	H	91	0	0	4	0
4	I	5	0	0	1	0
4	J	3	0	0	0	0
4	K	3	0	0	0	0
4	L	4	0	0	0	0
All	All	14240	0	13136	127	1

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 127 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:117:LYS:NZ	4:A:301:HOH:O	2.09	0.85
1:A:143:LYS:NZ	4:A:302:HOH:O	2.14	0.81
2:F:68:ARG:NH1	4:F:301:HOH:O	2.16	0.77
2:H:134:CYS:SG	4:H:301:HOH:O	2.43	0.76
2:H:110:ASP:OD2	2:H:199:GLN:NE2	2.17	0.71

All (1) 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)
1:C:203:SER:O	1:A:209:LYS:NZ[2_656]	2.09	0.11

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	217/223 (97%)	212 (98%)	5 (2%)	0	100	100
1	C	216/223 (97%)	210 (97%)	6 (3%)	0	100	100
1	E	217/223 (97%)	210 (97%)	7 (3%)	0	100	100
1	G	217/223 (97%)	213 (98%)	3 (1%)	1 (0%)	29	31
2	B	213/219 (97%)	208 (98%)	5 (2%)	0	100	100
2	D	212/219 (97%)	206 (97%)	6 (3%)	0	100	100
2	F	215/219 (98%)	209 (97%)	6 (3%)	0	100	100
2	H	216/219 (99%)	209 (97%)	6 (3%)	1 (0%)	29	31
3	I	7/13 (54%)	7 (100%)	0	0	100	100
3	J	8/13 (62%)	8 (100%)	0	0	100	100
3	K	7/13 (54%)	7 (100%)	0	0	100	100
3	L	7/13 (54%)	7 (100%)	0	0	100	100
All	All	1752/1820 (96%)	1706 (97%)	44 (2%)	2 (0%)	51	60

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	43	ASN
2	H	68	ARG

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	193/196 (98%)	191 (99%)	2 (1%)	76	86
1	C	192/196 (98%)	191 (100%)	1 (0%)	88	94
1	E	193/196 (98%)	192 (100%)	1 (0%)	88	94
1	G	193/196 (98%)	190 (98%)	3 (2%)	62	76
2	B	187/189 (99%)	187 (100%)	0	100	100
2	D	186/189 (98%)	185 (100%)	1 (0%)	88	94
2	F	187/189 (99%)	185 (99%)	2 (1%)	73	85
2	H	188/189 (100%)	186 (99%)	2 (1%)	73	85
3	I	8/12 (67%)	8 (100%)	0	100	100
3	J	9/12 (75%)	9 (100%)	0	100	100
3	K	8/12 (67%)	8 (100%)	0	100	100
3	L	8/12 (67%)	8 (100%)	0	100	100
All	All	1552/1588 (98%)	1540 (99%)	12 (1%)	81	90

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	15	SER
1	G	43	ASN
2	H	89	GLN
1	G	187	SER
1	A	199	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	35	ASN
2	D	90	GLN
1	E	35	ASN
2	F	90	GLN
1	G	35	ASN

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 monosaccharides 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, 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(Å²)	Q<0.9
1	A	219/223 (98%)	2.39	130 (59%)	0	0	11, 20, 32, 43	0
1	C	218/223 (97%)	2.43	135 (61%)	0	0	11, 20, 30, 38	0
1	E	219/223 (98%)	2.41	131 (59%)	0	0	13, 21, 31, 41	0
1	G	219/223 (98%)	2.50	146 (66%)	0	0	12, 21, 33, 46	0
2	B	217/219 (99%)	2.58	132 (60%)	0	0	10, 20, 30, 58	0
2	D	216/219 (98%)	2.48	136 (62%)	0	0	11, 21, 30, 39	0
2	F	217/219 (99%)	2.49	138 (63%)	0	0	13, 22, 31, 50	0
2	H	218/219 (99%)	2.62	144 (66%)	0	0	12, 21, 29, 42	0
3	I	9/13 (69%)	3.08	8 (88%)	0	0	20, 22, 26, 26	0
3	J	10/13 (76%)	2.94	10 (100%)	0	0	21, 23, 27, 29	0
3	K	9/13 (69%)	2.67	6 (66%)	0	0	21, 23, 25, 25	0
3	L	9/13 (69%)	2.77	6 (66%)	0	0	21, 24, 27, 27	0
All	All	1780/1820 (97%)	2.49	1122 (63%)	0	0	10, 21, 31, 58	0

The worst 5 of 1122 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1	GLU	10.6
2	B	214	CYS	9.9
2	H	86	TYR	9.0
2	B	13	VAL	8.1
2	F	57	GLY	7.3

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 ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.