



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 17, 2017 – 11:27 AM EDT

PDB ID : 5Y9C
Title : Crystal structure of HPV58 pentamer in complex with the Fab fragment of antibody A12A3
Authors : Li, S.W.; Li, Z.H.
Deposited on : unknown
Resolution : 3.44 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

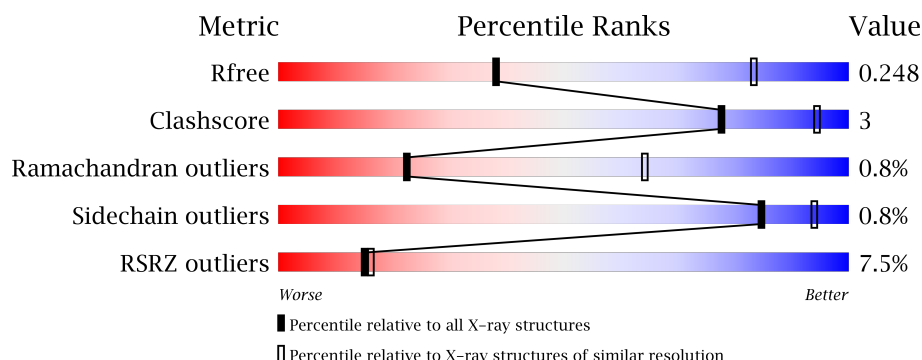
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.44 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}	100719	1082 (3.52-3.36)
Clashscore	112137	1025 (3.50-3.38)
Ramachandran outliers	110173	1155 (3.52-3.36)
Sidechain outliers	110143	1156 (3.52-3.36)
RSRZ outliers	101464	1107 (3.52-3.36)

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	490	<div> <div>2%</div> <div> <div></div> <div>76%</div> <div>9%</div> <div>15%</div> </div> </div>
1	B	490	<div> <div>2%</div> <div> <div></div> <div>78%</div> <div>7%</div> <div>14%</div> </div> </div>
1	C	490	<div> <div>6%</div> <div> <div></div> <div>77%</div> <div>8%</div> <div>15%</div> </div> </div>
1	D	490	<div> <div>3%</div> <div> <div></div> <div>77%</div> <div>7%</div> <div>16%</div> </div> </div>
1	E	490	<div> <div>3%</div> <div> <div></div> <div>76%</div> <div>8%</div> <div>16%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	216	<div><div></div><div>35%</div><div></div><div>76%</div><div></div><div>20%</div><div></div><div>• •</div></div>
3	L	213	<div><div></div><div>15%</div><div></div><div>84%</div><div></div><div>11%</div><div></div><div>• 5%</div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 19783 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 L1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	417	Total	C	N	O	S	0	0	0
			3324	2120	553	632	19			
1	B	420	Total	C	N	O	S	0	0	0
			3345	2131	558	637	19			
1	C	415	Total	C	N	O	S	0	0	0
			3311	2113	550	629	19			
1	D	414	Total	C	N	O	S	0	0	0
			3305	2110	549	627	19			
1	E	414	Total	C	N	O	S	0	0	0
			3305	2110	549	627	19			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	MET	-	initiating methionine	UNP P26535
A	176	SER	CYS	engineered mutation	UNP P26535
B	9	MET	-	initiating methionine	UNP P26535
B	176	SER	CYS	engineered mutation	UNP P26535
C	9	MET	-	initiating methionine	UNP P26535
C	176	SER	CYS	engineered mutation	UNP P26535
D	9	MET	-	initiating methionine	UNP P26535
D	176	SER	CYS	engineered mutation	UNP P26535
E	9	MET	-	initiating methionine	UNP P26535
E	176	SER	CYS	engineered mutation	UNP P26535

- Molecule 2 is a protein called heavy chain of Fab fragment of antibody A12A3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	210	Total	C	N	O	S	0	0	0
			1608	1018	264	319	7			

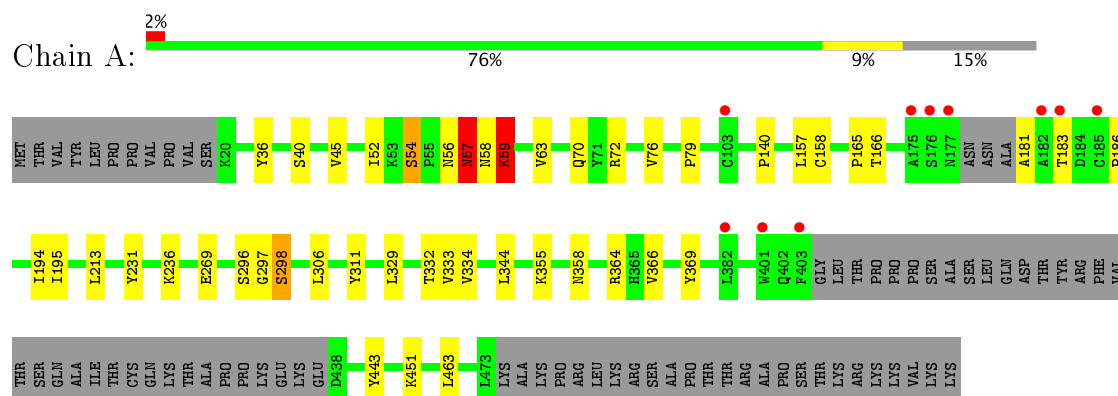
- Molecule 3 is a protein called light chain of Fab fragment of antibody A12A3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	L	203	Total	C	N	O	S	0	0	0
			1585	1003	264	312	6			

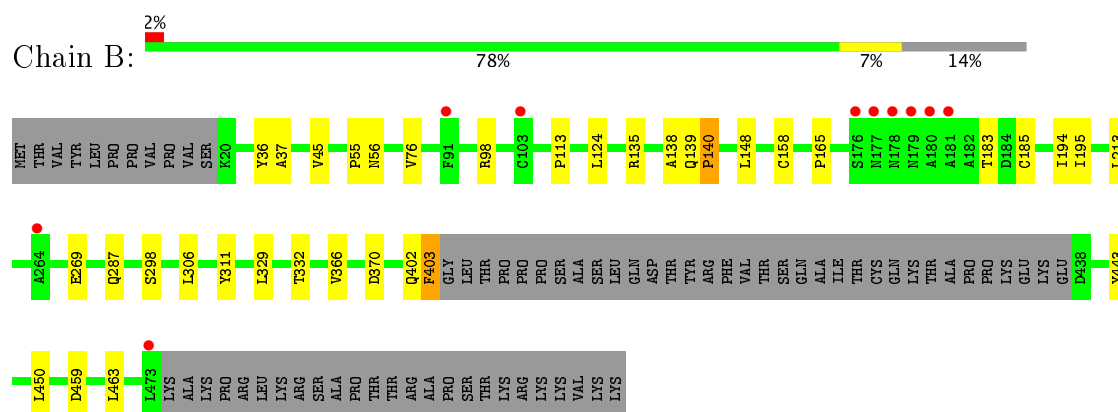
3 Residue-property plots [i](#)

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.

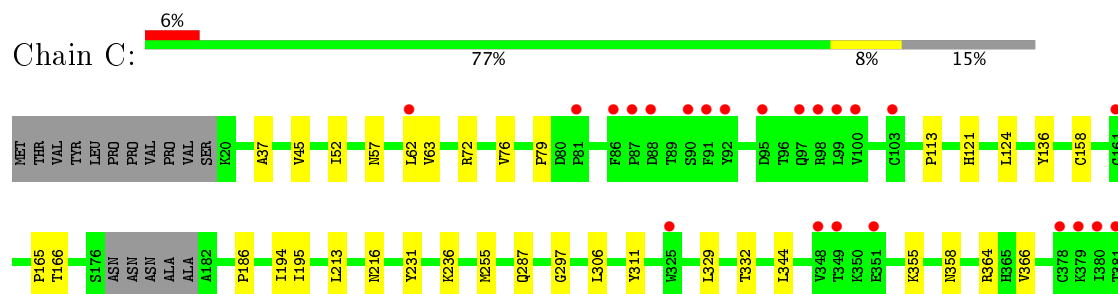
- Molecule 1: Major capsid protein L1

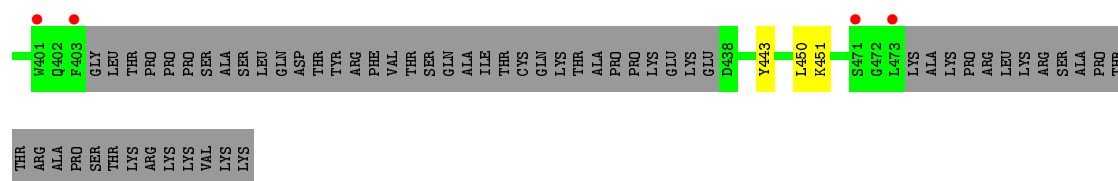


- Molecule 1: Major capsid protein L1

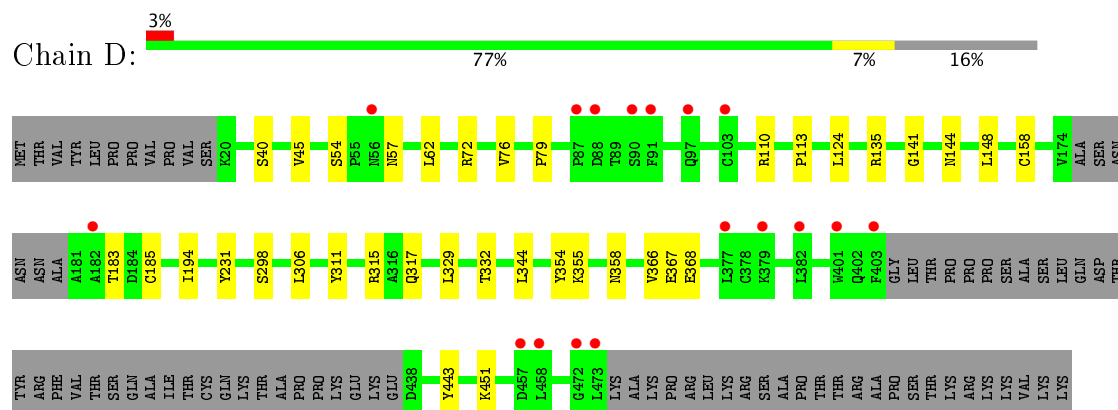


- Molecule 1: Major capsid protein L1

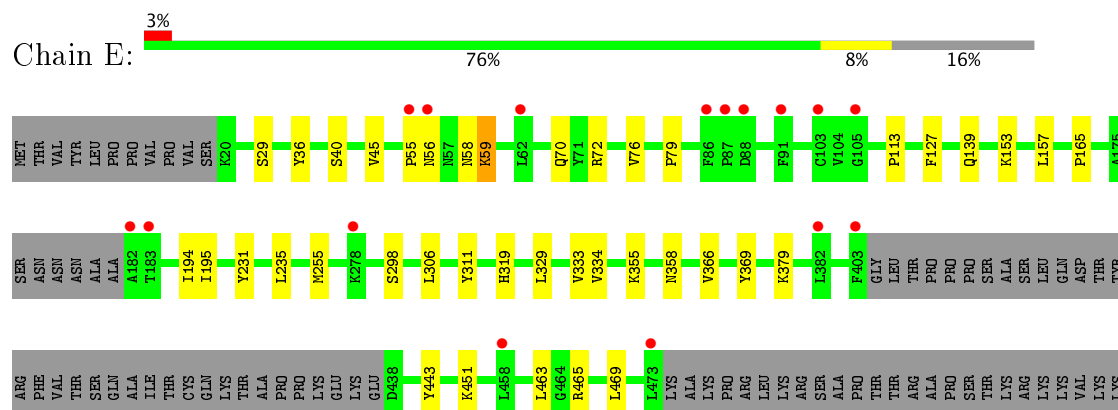




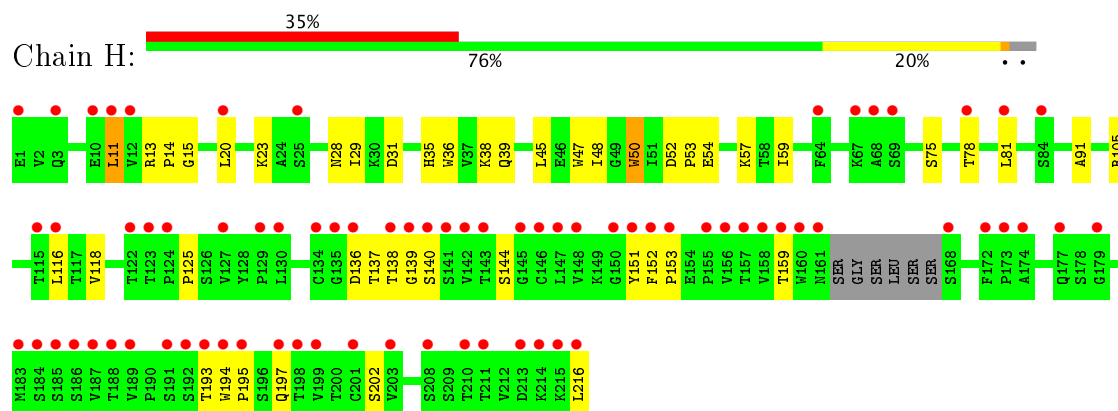
- Molecule 1: Major capsid protein L1



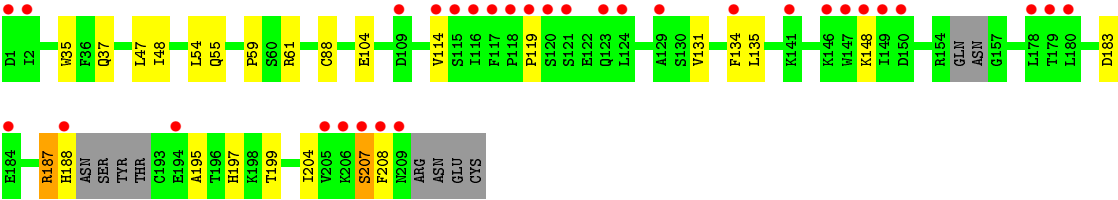
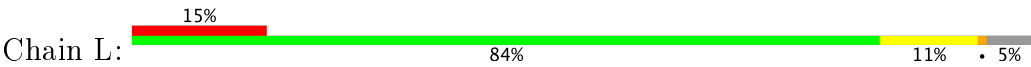
- Molecule 1: Major capsid protein L1



- Molecule 2: heavy chain of Fab fragment of antibody A12A3



- Molecule 3: light chain of Fab fragment of antibody A12A3



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	121.57Å 102.60Å 138.03Å 90.00° 114.52° 90.00°	Depositor
Resolution (Å)	47.86 – 3.44 47.86 – 3.44	Depositor EDS
% Data completeness (in resolution range)	99.5 (47.86-3.44) 99.6 (47.86-3.44)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.05 (at 3.48Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575)	Depositor
R, R_{free}	0.202 , 0.249 0.201 , 0.248	Depositor DCC
R_{free} test set	2025 reflections (4.94%)	DCC
Wilson B-factor (Å ²)	77.3	Xtriage
Anisotropy	0.840	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 51.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.022 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	19783	wwPDB-VP
Average B, all atoms (Å ²)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% 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 [i](#)

5.1 Standard geometry [i](#)

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	0/3410	0.44	0/4625
1	B	0.25	0/3432	0.44	0/4657
1	C	0.25	0/3397	0.44	0/4607
1	D	0.25	0/3391	0.43	0/4599
1	E	0.25	0/3391	0.44	0/4599
2	H	0.24	0/1652	0.44	0/2260
3	L	0.25	0/1624	0.44	0/2199
All	All	0.25	0/20297	0.44	0/27546

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 [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	3324	0	3204	24	0
1	B	3345	0	3222	20	0
1	C	3311	0	3193	24	0
1	D	3305	0	3188	23	0
1	E	3305	0	3188	23	0
2	H	1608	0	1551	23	0
3	L	1585	0	1531	14	0
All	All	19783	0	19077	131	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 3.

The worst 5 of 131 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:296:SER:O	1:A:298:SER:N	2.20	0.74
1:B:306:LEU:O	1:B:311:TYR:OH	2.08	0.69
3:L:183:ASP:OD2	3:L:187:ARG:NE	2.24	0.68
1:A:306:LEU:O	1:A:311:TYR:OH	2.11	0.67
1:D:306:LEU:O	1:D:311:TYR:OH	2.12	0.62

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	411/490 (84%)	387 (94%)	16 (4%)	8 (2%)	9	45
1	B	416/490 (85%)	394 (95%)	19 (5%)	3 (1%)	25	67
1	C	409/490 (84%)	390 (95%)	19 (5%)	0	100	100
1	D	408/490 (83%)	391 (96%)	15 (4%)	2 (0%)	32	72
1	E	408/490 (83%)	386 (95%)	19 (5%)	3 (1%)	25	67
2	H	206/216 (95%)	182 (88%)	21 (10%)	3 (2%)	12	51
3	L	197/213 (92%)	183 (93%)	14 (7%)	0	100	100
All	All	2455/2879 (85%)	2313 (94%)	123 (5%)	19 (1%)	22	64

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	57	ASN
1	A	58	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	297	GLY
1	D	298	SER
1	E	40	SER

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	370/435 (85%)	366 (99%)	4 (1%)	78	91
1	B	372/435 (86%)	370 (100%)	2 (0%)	91	96
1	C	369/435 (85%)	368 (100%)	1 (0%)	94	98
1	D	368/435 (85%)	368 (100%)	0	100	100
1	E	368/435 (85%)	367 (100%)	1 (0%)	94	98
2	H	181/186 (97%)	174 (96%)	7 (4%)	37	72
3	L	179/189 (95%)	176 (98%)	3 (2%)	66	87
All	All	2207/2550 (86%)	2189 (99%)	18 (1%)	85	94

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	11	LEU
2	H	13	ARG
2	H	138	THR
1	C	57	ASN
1	E	59	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	112	GLN
2	H	6	GLN
2	H	35	HIS
2	H	197	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 [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	417/490 (85%)	0.24	10 (2%) 59 54	54, 87, 137, 172	0
1	B	420/490 (85%)	0.12	10 (2%) 59 54	48, 81, 134, 203	0
1	C	415/490 (84%)	0.37	27 (6%) 20 19	60, 93, 156, 212	0
1	D	414/490 (84%)	0.26	17 (4%) 38 34	53, 91, 153, 186	0
1	E	414/490 (84%)	0.32	16 (3%) 40 36	55, 90, 147, 197	0
2	H	210/216 (97%)	1.61	76 (36%) 0 0	74, 125, 184, 215	0
3	L	203/213 (95%)	0.89	32 (15%) 2 3	79, 110, 166, 189	0
All	All	2493/2879 (86%)	0.42	188 (7%) 15 16	48, 92, 155, 215	0

The worst 5 of 188 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	L	124	LEU	5.7
2	H	130	LEU	5.7
1	D	473	LEU	5.5
1	E	182	ALA	5.4
2	H	160	TRP	5.0

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.