wwPDB X-ray Structure Validation Summary Report (i)

Sep 25, 2023 – 02:50 AM EDT

PDB ID	:	5W0K
Title	:	Crystal structure of EBV gHgL/CL40/gp42 N-domain
Authors	:	Sathiyamoorthy, K.; Jardetzky, T.S.
Deposited on	:	2017-05-31
Resolution	:	3.10 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:

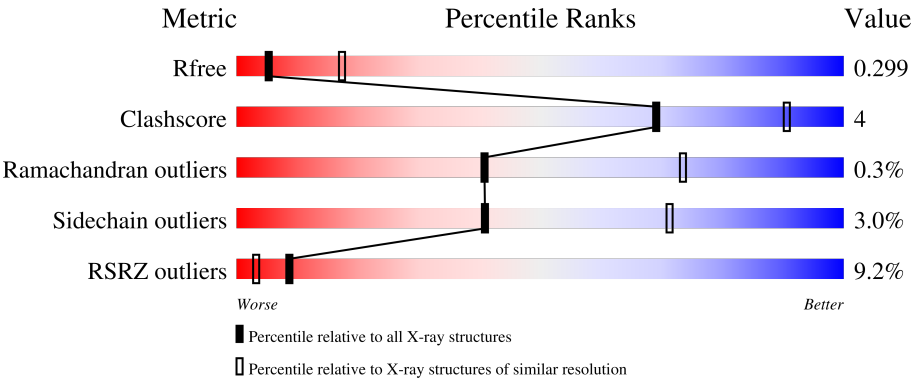
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.10 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.



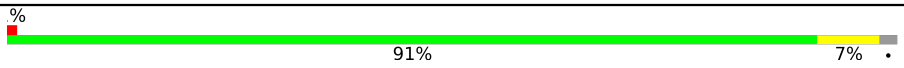
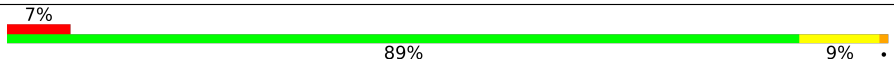
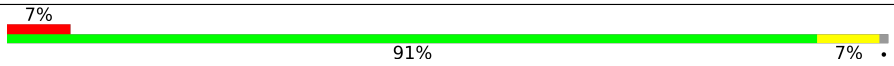
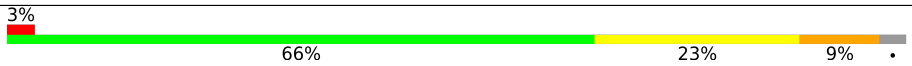
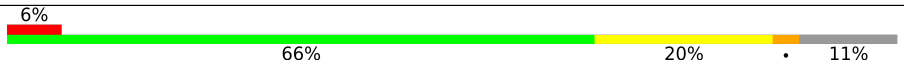
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	660	8% 90% 8%
1	C	660	17% 89% 9%
2	B	114	6% 75% 6% 18%
2	D	114	6% 70% 11% 19%
3	E	217	3% 91% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	217	
4	F	213	
4	L	213	
5	X	35	
5	Y	35	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 37028 atoms, of which 18365 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 Envelope glycoprotein H.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	655	Total	C	H	N	O	S	0	0	0
			10212	3268	5115	842	956	31			
1	C	655	Total	C	H	N	O	S	0	0	0
			10212	3268	5115	842	956	31			

- Molecule 2 is a protein called Envelope glycoprotein L.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	94	Total	C	H	N	O	S	0	0	0
			1422	454	704	120	140	4			
2	D	92	Total	C	H	N	O	S	0	0	0
			1391	442	690	118	137	4			

- Molecule 3 is a protein called CL40 IgG heavy chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	H	213	Total	C	H	N	O	S	0	0	0
			3182	1024	1568	258	323	9			
3	E	214	Total	C	H	N	O	S	0	0	0
			3191	1026	1574	259	323	9			

- Molecule 4 is a protein called CL40 IgG light chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	L	210	Total	C	H	N	O	S	0	0	0
			3185	1017	1555	272	334	7			
4	F	212	Total	C	H	N	O	S	0	0	0
			3223	1027	1574	278	337	7			

- Molecule 5 is a protein called Glycoprotein 42.

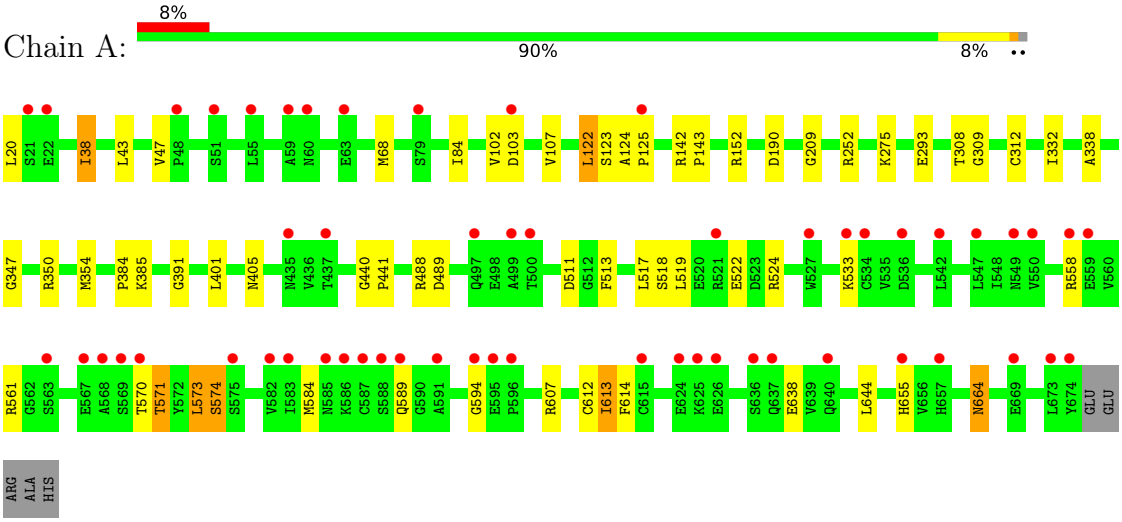
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	X	34	Total	C	H	N	O	0	0	0
			543	184	262	45	52			
5	Y	31	Total	C	H	N	O	0	0	0
			467	170	208	41	48			

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.

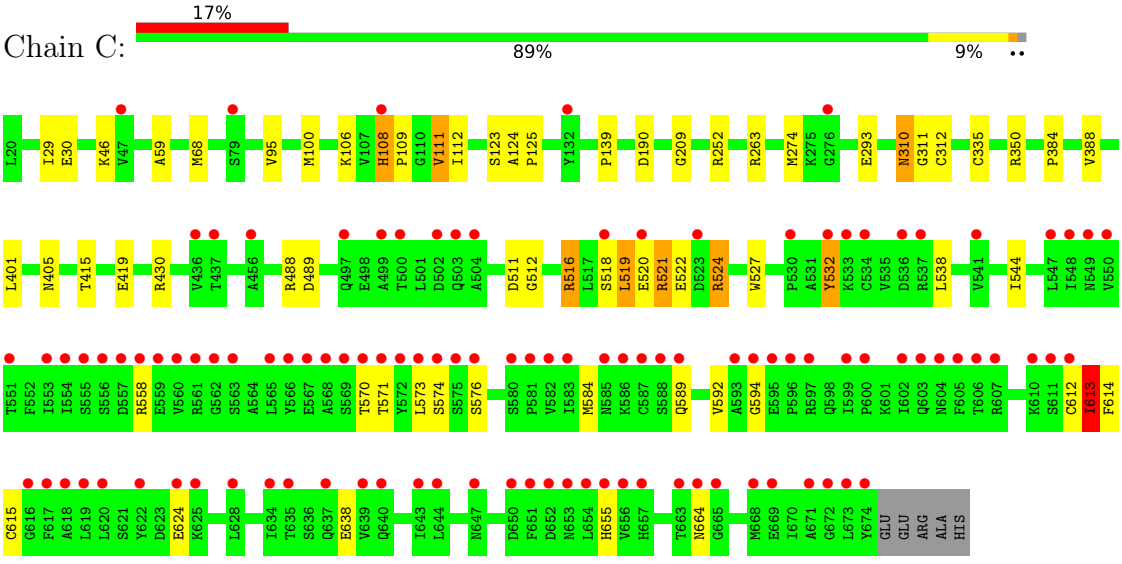
• Molecule 1: Envelope glycoprotein H

Chain A:



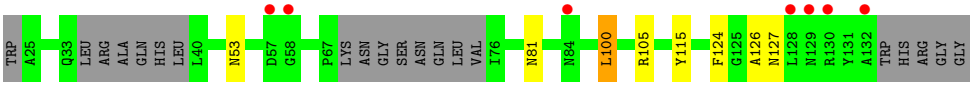
• Molecule 1: Envelope glycoprotein H

Chain C:

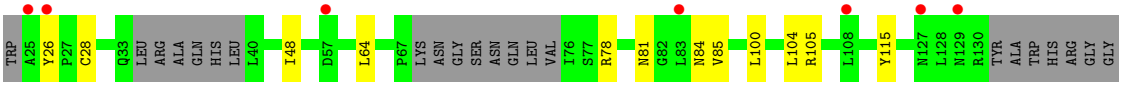


• Molecule 2: Envelope glycoprotein L

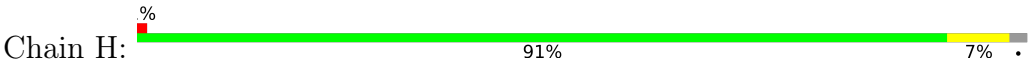
Chain B:

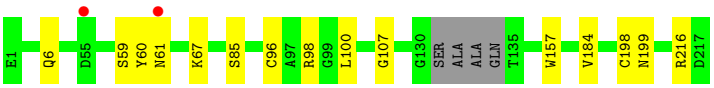


• Molecule 2: Envelope glycoprotein L

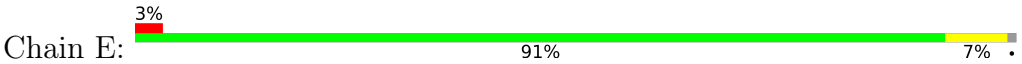


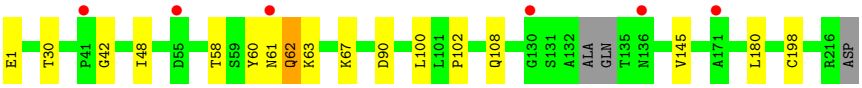
• Molecule 3: CL40 IgG heavy chain



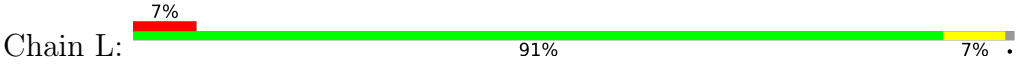


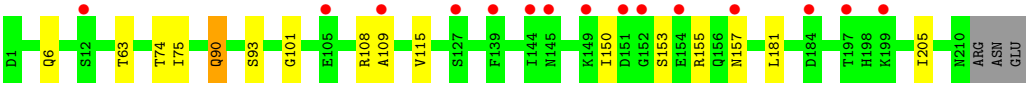
• Molecule 3: CL40 IgG heavy chain



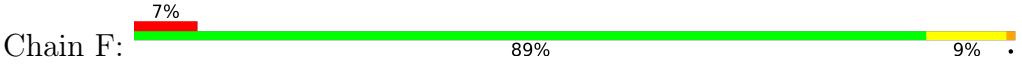


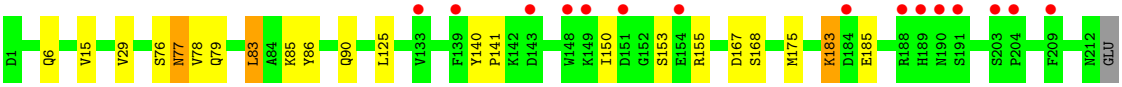
• Molecule 4: CL40 IgG light chain



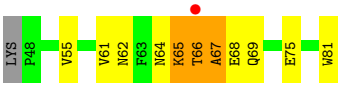


• Molecule 4: CL40 IgG light chain



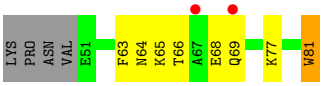


• Molecule 5: Glycoprotein 42



• Molecule 5: Glycoprotein 42

Chain Y:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	96.38Å 133.13Å 254.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	22.76 – 3.10 49.28 – 2.88	Depositor EDS
% Data completeness (in resolution range)	90.5 (22.76-3.10) 78.0 (49.28-2.88)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.54 (at 2.86Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.242 , 0.299 0.243 , 0.299	Depositor DCC
R_{free} test set	2923 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	59.5	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 53.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	37028	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.55% 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	A	0.25	0/5206	0.44	0/7066
1	C	0.25	0/5206	0.46	0/7066
2	B	0.24	0/728	0.43	0/986
2	D	0.25	0/710	0.46	0/961
3	E	0.25	0/1660	0.48	0/2269
3	H	0.26	0/1657	0.47	0/2265
4	F	0.26	0/1686	0.48	0/2287
4	L	0.25	0/1667	0.46	0/2262
5	X	0.27	0/294	0.51	0/406
5	Y	0.31	0/271	0.51	0/374
All	All	0.25	0/19085	0.46	0/25942

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	5097	5115	5115	39	0
1	C	5097	5115	5115	48	0
2	B	718	704	703	4	0
2	D	701	690	689	8	0
3	E	1617	1574	1574	5	0
3	H	1614	1568	1568	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	F	1649	1574	1574	13	0
4	L	1630	1555	1555	10	0
5	X	281	262	262	8	0
5	Y	259	208	239	10	0
All	All	18663	18365	18394	136	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 4.

The worst 5 of 136 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:612:CYS:O	1:A:614:PHE:N	1.99	0.95
4:L:63:THR:HG1	4:L:74:THR:HG1	1.15	0.88
1:A:488:ARG:NH1	1:A:511:ASP:OD1	2.10	0.84
2:B:100:LEU:O	2:B:105:ARG:NH1	2.12	0.82
1:C:516:ARG:O	1:C:524:ARG:NH2	2.14	0.81

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	653/660 (99%)	617 (94%)	32 (5%)	4 (1%)	25 59
1	C	653/660 (99%)	610 (93%)	41 (6%)	2 (0%)	41 73
2	B	88/114 (77%)	80 (91%)	8 (9%)	0	100 100
2	D	86/114 (75%)	81 (94%)	5 (6%)	0	100 100
3	E	210/217 (97%)	202 (96%)	8 (4%)	0	100 100
3	H	209/217 (96%)	194 (93%)	15 (7%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
4	F	210/213 (99%)	199 (95%)	11 (5%)	0	100 100
4	L	208/213 (98%)	199 (96%)	9 (4%)	0	100 100
5	X	32/35 (91%)	24 (75%)	6 (19%)	2 (6%)	1 8
5	Y	29/35 (83%)	22 (76%)	7 (24%)	0	100 100
All	All	2378/2478 (96%)	2228 (94%)	142 (6%)	8 (0%)	41 73

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	574	SER
1	A	613	ILE
5	X	67	ALA
1	C	520	GLU
1	A	573	LEU

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	561/565 (99%)	549 (98%)	12 (2%)	53 79
1	C	561/565 (99%)	540 (96%)	21 (4%)	34 66
2	B	82/98 (84%)	79 (96%)	3 (4%)	34 66
2	D	81/98 (83%)	80 (99%)	1 (1%)	71 88
3	E	185/187 (99%)	176 (95%)	9 (5%)	25 57
3	H	185/187 (99%)	178 (96%)	7 (4%)	33 66
4	F	188/189 (100%)	184 (98%)	4 (2%)	53 79
4	L	186/189 (98%)	184 (99%)	2 (1%)	73 89
5	X	32/33 (97%)	30 (94%)	2 (6%)	18 48
5	Y	29/33 (88%)	27 (93%)	2 (7%)	15 45
All	All	2090/2144 (98%)	2027 (97%)	63 (3%)	41 71

5 of 63 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	108	HIS
3	E	108	GLN
1	C	415	THR
3	E	63	LYS
4	F	175	MET

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	A	655/660 (99%)	0.48	56 (8%) 10 4	21, 56, 117, 144	0
1	C	655/660 (99%)	0.98	111 (16%) 1 0	19, 60, 149, 178	0
2	B	94/114 (82%)	0.50	7 (7%) 14 5	35, 62, 100, 117	0
2	D	92/114 (80%)	0.45	7 (7%) 13 5	32, 60, 104, 115	0
3	E	214/217 (98%)	0.15	6 (2%) 53 30	20, 40, 70, 119	0
3	H	213/217 (98%)	0.11	2 (0%) 84 69	23, 41, 72, 100	0
4	F	212/213 (99%)	0.44	15 (7%) 16 6	21, 56, 101, 128	0
4	L	210/213 (98%)	0.42	15 (7%) 16 6	26, 59, 88, 95	0
5	X	34/35 (97%)	0.17	1 (2%) 51 28	39, 61, 84, 99	0
5	Y	31/35 (88%)	0.39	2 (6%) 18 8	38, 64, 83, 93	0
All	All	2410/2478 (97%)	0.54	222 (9%) 9 3	19, 54, 125, 178	0

The worst 5 of 222 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	574	SER	13.8
1	C	671	ALA	12.7
1	C	569	SER	10.4
1	C	625	LYS	9.8
1	C	604	ASN	9.1

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 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.