Full wwPDB X-ray Structure Validation Report (i)

May 15, 2020 – 12:27 am BST

PDB ID	:	1GFL
Title	:	STRUCTURE OF GREEN FLUORESCENT PROTEIN
Authors	:	Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Deposited on	:	1996-08-23
Resolution	:	1.90 Å(reported)

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	:	2.11
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)	:	2.11

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 1.90 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		0.231
Clashscore		10
Ramachandran outliers		0.4%
Sidechain outliers		14.2%
RSRZ outliers		3.0%

Worse ← → *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
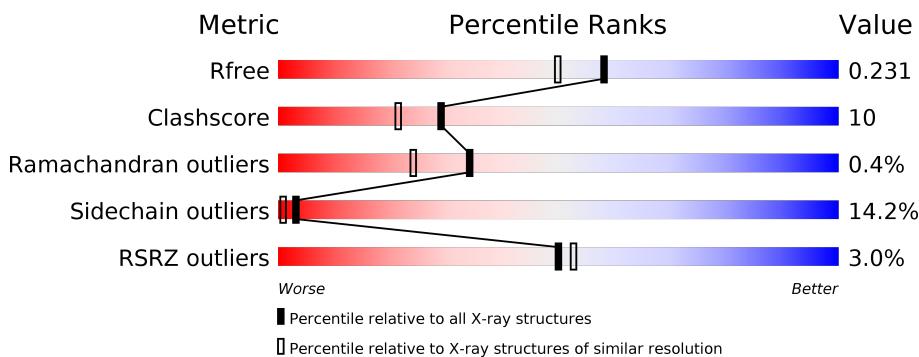
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 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	238	2% / 71% 21% · · ·
1	B	238	4% / 72% 19% · · ·
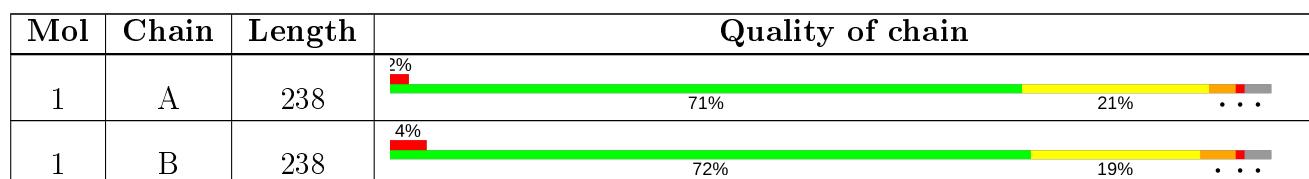
2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3950 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 GREEN FLUORESCENT PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	230	Total	C	N	O	S	0	0	0
			1825	1161	309	349	6			
1	B	230	Total	C	N	O	S	0	0	0
			1825	1161	309	349	6			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	142	Total	O	0	0
			142	142		
2	B	158	Total	O	0	0
			158	158		

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: GREEN FLUORESCENT PROTEIN
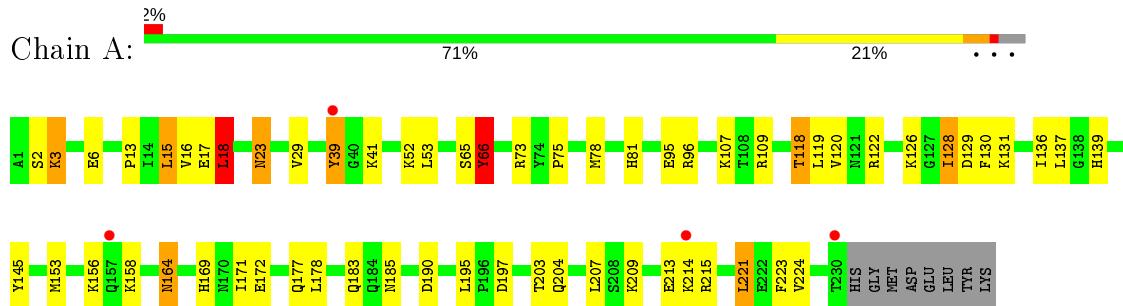
- Molecule 1: GREEN FLUORESCENT PROTEIN
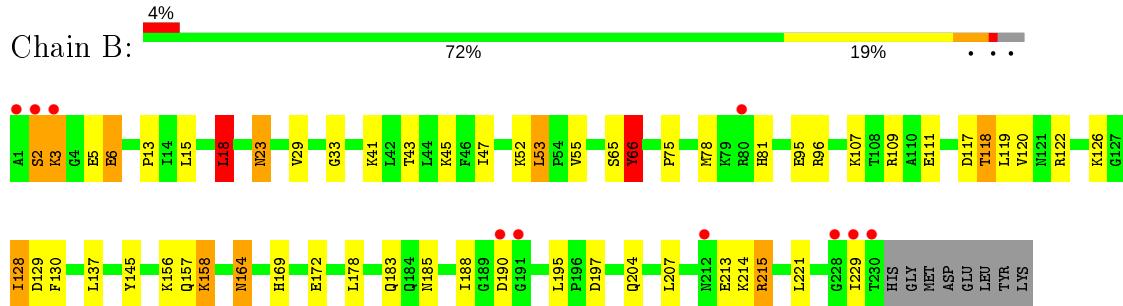
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	89.23Å 89.23Å 119.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 1.90 28.22 – 1.90	Depositor EDS
% Data completeness (in resolution range)	96.8 (10.00-1.90) 96.8 (28.22-1.90)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.56 (at 1.91Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.214 , 0.262 0.192 , 0.231	Depositor DCC
R_{free} test set	3906 reflections (10.15%)	wwPDB-VP
Wilson B-factor (Å^2)	24.3	Xtriage
Anisotropy	0.380	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 95.4	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3950	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.62% 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 (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	0.75	2/1868 (0.1%)	0.88	6/2524 (0.2%)
1	B	0.75	2/1868 (0.1%)	0.89	5/2524 (0.2%)
All	All	0.75	4/3736 (0.1%)	0.88	11/5048 (0.2%)

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
1	A	1	3
1	B	1	1
All	All	2	4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	66	TYR	CB-CG	-7.59	1.40	1.51
1	A	66	TYR	CA-CB	-7.05	1.38	1.53
1	A	66	TYR	CB-CG	-6.63	1.41	1.51
1	B	66	TYR	CA-CB	-5.86	1.41	1.53

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	66	TYR	N-CA-CB	12.52	133.13	110.60
1	A	66	TYR	N-CA-CB	11.10	130.57	110.60
1	A	66	TYR	CA-CB-CG	8.38	129.32	113.40
1	A	65	SER	C-N-CA	-8.38	100.76	121.70
1	B	65	SER	C-N-CA	-7.95	101.83	121.70
1	B	66	TYR	CA-CB-CG	7.22	127.11	113.40
1	A	18	LEU	CA-CB-CG	6.33	129.85	115.30
1	A	66	TYR	CA-C-N	-5.60	105.00	116.20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	TYR	CB-CA-C	5.40	121.20	110.40
1	B	18	LEU	CA-CB-CG	5.38	127.68	115.30
1	B	66	TYR	CA-C-N	-5.03	106.14	116.20

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	66	TYR	CA
1	B	66	TYR	CA

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	39	TYR	Sidechain
1	A	73	ARG	Sidechain
1	A	96	ARG	Peptide
1	B	96	ARG	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	1825	0	1778	36	0
1	B	1825	0	1778	37	0
2	A	142	0	0	10	0
2	B	158	0	0	15	0
All	All	3950	0	3556	69	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 10.

All (69) 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:183:GLN:HE21	1:A:185:ASN:HD21	1.10	0.97
1:B:183:GLN:HE21	1:B:185:ASN:HD21	1.09	0.92

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:2:SER:HB3	2:B:355:HOH:O	1.82	0.78
1:A:221:LEU:HB3	2:A:340:HOH:O	1.84	0.76
1:A:130:PHE:HB3	1:A:137:LEU:HD23	1.69	0.73
1:A:17:GLU:HG2	2:A:349:HOH:O	1.89	0.72
1:B:157:GLN:HB3	2:B:342:HOH:O	1.89	0.71
1:B:130:PHE:HB3	1:B:137:LEU:HD23	1.72	0.71
1:A:95:GLU:HG2	1:A:109:ARG:HG3	1.75	0.69
1:B:109:ARG:HD2	2:B:349:HOH:O	1.93	0.68
1:A:207:LEU:H	1:B:204:GLN:HE22	1.42	0.68
1:A:16:VAL:HG12	2:A:366:HOH:O	1.92	0.68
1:B:95:GLU:HG2	1:B:109:ARG:HG3	1.78	0.65
1:A:139:HIS:HD2	2:A:332:HOH:O	1.80	0.64
1:A:18:LEU:HB2	2:A:366:HOH:O	1.96	0.64
1:B:164:ASN:HB3	2:B:295:HOH:O	1.99	0.61
1:B:81:HIS:HD2	1:B:197:ASP:H	1.49	0.59
1:A:81:HIS:HD2	1:A:197:ASP:H	1.50	0.59
1:A:213:GLU:OE2	1:A:215:ARG:HD3	2.04	0.58
1:A:15:LEU:HB3	2:A:325:HOH:O	2.04	0.58
1:A:204:GLN:HE22	1:B:207:LEU:H	1.51	0.57
1:A:183:GLN:NE2	1:A:185:ASN:HD21	1.93	0.55
1:B:213:GLU:OE2	1:B:215:ARG:HD3	2.06	0.55
1:B:128:ILE:O	1:B:129:ASP:HB2	2.06	0.55
1:A:81:HIS:CD2	1:A:197:ASP:H	2.25	0.54
1:B:6:GLU:HG2	2:B:390:HOH:O	2.08	0.53
1:A:128:ILE:O	1:A:129:ASP:HB2	2.08	0.53
1:A:207:LEU:H	1:B:204:GLN:NE2	2.06	0.52
1:B:81:HIS:CD2	1:B:197:ASP:H	2.25	0.52
1:B:157:GLN:NE2	2:B:368:HOH:O	2.43	0.52
1:B:45:LYS:HE2	1:B:47:ILE:HD11	1.91	0.51
1:B:5:GLU:HG3	2:B:359:HOH:O	2.11	0.51
1:A:209:LYS:N	2:B:240:HOH:O	2.45	0.50
1:A:223:PHE:HE2	2:A:340:HOH:O	1.94	0.50
1:B:169:HIS:HD2	2:B:272:HOH:O	1.95	0.49
1:A:23:ASN:HD21	1:A:130:PHE:H	1.61	0.48
1:B:111:GLU:HG2	1:B:188:ILE:HD11	1.95	0.48
1:B:75:PRO:HG2	2:B:266:HOH:O	2.13	0.48
1:A:171:ILE:HD11	1:A:177:GLN:CB	2.44	0.48
1:A:13:PRO:HB2	1:A:118:THR:HB	1.94	0.48
1:A:130:PHE:HB3	1:A:137:LEU:CD2	2.40	0.47
1:A:122:ARG:HH11	1:A:122:ARG:HG2	1.80	0.47
1:B:130:PHE:HB3	1:B:137:LEU:CD2	2.43	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:75:PRO:HD2	1:A:78:MET:HG3	1.98	0.46
1:A:203:THR:HG23	1:A:224:VAL:HG22	1.98	0.45
1:A:164:ASN:ND2	2:A:328:HOH:O	2.50	0.45
1:A:3:LYS:HB3	1:A:3:LYS:HE3	1.76	0.45
1:A:209:LYS:C	2:B:240:HOH:O	2.56	0.44
1:A:204:GLN:NE2	1:B:207:LEU:H	2.16	0.44
1:B:183:GLN:NE2	1:B:185:ASN:HD21	1.93	0.44
1:A:169:HIS:HE1	2:A:260:HOH:O	1.98	0.44
1:A:39:TYR:HB2	2:A:327:HOH:O	2.17	0.44
1:B:229:ILE:HG21	2:B:327:HOH:O	2.16	0.44
1:B:13:PRO:HB2	1:B:118:THR:HB	2.00	0.43
1:B:183:GLN:HE21	1:B:185:ASN:ND2	1.93	0.43
1:B:122:ARG:HG2	1:B:122:ARG:HH11	1.84	0.43
1:B:157:GLN:CB	2:B:342:HOH:O	2.58	0.43
1:B:18:LEU:HD13	1:B:18:LEU:C	2.39	0.43
1:A:171:ILE:HD11	1:A:177:GLN:HB2	1.99	0.43
1:B:33:GLY:HA3	1:B:43:THR:O	2.18	0.43
1:A:136:ILE:HG22	1:A:137:LEU:HD22	2.01	0.42
1:B:3:LYS:HE3	1:B:3:LYS:HB3	1.75	0.42
1:A:131:LYS:O	1:A:137:LEU:HB2	2.19	0.42
1:B:158:LYS:HG2	2:B:367:HOH:O	2.20	0.42
1:B:75:PRO:HD2	1:B:78:MET:HG3	2.00	0.42
1:B:23:ASN:HD21	1:B:130:PHE:H	1.68	0.41
1:B:169:HIS:HE1	2:B:268:HOH:O	2.04	0.41
1:A:183:GLN:HE21	1:A:185:ASN:ND2	1.94	0.41
1:B:53:LEU:HD13	1:B:55:VAL:O	2.21	0.40

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	228/238 (96%)	223 (98%)	4 (2%)	1 (0%)	34	24
1	B	228/238 (96%)	223 (98%)	4 (2%)	1 (0%)	34	24
All	All	456/476 (96%)	446 (98%)	8 (2%)	2 (0%)	34	24

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	66	TYR
1	B	66	TYR

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	200/207 (97%)	172 (86%)	28 (14%)	3	1
1	B	200/207 (97%)	171 (86%)	29 (14%)	3	1
All	All	400/414 (97%)	343 (86%)	57 (14%)	3	1

All (57) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	SER
1	A	3	LYS
1	A	6	GLU
1	A	15	LEU
1	A	18	LEU
1	A	23	ASN
1	A	29	VAL
1	A	41	LYS
1	A	52	LYS
1	A	53	LEU
1	A	66	TYR
1	A	107	LYS
1	A	118	THR
1	A	119	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	120	VAL
1	A	126	LYS
1	A	128	ILE
1	A	145	TYR
1	A	153	MET
1	A	156	LYS
1	A	158	LYS
1	A	164	ASN
1	A	172	GLU
1	A	178	LEU
1	A	190	ASP
1	A	195	LEU
1	A	214	LYS
1	A	221	LEU
1	B	2	SER
1	B	3	LYS
1	B	6	GLU
1	B	15	LEU
1	B	18	LEU
1	B	23	ASN
1	B	29	VAL
1	B	41	LYS
1	B	52	LYS
1	B	53	LEU
1	B	66	TYR
1	B	107	LYS
1	B	117	ASP
1	B	118	THR
1	B	119	LEU
1	B	120	VAL
1	B	126	LYS
1	B	128	ILE
1	B	145	TYR
1	B	156	LYS
1	B	158	LYS
1	B	164	ASN
1	B	172	GLU
1	B	178	LEU
1	B	190	ASP
1	B	195	LEU
1	B	214	LYS
1	B	215	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	221	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (25) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	23	ASN
1	A	81	HIS
1	A	121	ASN
1	A	135	ASN
1	A	139	HIS
1	A	149	ASN
1	A	157	GLN
1	A	159	ASN
1	A	169	HIS
1	A	177	GLN
1	A	185	ASN
1	A	204	GLN
1	A	212	ASN
1	B	23	ASN
1	B	81	HIS
1	B	121	ASN
1	B	135	ASN
1	B	149	ASN
1	B	157	GLN
1	B	164	ASN
1	B	169	HIS
1	B	177	GLN
1	B	185	ASN
1	B	204	GLN
1	B	212	ASN

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(Å²)	Q<0.9
1	A	230/238 (96%)	0.11	4 (1%)	70	72	12, 24, 46, 62	0
1	B	230/238 (96%)	0.31	10 (4%)	35	38	11, 23, 46, 64	2 (0%)
All	All	460/476 (96%)	0.21	14 (3%)	50	53	11, 24, 47, 64	2 (0%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1	ALA	21.8
1	B	230	THR	6.0
1	B	2	SER	5.1
1	B	3	LYS	5.0
1	B	229	ILE	4.7
1	A	39	TYR	3.2
1	B	228	GLY	3.0
1	B	80	ARG	2.9
1	B	191	GLY	2.9
1	A	214	LYS	2.6
1	A	157	GLN	2.5
1	B	190	ASP	2.4
1	B	212	ASN	2.2
1	A	230	THR	2.1

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 (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.