



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 17, 2021 – 06:40 AM EDT

PDB ID : 1KP5
Title : Cyclic Green Fluorescent Protein
Authors : Hofmann, A.; Iwai, H.; Plueckthun, A.; Wlodawer, A.
Deposited on : 2001-12-28
Resolution : 2.60 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.23.2
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) : 2.23.2

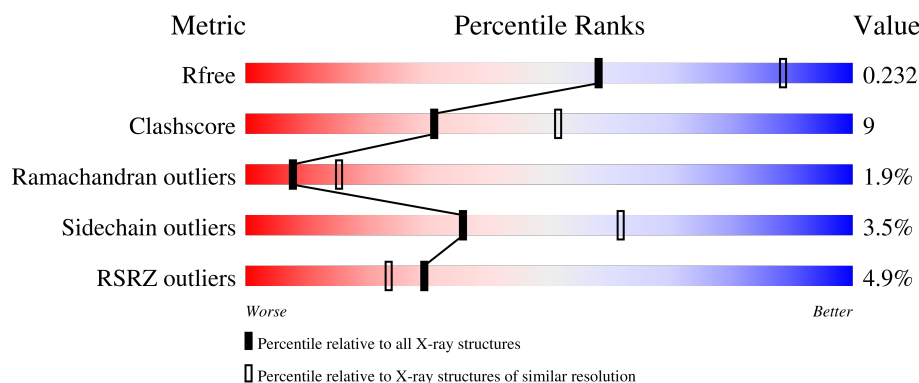
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 $\leq 5\%$. 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	246	<div> <div>7%</div> <div>76%</div> <div>21%</div> <div>..</div> </div>
1	B	246	<div> <div>3%</div> <div>81%</div> <div>15%</div> <div>..</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	GYS	A	76	X	-	-	-
1	GYS	B	376	X	-	-	-
2	SO4	B	923	-	-	-	X

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3936 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	243	Total	C	N	O	S	31	0	0
			1898	1200	327	366	5			
1	B	243	Total	C	N	O	S	59	0	0
			1904	1203	330	366	5			

There are 64 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	THR	-	expression tag	UNP P42212
A	2	GLY	-	expression tag	UNP P42212
A	3	SER	-	expression tag	UNP P42212
A	4	ARG	-	expression tag	UNP P42212
A	5	HIS	-	expression tag	UNP P42212
A	6	HIS	-	expression tag	UNP P42212
A	7	HIS	-	expression tag	UNP P42212
A	8	HIS	-	expression tag	UNP P42212
A	9	HIS	-	expression tag	UNP P42212
A	10	HIS	-	expression tag	UNP P42212
A	11	SER	MET	engineered mutation	UNP P42212
A	12	ARG	SER	engineered mutation	UNP P42212
A	35	HIS	GLN	engineered mutation	UNP P42212
A	76	GYS	SER	chromophore	UNP P42212
A	76	GYS	TYR	chromophore	UNP P42212
A	76	GYS	GLY	chromophore	UNP P42212
A	90	ARG	GLN	engineered mutation	UNP P42212
A	109	SER	PHE	engineered mutation	UNP P42212
A	110	PHE	TYR	engineered mutation	UNP P42212
A	151	LEU	MET	engineered mutation	UNP P42212
A	163	THR	MET	engineered mutation	UNP P42212
A	167	GLN	PRO	engineered mutation	UNP P42212
A	173	ALA	VAL	engineered mutation	UNP P42212
A	182	GLU	LYS	engineered mutation	UNP P42212
A	229	VAL	ILE	engineered mutation	UNP P42212

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	239	LEU	ILE	engineered mutation	UNP P42212
A	240	VAL	THR	engineered mutation	UNP P42212
A	241	PRO	HIS	engineered mutation	UNP P42212
A	242	ARG	GLY	engineered mutation	UNP P42212
A	243	GLY	MET	engineered mutation	UNP P42212
A	244	THR	ASP	engineered mutation	UNP P42212
A	245	GLY	GLU	engineered mutation	UNP P42212
B	301	THR	-	expression tag	UNP P42212
B	302	GLY	-	expression tag	UNP P42212
B	303	SER	-	expression tag	UNP P42212
B	304	ARG	-	expression tag	UNP P42212
B	305	HIS	-	expression tag	UNP P42212
B	306	HIS	-	expression tag	UNP P42212
B	307	HIS	-	expression tag	UNP P42212
B	308	HIS	-	expression tag	UNP P42212
B	309	HIS	-	expression tag	UNP P42212
B	310	HIS	-	expression tag	UNP P42212
B	311	SER	MET	engineered mutation	UNP P42212
B	312	ARG	SER	engineered mutation	UNP P42212
B	335	HIS	GLN	engineered mutation	UNP P42212
B	376	GYS	SER	chromophore	UNP P42212
B	376	GYS	TYR	chromophore	UNP P42212
B	376	GYS	GLY	chromophore	UNP P42212
B	390	ARG	GLN	engineered mutation	UNP P42212
B	409	SER	PHE	engineered mutation	UNP P42212
B	410	PHE	TYR	engineered mutation	UNP P42212
B	451	LEU	MET	engineered mutation	UNP P42212
B	463	THR	MET	engineered mutation	UNP P42212
B	467	GLN	PRO	engineered mutation	UNP P42212
B	473	ALA	VAL	engineered mutation	UNP P42212
B	482	GLU	LYS	engineered mutation	UNP P42212
B	529	VAL	ILE	engineered mutation	UNP P42212
B	539	LEU	ILE	engineered mutation	UNP P42212
B	540	VAL	THR	engineered mutation	UNP P42212
B	541	PRO	HIS	engineered mutation	UNP P42212
B	542	ARG	GLY	engineered mutation	UNP P42212
B	543	GLY	MET	engineered mutation	UNP P42212
B	544	THR	ASP	engineered mutation	UNP P42212
B	545	GLY	GLU	engineered mutation	UNP P42212

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

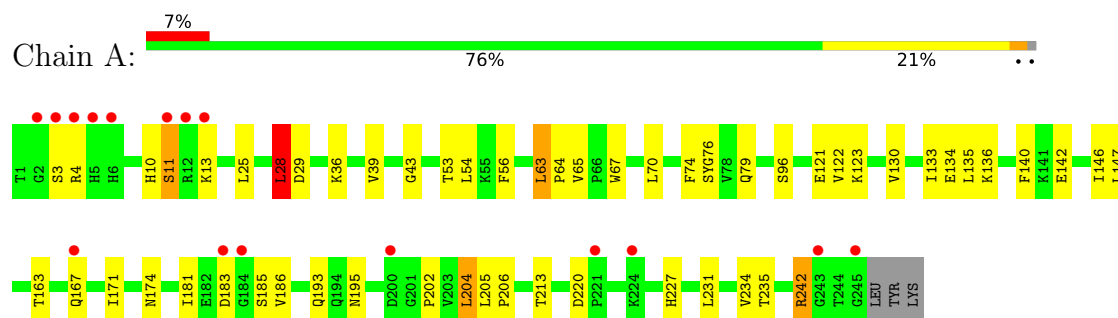
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	36	Total	O	0	0
			36	36		
3	B	58	Total	O	0	0
			58	58		

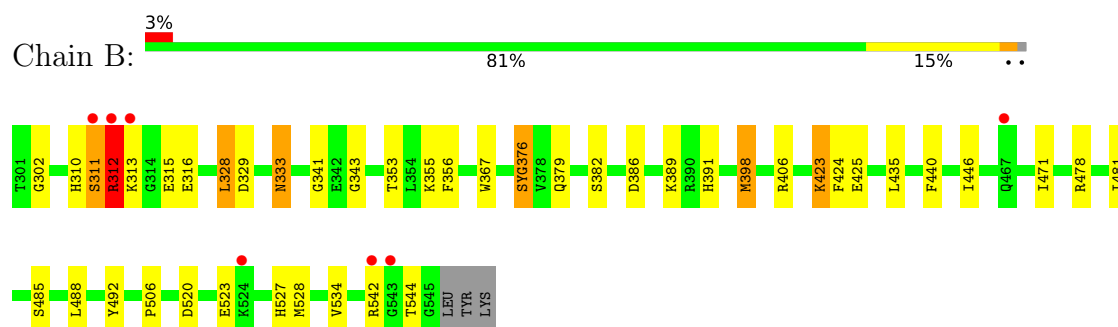
3 Residue-property plots

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: Green Fluorescent Protein



• Molecule 1: Green Fluorescent Protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α , β , γ	142.45Å 142.45Å 183.81Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.99 – 2.60 24.65 – 2.60	Depositor EDS
% Data completeness (in resolution range)	92.8 (19.99-2.60) 92.9 (24.65-2.60)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.55 (at 2.60Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.208 , 0.256 0.197 , 0.232	Depositor DCC
R_{free} test set	3194 reflections (10.01%)	wwPDB-VP
Wilson B-factor (Å ²)	37.0	Xtriage
Anisotropy	0.492	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 44.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3936	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.78% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GYS, SO4

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	0.38	0/1918	0.63	1/2593 (0.0%)
1	B	0.39	0/1924	0.62	0/2600
All	All	0.39	0/3842	0.62	1/5193 (0.0%)

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	0
1	B	1	0
All	All	2	0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	28	LEU	CA-CB-CG	5.82	128.69	115.30

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	76	GYS	CA1
1	B	376	GYS	CA1

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	1898	0	1814	37	0
1	B	1904	0	1823	32	0
2	A	20	0	0	0	0
2	B	20	0	0	0	0
3	A	36	0	0	0	0
3	B	58	0	0	2	0
All	All	3936	0	3637	67	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 9.

The worst 5 of 67 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:398:MET:HE1	1:B:423:LYS:HA	1.43	0.98
1:A:36:LYS:HB3	1:B:425:GLU:HG2	1.65	0.79
1:B:367:TRP:HB3	1:B:528:MET:HE3	1.71	0.73
1:B:478:ARG:HG2	1:B:488:LEU:HD23	1.74	0.70
1:B:398:MET:CE	1:B:423:LYS:HA	2.20	0.69

There are no symmetry-related clashes.

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	240/246 (98%)	216 (90%)	20 (8%)	4 (2%)	9 18

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	240/246 (98%)	226 (94%)	9 (4%)	5 (2%)	7	13
All	All	480/492 (98%)	442 (92%)	29 (6%)	9 (2%)	8	15

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	10	HIS
1	B	312	ARG
1	A	13	LYS
1	B	544	THR
1	A	3	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	200/213 (94%)	191 (96%)	9 (4%)	27	52
1	B	201/213 (94%)	196 (98%)	5 (2%)	47	73
All	All	401/426 (94%)	387 (96%)	14 (4%)	36	62

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	235	THR
1	A	242	ARG
1	B	423	LYS
1	B	333	ASN
1	B	398	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	208	ASN
1	B	487	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	333	ASN
1	B	494	GLN
1	B	459	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
1	GYS	B	376	1	22,22,23	1.99	10 (45%)	27,30,32	5.10	11 (40%)
1	GYS	A	76	1	22,22,23	1.95	9 (40%)	27,30,32	5.07	10 (37%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	GYS	B	376	1	1/1/5/7	3/9/29/30	0/2/2/2
1	GYS	A	76	1	1/1/5/7	2/9/29/30	0/2/2/2

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	76	GYS	C1-N2	-4.38	1.25	1.32
1	B	376	GYS	C1-N2	-4.04	1.26	1.32
1	B	376	GYS	CG2-CB2	3.48	1.53	1.46
1	A	76	GYS	CG2-CB2	3.34	1.53	1.46

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	376	GYS	C1-N3	-3.24	1.31	1.37

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	376	GYS	C2-CA2-N2	-15.24	98.26	108.93
1	A	76	GYS	C2-CA2-N2	-15.05	98.39	108.93
1	A	76	GYS	O2-C2-CA2	-12.72	123.82	130.96
1	B	376	GYS	O2-C2-CA2	-12.66	123.85	130.96
1	B	376	GYS	CA2-C2-N3	11.41	108.77	103.37

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	76	GYS	CA1
1	B	376	GYS	CA1

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	76	GYS	N-CA1-CB1-OG1
1	B	376	GYS	N-CA1-CB1-OG1
1	B	376	GYS	C1-CA1-CB1-OG1
1	A	76	GYS	C1-CA1-CB1-OG1
1	B	376	GYS	C-CA3-N3-C2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	376	GYS	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

8 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	SO4	A	902	-	4,4,4	0.26	0	6,6,6	0.13	0
2	SO4	B	903	-	4,4,4	0.26	0	6,6,6	0.09	0
2	SO4	A	922	-	4,4,4	0.26	0	6,6,6	0.07	0
2	SO4	B	924	-	4,4,4	0.29	0	6,6,6	0.09	0
2	SO4	A	901	-	4,4,4	0.28	0	6,6,6	0.06	0
2	SO4	B	921	-	4,4,4	0.24	0	6,6,6	0.08	0
2	SO4	A	904	-	4,4,4	0.27	0	6,6,6	0.15	0
2	SO4	B	923	-	4,4,4	0.33	0	6,6,6	0.06	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	236/246 (95%)	-0.22	16 (6%) 17 12	18, 33, 76, 91	0
1	B	230/246 (93%)	-0.45	7 (3%) 50 43	20, 30, 59, 90	0
All	All	466/492 (94%)	-0.34	23 (4%) 29 23	18, 31, 65, 91	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	5	HIS	10.0
1	A	2	GLY	7.9
1	A	3	SER	6.3
1	A	4	ARG	4.9
1	A	243	GLY	4.9

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
1	GYS	A	76	21/22	0.96	0.20	23,29,33,34	0
1	GYS	B	376	21/22	0.96	0.20	19,24,29,29	0

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q<0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	SO4	B	923	5/5	0.73	0.50	90,90,90,90	0
2	SO4	B	924	5/5	0.83	0.27	82,89,90,90	0
2	SO4	A	922	5/5	0.91	0.25	90,90,90,90	0
2	SO4	B	903	5/5	0.93	0.20	86,86,87,88	0
2	SO4	A	901	5/5	0.95	0.20	84,85,86,86	0
2	SO4	A	904	5/5	0.95	0.14	45,51,55,56	0
2	SO4	A	902	5/5	0.98	0.10	43,50,51,56	0
2	SO4	B	921	5/5	0.99	0.07	46,47,48,53	0

6.5 Other polymers [i](#)

There are no such residues in this entry.