



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 14, 2020 – 09:50 AM BST

PDB ID : 6TUS  
Title : human XPG, Apo2 form  
Authors : Ruiz, F.M.; Fernandez-Tornero, C.  
Deposited on : 2020-01-08  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 : 2.14.4.dev1  
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) : 2.14.4.dev1

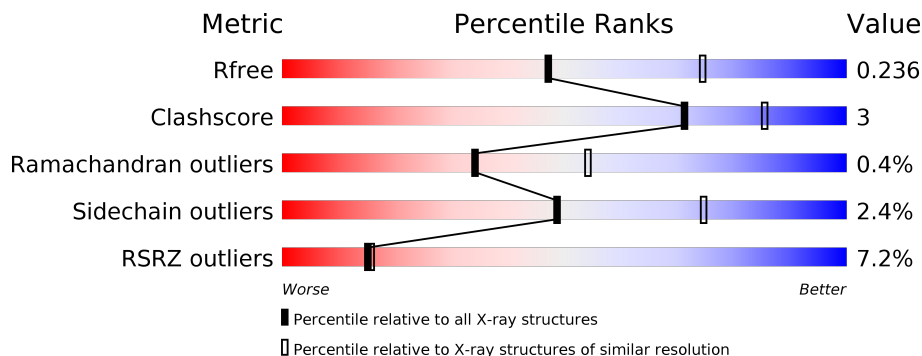
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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  $\geq 3$ , 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	356	<div style="display: flex; align-items: center;"> <div style="width: 5%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 72%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 9%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 14%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 5px;">5%      72%      9%      19%</p>
2	B	355	<div style="display: flex; align-items: center;"> <div style="width: 6%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 73%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 9%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 12%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 5px;">6%      73%      9%      17%</p>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4823 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 DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	289	2340	1526	389	416	9	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	748	GLY	-	linker	UNP P28715
A	749	THR	-	linker	UNP P28715
A	812	ALA	ASP	conflict	UNP P28715
A	991	ALA	-	expression tag	UNP P28715

- Molecule 2 is a protein called DNA repair protein complementing XP-G cells,DNA repair protein complementing XP-G cells.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	294	2386	1551	402	424	9	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

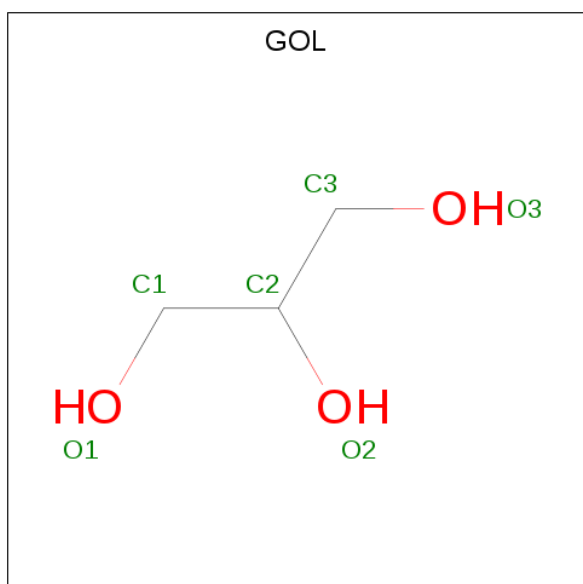
Chain	Residue	Modelled	Actual	Comment	Reference
B	748	GLY	-	linker	UNP P28715
B	749	THR	-	linker	UNP P28715
B	812	ALA	ASP	conflict	UNP P28715

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 6 3 3	0	0
4	B	1	Total C O 6 3 3	0	0
4	B	1	Total C O 6 3 3	0	0

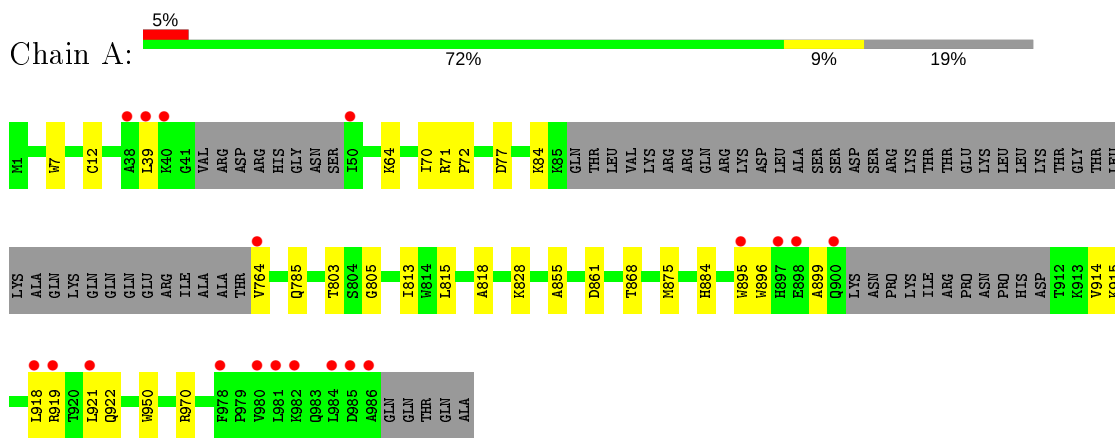
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	32	Total O 32 32	0	0
5	B	22	Total O 22 22	0	0

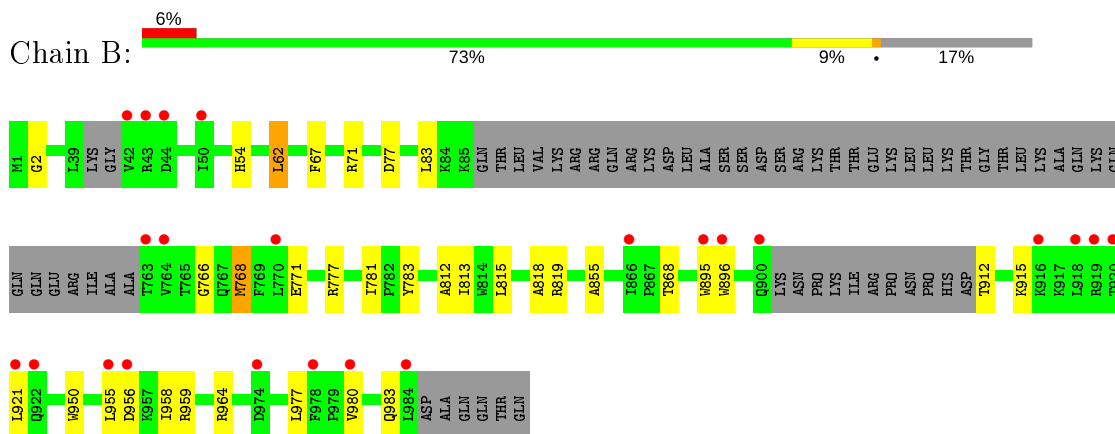
### 3 Residue-property plots

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: DNA repair protein complementing XP-G cells, DNA repair protein complementing XP-G cells



- Molecule 2: DNA repair protein complementing XP-G cells, DNA repair protein complementing XP-G cells



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	134.22Å 134.22Å 110.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.85 – 2.50 47.85 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.85-2.50) 99.7 (47.85-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.31 (at 2.51Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.203 , 0.237 0.203 , 0.236	Depositor DCC
$R_{free}$ test set	888 reflections (2.50%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	72.6	Xtrriage
Anisotropy	0.204	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 66.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4823	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	82.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, 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	0.25	0/2399	0.42	0/3247
2	B	0.25	0/2446	0.42	0/3311
All	All	0.25	0/4845	0.42	0/6558

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	2340	0	2344	18	0
2	B	2386	0	2386	16	0
3	A	15	0	0	0	0
3	B	10	0	0	0	0
4	A	6	0	8	1	0
4	B	12	0	16	0	0
5	A	32	0	0	0	0
5	B	22	0	0	0	0
All	All	4823	0	4754	33	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.



All (33) 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)
2:B:912:THR:HA	2:B:915:LYS:HB2	1.75	0.68
2:B:62:LEU:HD21	2:B:781:ILE:HG13	1.83	0.59
1:A:813:ILE:HG12	1:A:818:ALA:HB2	1.85	0.59
1:A:803:THR:HG23	1:A:805:GLY:H	1.67	0.58
2:B:815:LEU:HD21	2:B:855:ALA:HB2	1.84	0.58
2:B:896:TRP:NE1	2:B:921:LEU:O	2.34	0.58
1:A:815:LEU:HD21	1:A:855:ALA:HB2	1.87	0.57
1:A:915:LYS:HA	1:A:918:LEU:HB2	1.87	0.57
2:B:813:ILE:HG12	2:B:818:ALA:HB2	1.87	0.55
2:B:895:TRP:CZ2	2:B:915:LYS:HB3	2.43	0.53
1:A:12:CYS:HA	2:B:83:LEU:HD11	1.93	0.51
1:A:828:LYS:H	1:A:828:LYS:HD2	1.77	0.50
1:A:895:TRP:CH2	1:A:919:ARG:HB3	2.48	0.49
1:A:84:LYS:HG3	1:A:861:ASP:HB3	1.96	0.48
1:A:896:TRP:NE1	1:A:921:LEU:O	2.47	0.48
1:A:7:TRP:HD1	4:A:1004:GOL:H12	1.80	0.47
1:A:77:ASP:O	1:A:785:GLN:NE2	2.45	0.47
2:B:2:GLY:HA2	2:B:812:ALA:HB2	1.98	0.45
2:B:777:ARG:NH1	2:B:783:TYR:OH	2.50	0.45
2:B:83:LEU:HD12	2:B:83:LEU:H	1.82	0.45
2:B:54:HIS:CD2	2:B:768:MET:HB3	2.53	0.44
1:A:895:TRP:CZ2	1:A:915:LYS:HB3	2.51	0.44
1:A:64:LYS:HD3	1:A:64:LYS:HA	1.83	0.43
1:A:39:LEU:HD13	1:A:764:VAL:HG13	1.99	0.43
2:B:956:ASP:N	2:B:956:ASP:OD2	2.51	0.43
2:B:958:ILE:HG21	2:B:977:LEU:HD11	2.02	0.42
2:B:67:PHE:O	2:B:983:GLN:HG2	2.19	0.42
1:A:922:GLN:H	1:A:922:GLN:HG2	1.69	0.42
1:A:71:ARG:HB3	1:A:950:TRP:CE2	2.56	0.41
2:B:955:LEU:O	2:B:959:ARG:HG3	2.21	0.41
1:A:70:ILE:O	1:A:72:PRO:HD3	2.21	0.40
2:B:71:ARG:HB3	2:B:950:TRP:CE2	2.56	0.40
1:A:914:VAL:O	1:A:918:LEU:HG	2.22	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	281/356 (79%)	274 (98%)	6 (2%)	1 (0%)	34	54
2	B	286/355 (81%)	276 (96%)	9 (3%)	1 (0%)	41	61
All	All	567/711 (80%)	550 (97%)	15 (3%)	2 (0%)	34	54

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	899	ALA
2	B	766	GLY

### 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	252/312 (81%)	248 (98%)	4 (2%)	62	84
2	B	258/312 (83%)	250 (97%)	8 (3%)	40	67
All	All	510/624 (82%)	498 (98%)	12 (2%)	49	74

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	868	THR
1	A	875	MET
1	A	884	HIS
1	A	970	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	62	LEU
2	B	77	ASP
2	B	768	MET
2	B	771	GLU
2	B	819	ARG
2	B	868	THR
2	B	964	ARG
2	B	980	VAL

Some 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](#)

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	$\# Z  > 2$
3	SO4	B	1001	-	4,4,4	0.14	0	6,6,6	0.06	0
4	GOL	B	1003	-	5,5,5	0.91	0	5,5,5	1.01	0
3	SO4	A	1003	-	4,4,4	0.14	0	6,6,6	0.05	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	SO4	A	1001	-	4,4,4	0.14	0	6,6,6	0.06	0
4	GOL	A	1004	-	5,5,5	0.91	0	5,5,5	0.98	0
3	SO4	A	1002	-	4,4,4	0.14	0	6,6,6	0.05	0
3	SO4	B	1002	-	4,4,4	0.15	0	6,6,6	0.05	0
4	GOL	B	1004	-	5,5,5	0.94	0	5,5,5	0.91	0

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
4	GOL	B	1003	-	-	2/4/4/4	-
4	GOL	B	1004	-	-	2/4/4/4	-
4	GOL	A	1004	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	1003	GOL	C1-C2-C3-O3
4	B	1004	GOL	C1-C2-C3-O3
4	B	1003	GOL	O2-C2-C3-O3
4	B	1004	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1004	GOL	1	0

## 5.7 Other polymers [i](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	289/356 (81%)	0.39	19 (6%) 18 19	52, 72, 120, 144	0
2	B	294/355 (82%)	0.62	23 (7%) 13 13	54, 79, 126, 144	0
All	All	583/711 (81%)	0.50	42 (7%) 15 16	52, 76, 124, 144	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	42	VAL	5.0
2	B	900	GLN	4.9
1	A	919	ARG	4.8
1	A	918	LEU	4.5
1	A	38	ALA	4.3
2	B	763	THR	4.2
2	B	896	TRP	4.1
2	B	918	LEU	3.8
2	B	764	VAL	3.7
2	B	50	ILE	3.7
2	B	922	GLN	3.7
2	B	921	LEU	3.3
2	B	980	VAL	3.2
2	B	978	PHE	3.1
1	A	764	VAL	3.0
1	A	985	ASP	3.0
2	B	919	ARG	2.8
1	A	921	LEU	2.7
2	B	43	ARG	2.7
2	B	920	THR	2.7
1	A	981	LEU	2.7
2	B	895	TRP	2.6
1	A	50	ILE	2.6
1	A	39	LEU	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	984	LEU	2.5
2	B	984	LEU	2.5
1	A	978	PHE	2.5
2	B	866	ILE	2.4
2	B	974	ASP	2.4
2	B	956	ASP	2.4
1	A	900	GLN	2.4
1	A	40	LYS	2.3
1	A	898	GLU	2.3
1	A	897	HIS	2.3
1	A	895	TRP	2.2
2	B	770	LEU	2.2
1	A	982	LYS	2.1
1	A	986	ALA	2.1
2	B	955	LEU	2.1
2	B	916	LYS	2.0
1	A	980	VAL	2.0
2	B	44	ASP	2.0

## 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 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	SO4	A	1002	5/5	0.74	0.34	102,113,159,170	0
4	GOL	B	1004	6/6	0.77	0.30	80,90,94,107	0
3	SO4	A	1003	5/5	0.83	0.33	131,132,143,157	0
4	GOL	A	1004	6/6	0.88	0.16	74,80,82,83	0
4	GOL	B	1003	6/6	0.90	0.15	69,87,94,95	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	SO4	A	1001	5/5	0.92	0.17	97,99,138,139	0
3	SO4	B	1001	5/5	0.93	0.16	104,109,134,143	0
3	SO4	B	1002	5/5	0.95	0.30	134,156,171,174	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.