



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 19, 2024 – 08:27 PM EDT

PDB ID : 1YJ5  
Title : Molecular architecture of mammalian polynucleotide kinase, a DNA repair enzyme  
Authors : Bernstein, N.K.; Williams, R.S.; Rakovszky, M.L.; Cui, D.; Green, R.; Karimi-Busheri, F.; Mani, R.S.; Galicia, S.; Koch, C.A.; Cass, C.E.; Durocher, D.; Weinfeld, M.; Glover, J.N.M.  
Deposited on : 2005-01-13  
Resolution : 2.80 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.39

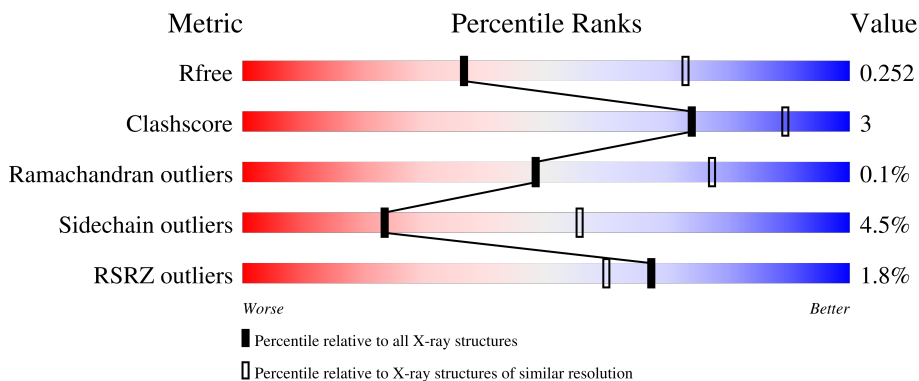
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	164625	3657 (2.80-2.80)
Clashscore	180529	4123 (2.80-2.80)
Ramachandran outliers	177936	4071 (2.80-2.80)
Sidechain outliers	177891	4073 (2.80-2.80)
RSRZ outliers	164620	3659 (2.80-2.80)

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  $\geq 3$ , 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	383	 5% 87% 10% ..
1	B	383	 5% 89% 10% .
2	C	143	 5% 66% 6% .. 27%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6844 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 5' polynucleotide kinase-3' phosphatase catalytic domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	N	O	S				Se
1	A	379	2955	1891	517	535	7	5	0	0	0
1	B	383	2979	1905	521	541	7	5	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	219	MSE	MET	modified residue	UNP Q9JLV6
A	264	MSE	MET	modified residue	UNP Q9JLV6
A	465	MSE	MET	modified residue	UNP Q9JLV6
A	476	MSE	MET	modified residue	UNP Q9JLV6
A	478	MSE	MET	modified residue	UNP Q9JLV6
B	219	MSE	MET	modified residue	UNP Q9JLV6
B	264	MSE	MET	modified residue	UNP Q9JLV6
B	465	MSE	MET	modified residue	UNP Q9JLV6
B	476	MSE	MET	modified residue	UNP Q9JLV6
B	478	MSE	MET	modified residue	UNP Q9JLV6

- Molecule 2 is a protein called 5' polynucleotide kinase-3' phosphatase FHA domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	105	772	491	137	143	1	0	0	0

- 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	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	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
3	B	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

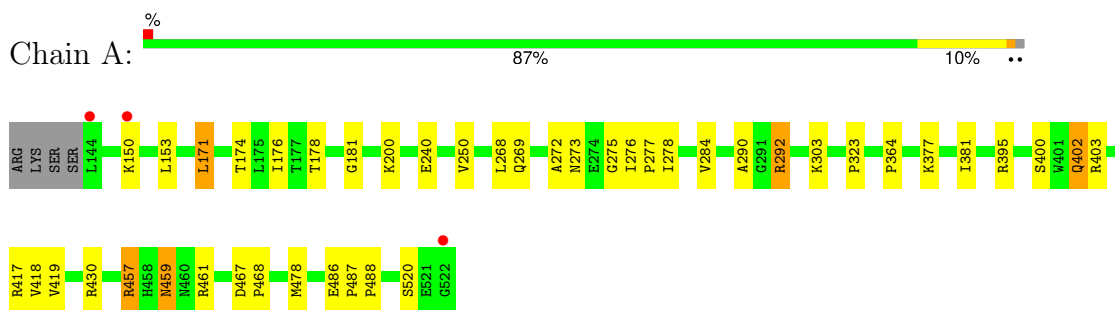
- Molecule 4 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	50	Total 50	O 50	0	0
4	B	26	Total 26	O 26	0	0
4	C	2	Total 2	O 2	0	0

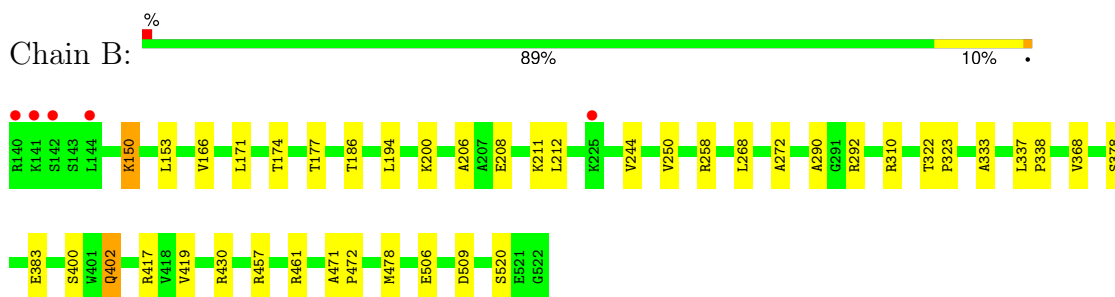
### 3 Residue-property plots [i](#)

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.

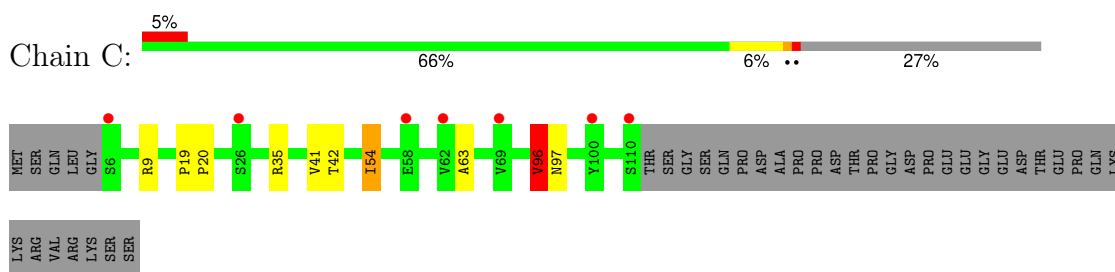
- Molecule 1: 5' polynucleotide kinase-3' phosphatase catalytic domain



- Molecule 1: 5' polynucleotide kinase-3' phosphatase catalytic domain



- Molecule 2: 5' polynucleotide kinase-3' phosphatase FHA domain



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.47Å 169.87Å 77.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.75 – 2.80 29.75 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.5 (29.75-2.80) 99.4 (29.75-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.85 (at 2.80Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.219 , 0.253 0.218 , 0.252	Depositor DCC
$R_{free}$ test set	1562 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.3	Xtrriage
Anisotropy	0.652	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 34.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	6844	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.80% 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: 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	0.30	0/3027	0.46	0/4099
1	B	0.31	0/3052	0.47	0/4132
2	C	0.28	0/789	0.48	0/1081
All	All	0.30	0/6868	0.47	0/9312

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	2955	0	2947	19	0
1	B	2979	0	2958	18	0
2	C	772	0	772	4	0
3	A	35	0	0	0	0
3	B	25	0	0	0	0
4	A	50	0	0	0	0
4	B	26	0	0	0	0
4	C	2	0	0	0	0
All	All	6844	0	6677	40	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 (40) 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:461:ARG:HH11	1:A:520:SER:HB2	1.41	0.85
1:A:403:ARG:NH2	1:B:338:PRO:O	2.26	0.69
1:A:269:GLN:HE22	1:A:278:ILE:H	1.41	0.67
1:B:400:SER:HB2	1:B:402:GLN:HE21	1.61	0.65
1:A:430:ARG:NH2	1:A:486:GLU:O	2.32	0.61
1:B:150:LYS:HD3	1:B:150:LYS:H	1.69	0.58
1:A:400:SER:HB2	1:A:402:GLN:HE21	1.68	0.57
1:A:290:ALA:HB3	1:A:292:ARG:HD2	1.87	0.57
1:A:457:ARG:HG3	1:A:478:MSE:HE2	1.92	0.51
1:A:377:LYS:O	1:A:381:ILE:HG12	2.12	0.50
1:A:178:THR:HG22	1:A:181:GLY:H	1.78	0.49
1:B:206:ALA:HB2	1:B:212:LEU:HD11	1.96	0.48
1:A:153:LEU:HB2	1:A:250:VAL:HB	1.95	0.48
1:A:171:LEU:HD22	1:A:176:ILE:HG12	1.95	0.48
2:C:54:ILE:HG23	2:C:63:ALA:HB3	1.96	0.47
1:B:153:LEU:HB2	1:B:250:VAL:HB	1.97	0.47
1:B:166:VAL:HG22	1:B:211:LYS:HB3	1.96	0.47
1:B:461:ARG:NH1	1:B:520:SER:OG	2.48	0.47
2:C:35:ARG:HA	2:C:41:VAL:HB	1.99	0.45
1:A:487:PRO:HA	1:A:488:PRO:HD3	1.87	0.45
1:A:268:LEU:HA	1:A:272:ALA:HB3	2.00	0.43
1:B:290:ALA:O	1:B:310:ARG:NH1	2.48	0.43
2:C:19:PRO:HA	2:C:20:PRO:HD3	1.90	0.43
1:A:276:ILE:HA	1:A:277:PRO:HD3	1.84	0.43
1:A:364:PRO:HA	1:A:417:ARG:HG2	2.00	0.43
1:B:177:THR:HB	1:B:194:LEU:HD13	2.01	0.42
1:B:292:ARG:HE	1:B:322:THR:HG21	1.83	0.42
1:A:467:ASP:HA	1:A:468:PRO:HD3	1.84	0.42
1:A:174:THR:HG23	1:A:323:PRO:HG3	2.00	0.42
1:B:174:THR:HG23	1:B:323:PRO:HG3	2.02	0.42
1:B:457:ARG:HG3	1:B:478:MSE:HE2	2.00	0.42
1:B:268:LEU:HA	1:B:272:ALA:HB3	2.00	0.42
1:B:471:ALA:HA	1:B:472:PRO:HD3	1.88	0.42
1:B:292:ARG:HD3	1:B:333:ALA:HB3	2.02	0.41
1:B:150:LYS:O	1:B:150:LYS:HG2	2.20	0.41
1:B:292:ARG:HH21	1:B:322:THR:HB	1.84	0.41
1:B:368:VAL:HG11	1:B:430:ARG:HB2	2.03	0.41
1:A:273:ASN:O	1:A:275:GLY:N	2.48	0.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:459:ASN:HD22	1:A:459:ASN:HA	1.65	0.41
2:C:96:VAL:HB	2:C:97:ASN:H	1.57	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	377/383 (98%)	365 (97%)	12 (3%)	0	100	100
1	B	381/383 (100%)	371 (97%)	10 (3%)	0	100	100
2	C	103/143 (72%)	97 (94%)	5 (5%)	1 (1%)	13	39
All	All	861/909 (95%)	833 (97%)	27 (3%)	1 (0%)	48	77

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	96	VAL

### 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	316/316 (100%)	303 (96%)	13 (4%)	26	59
1	B	318/316 (101%)	303 (95%)	15 (5%)	22	54

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	C	83/123 (68%)	79 (95%)	4 (5%)	21	53
All	All	717/755 (95%)	685 (96%)	32 (4%)	23	55

All (32) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	150	LYS
1	A	171	LEU
1	A	200	LYS
1	A	240	GLU
1	A	284	VAL
1	A	292	ARG
1	A	303	LYS
1	A	395	ARG
1	A	402	GLN
1	A	418	VAL
1	A	419	VAL
1	A	457	ARG
1	A	459	ASN
1	B	150	LYS
1	B	171	LEU
1	B	186	THR
1	B	200	LYS
1	B	208	GLU
1	B	244	VAL
1	B	258	ARG
1	B	337	LEU
1	B	378	SER
1	B	383	GLU
1	B	402	GLN
1	B	417	ARG
1	B	419	VAL
1	B	506	GLU
1	B	509	ASP
2	C	9	ARG
2	C	42	THR
2	C	54	ILE
2	C	96	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (11) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	253	HIS
1	A	269	GLN
1	A	402	GLN
1	A	459	ASN
1	B	163	GLN
1	B	217	ASN
1	B	295	ASN
1	B	402	GLN
1	B	409	GLN
1	B	460	ASN
1	B	505	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 oligosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

12 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	A	102	-	4,4,4	0.25	0	6,6,6	0.09	0
3	SO4	A	110	-	4,4,4	0.24	0	6,6,6	0.14	0
3	SO4	B	104	-	4,4,4	0.25	0	6,6,6	0.15	0
3	SO4	A	101	-	4,4,4	0.22	0	6,6,6	0.09	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	SO4	A	103	-	4,4,4	0.20	0	6,6,6	0.13	0
3	SO4	B	106	-	4,4,4	0.24	0	6,6,6	0.05	0
3	SO4	A	108	-	4,4,4	0.24	0	6,6,6	0.13	0
3	SO4	A	105	-	4,4,4	0.27	0	6,6,6	0.09	0
3	SO4	B	109	-	4,4,4	0.23	0	6,6,6	0.09	0
3	SO4	B	111	-	4,4,4	0.25	0	6,6,6	0.19	0
3	SO4	B	112	-	4,4,4	0.24	0	6,6,6	0.09	0
3	SO4	A	107	-	4,4,4	0.25	0	6,6,6	0.10	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, 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	374/383 (97%)	-0.34	3 (0%) 82 77	15, 22, 34, 41	0
1	B	378/383 (98%)	-0.06	5 (1%) 74 67	13, 25, 35, 47	0
2	C	105/143 (73%)	0.81	7 (6%) 25 19	38, 39, 42, 42	0
All	All	857/909 (94%)	-0.07	15 (1%) 67 60	13, 24, 39, 47	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	142	SER	5.7
2	C	6	SER	4.7
1	B	141	LYS	4.4
2	C	110	SER	4.0
1	B	140	ARG	2.9
2	C	69	VAL	2.6
1	A	522	GLY	2.6
2	C	58	GLU	2.5
1	A	144	LEU	2.4
2	C	100	TYR	2.3
2	C	62	VAL	2.2
1	B	144	LEU	2.2
2	C	26	SER	2.2
1	B	225	LYS	2.1
1	A	150	LYS	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 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(Å <sup>2</sup> )	Q<0.9
3	SO4	B	111	5/5	0.71	0.31	33,33,33,34	5
3	SO4	B	109	5/5	0.75	0.32	119,120,120,120	0
3	SO4	B	106	5/5	0.80	0.18	111,111,111,111	0
3	SO4	A	110	5/5	0.81	0.27	40,40,40,41	5
3	SO4	A	105	5/5	0.81	0.19	84,84,85,85	0
3	SO4	A	102	5/5	0.84	0.15	73,73,73,73	0
3	SO4	A	107	5/5	0.85	0.33	90,91,91,91	0
3	SO4	A	108	5/5	0.85	0.21	93,93,93,93	0
3	SO4	B	112	5/5	0.91	0.17	35,35,35,36	5
3	SO4	A	103	5/5	0.93	0.11	46,46,47,47	0
3	SO4	A	101	5/5	0.99	0.04	31,32,32,32	0
3	SO4	B	104	5/5	0.99	0.04	24,25,25,25	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.