



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 08:58 pm BST

PDB ID : 4CXK  
Title : G9 mutant of PAS, arylsulfatase from *Pseudomonas Aeruginosa*  
Authors : Miton, C.M.; Jonas, S.; Mohammed, M.F.; Fischer, G.; Loo, B.v.; Kintses, B.;  
Hyvonen, M.; Tokuriki, N.; Hollfelder, F.  
Deposited on : 2014-04-07  
Resolution : 1.86 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](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	:	2.11
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

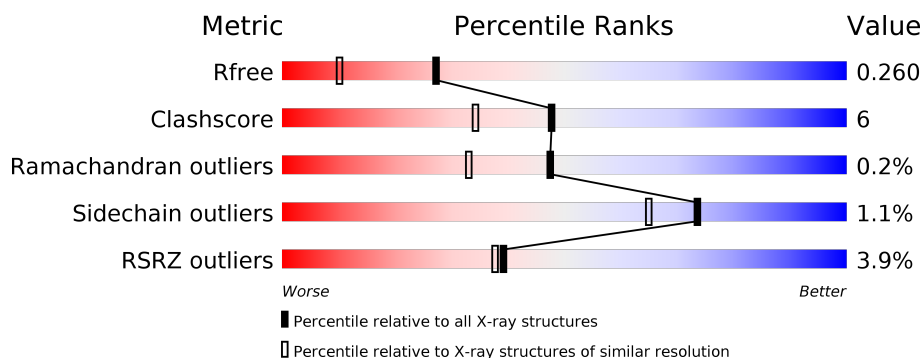
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.86 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	536	<div> <div>4%</div> <div>85%</div> <div>12%</div> <div>••</div> </div>
1	B	536	<div> <div>4%</div> <div>85%</div> <div>12%</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
3	SO4	A	1529	-	-	X	-
4	PEG	B	1529	-	-	X	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 9280 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 ARYLSULFATASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	525	Total	C	N	O	S	0	1	0
			4147	2639	730	770	8			
1	B	525	Total	C	N	O	S	0	2	0
			4134	2631	726	769	8			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	535	GLU	-	expression tag	UNP P51691
A	536	ARG	-	expression tag	UNP P51691
A	48	THR	ALA	engineered mutation	UNP P51691
A	50	ALA	THR	engineered mutation	UNP P51691
A	69	LEU	ILE	engineered mutation	UNP P51691
A	72	VAL	MET	engineered mutation	UNP P51691
A	337	ASP	GLY	engineered mutation	UNP P51691
A	352	SER	ARG	engineered mutation	UNP P51691
A	461	GLY	GLU	engineered mutation	UNP P51691
A	523	ASP	GLU	engineered mutation	UNP P51691
B	535	GLU	-	expression tag	UNP P51691
B	536	ARG	-	expression tag	UNP P51691
B	48	THR	ALA	engineered mutation	UNP P51691
B	50	ALA	THR	engineered mutation	UNP P51691
B	69	LEU	ILE	engineered mutation	UNP P51691
B	72	VAL	MET	engineered mutation	UNP P51691
B	337	ASP	GLY	engineered mutation	UNP P51691
B	352	SER	ARG	engineered mutation	UNP P51691
B	461	GLY	GLU	engineered mutation	UNP P51691
B	523	ASP	GLU	engineered mutation	UNP P51691

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Ca	0	0
			1	1		
2	A	1	Total	Ca	0	0
			1	1		

- 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	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			7	4	3		
4	B	1	Total	C	O	0	0
			7	4	3		

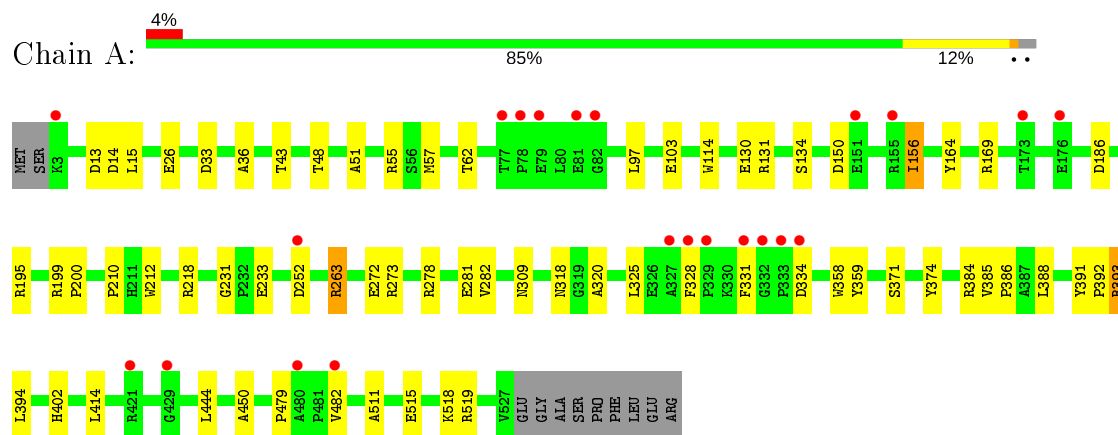
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	502	Total	O	0	0
			502	502		
5	B	471	Total	O	0	0
			471	471		

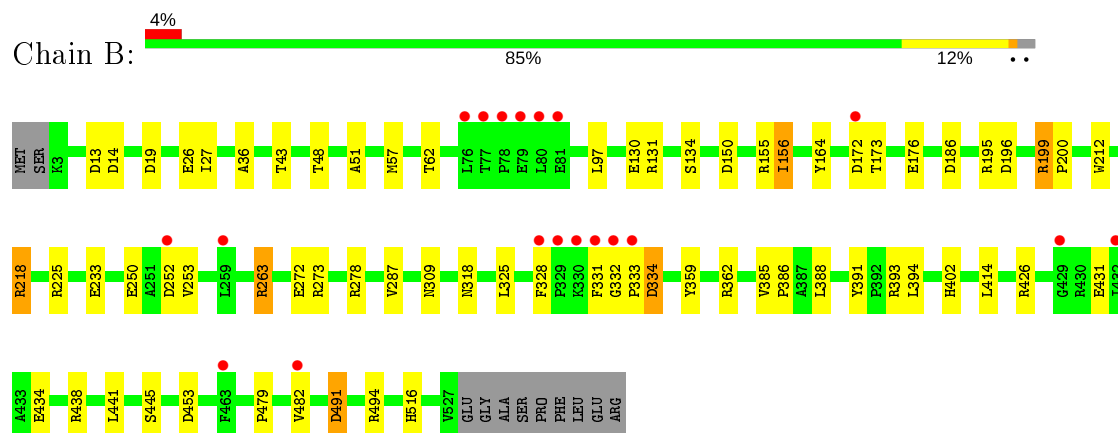
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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: ARYLSULFATASE



#### • Molecule 1: ARYLSULFATASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	188.30 Å 67.06 Å 90.05 Å 90.00° 94.04° 90.00°	Depositor
Resolution (Å)	93.92 – 1.86 67.33 – 1.86	Depositor EDS
% Data completeness (in resolution range)	98.2 (93.92-1.86) 98.2 (67.33-1.86)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 1.86 Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, $R_{free}$	0.208 , 0.254 0.215 , 0.260	Depositor DCC
$R_{free}$ test set	4617 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.6	Xtriage
Anisotropy	0.069	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 52.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	9280	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.49% 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: DDZ, CA, PEG, 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	0.95	3/4255 (0.1%)	0.97	10/5789 (0.2%)
1	B	0.97	2/4247 (0.0%)	0.98	15/5781 (0.3%)
All	All	0.96	5/8502 (0.1%)	0.98	25/11570 (0.2%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	26	GLU	CD-OE2	5.93	1.32	1.25
1	B	434	GLU	CD-OE2	-5.60	1.19	1.25
1	A	26	GLU	CD-OE2	5.53	1.31	1.25
1	A	231	GLY	C-O	-5.33	1.15	1.23
1	A	371	SER	CA-CB	5.22	1.60	1.52

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	273	ARG	NE-CZ-NH1	10.39	125.50	120.30
1	A	263	ARG	NE-CZ-NH2	-10.00	115.30	120.30
1	A	273	ARG	NE-CZ-NH2	-9.70	115.45	120.30
1	B	273	ARG	NE-CZ-NH1	9.52	125.06	120.30
1	B	263	ARG	NE-CZ-NH2	-8.83	115.88	120.30

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	4147	0	4010	54	0
1	B	4134	0	3985	44	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	5	0	0	3	0
3	B	5	0	0	1	0
4	A	7	0	10	0	0
4	B	7	0	10	6	0
5	A	502	0	0	28	2
5	B	471	0	0	20	2
All	All	9280	0	8015	100	2

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 6.

The worst 5 of 100 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:15:LEU:HA	5:A:2011:HOH:O	1.41	1.19
1:A:515:GLU:HA	5:A:2492:HOH:O	1.45	1.16
1:A:282:VAL:HA	5:A:2312:HOH:O	1.42	1.16
1:A:55:ARG:HD3	5:A:2062:HOH:O	1.46	1.13
1:A:402:HIS:CD2	5:A:2054:HOH:O	2.04	1.09

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:2256:HOH:O	5:B:2289:HOH:O[4_455]	1.34	0.86
5:A:2368:HOH:O	5:B:2289:HOH:O[4_455]	2.10	0.10

## 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	522/536 (97%)	508 (97%)	13 (2%)	1 (0%)	47	33
1	B	523/536 (98%)	504 (96%)	18 (3%)	1 (0%)	47	33
All	All	1045/1072 (98%)	1012 (97%)	31 (3%)	2 (0%)	47	33

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	334	ASP
1	B	334	ASP

### 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	419/429 (98%)	415 (99%)	4 (1%)	76	69
1	B	417/429 (97%)	412 (99%)	5 (1%)	71	62
All	All	836/858 (97%)	827 (99%)	9 (1%)	73	65

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	156	ILE
1	B	426	ARG
1	B	318	ASN
1	A	359	TYR
1	B	176	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	309	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	402	HIS
1	B	309	ASN
1	B	402	HIS

### 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	DDZ	B	51	1,2	4,6,7	1.20	0	3,7,9	3.24	2 (66%)
1	DDZ	A	51	1,2	4,6,7	1.28	0	3,7,9	3.05	1 (33%)

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	DDZ	B	51	1,2	-	0/2/6/8	-
1	DDZ	A	51	1,2	-	0/2/6/8	-

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	51	DDZ	OG2-CB-OG1	5.08	120.21	111.27
1	A	51	DDZ	OG2-CB-OG1	4.86	119.83	111.27
1	B	51	DDZ	O-C-CA	-2.19	119.05	124.78

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	51	DDZ	1	0
1	A	51	DDZ	3	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 for Mogul analysis.

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$
4	PEG	B	1529	-	6,6,6	0.63	0	5,5,5	0.65	0
4	PEG	A	1531	-	6,6,6	0.41	0	5,5,5	0.52	0
3	SO4	A	1529	-	4,4,4	0.81	0	6,6,6	1.23	1 (16%)
3	SO4	B	1530	-	4,4,4	0.50	0	6,6,6	1.13	1 (16%)

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	PEG	A	1531	-	-	3/4/4/4	-
4	PEG	B	1529	-	-	3/4/4/4	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1529	SO4	O4-S-O3	-2.46	98.55	109.06
3	B	1530	SO4	O2-S-O1	-2.05	94.27	109.43

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1531	PEG	O1-C1-C2-O2
4	B	1529	PEG	O1-C1-C2-O2
4	A	1531	PEG	C1-C2-O2-C3
4	B	1529	PEG	C4-C3-O2-C2
4	A	1531	PEG	C4-C3-O2-C2

There are no ring outliers.

3 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	1529	PEG	6	0
3	A	1529	SO4	3	0
3	B	1530	SO4	1	0

## 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	524/536 (97%)	0.49	22 (4%) 36 34	4, 12, 30, 53	3 (0%)
1	B	524/536 (97%)	0.53	19 (3%) 42 40	4, 12, 29, 56	3 (0%)
All	All	1048/1072 (97%)	0.51	41 (3%) 39 38	4, 12, 30, 56	6 (0%)

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	328	PHE	8.3
1	B	331	PHE	7.0
1	B	332	GLY	5.9
1	A	331	PHE	5.1
1	A	328	PHE	4.6

### 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, 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
1	DDZ	B	51	7/8	0.94	0.14	8,9,9,9	0
1	DDZ	A	51	7/8	0.96	0.12	8,9,9,10	0

### 6.3 Carbohydrates [i](#)

There are no carbohydrates 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
4	PEG	A	1531	7/7	0.77	0.14	26,26,29,31	0
3	SO4	A	1529	5/5	0.83	0.24	32,37,41,47	0
4	PEG	B	1529	7/7	0.87	0.20	24,24,28,31	0
3	SO4	B	1530	5/5	0.91	0.21	32,33,34,37	0
2	CA	B	1528	1/1	0.99	0.06	16,16,16,16	0
2	CA	A	1528	1/1	0.99	0.07	16,16,16,16	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.