



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 19, 2020 – 05:42 pm BST

PDB ID : 5DY6  
Title : Enhanced superfolder GFP with DBCO at 148  
Authors : Jones, D.D.; Rizkallah, P.J.; Worthy, H.L.  
Deposited on : 2015-09-24  
Resolution : 2.66 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

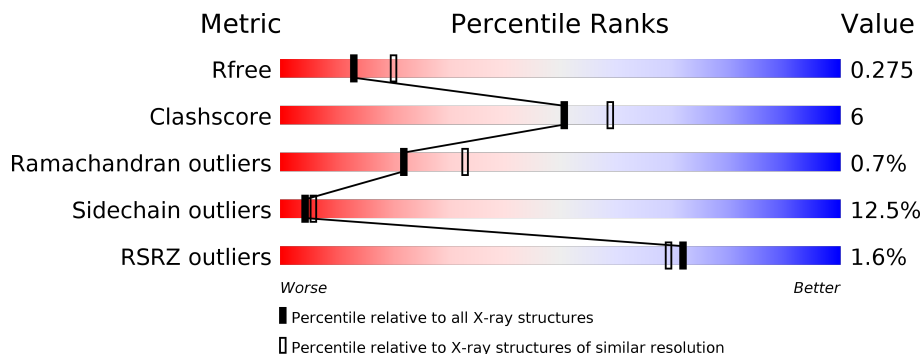
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.66 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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	226	 76% 20% .
1	B	226	 81% 16% .

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
2	SO4	A	302	-	-	-	X

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3656 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.

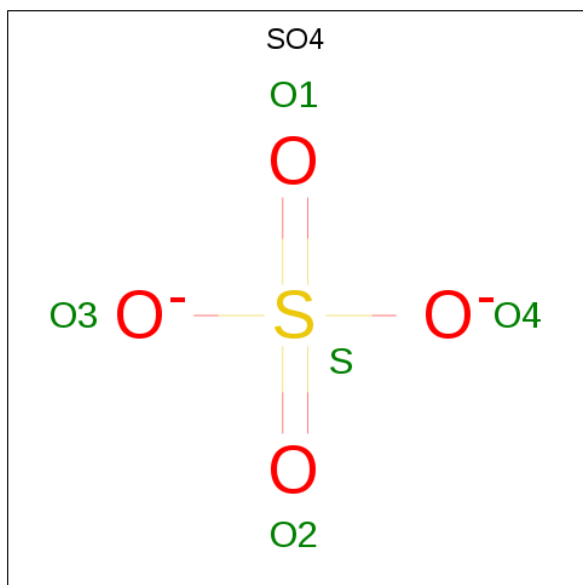
- Molecule 1 is a protein called Green fluorescent protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	226	1830	1164	314	347	5	0	0	0
1	B	225	1821	1159	313	344	5	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	ARG	SER	conflict	UNP A0A059PIQ0
A	?	CRO	THR	chromophore	UNP A0A059PIQ0
A	?	CRO	TYR	chromophore	UNP A0A059PIQ0
A	66	CRO	GLY	chromophore	UNP A0A059PIQ0
A	72	SER	ALA	conflict	UNP A0A059PIQ0
A	80	ARG	GLN	conflict	UNP A0A059PIQ0
A	148	66C	HIS	engineered mutation	UNP A0A059PIQ0
A	206	VAL	ALA	conflict	UNP A0A059PIQ0
B	30	ARG	SER	conflict	UNP A0A059PIQ0
B	?	-	THR	chromophore	UNP A0A059PIQ0
B	?	-	TYR	chromophore	UNP A0A059PIQ0
B	66	CRO	GLY	chromophore	UNP A0A059PIQ0
B	72	SER	ALA	conflict	UNP A0A059PIQ0
B	80	ARG	GLN	conflict	UNP A0A059PIQ0
B	148	66C	HIS	engineered mutation	UNP A0A059PIQ0
B	206	VAL	ALA	conflict	UNP A0A059PIQ0

- Molecule 2 is SULFATE ION (three-letter code: SO<sub>4</sub>) (formula: O<sub>4</sub>S).

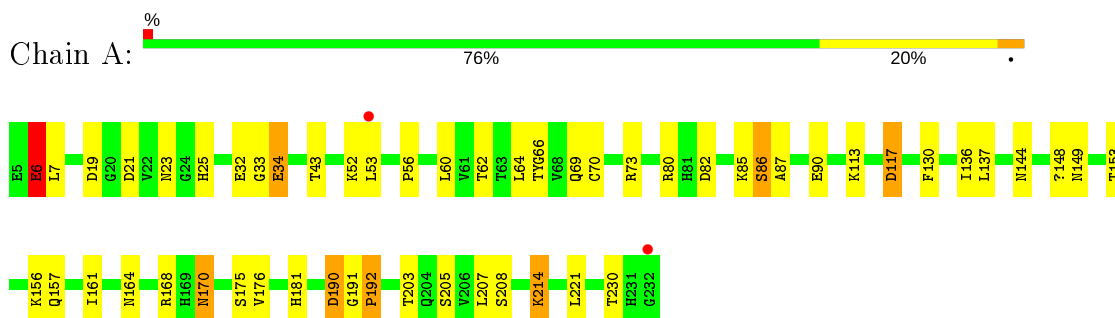


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O 1 1	0	0
2	A	1	Total O S 4 3 1	0	0

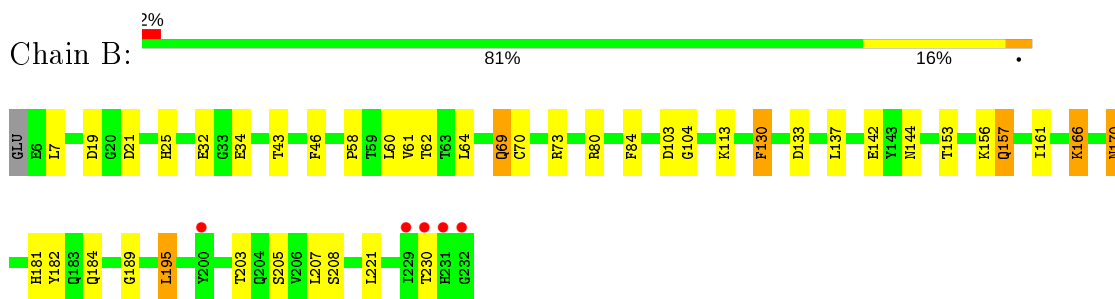
### 3 Residue-property plots [i](#)

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.98Å 89.38Å 122.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	72.18 – 2.66 72.18 – 2.66	Depositor EDS
% Data completeness (in resolution range)	99.4 (72.18-2.66) 99.4 (72.18-2.66)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.74 (at 2.65Å)	Xtrriage
Refinement program	REFMAC 5.8.0124	Depositor
R, $R_{free}$	0.194 , 0.290 0.199 , 0.275	Depositor DCC
$R_{free}$ test set	705 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	90.7	Xtrriage
Anisotropy	0.450	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 83.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3656	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	112.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 32.87 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.7689e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 66C, SO4, CRO

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	0.77	0/1810	0.94	1/2445 (0.0%)
1	B	0.80	0/1801	0.96	1/2433 (0.0%)
All	All	0.79	0/3611	0.95	2/4878 (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	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	195	LEU	CA-CB-CG	6.03	129.16	115.30
1	A	117	ASP	CB-CG-OD1	5.30	123.07	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	6	GLU	Peptide
1	B	189	GLY	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	1830	0	1749	22	0
1	B	1821	0	1743	19	0
2	A	5	0	0	0	0
All	All	3656	0	3492	41	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 6.

The worst 5 of 41 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:104:GLY:HA3	1:B:130:PHE:CD2	2.32	0.64
1:A:23:ASN:HD21	1:A:130:PHE:H	1.51	0.58
1:B:157:GLN:HE21	1:B:157:GLN:N	2.02	0.58
1:B:104:GLY:N	1:B:130:PHE:HB3	2.19	0.57
1:A:62:THR:HG21	1:A:181:HIS:NE2	2.22	0.55

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	220/226 (97%)	209 (95%)	9 (4%)	2 (1%)	17 26
1	B	219/226 (97%)	205 (94%)	13 (6%)	1 (0%)	29 43
All	All	439/452 (97%)	414 (94%)	22 (5%)	3 (1%)	22 33

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	ARG
1	B	80	ARG
1	A	192	PRO

### 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	196/196 (100%)	171 (87%)	25 (13%)	<b>4</b> <b>5</b>
1	B	195/196 (100%)	171 (88%)	24 (12%)	<b>4</b> <b>6</b>
All	All	391/392 (100%)	342 (88%)	49 (12%)	<b>4</b> <b>6</b>

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	208	SER
1	B	25	HIS
1	B	203	THR
1	A	230	THR
1	B	32	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	149	ASN
1	A	177	GLN
1	B	157	GLN
1	A	144	ASN
1	A	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](#)

4 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	CRO	B	66	1	23,23,24	3.74	5 (21%)	30,32,34	3.83	8 (26%)
1	CRO	A	66	1	23,23,24	4.59	6 (26%)	30,32,34	4.95	10 (33%)
1	66C	A	148	1	34,39,40	1.15	2 (5%)	37,55,57	2.14	10 (27%)
1	66C	B	148	1	34,39,40	1.28	3 (8%)	37,55,57	1.99	12 (32%)

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	CRO	B	66	1	-	1/12/31/32	0/2/2/2
1	CRO	A	66	1	-	3/12/31/32	0/2/2/2
1	66C	A	148	1	-	3/14/33/35	0/4/5/5
1	66C	B	148	1	-	4/14/33/35	0/4/5/5

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	CRO	CB2-CA2	19.48	1.51	1.35
1	B	66	CRO	CB2-CA2	15.79	1.48	1.35
1	A	66	CRO	CA1-C1	-5.79	1.43	1.51
1	A	66	CRO	C2-N3	-5.42	1.27	1.39
1	A	66	CRO	CA2-C2	-4.82	1.43	1.48

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	66	CRO	CA2-C2-N3	19.73	112.70	103.37

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	66	CRO	CA2-C2-N3	13.40	109.70	103.37
1	B	66	CRO	O2-C2-CA2	-12.24	124.09	130.96
1	A	66	CRO	O2-C2-CA2	-10.01	125.34	130.96
1	A	66	CRO	C2-CA2-N2	-8.61	102.90	108.93

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	66	CRO	N3-C1-CA1-CB1
1	A	148	66C	OBE-CBD-NAU-CAT
1	A	148	66C	N-CA-CB-CG
1	A	148	66C	C-CA-CB-CG
1	B	148	66C	OBE-CBD-NAU-CAT

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	66	CRO	2	0
1	A	148	66C	1	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 1 is modelled with single atom - leaving 1 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
2	SO4	A	302	2	1,3,4	1.03	0	0,3,6	0.00	-

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	224/226 (99%)	-0.01	2 (0%) 84 83	78, 108, 152, 196	0
1	B	223/226 (98%)	0.03	5 (2%) 62 57	75, 111, 153, 198	0
All	All	447/452 (98%)	0.01	7 (1%) 72 69	75, 110, 152, 198	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	232	GLY	11.5
1	A	232	GLY	6.6
1	B	231	HIS	4.1
1	B	230	THR	3.9
1	B	229	ILE	3.3

### 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	66C	B	148	35/36	0.92	0.20	87,129,162,173	0
1	66C	A	148	35/36	0.95	0.18	85,125,148,154	0
1	CRO	B	66	22/23	0.97	0.20	73,88,96,109	0
1	CRO	A	66	22/23	0.97	0.19	72,84,97,111	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(Å <sup>2</sup> )	Q<0.9
2	SO4	A	302	4/5	0.61	0.43	178,189,193,198	0
2	SO4	A	301	1/5	0.91	0.61	194,194,194,194	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.