Full wwPDB X-ray Structure Validation Report ⓘ

Apr 16, 2025 – 04:08 pm BST

PDB ID	:	9EP5 / pdb_00009ep5
Title	:	GH191 family Alpha-Galactosaminidase from Environmental sample (99.2% identity to Myxococcus fulvus enzyme)
Authors	:	Roth, C.; Moroz, O.V.; Miranda, S.A.D.; Jahn, L.; Blagova, E.; Lebedev, A.A.; Segura, D.R.; Stringer, M.A.; Friis, E.P.; Davies, G.J.; Wilson, K.S.
Deposited on	:	2024-03-17
Resolution	:	1.64 Å(reported)

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

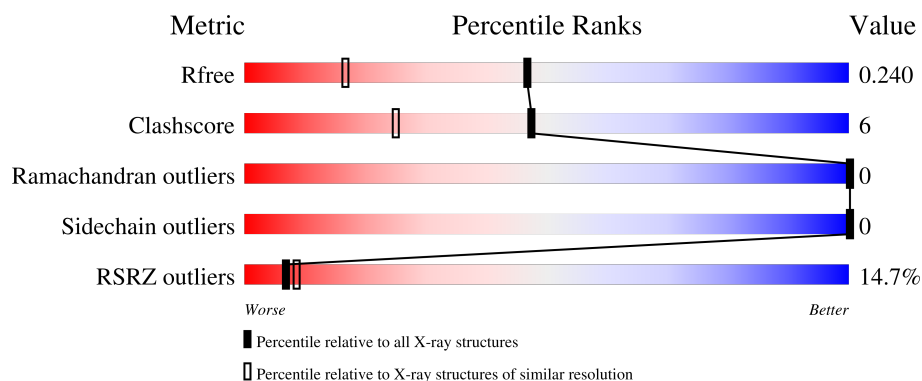|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

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 1.64 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	Percentile Ranks	Value
Rfree		0.240
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		14.7%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	164625	1015 (1.64-1.64)
Clashscore	180529	1093 (1.64-1.64)
Ramachandran outliers	177936	1077 (1.64-1.64)
Sidechain outliers	177891	1077 (1.64-1.64)
RSRZ outliers	164620	1015 (1.64-1.64)

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 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 <=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	300	% · 87% · 9%
1	D	300	26% · 84% 7% 9%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 8705 atoms, of which 3973 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 Glycoside-hydrolase family GH114 TIM-barrel domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	272	Total	C	H	N	O	S	0	4	0
			4145	1360	2002	364	412	7			
1	D	272	Total	C	H	N	O	S	0	2	0
			4085	1347	1971	361	399	7			

- Molecule 2 is SULFATE ION (CCD ID: SO4) (formula: O_4S).
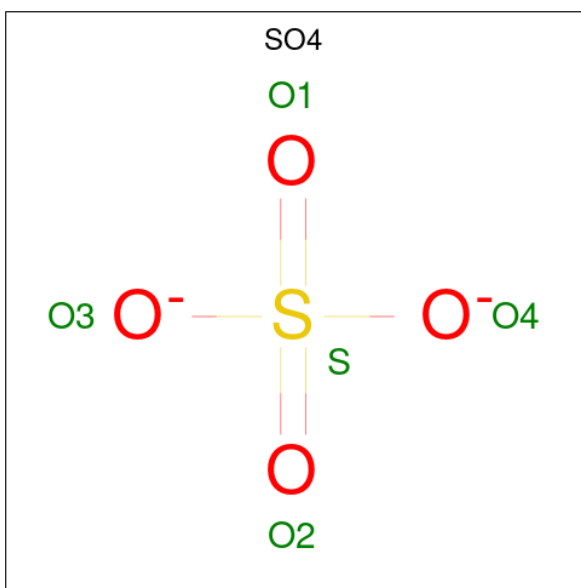
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	338	Total O 339 339	0	1
3	D	126	Total O 126 126	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: Glycoside-hydrolase family GH114 TIM-barrel domain-containing protein
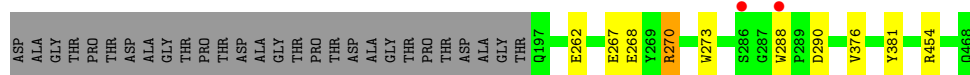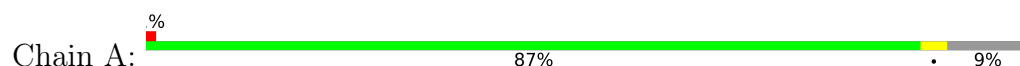
• Molecule 1: Glycoside-hydrolase family GH114 TIM-barrel domain-containing protein
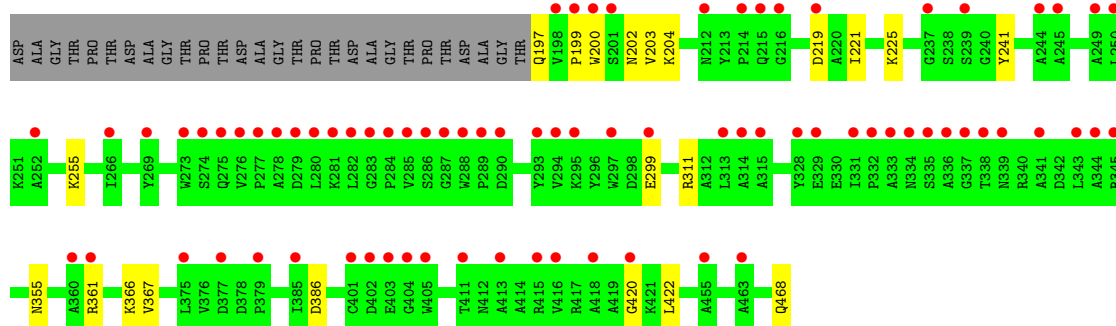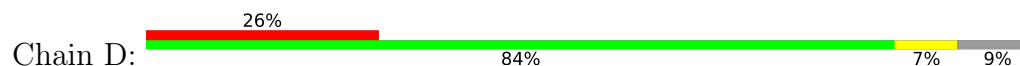
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	41.84Å 185.32Å 128.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.33 – 1.64 46.33 – 1.64	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.33-1.64) 100.0 (46.33-1.64)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.30 (at 1.64Å)	Xtriage
Refinement program	REFMAC 5.8.0425	Depositor
R, R_{free}	0.197 , 0.240 0.196 , 0.240	Depositor DCC
R_{free} test set	3200 reflections (5.18%)	wwPDB-VP
Wilson B-factor (Å^2)	20.7	Xtriage
Anisotropy	0.617	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 53.9	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8705	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.26% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.55	0/2216	0.83	1/3026 (0.0%)
1	D	0.36	0/2180	0.69	1/2979 (0.0%)
All	All	0.46	0/4396	0.76	2/6005 (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	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	361	ARG	NE-CZ-NH2	-5.25	117.68	120.30
1	A	270	ARG	NE-CZ-NH2	-5.18	117.71	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	270	ARG	Sidechain
1	A	454	ARG	Sidechain

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	2143	2002	2020	4	0
1	D	2114	1971	1997	42	0
2	A	5	0	0	0	0
2	D	5	0	0	0	0
3	A	339	0	0	0	0
3	D	126	0	0	40	0
All	All	4732	3973	4017	46	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.

All (46) 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:D:202:ASN:HA	3:D:609:HOH:O	1.15	1.26
1:D:386:ASP:HB2	3:D:620:HOH:O	1.33	1.24
1:D:420:GLY:O	3:D:608:HOH:O	1.58	1.20
1:D:204:LYS:CE	3:D:711:HOH:O	1.92	1.16
1:D:204:LYS:CG	3:D:711:HOH:O	1.97	1.12
1:D:202:ASN:CA	3:D:609:HOH:O	1.75	1.12
1:D:204:LYS:HG3	3:D:711:HOH:O	1.55	1.06
1:D:203:VAL:HG23	3:D:607:HOH:O	1.57	1.05
1:D:204:LYS:NZ	3:D:611:HOH:O	1.90	1.05
1:D:241:TYR:OH	3:D:610:HOH:O	1.75	1.04
1:D:241:TYR:CZ	3:D:610:HOH:O	2.10	1.03
1:D:422:LEU:HD22	3:D:626:HOH:O	1.56	1.03
1:D:202:ASN:OD1	3:D:609:HOH:O	1.74	1.02
1:D:386:ASP:CB	3:D:620:HOH:O	1.92	1.00
1:D:366:LYS:HD2	3:D:620:HOH:O	1.66	0.96
1:D:420:GLY:CA	3:D:608:HOH:O	2.15	0.95
1:D:366:LYS:HB3	3:D:620:HOH:O	1.69	0.91
1:D:199:PRO:HA	3:D:612:HOH:O	1.72	0.87
1:D:420:GLY:HA3	3:D:608:HOH:O	1.72	0.84
1:D:204:LYS:HE3	3:D:711:HOH:O	1.64	0.81
1:D:219:ASP:HB2	3:D:629:HOH:O	1.82	0.80

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:202:ASN:CG	3:D:609:HOH:O	2.05	0.79
1:D:366:LYS:CB	3:D:620:HOH:O	2.29	0.76
1:D:367:VAL:HA	3:D:706:HOH:O	1.87	0.74
1:D:200:TRP:N	3:D:612:HOH:O	2.17	0.74
1:D:203:VAL:CG2	3:D:607:HOH:O	2.26	0.74
1:D:221:ILE:O	1:D:255:LYS:HE2	1.88	0.73
1:D:366:LYS:CD	3:D:620:HOH:O	2.35	0.64
1:D:204:LYS:HE2	3:D:711:HOH:O	1.78	0.58
1:D:203:VAL:CB	3:D:607:HOH:O	2.50	0.57
1:A:288:TRP:CE2	1:A:290:ASP:HB2	2.41	0.55
1:D:241:TYR:CE2	3:D:610:HOH:O	2.53	0.51
1:D:367:VAL:HG22	3:D:706:HOH:O	2.09	0.50
1:D:299[B]:GLU:HG3	3:D:633:HOH:O	2.13	0.49
1:D:202:ASN:C	3:D:609:HOH:O	2.28	0.48
1:A:268:GLU:HA	1:A:273:TRP:CD2	2.49	0.47
1:D:311:ARG:NH1	3:D:622:HOH:O	2.48	0.47
1:D:386:ASP:CG	3:D:620:HOH:O	2.34	0.47
1:D:200:TRP:CD2	3:D:626:HOH:O	2.56	0.46
1:A:262:GLU:OE1	1:A:267:GLU:OE2	2.34	0.46
1:D:203:VAL:HB	3:D:607:HOH:O	2.16	0.45
1:D:197:GLN:C	3:D:602:HOH:O	2.51	0.43
1:A:376:VAL:HA	1:A:381:TYR:CD2	2.54	0.43
1:D:311:ARG:NE	3:D:610:HOH:O	2.51	0.43
1:D:225:LYS:HD2	1:D:468:GLN:C	2.40	0.42
1:D:355:ASN:ND2	3:D:624:HOH:O	2.50	0.42

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	274/300 (91%)	268 (98%)	6 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	272/300 (91%)	264 (97%)	8 (3%)	0	100	100
All	All	546/600 (91%)	532 (97%)	14 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	219/236 (93%)	219 (100%)	0	100	100
1	D	210/236 (89%)	210 (100%)	0	100	100
All	All	429/472 (91%)	429 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	209	GLN
1	D	209	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

2 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	D	501	-	4,4,4	0.39	0	6,6,6	0.08	0
2	SO4	A	501	-	4,4,4	0.35	0	6,6,6	0.08	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 ⓘ

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, 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(Å²)	Q<0.9
1	A	272/300 (90%)	-0.43	2 (0%)	84	87	10, 19, 32, 48	2 (0%)
1	D	272/300 (90%)	1.60	78 (28%)	1	1	3, 6, 11, 16	272 (100%)
All	All	544/600 (90%)	0.58	80 (14%)	7	8	3, 14, 29, 48	274 (50%)

All (80) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	201	SER	6.4
1	D	288	TRP	6.0
1	D	276	VAL	5.4
1	D	279	ASP	5.4
1	D	269	TYR	5.3
1	D	278	ALA	5.1
1	D	333	ALA	5.0
1	D	405	TRP	4.8
1	D	274	SER	4.6
1	D	275	GLN	4.6
1	D	282	LEU	4.6
1	D	343	LEU	4.6
1	D	250	LEU	4.5
1	D	199	PRO	4.5
1	D	287	GLY	4.5
1	D	198	VAL	4.4
1	D	299[A]	GLU	4.4
1	D	277	PRO	4.3
1	D	215	GLN	4.1
1	D	237	GLY	4.1
1	D	339	ASN	4.1
1	D	336	ALA	3.9
1	D	337	GLY	3.9
1	D	290	ASP	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	284	PRO	3.5
1	D	314	ALA	3.4
1	D	285	VAL	3.4
1	D	335	SER	3.4
1	D	418	ALA	3.4
1	D	283	GLY	3.4
1	D	329	GLU	3.3
1	D	281	LYS	3.3
1	D	294	VAL	3.2
1	D	280	LEU	3.2
1	D	332	PRO	3.2
1	D	273	TRP	3.2
1	D	420	GLY	3.1
1	D	266	ILE	3.0
1	D	338	THR	3.0
1	D	286	SER	2.9
1	D	245	ALA	2.9
1	D	455	ALA	2.9
1	D	404	GLY	2.9
1	A	288	TRP	2.8
1	D	297	TRP	2.8
1	D	200	TRP	2.8
1	D	331	ILE	2.8
1	D	244	ALA	2.8
1	D	345	ARG	2.8
1	D	385	ILE	2.7
1	D	216	GLY	2.7
1	D	377	ASP	2.6
1	D	416	VAL	2.6
1	D	463	ALA	2.5
1	D	402	ASP	2.5
1	D	413	ALA	2.5
1	D	361	ARG	2.5
1	D	379	PRO	2.5
1	D	289	PRO	2.4
1	D	334	ASN	2.4
1	D	375	LEU	2.4
1	D	252	ALA	2.3
1	D	344	ALA	2.3
1	D	403	GLU	2.3
1	A	286	SER	2.3
1	D	313	LEU	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	295	LYS	2.2
1	D	212	ASN	2.2
1	D	249	ALA	2.2
1	D	411	THR	2.2
1	D	214	PRO	2.2
1	D	401	CYS	2.2
1	D	360	ALA	2.1
1	D	415	ARG	2.1
1	D	328	TYR	2.1
1	D	315	ALA	2.1
1	D	341	ALA	2.1
1	D	293	TYR	2.1
1	D	219	ASP	2.0
1	D	239	SER	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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	D	501	5/5	0.91	0.10	32,33,36,38	5
2	SO4	A	501	5/5	0.98	0.06	27,29,31,33	0

6.5 Other polymers ⓘ

There are no such residues in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK