



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 12, 2023 – 03:48 pm GMT

PDB ID : 8BWK  
Title : Metagenomic derived PL6 alginate lyase  
Authors : Angen, E.F.; Wilkens, C.  
Deposited on : 2022-12-06  
Resolution : 2.29 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 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 : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

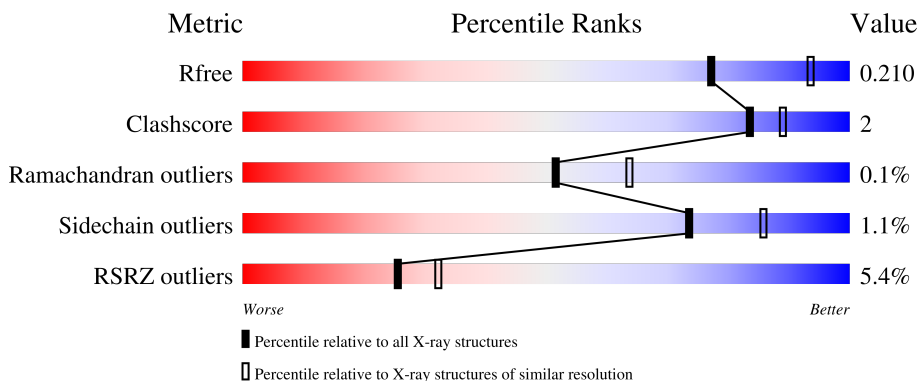
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.29 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	728	 2% 94% 5%
1	B	728	 1% 94% 6%
1	C	728	 3% 94% 6%
1	D	728	 15% 90% 7%

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	CL	B	806	-	-	X	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 45499 atoms, of which 22196 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 Alginate lyase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	725	11238	3554	5574	1005	1098	7	0	1	0
1	B	725	11251	3558	5576	1009	1101	7	14	2	0
1	C	727	11255	3558	5582	1008	1101	6	0	0	0
1	D	708	10984	3480	5444	983	1071	6	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP A0A0P7DTY9
A	2	GLY	-	expression tag	UNP A0A0P7DTY9
A	340	ASP	ASN	conflict	UNP A0A0P7DTY9
A	355	SER	PRO	conflict	UNP A0A0P7DTY9
A	387	GLN	LYS	conflict	UNP A0A0P7DTY9
A	554	ASP	ASN	conflict	UNP A0A0P7DTY9
A	641	LYS	ASN	conflict	UNP A0A0P7DTY9
A	655	LYS	GLN	conflict	UNP A0A0P7DTY9
A	706	TYR	HIS	conflict	UNP A0A0P7DTY9
A	708	GLU	LYS	conflict	UNP A0A0P7DTY9
B	1	MET	-	initiating methionine	UNP A0A0P7DTY9
B	2	GLY	-	expression tag	UNP A0A0P7DTY9
B	340	ASP	ASN	conflict	UNP A0A0P7DTY9
B	355	SER	PRO	conflict	UNP A0A0P7DTY9
B	387	GLN	LYS	conflict	UNP A0A0P7DTY9
B	554	ASP	ASN	conflict	UNP A0A0P7DTY9
B	641	LYS	ASN	conflict	UNP A0A0P7DTY9
B	655	LYS	GLN	conflict	UNP A0A0P7DTY9
B	706	TYR	HIS	conflict	UNP A0A0P7DTY9
B	708	GLU	LYS	conflict	UNP A0A0P7DTY9
C	1	MET	-	initiating methionine	UNP A0A0P7DTY9

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	2	GLY	-	expression tag	UNP A0A0P7DTY9
C	340	ASP	ASN	conflict	UNP A0A0P7DTY9
C	355	SER	PRO	conflict	UNP A0A0P7DTY9
C	387	GLN	LYS	conflict	UNP A0A0P7DTY9
C	554	ASP	ASN	conflict	UNP A0A0P7DTY9
C	641	LYS	ASN	conflict	UNP A0A0P7DTY9
C	655	LYS	GLN	conflict	UNP A0A0P7DTY9
C	706	TYR	HIS	conflict	UNP A0A0P7DTY9
C	708	GLU	LYS	conflict	UNP A0A0P7DTY9
D	1	MET	-	initiating methionine	UNP A0A0P7DTY9
D	2	GLY	-	expression tag	UNP A0A0P7DTY9
D	340	ASP	ASN	conflict	UNP A0A0P7DTY9
D	355	SER	PRO	conflict	UNP A0A0P7DTY9
D	387	GLN	LYS	conflict	UNP A0A0P7DTY9
D	554	ASP	ASN	conflict	UNP A0A0P7DTY9
D	641	LYS	ASN	conflict	UNP A0A0P7DTY9
D	655	LYS	GLN	conflict	UNP A0A0P7DTY9
D	706	TYR	HIS	conflict	UNP A0A0P7DTY9
D	708	GLU	LYS	conflict	UNP A0A0P7DTY9

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	5	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Cl	0	0
			1	1		
3	B	3	Total	Cl	0	0
			3	3		
3	C	2	Total	Cl	0	0
			2	2		

- 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	B	1	Total	C	H	O	0	0
			17	4	10	3		
4	D	1	Total	C	H	O	0	0
			17	4	10	3		

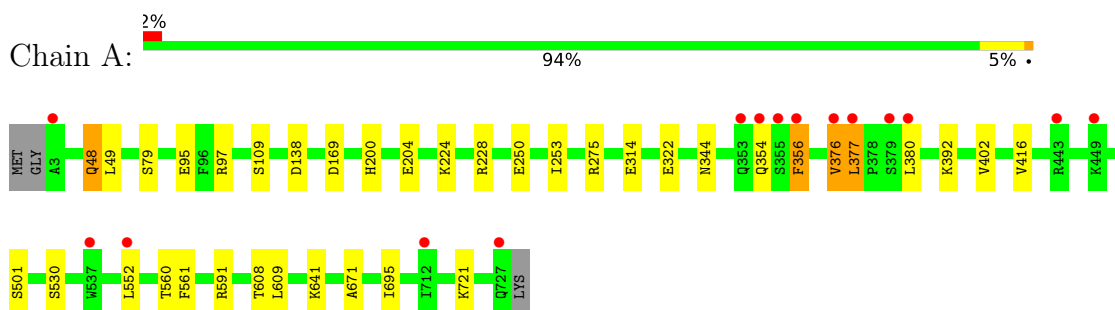
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	242	Total	O	0	0
			242	242		
5	B	205	Total	O	0	0
			205	205		
5	C	125	Total	O	0	0
			125	125		
5	D	94	Total	O	0	0
			94	94		

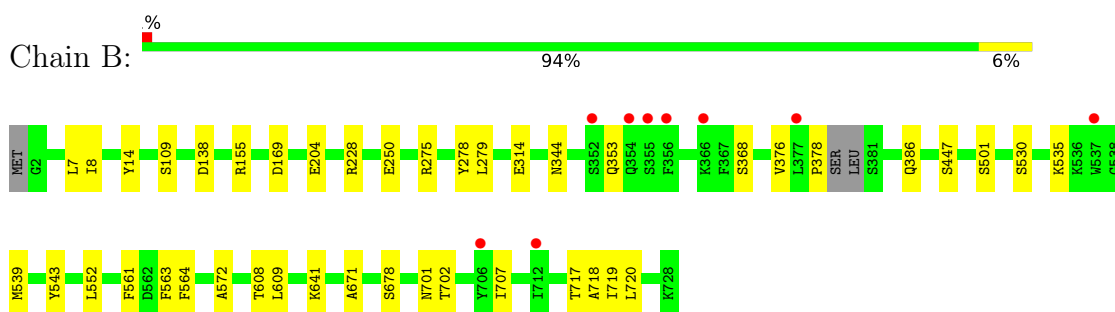
### 3 Residue-property plots [i](#)

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.

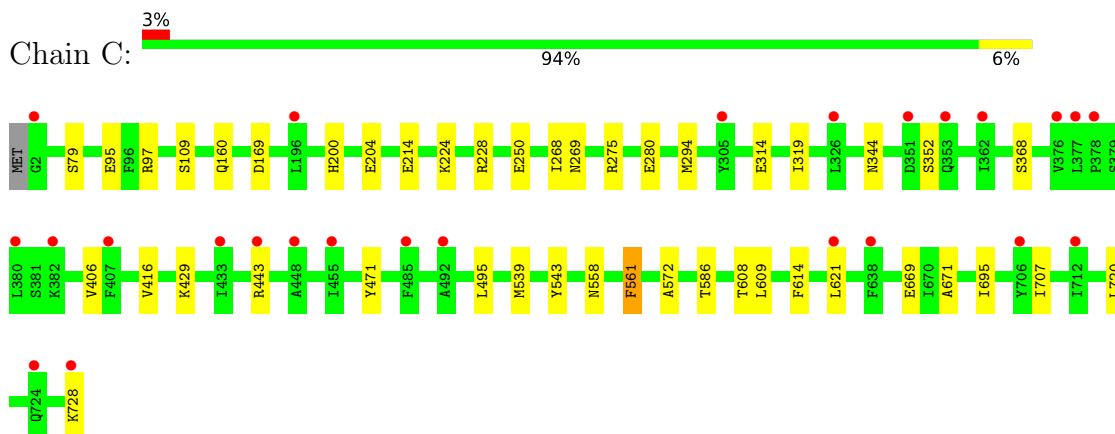
- Molecule 1: Alginate lyase



- Molecule 1: Alginate lyase

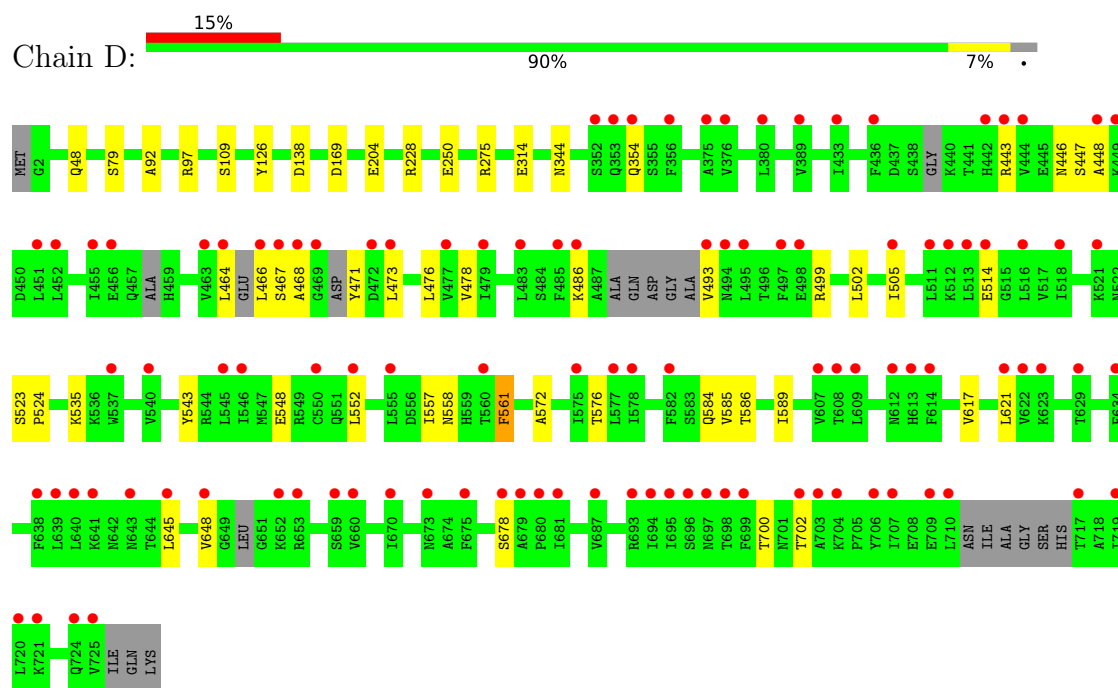


- Molecule 1: Alginate lyase



- Molecule 1: Alginate lyase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.75Å 228.85Å 113.73Å 90.00° 101.86° 90.00°	Depositor
Resolution (Å)	63.86 – 2.29 111.30 – 2.29	Depositor EDS
% Data completeness (in resolution range)	97.9 (63.86-2.29) 98.0 (111.30-2.29)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.35 (at 2.29Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.180 , 0.210 0.180 , 0.210	Depositor DCC
$R_{free}$ test set	7888 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.9	Xtrriage
Anisotropy	0.467	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 42.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	45499	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.18% 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: CL, SO4, PEG

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	0/5771	0.51	0/7812
1	B	0.25	0/5779	0.51	0/7820
1	C	0.25	0/5777	0.50	0/7818
1	D	0.25	0/5636	0.50	0/7618
All	All	0.25	0/22963	0.50	0/31068

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 [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	5664	5574	5574	21	0
1	B	5675	5576	5578	26	0
1	C	5673	5582	5581	25	0
1	D	5540	5444	5451	36	0
2	A	15	0	0	0	0
2	B	15	0	0	0	0
2	C	15	0	0	0	0
2	D	20	0	0	0	0
3	A	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	3	0	0	2	0
3	C	2	0	0	0	0
4	B	7	10	10	0	0
4	D	7	10	10	2	0
5	A	242	0	0	2	0
5	B	205	0	0	4	0
5	C	125	0	0	2	0
5	D	94	0	0	2	0
All	All	23303	22196	22204	109	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 2.

The worst 5 of 109 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:486:LYS:HG2	1:D:514:GLU:HB2	1.53	0.89
1:D:535:LYS:N	5:D:901:HOH:O	2.11	0.84
3:B:806:CL:CL	5:B:1049:HOH:O	2.40	0.77
1:D:621:LEU:HD11	1:D:645:LEU:HD21	1.67	0.76
1:D:466:LEU:N	1:D:471:TYR:HH	1.87	0.72

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	724/728 (100%)	698 (96%)	24 (3%)	2 (0%)	41	50
1	B	725/728 (100%)	699 (96%)	25 (3%)	1 (0%)	51	64
1	C	725/728 (100%)	694 (96%)	30 (4%)	1 (0%)	51	64

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	692/728 (95%)	658 (95%)	34 (5%)	0	100	100
All	All	2866/2912 (98%)	2749 (96%)	113 (4%)	4 (0%)	51	64

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	354	GLN
1	B	378	PRO
1	C	352	SER
1	A	376	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	619/620 (100%)	613 (99%)	6 (1%)	76	87
1	B	619/620 (100%)	615 (99%)	4 (1%)	86	94
1	C	619/620 (100%)	612 (99%)	7 (1%)	73	86
1	D	607/620 (98%)	596 (98%)	11 (2%)	59	75
All	All	2464/2480 (99%)	2436 (99%)	28 (1%)	73	86

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	539	MET
1	D	700	THR
1	D	97	ARG
1	D	557	ILE
1	C	669	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	306	HIS

### 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](#)

Of 21 ligands modelled in this entry, 6 are monoatomic - leaving 15 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	C	803	-	4,4,4	0.15	0	6,6,6	0.04	0
2	SO4	D	804	-	4,4,4	0.14	0	6,6,6	0.06	0
4	PEG	D	802	-	6,6,6	0.12	0	5,5,5	0.09	0
2	SO4	A	802	-	4,4,4	0.15	0	6,6,6	0.05	0
2	SO4	D	803	-	4,4,4	0.14	0	6,6,6	0.05	0
4	PEG	B	803	-	6,6,6	0.12	0	5,5,5	0.09	0
2	SO4	C	802	-	4,4,4	0.13	0	6,6,6	0.07	0
2	SO4	D	801	-	4,4,4	0.13	0	6,6,6	0.06	0
2	SO4	B	802	-	4,4,4	0.14	0	6,6,6	0.06	0
2	SO4	B	804	-	4,4,4	0.14	0	6,6,6	0.07	0
2	SO4	A	801	-	4,4,4	0.14	0	6,6,6	0.05	0
2	SO4	C	801	-	4,4,4	0.13	0	6,6,6	0.06	0
2	SO4	D	805	-	4,4,4	0.15	0	6,6,6	0.05	0
2	SO4	A	803	-	4,4,4	0.14	0	6,6,6	0.06	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	SO4	B	801	-	4,4,4	0.14	0	6,6,6	0.05	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	PEG	D	802	-	-	3/4/4/4	-
4	PEG	B	803	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	802	PEG	O1-C1-C2-O2
4	D	802	PEG	C1-C2-O2-C3
4	D	802	PEG	O2-C3-C4-O4
4	B	803	PEG	O2-C3-C4-O4
4	B	803	PEG	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	802	PEG	2	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	725/728 (99%)	0.40	15 (2%) 63 70	28, 39, 64, 133	0
1	B	725/728 (99%)	0.42	9 (1%) 79 83	30, 42, 71, 161	0
1	C	727/728 (99%)	0.43	25 (3%) 45 52	32, 55, 84, 129	0
1	D	708/728 (97%)	0.99	107 (15%) 2 3	34, 65, 117, 186	0
All	All	2885/2912 (99%)	0.56	156 (5%) 25 32	28, 48, 100, 186	0

The worst 5 of 156 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	467	SER	10.6
1	B	377	LEU	8.9
1	D	552	LEU	8.5
1	A	377	LEU	8.3
1	D	451	LEU	7.3

### 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](#)

LIGAND-RSR INFOmissingINFO



## 6.5 Other polymers [i](#)

There are no such residues in this entry.