Full wwPDB X-ray Structure Validation Report (i)

Jun 15, 2020 – 09:04 am BST

PDB ID	:	4ZLR
Title	:	Structure of the Brat-NHL domain bound to consensus RNA motif
Authors	:	Jakob, L.; Treiber, N.; Treiber, T.; Stotz, M.; Loedige, I.; Meister, G.
Deposited on	:	2015-05-01
Resolution	:	2.30 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

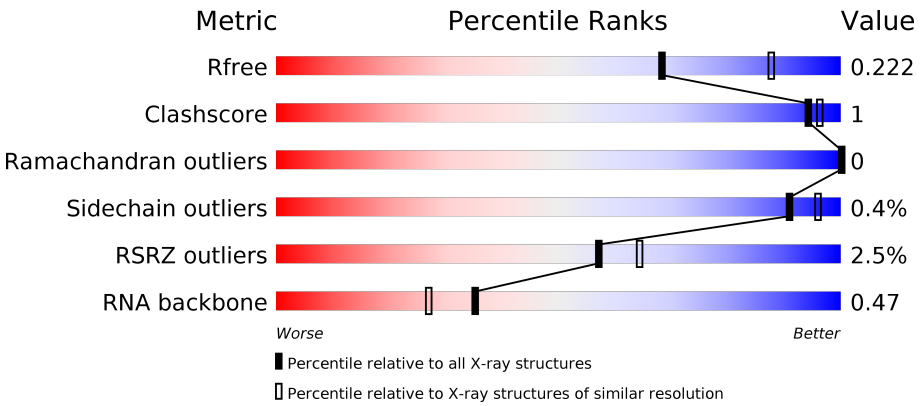
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.30 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)
RNA backbone	3102	1090 (2.70-1.90)

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 >=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	282	3% 96% • •
1	B	282	% 93% 6% •
2	C	15	13% 53% 40% 7%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 9677 atoms, of which 4544 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.

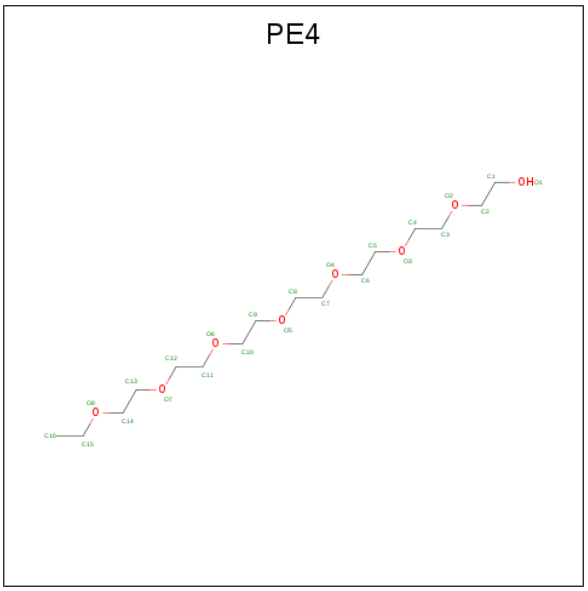
- Molecule 1 is a protein called Brain tumor protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	278	4379	1402	2167	396	404	10	0	2	0
1	B	278	4374	1399	2164	396	405	10	0	1	0

- Molecule 2 is a RNA chain called RNA (5'-R(*UP*UP*GP*UP*UP*GP*UP*UP*UP*UP*UP*UP*UP*U)-3').

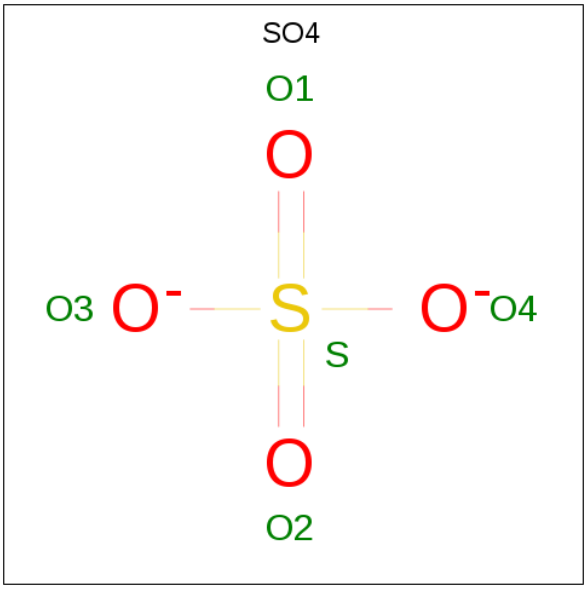
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	C	14	414	124	139	32	106	13	0	0	0

- Molecule 3 is 2-{2-[2-(2-{2-[2-(2-ETHOXY-ETHOXY)-ETHOXY]-ETHOXY}-ETHOXY)-ETHOXY]-ETHOXY}-ETHANOL (three-letter code: PE4) (formula: $C_{16}H_{34}O_8$).



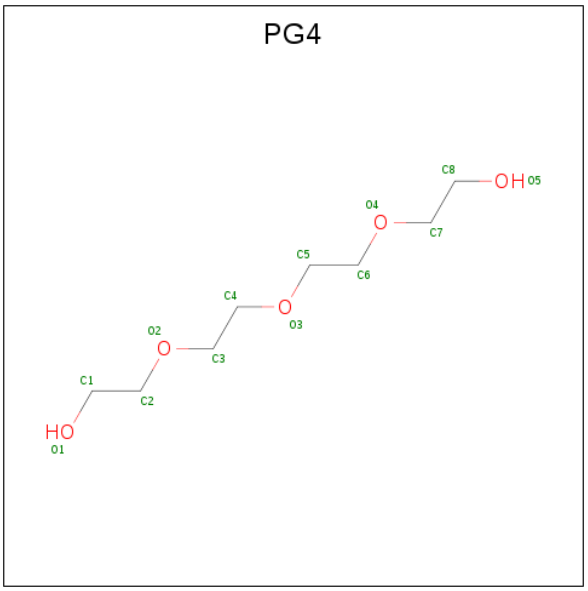
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
3	A	1	58	16	34	8	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



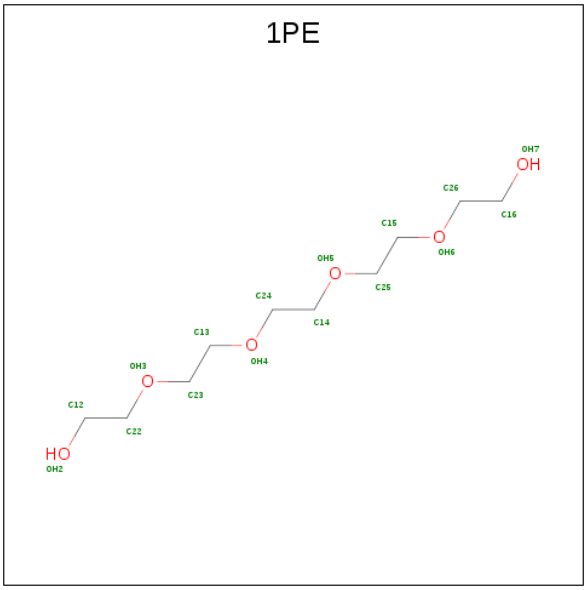
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	O	S			
4	A	1	5	4	1		0	0
			Total	O	S			
4	A	1	5	4	1		0	0
			Total	O	S			
4	A	1	5	4	1		0	0
			Total	O	S			
4	A	1	5	4	1		0	0
			Total	O	S			
4	A	1	5	4	1		0	0
			Total	O	S			
4	A	1	5	4	1		0	0
			Total	O	S			
4	B	1	5	4	1		0	0
			Total	O	S			
4	B	1	5	4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

- Molecule 5 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
5	B	1	31	8	18	5	0	0

- Molecule 6 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula: $C_{10}H_{22}O_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
6	B	1	38	10	22	6	0	0

- Molecule 7 is water.

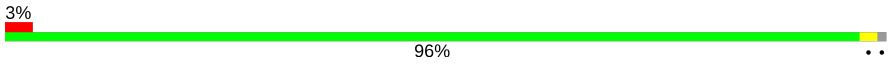
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	170	170	170	0	0
			Total	O		
7	C	11	11	11	0	0
			Total	O		
7	B	157	157	157	0	0

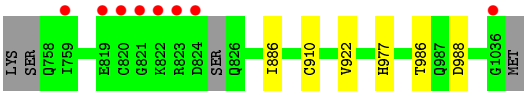
3 Residue-property plots (i)

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: Brain tumor protein

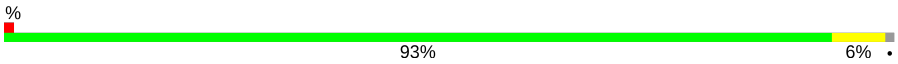
Chain A:

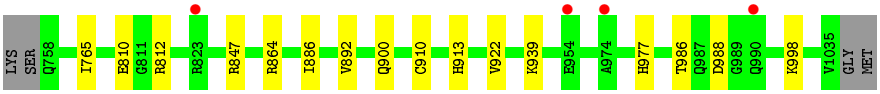




- Molecule 1: Brain tumor protein

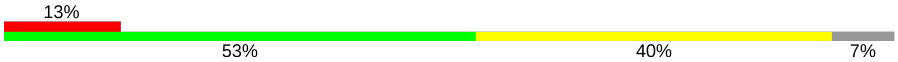
Chain B:

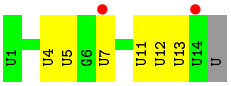




- Molecule 2: RNA (5'-R(*UP*UP*GP*UP*UP*GP*UP*UP*UP*UP*UP*UP*UP*U)-3')

Chain C:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	70.19Å 70.19Å 279.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.77 – 2.30 19.77 – 2.30	Depositor EDS
% Data completeness (in resolution range)	100.0 (19.77-2.30) 100.0 (19.77-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.23	Depositor
$< I/\sigma(I) >$ [1]	2.50 (at 2.30Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.188 , 0.219 0.194 , 0.222	Depositor DCC
R_{free} test set	1612 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	27.0	Xtriage
Anisotropy	0.091	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 44.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9677	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.08% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: PE4, PG4, SO4, 1PE

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	0.23	0/2260	0.44	0/3051
1	B	0.23	0/2256	0.44	0/3048
2	C	0.22	0/303	0.77	0/468
All	All	0.23	0/4819	0.47	0/6567

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	2212	2167	2166	3	0
1	B	2210	2164	2163	9	0
2	C	275	139	140	2	0
3	A	24	34	34	0	0
4	A	30	0	0	0	0
4	B	15	0	0	0	0
5	B	13	18	18	0	0
6	B	16	22	22	0	0
7	A	170	0	0	0	2
7	B	157	0	0	2	2

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	C	11	0	0	0	0
All	All	5133	4544	4543	12	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 1.

All (12) 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:986:THR:OG1	1:B:988:ASP:OD1	2.16	0.61
1:A:986:THR:OG1	1:A:988:ASP:OD1	2.10	0.61
1:B:864:ARG:NH1	1:B:900:GLN:O	2.43	0.51
1:B:886:ILE:HD13	1:B:922:VAL:HG11	1.95	0.49
1:A:886:ILE:HD13	1:A:922:VAL:HG11	1.96	0.47
1:B:847:ARG:NH1	7:B:1206:HOH:O	2.49	0.46
1:B:913:HIS:O	1:B:939:LYS:NZ	2.45	0.44
1:B:892:VAL:HG12	7:B:1335:HOH:O	2.17	0.44
1:B:810:GLU:OE1	1:B:812:ARG:NH1	2.52	0.41
1:B:765:ILE:HG23	1:B:998:LYS:HD2	2.02	0.41
1:A:977:HIS:HA	2:C:5:U:O2	2.21	0.41
2:C:13:U:O2	1:B:977:HIS:HA	2.21	0.41

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)
7:A:1279:HOH:O	7:B:1355:HOH:O[7_654]	2.04	0.16
7:A:1350:HOH:O	7:B:1335:HOH:O[6_544]	2.18	0.02

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	276/282 (98%)	268 (97%)	8 (3%)	0	100	100
1	B	277/282 (98%)	267 (96%)	10 (4%)	0	100	100
All	All	553/564 (98%)	535 (97%)	18 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	237/245 (97%)	236 (100%)	1 (0%)	91	96
1	B	238/245 (97%)	237 (100%)	1 (0%)	91	96
All	All	475/490 (97%)	473 (100%)	2 (0%)	91	96

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	910	CYS
1	B	910	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	13/15 (86%)	4 (30%)	0

All (4) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	4	U
2	C	7	U

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	C	11	U
2	C	12	U

There are no RNA pucker outliers to report.

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 carbohydrates 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	$\#\|Z\| > 2$
4	SO4	B	1105	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	A	1103	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	A	1104	-	4,4,4	0.13	0	6,6,6	0.05	0
5	PG4	B	1101	-	12,12,12	0.68	0	11,11,11	0.79	0
4	SO4	A	1106	-	4,4,4	0.15	0	6,6,6	0.05	0
6	1PE	B	1102	-	15,15,15	0.62	0	14,14,14	1.03	0
4	SO4	A	1102	-	4,4,4	0.15	0	6,6,6	0.06	0
4	SO4	A	1105	-	4,4,4	0.14	0	6,6,6	0.05	0
3	PE4	A	1101	-	23,23,23	1.24	0	22,22,22	0.84	0
4	SO4	B	1104	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	B	1103	-	4,4,4	0.14	0	6,6,6	0.05	0
4	SO4	A	1107	-	4,4,4	0.13	0	6,6,6	0.06	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
3	PE4	A	1101	-	-	11/21/21/21	-
6	1PE	B	1102	-	-	4/13/13/13	-
5	PG4	B	1101	-	-	3/10/10/10	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (18) torsion outliers are listed below:

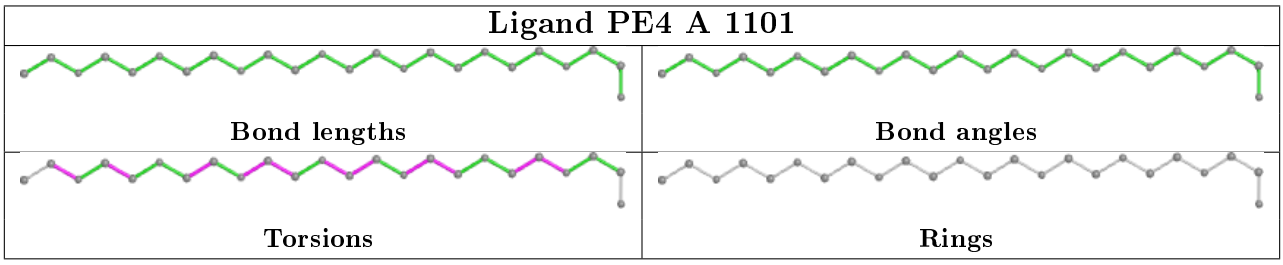
Mol	Chain	Res	Type	Atoms
5	B	1101	PG4	O4-C7-C8-O5
3	A	1101	PE4	O6-C10-C9-O5
3	A	1101	PE4	O4-C7-C8-O5
6	B	1102	1PE	OH7-C16-C26-OH6
3	A	1101	PE4	O7-C13-C14-O8
6	B	1102	1PE	OH5-C14-C24-OH4
5	B	1101	PG4	O2-C3-C4-O3
3	A	1101	PE4	O6-C11-C12-O7
3	A	1101	PE4	C16-C15-O8-C14
3	A	1101	PE4	C7-C8-O5-C9
3	A	1101	PE4	C5-C6-O4-C7
3	A	1101	PE4	C9-C10-O6-C11
6	B	1102	1PE	C16-C26-OH6-C15
3	A	1101	PE4	C4-C3-O2-C2
3	A	1101	PE4	O3-C5-C6-O4
5	B	1101	PG4	O1-C1-C2-O2
3	A	1101	PE4	O2-C3-C4-O3
6	B	1102	1PE	C24-C14-OH5-C25

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be

highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



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, 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(Å^2)	Q<0.9
1	A	278/282 (98%)	-0.15	8 (2%) 51 58	17, 25, 41, 57	0
1	B	278/282 (98%)	-0.18	4 (1%) 75 80	17, 27, 43, 54	0
2	C	14/15 (93%)	0.92	2 (14%) 2 3	27, 43, 61, 70	0
All	All	570/579 (98%)	-0.14	14 (2%) 57 64	17, 26, 43, 70	0

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	820	CYS	5.5
1	A	822	LYS	4.4
2	C	14	U	3.2
2	C	7	U	3.2
1	A	1036	GLY	2.9
1	A	823	ARG	2.6
1	B	954	GLU	2.6
1	A	819	GLU	2.6
1	A	759	ILE	2.6
1	B	823	ARG	2.4
1	A	821	GLY	2.3
1	B	974	ALA	2.2
1	A	824	ASP	2.2
1	B	990	GLN	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 carbohydrates in this entry.

6.4 Ligands (i)

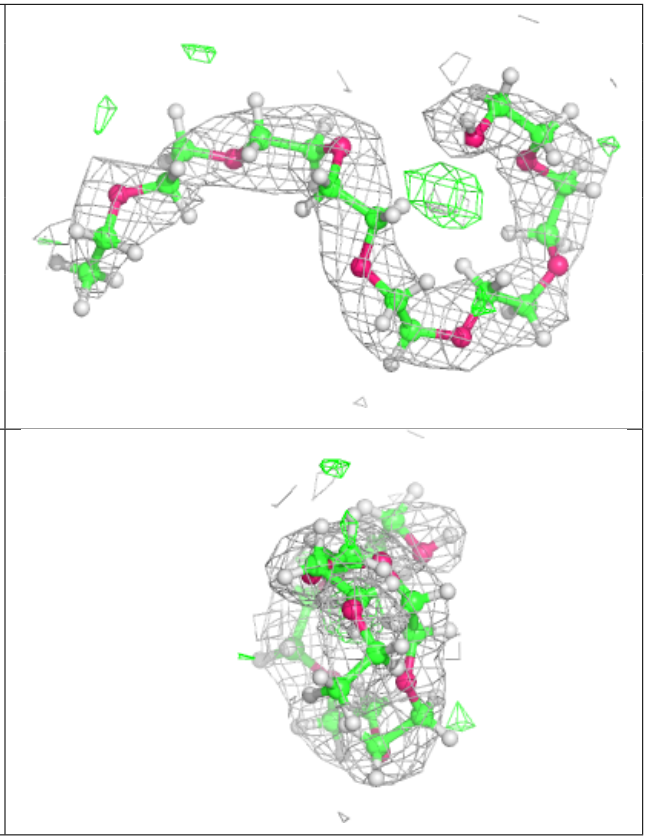
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	SO4	A	1105	5/5	0.83	0.40	62,70,86,95	0
3	PE4	A	1101	24/24	0.83	0.24	30,42,57,57	0
4	SO4	A	1107	5/5	0.89	0.29	49,59,69,78	0
4	SO4	A	1106	5/5	0.90	0.22	46,51,59,83	0
5	PG4	B	1101	13/13	0.92	0.13	26,32,36,38	0
6	1PE	B	1102	16/16	0.93	0.12	23,30,46,50	0
4	SO4	B	1105	5/5	0.94	0.28	52,58,73,82	0
4	SO4	A	1104	5/5	0.97	0.15	56,57,65,67	0
4	SO4	A	1103	5/5	0.97	0.15	37,37,41,46	0
4	SO4	B	1104	5/5	0.97	0.18	38,49,53,57	0
4	SO4	A	1102	5/5	0.97	0.11	32,34,40,41	0
4	SO4	B	1103	5/5	0.99	0.19	41,43,46,48	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around PE4 A 1101:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.