



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 26, 2017 – 10:08 PM EDT

PDB ID : 1EWQ
Title : CRYSTAL STRUCTURE TAQ MUTS COMPLEXED WITH A HET-
ERODUPLEX DNA AT 2.2 Å RESOLUTION
Authors : Obmolova, G.; Ban, C.; Hsieh, P.; Yang, W.
Deposited on : unknown
Resolution : 2.20 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

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : **NOT EXECUTED**
EDS : **NOT EXECUTED**
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

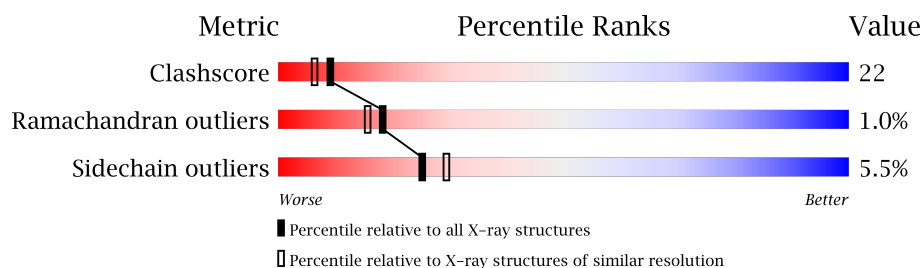
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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))
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)

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. 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	23	
2	D	22	
3	A	765	
3	B	765	

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
5	EDO	B	907	-	-	X	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 13724 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 DNA chain called DNA (5'-D(*GP*CP*GP*AP*CP*GP*CP*TP*AP*GP*CP*GP*TP*GP*CP*GP*GP*CP*TP*CP*GP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	23	Total	C	N	O	P	0	0	0
			469	222	87	138	22			

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*GP*AP*CP*GP*AP*GP*CP*CP*GP*CP*CP*GP*CP*TP*AP*GP*CP*GP*TP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	22	Total	C	N	O	P	0	0	0
			450	212	88	129	21			

- Molecule 3 is a protein called DNA MISMATCH REPAIR PROTEIN MUTS.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	A	759	Total	C	N	O	S	Se	0	0	0
			6007	3825	1068	1100	1	13			
3	B	749	Total	C	N	O	S	Se	0	0	0
			5936	3784	1055	1083	1	13			

There are 26 discrepancies between the modelled and reference sequences:

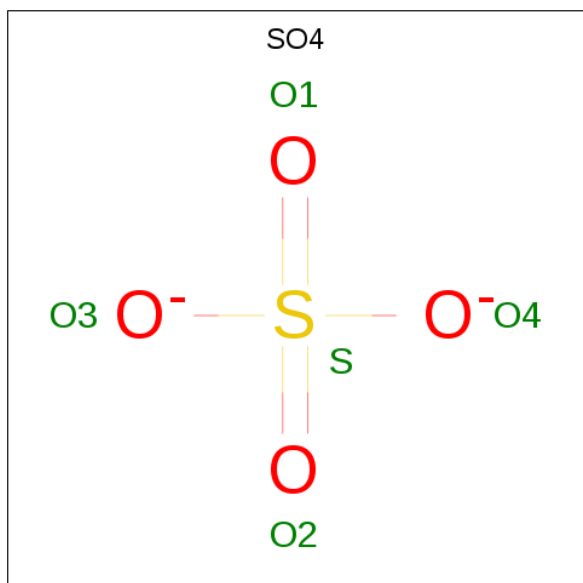
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	ENGINEERED	UNP Q56215
A	4	MSE	MET	ENGINEERED	UNP Q56215
A	70	MSE	MET	ENGINEERED	UNP Q56215
A	88	MSE	MET	ENGINEERED	UNP Q56215
A	201	MSE	MET	ENGINEERED	UNP Q56215
A	250	MSE	MET	ENGINEERED	UNP Q56215
A	481	MSE	MET	ENGINEERED	UNP Q56215
A	574	MSE	MET	ENGINEERED	UNP Q56215
A	586	MSE	MET	ENGINEERED	UNP Q56215
A	640	MSE	MET	ENGINEERED	UNP Q56215

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	643	MSE	MET	ENGINEERED	UNP Q56215
A	744	MSE	MET	ENGINEERED	UNP Q56215
A	762	MSE	MET	ENGINEERED	UNP Q56215
B	1001	MSE	MET	ENGINEERED	UNP Q56215
B	1004	MSE	MET	ENGINEERED	UNP Q56215
B	1070	MSE	MET	ENGINEERED	UNP Q56215
B	1088	MSE	MET	ENGINEERED	UNP Q56215
B	1201	MSE	MET	ENGINEERED	UNP Q56215
B	1250	MSE	MET	ENGINEERED	UNP Q56215
B	1481	MSE	MET	ENGINEERED	UNP Q56215
B	1574	MSE	MET	ENGINEERED	UNP Q56215
B	1586	MSE	MET	ENGINEERED	UNP Q56215
B	1640	MSE	MET	ENGINEERED	UNP Q56215
B	1643	MSE	MET	ENGINEERED	UNP Q56215
B	1744	MSE	MET	ENGINEERED	UNP Q56215
B	1762	MSE	MET	ENGINEERED	UNP Q56215

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

Continued on next page...

Continued from previous page...

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

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	24	Total	O	0	0
			24	24		

Continued on next page...

Continued from previous page...

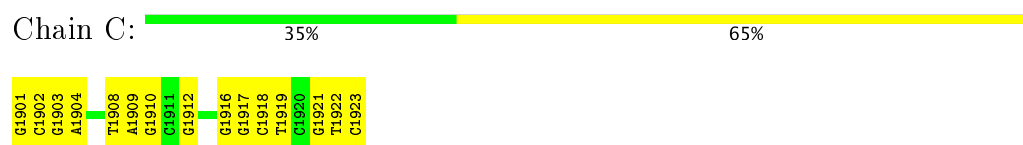
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	16	Total 16	O 16	0	0
6	A	407	Total 407	O 407	0	0
6	B	367	Total 367	O 367	0	0

3 Residue-property plots

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.

Note EDS was not executed.

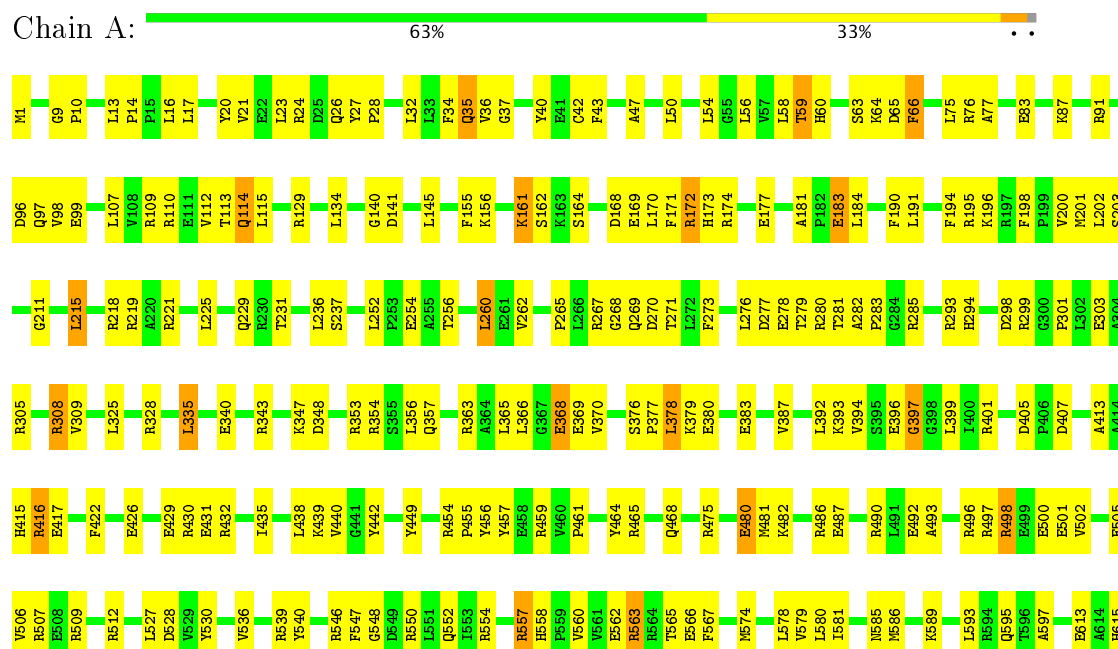
- Molecule 1: DNA (5'-D(*GP*CP*GP*AP*CP*GP*CP*TP*AP*GP*CP*GP*TP*GP*CP*GP*GP*CP*TP*CP*GP*TP*C)-3')

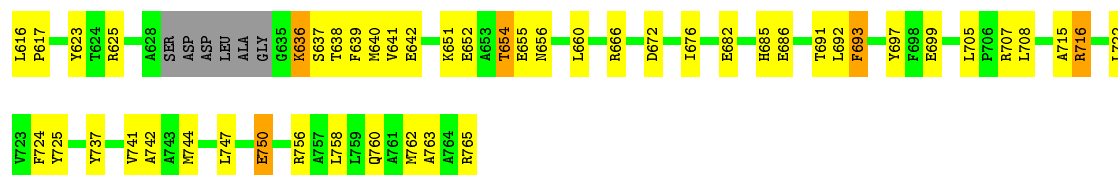


- Molecule 2: DNA (5'-D(*GP*GP*AP*CP*GP*AP*GP*CP*CP*GP*CP*CP*GP*CP*TP*AP*GP*CP*GP*TP*CP*G)-3')



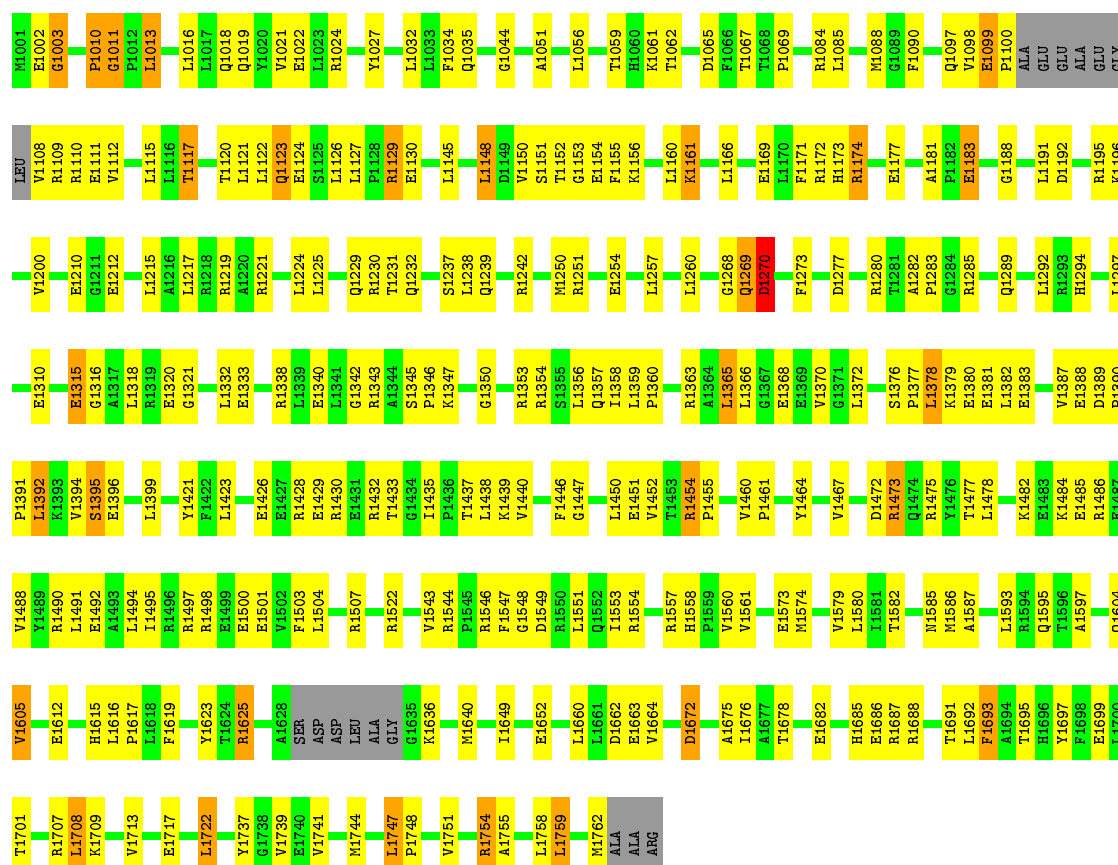
- Molecule 3: DNA MISMATCH REPAIR PROTEIN MUTS





• Molecule 3: DNA MISMATCH REPAIR PROTEIN MUTS

Chain B: 62% 32%



4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	103.17Å 114.22Å 160.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	26.60 – 2.20	Depositor
% Data completeness (in resolution range)	95.2 (26.60-2.20)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 0.9	Depositor
R, R_{free}	0.227 , 0.259	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	13724	wwPDB-VP
Average B, all atoms (Å ²)	38.0	wwPDB-VP

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: SO4, EDO

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	C	0.35	0/525	0.73	0/809
2	D	0.34	0/505	0.68	0/778
3	A	0.36	0/6112	0.62	1/8249 (0.0%)
3	B	0.35	0/6040	0.63	0/8152
All	All	0.36	0/13182	0.63	1/17988 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	134	LEU	N-CA-C	-5.41	96.41	111.00

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	C	469	0	259	22	0
2	D	450	0	246	23	0
3	A	6007	0	6093	268	0
3	B	5936	0	6024	272	0
4	A	10	0	0	0	0
4	B	10	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	4	0	6	3	0
5	B	24	0	36	9	0
6	A	407	0	0	19	0
6	B	367	0	0	24	0
6	C	24	0	0	1	0
6	D	16	0	0	3	0
All	All	13724	0	12664	568	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 22.

The worst 5 of 568 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)
3:B:1747:LEU:HD13	3:B:1748:PRO:HD2	1.37	1.06
3:B:1345:SER:HB2	3:B:1392:LEU:HD13	1.37	1.06
3:B:1010:PRO:HG2	3:B:1011:GLY:H	1.26	1.00
3:A:161:LYS:H	3:A:161:LYS:HZ2	1.15	0.93
2:D:1967:DG:H2"	2:D:1968:DC:H5"	1.50	0.93

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	
3	A	755/765 (99%)	720 (95%)	29 (4%)	6 (1%)	22	21
3	B	743/765 (97%)	713 (96%)	21 (3%)	9 (1%)	15	12
All	All	1498/1530 (98%)	1433 (96%)	50 (3%)	15 (1%)	18	16

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	270	ASP
3	A	397	GLY
3	B	1010	PRO
3	B	1011	GLY
3	B	1269	GLN

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	
3	A	613/604 (102%)	578 (94%)	35 (6%)	24	28
3	B	608/604 (101%)	576 (95%)	32 (5%)	26	31
All	All	1221/1208 (101%)	1154 (94%)	67 (6%)	25	29

5 of 67 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	636	LYS
3	B	1059	THR
3	B	1693	PHE
3	A	654	THR
3	A	750	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	595	GLN
3	B	1097	GLN
3	B	1604	GLN
3	A	685	HIS
3	B	1018	GLN

5.3.3 RNA ⓘ

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

11 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	A	851	-	4,4,4	0.33	0	6,6,6	0.15	0
4	SO4	A	852	-	4,4,4	0.31	0	6,6,6	0.12	0
5	EDO	A	902	-	3,3,3	0.41	0	2,2,2	0.34	0
4	SO4	B	1851	-	4,4,4	0.31	0	6,6,6	0.09	0
4	SO4	B	1852	-	4,4,4	0.33	0	6,6,6	0.10	0
5	EDO	B	901	-	3,3,3	0.49	0	2,2,2	0.32	0
5	EDO	B	903	-	3,3,3	0.42	0	2,2,2	0.37	0
5	EDO	B	904	-	3,3,3	0.41	0	2,2,2	0.37	0
5	EDO	B	905	-	3,3,3	0.41	0	2,2,2	0.37	0
5	EDO	B	906	-	3,3,3	0.35	0	2,2,2	0.40	0
5	EDO	B	907	-	3,3,3	0.35	0	2,2,2	0.36	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	SO4	A	851	-	-	0/0/0/0	0/0/0/0
4	SO4	A	852	-	-	0/0/0/0	0/0/0/0
5	EDO	A	902	-	-	0/1/1/1	0/0/0/0
4	SO4	B	1851	-	-	0/0/0/0	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	SO4	B	1852	-	-	0/0/0/0	0/0/0/0
5	EDO	B	901	-	-	0/1/1/1	0/0/0/0
5	EDO	B	903	-	-	0/1/1/1	0/0/0/0
5	EDO	B	904	-	-	0/1/1/1	0/0/0/0
5	EDO	B	905	-	-	0/1/1/1	0/0/0/0
5	EDO	B	906	-	-	0/1/1/1	0/0/0/0
5	EDO	B	907	-	-	0/1/1/1	0/0/0/0

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.

5 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	902	EDO	3	0
5	B	901	EDO	1	0
5	B	905	EDO	1	0
5	B	906	EDO	1	0
5	B	907	EDO	6	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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.