



# Full wwPDB X-ray Structure Validation Report ⓘ

May 26, 2020 – 07:01 am BST

PDB ID : 5JNT  
Title : Crystal structure of human low molecular weight protein tyrosine phosphatase (LMPTP) type A complexed with MES  
Authors : Stanford, S.M.; Aleshin, A.E.; Liddington, R.C.; Bankston, L.; Cadwell, G.; Bottini, N.  
Deposited on : 2016-04-30  
Resolution : 1.45 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](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



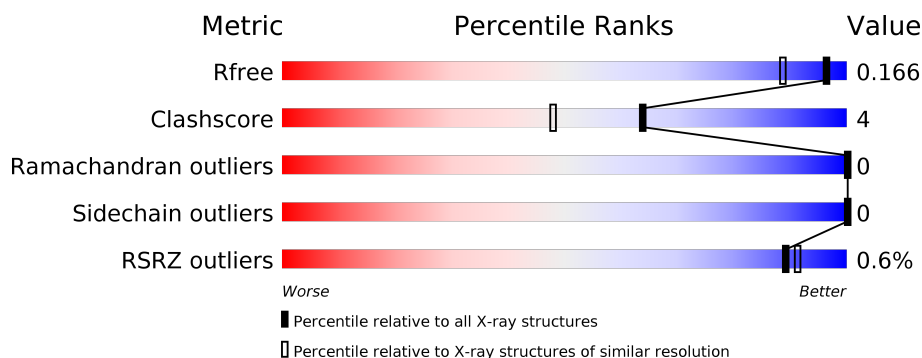
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.45 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	1156 (1.46-1.46)
Clashscore	141614	1202 (1.46-1.46)
Ramachandran outliers	138981	1178 (1.46-1.46)
Sidechain outliers	138945	1178 (1.46-1.46)
RSRZ outliers	127900	1139 (1.46-1.46)

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	160	<div> <div style="display: flex; align-items: center;"> <div style="width: 100%; height: 10px; position: relative;"> <div style="position: absolute; top: -5px; left: 0; width: 100%; height: 1px; background-color: red;"></div> <div style="position: absolute; bottom: -5px; left: 0; width: 100%; height: 1px; background-color: green;"></div> <div style="position: absolute; bottom: -5px; right: 0; width: 100%; height: 1px; background-color: yellow;"></div> <div style="position: absolute; bottom: -5px; right: 0; width: 100%; height: 1px; background-color: orange;"></div> <div style="position: absolute; bottom: -5px; right: 0; width: 100%; height: 1px; background-color: grey;"></div> </div> <div style="margin-left: 10px;"> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>73%</span> <span>21%</span> <span>• •</span> </div> </div> </div> </div>



## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 1591 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.

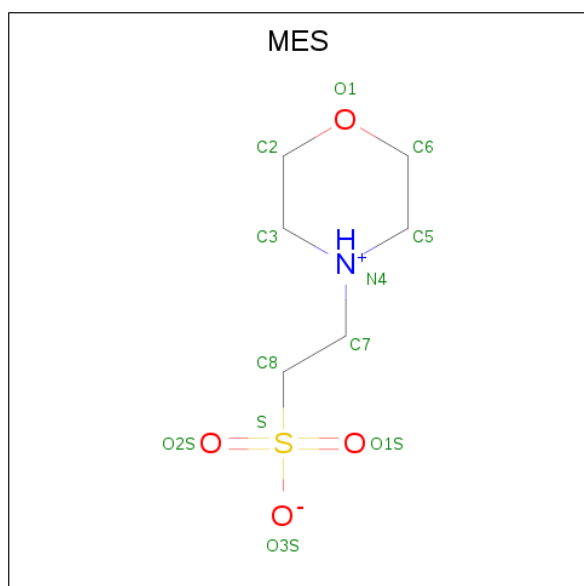
- Molecule 1 is a protein called Low molecular weight phosphotyrosine protein phosphatase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	154	1292	804	228	248	12	0	7	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P24666
A	-1	SER	-	expression tag	UNP P24666
A	4	VAL	ALA	conflict	UNP P24666

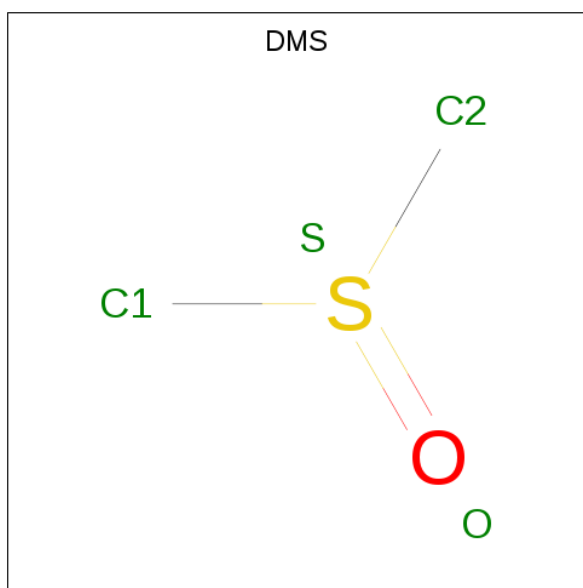
- Molecule 2 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	12	6	1	4	1	0	0

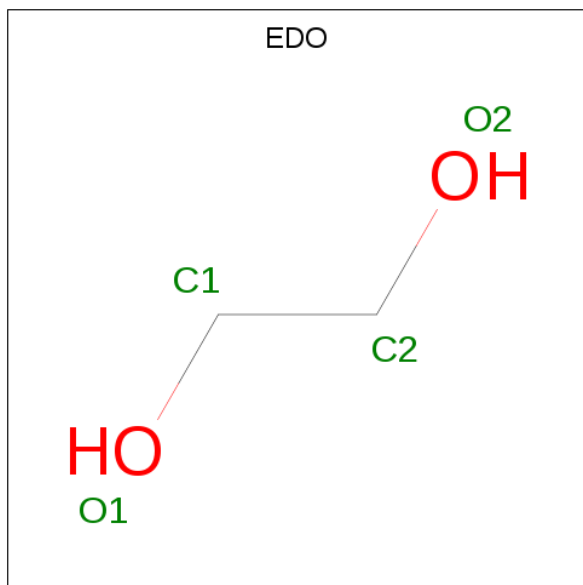


- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula:  $C_2H_6OS$ ).



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

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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



- Molecule 5 is water.

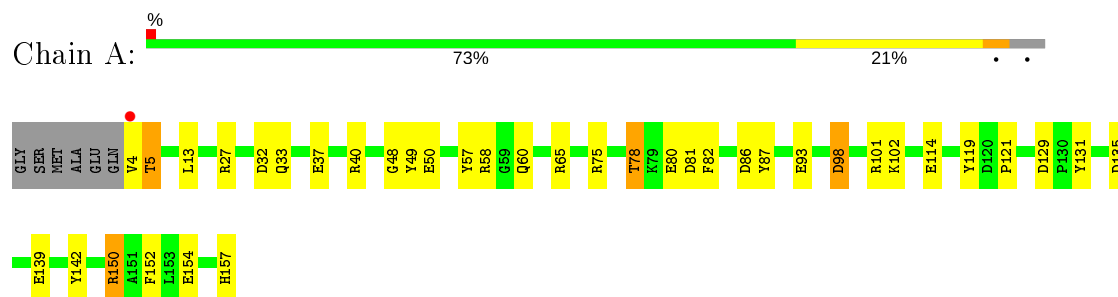
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	275	Total 275	O 275	0	0



### 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: Low molecular weight phosphotyrosine protein phosphatase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	33.91Å 55.02Å 96.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.90 – 1.45 28.86 – 1.36	Depositor EDS
% Data completeness (in resolution range)	99.8 (33.90-1.45) 97.7 (28.86-1.36)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.32 (at 1.36Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.135 , 0.166 0.135 , 0.166	Depositor DCC
$R_{free}$ test set	1930 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.9	Xtriage
Anisotropy	0.068	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 52.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	1591	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.51% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: DMS, EDO, MES

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	1.78	27/1315 (2.1%)	1.85	40/1770 (2.3%)

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

All (27) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	57	TYR	CE1-CZ	-11.03	1.24	1.38
1	A	101	ARG	NE-CZ	-10.31	1.19	1.33
1	A	57	TYR	CZ-OH	8.50	1.52	1.37
1	A	114	GLU	CD-OE1	-8.43	1.16	1.25
1	A	50	GLU	CD-OE2	7.98	1.34	1.25
1	A	60	GLN	CD-NE2	7.91	1.52	1.32
1	A	87	TYR	CD2-CE2	7.80	1.51	1.39
1	A	57	TYR	CG-CD1	-7.00	1.30	1.39
1	A	102	LYS	CD-CE	6.80	1.68	1.51
1	A	139	GLU	CD-OE2	6.58	1.32	1.25
1	A	87	TYR	CE1-CZ	-6.52	1.30	1.38
1	A	93[A]	GLU	CD-OE2	6.40	1.32	1.25
1	A	93[B]	GLU	CD-OE2	6.40	1.32	1.25
1	A	142	TYR	CE1-CZ	-6.06	1.30	1.38
1	A	87	TYR	CZ-OH	5.95	1.48	1.37
1	A	5	THR	CA-C	5.82	1.68	1.52
1	A	32	ASP	CB-CG	5.68	1.63	1.51
1	A	80	GLU	CB-CG	-5.66	1.41	1.52

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	82	PHE	CE2-CZ	-5.28	1.27	1.37
1	A	93[A]	GLU	CD-OE1	-5.24	1.19	1.25
1	A	93[B]	GLU	CD-OE1	-5.24	1.19	1.25
1	A	37	GLU	CG-CD	5.22	1.59	1.51
1	A	101	ARG	CZ-NH1	-5.20	1.26	1.33
1	A	33	GLN	CD-NE2	-5.16	1.20	1.32
1	A	78	THR	CA-CB	5.09	1.66	1.53
1	A	98	ASP	CG-OD1	5.08	1.37	1.25
1	A	75	ARG	CZ-NH1	-5.06	1.26	1.33

All (40) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	150[A]	ARG	NE-CZ-NH2	-14.02	113.29	120.30
1	A	150[B]	ARG	NE-CZ-NH2	-14.02	113.29	120.30
1	A	65	ARG	NE-CZ-NH1	13.94	127.27	120.30
1	A	57	TYR	CG-CD1-CE1	11.56	130.54	121.30
1	A	32	ASP	CB-CG-OD2	-10.86	108.52	118.30
1	A	98	ASP	CB-CG-OD2	-10.60	108.76	118.30
1	A	135	ASP	CB-CG-OD1	10.22	127.50	118.30
1	A	87	TYR	CD1-CE1-CZ	-9.74	111.04	119.80
1	A	58	ARG	NE-CZ-NH2	-9.46	115.57	120.30
1	A	119	TYR	CB-CG-CD1	9.30	126.58	121.00
1	A	129	ASP	CB-CG-OD1	8.98	126.38	118.30
1	A	75	ARG	NE-CZ-NH1	8.94	124.77	120.30
1	A	87	TYR	CG-CD2-CE2	-8.84	114.22	121.30
1	A	27	ARG	NE-CZ-NH1	8.79	124.69	120.30
1	A	150[A]	ARG	NE-CZ-NH1	8.67	124.64	120.30
1	A	150[B]	ARG	NE-CZ-NH1	8.67	124.64	120.30
1	A	87	TYR	CE1-CZ-CE2	8.27	133.04	119.80
1	A	40	ARG	NE-CZ-NH2	-8.09	116.26	120.30
1	A	57	TYR	CG-CD2-CE2	-8.08	114.84	121.30
1	A	135	ASP	CB-CG-OD2	-8.05	111.05	118.30
1	A	75	ARG	NE-CZ-NH2	-7.94	116.33	120.30
1	A	87	TYR	CD1-CG-CD2	7.78	126.46	117.90
1	A	57	TYR	CD1-CE1-CZ	-7.70	112.87	119.80
1	A	87	TYR	CZ-CE2-CD2	-7.58	112.98	119.80
1	A	87	TYR	CB-CG-CD1	-7.51	116.50	121.00
1	A	32	ASP	OD1-CG-OD2	6.90	136.42	123.30
1	A	87	TYR	CB-CG-CD2	-6.82	116.91	121.00
1	A	101	ARG	CD-NE-CZ	6.82	133.14	123.60
1	A	57	TYR	CB-CG-CD1	6.74	125.04	121.00

*Continued on next page...*



Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	87	TYR	CE1-CZ-OH	-6.62	102.21	120.10
1	A	86	ASP	CB-CG-OD2	5.91	123.62	118.30
1	A	129	ASP	CB-CG-OD2	-5.88	113.01	118.30
1	A	65	ARG	NH1-CZ-NH2	-5.68	113.16	119.40
1	A	157	HIS	CA-C-O	5.67	132.02	120.10
1	A	131	TYR	CZ-CE2-CD2	-5.63	114.74	119.80
1	A	152	PHE	CB-CG-CD1	5.37	124.56	120.80
1	A	48	GLY	O-C-N	-5.28	114.25	122.70
1	A	101	ARG	NE-CZ-NH1	-5.27	117.67	120.30
1	A	81	ASP	CB-CG-OD1	5.25	123.02	118.30
1	A	49	TYR	CB-CG-CD2	5.05	124.03	121.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	98	ASP	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	1292	0	1262	11	0
2	A	12	0	13	0	0
3	A	4	0	6	0	0
4	A	8	0	12	2	0
5	A	275	0	0	7	2
All	All	1591	0	1293	11	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 4.

All (11) 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:A:13[A]:LEU:HD23	5:A:366:HOH:O	1.74	0.86

Continued on next page...



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:13[B]:LEU:HD12	5:A:324:HOH:O	1.89	0.71
1:A:121:PRO:HG2	4:A:203:EDO:H11	1.75	0.68
1:A:154[B]:GLU:OE1	5:A:301:HOH:O	2.12	0.65
1:A:150[B]:ARG:NH2	4:A:203:EDO:O1	2.33	0.58
1:A:13[A]:LEU:CD2	5:A:366:HOH:O	2.42	0.56
1:A:78:THR:HB	5:A:459:HOH:O	2.12	0.48
1:A:4:VAL:HG12	1:A:5:THR:H	1.80	0.45
1:A:154[B]:GLU:CG	5:A:460:HOH:O	2.65	0.44
1:A:150[B]:ARG:O	1:A:154[B]:GLU:HG3	2.20	0.42
1:A:150[A]:ARG:NH1	5:A:309:HOH:O	2.52	0.42

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)
5:A:371:HOH:O	5:A:427:HOH:O[4_555]	2.12	0.08
5:A:441:HOH:O	5:A:491:HOH:O[1_655]	2.15	0.05

## 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	159/160 (99%)	156 (98%)	3 (2%)	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	145/142 (102%)	145 (100%)	0	100	100

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

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

Mol	Chain	Res	Type
1	A	104	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

4 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	EDO	A	204	-	3,3,3	1.69	1 (33%)	2,2,2	1.47	0
3	DMS	A	202	-	3,3,3	2.12	1 (33%)	3,3,3	0.88	0



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	EDO	A	203	-	3,3,3	0.93	0	2,2,2	2.53	1 (50%)
2	MES	A	201	-	12,12,12	1.91	2 (16%)	14,16,16	1.21	2 (14%)

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	EDO	A	204	-	-	1/1/1/1	-
4	EDO	A	203	-	-	1/1/1/1	-
2	MES	A	201	-	-	0/6/14/14	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	201	MES	C8-S	-5.31	1.70	1.77
3	A	202	DMS	O-S	3.44	1.73	1.50
2	A	201	MES	O1S-S	2.20	1.51	1.45
4	A	204	EDO	O1-C1	2.12	1.53	1.42

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	203	EDO	O1-C1-C2	3.09	134.11	111.91
2	A	201	MES	C2-C3-N4	2.13	113.34	110.10
2	A	201	MES	O1-C2-C3	-2.12	107.13	111.80

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	204	EDO	O1-C1-C2-O2
4	A	203	EDO	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	A	203	EDO	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	154/160 (96%)	-0.08	1 (0%) 89 91	8, 13, 24, 61	1 (0%)

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	4	VAL	13.2

### 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, 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
4	EDO	A	204	4/4	0.73	0.22	29,31,33,33	0
4	EDO	A	203	4/4	0.86	0.24	24,33,37,41	0
3	DMS	A	202	4/4	0.93	0.12	18,19,20,21	0
2	MES	A	201	12/12	0.98	0.07	9,10,14,14	0



## 6.5 Other polymers

There are no such residues in this entry.