



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 05:10 am BST

PDB ID : 4WAC  
Title : Crystal Structure of TarM  
Authors : Koc, C.; Stehle, T.; Xia, G.; Peschel, A.  
Deposited on : 2014-08-29  
Resolution : 2.40 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

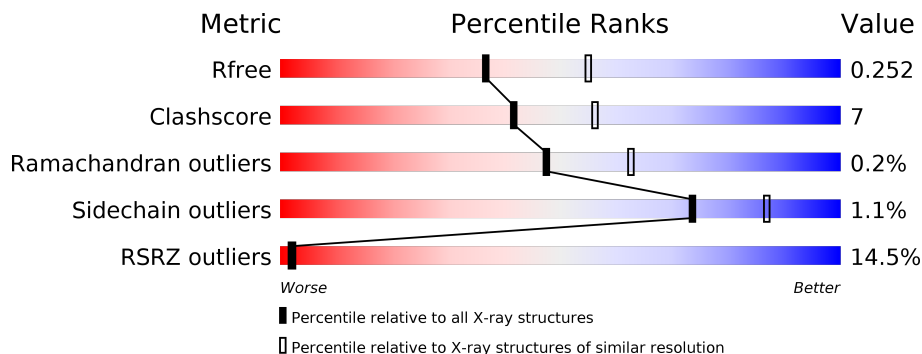
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	498	

## 2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 4194 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.

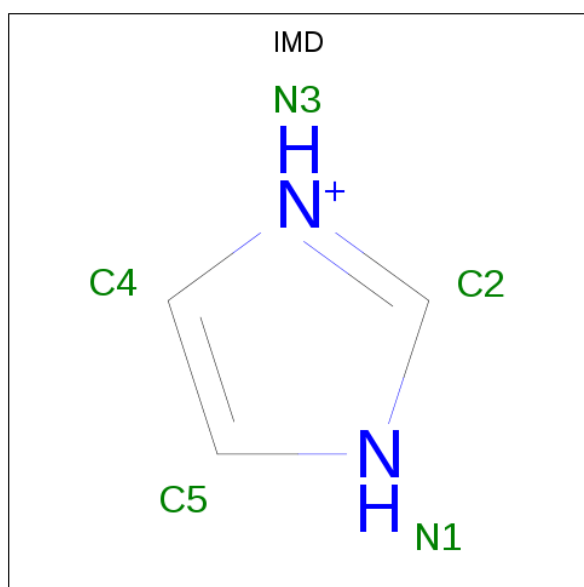
- Molecule 1 is a protein called Glycosyl transferase, group 1 family protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	498	4079	2612	687	761	19	0	1	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	ILE	-	expression tag	UNP Q2FI41
A	-3	ASP	-	expression tag	UNP Q2FI41
A	-2	PRO	-	expression tag	UNP Q2FI41
A	-1	PHE	-	expression tag	UNP Q2FI41
A	0	THR	-	expression tag	UNP Q2FI41

- Molecule 2 is IMIDAZOLE (three-letter code: IMD) (formula:  $C_3H_5N_2$ ).



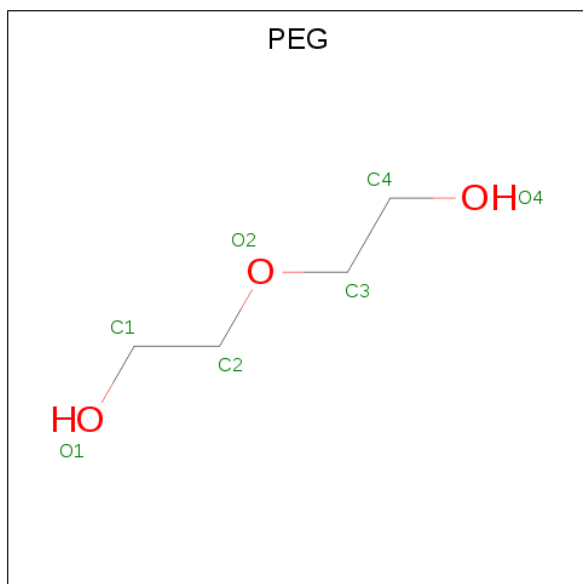
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	N		
2	A	1	5	3	2	0	0

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula:  $C_2H_3O_2$ ).



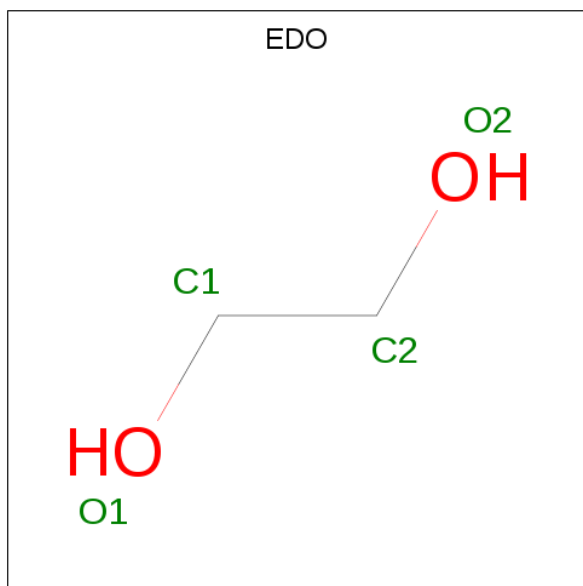
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



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

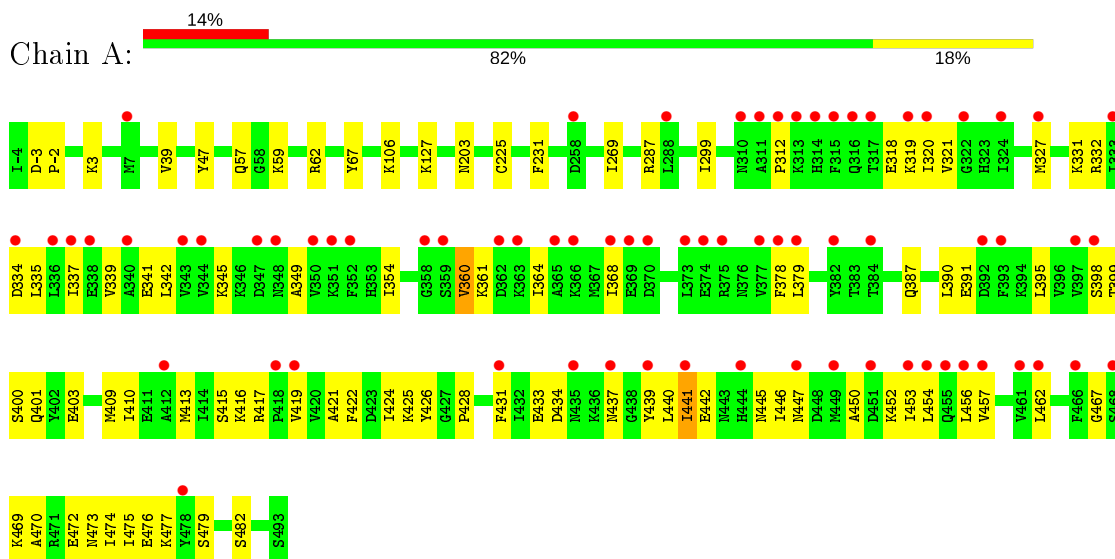
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	87	Total O 87 87	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: Glycosyl transferase, group 1 family protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	124.75Å 124.75Å 223.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.62 – 2.40 48.62 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.6 (48.62-2.40) 99.6 (48.62-2.40)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.60 (at 2.39Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.215 , 0.252 0.218 , 0.252	Depositor DCC
$R_{free}$ test set	2036 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.6	Xtrriage
Anisotropy	0.304	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 48.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4194	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.33% 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: ACT, PEG, EDO, IMD

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	0.58	0/4161	0.67	0/5587

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

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	442	GLU	Peptide

### 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	4079	0	4101	57	0
2	A	5	0	5	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	8	0	6	0	0
4	A	7	0	10	0	0
5	A	8	0	12	2	0
6	A	87	0	0	2	0
All	All	4194	0	4134	58	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 7.

All (58) 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:312:PRO:HB3	1:A:391:GLU:HG3	1.59	0.83
1:A:339:VAL:HG22	1:A:446:ILE:HB	1.61	0.83
1:A:203:ASN:HB3	5:A:506:EDO:H11	1.66	0.75
1:A:409:MET:HG2	1:A:419:VAL:HG11	1.70	0.74
1:A:401:GLN:HA	1:A:424:ILE:HG21	1.68	0.73
1:A:57:GLN:OE1	1:A:59:LYS:NZ	2.24	0.70
1:A:409:MET:HE1	1:A:428:PRO:HG3	1.74	0.69
1:A:472:GLU:HA	1:A:475:ILE:HD12	1.75	0.68
5:A:506:EDO:O1	6:A:601:HOH:O	2.12	0.68
1:A:334:ASP:OD1	1:A:335:LEU:N	2.29	0.65
1:A:62:ARG:HH21	1:A:62:ARG:HG2	1.63	0.62
1:A:395:LEU:HD21	1:A:457:VAL:HG12	1.82	0.61
1:A:361:LYS:HG3	1:A:379:LEU:HD21	1.83	0.59
1:A:331:LYS:NZ	1:A:403:GLU:OE2	2.37	0.58
1:A:453:ILE:O	1:A:457:VAL:HG13	2.04	0.58
1:A:337:ILE:HD11	1:A:364:ILE:HG23	1.86	0.58
1:A:341:GLU:HG3	1:A:345:LYS:HE2	1.87	0.56
1:A:332:ARG:HB2	1:A:399:THR:O	2.07	0.55
1:A:454:LEU:HA	1:A:457:VAL:HG22	1.88	0.54
1:A:364:ILE:O	1:A:368:ILE:HG13	2.08	0.53
1:A:434:ASP:HB2	1:A:440:LEU:CD2	2.39	0.53
1:A:341:GLU:O	1:A:345:LYS:HG2	2.09	0.52
1:A:473:ASN:O	1:A:477:LYS:HG2	2.10	0.51
1:A:327:MET:HB3	1:A:360:VAL:HG12	1.92	0.51
1:A:441:ILE:HD12	1:A:445:ASN:O	2.10	0.51
1:A:439:TYR:CD2	1:A:452:LYS:HB3	2.47	0.50
1:A:342:LEU:HD21	1:A:447:ASN:OD1	2.12	0.50
1:A:415:SER:O	1:A:417:ARG:NH1	2.45	0.49
1:A:470:ALA:O	1:A:474:ILE:HG12	2.11	0.49

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:416:LYS:O	1:A:467:GLY:HA3	2.12	0.49
1:A:433:GLU:OE2	1:A:437:ASN:ND2	2.46	0.49
1:A:361:LYS:CG	1:A:379:LEU:HD21	2.43	0.49
1:A:400:SER:O	1:A:424:ILE:HG12	2.13	0.49
1:A:335:LEU:O	1:A:339:VAL:HG23	2.13	0.49
1:A:425:LYS:HA	1:A:426:TYR:HA	1.55	0.48
1:A:231:PHE:CG	1:A:269:ILE:HG12	2.48	0.48
1:A:354:ILE:N	1:A:378:PHE:O	2.44	0.47
1:A:425:LYS:HB2	1:A:426:TYR:CD2	2.49	0.47
1:A:318:GLU:O	1:A:320:ILE:N	2.46	0.47
1:A:450:ALA:O	1:A:454:LEU:HG	2.15	0.46
1:A:410:ILE:HG13	1:A:431:PHE:CZ	2.51	0.46
1:A:473:ASN:O	1:A:476:GLU:HG2	2.17	0.45
1:A:398:SER:HB3	1:A:421:ALA:HB2	1.98	0.45
1:A:364:ILE:O	1:A:368:ILE:N	2.47	0.44
1:A:287:ARG:HG3	1:A:299:ILE:HG22	1.99	0.43
1:A:106:LYS:HG2	1:A:106:LYS:O	2.17	0.43
1:A:3:LYS:NZ	6:A:604:HOH:O	2.45	0.43
1:A:360:VAL:HG13	1:A:364:ILE:HD12	2.01	0.43
1:A:39:VAL:HA	1:A:67:TYR:O	2.19	0.42
1:A:456:LEU:HD11	1:A:462:LEU:HB2	2.02	0.42
1:A:413:MET:HE2	1:A:470:ALA:O	2.19	0.42
1:A:319:LYS:HE2	1:A:349:ALA:HB3	2.01	0.42
1:A:479:SER:O	1:A:482:SER:OG	2.29	0.42
1:A:456:LEU:HD12	1:A:456:LEU:HA	1.59	0.41
1:A:-3:ASP:HA	1:A:-2:PRO:HD2	1.88	0.41
1:A:387:GLN:HA	1:A:390:LEU:HD12	2.03	0.41
1:A:422:PHE:CE1	1:A:441:ILE:HD11	2.56	0.41
1:A:225:CYS:SG	1:A:231:PHE:HA	2.61	0.40

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	497/498 (100%)	473 (95%)	23 (5%)	1 (0%)	47 62

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	127	LYS

### 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	455/454 (100%)	450 (99%)	5 (1%)	73 87

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	47	TYR
1	A	321	VAL
1	A	360	VAL
1	A	441	ILE
1	A	469	LYS

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	304	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

6 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	IMD	A	501	-	3,5,5	0.40	0	4,5,5	0.92	0
3	ACT	A	505	-	1,3,3	2.99	1 (100%)	0,3,3	0.00	-
5	EDO	A	506	-	3,3,3	0.15	0	2,2,2	0.81	0
3	ACT	A	502	-	1,3,3	2.38	1 (100%)	0,3,3	0.00	-
4	PEG	A	503	-	6,6,6	0.48	0	5,5,5	0.91	0
5	EDO	A	504	-	3,3,3	0.53	0	2,2,2	0.19	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	A	503	-	-	1/4/4/4	-
5	EDO	A	504	-	-	0/1/1/1	-
5	EDO	A	506	-	-	1/1/1/1	-
2	IMD	A	501	-	-	-	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	505	ACT	CH3-C	2.99	1.52	1.48
3	A	502	ACT	CH3-C	2.38	1.51	1.48

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	503	PEG	O1-C1-C2-O2
5	A	506	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
5	A	506	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

### 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, 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	498/498 (100%)	0.69	72 (14%) <b>2</b> <b>2</b>	28, 69, 106, 124	0

All (72) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	454	LEU	6.7
1	A	373	LEU	6.6
1	A	365	ALA	6.5
1	A	382	TYR	6.5
1	A	461	VAL	6.4
1	A	340	ALA	6.4
1	A	453	ILE	6.1
1	A	314	HIS	6.0
1	A	313	LYS	5.9
1	A	374	GLU	5.9
1	A	337	ILE	5.7
1	A	359	SER	5.5
1	A	369	GLU	5.4
1	A	451	ASP	5.3
1	A	333	ILE	5.3
1	A	379	LEU	5.2
1	A	444	HIS	5.0
1	A	366	LYS	4.9
1	A	358	GLY	4.8
1	A	378	PHE	4.8
1	A	311	ALA	4.6
1	A	375	ARG	4.5
1	A	336	LEU	4.4
1	A	344	VAL	4.4
1	A	312	PRO	4.3
1	A	258	ASP	4.1
1	A	447	ASN	4.0

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	449	MET	3.7
1	A	384	THR	3.7
1	A	316	GLN	3.6
1	A	327	MET	3.5
1	A	468	SER	3.5
1	A	350	VAL	3.4
1	A	352	PHE	3.4
1	A	362	ASP	3.3
1	A	338	GLU	3.3
1	A	377	VAL	3.3
1	A	288	LEU	3.2
1	A	320	ILE	3.2
1	A	363	LYS	3.1
1	A	334	ASP	3.1
1	A	398	SER	3.0
1	A	462	LEU	3.0
1	A	431	PHE	3.0
1	A	310	ASN	2.8
1	A	412	ALA	2.8
1	A	317	THR	2.8
1	A	437	ASN	2.7
1	A	324	ILE	2.7
1	A	456	LEU	2.7
1	A	441	ILE	2.7
1	A	319	LYS	2.6
1	A	397	VAL	2.5
1	A	439	TYR	2.4
1	A	418	PRO	2.4
1	A	393	PHE	2.4
1	A	392	ASP	2.4
1	A	419	VAL	2.3
1	A	455	GLN	2.3
1	A	466	PHE	2.3
1	A	368	ILE	2.3
1	A	478	TYR	2.2
1	A	370	ASP	2.2
1	A	457	VAL	2.1
1	A	7	MET	2.1
1	A	435	ASN	2.1
1	A	343	VAL	2.1
1	A	351	LYS	2.1
1	A	348	ASN	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	347	ASP	2.0
1	A	315	PHE	2.0
1	A	322	GLY	2.0

## 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	PEG	A	503	7/7	0.70	0.21	60,77,89,100	0
3	ACT	A	505	4/4	0.81	0.30	48,52,54,82	0
5	EDO	A	504	4/4	0.93	0.15	73,78,81,91	0
5	EDO	A	506	4/4	0.94	0.23	48,48,60,76	0
3	ACT	A	502	4/4	0.95	0.17	41,53,60,65	0
2	IMD	A	501	5/5	0.98	0.20	38,38,39,41	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.