



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 20, 2024 – 03:18 pm GMT

PDB ID : 8P1R  
Title : Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant.  
Authors : Owen, C.D.; Penner, M.; Gascuena, A.N.; Wu, H.; Hernando, P.J.; Monaco, S.; Le Gall, G.; Gardner, R.; Ndeh, D.; Urbanowicz, P.A.; Spencer, D.I.R.; Walsh, M.A.; Angulo, J.; Juge, N.  
Deposited on : 2023-05-12  
Resolution : 1.66 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

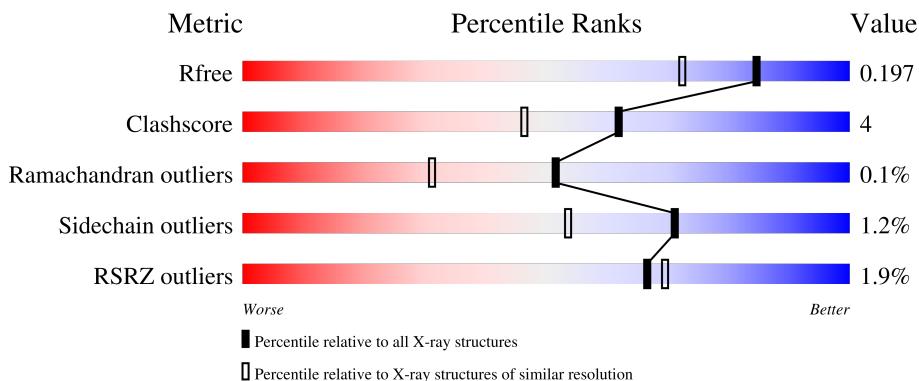
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.66 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	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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	350	
1	B	350	
1	C	350	
1	D	350	
1	E	350	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	350	 2% 79% 9% 12%

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
2	EDO	E	402	-	X	-	-

## 2 Entry composition [i](#)

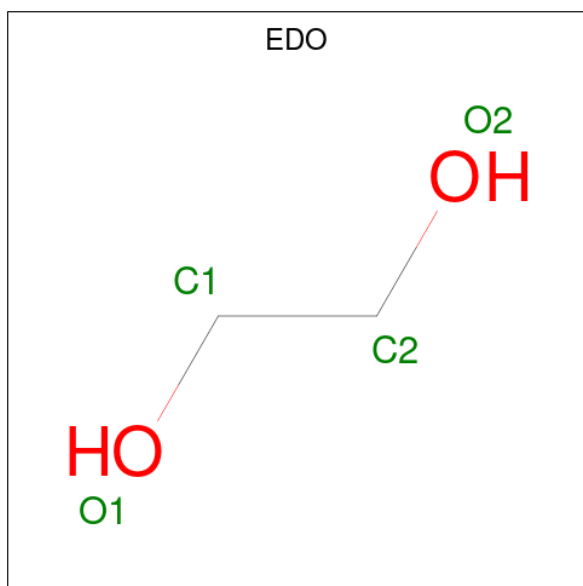
There are 4 unique types of molecules in this entry. The entry contains 18342 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	341	Total 2791	C 1764	N 466	O 550	S 11	0	7	0
1	B	338	Total 2758	C 1749	N 461	O 537	S 11	0	5	0
1	C	335	Total 2786	C 1763	N 470	O 542	S 11	0	9	0
1	D	335	Total 2737	C 1738	N 458	O 529	S 12	0	3	0
1	E	347	Total 2809	C 1782	N 471	O 544	S 12	0	3	0
1	F	309	Total 2537	C 1617	N 423	O 486	S 11	0	5	0

- Molecule 2 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
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0
2	C	1	Total C O 4 2 2	0	0
2	C	1	Total C O 4 2 2	0	0
2	D	1	Total C O 4 2 2	0	0
2	E	1	Total C O 4 2 2	0	0
2	E	1	Total C O 4 2 2	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	F	1	Total Mg 1 1	0	0

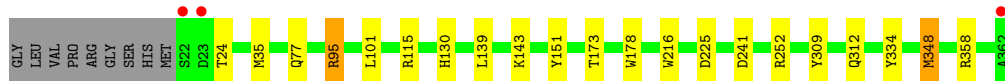
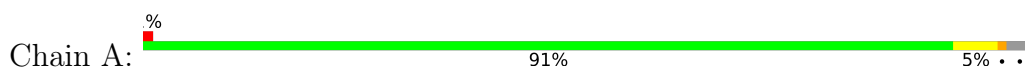
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	369	Total O 369 369	0	0
4	B	326	Total O 326 326	0	0
4	C	299	Total O 299 299	0	0
4	D	282	Total O 282 282	0	0
4	E	314	Total O 314 314	0	0
4	F	293	Total O 293 293	0	0

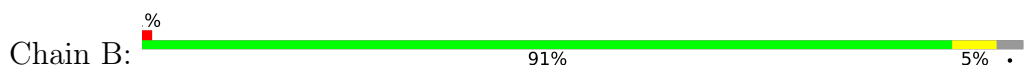
### 3 Residue-property plots [i](#)

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.

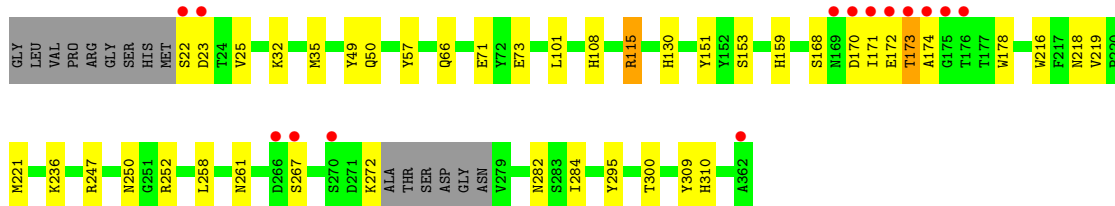
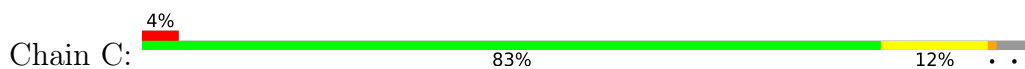
- Molecule 1: Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant



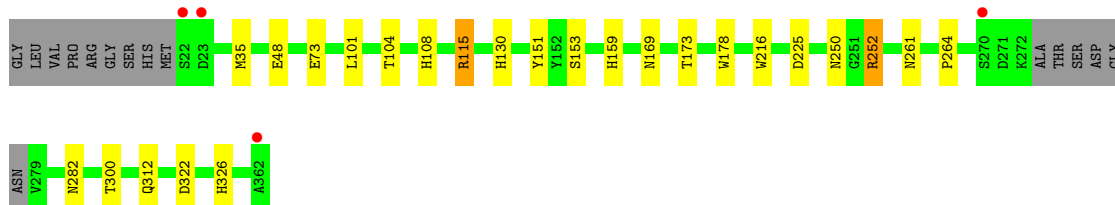
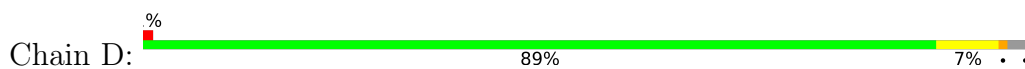
- Molecule 1: Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant



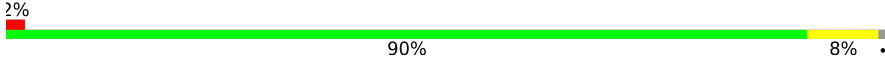
- Molecule 1: Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant

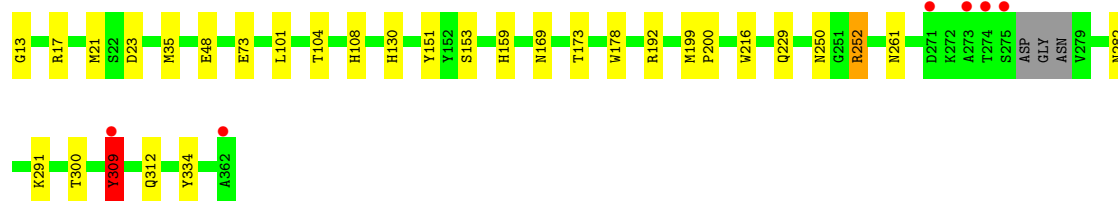


- Molecule 1: Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant




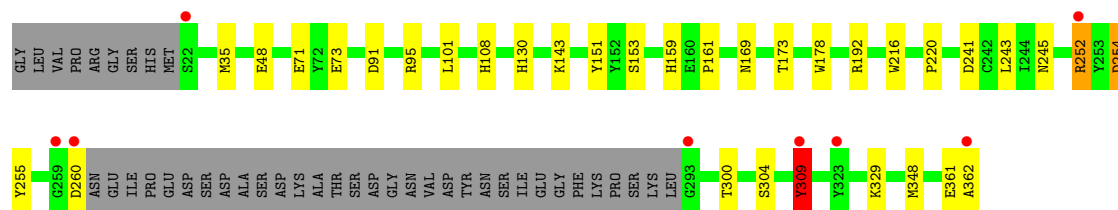
- Molecule 1: Bifidobacterium asteroides alpha-L-fucosidase (TT1819) catalytic mutant

Chain E:  2% 90% 8%



- Molecule 1: *Bifidobacterium asteroides* alpha-L-fucosidase (TT1819) catalytic mutant

Chain F:  2% 79% 9% 12%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.68Å 136.25Å 160.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	104.07 – 1.66 103.86 – 1.66	Depositor EDS
% Data completeness (in resolution range)	99.0 (104.07-1.66) 99.0 (103.86-1.66)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.42 (at 1.66Å)	Xtrriage
Refinement program	REFMAC 5.8.0405	Depositor
R, $R_{free}$	0.160 , 0.188 0.172 , 0.197	Depositor DCC
$R_{free}$ test set	11107 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	13.1	Xtrriage
Anisotropy	0.304	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 41.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	18342	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% 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: EDO, MG

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	0.59	0/2864	0.79	2/3884 (0.1%)
1	B	0.58	0/2831	0.80	2/3837 (0.1%)
1	C	0.58	0/2858	0.82	1/3873 (0.0%)
1	D	0.56	0/2810	0.81	3/3808 (0.1%)
1	E	0.59	0/2885	0.83	4/3912 (0.1%)
1	F	0.61	0/2606	0.89	5/3534 (0.1%)
All	All	0.59	0/16854	0.82	17/22848 (0.1%)

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	3
1	B	0	2
1	C	0	1
1	D	0	2
1	F	0	2
All	All	0	10

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	115[A]	ARG	CB-CG-CD	8.86	134.63	111.60
1	D	115[B]	ARG	CB-CG-CD	8.86	134.63	111.60
1	D	115[C]	ARG	CB-CG-CD	8.86	134.63	111.60
1	F	254	ASP	CB-CG-OD1	-7.36	111.67	118.30
1	A	95	ARG	NE-CZ-NH1	7.03	123.81	120.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	115	ARG	Sidechain
1	A	358	ARG	Sidechain
1	A	95	ARG	Sidechain
1	B	115[A]	ARG	Sidechain
1	B	252	ARG	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	2791	0	2609	20	0
1	B	2758	0	2588	24	0
1	C	2786	0	2616	32	0
1	D	2737	0	2576	20	0
1	E	2809	0	2640	24	0
1	F	2537	0	2390	27	0
2	A	8	0	11	2	0
2	B	12	0	14	3	0
2	C	8	0	12	0	0
2	D	4	0	6	0	0
2	E	8	0	10	2	0
3	F	1	0	0	0	0
4	A	369	0	0	5	2
4	B	326	0	0	9	1
4	C	299	0	0	14	2
4	D	282	0	0	5	0
4	E	314	0	0	5	1
4	F	293	0	0	12	0
All	All	18342	0	15472	142	3

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.

The worst 5 of 142 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:F:71[B]:GLU:OE2	4:F:501:HOH:O	1.85	0.95
1:C:218:ASN:OD1	4:C:501:HOH:O	1.84	0.93
1:C:115[A]:ARG:NH2	4:C:502:HOH:O	2.03	0.90
1:A:225[A]:ASP:OD1	1:A:252:ARG:CZ	2.20	0.90
1:B:309[B]:TYR:HE1	1:E:312:GLN:HE22	1.24	0.85

All (3) 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)
4:A:638:HOH:O	4:C:515:HOH:O[1_455]	2.13	0.07
4:A:638:HOH:O	4:C:574:HOH:O[1_455]	2.14	0.06
4:B:565:HOH:O	4:E:704:HOH:O[1_655]	2.14	0.06

## 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	346/350 (99%)	334 (96%)	12 (4%)	0	100	100
1	B	339/350 (97%)	327 (96%)	12 (4%)	0	100	100
1	C	341/350 (97%)	327 (96%)	13 (4%)	1 (0%)	41	22
1	D	335/350 (96%)	323 (96%)	12 (4%)	0	100	100
1	E	346/350 (99%)	333 (96%)	13 (4%)	0	100	100
1	F	310/350 (89%)	300 (97%)	10 (3%)	0	100	100
All	All	2017/2100 (96%)	1944 (96%)	72 (4%)	1 (0%)	51	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	173	THR

### 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	295/295 (100%)	293 (99%)	2 (1%)	84	73
1	B	290/295 (98%)	287 (99%)	3 (1%)	76	62
1	C	294/295 (100%)	288 (98%)	6 (2%)	55	32
1	D	288/295 (98%)	287 (100%)	1 (0%)	92	88
1	E	296/295 (100%)	290 (98%)	6 (2%)	55	32
1	F	265/295 (90%)	260 (98%)	5 (2%)	57	34
All	All	1728/1770 (98%)	1705 (99%)	23 (1%)	71	50

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	252	ARG
1	E	309[B]	TYR
1	E	309[A]	TYR
1	F	216	TRP
1	C	170	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	130	HIS
1	E	337	ASN
1	E	159	HIS
1	E	261	ASN
1	F	130	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 11 ligands modelled in this entry, 1 is monoatomic - leaving 10 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	EDO	C	402	-	3,3,3	0.16	0	2,2,2	0.34	0
2	EDO	B	402	-	3,3,3	2.17	1 (33%)	2,2,2	0.32	0
2	EDO	A	401	-	3,3,3	0.80	0	2,2,2	0.42	0
2	EDO	E	401	-	3,3,3	0.73	0	2,2,2	0.21	0
2	EDO	A	402	-	3,3,3	0.55	0	2,2,2	0.56	0
2	EDO	B	401	-	3,3,3	0.52	0	2,2,2	0.56	0
2	EDO	C	401	-	3,3,3	0.25	0	2,2,2	0.13	0
2	EDO	E	402	-	3,3,3	3.25	2 (66%)	2,2,2	1.07	0
2	EDO	D	401	-	3,3,3	0.69	0	2,2,2	0.24	0
2	EDO	B	403	-	3,3,3	1.93	1 (33%)	2,2,2	0.58	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
2	EDO	C	402	-	-	1/1/1/1	-
2	EDO	B	402	-	-	1/1/1/1	-
2	EDO	A	401	-	-	0/1/1/1	-
2	EDO	E	401	-	-	0/1/1/1	-

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	A	402	-	-	0/1/1/1	-
2	EDO	B	401	-	-	0/1/1/1	-
2	EDO	C	401	-	-	0/1/1/1	-
2	EDO	E	402	-	-	1/1/1/1	-
2	EDO	D	401	-	-	0/1/1/1	-
2	EDO	B	403	-	-	0/1/1/1	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	402	EDO	O1-C1	-4.70	1.17	1.42
2	B	402	EDO	O1-C1	-3.46	1.24	1.42
2	B	403	EDO	O1-C1	-2.61	1.28	1.42
2	E	402	EDO	O2-C2	-2.49	1.29	1.42

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	402	EDO	O1-C1-C2-O2
2	C	402	EDO	O1-C1-C2-O2
2	B	402	EDO	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	402	EDO	2	0
2	A	401	EDO	2	0
2	E	402	EDO	2	0
2	B	403	EDO	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	341/350 (97%)	-0.40	3 (0%) 84 86	8, 13, 27, 60	0
1	B	338/350 (96%)	-0.40	3 (0%) 84 86	8, 14, 34, 71	0
1	C	335/350 (95%)	-0.11	14 (4%) 36 35	8, 15, 43, 135	0
1	D	335/350 (95%)	-0.29	4 (1%) 79 81	9, 16, 33, 82	0
1	E	347/350 (99%)	-0.33	6 (1%) 70 73	8, 14, 32, 72	0
1	F	309/350 (88%)	-0.23	8 (2%) 56 56	8, 15, 36, 80	0
All	All	2005/2100 (95%)	-0.29	38 (1%) 66 69	8, 14, 35, 135	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	170	ASP	12.3
1	D	362	ALA	12.0
1	C	173	THR	10.6
1	F	362	ALA	7.8
1	C	171	ILE	7.5

### 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](#)

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
2	EDO	C	402	4/4	0.85	0.13	36,38,38,48	0
2	EDO	A	401	4/4	0.90	0.18	28,29,30,31	0
2	EDO	D	401	4/4	0.94	0.09	14,15,15,16	0
2	EDO	B	402	4/4	0.95	0.16	7,14,18,25	0
2	EDO	B	403	4/4	0.95	0.13	13,19,21,25	0
2	EDO	C	401	4/4	0.96	0.11	17,22,23,25	0
2	EDO	B	401	4/4	0.97	0.07	12,13,13,13	0
2	EDO	E	401	4/4	0.97	0.06	12,15,15,16	0
2	EDO	E	402	4/4	0.97	0.10	13,14,14,16	0
2	EDO	A	402	4/4	0.98	0.08	10,11,12,12	0
3	MG	F	401	1/1	0.99	0.13	16,16,16,16	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.