



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 09:29 am BST

PDB ID : 5OSS
Title : Beta-glucosidase from *Thermotoga maritima* in complex with Gluco-1H-imidazole
Authors : Offen, W.A.; Schroeder, S.P.; Davies, G.J.; Overkleeft, H.S.
Deposited on : 2017-08-18
Resolution : 1.70 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

<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

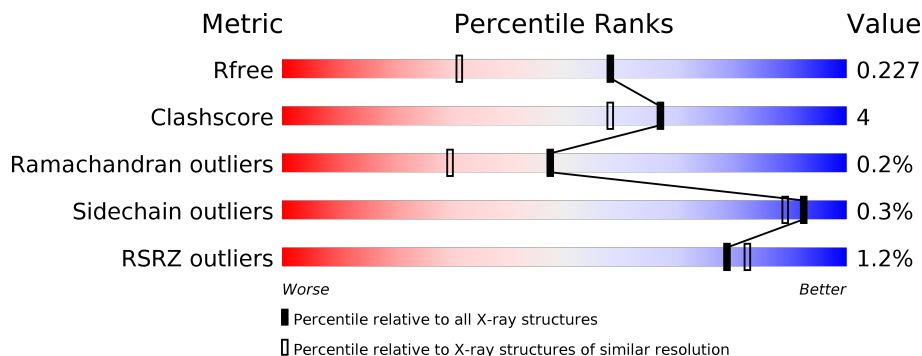
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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 $\leq 5\%$. 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	468	
1	B	468	

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	A	502	-	-	X	-
2	EDO	A	503	-	-	X	-
2	EDO	A	505	-	-	X	-

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 7660 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 Beta-glucosidase A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	443	3632	2365	612	649	6	0	6	0
1	B	443	3599	2341	609	643	6	0	6	0

There are 46 discrepancies between the modelled and reference sequences:

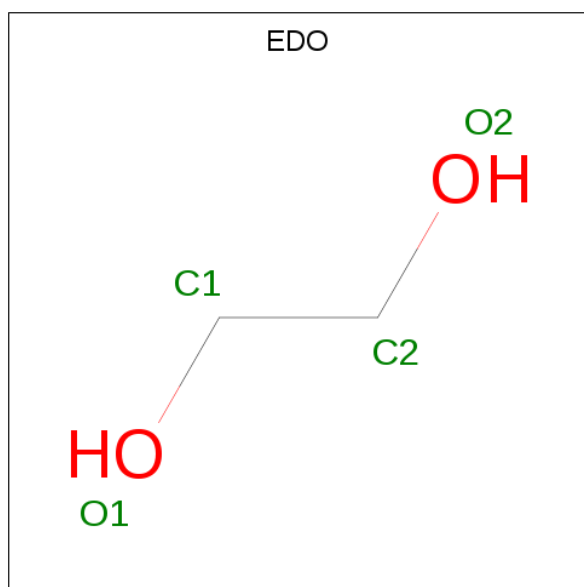
Chain	Residue	Modelled	Actual	Comment	Reference
A	-21	MET	-	initiating methionine	UNP Q08638
A	-20	GLY	-	expression tag	UNP Q08638
A	-19	SER	-	expression tag	UNP Q08638
A	-18	SER	-	expression tag	UNP Q08638
A	-17	HIS	-	expression tag	UNP Q08638
A	-16	HIS	-	expression tag	UNP Q08638
A	-15	HIS	-	expression tag	UNP Q08638
A	-14	HIS	-	expression tag	UNP Q08638
A	-13	HIS	-	expression tag	UNP Q08638
A	-12	HIS	-	expression tag	UNP Q08638
A	-11	SER	-	expression tag	UNP Q08638
A	-10	SER	-	expression tag	UNP Q08638
A	-9	GLY	-	expression tag	UNP Q08638
A	-8	LEU	-	expression tag	UNP Q08638
A	-7	VAL	-	expression tag	UNP Q08638
A	-6	PRO	-	expression tag	UNP Q08638
A	-5	ARG	-	expression tag	UNP Q08638
A	-4	GLY	-	expression tag	UNP Q08638
A	-3	SER	-	expression tag	UNP Q08638
A	-2	HIS	-	expression tag	UNP Q08638
A	-1	MET	-	expression tag	UNP Q08638
A	0	ALA	-	expression tag	UNP Q08638
A	1	SER	-	expression tag	UNP Q08638
B	-21	MET	-	initiating methionine	UNP Q08638
B	-20	GLY	-	expression tag	UNP Q08638

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-19	SER	-	expression tag	UNP Q08638
B	-18	SER	-	expression tag	UNP Q08638
B	-17	HIS	-	expression tag	UNP Q08638
B	-16	HIS	-	expression tag	UNP Q08638
B	-15	HIS	-	expression tag	UNP Q08638
B	-14	HIS	-	expression tag	UNP Q08638
B	-13	HIS	-	expression tag	UNP Q08638
B	-12	HIS	-	expression tag	UNP Q08638
B	-11	SER	-	expression tag	UNP Q08638
B	-10	SER	-	expression tag	UNP Q08638
B	-9	GLY	-	expression tag	UNP Q08638
B	-8	LEU	-	expression tag	UNP Q08638
B	-7	VAL	-	expression tag	UNP Q08638
B	-6	PRO	-	expression tag	UNP Q08638
B	-5	ARG	-	expression tag	UNP Q08638
B	-4	GLY	-	expression tag	UNP Q08638
B	-3	SER	-	expression tag	UNP Q08638
B	-2	HIS	-	expression tag	UNP Q08638
B	-1	MET	-	expression tag	UNP Q08638
B	0	ALA	-	expression tag	UNP Q08638
B	1	SER	-	expression tag	UNP Q08638

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



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

Continued on next page...

Continued from previous page...

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	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 8 4 4	0	1
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

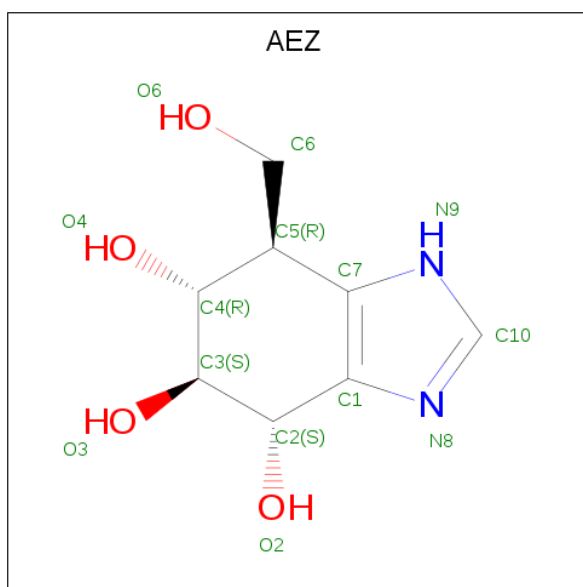
- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Cl 1 1	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Na 1 1	0	0

- Molecule 5 is (4 {S},5 {S},6 {R},7 {R})-7-(hydroxymethyl)-4,5,6,7-tetrahydro-1 {H}-benzimidazole-4,5,6-triol (three-letter code: AEZ) (formula: C₈H₁₂N₂O₄).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	B	1	14	8	2	4	0	0

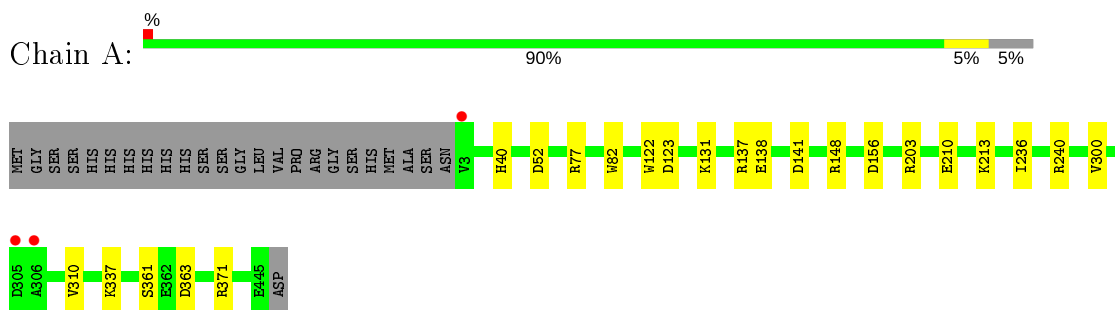
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	224	224	224	0	0
6	B	143	145	145	0	2

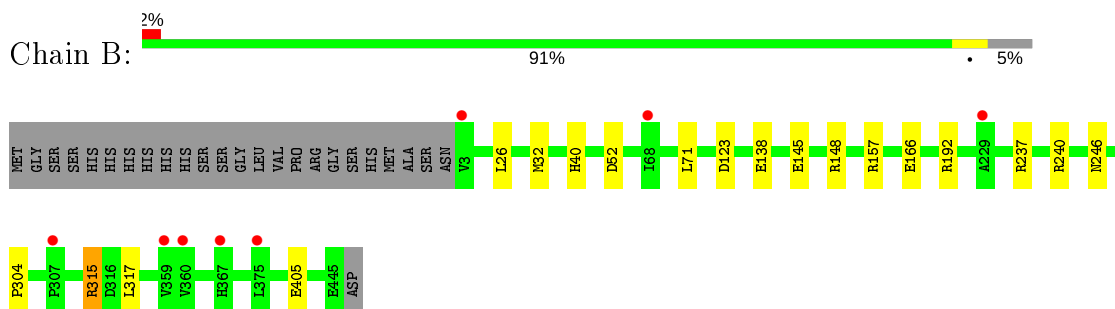
3 Residue-property plots [i](#)

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: Beta-glucosidase A



- Molecule 1: Beta-glucosidase A



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	94.32Å 94.62Å 113.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	72.77 – 1.70 72.66 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (72.77-1.70) 99.9 (72.66-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.00 (at 1.70Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.186 , 0.222 0.197 , 0.227	Depositor DCC
R_{free} test set	5629 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	32.8	Xtrriage
Anisotropy	0.039	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 44.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.000 for k,h,-l	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	7660	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NA, AEZ, EDO, CL

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	0.98	0/3751	0.95	6/5103 (0.1%)
1	B	0.89	0/3716	0.88	5/5061 (0.1%)
All	All	0.93	0/7467	0.91	11/10164 (0.1%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	123	ASP	CB-CG-OD2	-7.03	111.98	118.30
1	B	192	ARG	NE-CZ-NH1	6.04	123.32	120.30
1	A	137	ARG	NE-CZ-NH1	5.96	123.28	120.30
1	A	371	ARG	NE-CZ-NH1	5.52	123.06	120.30
1	B	157	ARG	NE-CZ-NH2	-5.50	117.55	120.30

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	3632	0	3455	42	0
1	B	3599	0	3370	14	0
2	A	32	0	47	28	0
2	B	12	0	18	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
4	B	1	0	0	0	0
5	B	14	0	0	2	0
6	A	224	0	0	6	0
6	B	145	0	0	8	0
All	All	7660	0	6890	56	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 4.

The worst 5 of 56 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:138[B]:GLU:OE2	6:B:602:HOH:O	1.61	1.16
1:A:337:LYS:NZ	2:A:505:EDO:C2	2.15	1.10
1:A:148[A]:ARG:HE	2:A:502:EDO:H22	1.19	1.08
1:A:203[B]:ARG:HG3	1:A:203[B]:ARG:NH2	1.65	1.00
1:A:131[A]:LYS:CD	2:A:503:EDO:H21	1.90	0.99

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	447/468 (96%)	437 (98%)	8 (2%)	2 (0%)	34	18
1	B	447/468 (96%)	435 (97%)	12 (3%)	0	100	100
All	All	894/936 (96%)	872 (98%)	20 (2%)	2 (0%)	47	30

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	213	LYS
1	A	122	TRP

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	367/399 (92%)	367 (100%)	0	100 100
1	B	356/399 (89%)	354 (99%)	2 (1%)	86 80
All	All	723/798 (91%)	721 (100%)	2 (0%)	92 89

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

Mol	Chain	Res	Type
1	B	71	LEU
1	B	246	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	40	HIS
1	B	246	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](#)

Of 14 ligands modelled in this entry, 2 are monoatomic - leaving 12 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	A	504	-	3,3,3	0.45	0	2,2,2	0.54	0
2	EDO	A	507[B]	-	3,3,3	0.60	0	2,2,2	0.26	0
2	EDO	B	501	-	3,3,3	0.37	0	2,2,2	0.33	0
2	EDO	A	507[A]	-	3,3,3	0.65	0	2,2,2	0.45	0
2	EDO	A	506	-	3,3,3	0.26	0	2,2,2	0.88	0
2	EDO	A	503	-	3,3,3	0.44	0	2,2,2	0.52	0
2	EDO	A	502	-	3,3,3	0.48	0	2,2,2	0.37	0
5	AEZ	B	505	-	10,15,15	1.51	2 (20%)	9,22,22	1.91	3 (33%)
2	EDO	B	502	-	3,3,3	0.44	0	2,2,2	0.38	0
2	EDO	A	505	-	3,3,3	0.54	0	2,2,2	0.09	0
2	EDO	A	501	-	3,3,3	0.42	0	2,2,2	0.92	0
2	EDO	B	503	-	3,3,3	0.36	0	2,2,2	0.88	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	A	504	-	-	0/1/1/1	-
2	EDO	A	507[B]	-	-	1/1/1/1	-
2	EDO	B	501	-	-	0/1/1/1	-
2	EDO	A	507[A]	-	-	1/1/1/1	-
2	EDO	A	506	-	-	1/1/1/1	-
2	EDO	A	503	-	-	0/1/1/1	-
2	EDO	A	502	-	-	1/1/1/1	-
5	AEZ	B	505	-	-	0/2/22/22	0/2/2/2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	B	502	-	-	0/1/1/1	-
2	EDO	A	505	-	-	1/1/1/1	-
2	EDO	A	501	-	-	1/1/1/1	-
2	EDO	B	503	-	-	1/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	505	AEZ	C1-C7	3.22	1.48	1.43
5	B	505	AEZ	C7-C5	-2.70	1.47	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	505	AEZ	C4-C3-C2	3.67	115.81	110.24
5	B	505	AEZ	C1-C2-C3	-3.14	106.41	112.60
5	B	505	AEZ	O2-C2-C3	2.19	113.14	108.55

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	507[A]	EDO	O1-C1-C2-O2
2	A	502	EDO	O1-C1-C2-O2
2	B	503	EDO	O1-C1-C2-O2
2	A	506	EDO	O1-C1-C2-O2
2	A	501	EDO	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 30 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	503	EDO	5	0
2	A	502	EDO	11	0
5	B	505	AEZ	2	0
2	A	505	EDO	12	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, 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	443/468 (94%)	-0.21	3 (0%) 87 90	21, 32, 51, 64	27 (6%)
1	B	443/468 (94%)	0.02	8 (1%) 68 72	22, 41, 60, 74	38 (8%)
All	All	886/936 (94%)	-0.09	11 (1%) 79 82	21, 36, 58, 74	65 (7%)

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	307	PRO	4.3
1	B	360	VAL	4.0
1	B	68	ILE	3.7
1	A	3	VAL	3.3
1	B	367	HIS	3.1

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(\AA^2)	Q<0.9
2	EDO	A	503	4/4	0.68	0.23	35,37,39,42	4
2	EDO	A	507[A]	4/4	0.75	0.17	32,33,34,39	4
2	EDO	A	507[B]	4/4	0.75	0.17	32,34,35,40	4
2	EDO	A	504	4/4	0.86	0.26	38,38,39,44	4
2	EDO	A	502	4/4	0.88	0.26	37,38,38,38	4
2	EDO	B	503	4/4	0.88	0.20	34,36,37,39	4
2	EDO	A	505	4/4	0.89	0.20	43,46,48,49	4
2	EDO	A	506	4/4	0.90	0.27	41,42,43,45	4
2	EDO	B	502	4/4	0.90	0.23	34,35,36,37	4
5	AEZ	B	505	14/14	0.91	0.14	29,39,44,48	14
2	EDO	A	501	4/4	0.93	0.12	36,46,48,49	0
2	EDO	B	501	4/4	0.96	0.18	34,40,40,44	4
3	CL	A	508	1/1	0.98	0.06	48,48,48,48	0
4	NA	B	504	1/1	0.99	0.09	31,31,31,31	0

6.5 Other polymers [i](#)

There are no such residues in this entry.