



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 08:23 PM UTC

PDB ID : 4I5U / pdb_00004i5u
Title : Crystal structure of a fungal chimeric cellobiohydrolase Cel6A
Authors : Arnold, F.H.; Wu, I.
Deposited on : 2012-11-29
Resolution : 1.22 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 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

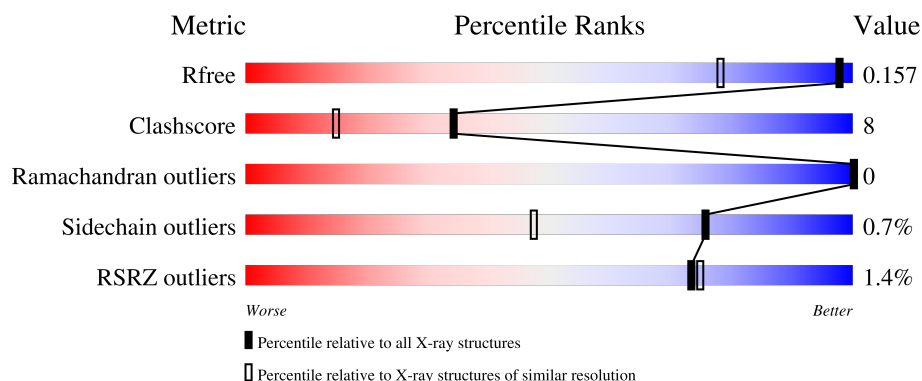
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.22 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}	180053	2002 (1.24-1.20)
Clashscore	190562	2061 (1.24-1.20)
Ramachandran outliers	187476	2009 (1.24-1.20)
Sidechain outliers	187428	2008 (1.24-1.20)
RSRZ outliers	180081	2000 (1.24-1.20)

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 of 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	364	

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
3	EDO	A	507	-	-	X	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	EDO	A	509	-	-	X	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3236 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.

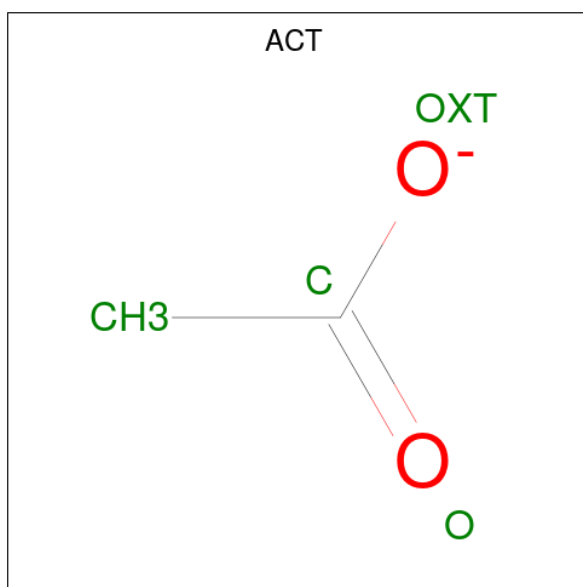
- Molecule 1 is a protein called Chimeric cel6A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	358	Total	C	N	O	S	0	17	0
			2833	1814	477	535	7			

There are 12 discrepancies between the modelled and reference sequences:

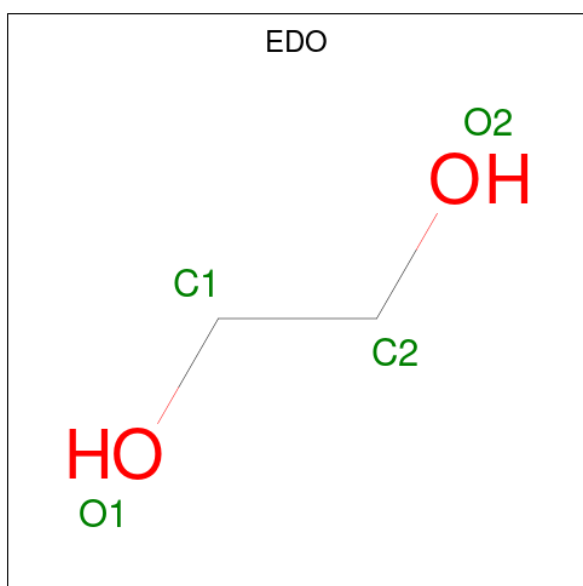
Chain	Residue	Modelled	Actual	Comment	Reference
A	128	ALA	VAL	engineered mutation	UNP Q9C1S9
A	135	LEU	MET	engineered mutation	UNP P07987
A	277	LEU	GLN	engineered mutation	UNP P07987
A	317	PRO	SER	engineered mutation	UNP Q5G2D4
A	406	PRO	SER	engineered mutation	UNP P07987
A	413	PRO	SER	engineered mutation	UNP P07987
A	448	HIS	-	expression tag	UNP P07987
A	449	HIS	-	expression tag	UNP P07987
A	450	HIS	-	expression tag	UNP P07987
A	451	HIS	-	expression tag	UNP P07987
A	452	HIS	-	expression tag	UNP P07987
A	453	HIS	-	expression tag	UNP P07987

- Molecule 2 is ACETATE ION (CCD ID: ACT) (formula: C₂H₃O₂).



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

- Molecule 3 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C₂H₆O₂).



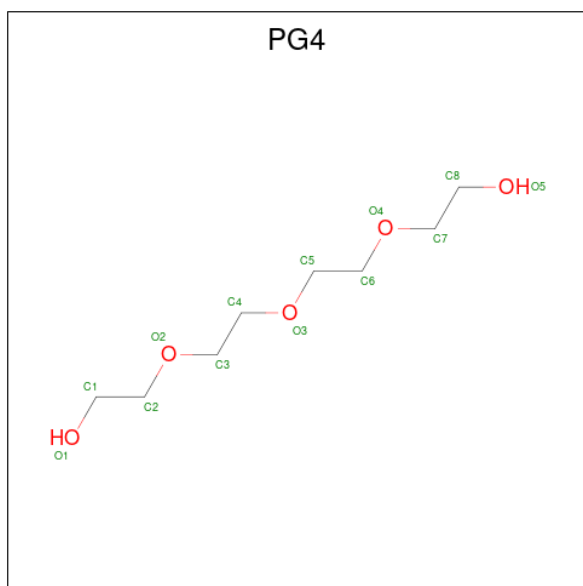
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

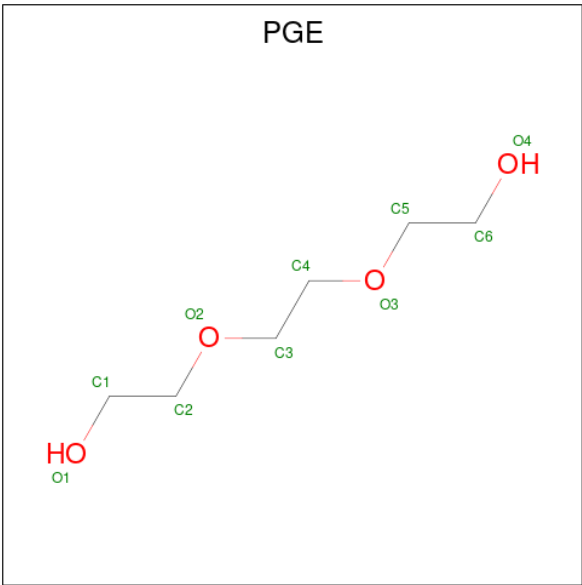
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 4 is TETRAETHYLENE GLYCOL (CCD ID: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	7	4		

- Molecule 5 is TRIETHYLENE GLYCOL (CCD ID: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			10	6	4		
5	A	1	Total	C	O	0	0
			7	4	3		

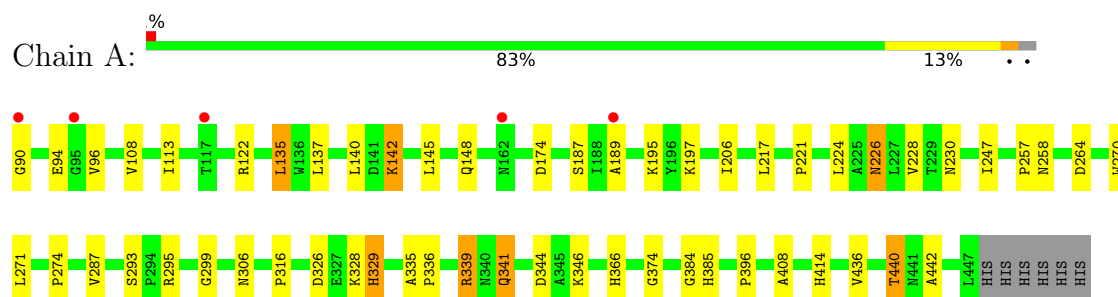
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	327	Total	O	0	0
			327	327		

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: Chimeric cel6A



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	158.60Å 45.41Å 58.19Å 90.00° 95.47° 90.00°	Depositor
Resolution (Å)	35.31 – 1.22 35.31 – 1.22	Depositor EDS
% Data completeness (in resolution range)	87.0 (35.31-1.22) 87.0 (35.31-1.22)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.03	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.77 (at 1.22Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.143 , 0.158 0.142 , 0.157	Depositor DCC
R_{free} test set	5345 reflections (4.35%)	wwPDB-VP
Wilson B-factor (Å ²)	10.3	Xtriage
Anisotropy	0.060	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 43.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	3236	wwPDB-VP
Average B, all atoms (Å ²)	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.13% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: PG4, PGE, EDO, ACT

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	1.70	29/2955 (1.0%)	1.34	8/4050 (0.2%)

The worst 5 of 29 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	440	THR	CB-OG1	-8.20	1.30	1.43
1	A	442	ALA	C-O	7.27	1.32	1.23
1	A	142	LYS	CA-C	-6.95	1.43	1.52
1	A	329	HIS	ND1-CE1	6.69	1.39	1.32
1	A	341	GLN	CG-CD	-6.14	1.36	1.52

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	341	GLN	CA-CB-CG	-6.54	101.02	114.10
1	A	122	ARG	NE-CZ-NH2	5.72	124.35	119.20
1	A	94	GLU	CB-CG-CD	5.42	121.81	112.60
1	A	113	ILE	CA-C-N	-5.23	114.43	119.87
1	A	113	ILE	C-N-CA	-5.23	114.43	119.87

There are no chirality outliers.

There are no planarity outliers.

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	2833	0	2740	33	0
2	A	8	0	6	1	0
3	A	40	0	60	12	0
4	A	11	0	13	6	0
5	A	17	0	23	7	0
6	A	327	0	0	14	0
All	All	3236	0	2842	44	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 8.

The worst 5 of 44 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:306[B]:ASN:ND2	1:A:306[B]:ASN:CG	1.71	1.48
5:A:514:PGE:H42	6:A:709:HOH:O	1.34	1.25
1:A:339:ARG:HE	3:A:509:EDO:H11	1.17	1.08
1:A:195:LYS:HZ2	4:A:513:PG4:H52	1.30	0.97
1:A:197[B]:LYS:NZ	6:A:642:HOH:O	1.98	0.95

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	372/364 (102%)	363 (98%)	9 (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	298/290 (103%)	296 (99%)	2 (1%)	76	47

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

Mol	Chain	Res	Type
1	A	187	SER
1	A	274	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	290	ASN
1	A	329	HIS
1	A	437	GLN
1	A	226	ASN
1	A	205	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry

15 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
3	EDO	A	505	-	3,3,3	0.70	0	2,2,2	0.51	0
4	PG4	A	513	-	10,10,12	1.46	3 (30%)	9,9,11	2.68	5 (55%)
3	EDO	A	508	-	3,3,3	0.70	0	2,2,2	0.91	0
3	EDO	A	507	-	3,3,3	0.67	0	2,2,2	0.92	0
3	EDO	A	510	-	3,3,3	1.25	0	2,2,2	0.79	0
3	EDO	A	503	-	3,3,3	0.56	0	2,2,2	0.41	0
3	EDO	A	509	-	3,3,3	1.99	1 (33%)	2,2,2	0.07	0
3	EDO	A	511	-	3,3,3	0.15	0	2,2,2	1.47	0
5	PGE	A	514	-	9,9,9	1.16	0	8,8,8	1.13	0
5	PGE	A	515	-	6,6,9	0.86	0	5,5,8	0.63	0
3	EDO	A	512	-	3,3,3	0.65	0	2,2,2	0.48	0
2	ACT	A	502	-	3,3,3	1.02	0	3,3,3	1.25	0
3	EDO	A	506	-	3,3,3	0.57	0	2,2,2	0.87	0
3	EDO	A	504	-	3,3,3	0.94	0	2,2,2	1.56	1 (50%)
2	ACT	A	501	-	3,3,3	1.54	1 (33%)	3,3,3	1.05	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
3	EDO	A	505	-	-	0/1/1/1	-
4	PG4	A	513	-	-	0/8/8/10	-
3	EDO	A	508	-	-	0/1/1/1	-
3	EDO	A	507	-	-	1/1/1/1	-
3	EDO	A	510	-	-	1/1/1/1	-
3	EDO	A	503	-	-	1/1/1/1	-
3	EDO	A	509	-	-	1/1/1/1	-
3	EDO	A	511	-	-	1/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	PGE	A	514	-	-	3/7/7/7	-
5	PGE	A	515	-	-	3/4/4/7	-
3	EDO	A	512	-	-	0/1/1/1	-
3	EDO	A	506	-	-	1/1/1/1	-
3	EDO	A	504	-	-	1/1/1/1	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	509	EDO	O1-C1	2.88	1.56	1.42
4	A	513	PG4	O3-C4	2.66	1.53	1.42
4	A	513	PG4	C6-C5	-2.10	1.38	1.49
4	A	513	PG4	O2-C3	2.08	1.52	1.40
2	A	501	ACT	CH3-C	2.01	1.57	1.49

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	513	PG4	O3-C4-C3	5.87	137.13	110.35
4	A	513	PG4	O4-C6-C5	3.15	124.70	110.35
4	A	513	PG4	O3-C5-C6	2.62	122.27	110.35
4	A	513	PG4	O2-C3-C4	2.58	129.91	111.00
3	A	504	EDO	O1-C1-C2	-2.16	95.91	112.39

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	514	PGE	C6-C5-O3-C4
5	A	515	PGE	O2-C3-C4-O3
3	A	509	EDO	O1-C1-C2-O2
3	A	503	EDO	O1-C1-C2-O2
3	A	506	EDO	O1-C1-C2-O2

There are no ring outliers.

6 monomers are involved in 26 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	513	PG4	6	0
3	A	507	EDO	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	509	EDO	4	0
5	A	514	PGE	5	0
5	A	515	PGE	2	0
2	A	501	ACT	1	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, 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	358/364 (98%)	-0.33	5 (1%) 73 75	4, 10, 21, 29	18 (5%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	90[A]	GLY	2.9
1	A	189	ALA	2.3
1	A	117	THR	2.3
1	A	162	ASN	2.2
1	A	95	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 oligosaccharides 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(Å ²)	Q<0.9
3	EDO	A	511	4/4	0.74	0.24	36,38,41,42	0
3	EDO	A	510	4/4	0.77	0.18	29,30,30,37	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\text{\AA}^2$)	Q<0.9
3	EDO	A	512	4/4	0.79	0.20	40,41,41,44	0
5	PGE	A	515	7/10	0.82	0.19	30,41,44,44	0
3	EDO	A	509	4/4	0.84	0.12	18,23,28,29	0
4	PG4	A	513	11/13	0.86	0.15	18,24,32,36	0
3	EDO	A	504	4/4	0.86	0.14	19,21,29,43	0
2	ACT	A	501	4/4	0.88	0.12	25,25,32,40	0
5	PGE	A	514	10/10	0.89	0.12	23,29,38,41	0
3	EDO	A	503	4/4	0.89	0.15	41,45,47,48	0
3	EDO	A	506	4/4	0.90	0.13	35,39,41,43	0
3	EDO	A	507	4/4	0.91	0.14	16,22,34,48	0
3	EDO	A	508	4/4	0.92	0.10	28,30,34,39	0
3	EDO	A	505	4/4	0.92	0.13	14,16,22,22	0
2	ACT	A	502	4/4	0.93	0.09	21,24,27,27	0

6.5 Other polymers ⓘ

There are no such residues in this entry.