



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 29, 2026 – 09:49 PM UTC

PDB ID : 5ZIG / pdb_00005zig
Title : The Structure of cellobiose 2-epimerase from Spirochaeta thermophila DSM 6192
Authors : Feng, Y.H.; Yang, R.J.; Andrew, J.F.
Deposited on : 2018-03-15
Resolution : 2.05 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
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
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

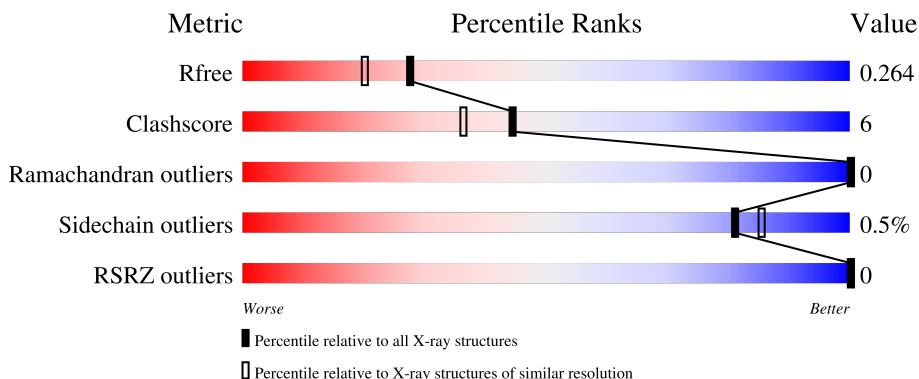
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION





The reported resolution of this entry is 2.05 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	2260 (2.04-2.04)
Clashscore	190562	2333 (2.04-2.04)
Ramachandran outliers	187476	2318 (2.04-2.04)
Sidechain outliers	187428	2318 (2.04-2.04)
RSRZ outliers	180081	2260 (2.04-2.04)

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	404	 90% 9%
1	B	404	 84% 16%
1	C	404	 79% 19% ..
1	D	404	 83% 15% .

2 Entry composition [i](#)

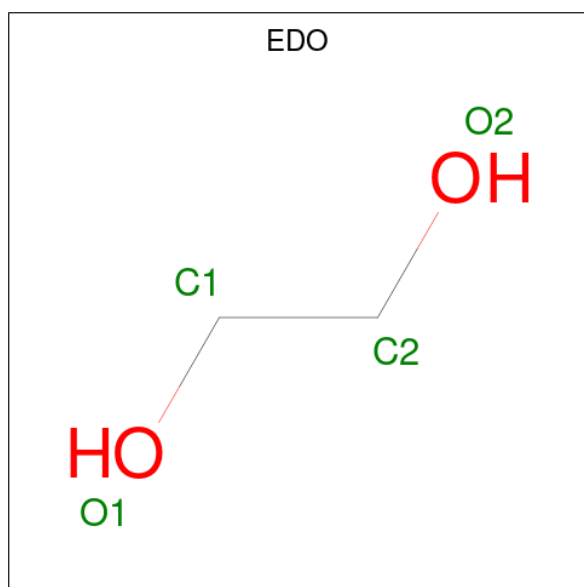
There are 3 unique types of molecules in this entry. The entry contains 13681 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Cellobiose 2-epimerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	404	Total	C	N	O	S	0	4	0
			3293	2114	570	598	11			
1	B	404	Total	C	N	O	S	0	3	0
			3281	2107	566	595	13			
1	C	399	Total	C	N	O	S	0	4	0
			3246	2086	560	589	11			
1	D	398	Total	C	N	O	S	0	4	0
			3244	2083	565	586	10			

- Molecule 2 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C₂H₆O₂) (labeled as "Ligand of Interest" by depositor).



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

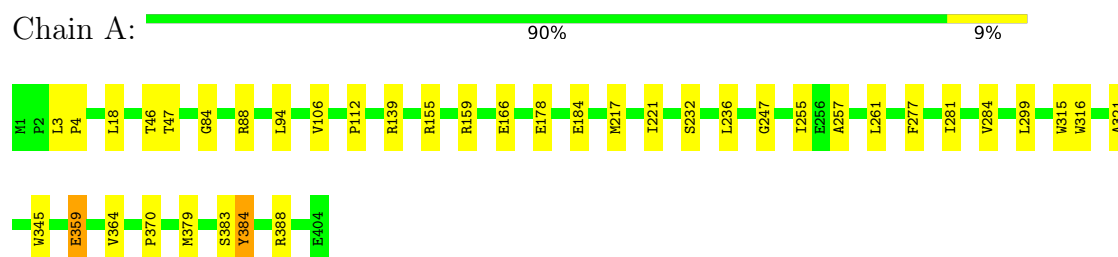
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	174	Total 174	O 174	0	0
3	B	159	Total 159	O 159	0	0
3	C	136	Total 136	O 136	0	0
3	D	140	Total 140	O 140	0	0

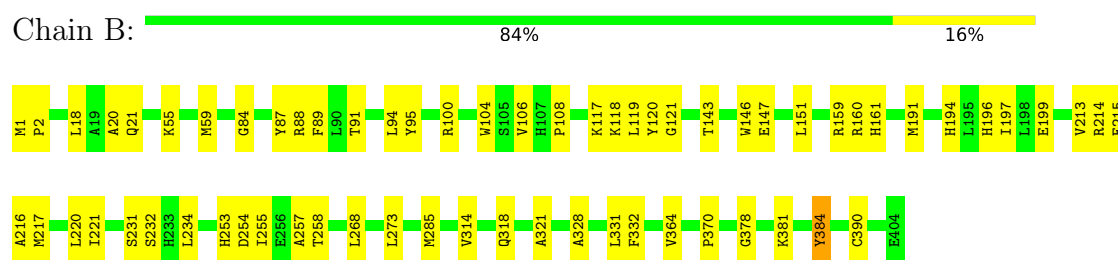
3 Residue-property plots

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.

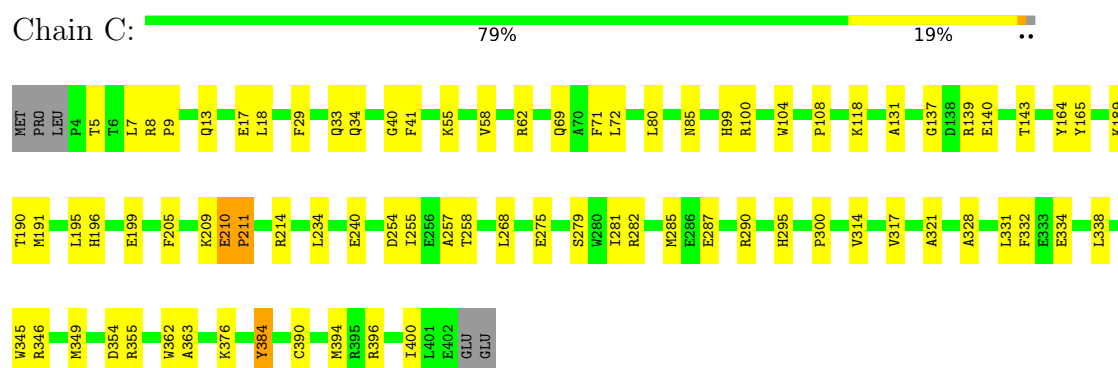
• Molecule 1: Cellobiose 2-epimerase



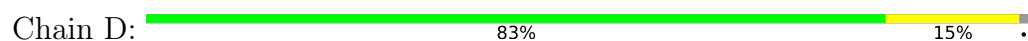
• Molecule 1: Cellobiose 2-epimerase



• Molecule 1: Cellobiose 2-epimerase



• Molecule 1: Cellobiose 2-epimerase



MET	PRO	LEU	PRO	T5	T6	L7	R8	P9	T10	L11	R12	Q13	I14		E17	L18		D50	F51	T52		L57	V58	M59		L94	Y95	D96		R100		V106		H133		L142	T143		L151	L152	E153	D154	R155		Y164		E182		L198		R214	E215	A216	M217	F218	H219	L220	I221	L222
I223		I255		S264		L268		L273		F277		L281		M285	E286	E287	T288	A289	R290		A321		A328		L331		R336		W345		D354		E359	W360		V364		P370		K376		K381	A382	S383	Y384	H385	N386	S387		C390		R395		F402	GLU	GLU			

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	67.53Å 205.16Å 67.72Å 90.00° 104.26° 90.00°	Depositor
Resolution (Å)	32.75 – 2.05 32.75 – 2.05	Depositor EDS
% Data completeness (in resolution range)	86.6 (32.75-2.05) 85.8 (32.75-2.05)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.00 (at 2.05Å)	Xtriage
Refinement program	PHENIX (1.12 _2829)	Depositor
R, R_{free}	0.196 , 0.267 0.195 , 0.264	Depositor DCC
R_{free} test set	4802 reflections (4.31%)	wwPDB-VP
Wilson B-factor (Å ²)	28.9	Xtriage
Anisotropy	0.759	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 26.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.366 for l,-k,h	Xtriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	13681	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.46% 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: EDO

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	0.67	4/3404 (0.1%)	0.62	1/4642 (0.0%)
1	B	0.79	12/3392 (0.4%)	0.68	1/4625 (0.0%)
1	C	0.60	6/3362 (0.2%)	0.57	3/4583 (0.1%)
1	D	0.44	5/3359 (0.1%)	0.54	2/4578 (0.0%)
All	All	0.64	27/13517 (0.2%)	0.60	7/18428 (0.0%)

The worst 5 of 27 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	194	HIS	C-O	-6.46	1.16	1.24
1	B	196	HIS	C-O	-6.41	1.15	1.24
1	A	384	TYR	C-O	-6.33	1.16	1.24
1	B	384	TYR	C-O	-6.28	1.16	1.24
1	B	220	LEU	C-O	-6.26	1.16	1.24

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	384	TYR	N-CA-C	9.16	121.20	111.03
1	A	384	TYR	N-CA-C	7.52	120.58	111.40
1	D	384	TYR	N-CA-C	5.75	118.42	111.40
1	C	210	GLU	CA-C-N	5.74	125.42	119.05
1	C	210	GLU	C-N-CA	5.74	125.42	119.05

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	3293	0	3126	23	0
1	B	3281	0	3115	39	0
1	C	3246	0	3082	54	0
1	D	3244	0	3085	40	0
2	A	4	0	6	0	0
2	D	4	0	6	0	0
3	A	174	0	0	4	0
3	B	159	0	0	2	0
3	C	136	0	0	2	0
3	D	140	0	0	2	0
All	All	13681	0	12420	153	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 6.

The worst 5 of 153 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:C:139:ARG:NH1	1:C:143:THR:OG1	1.68	1.27
1:D:50:ASP:OD1	1:D:52:THR:HG22	1.71	0.89
1:D:359:GLU:CD	1:D:360:TRP:H	1.87	0.82
1:D:359:GLU:OE2	1:D:381:LYS:NZ	2.16	0.77
1:C:211:PRO:HD2	3:C:591:HOH:O	1.86	0.75

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	406/404 (100%)	393 (97%)	13 (3%)	0	100	100
1	B	405/404 (100%)	390 (96%)	15 (4%)	0	100	100
1	C	401/404 (99%)	380 (95%)	21 (5%)	0	100	100
1	D	400/404 (99%)	387 (97%)	13 (3%)	0	100	100
All	All	1612/1616 (100%)	1550 (96%)	62 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	337/333 (101%)	334 (99%)	3 (1%)	70	75
1	B	336/333 (101%)	335 (100%)	1 (0%)	86	90
1	C	332/333 (100%)	331 (100%)	1 (0%)	86	90
1	D	331/333 (99%)	330 (100%)	1 (0%)	86	90
All	All	1336/1332 (100%)	1330 (100%)	6 (0%)	81	88

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	215	GLU
1	C	209	LYS
1	D	384	TYR
1	A	359	GLU
1	A	315	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	351	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	107	HIS
1	B	295	HIS
1	C	301	ASN
1	D	13	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 [i](#)

2 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	EDO	A	501	-	3,3,3	0.40	0	2,2,2	0.57	0
2	EDO	D	501	-	3,3,3	0.49	0	2,2,2	0.40	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	501	-	-	1/1/1/1	-
2	EDO	D	501	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

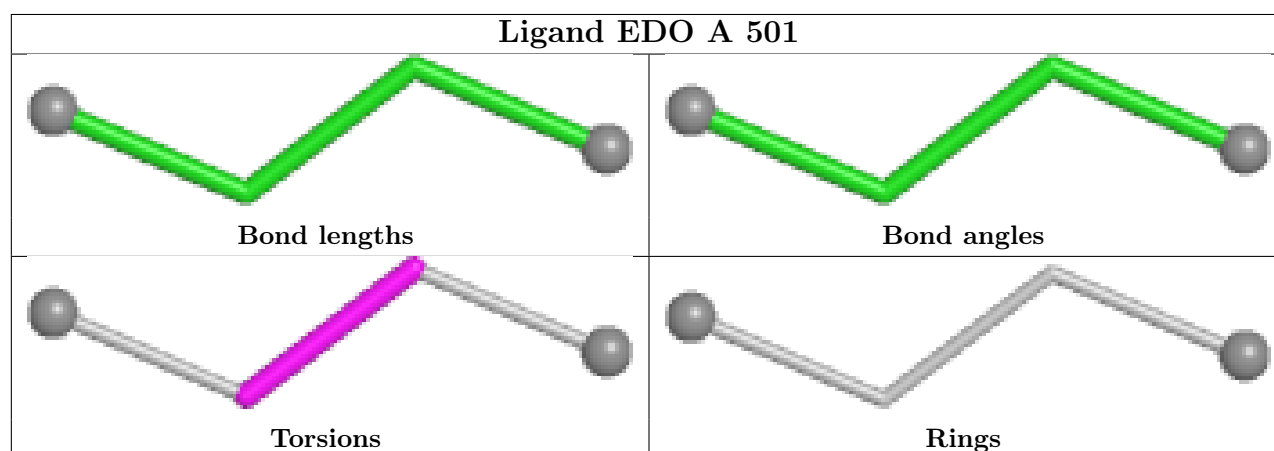
All (1) torsion outliers are listed below:

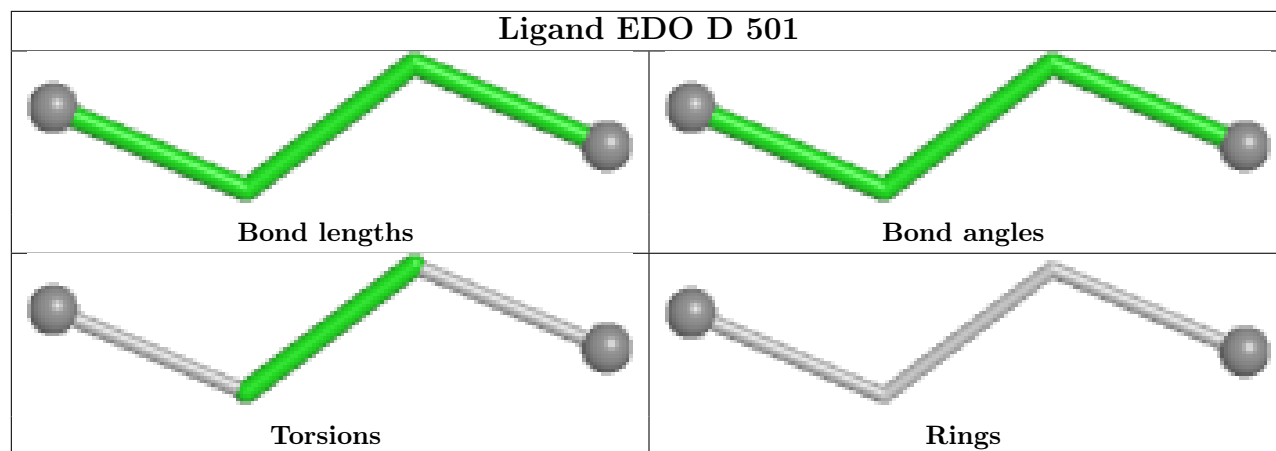
Mol	Chain	Res	Type	Atoms
2	A	501	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





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	404/404 (100%)	-1.46	0 100 100	12, 30, 48, 75	4 (0%)
1	B	404/404 (100%)	-1.47	0 100 100	13, 30, 46, 73	3 (0%)
1	C	399/404 (98%)	-1.42	0 100 100	18, 35, 53, 86	4 (1%)
1	D	398/404 (98%)	-1.40	0 100 100	15, 35, 54, 86	4 (1%)
All	All	1605/1616 (99%)	-1.44	0 100 100	12, 32, 51, 86	15 (0%)

There are no RSRZ outliers to report.

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.

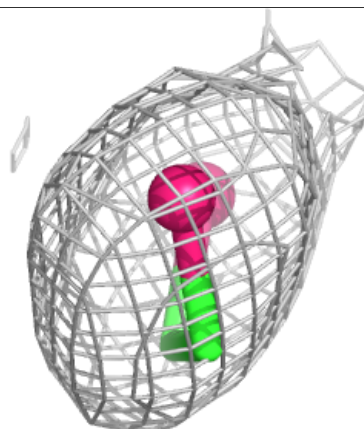
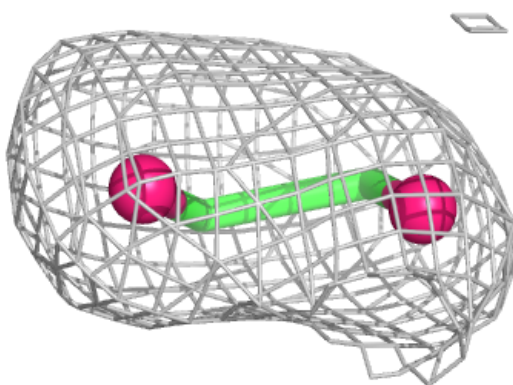
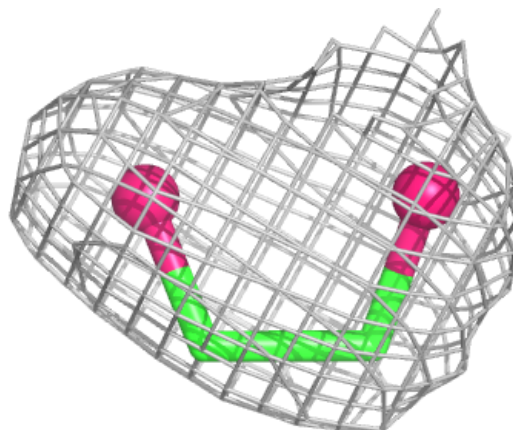
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	EDO	A	501	4/4	0.99	0.05	38,42,42,44	0
2	EDO	D	501	4/4	0.99	0.04	35,40,46,47	0

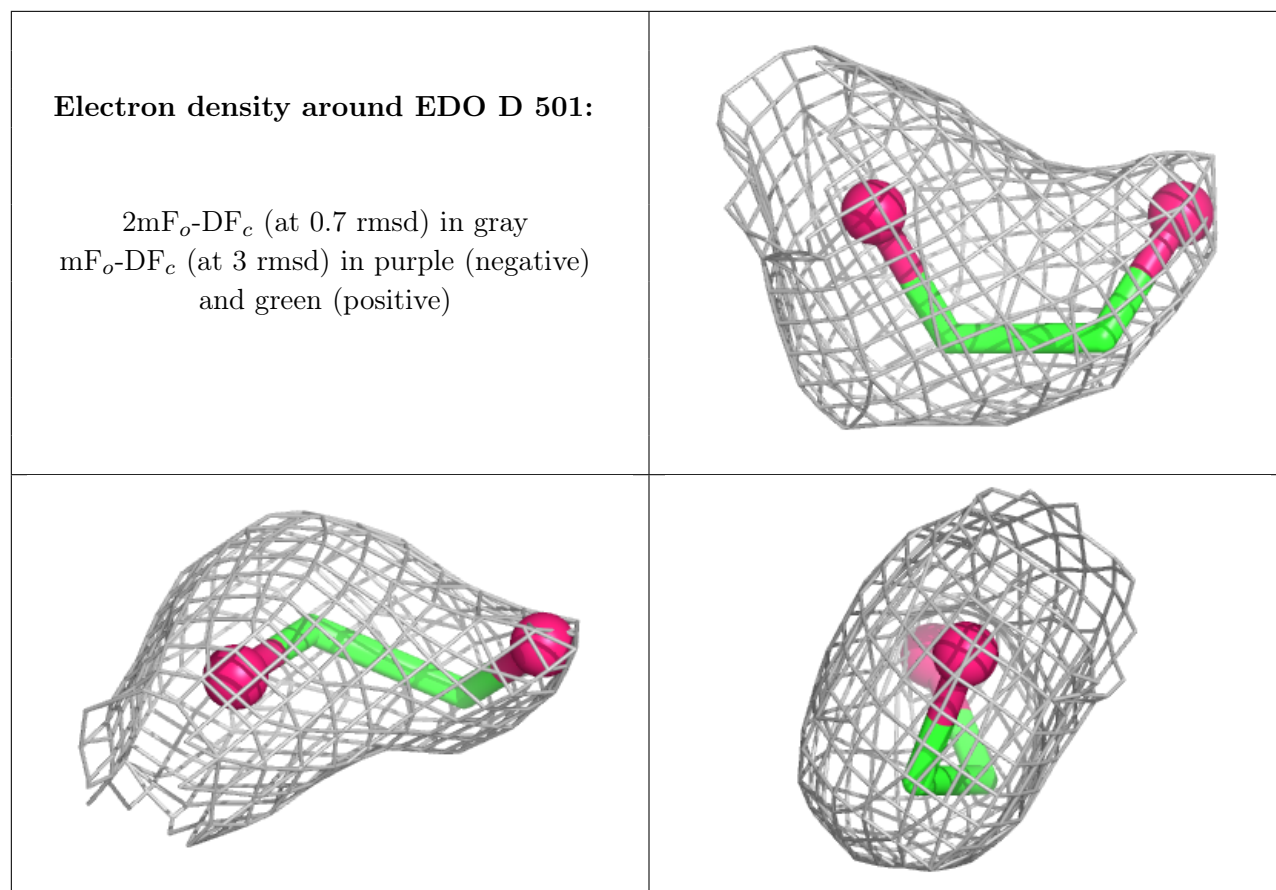
The following is a graphical depiction of the model fit to experimental electron density of all

instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around EDO A 501:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)





6.5 Other polymers [i](#)

There are no such residues in this entry.