Full wwPDB X-ray Structure Validation Report ⓘ

Dec 8, 2023 – 07:48 am GMT

PDB ID	:	8QEF
Title	:	A carbohydrate esterase family 15 (CE15) glucuronoyl esterase from Phocaeicola ATCC 8482 bound to novel ligand.
Authors	:	Banerjee, S.; Poulsen, J.N.; Mazurkewich, S.; Seveso, A.; Larsbrink, J.; Lo Leggio, L.
Deposited on	:	2023-08-31
Resolution	:	2.16 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*
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:

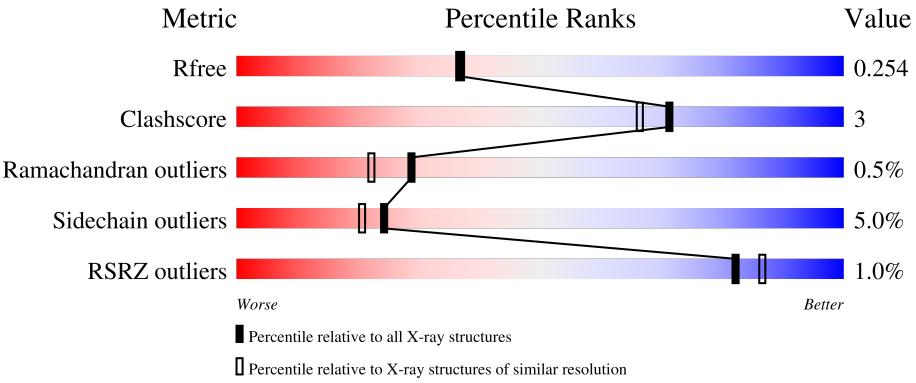
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.16 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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 <=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	426	80% 11% • 8%
1	B	426	2% 80% 11% 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6404 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 Putative acetyl xylan esterase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	390	Total	C	N	O	S	0	7	0
			3161	2027	531	585	18			
1	B	388	Total	C	N	O	S	0	3	0
			3117	1998	522	578	19			

There are 56 discrepancies between the modelled and reference sequences:

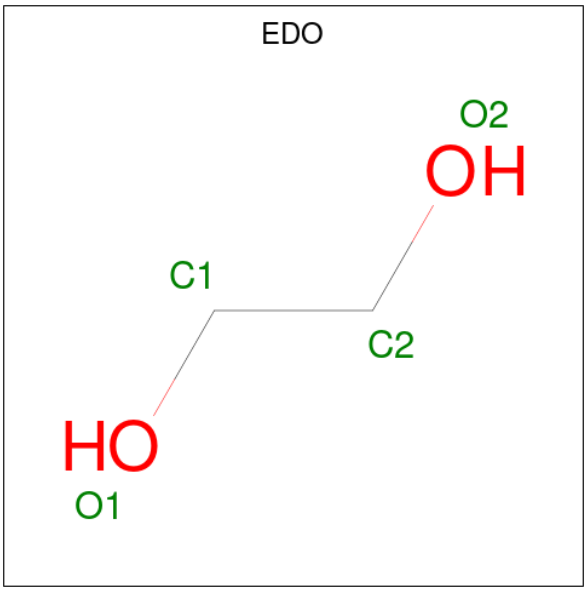
Chain	Residue	Modelled	Actual	Comment	Reference
A	2	MET	-	initiating methionine	UNP A6KWT9
A	3	GLY	-	expression tag	UNP A6KWT9
A	4	SER	-	expression tag	UNP A6KWT9
A	5	SER	-	expression tag	UNP A6KWT9
A	6	HIS	-	expression tag	UNP A6KWT9
A	7	HIS	-	expression tag	UNP A6KWT9
A	8	HIS	-	expression tag	UNP A6KWT9
A	9	HIS	-	expression tag	UNP A6KWT9
A	10	HIS	-	expression tag	UNP A6KWT9
A	11	HIS	-	expression tag	UNP A6KWT9
A	12	SER	-	expression tag	UNP A6KWT9
A	13	SER	-	expression tag	UNP A6KWT9
A	14	GLU	-	expression tag	UNP A6KWT9
A	15	ASN	-	expression tag	UNP A6KWT9
A	16	LEU	-	expression tag	UNP A6KWT9
A	17	TYR	-	expression tag	UNP A6KWT9
A	18	PHE	-	expression tag	UNP A6KWT9
A	19	GLN	-	expression tag	UNP A6KWT9
A	20	GLY	-	expression tag	UNP A6KWT9
A	21	HIS	-	expression tag	UNP A6KWT9
A	22	SER	-	expression tag	UNP A6KWT9
A	23	LYS	-	expression tag	UNP A6KWT9
A	24	SER	-	expression tag	UNP A6KWT9
A	25	PRO	-	expression tag	UNP A6KWT9
A	26	ARG	-	expression tag	UNP A6KWT9

Continued on next page...

Continued from previous page...

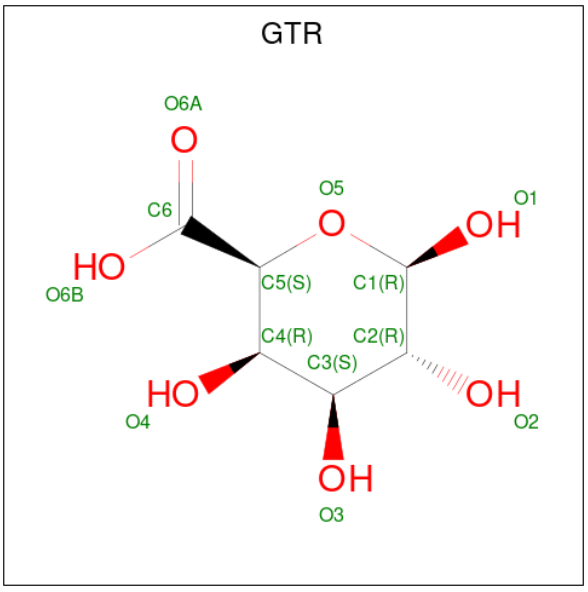
Chain	Residue	Modelled	Actual	Comment	Reference
A	27	LYS	-	expression tag	UNP A6KWT9
A	28	ASP	-	expression tag	UNP A6KWT9
A	29	TYR	-	expression tag	UNP A6KWT9
B	2	MET	-	initiating methionine	UNP A6KWT9
B	3	GLY	-	expression tag	UNP A6KWT9
B	4	SER	-	expression tag	UNP A6KWT9
B	5	SER	-	expression tag	UNP A6KWT9
B	6	HIS	-	expression tag	UNP A6KWT9
B	7	HIS	-	expression tag	UNP A6KWT9
B	8	HIS	-	expression tag	UNP A6KWT9
B	9	HIS	-	expression tag	UNP A6KWT9
B	10	HIS	-	expression tag	UNP A6KWT9
B	11	HIS	-	expression tag	UNP A6KWT9
B	12	SER	-	expression tag	UNP A6KWT9
B	13	SER	-	expression tag	UNP A6KWT9
B	14	GLU	-	expression tag	UNP A6KWT9
B	15	ASN	-	expression tag	UNP A6KWT9
B	16	LEU	-	expression tag	UNP A6KWT9
B	17	TYR	-	expression tag	UNP A6KWT9
B	18	PHE	-	expression tag	UNP A6KWT9
B	19	GLN	-	expression tag	UNP A6KWT9
B	20	GLY	-	expression tag	UNP A6KWT9
B	21	HIS	-	expression tag	UNP A6KWT9
B	22	SER	-	expression tag	UNP A6KWT9
B	23	LYS	-	expression tag	UNP A6KWT9
B	24	SER	-	expression tag	UNP A6KWT9
B	25	PRO	-	expression tag	UNP A6KWT9
B	26	ARG	-	expression tag	UNP A6KWT9
B	27	LYS	-	expression tag	UNP A6KWT9
B	28	ASP	-	expression tag	UNP A6KWT9
B	29	TYR	-	expression tag	UNP A6KWT9

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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	B	1	Total C O 4 2 2	0	0

- Molecule 3 is beta-D-galactopyranuronic acid (three-letter code: GTR) (formula: $C_6H_{10}O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C O 13 6 7	0	0

- Molecule 4 is water.

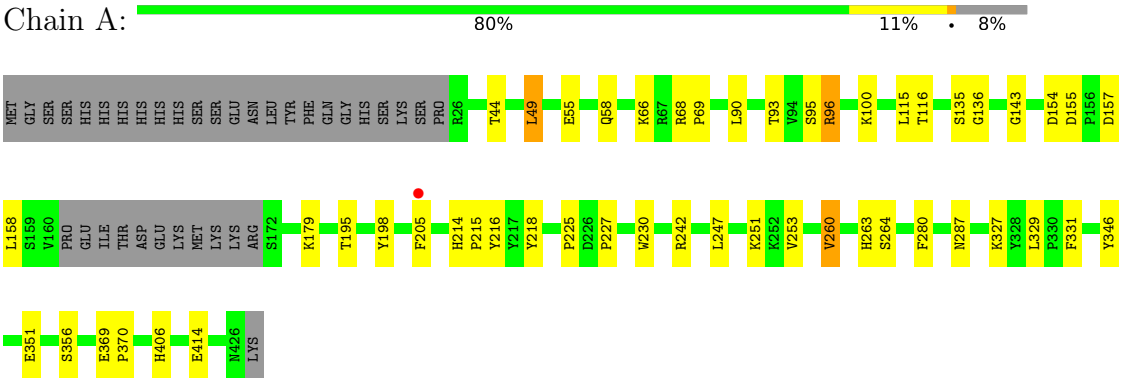
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	61	Total O 62 62	0	1
4	B	34	Total O 35 35	0	1

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.

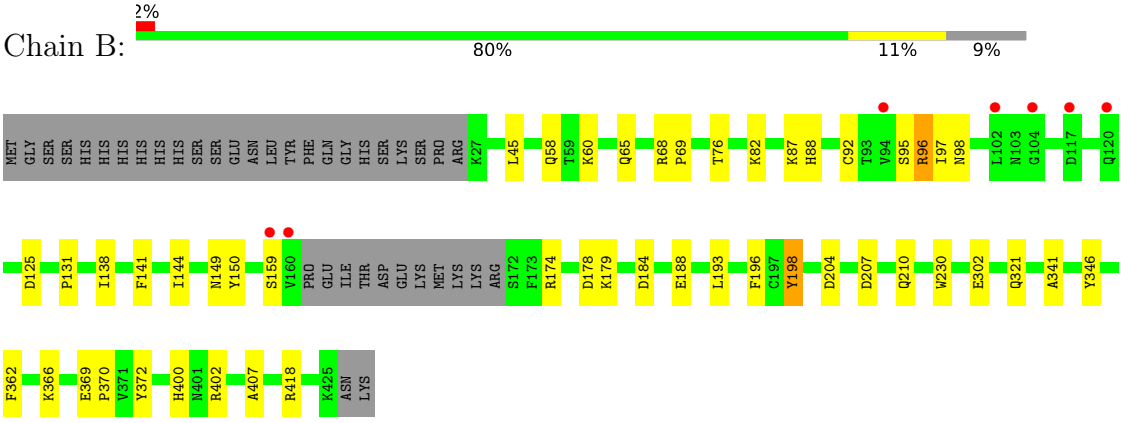
• Molecule 1: Putative acetyl xylan esterase

Chain A:



• Molecule 1: Putative acetyl xylan esterase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	74.22Å 59.16Å 87.16Å 90.00° 101.44° 90.00°	Depositor
Resolution (Å)	85.43 – 2.16 85.43 – 2.16	Depositor EDS
% Data completeness (in resolution range)	70.6 (85.43-2.16) 70.6 (85.43-2.16)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.54 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.179 , 0.252 0.185 , 0.254	Depositor DCC
R_{free} test set	1367 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	30.0	Xtriage
Anisotropy	0.092	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 52.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6404	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.30% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: GTR, 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	0.66	0/3271	0.83	0/4441
1	B	0.66	0/3215	0.81	0/4366
All	All	0.66	0/6486	0.82	0/8807

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	136	GLY	Peptide

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	3161	0	3026	25	0
1	B	3117	0	2961	19	0
2	A	12	0	18	0	0
2	B	4	0	6	0	0
3	B	13	0	9	0	0
4	A	62	0	0	3	0
4	B	35	0	0	2	0
All	All	6404	0	6020	43	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 3.

All (43) 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:215:PRO:HA	1:A:218:TYR:CD2	2.32	0.64
1:A:95:SER:O	1:A:96:ARG:HG2	2.01	0.61
1:A:264[A]:SER:HB3	4:A:611[A]:HOH:O	2.02	0.58
1:A:49:LEU:HB2	4:A:606:HOH:O	2.04	0.56
1:B:302:GLU:O	4:B:609[B]:HOH:O	2.17	0.56
1:A:90:LEU:HD13	1:A:115:LEU:HA	1.89	0.54
1:A:95:SER:C	1:A:96:ARG:HG2	2.28	0.53
1:A:55:GLU:OE2	1:A:66:LYS:CE	2.57	0.52
1:B:362:PHE:CZ	1:B:366:LYS:HD3	2.44	0.52
1:B:184:ASP:O	1:B:188:GLU:HG3	2.10	0.51
1:A:90:LEU:HD12	1:A:242:ARG:HD3	1.93	0.51
1:A:247:LEU:O	1:A:253[A]:VAL:CG2	2.59	0.51
1:A:369:GLU:N	1:A:370:PRO:CD	2.74	0.51
1:B:207:ASP:OD1	1:B:210:GLN:N	2.39	0.51
1:A:55:GLU:OE2	1:A:66:LYS:HE2	2.12	0.49
1:A:215:PRO:HA	1:A:218:TYR:CE2	2.48	0.49
1:B:341:ALA:HB2	1:B:372:TYR:CE1	2.49	0.48
1:A:155:ASP:HB3	1:A:158:LEU:HD12	1.95	0.48
1:B:68:ARG:HB3	1:B:69:PRO:HD3	1.94	0.48
1:B:131:PRO:HD2	1:B:138:ILE:HG21	1.97	0.47
1:B:76:THR:HG22	1:B:82:LYS:HG2	1.97	0.46
1:B:402:ARG:NH2	1:B:407:ALA:O	2.48	0.46
1:A:116:THR:HG21	1:A:216:TYR:CG	2.50	0.46
1:A:143:GLY:O	1:A:195:THR:HA	2.15	0.46
1:A:260:VAL:HG13	1:A:280:PHE:CD1	2.51	0.46
1:B:141:PHE:O	1:B:193:LEU:HD12	2.16	0.45
1:B:369:GLU:N	1:B:370:PRO:CD	2.79	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:329:LEU:HB3	1:A:331:PHE:O	2.17	0.45
1:B:184:ASP:HB2	4:B:631:HOH:O	2.17	0.44
1:B:150:TYR:O	1:B:174:ARG:HA	2.18	0.43
1:A:49:LEU:HD12	1:A:49:LEU:HA	1.89	0.43
1:A:356[B]:SER:HA	4:A:642:HOH:O	2.19	0.43
1:B:95:SER:O	1:B:96:ARG:HB2	2.18	0.43
1:A:263:HIS:CE1	1:A:264[B]:SER:HG	2.36	0.42
1:A:214:HIS:N	1:A:215:PRO:HD2	2.34	0.42
1:A:247:LEU:O	1:A:253[A]:VAL:HG22	2.19	0.42
1:A:68:ARG:HB3	1:A:69:PRO:HD3	2.02	0.41
1:B:125:ASP:O	1:B:149:ASN:ND2	2.53	0.41
1:B:97:ILE:HG22	1:B:98:ASN:N	2.34	0.41
1:B:198:TYR:CE2	1:B:204:ASP:HB2	2.56	0.41
1:A:227:PRO:HB3	1:B:321:GLN:OE1	2.19	0.41
1:B:144:ILE:HD12	1:B:196:PHE:CZ	2.56	0.41
1:A:218:TYR:OH	1:A:225:PRO:HD3	2.21	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	393/426 (92%)	376 (96%)	15 (4%)	2 (0%)	29	22
1	B	387/426 (91%)	365 (94%)	20 (5%)	2 (0%)	29	22
All	All	780/852 (92%)	741 (95%)	35 (4%)	4 (0%)	29	22

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	96	ARG
1	A	406	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	154	ASP
1	B	178	ASP

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	337/366 (92%)	318 (94%)	19 (6%)	21	16
1	B	329/366 (90%)	314 (95%)	15 (5%)	27	23
All	All	666/732 (91%)	632 (95%)	34 (5%)	24	20

All (34) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	44	THR
1	A	49	LEU
1	A	58	GLN
1	A	93	THR
1	A	96	ARG
1	A	100	LYS
1	A	135	SER
1	A	157	ASP
1	A	179	LYS
1	A	198	TYR
1	A	205	PHE
1	A	230	TRP
1	A	251	LYS
1	A	260	VAL
1	A	287	ASN
1	A	327	LYS
1	A	346	TYR
1	A	351	GLU
1	A	414	GLU
1	B	45	LEU
1	B	58	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	60	LYS
1	B	65	GLN
1	B	87	LYS
1	B	88	HIS
1	B	92[A]	CYS
1	B	92[B]	CYS
1	B	159	SER
1	B	179	LYS
1	B	198	TYR
1	B	230	TRP
1	B	346	TYR
1	B	400	HIS
1	B	418	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	73	ASN
1	B	62	GLN
1	B	103	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

5 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	502	-	3,3,3	0.40	0	2,2,2	0.47	0
2	EDO	A	501	-	3,3,3	0.14	0	2,2,2	0.19	0
2	EDO	B	502	-	3,3,3	0.29	0	2,2,2	0.20	0
3	GTR	B	501	-	13,13,13	1.44	2 (15%)	18,19,19	1.05	1 (5%)
2	EDO	A	503	-	3,3,3	0.10	0	2,2,2	0.14	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	502	-	-	0/1/1/1	-
2	EDO	A	501	-	-	0/1/1/1	-
2	EDO	B	502	-	-	0/1/1/1	-
3	GTR	B	501	-	-	0/4/24/24	0/1/1/1
2	EDO	A	503	-	-	0/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	501	GTR	C4-C5	-3.10	1.48	1.53
3	B	501	GTR	O1-C1	2.28	1.46	1.39

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	501	GTR	O4-C4-C5	-2.06	105.12	109.74

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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^{th} 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	390/426 (91%)	-0.11	1 (0%)	94	95	15, 28, 51, 73	0
1	B	388/426 (91%)	0.03	7 (1%)	68	75	18, 38, 64, 85	0
All	All	778/852 (91%)	-0.04	8 (1%)	82	86	15, 32, 61, 85	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	160	VAL	4.0
1	B	120	GLN	2.9
1	B	117	ASP	2.8
1	B	102	LEU	2.7
1	A	205	PHE	2.5
1	B	94	VAL	2.4
1	B	159	SER	2.2
1	B	104	GLY	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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^{th} 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	B	502	4/4	0.88	0.13	39,39,40,44	0
2	EDO	A	503	4/4	0.89	0.19	49,50,50,50	0
3	GTR	B	501	13/13	0.89	0.16	45,54,59,67	0
2	EDO	A	502	4/4	0.93	0.19	22,25,29,30	0
2	EDO	A	501	4/4	0.93	0.08	49,50,51,52	0

6.5 Other polymers (i)

There are no such residues in this entry.