wwPDB X-ray Structure Validation Summary Report (i)

Jan 4, 2024 – 11:06 pm GMT

PDB ID	:	5LDR
Title	:	Crystal structure of a cold-adapted dimeric beta-D-galactosidase from Paracoccus sp. 32d strain in complex with galactose
Authors	:	Rutkiewicz-Krotewicz, M.; Bujacz, A.; Pietrzyk, A.J.; Sekula, B.; Bujacz, G.
Deposited on	:	2016-06-27
Resolution	:	3.15 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 (i) 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 (i)) were used in the production of this report:

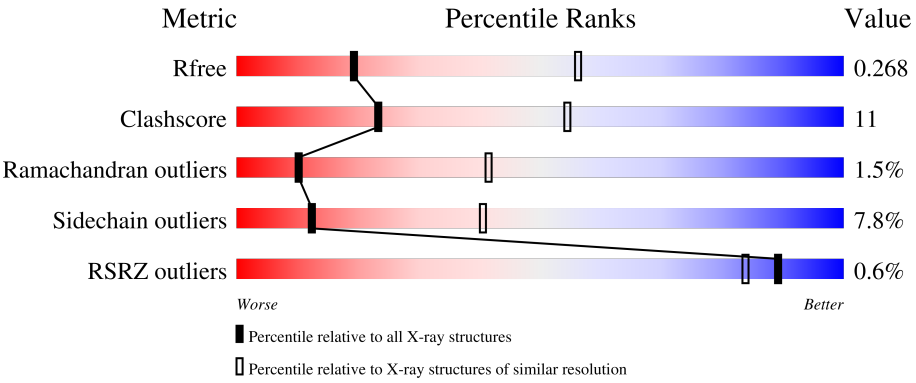
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 3.15 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	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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	732	% 71% 26% .
2	B	731	% 71% 26% .

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
5	PEG	B	805	-	-	X	X

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 11724 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-D-galactosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	731	5764	3632	1038	1068	26	0	0	0

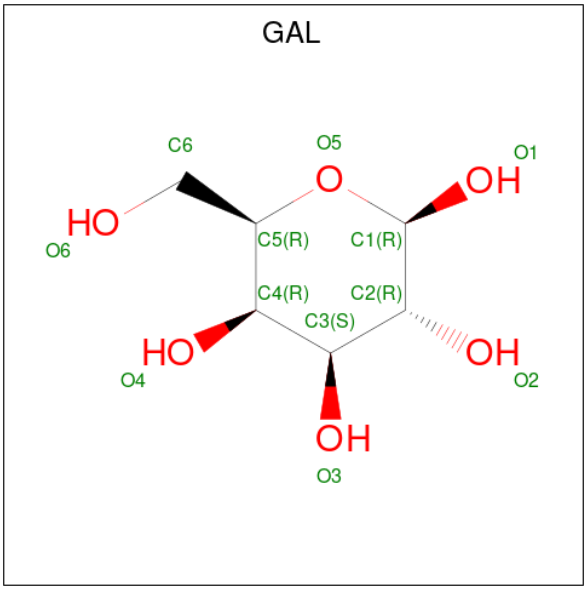
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	ASN	-	expression tag	UNP D1LZK0

- Molecule 2 is a protein called Beta-D-galactosidase.

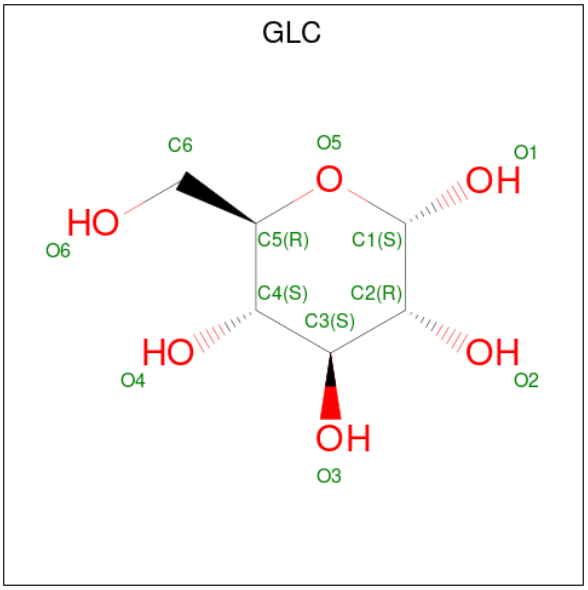
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	731	5763	3632	1038	1067	26	0	0	0

- Molecule 3 is beta-D-galactopyranose (three-letter code: GAL) (formula: $C_6H_{12}O_6$).



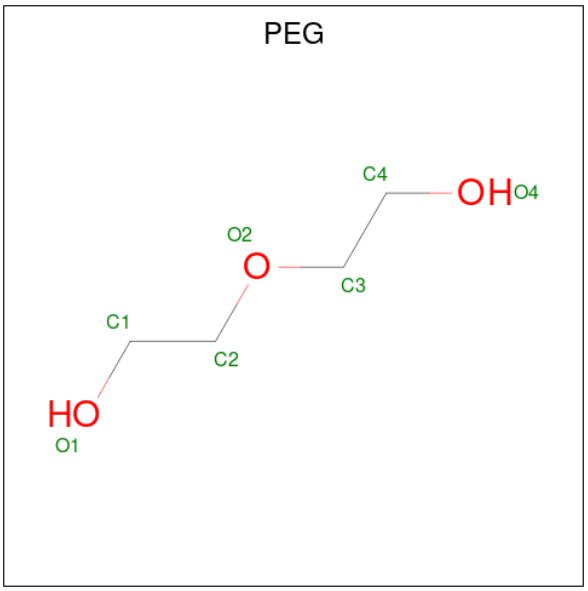
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 12 6 6	0	0
3	B	1	Total C O 12 6 6	0	0

- Molecule 4 is alpha-D-glucopyranose (three-letter code: GLC) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 12 6 6	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

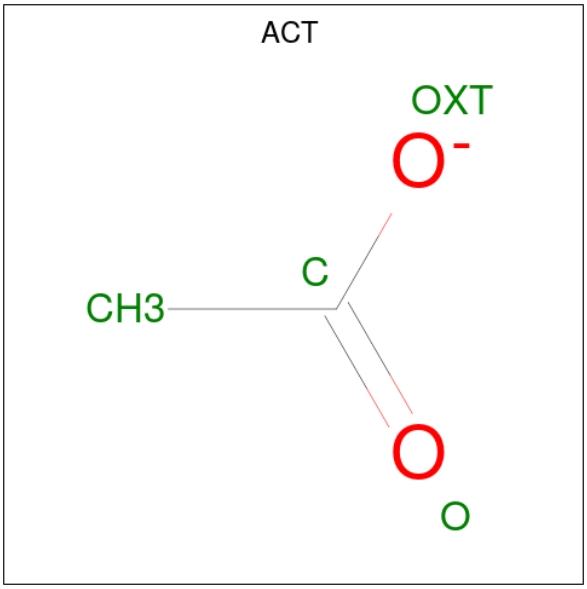


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 7 4 3	0	0
5	B	1	Total C O 7 4 3	0	0
5	B	1	Total C O 7 4 3	0	0
5	B	1	Total C O 7 4 3	0	0
5	B	1	Total C O 7 4 3	0	0

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	2	Total Cl 2 2	0	0
6	B	2	Total Cl 2 2	0	0

- Molecule 7 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



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

- Molecule 8 is water.

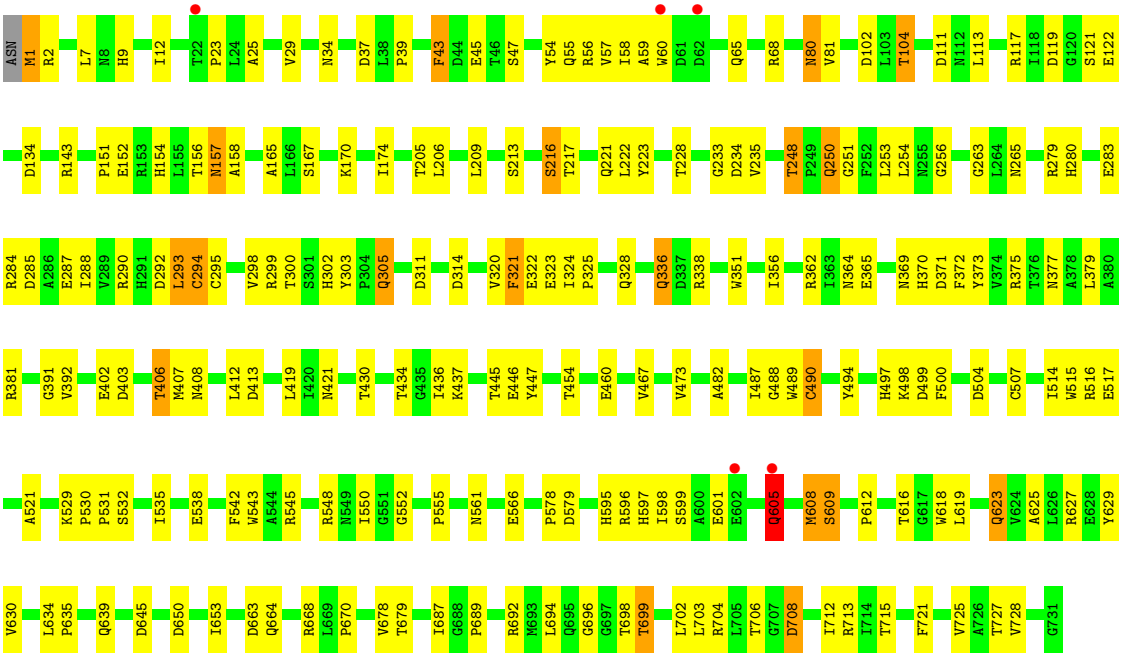
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	66	Total O 66 66	0	0
8	B	48	Total O 48 48	0	0

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.

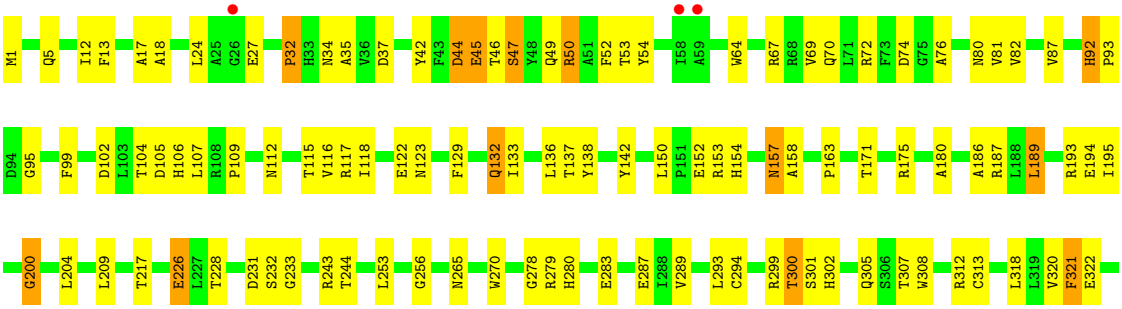
• Molecule 1: Beta-D-galactosidase

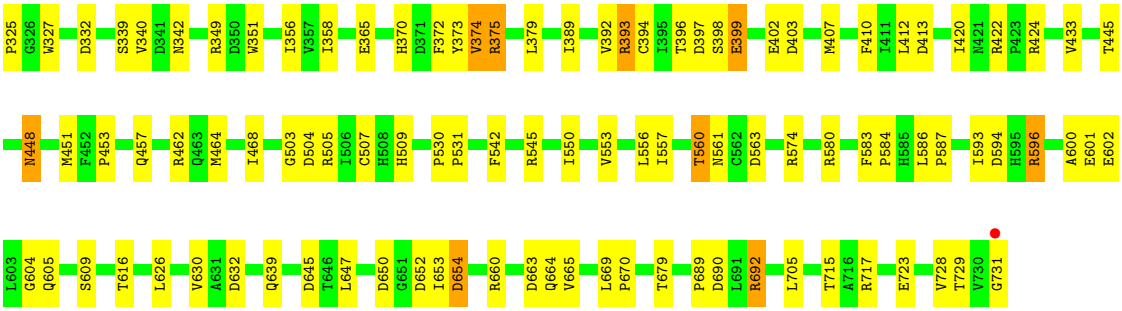
Chain A:



• Molecule 2: Beta-D-galactosidase

Chain B:




P325 G326 W327 D332 S339 V340 D341 N342 R349 D350 W351 I356 V357 I358 E365 H370 D371 F372 Y373 V374 R375 L379 I389 V392 R393 C394 I395 T396 D397 S398 E399 E402 D403 M407 F410 I411 L412 D413 I420 M421 R422 P423 R424 V433 T445
N448 M451 F452 P453 Q457 R462 Q463 M464 I468 G503 D504 R505 I506 C507 H508 H509 P530 P531 F542 R545 I550 V553 L556 I557 T560 N561 C562 D563 R574 R580 F583 P584 H585 L586 P587 I593 D594 H595 R596 A600 E601 E602
L603 G604 Q605 S609 T616 L626 V630 A631 D632 Q639 D645 T646 L647 D650 G651 D652 I653 D654 R660 D663 Q664 V665 L669 P670 T679 P689 D690 L691 R692 L705 T715 A716 R717 E723 V728 T729 V730 G731

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	79.26Å 107.58Å 201.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.27 – 3.15 46.27 – 3.15	Depositor EDS
% Data completeness (in resolution range)	96.7 (46.27-3.15) 96.7 (46.27-3.15)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.184 , 0.267 0.195 , 0.268	Depositor DCC
R_{free} test set	1480 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	65.9	Xtriage
Anisotropy	0.069	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 62.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	11724	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% 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: PEG, CL, ACT, GAL, GLC

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	0.86	0/5919	1.04	0/8084
2	B	0.87	1/5918 (0.0%)	1.03	0/8084
All	All	0.86	1/11837 (0.0%)	1.04	0/16168

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	200	GLY	N-CA	5.59	1.54	1.46

There are no bond angle outliers.

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	5764	0	5591	139	0
2	B	5763	0	5591	122	1
3	A	12	0	12	3	0
3	B	12	0	12	0	0
4	A	12	0	12	0	0
5	A	7	0	10	0	0
5	B	28	0	40	5	0
6	A	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	2	0	0	1	0
7	B	8	0	6	0	0
8	A	66	0	0	4	0
8	B	48	0	0	4	0
All	All	11724	0	11274	259	1

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 11.

The worst 5 of 259 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:678:VAL:HG23	1:A:715:THR:O	1.07	1.23
1:A:678:VAL:CG2	1:A:715:THR:O	2.02	1.07
1:A:157:ASN:HD22	1:A:158:ALA:H	1.19	0.89
1:A:460:GLU:OE2	1:A:548:ARG:NH1	2.12	0.81
1:A:165:ALA:O	1:A:170:LYS:NZ	2.15	0.79

All (1) 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)
2:B:175:ARG:NH1	2:B:679:THR:OG1[3_745]	2.01	0.19

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	729/732 (100%)	621 (85%)	100 (14%)	8 (1%)	14 48
2	B	729/731 (100%)	631 (87%)	84 (12%)	14 (2%)	8 36

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1458/1463 (100%)	1252 (86%)	184 (13%)	22 (2%)	10 41

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	609	SER
2	B	104	THR
2	B	332	ASP
1	A	532	SER
1	A	608	MET

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	611/612 (100%)	561 (92%)	50 (8%)	11 38
2	B	611/611 (100%)	566 (93%)	45 (7%)	13 43
All	All	1222/1223 (100%)	1127 (92%)	95 (8%)	12 40

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	67	ARG
2	B	231	ASP
2	B	74	ASP
2	B	157	ASN
2	B	321	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	377	ASN
1	A	408	ASN
2	B	611	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	282	GLN
2	B	509	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 14 ligands modelled in this entry, 4 are 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	$\#\|Z\| > 2$
5	PEG	A	803	-	6,6,6	0.55	0	5,5,5	0.14	0
4	GLC	A	802	-	12,12,12	1.07	0	17,17,17	1.70	4 (23%)
5	PEG	B	804	-	6,6,6	0.40	0	5,5,5	0.41	0
7	ACT	B	808	-	3,3,3	0.95	0	3,3,3	0.81	0
3	GAL	B	801	-	12,12,12	1.44	2 (16%)	17,17,17	2.77	8 (47%)
3	GAL	A	801	-	12,12,12	1.01	0	17,17,17	1.79	5 (29%)
7	ACT	B	809	-	3,3,3	0.93	0	3,3,3	0.89	0
5	PEG	B	805	-	6,6,6	0.45	0	5,5,5	0.93	0
5	PEG	B	802	-	6,6,6	0.53	0	5,5,5	0.42	0
5	PEG	B	803	-	6,6,6	0.68	0	5,5,5	0.89	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
5	PEG	A	803	-	-	1/4/4/4	-
4	GLC	A	802	-	-	2/2/22/22	0/1/1/1
5	PEG	B	804	-	-	0/4/4/4	-
3	GAL	B	801	-	-	2/2/22/22	0/1/1/1
3	GAL	A	801	-	-	0/2/22/22	0/1/1/1
5	PEG	B	805	-	-	1/4/4/4	-
5	PEG	B	802	-	-	1/4/4/4	-
5	PEG	B	803	-	-	2/4/4/4	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	801	GAL	C6-C5	2.51	1.60	1.51
3	B	801	GAL	C4-C5	2.41	1.58	1.53

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	801	GAL	C1-O5-C5	5.69	124.41	113.66
3	B	801	GAL	C3-C4-C5	5.22	119.55	110.24
3	A	801	GAL	O3-C3-C2	4.75	121.33	110.35
3	B	801	GAL	O5-C5-C6	4.67	118.05	106.44
4	A	802	GLC	C1-O5-C5	4.33	121.82	113.66

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	801	GAL	O5-C5-C6-O6
4	A	802	GLC	O5-C5-C6-O6
4	A	802	GLC	C4-C5-C6-O6
3	B	801	GAL	C4-C5-C6-O6
5	B	802	PEG	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	801	GAL	3	0
5	B	805	PEG	4	0
5	B	802	PEG	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

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	731/732 (99%)	-0.37	5 (0%)	87	81	31, 58, 84, 171	0
2	B	731/731 (100%)	-0.36	4 (0%)	91	86	36, 63, 89, 145	0
All	All	1462/1463 (99%)	-0.37	9 (0%)	89	84	31, 60, 87, 171	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	602	GLU	3.0
2	B	59	ALA	2.9
2	B	58	ILE	2.8
1	A	605	GLN	2.7
1	A	62	ASP	2.7

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
5	PEG	B	805	7/7	0.79	0.69	113,117,121,122	0
5	PEG	B	803	7/7	0.81	0.34	67,71,80,90	0
4	GLC	A	802	12/12	0.82	0.28	72,91,102,103	0
6	CL	B	807	1/1	0.84	0.22	75,75,75,75	0
6	CL	B	806	1/1	0.85	0.14	42,42,42,42	0
5	PEG	A	803	7/7	0.85	0.28	54,66,80,82	0
5	PEG	B	802	7/7	0.86	0.25	59,74,80,80	0
6	CL	A	804	1/1	0.86	0.21	57,57,57,57	0
5	PEG	B	804	7/7	0.92	0.17	67,70,72,72	0
7	ACT	B	809	4/4	0.92	0.22	47,54,55,55	0
7	ACT	B	808	4/4	0.93	0.33	65,66,70,71	0
6	CL	A	805	1/1	0.93	0.24	61,61,61,61	0
3	GAL	A	801	12/12	0.94	0.21	66,75,79,79	0
3	GAL	B	801	12/12	0.94	0.17	49,56,59,59	0

6.5 Other polymers (i)

There are no such residues in this entry.