wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 10:15 AM EDT

PDB ID	:	3QE5
Title	:	Complete structure of Streptococcus mutans Antigen I/II carboxy-terminus
Authors	:	Larson, M.R.; Rajashankar, K.R.; Crowley, P.J.; Kelly, C.; Mitchell, T.J.; Brady, L.J.; Deivanayagam, C.
Deposited on	:	2011-01-19
Resolution	:	2.50 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:

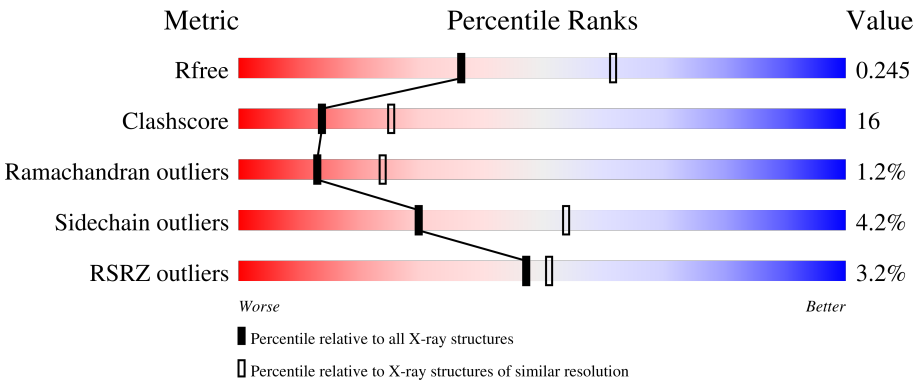
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	516	3% 67% 26% • 6%
1	B	516	3% 63% 28% • 6%

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
4	GLC	A	1496	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8029 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 Major cell-surface adhesin PAc.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	487	3801	2400	629	769	3	0	0	0
1	B	487	3801	2400	629	769	3	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	979	MET	-	expression tag	UNP P11657
A	980	ALA	-	expression tag	UNP P11657
A	981	SER	-	expression tag	UNP P11657
A	982	MET	-	expression tag	UNP P11657
A	983	THR	-	expression tag	UNP P11657
A	984	GLY	-	expression tag	UNP P11657
A	985	GLY	-	expression tag	UNP P11657
A	986	GLN	-	expression tag	UNP P11657
A	987	GLN	-	expression tag	UNP P11657
A	988	MET	-	expression tag	UNP P11657
A	989	GLY	-	expression tag	UNP P11657
A	990	ARG	-	expression tag	UNP P11657
A	991	ILE	-	expression tag	UNP P11657
A	1382	LYS	GLU	cloning artifact	UNP P11657
A	1487	LEU	-	expression tag	UNP P11657
A	1488	GLU	-	expression tag	UNP P11657
A	1489	HIS	-	expression tag	UNP P11657
A	1490	HIS	-	expression tag	UNP P11657
A	1491	HIS	-	expression tag	UNP P11657
A	1492	HIS	-	expression tag	UNP P11657
A	1493	HIS	-	expression tag	UNP P11657
A	1494	HIS	-	expression tag	UNP P11657
B	979	MET	-	expression tag	UNP P11657
B	980	ALA	-	expression tag	UNP P11657
B	981	SER	-	expression tag	UNP P11657

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	982	MET	-	expression tag	UNP P11657
B	983	THR	-	expression tag	UNP P11657
B	984	GLY	-	expression tag	UNP P11657
B	985	GLY	-	expression tag	UNP P11657
B	986	GLN	-	expression tag	UNP P11657
B	987	GLN	-	expression tag	UNP P11657
B	988	MET	-	expression tag	UNP P11657
B	989	GLY	-	expression tag	UNP P11657
B	990	ARG	-	expression tag	UNP P11657
B	991	ILE	-	expression tag	UNP P11657
B	1382	LYS	GLU	cloning artifact	UNP P11657
B	1487	LEU	-	expression tag	UNP P11657
B	1488	GLU	-	expression tag	UNP P11657
B	1489	HIS	-	expression tag	UNP P11657
B	1490	HIS	-	expression tag	UNP P11657
B	1491	HIS	-	expression tag	UNP P11657
B	1492	HIS	-	expression tag	UNP P11657
B	1493	HIS	-	expression tag	UNP P11657
B	1494	HIS	-	expression tag	UNP P11657

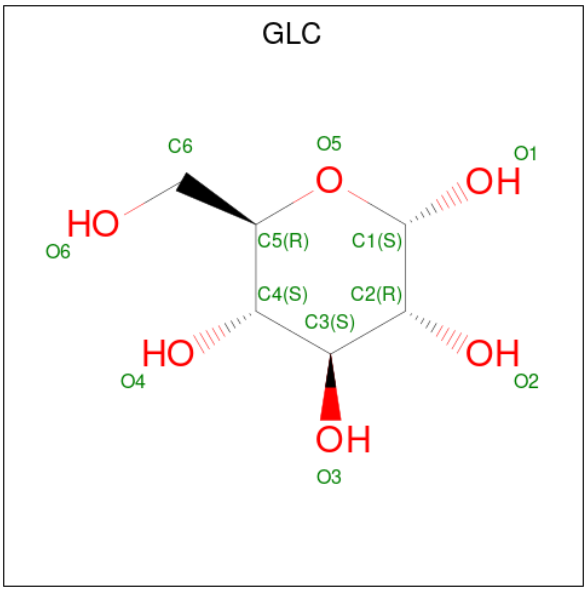
- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
2	A	3	3	3	0	0
			Total	Ca		
2	B	3	3	3	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

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



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

- Molecule 5 is water.

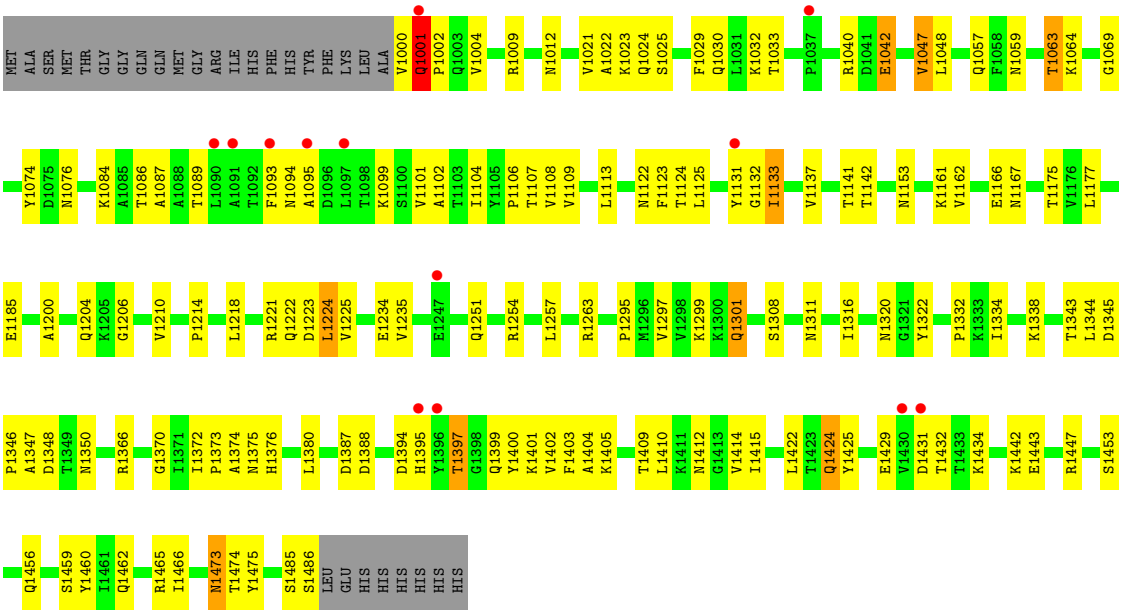
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	189	189	189	0	0
			Total	O		
5	B	135	135	135	0	0

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.

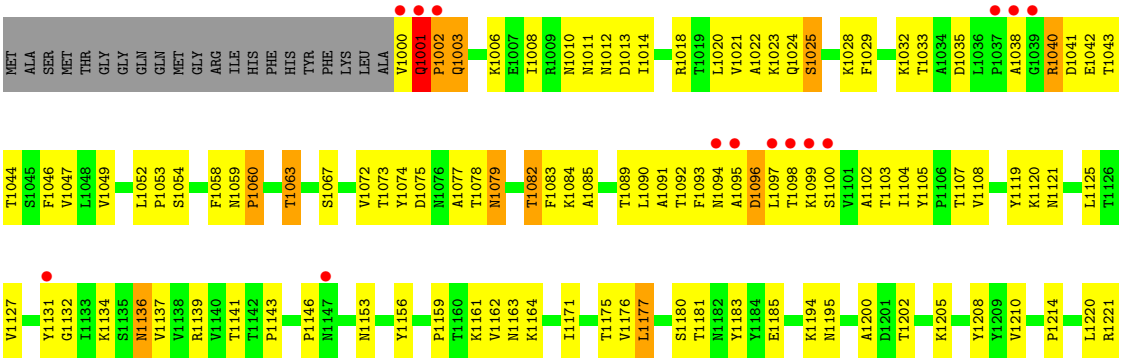
• Molecule 1: Major cell-surface adhesin PAc

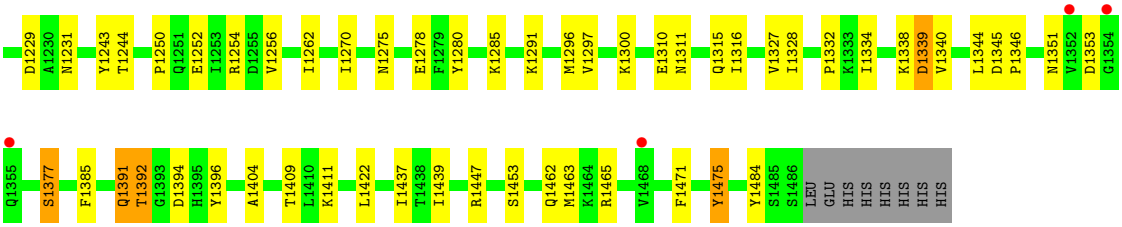
Chain A: 3% 67% 26% • 6%



• Molecule 1: Major cell-surface adhesin PAc

Chain B: 3% 63% 28% • 6%




D1229
A1230
N1231
Y1243
T1244
P1250
Q1251
E1252
T1253
R1254
D1255
V1256
I1262
I1270
N1275
E1278
F1279
Y1280
K1285
K1291
M1296
V1297
K1300
E1310
N1311
Q1315
I1316
V1327
I1328
P1332
K1333
I1334
K1338
D1339
V1340
L1344
D1345
P1346
N1351
V1352
D1353
G1354
Q1355
S1377
F1385
Q1391
T1392
G1393
D1394
H1395
Y1396
A1404
T1409
L1410
K1411
L1422
I1437
T1438
I1439
R1447
S1453
Q1462
M1463
K1464
R1465
V1468
F1471
Y1475
Y1484
S1485
S1486
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	108.50Å 156.36Å 213.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.50 39.76 – 2.51	Depositor EDS
% Data completeness (in resolution range)	82.7 (40.00-2.50) 91.6 (39.76-2.51)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	3.52 (at 2.51Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.208 , 0.244 0.212 , 0.245	Depositor DCC
R_{free} test set	6093 reflections (10.12%)	wwPDB-VP
Wilson B-factor (Å^2)	42.5	Xtriage
Anisotropy	0.640	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 53.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	8029	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, GLC, MG

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	0.42	0/3874	0.67	0/5277
1	B	0.44	0/3874	0.68	1/5277 (0.0%)
All	All	0.43	0/7748	0.68	1/10554 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1377	SER	N-CA-C	5.17	124.96	111.00

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	3801	0	3707	113	0
1	B	3801	0	3709	129	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	1	0	0	0	0
4	A	60	0	60	4	0
4	B	36	0	36	2	0
5	A	189	0	0	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	135	0	0	1	0
All	All	8029	0	7512	242	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 16.

The worst 5 of 242 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:1221:ARG:HD2	1:A:1297:VAL:HG23	1.27	1.15
1:A:1001:GLN:HB2	1:A:1002:PRO:CD	1.89	1.02
1:B:1394:ASP:HB3	1:B:1463:MET:HE2	1.44	0.99
1:B:1096:ASP:OD1	1:B:1098:THR:HG22	1.65	0.96
1:B:1059:ASN:O	1:B:1063:THR:HG22	1.67	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	485/516 (94%)	460 (95%)	24 (5%)	1 (0%)	47 68
1	B	485/516 (94%)	443 (91%)	31 (6%)	11 (2%)	6 10
All	All	970/1032 (94%)	903 (93%)	55 (6%)	12 (1%)	13 24

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1001	GLN
1	B	1095	ALA
1	B	1096	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	1339	ASP
1	B	1391	GLN

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	421/445 (95%)	403 (96%)	18 (4%)	29	53
1	B	421/445 (95%)	404 (96%)	17 (4%)	31	56
All	All	842/890 (95%)	807 (96%)	35 (4%)	30	54

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1136	ASN
1	B	1177	LEU
1	B	1344	LEU
1	A	1397	THR
1	A	1343	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1011	ASN
1	B	1079	ASN
1	B	1355	GLN
1	B	1030	GLN
1	B	1111	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 15 ligands modelled in this entry, 7 are monoatomic - leaving 8 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	$\#\|Z\| > 2$
4	GLC	A	1495	-	12,12,12	0.53	0	17,17,17	0.37	0
4	GLC	B	8	-	12,12,12	0.66	0	17,17,17	0.78	0
4	GLC	A	1497	-	12,12,12	0.73	0	17,17,17	0.62	0
4	GLC	A	1496	-	12,12,12	0.91	0	17,17,17	0.47	0
4	GLC	A	5	-	12,12,12	0.76	0	17,17,17	0.83	1 (5%)
4	GLC	B	7	-	12,12,12	0.79	1 (8%)	17,17,17	0.82	0
4	GLC	B	1495	-	12,12,12	0.60	0	17,17,17	0.68	0
4	GLC	A	4	-	12,12,12	0.39	0	17,17,17	0.46	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
4	GLC	A	1495	-	-	0/2/22/22	0/1/1/1
4	GLC	B	8	-	-	1/2/22/22	0/1/1/1
4	GLC	A	1497	-	-	0/2/22/22	0/1/1/1
4	GLC	A	1496	-	-	1/2/22/22	0/1/1/1
4	GLC	A	5	-	-	2/2/22/22	0/1/1/1
4	GLC	B	7	-	-	2/2/22/22	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	GLC	B	1495	-	-	2/2/22/22	0/1/1/1
4	GLC	A	4	-	-	0/2/22/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	7	GLC	C1-C2	2.03	1.57	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	5	GLC	C3-C4-C5	-2.21	106.23	110.23

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	5	GLC	O5-C5-C6-O6
4	A	5	GLC	C4-C5-C6-O6
4	B	7	GLC	O5-C5-C6-O6
4	B	7	GLC	C4-C5-C6-O6
4	B	1495	GLC	C4-C5-C6-O6

There are no ring outliers.

5 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1495	GLC	1	0
4	B	8	GLC	1	0
4	A	1496	GLC	1	0
4	A	5	GLC	2	0
4	B	7	GLC	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	487/516 (94%)	0.17	13 (2%)	54	58	25, 44, 62, 69	0
1	B	487/516 (94%)	0.21	18 (3%)	41	45	28, 43, 62, 73	0
All	All	974/1032 (94%)	0.19	31 (3%)	47	51	25, 43, 62, 73	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1000	VAL	7.3
1	B	1098	THR	4.1
1	B	1099	LYS	3.9
1	B	1038	ALA	3.9
1	A	1430	VAL	3.5

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 monosaccharides 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	CA	A	3	1/1	0.44	0.13	58,58,58,58	1
4	GLC	A	1496	12/12	0.51	0.48	68,72,73,73	0
4	GLC	B	7	12/12	0.73	0.37	75,76,76,77	0
4	GLC	B	8	12/12	0.75	0.35	72,75,75,76	0
4	GLC	A	1497	12/12	0.80	0.26	71,73,74,75	0
4	GLC	B	1495	12/12	0.84	0.17	60,62,64,65	0
4	GLC	A	1495	12/12	0.86	0.17	58,63,63,64	0
4	GLC	A	5	12/12	0.88	0.18	65,67,68,69	0
2	CA	B	6	1/1	0.92	0.11	41,41,41,41	1
2	CA	A	1	1/1	0.93	0.18	35,35,35,35	0
2	CA	A	2	1/1	0.95	0.22	51,51,51,51	0
4	GLC	A	4	12/12	0.95	0.12	44,49,50,51	0
3	MG	A	7	1/1	0.96	0.22	39,39,39,39	0
2	CA	B	4	1/1	0.97	0.10	44,44,44,44	0
2	CA	B	5	1/1	0.97	0.21	39,39,39,39	0

6.5 Other polymers (i)

There are no such residues in this entry.