Full wwPDB X-ray Structure Validation Report (i)

Jun 18, 2024 – 12:36 PM EDT

PDB ID	:	4D94
Title	:	Crystal Structure of TEP1r
Authors	:	Le, B.V.; Williams, M.; Logarajah, S.; Baxter, R.H.G.
Deposited on	:	2012-01-11
Resolution	:	2.70 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 (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:

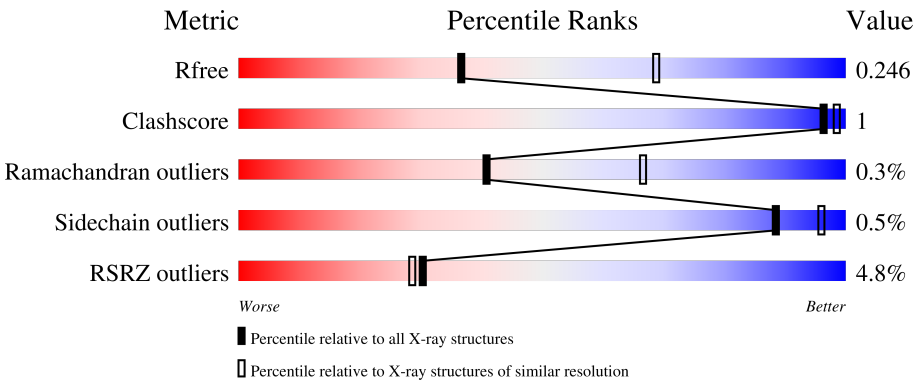
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	1325	5% 93%
2	B	2	100%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 10466 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 Thioester-containing protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1274	10246	6566	1708	1928	44	0	4	0

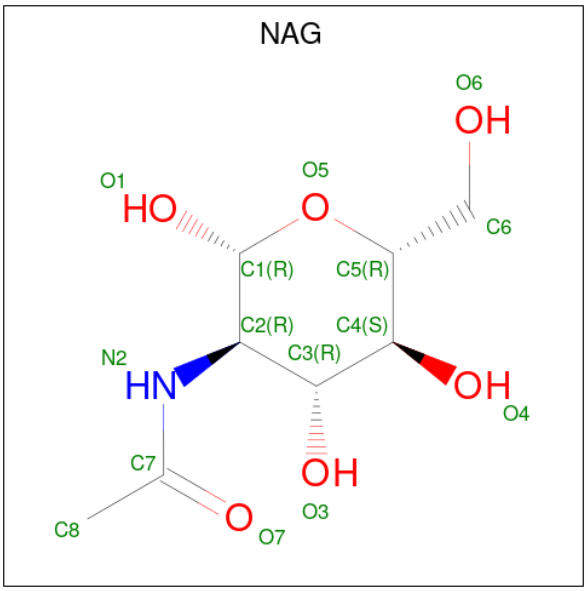
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1339	GLY	-	linker	UNP C9XI66
A	1340	GLY	-	linker	UNP C9XI66
A	1341	HIS	-	expression tag	UNP C9XI66
A	1342	HIS	-	expression tag	UNP C9XI66
A	1343	HIS	-	expression tag	UNP C9XI66
A	1344	HIS	-	expression tag	UNP C9XI66
A	1345	HIS	-	expression tag	UNP C9XI66
A	1346	HIS	-	expression tag	UNP C9XI66

- Molecule 2 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	2	28	16	2	10	0	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C N O 14 8 1 5	0	0
3	A	1	Total C N O 14 8 1 5	0	0
3	A	1	Total C N O 14 8 1 5	0	0
3	A	1	Total C N O 14 8 1 5	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Cl 1 1	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total Na 2 2	0	0

- Molecule 6 is water.

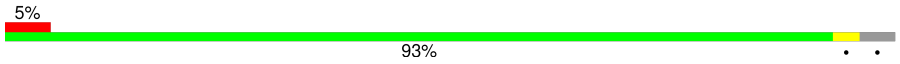
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	133	Total O 133 133	0	0

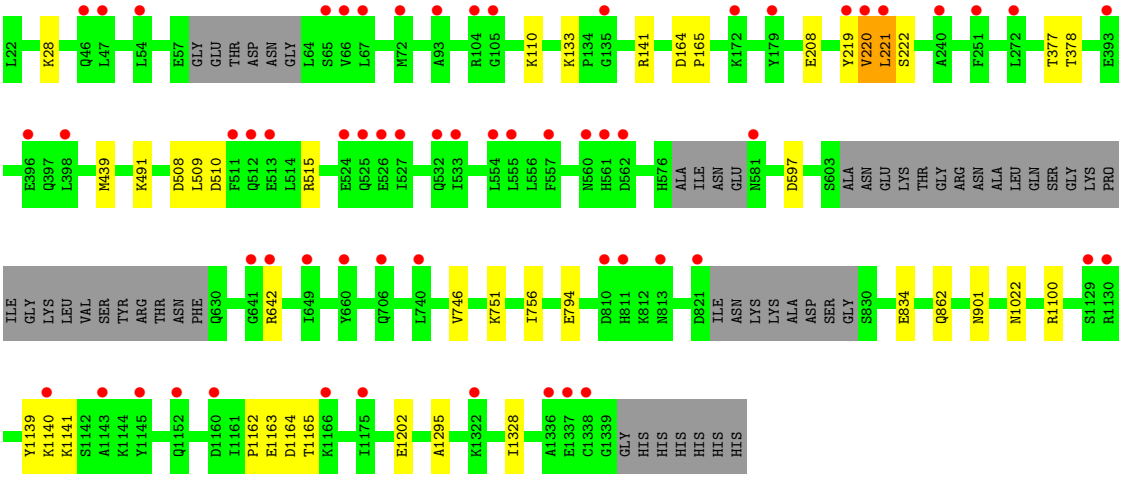
3 Residue-property plots (i)

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: Thioester-containing protein 1

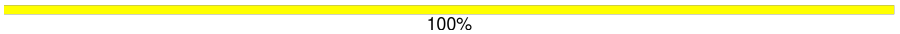
Chain A:





• Molecule 2: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	150.51Å 150.51Å 226.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.87 – 2.70 43.87 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.8 (45.87-2.70) 99.8 (43.87-2.70)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.7.3_928, REFMAC 5.1	Depositor
R, R_{free}	0.219 , 0.244 0.223 , 0.246	Depositor DCC
R_{free} test set	3631 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	57.4	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 43.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10466	wwPDB-VP
Average B, all atoms (Å^2)	70.0	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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CL, NAG

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	0.29	0/10455	0.54	1/14154 (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	A	141	ARG	NE-CZ-NH1	5.93	123.26	120.30

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	10246	0	10230	20	0
2	B	28	0	25	0	0
3	A	56	0	52	2	0
4	A	1	0	0	0	0
5	A	2	0	0	0	0
6	A	133	0	0	4	0
All	All	10466	0	10307	20	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 1.

All (20) 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:220:VAL:O	1:A:222:SER:N	2.33	0.62
1:A:110:LYS:NZ	1:A:597:ASP:OD1	2.33	0.61
1:A:1022:ASN:OD1	6:A:2279:HOH:O	2.16	0.61
1:A:491:LYS:NZ	1:A:508:ASP:OD2	2.35	0.59
1:A:1139:TYR:O	1:A:1141:LYS:N	2.35	0.58
1:A:219:TYR:O	1:A:221:LEU:N	2.38	0.56
1:A:510:ASP:OD1	1:A:515:ARG:NH1	2.40	0.55
1:A:901:ASN:ND2	6:A:2203:HOH:O	2.39	0.55
1:A:28:LYS:NZ	6:A:2227:HOH:O	2.39	0.54
1:A:377:THR:OG1	1:A:378:THR:N	2.45	0.50
1:A:164:ASP:HB2	1:A:165:PRO:HD2	1.97	0.46
1:A:208:GLU:CD	3:A:2001:NAG:H62	2.38	0.43
1:A:439:MET:HE1	1:A:509:LEU:HD13	2.00	0.43
1:A:1100:ARG:NH2	6:A:2252:HOH:O	2.50	0.43
1:A:208:GLU:OE1	3:A:2001:NAG:H62	2.19	0.43
1:A:1163:GLU:O	1:A:1165:THR:N	2.52	0.42
1:A:1295:ALA:HB2	1:A:1328:ILE:HG13	2.01	0.42
1:A:164:ASP:HB2	1:A:165:PRO:CD	2.50	0.41
1:A:834:GLU:HA	1:A:1162:PRO:HD2	2.02	0.41
1:A:746:VAL:HG23	1:A:756:ILE:HD13	2.02	0.41

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	1268/1325 (96%)	1206 (95%)	58 (5%)	4 (0%)	41 66

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	220	VAL
1	A	221	LEU
1	A	1140	LYS
1	A	1164	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	1138/1174 (97%)	1132 (100%)	6 (0%)	88 96

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	133	LYS
1	A	642	ARG
1	A	751	LYS
1	A	794	GLU
1	A	862	GLN
1	A	1202	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

2 monosaccharides 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	NAG	B	1	1,2	14,14,15	0.81	0	17,19,21	1.33	1 (5%)
2	NAG	B	2	2	14,14,15	0.92	0	17,19,21	1.36	3 (17%)

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	NAG	B	1	1,2	-	1/6/23/26	0/1/1/1
2	NAG	B	2	2	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	2	NAG	O5-C1-C2	-2.90	106.80	111.29
2	B	1	NAG	O5-C1-C2	-2.82	106.93	111.29
2	B	2	NAG	C2-N2-C7	-2.66	119.33	122.90
2	B	2	NAG	C6-C5-C4	-2.26	107.47	113.02

There are no chirality outliers.

All (3) torsion outliers are listed below:

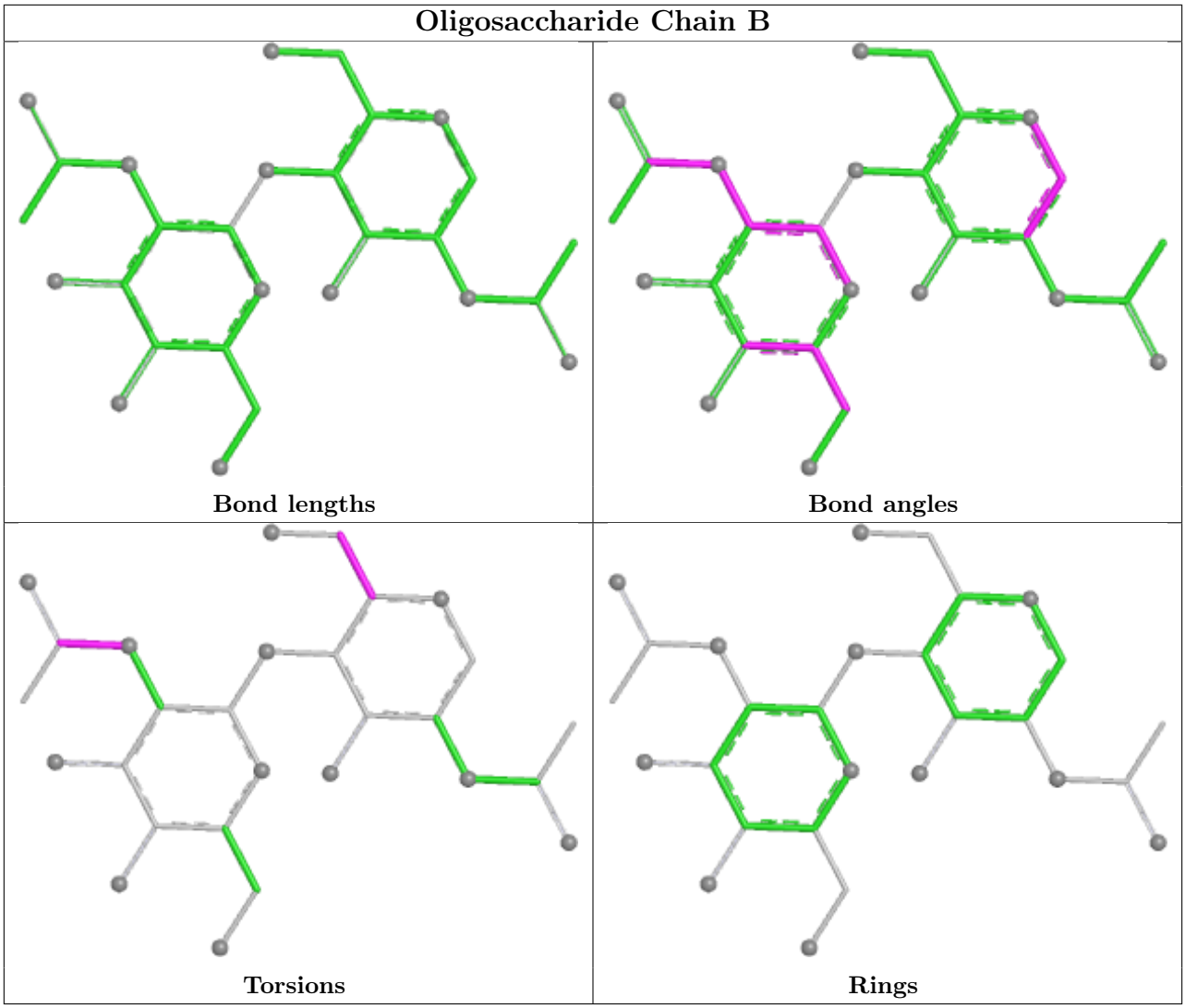
Mol	Chain	Res	Type	Atoms
2	B	2	NAG	C8-C7-N2-C2
2	B	2	NAG	O7-C7-N2-C2
2	B	1	NAG	C4-C5-C6-O6

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 oligosaccharide.



5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 3 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
3	NAG	A	2006	1	14,14,15	0.29	0	17,19,21	0.56	0
3	NAG	A	2002	1	14,14,15	0.29	0	17,19,21	0.57	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	NAG	A	2001	1	14,14,15	0.29	0	17,19,21	0.56	0
3	NAG	A	2003	1	14,14,15	0.30	0	17,19,21	0.56	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	NAG	A	2006	1	-	0/6/23/26	0/1/1/1
3	NAG	A	2002	1	-	0/6/23/26	0/1/1/1
3	NAG	A	2001	1	-	2/6/23/26	0/1/1/1
3	NAG	A	2003	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2003	NAG	O5-C5-C6-O6
3	A	2003	NAG	C4-C5-C6-O6
3	A	2001	NAG	C8-C7-N2-C2
3	A	2001	NAG	O7-C7-N2-C2

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2001	NAG	2	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	1274/1325 (96%)	0.28	61 (4%) 30 28	29, 64, 123, 181	1 (0%)

All (61) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	557	PHE	5.3
1	A	1336	ALA	5.1
1	A	1130	ARG	4.9
1	A	66	VAL	4.8
1	A	561	HIS	4.8
1	A	512	GLN	4.5
1	A	560	ASN	4.4
1	A	393	GLU	4.3
1	A	1140	LYS	4.3
1	A	1152	GLN	4.1
1	A	104	ARG	4.0
1	A	562	ASP	3.9
1	A	72	MET	3.8
1	A	221	LEU	3.7
1	A	1160	ASP	3.7
1	A	642	ARG	3.6
1	A	581	ASN	3.5
1	A	1338	CYS	3.5
1	A	1175	ILE	3.3
1	A	220	VAL	3.3
1	A	219	TYR	3.3
1	A	813	ASN	3.2
1	A	524	GLU	3.2
1	A	65	SER	3.1
1	A	135	GLY	3.0
1	A	1143	ALA	3.0
1	A	105	GLY	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	526	GLU	3.0
1	A	46	GLN	2.9
1	A	251	PHE	2.9
1	A	1337	GLU	2.8
1	A	513	GLU	2.8
1	A	641	GLY	2.8
1	A	67	LEU	2.7
1	A	706	GLN	2.7
1	A	660	TYR	2.6
1	A	1322	LYS	2.6
1	A	179	TYR	2.6
1	A	811	HIS	2.6
1	A	533	ILE	2.5
1	A	54	LEU	2.5
1	A	527	ILE	2.5
1	A	532	GLN	2.5
1	A	172	LYS	2.4
1	A	93	ALA	2.4
1	A	740	LEU	2.2
1	A	1145	TYR	2.2
1	A	810	ASP	2.2
1	A	821	ASP	2.2
1	A	525	GLN	2.2
1	A	398	LEU	2.2
1	A	511	PHE	2.2
1	A	240	ALA	2.2
1	A	272	LEU	2.2
1	A	1129	SER	2.1
1	A	47	LEU	2.1
1	A	555	LEU	2.1
1	A	554	LEU	2.1
1	A	396	GLU	2.1
1	A	1166	LYS	2.1
1	A	649	ILE	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)

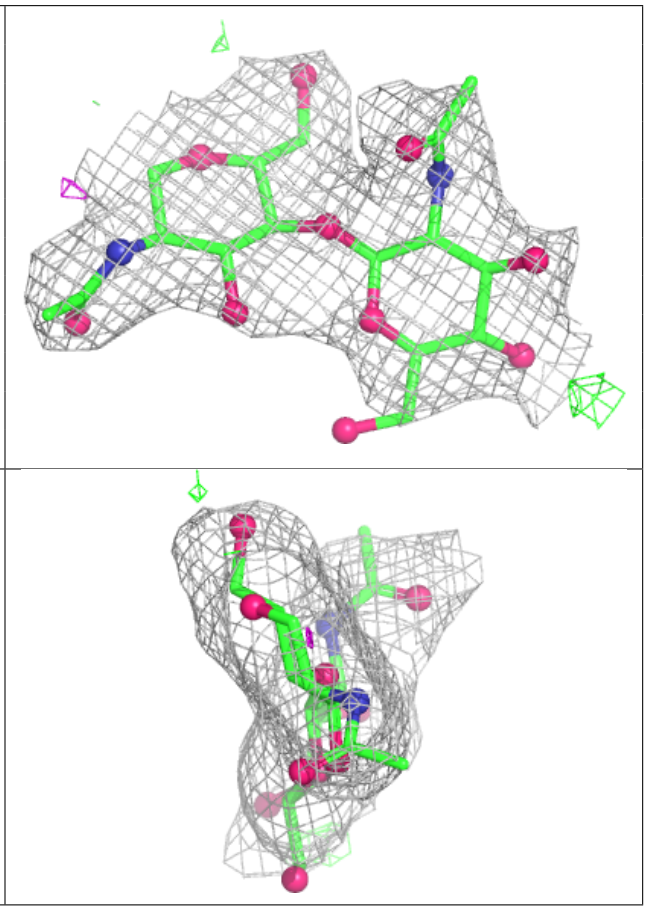
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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	NAG	B	1	14/15	0.88	0.19	89,100,109,117	0
2	NAG	B	2	14/15	0.90	0.16	105,126,135,137	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around Chain B:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)



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	NAG	A	2001	14/15	0.72	0.34	94,115,124,124	0
3	NAG	A	2002	14/15	0.82	0.21	58,75,98,109	0
3	NAG	A	2003	14/15	0.83	0.18	57,75,86,88	0
3	NAG	A	2006	14/15	0.86	0.16	86,109,117,121	0
5	NA	A	2008	1/1	0.86	0.57	47,47,47,47	0
5	NA	A	2009	1/1	0.87	0.18	78,78,78,78	0
4	CL	A	2007	1/1	0.99	0.11	43,43,43,43	0

6.5 Other polymers (i)

There are no such residues in this entry.