WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Nov 26, 2017 – 12:38 AM EST

PDB ID	:	5LR0
Title	:	Binding domain of Botulinum Neurotoxin DC in complex with SialylT
Authors	:	Berntsson, R.P.-A.; Stenmark, P.
Deposited on	:	unknown
Resolution	:	2.59 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

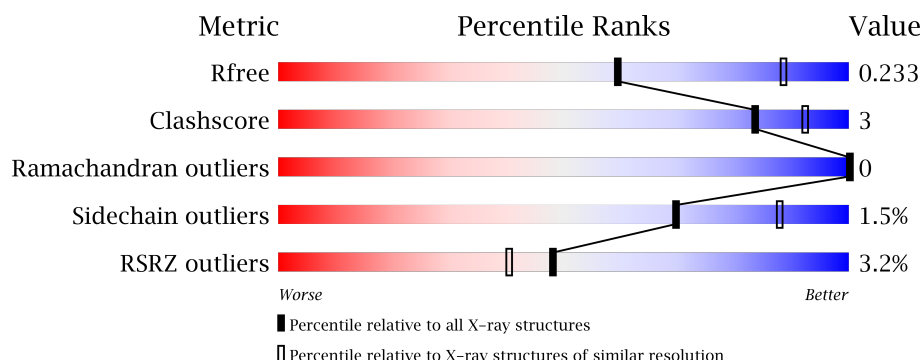|		
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

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 2.59 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	Percentile Ranks	Value
Rfree		0.233
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		1.5%
RSRZ outliers		3.2%

Worse — *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (2.60-2.60)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	431	5% 88% 6% 6%
1	B	431	.% 88% 8% •

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
2	SIA	B	1301	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6944 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 Botulinum neurotoxin D/C protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	407	Total 3338	C 2144	N 551	O 629	S 14	0	0	0
1	B	412	Total 3371	C 2163	N 557	O 637	S 14	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	854	MET	-	initiating methionine	UNP A5JGM8
A	855	HIS	-	expression tag	UNP A5JGM8
A	856	HIS	-	expression tag	UNP A5JGM8
A	857	HIS	-	expression tag	UNP A5JGM8
A	858	HIS	-	expression tag	UNP A5JGM8
A	859	HIS	-	expression tag	UNP A5JGM8
A	860	HIS	-	expression tag	UNP A5JGM8
A	861	TYR	-	expression tag	UNP A5JGM8
A	862	PHE	-	expression tag	UNP A5JGM8
B	854	MET	-	initiating methionine	UNP A5JGM8
B	855	HIS	-	expression tag	UNP A5JGM8
B	856	HIS	-	expression tag	UNP A5JGM8
B	857	HIS	-	expression tag	UNP A5JGM8
B	858	HIS	-	expression tag	UNP A5JGM8
B	859	HIS	-	expression tag	UNP A5JGM8
B	860	HIS	-	expression tag	UNP A5JGM8
B	861	TYR	-	expression tag	UNP A5JGM8
B	862	PHE	-	expression tag	UNP A5JGM8

- Molecule 2 is O-SIALIC ACID (three-letter code: SIA) (formula: $C_{11}H_{19}NO_9$).
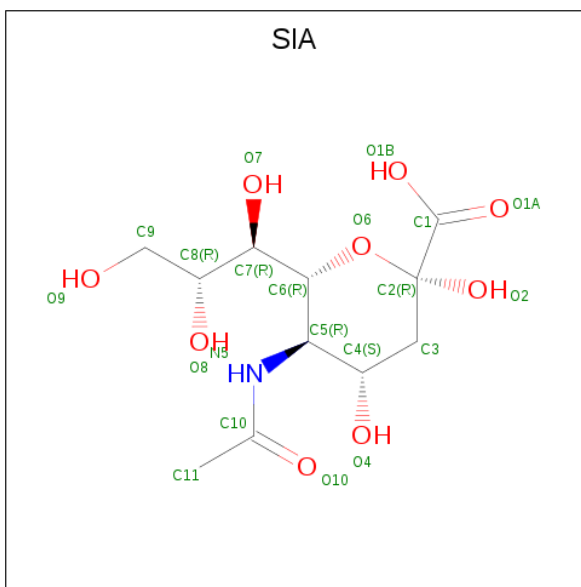
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 20	C 11	N 1	O 8	0	0
2	B	1	Total 20	C 11	N 1	O 8	0	0

- Molecule 3 is BETA-D-GALACTOSE (three-letter code: GAL) (formula: $C_6H_{12}O_6$).
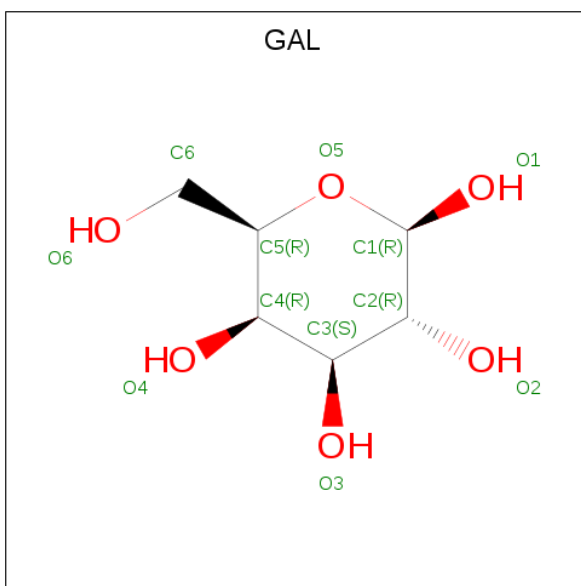
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 11	C 6	O 5	0	0
3	B	1	Total 11	C 6	O 5	0	0

- Molecule 4 is N-ACETYL-D-GALACTOSAMINE (three-letter code: NGA) (formula: $C_8H_{15}NO_6$).
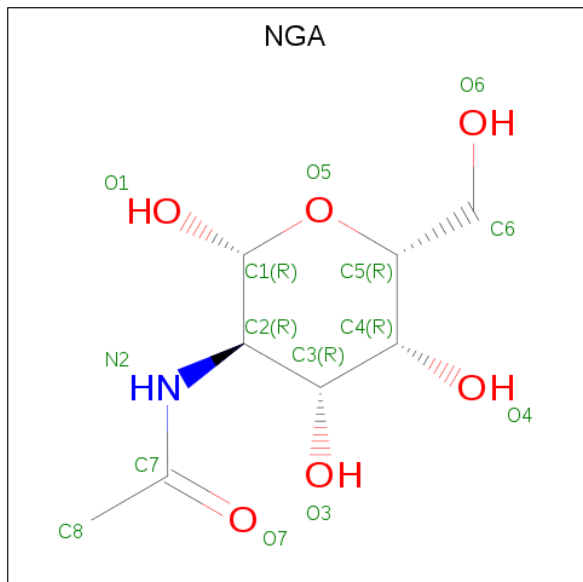
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 15	C 8	N 1	O 6	0	0
4	B	1	Total 15	C 8	N 1	O 6	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	70	Total 70	O 70	0	0
5	B	73	Total 73	O 73	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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: Botulinum neurotoxin D/C protein
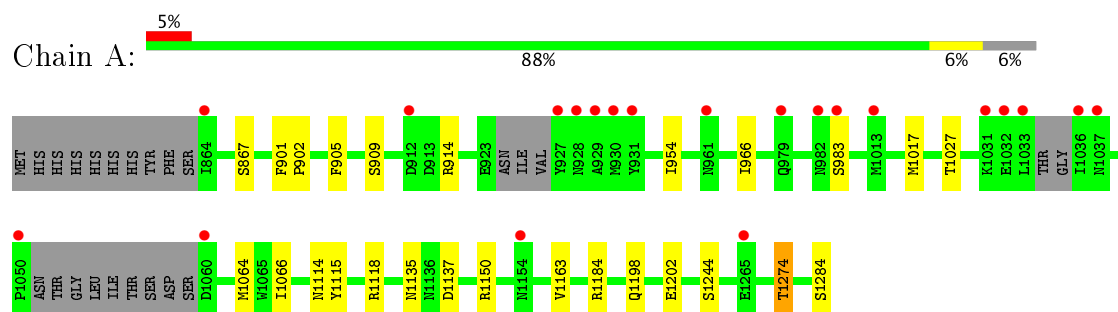
- Molecule 1: Botulinum neurotoxin D/C protein
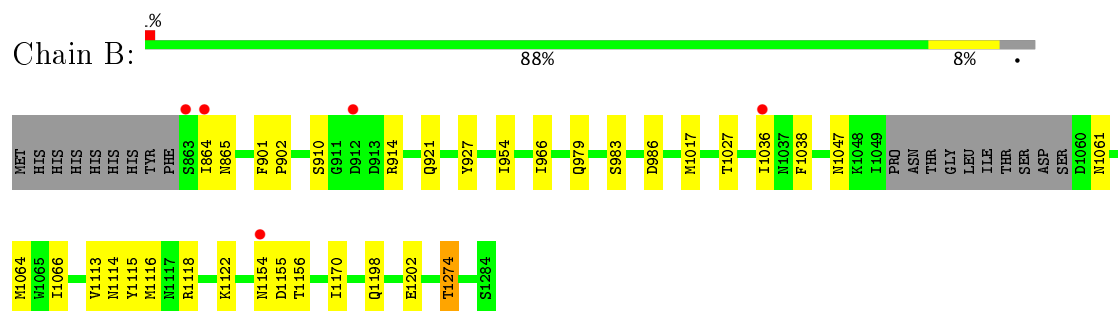
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	109.70Å 109.70Å 210.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.64 – 2.59 48.64 – 2.59	Depositor EDS
% Data completeness (in resolution range)	99.2 (48.64-2.59) 99.2 (48.64-2.59)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.03 (at 2.58Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.207 , 0.231 0.211 , 0.233	Depositor DCC
R_{free} test set	2028 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	38.6	Xtriage
Anisotropy	0.533	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 32.1	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6944	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% 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: SIA, GAL, NGA

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	0.35	0/3413	0.58	0/4614
1	B	0.35	0/3447	0.59	0/4663
All	All	0.35	0/6860	0.58	0/9277

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	B	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	B	1154	ASN	Peptide

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.

Full wwPDB X-ray Structure Validation Report

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3338	0	3261	14	0
1	B	3371	0	3297	19	0
2	A	20	0	17	1	0
2	B	20	0	17	2	0
3	A	11	0	9	1	0
3	B	11	0	9	2	0
4	A	15	0	13	1	0
4	B	15	0	14	0	0
5	A	70	0	0	1	0
5	B	73	0	0	0	0
All	All	6944	0	6637	37	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 (37) 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:B:914:ARG:NH1	1:B:1047:ASN:OD1	2.18	0.76
1:B:1114:ASN:OD1	1:B:1274:THR:HG23	1.88	0.74
1:A:1150:ARG:HB3	1:A:1163:VAL:HG12	1.69	0.72
1:B:979:GLN:HG3	1:B:1036:ILE:HG21	1.73	0.69
1:B:1122:LYS:HD2	1:B:1170:ILE:HD13	1.77	0.65
1:A:1114:ASN:OD1	1:A:1274:THR:HG23	2.02	0.60
3:A:1302:GAL:O5	4:A:1303:NGA:H4	2.01	0.59
1:A:1150:ARG:HB3	1:A:1163:VAL:CG1	2.35	0.56
1:A:954:ILE:HD12	1:A:1064:MET:HE3	1.88	0.56
1:A:1198:GLN:HB2	1:A:1202:GLU:HB2	1.90	0.54
2:B:1301:SIA:C1	3:B:1302:GAL:H4	2.37	0.54
1:B:1198:GLN:HB2	1:B:1202:GLU:HB2	1.89	0.53
1:B:1114:ASN:OD1	1:B:1274:THR:CG2	2.57	0.53
1:B:1036:ILE:HD11	1:B:1038:PHE:CZ	2.44	0.51
1:B:979:GLN:HG3	1:B:1036:ILE:CG2	2.41	0.50
1:B:901:PHE:CG	1:B:902:PRO:HA	2.47	0.50
1:B:954:ILE:HD12	1:B:1064:MET:HE3	1.92	0.49
1:B:921:GLN:HB3	1:B:927:TYR:CE2	2.48	0.49
1:A:901:PHE:CG	1:A:902:PRO:HA	2.48	0.48
1:B:864:ILE:HG13	1:B:865:ASN:OD1	2.14	0.48
1:A:1284:SER:C	5:A:1411:HOH:O	2.52	0.47
2:A:1301:SIA:O1A	2:A:1301:SIA:H6	2.14	0.47
1:A:1184:ARG:NH2	1:A:1244:SER:O	2.48	0.47
1:B:1115:TYR:HB3	1:B:1118:ARG:HG3	1.98	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:910:SER:HB3	1:B:1061:ASN:OD1	2.15	0.46
1:A:1114:ASN:OD1	1:A:1274:THR:CG2	2.63	0.46
1:B:1017:MET:HG2	1:B:1027:THR:HG23	1.99	0.45
1:A:1115:TYR:HB3	1:A:1118:ARG:HG3	1.99	0.44
2:B:1301:SIA:C1	3:B:1302:GAL:C4	2.92	0.43
1:A:1017:MET:HG2	1:A:1027:THR:HG23	2.01	0.43
1:A:909:SER:HA	1:A:914:ARG:HD2	2.02	0.42
1:B:1113:VAL:HA	1:B:1116:MET:HE2	2.01	0.42
1:A:905:PHE:CZ	1:A:1066:ILE:HB	2.55	0.42
1:B:1156:THR:O	1:B:1156:THR:HG22	2.20	0.41
1:B:1156:THR:O	1:B:1156:THR:CG2	2.68	0.41
1:B:1064:MET:HE2	1:B:1066:ILE:HD11	2.03	0.41
1:A:954:ILE:HD12	1:A:1064:MET:CE	2.51	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	399/431 (93%)	380 (95%)	19 (5%)	0	100	100
1	B	408/431 (95%)	384 (94%)	24 (6%)	0	100	100
All	All	807/862 (94%)	764 (95%)	43 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	373/395 (94%)	367 (98%)	6 (2%)	68	87
1	B	377/395 (95%)	372 (99%)	5 (1%)	73	90
All	All	750/790 (95%)	739 (98%)	11 (2%)	70	88

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	867	SER
1	A	966	ILE
1	A	983	SER
1	A	1135	ASN
1	A	1137	ASP
1	A	1274	THR
1	B	966	ILE
1	B	983	SER
1	B	986	ASP
1	B	1155	ASP
1	B	1274	THR

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

Mol	Chain	Res	Type
1	B	960	ASN
1	B	1037	ASN
1	B	1063	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

6 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	SIA	A	1301	3	17,20,21	1.83	4 (23%)	19,28,31	1.48	6 (31%)
3	GAL	A	1302	2,4	11,11,12	1.28	1 (9%)	13,15,17	1.25	2 (15%)
4	NGA	A	1303	3	15,15,15	1.32	2 (13%)	21,21,21	1.87	5 (23%)
2	SIA	B	1301	-	17,20,21	0.59	0	19,28,31	0.97	1 (5%)
3	GAL	B	1302	4	11,11,12	0.26	0	13,15,17	0.55	0
4	NGA	B	1303	3	15,15,15	0.40	0	21,21,21	0.50	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	SIA	A	1301	3	-	0/14/34/38	0/1/1/1
3	GAL	A	1302	2,4	-	0/2/19/22	0/1/1/1
4	NGA	A	1303	3	-	0/6/26/26	0/1/1/1
2	SIA	B	1301	-	-	0/14/34/38	0/1/1/1
3	GAL	B	1302	4	-	0/2/19/22	0/1/1/1
4	NGA	B	1303	3	-	0/6/26/26	0/1/1/1

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1301	SIA	C4-C5	-4.78	1.48	1.53
3	A	1302	GAL	O5-C1	-2.72	1.39	1.43
2	A	1301	SIA	C3-C4	-2.59	1.48	1.52
2	A	1301	SIA	O10-C10	-2.44	1.17	1.23
4	A	1303	NGA	O7-C7	-2.19	1.18	1.23
4	A	1303	NGA	C1-C2	-2.09	1.50	1.52
2	A	1301	SIA	C3-C2	2.36	1.56	1.52

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1303	NGA	C1-C2-C3	-4.78	104.03	110.54
4	A	1303	NGA	O1-C1-O5	-3.69	99.30	110.20
3	A	1302	GAL	O5-C1-C2	-2.91	106.23	110.79
4	A	1303	NGA	O5-C1-C2	-2.87	106.64	109.52
4	A	1303	NGA	C3-C4-C5	-2.73	105.40	110.22
2	A	1301	SIA	O6-C2-C3	-2.56	105.21	109.82
2	A	1301	SIA	O4-C4-C5	-2.55	105.02	110.40
4	A	1303	NGA	O4-C4-C5	-2.52	102.92	109.28
2	A	1301	SIA	O4-C4-C3	-2.42	104.17	110.02
2	A	1301	SIA	C4-C5-N5	-2.23	105.79	110.40
3	A	1302	GAL	O3-C3-C2	-2.04	106.31	110.02
2	A	1301	SIA	C3-C4-C5	2.05	113.94	111.46
2	B	1301	SIA	C4-C3-C2	2.13	113.75	109.75
2	A	1301	SIA	C4-C3-C2	2.85	115.10	109.75

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1301	SIA	1	0
3	A	1302	GAL	1	0
4	A	1303	NGA	1	0
2	B	1301	SIA	2	0
3	B	1302	GAL	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(Å²)	Q<0.9
1	A	407/431 (94%)	0.27	21 (5%)	28	21	27, 39, 75, 132	0
1	B	412/431 (95%)	0.04	5 (1%)	79	75	24, 36, 61, 96	0
All	All	819/862 (95%)	0.15	26 (3%)	48	40	24, 38, 67, 132	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	930	MET	8.1
1	A	927	TYR	6.3
1	A	931	TYR	4.7
1	A	929	ALA	4.7
1	A	1036	ILE	4.1
1	B	863	SER	4.1
1	A	1032	GLU	3.6
1	A	1050	PRO	3.4
1	A	864	ILE	3.4
1	B	864	ILE	3.1
1	A	982	ASN	3.1
1	B	1154	ASN	3.0
1	B	912	ASP	3.0
1	A	1037	ASN	2.8
1	A	928	ASN	2.8
1	A	983	SER	2.8
1	A	1013	MET	2.8
1	A	979	GLN	2.6
1	A	1031	LYS	2.6
1	A	1033	LEU	2.4
1	A	1154	ASN	2.3
1	A	912	ASP	2.2
1	A	961	ASN	2.1
1	A	1265	GLU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1060	ASP	2.1
1	B	1036	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)

There are no carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	SIA	B	1301	20/21	0.74	0.34	8.92	61,69,84,85	0
4	NGA	A	1303	15/15	0.95	0.14	-0.42	39,42,45,45	0
2	SIA	A	1301	20/21	0.96	0.11	-2.47	29,32,38,39	0
3	GAL	A	1302	11/12	0.96	0.12	-	36,37,38,38	0
3	GAL	B	1302	11/12	0.76	0.41	-	94,99,106,107	0
4	NGA	B	1303	15/15	0.31	0.70	-	110,117,121,122	0

6.5 Other polymers (i)

There are no such residues in this entry.