



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 11:57 pm BST

PDB ID : 4AZC
Title : Differential inhibition of the tandem GH20 catalytic modules in the pneumococcal exo-beta-D-N-acetylglucosaminidase, StrH
Authors : Pluvinae, B.; Stubbs, K.A.; Vocadlo, D.J.; Boraston, A.B.
Deposited on : 2012-06-25
Resolution : 2.09 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

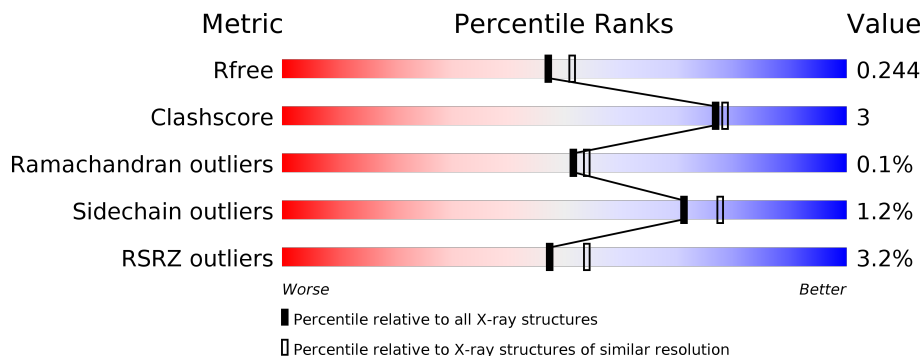
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.09 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 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 $\leq 5\%$. 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	442	
2	B	442	
2	C	442	
2	D	442	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 15613 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-N-ACETYLHEXOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	416	3346	2144	547	643	12	2	6	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	623	GLY	-	expression tag	UNP P49610
A	624	SER	-	expression tag	UNP P49610
A	625	HIS	-	expression tag	UNP P49610
A	626	MET	-	expression tag	UNP P49610
A	1041	LYS	LEU	conflict	UNP P49610

- Molecule 2 is a protein called BETA-N-ACETYLHEXOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	421	3399	2178	554	655	12	6	7	0
2	C	416	3374	2159	553	650	12	4	9	0
2	D	415	3357	2147	545	653	12	8	7	0

There are 12 discrepancies between the modelled and reference sequences:

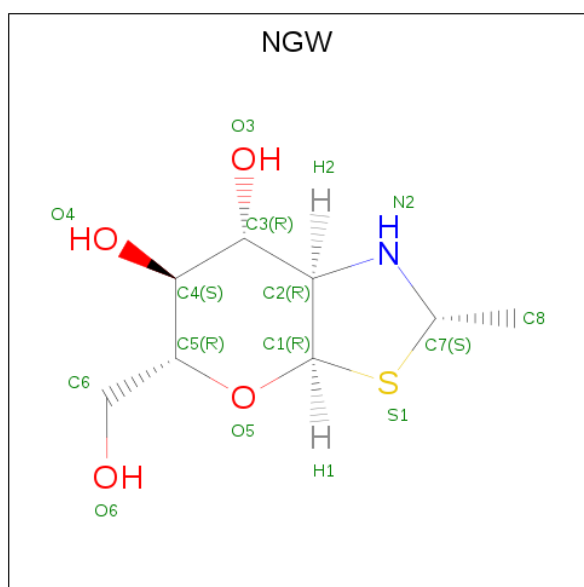
Chain	Residue	Modelled	Actual	Comment	Reference
B	623	GLY	-	expression tag	UNP P49610
B	624	SER	-	expression tag	UNP P49610
B	625	HIS	-	expression tag	UNP P49610
B	626	MET	-	expression tag	UNP P49610
C	623	GLY	-	expression tag	UNP P49610
C	624	SER	-	expression tag	UNP P49610
C	625	HIS	-	expression tag	UNP P49610

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	626	MET	-	expression tag	UNP P49610
D	623	GLY	-	expression tag	UNP P49610
D	624	SER	-	expression tag	UNP P49610
D	625	HIS	-	expression tag	UNP P49610
D	626	MET	-	expression tag	UNP P49610

- Molecule 3 is (2S,3aR,5R,6S,7R,7aR)-5-(hydroxymethyl)-2-methyl-2,3a,5,6,7,7a-hexahydro-1H-pyrano[3,2-d][1,3]thiazole-6,7-diol (three-letter code: NGW) (formula: C₈H₁₅NO₄S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	Total	C	N	O	S	0	0
			14	8	1	4	1		
3	B	1	Total	C	N	O	S	0	0
			14	8	1	4	1		
3	C	1	Total	C	N	O	S	0	0
			14	8	1	4	1		
3	D	1	Total	C	N	O	S	0	0
			14	8	1	4	1		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

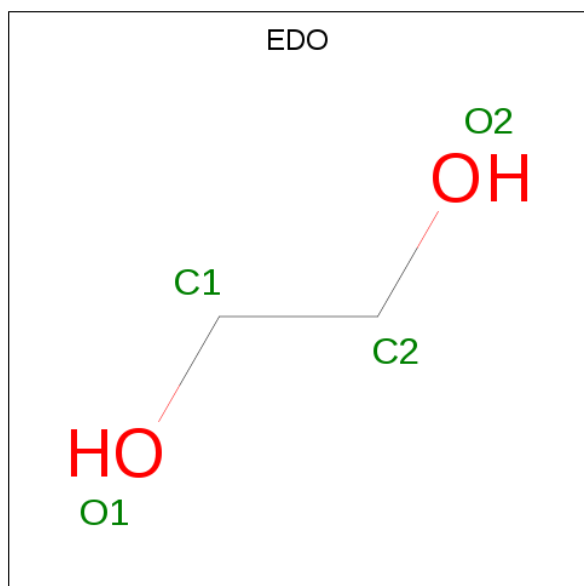
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
4	B	1	Total	Mg	0	0
			1	1		
4	A	4	Total	Mg	0	0
			4	4		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	3	Total	Mg	0	0
			3	3		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	2	2		
5	C	1	Total	C	O	0	0
			4	2	2		
5	C	1	Total	C	O	0	0
			4	2	2		
5	C	1	Total	C	O	0	0
			4	2	2		

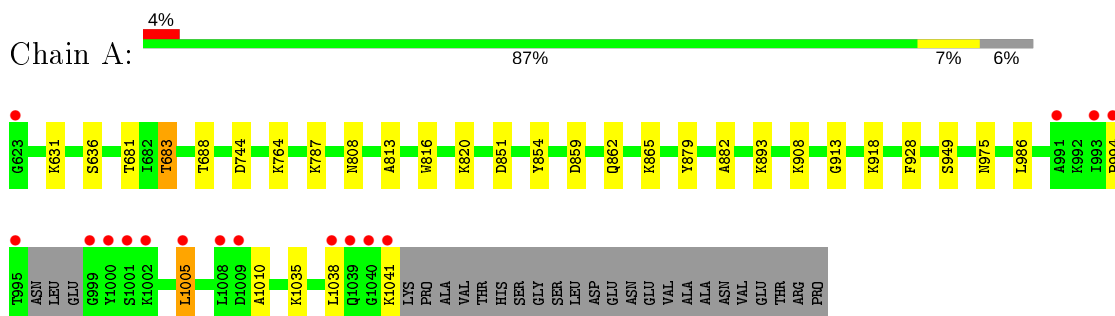
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	594	Total	O	0	2
			596	596		
6	B	472	Total	O	0	1
			473	473		
6	C	547	Total	O	0	2
			549	549		
6	D	438	Total	O	0	1
			439	439		

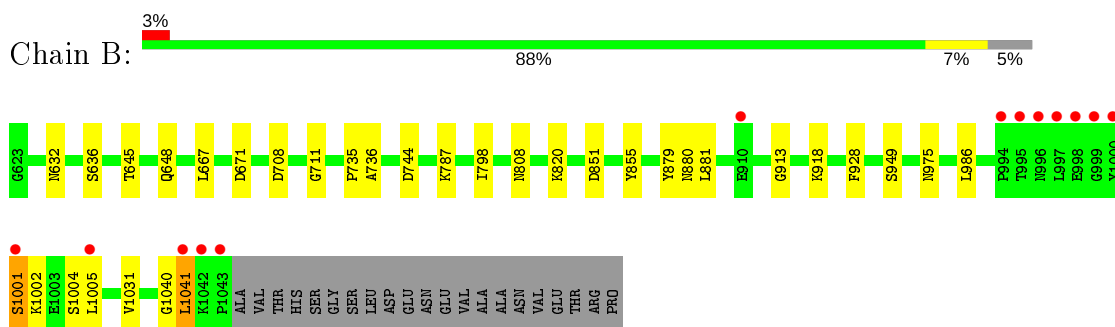
3 Residue-property plots [i](#)

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.

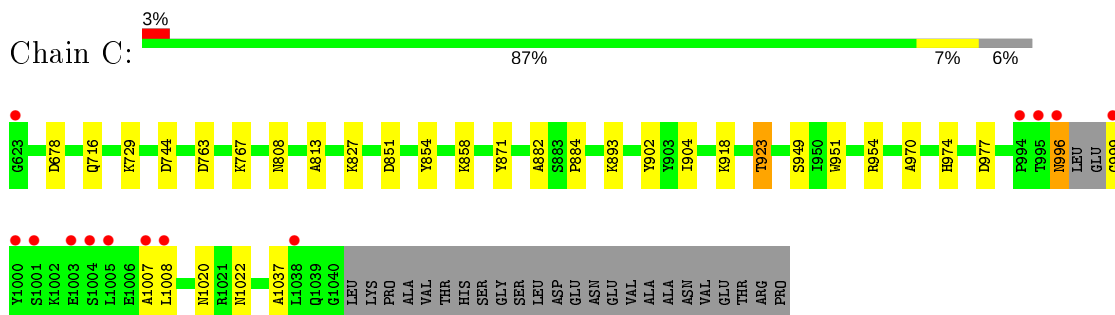
- Molecule 1: BETA-N-ACETYLHEXOSAMINIDASE



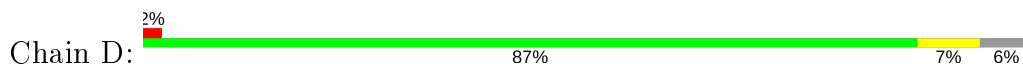
- Molecule 2: BETA-N-ACETYLHEXOSAMINIDASE

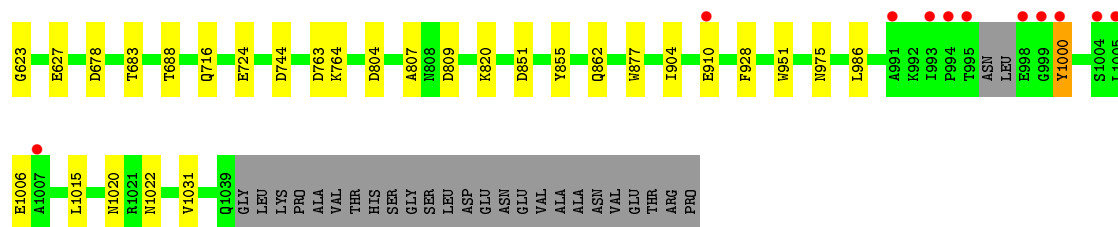


- Molecule 2: BETA-N-ACETYLHEXOSAMINIDASE



- Molecule 2: BETA-N-ACETYLHEXOSAMINIDASE





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	67.20Å 115.60Å 132.00Å 90.00° 99.40° 90.00°	Depositor
Resolution (Å)	52.81 – 2.09 43.57 – 2.09	Depositor EDS
% Data completeness (in resolution range)	100.0 (52.81-2.09) 99.9 (43.57-2.09)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.00 (at 2.08Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.190 , 0.240 0.190 , 0.244	Depositor DCC
R_{free} test set	5888 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	19.1	Xtrriage
Anisotropy	0.052	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 57.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	15613	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 69.84 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.4549e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: MG, NGW, 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	0.68	2/3423 (0.1%)	0.61	2/4623 (0.0%)
2	B	0.49	1/3478 (0.0%)	0.58	1/4699 (0.0%)
2	C	0.88	4/3448 (0.1%)	1.79	9/4655 (0.2%)
2	D	0.48	0/3434	0.57	0/4638
All	All	0.65	7/13783 (0.1%)	1.03	12/18615 (0.1%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1035	LYS	CE-NZ	-25.41	0.85	1.49
2	C	767	LYS	CD-CE	-25.22	0.88	1.51
2	C	954[A]	ARG	NE-CZ	-23.62	1.02	1.33
2	C	954[B]	ARG	NE-CZ	-23.62	1.02	1.33
2	B	1002	LYS	CG-CD	-7.08	1.28	1.52

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	C	954[A]	ARG	NE-CZ-NH2	-56.80	91.90	120.30
2	C	954[B]	ARG	NE-CZ-NH2	-56.80	91.90	120.30
2	C	954[A]	ARG	NE-CZ-NH1	56.14	148.37	120.30
2	C	954[B]	ARG	NE-CZ-NH1	56.14	148.37	120.30
2	C	767	LYS	CG-CD-CE	14.46	155.28	111.90

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	3346	0	3263	20	0
2	B	3399	0	3324	21	0
2	C	3374	0	3286	22	0
2	D	3357	0	3258	18	0
3	A	14	0	15	0	0
3	B	14	0	15	0	0
3	C	14	0	15	0	0
3	D	14	0	15	1	0
4	A	4	0	0	0	0
4	B	1	0	0	0	0
4	C	3	0	0	0	0
5	A	4	0	6	0	0
5	C	12	0	18	0	0
6	A	596	0	0	7	0
6	B	473	0	0	4	1
6	C	549	0	0	5	0
6	D	439	0	0	3	1
All	All	15613	0	13215	80	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 3.

The worst 5 of 80 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)
2:B:986[A]:LEU:HD12	2:B:986[A]:LEU:O	1.79	0.82
2:B:787[A]:LYS:HE2	6:B:3059:HOH:O	1.81	0.78
1:A:683:THR:HG23	1:A:688:THR:HG22	1.68	0.76
2:B:808:ASN:OD1	6:B:3163:HOH:O	2.12	0.68
1:A:913:GLY:O	1:A:918:LYS:HE2	1.92	0.68

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)
6:B:3249:HOH:O	6:D:3175:HOH:O[1_455]	2.09	0.11

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	418/442 (95%)	404 (97%)	14 (3%)	0	100	100
2	B	426/442 (96%)	411 (96%)	14 (3%)	1 (0%)	47	49
2	C	421/442 (95%)	410 (97%)	11 (3%)	0	100	100
2	D	418/442 (95%)	404 (97%)	13 (3%)	1 (0%)	47	49
All	All	1683/1768 (95%)	1629 (97%)	52 (3%)	2 (0%)	51	54

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	807	ALA
2	B	671	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	344/363 (95%)	340 (99%)	4 (1%)	71	77
2	B	352/363 (97%)	348 (99%)	4 (1%)	73	79
2	C	347/363 (96%)	343 (99%)	4 (1%)	71	77

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	D	347/363 (96%)	341 (98%)	6 (2%)	60	67
All	All	1390/1452 (96%)	1372 (99%)	18 (1%)	71	75

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	744	ASP
2	C	851	ASP
2	D	744	ASP
2	B	1001	SER
2	B	1041	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	716	GLN
2	C	761	HIS
2	D	716	GLN
2	B	929	ASN
2	B	975	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](#)

Of 16 ligands modelled in this entry, 8 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
5	EDO	C	2045	-	3,3,3	0.47	0	2,2,2	0.31	0
3	NGW	A	2042	-	11,15,15	1.47	1 (9%)	13,22,22	1.19	1 (7%)
3	NGW	C	2041	-	11,15,15	1.50	1 (9%)	13,22,22	1.31	2 (15%)
3	NGW	D	2040	-	11,15,15	1.35	1 (9%)	13,22,22	1.21	1 (7%)
3	NGW	B	2044	-	11,15,15	1.12	1 (9%)	13,22,22	1.06	1 (7%)
5	EDO	A	2047	-	3,3,3	0.48	0	2,2,2	0.32	0
5	EDO	C	2047	-	3,3,3	0.50	0	2,2,2	0.22	0
5	EDO	C	2046	-	3,3,3	0.42	0	2,2,2	0.31	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	EDO	C	2045	-	-	0/1/1/1	-
3	NGW	A	2042	-	-	0/2/30/30	0/2/2/2
3	NGW	C	2041	-	-	0/2/30/30	0/2/2/2
3	NGW	D	2040	-	-	0/2/30/30	0/2/2/2
3	NGW	B	2044	-	-	0/2/30/30	0/2/2/2
5	EDO	A	2047	-	-	1/1/1/1	-
5	EDO	C	2047	-	-	1/1/1/1	-
5	EDO	C	2046	-	-	0/1/1/1	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	2041	NGW	O5-C1	4.65	1.49	1.42
3	A	2042	NGW	O5-C1	4.41	1.49	1.42
3	D	2040	NGW	O5-C1	4.06	1.48	1.42
3	B	2044	NGW	O5-C1	3.30	1.47	1.42

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	2041	NGW	C1-C2-N2	3.26	106.33	102.74
3	D	2040	NGW	C1-C2-N2	2.70	105.71	102.74
3	A	2042	NGW	C1-C2-N2	2.63	105.64	102.74
3	B	2044	NGW	C1-C2-N2	2.52	105.51	102.74
3	C	2041	NGW	C7-S1-C1	2.12	96.83	93.55

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	2047	EDO	O1-C1-C2-O2
5	C	2047	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	2040	NGW	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	416/442 (94%)	-0.36	16 (3%) 40 46	8, 15, 44, 59	14 (3%)
2	B	421/442 (95%)	-0.25	13 (3%) 49 55	12, 21, 42, 62	9 (2%)
2	C	416/442 (94%)	-0.30	13 (3%) 49 55	9, 16, 45, 63	16 (3%)
2	D	415/442 (93%)	-0.24	11 (2%) 54 60	13, 21, 47, 72	11 (2%)
All	All	1668/1768 (94%)	-0.29	53 (3%) 47 54	8, 18, 45, 72	50 (2%)

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	999	GLY	7.1
2	C	995	THR	7.0
2	D	1000	TYR	6.8
1	A	999	GLY	6.0
2	D	995	THR	5.4

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. 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(\AA^2)	Q<0.9
5	EDO	C	2045	4/4	0.73	0.19	41,43,43,44	0
5	EDO	C	2047	4/4	0.74	0.17	44,44,44,45	0
5	EDO	A	2047	4/4	0.82	0.23	58,58,59,59	0
5	EDO	C	2046	4/4	0.92	0.17	42,42,42,42	0
4	MG	A	2046	1/1	0.94	0.12	25,25,25,25	0
3	NGW	D	2040	14/14	0.96	0.09	17,20,21,21	0
3	NGW	C	2041	14/14	0.96	0.09	12,14,16,16	0
3	NGW	B	2044	14/14	0.97	0.09	17,19,20,20	0
4	MG	C	2044	1/1	0.97	0.06	26,26,26,26	0
3	NGW	A	2042	14/14	0.97	0.09	9,10,11,11	0
4	MG	C	2043	1/1	0.98	0.11	16,16,16,16	0
4	MG	A	2043	1/1	0.98	0.15	15,15,15,15	0
4	MG	A	2044	1/1	0.98	0.06	16,16,16,16	0
4	MG	B	2045	1/1	0.98	0.09	16,16,16,16	0
4	MG	C	2042	1/1	0.99	0.08	13,13,13,13	0
4	MG	A	2045	1/1	0.99	0.07	22,22,22,22	0

6.5 Other polymers [i](#)

There are no such residues in this entry.