



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 28, 2024 – 11:50 pm BST

PDB ID : 2YL5
Title : Inhibition of the pneumococcal virulence factor StrH and molecular insights into N-glycan recognition and hydrolysis
Authors : Pluvinae, B.; Higgins, M.A.; Abbott, D.W.; Robb, C.; Dalia, A.B.; Deng, L.; Weiser, J.N.; Parsons, T.B.; Fairbanks, A.J.; Vocadlo, D.J.; Boraston, A.B.
Deposited on : 2011-05-31
Resolution : 2.15 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtrriage (Phenix) : 1.13
EDS : 2.36.2
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) : 2.36.2

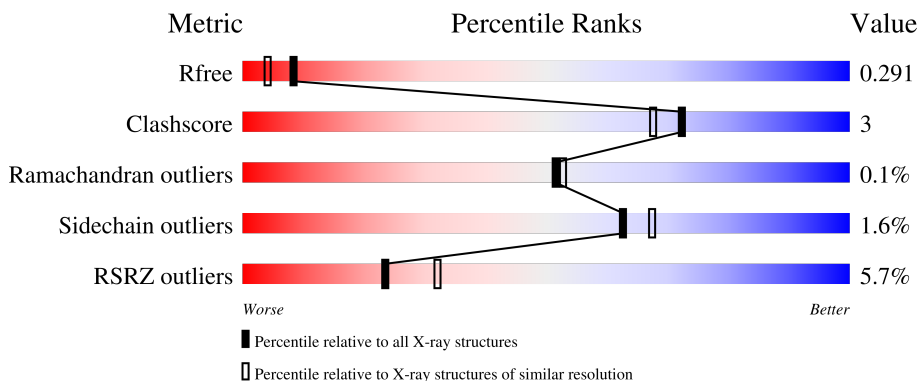
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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 $\leq 5\%$. 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	 4% 84% 8% 7%
1	B	442	 6% 87% 7% 7%
1	C	442	 5% 85% 8% 6%
1	D	442	 6% 83% 10% 8%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 15664 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	409	3320	2133	539	635	13	4	8	0
1	B	413	3340	2139	546	643	12	15	7	0
1	C	414	3343	2146	543	642	12	11	7	0
1	D	408	3282	2107	532	631	12	22	4	0

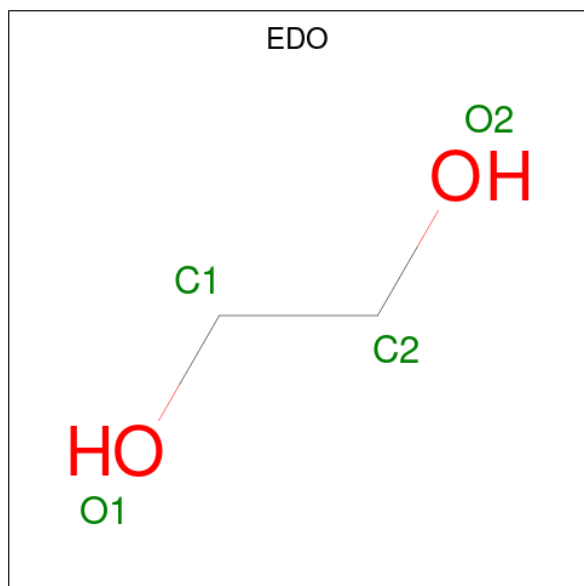
There are 16 discrepancies between the modelled and reference sequences:

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	5	Total Mg 5 5	0	0
2	B	2	Total Mg 2 2	0	0
2	C	3	Total Mg 3 3	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 8 4 4	0	1
3	A	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	C	1	Total C O 4 2 2	0	0
3	C	1	Total C O 4 2 2	0	0
3	C	1	Total C O 4 2 2	0	0
3	C	1	Total C O 4 2 2	0	0
3	D	1	Total C O 4 2 2	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	D	1	Total C O 4 2 2	0	0
3	D	1	Total C O 4 2 2	0	0
3	D	1	Total C O 4 2 2	0	0

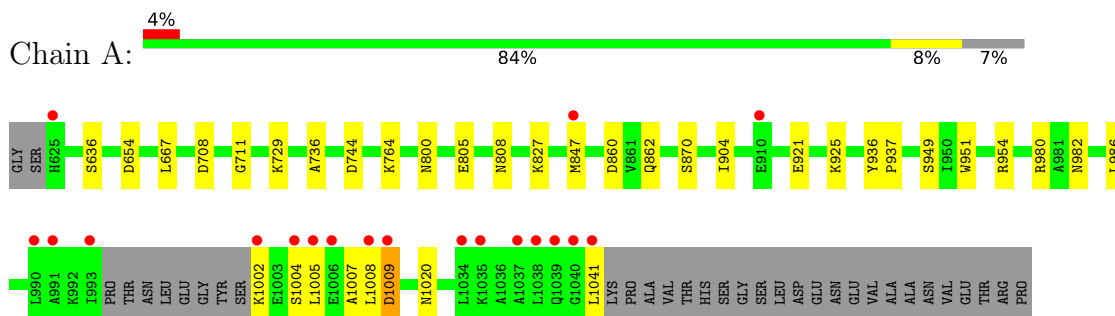
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	673	Total O 673 673	0	0
4	B	542	Total O 542 542	0	0
4	C	613	Total O 613 613	0	0
4	D	489	Total O 489 489	0	0

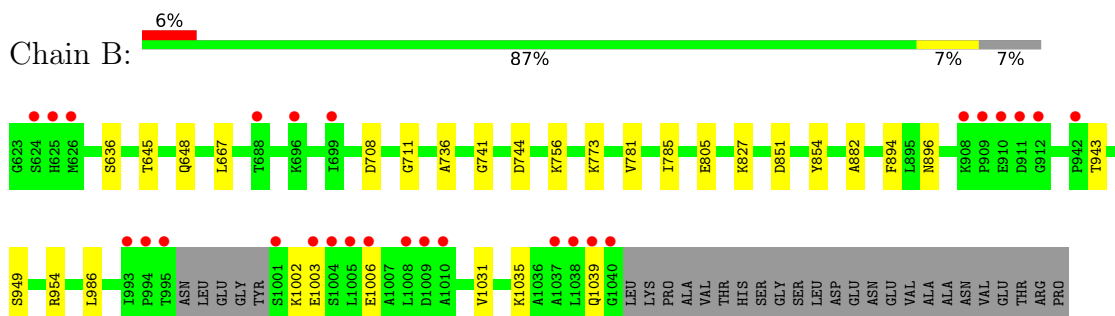
3 Residue-property plots [i](#)

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.

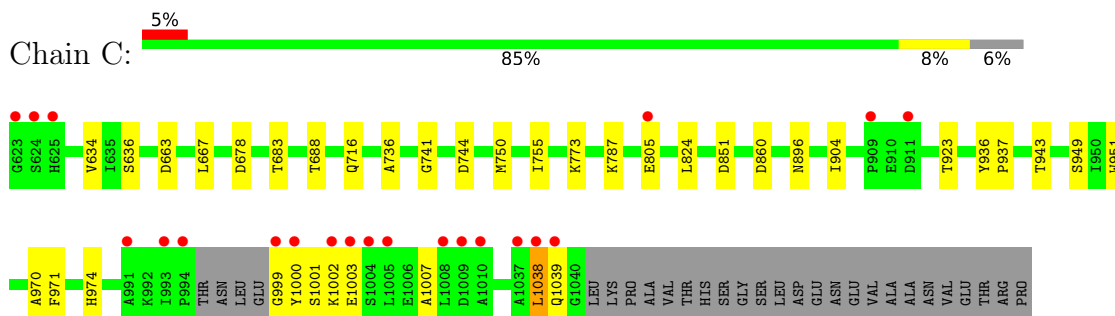
- Molecule 1: BETA-N-ACETYLHEXOSAMINIDASE



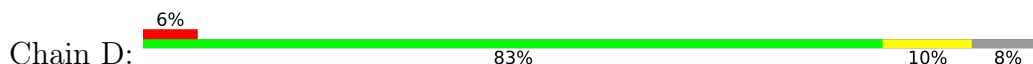
- Molecule 1: BETA-N-ACETYLHEXOSAMINIDASE

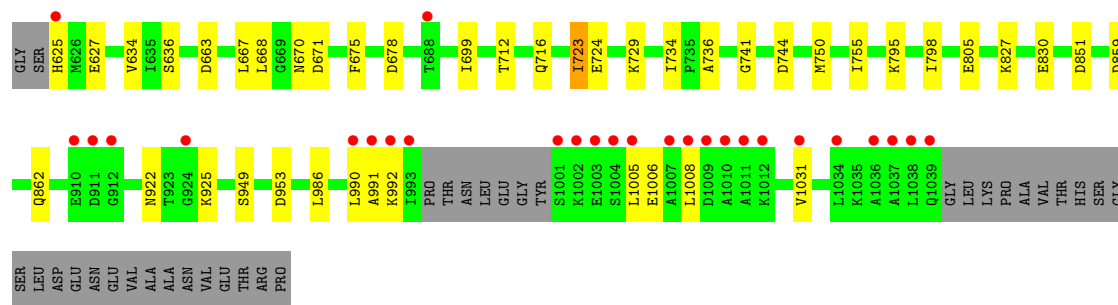


- Molecule 1: BETA-N-ACETYLHEXOSAMINIDASE



- Molecule 1: 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	67.25Å 115.86Å 132.26Å 90.00° 99.62° 90.00°	Depositor
Resolution (Å)	38.25 – 2.15 38.25 – 2.15	Depositor EDS
% Data completeness (in resolution range)	97.0 (38.25-2.15) 97.0 (38.25-2.15)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.44 (at 2.16Å)	Xtrriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.237 , 0.293 0.237 , 0.291	Depositor DCC
R_{free} test set	5261 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	13.1	Xtrriage
Anisotropy	0.204	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 53.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	15664	wwPDB-VP
Average B, all atoms (Å ²)	16.0	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 66.43 % 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 6.0551e-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, 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	0.50	3/3398 (0.1%)	0.46	1/4584 (0.0%)
1	B	0.86	5/3416 (0.1%)	0.61	4/4610 (0.1%)
1	C	0.44	2/3423 (0.1%)	0.48	4/4619 (0.1%)
1	D	0.42	5/3357 (0.1%)	0.55	5/4532 (0.1%)
All	All	0.58	15/13594 (0.1%)	0.53	14/18345 (0.1%)

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	756	LYS	CE-NZ	-38.99	0.51	1.49
1	A	764	LYS	CE-NZ	-20.01	0.99	1.49
1	B	1003	GLU	CB-CG	-15.22	1.23	1.52
1	C	1003	GLU	CG-CD	-15.20	1.29	1.51
1	B	1035	LYS	CG-CD	-12.99	1.08	1.52

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	756	LYS	CD-CE-NZ	25.23	169.73	111.70
1	D	1008	LEU	CB-CG-CD2	-11.06	92.20	111.00
1	D	1008	LEU	CA-CB-CG	-10.14	91.98	115.30
1	D	992	LYS	CD-CE-NZ	9.54	133.65	111.70
1	B	1003	GLU	CB-CG-CD	-8.57	91.06	114.20

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	3320	0	3265	26	0
1	B	3340	0	3272	18	0
1	C	3343	0	3282	24	0
1	D	3282	0	3216	21	0
2	A	5	0	0	0	0
2	B	2	0	0	0	0
2	C	3	0	0	0	0
3	A	16	0	24	3	0
3	B	4	0	6	0	0
3	C	16	0	24	0	0
3	D	16	0	24	0	0
4	A	673	0	0	0	0
4	B	542	0	0	0	0
4	C	613	0	0	0	0
4	D	489	0	0	0	0
All	All	15664	0	13113	89	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.

The worst 5 of 89 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:C:999:GLY:HA2	1:C:1000:TYR:HB2	1.26	1.10
1:C:787[B]:LYS:HA	1:C:787[B]:LYS:HE2	1.50	0.93
1:C:999:GLY:CA	1:C:1000:TYR:HB2	1.99	0.91
1:B:773[A]:LYS:HD3	1:B:827[A]:LYS:HZ2	1.41	0.86
1:D:625:HIS:HD2	1:D:627:GLU:H	1.26	0.83

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	413/442 (93%)	400 (97%)	13 (3%)	0	100	100
1	B	416/442 (94%)	400 (96%)	15 (4%)	1 (0%)	47	46
1	C	417/442 (94%)	402 (96%)	15 (4%)	0	100	100
1	D	408/442 (92%)	398 (98%)	10 (2%)	0	100	100
All	All	1654/1768 (94%)	1600 (97%)	53 (3%)	1 (0%)	51	53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	896	ASN

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%)	337 (98%)	7 (2%)	55	59
1	B	346/363 (95%)	340 (98%)	6 (2%)	60	65
1	C	346/363 (95%)	341 (99%)	5 (1%)	67	72
1	D	340/363 (94%)	333 (98%)	7 (2%)	53	57
All	All	1376/1452 (95%)	1351 (98%)	25 (2%)	62	63

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	851	ASP
1	C	1039	GLN
1	D	953	ASP
1	C	1038	LEU
1	D	723[A]	ILE

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	C	814	GLN
1	C	1039	GLN
1	D	982	ASN
1	C	929	ASN
1	D	625	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 23 ligands modelled in this entry, 10 are monoatomic - leaving 13 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	EDO	D	2043	-	3,3,3	0.45	0	2,2,2	0.34	0
3	EDO	A	2047	-	3,3,3	0.47	0	2,2,2	0.33	0
3	EDO	A	2049	-	3,3,3	0.44	0	2,2,2	0.35	0
3	EDO	A	2048[B]	-	3,3,3	0.44	0	2,2,2	0.35	0
3	EDO	C	2045	-	3,3,3	0.45	0	2,2,2	0.36	0
3	EDO	C	2044	-	3,3,3	0.45	0	2,2,2	0.39	0
3	EDO	A	2048[A]	-	3,3,3	0.45	0	2,2,2	0.34	0
3	EDO	D	2040	-	3,3,3	0.45	0	2,2,2	0.35	0
3	EDO	C	2046	-	3,3,3	0.45	0	2,2,2	0.34	0
3	EDO	D	2041	-	3,3,3	0.46	0	2,2,2	0.33	0
3	EDO	C	2047	-	3,3,3	0.46	0	2,2,2	0.34	0
3	EDO	B	2043	-	3,3,3	0.45	0	2,2,2	0.39	0
3	EDO	D	2042	-	3,3,3	0.46	0	2,2,2	0.32	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	EDO	D	2043	-	-	1/1/1/1	-
3	EDO	A	2047	-	-	1/1/1/1	-
3	EDO	A	2049	-	-	1/1/1/1	-
3	EDO	A	2048[B]	-	-	1/1/1/1	-
3	EDO	C	2045	-	-	0/1/1/1	-
3	EDO	C	2044	-	-	0/1/1/1	-
3	EDO	A	2048[A]	-	-	1/1/1/1	-
3	EDO	D	2040	-	-	1/1/1/1	-
3	EDO	C	2046	-	-	1/1/1/1	-
3	EDO	D	2041	-	-	0/1/1/1	-
3	EDO	C	2047	-	-	1/1/1/1	-
3	EDO	B	2043	-	-	0/1/1/1	-
3	EDO	D	2042	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2048[A]	EDO	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	2048[B]	EDO	O1-C1-C2-O2
3	A	2049	EDO	O1-C1-C2-O2
3	D	2043	EDO	O1-C1-C2-O2
3	A	2047	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2047	EDO	1	0
3	A	2049	EDO	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	409/442 (92%)	0.39	19 (4%) 32 42	3, 7, 29, 43	16 (3%)
1	B	413/442 (93%)	0.68	27 (6%) 18 25	13, 18, 34, 47	11 (2%)
1	C	414/442 (93%)	0.44	21 (5%) 28 36	3, 7, 33, 56	21 (5%)
1	D	408/442 (92%)	0.77	27 (6%) 18 24	12, 19, 42, 57	15 (3%)
All	All	1644/1768 (92%)	0.57	94 (5%) 23 32	3, 15, 35, 57	63 (3%)

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1000	TYR	6.6
1	D	1005	LEU	6.0
1	A	991	ALA	5.9
1	C	1004	SER	5.8
1	C	999	GLY	5.8

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(\AA^2)	Q<0.9
3	EDO	D	2041	4/4	0.49	0.33	58,58,58,58	0
2	MG	B	2041	1/1	0.73	0.22	35,35,35,35	0
3	EDO	D	2043	4/4	0.76	0.34	50,50,50,50	0
3	EDO	C	2047	4/4	0.78	0.34	34,34,34,34	0
3	EDO	A	2049	4/4	0.79	0.27	7,7,7,7	1
3	EDO	C	2046	4/4	0.82	0.24	32,32,32,32	0
3	EDO	D	2042	4/4	0.83	0.23	41,41,41,41	0
3	EDO	C	2044	4/4	0.84	0.23	10,10,10,11	0
3	EDO	A	2047	4/4	0.86	0.21	36,36,36,36	0
3	EDO	C	2045	4/4	0.87	0.18	16,16,16,16	0
3	EDO	A	2048[A]	4/4	0.87	0.29	10,10,10,10	4
3	EDO	A	2048[B]	4/4	0.87	0.29	2,2,2,2	4
2	MG	A	2045	1/1	0.89	0.17	27,27,27,27	0
3	EDO	D	2040	4/4	0.90	0.23	22,22,22,22	0
3	EDO	B	2043	4/4	0.93	0.26	26,26,26,26	0
2	MG	A	2044	1/1	0.95	0.15	4,4,4,4	0
2	MG	C	2041	1/1	0.96	0.15	2,2,2,2	0
2	MG	A	2046	1/1	0.96	0.08	9,9,9,9	0
2	MG	C	2043	1/1	0.97	0.05	20,20,20,20	0
2	MG	A	2043	1/1	0.97	0.07	2,2,2,2	0
2	MG	C	2042	1/1	0.99	0.12	6,6,6,6	0
2	MG	B	2042	1/1	0.99	0.16	14,14,14,14	0
2	MG	A	2042	1/1	0.99	0.16	6,6,6,6	0

6.5 Other polymers [i](#)

There are no such residues in this entry.