Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 12:47 pm BST

PDB ID	:	5UY1
Title	:	X-ray crystal structure of apo Halotag
Authors	:	Dunham, N.P.; Boal, A.K.
Deposited on	:	2017-02-23
Resolution	:	1.35 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 ⓘ symbol.

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

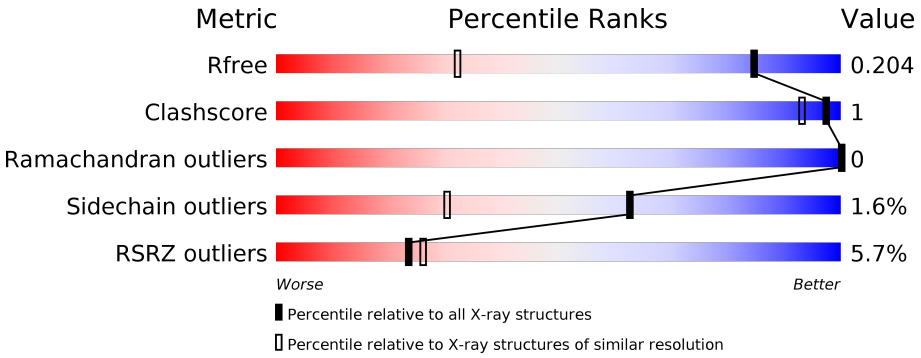
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.35 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	1509 (1.38-1.34)
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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 <=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	302	8% 93%
1	B	302	3% 93%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5124 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 Haloalkane dehalogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	290	Total	C	N	O	S	0	2	0
			2336	1523	392	412	9			
1	B	290	Total	C	N	O	S	0	0	0
			2328	1515	392	412	9			

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	47	VAL	LEU	conflict	UNP P0A3G2
A	58	THR	SER	conflict	UNP P0A3G2
A	78	GLY	ASP	conflict	UNP P0A3G2
A	87	PHE	TYR	conflict	UNP P0A3G2
A	88	MET	LEU	conflict	UNP P0A3G2
A	128	PHE	CYS	conflict	UNP P0A3G2
A	155	THR	ALA	conflict	UNP P0A3G2
A	160	LYS	GLU	conflict	UNP P0A3G2
A	167	VAL	ALA	conflict	UNP P0A3G2
A	172	THR	ALA	conflict	UNP P0A3G2
A	175	MET	LYS	conflict	UNP P0A3G2
A	176	GLY	CYS	conflict	UNP P0A3G2
A	195	ASN	LYS	conflict	UNP P0A3G2
A	224	GLU	ALA	conflict	UNP P0A3G2
A	227	ASP	ASN	conflict	UNP P0A3G2
A	257	LYS	GLU	conflict	UNP P0A3G2
A	264	ALA	THR	conflict	UNP P0A3G2
A	272	ASN	HIS	conflict	UNP P0A3G2
A	273	LEU	TYR	conflict	UNP P0A3G2
A	291	SER	-	expression tag	UNP P0A3G2
A	292	THR	-	expression tag	UNP P0A3G2
A	293	LEU	-	expression tag	UNP P0A3G2
A	294	GLU	-	expression tag	UNP P0A3G2
A	295	ILE	-	expression tag	UNP P0A3G2
A	296	SER	-	expression tag	UNP P0A3G2

Continued on next page...

Continued from previous page...

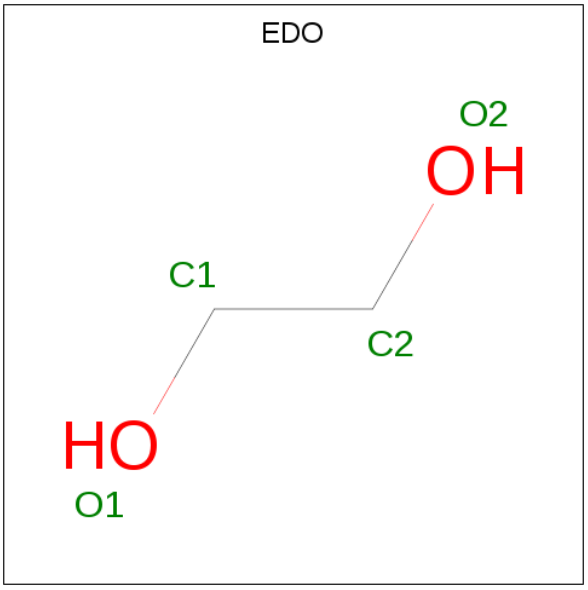
Chain	Residue	Modelled	Actual	Comment	Reference
A	297	GLY	-	expression tag	UNP P0A3G2
A	298	LEU	-	expression tag	UNP P0A3G2
A	299	GLU	-	expression tag	UNP P0A3G2
A	300	HIS	-	expression tag	UNP P0A3G2
A	301	HIS	-	expression tag	UNP P0A3G2
A	302	HIS	-	expression tag	UNP P0A3G2
A	303	HIS	-	expression tag	UNP P0A3G2
A	304	HIS	-	expression tag	UNP P0A3G2
A	305	HIS	-	expression tag	UNP P0A3G2
B	47	VAL	LEU	conflict	UNP P0A3G2
B	58	THR	SER	conflict	UNP P0A3G2
B	78	GLY	ASP	conflict	UNP P0A3G2
B	87	PHE	TYR	conflict	UNP P0A3G2
B	88	MET	LEU	conflict	UNP P0A3G2
B	128	PHE	CYS	conflict	UNP P0A3G2
B	155	THR	ALA	conflict	UNP P0A3G2
B	160	LYS	GLU	conflict	UNP P0A3G2
B	167	VAL	ALA	conflict	UNP P0A3G2
B	172	THR	ALA	conflict	UNP P0A3G2
B	175	MET	LYS	conflict	UNP P0A3G2
B	176	GLY	CYS	conflict	UNP P0A3G2
B	195	ASN	LYS	conflict	UNP P0A3G2
B	224	GLU	ALA	conflict	UNP P0A3G2
B	227	ASP	ASN	conflict	UNP P0A3G2
B	257	LYS	GLU	conflict	UNP P0A3G2
B	264	ALA	THR	conflict	UNP P0A3G2
B	272	ASN	HIS	conflict	UNP P0A3G2
B	273	LEU	TYR	conflict	UNP P0A3G2
B	291	SER	-	expression tag	UNP P0A3G2
B	292	THR	-	expression tag	UNP P0A3G2
B	293	LEU	-	expression tag	UNP P0A3G2
B	294	GLU	-	expression tag	UNP P0A3G2
B	295	ILE	-	expression tag	UNP P0A3G2
B	296	SER	-	expression tag	UNP P0A3G2
B	297	GLY	-	expression tag	UNP P0A3G2
B	298	LEU	-	expression tag	UNP P0A3G2
B	299	GLU	-	expression tag	UNP P0A3G2
B	300	HIS	-	expression tag	UNP P0A3G2
B	301	HIS	-	expression tag	UNP P0A3G2
B	302	HIS	-	expression tag	UNP P0A3G2
B	303	HIS	-	expression tag	UNP P0A3G2
B	304	HIS	-	expression tag	UNP P0A3G2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	305	HIS	-	expression tag	UNP P0A3G2

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Cl	0	0
			2	2		
3	A	1	Total	Cl	0	0
			1	1		

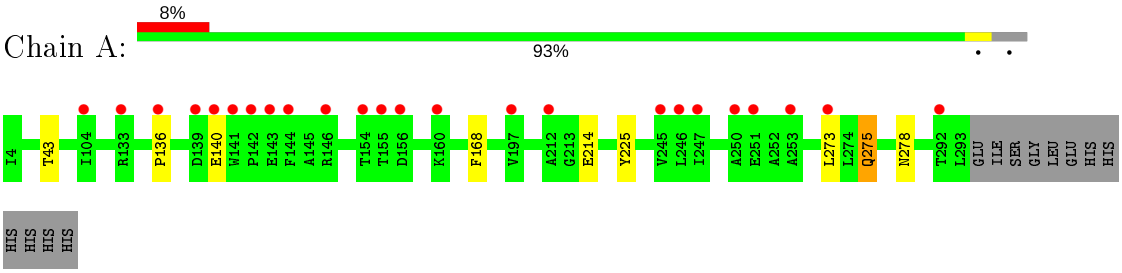
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	213	Total O 213 213	0	0
4	B	224	Total O 224 224	0	0

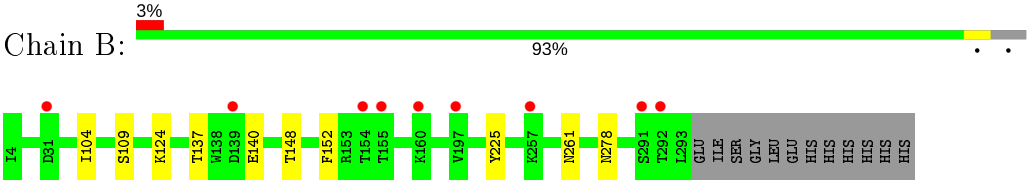
3 Residue-property plots (i)

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: Haloalkane dehalogenase



- Molecule 1: Haloalkane dehalogenase



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	69.08Å 94.63Å 99.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 1.35 28.69 – 1.35	Depositor EDS
% Data completeness (in resolution range)	97.0 (50.00-1.35) 97.1 (28.69-1.35)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.75 (at 1.35Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.188 , 0.196 0.196 , 0.204	Depositor DCC
R_{free} test set	7029 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	13.4	Xtriage
Anisotropy	0.024	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 38.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5124	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 52.90 % 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 4.5515e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

5.1 Standard geometry

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

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	0.36	2/2424 (0.1%)	0.49	0/3316
1	B	0.27	0/2407	0.48	0/3292
All	All	0.32	2/4831 (0.0%)	0.48	0/6608

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	214[A]	GLU	CG-CD	-8.07	1.39	1.51
1	A	214[B]	GLU	CG-CD	-8.07	1.39	1.51

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	2336	0	2288	3	0
1	B	2328	0	2271	4	0
2	A	12	0	18	0	0
2	B	8	0	12	0	0
3	A	1	0	0	0	0
3	B	2	0	0	0	0
4	A	213	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	224	0	0	1	0
All	All	5124	0	4589	7	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 (7) 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:137:THR:OG1	1:B:140:GLU:OE1	1.90	0.89
1:A:275:GLN:HE21	1:A:275:GLN:H	1.36	0.69
1:A:43:THR:HG21	1:A:273:LEU:HD21	1.92	0.51
1:B:148:THR:HG22	1:B:152:PHE:CZ	2.49	0.47
1:B:148:THR:HG21	4:B:531:HOH:O	2.14	0.46
1:A:136:PRO:HG2	1:A:140:GLU:OE2	2.18	0.44
1:B:104:ILE:HB	1:B:109:SER:HA	2.04	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	290/302 (96%)	281 (97%)	9 (3%)	0	100	100
1	B	288/302 (95%)	278 (96%)	10 (4%)	0	100	100
All	All	578/604 (96%)	559 (97%)	19 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains

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	252/261 (97%)	248 (98%)	4 (2%)	62 30
1	B	250/261 (96%)	246 (98%)	4 (2%)	62 30
All	All	502/522 (96%)	494 (98%)	8 (2%)	62 30

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	168	PHE
1	A	225	TYR
1	A	275	GLN
1	A	278	ASN
1	B	124	LYS
1	B	225	TYR
1	B	261	ASN
1	B	278	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	50	ASN
1	A	272	ASN
1	A	275	GLN
1	A	278	ASN
1	B	50	ASN
1	B	230	HIS
1	B	261	ASN
1	B	272	ASN
1	B	278	ASN

5.3.3 RNA

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 8 ligands modelled in this entry, 3 are monoatomic - leaving 5 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	$\#\|Z\| > 2$
2	EDO	A	403	-	3,3,3	0.46	0	2,2,2	0.30	0
2	EDO	A	401	-	3,3,3	0.47	0	2,2,2	0.16	0
2	EDO	B	401	-	3,3,3	0.49	0	2,2,2	0.19	0
2	EDO	A	402	-	3,3,3	0.45	0	2,2,2	0.32	0
2	EDO	B	402	-	3,3,3	0.46	0	2,2,2	0.11	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	EDO	A	403	-	-	0/1/1/1	-
2	EDO	A	401	-	-	0/1/1/1	-
2	EDO	B	401	-	-	0/1/1/1	-
2	EDO	A	402	-	-	0/1/1/1	-
2	EDO	B	402	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	402	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	290/302 (96%)	0.36	24 (8%) 11 12	9, 13, 24, 28	37 (12%)
1	B	290/302 (96%)	0.11	9 (3%) 49 56	9, 14, 18, 23	38 (13%)
All	All	580/604 (96%)	0.24	33 (5%) 23 26	9, 13, 22, 28	75 (12%)

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	155	THR	7.5
1	A	139	ASP	4.6
1	B	155	THR	4.3
1	A	144	PHE	4.3
1	A	154	THR	3.8
1	A	156	ASP	3.7
1	B	292	THR	3.3
1	A	133	ARG	3.3
1	A	292	THR	3.1
1	A	250	ALA	3.0
1	A	140	GLU	2.9
1	A	141	TRP	2.7
1	A	245	VAL	2.7
1	A	146	ARG	2.6
1	A	142	PRO	2.6
1	A	160	LYS	2.6
1	A	197	VAL	2.6
1	A	246	LEU	2.5
1	A	212	ALA	2.5
1	B	160	LYS	2.4
1	B	197	VAL	2.4
1	A	247	ILE	2.4
1	A	251	GLU	2.3
1	B	154	THR	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	253	ALA	2.2
1	A	104	ILE	2.1
1	A	273	LEU	2.1
1	A	143	GLU	2.1
1	B	31	ASP	2.1
1	B	139	ASP	2.0
1	B	257	LYS	2.0
1	B	291	SER	2.0
1	A	136	PRO	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. The B-factors column lists the minimum, median, 95^{th} 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	EDO	B	401	4/4	0.61	0.16	27,27,27,27	0
2	EDO	B	402	4/4	0.65	0.26	23,23,23,24	0
2	EDO	A	403	4/4	0.75	0.19	28,28,28,28	0
2	EDO	A	402	4/4	0.78	0.21	28,29,29,29	0
3	CL	B	404	1/1	0.85	0.06	37,37,37,37	0
2	EDO	A	401	4/4	0.94	0.08	14,14,14,14	0
3	CL	A	404	1/1	1.00	0.04	11,11,11,11	0
3	CL	B	403	1/1	1.00	0.03	12,12,12,12	0

6.5 Other polymers (i)

There are no such residues in this entry.