wwPDB X-ray Structure Validation Summary Report (i)

Mar 9, 2026 – 06:17 AM UTC

PDB ID	:	5UHK / pdb_00005uhk
Title	:	Crystal structure of the core catalytic domain of Human O-GlcNAcase
Authors	:	Klein, D.J.; Elsen, N.L.
Deposited on	:	2017-01-11
Resolution	:	2.97 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 (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 (1)) were used in the production of this report:

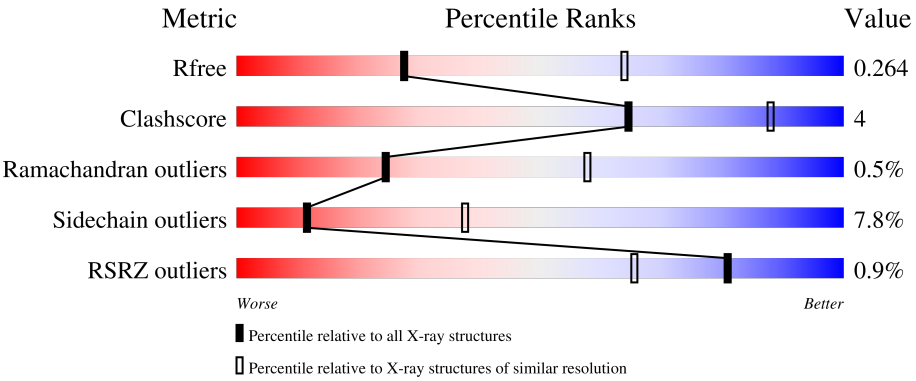
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

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 2.97 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}	180053	3580 (3.00-2.96)
Clashscore	190562	3904 (3.00-2.96)
Ramachandran outliers	187476	3761 (3.00-2.96)
Sidechain outliers	187428	3764 (3.00-2.96)
RSRZ outliers	180081	3579 (3.00-2.96)

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	345	73% 14% • 12%
1	C	345	% 69% 18% • 12%
2	B	163	% 54% 18% • 26%
2	D	163	2% 55% 17% • 26%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6936 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 O-GlcNAcase TIM-barrel domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	304	Total	C	N	O	S	0	0	0
			2498	1616	413	455	14			
1	C	302	Total	C	N	O	S	0	0	0
			2471	1601	405	451	14			

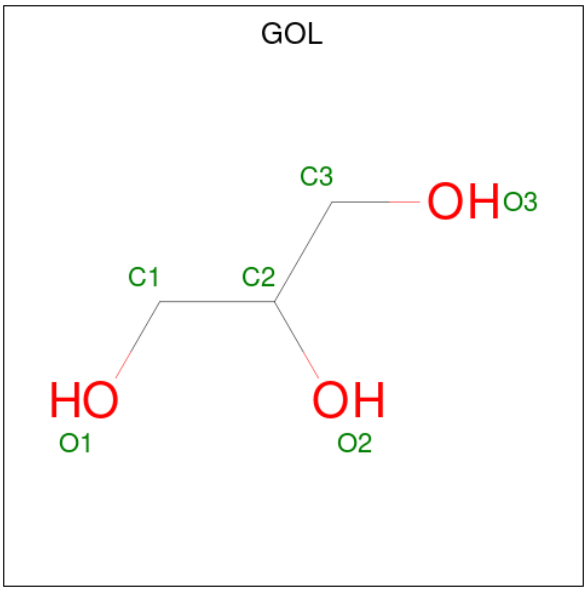
- Molecule 2 is a protein called O-GlcNAcase stalk domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	120	Total	C	N	O	S	0	0	0
			970	627	160	172	11			
2	D	120	Total	C	N	O	S	0	0	0
			973	629	160	173	11			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	543	MET	-	initiating methionine	UNP O60502
D	543	MET	-	initiating methionine	UNP O60502

- Molecule 3 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

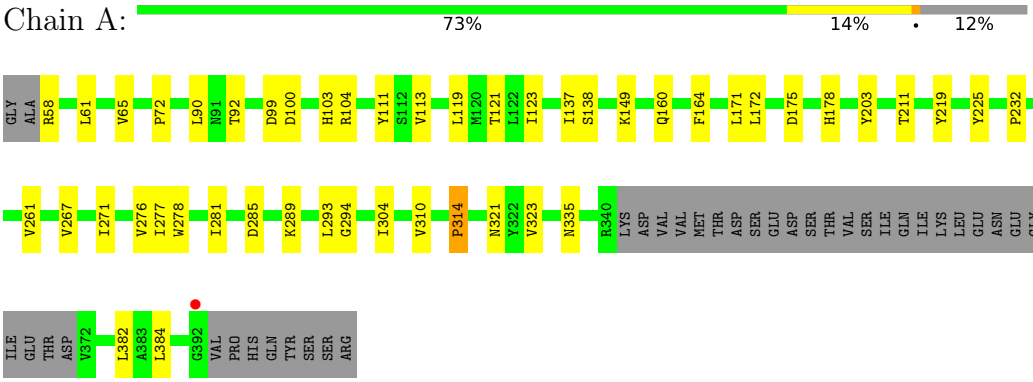
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O 1 1	0	0
4	B	1	Total O 1 1	0	0
4	C	4	Total O 4 4	0	0

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.

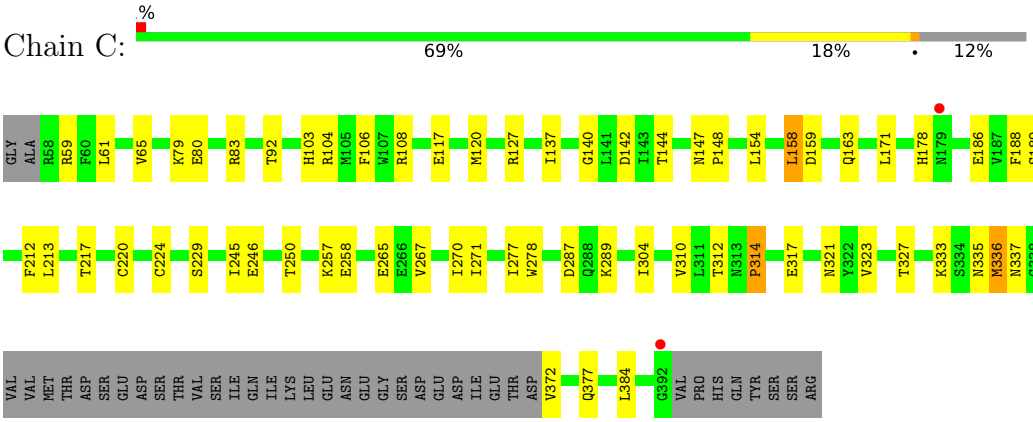
• Molecule 1: O-GlcNAcase TIM-barrel domain

Chain A:



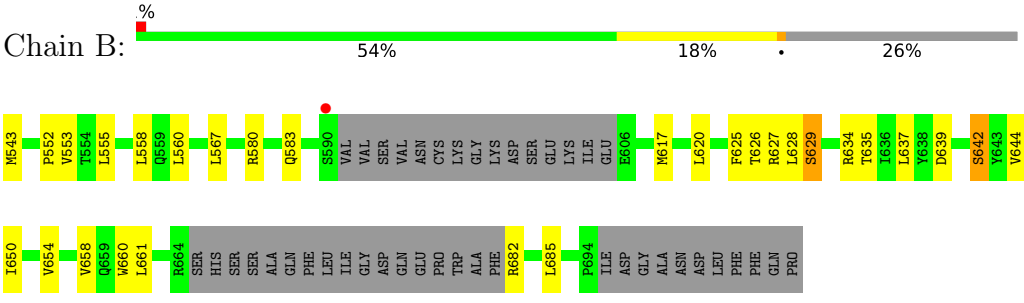
• Molecule 1: O-GlcNAcase TIM-barrel domain

Chain C:



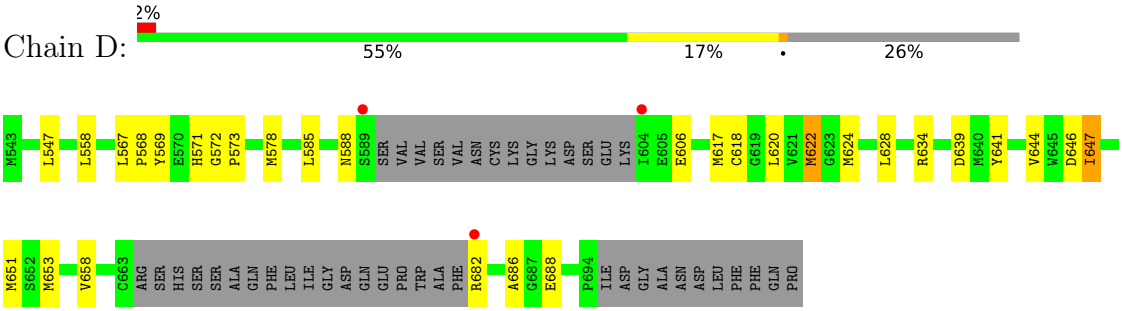
• Molecule 2: O-GlcNAcase stalk domain

Chain B:



• Molecule 2: O-GlcNAcase stalk domain

Chain D:


2%
55%
17%
26%
M543 L547 L558 L567 P568 Y569 E570 H571 G572 P573 M578 L585 N588 S589 SER VAL VAL SER VAL ASN CYS LYS GLY LYS ASP SER GLU LYS I604 E605 E606 M617 C618 G619 L620 V621 M622 G623 M624 L628 R634 D639 M640 Y641 V644 W645 D646 I647
M651 S652 M653 V658 C663 ARG SER HIS SER SER ALA GLN PHE LEU ILE GLY ASP GLN GLU PRO TRP ALA PHE R682 A686 G687 E688 P694 ILE ASP GLY ALA ASN ASP LEU PHE PHE GLN PRO

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	96.80Å 96.80Å 256.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.24 – 2.97 34.24 – 2.97	Depositor EDS
% Data completeness (in resolution range)	95.2 (34.24-2.97) 95.5 (34.24-2.97)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.08 (at 2.95Å)	Xtriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.194 , 0.251 0.208 , 0.264	Depositor DCC
R_{free} test set	1264 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	91.6	Xtriage
Anisotropy	0.457	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 75.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6936	wwPDB-VP
Average B, all atoms (Å^2)	102.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.18% 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

5.1 Standard geometry

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

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	0.90	1/2565 (0.0%)	1.41	17/3475 (0.5%)
1	C	0.90	1/2538 (0.0%)	1.44	18/3441 (0.5%)
2	B	0.84	0/993	1.46	5/1337 (0.4%)
2	D	0.82	0/996	1.41	3/1341 (0.2%)
All	All	0.88	2/7092 (0.0%)	1.43	43/9594 (0.4%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	104	ARG	CA-C	7.72	1.56	1.52
1	C	104	ARG	CA-C	5.06	1.54	1.52

The worst 5 of 43 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	278	TRP	N-CA-C	-8.60	94.57	108.41
1	A	289	LYS	N-CA-C	-6.78	104.63	113.17
1	A	164	PHE	CA-CB-CG	6.75	120.55	113.80
1	C	336	MET	CA-C-N	6.35	131.21	122.77
1	C	336	MET	C-N-CA	6.35	131.21	122.77

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	2498	0	2453	15	0
1	C	2471	0	2420	19	0
2	B	970	0	948	14	0
2	D	973	0	949	12	0
3	A	12	0	16	2	0
3	C	6	0	8	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	4	0	0	0	0
All	All	6936	0	6794	54	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 4.

The worst 5 of 54 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:333:LYS:HA	1:C:336:MET:HG3	1.63	0.80
1:A:242:LEU:HD23	1:A:243:PRO:HD2	1.72	0.71
1:C:257:LYS:HB2	2:D:569:TYR:CD2	2.35	0.62
1:C:304:ILE:HD12	1:C:335:ASN:HB3	1.82	0.61
2:B:617:MET:HA	2:B:620:LEU:HD12	1.83	0.61

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	300/345 (87%)	280 (93%)	19 (6%)	1 (0%)	36 67
1	C	298/345 (86%)	274 (92%)	22 (7%)	2 (1%)	18 50

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	B	114/163 (70%)	109 (96%)	5 (4%)	0	100 100
2	D	114/163 (70%)	108 (95%)	5 (4%)	1 (1%)	14 45
All	All	826/1016 (81%)	771 (93%)	51 (6%)	4 (0%)	24 58

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	641	TYR
1	A	314	PRO
1	C	314	PRO
1	C	140	GLY

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	272/310 (88%)	256 (94%)	16 (6%)	18 48
1	C	268/310 (86%)	248 (92%)	20 (8%)	12 39
2	B	102/141 (72%)	90 (88%)	12 (12%)	5 21
2	D	102/141 (72%)	92 (90%)	10 (10%)	7 28
All	All	744/902 (82%)	686 (92%)	58 (8%)	11 37

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	59	ARG
2	D	646	ASP
1	C	137	ILE
2	D	644	VAL
2	D	588	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	193	HIS
1	C	227	ASN
2	D	690	GLN
2	D	583	GLN
2	D	630	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

3 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	$\#\|Z\| > 2$
3	GOL	A	501	-	5,5,5	0.17	0	5,5,5	0.26	0
3	GOL	C	501	-	5,5,5	0.14	0	5,5,5	0.31	0
3	GOL	A	502	-	5,5,5	0.11	0	5,5,5	0.25	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	GOL	A	501	-	-	2/4/4/4	-
3	GOL	C	501	-	-	3/4/4/4	-
3	GOL	A	502	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	501	GOL	O1-C1-C2-C3
3	A	502	GOL	C1-C2-C3-O3
3	C	501	GOL	C1-C2-C3-O3
3	A	501	GOL	O1-C1-C2-O2
3	A	502	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	GOL	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

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	304/345 (88%)	-0.26	1 (0%)	90	82	64, 96, 126, 169	0
1	C	302/345 (87%)	-0.19	3 (0%)	79	63	66, 104, 137, 160	0
2	B	120/163 (73%)	0.05	1 (0%)	82	67	69, 104, 151, 172	0
2	D	120/163 (73%)	-0.00	3 (2%)	58	39	69, 103, 146, 162	0
All	All	846/1016 (83%)	-0.15	8 (0%)	81	65	64, 101, 142, 172	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	604	ILE	3.2
2	B	590	SER	2.8
2	D	589	SER	2.5
1	C	392	GLY	2.4
1	A	392	GLY	2.3

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

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
3	GOL	C	501	6/6	0.88	0.17	89,100,106,109	0
3	GOL	A	502	6/6	0.93	0.12	83,91,95,97	0
3	GOL	A	501	6/6	0.93	0.14	84,88,92,94	0

6.5 Other polymers (i)

There are no such residues in this entry.