



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 20, 2017 – 03:54 PM EDT

PDB ID : 5UHK
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 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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-20029077
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-20029077

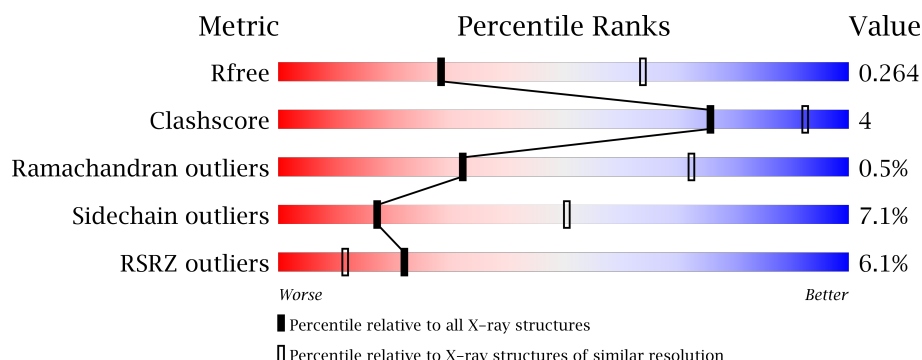
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.97 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}	100719	2168 (3.00-2.96)
Clashscore	112137	2535 (3.00-2.96)
Ramachandran outliers	110173	2451 (3.00-2.96)
Sidechain outliers	110143	2454 (3.00-2.96)
RSRZ outliers	101464	2192 (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 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 $\leq 5\%$. 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	345	<div> <div>2%</div> <div>77%</div> <div>10%</div> <div>•</div> <div>12%</div> </div>
1	C	345	<div> <div>3%</div> <div>74%</div> <div>13%</div> <div>•</div> <div>12%</div> </div>
2	B	163	<div> <div>11%</div> <div>56%</div> <div>16%</div> <div>•</div> <div>26%</div> </div>
2	D	163	<div> <div>11%</div> <div>59%</div> <div>14%</div> <div>•</div> <div>26%</div> </div>

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	GOL	C	501	-	-	-	X

2 Entry composition

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 (three-letter code: GOL) (formula: C₃H₈O₃).



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

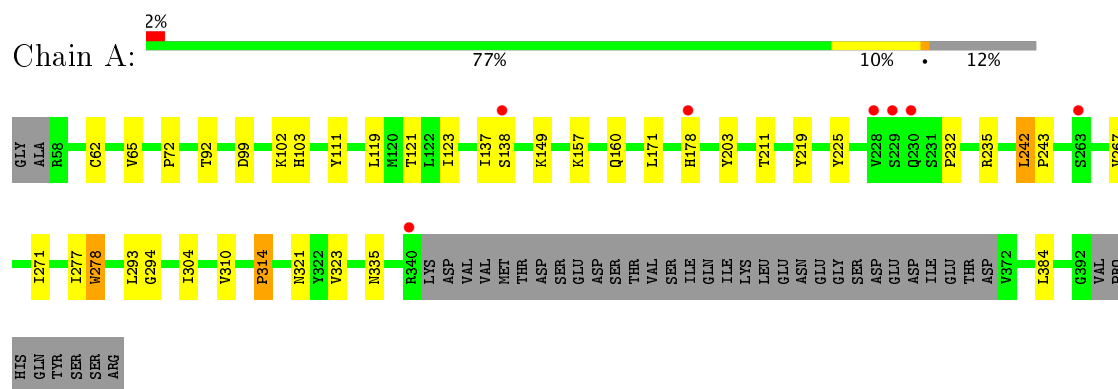
- Molecule 4 is water.

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

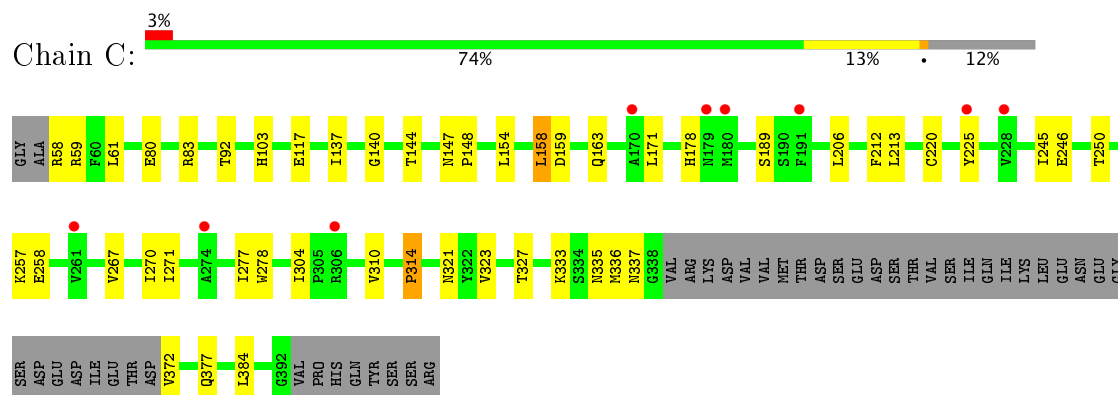
3 Residue-property plots [i](#)

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.

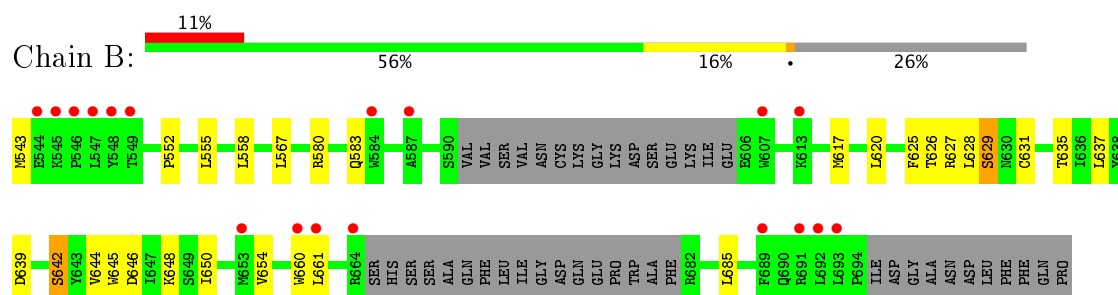
• Molecule 1: O-GlcNAcase TIM-barrel domain



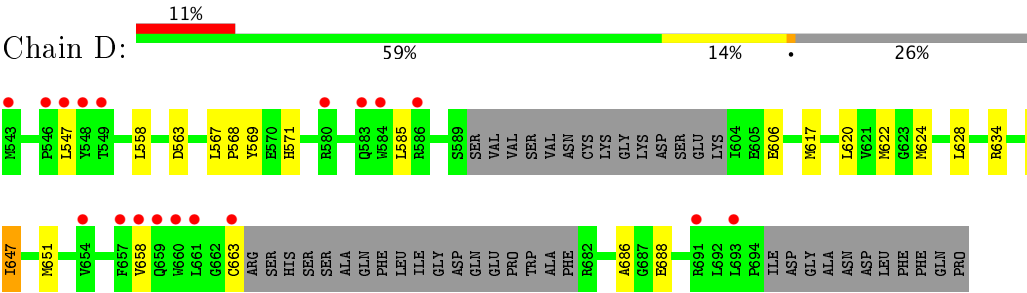
• Molecule 1: O-GlcNAcase TIM-barrel domain



• Molecule 2: O-GlcNAcase stalk domain



• Molecule 2: O-GlcNAcase stalk domain



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	96.80 Å 96.80 Å 256.96 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.24 – 2.97 34.22 – 2.97	Depositor EDS
% Data completeness (in resolution range)	95.2 (34.24-2.97) 95.6 (34.22-2.97)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.08 (at 2.95 Å)	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 (5.36%)	DCC
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 , 77.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	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.*

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: 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	0.54	0/2565	0.72	0/3475
1	C	0.53	0/2538	0.72	1/3441 (0.0%)
2	B	0.52	0/993	0.67	0/1337
2	D	0.50	0/996	0.65	0/1341
All	All	0.53	0/7092	0.70	1/9594 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	278	TRP	N-CA-C	-6.09	94.57	111.00

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	2498	0	2453	16	0
1	C	2471	0	2420	17	0
2	B	970	0	948	14	0
2	D	973	0	949	9	0
3	A	12	0	16	2	0
3	C	6	0	8	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
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	50	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.

All (50) 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
1:C:80:GLU:HA	1:C:83:ARG:HD2	1.83	0.60
1:C:314:PRO:HB2	1:C:321:ASN:OD1	2.04	0.57
2:D:585:LEU:HD13	2:D:658:VAL:HG11	1.88	0.55
2:B:629:SER:HA	2:B:637:LEU:HD22	1.88	0.55
1:A:111:TYR:H	1:A:160:GLN:HE22	1.58	0.51
2:B:625:PHE:CZ	2:B:645:TRP:HE3	2.28	0.51
2:B:650:ILE:HG13	2:D:686:ALA:HB1	1.92	0.51
1:C:220:CYS:HB3	1:C:250:THR:OG1	2.11	0.50
1:A:304:ILE:HD12	1:A:335:ASN:HB3	1.95	0.49
1:C:158:LEU:HD22	1:C:206:LEU:HD11	1.95	0.49
2:B:552:PRO:O	2:B:627:ARG:HD2	2.13	0.48
1:A:225:TYR:O	1:A:232:PRO:HD2	2.13	0.48
2:D:617:MET:HA	2:D:620:LEU:HD12	1.96	0.48
1:A:219:TYR:CE1	3:A:501:GOL:H32	2.48	0.48
2:D:647:ILE:O	2:D:651:MET:HB2	2.14	0.48
2:B:580:ARG:HA	2:B:583:GLN:HE21	1.79	0.47
1:A:72:PRO:HG3	1:A:99:ASP:HB3	1.95	0.47
2:B:625:PHE:HD1	2:B:644:VAL:HG12	1.80	0.47
1:C:154:LEU:O	1:C:158:LEU:HD12	2.15	0.47
1:C:212:PHE:HD1	1:C:245:ILE:HG23	1.80	0.46
1:C:277:ILE:HB	1:C:310:VAL:HG22	1.98	0.46
1:A:203:TYR:HD2	1:A:242:LEU:HD13	1.80	0.46
1:A:267:VAL:O	1:A:271:ILE:HG12	2.15	0.46
2:D:567:LEU:HB3	2:D:568:PRO:HD2	1.97	0.46
1:C:213:LEU:HD23	1:C:246:GLU:HB2	1.97	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:629:SER:HA	2:B:637:LEU:CD2	2.47	0.45
1:A:278:TRP:CZ2	3:A:501:GOL:H12	2.52	0.45
2:B:642:SER:O	2:B:646:ASP:HB2	2.17	0.44
1:C:159:ASP:O	1:C:163:GLN:HG2	2.19	0.43
1:A:384:LEU:HD22	2:B:558:LEU:HB3	2.00	0.43
2:B:685:LEU:HD13	2:D:571:HIS:CE1	2.53	0.43
1:A:119:LEU:O	1:A:123:ILE:HG13	2.19	0.43
1:A:294:GLY:HA3	2:B:567:LEU:HG	2.01	0.43
1:A:102:LYS:HB3	1:A:157:LYS:HZ1	1.83	0.43
1:C:314:PRO:HB2	1:C:321:ASN:CG	2.39	0.42
1:C:267:VAL:O	1:C:271:ILE:HG12	2.19	0.42
1:C:384:LEU:HD22	2:D:558:LEU:HB3	2.00	0.42
2:D:624:MET:HG2	2:D:628:LEU:HD12	2.02	0.42
1:A:277:ILE:HB	1:A:310:VAL:HG22	2.01	0.41
1:A:314:PRO:HB2	1:A:321:ASN:CG	2.40	0.41
1:C:147:ASN:HA	1:C:148:PRO:HD3	1.92	0.41
1:A:293:LEU:O	1:A:323:VAL:HG11	2.21	0.40
1:C:323:VAL:O	1:C:327:THR:HG23	2.20	0.40
2:B:646:ASP:O	2:B:650:ILE:HG12	2.21	0.40
2:B:660:TRP:HE3	2:B:661:LEU:HG	1.86	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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%)	44	80
1	C	298/345 (86%)	274 (92%)	22 (7%)	2 (1%)	25	65
2	B	114/163 (70%)	109 (96%)	5 (4%)	0	100	100
2	D	114/163 (70%)	108 (95%)	5 (4%)	1 (1%)	20	60
All	All	826/1016 (81%)	771 (93%)	51 (6%)	4 (0%)	32	72

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 ⓘ

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%)	258 (95%)	14 (5%)	28	65
1	C	268/310 (86%)	250 (93%)	18 (7%)	19	53
2	B	102/141 (72%)	91 (89%)	11 (11%)	7	28
2	D	102/141 (72%)	92 (90%)	10 (10%)	9	33
All	All	744/902 (82%)	691 (93%)	53 (7%)	17	50

All (53) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	62	CYS
1	A	65	VAL
1	A	92	THR
1	A	103	HIS
1	A	121	THR
1	A	137	ILE
1	A	138	SER
1	A	149	LYS
1	A	171	LEU
1	A	178	HIS
1	A	211	THR
1	A	235	ARG
1	A	242	LEU
1	A	278	TRP
2	B	543	MET
2	B	555	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	626	THR
2	B	628	LEU
2	B	629	SER
2	B	631	CYS
2	B	635	THR
2	B	639	ASP
2	B	642	SER
2	B	648	LYS
2	B	654	VAL
1	C	58	ARG
1	C	59	ARG
1	C	61	LEU
1	C	92	THR
1	C	103	HIS
1	C	117	GLU
1	C	137	ILE
1	C	144	THR
1	C	158	LEU
1	C	171	LEU
1	C	178	HIS
1	C	189	SER
1	C	225	TYR
1	C	258	GLU
1	C	270	ILE
1	C	337	ASN
1	C	372	VAL
1	C	377	GLN
2	D	547	LEU
2	D	563	ASP
2	D	606	GLU
2	D	622	MET
2	D	634	ARG
2	D	639	ASP
2	D	646	ASP
2	D	647	ILE
2	D	663	CYS
2	D	688	GLU

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	86	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	160	GLN
1	A	193	HIS
2	B	583	GLN
1	C	118	GLN
1	C	193	HIS
1	C	227	ASN
1	C	230	GLN
2	D	690	GLN

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](#)

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	0.21	0	5,5,5	0.25	0
3	GOL	A	502	-	5,5,5	0.15	0	5,5,5	0.24	0
3	GOL	C	501	-	5,5,5	0.18	0	5,5,5	0.29	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	-	-	0/4/4/4	0/0/0/0
3	GOL	A	502	-	-	0/4/4/4	0/0/0/0
3	GOL	C	501	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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, 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	304/345 (88%)	0.24	7 (2%) 61 39	64, 96, 126, 169	0
1	C	302/345 (87%)	0.30	9 (2%) 51 30	66, 104, 137, 160	0
2	B	120/163 (73%)	0.61	18 (15%) 3 1	69, 104, 151, 172	0
2	D	120/163 (73%)	0.75	18 (15%) 3 1	69, 103, 146, 162	0
All	All	846/1016 (83%)	0.39	52 (6%) 22 12	64, 101, 142, 172	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	179	ASN	6.6
2	D	548	TYR	5.5
2	B	692	LEU	5.4
2	D	549	THR	5.1
2	D	547	LEU	4.7
2	D	658	VAL	4.6
1	C	180	MET	4.4
2	D	583	GLN	4.2
2	B	691	ARG	3.8
2	D	657	PHE	3.8
2	D	543	MET	3.6
2	B	584	TRP	3.6
2	D	546	PRO	3.6
2	B	653	MET	3.5
1	A	340	ARG	3.5
2	B	549	THR	3.4
2	D	663	CYS	3.4
2	B	660	TRP	3.3
2	B	548	TYR	3.2
2	B	693	LEU	3.1
2	B	545	LYS	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	544	GLU	3.1
2	D	693	LEU	2.9
2	B	607	TRP	2.9
1	C	191	PHE	2.9
1	C	261	VAL	2.8
2	D	661	LEU	2.8
1	C	306	ARG	2.8
1	A	230	GLN	2.7
1	A	138	SER	2.7
2	B	547	LEU	2.6
2	D	580	ARG	2.6
1	A	263	SER	2.6
2	D	654	VAL	2.5
1	A	228	VAL	2.5
2	B	613	LYS	2.5
2	D	660	TRP	2.5
1	A	178	HIS	2.4
2	D	584	TRP	2.4
1	A	229	SER	2.4
2	B	546	PRO	2.4
1	C	274	ALA	2.3
1	C	228	VAL	2.3
1	C	225	TYR	2.3
1	C	170	ALA	2.2
2	D	659	GLN	2.2
2	D	691	ARG	2.2
2	B	587	ALA	2.1
2	B	661	LEU	2.0
2	B	689	PHE	2.1
2	D	586	ARG	2.0
2	B	664	ARG	2.0

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 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
3	GOL	C	501	6/6	0.89	0.45	4.01	89,100,106,109	0
3	GOL	A	501	6/6	0.94	0.33	1.27	84,88,92,94	0
3	GOL	A	502	6/6	0.90	0.31	1.11	83,91,95,97	0

6.5 Other polymers [i](#)

There are no such residues in this entry.