wwPDB X-ray Structure Validation Summary Report (i)

Jun 13, 2020 – 09:52 pm BST

PDB ID	:	5UHP
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	:	2.79 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 following versions of software and data (see references (1)) were used in the production of this report:

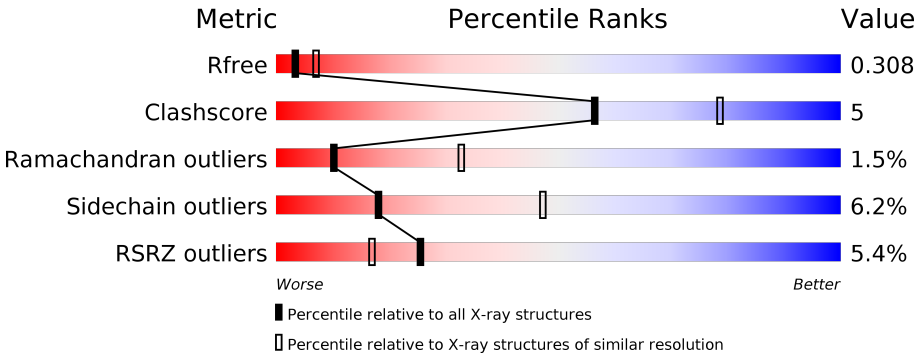
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.79 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.



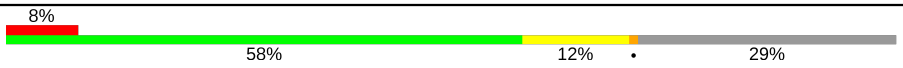
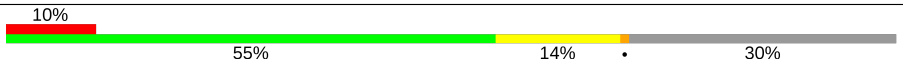
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	388	2% 65% 13% • 21%
1	B	388	% 67% 12% • 21%
1	C	388	2% 65% 14% • 20%
1	D	388	4% 60% 17% • 20%
2	E	161	9% 53% 15% • 30%
2	F	161	7% 53% 13% • 32%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	161	
2	H	161	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13813 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	307	Total	C	N	O	S	0	0	0
			2522	1634	414	460	14			
1	B	307	Total	C	N	O	S	0	0	0
			2518	1631	413	460	14			
1	C	310	Total	C	N	O	S	0	0	0
			2515	1626	416	459	14			
1	D	310	Total	C	N	O	S	0	0	0
			2512	1624	414	460	14			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	13	GLY	-	expression tag	UNP O60502
B	13	GLY	-	expression tag	UNP O60502
C	13	GLY	-	expression tag	UNP O60502
D	13	GLY	-	expression tag	UNP O60502

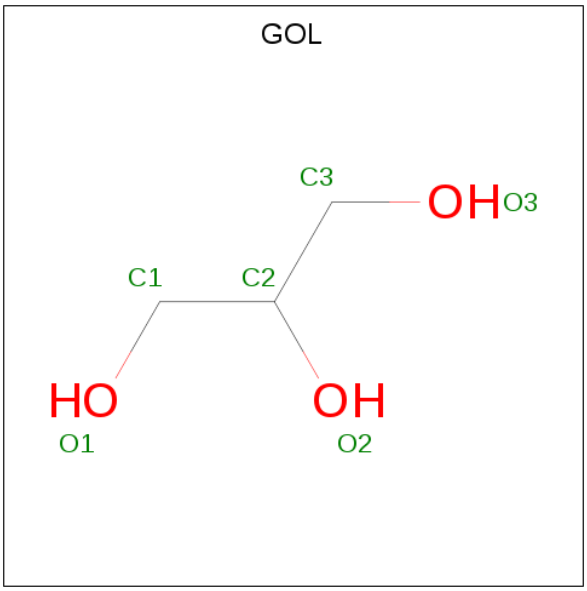
- Molecule 2 is a protein called O-GlcNAcase stalk domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	112	Total	C	N	O	S	0	0	0
			912	588	152	163	9			
2	F	109	Total	C	N	O	S	0	0	0
			896	580	149	158	9			
2	G	114	Total	C	N	O	S	0	0	0
			949	616	159	165	9			
2	H	112	Total	C	N	O	S	0	0	0
			928	600	156	163	9			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	553	MET	-	initiating methionine	UNP O60502
E	706	HIS	-	expression tag	UNP O60502
E	707	HIS	-	expression tag	UNP O60502
E	708	HIS	-	expression tag	UNP O60502
E	709	HIS	-	expression tag	UNP O60502
E	710	HIS	-	expression tag	UNP O60502
E	711	HIS	-	expression tag	UNP O60502
E	712	HIS	-	expression tag	UNP O60502
E	713	HIS	-	expression tag	UNP O60502
F	553	MET	-	initiating methionine	UNP O60502
F	706	HIS	-	expression tag	UNP O60502
F	707	HIS	-	expression tag	UNP O60502
F	708	HIS	-	expression tag	UNP O60502
F	709	HIS	-	expression tag	UNP O60502
F	710	HIS	-	expression tag	UNP O60502
F	711	HIS	-	expression tag	UNP O60502
F	712	HIS	-	expression tag	UNP O60502
F	713	HIS	-	expression tag	UNP O60502
G	553	MET	-	initiating methionine	UNP O60502
G	706	HIS	-	expression tag	UNP O60502
G	707	HIS	-	expression tag	UNP O60502
G	708	HIS	-	expression tag	UNP O60502
G	709	HIS	-	expression tag	UNP O60502
G	710	HIS	-	expression tag	UNP O60502
G	711	HIS	-	expression tag	UNP O60502
G	712	HIS	-	expression tag	UNP O60502
G	713	HIS	-	expression tag	UNP O60502
H	553	MET	-	initiating methionine	UNP O60502
H	706	HIS	-	expression tag	UNP O60502
H	707	HIS	-	expression tag	UNP O60502
H	708	HIS	-	expression tag	UNP O60502
H	709	HIS	-	expression tag	UNP O60502
H	710	HIS	-	expression tag	UNP O60502
H	711	HIS	-	expression tag	UNP O60502
H	712	HIS	-	expression tag	UNP O60502
H	713	HIS	-	expression tag	UNP O60502

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is water.

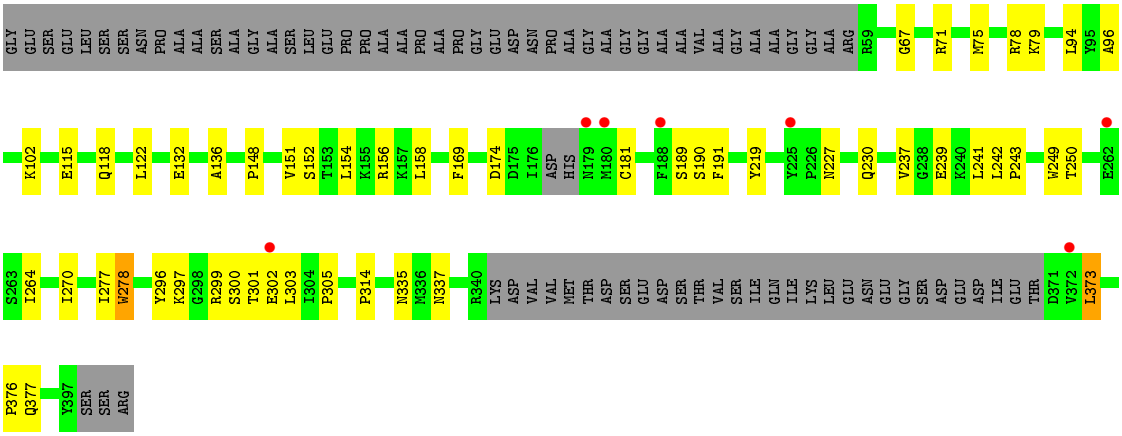
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	5	Total O 5 5	0	0
4	B	11	Total O 11 11	0	0
4	C	8	Total O 8 8	0	0
4	D	3	Total O 3 3	0	0
4	E	3	Total O 3 3	0	0
4	F	3	Total O 3 3	0	0
4	H	4	Total O 4 4	0	0

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.

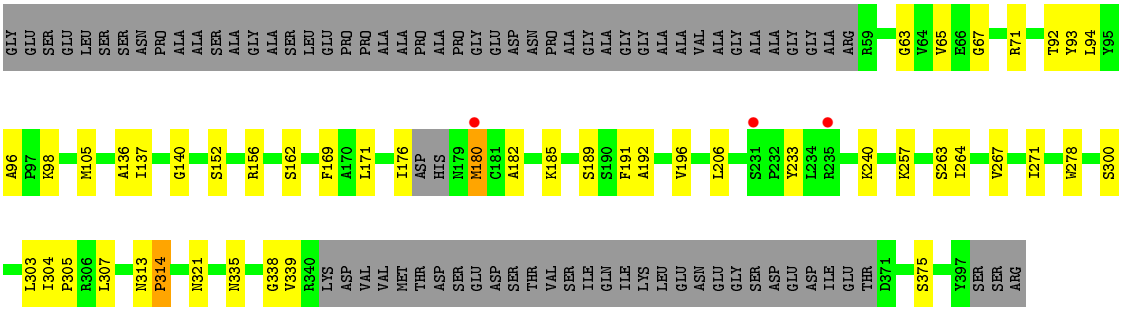
- Molecule 1: O-GlcNAcase TIM-barrel domain

Chain A:



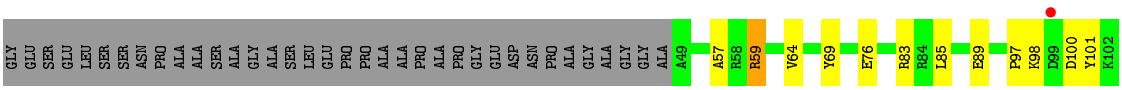
- Molecule 1: O-GlcNAcase TIM-barrel domain

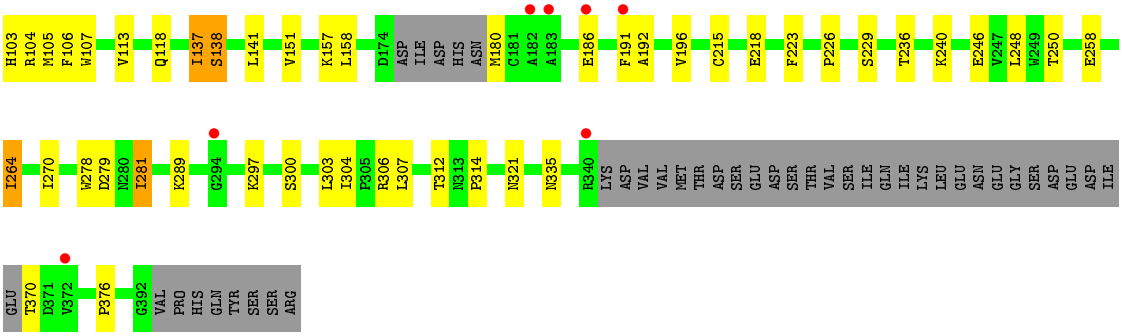
Chain B:



- Molecule 1: O-GlcNAcase TIM-barrel domain

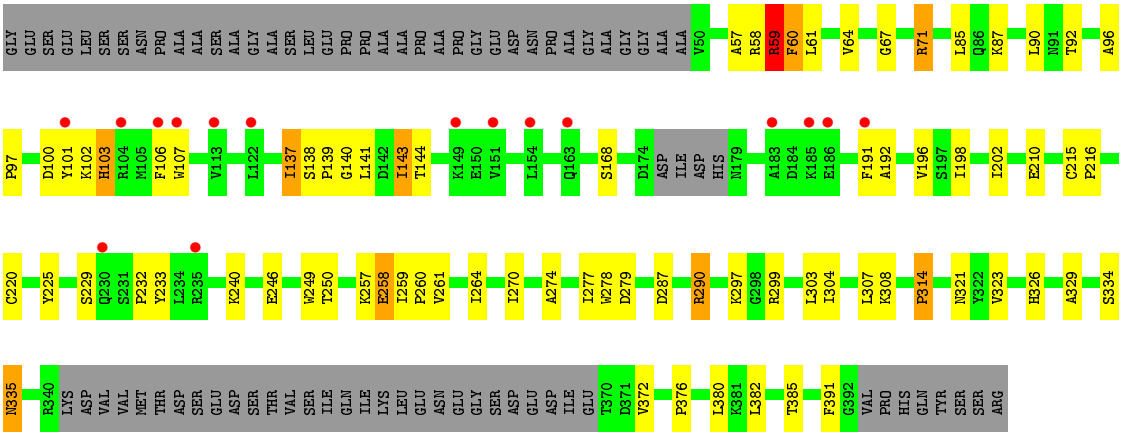
Chain C:





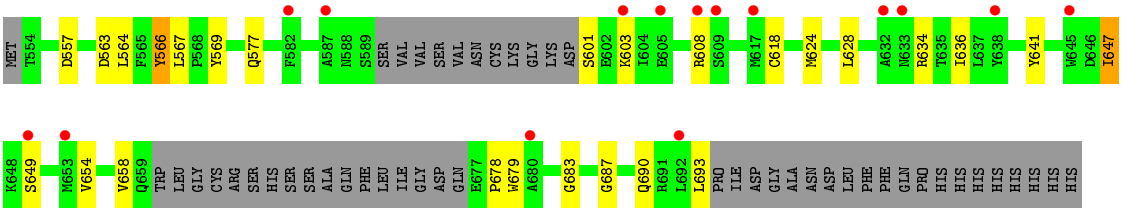
• Molecule 1: O-GlcNAcase TIM-barrel domain

Chain D:



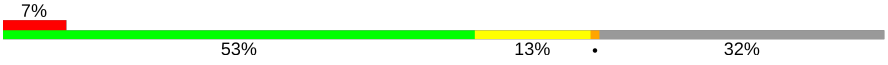
• Molecule 2: O-GlcNAcase stalk domain

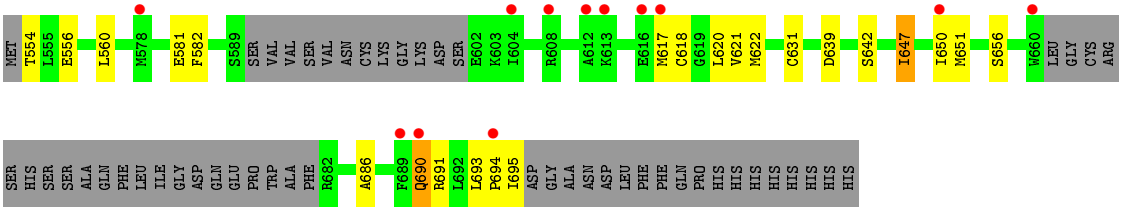
Chain E:



• Molecule 2: O-GlcNAcase stalk domain

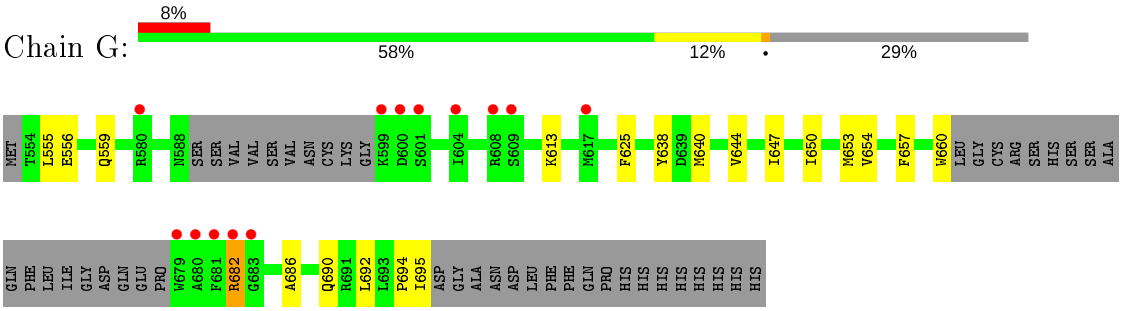
Chain F:





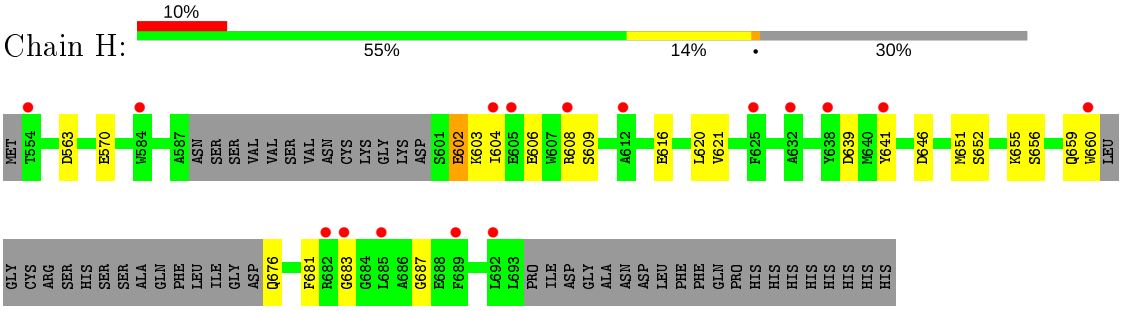
• Molecule 2: O-GlcNAcase stalk domain

Chain G:



- Molecule 2: O-GlcNAcase stalk domain

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	89.72Å 91.53Å 94.52Å 77.27° 62.81° 63.03°	Depositor
Resolution (Å)	40.78 – 2.79 40.78 – 2.79	Depositor EDS
% Data completeness (in resolution range)	100.0 (40.78-2.79) 100.0 (40.78-2.79)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.06 (at 2.81Å)	Xtriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.239 , 0.292 0.252 , 0.308	Depositor DCC
R_{free} test set	3077 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å^2)	55.0	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 70.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.000 for h,l,h-k 0.000 for h,h-l,k 0.000 for h,h-k,h-l 0.000 for -h,-h+k,-l 0.000 for -h,-k,-h+l 0.000 for -h,-l,-k 0.022 for -h,-h+l,-h+k	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	13813	wwPDB-VP
Average B, all atoms (Å^2)	58.0	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 28.59 % 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 1.7936e-03. 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: 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	0.50	0/2590	0.72	0/3510
1	B	0.52	0/2586	0.74	1/3506 (0.0%)
1	C	0.54	0/2580	0.76	0/3494
1	D	0.51	0/2577	0.76	1/3491 (0.0%)
2	E	0.53	0/932	0.72	0/1251
2	F	0.52	0/916	0.70	0/1230
2	G	0.53	0/972	0.76	0/1305
2	H	0.52	0/950	0.72	0/1276
All	All	0.52	0/14103	0.74	2/19063 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	59	ARG	C-N-CA	6.78	138.66	121.70
1	B	278	TRP	N-CA-C	-5.02	97.44	111.00

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	2522	0	2472	22	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2518	0	2461	20	0
1	C	2515	0	2474	26	0
1	D	2512	0	2464	40	0
2	E	912	0	887	10	0
2	F	896	0	880	11	0
2	G	949	0	932	11	0
2	H	928	0	905	9	0
3	A	6	0	8	0	0
3	B	6	0	8	0	0
3	C	6	0	8	1	0
3	D	6	0	8	0	0
4	A	5	0	0	0	0
4	B	11	0	0	0	0
4	C	8	0	0	1	0
4	D	3	0	0	0	0
4	E	3	0	0	0	0
4	F	3	0	0	0	0
4	H	4	0	0	0	0
All	All	13813	0	13507	137	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 5.

The worst 5 of 137 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:104:ARG:HG3	1:C:138:SER:HB2	1.64	0.79
1:D:58:ARG:O	1:D:59:ARG:HB2	1.82	0.76
1:C:97:PRO:HB2	1:C:100:ASP:HB2	1.74	0.69
1:D:71:ARG:HE	2:H:641:TYR:HD1	1.41	0.68
1:B:180:MET:HB3	1:B:185:LYS:HD3	1.76	0.67

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	301/388 (78%)	280 (93%)	20 (7%)	1 (0%)	41 72
1	B	301/388 (78%)	281 (93%)	16 (5%)	4 (1%)	12 36
1	C	304/388 (78%)	279 (92%)	19 (6%)	6 (2%)	7 24
1	D	304/388 (78%)	273 (90%)	24 (8%)	7 (2%)	6 21
2	E	106/161 (66%)	94 (89%)	9 (8%)	3 (3%)	5 17
2	F	103/161 (64%)	100 (97%)	3 (3%)	0	100 100
2	G	108/161 (67%)	97 (90%)	9 (8%)	2 (2%)	8 26
2	H	106/161 (66%)	98 (92%)	7 (7%)	1 (1%)	17 46
All	All	1633/2196 (74%)	1502 (92%)	107 (7%)	24 (2%)	10 33

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	182	ALA
1	D	59	ARG
1	D	101	TYR
2	E	678	PRO
2	G	694	PRO

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	275/331 (83%)	262 (95%)	13 (5%)	26 59
1	B	274/331 (83%)	267 (97%)	7 (3%)	46 79
1	C	270/331 (82%)	252 (93%)	18 (7%)	16 43
1	D	270/331 (82%)	250 (93%)	20 (7%)	13 37
2	E	95/140 (68%)	86 (90%)	9 (10%)	8 25
2	F	94/140 (67%)	85 (90%)	9 (10%)	8 24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	G	99/140 (71%)	93 (94%)	6 (6%)	18 48
2	H	96/140 (69%)	86 (90%)	10 (10%)	7 21
All	All	1473/1884 (78%)	1381 (94%)	92 (6%)	18 47

5 of 92 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	87	LYS
1	D	246	GLU
2	H	609	SER
1	D	102	LYS
1	D	138	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	103	HIS
1	D	288	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry

4 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	C	501	-	5,5,5	0.20	0	5,5,5	0.37	0
3	GOL	B	501	-	5,5,5	0.13	0	5,5,5	0.32	0
3	GOL	D	501	-	5,5,5	0.11	0	5,5,5	0.46	0
3	GOL	A	501	-	5,5,5	0.14	0	5,5,5	0.54	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	C	501	-	-	2/4/4/4	-
3	GOL	B	501	-	-	0/4/4/4	-
3	GOL	D	501	-	-	0/4/4/4	-
3	GOL	A	501	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

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

There are no ring outliers.

1 monomer is involved in 1 short contact:

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

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	307/388 (79%)	-0.01	7 (2%) 60 51	35, 49, 67, 80	0
1	B	307/388 (79%)	-0.01	3 (0%) 82 77	32, 45, 61, 74	0
1	C	310/388 (79%)	0.10	8 (2%) 56 46	26, 52, 77, 100	0
1	D	310/388 (79%)	0.37	16 (5%) 27 18	32, 63, 96, 115	0
2	E	112/161 (69%)	0.71	15 (13%) 3 1	49, 79, 104, 118	0
2	F	109/161 (67%)	0.48	12 (11%) 5 3	35, 57, 78, 91	0
2	G	114/161 (70%)	0.58	13 (11%) 5 3	34, 58, 86, 108	0
2	H	112/161 (69%)	0.88	16 (14%) 2 1	46, 78, 109, 121	0
All	All	1681/2196 (76%)	0.26	90 (5%) 25 17	26, 54, 93, 121	0

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	G	682	ARG	5.9
2	G	680	ALA	5.3
2	F	612	ALA	4.8
1	C	191	PHE	4.5
2	F	608	ARG	4.3

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 carbohydrates 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	D	501	6/6	0.83	0.19	46,51,51,52	0
3	GOL	A	501	6/6	0.88	0.16	52,54,55,56	0
3	GOL	B	501	6/6	0.89	0.18	58,60,62,63	0
3	GOL	C	501	6/6	0.91	0.14	53,55,56,57	0

6.5 Other polymers

There are no such residues in this entry.