



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 30, 2021 – 06:16 PM EDT

PDB ID : 7JVH
Title : Crystal structure of a GH43_12 retrieved from capybara gut metagenome
Authors : Cabral, L.; Domingues, M.N.; Martins, M.P.; Persinoti, G.F.; Morais, M.A.B.;
Murakami, M.T.
Deposited on : 2020-08-21
Resolution : 2.48 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 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 : 2.23.1
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) : 2.23.1

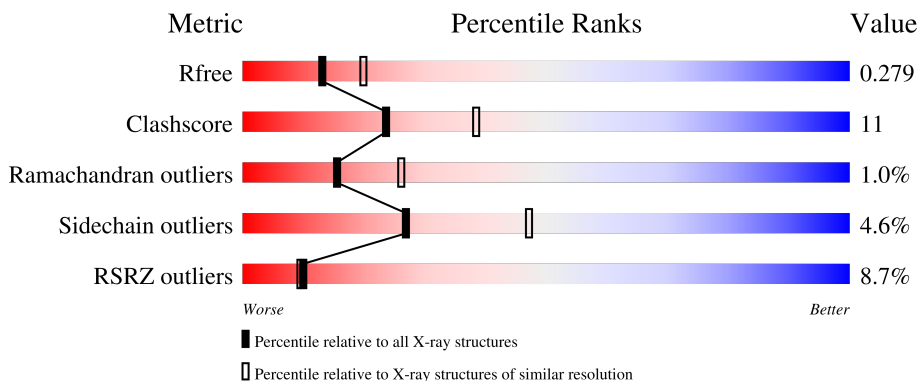
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.48 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.



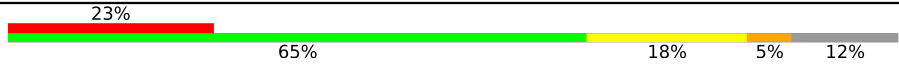
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5857 (2.50-2.46)
Clashscore	141614	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)
RSRZ outliers	127900	5738 (2.50-2.46)

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	567	 81% 12% • 6%
1	B	567	 80% 13% • 6%
1	C	567	 2% 75% 17% • 5%
1	D	567	 2% 77% 15% •• 6%
1	E	567	 20% 63% 20% •• 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain												
1	F	567	 <p>A horizontal bar chart showing the quality distribution of chain F. The bar is divided into five segments: red (23%), green (65%), yellow (18%), orange (5%), and grey (12%).</p> <table border="1"><thead><tr><th>Quality Category</th><th>Percentage</th></tr></thead><tbody><tr><td>Red</td><td>23%</td></tr><tr><td>Green</td><td>65%</td></tr><tr><td>Yellow</td><td>18%</td></tr><tr><td>Orange</td><td>5%</td></tr><tr><td>Grey</td><td>12%</td></tr></tbody></table>	Quality Category	Percentage	Red	23%	Green	65%	Yellow	18%	Orange	5%	Grey	12%
Quality Category	Percentage														
Red	23%														
Green	65%														
Yellow	18%														
Orange	5%														
Grey	12%														

2 Entry composition [i](#)

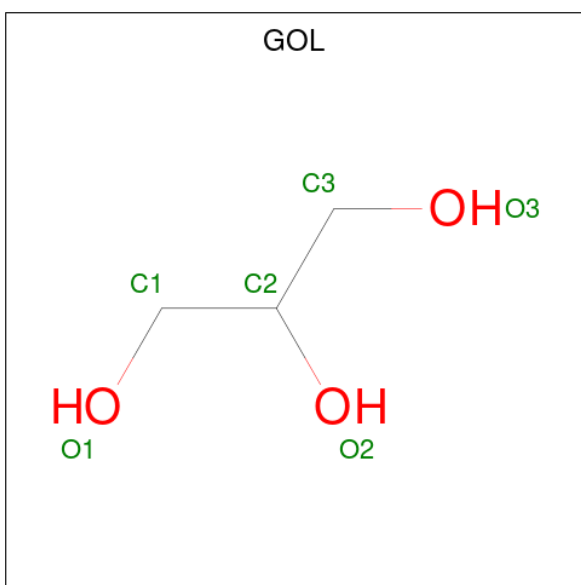
There are 3 unique types of molecules in this entry. The entry contains 25078 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Glycoside Hydrolase Family 43_12.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	535	Total 4232	C 2721	N 700	O 793	S 9	Se 9	0	0	0
1	B	535	Total 4231	C 2719	N 700	O 794	S 9	Se 9	0	0	0
1	C	536	Total 4239	C 2725	N 702	O 794	S 9	Se 9	0	0	0
1	D	534	Total 4221	C 2713	N 698	O 792	S 9	Se 9	0	0	0
1	E	497	Total 3943	C 2543	N 650	O 733	S 8	Se 9	0	0	0
1	F	499	Total 3964	C 2556	N 651	O 740	S 8	Se 9	0	0	0

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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


- Molecule 3 is water.

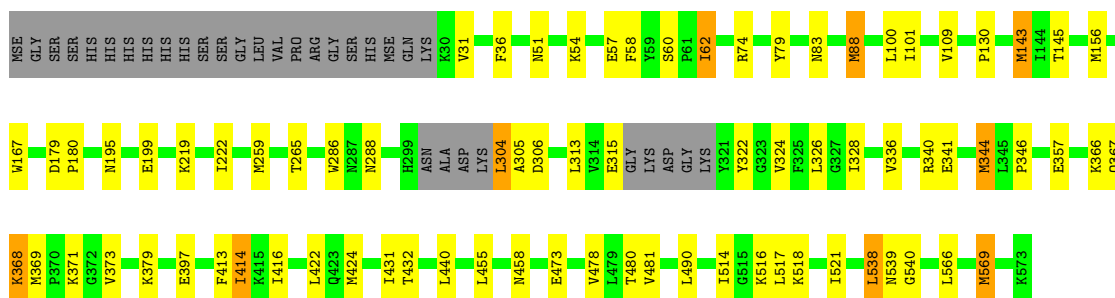
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	64	Total	O	0	0
			64	64		
3	B	52	Total	O	0	0
			52	52		
3	C	32	Total	O	0	0
			32	32		
3	D	31	Total	O	0	0
			31	31		
3	E	4	Total	O	0	0
			4	4		
3	F	5	Total	O	0	0
			5	5		

3 Residue-property plots


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.

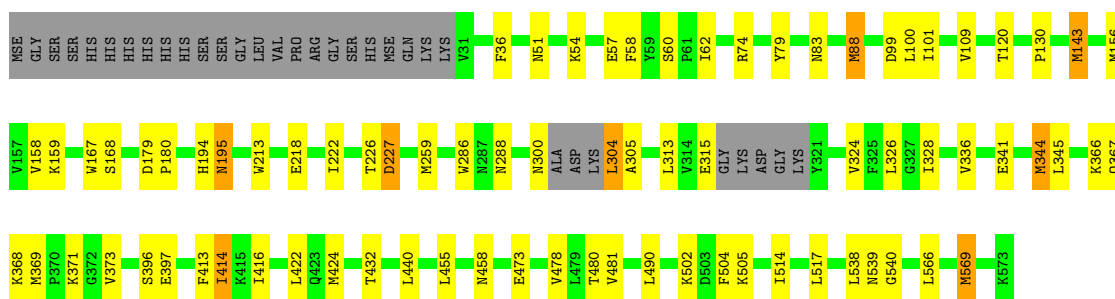
- Molecule 1: Glycoside Hydrolase Family 43_12

Chain A: 




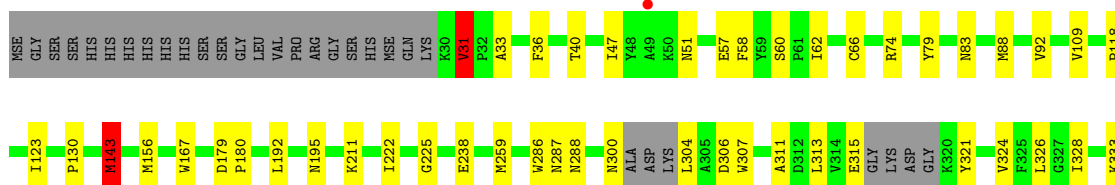
- Molecule 1: Glycoside Hydrolase Family 43_12

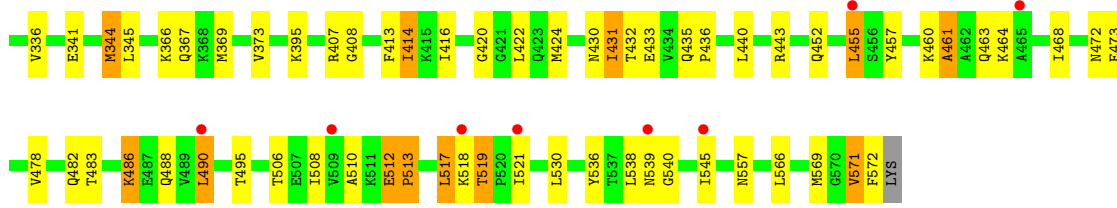
Chain B: 



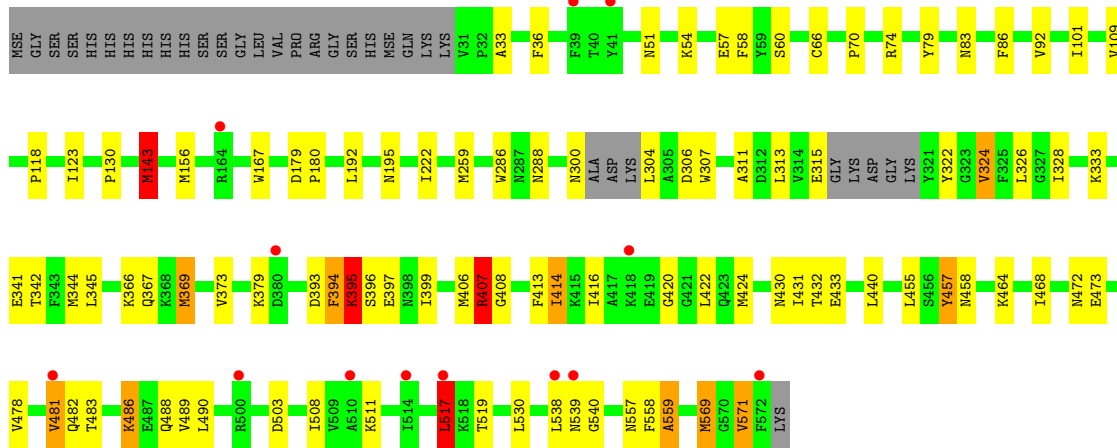
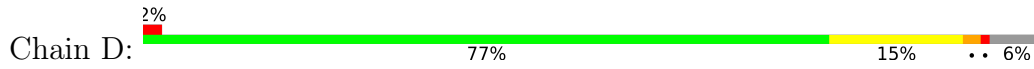
- Molecule 1: Glycoside Hydrolase Family 43_12

Chain C: 

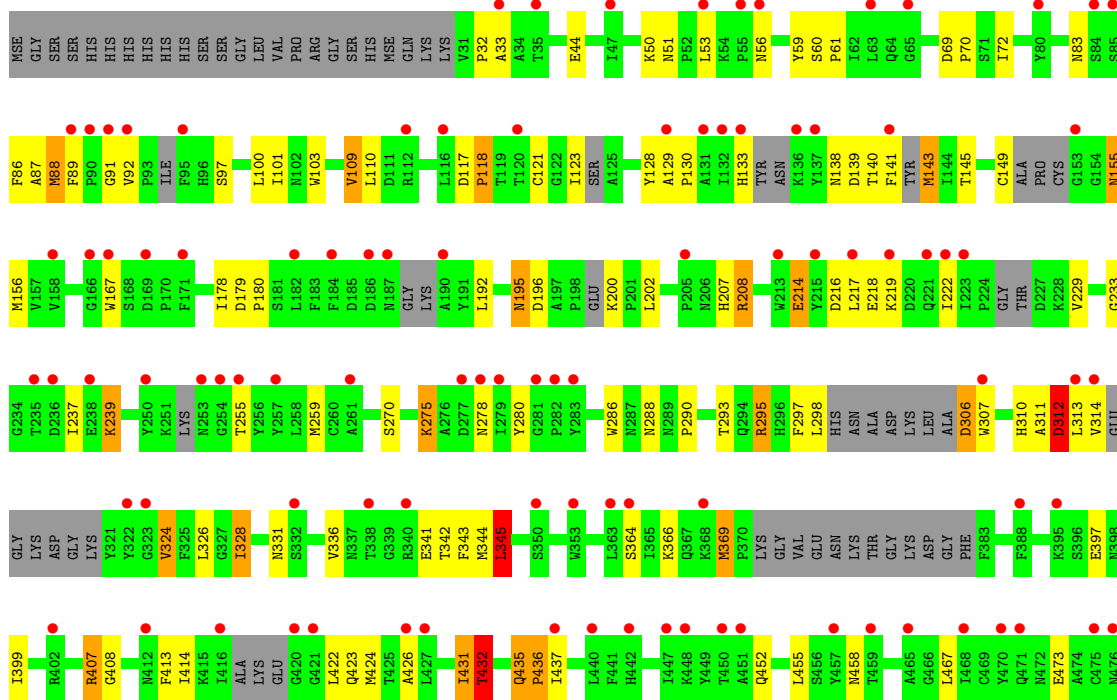


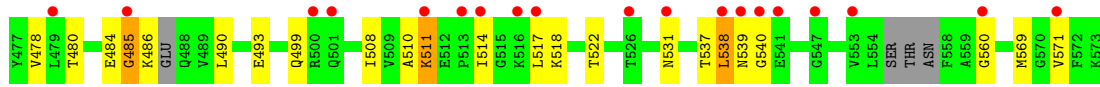


• Molecule 1: Glycoside Hydrolase Family 43_12

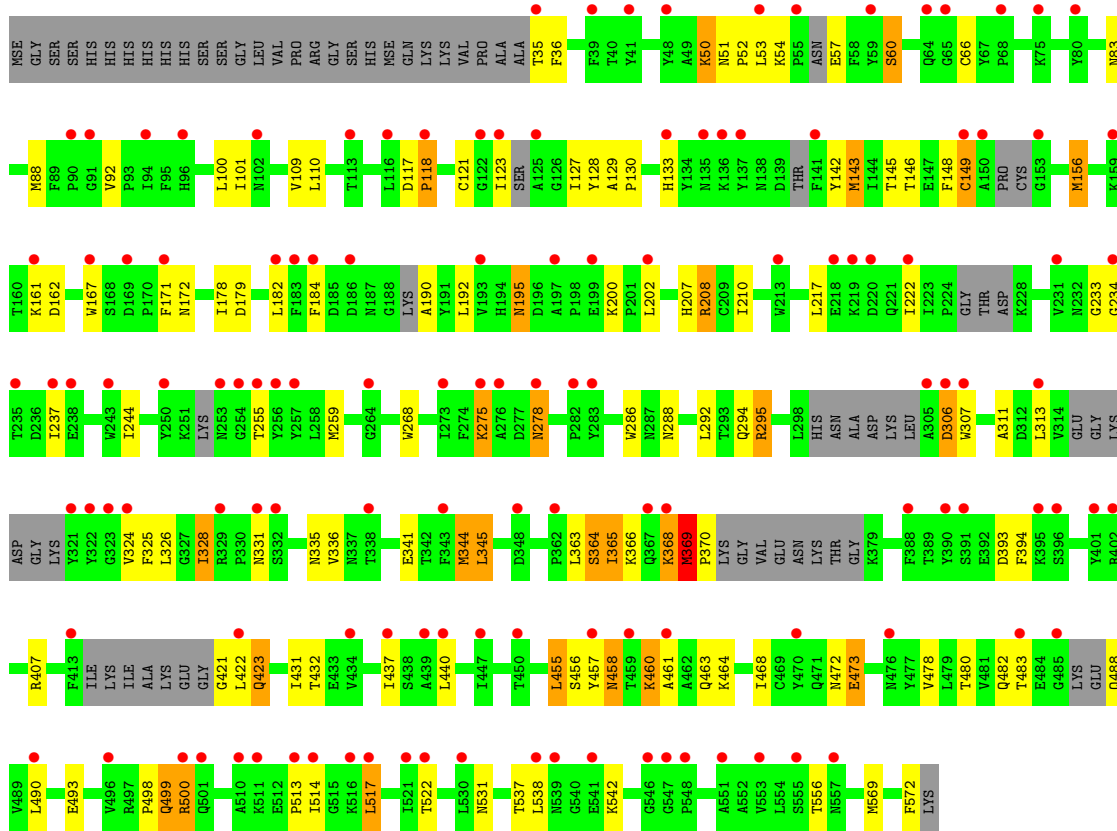


• Molecule 1: Glycoside Hydrolase Family 43_12





● Molecule 1: Glycoside Hydrolase Family 43_12



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	107.22Å 107.38Å 132.53Å 99.03° 113.38° 108.45°	Depositor
Resolution (Å)	49.20 – 2.48 49.20 – 2.48	Depositor EDS
% Data completeness (in resolution range)	96.0 (49.20-2.48) 95.8 (49.20-2.48)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.20 (at 2.48Å)	Xtrriage
Refinement program	PHENIX 1.8	Depositor
R, R_{free}	0.246 , 0.278 0.249 , 0.279	Depositor DCC
R_{free} test set	8303 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	53.8	Xtrriage
Anisotropy	0.711	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 41.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.437 for k,h,-h-k-l	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	25078	wwPDB-VP
Average B, all atoms (Å ²)	85.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.78% 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	0.49	1/4343 (0.0%)	0.70	2/5887 (0.0%)
1	B	0.50	2/4342 (0.0%)	0.71	2/5887 (0.0%)
1	C	0.57	6/4350 (0.1%)	0.77	5/5898 (0.1%)
1	D	0.52	4/4332 (0.1%)	0.76	7/5876 (0.1%)
1	E	0.59	8/4037 (0.2%)	0.80	9/5459 (0.2%)
1	F	0.52	4/4062 (0.1%)	0.79	10/5498 (0.2%)
All	All	0.53	25/25466 (0.1%)	0.76	35/34505 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	118	PRO	N-CA	13.51	1.70	1.47
1	E	436	PRO	N-CA	13.20	1.69	1.47
1	C	513	PRO	N-CA	13.03	1.69	1.47
1	E	118	PRO	N-CA	11.76	1.67	1.47
1	C	31	VAL	C-N	9.12	1.51	1.34

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	161	LYS	N-CA-C	-9.82	84.49	111.00
1	E	431	ILE	N-CA-C	-8.60	87.78	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	143	MSE	CG-SE-CE	8.18	116.89	98.90
1	D	143	MSE	CG-SE-CE	8.02	116.53	98.90
1	A	344	MSE	CG-SE-CE	7.90	116.27	98.90

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	226	THR	Mainchain
1	B	227	ASP	Mainchain

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	4232	0	4103	52	0
1	B	4231	0	4096	54	0
1	C	4239	0	4109	96	0
1	D	4221	0	4083	86	0
1	E	3943	0	3807	133	0
1	F	3964	0	3804	137	0
2	A	18	0	24	1	0
2	B	12	0	16	0	0
2	C	12	0	16	3	0
2	D	12	0	16	1	0
2	F	6	0	8	0	0
3	A	64	0	0	0	0
3	B	52	0	0	0	0
3	C	32	0	0	0	0
3	D	31	0	0	0	0
3	E	4	0	0	0	0
3	F	5	0	0	0	0
All	All	25078	0	24082	550	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 11.

The worst 5 of 550 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:E:436:PRO:N	1:E:436:PRO:CA	1.69	1.42
1:C:513:PRO:N	1:C:513:PRO:CA	1.69	1.38
1:F:118:PRO:N	1:F:118:PRO:CA	1.70	1.34
1:F:36:PHE:CE1	1:F:455:LEU:HD13	1.69	1.25
1:E:118:PRO:HB2	1:E:407:ARG:HA	1.15	1.11

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	529/567 (93%)	498 (94%)	28 (5%)	3 (1%)	25	40
1	B	529/567 (93%)	497 (94%)	30 (6%)	2 (0%)	34	52
1	C	530/567 (94%)	500 (94%)	26 (5%)	4 (1%)	19	33
1	D	528/567 (93%)	489 (93%)	31 (6%)	8 (2%)	10	16
1	E	465/567 (82%)	425 (91%)	34 (7%)	6 (1%)	12	19
1	F	473/567 (83%)	438 (93%)	28 (6%)	7 (2%)	10	16
All	All	3054/3402 (90%)	2847 (93%)	177 (6%)	30 (1%)	15	26

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	31	VAL
1	A	539	ASN
1	B	539	ASN
1	C	539	ASN
1	D	395	LYS

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	455/470 (97%)	440 (97%)	15 (3%)	38	61
1	B	455/470 (97%)	442 (97%)	13 (3%)	42	66
1	C	456/470 (97%)	441 (97%)	15 (3%)	38	61
1	D	454/470 (97%)	436 (96%)	18 (4%)	31	53
1	E	425/470 (90%)	388 (91%)	37 (9%)	10	18
1	F	426/470 (91%)	400 (94%)	26 (6%)	18	34
All	All	2671/2820 (95%)	2547 (95%)	124 (5%)	27	47

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	517	LEU
1	F	369	MSE
1	E	214	GLU
1	F	368	LYS
1	F	499	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	488	GLN
1	C	488	GLN
1	D	488	GLN
1	E	488	GLN
1	F	488	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

10 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	$\# Z > 2$
2	GOL	A	601	-	5,5,5	0.10	0	5,5,5	0.28	0
2	GOL	D	602	-	5,5,5	0.11	0	5,5,5	0.37	0
2	GOL	F	601	-	5,5,5	0.10	0	5,5,5	0.27	0
2	GOL	B	602	-	5,5,5	0.12	0	5,5,5	0.36	0
2	GOL	A	603	-	5,5,5	0.09	0	5,5,5	0.26	0
2	GOL	D	601	-	5,5,5	0.10	0	5,5,5	0.33	0
2	GOL	B	601	-	5,5,5	0.13	0	5,5,5	0.40	0
2	GOL	C	601	-	5,5,5	0.11	0	5,5,5	0.34	0
2	GOL	A	602	-	5,5,5	0.09	0	5,5,5	0.30	0
2	GOL	C	602	-	5,5,5	0.16	0	5,5,5	0.38	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	GOL	A	601	-	-	1/4/4/4	-
2	GOL	D	602	-	-	3/4/4/4	-
2	GOL	F	601	-	-	0/4/4/4	-
2	GOL	B	602	-	-	2/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	603	-	-	3/4/4/4	-
2	GOL	D	601	-	-	2/4/4/4	-
2	GOL	B	601	-	-	0/4/4/4	-
2	GOL	C	601	-	-	2/4/4/4	-
2	GOL	A	602	-	-	2/4/4/4	-
2	GOL	C	602	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 19 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	601	GOL	C1-C2-C3-O3
2	D	601	GOL	O1-C1-C2-C3
2	D	601	GOL	O1-C1-C2-O2
2	A	602	GOL	C1-C2-C3-O3
2	A	603	GOL	O1-C1-C2-C3

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	601	GOL	1	0
2	A	602	GOL	1	0
2	C	602	GOL	3	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	526/567 (92%)	-0.02	0 100 100	41, 58, 83, 112	0
1	B	526/567 (92%)	-0.01	0 100 100	41, 58, 85, 115	0
1	C	527/567 (92%)	0.14	9 (1%) 70 71	45, 74, 120, 155	0
1	D	525/567 (92%)	0.08	13 (2%) 57 59	44, 75, 121, 167	0
1	E	488/567 (86%)	1.30	116 (23%) 0 0	101, 118, 147, 190	0
1	F	490/567 (86%)	1.39	130 (26%) 0 0	102, 118, 151, 178	0
All	All	3082/3402 (90%)	0.46	268 (8%) 10 9	41, 77, 136, 190	0

The worst 5 of 268 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	500	ARG	9.4
1	E	133	HIS	7.6
1	E	255	THR	7.6
1	F	539	ASN	7.5
1	E	539	ASN	6.8

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(Å ²)	Q<0.9
2	GOL	F	601	6/6	0.82	0.12	74,86,91,92	0
2	GOL	B	602	6/6	0.90	0.14	71,75,80,82	0
2	GOL	B	601	6/6	0.90	0.24	51,61,64,69	0
2	GOL	A	602	6/6	0.91	0.17	71,79,86,91	0
2	GOL	C	602	6/6	0.92	0.16	50,52,63,68	0
2	GOL	A	603	6/6	0.93	0.18	59,70,74,74	0
2	GOL	D	601	6/6	0.94	0.20	51,57,61,68	0
2	GOL	C	601	6/6	0.94	0.21	59,62,66,75	0
2	GOL	D	602	6/6	0.95	0.18	40,52,58,60	0
2	GOL	A	601	6/6	0.97	0.13	45,51,55,58	0

6.5 Other polymers [i](#)

There are no such residues in this entry.