wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2026 – 03:55 PM EDT

PDB ID	:	39BB / pdb_000039bb
Title	:	Crystal structure of GuMH, a flavin-dependent halogenase from the human gut microbiome
Authors	:	Okwuosa, C.C.; Bellizzi, J.J.
Deposited on	:	2026-09-28
Resolution	:	2.38 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 (i)) were used in the production of this report:

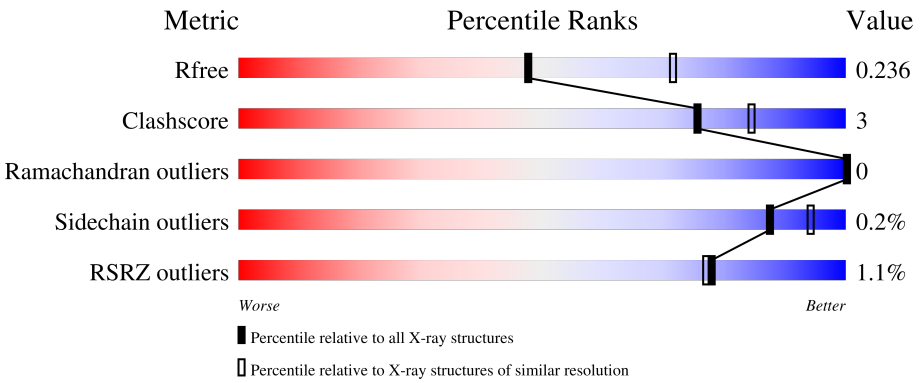
MolProbity	:	4-5-2 with Phenix2.2
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.2
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.51

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 2.38 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	7164 (2.40-2.36)
Clashscore	190562	7722 (2.40-2.36)
Ramachandran outliers	187476	7626 (2.40-2.36)
Sidechain outliers	187428	7627 (2.40-2.36)
RSRZ outliers	180081	7170 (2.40-2.36)

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	527	87% 6% 7%
1	B	527	% 86% 7% 7%
1	C	527	3% 82% 10% 7%
1	D	527	86% 7% 7%

2 Entry composition (i)

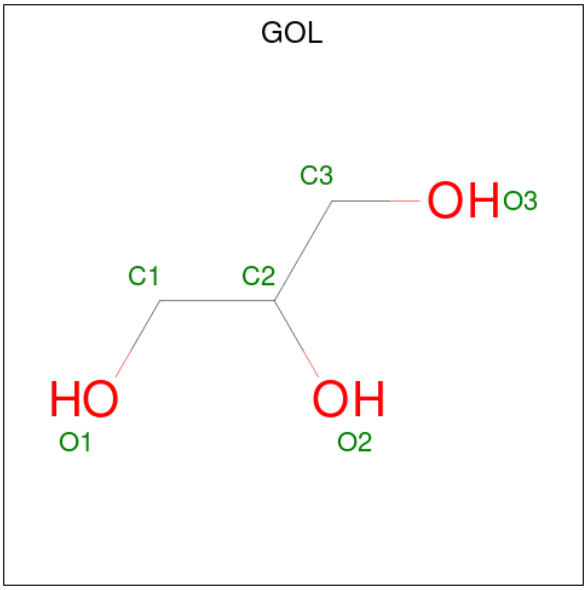
There are 3 unique types of molecules in this entry. The entry contains 15829 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 GuMH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	489	3771	2392	691	677	11	0	1	0
1	B	491	3791	2404	697	679	11	0	2	0
1	C	489	3763	2387	690	676	10	0	0	0
1	D	489	3762	2387	691	673	11	0	1	0

- Molecule 2 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



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

Continued on next page...

Continued from previous page...

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

- Molecule 3 is water.

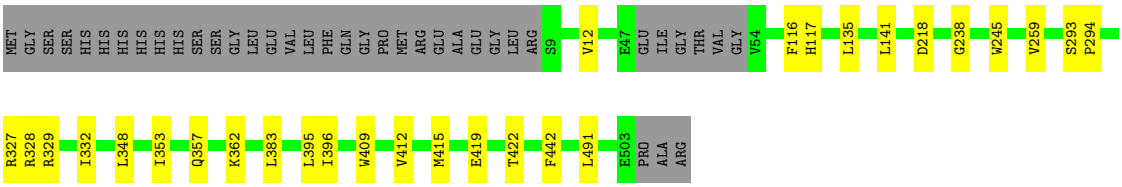
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	202	202	202	0	0
3	B	167	167	167	0	0
3	C	119	119	119	0	0
3	D	164	164	164	0	0

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.

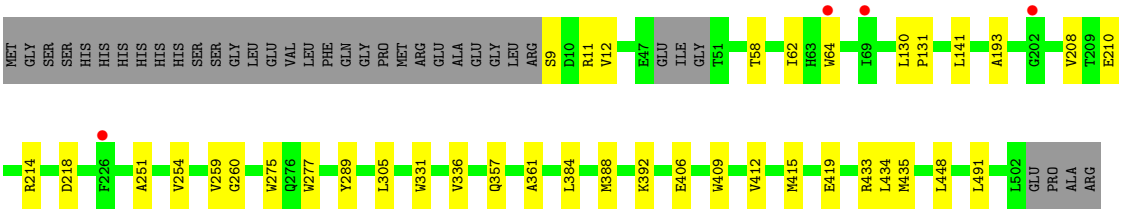
- Molecule 1: GuMH

Chain A:



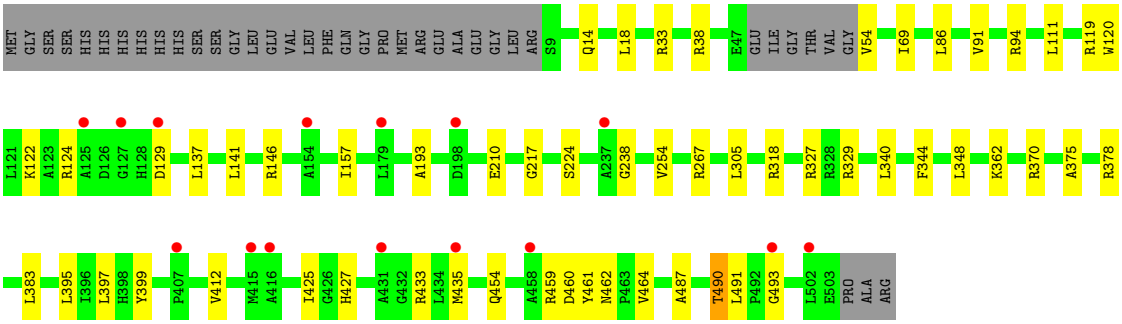
- Molecule 1: GuMH

Chain B:



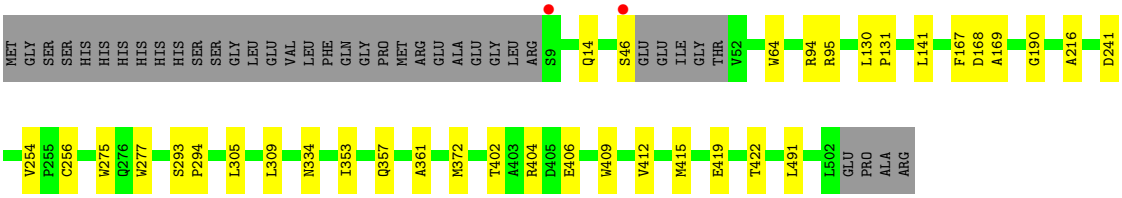
- Molecule 1: GuMH

Chain C:



- Molecule 1: GuMH

Chain D:


MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU GLU VAL LEU PHE GLN GLY PRO MET ARG GLU ALA GLU GLY LEU ARG S9 Q14 S46 GLU GLU ILE GLY THR V52 W64 R94 R95 L130 P131 L141 F167 D168 A169 G190 A216 D241
V254 P255 C256 W275 Q276 W277 S293 P294 L305 L309 N334 I353 Q357 A361 M372 T402 A403 R404 D405 E406 W409 V412 M415 E419 T422 L491 L502 GLU PRO ALA ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	191.45Å 107.17Å 107.38Å 90.00° 104.12° 90.00°	Depositor
Resolution (Å)	47.98 – 2.38 47.98 – 2.38	Depositor EDS
% Data completeness (in resolution range)	97.8 (47.98-2.38) 97.8 (47.98-2.38)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 2.37Å)	Xtriage
Refinement program	PHENIX 2.1_6048	Depositor
R, R_{free}	0.189 , 0.235 0.190 , 0.236	Depositor DCC
R_{free} test set	2000 reflections (2.37%)	wwPDB-VP
Wilson B-factor (Å^2)	45.2	Xtriage
Anisotropy	0.176	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 42.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	15829	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.98% 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 (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	0.11	0/3861	0.28	0/5251
1	B	0.11	0/3881	0.28	0/5278
1	C	0.12	0/3853	0.28	0/5241
1	D	0.11	0/3852	0.29	0/5240
All	All	0.11	0/15447	0.28	0/21010

There are no bond length outliers.

There are no bond angle outliers.

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	3771	0	3728	16	0
1	B	3791	0	3753	23	0
1	C	3763	0	3720	34	0
1	D	3762	0	3724	23	0
2	A	18	0	24	2	0
2	B	18	0	24	0	0
2	C	18	0	24	0	0
2	D	36	0	48	0	0
3	A	202	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	167	0	0	0	0
3	C	119	0	0	0	0
3	D	164	0	0	0	0
All	All	15829	0	15045	92	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 3.

The worst 5 of 92 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:A:353:ILE:HG22	1:A:357:GLN:HE21	1.49	0.77
1:C:54:VAL:HG21	1:C:318:ARG:HH22	1.55	0.71
1:B:433:ARG:HB3	1:B:435:MET:HE3	1.72	0.71
1:D:130:LEU:HD12	1:D:131:PRO:HD2	1.71	0.71
1:D:167:PHE:HE1	1:D:169:ALA:HA	1.61	0.66

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	486/527 (92%)	479 (99%)	7 (1%)	0	100	100
1	B	489/527 (93%)	478 (98%)	11 (2%)	0	100	100
1	C	485/527 (92%)	469 (97%)	16 (3%)	0	100	100
1	D	486/527 (92%)	480 (99%)	6 (1%)	0	100	100
All	All	1946/2108 (92%)	1906 (98%)	40 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	372/401 (93%)	371 (100%)	1 (0%)	86	93
1	B	374/401 (93%)	374 (100%)	0	100	100
1	C	371/401 (92%)	369 (100%)	2 (0%)	81	90
1	D	371/401 (92%)	371 (100%)	0	100	100
All	All	1488/1604 (93%)	1485 (100%)	3 (0%)	87	94

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	259	VAL
1	C	86	LEU
1	C	490	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	427	HIS
1	D	357	GLN
1	B	454	GLN
1	D	423	GLN
1	C	478	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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

15 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	D	601	-	5,5,5	0.38	0	5,5,5	0.40	0
2	GOL	A	603	-	5,5,5	0.32	0	5,5,5	0.46	0
2	GOL	D	604	-	5,5,5	0.35	0	5,5,5	0.39	0
2	GOL	D	605	-	5,5,5	0.35	0	5,5,5	0.46	0
2	GOL	D	603	-	5,5,5	0.33	0	5,5,5	0.41	0
2	GOL	D	606	-	5,5,5	0.31	0	5,5,5	0.45	0
2	GOL	C	603	-	5,5,5	0.33	0	5,5,5	0.38	0
2	GOL	B	603	-	5,5,5	0.33	0	5,5,5	0.43	0
2	GOL	C	601	-	5,5,5	0.33	0	5,5,5	0.49	0
2	GOL	A	601	-	5,5,5	0.33	0	5,5,5	0.32	0
2	GOL	B	602	-	5,5,5	0.37	0	5,5,5	0.27	0
2	GOL	C	602	-	5,5,5	0.33	0	5,5,5	0.42	0
2	GOL	D	602	-	5,5,5	0.34	0	5,5,5	0.36	0
2	GOL	B	601	-	5,5,5	0.34	0	5,5,5	0.40	0
2	GOL	A	602	-	5,5,5	0.38	0	5,5,5	0.56	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	D	601	-	-	2/4/4/4	-
2	GOL	A	603	-	-	0/4/4/4	-
2	GOL	D	604	-	-	2/4/4/4	-
2	GOL	D	605	-	-	3/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	D	603	-	-	2/4/4/4	-
2	GOL	D	606	-	-	0/4/4/4	-
2	GOL	C	603	-	-	0/4/4/4	-
2	GOL	B	603	-	-	3/4/4/4	-
2	GOL	C	601	-	-	2/4/4/4	-
2	GOL	A	601	-	-	4/4/4/4	-
2	GOL	B	602	-	-	2/4/4/4	-
2	GOL	C	602	-	-	2/4/4/4	-
2	GOL	D	602	-	-	2/4/4/4	-
2	GOL	B	601	-	-	0/4/4/4	-
2	GOL	A	602	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 25 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	GOL	O1-C1-C2-C3
2	D	602	GOL	C1-C2-C3-O3
2	D	604	GOL	O1-C1-C2-O2
2	D	604	GOL	O1-C1-C2-C3
2	A	601	GOL	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	GOL	1	0
2	A	602	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 (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	489/527 (92%)	-0.02	0	100	100	25, 51, 74, 97	1 (0%)
1	B	491/527 (93%)	0.04	4 (0%)	82	82	23, 52, 71, 103	2 (0%)
1	C	489/527 (92%)	0.24	15 (3%)	51	51	40, 58, 93, 122	0
1	D	489/527 (92%)	0.02	2 (0%)	88	88	29, 52, 70, 109	1 (0%)
All	All	1958/2108 (92%)	0.07	21 (1%)	78	77	23, 53, 81, 122	4 (0%)

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	431	ALA	3.6
1	B	226	PHE	3.3
1	B	202	GLY	2.9
1	D	9	SER	2.8
1	C	198	ASP	2.7

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 oligosaccharides 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	C	603	6/6	0.78	0.16	61,70,74,76	0
2	GOL	A	603	6/6	0.82	0.13	61,69,73,74	0
2	GOL	A	602	6/6	0.82	0.13	48,49,58,58	0
2	GOL	D	606	6/6	0.82	0.15	61,67,70,73	0
2	GOL	C	601	6/6	0.83	0.16	70,73,79,79	0
2	GOL	D	605	6/6	0.84	0.12	51,52,68,77	0
2	GOL	D	603	6/6	0.85	0.14	74,76,78,78	0
2	GOL	D	602	6/6	0.86	0.15	62,68,72,79	0
2	GOL	D	604	6/6	0.86	0.16	51,53,56,62	0
2	GOL	D	601	6/6	0.87	0.11	60,63,66,71	0
2	GOL	B	601	6/6	0.87	0.16	47,49,55,64	0
2	GOL	C	602	6/6	0.89	0.13	50,64,64,65	0
2	GOL	B	603	6/6	0.90	0.14	64,71,75,77	0
2	GOL	A	601	6/6	0.91	0.11	52,53,56,56	0
2	GOL	B	602	6/6	0.92	0.11	48,50,52,57	0

6.5 Other polymers (i)

There are no such residues in this entry.