



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 03:49 pm GMT

PDB ID : 5T9J
Title : Crystal Structure of human GEN1 in complex with Holliday junction DNA in the upper interface
Authors : Lee, S.-H.; Biertumpfel, C.
Deposited on : 2016-09-09
Resolution : 3.00 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

<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 : trunk28620
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) : recalc28949

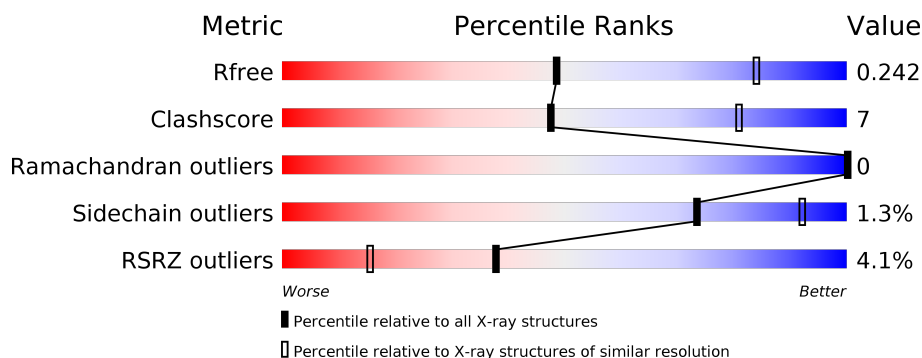
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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	506	<div> <div>3%</div> <div> <div></div> <div>62%</div> <div>15%</div> <div>23%</div> </div> </div>
1	B	506	<div> <div>4%</div> <div> <div></div> <div>68%</div> <div>9%</div> <div>23%</div> </div> </div>
2	C	20	<div> <div></div> <div> <div></div> <div>65%</div> <div>35%</div> </div> </div>
3	D	20	<div> <div></div> <div> <div></div> <div>50%</div> <div>50%</div> </div> </div>
4	E	20	<div> <div></div> <div> <div></div> <div>70%</div> <div>25%</div> <div>5%</div> </div> </div>
5	F	20	<div> <div></div> <div> <div></div> <div>65%</div> <div>30%</div> <div>5%</div> </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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	GOL	A	601	-	-	-	X
6	GOL	A	602	-	-	-	X
6	GOL	A	604	-	-	-	X
6	GOL	A	605	-	-	-	X
6	GOL	A	607	-	-	-	X
6	GOL	B	605	-	-	-	X
6	GOL	C	102	-	-	-	X
6	GOL	C	103	-	-	-	X

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 Flap endonuclease GEN homolog 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	390	Total 3144	C 2013	N 547	O 560	S 24	0	0	0
1	B	391	Total 3154	C 2020	N 549	O 561	S 24	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	ASN	ASP	engineered mutation	UNP Q17RS7
A	92	THR	SER	variant	UNP Q17RS7
A	310	ASN	SER	variant	UNP Q17RS7
A	506	LEU	-	expression tag	UNP Q17RS7
B	30	ASN	ASP	engineered mutation	UNP Q17RS7
B	92	THR	SER	variant	UNP Q17RS7
B	310	ASN	SER	variant	UNP Q17RS7
B	506	LEU	-	expression tag	UNP Q17RS7

- Molecule 2 is a DNA chain called DNA (5'-D(*DGP*DAP*DAP*DTP*DTP*DCP*DCP*DCP*DGP*DAP*DTP*DTP*DAP*DGP*DGP*DGP*DAP*DTP*DGP*DC)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	20	Total	C	N	O	P	0	0	0
			407	194	76	118	19			

- Molecule 3 is a DNA chain called DNA (5'-D(*DGP*DCP*DAP*DTP*DCP*DCP*DCP*DT*
TP*DAP*DAP*DGP*DCP*DTP*DCP*DCP*DAP*DTP*DCP*DGP*DT)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	20	Total	C	N	O	P	0	0	0
			409	194	79	117	19			

- Molecule 4 is a DNA chain called DNA (5'-D(*DAP*DCP*DGP*DAP*DTP*DGP*DGP*

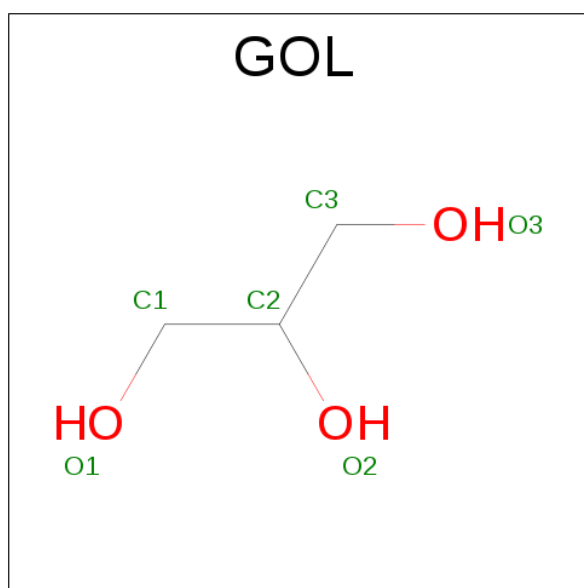
DAP*DGP*DCP*DCP*DGP*DCP*DTP*DAP*DGP*DGP*DCP*DTP*DC)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	19	Total	C	N	O	P	0	0	0
			379	182	67	112	18			

- Molecule 5 is a DNA chain called DNA (5'-D(*DGP*DAP*DGP*DCP*DCP*DTP*DAP*DGP*DCP*DGP*DTP*DCP*DCP*DGP*DGP*DAP*DAP*DTP*DTP*DC)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	19	Total	C	N	O	P	0	0	0
			394	187	74	114	19			

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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

Continued on next page...

Continued from previous page...


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

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	B	1	Total Mg 1 1	0	0
7	A	1	Total Mg 1 1	0	0

- Molecule 8 is water.

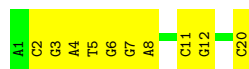
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	10	Total O 10 10	0	0
8	B	11	Total O 11 11	0	0
8	D	1	Total O 1 1	0	0
8	F	1	Total O 1 1	0	0

Chain C:  65% 35%



- Molecule 3: DNA (5'-D(*DGP*DCP*DAP*DTP*DCP*DCP*DCP*DTP*DAP*DAP*DGP*DCP*DTP*DCP*DCP*DAP*DTP*DCP*DGP*DT)-3')

Chain D:  50% 50%



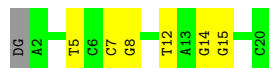
- Molecule 4: DNA (5'-D(*DAP*DCP*DGP*DAP*DTP*DGP*DGP*DAP*DGP*DCP*DCP*DCP*DCP*DTP*DAP*DGP*DGP*DCP*DTP*DC)-3')

Chain E:  70% 25% 5%



- Molecule 5: DNA (5'-D(*DGP*DAP*DGP*DCP*DCP*DTP*DAP*DGP*DCP*DGP*DTP*DCP*DCP*DGP*DGP*DAP*DAP*DTP*DTP*DC)-3')

Chain F:  65% 30% 5%



4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α , β , γ	86.94Å 86.94Å 200.72Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	75.29 – 3.00 75.29 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.9 (75.29-3.00) 100.0 (75.29-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.59 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.10.1_2155)	Depositor
R, R_{free}	0.199 , 0.241 0.197 , 0.242	Depositor DCC
R_{free} test set	1720 reflections (5.07%)	DCC
Wilson B-factor (Å ²)	93.2	Xtriage
Anisotropy	0.009	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 80.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.038 for -h,-k,l 0.478 for h,-h-k,-l 0.037 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8002	wwPDB-VP
Average B, all atoms (Å ²)	128.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.38% 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, MG

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	0.26	0/3210	0.44	0/4323
1	B	0.26	0/3220	0.42	0/4336
2	C	0.53	0/456	0.90	0/702
3	D	0.56	0/459	0.95	0/707
4	E	0.52	0/423	0.90	0/649
5	F	0.58	0/442	0.92	0/681
All	All	0.34	0/8210	0.59	0/11398

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	3144	0	3188	49	0
1	B	3154	0	3199	34	0
2	C	407	0	226	9	0
3	D	409	0	225	12	0
4	E	379	0	215	5	0
5	F	394	0	215	8	0
6	A	42	0	56	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	30	0	39	1	0
6	C	18	0	24	0	0
7	A	1	0	0	0	0
7	B	1	0	0	0	0
8	A	10	0	0	3	0
8	B	11	0	0	1	0
8	D	1	0	0	0	0
8	F	1	0	0	0	0
All	All	8002	0	7387	106	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 7.

The worst 5 of 106 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)
2:C:16:DA:H2	5:F:5:DT:H3	1.12	0.93
1:A:200:GLY:HA2	1:A:203:ILE:HD12	1.66	0.77
2:C:16:DA:C2	5:F:5:DT:N3	2.51	0.75
2:C:16:DA:N1	5:F:5:DT:O4	2.26	0.69
1:A:15:GLN:HG3	1:A:181:CYS:HB3	1.76	0.68

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	384/506 (76%)	375 (98%)	9 (2%)	0	100	100
1	B	385/506 (76%)	378 (98%)	7 (2%)	0	100	100
All	All	769/1012 (76%)	753 (98%)	16 (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	344/456 (75%)	337 (98%)	7 (2%)	60	87
1	B	345/456 (76%)	343 (99%)	2 (1%)	89	96
All	All	689/912 (76%)	680 (99%)	9 (1%)	73	92

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	93	ARG
1	B	328	PHE
1	A	329	HIS
1	A	48	VAL
1	A	94	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	314	ASN

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

Of 17 ligands modelled in this entry, 2 are monoatomic - leaving 15 for Mogul analysis.

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	$\# Z > 2$
6	GOL	A	601	-	5,5,5	0.36	0	5,5,5	0.22	0
6	GOL	A	602	-	5,5,5	0.36	0	5,5,5	0.18	0
6	GOL	A	603	-	5,5,5	0.36	0	5,5,5	0.20	0
6	GOL	A	604	-	5,5,5	0.34	0	5,5,5	0.20	0
6	GOL	A	605	-	5,5,5	0.35	0	5,5,5	0.22	0
6	GOL	A	606	-	5,5,5	0.34	0	5,5,5	0.27	0
6	GOL	A	607	-	5,5,5	0.36	0	5,5,5	0.17	0
6	GOL	B	601	-	5,5,5	0.35	0	5,5,5	0.24	0
6	GOL	B	602	-	5,5,5	0.35	0	5,5,5	0.22	0
6	GOL	B	603	-	5,5,5	0.31	0	5,5,5	0.25	0
6	GOL	B	604	-	5,5,5	0.34	0	5,5,5	0.26	0
6	GOL	B	605	-	5,5,5	0.34	0	5,5,5	0.25	0
6	GOL	C	101	-	5,5,5	0.35	0	5,5,5	0.19	0
6	GOL	C	102	-	5,5,5	0.35	0	5,5,5	0.16	0
6	GOL	C	103	-	5,5,5	0.36	0	5,5,5	0.21	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
6	GOL	A	601	-	-	0/4/4/4	0/0/0/0
6	GOL	A	602	-	-	0/4/4/4	0/0/0/0
6	GOL	A	603	-	-	0/4/4/4	0/0/0/0
6	GOL	A	604	-	-	0/4/4/4	0/0/0/0
6	GOL	A	605	-	-	0/4/4/4	0/0/0/0
6	GOL	A	606	-	-	0/4/4/4	0/0/0/0
6	GOL	A	607	-	-	0/4/4/4	0/0/0/0
6	GOL	B	601	-	-	0/4/4/4	0/0/0/0
6	GOL	B	602	-	-	0/4/4/4	0/0/0/0
6	GOL	B	603	-	-	0/4/4/4	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	GOL	B	604	-	-	0/4/4/4	0/0/0/0
6	GOL	B	605	-	-	0/4/4/4	0/0/0/0
6	GOL	C	101	-	-	0/4/4/4	0/0/0/0
6	GOL	C	102	-	-	0/4/4/4	0/0/0/0
6	GOL	C	103	-	-	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.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	605	GOL	1	0
6	A	606	GOL	1	0
6	B	605	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	390/506 (77%)	0.12	17 (4%)	35	14	60, 106, 226, 284	0
1	B	391/506 (77%)	0.13	18 (4%)	33	13	61, 107, 228, 321	0
2	C	20/20 (100%)	-0.38	0	100	100	85, 110, 273, 274	0
3	D	20/20 (100%)	-0.52	0	100	100	88, 113, 289, 303	0
4	E	19/20 (95%)	-0.48	0	100	100	81, 110, 266, 282	0
5	F	19/20 (95%)	-0.53	0	100	100	89, 113, 258, 278	0
All	All	859/1092 (78%)	0.07	35 (4%)	38	15	60, 107, 244, 321	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	95	GLY	6.6
1	A	211	PRO	6.0
1	B	95	GLY	4.6
1	A	81	LEU	4.6
1	B	96	SER	3.7

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 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. 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
6	GOL	A	607	6/6	0.72	0.57	8.08	99,122,141,155	0
6	GOL	A	604	6/6	0.69	0.54	8.04	94,125,138,145	0
6	GOL	A	605	6/6	0.87	0.46	6.58	127,143,151,157	0
6	GOL	C	103	6/6	0.79	0.26	5.13	126,139,143,144	0
6	GOL	C	102	6/6	0.82	0.23	4.25	114,133,137,143	0
6	GOL	A	601	6/6	0.80	0.41	3.59	108,128,130,133	0
6	GOL	B	605	6/6	0.92	0.43	2.69	77,111,125,136	0
6	GOL	A	602	6/6	0.79	0.30	2.42	117,125,132,136	0
6	GOL	B	604	6/6	0.89	0.20	0.60	116,124,135,143	0
6	GOL	B	602	6/6	0.82	0.22	0.43	96,116,122,128	0
6	GOL	C	101	6/6	0.92	0.23	0.16	93,113,126,140	0
7	MG	B	606	1/1	0.73	0.13	-1.12	133,133,133,133	0
6	GOL	B	603	6/6	0.63	0.17	-1.19	122,137,145,146	0
7	MG	A	608	1/1	0.56	0.14	-3.35	122,122,122,122	0
6	GOL	B	601	6/6	0.81	0.43	-	105,130,132,142	0
6	GOL	A	606	6/6	0.84	0.18	-	116,122,137,142	0
6	GOL	A	603	6/6	0.57	0.28	-	135,140,151,154	0

6.5 Other polymers

There are no such residues in this entry.