Full wwPDB X-ray Structure Validation Report (i)

May 14, 2020 – 08:20 pm BST

PDB ID	:	5KON
Title	:	Protruding domain of GII.4 human norovirus isolate 08-14
Authors	:	Singh, B.K.; Hansman, G.S.
Deposited on	:	2016-07-01
Resolution	:	1.51 Å(reported)

The following versions of software and data (see references (i)) 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 1.51 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.
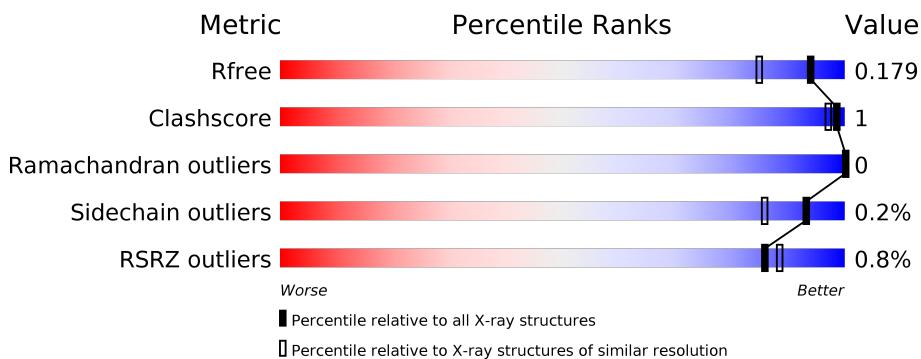
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4009 (1.54-1.50)
Clashscore	141614	4249 (1.54-1.50)
Ramachandran outliers	138981	4148 (1.54-1.50)
Sidechain outliers	138945	4146 (1.54-1.50)
RSRZ outliers	127900	3943 (1.54-1.50)

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	310	100%
1	B	310	99%
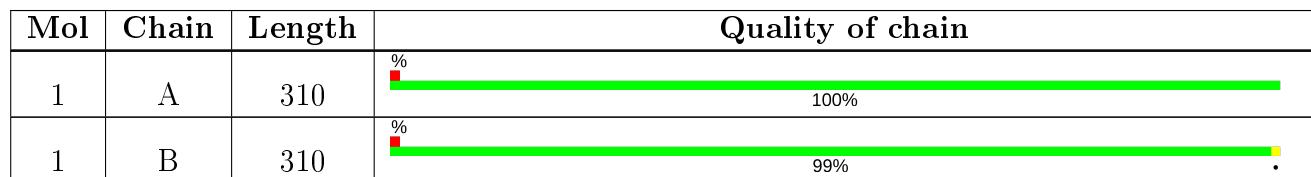
2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9961 atoms, of which 4350 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 VP1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	310	Total 4606	C 1535	H 2186	N 411	O 463	S 11	0	4	0
1	B	310	Total 4593	C 1541	H 2164	N 414	O 463	S 11	0	3	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	221	GLY	-	expression tag	UNP A0A0N9DUX3
A	222	PRO	-	expression tag	UNP A0A0N9DUX3
A	223	GLY	-	expression tag	UNP A0A0N9DUX3
A	224	SER	-	expression tag	UNP A0A0N9DUX3
B	221	GLY	-	expression tag	UNP A0A0N9DUX3
B	222	PRO	-	expression tag	UNP A0A0N9DUX3
B	223	GLY	-	expression tag	UNP A0A0N9DUX3
B	224	SER	-	expression tag	UNP A0A0N9DUX3

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

O2

OH

C1

C2

HO

O1
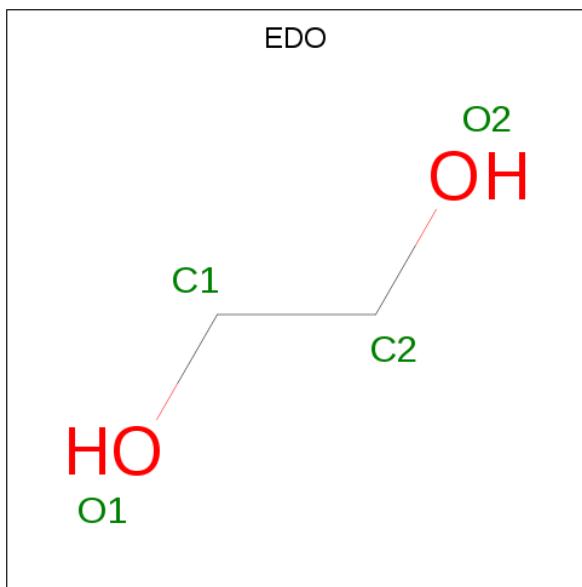
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total C O			0	0
			4	2	2		

- Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).
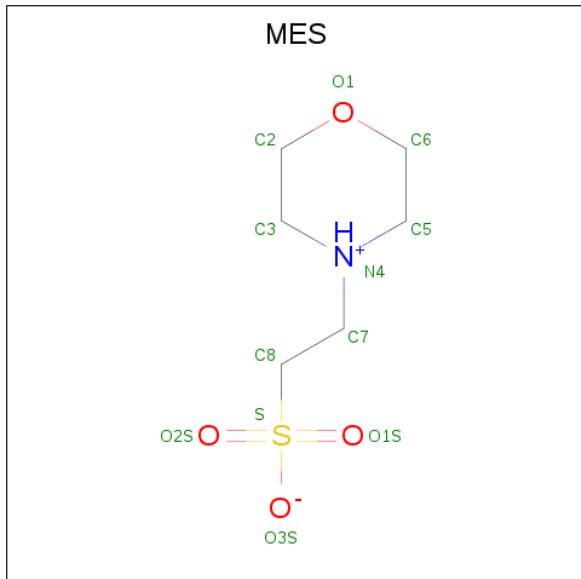
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
3	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
3	B	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	324	Total	O	0	0
			324	324		
4	B	342	Total	O	0	0
			342	342		

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: VP1
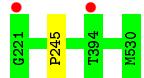
- Molecule 1: VP1

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	63.64Å　55.27Å　86.81Å 90.00°　99.56°　90.00°	Depositor
Resolution (Å)	42.80　–　1.51 42.80　–　1.50	Depositor EDS
% Data completeness (in resolution range)	99.4 (42.80-1.51) 99.0 (42.80-1.50)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.63 (at 1.50Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.126　,　0.175 0.133　,　0.179	Depositor DCC
R_{free} test set	4706 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	10.3	Xtriage
Anisotropy	0.179	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 54.4	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	9961	wwPDB-VP
Average B, all atoms (Å^2)	12.0	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 77.51 % 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 8.3767e-07. 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, MES

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	0.52	0/2501	0.66	0/3421
1	B	0.51	0/2508	0.64	0/3431
All	All	0.51	0/5009	0.65	0/6852

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	2420	2186	2307	1	0
1	B	2429	2164	2312	4	0
2	A	40	0	60	0	0
2	B	20	0	30	3	0
3	A	24	0	25	0	0
3	B	12	0	12	1	0
4	A	324	0	0	0	0
4	B	342	0	0	0	0
All	All	5611	4350	4746	5	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 1.

All (5) 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:B:257:PHE:HZ	2:B:602:EDO:H22	1.54	0.72
1:B:257:PHE:CZ	2:B:602:EDO:H22	2.30	0.66
1:A:245:PRO:HG3	1:B:282:ASN:OD1	2.19	0.43
3:B:606:MES:H81	3:B:606:MES:H51	1.75	0.42
1:B:232:LEU:HD11	2:B:605:EDO:H11	2.03	0.40

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	312/310 (101%)	307 (98%)	5 (2%)	0	100	100
1	B	311/310 (100%)	304 (98%)	7 (2%)	0	100	100
All	All	623/620 (100%)	611 (98%)	12 (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	268/267 (100%)	268 (100%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	268/267 (100%)	266 (99%)	2 (1%)	84	69
All	All	536/534 (100%)	534 (100%)	2 (0%)	93	82

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	289[A]	ASP
1	B	289[B]	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

18 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	EDO	A	605	-	3,3,3	0.48	0	2,2,2	0.73	0
2	EDO	A	604	-	3,3,3	0.48	0	2,2,2	0.45	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#|Z| > 2	Counts	RMSZ	#|Z| > 2
2	EDO	A	601	-	3,3,3	0.42	0	2,2,2	0.46	0
2	EDO	B	604	-	3,3,3	0.54	0	2,2,2	0.25	0
3	MES	A	612	-	12,12,12	1.73	1 (8%)	14,16,16	1.93	6 (42%)
2	EDO	A	609	-	3,3,3	0.57	0	2,2,2	0.20	0
2	EDO	A	607	-	3,3,3	0.60	0	2,2,2	0.58	0
2	EDO	A	608	-	3,3,3	0.64	0	2,2,2	0.09	0
2	EDO	B	605	-	3,3,3	0.77	0	2,2,2	0.08	0
2	EDO	A	603	-	3,3,3	0.48	0	2,2,2	0.08	0
2	EDO	B	601	-	3,3,3	0.50	0	2,2,2	0.40	0
2	EDO	A	610	-	3,3,3	0.45	0	2,2,2	0.53	0
3	MES	B	606	-	12,12,12	1.83	1 (8%)	14,16,16	2.28	6 (42%)
3	MES	A	611	-	12,12,12	1.96	1 (8%)	14,16,16	1.79	4 (28%)
2	EDO	A	606	-	3,3,3	0.57	0	2,2,2	0.34	0
2	EDO	B	602	-	3,3,3	0.51	0	2,2,2	0.26	0
2	EDO	A	602	-	3,3,3	0.41	0	2,2,2	0.67	0
2	EDO	B	603	-	3,3,3	0.47	0	2,2,2	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	EDO	A	605	-	-	0/1/1/1	-
2	EDO	A	604	-	-	0/1/1/1	-
2	EDO	A	601	-	-	0/1/1/1	-
2	EDO	B	604	-	-	0/1/1/1	-
3	MES	A	612	-	-	1/6/14/14	0/1/1/1
2	EDO	A	609	-	-	0/1/1/1	-
2	EDO	A	607	-	-	0/1/1/1	-
2	EDO	A	608	-	-	0/1/1/1	-
2	EDO	B	605	-	-	1/1/1/1	-
2	EDO	A	603	-	-	0/1/1/1	-
2	EDO	B	601	-	-	0/1/1/1	-
2	EDO	A	610	-	-	0/1/1/1	-
3	MES	B	606	-	-	0/6/14/14	0/1/1/1
3	MES	A	611	-	-	3/6/14/14	0/1/1/1
2	EDO	A	606	-	-	0/1/1/1	-
2	EDO	B	602	-	-	0/1/1/1	-
2	EDO	A	602	-	-	0/1/1/1	-
2	EDO	B	603	-	-	0/1/1/1	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	611	MES	C8-S	-6.47	1.68	1.77
3	B	606	MES	C8-S	-5.91	1.69	1.77
3	A	612	MES	C8-S	-5.57	1.69	1.77

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	606	MES	C5-N4-C3	4.65	119.30	108.83
3	A	611	MES	O3S-S-C8	4.28	112.68	105.77
3	B	606	MES	O2S-S-C8	-3.83	102.30	106.92
3	A	612	MES	C5-N4-C3	3.72	117.20	108.83
3	A	612	MES	O1S-S-C8	3.36	110.96	106.92
3	B	606	MES	O3S-S-C8	3.30	111.10	105.77
3	A	612	MES	O2S-S-C8	3.19	110.76	106.92
3	B	606	MES	O1S-S-C8	3.12	110.67	106.92
3	A	611	MES	C5-N4-C3	3.07	115.74	108.83
3	A	611	MES	O2S-S-C8	2.78	110.26	106.92
3	B	606	MES	C7-N4-C5	2.32	117.16	111.23
3	A	611	MES	O2S-S-O1S	-2.31	105.95	113.95
3	A	612	MES	C7-N4-C5	2.11	116.63	111.23
3	A	612	MES	C6-C5-N4	-2.08	106.94	110.10
3	A	612	MES	O3S-S-C8	-2.07	102.42	105.77
3	B	606	MES	C6-C5-N4	-2.07	106.97	110.10

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	611	MES	C7-C8-S-O1S
3	A	611	MES	C7-C8-S-O3S
2	B	605	EDO	O1-C1-C2-O2
3	A	611	MES	C7-C8-S-O2S
3	A	612	MES	C8-C7-N4-C5

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	605	EDO	1	0
3	B	606	MES	1	0

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	602	EDO	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 (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	310/310 (100%)	-0.49	2 (0%)	89	91	5, 9, 20, 30	0
1	B	310/310 (100%)	-0.47	3 (0%)	82	85	5, 10, 21, 28	0
All	All	620/620 (100%)	-0.48	5 (0%)	86	88	5, 9, 21, 30	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	352[A]	TYR	4.4
1	A	221	GLY	3.2
1	B	294	THR	2.5
1	B	222	PRO	2.4
1	A	394	THR	2.0

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 (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(Å²)	Q<0.9
2	EDO	A	607	4/4	0.86	0.19	19,21,23,26	0
2	EDO	A	608	4/4	0.92	0.11	20,23,34,36	0
2	EDO	B	605	4/4	0.93	0.11	12,18,20,37	0
3	MES	B	606	12/12	0.95	0.16	21,31,48,53	0
3	MES	A	611	12/12	0.95	0.11	17,24,36,45	0
2	EDO	A	605	4/4	0.96	0.07	12,12,23,25	0
2	EDO	A	602	4/4	0.96	0.07	12,17,17,22	0
2	EDO	A	609	4/4	0.97	0.09	9,17,21,25	0
3	MES	A	612	12/12	0.97	0.11	18,33,36,46	0
2	EDO	A	610	4/4	0.97	0.09	11,19,23,27	0
2	EDO	B	604	4/4	0.98	0.10	8,15,22,23	0
2	EDO	A	601	4/4	0.98	0.06	11,11,19,21	0
2	EDO	A	603	4/4	0.98	0.05	11,15,17,23	0
2	EDO	A	606	4/4	0.98	0.09	14,14,15,24	0
2	EDO	B	602	4/4	0.98	0.08	14,18,23,29	0
2	EDO	B	601	4/4	0.98	0.05	12,19,21,25	0
2	EDO	B	603	4/4	0.98	0.06	15,16,21,22	0
2	EDO	A	604	4/4	0.99	0.04	11,11,11,13	0

6.5 Other polymers (i)

There are no such residues in this entry.