wwPDB X-ray Structure Validation Summary Report (i)

Feb 16, 2023 – 12:18 pm GMT

PDB ID	:	8AGH
Title	:	BK Polyomavirus VP1 mutant E73A
Authors	:	Sorin, M.N.; Di Maio, A.; Silva, L.M.; Ebert, D.; Delannoy, C.; Nguyen, N.-K.; Guerardel, Y.; Chai, W.; Halary, F.; Renaudin-Autain, K.; Liu, Y.; Bressollette-Bodin, C.; Stehle, T.; McIlroy, D.
Deposited on	:	2022-07-20
Resolution	:	1.89 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 (1)) were used in the production of this report:

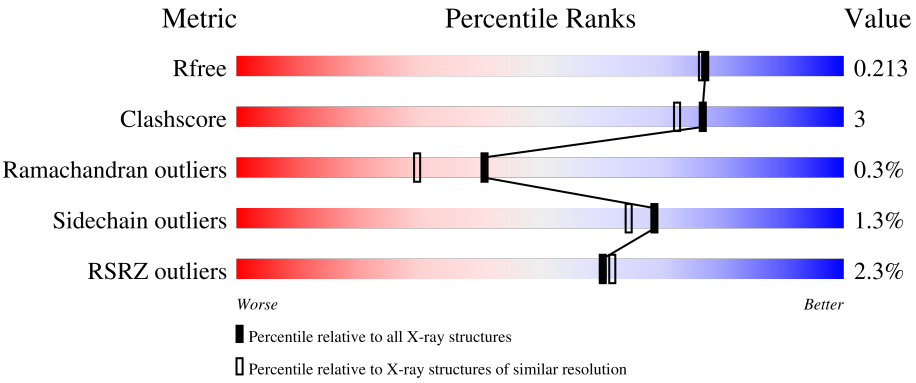
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.32.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

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 1.89 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	9470 (1.90-1.86)
Clashscore	141614	10282 (1.90-1.86)
Ramachandran outliers	138981	10152 (1.90-1.86)
Sidechain outliers	138945	10152 (1.90-1.86)
RSRZ outliers	127900	9303 (1.90-1.86)

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	AAA	271	89% 6% 5%
1	BBB	271	3% 88% 7%
1	CCC	271	4% 85% 8% 7%
1	DDD	271	90% 7%
1	EEE	271	2% 88% 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 10521 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.

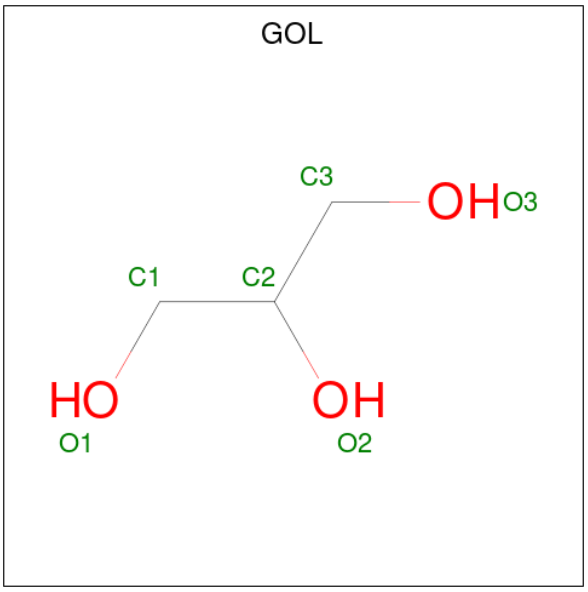
- Molecule 1 is a protein called Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	258	Total	C	N	O	S	0	0	0
			1956	1232	336	376	12			
1	BBB	261	Total	C	N	O	S	0	1	0
			1971	1242	341	375	13			
1	CCC	253	Total	C	N	O	S	0	2	0
			1900	1197	327	363	13			
1	DDD	266	Total	C	N	O	S	0	0	0
			2003	1262	346	383	12			
1	EEE	262	Total	C	N	O	S	1	8	0
			2030	1276	350	392	12			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	72	ALA	GLU	engineered mutation	UNP P03088
AAA	104	SER	CYS	engineered mutation	UNP P03088
BBB	72	ALA	GLU	engineered mutation	UNP P03088
BBB	104	SER	CYS	engineered mutation	UNP P03088
CCC	72	ALA	GLU	engineered mutation	UNP P03088
CCC	104	SER	CYS	engineered mutation	UNP P03088
DDD	72	ALA	GLU	engineered mutation	UNP P03088
DDD	104	SER	CYS	engineered mutation	UNP P03088
EEE	72	ALA	GLU	engineered mutation	UNP P03088
EEE	104	SER	CYS	engineered mutation	UNP P03088

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	1	Total C O 6 3 3	0	0
2	BBB	1	Total C O 6 3 3	0	0
2	CCC	1	Total C O 6 3 3	0	0
2	DDD	1	Total C O 6 3 3	0	0
2	EEE	1	Total C O 6 3 3	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total Cl 1 1	0	0
3	BBB	1	Total Cl 1 1	0	0
3	CCC	1	Total Cl 1 1	0	0
3	DDD	1	Total Cl 1 1	0	0
3	EEE	1	Total Cl 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	130	Total O 130 130	0	0
4	BBB	113	Total O 113 113	0	0
4	CCC	121	Total O 121 121	0	0
4	DDD	126	Total O 126 126	0	0
4	EEE	136	Total O 136 136	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: Major capsid protein VP1

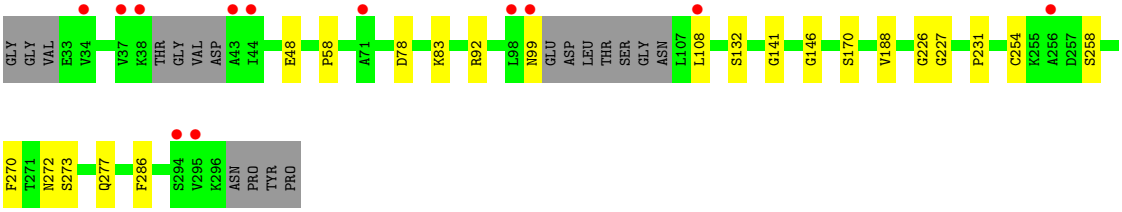
Chain AAA:

• Molecule 1: Major capsid protein VP1

Chain BBB:

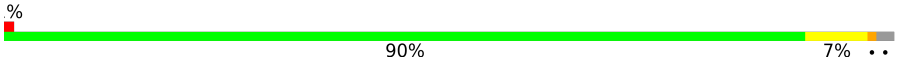
• Molecule 1: Major capsid protein VP1

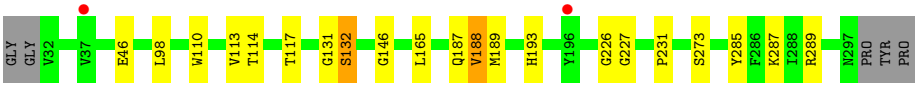
Chain CCC:



• Molecule 1: Major capsid protein VP1

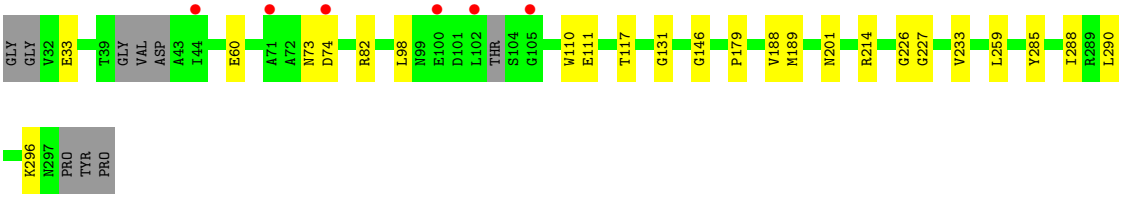
Chain DDD:





• Molecule 1: Major capsid protein VP1

Chain EEE:


GLY
GLY
V32
E33
T39
GLY
VAL
ASP
A43
I44
E60
A71
A72
N73
D74
R82
L98
N99
E100
D101
L102
THR
S104
G105
W110
E111
T117
G131
G146
P179
V188
M189
N201
R214
G226
G227
V233
L259
Y285
I288
R289
L290

K296
N297
PRO
TYR
PRO

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	144.49Å 152.31Å 62.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.90 – 1.89 47.90 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.90-1.89) 99.9 (47.90-1.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.32 (at 1.88Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.174 , 0.211 0.182 , 0.213	Depositor DCC
R_{free} test set	5587 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	28.6	Xtriage
Anisotropy	0.153	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 39.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10521	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.41% 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

5.1 Standard geometry

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

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	AAA	0.69	0/2001	0.91	0/2725
1	BBB	0.69	0/2016	0.87	0/2746
1	CCC	0.70	0/1944	0.88	0/2650
1	DDD	0.71	0/2049	0.89	0/2794
1	EEE	0.70	0/2083	0.90	2/2836 (0.1%)
All	All	0.70	0/10093	0.89	2/13751 (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	EEE	214[A]	ARG	NE-CZ-NH2	-6.90	116.85	120.30
1	EEE	214[B]	ARG	NE-CZ-NH2	-6.90	116.85	120.30

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	AAA	1956	0	1869	11	0
1	BBB	1971	0	1871	12	0
1	CCC	1900	0	1772	13	0
1	DDD	2003	0	1907	15	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	EEE	2030	0	1924	14	0
2	AAA	6	0	8	0	0
2	BBB	6	0	8	0	0
2	CCC	6	0	8	0	0
2	DDD	6	0	8	0	0
2	EEE	6	0	8	0	0
3	AAA	1	0	0	0	0
3	BBB	1	0	0	0	0
3	CCC	1	0	0	0	0
3	DDD	1	0	0	1	0
3	EEE	1	0	0	0	0
4	AAA	130	0	0	1	0
4	BBB	113	0	0	0	0
4	CCC	121	0	0	1	0
4	DDD	126	0	0	3	0
4	EEE	136	0	0	2	0
All	All	10521	0	9383	57	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 57 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:CCC:99:ASN:HD21	1:CCC:108:LEU:H	1.32	0.76
1:DDD:132:SER:HB3	4:DDD:505:HOH:O	1.93	0.68
1:EEE:74[A]:ASP:OD1	4:EEE:501:HOH:O	2.10	0.67
1:BBB:297:ASN:N	1:BBB:297:ASN:OD1	2.30	0.63
1:DDD:131:GLY:N	1:EEE:73[B]:ASN:O	2.32	0.63

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	AAA	254/271 (94%)	244 (96%)	10 (4%)	0	100 100
1	BBB	258/271 (95%)	242 (94%)	14 (5%)	2 (1%)	19 9
1	CCC	249/271 (92%)	239 (96%)	9 (4%)	1 (0%)	34 22
1	DDD	264/271 (97%)	256 (97%)	7 (3%)	1 (0%)	34 22
1	EEE	264/271 (97%)	250 (95%)	14 (5%)	0	100 100
All	All	1289/1355 (95%)	1231 (96%)	54 (4%)	4 (0%)	41 30

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BBB	188	VAL
1	CCC	188	VAL
1	DDD	188	VAL
1	BBB	74	ASP

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	AAA	209/231 (90%)	205 (98%)	4 (2%)	57 49
1	BBB	207/231 (90%)	202 (98%)	5 (2%)	49 39
1	CCC	196/231 (85%)	195 (100%)	1 (0%)	88 88
1	DDD	211/231 (91%)	209 (99%)	2 (1%)	78 76
1	EEE	214/231 (93%)	213 (100%)	1 (0%)	88 88
All	All	1037/1155 (90%)	1024 (99%)	13 (1%)	69 64

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	273	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	BBB	297	ASN
1	EEE	179	PRO
1	DDD	132	SER
1	DDD	273	SER

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 5 are monoatomic - leaving 5 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	$\#\|Z\| > 2$
2	GOL	AAA	501	-	5,5,5	0.13	0	5,5,5	0.41	0
2	GOL	BBB	501	-	5,5,5	0.15	0	5,5,5	0.43	0
2	GOL	DDD	401	-	5,5,5	0.16	0	5,5,5	0.42	0
2	GOL	EEE	401	-	5,5,5	0.13	0	5,5,5	0.41	0
2	GOL	CCC	401	-	5,5,5	0.17	0	5,5,5	0.49	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	AAA	501	-	-	1/4/4/4	-
2	GOL	BBB	501	-	-	0/4/4/4	-
2	GOL	DDD	401	-	-	4/4/4/4	-
2	GOL	EEE	401	-	-	2/4/4/4	-
2	GOL	CCC	401	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	CCC	401	GOL	C1-C2-C3-O3
2	DDD	401	GOL	O1-C1-C2-C3
2	EEE	401	GOL	C1-C2-C3-O3
2	EEE	401	GOL	O2-C2-C3-O3
2	DDD	401	GOL	C1-C2-C3-O3

There are no ring outliers.

No monomer is involved in short contacts.

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	AAA	258/271 (95%)	-0.16	3 (1%)	79	80	22, 30, 49, 67	0
1	BBB	261/271 (96%)	-0.05	7 (2%)	54	56	22, 31, 57, 86	0
1	CCC	253/271 (93%)	0.04	12 (4%)	31	33	22, 33, 58, 77	0
1	DDD	266/271 (98%)	-0.10	2 (0%)	86	87	20, 30, 55, 66	0
1	EEE	262/271 (96%)	-0.13	6 (2%)	60	62	21, 28, 48, 92	0
All	All	1300/1355 (95%)	-0.08	30 (2%)	60	62	20, 31, 55, 92	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	CCC	108	LEU	4.5
1	CCC	99	ASN	4.4
1	EEE	105	GLY	4.1
1	BBB	41	VAL	3.6
1	BBB	72	ALA	3.4

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 monosaccharides 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
2	GOL	DDD	401	6/6	0.82	0.23	43,57,65,67	0
2	GOL	EEE	401	6/6	0.83	0.21	38,57,57,63	0
2	GOL	BBB	501	6/6	0.84	0.17	43,53,56,58	0
2	GOL	AAA	501	6/6	0.85	0.24	50,62,68,73	0
2	GOL	CCC	401	6/6	0.85	0.19	42,50,51,58	0
3	CL	CCC	402	1/1	0.85	0.10	51,51,51,51	0
3	CL	DDD	402	1/1	0.97	0.05	45,45,45,45	0
3	CL	EEE	402	1/1	0.97	0.09	45,45,45,45	0
3	CL	AAA	502	1/1	0.98	0.07	39,39,39,39	0
3	CL	BBB	502	1/1	0.99	0.04	41,41,41,41	0

6.5 Other polymers (i)

There are no such residues in this entry.