wwPDB X-ray Structure Validation Summary Report (i)

Oct 4, 2023 – 01:11 pm BST

PDB ID	:	8BTJ
Title	:	Murine cytomegalovirus protein M35
Authors	:	Schmelz, S.; Van den Heuvel, J.; Blankenfeldt, W.
Deposited on	:	2022-11-29
Resolution	:	1.94 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:

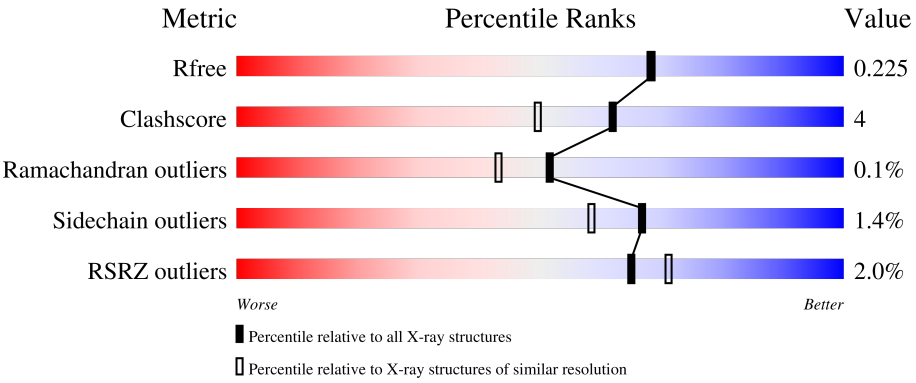
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.94 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	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	453	2% 80% 9% 11%
1	B	453	2% 78% 10% 11%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 13479 atoms, of which 6620 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.

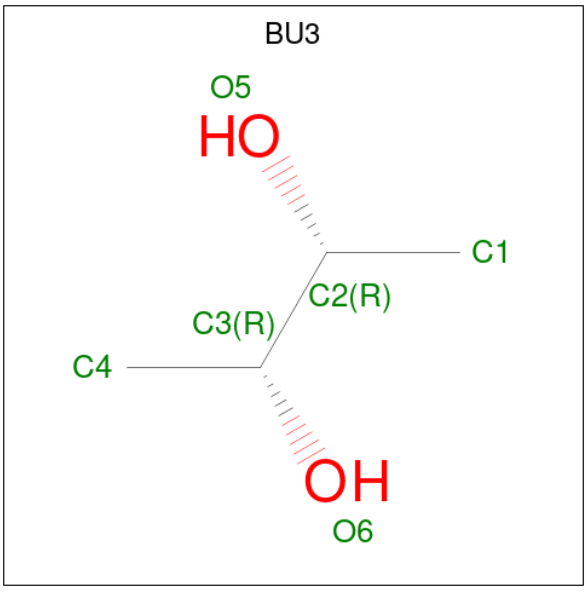
- Molecule 1 is a protein called Protein M35.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	403	Total	C	H	N	O	S	0	15	0
			6530	2053	3273	578	603	23			
1	B	402	Total	C	H	N	O	S	0	14	0
			6481	2041	3245	571	602	22			

There are 4 discrepancies between the modelled and reference sequences:

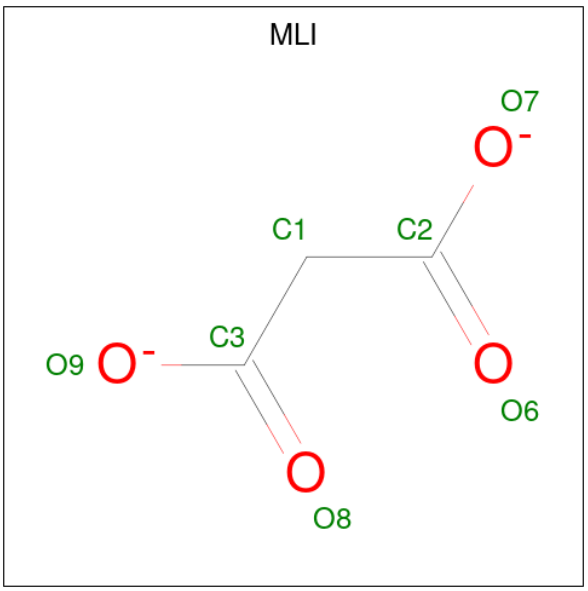
Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP A8E1G1
A	1	SER	-	expression tag	UNP A8E1G1
B	0	GLY	-	expression tag	UNP A8E1G1
B	1	SER	-	expression tag	UNP A8E1G1

- Molecule 2 is (R,R)-2,3-BUTANEDIOL (three-letter code: BU3) (formula: $C_4H_{10}O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C H O 16 4 10 2	0	0
2	A	1	Total C H O 16 4 10 2	0	0
2	A	1	Total C H O 16 4 10 2	0	0
2	A	1	Total C H O 16 4 10 2	0	0
2	A	1	Total C H O 16 4 10 2	0	0
2	A	1	Total C H O 16 4 10 2	0	0
2	B	1	Total C H O 16 4 10 2	0	0
2	B	1	Total C H O 16 4 10 2	0	0
2	B	1	Total C H O 16 4 10 2	0	0
2	B	1	Total C H O 16 4 10 2	0	0

- Molecule 3 is MALONATE ION (three-letter code: MLI) (formula: $C_3H_2O_4$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C H O 9 3 2 4	0	0

- Molecule 4 is water.

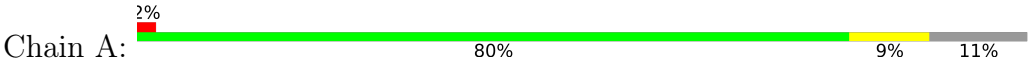
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	153	Total O 153 153	0	0
4	B	146	Total O 146 146	0	0

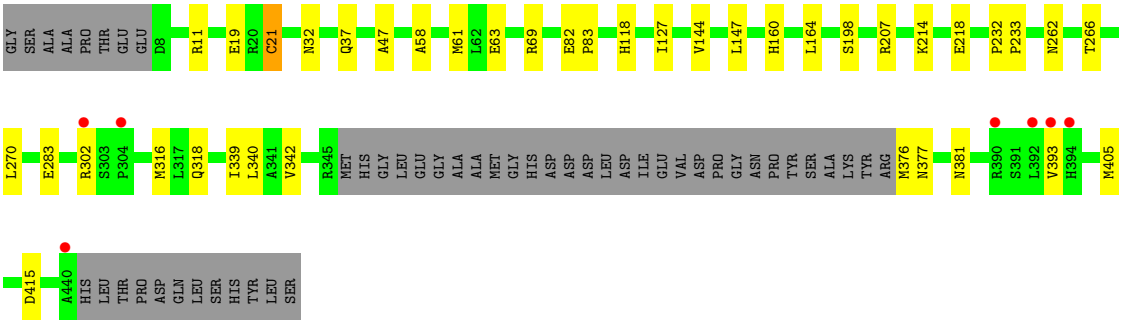
3 Residue-property plots (i)

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: Protein M35

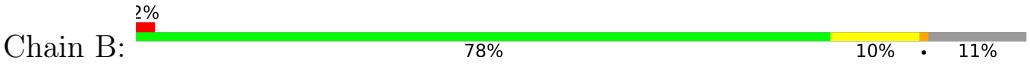
Chain A:

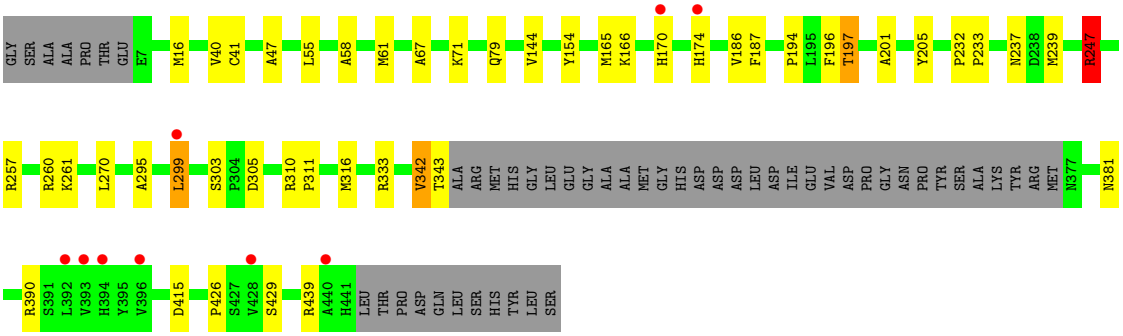




• Molecule 1: Protein M35

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	59.09Å 130.71Å 67.83Å 90.00° 101.26° 90.00°	Depositor
Resolution (Å)	57.95 – 1.94 66.52 – 1.94	Depositor EDS
% Data completeness (in resolution range)	67.8 (57.95-1.94) 63.0 (66.52-1.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	-0.33 (at 1.94Å)	Xtriage
Refinement program	PHENIX 1.20-4459	Depositor
R, R_{free}	0.174 , 0.226 0.174 , 0.225	Depositor DCC
R_{free} test set	2544 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	26.1	Xtriage
Anisotropy	0.114	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 51.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	13479	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.48% 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: MLI, BU3

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	0.60	1/3372 (0.0%)	0.75	0/4572
1	B	0.62	0/3349	0.76	0/4544
All	All	0.61	1/6721 (0.0%)	0.76	0/9116

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	A	0	1
1	B	0	3
All	All	0	4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	283	GLU	CD-OE1	5.14	1.31	1.25

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	302	ARG	Sidechain
1	B	247[A]	ARG	Sidechain
1	B	257	ARG	Sidechain
1	B	260	ARG	Sidechain

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	3257	3273	3214	29	0
1	B	3236	3245	3183	32	0
2	A	36	60	60	7	0
2	B	24	40	40	3	0
3	A	7	2	2	0	0
4	A	153	0	0	3	0
4	B	146	0	0	2	0
All	All	6859	6620	6499	58	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 4.

The worst 5 of 58 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:261:LYS:NZ	4:B:601:HOH:O	2.15	0.79
1:B:270:LEU:HD13	1:B:316:MET:SD	2.24	0.78
1:B:16:MET:O	2:B:501:BU3:H41	1.87	0.73
1:B:186:VAL:HG23	1:B:187:PHE:CD2	2.28	0.69
1:A:118:HIS:ND1	4:A:603:HOH:O	2.26	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	414/453 (91%)	403 (97%)	11 (3%)	0	100	100
1	B	412/453 (91%)	405 (98%)	6 (2%)	1 (0%)	47	39
All	All	826/906 (91%)	808 (98%)	17 (2%)	1 (0%)	51	43

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	426	PRO

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	366/393 (93%)	362 (99%)	4 (1%)	73	67
1	B	366/393 (93%)	358 (98%)	8 (2%)	52	39
All	All	732/786 (93%)	720 (98%)	12 (2%)	67	52

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	247[B]	ARG
1	B	299[A]	LEU
1	B	429	SER
1	B	299[B]	LEU
1	A	415	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	42	ASN
1	B	383	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)

11 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	BU3	B	503	-	4,5,5	0.35	0	6,6,6	0.27	0
2	BU3	B	502	-	4,5,5	0.24	0	6,6,6	0.61	0
2	BU3	B	500	-	4,5,5	0.51	0	6,6,6	0.88	0
2	BU3	A	503	-	4,5,5	0.29	0	6,6,6	0.76	0
2	BU3	B	501	-	4,5,5	0.24	0	6,6,6	0.55	0
2	BU3	A	504	-	4,5,5	0.47	0	6,6,6	0.98	0
2	BU3	A	500	-	4,5,5	0.13	0	6,6,6	0.88	0
2	BU3	A	505	-	4,5,5	0.29	0	6,6,6	0.68	0
2	BU3	A	502	-	4,5,5	0.25	0	6,6,6	0.33	0
2	BU3	A	506	-	4,5,5	0.49	0	6,6,6	0.73	0
3	MLI	A	501	-	6,6,6	1.92	1 (16%)	7,7,7	1.15	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	BU3	B	503	-	-	4/4/4/4	-
2	BU3	B	502	-	-	4/4/4/4	-
2	BU3	B	500	-	-	1/4/4/4	-
2	BU3	A	503	-	-	4/4/4/4	-
2	BU3	B	501	-	-	0/4/4/4	-
2	BU3	A	504	-	-	2/4/4/4	-
2	BU3	A	500	-	-	0/4/4/4	-
2	BU3	A	505	-	-	4/4/4/4	-
2	BU3	A	502	-	-	0/4/4/4	-
2	BU3	A	506	-	-	0/4/4/4	-
3	MLI	A	501	-	-	2/4/4/4	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	501	MLI	C1-C3	3.52	1.56	1.51

There are no bond angle outliers.

There are no chirality outliers.

5 of 21 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	503	BU3	O5-C2-C3-O6
2	A	503	BU3	C1-C2-C3-O6
2	A	503	BU3	O5-C2-C3-C4
2	A	505	BU3	O5-C2-C3-O6
2	A	505	BU3	C1-C2-C3-O6

There are no ring outliers.

6 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	500	BU3	1	0
2	B	501	BU3	2	0
2	A	504	BU3	2	0
2	A	500	BU3	1	0
2	A	505	BU3	3	0
2	A	502	BU3	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

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	A	403/453 (88%)	0.06	7 (1%)	70	75	20, 32, 62, 98	0
1	B	402/453 (88%)	0.07	9 (2%)	62	69	19, 32, 65, 89	0
All	All	805/906 (88%)	0.07	16 (1%)	65	71	19, 32, 65, 98	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	394	HIS	5.1
1	A	392	LEU	4.5
1	B	393	VAL	3.9
1	B	396	VAL	3.8
1	A	394	HIS	3.7

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	BU3	B	501	6/6	0.83	0.27	55,67,84,88	0
3	MLI	A	501	7/7	0.85	0.19	51,61,73,76	0
2	BU3	A	506	6/6	0.86	0.25	50,65,83,83	0
2	BU3	B	502	6/6	0.87	0.20	54,69,87,87	0
2	BU3	A	500	6/6	0.90	0.21	45,54,65,68	0
2	BU3	A	505	6/6	0.91	0.39	41,63,76,76	0
2	BU3	A	504	6/6	0.91	0.27	37,51,76,77	0
2	BU3	B	500	6/6	0.91	0.27	41,58,77,93	0
2	BU3	B	503	6/6	0.93	0.22	33,47,64,64	0
2	BU3	A	503	6/6	0.93	0.24	50,68,88,88	0
2	BU3	A	502	6/6	0.94	0.25	49,59,74,89	0

6.5 Other polymers (i)

There are no such residues in this entry.