wwPDB X-ray Structure Validation Summary Report (i)

Nov 24, 2021 – 01:14 pm GMT

PDB ID	:	7NFQ
Title	:	Fujian capmidlink domain in complex with Nb8193
Authors	:	Keown, J.R.; Grimes, J.M.; Fodor, E.
Deposited on	:	2021-02-07
Resolution	:	1.68 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 following versions of software and data (see references (1)) were used in the production of this report:

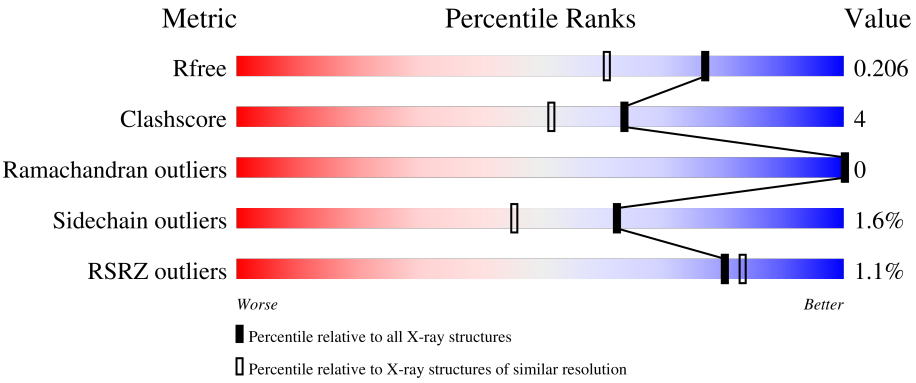
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 1.68 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	6780 (1.70-1.66)
Clashscore	141614	7310 (1.70-1.66)
Ramachandran outliers	138981	7173 (1.70-1.66)
Sidechain outliers	138945	7172 (1.70-1.66)
RSRZ outliers	127900	6661 (1.70-1.66)

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	304	84% 8% 8%
1	B	304	80% 12% 7%
2	C	127	79% 12% 9%
2	D	127	85% 6% 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12989 atoms, of which 6295 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 Polymerase basic protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	279	Total	C	H	N	O	S	0	1	0
			4431	1358	2240	401	416	16			
1	B	282	Total	C	H	N	O	S	0	3	0
			4511	1384	2280	408	423	16			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	233	HIS	-	expression tag	UNP Q6E3N3
A	234	HIS	-	expression tag	UNP Q6E3N3
A	235	HIS	-	expression tag	UNP Q6E3N3
A	236	HIS	-	expression tag	UNP Q6E3N3
A	237	HIS	-	expression tag	UNP Q6E3N3
A	238	HIS	-	expression tag	UNP Q6E3N3
A	239	GLY	-	expression tag	UNP Q6E3N3
A	240	GLU	-	expression tag	UNP Q6E3N3
A	241	ASN	-	expression tag	UNP Q6E3N3
A	242	LEU	-	expression tag	UNP Q6E3N3
A	243	TYR	-	expression tag	UNP Q6E3N3
A	244	PHE	-	expression tag	UNP Q6E3N3
A	245	GLN	-	expression tag	UNP Q6E3N3
A	246	GLY	-	expression tag	UNP Q6E3N3
B	233	HIS	-	expression tag	UNP Q6E3N3
B	234	HIS	-	expression tag	UNP Q6E3N3
B	235	HIS	-	expression tag	UNP Q6E3N3
B	236	HIS	-	expression tag	UNP Q6E3N3
B	237	HIS	-	expression tag	UNP Q6E3N3
B	238	HIS	-	expression tag	UNP Q6E3N3
B	239	GLY	-	expression tag	UNP Q6E3N3
B	240	GLU	-	expression tag	UNP Q6E3N3
B	241	ASN	-	expression tag	UNP Q6E3N3
B	242	LEU	-	expression tag	UNP Q6E3N3
B	243	TYR	-	expression tag	UNP Q6E3N3

Continued on next page...

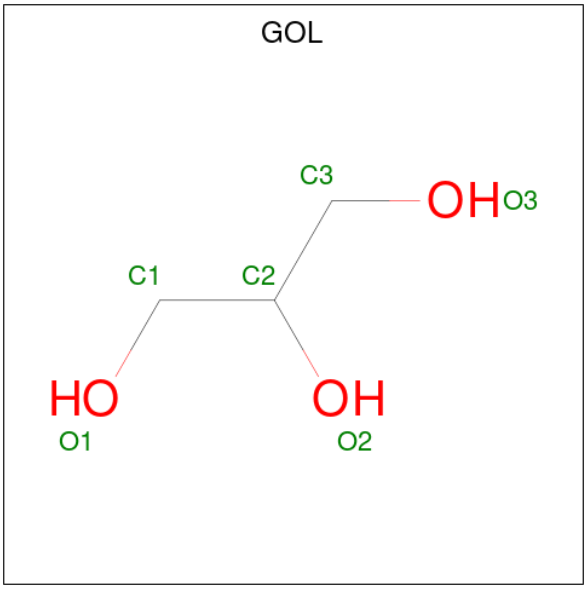
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	244	PHE	-	expression tag	UNP Q6E3N3
B	245	GLN	-	expression tag	UNP Q6E3N3
B	246	GLY	-	expression tag	UNP Q6E3N3

- Molecule 2 is a protein called Nb8193.

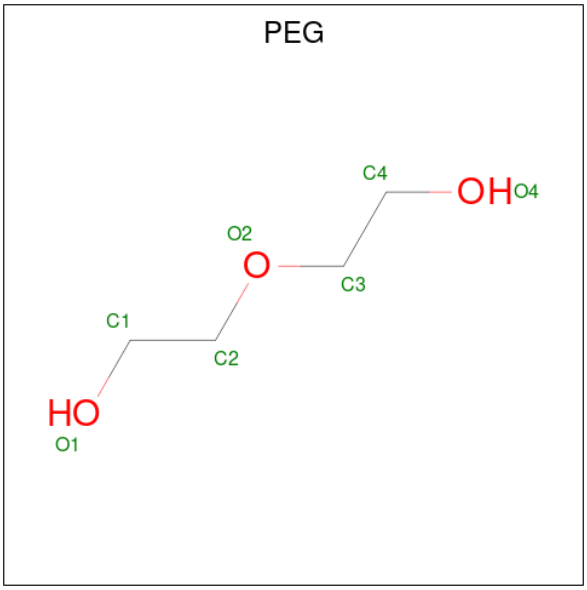
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	116	Total	C	H	N	O	S	0	1	0
			1762	559	867	157	173	6			
2	D	116	Total	C	H	N	O	S	0	0	0
			1745	554	858	156	172	5			

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			14	3	8	3		
3	B	1	Total	C	H	O	0	0
			14	3	8	3		
3	B	1	Total	C	H	O	0	0
			14	3	8	3		
3	B	1	Total	C	H	O	0	0
			14	3	8	3		
3	D	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	H	O	0	0
			17	4	10	3		

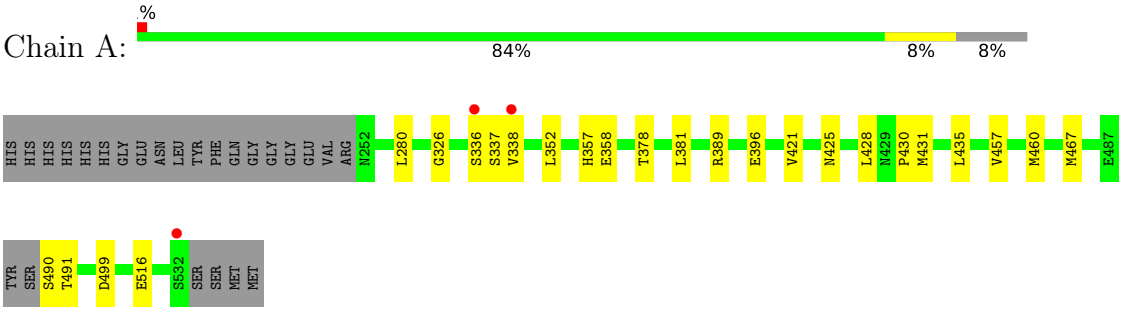
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	133	Total	O	0	0
			133	133		
5	C	54	Total	O	0	0
			54	54		
5	B	177	Total	O	0	0
			177	177		
5	D	89	Total	O	0	0
			89	89		

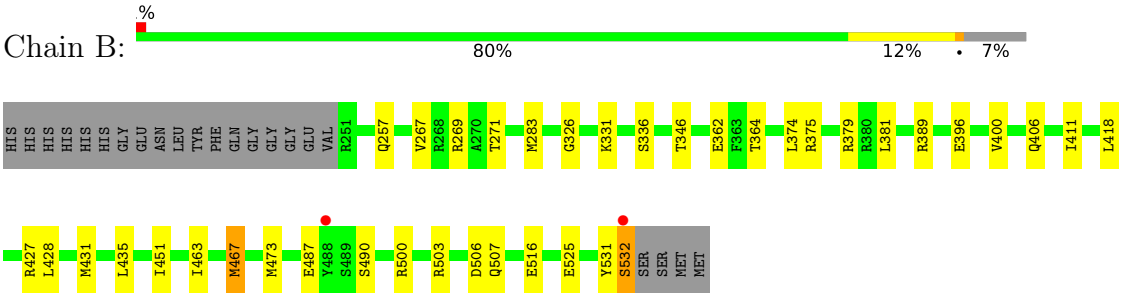
3 Residue-property plots (i)

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.

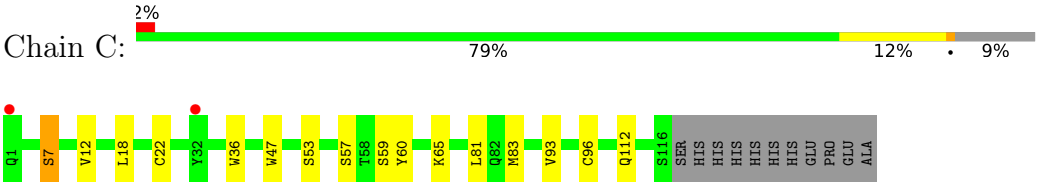
• Molecule 1: Polymerase basic protein 2



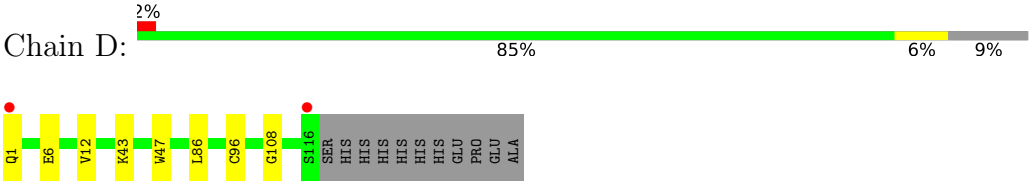
• Molecule 1: Polymerase basic protein 2



• Molecule 2: Nb8193



• Molecule 2: Nb8193



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	42.58Å 106.54Å 98.41Å 90.00° 98.01° 90.00°	Depositor
Resolution (Å)	44.31 – 1.68 53.27 – 1.68	Depositor EDS
% Data completeness (in resolution range)	70.3 (44.31-1.68) 67.1 (53.27-1.68)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.59 (at 1.68Å)	Xtriage
Refinement program	PHENIX 1.18rc2_3794	Depositor
R, R_{free}	0.167 , 0.205 0.167 , 0.206	Depositor DCC
R_{free} test set	3442 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	25.6	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	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	0.97	EDS
Total number of atoms	12989	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.13% 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: PEG, 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	0.54	0/2217	0.73	2/2982 (0.1%)
1	B	0.59	0/2262	0.76	2/3045 (0.1%)
2	C	0.58	1/914 (0.1%)	0.75	1/1236 (0.1%)
2	D	0.67	1/906 (0.1%)	0.81	0/1226
All	All	0.58	2/6299 (0.0%)	0.75	5/8489 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	96	CYS	CB-SG	-8.08	1.68	1.82
2	C	96	CYS	CB-SG	-7.03	1.70	1.82

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	431	MET	CG-SD-CE	9.87	115.99	100.20
2	C	83	MET	CG-SD-CE	7.96	112.93	100.20
1	B	506	ASP	CB-CG-OD1	6.21	123.89	118.30
1	A	499	ASP	CB-CG-OD2	5.99	123.69	118.30
1	B	467	MET	CG-SD-CE	5.89	109.62	100.20

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	A	2191	2240	2241	15	0
1	B	2231	2280	2284	27	0
2	C	895	867	866	7	0
2	D	887	858	858	4	0
3	A	6	8	8	1	0
3	B	18	24	24	3	0
3	D	6	8	8	1	0
4	B	7	10	10	0	0
5	A	133	0	0	1	0
5	B	177	0	0	5	0
5	C	54	0	0	1	0
5	D	89	0	0	3	0
All	All	6694	6295	6299	52	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 52 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:269:ARG:HB2	1:B:283:MET:HE1	1.74	0.70
1:B:418:LEU:O	1:B:427:ARG:NH2	2.26	0.67
1:A:352:LEU:HD22	1:A:430:PRO:HB3	1.76	0.65
2:D:1:GLN:NE2	5:D:402:HOH:O	2.29	0.64
1:B:379:ARG:HG3	3:B:703:GOL:H31	1.78	0.64

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	276/304 (91%)	272 (99%)	4 (1%)	0	100	100
1	B	283/304 (93%)	280 (99%)	3 (1%)	0	100	100
2	C	115/127 (91%)	114 (99%)	1 (1%)	0	100	100
2	D	114/127 (90%)	113 (99%)	1 (1%)	0	100	100
All	All	788/862 (91%)	779 (99%)	9 (1%)	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	246/266 (92%)	244 (99%)	2 (1%)	81	72
1	B	251/266 (94%)	247 (98%)	4 (2%)	62	46
2	C	96/105 (91%)	92 (96%)	4 (4%)	30	10
2	D	95/105 (90%)	94 (99%)	1 (1%)	73	61
All	All	688/742 (93%)	677 (98%)	11 (2%)	62	46

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	487	GLU
1	B	490	SER
2	D	43	LYS
1	B	532	SER
2	C	53	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	288	GLN
1	A	507	GLN
1	B	426	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	39	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)

6 ligands are modelled in this entry.

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.

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	A	279/304 (91%)	-0.10	3 (1%)	80	83	23, 35, 56, 84	0
1	B	282/304 (92%)	-0.16	2 (0%)	87	89	21, 30, 48, 80	0
2	C	116/127 (91%)	-0.07	2 (1%)	70	74	26, 37, 51, 71	0
2	D	116/127 (91%)	-0.24	2 (1%)	70	74	21, 29, 43, 52	0
All	All	793/862 (91%)	-0.14	9 (1%)	80	83	21, 33, 51, 84	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	488	TYR	4.0
1	A	338	VAL	3.7
2	C	1	GLN	3.6
1	A	532	SER	2.6
2	D	1	GLN	2.5

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
3	GOL	D	301	6/6	0.77	0.32	55,67,79,79	0
4	PEG	B	704	7/7	0.82	0.33	39,49,69,71	0
3	GOL	B	703	6/6	0.83	0.13	46,56,70,85	0
3	GOL	B	701	6/6	0.87	0.09	33,44,59,62	0
3	GOL	A	601	6/6	0.95	0.10	36,47,68,82	0
3	GOL	B	702	6/6	0.96	0.14	26,42,66,80	0

6.5 Other polymers (i)

There are no such residues in this entry.