Full wwPDB X-ray Structure Validation Report ⓘ

Mar 6, 2024 – 09:11 pm GMT

PDB ID	:	8Q28
Title	:	Se-Met labelled TtX122A - A domain of unknown function from the Teredinibacter turnerae protein TERTU_3803
Authors	:	Rajagopal, B.S.; Hemsworth, G.R.
Deposited on	:	2023-08-01
Resolution	:	1.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

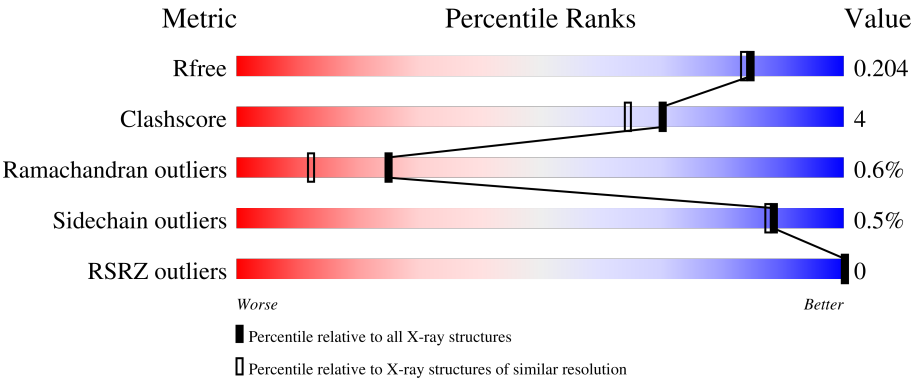
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	248	85% 11%
1	B	248	86% 10%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7670 atoms, of which 3492 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 Gluconolactonase domain protein.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	A	241	Total	C	H	N	O	S	Se	53	1	0
			3621	1175	1736	334	369	2	5			
1	B	240	Total	C	H	N	O	S	Se	52	2	0
			3612	1173	1732	331	369	2	5			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	241	LEU	-	expression tag	UNP C5BSV8
A	242	GLU	-	expression tag	UNP C5BSV8
A	243	HIS	-	expression tag	UNP C5BSV8
A	244	HIS	-	expression tag	UNP C5BSV8
A	245	HIS	-	expression tag	UNP C5BSV8
A	246	HIS	-	expression tag	UNP C5BSV8
A	247	HIS	-	expression tag	UNP C5BSV8
A	248	HIS	-	expression tag	UNP C5BSV8
B	241	LEU	-	expression tag	UNP C5BSV8
B	242	GLU	-	expression tag	UNP C5BSV8
B	243	HIS	-	expression tag	UNP C5BSV8
B	244	HIS	-	expression tag	UNP C5BSV8
B	245	HIS	-	expression tag	UNP C5BSV8
B	246	HIS	-	expression tag	UNP C5BSV8
B	247	HIS	-	expression tag	UNP C5BSV8
B	248	HIS	-	expression tag	UNP C5BSV8

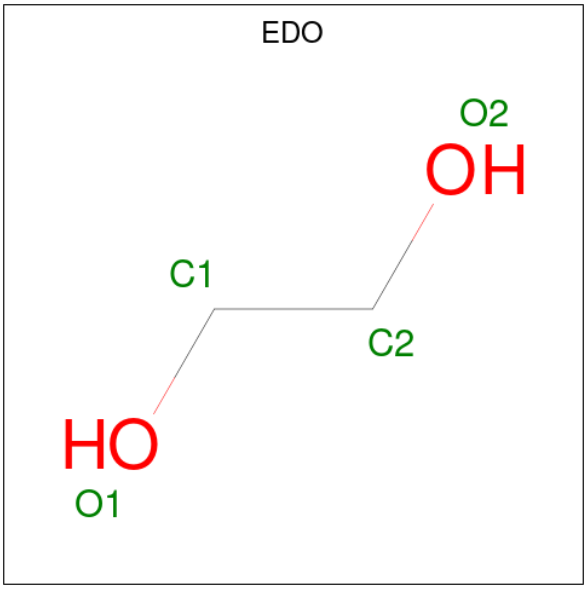
- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ca	0	0
			1	1		
2	B	1	Total	Ca	0	0
			1	1		

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Mg 2 2	0	0
3	B	2	Total Mg 2 2	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C H O 10 2 6 2	1	0
4	A	1	Total C H O 10 2 6 2	1	0
4	B	1	Total C H O 10 2 6 2	1	0
4	B	1	Total C H O 10 2 6 2	1	0

- Molecule 5 is water.

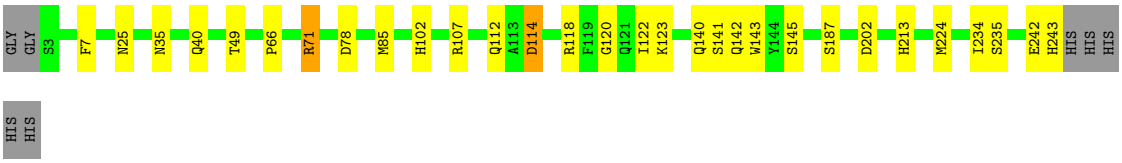
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	165	Total O 165 165	0	0
5	B	226	Total O 226 226	0	0

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.

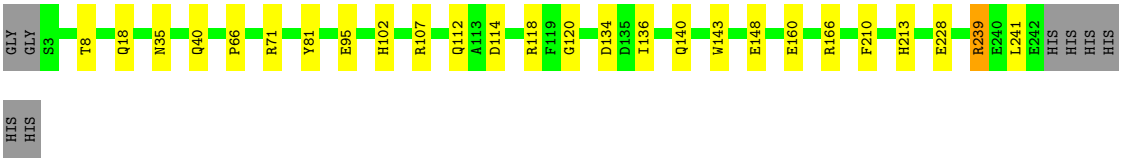
• Molecule 1: Gluconolactonase domain protein

Chain A:



• Molecule 1: Gluconolactonase domain protein

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	48.00Å 75.44Å 69.75Å 90.00° 107.62° 90.00°	Depositor
Resolution (Å)	49.88 – 1.80 49.88 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.88-1.80) 98.9 (49.88-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.22 (at 1.79Å)	Xtriage
Refinement program	REFMAC 5.8.0419	Depositor
R, R_{free}	0.157 , 0.202 0.158 , 0.204	Depositor DCC
R_{free} test set	2143 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	15.7	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 45.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.011 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7670	wwPDB-VP
Average B, all atoms (Å^2)	22.0	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 34.78 % 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 6.4712e-04. 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

5.1 Standard geometry

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

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	0.77	2/1935 (0.1%)	1.10	4/2627 (0.2%)
1	B	0.85	3/1932 (0.2%)	1.13	7/2623 (0.3%)
All	All	0.81	5/3867 (0.1%)	1.11	11/5250 (0.2%)

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	2
All	All	0	3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	148	GLU	CD-OE2	-7.24	1.17	1.25
1	B	148	GLU	CD-OE1	-6.63	1.18	1.25
1	A	242	GLU	CD-OE2	-5.79	1.19	1.25
1	A	242	GLU	CD-OE1	5.64	1.31	1.25
1	B	95	GLU	CD-OE2	-5.24	1.19	1.25

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	243	HIS	CA-C-O	-12.88	93.05	120.10
1	B	71	ARG	CG-CD-NE	-7.29	96.50	111.80
1	B	107	ARG	NE-CZ-NH1	-7.16	116.72	120.30
1	A	107	ARG	NE-CZ-NH1	-6.40	117.10	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	134	ASP	CB-CG-OD1	6.22	123.89	118.30
1	B	239	ARG	CG-CD-NE	-5.62	99.99	111.80
1	A	114	ASP	N-CA-CB	-5.59	100.54	110.60
1	B	166	ARG	NE-CZ-NH1	5.52	123.06	120.30
1	B	210	PHE	CB-CG-CD1	5.37	124.56	120.80
1	A	49	THR	CA-CB-OG1	5.22	119.97	109.00
1	B	239	ARG	NE-CZ-NH2	-5.20	117.70	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	71	ARG	Sidechain
1	B	239	ARG	Sidechain
1	B	241	LEU	Peptide

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	1885	1736	1728	15	0
1	B	1880	1732	1724	8	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	8	12	12	1	0
4	B	8	12	12	3	0
5	A	165	0	0	2	1
5	B	226	0	0	4	1
All	All	4178	3492	3476	26	1

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.

All (26) 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)
4:B:304:EDO:H22	5:B:600:HOH:O	1.88	0.74
4:B:304:EDO:C2	5:B:600:HOH:O	2.38	0.71
1:B:18[B]:GLN:OE1	5:B:401:HOH:O	2.09	0.70
1:A:142:GLN:O	1:A:145:SER:OG	2.16	0.62
1:A:202:ASP:HA	4:A:304:EDO:O1	2.02	0.58
1:B:112:GLN:HB3	1:B:114:ASP:HB2	1.86	0.57
1:A:112:GLN:OE1	1:A:114:ASP:HB2	2.07	0.54
1:A:140:GLN:HA	1:A:143:TRP:CE2	2.42	0.54
1:A:112:GLN:HB3	1:A:114:ASP:HB2	1.89	0.53
1:B:102:HIS:CG	1:B:118:ARG:HB3	2.46	0.51
1:A:78:ASP:OD2	5:A:401:HOH:O	2.19	0.51
4:B:304:EDO:H21	5:B:600:HOH:O	2.07	0.50
1:A:122:ILE:O	1:A:123:LYS:C	2.52	0.48
1:A:187:SER:OG	5:A:402:HOH:O	2.20	0.47
1:A:66:PRO:HB3	1:A:213:HIS:CG	2.51	0.46
1:B:140:GLN:HA	1:B:143:TRP:CE2	2.51	0.45
1:B:66:PRO:HB3	1:B:213:HIS:CG	2.51	0.45
1:A:140:GLN:HA	1:A:143:TRP:CD2	2.52	0.45
1:B:8:THR:HG22	1:B:228:GLU:HG2	2.00	0.43
1:B:35:ASN:O	1:B:40:GLN:HG3	2.19	0.42
1:A:7:PHE:CZ	1:A:71:ARG:NH2	2.87	0.42
1:B:81:TYR:CE2	1:B:160:GLU:HG3	2.55	0.41
1:A:35:ASN:O	1:A:40:GLN:HG3	2.20	0.41
1:A:85:MSE:O	1:A:224:MSE:HA	2.20	0.41
1:A:234:ILE:HG22	1:A:235:SER:N	2.36	0.41
1:A:102:HIS:CG	1:A:118:ARG:HB3	2.56	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:555:HOH:O	5:B:609:HOH:O[1_554]	2.06	0.14

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	240/248 (97%)	229 (95%)	9 (4%)	2 (1%)	19 7
1	B	240/248 (97%)	229 (95%)	10 (4%)	1 (0%)	34 21
All	All	480/496 (97%)	458 (95%)	19 (4%)	3 (1%)	25 12

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	141	SER
1	A	120	GLY
1	B	120	GLY

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	202/201 (100%)	201 (100%)	1 (0%)	88 87
1	B	201/201 (100%)	200 (100%)	1 (0%)	88 87
All	All	403/402 (100%)	401 (100%)	2 (0%)	88 87

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

Mol	Chain	Res	Type
1	A	25	ASN
1	B	136	ILE

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

Mol	Chain	Res	Type
1	B	55	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 6 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
4	EDO	B	304	-	3,3,3	0.51	0	2,2,2	0.71	0
4	EDO	A	305	-	3,3,3	0.63	0	2,2,2	0.46	0
4	EDO	B	305	-	3,3,3	0.64	0	2,2,2	1.08	0
4	EDO	A	304	-	3,3,3	0.31	0	2,2,2	0.45	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
4	EDO	B	304	-	-	1/1/1/1	-
4	EDO	A	305	-	-	1/1/1/1	-
4	EDO	B	305	-	-	1/1/1/1	-
4	EDO	A	304	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	305	EDO	O1-C1-C2-O2
4	B	305	EDO	O1-C1-C2-O2
4	A	304	EDO	O1-C1-C2-O2
4	B	304	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	304	EDO	3	0
4	A	304	EDO	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	236/248 (95%)	-0.51	0	100	100	14, 24, 40, 53	0
1	B	235/248 (94%)	-0.67	0	100	100	11, 17, 28, 47	0
All	All	471/496 (94%)	-0.59	0	100	100	11, 20, 36, 53	0

There are no RSRZ outliers to report.

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
4	EDO	A	305	4/4	0.94	0.13	28,34,37,37	1
4	EDO	B	305	4/4	0.95	0.10	21,27,28,28	1
4	EDO	B	304	4/4	0.96	0.07	30,41,45,45	1
4	EDO	A	304	4/4	0.96	0.07	28,31,37,37	1
3	MG	B	302	1/1	0.97	0.05	20,20,20,20	0
3	MG	A	303	1/1	0.99	0.05	20,20,20,20	0
2	CA	A	301	1/1	0.99	0.06	21,21,21,21	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	MG	A	302	1/1	0.99	0.09	18,18,18,18	0
2	CA	B	301	1/1	1.00	0.11	17,17,17,17	0
3	MG	B	303	1/1	1.00	0.04	18,18,18,18	0

6.5 Other polymers (i)

There are no such residues in this entry.