Full wwPDB X-ray Structure Validation Report (i)

Jan 7, 2024 – 07:05 pm GMT

PDB ID	:	6HBM
Title	:	Crystal structure of MSMEG_1712 from Mycobacterium smegmatis in complex with alpha-L-arabinofuranose
Authors	:	Li, M.; Mueller, C.; Einsle, O.; Jessen-Trefzer, C.
Deposited on	:	2018-08-10
Resolution	:	2.76 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	:	1.8.4, 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 ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.76 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.
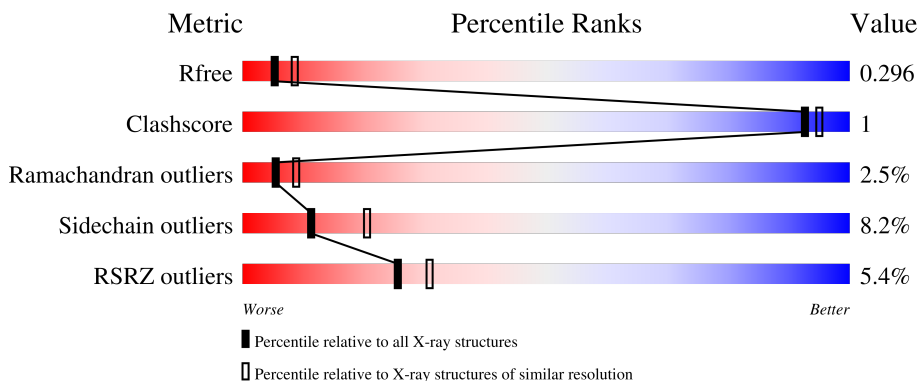
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	320	85% 10% 5%
1	B	320	82% 12% • 5%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 9093 atoms, of which 4482 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 ABC transporter periplasmic-binding protein YtfQ.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	304	Total 4510	C 1426	H 2230	N 398	O 450	S 6	386	0	0
1	B	304	Total 4516	C 1427	H 2233	N 398	O 452	S 6	634	0	0

There are 28 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP A0QT50
A	308	LYS	-	expression tag	UNP A0QT50
A	309	LEU	-	expression tag	UNP A0QT50
A	310	ALA	-	expression tag	UNP A0QT50
A	311	ALA	-	expression tag	UNP A0QT50
A	312	ALA	-	expression tag	UNP A0QT50
A	313	LEU	-	expression tag	UNP A0QT50
A	314	GLU	-	expression tag	UNP A0QT50
A	315	HIS	-	expression tag	UNP A0QT50
A	316	HIS	-	expression tag	UNP A0QT50
A	317	HIS	-	expression tag	UNP A0QT50
A	318	HIS	-	expression tag	UNP A0QT50
A	319	HIS	-	expression tag	UNP A0QT50
A	320	HIS	-	expression tag	UNP A0QT50
B	1	MET	-	initiating methionine	UNP A0QT50
B	308	LYS	-	expression tag	UNP A0QT50
B	309	LEU	-	expression tag	UNP A0QT50
B	310	ALA	-	expression tag	UNP A0QT50
B	311	ALA	-	expression tag	UNP A0QT50
B	312	ALA	-	expression tag	UNP A0QT50
B	313	LEU	-	expression tag	UNP A0QT50
B	314	GLU	-	expression tag	UNP A0QT50
B	315	HIS	-	expression tag	UNP A0QT50
B	316	HIS	-	expression tag	UNP A0QT50
B	317	HIS	-	expression tag	UNP A0QT50

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	318	HIS	-	expression tag	UNP A0QT50
B	319	HIS	-	expression tag	UNP A0QT50
B	320	HIS	-	expression tag	UNP A0QT50

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	7	Total 7	Zn 7	0	0
2	B	5	Total 5	Zn 5	0	0

- Molecule 3 is alpha-L-arabinofuranose (three-letter code: AHR) (formula: $C_5H_{10}O_5$).
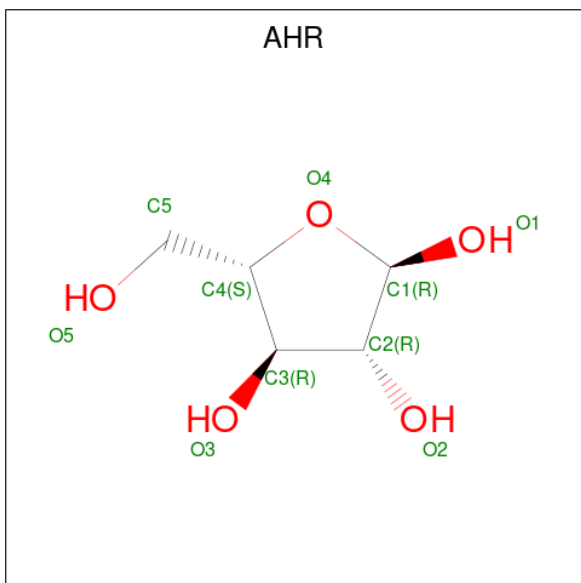
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 19	C 5	H 9	O 5	0	0
3	B	1	Total 20	C 5	H 10	O 5	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	8	Total 8	O 8	0	0

Continued on next page...

Continued from previous page...

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	8	Total 8	O 8	0	0

3 Residue-property plots ⓘ

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: ABC transporter periplasmic-binding protein YtfQ
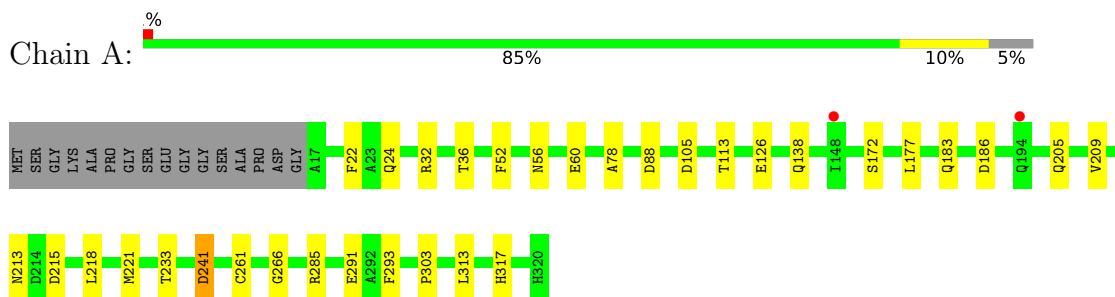
• Molecule 1: ABC transporter periplasmic-binding protein YtfQ
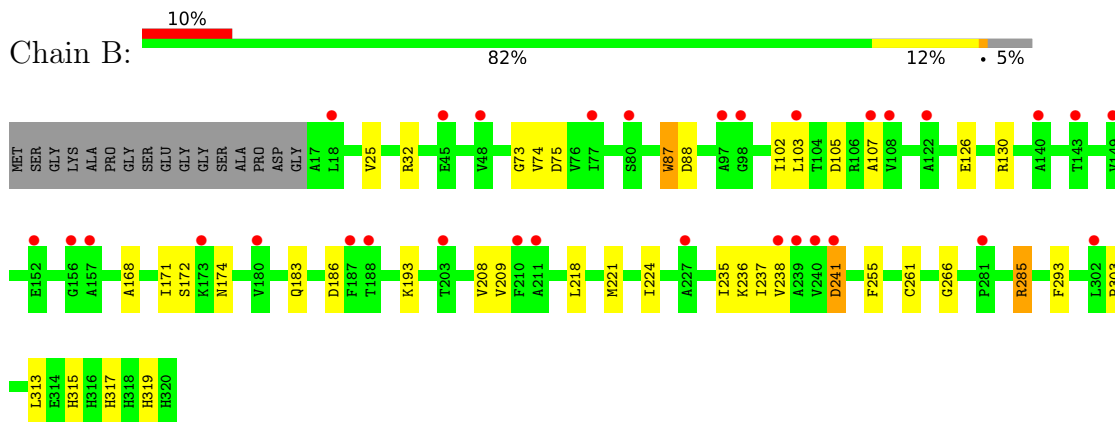
4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	119.50Å 119.50Å 234.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.95 – 2.76 28.95 – 2.76	Depositor EDS
% Data completeness (in resolution range)	50.5 (28.95-2.76) 50.5 (28.95-2.76)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.11 (at 2.76Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.218 , 0.279 0.239 , 0.296	Depositor DCC
R_{free} test set	522 reflections (4.63%)	wwPDB-VP
Wilson B-factor (Å^2)	91.6	Xtriage
Anisotropy	0.097	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 105.6	EDS
L-test for twinning [2]	$< |L| > = 0.46, < L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	9093	wwPDB-VP
Average B, all atoms (Å^2)	96.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.04% 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: AHR, ZN

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	0.51	0/2321	0.68	0/3148
1	B	0.50	0/2324	0.66	0/3152
All	All	0.50	0/4645	0.67	0/6300

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 ⓘ

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	2280	2230	2233	6	0
1	B	2283	2233	2234	8	0
2	A	7	0	0	0	0
2	B	5	0	0	0	0
3	A	10	9	0	0	0
3	B	10	10	0	0	0
4	A	8	0	0	0	0
4	B	8	0	0	0	0
All	All	4611	4482	4467	12	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 (12) 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:103:LEU:HD21	1:B:107:ALA:HA	1.87	0.55
1:A:22:PHE:HA	1:A:78:ALA:O	2.09	0.51
1:B:237:ILE:HG23	1:B:255:PHE:HA	1.97	0.45
1:A:241:ASP:HA	1:A:261:CYS:HB3	1.98	0.45
1:A:291:GLU:OE1	1:B:315:HIS:CE1	2.70	0.44
1:A:36:THR:HG23	1:A:52:PHE:CZ	2.53	0.44
1:A:317:HIS:CE1	1:B:317:HIS:CE1	3.05	0.44
1:B:241:ASP:HA	1:B:261:CYS:HB3	1.99	0.43
1:A:218:LEU:HA	1:A:221:MET:HG2	2.02	0.42
1:B:218:LEU:HA	1:B:221:MET:HG2	2.02	0.40
1:B:285:ARG:HB3	1:B:319:HIS:HB2	2.03	0.40
1:B:168:ALA:O	1:B:171:ILE:HG13	2.21	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	302/320 (94%)	274 (91%)	23 (8%)	5 (2%)	9	16
1	B	302/320 (94%)	270 (89%)	22 (7%)	10 (3%)	4	5
All	All	604/640 (94%)	544 (90%)	45 (8%)	15 (2%)	5	9

All (15) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	105	ASP
1	A	241	ASP
1	A	303	PRO
1	B	105	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	241	ASP
1	B	303	PRO
1	B	25	VAL
1	B	73	GLY
1	B	74	VAL
1	B	87	TRP
1	B	236	LYS
1	A	172	SER
1	B	172	SER
1	A	266	GLY
1	B	266	GLY

5.3.2   Protein sidechains ⓘ

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	232/242 (96%)	213 (92%)	19 (8%)	11	20
1	B	233/242 (96%)	214 (92%)	19 (8%)	11	20
All	All	465/484 (96%)	427 (92%)	38 (8%)	11	20

All (38) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	24	GLN
1	A	32	ARG
1	A	56	ASN
1	A	60	GLU
1	A	88	ASP
1	A	113	THR
1	A	126	GLU
1	A	138	GLN
1	A	177	LEU
1	A	183	GLN
1	A	186	ASP
1	A	205	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	209	VAL
1	A	213	ASN
1	A	215	ASP
1	A	233	THR
1	A	285	ARG
1	A	293	PHE
1	A	313	LEU
1	B	32	ARG
1	B	75	ASP
1	B	87	TRP
1	B	88	ASP
1	B	102	ILE
1	B	126	GLU
1	B	130	ARG
1	B	174	ASN
1	B	183	GLN
1	B	186	ASP
1	B	193	LYS
1	B	208	VAL
1	B	209	VAL
1	B	224	ILE
1	B	235	ILE
1	B	238	VAL
1	B	285	ARG
1	B	293	PHE
1	B	313	LEU

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

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6   Ligand geometry ⓘ

Of 14 ligands modelled in this entry, 12 are monoatomic - leaving 2 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	#$|Z| > 2$
3	AHR	B	406	-	10,10,10	0.58	0	13,14,14	0.62	0
3	AHR	A	408	-	10,10,10	0.79	0	13,14,14	1.04	1 (7%)

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
3	AHR	B	406	-	-	2/2/18/18	0/1/1/1
3	AHR	A	408	-	-	0/2/18/18	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	408	AHR	O2-C2-C1	2.08	117.55	111.82

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	406	AHR	C3-C4-C5-O5
3	B	406	AHR	O4-C4-C5-O5

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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	304/320 (95%)	0.02	2 (0%)	87	91	45, 79, 109, 124	37 (12%)
1	B	304/320 (95%)	0.38	31 (10%)	6	7	60, 116, 141, 152	59 (19%)
All	All	608/640 (95%)	0.20	33 (5%)	25	31	45, 93, 137, 152	96 (15%)

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	45	GLU	3.5
1	B	107	ALA	3.3
1	B	239	ALA	3.2
1	B	108	VAL	3.0
1	B	281	PRO	3.0
1	B	203	THR	3.0
1	B	48	VAL	3.0
1	B	80	SER	2.9
1	B	156	GLY	2.8
1	B	149	VAL	2.8
1	B	18	LEU	2.8
1	B	241	ASP	2.8
1	B	173	LYS	2.7
1	B	97	ALA	2.7
1	B	187	PHE	2.5
1	B	103	LEU	2.5
1	B	240	VAL	2.5
1	B	227	ALA	2.4
1	B	211	ALA	2.4
1	B	180	VAL	2.3
1	B	302	LEU	2.3
1	B	98	GLY	2.2
1	B	122	ALA	2.2
1	B	143	THR	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	77	ILE	2.1
1	B	140	ALA	2.1
1	A	148	ILE	2.1
1	A	194	GLN	2.0
1	B	188	THR	2.0
1	B	210	PHE	2.0
1	B	152	GLU	2.0
1	B	157	ALA	2.0
1	B	238	VAL	2.0

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, 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(Å^2)	Q<0.9
2	ZN	A	406	1/1	0.91	0.17	134,134,134,134	0
3	AHR	B	406	10/10	0.94	0.15	19,19,26,31	20
2	ZN	A	401	1/1	0.98	0.20	56,56,56,56	0
2	ZN	B	403	1/1	0.98	0.20	89,89,89,89	0
2	ZN	B	405	1/1	0.98	0.23	98,98,98,98	0
3	AHR	A	408	10/10	0.98	0.15	19,19,24,27	0
2	ZN	A	403	1/1	0.98	0.23	72,72,72,72	0
2	ZN	B	402	1/1	0.99	0.14	58,58,58,58	0
2	ZN	A	404	1/1	0.99	0.14	80,80,80,80	0
2	ZN	B	404	1/1	0.99	0.21	89,89,89,89	0
2	ZN	A	405	1/1	0.99	0.13	67,67,67,67	0
2	ZN	A	402	1/1	0.99	0.17	63,63,63,63	0
2	ZN	A	407	1/1	0.99	0.11	78,78,78,78	0
2	ZN	B	401	1/1	1.00	0.17	77,77,77,77	0

6.5 Other polymers (i)

There are no such residues in this entry.