



Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2017 – 11:41 pm GMT

PDB ID : 3OSG
Title : The structure of protozoan parasite *Trichomonas vaginalis* Myb2 in complex with MRE-1-12 DNA
Authors : Jiang, I.; Tsai, C.K.; Chen, S.C.; Wang, S.H.; Amiraslanov, I.; Chang, C.F.; Wu, W.J.; Tai, J.H.; Liaw, Y.C.; Huang, T.H.
Deposited on : 2010-09-09
Resolution : 2.00 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

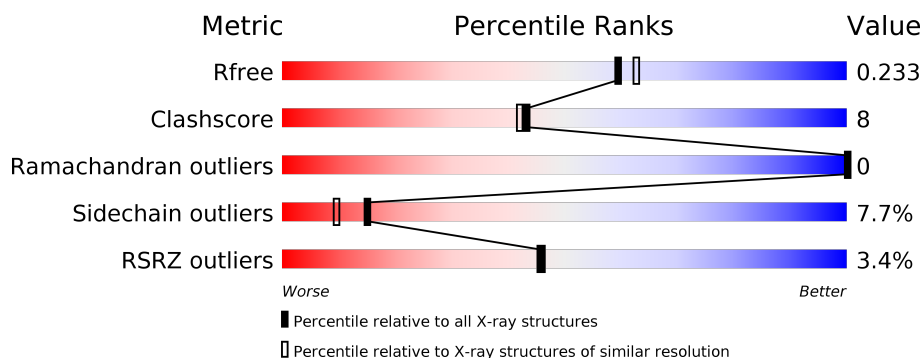
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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}	100719	6609 (2.00-2.00)
Clashscore	112137	7775 (2.00-2.00)
Ramachandran outliers	110173	7679 (2.00-2.00)
Sidechain outliers	110143	7678 (2.00-2.00)
RSRZ outliers	101464	6696 (2.00-2.00)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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 $\leq 5\%$. 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	126	<div> <div>3%</div> <div> <div></div> <div>70%</div> <div>13%</div> <div>• •</div> <div>13%</div> </div> </div>
1	D	126	<div> <div>4%</div> <div> <div></div> <div>63%</div> <div>16%</div> <div>•</div> <div>18%</div> </div> </div>
2	B	12	<div> <div></div> <div> <div>42%</div> <div>50%</div> <div>8%</div> </div> </div>
2	E	12	<div> <div></div> <div> <div>58%</div> <div>25%</div> <div>17%</div> </div> </div>
3	C	12	<div> <div></div> <div> <div>50%</div> <div>42%</div> <div>8%</div> </div> </div>
3	F	12	<div> <div></div> <div> <div>58%</div> <div>25%</div> <div>17%</div> </div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3034 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 MYB21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	110	Total	C	N	O	S	0	0	0
			906	577	169	157	3			
1	D	103	Total	C	N	O	S	0	0	0
			856	546	159	148	3			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	39	MET	-	EXPRESSION TAG	UNP Q58HP3
A	157	LEU	-	EXPRESSION TAG	UNP Q58HP3
A	158	GLU	-	EXPRESSION TAG	UNP Q58HP3
A	159	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	160	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	161	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	162	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	163	HIS	-	EXPRESSION TAG	UNP Q58HP3
A	164	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	39	MET	-	EXPRESSION TAG	UNP Q58HP3
D	157	LEU	-	EXPRESSION TAG	UNP Q58HP3
D	158	GLU	-	EXPRESSION TAG	UNP Q58HP3
D	159	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	160	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	161	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	162	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	163	HIS	-	EXPRESSION TAG	UNP Q58HP3
D	164	HIS	-	EXPRESSION TAG	UNP Q58HP3

- Molecule 2 is a DNA chain called 5'-D(*AP*AP*AP*TP*AP*TP*CP*GP*TP*TP*AP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	12	Total	C	N	O	P	0	0	0
			243	119	43	70	11			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	12	Total 243	C 119	N 43	O 70	P 11	0	0	0

- Molecule 3 is a DNA chain called 5'-D(*AP*TP*AP*AP*CP*GP*AP*TP*AP*TP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total 243	C 119	N 43	O 70	P 11	0	0	0
3	F	12	Total 243	C 119	N 43	O 70	P 11	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	83	Total 83	O 83	0	0
4	B	44	Total 44	O 44	0	0
4	C	42	Total 42	O 42	0	0
4	D	82	Total 82	O 82	0	0
4	E	24	Total 24	O 24	0	0
4	F	25	Total 25	O 25	0	0

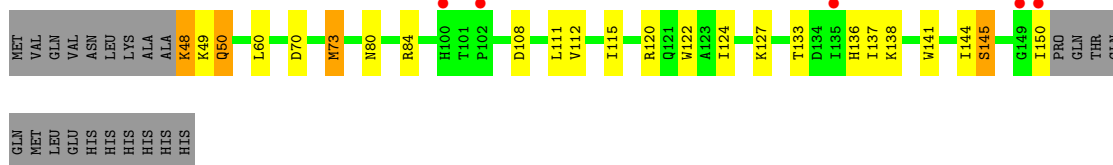
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: MYB21



- Molecule 1: MYB21



- Molecule 2: 5'-D(*AP*AP*AP*TP*AP*TP*CP*GP*TP*TP*AP*T)-3'



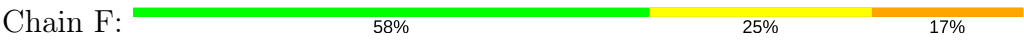
- Molecule 2: 5'-D(*AP*AP*AP*TP*AP*TP*CP*GP*TP*TP*AP*T)-3'



- Molecule 3: 5'-D(*AP*TP*AP*AP*CP*GP*AP*TP*AP*TP*TP*T)-3'



- Molecule 3: 5'-D(*AP*TP*AP*AP*CP*GP*AP*TP*AP*TP*TP*T)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	40.12Å 127.53Å 41.47Å 90.00° 100.23° 90.00°	Depositor
Resolution (Å)	24.19 – 2.00 24.19 – 2.00	Depositor EDS
% Data completeness (in resolution range)	93.8 (24.19-2.00) 93.9 (24.19-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.59 (at 1.99Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6_289)	Depositor
R, R_{free}	0.191 , 0.233 0.185 , 0.233	Depositor DCC
R_{free} test set	1293 reflections (4.97%)	DCC
Wilson B-factor (Å ²)	28.5	Xtriage
Anisotropy	0.531	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 50.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.041 for l,-k,h	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3034	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.08% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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

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	0.41	0/928	0.58	0/1252
1	D	0.41	0/879	0.56	0/1187
2	B	0.83	0/272	1.61	6/418 (1.4%)
2	E	0.69	0/272	1.48	7/418 (1.7%)
3	C	0.84	0/272	1.60	7/418 (1.7%)
3	F	0.74	0/272	1.61	7/418 (1.7%)
All	All	0.58	0/2895	1.10	27/4111 (0.7%)

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

There are no bond length outliers.

All (27) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	10	DT	C1'-O4'-C4'	-7.38	102.72	110.10
2	B	10	DT	O4'-C1'-N1	-7.09	103.03	108.00
2	B	3	DA	O4'-C1'-N9	-6.95	103.14	108.00
3	C	9	DA	O4'-C1'-N9	-6.78	103.26	108.00
3	F	2	DT	O4'-C1'-C2'	-6.16	100.97	105.90
2	B	6	DT	O4'-C1'-N1	-6.00	103.80	108.00
2	E	6	DT	N3-C4-O4	5.94	123.47	119.90
3	F	9	DA	P-O3'-C3'	5.94	126.83	119.70
3	F	8	DT	O4'-C1'-N1	-5.84	103.91	108.00
3	F	5	DC	O4'-C1'-N1	5.82	112.08	108.00
3	F	2	DT	C3'-C2'-C1'	-5.76	95.58	102.50
3	F	2	DT	C1'-O4'-C4'	-5.75	104.35	110.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	5	DC	P-O3'-C3'	5.65	126.48	119.70
2	E	10	DT	C5-C4-O4	-5.57	121.00	124.90
3	C	7	DA	P-O3'-C3'	5.51	126.31	119.70
2	E	6	DT	C5-C4-O4	-5.50	121.05	124.90
3	C	3	DA	O4'-C1'-N9	-5.49	104.16	108.00
2	E	12	DT	O4'-C1'-N1	5.37	111.76	108.00
2	B	2	DA	O4'-C1'-N9	-5.29	104.29	108.00
3	C	6	DG	O4'-C1'-N9	-5.29	104.30	108.00
2	B	12	DT	C5-C4-O4	-5.20	121.26	124.90
2	E	1	DA	O4'-C4'-C3'	-5.15	102.44	104.50
3	C	2	DT	C3'-C2'-C1'	-5.13	96.34	102.50
3	C	7	DA	C1'-O4'-C4'	-5.12	104.98	110.10
2	B	8	DG	O4'-C1'-N9	-5.12	104.42	108.00
2	E	10	DT	O4'-C1'-N1	-5.07	104.45	108.00
2	E	10	DT	N3-C4-O4	5.07	122.94	119.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	40	VAL	Peptide

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	906	0	909	18	0
1	D	856	0	855	18	0
2	B	243	0	139	2	0
2	E	243	0	139	2	0
3	C	243	0	139	1	0
3	F	243	0	139	1	0
4	A	83	0	0	4	0
4	B	44	0	0	0	0
4	C	42	0	0	0	0
4	D	82	0	0	0	0
4	E	24	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	F	25	0	0	0	0
All	All	3034	0	2320	39	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 8.

All (39) 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:A:129:PHE:HB3	1:A:132:ARG:HG3	1.55	0.87
1:D:48:LYS:HB3	1:D:48:LYS:NZ	1.95	0.81
1:A:42:VAL:O	1:A:43:ASN:HB2	1.93	0.68
3:F:8:DT:H2''	3:F:9:DA:C8	2.30	0.67
1:D:70:ASP:OD2	1:D:73:MET:HE1	1.96	0.66
1:D:48:LYS:HZ3	1:D:48:LYS:HB3	1.60	0.66
1:A:78:PHE:O	4:A:234:HOH:O	2.15	0.64
1:A:109:ALA:O	1:A:112:VAL:HG22	2.02	0.60
1:A:133:THR:H	1:A:136:HIS:HD2	1.52	0.56
1:A:133:THR:H	1:A:136:HIS:CD2	2.24	0.55
1:D:141:TRP:O	1:D:145:SER:HB2	2.08	0.54
1:D:111:LEU:HD23	1:D:144:ILE:HD12	1.90	0.52
1:D:150:ILE:HG22	1:D:150:ILE:O	2.10	0.52
1:D:124:ILE:HA	1:D:127:LYS:HD2	1.91	0.52
1:D:120:ARG:CZ	1:D:122:TRP:HZ2	2.23	0.51
1:A:107:GLU:OE2	1:A:132:ARG:HD2	2.09	0.51
1:A:43:ASN:N	1:A:45:LYS:O	2.41	0.51
1:D:49:LYS:HE3	2:E:6:DT:O2	2.10	0.51
1:D:133:THR:H	1:D:136:HIS:CD2	2.29	0.50
1:A:51:LYS:HD3	2:B:6:DT:H5'	1.95	0.49
1:D:108:ASP:OD2	1:D:144:ILE:HD11	2.14	0.47
1:A:75:ALA:O	4:A:234:HOH:O	2.21	0.46
1:A:45:LYS:HD2	1:A:45:LYS:HA	1.68	0.45
1:D:49:LYS:HG2	4:E:26:HOH:O	2.16	0.45
1:A:150:ILE:O	1:A:150:ILE:HG22	2.17	0.45
2:E:11:DA:H1'	2:E:12:DT:H5''	1.98	0.45
1:D:48:LYS:HZ2	1:D:48:LYS:HB3	1.75	0.45
1:A:40:VAL:HG22	1:A:51:LYS:O	2.18	0.44
2:B:11:DA:C2	3:C:3:DA:C2	3.06	0.44
1:A:43:ASN:OD1	4:A:227:HOH:O	2.21	0.44
1:A:112:VAL:O	1:A:116:GLN:HG3	2.18	0.43
1:D:112:VAL:O	1:D:115:ILE:HG22	2.18	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:60:LEU:HD13	1:A:94:LEU:HD12	2.02	0.41
1:D:137:ILE:HG22	1:D:138:LYS:N	2.36	0.41
1:D:50:GLN:NE2	1:D:80:ASN:OD1	2.53	0.41
1:D:48:LYS:CB	1:D:48:LYS:NZ	2.75	0.41
1:D:120:ARG:NH2	1:D:122:TRP:HZ2	2.19	0.41
1:A:91:LYS:HD3	4:A:197:HOH:O	2.21	0.40
1:A:113:GLN:HA	1:A:116:GLN:HE21	1.86	0.40

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	106/126 (84%)	102 (96%)	4 (4%)	0	100	100
1	D	101/126 (80%)	97 (96%)	4 (4%)	0	100	100
All	All	207/252 (82%)	199 (96%)	8 (4%)	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	94/110 (86%)	86 (92%)	8 (8%)	12	7

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	89/110 (81%)	83 (93%)	6 (7%)	19	13
All	All	183/220 (83%)	169 (92%)	14 (8%)	15	9

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	40	VAL
1	A	41	GLN
1	A	42	VAL
1	A	43	ASN
1	A	45	LYS
1	A	60	LEU
1	A	84	ARG
1	A	132	ARG
1	D	48	LYS
1	D	50	GLN
1	D	60	LEU
1	D	73	MET
1	D	84	ARG
1	D	145	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	43	ASN
1	A	66	GLN
1	A	113	GLN
1	A	116	GLN
1	A	136	HIS
1	D	50	GLN
1	D	67	HIS
1	D	113	GLN
1	D	136	HIS

5.3.3 RNA ⓘ

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	110/126 (87%)	0.37	4 (3%) 43 44	21, 34, 47, 54	0
1	D	103/126 (81%)	0.38	5 (4%) 30 30	19, 34, 48, 61	0
2	B	12/12 (100%)	-0.45	0 100 100	20, 26, 32, 33	0
2	E	12/12 (100%)	-0.20	0 100 100	25, 35, 40, 41	0
3	C	12/12 (100%)	-0.64	0 100 100	19, 24, 29, 30	0
3	F	12/12 (100%)	-0.42	0 100 100	31, 34, 38, 39	0
All	All	261/300 (87%)	0.22	9 (3%) 46 46	19, 34, 47, 61	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	150	ILE	5.2
1	A	135	ILE	4.3
1	A	150	ILE	4.0
1	D	149	GLY	3.6
1	D	135	ILE	3.2
1	A	40	VAL	2.5
1	D	100	HIS	2.3
1	A	128	PHE	2.2
1	D	102	PRO	2.2

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.