



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 12:57 am BST

PDB ID : 2P9J
Title : Crystal structure of AQ2171 from Aquifex aeolicus
Authors : Yang, H.; Chen, L.; Agari, Y.; Ebihara, A.; Shinkai, A.; Kuramitsu, S.; Yokoyama, S.; Rose, J.P.; Wang, B.-C.; Southeast Collaboratory for Structural Genomics (SECSG); RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on : 2007-03-26
Resolution : 2.40 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

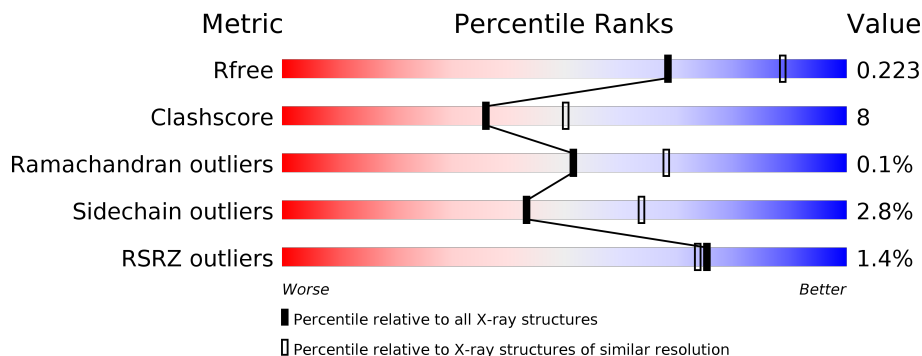
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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.



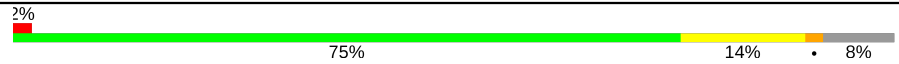

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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 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 $\leq 5\%$. 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	162	 77% 19%
1	B	162	 86% 13%
1	C	162	 81% 19%
1	D	162	 85% 14%
1	E	162	 71% 23% 6%
1	F	162	 75% 18% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	162	 <p>2% 75% 14% • 8%</p>
1	H	162	 <p>2% 83% 13% ••</p>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 10046 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 Hypothetical protein AQ2171.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	158	1241	800	208	230	3	4	0	0
1	B	162	1282	825	214	240	3	0	0	0
1	C	162	1279	824	214	238	3	0	0	0
1	D	161	1274	821	213	237	3	0	0	0
1	E	153	1187	769	195	220	3	0	0	0
1	F	151	1179	769	192	215	3	0	0	0
1	G	149	1167	755	194	215	3	0	0	0
1	H	156	1219	789	204	223	3	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
A	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
A	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
B	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
B	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
B	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
C	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
C	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
C	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
D	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
D	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
D	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
E	14	MSE	MET	MODIFIED RESIDUE	UNP O67920

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
E	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
F	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
F	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
F	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
G	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
G	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
G	116	MSE	MET	MODIFIED RESIDUE	UNP O67920
H	14	MSE	MET	MODIFIED RESIDUE	UNP O67920
H	51	MSE	MET	MODIFIED RESIDUE	UNP O67920
H	116	MSE	MET	MODIFIED RESIDUE	UNP O67920

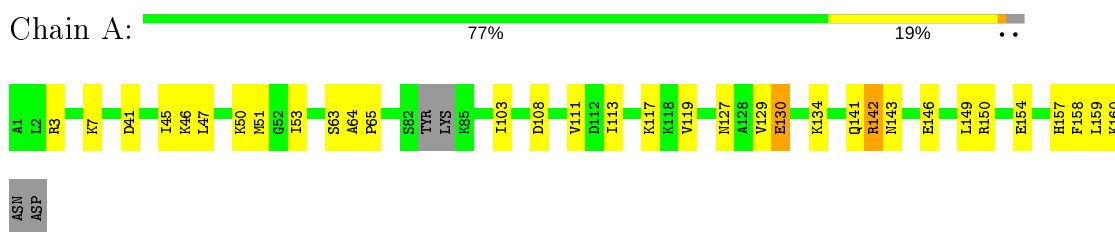
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	28	Total O 28 28	0	0
2	B	22	Total O 22 22	0	0
2	C	34	Total O 34 34	0	0
2	D	30	Total O 30 30	0	0
2	E	29	Total O 29 29	0	0
2	F	26	Total O 26 26	0	0
2	G	28	Total O 28 28	0	0
2	H	21	Total O 21 21	0	0

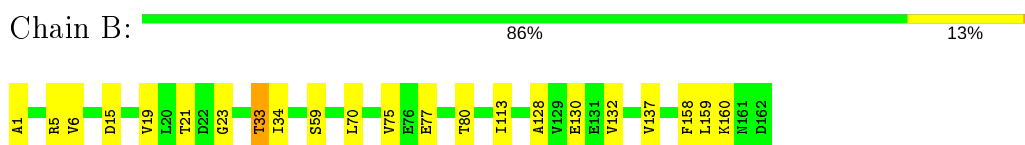
3 Residue-property plots [i](#)

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.

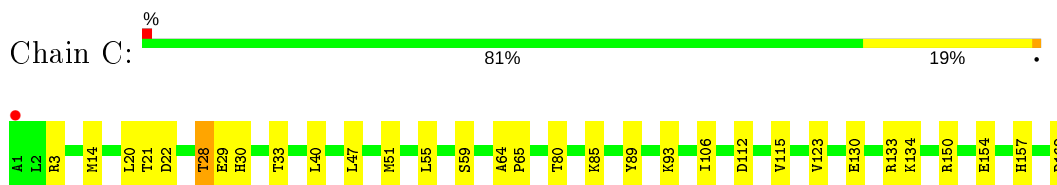
- Molecule 1: Hypothetical protein AQ2171



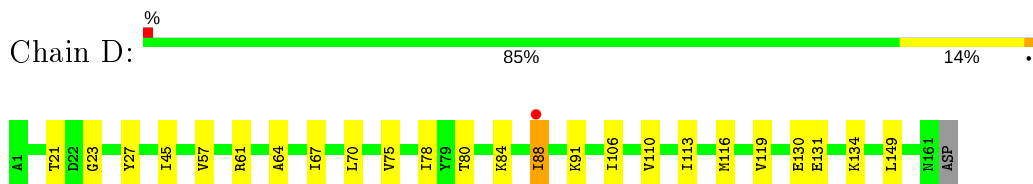
- Molecule 1: Hypothetical protein AQ2171



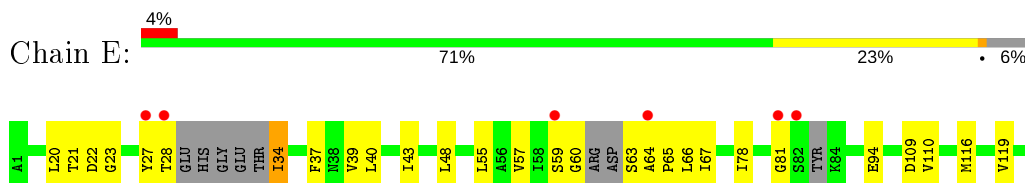
- Molecule 1: Hypothetical protein AQ2171



- Molecule 1: Hypothetical protein AQ2171

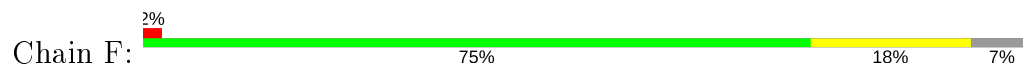


- Molecule 1: Hypothetical protein AQ2171

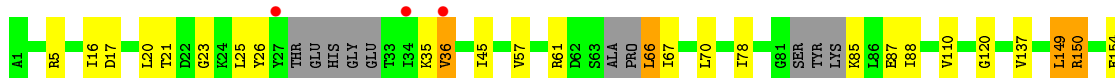
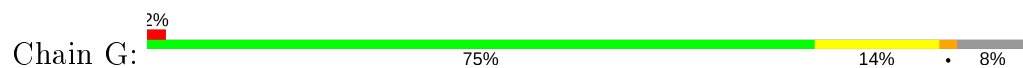




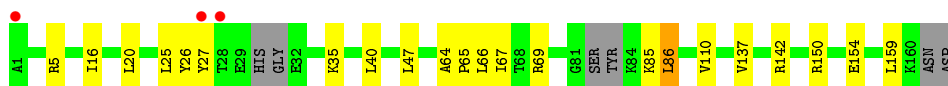
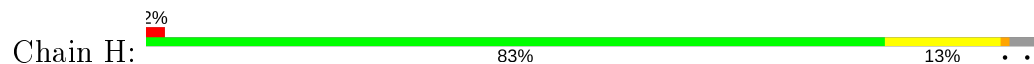
- Molecule 1: Hypothetical protein AQ2171



- Molecule 1: Hypothetical protein AQ2171



- Molecule 1: Hypothetical protein AQ2171



4 Data and refinement statistics

Property	Value	Source
Space group	P 42	Depositor
Cell constants a, b, c, α , β , γ	121.33Å 121.33Å 86.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.92 – 2.40 33.91 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (33.92-2.40) 96.3 (33.91-2.40)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.74 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.216 , 0.268 0.222 , 0.223	Depositor DCC
R_{free} test set	2418 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	29.2	Xtriage
Anisotropy	0.042	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 8.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.129 for h,-k,-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	10046	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.47% 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 [i](#)

5.1 Standard geometry [i](#)

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	0.52	1/1250 (0.1%)	0.62	0/1671
1	B	0.61	2/1293 (0.2%)	0.62	0/1730
1	C	0.54	0/1290	0.60	0/1726
1	D	0.48	0/1285	0.59	0/1719
1	E	0.56	0/1193	0.64	0/1595
1	F	0.51	0/1186	0.62	0/1585
1	G	0.52	0/1171	0.64	0/1562
1	H	0.51	0/1226	0.67	1/1638 (0.1%)
All	All	0.53	3/9894 (0.0%)	0.62	1/13226 (0.0%)

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	B	0	1
1	E	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	77	GLU	CD-OE2	-5.55	1.19	1.25
1	B	77	GLU	CD-OE1	-5.29	1.19	1.25
1	A	130	GLU	CG-CD	5.17	1.59	1.51

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	86	LEU	CB-CA-C	-8.61	93.84	110.20

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	75	VAL	Peptide
1	E	143	ASN	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	1241	0	1318	26	0
1	B	1282	0	1355	12	0
1	C	1279	0	1353	28	0
1	D	1274	0	1351	14	0
1	E	1187	0	1249	25	0
1	F	1179	0	1245	19	0
1	G	1167	0	1240	37	0
1	H	1219	0	1298	35	0
2	A	28	0	0	0	0
2	B	22	0	0	0	0
2	C	34	0	0	0	0
2	D	30	0	0	0	0
2	E	29	0	0	0	0
2	F	26	0	0	0	0
2	G	28	0	0	0	0
2	H	21	0	0	0	0
All	All	10046	0	10409	162	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.

The worst 5 of 162 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:G:36:VAL:HG21	1:H:27:TYR:H	0.96	1.12
1:A:3:ARG:HH21	1:E:161:ASN:CB	1.64	1.10
1:G:36:VAL:HG21	1:H:27:TYR:N	1.70	1.06
1:H:86:LEU:O	1:H:86:LEU:HG	1.69	0.92
1:H:40:LEU:HB3	1:H:150:ARG:HD2	1.58	0.86

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	154/162 (95%)	149 (97%)	5 (3%)	0	100	100
1	B	160/162 (99%)	154 (96%)	6 (4%)	0	100	100
1	C	160/162 (99%)	156 (98%)	4 (2%)	0	100	100
1	D	159/162 (98%)	155 (98%)	4 (2%)	0	100	100
1	E	145/162 (90%)	137 (94%)	7 (5%)	1 (1%)	22	32
1	F	143/162 (88%)	139 (97%)	4 (3%)	0	100	100
1	G	141/162 (87%)	133 (94%)	8 (6%)	0	100	100
1	H	150/162 (93%)	143 (95%)	7 (5%)	0	100	100
All	All	1212/1296 (94%)	1166 (96%)	45 (4%)	1 (0%)	51	68

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	144	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	134/136 (98%)	131 (98%)	3 (2%)	52	71

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	139/136 (102%)	137 (99%)	2 (1%)	67	82
1	C	138/136 (102%)	137 (99%)	1 (1%)	84	92
1	D	138/136 (102%)	135 (98%)	3 (2%)	52	71
1	E	126/136 (93%)	119 (94%)	7 (6%)	21	34
1	F	125/136 (92%)	121 (97%)	4 (3%)	39	59
1	G	124/136 (91%)	119 (96%)	5 (4%)	31	49
1	H	130/136 (96%)	126 (97%)	4 (3%)	40	60
All	All	1054/1088 (97%)	1025 (97%)	29 (3%)	43	63

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	59	SER
1	F	36	VAL
1	H	47	LEU
1	E	94	GLU
1	F	51	MSE

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

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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 [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	155/162 (95%)	-0.28	0 100 100	13, 23, 31, 34	2 (1%)
1	B	159/162 (98%)	-0.36	0 100 100	13, 21, 29, 35	0
1	C	159/162 (98%)	-0.29	1 (0%) 89 88	13, 20, 30, 40	0
1	D	158/162 (97%)	-0.24	1 (0%) 89 88	14, 25, 37, 41	0
1	E	150/162 (92%)	0.09	6 (4%) 38 37	17, 27, 45, 50	0
1	F	148/162 (91%)	-0.20	3 (2%) 65 63	14, 22, 37, 47	0
1	G	146/162 (90%)	-0.01	3 (2%) 63 61	13, 26, 43, 53	0
1	H	153/162 (94%)	-0.10	3 (1%) 65 63	18, 25, 36, 43	0
All	All	1228/1296 (94%)	-0.18	17 (1%) 75 73	13, 24, 38, 53	2 (0%)

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	27	TYR	6.0
1	F	27	TYR	5.5
1	F	33	THR	5.4
1	E	64	ALA	5.0
1	G	34	ILE	3.9

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.