



# Full wwPDB X-ray Structure Validation Report ⓘ

Apr 2, 2014 – 08:19 PM EDT

PDB ID : 3F9Y  
Title : Structural Insights into Lysine Multiple Methylation by SET Domain Methyltransferases, SET8-Y334F / H4-Lys20me1 / AdoHcy  
Authors : Couture, J-F.; Dirk, L.M.A.; Brunzelle, J.S.; Houtz, R.L.; Trievel, R.C.  
Deposited on : 2008-11-14  
Resolution : 1.50 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

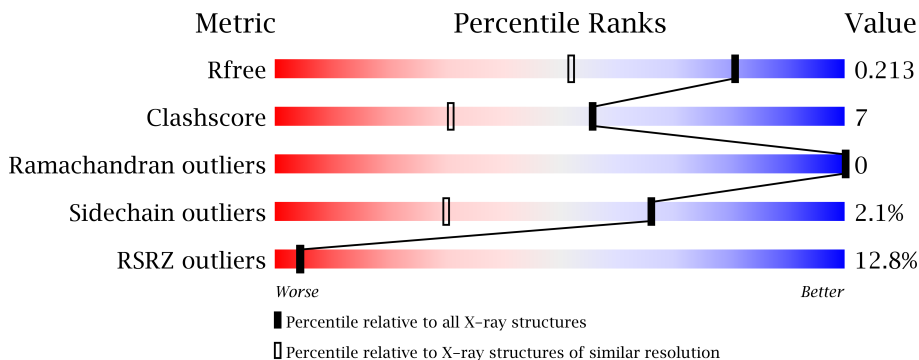
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : **FAILED**  
Xtriage (Phenix) : dev-1439  
EDS : stable22978  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22978

# 1 Overall quality at a glance

The reported resolution of this entry is 1.50 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))
$R_{free}$	66092	1513 (1.50-1.50)
Clashscore	79885	1768 (1.50-1.50)
Ramachandran outliers	78287	1720 (1.50-1.50)
Sidechain outliers	78261	1718 (1.50-1.50)
RSRZ outliers	66119	1514 (1.50-1.50)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	166	
1	B	166	
2	E	10	
2	F	10	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2978 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Histone-lysine N-methyltransferase SETD8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	159	Total	C	N	O	S	0	1	0
			1271	795	223	247	6			
1	B	160	Total	C	N	O	S	0	6	0
			1295	807	230	252	6			

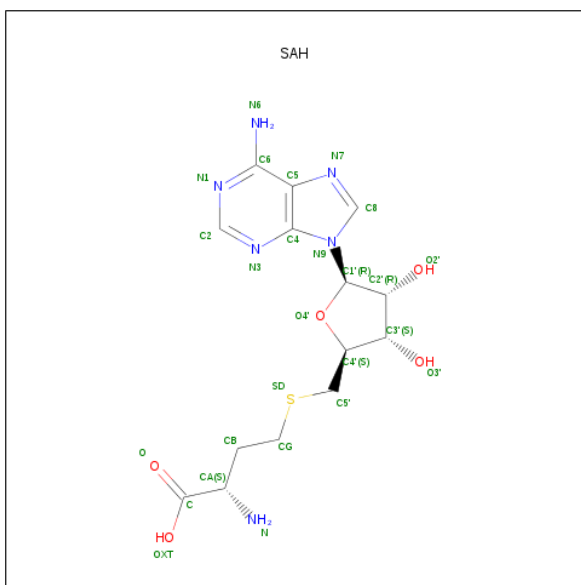
There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	187	GLY	-	EXPRESSION TAG	UNP Q9NQR1
A	188	ALA	-	EXPRESSION TAG	UNP Q9NQR1
A	189	MET	-	EXPRESSION TAG	UNP Q9NQR1
A	190	GLY	-	EXPRESSION TAG	UNP Q9NQR1
A	334	PHE	TYR	ENGINEERED	UNP Q9NQR1
B	187	GLY	-	EXPRESSION TAG	UNP Q9NQR1
B	188	ALA	-	EXPRESSION TAG	UNP Q9NQR1
B	189	MET	-	EXPRESSION TAG	UNP Q9NQR1
B	190	GLY	-	EXPRESSION TAG	UNP Q9NQR1
B	334	PHE	TYR	ENGINEERED	UNP Q9NQR1

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	7	Total	C	N	O	0	0	0
			66	42	17	7			
2	F	8	Total	C	N	O	0	0	0
			67	42	17	8			

- Molecule 3 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: C<sub>14</sub>H<sub>20</sub>N<sub>6</sub>O<sub>5</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total 26	C 14	N 6	O 5	S 1	0	0
3	B	1	Total 26	C 14	N 6	O 5	S 1	0	0

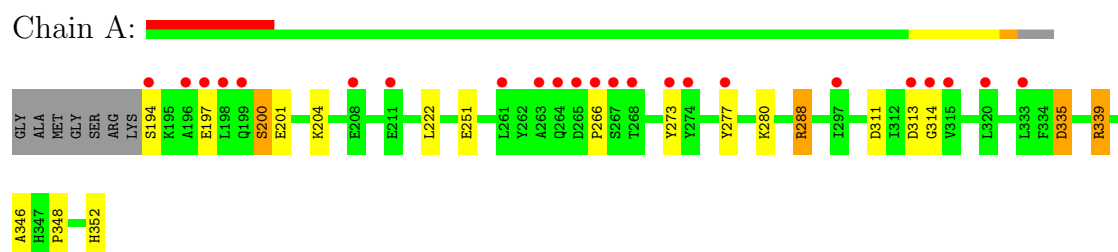
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	227	Total O 227 227	0	0

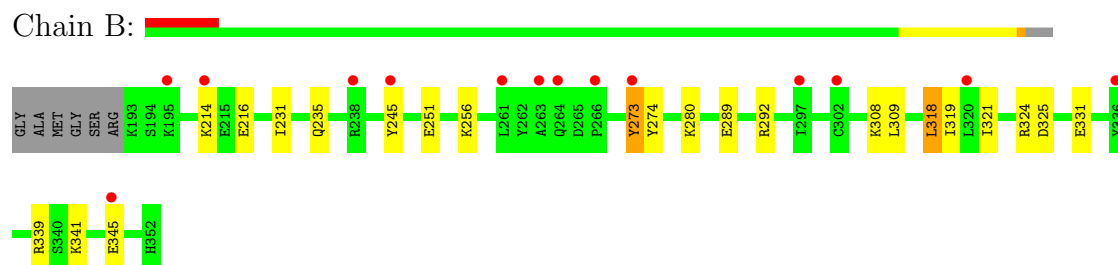
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Histone-lysine N-methyltransferase SETD8



- Molecule 1: Histone-lysine N-methyltransferase SETD8



- Molecule 2: Histone H4



- Molecule 2: Histone H4



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.40Å 45.00Å 52.40Å 91.40° 115.30° 89.80°	Depositor
Resolution (Å)	23.68 – 1.50 23.68 – 1.65	Depositor EDS
% Data completeness (in resolution range)	100.0 (23.68-1.50) 96.3 (23.68-1.65)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	12.01 (at 1.65Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.194 , 0.220 0.187 , 0.213	Depositor DCC
$R_{free}$ test set	2125 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	28.1	Xtriage
Anisotropy	0.073	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 55.9	EDS
Estimated twinning fraction	0.018 for h,-k,-h-l 0.014 for -h,k,-l 0.009 for -h,-k,h+l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 42488 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2978	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.34% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MLZ, SAH

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	0.94	1/1299 (0.1%)	1.01	3/1743 (0.2%)
1	B	0.92	1/1354 (0.1%)	1.01	6/1813 (0.3%)
2	E	1.01	0/55	0.84	0/69
2	F	0.90	0/56	0.97	0/72
All	All	0.93	2/2764 (0.1%)	1.00	9/3697 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	194	SER	CB-OG	6.29	1.50	1.42
1	B	273	TYR	CE2-CZ	6.00	1.46	1.38

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	339	ARG	NE-CZ-NH1	9.43	125.01	120.30
1	A	339	ARG	NE-CZ-NH2	-8.84	115.88	120.30
1	B	324	ARG	NE-CZ-NH2	7.29	123.94	120.30
1	A	335	ASP	CB-CG-OD2	-5.82	113.07	118.30
1	B	318	LEU	CB-CG-CD1	-5.68	101.34	111.00
1	B	289	GLU	OE1-CD-OE2	-5.35	116.89	123.30
1	B	245	TYR	CZ-CE2-CD2	-5.29	115.04	119.80
1	B	292	ARG	NE-CZ-NH2	-5.21	117.69	120.30
1	B	325	ASP	CB-CG-OD1	5.01	122.81	118.30

There are no chirality outliers.

There are no planarity outliers.

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1271	0	1249	19	0
1	B	1295	0	1267	17	0
2	E	66	0	78	2	0
2	F	67	0	69	0	0
3	A	26	0	19	0	0
3	B	26	0	19	0	0
4	B	227	0	0	16	1
All	All	2978	0	2701	36	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 7.

All (36) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:313:ASP:HB2	4:B:378:HOH:O	1.59	1.03
1:B:214:LYS:HB3	4:B:373:HOH:O	1.63	0.98
1:A:313:ASP:HB2	4:B:111:HOH:O	1.68	0.92
1:A:288:ARG:CG	1:A:288:ARG:HH11	1.86	0.88
1:A:288:ARG:HH11	1:A:288:ARG:HG3	1.40	0.84
1:A:288:ARG:HD3	4:B:170:HOH:O	1.78	0.82
1:A:346:ALA:HB2	2:E:16:LYS:HE3	1.62	0.80
1:B:309:LEU:HD23	1:B:318:LEU:HD23	1.67	0.75
4:B:374:HOH:O	2:E:19:ARG:HD3	1.91	0.70
1:B:339[B]:ARG:NH2	4:B:183:HOH:O	2.27	0.67
1:A:288:ARG:NH1	1:A:288:ARG:HG3	2.12	0.65
1:B:235:GLN:OE1	4:B:112:HOH:O	2.16	0.62
1:A:197:GLU:O	1:A:201:GLU:HG2	2.04	0.57
1:B:214:LYS:HG2	1:B:216:GLU:HG3	1.87	0.55
1:A:222:LEU:HD13	1:B:274:TYR:CG	2.40	0.55
1:A:277:TYR:OH	1:A:311:ASP:OD2	2.14	0.54
1:B:319:ILE:HD12	1:B:321:ILE:HD11	1.90	0.54
1:B:214:LYS:CB	4:B:373:HOH:O	2.36	0.53
1:B:308[B]:LYS:HG2	1:B:319:ILE:HG13	1.90	0.53
1:A:280:LYS:HD3	4:B:381:HOH:O	2.09	0.52

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:288:ARG:CG	1:A:288:ARG:NH1	2.57	0.51
1:B:231:ILE:HD13	1:B:331:GLU:HA	1.93	0.51
1:B:308[B]:LYS:HE2	1:B:309:LEU:O	2.11	0.50
1:A:266:PRO:HG3	4:B:173:HOH:O	2.12	0.49
1:A:335:ASP:OD2	1:A:339:ARG:NH2	2.30	0.49
1:A:200:SER:O	1:A:204:LYS:HG2	2.15	0.47
1:B:309:LEU:HD23	1:B:309:LEU:HA	1.77	0.47
1:B:256:LYS:HD2	4:B:150:HOH:O	2.15	0.47
1:B:235:GLN:CD	4:B:112:HOH:O	2.55	0.45
1:B:251:GLU:HB2	4:B:361:HOH:O	2.16	0.45
1:A:352:HIS:CD2	4:B:88:HOH:O	2.71	0.43
1:A:313:ASP:CB	4:B:111:HOH:O	2.46	0.43
1:B:280:LYS:HE2	4:B:46:HOH:O	2.18	0.41
1:A:346:ALA:O	1:A:348:PRO:HD3	2.20	0.40
1:A:277:TYR:OH	1:A:314:GLY:HA2	2.21	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	Distance(Å)	Clash(Å)
4:B:121:HOH:O	4:B:127:HOH:O[1_556]	2.19	0.01

## 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	158/166 (95%)	154 (98%)	4 (2%)	0	100	100
1	B	164/166 (99%)	160 (98%)	4 (2%)	0	100	100
2	E	4/10 (40%)	4 (100%)	0	0	100	100
2	F	5/10 (50%)	5 (100%)	0	0	100	100
All	All	331/352 (94%)	323 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	136/139 (98%)	132 (97%)	4 (3%)	55	18
1	B	142/139 (102%)	140 (99%)	2 (1%)	78	51
2	E	6/8 (75%)	6 (100%)	0	100	100
2	F	5/8 (62%)	5 (100%)	0	100	100
All	All	289/294 (98%)	283 (98%)	6 (2%)	66	30

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	200	SER
1	A	251	GLU
1	A	273	TYR
1	A	288	ARG
1	B	273	TYR
1	B	341	LYS

Some 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 chains in this entry.

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

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	159/166 (95%)	0.73	23 (14%) 3 3	26, 31, 42, 47	1 (0%)
1	B	160/166 (96%)	0.53	14 (8%) 10 11	25, 31, 40, 44	0
2	E	7/10 (70%)	1.45	3 (42%) 1 0	30, 32, 35, 42	0
2	F	8/10 (80%)	2.03	3 (37%) 1 1	29, 33, 39, 41	0
All	All	334/352 (94%)	0.68	43 (12%) 4 4	25, 31, 41, 47	1 (0%)

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	266	PRO	6.3
1	A	194	SER	5.1
2	F	21	VAL	4.6
1	A	320[A]	LEU	4.4
2	F	16	LYS	4.1
1	A	314	GLY	3.9
1	A	267	SER	3.9
1	A	313	ASP	3.8
2	F	20	MLZ	3.7
1	B	266	PRO	3.7
1	A	197	GLU	3.6
1	B	214	LYS	3.6
1	A	264	GLN	3.3
2	E	16	LYS	3.3
1	B	263	ALA	3.3
1	A	199	GLN	3.2
1	A	263	ALA	3.2
1	B	273	TYR	3.1
1	A	261	LEU	3.1
1	A	277	TYR	3.1
1	A	297	ILE	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	345[A]	GLU	2.9
1	B	261	LEU	2.8
1	B	245	TYR	2.8
1	B	297	ILE	2.7
1	A	268	THR	2.7
1	B	238	ARG	2.7
1	B	195	LYS	2.6
1	A	198	LEU	2.6
1	A	196	ALA	2.5
2	E	21	VAL	2.4
1	A	265	ASP	2.4
2	E	20	MLZ	2.4
1	A	315	VAL	2.4
1	B	336	TYR	2.3
1	B	302	CYS	2.3
1	B	320[A]	LEU	2.3
1	A	211	GLU	2.2
1	A	273	TYR	2.2
1	A	274	TYR	2.1
1	A	333	LEU	2.1
1	A	208	GLU	2.1
1	B	264	GLN	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	MLZ	F	20	10/11	0.32	3.72	27,29,32,32	0
2	MLZ	E	20	10/11	0.25	3.33	28,29,32,32	0

## 6.3 Carbohydrates ⓘ

There are no carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	SAH	B	801	26/26	0.10	-0.20	20,24,27,29	0
3	SAH	A	801	26/26	0.07	-0.83	17,21,26,27	0

## 6.5 Other polymers

There are no such residues in this entry.