



# Full wwPDB X-ray Structure Validation Report ⓘ

Dec 8, 2025 – 12:33 PM EST

PDB ID : 9MHI / pdb\_00009mhi  
Title : The structure of SeMet substituted Zcp  
Authors : Bera, A.K.; Liyayi, I.K.; Criss, A.K.; Noinaj, N.  
Deposited on : 2024-12-12  
Resolution : 2.39 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](mailto: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 : 4-5-2 with Phenix2.0  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.47

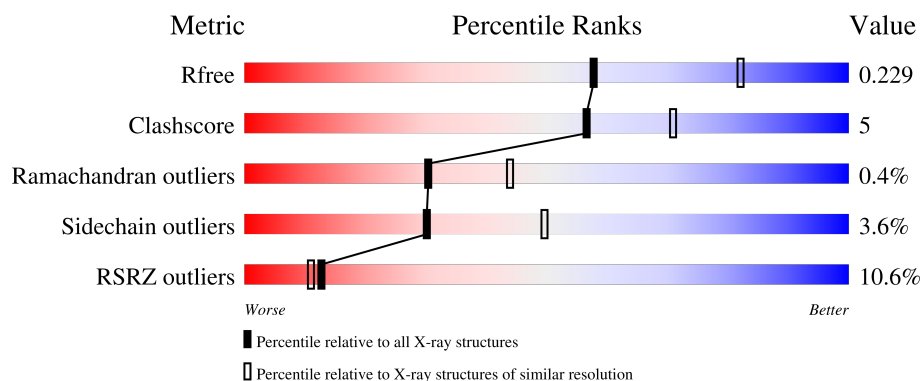
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.39 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))
$R_{free}$	164625	4642 (2.40-2.40)
Clashscore	180529	5218 (2.40-2.40)
Ramachandran outliers	177936	5158 (2.40-2.40)
Sidechain outliers	177891	5159 (2.40-2.40)
RSRZ outliers	164620	4642 (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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	274	<div> <div>8%</div> <div> <div></div> <div>75%</div> <div>13%</div> <div>• 11%</div> </div> </div>
1	B	274	<div> <div>10%</div> <div> <div></div> <div>79%</div> <div>9%</div> <div>• 11%</div> </div> </div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 3904 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 Phosphoribosylglycinamide formyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	245	Total	C	N	O	S	Se	0	0	0
			1892	1191	329	367	2	3			
1	B	245	Total	C	N	O	S	Se	0	0	0
			1883	1187	326	365	2	3			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	SER	-	expression tag	UNP Q5F7W6
A	-4	TYR	-	expression tag	UNP Q5F7W6
A	-3	TYR	-	expression tag	UNP Q5F7W6
A	-2	HIS	-	expression tag	UNP Q5F7W6
A	-1	HIS	-	expression tag	UNP Q5F7W6
A	0	HIS	-	expression tag	UNP Q5F7W6
A	1	HIS	-	expression tag	UNP Q5F7W6
A	2	HIS	-	expression tag	UNP Q5F7W6
A	3	HIS	-	expression tag	UNP Q5F7W6
A	4	ASP	-	expression tag	UNP Q5F7W6
A	5	TYR	-	expression tag	UNP Q5F7W6
A	6	ASP	-	expression tag	UNP Q5F7W6
A	7	ILE	-	expression tag	UNP Q5F7W6
A	8	PRO	-	expression tag	UNP Q5F7W6
A	9	THR	-	expression tag	UNP Q5F7W6
A	10	THR	-	expression tag	UNP Q5F7W6
A	11	GLU	-	expression tag	UNP Q5F7W6
A	12	ASN	-	expression tag	UNP Q5F7W6
A	13	LEU	-	expression tag	UNP Q5F7W6
A	14	TYR	-	expression tag	UNP Q5F7W6
A	15	PHE	-	expression tag	UNP Q5F7W6
A	16	GLN	-	expression tag	UNP Q5F7W6
A	17	GLY	-	expression tag	UNP Q5F7W6
A	18	ALA	-	expression tag	UNP Q5F7W6
A	19	MSE	-	expression tag	UNP Q5F7W6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	VAL	-	expression tag	UNP Q5F7W6
B	-5	SER	-	expression tag	UNP Q5F7W6
B	-4	TYR	-	expression tag	UNP Q5F7W6
B	-3	TYR	-	expression tag	UNP Q5F7W6
B	-2	HIS	-	expression tag	UNP Q5F7W6
B	-1	HIS	-	expression tag	UNP Q5F7W6
B	0	HIS	-	expression tag	UNP Q5F7W6
B	1	HIS	-	expression tag	UNP Q5F7W6
B	2	HIS	-	expression tag	UNP Q5F7W6
B	3	HIS	-	expression tag	UNP Q5F7W6
B	4	ASP	-	expression tag	UNP Q5F7W6
B	5	TYR	-	expression tag	UNP Q5F7W6
B	6	ASP	-	expression tag	UNP Q5F7W6
B	7	ILE	-	expression tag	UNP Q5F7W6
B	8	PRO	-	expression tag	UNP Q5F7W6
B	9	THR	-	expression tag	UNP Q5F7W6
B	10	THR	-	expression tag	UNP Q5F7W6
B	11	GLU	-	expression tag	UNP Q5F7W6
B	12	ASN	-	expression tag	UNP Q5F7W6
B	13	LEU	-	expression tag	UNP Q5F7W6
B	14	TYR	-	expression tag	UNP Q5F7W6
B	15	PHE	-	expression tag	UNP Q5F7W6
B	16	GLN	-	expression tag	UNP Q5F7W6
B	17	GLY	-	expression tag	UNP Q5F7W6
B	18	ALA	-	expression tag	UNP Q5F7W6
B	19	MSE	-	expression tag	UNP Q5F7W6
B	20	VAL	-	expression tag	UNP Q5F7W6

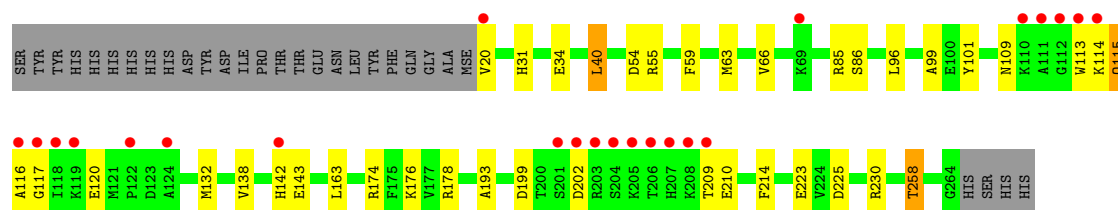
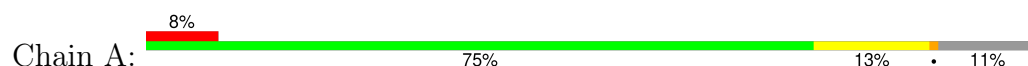
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	91	Total O 91 91	0	0
2	B	38	Total O 38 38	0	0

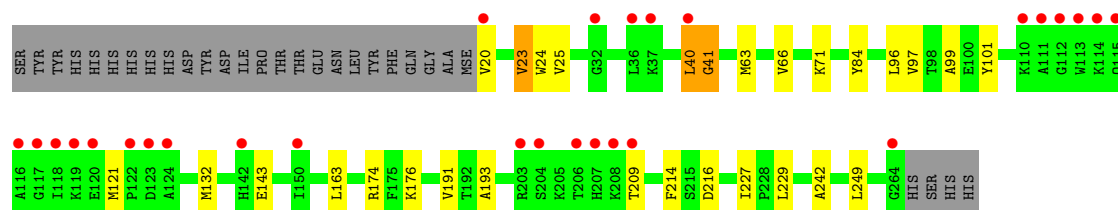
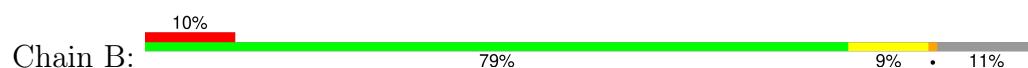
### 3 Residue-property plots

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: Phosphoribosylglycinamide formyltransferase



- Molecule 1: Phosphoribosylglycinamide formyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.76Å 92.30Å 127.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.15 – 2.39 46.15 – 2.39	Depositor EDS
% Data completeness (in resolution range)	97.0 (46.15-2.39) 97.0 (46.15-2.39)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.13 (at 2.39Å)	Xtriage
Refinement program	PHENIX (???)	Depositor
R, $R_{free}$	0.195 , 0.230 0.196 , 0.229	Depositor DCC
$R_{free}$ test set	1950 reflections (7.30%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.6	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 61.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3904	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical 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	0.16	0/1931	0.46	1/2615 (0.0%)
1	B	0.14	0/1922	0.35	0/2604
All	All	0.15	0/3853	0.41	1/5219 (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	A	0	4
1	B	0	2
All	All	0	6

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	116	ALA	N-CA-C	6.27	124.16	110.80

There are no chirality outliers.

All (6) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	113	TRP	Peptide
1	A	114	LYS	Peptide
1	A	115	GLN	Peptide
1	A	120	GLU	Peptide
1	B	121	MSE	Peptide
1	B	40	LEU	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	1892	0	1779	17	0
1	B	1883	0	1766	18	0
2	A	91	0	0	3	0
2	B	38	0	0	4	0
All	All	3904	0	3545	33	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 5.

All (33) 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:142:HIS:O	2:A:301:HOH:O	1.84	0.93
1:B:249:LEU:O	2:B:301:HOH:O	2.04	0.75
1:A:258:THR:HG22	2:A:362:HOH:O	1.97	0.63
1:A:101:TYR:HB3	1:A:132:MSE:HB2	1.80	0.61
1:B:227:ILE:O	2:B:302:HOH:O	2.17	0.59
1:B:63:MSE:HE3	1:B:99:ALA:HB2	1.87	0.56
1:A:193:ALA:HB3	1:A:214:PHE:HB3	1.88	0.56
1:B:193:ALA:HB3	1:B:214:PHE:HB3	1.90	0.52
1:A:54:ASP:OD2	1:A:55:ARG:NH1	2.43	0.51
1:B:66:VAL:HB	1:B:96:LEU:HB2	1.95	0.49
1:A:174:ARG:HB3	1:A:225:ASP:HB3	1.96	0.48
1:B:25:VAL:HG13	1:B:63:MSE:HE1	1.96	0.48
1:B:101:TYR:CD2	1:B:132:MSE:HE2	2.49	0.48
1:A:163:LEU:HD12	1:A:176:LYS:HG2	1.96	0.47
1:B:63:MSE:HB3	1:B:63:MSE:HE2	1.75	0.47
1:B:84:TYR:O	2:B:303:HOH:O	2.20	0.46
1:A:66:VAL:HB	1:A:96:LEU:HB2	1.98	0.46
1:A:117:GLY:O	1:B:242:ALA:HB1	2.16	0.46
1:B:163:LEU:HD12	1:B:176:LYS:HG2	1.97	0.45
1:A:63:MSE:SE	1:A:99:ALA:HB2	2.66	0.45
1:A:31:HIS:O	1:A:34:GLU:HG2	2.17	0.45
1:A:199:ASP:OD1	1:B:174:ARG:NH2	2.39	0.44
1:A:40:LEU:HD11	1:A:59:PHE:CD1	2.52	0.44

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:178:ARG:NH2	1:A:223:GLU:OE2	2.41	0.43
1:B:63:MSE:HE2	1:B:97:VAL:HG12	1.99	0.43
1:A:142:HIS:O	1:A:143:GLU:HG2	2.19	0.43
1:B:143:GLU:OE2	1:B:143:GLU:N	2.43	0.42
1:B:216:ASP:OD2	2:B:304:HOH:O	2.22	0.42
1:A:230:ARG:HD2	2:A:387:HOH:O	2.20	0.42
1:B:24:TRP:O	1:B:41:GLY:HA3	2.20	0.41
1:A:85:ARG:HG2	1:A:86:SER:O	2.20	0.41
1:B:23:VAL:HA	1:B:41:GLY:O	2.22	0.40
1:B:66:VAL:HG22	1:B:71:LYS:HG2	2.02	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	243/274 (89%)	222 (91%)	20 (8%)	1 (0%)	30	44
1	B	243/274 (89%)	230 (95%)	12 (5%)	1 (0%)	30	44
All	All	486/548 (89%)	452 (93%)	32 (7%)	2 (0%)	30	44

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	115	GLN
1	B	41	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	194/231 (84%)	186 (96%)	8 (4%)	26	44
1	B	192/231 (83%)	186 (97%)	6 (3%)	35	56
All	All	386/462 (84%)	372 (96%)	14 (4%)	30	49

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

Mol	Chain	Res	Type
1	A	20	VAL
1	A	40	LEU
1	A	109	ASN
1	A	138	VAL
1	A	202	ASP
1	A	209	THR
1	A	210	GLU
1	A	258	THR
1	B	20	VAL
1	B	23	VAL
1	B	40	LEU
1	B	191	VAL
1	B	209	THR
1	B	229	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	212	GLN
1	B	64	GLN
1	B	109	ASN
1	B	168	ASN
1	B	212	GLN

### 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 oligosaccharides 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

### 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	241/274 (87%)	0.30	23 (9%) 15 13	31, 50, 148, 195	0
1	B	241/274 (87%)	0.65	28 (11%) 11 9	38, 64, 148, 195	0
All	All	482/548 (87%)	0.47	51 (10%) 13 11	31, 57, 151, 195	0

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	209	THR	6.1
1	B	20	VAL	5.5
1	B	116	ALA	5.4
1	A	117	GLY	5.2
1	A	208	LYS	5.1
1	A	204	SER	5.0
1	A	203	ARG	4.7
1	B	118	ILE	4.3
1	B	207	HIS	4.2
1	A	118	ILE	4.2
1	A	111	ALA	4.2
1	A	116	ALA	4.1
1	B	111	ALA	4.1
1	A	206	THR	4.0
1	A	124	ALA	4.0
1	B	206	THR	3.9
1	B	115	GLN	3.8
1	B	204	SER	3.7
1	A	207	HIS	3.6
1	B	122	PRO	3.6
1	A	20	VAL	3.5
1	A	113	TRP	3.5
1	B	124	ALA	3.5
1	B	123	ASP	3.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	201	SER	3.3
1	B	113	TRP	3.2
1	B	40	LEU	3.2
1	A	202	ASP	3.2
1	A	122	PRO	3.2
1	B	208	LYS	3.1
1	B	120	GLU	3.1
1	B	119	LYS	3.0
1	A	110	LYS	2.9
1	B	203	ARG	2.8
1	A	69	LYS	2.8
1	B	32	GLY	2.7
1	A	119	LYS	2.7
1	A	142	HIS	2.6
1	B	110	LYS	2.6
1	B	114	LYS	2.6
1	A	205	LYS	2.6
1	A	114	LYS	2.4
1	B	117	GLY	2.3
1	B	209	THR	2.2
1	B	150	ILE	2.2
1	B	142	HIS	2.2
1	A	112	GLY	2.1
1	B	36	LEU	2.1
1	B	112	GLY	2.1
1	B	264	GLY	2.0
1	B	37	LYS	2.0

## 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 oligosaccharides 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.