



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 13, 2017 – 01:54 PM EST

PDB ID : 3WTY  
Title : Crystal structure of the complex comprised of ETS1(G333P), RUNX1, CBF-BETA, and the tcralpha gene enhancer DNA  
Authors : Shiina, M.; Hamada, K.; Ogata, K.  
Deposited on : unknown  
Resolution : 2.70 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](mailto: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	:	rb-20030345
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)	:	rb-20030345

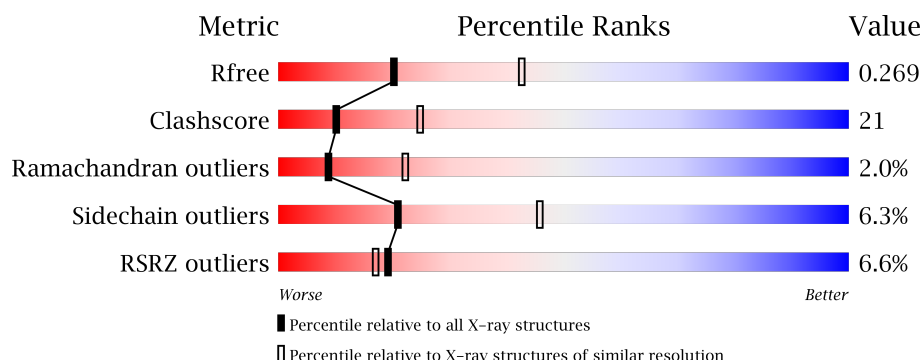
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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. 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	204	
1	F	204	
2	B	142	
2	G	142	
3	C	166	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	H	166	
4	D	15	
4	I	15	
5	E	15	
5	J	15	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7039 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 Runt-related transcription factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	118	Total	C	N	O	S	0	0	0
			914	574	169	167	4			
1	F	118	Total	C	N	O	S	0	0	0
			914	574	169	167	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	LYS	LEU	ENGINEERED MUTATION	UNP Q03347
F	94	LYS	LEU	ENGINEERED MUTATION	UNP Q03347

- Molecule 2 is a protein called Core-binding factor subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	130	Total	C	N	O	S	0	0	0
			1071	671	195	199	6			
2	G	129	Total	C	N	O	S	0	0	0
			1062	666	193	197	6			

- Molecule 3 is a protein called Protein C-ets-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	118	Total	C	N	O	S	0	0	0
			970	630	165	171	4			
3	H	101	Total	C	N	O	S	0	0	0
			856	556	147	149	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	333	PRO	GLY	ENGINEERED MUTATION	UNP P14921
H	333	PRO	GLY	ENGINEERED MUTATION	UNP P14921

- Molecule 4 is a DNA chain called DNA (5'-D(\*GP\*AP\*AP\*GP\*CP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			
4	I	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			

- Molecule 5 is a DNA chain called DNA (5'-D(\*AP\*GP\*AP\*GP\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*TP\*TP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			
5	J	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			

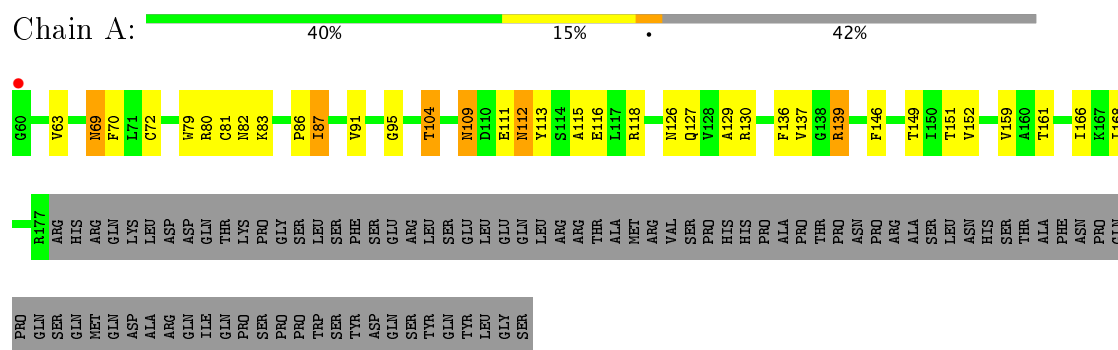
- Molecule 6 is water.

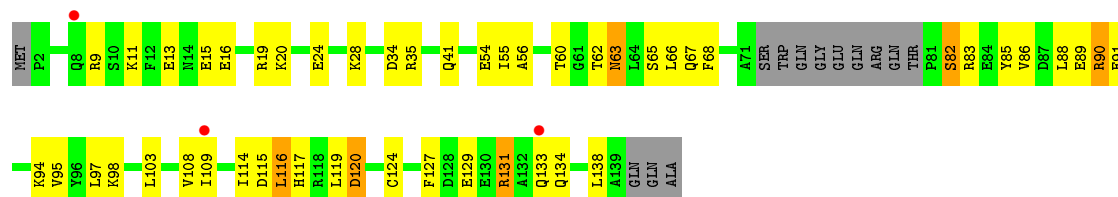
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	3	Total	O	0	0
			3	3		
6	B	2	Total	O	0	0
			2	2		
6	C	10	Total	O	0	0
			10	10		
6	F	5	Total	O	0	0
			5	5		
6	G	3	Total	O	0	0
			3	3		
6	H	3	Total	O	0	0
			3	3		
6	D	2	Total	O	0	0
			2	2		
6	E	4	Total	O	0	0
			4	4		
6	J	2	Total	O	0	0
			2	2		

### 3 Residue-property plots [i](#)

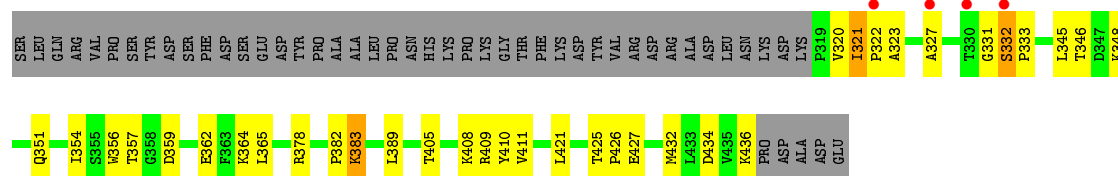
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Runt-related transcription factor 1

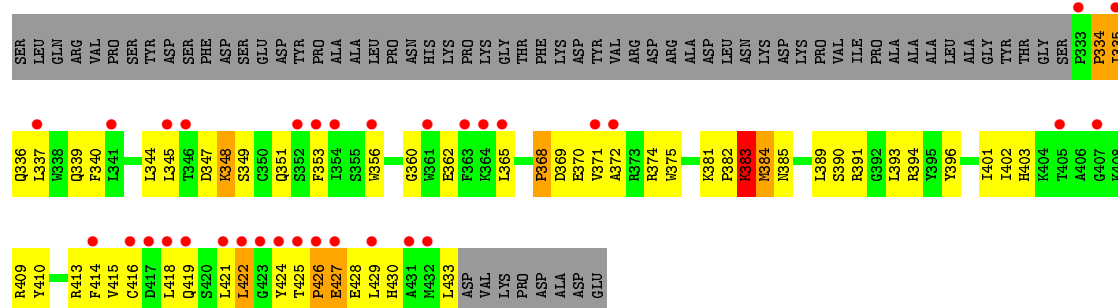
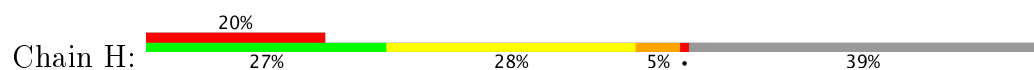




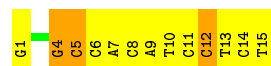
• Molecule 3: Protein C-ets-1



• Molecule 3: Protein C-ets-1



• Molecule 4: DNA (5'-D(\*GP\*AP\*AP\*GP\*CP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*T)-3')

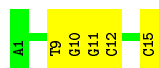


• Molecule 4: DNA (5'-D(\*GP\*AP\*AP\*GP\*CP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*T)-3')



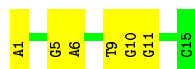
• Molecule 5: DNA (5'-D(\*AP\*GP\*AP\*GP\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*TP\*TP\*C)-3')

Chain E:  67% 33%



- Molecule 5: DNA (5'-D(\*AP\*GP\*AP\*GP\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*TP\*TP\*C)-3')

Chain J:  60% 40%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.80Å 102.50Å 194.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.69 – 2.70 48.69 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.7 (48.69-2.70) 98.8 (48.69-2.70)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.52 (at 2.69Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.227 , 0.268 0.227 , 0.269	Depositor DCC
$R_{free}$ test set	4390 reflections (10.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	78.8	Xtriage
Anisotropy	0.203	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 60.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7039	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.69% 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

### 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	0.58	1/933 (0.1%)	0.69	0/1268
1	F	0.58	0/933	0.76	0/1268
2	B	0.45	0/1093	0.57	0/1466
2	G	0.49	0/1084	0.58	0/1454
3	C	0.54	0/998	0.69	0/1349
3	H	0.39	0/881	0.50	0/1187
4	D	0.70	0/334	0.99	1/512 (0.2%)
4	I	0.65	0/334	0.90	1/512 (0.2%)
5	E	0.66	0/348	0.83	0/537
5	J	0.58	0/348	0.79	0/537
All	All	0.54	1/7286 (0.0%)	0.69	2/10090 (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
4	D	0	4
4	I	0	2
All	All	0	6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	81	CYS	CB-SG	-5.90	1.72	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	4	DG	N9-C1'-C2'	-6.31	100.61	112.60
4	I	4	DG	N9-C1'-C2'	-5.73	101.71	112.60

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	12	DC	Sidechain
4	D	15	DT	Sidechain
4	D	4	DG	Sidechain
4	D	5	DC	Sidechain
4	I	4	DG	Sidechain

## 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	914	0	922	41	0
1	F	914	0	922	25	0
2	B	1071	0	1034	54	0
2	G	1062	0	1026	47	0
3	C	970	0	970	29	0
3	H	856	0	852	53	0
4	D	299	0	170	19	0
4	I	299	0	170	17	0
5	E	310	0	171	3	0
5	J	310	0	171	4	0
6	A	3	0	0	0	0
6	B	2	0	0	0	0
6	C	10	0	0	0	0
6	D	2	0	0	0	0
6	E	4	0	0	0	0
6	F	5	0	0	0	0
6	G	3	0	0	0	0
6	H	3	0	0	0	0
6	J	2	0	0	0	0
All	All	7039	0	6408	282	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 21.

The worst 5 of 282 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:82:ASN:HD21	1:A:118:ARG:HH11	1.07	0.98
1:F:130:ARG:HH12	1:F:132:ASN:HD22	1.04	0.94
1:A:82:ASN:HD21	1:A:118:ARG:NH1	1.79	0.81
1:A:82:ASN:ND2	1:A:118:ARG:HH11	1.78	0.81
4:D:10:DT:H2''	4:D:11:DC:H5'	1.64	0.79

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	116/204 (57%)	109 (94%)	7 (6%)	0	100	100
1	F	116/204 (57%)	105 (90%)	10 (9%)	1 (1%)	20	46
2	B	126/142 (89%)	102 (81%)	21 (17%)	3 (2%)	7	17
2	G	125/142 (88%)	111 (89%)	12 (10%)	2 (2%)	11	28
3	C	116/166 (70%)	107 (92%)	8 (7%)	1 (1%)	20	46
3	H	99/166 (60%)	74 (75%)	18 (18%)	7 (7%)	1	1
All	All	698/1024 (68%)	608 (87%)	76 (11%)	14 (2%)	9	22

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	90	ARG
3	C	332	SER
2	G	90	ARG
3	H	334	PRO
3	H	383	LYS

### 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	100/179 (56%)	93 (93%)	7 (7%)	18	40
1	F	100/179 (56%)	96 (96%)	4 (4%)	36	67
2	B	113/123 (92%)	102 (90%)	11 (10%)	9	22
2	G	112/123 (91%)	104 (93%)	8 (7%)	17	39
3	C	103/146 (70%)	99 (96%)	4 (4%)	37	68
3	H	92/146 (63%)	87 (95%)	5 (5%)	26	54
All	All	620/896 (69%)	581 (94%)	39 (6%)	21	46

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	140	GLN
3	C	427	GLU
3	H	383	LYS
3	C	321	ILE
3	C	378	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	134	GLN
2	B	140	GLN
2	G	67	GLN
2	B	41	GLN
2	B	104	ASN

### 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 ⓘ

### 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	118/204 (57%)	0.17	1 (0%) 86 86	54, 71, 94, 104	0
1	F	118/204 (57%)	0.18	1 (0%) 86 86	50, 64, 84, 104	0
2	B	130/142 (91%)	0.59	9 (6%) 18 16	70, 98, 129, 136	0
2	G	129/142 (90%)	0.44	3 (2%) 61 61	70, 91, 117, 122	0
3	C	118/166 (71%)	0.35	4 (3%) 46 45	51, 71, 104, 109	0
3	H	101/166 (60%)	1.55	33 (32%) 0 0	74, 123, 164, 167	0
4	D	15/15 (100%)	-0.19	0 100 100	58, 62, 72, 78	0
4	I	15/15 (100%)	-0.22	0 100 100	62, 70, 84, 87	0
5	E	15/15 (100%)	-0.24	0 100 100	54, 61, 75, 77	0
5	J	15/15 (100%)	-0.35	0 100 100	56, 68, 85, 88	0
All	All	774/1084 (71%)	0.46	51 (6%) 19 17	50, 81, 130, 167	0

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	H	424	TYR	6.1
3	H	423	GLY	5.1
3	H	416	CYS	4.9
3	H	333	PRO	4.7
3	H	361	TRP	4.0

### 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 [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.