



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 11:25 am GMT

PDB ID : 1IKY  
Title : HIV-1 Reverse Transcriptase in Complex with the Inhibitor MSC194  
Authors : Lindberg, J.; Unge, T.  
Deposited on : 2001-05-07  
Resolution : 3.00 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
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
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

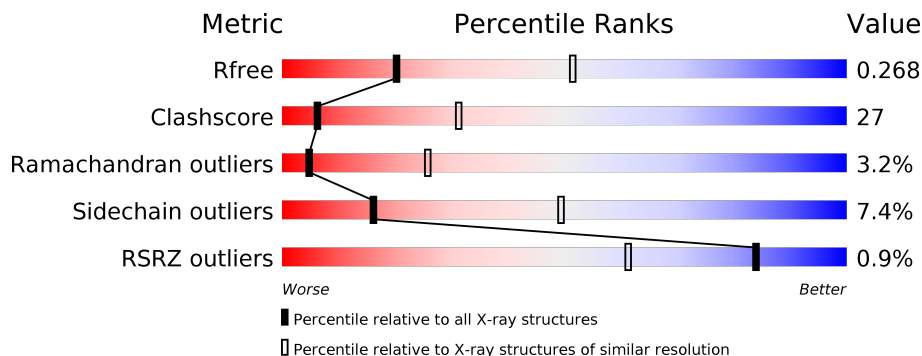
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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  $\geq 3$ , 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	560	<div> <div></div> <div> <div></div> <div>51%</div> <div>42%</div> <div>6%</div> <div></div> </div> </div>
2	B	427	<div> <div></div> <div> <div></div> <div>52%</div> <div>37%</div> <div>6%</div> <div>5%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7918 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 POL POLYPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	557	Total	C	N	O	S	2	0	0
			4522	2925	754	835	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	103	ASN	LYS	ENGINEERED	UNP P03366
A	478	GLN	GLU	ENGINEERED	UNP P03366

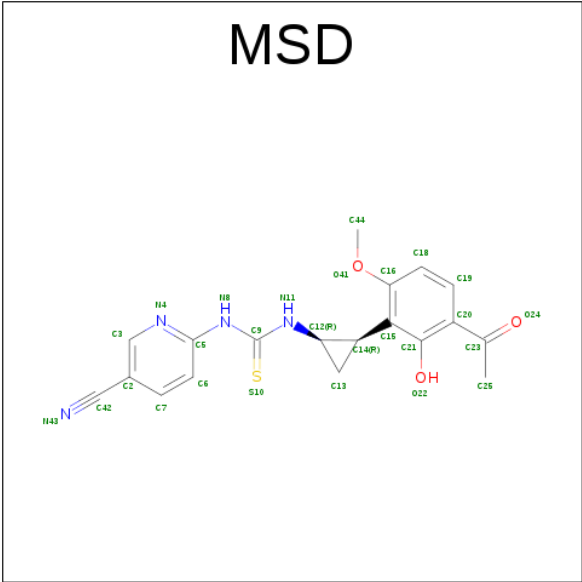
- Molecule 2 is a protein called POL POLYPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	405	Total	C	N	O	S	30	0	0
			3345	2178	550	611	6			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1103	ASN	LYS	ENGINEERED	UNP P03366

- Molecule 3 is 1-[2-(3-ACETYL-2-HYDROXY-6-METHOXY-PHENYL)-CYCLOPROPYL]-3-(5-CYANO-PYRIDIN-2-YL)-THIOUREA (three-letter code: MSD) (formula: C<sub>19</sub>H<sub>18</sub>N<sub>4</sub>O<sub>3</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	S	0	0
			27	19	4	3	1		

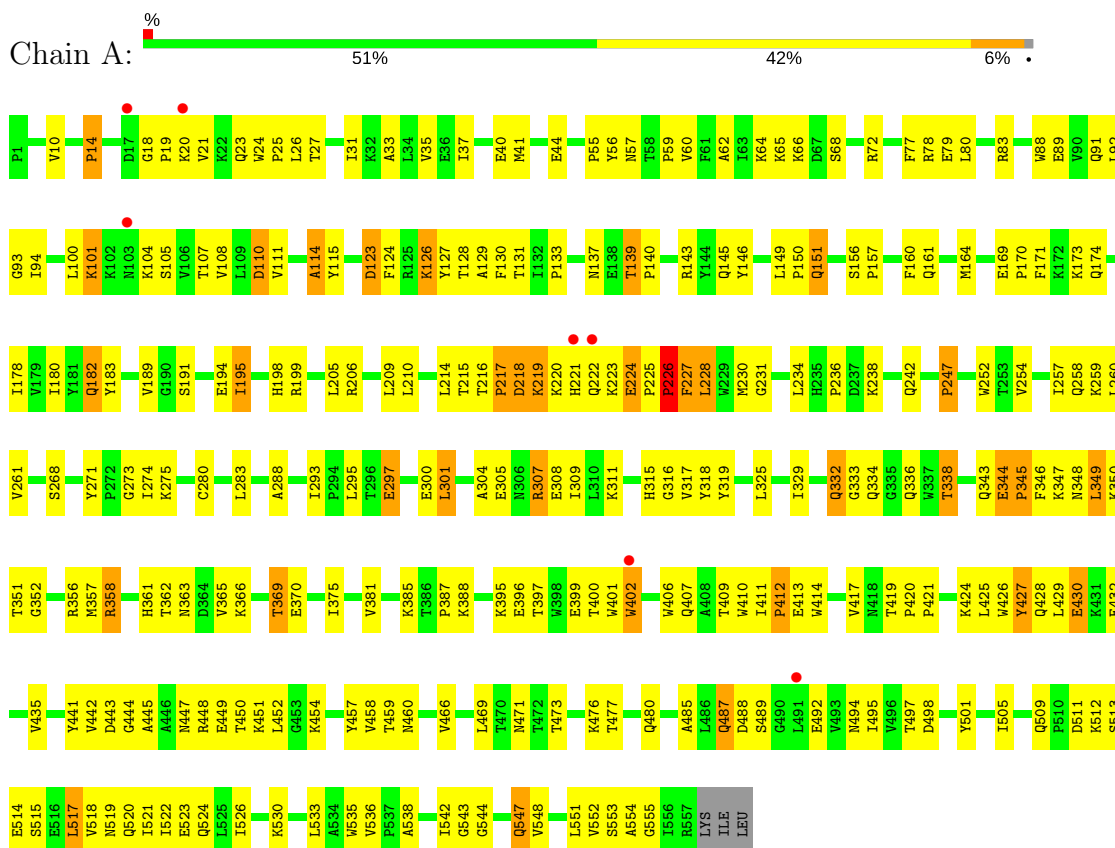
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	8	Total	O	0	0
			8	8		
4	B	16	Total	O	0	0
			16	16		

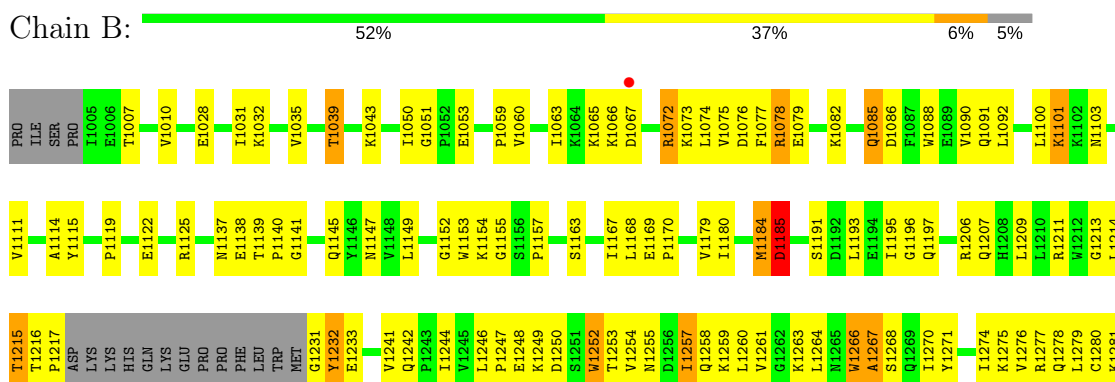
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: POL POLYPROTEIN



#### • Molecule 2: POL POLYPROTEIN



A1355	L1282
R1356	L1283
NET	R1284
ARG	G1285
GLY	T1286
ALA	K1287
HIS	A1288
T1362	L1289
N1363	T1290
D1364	E1291
V1365	V1292
	I1293
L1368	P1294
T1369	L1295
E1370	T1296
A1371	E1297
V1372	E1298
	A1299
I1375	E1300
	L1301
S1379	E1302
T1380	L1303
V1381	A1304
	E1305
P1387	N1306
	A1307
I1393	E1308
Q1394	T1309
K1395	L1310
W1398	V1314
E1399	
T1400	V1317
W1401	
	D1320
Q1407	
	K1323
W1410	D1324
	L1325
W1414	I1326
V1417	Q1330
N1418	K1331
T1419	Q1332
P1420	
P1421	Q1336
L1422	W1337
	T1338
	Y1339
L1425	Q1340
W1426	I1341
Y1427	Y1342
	Q1343
	E1344
	P1345
	F1346
	K1347
	K1353
	Y1354

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	120.34Å 156.54Å 156.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.78 – 3.00 45.78 – 2.94	Depositor EDS
% Data completeness (in resolution range)	90.8 (45.78-3.00) 86.0 (45.78-2.94)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.26 (at 2.96Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.214 , 0.273 0.206 , 0.268	Depositor DCC
$R_{free}$ test set	1374 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	69.4	Xtriage
Anisotropy	0.216	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 56.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	7918	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% 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

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

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	0/4640	0.65	0/6305
2	B	0.41	0/3439	0.65	0/4674
All	All	0.41	0/8079	0.65	0/10979

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	4522	0	4572	245	1
2	B	3345	0	3372	193	0
3	A	27	0	17	4	0
4	A	8	0	0	0	0
4	B	16	0	0	1	0
All	All	7918	0	7961	426	1

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 27.

The worst 5 of 426 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:131:THR:HG22	1:A:143:ARG:HG2	1.36	1.03
2:B:1215:THR:HG22	2:B:1216:THR:H	1.24	1.01
1:A:458:VAL:HG21	1:A:547:GLN:HE22	1.32	0.94
1:A:139:THR:HG22	1:A:140:PRO:HD2	1.53	0.90
1:A:435:VAL:HG13	2:B:1290:THR:HG21	1.55	0.89

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	Interatomic distance (Å)	Clash overlap (Å)
1:A:448:ARG:NH2	1:A:448:ARG:NH2[3_557]	1.64	0.56

## 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	555/560 (99%)	475 (86%)	60 (11%)	20 (4%)	4	22
2	B	399/427 (93%)	342 (86%)	46 (12%)	11 (3%)	6	29
All	All	954/987 (97%)	817 (86%)	106 (11%)	31 (3%)	5	26

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	PRO
2	B	1065	LYS
2	B	1085	GLN
2	B	1294	PRO
1	A	18	GLY

### 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	495/500 (99%)	456 (92%)	39 (8%)	14	46
2	B	369/389 (95%)	344 (93%)	25 (7%)	18	54
All	All	864/889 (97%)	800 (93%)	64 (7%)	16	49

5 of 64 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	338	THR
1	A	459	THR
2	B	1410	TRP
1	A	344	GLU
1	A	369	THR

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

Mol	Chain	Res	Type
1	A	524	GLN
1	A	539	HIS
2	B	1255	ASN
1	A	407	GLN
1	A	509	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	MSD	A	2000	-	27,29,29	4.07	18 (66%)	34,41,41	2.12	12 (35%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MSD	A	2000	-	-	0/20/25/25	0/2/3/3

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	2000	MSD	C9-S10	-15.37	1.29	1.68
3	A	2000	MSD	O41-C44	-5.58	1.26	1.42
3	A	2000	MSD	C16-C15	2.04	1.43	1.39
3	A	2000	MSD	C9-N11	2.18	1.38	1.34
3	A	2000	MSD	C3-N4	2.39	1.39	1.34

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2000	MSD	C5-N8-C9	-3.69	126.59	130.71
3	A	2000	MSD	C2-C3-N4	-3.49	118.72	123.71
3	A	2000	MSD	C13-C14-C15	-3.14	116.03	122.26
3	A	2000	MSD	C12-N11-C9	-2.52	121.22	125.80

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2000	MSD	O41-C16-C18	-2.10	120.84	124.37

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2000	MSD	4	0

## 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, 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	557/560 (99%)	-0.16	7 (1%)	77 51	20, 56, 92, 130	2 (0%)
2	B	405/427 (94%)	-0.38	2 (0%)	90 74	23, 49, 109, 133	9 (2%)
All	All	962/987 (97%)	-0.25	9 (0%)	84 61	20, 54, 105, 133	11 (1%)

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	222	GLN	2.6
2	B	1306	ASN	2.5
1	A	402	TRP	2.4
1	A	221	HIS	2.3
1	A	17	ASP	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](#)

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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	MSD	A	2000	27/27	0.96	0.24	-0.01	28,34,39,41	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.