



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 20, 2023 – 03:03 PM EDT

PDB ID : 5KIS  
Title : YenB/RHS2 complex  
Authors : Busby, J.N.; Hurst, M.R.H.; Lott, J.S.  
Deposited on : 2016-06-16  
Resolution : 2.40 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 : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.35.1  
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) : 2.35.1

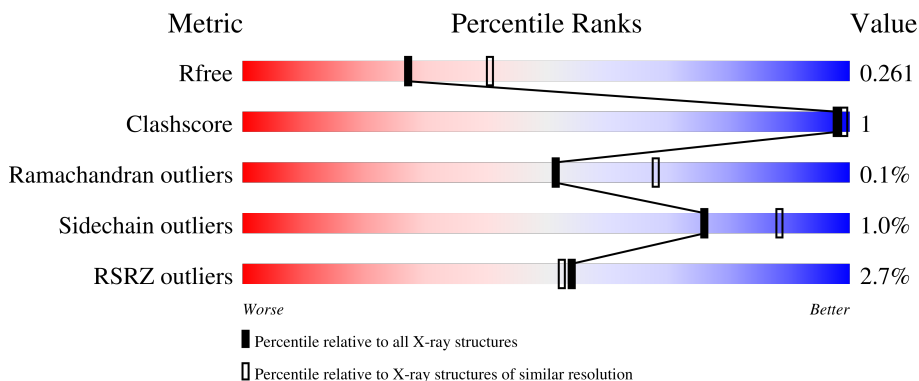
# 1 Overall quality at a glance

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 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	1491	
2	B	965	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 16842 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 YenB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1446	11438	7236	1966	2199	37	0	7	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP B6A880
A	-2	SER	-	expression tag	UNP B6A880
A	-1	GLY	-	expression tag	UNP B6A880
A	0	ALA	-	expression tag	UNP B6A880

- Molecule 2 is a protein called RHS2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	675	5177	3211	913	1048	5	0	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Ca	0	0
			2	2		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Na	0	0
			2	2		

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total Cl 2 2	0	0
5	B	2	Total Cl 2 2	0	0

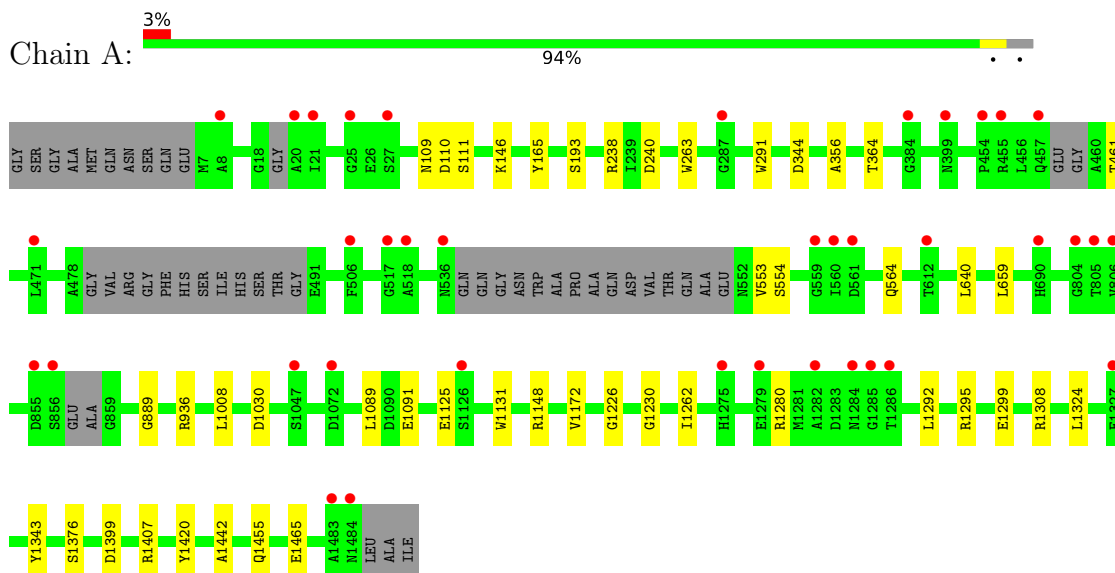
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	161	Total O 161 161	0	0
6	B	58	Total O 58 58	0	0

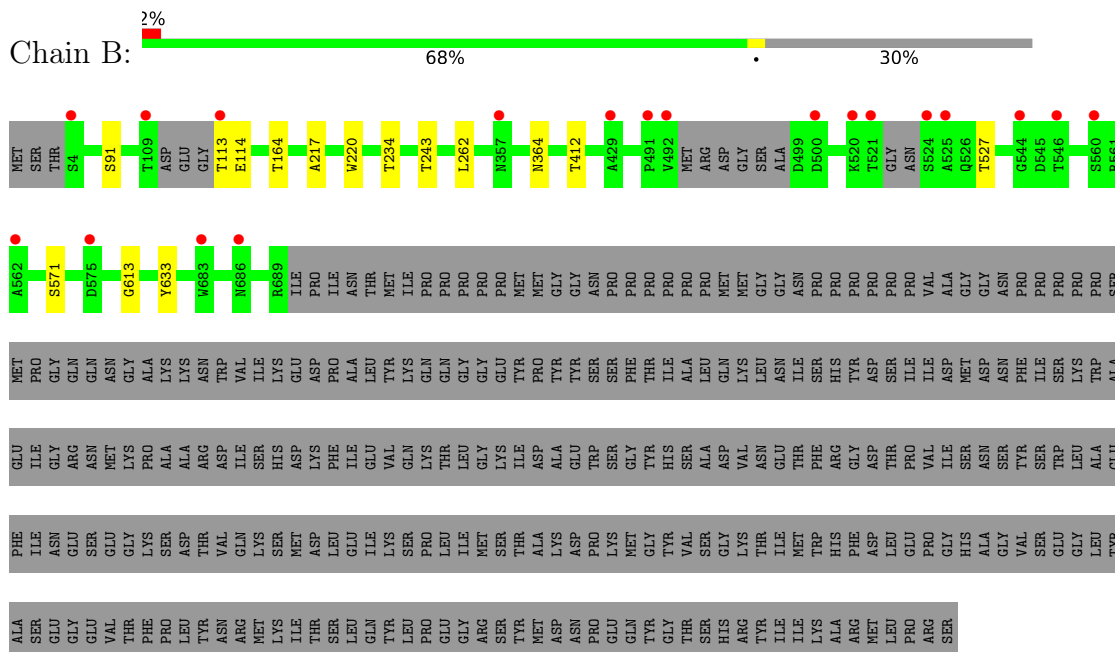
### 3 Residue-property plots [i](#)

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: YenB



- Molecule 2: RHS2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	148.61Å 132.76Å 155.36Å 90.00° 103.85° 90.00°	Depositor
Resolution (Å)	32.51 – 2.40 32.51 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.0 (32.51-2.40) 99.1 (32.51-2.40)	Depositor EDS
$R_{merge}$	0.30	Depositor
$R_{sym}$	0.30	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 2.39Å)	Xtrriage
Refinement program	REFMAC 5.8.0151	Depositor
R, $R_{free}$	0.216 , 0.260 0.219 , 0.261	Depositor DCC
$R_{free}$ test set	5641 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.8	Xtrriage
Anisotropy	0.149	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 27.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	16842	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.66% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, NA, CL

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	0.45	0/11761	0.68	3/16036 (0.0%)
2	B	0.46	0/5276	0.69	1/7200 (0.0%)
All	All	0.45	0/17037	0.68	4/23236 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1407	ARG	NE-CZ-NH1	5.54	123.07	120.30
1	A	238	ARG	NE-CZ-NH2	5.37	122.98	120.30
2	B	262	LEU	CA-CB-CG	5.23	127.32	115.30
1	A	1148	ARG	NE-CZ-NH2	5.14	122.87	120.30

There are no chirality outliers.

There are no planarity outliers.

### 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	11438	0	10838	19	0
2	B	5177	0	4917	5	0
3	A	2	0	0	0	0
4	A	2	0	0	0	0
5	A	2	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	2	0	0	0	0
6	A	161	0	0	0	0
6	B	58	0	0	0	0
All	All	16842	0	15755	23	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 1.

All (23) 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:1295[B]:ARG:HH21	1:A:1295[B]:ARG:HB2	1.34	0.92
1:A:640:LEU:HD13	1:A:659:LEU:HD21	1.72	0.71
1:A:553:VAL:HG12	1:A:554:SER:N	2.26	0.50
1:A:1292:LEU:HD11	1:A:1299:GLU:HA	1.94	0.49
1:A:1230:GLY:HA2	1:A:1343:TYR:CE1	2.49	0.47
1:A:1295[B]:ARG:HH21	1:A:1295[B]:ARG:CB	2.17	0.47
2:B:217:ALA:HA	2:B:220:TRP:CE3	2.50	0.46
1:A:1262:ILE:N	1:A:1262:ILE:HD12	2.32	0.45
1:A:1295[B]:ARG:HB2	1:A:1295[B]:ARG:NH2	2.16	0.45
1:A:344:ASP:HB2	1:A:356:ALA:HB3	1.99	0.44
1:A:889:GLY:O	1:A:936:ARG:NH1	2.50	0.44
1:A:1089:LEU:HB2	1:A:1131:TRP:HB2	1.99	0.44
1:A:1226:GLY:HA2	2:B:164:THR:CG2	2.48	0.44
2:B:613:GLY:HA2	2:B:633:TYR:CD1	2.52	0.44
1:A:1455:GLN:HE21	1:A:1465:GLU:HG2	1.83	0.44
1:A:109:ASN:OD1	1:A:111:SER:N	2.52	0.43
1:A:240:ASP:HB3	1:A:263:TRP:CG	2.54	0.43
2:B:113:THR:HG22	2:B:114:GLU:HG3	2.02	0.42
1:A:1280:ARG:HG3	1:A:1324:LEU:HD22	2.01	0.42
2:B:234:THR:HB	2:B:243:THR:OG1	2.21	0.41
1:A:146:LYS:HB3	1:A:165:TYR:HB2	2.01	0.41
1:A:1420:TYR:CD1	1:A:1442:ALA:HB2	2.56	0.41
1:A:1230:GLY:HA2	1:A:1343:TYR:CD1	2.56	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	1441/1491 (97%)	1402 (97%)	38 (3%)	1 (0%)	51	68
2	B	667/965 (69%)	645 (97%)	21 (3%)	1 (0%)	51	68
All	All	2108/2456 (86%)	2047 (97%)	59 (3%)	2 (0%)	51	68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1125	GLU
2	B	364	ASN

### 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	1214/1267 (96%)	1201 (99%)	13 (1%)	73	87
2	B	539/810 (66%)	535 (99%)	4 (1%)	84	92
All	All	1753/2077 (84%)	1736 (99%)	17 (1%)	76	88

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	110	ASP
1	A	193	SER
1	A	291	TRP
1	A	364	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	461	THR
1	A	564	GLN
1	A	1008	LEU
1	A	1030	ASP
1	A	1091	GLU
1	A	1172	VAL
1	A	1308	ARG
1	A	1376	SER
1	A	1399	ASP
2	B	91	SER
2	B	412	THR
2	B	527	THR
2	B	571	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	390	GLN
1	A	994	GLN
1	A	1001	GLN
1	A	1022	GLN
1	A	1024	ASN
1	A	1455	GLN

### 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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	1446/1491 (96%)	-0.17	38 (2%) 56 54	25, 41, 73, 106	0
2	B	675/965 (69%)	-0.20	19 (2%) 53 51	27, 44, 76, 104	0
All	All	2121/2456 (86%)	-0.18	57 (2%) 54 52	25, 42, 73, 106	0

All (57) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	560	ILE	6.7
1	A	806	VAL	5.2
1	A	561	ASP	4.7
1	A	856	SER	4.6
2	B	492	VAL	4.5
2	B	521	THR	4.5
2	B	109	THR	4.3
1	A	8	ALA	4.2
1	A	559	GLY	4.1
1	A	1126	SER	4.0
2	B	500	ASP	3.8
2	B	113	THR	3.8
1	A	1484	ASN	3.7
2	B	524	SER	3.6
1	A	518	ALA	3.3
1	A	1284	ASN	3.3
1	A	690	HIS	3.2
2	B	429	ALA	3.2
1	A	1282	ALA	3.2
2	B	546	THR	3.2
1	A	1285	GLY	3.2
1	A	27	SER	3.1
1	A	805	THR	3.1
1	A	454	PRO	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	560	SER	3.1
1	A	1275	HIS	2.9
1	A	455	ARG	2.8
1	A	1047	SER	2.8
2	B	357	ASN	2.8
1	A	612	THR	2.7
1	A	804	GLY	2.7
1	A	399	ASN	2.6
2	B	525	ALA	2.6
1	A	506	PHE	2.6
2	B	544	GLY	2.6
1	A	1327	GLU	2.5
2	B	575	ASP	2.5
1	A	471	LEU	2.4
2	B	520	LYS	2.4
1	A	1483	ALA	2.4
2	B	4	SER	2.4
1	A	287	GLY	2.4
2	B	683	TRP	2.4
1	A	384	GLY	2.4
1	A	855	ASP	2.4
1	A	457	GLN	2.4
1	A	1279	GLU	2.3
2	B	562	ALA	2.3
1	A	25	GLY	2.3
1	A	536	ASN	2.2
2	B	491	PRO	2.2
1	A	517	GLY	2.2
1	A	21	ILE	2.2
1	A	20	ALA	2.2
1	A	1072	ASP	2.1
2	B	686	ASN	2.1
1	A	1286	THR	2.1

## 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 monosaccharides 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NA	A	1503	1/1	0.81	0.11	65,65,65,65	0
5	CL	A	1505	1/1	0.94	0.05	58,58,58,58	0
3	CA	A	1506	1/1	0.96	0.12	50,50,50,50	0
5	CL	B	1001	1/1	0.96	0.07	70,70,70,70	0
3	CA	A	1501	1/1	0.97	0.13	48,48,48,48	0
5	CL	A	1504	1/1	0.97	0.06	39,39,39,39	0
4	NA	A	1502	1/1	0.98	0.12	28,28,28,28	0
5	CL	B	1002	1/1	0.99	0.10	46,46,46,46	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.