



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 29, 2017 – 08:12 AM EDT

PDB ID : 2J7N  
Title : Structure of the RNAi polymerase from *Neurospora crassa*  
Authors : Salgado, P.S.; Koivunen, M.R.L.; Makeyev, E.V.; Bamford, D.H.; Stuart, D.I.; Grimes, J.M.  
Deposited on : 2006-10-13  
Resolution : 2.30 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	:	rb-20029077
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-20029077

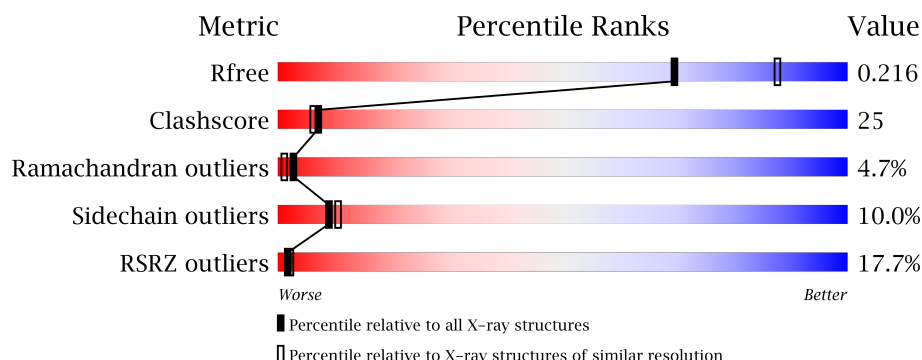
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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	1022	<div> <div>14%</div> <div> <div></div> <div>62%</div> <div>23%</div> <div>6%</div> <div>9%</div> </div> </div>
1	B	1022	<div> <div>19%</div> <div> <div></div> <div>57%</div> <div>24%</div> <div>8%</div> <div>9%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	GOL	B	3375	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15945 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 RNA-DEPENDENT RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	935	Total	C	N	O	S	0	0	1
			7520	4814	1304	1368	34			
1	B	932	Total	C	N	O	S	0	0	1
			7498	4798	1300	1366	34			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	559	ALA	GLY	conflict	UNP Q9Y7G6
B	559	ALA	GLY	conflict	UNP Q9Y7G6

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Mg	0	0
			1	1		
2	A	1	Total	Mg	0	0
			1	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			5	3	2		

- Molecule 4 is water.

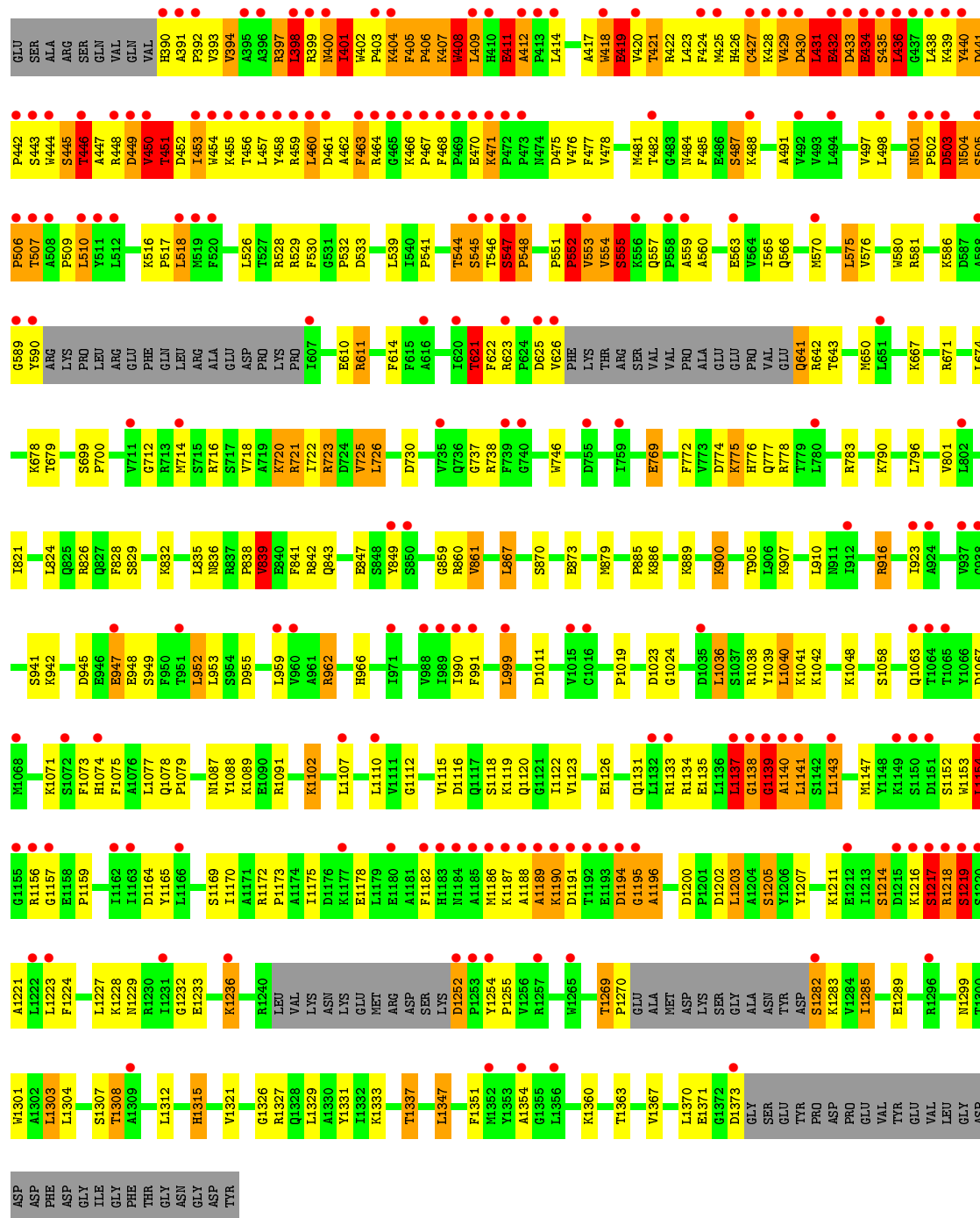
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	499	Total	O	0	0
			499	499		
4	B	421	Total	O	0	0
			421	421		



TYR  
PRO  
ASP  
PRO  
GLU  
VAL  
GLN  
TYR  
GLU  
VAL  
LEU  
GLY  
GLY  
ASP  
ASP  
PHE  
ASP  
GLY  
ILE  
GLY  
PHE  
THR  
GLY  
ASU  
GLY  
ASP  
TYR

• Molecule 1: RNA-DEPENDENT RNA POLYMERASE

Chain B: 19% 57% 24% 8% 9%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.02Å 122.55Å 114.70Å 90.00° 108.90° 90.00°	Depositor
Resolution (Å)	19.98 – 2.30 19.97 – 2.30	Depositor EDS
% Data completeness (in resolution range)	97.7 (19.98-2.30) 97.7 (19.97-2.30)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 2.30Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.217 , 0.264 0.218 , 0.216	Depositor DCC
$R_{free}$ test set	5728 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.5	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 52.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	15945	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 27.58 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.1432e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: GOL, MG

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	0.89	10/7713 (0.1%)	0.76	10/10439 (0.1%)
1	B	0.61	12/7689 (0.2%)	0.73	8/10407 (0.1%)
All	All	0.77	22/15402 (0.1%)	0.74	18/20846 (0.1%)

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	16
1	B	0	16
All	All	0	32

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	434	GLU	CD-OE2	50.04	1.80	1.25
1	A	434	GLU	CD-OE1	25.38	1.53	1.25
1	A	432	GLU	CD-OE1	16.68	1.44	1.25
1	A	436	LEU	C-N	16.11	1.62	1.33
1	B	435	SER	CB-OG	-13.76	1.24	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	434	GLU	OE1-CD-OE2	13.01	138.91	123.30
1	A	436	LEU	O-C-N	7.02	135.13	123.20
1	A	952	LEU	CA-CB-CG	6.91	131.20	115.30
1	A	1316	LYS	N-CA-C	-6.56	93.28	111.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1283	LYS	N-CA-C	-6.52	93.40	111.00

There are no chirality outliers.

5 of 32 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	456	THR	Peptide
1	A	504	ASN	Peptide
1	A	506	PRO	Peptide
1	A	545	SER	Peptide
1	A	547	SER	Peptide

## 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	7520	0	7468	365	1
1	B	7498	0	7440	401	1
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	B	5	0	5	1	0
4	A	499	0	0	20	2
4	B	421	0	0	22	2
All	All	15945	0	14913	757	3

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 25.

The worst 5 of 757 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:B:431:LEU:CB	1:B:432:GLU:HB2	1.34	1.51
1:B:407:LYS:CB	1:B:408:TRP:HB3	1.41	1.51
1:A:641:GLN:CB	1:A:642:ARG:HB2	1.50	1.41
1:B:407:LYS:HB2	1:B:408:TRP:CB	1.47	1.41
1:B:431:LEU:HB3	1:B:432:GLU:CB	1.48	1.40

All (3) 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)
4:A:2389:HOH:O	4:B:2333:HOH:O[1_455]	1.90	0.30
1:A:1156:ARG:NH1	1:B:1126:GLU:OE2[1_455]	2.07	0.13
4:A:2109:HOH:O	4:B:2384:HOH:O[2_746]	2.16	0.04

## 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	925/1022 (90%)	836 (90%)	51 (6%)	38 (4%)	3	1
1	B	922/1022 (90%)	811 (88%)	63 (7%)	48 (5%)	2	1
All	All	1847/2044 (90%)	1647 (89%)	114 (6%)	86 (5%)	3	1

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	399	ARG
1	A	412	ALA
1	A	439	LYS
1	A	440	TYR
1	A	459	ARG

### 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	815/891 (92%)	748 (92%)	67 (8%)	13	16
1	B	812/891 (91%)	716 (88%)	96 (12%)	6	6
All	All	1627/1782 (91%)	1464 (90%)	163 (10%)	9	10

5 of 163 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	409	LEU
1	B	518	LEU
1	B	1269	THR
1	B	419	GLU
1	B	449	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	501	ASN
1	B	566	GLN
1	B	822	ASN
1	B	390	HIS
1	B	1113	ASN

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

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	GOL	B	3375	-	4,4,5	0.49	0	2,4,5	0.89	0

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	GOL	B	3375	-	-	0/2/2/4	0/0/0/0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	3375	GOL	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	436:LEU	C	437:GLY	N	1.62

## 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	935/1022 (91%)	0.83	140 (14%) <b>3</b> <b>4</b>	39, 50, 65, 78	0
1	B	932/1022 (91%)	1.09	191 (20%) <b>1</b> <b>1</b>	38, 51, 64, 77	0
All	All	1867/2044 (91%)	0.96	331 (17%) <b>2</b> <b>2</b>	38, 51, 64, 78	0

The worst 5 of 331 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	590	TYR	12.6
1	B	1191	ASP	11.1
1	B	507	THR	10.2
1	B	466	LYS	9.6
1	A	626	VAL	9.3

### 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	GOL	B	3375	5/6	0.89	0.21	3.11	66,67,67,67	0
2	MG	B	3374	1/1	0.98	0.10	-	31,31,31,31	0
2	MG	A	3374	1/1	0.95	0.05	-	50,50,50,50	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.