



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 11:01 am BST

PDB ID : 5M3J  
Title : Influenza B polymerase bound to four heptad repeats of serine 5 phosphorylated Pol II CTD  
Authors : Lukarska, M.; Pflug, A.; Cusack, S.  
Deposited on : 2016-10-14  
Resolution : 3.50 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

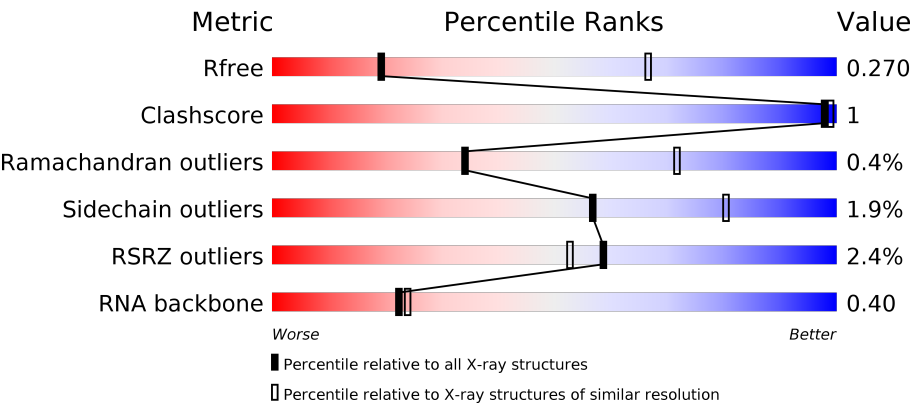
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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 <sub>free</sub>	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)
RNA backbone	3102	1002 (4.00-3.00)

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 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	751	<div><div></div><div><div></div><div></div><div></div><div></div></div><div>92% • 5%</div></div>
2	B	772	<div><div></div><div><div></div><div></div><div></div><div></div></div><div>92% • 5%</div></div>
3	C	798	<div><div>5%</div><div></div><div></div><div></div><div></div></div> <div>87% • 9%</div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	V	14	<div><div></div><div>79%</div><div>21%</div></div>
6	X	28	<div><div></div><div>32%</div><div>68%</div></div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 18013 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	716	5747	3651	962	1094	40	0	0	0

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	GLY	-	expression tag	UNP Q5V8Z9
A	-12	SER	-	expression tag	UNP Q5V8Z9
A	-11	HIS	-	expression tag	UNP Q5V8Z9
A	-10	HIS	-	expression tag	UNP Q5V8Z9
A	-9	HIS	-	expression tag	UNP Q5V8Z9
A	-8	HIS	-	expression tag	UNP Q5V8Z9
A	-7	HIS	-	expression tag	UNP Q5V8Z9
A	-6	HIS	-	expression tag	UNP Q5V8Z9
A	-5	HIS	-	expression tag	UNP Q5V8Z9
A	-4	HIS	-	expression tag	UNP Q5V8Z9
A	-3	GLY	-	expression tag	UNP Q5V8Z9
A	-2	SER	-	expression tag	UNP Q5V8Z9
A	-1	GLY	-	expression tag	UNP Q5V8Z9
A	0	SER	-	expression tag	UNP Q5V8Z9
A	727	GLY	-	expression tag	UNP Q5V8Z9
A	728	SER	-	expression tag	UNP Q5V8Z9
A	729	GLY	-	expression tag	UNP Q5V8Z9
A	730	SER	-	expression tag	UNP Q5V8Z9
A	731	GLY	-	expression tag	UNP Q5V8Z9
A	732	GLU	-	expression tag	UNP Q5V8Z9
A	733	ASN	-	expression tag	UNP Q5V8Z9
A	734	LEU	-	expression tag	UNP Q5V8Z9
A	735	TYR	-	expression tag	UNP Q5V8Z9
A	736	PHE	-	expression tag	UNP Q5V8Z9
A	737	GLN	-	expression tag	UNP Q5V8Z9

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	733	Total	C	N	O	S	0	0	0
			5761	3636	999	1074	52			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-8	GLY	-	expression tag	UNP Q5V8Y6
B	-7	SER	-	expression tag	UNP Q5V8Y6
B	-6	GLY	-	expression tag	UNP Q5V8Y6
B	-5	SER	-	expression tag	UNP Q5V8Y6
B	-4	GLY	-	expression tag	UNP Q5V8Y6
B	-3	SER	-	expression tag	UNP Q5V8Y6
B	-2	GLY	-	expression tag	UNP Q5V8Y6
B	-1	SER	-	expression tag	UNP Q5V8Y6
B	0	GLY	-	expression tag	UNP Q5V8Y6
B	753	GLY	-	expression tag	UNP Q5V8Y6
B	754	SER	-	expression tag	UNP Q5V8Y6
B	755	GLY	-	expression tag	UNP Q5V8Y6
B	756	SER	-	expression tag	UNP Q5V8Y6
B	757	GLY	-	expression tag	UNP Q5V8Y6
B	758	GLU	-	expression tag	UNP Q5V8Y6
B	759	ASN	-	expression tag	UNP Q5V8Y6
B	760	LEU	-	expression tag	UNP Q5V8Y6
B	761	TYR	-	expression tag	UNP Q5V8Y6
B	762	PHE	-	expression tag	UNP Q5V8Y6
B	763	GLN	-	expression tag	UNP Q5V8Y6

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	730	Total	C	N	O	S	0	0	0
			5843	3715	1023	1065	40			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-8	GLY	-	expression tag	UNP Q5V8X3
C	-7	SER	-	expression tag	UNP Q5V8X3
C	-6	GLY	-	expression tag	UNP Q5V8X3
C	-5	SER	-	expression tag	UNP Q5V8X3
C	-4	GLY	-	expression tag	UNP Q5V8X3
C	-3	SER	-	expression tag	UNP Q5V8X3
C	-2	GLY	-	expression tag	UNP Q5V8X3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	SER	-	expression tag	UNP Q5V8X3
C	0	GLY	-	expression tag	UNP Q5V8X3
C	771	GLY	-	expression tag	UNP Q5V8X3
C	772	TRP	-	expression tag	UNP Q5V8X3
C	773	SER	-	expression tag	UNP Q5V8X3
C	774	HIS	-	expression tag	UNP Q5V8X3
C	775	PRO	-	expression tag	UNP Q5V8X3
C	776	GLN	-	expression tag	UNP Q5V8X3
C	777	PHE	-	expression tag	UNP Q5V8X3
C	778	GLU	-	expression tag	UNP Q5V8X3
C	779	LYS	-	expression tag	UNP Q5V8X3
C	780	GLY	-	expression tag	UNP Q5V8X3
C	781	SER	-	expression tag	UNP Q5V8X3
C	782	GLY	-	expression tag	UNP Q5V8X3
C	783	SER	-	expression tag	UNP Q5V8X3
C	784	GLU	-	expression tag	UNP Q5V8X3
C	785	ASN	-	expression tag	UNP Q5V8X3
C	786	LEU	-	expression tag	UNP Q5V8X3
C	787	TYR	-	expression tag	UNP Q5V8X3
C	788	PHE	-	expression tag	UNP Q5V8X3
C	789	GLN	-	expression tag	UNP Q5V8X3

- Molecule 4 is a RNA chain called RNA (5'-R(\*UP\*AP\*UP\*AP\*CP\*CP\*UP\*CP\*UP\*GP\*CP\*UP\*UP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	R	14	Total	C	N	O	P	0	0	0
			282	128	42	99	13			

- Molecule 5 is a RNA chain called RNA (5'-R(P\*AP\*GP\*UP\*AP\*GP\*UP\*AP\*AP\*CP\*AP\*AP\*GP\*AP\*G)-3').

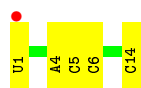
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	V	14	Total	C	N	O	P	0	0	0
			307	137	62	94	14			

- Molecule 6 is a protein called DNA-directed RNA polymerase subunit.

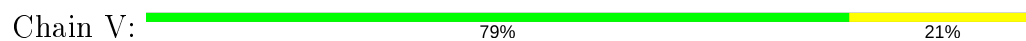
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	X	9	Total	C	N	O	P	0	0	0
			73	44	9	19	1			



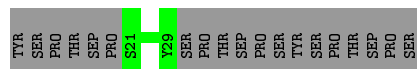
- Molecule 4: RNA (5'-R(\*UP\*AP\*UP\*AP\*CP\*CP\*UP\*CP\*UP\*GP\*CP\*UP\*UP\*C)-3')



- Molecule 5: RNA (5'-R(P\*AP\*GP\*UP\*AP\*GP\*UP\*AP\*AP\*CP\*AP\*AP\*GP\*AP\*G)-3')



- Molecule 6: DNA-directed RNA polymerase subunit





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	200.16Å 200.16Å 250.38Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.07 – 3.50 49.07 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.07-3.50) 100.0 (49.07-3.50)	Depositor EDS
$R_{merge}$	0.19	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.55 (at 3.48Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, $R_{free}$	0.239 , 0.268 0.237 , 0.270	Depositor DCC
$R_{free}$ test set	3523 reflections (4.80%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	119.5	Xtriage
Anisotropy	0.546	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 96.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.066 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	18013	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	169.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.25% 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: SEP

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	0.24	0/5863	0.40	0/7907
2	B	0.25	0/5873	0.41	0/7915
3	C	0.25	0/5943	0.41	0/7987
4	R	0.23	0/312	0.75	0/482
5	V	0.61	1/345 (0.3%)	0.67	0/535
6	X	0.25	0/65	0.35	0/87
All	All	0.26	1/18401 (0.0%)	0.42	0/24913

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	V	1	A	OP3-P	-10.69	1.48	1.61

There are no bond angle outliers.

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	5747	0	5719	9	0
2	B	5761	0	5783	8	0
3	C	5843	0	6015	14	0
4	R	282	0	146	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	V	307	0	153	0	0
6	X	73	0	56	0	0
All	All	18013	0	17872	25	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.

The worst 5 of 25 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)
3:C:565:PHE:O	3:C:605:ARG:NH1	2.35	0.59
2:B:195:ASN:O	2:B:197:LYS:N	2.35	0.59
1:A:509:GLY:N	1:A:512:ASP:OD2	2.38	0.57
3:C:565:PHE:CE2	3:C:605:ARG:HB2	2.44	0.53
3:C:567:LEU:O	3:C:605:ARG:NH2	2.38	0.52

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	712/751 (95%)	667 (94%)	43 (6%)	2 (0%)	41	75
2	B	729/772 (94%)	686 (94%)	41 (6%)	2 (0%)	41	75
3	C	726/798 (91%)	681 (94%)	41 (6%)	4 (1%)	25	64
6	X	6/28 (21%)	4 (67%)	2 (33%)	0	100	100
All	All	2173/2349 (92%)	2038 (94%)	127 (6%)	8 (0%)	34	72

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	569	LYS
2	B	196	ARG
3	C	307	ASP
3	C	730	LEU
3	C	292	ASP

### 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	637/664 (96%)	627 (98%)	10 (2%)	62	83
2	B	631/657 (96%)	617 (98%)	14 (2%)	52	78
3	C	638/694 (92%)	625 (98%)	13 (2%)	55	79
6	X	8/24 (33%)	8 (100%)	0	100	100
All	All	1914/2039 (94%)	1877 (98%)	37 (2%)	57	80

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	327	ARG
2	B	616	ASP
3	C	610	MET
2	B	399	GLU
2	B	420	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	R	13/14 (92%)	4 (30%)	1 (7%)
5	V	13/14 (92%)	2 (15%)	0
All	All	26/28 (92%)	6 (23%)	1 (3%)

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	R	4	A
4	R	5	C
4	R	6	C
4	R	14	C
5	V	7	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
4	R	1	U

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

1 non-standard protein/DNA/RNA residue 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	$\# Z  > 2$
6	SEP	X	26	6	8,9,10	1.58	1 (12%)	8,12,14	1.73	2 (25%)

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
6	SEP	X	26	6	-	1/5/8/10	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	X	26	SEP	P-O1P	3.45	1.61	1.50

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	X	26	SEP	OG-CB-CA	3.82	111.86	108.14
6	X	26	SEP	P-OG-CB	-2.72	110.79	118.30

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	X	26	SEP	N-CA-CB-OG

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [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	716/751 (95%)	-0.11	11 (1%) 73 68	109, 162, 262, 353	0
2	B	733/772 (94%)	-0.13	5 (0%) 87 83	105, 150, 198, 247	0
3	C	730/798 (91%)	0.09	36 (4%) 29 26	104, 170, 269, 355	0
4	R	14/14 (100%)	0.24	1 (7%) 16 15	142, 167, 202, 217	0
5	V	14/14 (100%)	-0.16	0 100 100	129, 142, 149, 155	0
6	X	8/28 (28%)	-0.61	0 100 100	138, 167, 200, 203	0
All	All	2215/2377 (93%)	-0.05	53 (2%) 59 53	104, 159, 254, 355	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	425	ARG	8.9
3	C	421	ASN	5.2
3	C	420	ILE	4.6
3	C	386	LEU	4.3
1	A	148	PHE	4.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [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(Å <sup>2</sup> )	Q<0.9
6	SEP	X	26	10/11	0.94	0.08	153,156,160,161	0

### 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.