



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2017 – 05:10 PM EST

PDB ID : 5MHS  
Title : T1L reovirus sigma1 complexed with 5C6 Fab fragments  
Authors : Stehle, T.; Dietrich, M.H.  
Deposited on : 2016-11-25  
Resolution : 3.70 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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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-20028442

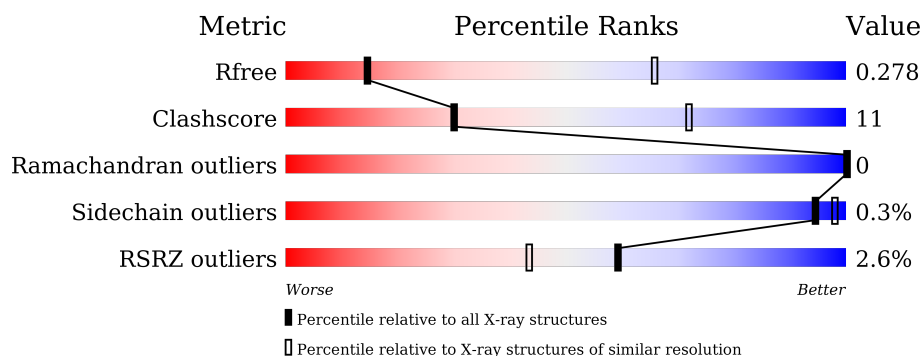
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.70 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))
$R_{free}$	91344	1101 (3.90-3.50)
Clashscore	102246	1224 (3.90-3.50)
Ramachandran outliers	100387	1172 (3.90-3.50)
Sidechain outliers	100360	1170 (3.90-3.50)
RSRZ outliers	91569	1108 (3.90-3.50)

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	172	<div> <div>83%</div> <div>12%</div> <div>5%</div> </div>
1	B	172	<div> <div>74%</div> <div>21%</div> <div>5%</div> </div>
1	C	172	<div> <div>83%</div> <div>12%</div> <div>5%</div> </div>
2	D	217	<div> <div>2%</div> <div>81%</div> <div>18%</div> <div>.</div> </div>
2	F	217	<div> <div>6%</div> <div>74%</div> <div>23%</div> <div>.</div> </div>
2	G	217	<div> <div>2%</div> <div>75%</div> <div>20%</div> <div>6%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	221	<div><div>%</div><div><div></div><div>77%</div><div>17%</div><div>5%</div></div></div>
3	H	221	<div><div>7%</div><div><div></div><div>71%</div><div>20%</div><div>9%</div></div></div>
3	N	221	<div><div>3%</div><div><div></div><div>75%</div><div>18%</div><div>7%</div></div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11874 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 Outer capsid protein sigma-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	163	Total	C	N	O	S	0	0	0
			1252	795	210	240	7			
1	B	163	Total	C	N	O	S	0	0	0
			1258	804	214	234	6			
1	C	163	Total	C	N	O	S	0	0	0
			1283	819	213	244	7			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	299	GLY	-	expression tag	UNP P04506
A	300	VAL	-	expression tag	UNP P04506
A	301	LEU	-	expression tag	UNP P04506
A	302	ASN	-	expression tag	UNP P04506
A	303	GLN	-	expression tag	UNP P04506
A	304	GLY	-	expression tag	UNP P04506
A	305	VAL	-	expression tag	UNP P04506
A	306	THR	-	expression tag	UNP P04506
A	307	SER	THR	conflict	UNP P04506
B	299	GLY	-	expression tag	UNP P04506
B	300	VAL	-	expression tag	UNP P04506
B	301	LEU	-	expression tag	UNP P04506
B	302	ASN	-	expression tag	UNP P04506
B	303	GLN	-	expression tag	UNP P04506
B	304	GLY	-	expression tag	UNP P04506
B	305	VAL	-	expression tag	UNP P04506
B	306	THR	-	expression tag	UNP P04506
B	307	SER	THR	conflict	UNP P04506
C	299	GLY	-	expression tag	UNP P04506
C	300	VAL	-	expression tag	UNP P04506
C	301	LEU	-	expression tag	UNP P04506
C	302	ASN	-	expression tag	UNP P04506
C	303	GLN	-	expression tag	UNP P04506

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	304	GLY	-	expression tag	UNP P04506
C	305	VAL	-	expression tag	UNP P04506
C	306	THR	-	expression tag	UNP P04506
C	307	SER	THR	conflict	UNP P04506

- Molecule 2 is a protein called 5C6 Fab light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	214	Total	C	N	O	S	0	0	0
			1392	870	244	272	6			
2	F	212	Total	C	N	O	S	0	0	0
			1375	852	241	276	6			
2	G	205	Total	C	N	O	S	0	0	0
			1352	843	240	264	5			


- Molecule 3 is a protein called 5C6 Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	210	Total	C	N	O	S	0	0	0
			1332	827	235	265	5			
3	H	201	Total	C	N	O	S	0	0	0
			1271	799	221	247	4			
3	N	205	Total	C	N	O	S	0	0	0
			1359	860	231	263	5			

### 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 errors 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: Outer capsid protein sigma-1

Chain A: 




- Molecule 1: Outer capsid protein sigma-1

Chain B: 




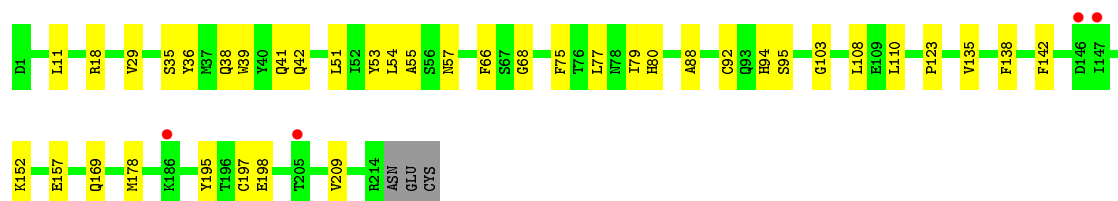
- Molecule 1: Outer capsid protein sigma-1

Chain C: 



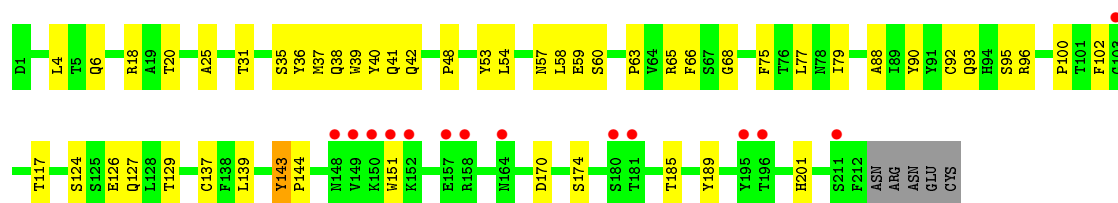
- Molecule 2: 5C6 Fab light chain

Chain D: 

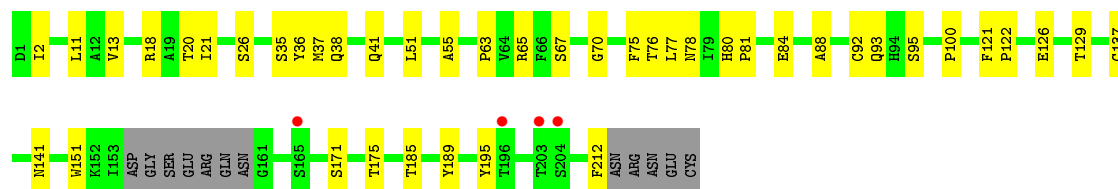
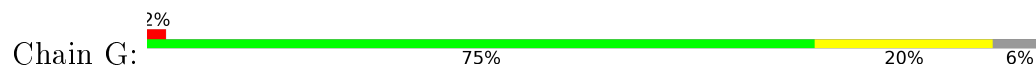


- Molecule 2: 5C6 Fab light chain

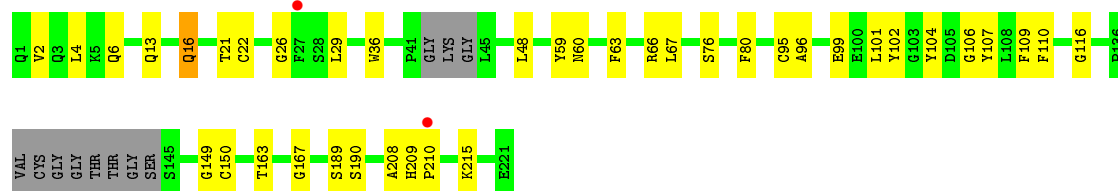
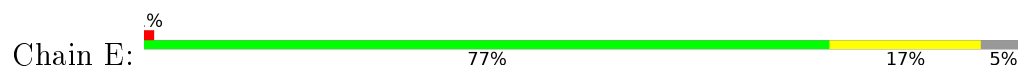
Chain F: 



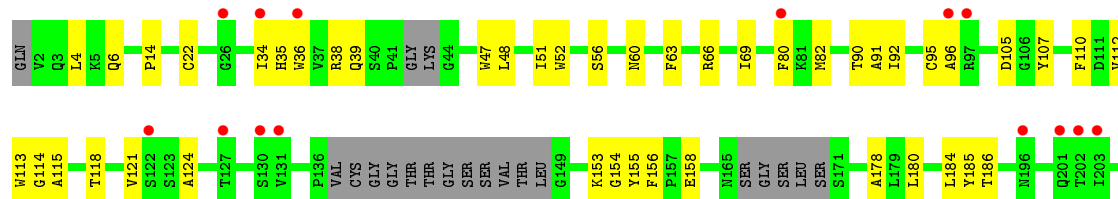
• Molecule 2: 5C6 Fab light chain



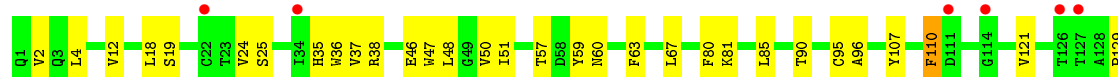
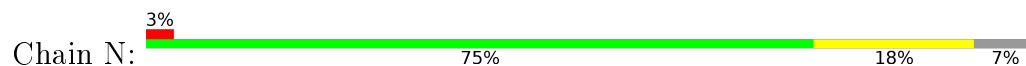
• Molecule 3: 5C6 Fab heavy chain

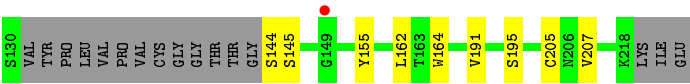


• Molecule 3: 5C6 Fab heavy chain



• Molecule 3: 5C6 Fab heavy chain







## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	223.61Å 129.08Å 87.86Å 90.00° 101.24° 90.00°	Depositor
Resolution (Å)	48.99 – 3.70 48.99 – 3.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (48.99-3.70) 100.0 (48.99-3.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.49 (at 3.67Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, $R_{free}$	0.261 , 0.278 0.260 , 0.278	Depositor DCC
$R_{free}$ test set	1312 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	112.9	Xtriage
Anisotropy	0.497	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 112.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	11874	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	143.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.17% 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](#)

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	0.25	0/1288	0.50	0/1765
1	B	0.27	0/1294	0.53	0/1769
1	C	0.27	0/1319	0.50	0/1805
2	D	0.27	0/1427	0.52	0/1968
2	F	0.28	0/1407	0.51	0/1943
2	G	0.26	0/1384	0.50	0/1908
3	E	0.33	0/1365	0.56	0/1888
3	H	0.30	0/1303	0.54	0/1799
3	N	0.33	0/1394	0.59	0/1922
All	All	0.29	0/12181	0.53	0/16767

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 [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	1252	0	1128	14	1
1	B	1258	0	1156	26	0
1	C	1283	0	1206	16	1
2	D	1392	0	1072	33	0
2	F	1375	0	1060	40	0
2	G	1352	0	1065	27	0
3	E	1332	0	947	36	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	H	1271	0	902	34	1
3	N	1359	0	1065	32	1
All	All	11874	0	9601	238	2

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 11.

The worst 5 of 238 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:E:6:GLN:OE1	3:E:6:GLN:N	2.10	0.83
3:H:4:LEU:HB3	3:H:114:GLY:HA3	1.62	0.80
1:B:426:ASP:OD1	3:E:104:TYR:OH	2.03	0.77
3:H:4:LEU:HD21	3:H:22:CYS:HB3	1.67	0.77
2:G:36:TYR:HB3	2:G:95:SER:HB2	1.69	0.75

All (2) 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:422:GLY:O	3:H:107:TYR:OH[2_756]	2.17	0.03
1:C:422:GLY:O	3:N:107:TYR:OH[4_756]	2.18	0.02

## 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	161/172 (94%)	160 (99%)	1 (1%)	0	100	100
1	B	161/172 (94%)	160 (99%)	1 (1%)	0	100	100
1	C	161/172 (94%)	160 (99%)	1 (1%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	212/217 (98%)	208 (98%)	4 (2%)	0	100	100
2	F	210/217 (97%)	205 (98%)	5 (2%)	0	100	100
2	G	201/217 (93%)	195 (97%)	6 (3%)	0	100	100
3	E	204/221 (92%)	200 (98%)	4 (2%)	0	100	100
3	H	193/221 (87%)	190 (98%)	3 (2%)	0	100	100
3	N	201/221 (91%)	196 (98%)	5 (2%)	0	100	100
All	All	1704/1830 (93%)	1674 (98%)	30 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	126/149 (85%)	126 (100%)	0	100	100
1	B	126/149 (85%)	126 (100%)	0	100	100
1	C	137/149 (92%)	137 (100%)	0	100	100
2	D	104/191 (54%)	104 (100%)	0	100	100
2	F	105/191 (55%)	104 (99%)	1 (1%)	82	92
2	G	105/191 (55%)	105 (100%)	0	100	100
3	E	90/190 (47%)	89 (99%)	1 (1%)	80	92
3	H	83/190 (44%)	83 (100%)	0	100	100
3	N	104/190 (55%)	103 (99%)	1 (1%)	82	92
All	All	980/1590 (62%)	977 (100%)	3 (0%)	94	98

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	16	GLN
2	F	143	TYR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	N	110	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	455	GLN
2	F	38	GLN
2	F	93	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 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	163/172 (94%)	-0.20	0 100 100	67, 95, 152, 172	0
1	B	163/172 (94%)	-0.25	0 100 100	64, 92, 145, 172	0
1	C	163/172 (94%)	-0.23	0 100 100	57, 88, 150, 185	0
2	D	214/217 (98%)	-0.05	4 (1%) 70 55	86, 166, 239, 278	0
2	F	212/217 (97%)	0.24	14 (6%) 22 12	93, 167, 266, 283	0
2	G	205/217 (94%)	-0.15	4 (1%) 68 53	76, 142, 258, 328	0
3	E	210/221 (95%)	-0.28	2 (0%) 84 72	80, 158, 201, 217	0
3	H	201/221 (90%)	0.23	15 (7%) 17 11	90, 190, 255, 297	0
3	N	205/221 (92%)	-0.09	7 (3%) 49 34	77, 148, 241, 259	0
All	All	1736/1830 (94%)	-0.08	46 (2%) 59 43	57, 137, 247, 328	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	148	ASN	7.1
2	F	157	GLU	6.7
2	F	196	THR	6.4
2	F	152	LYS	5.7
3	H	203	ILE	5.2

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.