



# Full wwPDB X-ray Structure Validation Report ⓘ

May 17, 2020 – 05:43 am BST

PDB ID : 5IBM  
Title : Structure of S502P, a Cancer-Associated Mutation of the Oncogenic Phosphatase SHP2  
Authors : Blacklow, S.C.; Stams, T.; Fodor, M.; LaRochelle, J.R.  
Deposited on : 2016-02-22  
Resolution : 2.18 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 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	:	2.11
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

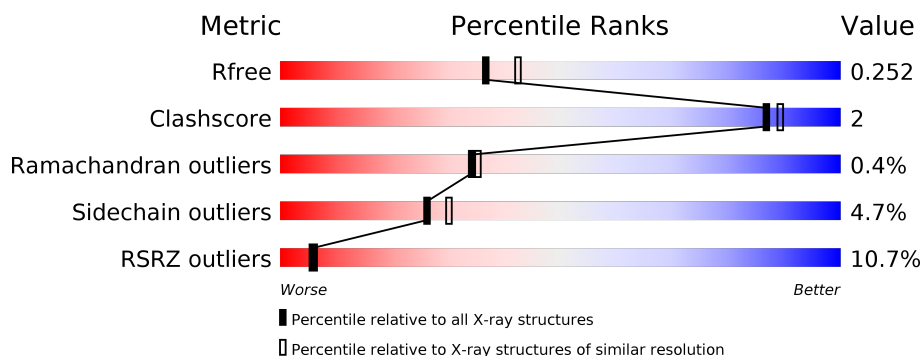
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.18 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	6864 (2.20-2.16)
Clashscore	141614	7689 (2.20-2.16)
Ramachandran outliers	138981	7564 (2.20-2.16)
Sidechain outliers	138945	7564 (2.20-2.16)
RSRZ outliers	127900	6738 (2.20-2.16)

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	526	<div> <div>10%</div> <div>84%</div> <div>7%</div> <div>9%</div> </div>
1	B	526	<div> <div>10%</div> <div>80%</div> <div>11%</div> <div>8%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8166 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 Tyrosine-protein phosphatase non-receptor type 11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	480	Total	C	N	O	S	0	1	0
			3899	2466	690	724	19			
1	B	482	Total	C	N	O	S	0	0	0
			3900	2463	692	727	18			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP Q06124
A	502	PRO	SER	engineered mutation	UNP Q06124
B	0	SER	-	expression tag	UNP Q06124
B	502	PRO	SER	engineered mutation	UNP Q06124

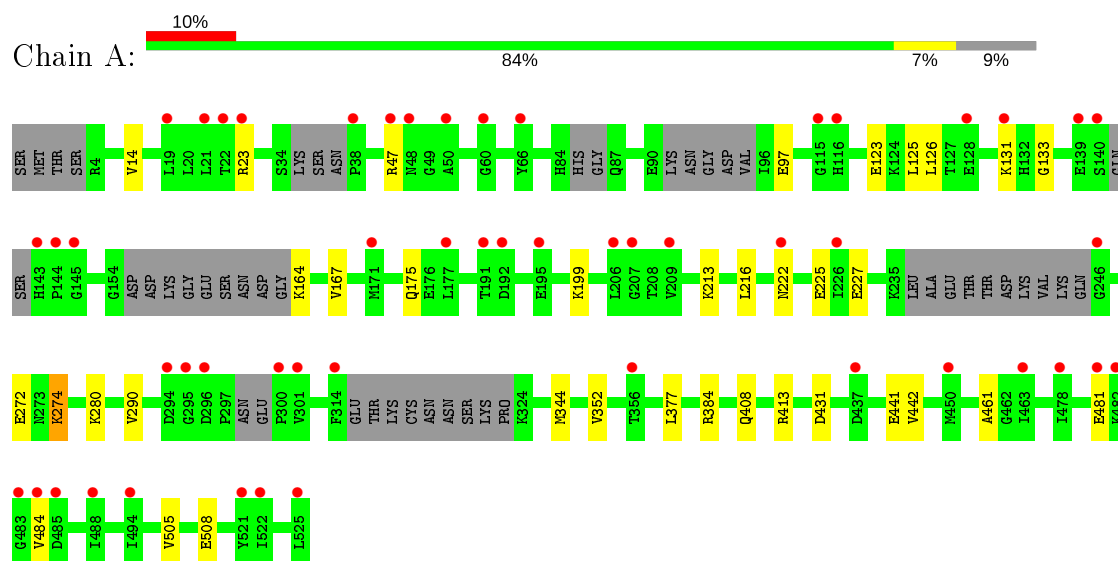
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	191	Total	O	0	0
			191	191		
2	B	176	Total	O	0	0
			176	176		

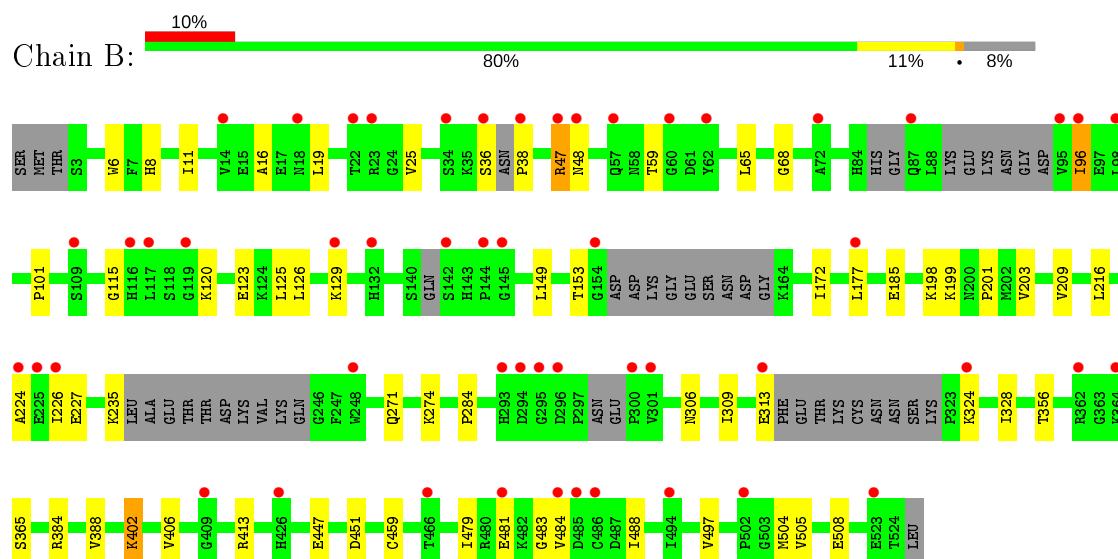
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Tyrosine-protein phosphatase non-receptor type 11



- Molecule 1: Tyrosine-protein phosphatase non-receptor type 11



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.19 Å   214.24 Å   55.47 Å 90.00°   95.55°   90.00°	Depositor
Resolution (Å)	43.60 – 2.18 41.47 – 2.18	Depositor EDS
% Data completeness (in resolution range)	99.0 (43.60-2.18) 99.1 (41.47-2.18)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.39 (at 2.18 Å)	Xtriage
Refinement program	BUSTER 2.11.5	Depositor
R, $R_{free}$	0.189   ,   0.240 0.197   ,   0.252	Depositor DCC
$R_{free}$ test set	2651 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	37.2	Xtriage
Anisotropy	0.471	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 52.9	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	0.95	EDS
Total number of atoms	8166	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.80% 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

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	0.51	0/3982	0.67	0/5365
1	B	0.51	0/3980	0.68	0/5363
All	All	0.51	0/7962	0.68	0/10728

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	3899	0	3847	11	0
1	B	3900	0	3848	22	0
2	A	191	0	0	1	0
2	B	176	0	0	0	0
All	All	8166	0	7695	32	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 2.

All (32) 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:65:LEU:HD12	1:B:68:GLY:HA3	1.66	0.78

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:126:LEU:HD23	1:A:216:LEU:HD13	1.82	0.60
1:B:47:ARG:HH11	1:B:48:ASN:HD22	1.50	0.59
1:B:271:GLN:HA	1:B:274:LYS:HE2	1.84	0.59
1:A:133:GLY:HA3	1:A:213:LYS:HB2	1.84	0.58
1:B:125:LEU:HB3	1:B:216:LEU:HD21	1.88	0.54
1:A:125:LEU:HB3	1:A:216:LEU:HD21	1.89	0.53
1:B:126:LEU:HD23	1:B:216:LEU:HD13	1.92	0.51
1:B:309:ILE:HD13	1:B:328:ILE:HG12	1.93	0.51
1:B:203:VAL:HG22	1:B:209:VAL:HG22	1.94	0.49
1:A:222:ASN:O	1:A:225:GLU:HB2	2.14	0.48
1:B:388:VAL:HG21	1:B:402:LYS:HE3	1.96	0.48
1:B:384:ARG:HB2	1:B:406:VAL:HG12	1.95	0.48
1:B:11:ILE:HD12	1:B:16:ALA:HB2	1.96	0.48
1:A:352:VAL:HG11	1:A:442:VAL:HG13	1.97	0.46
1:B:149:LEU:HB2	1:B:172:ILE:HD11	1.98	0.46
1:A:123:GLU:HG2	1:A:167:VAL:HG11	1.99	0.44
1:A:175:GLN:HB3	1:B:25:VAL:HG11	1.99	0.44
1:B:479:ILE:O	1:B:483:GLY:HA2	2.18	0.44
1:B:47:ARG:HG2	1:B:96:ILE:CD1	2.48	0.44
1:B:36:SER:O	1:B:38:PRO:HD3	2.19	0.43
1:A:290:VAL:HG11	1:A:344:MET:HG3	2.01	0.43
1:B:6:TRP:HB3	1:B:101:PRO:HB3	2.00	0.43
1:B:199:LYS:O	1:B:201:PRO:HD3	2.18	0.42
1:A:377:LEU:HD11	1:A:384:ARG:HG2	2.01	0.42
1:B:497:VAL:HG12	1:B:504:MET:HG3	2.00	0.42
1:A:274:LYS:HD3	1:A:280:LYS:HB2	2.00	0.42
1:A:461:ALA:HB3	2:A:632:HOH:O	2.20	0.42
1:B:224:ALA:HB2	1:B:484:VAL:CG2	2.50	0.41
1:B:8:HIS:HB3	1:B:11:ILE:HG12	2.02	0.41
1:B:284:PRO:HG3	1:B:306:ASN:HA	2.03	0.41
1:B:356:THR:OG1	1:B:459:CYS:HB3	2.21	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	463/526 (88%)	450 (97%)	12 (3%)	1 (0%)	47	52
1	B	464/526 (88%)	444 (96%)	17 (4%)	3 (1%)	25	24
All	All	927/1052 (88%)	894 (96%)	29 (3%)	4 (0%)	34	35

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	324	LYS
1	A	505	VAL
1	B	115	GLY
1	B	505	VAL

### 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	428/468 (92%)	411 (96%)	17 (4%)	31	37
1	B	429/468 (92%)	406 (95%)	23 (5%)	22	24
All	All	857/936 (92%)	817 (95%)	40 (5%)	26	30

All (40) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	14	VAL
1	A	23	ARG
1	A	47	ARG
1	A	97	GLU
1	A	131	LYS
1	A	164	LYS
1	A	199	LYS
1	A	227	GLU
1	A	272	GLU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	274	LYS
1	A	408	GLN
1	A	413	ARG
1	A	431	ASP
1	A	441	GLU
1	A	481	GLU
1	A	484	VAL
1	A	508	GLU
1	B	19	LEU
1	B	47	ARG
1	B	59	THR
1	B	96	ILE
1	B	120	LYS
1	B	123	GLU
1	B	129	LYS
1	B	153	THR
1	B	177	LEU
1	B	185	GLU
1	B	198	LYS
1	B	226	ILE
1	B	227	GLU
1	B	235	LYS
1	B	313	GLU
1	B	365	SER
1	B	402	LYS
1	B	413	ARG
1	B	447	GLU
1	B	451	ASP
1	B	481	GLU
1	B	488	ILE
1	B	508	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	18	ASN
1	A	114	HIS
1	A	256	GLN
1	A	520	HIS
1	B	48	ASN
1	B	114	HIS
1	B	256	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	281	ASN

### 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 ⓘ

### 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	480/526 (91%)	0.89	51 (10%) 6 6	23, 48, 77, 99	0
1	B	482/526 (91%)	0.92	52 (10%) 5 6	28, 50, 79, 114	0
All	All	962/1052 (91%)	0.91	103 (10%) 6 6	23, 49, 79, 114	0

All (103) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	177	LEU	6.4
1	B	409	GLY	5.9
1	B	36	SER	5.8
1	A	48	ASN	5.5
1	A	144	PRO	5.2
1	A	296	ASP	5.2
1	B	95	VAL	5.0
1	B	301	VAL	4.9
1	B	145	GLY	4.9
1	B	295	GLY	4.6
1	B	177	LEU	4.5
1	B	296	ASP	4.5
1	B	116	HIS	4.2
1	B	364	LYS	4.1
1	A	246	GLY	4.0
1	B	300	PRO	3.9
1	B	486	CYS	3.9
1	B	294	ASP	3.8
1	A	226	ILE	3.8
1	B	248	TRP	3.7
1	A	47	ARG	3.7
1	A	525	LEU	3.7
1	B	23	ARG	3.6
1	B	96	ILE	3.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	47	ARG	3.6
1	B	485	ASP	3.5
1	A	66	TYR	3.5
1	B	87	GLN	3.5
1	A	206	LEU	3.5
1	A	128	GLU	3.4
1	A	301	VAL	3.3
1	B	144	PRO	3.3
1	A	294	ASP	3.3
1	B	48	ASN	3.3
1	B	293	HIS	3.2
1	A	116	HIS	3.2
1	A	171	MET	3.2
1	A	140	SER	3.2
1	A	314	PHE	3.1
1	A	38	PRO	3.1
1	A	22	THR	3.0
1	A	488	ILE	3.0
1	A	484	VAL	3.0
1	B	226	ILE	3.0
1	B	481	GLU	2.9
1	A	23	ARG	2.8
1	A	522	ILE	2.8
1	B	324	LYS	2.8
1	B	154	GLY	2.8
1	A	485	ASP	2.8
1	A	481	GLU	2.8
1	A	60	GLY	2.7
1	A	300	PRO	2.7
1	A	494	ILE	2.7
1	A	145	GLY	2.7
1	B	60	GLY	2.7
1	A	478	ILE	2.7
1	B	62	TYR	2.6
1	B	38	PRO	2.6
1	A	143	HIS	2.6
1	A	139	GLU	2.6
1	A	483	GLY	2.6
1	A	295	GLY	2.6
1	B	142	SER	2.5
1	A	521	TYR	2.5
1	B	362	ARG	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	109	SER	2.5
1	B	132	HIS	2.4
1	B	426	HIS	2.4
1	B	72	ALA	2.4
1	A	463	ILE	2.4
1	A	21	LEU	2.4
1	B	224	ALA	2.4
1	A	195	GLU	2.4
1	A	356	THR	2.4
1	B	225	GLU	2.3
1	A	222	ASN	2.3
1	B	98	LEU	2.3
1	B	484	VAL	2.3
1	A	19	LEU	2.3
1	B	14	VAL	2.3
1	B	502	PRO	2.2
1	A	50	ALA	2.2
1	B	466	THR	2.2
1	A	437	ASP	2.2
1	B	313	GLU	2.2
1	A	207	GLY	2.2
1	B	119	GLY	2.2
1	B	523	GLU	2.1
1	B	117	LEU	2.1
1	B	129	LYS	2.1
1	A	192	ASP	2.1
1	B	57	GLN	2.1
1	B	494	ILE	2.1
1	A	482	LYS	2.1
1	A	191	THR	2.1
1	A	131	LYS	2.0
1	B	18	ASN	2.0
1	A	209	VAL	2.0
1	B	34	SER	2.0
1	A	450	MET	2.0
1	B	22	THR	2.0
1	A	115	GLY	2.0

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

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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.