



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 24, 2025 – 05:08 pm BST

PDB ID : 9FF5 / pdb\_00009ff5  
Title : The structure of G.kaustophilus T-1 ScoC-23bp dsDNA complex  
Authors : Hadad, N.; Shulami, S.; Pomyalov, S.; Shoham, Y.; Shoham, G.  
Deposited on : 2024-05-22  
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 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 : 4-5-2 with Phenix2.0rc1  
Xtriage (Phenix) : 2.0rc1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.43.1

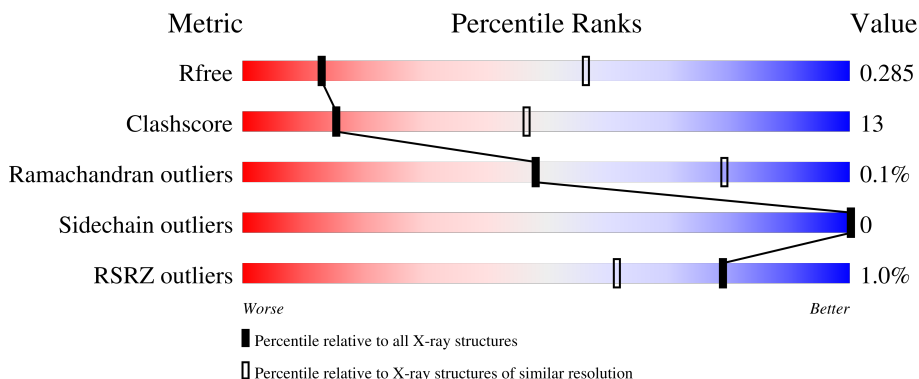
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
$R_{free}$	164625	1094 (3.56-3.44)
Clashscore	180529	1045 (3.54-3.46)
Ramachandran outliers	177936	1032 (3.54-3.46)
Sidechain outliers	177891	1033 (3.54-3.46)
RSRZ outliers	164620	1093 (3.56-3.44)

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	E	23	
1	F	23	
1	I	23	
2	H	23	
2	J	23	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	A	207	<div><div></div><div>61%</div><div>25%</div><div>14%</div></div>
3	B	207	<div><div></div><div>67%</div><div>19%</div><div>14%</div></div>
3	C	207	<div><div></div><div>%</div><div>62%</div><div>23%</div><div>14%</div></div>
3	D	207	<div><div></div><div>66%</div><div>19%</div><div>14%</div></div>
4	G	8	<div><div></div><div>38%</div><div>62%</div></div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 8400 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 DNA chain called DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*AP\*AP\*CP\*AP\*AP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	I	23	Total	C	N	O	P	0	0	0
			469	229	86	132	22			
1	E	21	Total	C	N	O	P	0	0	0
			429	210	78	121	20			
1	F	23	Total	C	N	O	P	0	0	0
			469	229	86	132	22			

- Molecule 2 is a DNA chain called DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*TP\*TP\*GP\*TP\*TP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	J	23	Total	C	N	O	P	0	0	0
			468	230	76	140	22			
2	H	23	Total	C	N	O	P	0	0	0
			468	230	76	140	22			

- Molecule 3 is a protein called HTH-type transcriptional regulator Hpr.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	178	Total	C	N	O	S	0	0	0
			1484	965	243	268	8			
3	C	178	Total	C	N	O	S	0	0	0
			1484	965	243	268	8			
3	A	178	Total	C	N	O	S	0	0	0
			1484	965	243	268	8			
3	D	178	Total	C	N	O	S	0	0	0
			1484	965	243	268	8			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	202	HIS	-	expression tag	UNP Q5L293
B	203	HIS	-	expression tag	UNP Q5L293
B	204	HIS	-	expression tag	UNP Q5L293
B	205	HIS	-	expression tag	UNP Q5L293
B	206	HIS	-	expression tag	UNP Q5L293
B	207	HIS	-	expression tag	UNP Q5L293
C	202	HIS	-	expression tag	UNP Q5L293
C	203	HIS	-	expression tag	UNP Q5L293
C	204	HIS	-	expression tag	UNP Q5L293
C	205	HIS	-	expression tag	UNP Q5L293
C	206	HIS	-	expression tag	UNP Q5L293
C	207	HIS	-	expression tag	UNP Q5L293
A	202	HIS	-	expression tag	UNP Q5L293
A	203	HIS	-	expression tag	UNP Q5L293
A	204	HIS	-	expression tag	UNP Q5L293
A	205	HIS	-	expression tag	UNP Q5L293
A	206	HIS	-	expression tag	UNP Q5L293
A	207	HIS	-	expression tag	UNP Q5L293
D	202	HIS	-	expression tag	UNP Q5L293
D	203	HIS	-	expression tag	UNP Q5L293
D	204	HIS	-	expression tag	UNP Q5L293
D	205	HIS	-	expression tag	UNP Q5L293
D	206	HIS	-	expression tag	UNP Q5L293
D	207	HIS	-	expression tag	UNP Q5L293

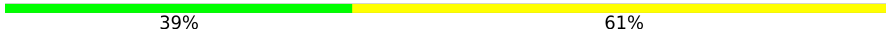
- Molecule 4 is a DNA chain called DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*AP\*TP\*T)-3').

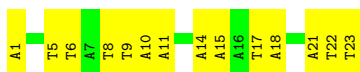
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	G	8	Total 161	C 80	N 28	O 46	P 7	0	0	0

### 3 Residue-property plots

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.

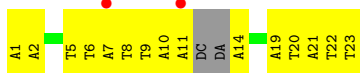
- Molecule 1: DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*AP\*AP\*CP\*AP\*AP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3')

Chain I: 



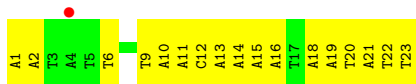
- Molecule 1: DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*AP\*AP\*CP\*AP\*AP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3')

Chain E: 



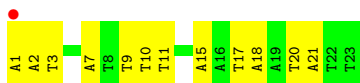
- Molecule 1: DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*AP\*AP\*CP\*AP\*AP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3')

Chain F: 



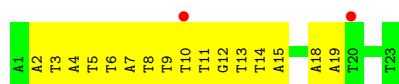
- Molecule 2: DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*TP\*TP\*GP\*TP\*TP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3')

Chain J: 



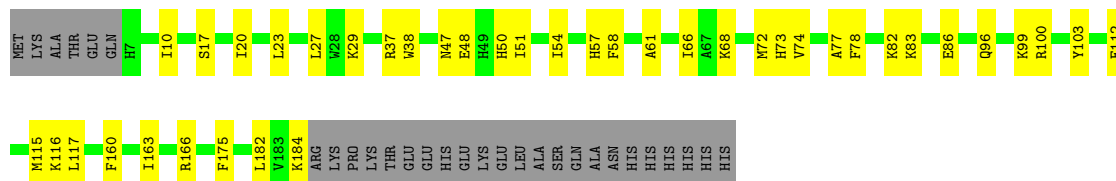
- Molecule 2: DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*TP\*AP\*TP\*TP\*TP\*TP\*GP\*TP\*TP\*AP\*AP\*TP\*AP\*AP\*TP\*AP\*TP\*T)-3')

Chain H: 



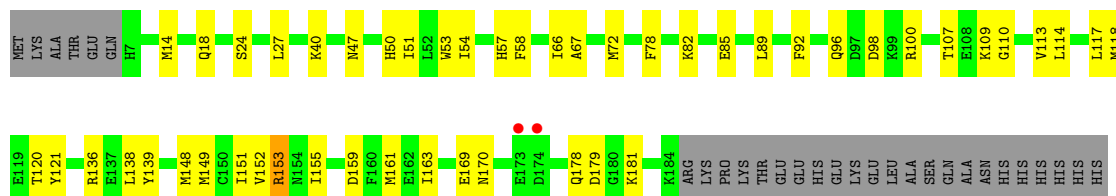
- Molecule 3: HTH-type transcriptional regulator Hpr

Chain B: 67% 19% 14%



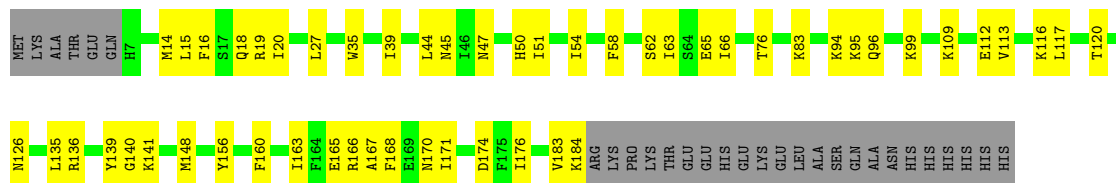
- Molecule 3: HTH-type transcriptional regulator Hpr

Chain C: 62% 23% 14%



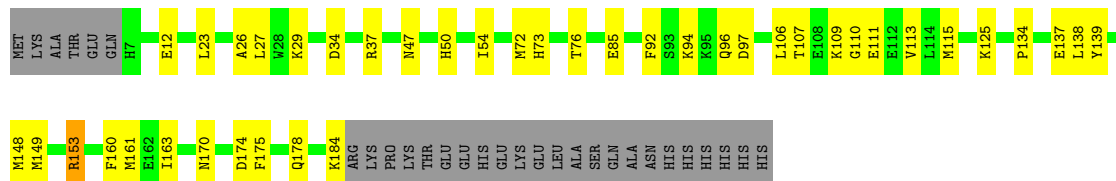
- Molecule 3: HTH-type transcriptional regulator Hpr

Chain A: 61% 25% 14%



- Molecule 3: HTH-type transcriptional regulator Hpr

Chain D: 66% 19% 14%



- Molecule 4: DNA (5'-D(\*AP\*AP\*TP\*AP\*TP\*AP\*TP\*T)-3')

Chain G: 38% 62%

A1	A2	A3	A4	T22	T23
----	----	----	----	-----	-----



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.20Å 100.15Å 96.61Å 90.00° 112.33° 90.00°	Depositor
Resolution (Å)	42.68 – 3.50 42.68 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (42.68-3.50) 99.4 (42.68-3.50)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.77 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.233 , 0.282 0.235 , 0.285	Depositor DCC
$R_{free}$ test set	898 reflections (4.53%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	115.2	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 88.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8400	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% 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	E	0.38	0/481	0.60	0/738
1	F	0.45	0/527	0.73	0/811
1	I	0.58	0/527	0.79	0/811
2	H	0.52	0/523	0.73	0/806
2	J	0.54	0/523	0.80	0/806
3	A	0.57	0/1519	0.83	0/2037
3	B	0.61	0/1519	0.84	0/2037
3	C	0.61	0/1519	0.86	1/2037 (0.0%)
3	D	0.63	1/1519 (0.1%)	0.87	1/2037 (0.0%)
4	G	0.34	0/179	0.57	0/272
All	All	0.57	1/8836 (0.0%)	0.81	2/12392 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	153	ARG	NE-CZ	5.65	1.39	1.33

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	153	ARG	CG-CD-NE	-5.49	99.92	112.00
3	D	153	ARG	NE-CZ-NH2	5.13	123.82	119.20

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	E	429	0	243	14	1
1	F	469	0	264	21	2
1	I	469	0	264	16	0
2	H	468	0	268	25	0
2	J	468	0	268	9	1
3	A	1484	0	1482	40	0
3	B	1484	0	1482	35	0
3	C	1484	0	1482	38	0
3	D	1484	0	1482	35	0
4	G	161	0	95	4	0
All	All	8400	0	7330	209	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 13.

The worst 5 of 209 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:F:18:DA:H2''	1:F:19:DA:C8	2.08	0.88
1:E:22:DT:H2'	1:E:23:DT:C6	2.09	0.87
1:F:18:DA:H2''	1:F:19:DA:H8	1.41	0.85
2:H:14:DT:H2'	2:H:15:DA:C8	2.13	0.84
2:H:14:DT:H2'	2:H:15:DA:H8	1.45	0.81

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)
2:J:1:DA:O5'	1:F:23:DT:O2[2_445]	2.01	0.19
1:E:23:DT:OP2	1:F:1:DA:O5'[2_565]	2.17	0.03

## 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	
3	A	176/207 (85%)	169 (96%)	6 (3%)	1 (1%)	22	56
3	B	176/207 (85%)	173 (98%)	3 (2%)	0	100	100
3	C	176/207 (85%)	168 (96%)	8 (4%)	0	100	100
3	D	176/207 (85%)	171 (97%)	5 (3%)	0	100	100
All	All	704/828 (85%)	681 (97%)	22 (3%)	1 (0%)	48	79

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	140	GLY

### 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	
3	A	160/186 (86%)	160 (100%)	0	100	100
3	B	160/186 (86%)	160 (100%)	0	100	100
3	C	160/186 (86%)	160 (100%)	0	100	100
3	D	160/186 (86%)	160 (100%)	0	100	100
All	All	640/744 (86%)	640 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	178	GLN
3	D	7	HIS
3	D	96	GLN
3	B	101	ASN
3	B	73	HIS

### 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 oligosaccharides 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
4	G	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	G	4:DA	O3'	20:DT	P	35.02

## 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	E	21/23 (91%)	0.85	2 (9%) 15 12	150, 195, 230, 237	0
1	F	23/23 (100%)	0.70	1 (4%) 40 27	107, 143, 161, 179	0
1	I	23/23 (100%)	0.15	0 100 100	95, 112, 123, 124	0
2	H	23/23 (100%)	0.82	2 (8%) 17 13	102, 142, 161, 173	0
2	J	23/23 (100%)	0.20	1 (4%) 40 27	90, 111, 127, 143	0
3	A	178/207 (85%)	-0.26	0 100 100	58, 93, 141, 167	0
3	B	178/207 (85%)	-0.23	0 100 100	56, 87, 148, 188	0
3	C	178/207 (85%)	-0.14	2 (1%) 77 59	57, 94, 131, 167	0
3	D	178/207 (85%)	-0.21	0 100 100	56, 88, 143, 164	0
4	G	8/8 (100%)	0.51	0 100 100	130, 151, 192, 196	0
All	All	833/951 (87%)	-0.10	8 (0%) 79 61	56, 94, 157, 237	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	7	DA	3.0
3	C	173	GLU	2.8
2	J	1	DA	2.4
1	F	4	DA	2.3
2	H	10	DT	2.3

### 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 monosaccharides 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.