



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 17, 2021 – 12:27 AM EDT

PDB ID : 1RGU  
Title : The crystal structure of human Tyrosyl-DNA Phosphodiesterase complexed with vanadate, octopamine, and tetranucleotide AGTG  
Authors : Davies, D.R.; Interthal, H.; Champoux, J.J.; Hol, W.G.  
Deposited on : 2003-11-13  
Resolution : 2.22 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 : 2.23.2  
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) : 2.23.2

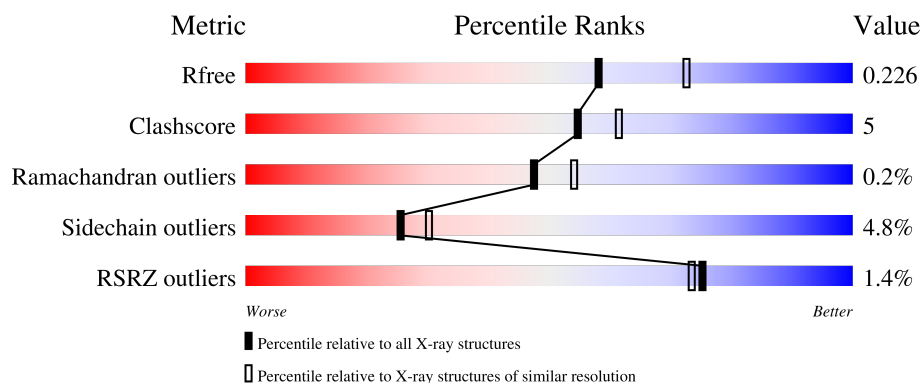
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.22 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	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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	D	4	<div> <div></div> <div>75%</div> <div>25%</div> </div>
1	F	4	<div> <div>25%</div> <div>25%</div> <div>50%</div> </div>
2	A	485	<div> <div>2%</div> <div>75%</div> <div>12%</div> <div>12%</div> </div>
2	B	485	<div> <div>%</div> <div>74%</div> <div>14%</div> <div>12%</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	VO4	A	699	-	-	X	-
3	VO4	B	699	-	-	X	-

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 7141 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 5'-D(\*AP\*GP\*TP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	3	Total	C	N	O	P	0	0	0
			53	25	7	18	3			
1	F	2	Total	C	N	O	P	0	0	0
			42	20	7	13	2			

- Molecule 2 is a protein called Tyrosyl-DNA phosphodiesterase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	425	Total	C	N	O	S	0	0	0
			3394	2208	573	602	11			
2	B	429	Total	C	N	O	S	0	0	0
			3426	2230	578	607	11			

There are 56 discrepancies between the modelled and reference sequences:

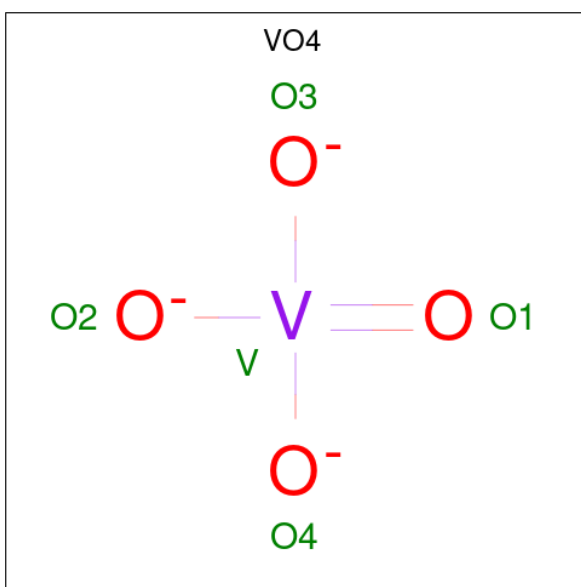
Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	cloning artifact	UNP Q9NUW8
A	125	GLY	-	cloning artifact	UNP Q9NUW8
A	126	SER	-	cloning artifact	UNP Q9NUW8
A	127	SER	-	cloning artifact	UNP Q9NUW8
A	128	HIS	-	cloning artifact	UNP Q9NUW8
A	129	HIS	-	cloning artifact	UNP Q9NUW8
A	130	HIS	-	cloning artifact	UNP Q9NUW8
A	131	HIS	-	cloning artifact	UNP Q9NUW8
A	132	HIS	-	cloning artifact	UNP Q9NUW8
A	133	HIS	-	cloning artifact	UNP Q9NUW8
A	134	SER	-	cloning artifact	UNP Q9NUW8
A	135	SER	-	cloning artifact	UNP Q9NUW8
A	136	GLY	-	cloning artifact	UNP Q9NUW8
A	137	LEU	-	cloning artifact	UNP Q9NUW8
A	138	VAL	-	cloning artifact	UNP Q9NUW8
A	139	PRO	-	cloning artifact	UNP Q9NUW8

*Continued on next page...*

*Continued from previous page...*

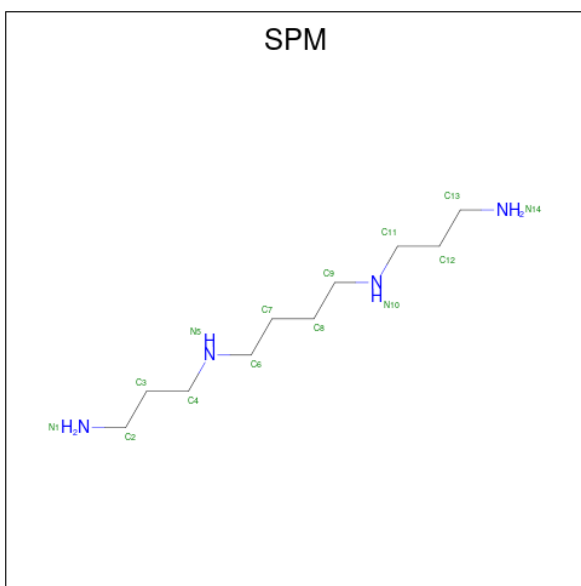
Chain	Residue	Modelled	Actual	Comment	Reference
A	140	ARG	-	cloning artifact	UNP Q9NUW8
A	141	GLY	-	cloning artifact	UNP Q9NUW8
A	142	SER	-	cloning artifact	UNP Q9NUW8
A	143	HIS	-	cloning artifact	UNP Q9NUW8
A	144	MET	-	cloning artifact	UNP Q9NUW8
A	145	LEU	-	cloning artifact	UNP Q9NUW8
A	146	GLU	-	cloning artifact	UNP Q9NUW8
A	147	ASP	-	cloning artifact	UNP Q9NUW8
A	148	PRO	-	cloning artifact	UNP Q9NUW8
A	322	ASN	ASP	engineered mutation	UNP Q9NUW8
A	328	THR	MET	engineered mutation	UNP Q9NUW8
A	548	LEU	PHE	engineered mutation	UNP Q9NUW8
B	124	MET	-	cloning artifact	UNP Q9NUW8
B	125	GLY	-	cloning artifact	UNP Q9NUW8
B	126	SER	-	cloning artifact	UNP Q9NUW8
B	127	SER	-	cloning artifact	UNP Q9NUW8
B	128	HIS	-	cloning artifact	UNP Q9NUW8
B	129	HIS	-	cloning artifact	UNP Q9NUW8
B	130	HIS	-	cloning artifact	UNP Q9NUW8
B	131	HIS	-	cloning artifact	UNP Q9NUW8
B	132	HIS	-	cloning artifact	UNP Q9NUW8
B	133	HIS	-	cloning artifact	UNP Q9NUW8
B	134	SER	-	cloning artifact	UNP Q9NUW8
B	135	SER	-	cloning artifact	UNP Q9NUW8
B	136	GLY	-	cloning artifact	UNP Q9NUW8
B	137	LEU	-	cloning artifact	UNP Q9NUW8
B	138	VAL	-	cloning artifact	UNP Q9NUW8
B	139	PRO	-	cloning artifact	UNP Q9NUW8
B	140	ARG	-	cloning artifact	UNP Q9NUW8
B	141	GLY	-	cloning artifact	UNP Q9NUW8
B	142	SER	-	cloning artifact	UNP Q9NUW8
B	143	HIS	-	cloning artifact	UNP Q9NUW8
B	144	MET	-	cloning artifact	UNP Q9NUW8
B	145	LEU	-	cloning artifact	UNP Q9NUW8
B	146	GLU	-	cloning artifact	UNP Q9NUW8
B	147	ASP	-	cloning artifact	UNP Q9NUW8
B	148	PRO	-	cloning artifact	UNP Q9NUW8
B	322	ASN	ASP	engineered mutation	UNP Q9NUW8
B	328	THR	MET	engineered mutation	UNP Q9NUW8
B	548	LEU	PHE	engineered mutation	UNP Q9NUW8

- Molecule 3 is VANADATE ION (three-letter code: VO4) (formula: O<sub>4</sub>V).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	V	0	0
			3	2	1		
3	B	1	Total	O	V	0	0
			3	2	1		

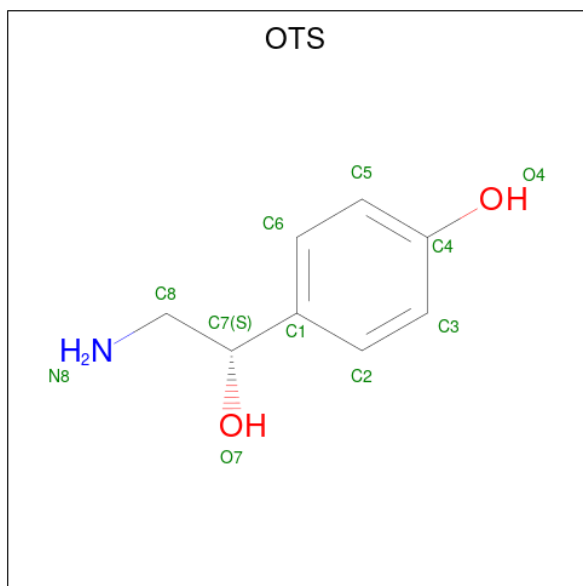
- Molecule 4 is SPERMINE (three-letter code: SPM) (formula:  $C_{10}H_{26}N_4$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	N	0	0
			14	10	4		

- Molecule 5 is 4-(2S-AMINO-1-HYDROXYETHYL)PHENOL (three-letter code: OTS)

(formula: C<sub>8</sub>H<sub>11</sub>NO<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			11	8	1	2		
5	B	1	Total	C	N	O	0	0
			11	8	1	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	2	Total	O	0	0
			2	2		
6	F	4	Total	O	0	0
			4	4		
6	A	82	Total	O	0	0
			82	82		
6	B	96	Total	O	0	0
			96	96		

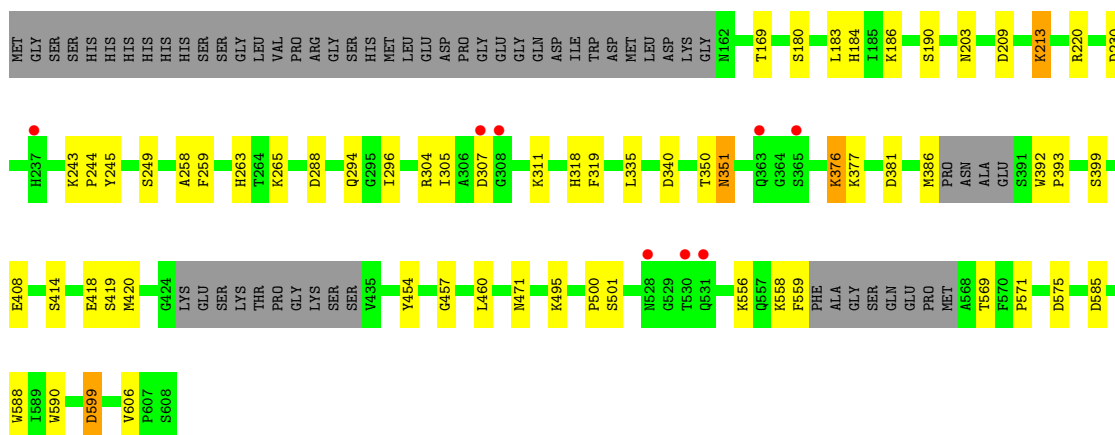



- Molecule 1: 5'-D(\*AP\*GP\*TP\*G)-3'

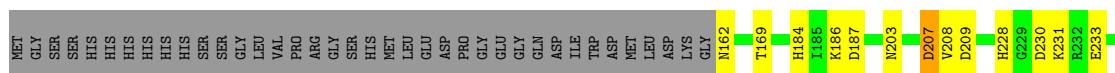


- |    |      |      |
|----|------|------|
| DA | T805 | G806 |
| DG |      |      |

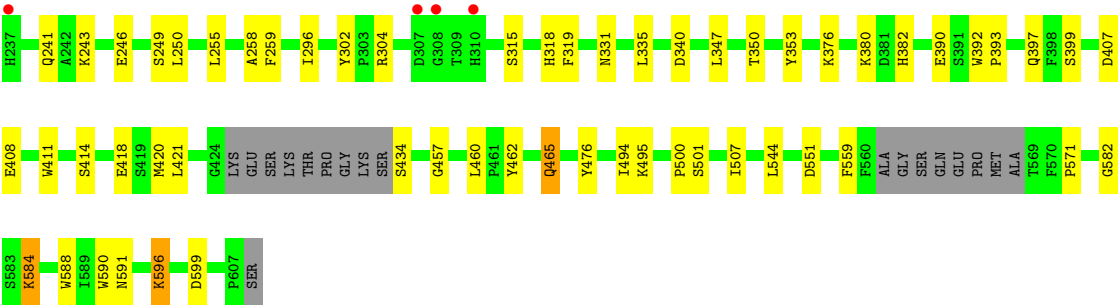
- Chain A:  2% 75% 12% 12%



- Chain B:  74% 14% 12% 2%







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.72Å 104.56Å 193.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.22 48.30 – 2.22	Depositor EDS
% Data completeness (in resolution range)	92.9 (50.00-2.22) 92.9 (48.30-2.22)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.06 (at 2.22Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.187 , 0.231 0.185 , 0.226	Depositor DCC
$R_{free}$ test set	2398 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.3	Xtriage
Anisotropy	0.385	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 42.7	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	7141	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.02% 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

Bond lengths and bond angles in the following residue types are not validated in this section: SPM, VO4, OTS

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	D	0.92	0/58	1.38	0/87
1	F	0.86	0/46	1.77	3/69 (4.3%)
2	A	0.47	0/3505	0.69	6/4765 (0.1%)
2	B	0.51	0/3540	0.70	7/4816 (0.1%)
All	All	0.50	0/7149	0.72	16/9737 (0.2%)

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	209	ASP	CB-CG-OD2	7.43	124.99	118.30
2	B	599	ASP	CB-CG-OD2	7.00	124.60	118.30
2	A	340	ASP	CB-CG-OD2	6.97	124.57	118.30
2	B	407	ASP	CB-CG-OD2	6.56	124.21	118.30
2	B	551	ASP	CB-CG-OD2	6.01	123.71	118.30

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	D	53	0	29	0	0
1	F	42	0	23	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	3394	0	3330	30	0
2	B	3426	0	3354	41	0
3	A	3	0	0	3	0
3	B	3	0	0	2	0
4	A	14	0	26	2	0
5	A	11	0	10	0	0
5	B	11	0	10	0	0
6	A	82	0	0	1	0
6	B	96	0	0	1	0
6	D	2	0	0	0	0
6	F	4	0	0	0	0
All	All	7141	0	6782	72	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 5.

The worst 5 of 72 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)
2:B:228:HIS:HD2	2:B:230:ASP:H	1.21	0.89
2:B:465:GLN:HE22	2:B:591:ASN:HD21	1.34	0.76
2:B:408:GLU:OE2	2:B:476:TYR:OH	2.04	0.72
2:B:250:LEU:HD12	2:B:250:LEU:N	2.05	0.71
2:A:213:LYS:N	2:A:213:LYS:HD3	2.06	0.71

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
2	A	417/485 (86%)	400 (96%)	16 (4%)	1 (0%)	47 54

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	423/485 (87%)	413 (98%)	9 (2%)	1 (0%)	47	54
All	All	840/970 (87%)	813 (97%)	25 (3%)	2 (0%)	47	54

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	307	ASP
2	B	411	TRP

### 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	
2	A	366/421 (87%)	348 (95%)	18 (5%)	25	29
2	B	369/421 (88%)	352 (95%)	17 (5%)	27	32
All	All	735/842 (87%)	700 (95%)	35 (5%)	25	30

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	434	SER
2	B	462	TYR
2	B	584	LYS
2	A	386	MET
2	A	376	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	397	GLN
2	B	528	ASN
2	B	165	GLN
2	B	184	HIS
2	B	228	HIS

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

5 ligands are 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	$\# Z  > 2$
5	OTS	A	995	3	11,11,11	0.82	0	13,14,14	0.74	0
4	SPM	A	999	-	13,13,13	0.38	0	12,12,12	0.87	0
3	VO4	B	699	1,5,2	0,2,4	-	-	-		
3	VO4	A	699	1,5,2	0,2,4	-	-	-		
5	OTS	B	996	3	11,11,11	0.65	0	13,14,14	0.54	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
5	OTS	B	996	3	-	6/6/6/6	0/1/1/1
5	OTS	A	995	3	-	2/6/6/6	0/1/1/1
4	SPM	A	999	-	-	6/11/11/11	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	B	996	OTS	C1-C7-C8-N8
5	B	996	OTS	O7-C7-C8-N8
4	A	999	SPM	C2-C3-C4-N5
4	A	999	SPM	N10-C11-C12-C13
5	B	996	OTS	C2-C1-C7-C8

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	999	SPM	2	0
3	B	699	VO4	2	0
3	A	699	VO4	3	0

## 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	D	3/4 (75%)	-0.70	0 100 100	29, 29, 38, 55	0
1	F	2/4 (50%)	-0.95	0 100 100	27, 27, 27, 34	0
2	A	425/485 (87%)	-0.30	8 (1%) 66 65	17, 30, 51, 66	0
2	B	429/485 (88%)	-0.37	4 (0%) 84 83	13, 25, 45, 60	0
All	All	859/978 (87%)	-0.34	12 (1%) 75 73	13, 28, 49, 66	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	528	ASN	4.3
2	B	307	ASP	3.2
2	B	308	GLY	2.9
2	A	531	GLN	2.6
2	A	308	GLY	2.5

### 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 monosaccharides 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. 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( $\text{\AA}^2$ )	Q<0.9
4	SPM	A	999	14/14	0.86	0.21	33,44,54,54	0
5	OTS	A	995	11/11	0.97	0.10	18,28,40,44	0
5	OTS	B	996	11/11	0.97	0.11	27,30,43,45	0
3	VO4	B	699	3/5	1.00	0.09	17,17,21,22	0
3	VO4	A	699	3/5	1.00	0.08	15,15,20,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.