



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 12:03 PM UTC

PDB ID : 1RFI / pdb_00001rfi
Title : Crystal structure of human Tyrosyl-DNA Phosphodiesterase complexed with vanadate, pentapeptide KLNYK, and tetranucleotide AGTC
Authors : Davies, D.R.; Interthal, H.; Champoux, J.J.; Hol, W.G.
Deposited on : 2003-11-10
Resolution : 2.20 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

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 Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

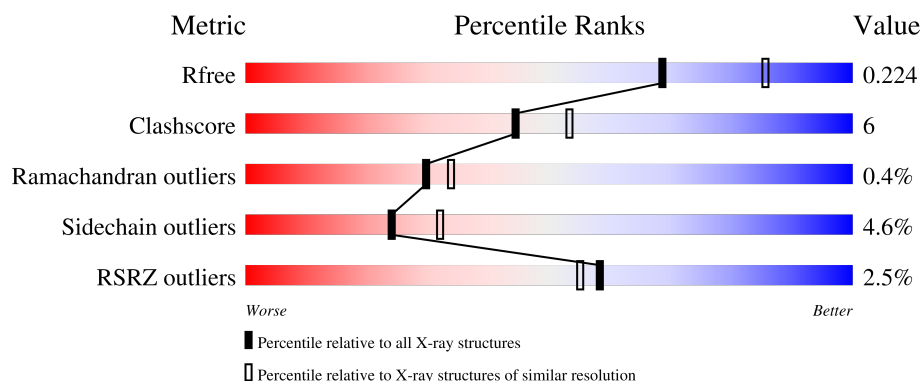
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	180053	6164 (2.20-2.20)
Clashscore	190562	6851 (2.20-2.20)
Ramachandran outliers	187476	6768 (2.20-2.20)
Sidechain outliers	187428	6769 (2.20-2.20)
RSRZ outliers	180081	6166 (2.20-2.20)

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 ≥ 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	
1	F	4	
2	A	485	
2	B	485	
3	C	5	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	5	 60% 40%

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
4	VO4	A	699	-	-	X	-
4	VO4	B	699	-	-	X	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7258 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*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	3	Total	C	N	O	P	0	0	0
			50	24	5	18	3			
1	F	3	Total	C	N	O	P	0	0	0
			50	24	5	18	3			

- 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...

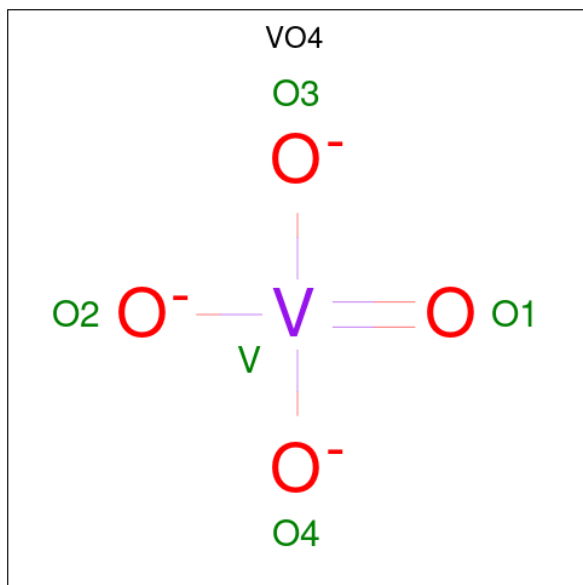
Continued from previous page...

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 a protein called Topoisomerase I-Derived Peptide.

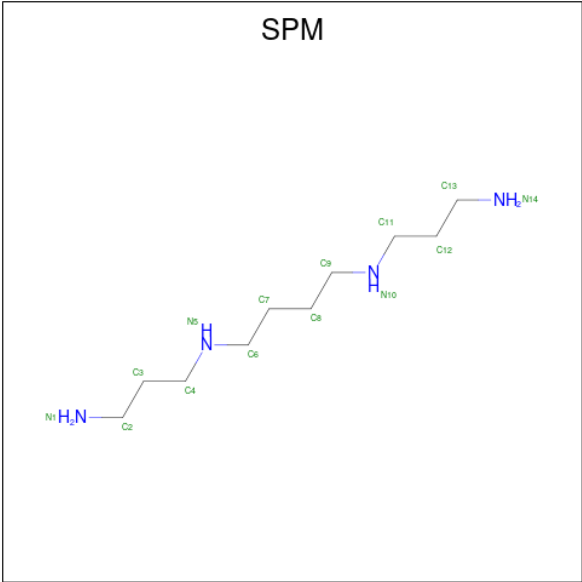
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	5	Total	C	N	O	0	0	0
			47	31	8	8			
3	E	3	Total	C	N	O	0	0	0
			28	19	4	5			

- Molecule 4 is VANADATE ION (CCD ID: VO4) (formula: O_4V).



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

- Molecule 5 is SPERMINE (CCD ID: SPM) (formula: $C_{10}H_{26}N_4$).



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

- Molecule 6 is water.

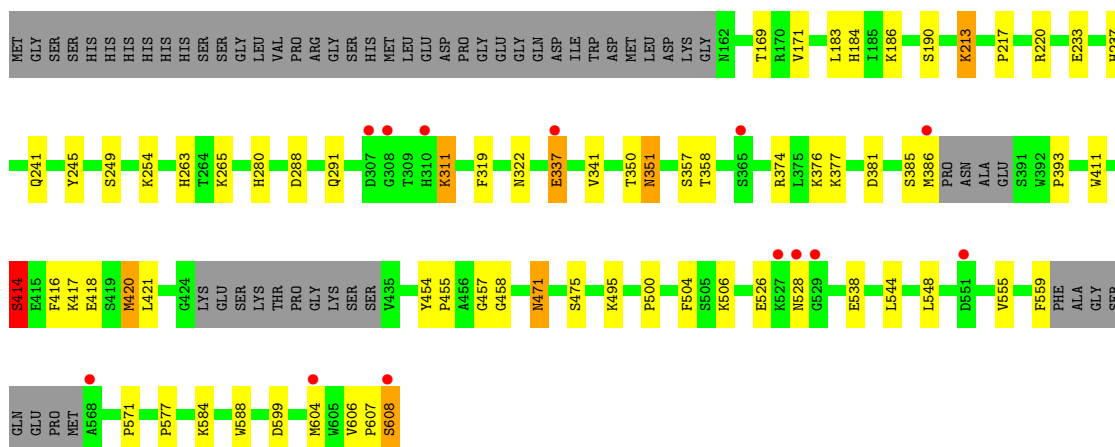
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	5	Total	O	0	0
			5	5		
6	F	2	Total	O	0	0
			2	2		
6	A	112	Total	O	0	0
			112	112		
6	B	123	Total	O	0	0
			123	123		
6	C	1	Total	O	0	0
			1	1		

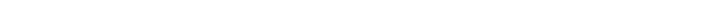
- Molecule 1: 5'-D(*AP*GP*TP*C)-3'

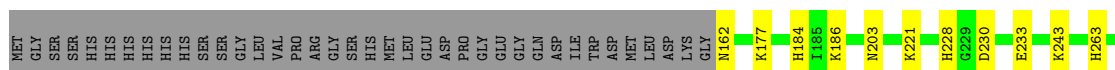


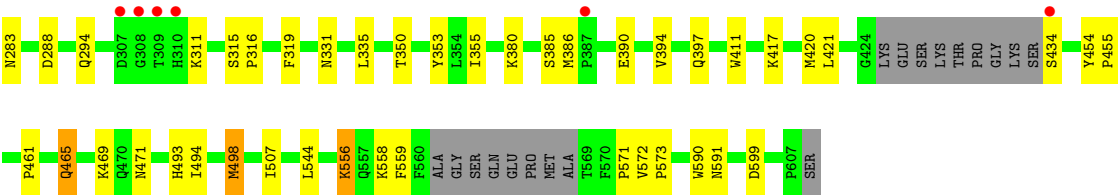
- | | |
|------|--|
| DA | |
| G804 | |
| T805 | |
| C806 | |

- Chain A: 3% 73% 13% 12%



- Chain B:  77% 11% 12%





• Molecule 3: Topoisomerase I-Derived Peptide



• Molecule 3: Topoisomerase I-Derived Peptide



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	49.82Å 104.83Å 193.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 50.00 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.9 (50.00-2.20) 99.9 (50.00-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.72 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.188 , 0.227 0.186 , 0.224	Depositor DCC
R_{free} test set	2674 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å ²)	24.5	Xtriage
Anisotropy	0.502	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 35.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7258	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.07% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: SPM, VO4

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	0.36	0/54	0.86	0/80
1	F	0.42	0/54	1.25	0/80
2	A	0.55	1/3505 (0.0%)	0.83	0/4765
2	B	0.59	1/3540 (0.0%)	0.84	0/4816
3	C	0.57	0/47	0.61	0/59
3	E	0.54	0/28	0.44	0/37
All	All	0.57	2/7228 (0.0%)	0.84	0/9837

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	604	MET	SD-CE	6.27	1.95	1.79
2	B	498	MET	SD-CE	-5.43	1.66	1.79

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	D	50	0	29	0	0
1	F	50	0	29	2	0
2	A	3394	0	3330	44	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	3426	0	3354	33	0
3	C	47	0	50	5	0
3	E	28	0	24	0	0
4	A	3	0	0	3	0
4	B	3	0	0	3	0
5	A	14	0	26	2	0
6	A	112	0	0	5	0
6	B	123	0	0	0	0
6	C	1	0	0	1	0
6	D	5	0	0	0	0
6	F	2	0	0	0	0
All	All	7258	0	6842	83	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 6.

The worst 5 of 83 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:498:MET:CE	2:B:507:ILE:HG21	1.77	1.14
2:B:498:MET:HE2	2:B:507:ILE:HG21	1.40	1.03
2:B:498:MET:HE2	2:B:507:ILE:CG2	2.00	0.90
2:A:607:PRO:O	2:A:608:SER:HB2	1.74	0.86
2:B:228:HIS:HD2	2:B:230:ASP:H	1.26	0.81

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%)	407 (98%)	8 (2%)	2 (0%)	24 27

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	423/485 (87%)	415 (98%)	7 (2%)	1 (0%)	43	51
3	C	3/5 (60%)	2 (67%)	1 (33%)	0	100	100
3	E	1/5 (20%)	1 (100%)	0	0	100	100
All	All	844/980 (86%)	825 (98%)	16 (2%)	3 (0%)	30	34

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	414	SER
2	A	411	TRP
2	B	411	TRP

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	
2	A	366/421 (87%)	348 (95%)	18 (5%)	22	29
2	B	369/421 (88%)	353 (96%)	16 (4%)	26	35
3	C	5/5 (100%)	5 (100%)	0	100	100
3	E	3/5 (60%)	3 (100%)	0	100	100
All	All	743/852 (87%)	709 (95%)	34 (5%)	24	32

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	385	SER
2	B	390	GLU
2	B	471	ASN
2	A	471	ASN
2	A	420	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	369	ASN
2	B	489	ASN
2	B	531	GLN
2	B	528	ASN
2	A	397	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

3 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	$\# Z > 2$
4	VO4	B	699	2,1,3	0,2,4	-	-	-		
4	VO4	A	699	2,1,3	0,2,4	-	-	-		
5	SPM	A	999	-	13,13,13	0.36	0	12,12,12	0.79	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	SPM	A	999	-	-	11/11/11/11	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	999	SPM	C7-C8-C9-N10
5	A	999	SPM	C2-C3-C4-N5
5	A	999	SPM	N5-C6-C7-C8
5	A	999	SPM	C12-C11-N10-C9
5	A	999	SPM	C6-C7-C8-C9

There are no ring outliers.

3 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	699	VO4	3	0
4	A	699	VO4	3	0
5	A	999	SPM	2	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, 95th 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 ²)	Q<0.9
1	D	3/4 (75%)	0.12	0 100 100	26, 26, 40, 48	0
1	F	3/4 (75%)	0.26	0 100 100	24, 24, 37, 47	0
2	A	425/485 (87%)	-0.05	13 (3%) 51 48	12, 26, 45, 63	0
2	B	429/485 (88%)	-0.32	6 (1%) 73 71	8, 20, 39, 53	0
3	C	5/5 (100%)	2.02	3 (60%) 0 0	36, 41, 49, 54	0
3	E	3/5 (60%)	1.05	0 100 100	34, 34, 43, 43	0
All	All	868/988 (87%)	-0.17	22 (2%) 58 55	8, 23, 44, 63	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	307	ASP	3.6
2	B	309	THR	3.4
2	B	307	ASP	3.4
2	A	568	ALA	3.2
2	A	308	GLY	3.0

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 oligosaccharides 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, 95th 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 ²)	Q<0.9
5	SPM	A	999	14/14	0.85	0.18	36,39,47,47	0
4	VO4	A	699	3/5	0.99	0.05	16,16,20,20	0
4	VO4	B	699	3/5	1.00	0.05	15,15,17,17	0

6.5 Other polymers [i](#)

There are no such residues in this entry.