wwPDB X-ray Structure Validation Summary Report (i)

Aug 5, 2026 – 06:32 AM EDT

PDB ID	:	1RGU / pdb_00001rgu
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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

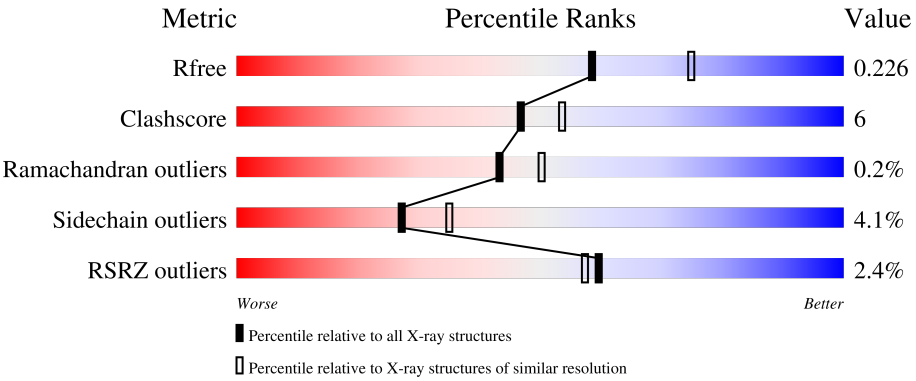
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.50

1 Overall quality at a glance (i)

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}	180053	7682 (2.24-2.20)
Clashscore	190562	8402 (2.24-2.20)
Ramachandran outliers	187476	8303 (2.24-2.20)
Sidechain outliers	187428	8304 (2.24-2.20)
RSRZ outliers	180081	7683 (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 >=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 <=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	75% 25%
1	F	4	50% 50%
2	A	485	2% 75% 12% • 12%
2	B	485	2% 75% 13% • 12%

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:

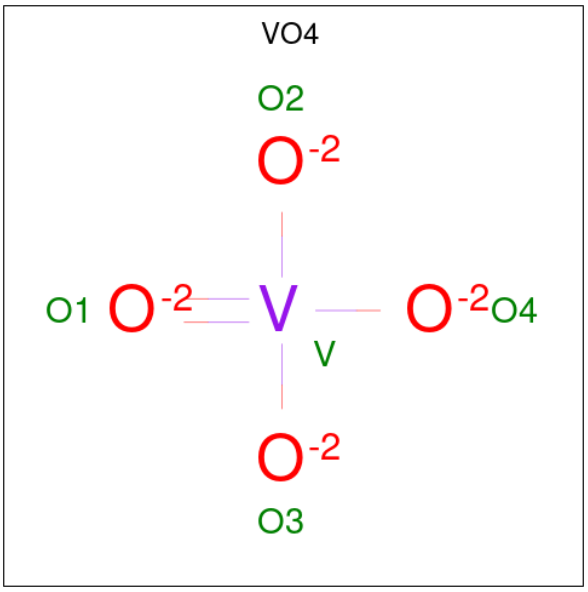
Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	cloning artifact	UNP Q9NUW8
A	125	GLY	-	cloning artifact	UNP Q9NUW8
A	126	SER	-	cloning artifact	UNP Q9NUW8
A	127	SER	-	cloning artifact	UNP Q9NUW8
A	128	HIS	-	cloning artifact	UNP Q9NUW8
A	129	HIS	-	cloning artifact	UNP Q9NUW8
A	130	HIS	-	cloning 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...

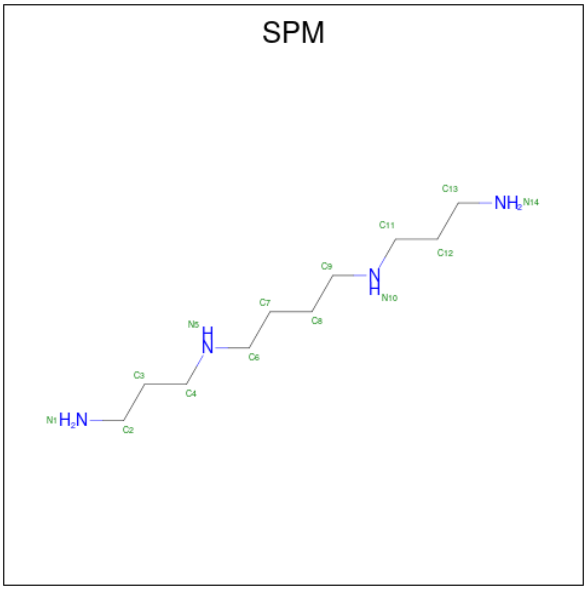
Chain	Residue	Modelled	Actual	Comment	Reference
A	140	ARG	-	cloning artifact	UNP Q9NUW8
A	141	GLY	-	cloning artifact	UNP Q9NUW8
A	142	SER	-	cloning artifact	UNP Q9NUW8
A	143	HIS	-	cloning artifact	UNP Q9NUW8
A	144	MET	-	cloning artifact	UNP Q9NUW8
A	145	LEU	-	cloning artifact	UNP Q9NUW8
A	146	GLU	-	cloning 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 (CCD ID: VO4) (formula: O_4V).



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

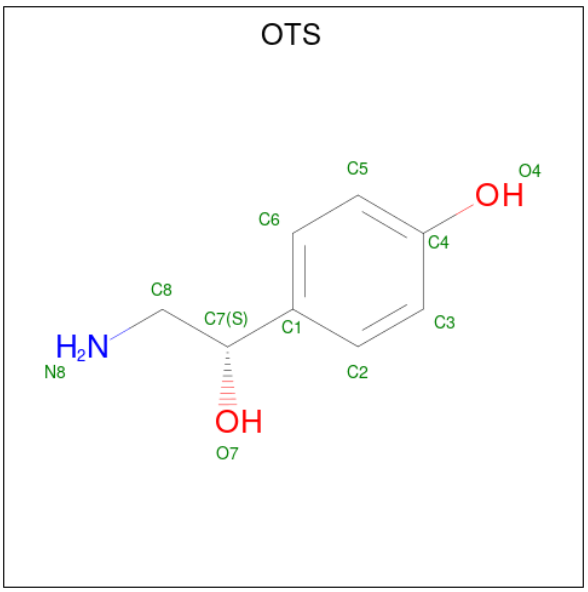
- Molecule 4 is SPERMINE (CCD ID: SPM) (formula: $C_{10}H_{26}N_4$).



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

- Molecule 5 is 4-(2S-AMINO-1-HYDROXYETHYL)PHENOL (CCD ID: OTS) (formula:

$C_8H_{11}NO_2$).



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

- Molecule 6 is water.

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

3 Residue-property plots (i)

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

Chain D:





• Molecule 1: 5'-D(*AP*GP*TP*G)-3'

Chain F:

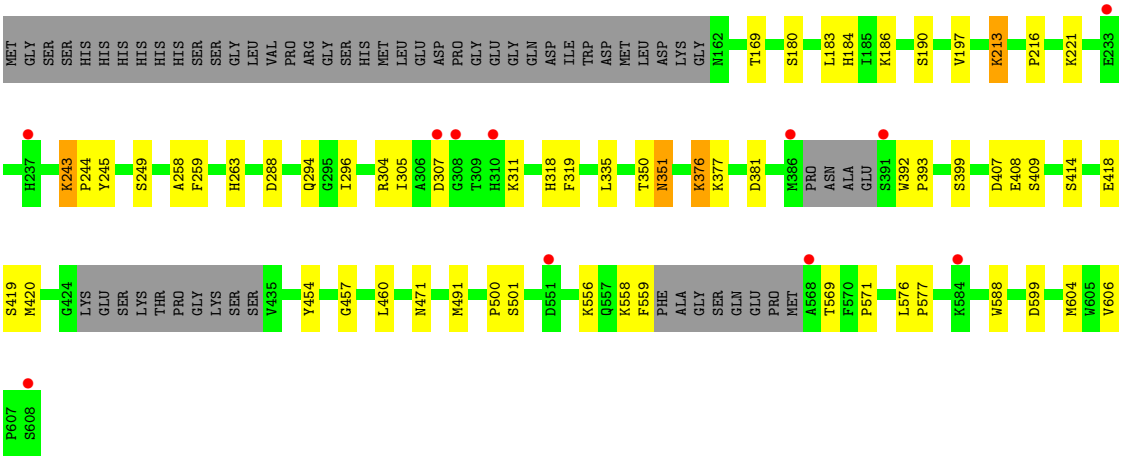




• Molecule 2: Tyrosyl-DNA phosphodiesterase 1

Chain A:



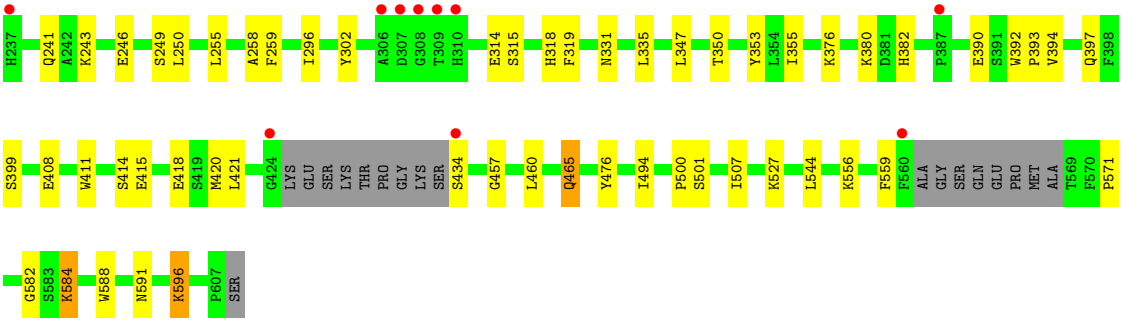


• Molecule 2: Tyrosyl-DNA phosphodiesterase 1

Chain B:






H237
Q241
A242
K243
E246
S249
L250
L255
A258
F259
I296
Y302
A306
D307
G308
T309
H310
E314
S315
H318
F319
N331
L335
L347
T350
Y353
L354
I355
K376
K380
D381
H382
P387
E390
S391
W392
P393
V394
Q397
F398
S399
E408
W411
S414
E415
E418
S419
M420
L421
G424
LYS
GLU
SER
LYS
THR
PRO
GLY
LYS
SER
S434
G457
L460
Q465
Y476
I494
P500
S501
I507
K527
L544
K556
F559
F560
ALA
GLY
SER
GLN
GLU
PRO
MET
ALA
T569
F570
P571
G582
S583
K584
W588
N591
K596
P607
SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	49.72Å 104.56Å 193.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.22 50.00 – 2.22	Depositor EDS
% Data completeness (in resolution range)	92.9 (50.00-2.22) 92.9 (50.00-2.22)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	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^2)	29.3	Xtriage
Anisotropy	0.385	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 35.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	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^2)	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.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: OTS, VO4, SPM

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	0.38	0/58	0.92	0/87
1	F	0.45	0/46	1.32	0/69
2	A	0.57	1/3505 (0.0%)	0.84	1/4765 (0.0%)
2	B	0.62	0/3540	0.86	2/4816 (0.0%)
All	All	0.59	1/7149 (0.0%)	0.86	3/9737 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	604	MET	SD-CE	5.92	1.94	1.79

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	216	PRO	O-C-N	5.82	123.99	121.31
2	A	216	PRO	O-C-N	5.10	123.55	121.15
2	B	415	GLU	N-CA-C	5.06	116.60	111.14

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
2	A	3394	0	3330	33	0
2	B	3426	0	3354	44	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	78	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 78 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%)	43	50
2	B	423/485 (87%)	413 (98%)	9 (2%)	1 (0%)	43	50
All	All	840/970 (87%)	813 (97%)	25 (3%)	2 (0%)	43	50

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%)	351 (96%)	15 (4%)	27	35
2	B	369/421 (88%)	354 (96%)	15 (4%)	27	35
All	All	735/842 (87%)	705 (96%)	30 (4%)	27	35

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	471	ASN
2	B	556	LYS
2	B	197	VAL
2	B	596	LYS
2	B	434	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	331	ASN
2	B	465	GLN
2	B	528	ASN
2	B	397	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	184	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)

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	B	996	3	11,11,11	0.66	0	14,14,14	0.63	0
3	VO4	A	699	2,5,1	0,2,4	-	-	-		
5	OTS	A	995	3	11,11,11	0.84	1 (9%)	14,14,14	0.71	0
4	SPM	A	999	-	13,13,13	0.38	0	12,12,12	0.87	0
3	VO4	B	699	2,5,1	0,2,4	-	-	-		

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
4	SPM	A	999	-	-	6/11/11/11	-
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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	995	OTS	C1-C7	2.08	1.55	1.51

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
3	A	699	VO4	3	0
4	A	999	SPM	2	0
3	B	699	VO4	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

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^{th} 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%)	-0.22	0	100	100	29, 29, 38, 55	0
1	F	2/4 (50%)	-0.84	0	100	100	27, 27, 27, 34	0
2	A	425/485 (87%)	-0.04	11 (2%)	57	55	17, 30, 51, 66	0
2	B	429/485 (88%)	-0.27	10 (2%)	61	59	13, 25, 45, 60	0
All	All	859/978 (87%)	-0.16	21 (2%)	59	57	13, 28, 49, 66	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	307	ASP	4.6
2	B	308	GLY	4.2
2	B	310	HIS	3.7
2	B	306	ALA	3.2
2	B	309	THR	3.2

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

There are no oligosaccharides in this entry.

6.4 Ligands

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^{th} 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
4	SPM	A	999	14/14	0.88	0.15	33,44,54,54	0
5	OTS	A	995	11/11	0.92	0.09	18,28,40,44	0
5	OTS	B	996	11/11	0.93	0.09	27,30,43,45	0
3	VO4	B	699	3/5	1.00	0.04	17,17,21,22	0
3	VO4	A	699	3/5	1.00	0.05	15,15,20,20	0

6.5 Other polymers (i)

There are no such residues in this entry.