wwPDB X-ray Structure Validation Summary Report (i)

Aug 14, 2023 – 09:24 PM EDT

PDB ID	:	1RG2
Title	:	Crystal structure of human Tyrosyl-DNA Phosphodiesterase complexed with vanadate, octopamine, and tetranucleotide AGTA
Authors	:	Davies, D.R.; Interthal, H.; Champoux, J.J.; Hol, W.G.
Deposited on	:	2003-11-11
Resolution	:	2.10 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:

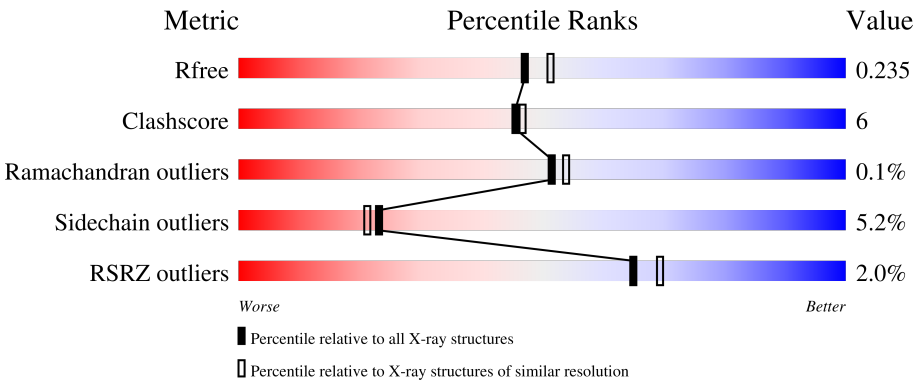
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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% 71% 16% • 12%
2	B	485	2% 74% 14% • 11%

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 7 unique types of molecules in this entry. The entry contains 7225 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*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	D	3	52	25	7	17	3	0	0	0
			Total	C	N	O	P			
1	F	2	41	20	7	12	2	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	425	3394	2208	573	602	11	0	0	0
			Total	C	N	O	S			
2	B	434	3452	2242	584	615	11	0	0	0

There are 56 discrepancies between the modelled and reference sequences:

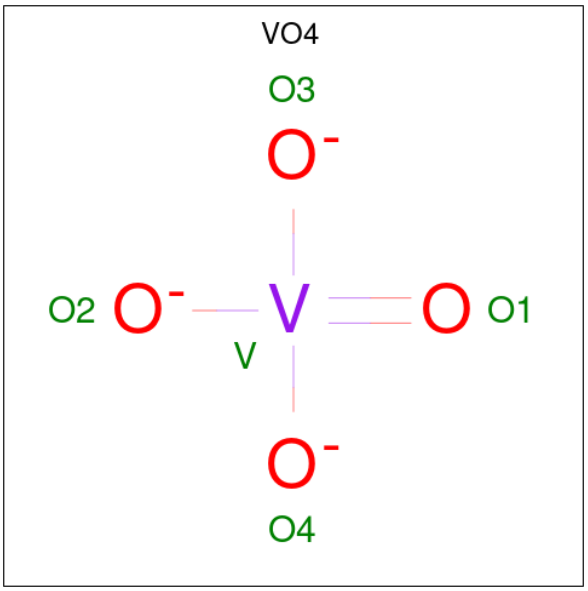
Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	cloning artifact	UNP Q9NUW8
A	125	GLY	-	cloning artifact	UNP Q9NUW8
A	126	SER	-	cloning artifact	UNP Q9NUW8
A	127	SER	-	cloning artifact	UNP Q9NUW8
A	128	HIS	-	cloning artifact	UNP Q9NUW8
A	129	HIS	-	cloning artifact	UNP Q9NUW8
A	130	HIS	-	cloning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

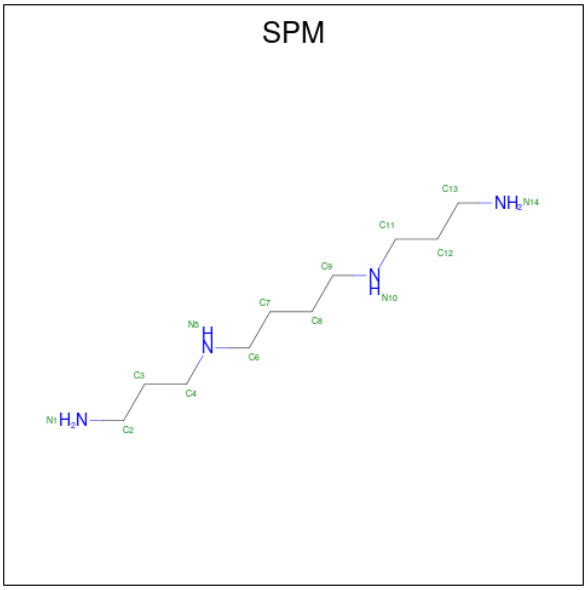
Chain	Residue	Modelled	Actual	Comment	Reference
A	140	ARG	-	cloning artifact	UNP Q9NUW8
A	141	GLY	-	cloning artifact	UNP Q9NUW8
A	142	SER	-	cloning artifact	UNP Q9NUW8
A	143	HIS	-	cloning artifact	UNP Q9NUW8
A	144	MET	-	cloning artifact	UNP Q9NUW8
A	145	LEU	-	cloning artifact	UNP Q9NUW8
A	146	GLU	-	cloning 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_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

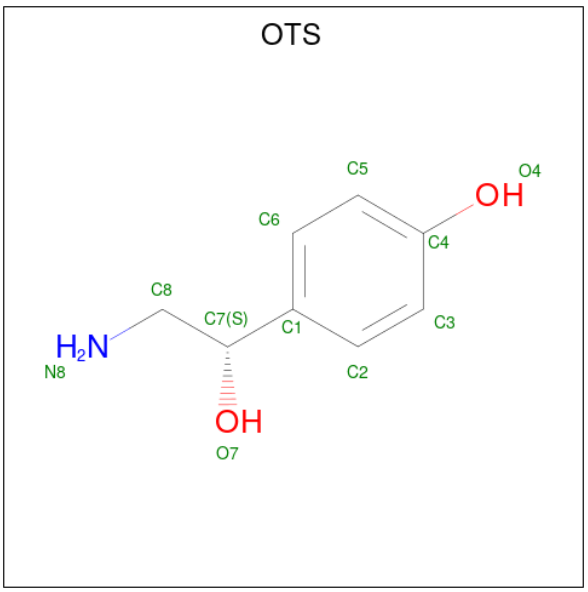
- Molecule 4 is SPERMINE (three-letter code: SPM) (formula: $C_{10}H_{26}N_4$).



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

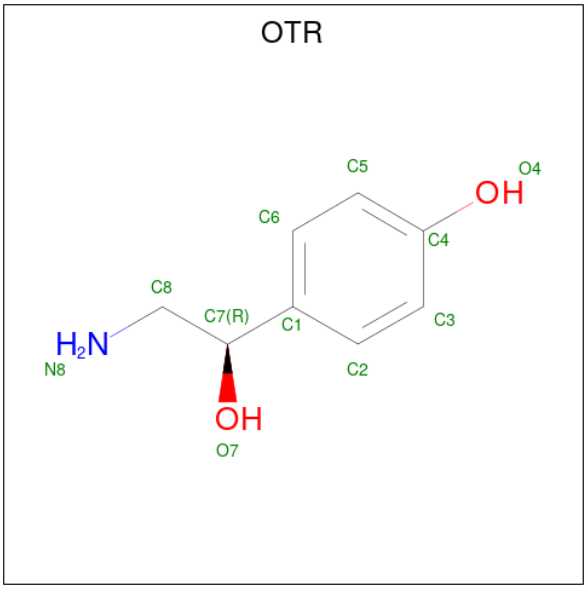
- Molecule 5 is 4-(2S-AMINO-1-HYDROXYETHYL)PHENOL (three-letter code: OTS)

(formula: $C_8H_{11}NO_2$).



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

- Molecule 6 is 4-(2R-AMINO-1-HYDROXYETHYL)PHENOL (three-letter code: OTR) (formula: $C_8H_{11}NO_2$).



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

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	D	4	4	4	0	0
7	F	3	3	3	0	0
7	A	93	93	93	0	0
7	B	122	122	122	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*A)-3'

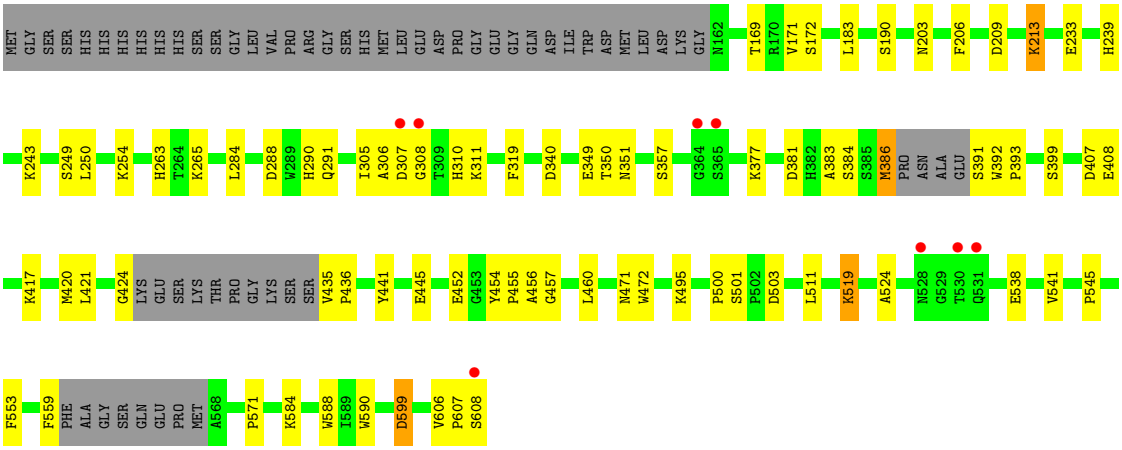
Chain D:

• Molecule 1: 5'-D(*AP*GP*TP*A)-3'

Chain F:

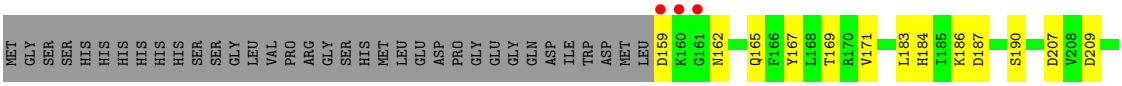
• Molecule 2: Tyrosyl-DNA phosphodiesterase 1

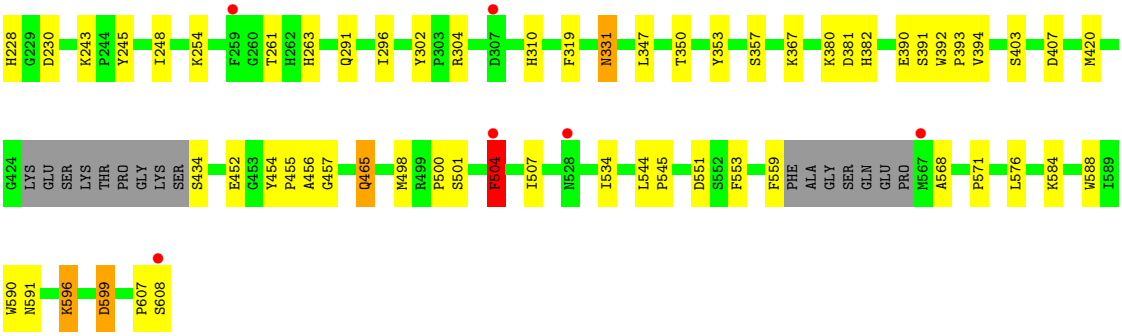
Chain A:



• Molecule 2: Tyrosyl-DNA phosphodiesterase 1

Chain B:




H228
G229
D230
K243
P244
Y245
I248
K254
F259
G260
T261
H262
H263
Q291
I296
Y302
P303
R304
D307
H310
F319
N331
L347
T350
Y353
S357
K367
K380
D381
H382
E390
S391
W392
P393
V394
S403
D407
M420
G424
LYS
GLU
SER
LYS
THR
PRO
GLY
LYS
SER
S434
E452
G453
Y454
P455
A456
G457
Q465
M498
R499
P500
S501
F504
I507
N528
I534
L544
P545
D551
S552
F553
F559
PHE
ALA
GLY
SER
GLN
GLU
PRO
M567
A568
P571
L576
K584
W588
I589
W590
N591
K596
D599
P607
S608

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	49.81Å 104.69Å 193.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.10 48.33 – 2.10	Depositor EDS
% Data completeness (in resolution range)	95.8 (50.00-2.10) 95.8 (48.33-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	3.16 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.195 , 0.233 0.197 , 0.235	Depositor DCC
R_{free} test set	2917 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	21.4	Xtriage
Anisotropy	0.604	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 51.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7225	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: VO4, OTR, SPM, 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	0.75	0/57	1.41	0/85
1	F	0.84	0/45	1.30	0/67
2	A	0.43	0/3505	0.67	5/4765 (0.1%)
2	B	0.69	4/3565 (0.1%)	1.12	14/4849 (0.3%)
All	All	0.58	4/7172 (0.1%)	0.93	19/9766 (0.2%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	504	PHE	CE1-CZ	21.02	1.77	1.37
2	B	504	PHE	CE2-CZ	13.62	1.63	1.37
2	B	504	PHE	CG-CD1	-13.09	1.19	1.38
2	B	504	PHE	CG-CD2	-5.88	1.29	1.38

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	504	PHE	CG-CD1-CE1	-41.50	75.15	120.80
2	B	504	PHE	CG-CD2-CE2	-27.66	90.37	120.80
2	B	504	PHE	CB-CG-CD1	-27.08	101.84	120.80
2	B	504	PHE	CZ-CE2-CD2	-14.38	102.85	120.10
2	B	504	PHE	CD1-CG-CD2	12.63	134.72	118.30

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	52	0	29	0	0
1	F	41	0	23	0	0
2	A	3394	0	3330	39	0
2	B	3452	0	3377	46	0
3	A	3	0	0	3	0
3	B	3	0	0	3	0
4	A	14	0	26	1	0
5	A	11	0	10	0	0
5	B	11	0	10	0	0
6	A	11	0	10	0	0
6	B	11	0	10	0	0
7	A	93	0	0	4	0
7	B	122	0	0	5	0
7	D	4	0	0	0	0
7	F	3	0	0	0	0
All	All	7225	0	6825	88	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 88 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	1.69	1.22
2:B:167:TYR:CD2	7:B:625:HOH:O	2.13	0.98
2:B:167:TYR:HB3	7:B:625:HOH:O	1.65	0.95
2:B:228:HIS:HD2	2:B:230:ASP:H	1.16	0.91
2:B:165:GLN:HA	7:B:635:HOH:O	1.74	0.86

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%)	404 (97%)	12 (3%)	1 (0%)	47	49
2	B	428/485 (88%)	415 (97%)	13 (3%)	0	100	100
All	All	845/970 (87%)	819 (97%)	25 (3%)	1 (0%)	51	54

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	307	ASP

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%)	345 (94%)	21 (6%)	20	18
2	B	371/421 (88%)	354 (95%)	17 (5%)	27	26
All	All	737/842 (88%)	699 (95%)	38 (5%)	23	21

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	367	LYS
2	B	584	LYS
2	B	380	LYS
2	B	434	SER
2	B	596	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	184	HIS
2	B	465	GLN
2	B	228	HIS
2	B	528	ASN
2	B	366	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

7 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	997[A]	3	11,11,11	0.80	1 (9%)	13,14,14	0.70	0
6	OTR	A	996[B]	3	11,11,11	0.56	0	13,14,14	0.64	0
3	VO4	B	699	5,1,2,6	0,2,4	-	-	-		
5	OTS	A	995[A]	3	11,11,11	0.54	0	13,14,14	0.75	0
3	VO4	A	699	5,1,2,6	0,2,4	-	-	-		
4	SPM	A	999	-	13,13,13	0.37	0	12,12,12	0.85	0
6	OTR	B	998[B]	3	11,11,11	0.80	1 (9%)	13,14,14	0.70	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	997[A]	3	-	3/6/6/6	0/1/1/1
6	OTR	A	996[B]	3	-	6/6/6/6	0/1/1/1
5	OTS	A	995[A]	3	-	3/6/6/6	0/1/1/1
4	SPM	A	999	-	-	7/11/11/11	-
6	OTR	B	998[B]	3	-	6/6/6/6	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	997[A]	OTS	C1-C7	2.19	1.55	1.51
6	B	998[B]	OTR	C1-C7	2.19	1.55	1.51

There are no bond angle outliers.

There are no chirality outliers.

5 of 25 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	995[A]	OTS	C1-C7-C8-N8
5	B	997[A]	OTS	C1-C7-C8-N8
6	A	996[B]	OTR	C1-C7-C8-N8
6	A	996[B]	OTR	O7-C7-C8-N8
6	B	998[B]	OTR	C1-C7-C8-N8

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	699	VO4	3	0
3	A	699	VO4	3	0
4	A	999	SPM	1	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%)	-0.17	0	100	100	25, 25, 34, 49	0
1	F	2/4 (50%)	-0.50	0	100	100	25, 25, 25, 32	0
2	A	425/485 (87%)	-0.17	8 (1%)	66	71	10, 24, 45, 63	0
2	B	434/485 (89%)	-0.25	9 (2%)	63	68	8, 19, 40, 67	0
All	All	864/978 (88%)	-0.21	17 (1%)	65	69	8, 22, 43, 67	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	161	GLY	6.3
2	B	160	LYS	5.9
2	B	608	SER	5.7
2	A	364	GLY	3.8
2	A	528	ASN	3.6

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, 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
4	SPM	A	999	14/14	0.81	0.21	26,38,49,49	0
6	OTR	B	998[B]	11/11	0.92	0.13	19,21,26,30	11
6	OTR	A	996[B]	11/11	0.93	0.12	18,21,28,32	11
5	OTS	B	997[A]	11/11	0.93	0.14	19,21,26,30	11
5	OTS	A	995[A]	11/11	0.96	0.13	18,21,28,32	11
3	VO4	B	699	3/5	1.00	0.10	16,16,16,17	0
3	VO4	A	699	3/5	1.00	0.07	10,10,15,15	0

6.5 Other polymers (i)

There are no such residues in this entry.