



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 09:01 PM GMT

PDB ID : 1N5Y
Title : HIV-1 Reverse Transcriptase Crosslinked to Post-Translocation AZTMP-Terminated DNA (Complex P)
Authors : Sarafianos, S.G.; Clark Jr., A.D.; Das, K.; Tuske, S.; Birktoft, J.J.; Ilankumar, P.; Ramesha, A.R.; Sayer, J.M.; Jerina, D.M.; Boyer, P.L.; Hughes, S.H.; Arnold, E.
Deposited on : 2002-11-07
Resolution : 3.10 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
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

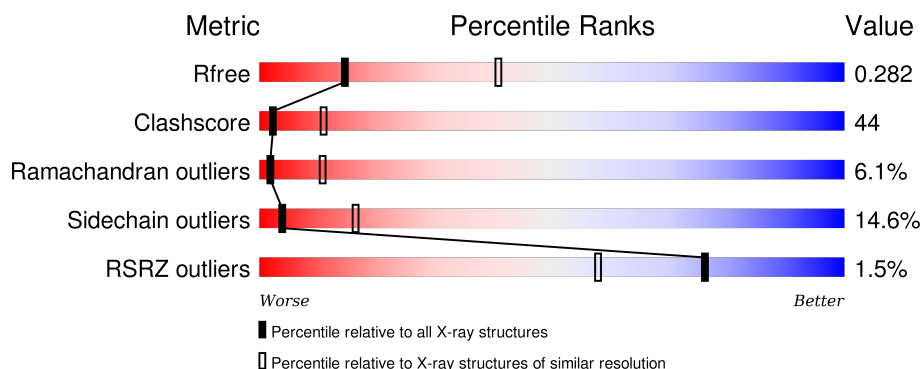
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.10 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}	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	T	27	<div> <div></div> <div>37%</div> <div>37%</div> <div>22%</div> </div>
2	P	21	<div> <div>5%</div> <div>33%</div> <div>57%</div> <div>5%</div> </div>
3	A	558	<div> <div>2%</div> <div>39%</div> <div>48%</div> <div>13%</div> </div>
4	B	430	<div> <div>%</div> <div>33%</div> <div>52%</div> <div>13%</div> </div>
5	L	211	<div> <div>36%</div> <div>55%</div> <div>9%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	H	225	

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
2	MRG	P	817	-	-	X	-
7	MG	A	1001	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 12185 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*TP*GP*C*TP*AP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*GP*TP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	21	Total	C	N	O	P	0	0	0
			432	204	87	121	20			

- Molecule 2 is a DNA chain called 5'-D(*A*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(ATM))-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	P	20	Total	C	N	O	P	S	0	0	0
			408	195	72	121	19	1			

- Molecule 3 is a protein called REVERSE TRANSCRIPTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	558	Total	C	N	O	S	15	0	0
			4482	2901	741	832	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	258	CYS	GLN	ENGINEERED	UNP P03366
A	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 4 is a protein called REVERSE TRANSCRIPTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	429	Total	C	N	O	S	18	0	0
			3534	2304	586	637	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 5 is a protein called monoclonal antibody (light chain).

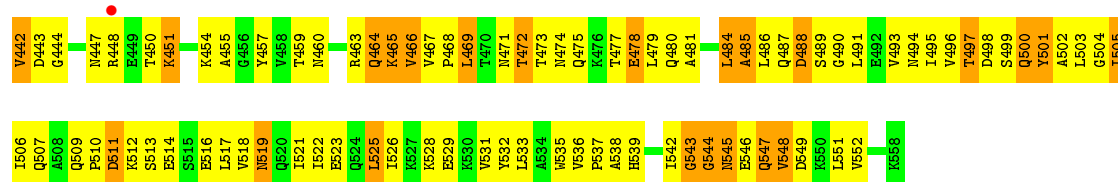
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	L	211	Total	C	N	O	S	0	0	0
			1643	1025	270	342	6			

- Molecule 6 is a protein called monoclonal antibody (heavy chain).

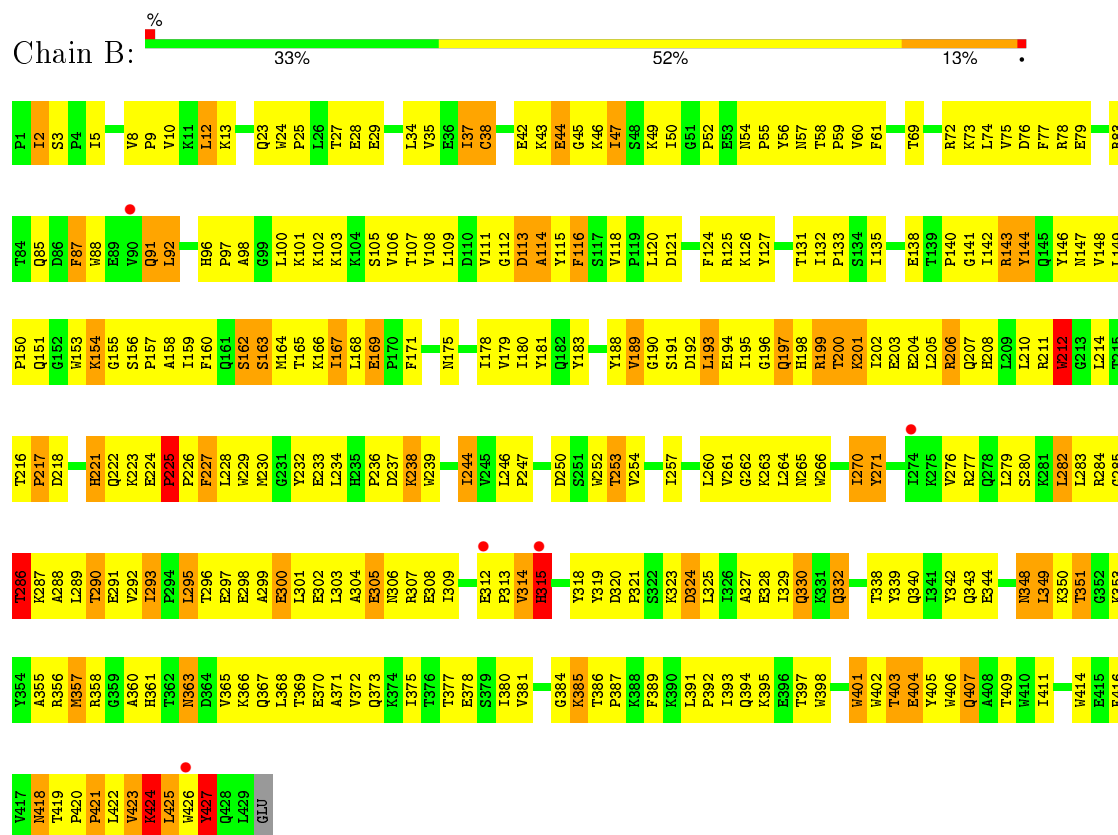
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	H	225	Total	C	N	O	S	0	0	0
			1685	1060	276	340	9			

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

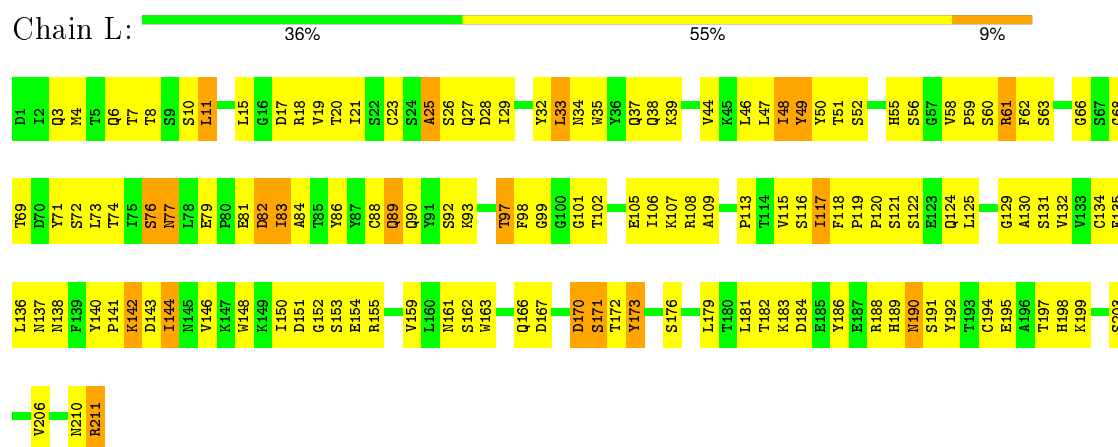
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	Mg	0	0
			1	1		



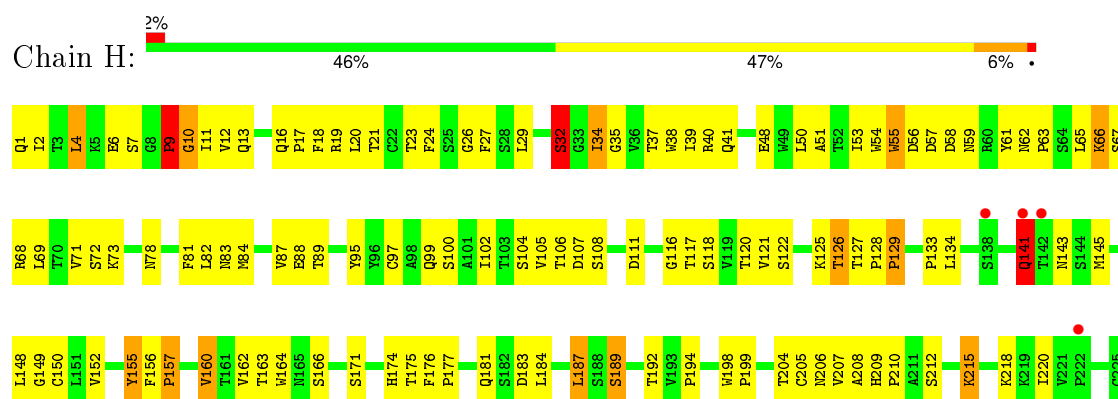
• Molecule 4: REVERSE TRANSCRIPTASE



• Molecule 5: monoclonal antibody (light chain)



• Molecule 6: monoclonal antibody (heavy chain)



4 Data and refinement statistics

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, α , β , γ	166.70Å 166.70Å 221.12Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 3.10 34.36 – 3.10	Depositor EDS
% Data completeness (in resolution range)	91.8 (20.00-3.10) 91.7 (34.36-3.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.54 (at 3.12Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.255 , 0.285 0.253 , 0.282	Depositor DCC
R_{free} test set	2376 reflections (4.07%)	DCC
Wilson B-factor (Å ²)	85.4	Xtriage
Anisotropy	0.042	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 56.4	EDS
Estimated twinning fraction	0.054 for -h,-k,l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 61847 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	12185	wwPDB-VP
Average B, all atoms (Å ²)	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.91% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ATM, MG, MRG

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	T	2.09	16/486 (3.3%)	1.80	20/749 (2.7%)
2	P	2.68	28/400 (7.0%)	2.60	44/612 (7.2%)
3	A	0.62	1/4600 (0.0%)	0.77	1/6259 (0.0%)
4	B	0.69	0/3639	0.83	3/4949 (0.1%)
5	L	0.58	0/1681	0.78	0/2283
6	H	0.63	0/1729	0.84	1/2372 (0.0%)
All	All	0.88	45/12535 (0.4%)	0.98	69/17224 (0.4%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	P	1	0
3	A	0	1
5	L	0	1
All	All	1	2

The worst 5 of 45 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	812	DT	C4-C5	19.16	1.62	1.45
2	P	806	DT	C4-C5	18.46	1.61	1.45
1	T	720	DG	C5-C6	15.98	1.58	1.42
1	T	721	DG	C5-C6	11.98	1.54	1.42
2	P	813	DT	C4-C5	10.86	1.54	1.45

The worst 5 of 69 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	P	806	DT	N3-C4-O4	-17.47	109.42	119.90
2	P	806	DT	C5-C4-O4	15.05	135.44	124.90
2	P	812	DT	N3-C4-O4	-13.97	111.52	119.90
2	P	812	DT	C5-C4-O4	12.77	133.84	124.90
2	P	816	DG	N9-C1'-C2'	12.07	135.53	112.60

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	P	815	DG	C1'

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	354	TYR	Sidechain
5	L	49	TYR	Sidechain

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	T	432	0	235	38	0
2	P	408	0	230	35	0
3	A	4482	0	4484	419	0
4	B	3534	0	3568	359	1
5	L	1643	0	1565	137	0
6	H	1685	0	1640	103	0
7	A	1	0	0	0	0
All	All	12185	0	11722	1038	1

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 44.

The worst 5 of 1038 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)
5:L:61:ARG:HB2	5:L:76:SER:HB3	1.25	1.17

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:H:166:SER:H	6:H:206:ASN:ND2	1.51	1.07
6:H:166:SER:N	6:H:206:ASN:HD21	1.52	1.05
3:A:441:TYR:CE2	3:A:544:GLY:HA3	1.91	1.05
4:B:244:ILE:HD13	4:B:244:ILE:H	1.18	1.04

All (1) 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)
4:B:2:ILE:CD1	4:B:2:ILE:CD1[6_565]	1.62	0.58

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	556/558 (100%)	424 (76%)	102 (18%)	30 (5%)	2	14
4	B	427/430 (99%)	317 (74%)	77 (18%)	33 (8%)	1	6
5	L	209/211 (99%)	165 (79%)	32 (15%)	12 (6%)	2	12
6	H	223/225 (99%)	191 (86%)	21 (9%)	11 (5%)	3	16
All	All	1415/1424 (99%)	1097 (78%)	232 (16%)	86 (6%)	2	11

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	137	ASN
3	A	139	THR
3	A	249	LYS
3	A	254	VAL
3	A	273	GLY

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	
3	A	485/498 (97%)	414 (85%)	71 (15%)	4	16
4	B	388/392 (99%)	324 (84%)	64 (16%)	3	12
5	L	190/190 (100%)	168 (88%)	22 (12%)	7	27
6	H	196/196 (100%)	169 (86%)	27 (14%)	4	19
All	All	1259/1276 (99%)	1075 (85%)	184 (15%)	4	16

5 of 184 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	113	ASP
4	B	244	ILE
6	H	120	THR
4	B	144	TYR
4	B	199	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	182	GLN
4	B	363	ASN
6	H	174	HIS
4	B	235	HIS
4	B	330	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](#)

2 non-standard protein/DNA/RNA residues 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
2	MRG	P	817	1,3,2	20,28,29	2.59	7 (35%)	25,39,42	3.19	11 (44%)
2	ATM	P	822	1,2	13,23,24	1.98	5 (38%)	17,32,35	3.65	7 (41%)

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
2	MRG	P	817	1,3,2	-	0/8/26/27	0/3/3/3
2	ATM	P	822	1,2	-	0/6/24/25	0/2/2/2

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	817	MRG	C8-N7	-4.53	1.25	1.34
2	P	817	MRG	C5-C4	-3.08	1.33	1.40
2	P	817	MRG	C21-N2	-2.91	1.39	1.46
2	P	817	MRG	C4-N3	-2.68	1.31	1.35
2	P	817	MRG	C2-N3	-2.27	1.27	1.34

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	P	817	MRG	C5-C6-N1	-9.69	110.34	123.59
2	P	822	ATM	C5-C4-N3	-4.94	119.64	125.14
2	P	817	MRG	C23-C22-C21	-4.90	96.94	113.12
2	P	817	MRG	C2-N3-C4	-3.58	110.78	115.09
2	P	817	MRG	C6-C5-C4	-2.73	117.64	120.90

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	P	817	MRG	9	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 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	T	21/27 (77%)	0.19	0 100 100	64, 84, 98, 107	0
2	P	18/21 (85%)	0.05	0 100 100	75, 83, 100, 103	0
3	A	556/558 (99%)	-0.10	13 (2%) 64 40	44, 91, 110, 110	1 (0%)
4	B	428/430 (99%)	-0.12	5 (1%) 81 64	35, 80, 110, 110	2 (0%)
5	L	211/211 (100%)	-0.08	0 100 100	56, 88, 110, 110	0
6	H	225/225 (100%)	-0.22	4 (1%) 71 50	51, 78, 105, 110	0
All	All	1459/1472 (99%)	-0.12	22 (1%) 76 58	35, 85, 110, 110	3 (0%)

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
6	H	141	GLN	5.2
3	A	67	ASP	4.5
4	B	315	HIS	3.8
3	A	448	ARG	3.7
3	A	64	LYS	3.6

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
2	MRG	P	817	26/27	0.87	0.20	-	92,92,92,92	0
2	ATM	P	822	22/23	0.92	0.20	-	62,68,87,87	0

6.3 Carbohydrates [i](#)

There are no carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
7	MG	A	1001	1/1	0.98	0.61	6.40	62,62,62,62	0

6.5 Other polymers [i](#)

There are no such residues in this entry.