wwPDB X-ray Structure Validation Summary Report (i)

Oct 17, 2021 – 01:03 AM EDT

PDB ID	:	1N6Q
Title	:	HIV-1 Reverse Transcriptase Crosslinked to pre-translocation AZTMP-terminated DNA (complex N)
Authors	:	Sarafianos, S.G.; Clark Jr., A.D.; Das, K.; Tuske, S.; Birktoft, J.J.; Ilankumaran, I.; Ramesha, A.R.; Sayer, J.M.; Jerina, D.M.; Boyer, P.L.; Hughes, S.H.; Arnold, E.
Deposited on	:	2002-11-11
Resolution	:	3.00 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 following versions of software and data (see references (i)) were used in the production of this report:

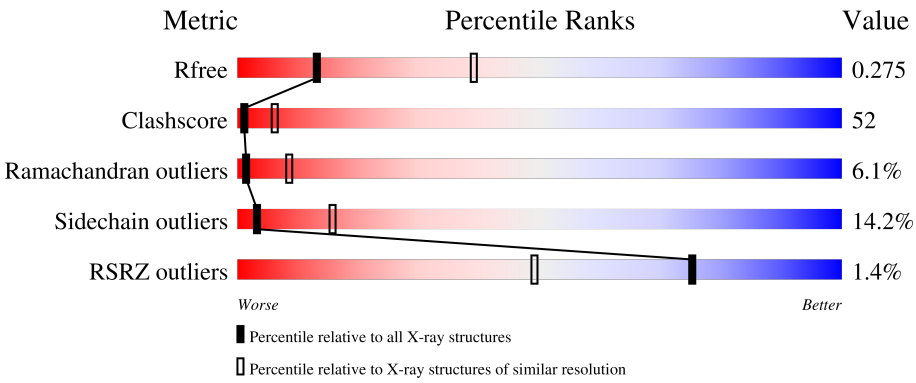
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	T	27	4%; 11%, 70%, •, 15%
2	P	22	14%, 64%, 18%, 5%
3	A	558	3%; 31%, 53%, 15%, •
4	B	430	%; 28%, 58%, 12%, •
5	L	211	28%, 59%, 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	H	225	40% 52% 7%

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	ATM	P	823	-	-	X	-
7	MG	A	1002	-	-	-	X

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 12250 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*T*GP*CP*AP*TP*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
			Total	C	N	O	P			
1	T	23	473	223	95	133	22	0	0	0

- 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*AP*(ATM))-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
2	P	21	429	205	77	126	20	1	0	0	0

- Molecule 3 is a protein called Reverse Transcriptase.

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

There are 2 discrepancies between the modelled and reference sequences:

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

- Molecule 4 is a protein called Reverse Transcriptase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	429	3534	2304	586	637	7	12	0	0

There is a discrepancy between the modelled and reference sequences:

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

- Molecule 5 is a protein called Monoclonal Antibody (Light Chain).

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

- Molecule 6 is a protein called Monoclonal Antibody (Heavy Chain).

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

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
7	A	2	2	2	0	0

- Molecule 8 is water.

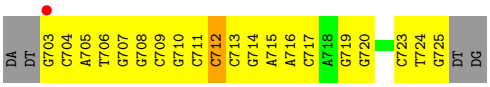
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	A	2	2	2	0	0

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*T*GP*CP*AP*TP*GP*GP*CP*GP*CP*CP*CP*GP*AP*AP*CP*AP*GP*GP*GP*AP*CP*TP*GP*TP*G)-3'

Chain T: 4% 11% 70% • 15%

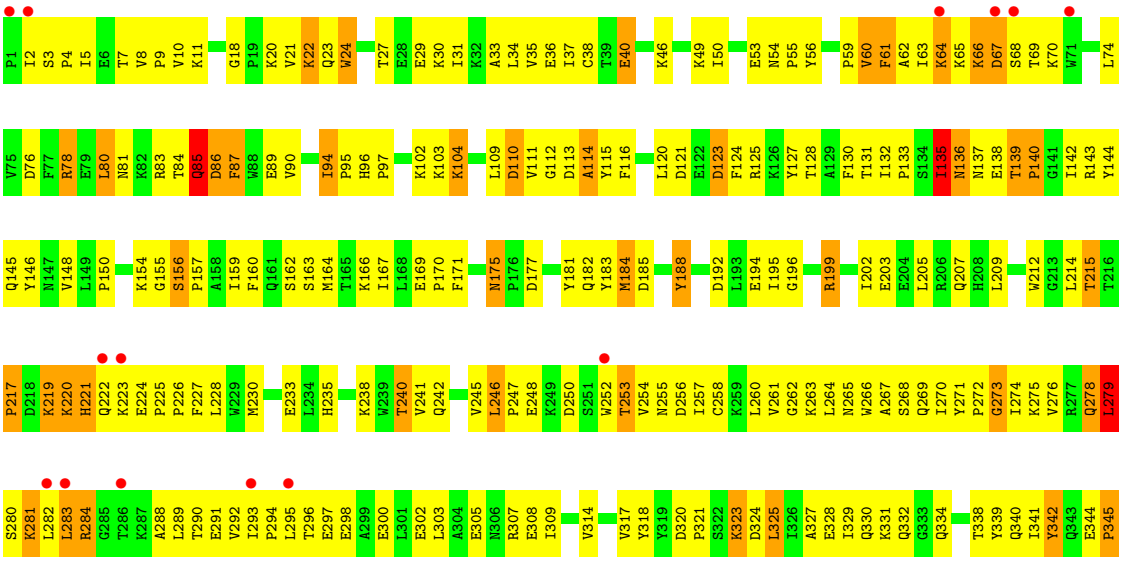


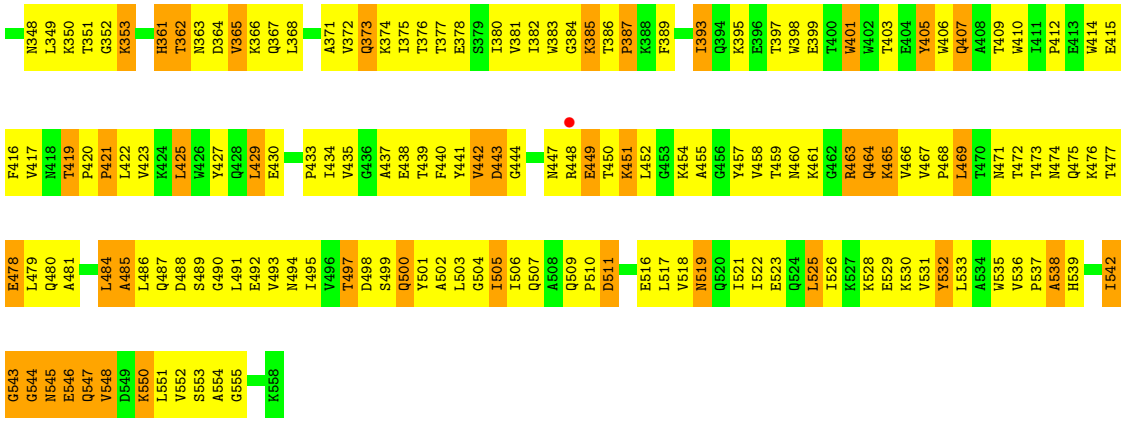
• Molecule 2: 5'-D(*A*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*AP*(ATM))-3'

Chain P: 14% 64% 18% 5%

• Molecule 3: Reverse Transcriptase

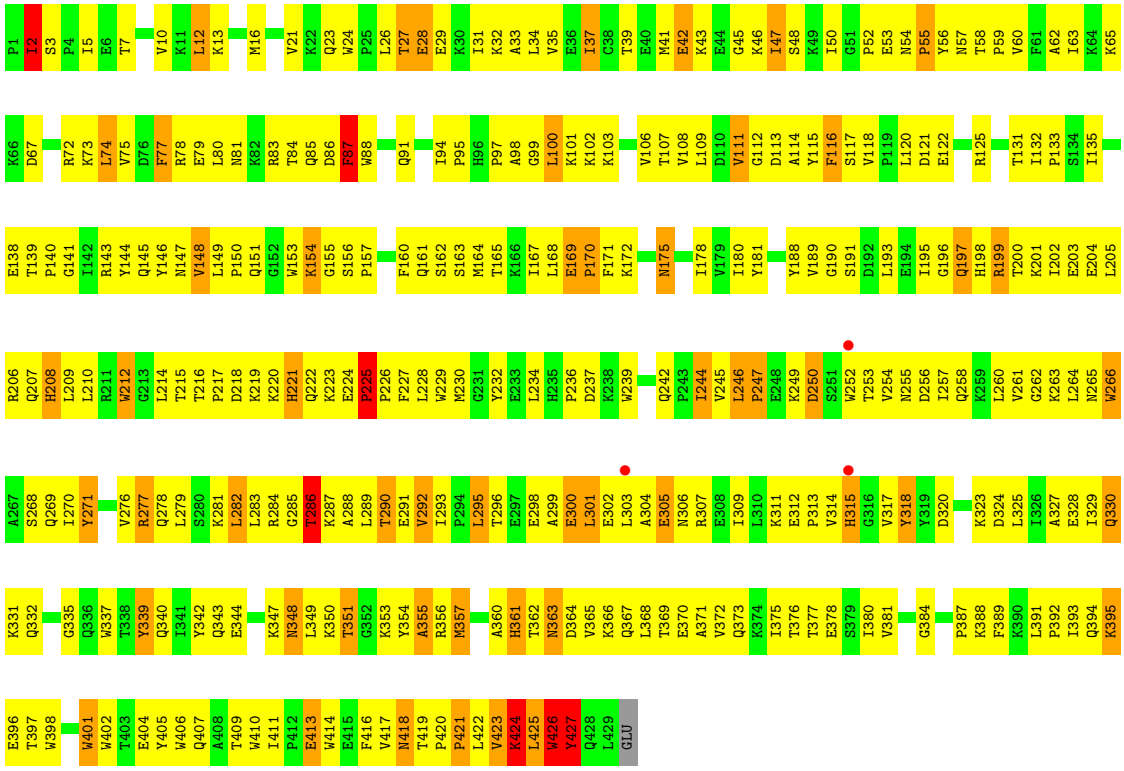
Chain A: 3% 31% 53% 15% •





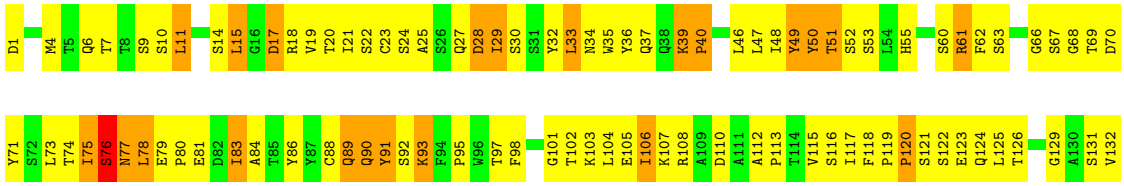
- Molecule 4: Reverse Transcriptase

Chain B:



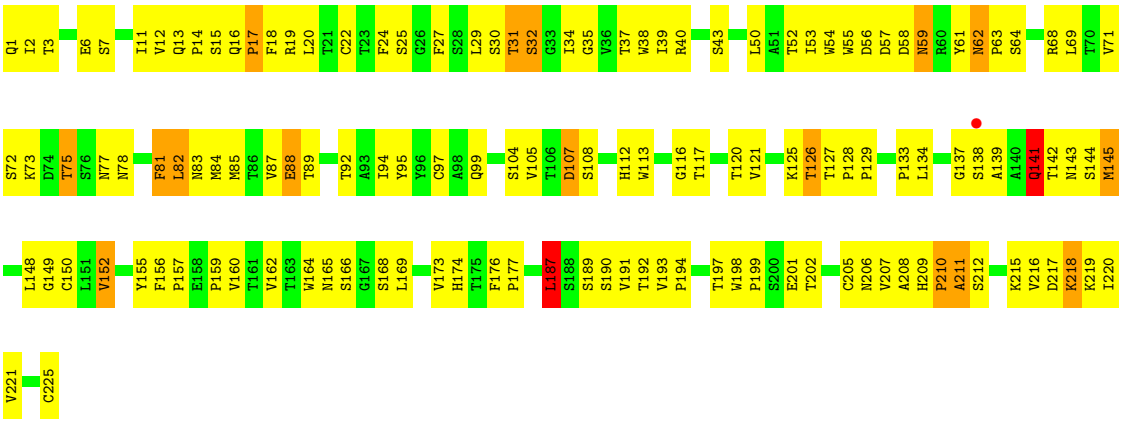
- Molecule 5: Monoclonal Antibody (Light Chain)

Chain L:



• Molecule 6: Monoclonal Antibody (Heavy Chain)

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, α, β, γ	166.35Å 166.35Å 220.96Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 3.00 39.96 – 3.00	Depositor EDS
% Data completeness (in resolution range)	85.0 (20.00-3.00) 85.0 (39.96-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.32 (at 3.01Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.247 , 0.284 0.239 , 0.275	Depositor DCC
R_{free} test set	2631 reflections (4.06%)	wwPDB-VP
Wilson B-factor (Å^2)	67.1	Xtriage
Anisotropy	0.053	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 63.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.052 for -h,-k,l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	12250	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.92% 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: MG, MRG, ATM

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	0.76	0/532	1.04	1/820 (0.1%)
2	P	0.66	0/424	1.13	5/649 (0.8%)
3	A	0.48	2/4600 (0.0%)	0.72	0/6259
4	B	0.59	2/3639 (0.1%)	0.83	6/4949 (0.1%)
5	L	0.50	1/1681 (0.1%)	0.74	0/2283
6	H	0.49	0/1729	0.83	2/2372 (0.1%)
All	All	0.54	5/12605 (0.0%)	0.80	14/17332 (0.1%)

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
4	B	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	425	LEU	C-N	-7.72	1.16	1.34
3	A	248	GLU	CD-OE2	7.38	1.33	1.25
4	B	426	TRP	N-CA	-6.47	1.33	1.46
5	L	50	TYR	C-N	-5.06	1.22	1.34
3	A	421	PRO	C-N	-5.06	1.22	1.34

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	P	815	DG	N9-C1'-C2'	9.97	131.55	112.60
2	P	816	DG	N9-C1'-C2'	7.73	127.28	112.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	427	TYR	N-CA-C	7.72	131.84	111.00
4	B	425	LEU	O-C-N	7.16	134.16	122.70
2	P	815	DG	O4'-C1'-C2'	6.40	111.02	105.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	427	TYR	Sidechain

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	T	473	0	257	34	0
2	P	429	0	242	28	0
3	A	4482	0	4484	506	0
4	B	3534	0	3567	407	0
5	L	1643	0	1564	208	0
6	H	1685	0	1640	137	0
7	A	2	0	0	0	0
8	A	2	0	0	0	0
All	All	12250	0	11754	1256	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 52.

The worst 5 of 1256 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)
6:H:92:THR:HG23	6:H:120:THR:HA	1.34	1.07
3:A:22:LYS:H	3:A:22:LYS:HD3	1.13	1.06
3:A:441:TYR:CE2	3:A:544:GLY:HA3	1.91	1.05
4:B:60:VAL:HG12	4:B:75:VAL:HG22	1.37	1.04
3:A:501:TYR:CE1	3:A:505:ILE:HD11	1.93	1.04

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
3	A	556/558 (100%)	408 (73%)	110 (20%)	38 (7%)	1 6
4	B	427/430 (99%)	333 (78%)	62 (14%)	32 (8%)	1 5
5	L	209/211 (99%)	167 (80%)	31 (15%)	11 (5%)	2 11
6	H	223/225 (99%)	187 (84%)	30 (14%)	6 (3%)	5 26
All	All	1415/1424 (99%)	1095 (77%)	233 (16%)	87 (6%)	1 8

5 of 87 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	66	LYS
3	A	195	ILE
3	A	278	GLN
3	A	345	PRO
3	A	393	ILE

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%)	416 (86%)	69 (14%)	3 16
4	B	388/392 (99%)	328 (84%)	60 (16%)	2 13
5	L	190/190 (100%)	167 (88%)	23 (12%)	5 21

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
6	H	196/196 (100%)	169 (86%)	27 (14%)	3	17
All	All	1259/1276 (99%)	1080 (86%)	179 (14%)	3	16

5 of 179 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	351	THR
5	L	91	TYR
4	B	364	ASP
5	L	10	SER
5	L	180	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	394	GLN
5	L	210	ASN
4	B	418	ASN
5	L	90	GLN
6	H	62	ASN

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	3,1,2	22,28,29	2.25	3 (13%)	23,39,42	4.05	11 (47%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	ATM	P	823	1,2	16,23,24	1.46	4 (25%)	17,32,35	3.21	3 (17%)

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	3,1,2	-	2/8/26/27	0/3/3/3
2	ATM	P	823	1,2	-	2/7/24/25	0/2/2/2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	817	MRG	C2-N2	7.29	1.45	1.34
2	P	817	MRG	C21-N2	-5.56	1.34	1.45
2	P	823	ATM	N4'-N3'	3.07	1.31	1.23
2	P	823	ATM	C3'-N3'	-2.56	1.41	1.48
2	P	823	ATM	C4-N3	2.44	1.37	1.33

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	P	817	MRG	C21-N2-C2	-14.08	99.41	123.75
2	P	823	ATM	C4-N3-C2	11.80	125.11	115.14
2	P	817	MRG	C5-C6-N1	-7.75	112.83	123.43
2	P	817	MRG	C6-N1-C2	5.85	125.65	115.18
2	P	817	MRG	C23-C22-C21	-4.18	99.22	112.65

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	P	817	MRG	N2-C21-C22-C23
2	P	823	ATM	C3'-N3'-N4'-N5'
2	P	823	ATM	O4'-C1'-N1-C6
2	P	817	MRG	C21-C22-C23-S24

There are no ring outliers.

2 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	P	817	MRG	1	0
2	P	823	ATM	8	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
4	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	425:LEU	C	426:TRP	N	1.16

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	T	23/27 (85%)	0.02	1 (4%) 35 13	57, 99, 110, 121	0
2	P	19/22 (86%)	-0.06	0 100 100	74, 88, 106, 110	0
3	A	556/558 (99%)	-0.06	15 (2%) 54 26	35, 80, 110, 110	1 (0%)
4	B	428/430 (99%)	-0.27	3 (0%) 87 69	27, 62, 108, 110	1 (0%)
5	L	211/211 (100%)	-0.24	0 100 100	39, 71, 106, 110	0
6	H	225/225 (100%)	-0.34	1 (0%) 92 79	35, 62, 99, 110	0
All	All	1462/1473 (99%)	-0.19	20 (1%) 75 49	27, 71, 109, 121	2 (0%)

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	283	LEU	3.8
4	B	315	HIS	3.5
3	A	252	TRP	3.1
3	A	448	ARG	2.9
3	A	2	ILE	2.8

6.2 Non-standard residues in protein, DNA, RNA chains (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
2	MRG	P	817	26/27	0.91	0.16	79,79,79,80	0
2	ATM	P	823	22/23	0.94	0.16	62,68,81,86	0

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
7	MG	A	1002	1/1	0.70	0.43	58,58,58,58	0
7	MG	A	1001	1/1	0.98	0.35	54,54,54,54	0

6.5 Other polymers (i)

There are no such residues in this entry.