wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:18 PM GMT

PDB ID	:	1T03
Title	:	HIV-1 reverse transcriptase crosslinked to tenofovir terminated template-primer (complex P)
Authors	:	Tuske, S.; Sarafianos, S.G.; Ding, J.; Arnold, E.
Deposited on	:	2004-04-07
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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

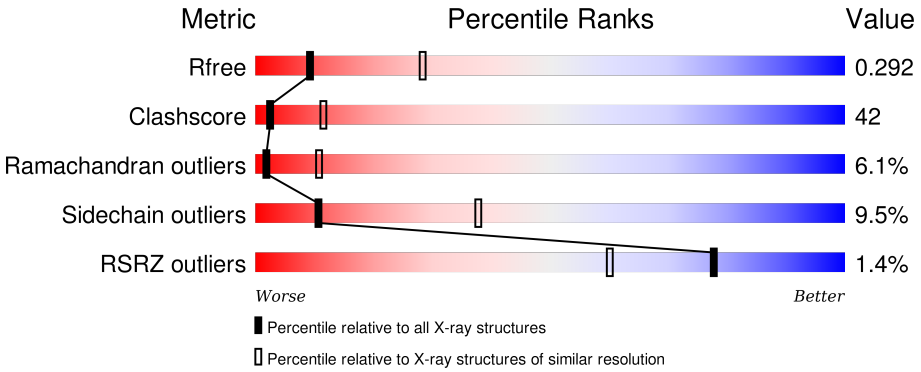
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

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.



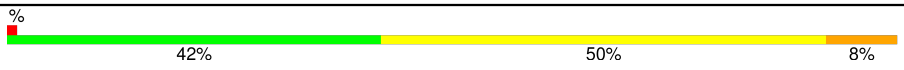
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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 <=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% 33% 44% 7% 15%
2	P	21	24% 62% 10% 5%
3	A	558	3% 41% 51% 8%
4	B	437	36% 51% 10%
5	L	211	39% 53% 7%

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
7	MG	A	559	-	-	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 12252 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 Synthetic oligonucleotide template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	23	Total	C	N	O	P	0	0	0
			473	223	95	133	22			

- Molecule 2 is a DNA chain called Synthetic oligonucleotide primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	20	Total	C	N	O	P	0	2	0
			434	209	80	125	20			

- Molecule 3 is a protein called POL polyprotein.

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 POL polyprotein.

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

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366
B	430	GLY	-	CLONING ARTIFACT	UNP P03366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	431	GLY	-	CLONING ARTIFACT	UNP P03366
B	432	HIS	-	EXPRESSION TAG	UNP P03366
B	433	HIS	-	EXPRESSION TAG	UNP P03366
B	434	HIS	-	EXPRESSION TAG	UNP P03366
B	435	HIS	-	EXPRESSION TAG	UNP P03366
B	436	HIS	-	EXPRESSION TAG	UNP P03366
B	437	HIS	-	EXPRESSION TAG	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).

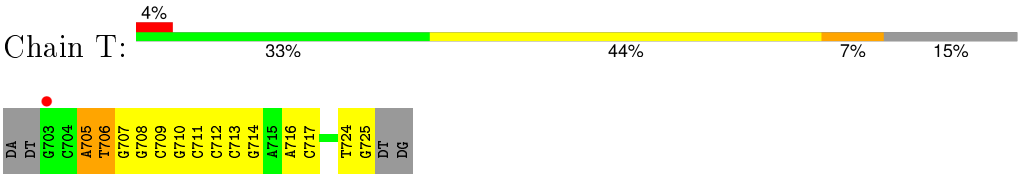
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	Mg	0	0
			1	1		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Synthetic oligonucleotide template

Chain T:

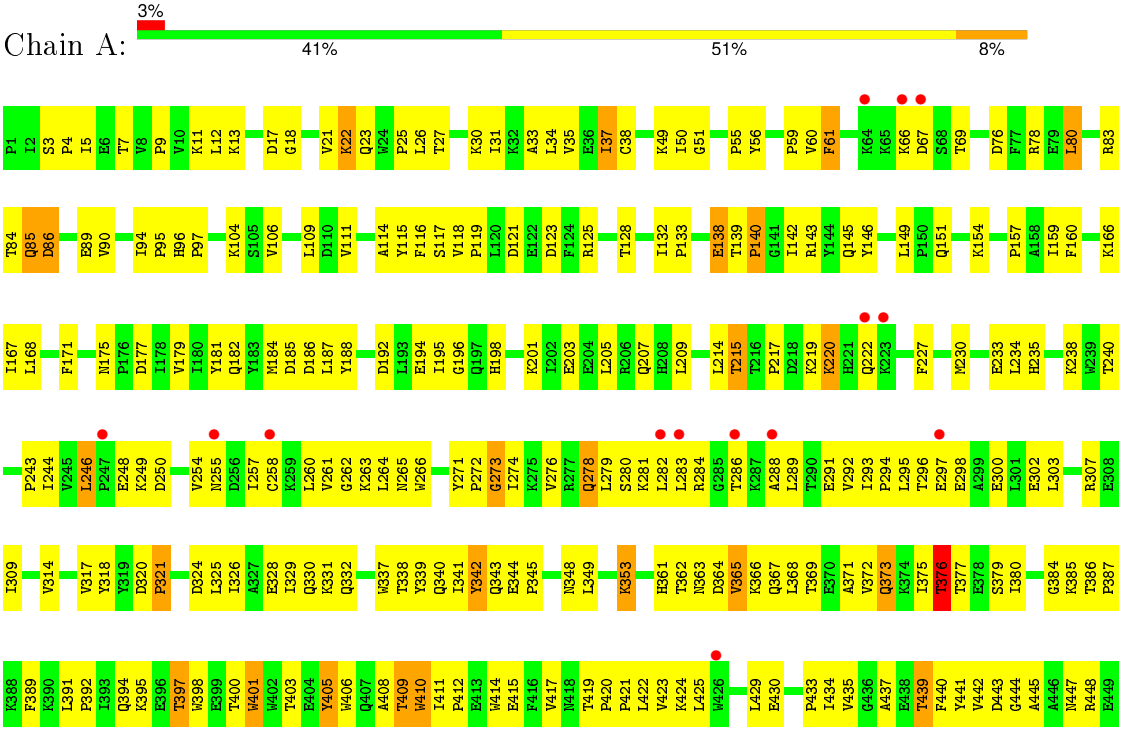


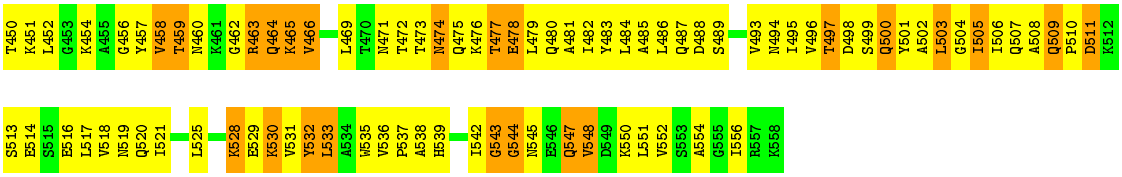
- Molecule 2: Synthetic oligonucleotide primer

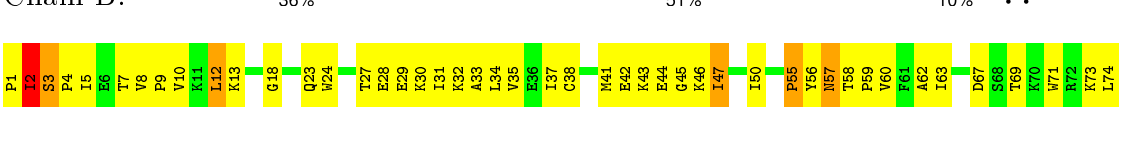
Chain P:

- Molecule 3: POL polyprotein

Chain A:

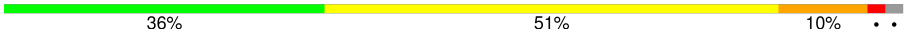


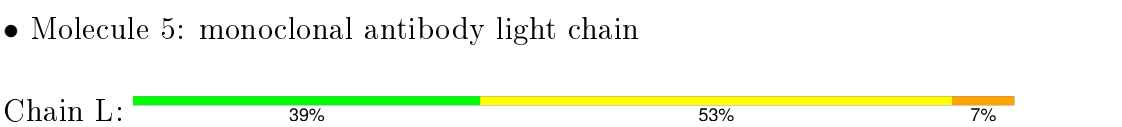




• Molecule 4: POL polyprotein

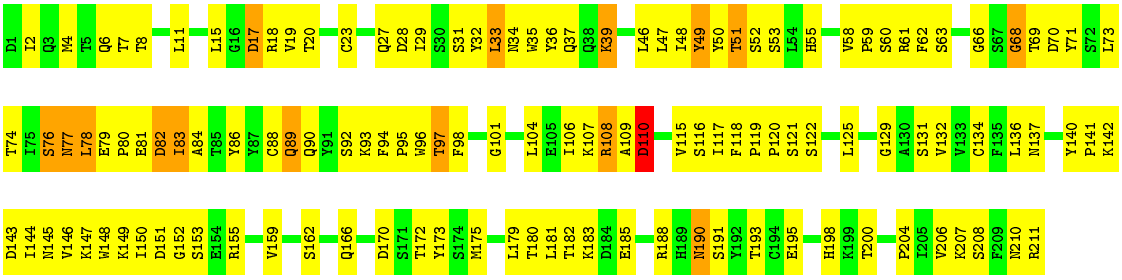
Chain B:





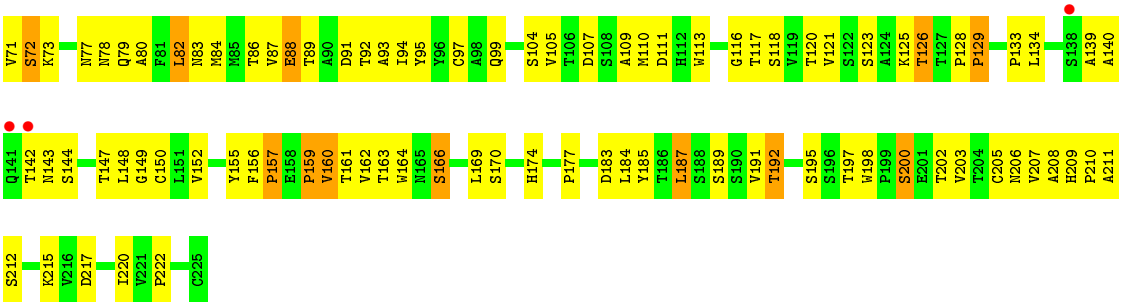
• Molecule 5: monoclonal antibody light chain

Chain L:



• Molecule 6: monoclonal antibody heavy chain

Chain H:


V71 S72 K73 W77 N78 Q79 A80 F81 L82 N83 M84 M85 T86 V87 E88 T89 A90 D91 T92 A93 I94 Y95 Y96 C97 A98 Q99 S104 V105 T106 D107 S108 A109 M110 D111 H112 W113 G116 T117 S118 V119 T120 V121 S122 S123 A124 K125 T126 T127 P128 P129 P133 L134 S138 A139 A140
Q141 T142 N143 S144 T147 L148 G149 C150 L151 V152 Y155 F156 P157 E158 P159 V160 T161 V162 T163 W164 N165 S166 L169 S170 H174 P177 D183 L184 Y185 T186 L187 S188 S189 S190 V191 T192 S195 S196 T197 W198 P199 S200 E201 T202 V203 T204 C205 N206 V207 A208 H209 P210 A211
S212 K215 V216 D217 I220 V221 P222 C225

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, α, β, γ	166.78Å 166.78Å 221.13Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 3.10 32.57 – 3.10	Depositor EDS
% Data completeness (in resolution range)	93.9 (20.00-3.10) 95.0 (32.57-3.10)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.01 (at 3.12Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.256 , 0.295 0.251 , 0.292	Depositor DCC
R_{free} test set	2446 reflections (4.04%)	DCC
Wilson B-factor (Å^2)	70.1	Xtriage
Anisotropy	0.010	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 45.5	EDS
Estimated twinning fraction	0.057 for -h,-k,l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Outliers	0 of 60707 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	12252	wwPDB-VP
Average B, all atoms (Å^2)	64.0	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, 0.375 respectively for untwinned datasets, and 0.333, 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: MG, MRG, TFO

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	0.73	2/532 (0.4%)	1.04	4/820 (0.5%)
2	P	0.90	2/420 (0.5%)	1.01	3/640 (0.5%)
3	A	0.45	1/4600 (0.0%)	0.65	0/6259
4	B	0.52	1/3639 (0.0%)	0.74	3/4949 (0.1%)
5	L	0.43	0/1681	0.71	1/2283 (0.0%)
6	H	0.48	0/1729	0.77	0/2372
All	All	0.51	6/12601 (0.0%)	0.74	11/17323 (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

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	248	GLU	CD-OE2	7.06	1.33	1.25
2	P	821[A]	DC	C2'-C1'	6.65	1.58	1.52
2	P	821[B]	DC	C2'-C1'	6.65	1.58	1.52
1	T	705	DA	C3'-C2'	-6.58	1.44	1.52
4	B	212	TRP	CB-CG	-5.57	1.40	1.50

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	T	706	DT	O5'-P-OP1	8.87	121.34	110.70
1	T	705	DA	OP1-P-O3'	-8.69	86.08	105.20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	T	705	DA	N9-C1'-C2'	-7.34	98.66	112.60
2	P	821[A]	DC	N1-C1'-C2'	-7.32	98.69	112.60
2	P	821[B]	DC	N1-C1'-C2'	-7.32	98.69	112.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	188	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	32	0
2	P	434	0	248	23	0
3	A	4482	0	4485	398	0
4	B	3534	0	3568	318	1
5	L	1643	0	1565	160	0
6	H	1685	0	1640	119	0
7	A	1	0	0	0	0
All	All	12252	0	11763	1008	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 42.

The worst 5 of 1008 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)
3:A:501:TYR:CE1	3:A:505:ILE:HD11	1.86	1.10
3:A:50:ILE:HD11	3:A:145:GLN:HB2	1.10	1.09
3:A:344:GLU:HG3	3:A:345:PRO:HD2	1.36	1.07
3:A:138:GLU:HG2	3:A:139:THR:H	1.10	1.06
5:L:61:ARG:HB2	5:L:76:SER:HB3	1.36	1.06

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:2:ILE:O	4:B:2:ILE:O[6_565]	1.67	0.53

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%)	431 (78%)	92 (16%)	33 (6%)	2 12
4	B	427/437 (98%)	324 (76%)	75 (18%)	28 (7%)	1 9
5	L	209/211 (99%)	165 (79%)	34 (16%)	10 (5%)	3 17
6	H	223/225 (99%)	185 (83%)	23 (10%)	15 (7%)	1 9
All	All	1415/1431 (99%)	1105 (78%)	224 (16%)	86 (6%)	2 11

5 of 86 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	67	ASP
3	A	138	GLU
3	A	273	GLY
3	A	466	VAL
3	A	474	ASN

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%)	444 (92%)	41 (8%)	13 45
4	B	388/397 (98%)	343 (88%)	45 (12%)	7 27
5	L	190/190 (100%)	178 (94%)	12 (6%)	22 58
6	H	196/196 (100%)	174 (89%)	22 (11%)	7 29
All	All	1259/1281 (98%)	1139 (90%)	120 (10%)	11 38

5 of 120 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	215	THR
4	B	277	ARG
6	H	123	SER
4	B	216	THR
4	B	234	LEU

Some 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	151	GLN
4	B	242	GLN
6	H	62	ASN
4	B	161	GLN
4	B	175	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)

3 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	1,2	16,24,29	1.16	1 (6%)	20,35,42	2.72	4 (20%)
2	TFO	P	822[A]	2	12,19,20	3.88	6 (50%)	8,26,29	4.01	3 (37%)
2	TFO	P	822[B]	2	12,19,20	0.98	1 (8%)	8,26,29	1.05	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
2	MRG	P	817	1,2	-	0/3/21/27	0/3/3/3
2	TFO	P	822[A]	2	-	0/5/9/10	0/2/2/2
2	TFO	P	822[B]	2	-	0/5/9/10	0/2/2/2

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	822[A]	TFO	C6'-N9	-8.31	1.39	1.48
2	P	822[A]	TFO	O9'-C7'	-7.02	1.36	1.43
2	P	822[A]	TFO	C4-N3	-3.50	1.30	1.35
2	P	822[A]	TFO	C5-C4	-2.77	1.34	1.40
2	P	822[A]	TFO	C2-N3	-2.23	1.28	1.32

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	P	817	MRG	C5-C6-N1	-8.91	111.40	123.59
2	P	822[A]	TFO	C9'-O9'-C7'	-8.01	107.93	114.35
2	P	822[A]	TFO	N3-C2-N1	-4.49	125.45	128.89
2	P	817	MRG	O3'-C3'-C2'	-2.94	101.00	110.74
2	P	817	MRG	N3-C2-N1	-2.49	123.65	127.44

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	P	817	MRG	3	0
2	P	822[A]	TFO	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	P	822[B]	TFO	5	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, 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	T	23/27 (85%)	0.15	1 (4%) 39 18	53, 99, 132, 135	0
2	P	18/21 (85%)	0.12	0 100 100	61, 87, 109, 121	0
3	A	556/558 (99%)	-0.24	14 (2%) 61 37	16, 69, 117, 131	1 (0%)
4	B	429/437 (98%)	-0.40	2 (0%) 91 83	15, 48, 112, 128	1 (0%)
5	L	211/211 (100%)	-0.31	0 100 100	26, 60, 105, 116	0
6	H	225/225 (100%)	-0.43	3 (1%) 79 62	16, 48, 92, 129	0
All	All	1462/1479 (98%)	-0.31	20 (1%) 78 60	15, 58, 113, 135	2 (0%)

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	315	HIS	4.3
6	H	141	GLN	3.9
3	A	222	GLN	3.6
3	A	258	CYS	3.1
6	H	138	SER	2.8

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, 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	LLDF	B-factors(Å^2)	Q<0.9
2	MRG	P	817	22/27	0.94	0.15	-	97,109,112,115	0
2	TFO	P	822[B]	18/19	0.80	0.42	-	86,98,98,98	18

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
2	TFO	P	822[A]	18/19	0.80	0.42	-	85,98,98,98	18

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, 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	LLDF	B-factors(Å^2)	Q<0.9
7	MG	A	559	1/1	0.98	0.26	2.55	18,18,18,18	0

6.5 Other polymers (i)

There are no such residues in this entry.