WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 07:27 AM GMT

PDB ID	:	3AVW
Title	:	Structure of viral RNA polymerase complex 4
Authors	:	Takeshita, D.; Tomita, K.
Deposited on	:	2011-03-08
Resolution	:	2.60 Å(reported)

The following versions of software and data (see references (i)) 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 2.60 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.
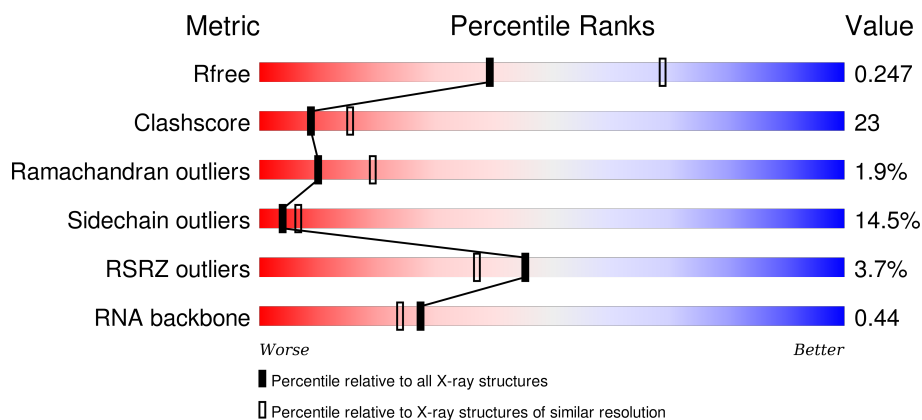
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)
RNA backbone	2183	1022 (3.00-2.20)

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 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	A	1289	3% — 53% / 32% / 8% / 7%
2	G	8	13% / 63% / 25%
3	T	12	17% — 50% / 42% / 8%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 9843 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 protein called Elongation factor Ts, Elongation factor Tu, LINKER, Q beta replicase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1203	Total 9287	C 5865	N 1605	O 1772	S 45	0	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	284	HIS	-	LINKER	UNP P0A6P3
A	1284	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1285	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1286	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1287	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1288	HIS	-	EXPRESSION TAG	UNP Q8LTE0
A	1289	HIS	-	EXPRESSION TAG	UNP Q8LTE0

- Molecule 2 is a RNA chain called RNA (5'-R(*GP*GP*GP*UP*CP*CP*AP*C)-3').

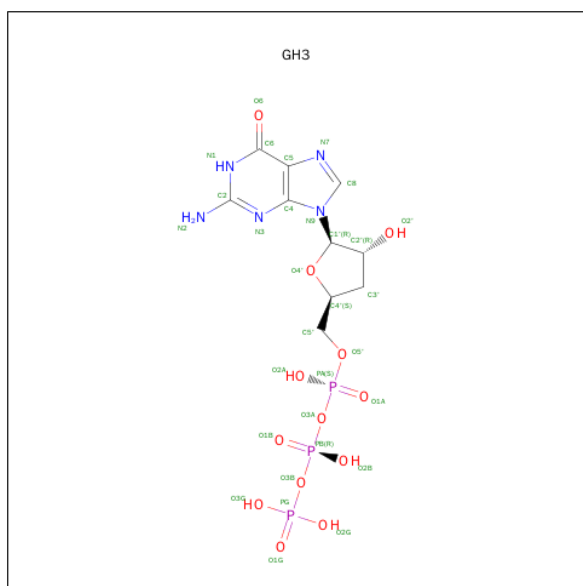| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	8	Total 167	C 76	N 31	O 53	P 7	0	0	0

- Molecule 3 is a RNA chain called RNA (5'-R(*AP*UP*CP*GP*UP*GP*GP*AP*CP*CP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	12	Total 252	C 114	N 46	O 81	P 11	0	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total 2	Ca 2	0	0

- Molecule 5 is 3'-DEOXY-GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GH3) (formula: $C_{10}H_{16}N_5O_{13}P_3$).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
5	T	1	31	10	5	13	3	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	92	Total 92	O 92	0	0
6	G	3	Total 3	O 3	0	0
6	T	9	Total 9	O 9	0	0

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: Elongation factor Ts, Elongation factor Tu, LINKER, Q beta replicase

D947 Q948 T949 I950 N951 Q952 R953 R954 A955 H956 E957 V960 T961 N962 N963 D68 L69 S970 A971 A972 C981 L984 W89 L96 R97 S998 P999 L1003 P1004 E1012 K1013 I1014 S1015 S1016 N1019 G1020 Y1021 T1022 F1023 E1024 L1025 E1026 I1029 L1033 A1034 R1035 I1040

T1049 D1053 D1054 I1055 I1056 L1057 E1067 V1068 F1069 K1070 Y1071 V1072 G1073 F1074 T1075 T1076 N1077 T1081 E1089 S1090 C1091 G1092 K1093 H1094 V1099 D1100 V1101 T1102 Y1105 I1106 R1109 L1116 I1117 L1118 N1121 N1122 L1123 Y1124 T1128 R1145 L1148 P1149 K1150 Q1151 L1152 Q1153 R1154

G1160 V1167 I1172 N1173 P1174 F1175 A1176 K1177 N1178 R1179 I1182 R1183 Y1184 V1185 P1186 R1193 D1194 L1200 G1201 L1204 Y1205 D1206 R1210 N1216 ASP GLY LEU PRO LEU ARG GLY PRO SER GLY CYS ASP SER ALA ASP LEU PHE A1234 I1235 D1236 Q1237 L1238 I1239 C1240 R1241 S1242 N1243 P1244

T1245 K1246 S1250 T1251 F1254 Q1257 C1261 S1262 S1263 R1264 VAL LEU ALA PRO TYR GLY VAL PHE GLN GLY THR LYS VAL ALA SER LEU HIS GLU ALA HIS HIS HIS HIS HIS HIS

- Molecule 2: RNA (5'-R(*GP*GP*GP*UP*CP*CP*AP*C)-3')

Chain G: 13% 63% 25%

G2001 G2002 G2003 U2004 C2005 C2006 A2007 C2008
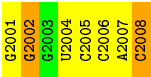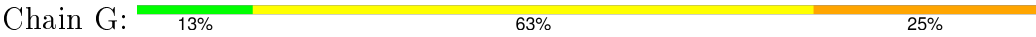
- Molecule 3: RNA (5'-R(*AP*UP*CP*GP*UP*GP*GP*AP*CP*CP*CP*A)-3')

Chain T: 17% 50% 42% 8%

A2101 U2102 C2103 C2109 C2110 C2111 A2112

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	140.02Å 256.83Å 101.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.99 – 2.60 41.84 – 2.60	Depositor EDS
% Data completeness (in resolution range)	94.1 (19.99-2.60) 98.8 (41.84-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 2.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, R_{free}	0.201 , 0.253 0.201 , 0.247	Depositor DCC
R_{free} test set	2824 reflections (5.34%)	DCC
Wilson B-factor (Å²)	52.4	Xtriage
Anisotropy	0.366	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.32 , 51.0	EDS
Estimated twinning fraction	0.017 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.024 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Outliers	0 of 55818 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9843	wwPDB-VP
Average B, all atoms (Å²)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.72% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, GH3

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	A	0.49	1/9456 (0.0%)	0.62	0/12787
2	G	0.59	0/186	1.03	1/288 (0.3%)
3	T	0.73	0/281	1.09	1/436 (0.2%)
All	All	0.50	1/9923 (0.0%)	0.65	2/13511 (0.0%)

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
1	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	981	CYS	CB-SG	-5.05	1.73	1.81

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	2008	C	O4'-C1'-N1	5.67	112.74	108.20
3	T	2103	C	C6-N1-C2	-5.11	118.26	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	849	SER	Peptide

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	A	9287	0	9273	445	0
2	G	167	0	87	6	0
3	T	252	0	132	4	0
4	A	2	0	0	0	0
5	T	31	0	12	1	0
6	A	92	0	0	2	0
6	G	3	0	0	0	0
6	T	9	0	0	0	0
All	All	9843	0	9504	452	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 23.

The worst 5 of 452 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)
1:A:1019:ASN:HB3	1:A:1022:THR:HG22	1.25	1.15
1:A:491:ILE:HD12	1:A:491:ILE:H	1.17	1.06
1:A:968:ASP:H	1:A:1081:THR:HG22	1.26	0.98
1:A:949:THR:HG23	1:A:953:ARG:HH12	1.27	0.97
1:A:961:THR:HG22	1:A:963:ASN:H	1.26	0.97

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
1	A	1193/1289 (93%)	1070 (90%)	100 (8%)	23 (2%)	10 19

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	34	ILE
1	A	104	LYS
1	A	250	ASN
1	A	381	ALA
1	A	806	ARG

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
1	A	995/1060 (94%)	851 (86%)	144 (14%)	4 6

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	622	VAL
1	A	791	LEU
1	A	1182	ILE
1	A	628	LEU
1	A	717	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	763	ASN
1	A	901	ASN
1	A	1237	GLN
1	A	862	HIS
1	A	307	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	G	6/8 (75%)	1 (16%)	0
3	T	11/12 (91%)	2 (18%)	0
All	All	17/20 (85%)	3 (17%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	G	2002	G
3	T	2102	U
3	T	2111	C

There are no RNA pucker outliers to report.

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

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	GH3	T	2501	4	25,33,33	2.60	7 (28%)	29,52,52	2.56	11 (37%)

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	GH3	T	2501	4	-	0/18/34/34	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	T	2501	GH3	C8-N7	-5.37	1.24	1.34
5	T	2501	GH3	C2'-C1'	-3.05	1.48	1.53
5	T	2501	GH3	C6-C5	-2.25	1.36	1.41
5	T	2501	GH3	C6-N1	2.24	1.37	1.33
5	T	2501	GH3	C5-C4	2.61	1.46	1.40

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	T	2501	GH3	C1'-N9-C4	-9.86	112.06	126.94
5	T	2501	GH3	N3-C2-N1	-3.42	122.23	127.44
5	T	2501	GH3	C5-C6-N1	-2.75	119.83	123.59
5	T	2501	GH3	C6-C5-C4	-2.59	117.80	120.90
5	T	2501	GH3	PA-O3A-PB	-2.47	125.80	132.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	T	2501	GH3	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	A	1203/1289 (93%)	-0.08	43 (3%)	46	38	20, 51, 100, 127	0
2	G	8/8 (100%)	-0.09	0	100	100	35, 52, 83, 84	0
3	T	12/12 (100%)	0.19	2 (16%)	2	1	32, 49, 91, 118	0
All	All	1223/1309 (93%)	-0.08	45 (3%)	45	37	20, 51, 100, 127	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	701	SER	7.0
1	A	191	VAL	6.1
1	A	703	ASN	5.6
3	T	2101	A	5.6
1	A	702	ARG	5.3

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 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(Å^2)	Q<0.9
5	GH3	T	2501	31/31	0.98	0.20	-0.15	29,33,39,39	0
4	CA	A	3002	1/1	0.92	0.40	-	61,61,61,61	0
4	CA	A	3001	1/1	0.98	0.21	-	33,33,33,33	0

6.5 Other polymers (i)

There are no such residues in this entry.