wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 07:21 AM GMT

PDB ID	:	3AGP
Title	:	Structure of viral polymerase form I
Authors	:	Takeshita, D.; Tomita, K.
Deposited on	:	2010-04-06
Resolution	:	2.80 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:

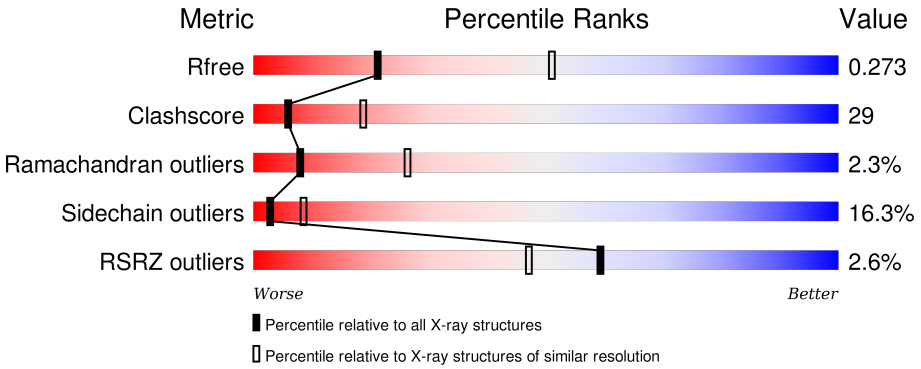
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	2% 47% 37% 9% 7%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9311 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	1202	Total	C	N	O	S	0	0	0
			9274	5857	1603	1769	45			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ca	0	0
			1	1		

- Molecule 3 is water.

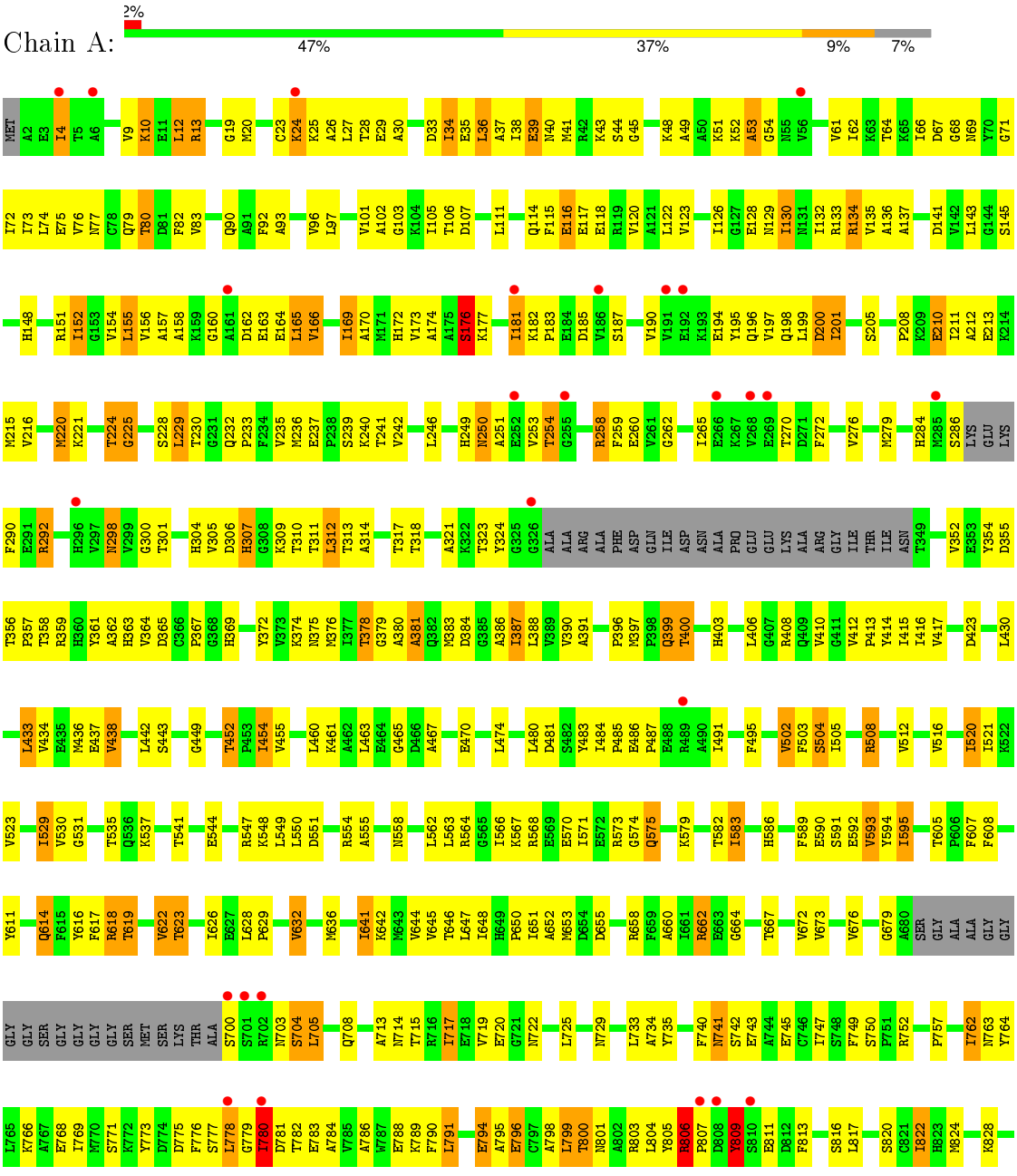
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	36	Total	O	0	0
			36	36		

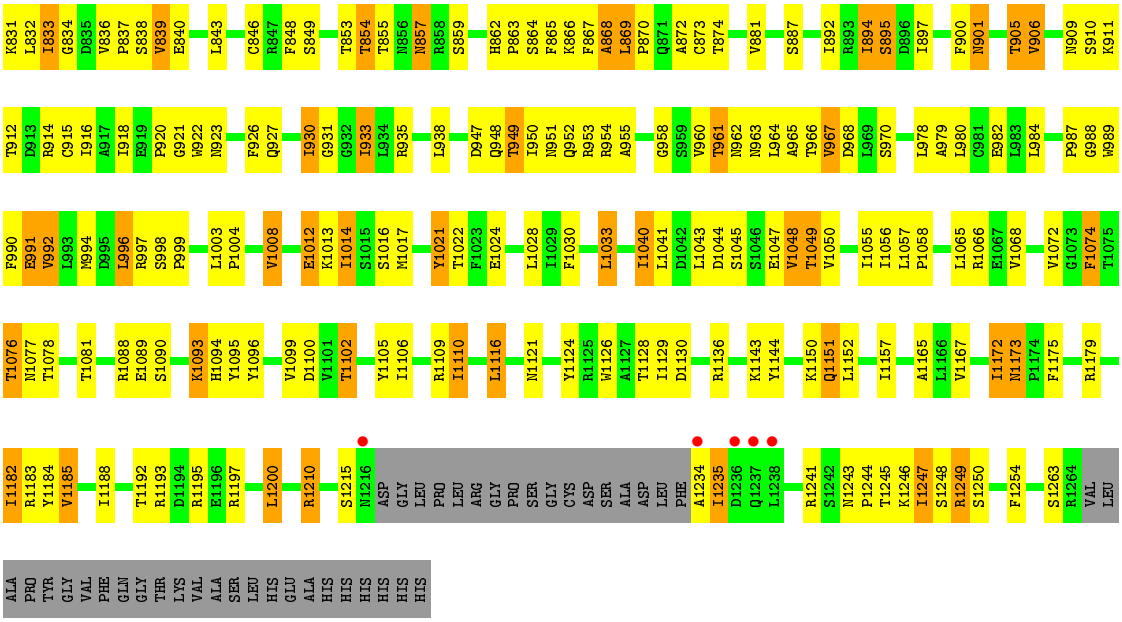
3 Residue-property plots (i)

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

Chain A:




K831 L832 I833 G834 D835 V836 P837 S838 V839 E840 L843 C846 R847 F848 S849 T853 T854 T855 N856 N857 R858 S859 H862 P863 S864 F865 K866 F867 A868 L869 P870 Q871 A872 C873 T874 V881 S887 I892 R893 I894 S895 D896 I897 F900 N901 T905 V906 N909 S910 K911
T912 D913 R914 C915 I916 A917 I918 E919 P920 G921 W922 N923 F926 Q927 I930 G931 G932 I933 L934 R935 L938 D947 Q949 T949 I950 N951 Q952 R953 R954 A955 G958 S959 V960 T961 N962 N963 L964 A965 T966 V967 D968 L969 S970 L978 A979 L980 C981 E982 L983 L984 P987 G988 W989
F990 E991 V992 L993 M994 D995 L996 R997 S998 P999 L1003 P1004 V1008 E1012 K1013 I1014 S1015 S1016 M1017 Y1021 T1022 F1023 E1024 L1028 I1029 F1030 L1033 I1040 L1041 D1042 L1043 D1044 S1045 S1046 E1047 V1048 T1049 V1050 I1055 I1056 L1057 P1058 L1065 R1066 E1067 V1068 V1072 G1073 F1074 T1075
T1076 N1077 T1078 T1081 R1088 E1089 S1090 K1093 H1094 Y1095 V1096 V1099 D1100 V1101 T1102 Y1105 I1106 R1109 I1110 L1116 N1121 Y1124 R1125 W1126 A1127 T1128 I1129 D1130 R1136 K1143 Y1144 K1150 Q1151 L1152 I1157 A1165 L1166 V1167 I1172 N1173 P1174 F1175 R1179
I1182 R1183 Y1184 V1185 I1188 T1192 R1193 D1194 R1195 E1196 R1197 L1200 R1210 S1215 N1216 ASP GLY LEU PRO LEU ARG GLY PRO SER GLY CYS ASP SER ALA ASP LEU PHE A1234 I1235 D1236 Q1237 L1238 R1241 S1242 N1243 P1244 T1245 K1246 I1247 S1248 R1249 S1250 F1254 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

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	139.50Å 256.43Å 100.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.90 – 2.80 29.55 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.2 (19.90-2.80) 98.2 (29.55-2.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 2.80Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.215 , 0.283 0.208 , 0.273	Depositor DCC
R_{free} test set	2231 reflections (5.06%)	DCC
Wilson B-factor (Å^2)	60.4	Xtriage
Anisotropy	0.454	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 42.0	EDS
Estimated twinning fraction	0.022 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.023 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 44108 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9311	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.93% 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: CA

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	0.39	0/9443	0.56	0/12771

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	A	9274	0	9256	530	0
2	A	1	0	0	0	0
3	A	36	0	0	3	0
All	All	9311	0	9256	530	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 29.

The worst 5 of 530 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:491:ILE:HG22	1:A:555:ALA:HB3	1.36	1.07

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1100:ASP:OD1	1:A:1102:THR:HG22	1.59	1.02
1:A:133:ARG:HG2	1:A:134:ARG:H	1.25	1.01
1:A:854:THR:HG21	1:A:872:ALA:HB3	1.42	1.01
1:A:806:ARG:H	1:A:807:PRO:HD2	1.31	0.96

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	1192/1289 (92%)	1045 (88%)	120 (10%)	27 (2%)	8 26

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	ALA
1	A	778	LEU
1	A	806	ARG
1	A	970	SER
1	A	19	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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	993/1060 (94%)	831 (84%)	162 (16%)	3 8

5 of 162 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	614	GLN
1	A	762	ILE
1	A	1116	LEU
1	A	622	VAL
1	A	705	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	741	ASN
1	A	862	HIS
1	A	1155	ASN
1	A	857	ASN
1	A	871	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

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

There are no such residues in this entry.

5.8 Polymer linkage issues

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	A	1202/1289 (93%)	-0.15	31 (2%) 59 47	35, 75, 124, 148	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	807	PRO	4.6
1	A	1234	ALA	4.3
1	A	701	SER	4.0
1	A	161	ALA	3.8
1	A	810	SER	3.7

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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	CA	A	2001	1/1	0.97	0.22	-	82,82,82,82	0

6.5 Other polymers (i)

There are no such residues in this entry.