W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 27, 2014 – 03:39 PM GMT

PDB ID : 2G33
Title : Human Hepatitis B Virus T=4 capsid, strain adyw
Authors : Bourne, C.R.; Zlotnick, A.
Deposited on : 2006-02-17
Resolution : 3.96 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
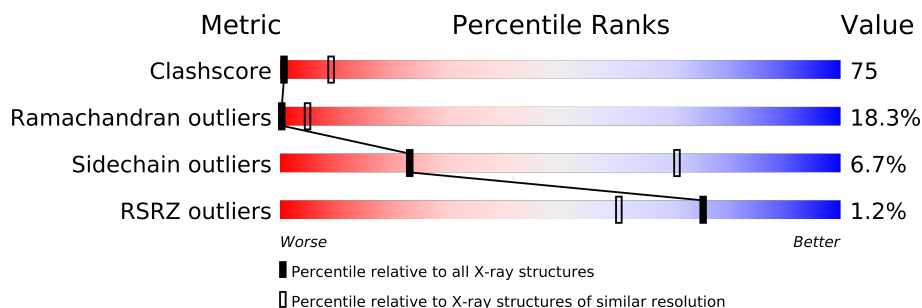A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

---

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

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance ⓘ

The reported resolution of this entry is 3.96 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	Percentile Ranks	Value
Clashscore		75
Ramachandran outliers		18.3%
Sidechain outliers		6.7%
RSRZ outliers		1.2%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	79885	1195 (4.42-3.50)
Ramachandran outliers	78287	1137 (4.42-3.50)
Sidechain outliers	78261	1124 (4.42-3.50)
RSRZ outliers	66119	1005 (4.46-3.46)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	150	
1	B	150	
1	C	150	
1	D	150	

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 4658 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Core antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	147	Total	C	N	O	S	0	0	0
			1167	759	191	215	2			
1	D	146	Total	C	N	O	S	0	0	0
			1160	755	190	213	2			
1	B	147	Total	C	N	O	S	0	0	0
			1167	759	191	215	2			
1	A	148	Total	C	N	O	S	0	0	0
			1164	758	192	212	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	48	ALA	CYS	ENGINEERED	UNP P03147
C	61	ALA	CYS	ENGINEERED	UNP P03147
C	107	ALA	CYS	ENGINEERED	UNP P03147
C	150	CYS	-	INSERTION	UNP P03147
D	48	ALA	CYS	ENGINEERED	UNP P03147
D	61	ALA	CYS	ENGINEERED	UNP P03147
D	107	ALA	CYS	ENGINEERED	UNP P03147
D	150	CYS	-	INSERTION	UNP P03147
B	48	ALA	CYS	ENGINEERED	UNP P03147
B	61	ALA	CYS	ENGINEERED	UNP P03147
B	107	ALA	CYS	ENGINEERED	UNP P03147
B	150	CYS	-	INSERTION	UNP P03147
A	48	ALA	CYS	ENGINEERED	UNP P03147
A	61	ALA	CYS	ENGINEERED	UNP P03147
A	107	ALA	CYS	ENGINEERED	UNP P03147
A	150	CYS	-	INSERTION	UNP P03147

3 Residue-property plots ⓘ

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.
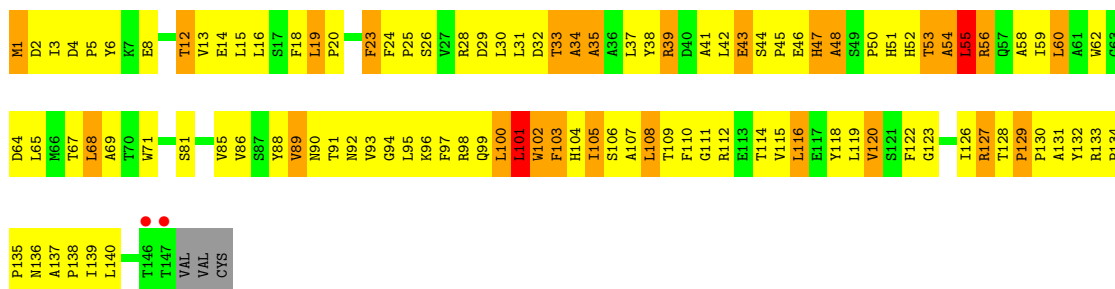
• Molecule 1: Core antigen

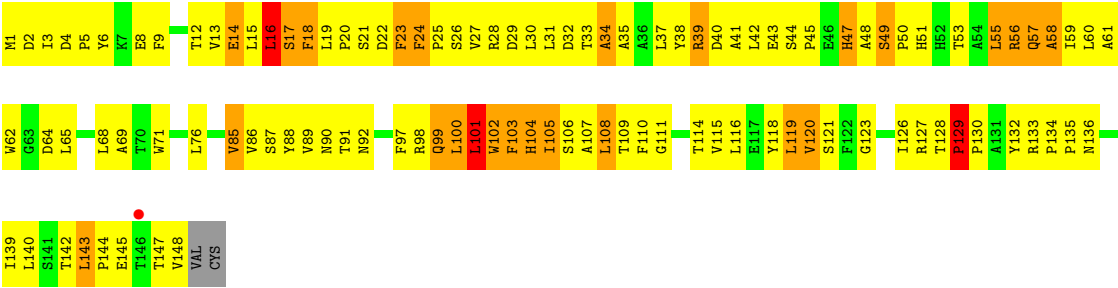Chain C:
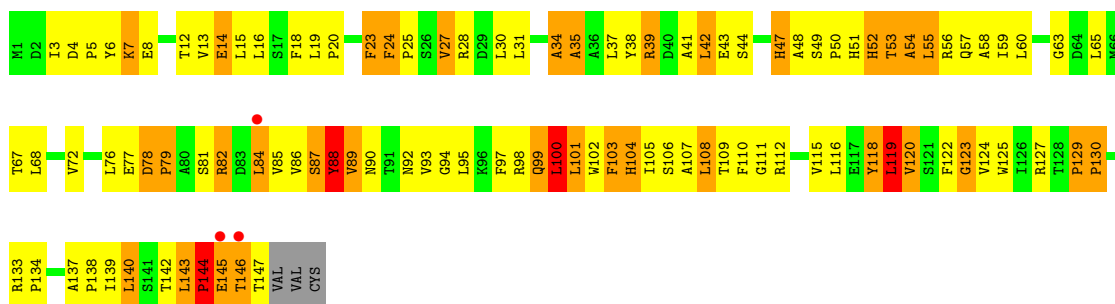
• Molecule 1: Core antigen

Chain D:

• Molecule 1: Core antigen

Chain B:

• Molecule 1: Core antigen

Chain A:

4   Data and refinement statistics   (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	558.40Å 327.14Å 562.24Å 90.00° 109.12° 90.00°	Depositor
Resolution (Å)	40.00 – 3.96 39.95 – 4.00	Depositor EDS
% Data completeness (in resolution range)	79.2 (40.00-3.96) 81.5 (39.95-4.00)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.64 (at 4.00Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.360 , 0.372 0.331 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	133.8	Xtriage
Anisotropy	0.325	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.22 , 23.6	EDS
Estimated twinning fraction	0.000 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h-1/2*k-l 0.000 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h+1/2*k-l	Xtriage
L-test for twinning	$< |L| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Outliers	0 of 653450 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	4658	wwPDB-VP
Average B, all atoms (Å^2)	106.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.32% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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	0.62	0/1200	0.88	1/1648 (0.1%)
1	B	0.64	0/1203	0.86	1/1652 (0.1%)
1	C	0.62	0/1203	0.98	6/1652 (0.4%)
1	D	0.66	0/1196	0.92	1/1642 (0.1%)
All	All	0.63	0/4802	0.91	9/6594 (0.1%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	145	GLU	N-CA-C	9.47	136.58	111.00
1	C	144	PRO	N-CA-C	7.60	131.85	112.10
1	D	143	LEU	N-CA-C	7.07	130.09	111.00
1	C	82	ARG	N-CA-C	-6.53	93.37	111.00
1	C	140	LEU	N-CA-C	5.73	126.46	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1164	0	1136	187	0
1	B	1167	0	1148	166	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1167	0	1148	199	0
1	D	1160	0	1141	195	0
All	All	4658	0	4573	696	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 75.

The worst 5 of 696 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:139:ILE:HG22	1:A:140:LEU:H	0.93	1.07
1:C:78:ASP:HB3	1:C:79:PRO:CD	1.80	1.06
1:D:139:ILE:HG22	1:D:140:LEU:H	1.18	1.04
1:C:78:ASP:HB3	1:C:79:PRO:HD3	1.10	1.04
1:D:55:LEU:HD12	1:D:55:LEU:H	1.25	1.01

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	146/150 (97%)	78 (53%)	41 (28%)	27 (18%)	0	4
1	B	145/150 (97%)	86 (59%)	36 (25%)	23 (16%)	0	7
1	C	145/150 (97%)	76 (52%)	37 (26%)	32 (22%)	0	2
1	D	144/150 (96%)	80 (56%)	40 (28%)	24 (17%)	0	7
All	All	580/600 (97%)	320 (55%)	154 (27%)	106 (18%)	0	5

5 of 106 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	34	ALA
1	C	39	ARG
1	C	48	ALA

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	54	ALA
1	C	78	ASP

5.3.2 Protein sidechains ⓘ

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	124/130 (95%)	118 (95%)	6 (5%)	35	81
1	B	127/130 (98%)	117 (92%)	10 (8%)	18	64
1	C	127/130 (98%)	119 (94%)	8 (6%)	25	73
1	D	126/130 (97%)	116 (92%)	10 (8%)	18	64
All	All	504/520 (97%)	470 (93%)	34 (7%)	23	71

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	129	PRO
1	B	12	THR
1	A	49	SER
1	D	144	PRO
1	C	119	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	90	ASN
1	C	99	GLN
1	B	90	ASN
1	A	57	GLN

5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

There are no ligands in this entry.

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, 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	148/150 (98%)	-0.24	1 (0%)	84	72	44, 95, 160, 196	0
1	B	147/150 (98%)	-0.18	2 (1%)	72	57	43, 94, 159, 200	0
1	C	147/150 (98%)	-0.01	3 (2%)	62	47	35, 105, 176, 200	0
1	D	146/150 (97%)	-0.04	1 (0%)	84	72	40, 106, 171, 194	0
All	All	588/600 (98%)	-0.12	7 (1%)	75	60	35, 100, 172, 200	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	145	GLU	3.0
1	B	146	THR	2.5
1	C	146	THR	2.4
1	D	80	ALA	2.4
1	B	147	THR	2.3

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.