



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 07:38 pm GMT

PDB ID : 3V6Z  
Title : Crystal Structure of Hepatitis B Virus e-antigen  
Authors : Dimattia, M.A.; Watts, N.R.; Stahl, S.J.; Grimes, J.M.; Steven, A.C.; Stuart, D.I.; Wingfield, P.T.  
Deposited on : 2011-12-20  
Resolution : 3.34 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

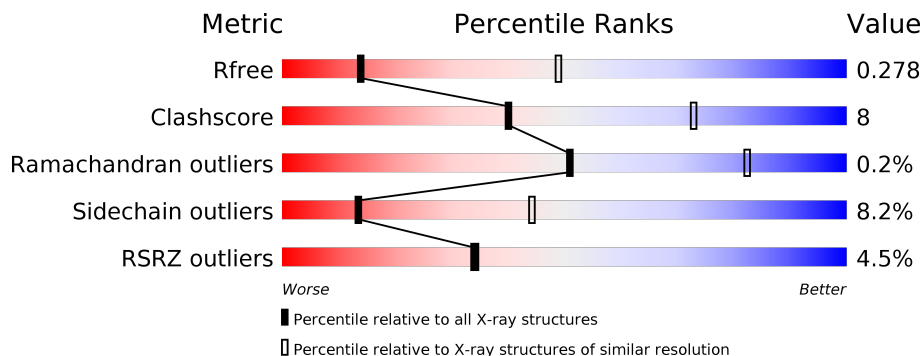
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.34 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))
$R_{free}$	100719	1167 (3.40-3.28)
Clashscore	112137	1239 (3.40-3.28)
Ramachandran outliers	110173	1219 (3.40-3.28)
Sidechain outliers	110143	1218 (3.40-3.28)
RSRZ outliers	101464	1176 (3.40-3.28)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	224	<div> <div>2%</div> <div> <div></div> <div>80%</div> <div>17%</div> <div>..</div> </div> </div>
1	C	224	<div> <div>3%</div> <div> <div></div> <div>80%</div> <div>17%</div> <div>..</div> </div> </div>
2	B	219	<div> <div>%</div> <div> <div></div> <div>79%</div> <div>20%</div> <div>.</div> </div> </div>
2	D	219	<div> <div>3%</div> <div> <div></div> <div>79%</div> <div>20%</div> <div>.</div> </div> </div>
3	E	159	<div> <div>12%</div> <div> <div></div> <div>69%</div> <div>19%</div> <div>5%</div> <div>7%</div> </div> </div>
3	F	159	<div> <div>9%</div> <div> <div></div> <div>69%</div> <div>19%</div> <div>5%</div> <div>7%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9082 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 Fab e6 Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	221	Total	C	N	O	S	0	0	0
			1652	1039	265	339	9			
1	C	221	Total	C	N	O	S	0	0	0
			1652	1039	265	339	9			

- Molecule 2 is a protein called Fab e6 Light Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	219	Total	C	N	O	S	0	0	0
			1707	1066	284	348	9			
2	D	219	Total	C	N	O	S	0	0	0
			1707	1066	284	348	9			

- Molecule 3 is a protein called e-antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	148	Total	C	N	O	S	0	0	0
			1182	773	194	211	4			
3	F	148	Total	C	N	O	S	0	0	0
			1182	773	194	211	4			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	58	ALA	CYS	engineered mutation	UNP Q9QMH8
E	74	ASP	GLU	natural variant	UNP Q9QMH8
E	103	VAL	MET	natural variant	UNP Q9QMH8
E	117	ALA	CYS	engineered mutation	UNP Q9QMH8
E	133	ALA	GLY	engineered mutation	UNP Q9QMH8
F	58	ALA	CYS	engineered mutation	UNP Q9QMH8
F	74	ASP	GLU	natural variant	UNP Q9QMH8
F	103	VAL	MET	natural variant	UNP Q9QMH8

*Continued on next page...*

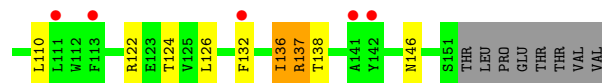
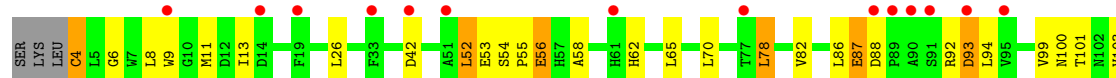
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	117	ALA	CYS	engineered mutation	UNP Q9QMH8
F	133	ALA	GLY	engineered mutation	UNP Q9QMH8

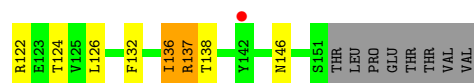




• Molecule 3: e-antigen



• Molecule 3: e-antigen



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.66Å 75.76Å 88.70Å 96.77° 103.81° 116.04°	Depositor
Resolution (Å)	46.20 – 3.34 46.22 – 3.34	Depositor EDS
% Data completeness (in resolution range)	92.6 (46.20-3.34) 84.0 (46.22-3.34)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.37 (at 3.32Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, $R_{free}$	0.234 , 0.233 0.266 , 0.278	Depositor DCC
$R_{free}$ test set	1003 reflections (5.11%)	DCC
Wilson B-factor (Å <sup>2</sup> )	90.7	Xtriage
Anisotropy	0.038	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 90.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	9082	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.15% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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](#)

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	0.43	0/1694	0.79	2/2312 (0.1%)
1	C	0.43	0/1694	0.79	2/2312 (0.1%)
2	B	0.39	0/1746	0.74	0/2369
2	D	0.39	0/1746	0.74	0/2369
3	E	0.43	0/1221	0.63	1/1676 (0.1%)
3	F	0.43	0/1221	0.63	1/1676 (0.1%)
All	All	0.41	0/9322	0.73	6/12714 (0.0%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	11	LEU	CD1-CG-CD2	-6.67	90.49	110.50
1	A	11	LEU	CD1-CG-CD2	-6.34	91.48	110.50
1	C	171	SER	N-CA-C	5.24	125.14	111.00
3	E	93	ASP	CB-CG-OD2	5.23	123.01	118.30
1	A	171	SER	N-CA-C	5.20	125.04	111.00

There are no chirality outliers.

There are no planarity outliers.

### 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	1652	0	1594	32	0
1	C	1652	0	1594	34	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1707	0	1637	36	0
2	D	1707	0	1637	34	0
3	E	1182	0	1153	17	0
3	F	1182	0	1153	18	0
All	All	9082	0	8768	148	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 8.

The worst 5 of 148 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:E:87:GLU:O	3:E:92:ARG:HG3	1.43	1.19
3:F:87:GLU:O	3:F:92:ARG:HG3	1.43	1.18
1:A:11:LEU:HD23	1:A:157:PRO:HG3	1.13	1.12
1:C:11:LEU:HD23	1:C:157:PRO:HG3	1.11	1.09
1:C:11:LEU:HD23	1:C:157:PRO:CG	1.88	1.03

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	217/224 (97%)	215 (99%)	1 (0%)	1 (0%)	32	68
1	C	217/224 (97%)	215 (99%)	1 (0%)	1 (0%)	32	68
2	B	217/219 (99%)	208 (96%)	9 (4%)	0	100	100
2	D	217/219 (99%)	208 (96%)	9 (4%)	0	100	100
3	E	146/159 (92%)	130 (89%)	16 (11%)	0	100	100
3	F	146/159 (92%)	130 (89%)	16 (11%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1160/1204 (96%)	1106 (95%)	52 (4%)	2 (0%)	51 82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	104	GLY
1	C	104	GLY

### 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	191/192 (100%)	184 (96%)	7 (4%)	39 71
1	C	191/192 (100%)	184 (96%)	7 (4%)	39 71
2	B	196/196 (100%)	181 (92%)	15 (8%)	15 47
2	D	196/196 (100%)	181 (92%)	15 (8%)	15 47
3	E	127/138 (92%)	107 (84%)	20 (16%)	3 14
3	F	127/138 (92%)	107 (84%)	20 (16%)	3 14
All	All	1028/1052 (98%)	944 (92%)	84 (8%)	13 44

5 of 84 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	112	GLU
3	E	11	MET
3	F	94	LEU
2	D	143	LEU
2	D	201	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	39	GLN
1	C	209	HIS
2	D	163	GLN
2	B	163	GLN
2	D	131	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 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](#)

There are no ligands in this entry.

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	221/224 (98%)	0.26	5 (2%) 61 61	59, 89, 134, 157	0
1	C	221/224 (98%)	0.32	7 (3%) 48 48	56, 93, 145, 166	0
2	B	219/219 (100%)	0.42	2 (0%) 84 85	67, 98, 136, 146	0
2	D	219/219 (100%)	0.44	6 (2%) 55 55	69, 101, 141, 157	0
3	E	148/159 (93%)	0.72	19 (12%) 4 4	64, 149, 201, 226	0
3	F	148/159 (93%)	0.71	14 (9%) 9 9	63, 144, 193, 218	0
All	All	1176/1204 (97%)	0.45	53 (4%) 34 34	56, 103, 169, 226	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	89	PRO	10.1
3	E	89	PRO	8.7
3	E	88	ASP	6.7
1	A	170	SER	6.4
3	F	88	ASP	6.1

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.