



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 07:12 pm BST

PDB ID : 2QIJ
Title : Hepatitis B Capsid Protein with an N-terminal extension modelled into 8.9 Å data.
Authors : Tan, W.S.; McNae, I.W.; Ho, K.L.; Walkinshaw, M.D.
Deposited on : 2007-07-04
Resolution : 8.90 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

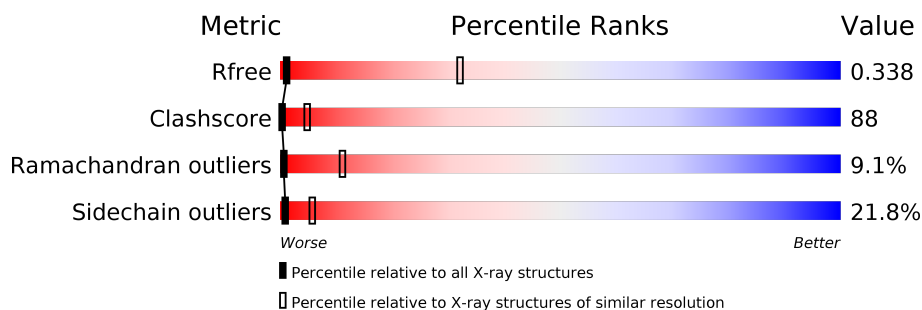
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 8.90 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}	130704	1005 (11.50-3.90)
Clashscore	141614	1070 (11.50-3.90)
Ramachandran outliers	138981	1003 (11.50-3.90)
Sidechain outliers	138945	1003 (11.50-3.86)

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 respectively. 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	157	
1	B	157	
1	C	157	
1	D	157	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4625 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 Core antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	142	Total	C	N	O	S	0	0	0
			1137	741	188	204	4			
1	D	151	Total	C	N	O	S	0	0	0
			1206	784	197	220	5			
1	B	143	Total	C	N	O	S	0	0	0
			1145	747	189	205	4			
1	A	142	Total	C	N	O	S	0	0	0
			1137	741	188	204	4			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-8	MET	-	EXPRESSION TAG	UNP P03147
C	-7	THR	-	EXPRESSION TAG	UNP P03147
C	-6	MET	-	EXPRESSION TAG	UNP P03147
C	-5	ILE	-	EXPRESSION TAG	UNP P03147
C	-4	THR	-	EXPRESSION TAG	UNP P03147
C	-3	ASP	-	EXPRESSION TAG	UNP P03147
C	-2	SER	-	EXPRESSION TAG	UNP P03147
C	-1	LEU	-	EXPRESSION TAG	UNP P03147
C	0	GLU	-	EXPRESSION TAG	UNP P03147
C	1	PHE	-	EXPRESSION TAG	UNP P03147
C	2	HIS	-	EXPRESSION TAG	UNP P03147
D	-8	MET	-	EXPRESSION TAG	UNP P03147
D	-7	THR	-	EXPRESSION TAG	UNP P03147
D	-6	MET	-	EXPRESSION TAG	UNP P03147
D	-5	ILE	-	EXPRESSION TAG	UNP P03147
D	-4	THR	-	EXPRESSION TAG	UNP P03147
D	-3	ASP	-	EXPRESSION TAG	UNP P03147
D	-2	SER	-	EXPRESSION TAG	UNP P03147
D	-1	LEU	-	EXPRESSION TAG	UNP P03147
D	0	GLU	-	EXPRESSION TAG	UNP P03147
D	1	PHE	-	EXPRESSION TAG	UNP P03147

Continued on next page...

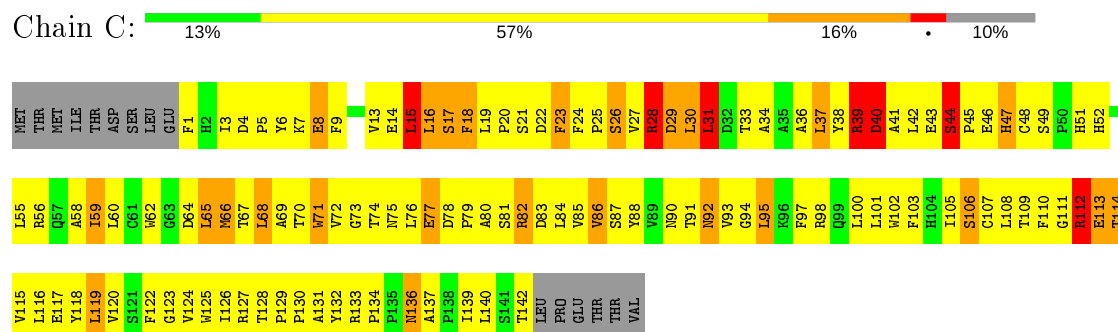
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	2	HIS	-	EXPRESSION TAG	UNP P03147
B	-8	MET	-	EXPRESSION TAG	UNP P03147
B	-7	THR	-	EXPRESSION TAG	UNP P03147
B	-6	MET	-	EXPRESSION TAG	UNP P03147
B	-5	ILE	-	EXPRESSION TAG	UNP P03147
B	-4	THR	-	EXPRESSION TAG	UNP P03147
B	-3	ASP	-	EXPRESSION TAG	UNP P03147
B	-2	SER	-	EXPRESSION TAG	UNP P03147
B	-1	LEU	-	EXPRESSION TAG	UNP P03147
B	0	GLU	-	EXPRESSION TAG	UNP P03147
B	1	PHE	-	EXPRESSION TAG	UNP P03147
B	2	HIS	-	EXPRESSION TAG	UNP P03147
A	-8	MET	-	EXPRESSION TAG	UNP P03147
A	-7	THR	-	EXPRESSION TAG	UNP P03147
A	-6	MET	-	EXPRESSION TAG	UNP P03147
A	-5	ILE	-	EXPRESSION TAG	UNP P03147
A	-4	THR	-	EXPRESSION TAG	UNP P03147
A	-3	ASP	-	EXPRESSION TAG	UNP P03147
A	-2	SER	-	EXPRESSION TAG	UNP P03147
A	-1	LEU	-	EXPRESSION TAG	UNP P03147
A	0	GLU	-	EXPRESSION TAG	UNP P03147
A	1	PHE	-	EXPRESSION TAG	UNP P03147
A	2	HIS	-	EXPRESSION TAG	UNP P03147

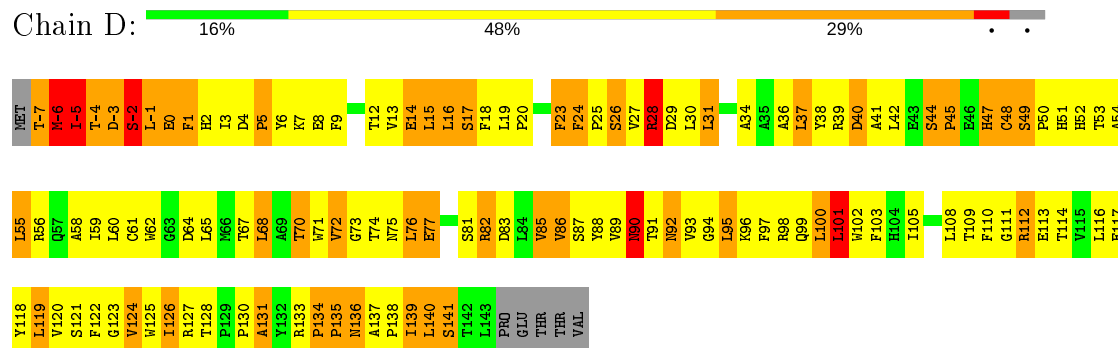
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

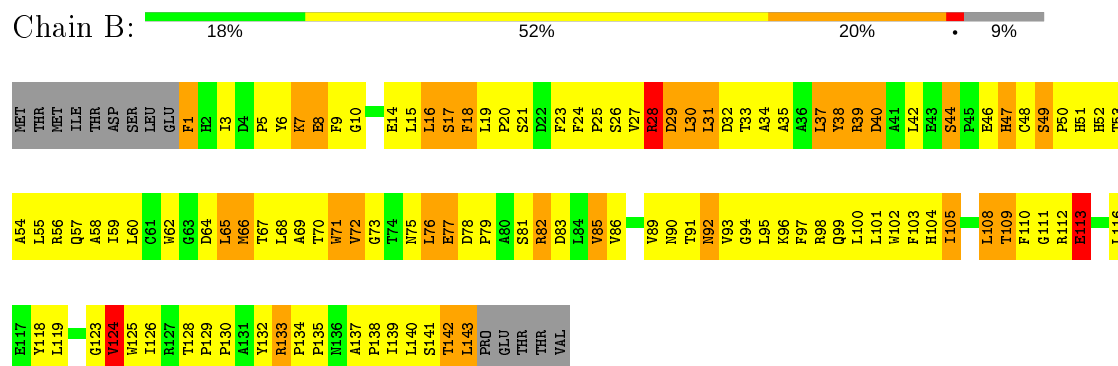
- Molecule 1: Core antigen



- Molecule 1: Core antigen

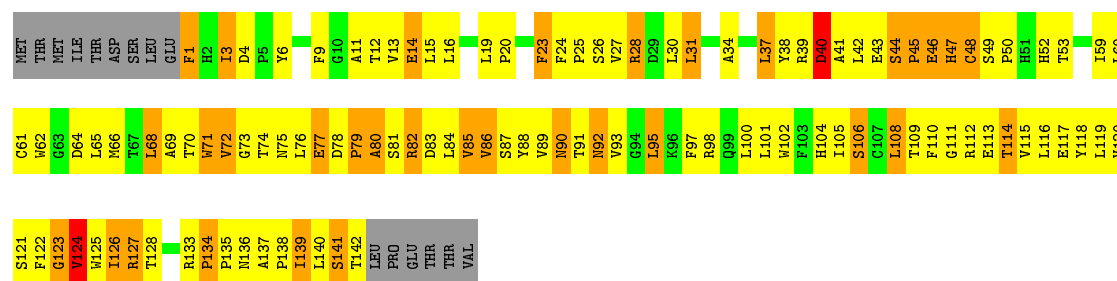


- Molecule 1: Core antigen



- Molecule 1: Core antigen

Frequency	Percentage
Daily	20%
Often	48%
Sometimes	21%
Never	10%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	352.32Å 465.52Å 645.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	60.41 – 8.90 60.06 – 8.90	Depositor EDS
% Data completeness (in resolution range)	(Not available) (60.41-8.90) 99.9 (60.06-8.90)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.57 (at 8.37Å)	Xtriage
Refinement program		Depositor
R, R_{free}	(Not available) , (Not available) 0.331 , 0.338	Depositor DCC
R_{free} test set	4087 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	273.2	Xtriage
Anisotropy	0.664	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 302.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.35$, $\langle L^2 \rangle = 0.18$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.62	EDS
Total number of atoms	4625	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.66% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

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	0.74	0/1174	0.91	5/1610 (0.3%)
1	B	0.78	0/1182	0.98	5/1621 (0.3%)
1	C	0.79	2/1174 (0.2%)	0.95	6/1610 (0.4%)
1	D	0.71	1/1243 (0.1%)	0.90	6/1704 (0.4%)
All	All	0.76	3/4773 (0.1%)	0.93	22/6545 (0.3%)

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	D	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	31	LEU	CG-CD1	-7.17	1.25	1.51
1	C	31	LEU	CG-CD2	-7.15	1.25	1.51
1	D	48	CYS	CB-SG	-5.07	1.73	1.81

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	31	LEU	CB-CG-CD2	7.05	122.99	111.00
1	C	15	LEU	CA-CB-CG	-6.19	101.06	115.30
1	D	119	LEU	CA-CB-CG	-5.96	101.58	115.30
1	D	31	LEU	CA-CB-CG	-5.92	101.69	115.30
1	B	47	HIS	N-CA-C	-5.87	95.14	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	-5	ILE	Peptide
1	D	-6	MET	Peptide
1	D	-7	THR	Peptide
1	D	1	PHE	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	1137	0	1112	202	0
1	B	1145	0	1123	198	0
1	C	1137	0	1112	210	0
1	D	1206	0	1183	255	0
All	All	4625	0	4530	802	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 88.

The worst 5 of 802 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:65:LEU:CD1	1:A:93:VAL:HG11	1.56	1.34
1:D:-7:THR:CG2	1:D:-5:ILE:HG22	1.61	1.31
1:B:93:VAL:CG1	1:B:97:PHE:HE1	1.54	1.18
1:D:14:GLU:O	1:D:17:SER:HB3	1.44	1.15
1:A:62:TRP:HE3	1:A:97:PHE:CD2	1.64	1.15

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	140/157 (89%)	98 (70%)	31 (22%)	11 (8%)	1	13
1	B	141/157 (90%)	84 (60%)	46 (33%)	11 (8%)	1	13
1	C	140/157 (89%)	101 (72%)	30 (21%)	9 (6%)	1	16
1	D	149/157 (95%)	90 (60%)	38 (26%)	21 (14%)	0	4
All	All	570/628 (91%)	373 (65%)	145 (25%)	52 (9%)	1	11

5 of 52 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	-5	ILE
1	D	0	GLU
1	D	16	LEU
1	A	124	VAL
1	C	18	PHE

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	125/140 (89%)	102 (82%)	23 (18%)	1	9
1	B	126/140 (90%)	100 (79%)	26 (21%)	1	7
1	C	125/140 (89%)	97 (78%)	28 (22%)	1	5
1	D	134/140 (96%)	100 (75%)	34 (25%)	0	3
All	All	510/560 (91%)	399 (78%)	111 (22%)	1	6

5 of 111 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	82	ARG
1	B	1	PHE
1	A	92	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	85	VAL
1	D	121	SER

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	B	47	HIS
1	B	52	HIS
1	A	52	HIS
1	D	92	ASN
1	A	47	HIS

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.