



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 31, 2016 – 10:51 PM GMT

PDB ID : 1VHI  
Title : EPSTEIN BARR VIRUS NUCLEAR ANTIGEN-1 DNA-BINDING DOMAIN, RESIDUES 470-607  
Authors : Bochkarev, A.; Barwell, J.; Pfuetzner, R.; Furey, W.; Edwards, A.; Frappier, L.  
Deposited on : 1996-10-05  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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  
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

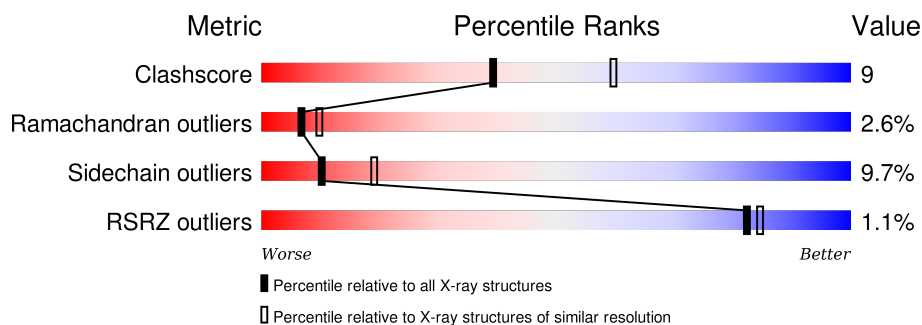
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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	142	<div> <div>2%</div> <div>65%</div> <div>26%</div> <div>6%</div> <div>••</div> </div>
1	B	142	<div> <div>62%</div> <div>25%</div> <div>5%</div> <div>•</div> <div>7%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2727 atoms, of which 585 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 EPSTEIN BARR VIRUS NUCLEAR ANTIGEN-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	139	Total	C	H	N	O	S	0	0	0
			1284	673	227	184	192	8			
1	B	132	Total	C	H	N	O	S	0	0	0
			1230	650	216	175	182	7			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	467	SER	LYS	CONFLICT	UNP P03211
A	469	MET	ARG	CONFLICT	UNP P03211
B	467	SER	LYS	CONFLICT	UNP P03211
B	469	MET	ARG	CONFLICT	UNP P03211

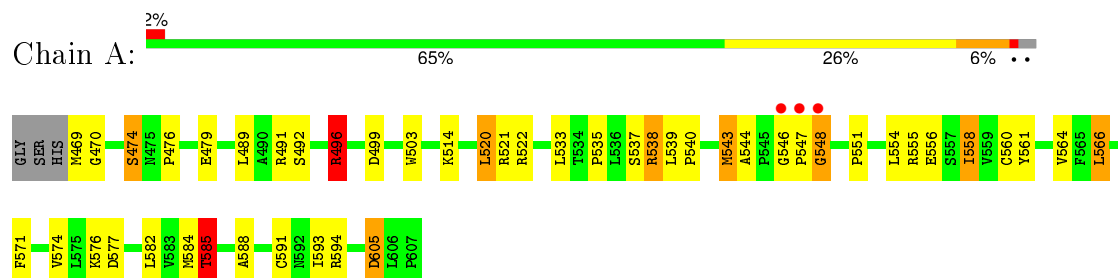
- Molecule 2 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	38	Total	H	O	0	0
			114	76	38		
2	B	33	Total	H	O	0	0
			99	66	33		

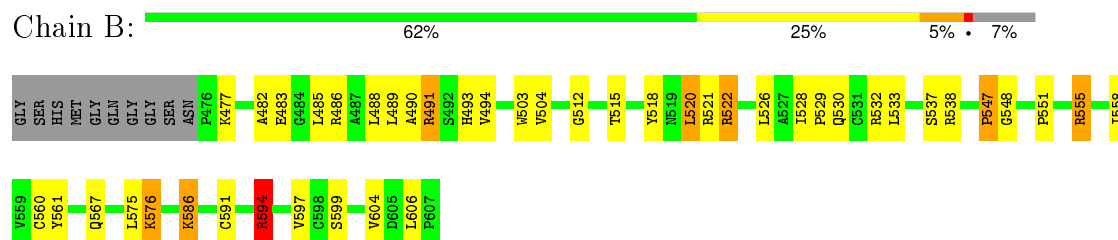
### 3 Residue-property plots

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: EPSTEIN BARR VIRUS NUCLEAR ANTIGEN-1



#### • Molecule 1: EPSTEIN BARR VIRUS NUCLEAR ANTIGEN-1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.99Å 69.08Å 70.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 2.50 34.54 – 2.50	Depositor EDS
% Data completeness (in resolution range)	95.5 (8.00-2.50) 95.6 (34.54-2.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.50 (at 2.51Å)	Xtriage
Refinement program	X-PLOR 3.0	Depositor
R, $R_{free}$	0.179 , (Not available) 0.168 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	24.7	Xtriage
Anisotropy	0.112	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 66.7	EDS
Estimated twinning fraction	0.015 for -h,l,k	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	2 of 10096 reflections (0.020%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2727	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.03% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.91	0/1082	1.66	19/1469 (1.3%)
1	B	0.90	0/1039	1.70	15/1412 (1.1%)
All	All	0.91	0/2121	1.68	34/2881 (1.2%)

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	B	0	1

There are no bond length outliers.

All (34) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	594	ARG	NE-CZ-NH1	12.62	126.61	120.30
1	B	522	ARG	NE-CZ-NH2	-8.99	115.81	120.30
1	A	594	ARG	NE-CZ-NH2	-8.78	115.91	120.30
1	A	543	MET	CG-SD-CE	-8.73	86.24	100.20
1	B	594	ARG	NE-CZ-NH2	-8.29	116.16	120.30
1	B	532	ARG	NE-CZ-NH1	8.13	124.37	120.30
1	A	594	ARG	NE-CZ-NH1	8.10	124.35	120.30
1	A	469	MET	CG-SD-CE	8.02	113.03	100.20
1	A	522	ARG	NE-CZ-NH1	7.50	124.05	120.30
1	B	555	ARG	NE-CZ-NH2	-7.37	116.61	120.30
1	B	503	TRP	CD1-CG-CD2	7.00	111.90	106.30
1	A	503	TRP	CD1-CG-CD2	6.84	111.78	106.30
1	A	521	ARG	NE-CZ-NH2	-6.82	116.89	120.30
1	A	585	THR	N-CA-CB	-6.82	97.35	110.30
1	B	503	TRP	CE2-CD2-CG	-6.59	102.03	107.30
1	B	515	THR	N-CA-CB	-6.57	97.82	110.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	566	LEU	CA-CB-CG	6.53	130.33	115.30
1	A	503	TRP	CE2-CD2-CG	-6.48	102.12	107.30
1	A	543	MET	CA-CB-CG	-6.28	102.62	113.30
1	A	605	ASP	CB-CG-OD2	-6.10	112.81	118.30
1	A	496	ARG	NE-CZ-NH1	6.07	123.34	120.30
1	B	560	CYS	CA-CB-SG	-6.05	103.10	114.00
1	A	496	ARG	NE-CZ-NH2	-6.05	117.28	120.30
1	A	605	ASP	CB-CG-OD1	6.03	123.72	118.30
1	B	493	HIS	N-CA-C	-5.96	94.92	111.00
1	A	538	ARG	NE-CZ-NH1	5.68	123.14	120.30
1	A	522	ARG	NE-CZ-NH2	-5.64	117.48	120.30
1	B	520	LEU	CA-CB-CG	5.62	128.24	115.30
1	A	584	MET	CA-CB-CG	5.32	122.34	113.30
1	B	551	PRO	N-CA-C	5.14	125.47	112.10
1	B	491	ARG	CA-CB-CG	5.13	124.69	113.40
1	B	512	GLY	CA-C-N	5.08	128.38	117.20
1	A	577	ASP	CB-CG-OD1	5.04	122.83	118.30
1	B	522	ARG	NE-CZ-NH1	5.02	122.81	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	594	ARG	Sidechain

## 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	1057	227	1067	18	1
1	B	1014	216	1031	20	0
2	A	38	76	0	1	0
2	B	33	66	0	1	0
All	All	2142	585	2098	37	1

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 9.

All (37) 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:B:594:ARG:HB3	1:B:594:ARG:HH11	1.51	0.76
1:A:476:PRO:HG2	1:A:479:GLU:HB2	1.73	0.69
1:A:520:LEU:HD12	1:A:593:ILE:HD13	1.74	0.69
1:A:538:ARG:HH21	1:A:558:ILE:HG22	1.60	0.66
1:A:551:PRO:HG2	1:A:554:LEU:HD12	1.82	0.61
1:B:538:ARG:HH21	1:B:558:ILE:HD12	1.66	0.60
1:B:488:LEU:O	1:B:491:ARG:HG2	2.04	0.57
1:A:582:LEU:O	1:A:585:THR:HB	2.06	0.56
1:A:571:PHE:HA	1:A:574:VAL:HG22	1.93	0.51
1:B:537:SER:HB3	1:B:561:TYR:CE2	2.47	0.50
1:B:529:PRO:HD2	1:B:530:GLN:OE1	2.14	0.48
1:B:518:TYR:O	1:B:522:ARG:HG2	2.14	0.47
1:B:485:LEU:O	1:B:489:LEU:HB2	2.14	0.47
1:B:528:ILE:HD11	1:B:575:LEU:HD13	1.96	0.46
1:B:586:LYS:O	1:B:591:CYS:HB3	2.14	0.46
1:A:537:SER:HB3	1:A:561:TYR:CE1	2.50	0.46
1:A:540:PRO:HA	1:A:556:GLU:HA	1.98	0.45
1:A:588:ALA:HA	1:A:591:CYS:SG	2.57	0.45
1:B:538:ARG:NH2	1:B:558:ILE:HD12	2.32	0.44
1:B:576:LYS:HB3	1:B:576:LYS:HE2	1.81	0.44
1:B:604:VAL:HG12	1:B:606:LEU:HD12	1.98	0.44
1:B:538:ARG:HH11	1:B:538:ARG:HG3	1.83	0.44
1:B:522:ARG:O	1:B:526:LEU:HD13	2.17	0.43
1:B:567:GLN:HG3	2:B:13:HOH:O	2.18	0.43
1:A:496:ARG:HD3	2:A:46:HOH:O	2.18	0.43
1:A:588:ALA:N	1:A:591:CYS:SG	2.92	0.43
1:A:548:GLY:HA3	1:A:555:ARG:NE	2.34	0.42
1:B:521:ARG:HG3	1:B:533:LEU:HD22	2.01	0.42
1:A:470:GLY:HA3	1:A:474:SER:O	2.20	0.42
1:B:482:ALA:O	1:B:486:ARG:HB2	2.20	0.42
1:A:548:GLY:HA3	1:A:555:ARG:CZ	2.50	0.42
1:B:538:ARG:NH1	1:B:538:ARG:HG3	2.35	0.41
1:A:514:LYS:HD2	1:A:560:CYS:SG	2.61	0.41
1:A:605:ASP:OD2	1:B:594:ARG:NH2	2.54	0.41
1:B:576:LYS:HG3	1:B:597:VAL:HG23	2.03	0.41
1:A:539:LEU:HD13	1:A:561:TYR:CD1	2.56	0.41
1:A:533:LEU:HD23	1:A:564:VAL:HG22	2.02	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.



Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:491:ARG:HH22	1:A:544:ALA:H[3_544]	1.30	0.30

## 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	137/142 (96%)	127 (93%)	6 (4%)	4 (3%)	6	8
1	B	130/142 (92%)	121 (93%)	6 (5%)	3 (2%)	8	12
All	All	267/284 (94%)	248 (93%)	12 (4%)	7 (3%)	7	10

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	546	GLY
1	B	547	PRO
1	B	548	GLY
1	A	535	PRO
1	A	547	PRO
1	B	490	ALA
1	A	548	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.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	115/117 (98%)	104 (90%)	11 (10%)	10	19

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	111/117 (95%)	100 (90%)	11 (10%)	10	18
All	All	226/234 (97%)	204 (90%)	22 (10%)	10	19

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	474	SER
1	A	489	LEU
1	A	492	SER
1	A	496	ARG
1	A	499	ASP
1	A	520	LEU
1	A	543	MET
1	A	558	ILE
1	A	566	LEU
1	A	576	LYS
1	A	585	THR
1	B	477	LYS
1	B	483	GLU
1	B	494	VAL
1	B	504	VAL
1	B	520	LEU
1	B	547	PRO
1	B	555	ARG
1	B	576	LYS
1	B	586	LYS
1	B	594	ARG
1	B	599	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	569	HIS

### 5.3.3 RNA ⓘ

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	139/142 (97%)	-0.62	3 (2%) 65 69	3, 11, 29, 39	0
1	B	132/142 (92%)	-0.49	0 100 100	4, 12, 32, 36	0
All	All	271/284 (95%)	-0.56	3 (1%) 82 84	3, 12, 31, 39	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	547	PRO	3.2
1	A	546	GLY	2.3
1	A	548	GLY	2.2

### 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.