



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 26, 2023 – 03:24 AM EDT

PDB ID : 5WMF  
Title : Crystal structure of the Hexameric Ring of Epstein-Barr Virus Nuclear Antigen-1, EBNA1  
Authors : Messick, T.E.  
Deposited on : 2017-07-28  
Resolution : 1.90 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](mailto: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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35.1  
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) : 2.35.1

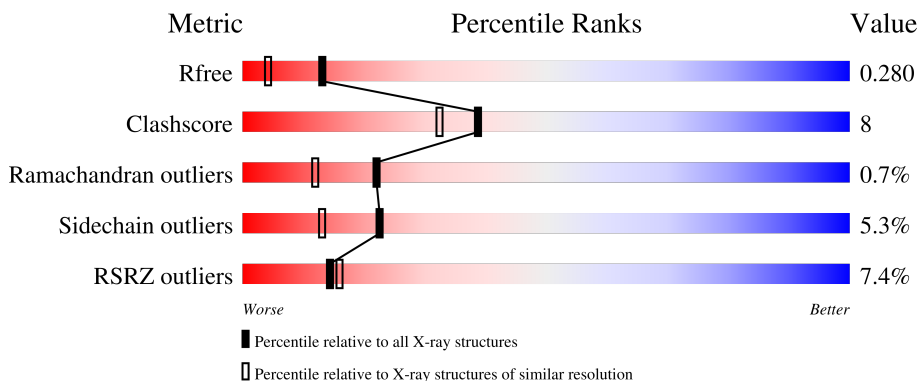
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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.



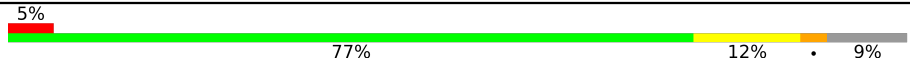
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	154	 9% 74% 12% • 11%
1	B	154	 5% 71% 18% • 11%
1	C	154	 5% 77% 11% • 10%
1	D	154	 8% 77% 11% • 11%
1	E	154	 7% 69% 16% • 12%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	154	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a small red segment at the beginning labeled '5%', a large green segment labeled '77%', a yellow segment labeled '12%', and a small grey segment at the end labeled '9%'.</p>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13143 atoms, of which 6431 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 nuclear antigen 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	137	2129	684	1069	182	187	7	0	0	0
1	B	137	2128	684	1069	182	186	7	0	0	0
1	C	138	2146	690	1080	182	186	8	0	0	0
1	D	137	2126	684	1067	182	186	7	0	0	0
1	E	135	2100	675	1056	179	183	7	0	0	0
1	F	140	2170	697	1090	185	190	8	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	466	GLY	-	expression tag	UNP P03211
A	467	SER	-	expression tag	UNP P03211
A	468	HIS	-	expression tag	UNP P03211
A	469	MET	-	expression tag	UNP P03211
B	466	GLY	-	expression tag	UNP P03211
B	467	SER	-	expression tag	UNP P03211
B	468	HIS	-	expression tag	UNP P03211
B	469	MET	-	expression tag	UNP P03211
C	466	GLY	-	expression tag	UNP P03211
C	467	SER	-	expression tag	UNP P03211
C	468	HIS	-	expression tag	UNP P03211
C	469	MET	-	expression tag	UNP P03211
D	466	GLY	-	expression tag	UNP P03211
D	467	SER	-	expression tag	UNP P03211
D	468	HIS	-	expression tag	UNP P03211
D	469	MET	-	expression tag	UNP P03211
E	466	GLY	-	expression tag	UNP P03211

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	467	SER	-	expression tag	UNP P03211
E	468	HIS	-	expression tag	UNP P03211
E	469	MET	-	expression tag	UNP P03211
F	466	GLY	-	expression tag	UNP P03211
F	467	SER	-	expression tag	UNP P03211
F	468	HIS	-	expression tag	UNP P03211
F	469	MET	-	expression tag	UNP P03211

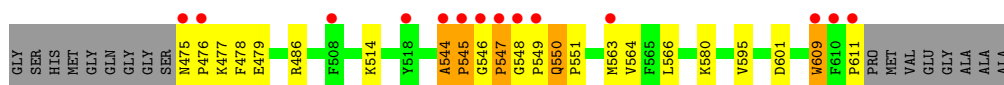
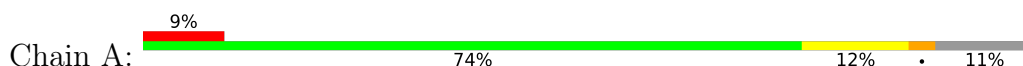
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	57	Total O 57 57	0	0
2	B	58	Total O 58 58	0	0
2	C	58	Total O 58 58	0	0
2	D	57	Total O 57 57	0	0
2	E	57	Total O 57 57	0	0
2	F	57	Total O 57 57	0	0

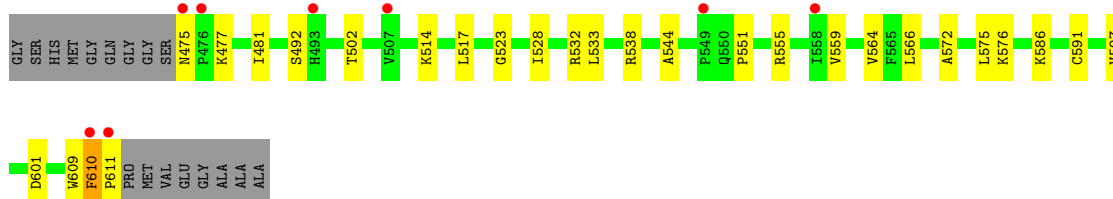
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 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.

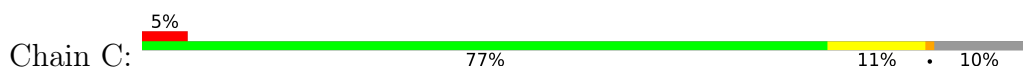
- Molecule 1: Epstein-Barr nuclear antigen 1



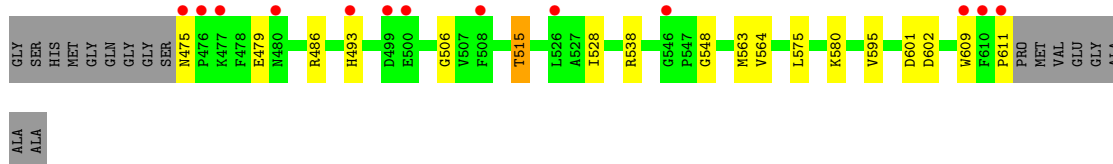
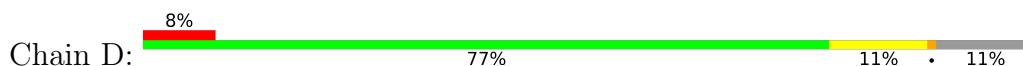
- Molecule 1: Epstein-Barr nuclear antigen 1



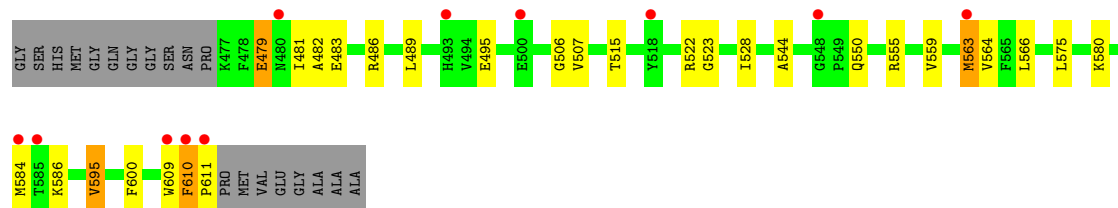
- Molecule 1: Epstein-Barr nuclear antigen 1



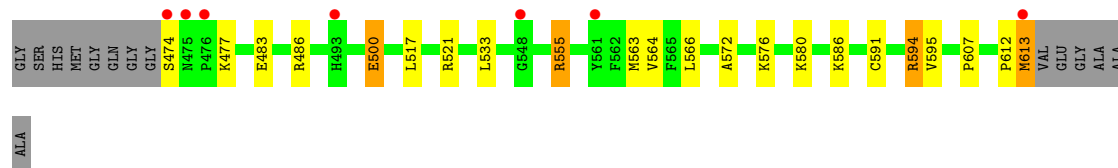
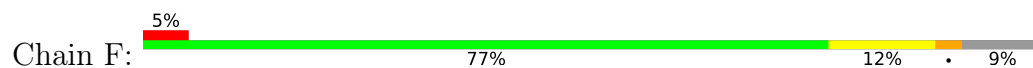
- Molecule 1: Epstein-Barr nuclear antigen 1



- Molecule 1: Epstein-Barr nuclear antigen 1



- Molecule 1: Epstein-Barr nuclear antigen 1



ALA

## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	31.28Å 84.77Å 84.71Å 60.01° 87.31° 88.53°	Depositor
Resolution (Å)	29.19 – 1.90 29.19 – 1.90	Depositor EDS
% Data completeness (in resolution range)	76.5 (29.19-1.90) 76.5 (29.19-1.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.74 (at 1.89Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.198 , 0.276 0.202 , 0.280	Depositor DCC
$R_{free}$ test set	1993 reflections (4.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.1	Xtriage
Anisotropy	0.375	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.45 , 63.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.047 for h,l,-k+1 0.047 for h,k-l,k 0.135 for h,-k+1,-k 0.135 for h,-l,k-l 0.066 for h,-k,-l 0.025 for -h,-k+1,l 0.016 for -h,-l,-k 0.014 for -h,-k,-k+1 0.023 for -h,k,k-l 0.023 for -h,k-l,l 0.016 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	13143	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.39% 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	0.68	0/1090	0.76	0/1487
1	B	0.73	0/1089	0.77	0/1485
1	C	0.74	0/1097	0.76	0/1495
1	D	0.65	0/1089	0.73	0/1485
1	E	0.68	0/1073	0.80	0/1462
1	F	0.75	0/1111	0.79	0/1515
All	All	0.70	0/6549	0.77	0/8929

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 [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	1060	1069	1069	24	0
1	B	1059	1069	1069	20	0
1	C	1066	1080	1080	11	0
1	D	1059	1067	1069	11	0
1	E	1044	1056	1056	22	0
1	F	1080	1090	1090	17	0
2	A	57	0	0	4	1
2	B	58	0	0	2	0
2	C	58	0	0	2	1
2	D	57	0	0	6	1

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	57	0	0	6	1
2	F	57	0	0	5	2
All	All	6712	6431	6433	96	4

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 96 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:D:475:ASN:N	2:D:701:HOH:O	2.02	0.92
1:D:486:ARG:NH2	2:D:702:HOH:O	2.04	0.89
1:B:538:ARG:NH2	2:B:701:HOH:O	2.14	0.79
1:E:479:GLU:O	2:E:701:HOH:O	2.01	0.77
1:C:500:GLU:N	1:C:500:GLU:OE1	2.18	0.76

All (4) 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)
2:F:754:HOH:O	2:F:757:HOH:O[1_655]	1.94	0.26
2:F:753:HOH:O	2:F:757:HOH:O[1_655]	2.02	0.18
2:A:746:HOH:O	2:E:739:HOH:O[1_556]	2.03	0.17
2:C:732:HOH:O	2:D:755:HOH:O[1_455]	2.17	0.03

## 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	135/154 (88%)	131 (97%)	1 (1%)	3 (2%)	<b>6</b> <b>1</b>
1	B	135/154 (88%)	130 (96%)	4 (3%)	1 (1%)	<b>22</b> <b>12</b>

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	136/154 (88%)	135 (99%)	1 (1%)	0	100	100
1	D	135/154 (88%)	134 (99%)	0	1 (1%)	22	12
1	E	133/154 (86%)	128 (96%)	4 (3%)	1 (1%)	19	9
1	F	138/154 (90%)	136 (99%)	2 (1%)	0	100	100
All	All	812/924 (88%)	794 (98%)	12 (2%)	6 (1%)	22	12

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	544	ALA
1	A	547	PRO
1	B	610	PHE
1	D	548	GLY
1	E	610	PHE

### 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	116/125 (93%)	111 (96%)	5 (4%)	29	19
1	B	116/125 (93%)	109 (94%)	7 (6%)	19	9
1	C	117/125 (94%)	112 (96%)	5 (4%)	29	19
1	D	116/125 (93%)	113 (97%)	3 (3%)	46	39
1	E	114/125 (91%)	105 (92%)	9 (8%)	12	5
1	F	119/125 (95%)	111 (93%)	8 (7%)	16	7
All	All	698/750 (93%)	661 (95%)	37 (5%)	22	13

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	595	VAL
1	F	594	ARG

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	F	474	SER
1	F	517	LEU
1	C	477	LYS

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

Mol	Chain	Res	Type
1	D	519	ASN

### 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 monosaccharides 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	137/154 (88%)	0.69	14 (10%) 6 8	18, 27, 57, 76	0
1	B	137/154 (88%)	0.55	8 (5%) 23 25	14, 25, 47, 61	0
1	C	138/154 (89%)	0.48	8 (5%) 23 25	15, 26, 45, 56	0
1	D	137/154 (88%)	0.68	13 (9%) 8 9	17, 28, 51, 74	0
1	E	135/154 (87%)	0.72	11 (8%) 12 13	16, 29, 53, 70	0
1	F	140/154 (90%)	0.45	7 (5%) 28 32	15, 25, 44, 71	0
All	All	824/924 (89%)	0.59	61 (7%) 14 16	14, 27, 51, 76	0

The worst 5 of 61 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	611	PRO	8.4
1	E	611	PRO	7.9
1	A	611	PRO	6.5
1	F	474	SER	6.1
1	A	547	PRO	5.3

### 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 monosaccharides 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.