wwPDB X-ray Structure Validation Summary Report (i)

Nov 22, 2021 – 10:06 am GMT

PDB ID	:	7NX5
Title	:	Crystal structure of the Epstein-Barr Virus protein ZEBRA (BZLF1, Zta) bound to a methylated DNA duplex
Authors	:	Bernaudat, F.; Petosa, C.
Deposited on	:	2021-03-17
Resolution	:	2.50 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 (i) symbol.

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

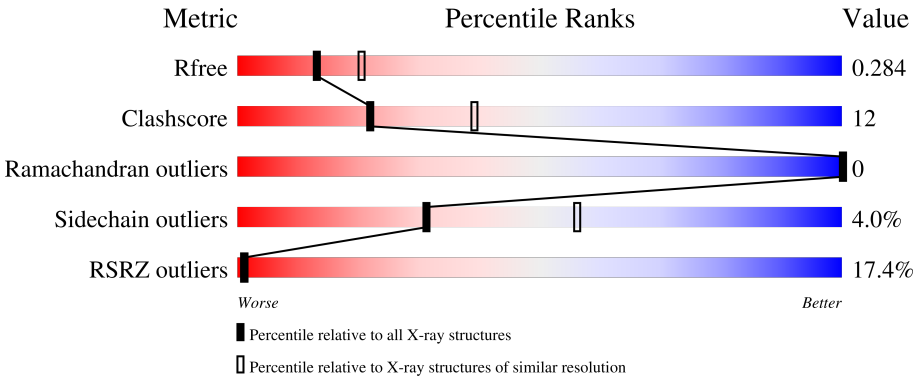
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (i)

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.



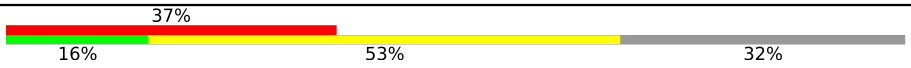
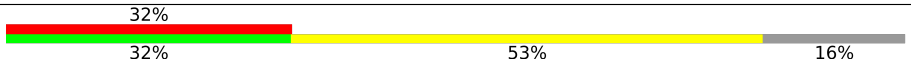
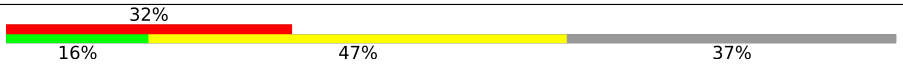
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 of 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 <=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	63	2%, 65%, 30%
1	B	63	13%, 73%, 21%
1	E	63	13%, 71%, 25%
1	F	63	16%, 71%, 27%
2	C	19	32%, 58%, 21%, 5%, 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	19	
3	D	19	
3	H	19	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3158 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 Trans-activator protein BZLF1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	62	Total	C	N	O	S	0	0	0
			498	311	100	85	2			
1	B	61	Total	C	N	O	S	0	0	0
			493	305	99	87	2			
1	E	61	Total	C	N	O	S	0	0	0
			467	294	92	79	2			
1	F	62	Total	C	N	O	S	0	0	0
			485	302	96	84	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	174	MET	-	initiating methionine	UNP P03206
A	189	SER	CYS	engineered mutation	UNP P03206
B	174	MET	-	initiating methionine	UNP P03206
B	189	SER	CYS	engineered mutation	UNP P03206
E	174	MET	-	initiating methionine	UNP P03206
E	189	SER	CYS	engineered mutation	UNP P03206
F	174	MET	-	initiating methionine	UNP P03206
F	189	SER	CYS	engineered mutation	UNP P03206

- Molecule 2 is a DNA chain called meZRE2 DNA (bottom strand).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	16	Total	C	N	O	P	0	0	0
			318	154	53	96	15			
2	G	13	Total	C	N	O	P	0	0	0
			251	122	40	77	12			

- Molecule 3 is a DNA chain called meZRE2 DNA (top strand).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	16	Total 332	C 158	N 68	O 91	P 15	0	0	0
3	H	12	Total 249	C 118	N 48	O 71	P 12	0	0	0

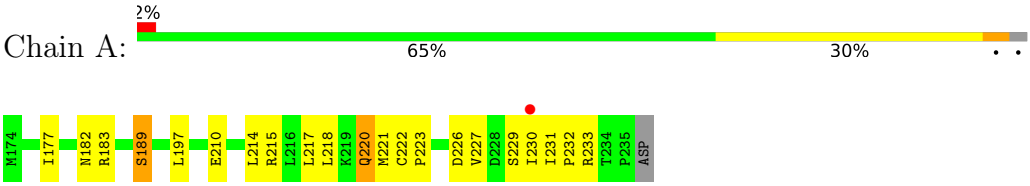
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	10	Total 10	O 10	0	0
4	B	15	Total 15	O 15	0	0
4	C	4	Total 4	O 4	0	0
4	D	9	Total 9	O 9	0	0
4	E	10	Total 10	O 10	0	0
4	F	4	Total 4	O 4	0	0
4	G	7	Total 7	O 7	0	0
4	H	6	Total 6	O 6	0	0

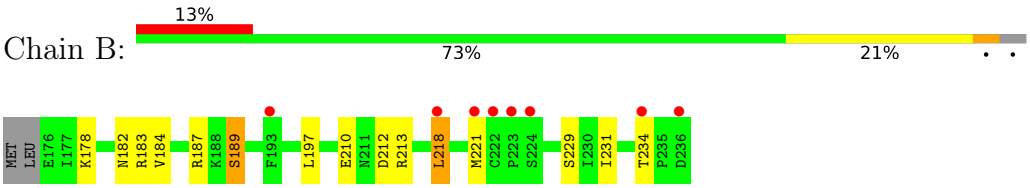
3 Residue-property plots (i)

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.

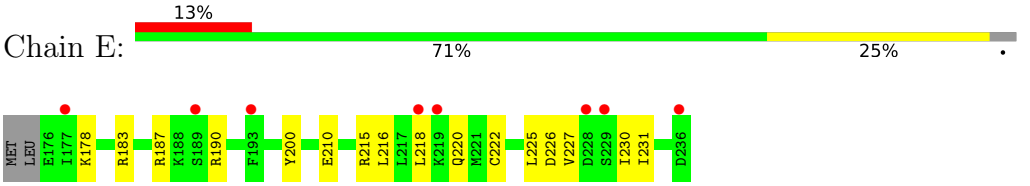
• Molecule 1: Trans-activator protein BZLF1



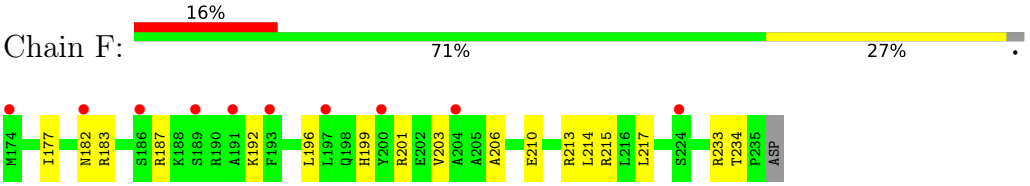
• Molecule 1: Trans-activator protein BZLF1



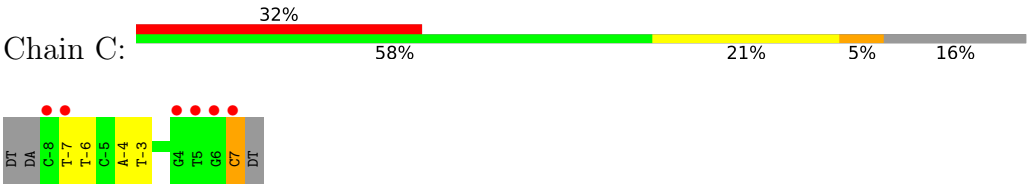
• Molecule 1: Trans-activator protein BZLF1



• Molecule 1: Trans-activator protein BZLF1

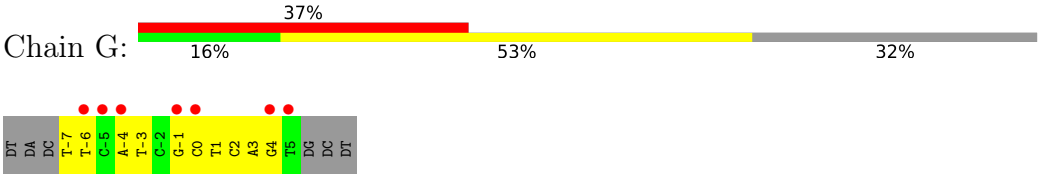


• Molecule 2: meZRE2 DNA (bottom strand)



• Molecule 2: meZRE2 DNA (bottom strand)

Chain G:

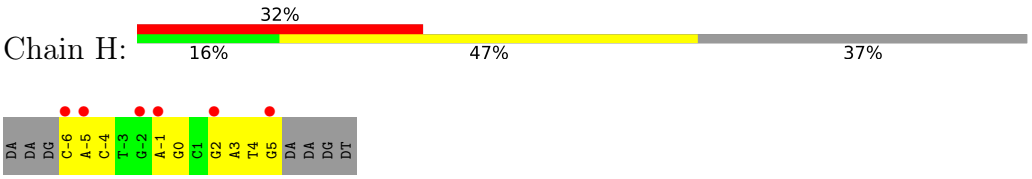


• Molecule 3: meZRE2 DNA (top strand)

Chain D:

• Molecule 3: meZRE2 DNA (top strand)

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	207.99Å 26.56Å 80.85Å 90.00° 103.10° 90.00°	Depositor
Resolution (Å)	16.00 – 2.50 34.17 – 2.50	Depositor EDS
% Data completeness (in resolution range)	95.5 (16.00-2.50) 98.4 (34.17-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.55 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.277 , 0.339 0.288 , 0.284	Depositor DCC
R_{free} test set	804 reflections (5.23%)	wwPDB-VP
Wilson B-factor (Å^2)	42.2	Xtriage
Anisotropy	0.250	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	3158	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 68.99 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.9885e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

Bond lengths and bond angles in the following residue types are not validated in this section: 5CM

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	0/504	0.60	0/673
1	B	0.53	0/499	0.74	2/664 (0.3%)
1	E	0.48	0/473	0.67	0/633
1	F	0.41	0/490	0.57	0/656
2	C	1.16	1/331 (0.3%)	1.12	1/506 (0.2%)
2	G	0.90	0/256	1.17	0/390
3	D	1.02	0/351	1.07	2/539 (0.4%)
3	H	0.94	0/256	1.00	0/391
All	All	0.73	1/3160 (0.0%)	0.86	5/4452 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	7	DC	C4-N4	-8.14	1.26	1.33

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	4	DT	O4'-C1'-N1	6.88	112.82	108.00
1	B	218	LEU	CB-CG-CD2	-5.48	101.68	111.00
3	D	4	DT	C3'-C2'-C1'	-5.18	96.28	102.50
1	B	218	LEU	CA-CB-CG	5.06	126.94	115.30
2	C	7	DC	N3-C4-C5	-5.00	119.90	121.90

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	498	0	524	18	0
1	B	493	0	508	13	0
1	E	467	0	474	14	1
1	F	485	0	503	16	0
2	C	318	0	179	5	0
2	G	251	0	146	8	0
3	D	332	0	182	7	0
3	H	249	0	137	6	1
4	A	10	0	0	3	0
4	B	15	0	0	1	0
4	C	4	0	0	0	0
4	D	9	0	0	0	0
4	E	10	0	0	1	0
4	F	4	0	0	1	0
4	G	7	0	0	4	1
4	H	6	0	0	2	0
All	All	3158	0	2653	68	2

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 12.

The worst 5 of 68 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)
2:G:-7:DT:H2"	2:G:-6:DT:C6	2.25	0.72
2:C:-7:DT:H2'	2:C:-6:DT:C6	2.24	0.72
1:A:217:LEU:HD23	1:B:218:LEU:HD21	1.72	0.72
1:A:226:ASP:OD2	4:A:301:HOH:O	2.09	0.70
3:D:-6:DC:H2"	3:D:-5:DA:C8	2.28	0.69

All (2) 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)
4:G:106:HOH:O	4:G:106:HOH:O[2_555]	1.93	0.27
1:E:183:ARG:NH2	3:H:-1:DA:OP1[1_565]	2.08	0.12

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	60/63 (95%)	57 (95%)	3 (5%)	0	100	100
1	B	59/63 (94%)	57 (97%)	2 (3%)	0	100	100
1	E	59/63 (94%)	57 (97%)	2 (3%)	0	100	100
1	F	60/63 (95%)	59 (98%)	1 (2%)	0	100	100
All	All	238/252 (94%)	230 (97%)	8 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	53/58 (91%)	50 (94%)	3 (6%)	20	39
1	B	52/58 (90%)	50 (96%)	2 (4%)	33	58
1	E	46/58 (79%)	45 (98%)	1 (2%)	52	77
1	F	51/58 (88%)	49 (96%)	2 (4%)	32	57
All	All	202/232 (87%)	194 (96%)	8 (4%)	31	56

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	183	ARG
1	F	177	ILE
1	B	229	SER
1	B	189	SER
1	E	178	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

4 non-standard protein/DNA/RNA residues are modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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, 95^{th} 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	62/63 (98%)	0.72	1 (1%)	72	74	32, 47, 101, 121	0
1	B	61/63 (96%)	0.97	8 (13%)	3	3	30, 46, 80, 89	0
1	E	61/63 (96%)	1.26	8 (13%)	3	3	53, 66, 102, 124	0
1	F	62/63 (98%)	1.11	10 (16%)	1	1	50, 63, 93, 100	0
2	C	15/19 (78%)	1.90	6 (40%)	0	0	35, 62, 102, 113	0
2	G	12/19 (63%)	2.37	7 (58%)	0	0	70, 89, 147, 150	0
3	D	15/19 (78%)	1.96	6 (40%)	0	0	42, 63, 97, 109	0
3	H	11/19 (57%)	1.92	6 (54%)	0	0	82, 87, 109, 128	0
All	All	299/328 (91%)	1.20	52 (17%)	1	1	30, 61, 102, 150	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	7	DC	6.0
3	D	8	DG	5.0
3	D	7	DA	4.9
2	C	-8	DC	4.8
3	D	-7	DG	4.8

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	5CM	G	-2	20/21	0.87	0.25	81,86,95,96	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	5CM	H	1	20/21	0.87	0.27	78,80,84,85	0
3	5CM	D	1	20/21	0.93	0.22	35,39,50,65	0
2	5CM	C	-2	20/21	0.94	0.19	30,36,49,50	0

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.