wwPDB X-ray Structure Validation Summary Report (i)

Feb 16, 2023 – 12:19 pm GMT

PDB ID	:	8AGO
Title	:	BK Polyomavirus VP1 mutant E73Q
Authors	:	Sorin, M.N.; Di Maio, A.; Silva, L.M.; Ebert, D.; Delannoy, C.; Nguyen, N.-K.; Guerardel, Y.; Chai, W.; Halary, F.; Renaudin-Autain, K.; Liu, Y.; Bressollette-Bodin, C.; Stehle, T.; McIlroy, D.
Deposited on	:	2022-07-20
Resolution	:	1.85 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.32.1
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)	:	2.32.1

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 1.85 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.
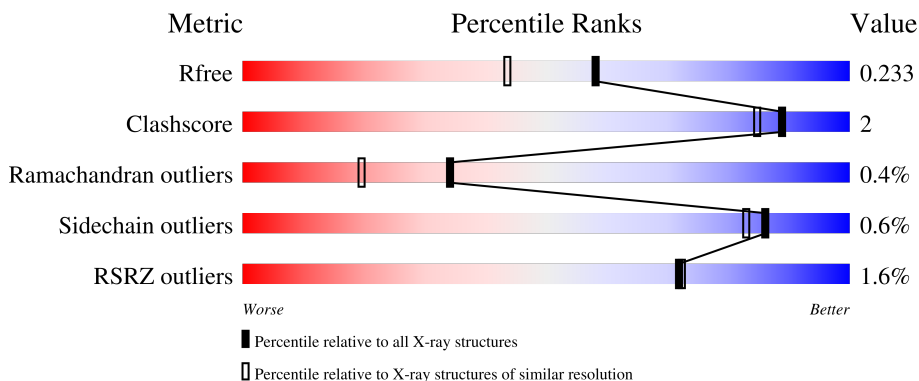
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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 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	AAA	271	91% 5% ·
1	BBB	271	.% 88% 5% 7%
1	CCC	271	4% 86% 6% 8%
1	DDD	271	.% 92% · ·
1	EEE	271	2% 89% 7% ··

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 10693 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 Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	260	Total 1980	C 1244	N 342	O 382	S 12	0	0	0
1	BBB	253	Total 1946	C 1227	N 339	O 367	S 13	0	3	0
1	CCC	249	Total 1891	C 1190	N 327	O 361	S 13	0	1	0
1	DDD	261	Total 1988	C 1250	N 345	O 380	S 13	0	2	0
1	EEE	262	Total 2024	C 1276	N 349	O 386	S 13	0	6	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	72	GLN	GLU	engineered mutation	UNP P03088
AAA	104	SER	CYS	engineered mutation	UNP P03088
BBB	72	GLN	GLU	engineered mutation	UNP P03088
BBB	104	SER	CYS	engineered mutation	UNP P03088
CCC	72	GLN	GLU	engineered mutation	UNP P03088
CCC	104	SER	CYS	engineered mutation	UNP P03088
DDD	72	GLN	GLU	engineered mutation	UNP P03088
DDD	104	SER	CYS	engineered mutation	UNP P03088
EEE	72	GLN	GLU	engineered mutation	UNP P03088
EEE	104	SER	CYS	engineered mutation	UNP P03088

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total 6	C 3	O 3	0	0
2	BBB	1	Total 6	C 3	O 3	0	0
2	EEE	1	Total 6	C 3	O 3	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	1	Total 1	Cl 1	0	0
3	BBB	1	Total 1	Cl 1	0	0
3	CCC	1	Total 1	Cl 1	0	0
3	DDD	1	Total 1	Cl 1	0	0
3	EEE	1	Total 1	Cl 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AAA	186	Total 186	O 186	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	BBB	168	Total 168	O 168	0	0
4	CCC	137	Total 137	O 137	0	0
4	DDD	156	Total 156	O 156	0	0
4	EEE	194	Total 194	O 194	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: Major capsid protein VP1

Chain AAA:
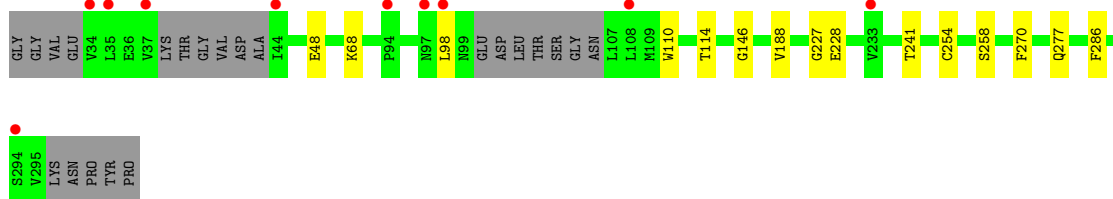
- Molecule 1: Major capsid protein VP1

Chain BBB:

- Molecule 1: Major capsid protein VP1

Chain CCC:

- Molecule 1: Major capsid protein VP1

Chain DDD:

- Molecule 1: Major capsid protein VP1

Chain EEE:

GLY GLY V32 V41 D42 ALA I44 C49 R63 R82 E100 D101 L102 THR SER GLY N106 T117 K142 E160 N181 P182 V188 M189 H193 N201 R214 L259 Y260 V261 Y285 I288 R289 L290 N297 PRO TYR PRO

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	145.25Å 153.04Å 62.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.55 – 1.85 48.54 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.6 (48.55-1.85) 99.6 (48.54-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.27 (at 1.86Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.190 , 0.229 0.197 , 0.233	Depositor DCC
R_{free} test set	5959 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	25.8	Xtriage
Anisotropy	0.322	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 35.1	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.011 for k,h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10693	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.47% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[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: GOL, CL

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	AAA	0.70	0/2025	0.88	0/2755
1	BBB	0.73	0/1999	0.86	0/2717
1	CCC	0.70	0/1938	0.86	0/2638
1	DDD	0.72	0/2039	0.87	1/2776 (0.0%)
1	EEE	0.74	0/2086	0.90	2/2837 (0.1%)
All	All	0.72	0/10087	0.88	3/13723 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	DDD	82	ARG	NE-CZ-NH2	-5.72	117.44	120.30
1	EEE	214[A]	ARG	NE-CZ-NH2	-5.42	117.59	120.30
1	EEE	214[B]	ARG	NE-CZ-NH2	-5.42	117.59	120.30

There are no chirality outliers.

There are no planarity outliers.

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	AAA	1980	0	1893	7	0
1	BBB	1946	0	1877	8	0
1	CCC	1891	0	1794	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	DDD	1988	0	1906	4	0
1	EEE	2024	0	1944	13	0
2	AAA	6	0	8	0	0
2	BBB	6	0	8	0	0
2	EEE	6	0	8	0	0
3	AAA	1	0	0	0	0
3	BBB	1	0	0	1	0
3	CCC	1	0	0	0	0
3	DDD	1	0	0	1	0
3	EEE	1	0	0	1	0
4	AAA	186	0	0	0	0
4	BBB	168	0	0	2	0
4	CCC	137	0	0	0	0
4	DDD	156	0	0	0	0
4	EEE	194	0	0	1	0
All	All	10693	0	9438	37	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 2.

The worst 5 of 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:EEE:44:ILE:HA	1:EEE:290:LEU:O	1.91	0.69
1:AAA:185:GLN:HE22	1:EEE:63:ARG:HE	1.40	0.67
1:EEE:82[B]:ARG:NH2	1:EEE:201:ASN:O	2.30	0.63
1:BBB:82[A]:ARG:NH1	4:BBB:501:HOH:O	2.33	0.62
1:EEE:188:VAL:HG23	1:EEE:189:MET:H	1.76	0.51

There are no symmetry-related clashes.

5.3   Torsion angles ⓘ

5.3.1   Protein backbone ⓘ

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	AAA	256/271 (94%)	246 (96%)	9 (4%)	1 (0%)	34	19
1	BBB	250/271 (92%)	239 (96%)	10 (4%)	1 (0%)	34	19
1	CCC	244/271 (90%)	231 (95%)	12 (5%)	1 (0%)	34	19
1	DDD	259/271 (96%)	249 (96%)	9 (4%)	1 (0%)	34	19
1	EEE	262/271 (97%)	251 (96%)	10 (4%)	1 (0%)	34	19
All	All	1271/1355 (94%)	1216 (96%)	50 (4%)	5 (0%)	34	19

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	188	VAL
1	DDD	188	VAL
1	BBB	188	VAL
1	CCC	188	VAL
1	EEE	188	VAL

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	AAA	213/232 (92%)	211 (99%)	2 (1%)	78	72
1	BBB	210/232 (90%)	209 (100%)	1 (0%)	88	86
1	CCC	202/232 (87%)	201 (100%)	1 (0%)	88	86
1	DDD	214/232 (92%)	212 (99%)	2 (1%)	78	72
1	EEE	218/232 (94%)	218 (100%)	0	100	100
All	All	1057/1160 (91%)	1051 (99%)	6 (1%)	86	83

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	CCC	228	GLU
1	DDD	228	GLU
1	DDD	248	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	AAA	228	GLU
1	AAA	179	PRO

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

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 5 are monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
2	GOL	BBB	401	-	5,5,5	0.16	0	5,5,5	0.48	0
2	GOL	AAA	401	-	5,5,5	0.19	0	5,5,5	0.39	0
2	GOL	EEE	401	-	5,5,5	0.15	0	5,5,5	0.50	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	BBB	401	-	-	4/4/4/4	-
2	GOL	AAA	401	-	-	0/4/4/4	-
2	GOL	EEE	401	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	BBB	401	GOL	O1-C1-C2-O2
2	BBB	401	GOL	O1-C1-C2-C3
2	EEE	401	GOL	C1-C2-C3-O3
2	EEE	401	GOL	O2-C2-C3-O3
2	BBB	401	GOL	C1-C2-C3-O3

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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	AAA	260/271 (95%)	-0.38	0	100	100	19, 27, 48, 71	0
1	BBB	253/271 (93%)	-0.28	3 (1%)	79	79	20, 29, 50, 74	0
1	CCC	249/271 (91%)	-0.13	10 (4%)	38	36	22, 32, 56, 78	0
1	DDD	261/271 (96%)	-0.20	2 (0%)	86	86	20, 29, 51, 77	0
1	EEE	262/271 (96%)	-0.33	5 (1%)	66	66	19, 26, 44, 77	0
All	All	1285/1355 (94%)	-0.26	20 (1%)	72	72	19, 29, 51, 78	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	DDD	34	VAL	5.1
1	EEE	102	LEU	4.2
1	EEE	44	ILE	4.1
1	CCC	108	LEU	4.0
1	EEE	100	GLU	3.9

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3   Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4   Ligands ⓘ

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, 95th 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	GOL	BBB	401	6/6	0.75	0.21	52,64,68,76	0
2	GOL	AAA	401	6/6	0.82	0.17	43,55,59,62	0
2	GOL	EEE	401	6/6	0.88	0.15	41,52,56,57	0
3	CL	DDD	401	1/1	0.95	0.08	43,43,43,43	0
3	CL	CCC	401	1/1	0.96	0.05	44,44,44,44	0
3	CL	AAA	402	1/1	0.99	0.05	33,33,33,33	0
3	CL	BBB	402	1/1	0.99	0.05	34,34,34,34	0
3	CL	EEE	402	1/1	0.99	0.06	37,37,37,37	0

6.5　Other polymers ⓘ

There are no such residues in this entry.