W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 2, 2021 – 03:11 pm GMT

PDB ID	:	7BNY
Title	:	Structure of 2A protein from encephalomyocarditis virus (EMCV)
Authors	:	Hill, C.H.; Napthine, S.; Pekarek, L.; Kibe, A.; Firth, A.E.; Graham, S.C.; Caliskan, N.; Brierley, I.
Deposited on	:	2021-01-22
Resolution	:	2.62 Å(reported)

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

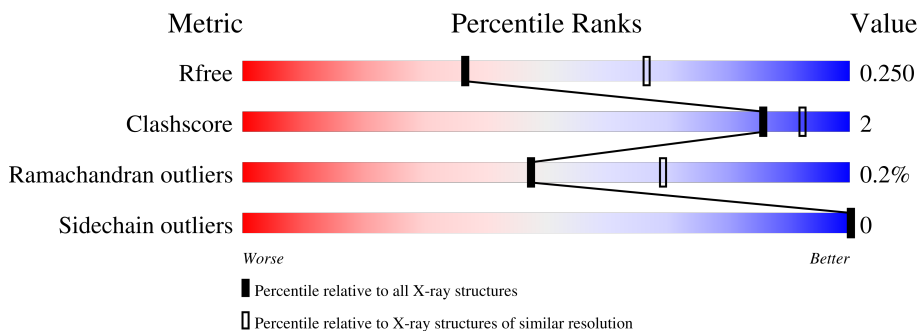|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry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

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.62 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	Percentile Ranks	Value
Rfree		0.250
Clashscore		2
Ramachandran outliers		0.2%
Sidechain outliers		0

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)

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%

Mol	Chain	Length	Quality of chain
1	A	152	86% · 13%
1	B	152	89% · 8%
1	C	152	73% 10% 17%
1	D	152	75% 8% 17%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 4596 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 Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	132	Total	C	N	O	S	0	0	0
			1106	712	198	193	3			
1	B	140	Total	C	N	O	S	0	1	0
			1175	755	209	208	3			
1	C	126	Total	C	N	O	S	0	0	0
			1056	681	191	181	3			
1	D	126	Total	C	N	O	S	0	1	0
			1067	687	195	182	3			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	144	GLY	-	expression tag	UNP P12296
A	145	SER	-	expression tag	UNP P12296
A	146	LYS	-	expression tag	UNP P12296
A	147	HIS	-	expression tag	UNP P12296
A	148	HIS	-	expression tag	UNP P12296
A	149	HIS	-	expression tag	UNP P12296
A	150	HIS	-	expression tag	UNP P12296
A	151	HIS	-	expression tag	UNP P12296
A	152	HIS	-	expression tag	UNP P12296
B	144	GLY	-	expression tag	UNP P12296
B	145	SER	-	expression tag	UNP P12296
B	146	LYS	-	expression tag	UNP P12296
B	147	HIS	-	expression tag	UNP P12296
B	148	HIS	-	expression tag	UNP P12296
B	149	HIS	-	expression tag	UNP P12296
B	150	HIS	-	expression tag	UNP P12296
B	151	HIS	-	expression tag	UNP P12296
B	152	HIS	-	expression tag	UNP P12296
C	144	GLY	-	expression tag	UNP P12296
C	145	SER	-	expression tag	UNP P12296
C	146	LYS	-	expression tag	UNP P12296

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	147	HIS	-	expression tag	UNP P12296
C	148	HIS	-	expression tag	UNP P12296
C	149	HIS	-	expression tag	UNP P12296
C	150	HIS	-	expression tag	UNP P12296
C	151	HIS	-	expression tag	UNP P12296
C	152	HIS	-	expression tag	UNP P12296
D	144	GLY	-	expression tag	UNP P12296
D	145	SER	-	expression tag	UNP P12296
D	146	LYS	-	expression tag	UNP P12296
D	147	HIS	-	expression tag	UNP P12296
D	148	HIS	-	expression tag	UNP P12296
D	149	HIS	-	expression tag	UNP P12296
D	150	HIS	-	expression tag	UNP P12296
D	151	HIS	-	expression tag	UNP P12296
D	152	HIS	-	expression tag	UNP P12296

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
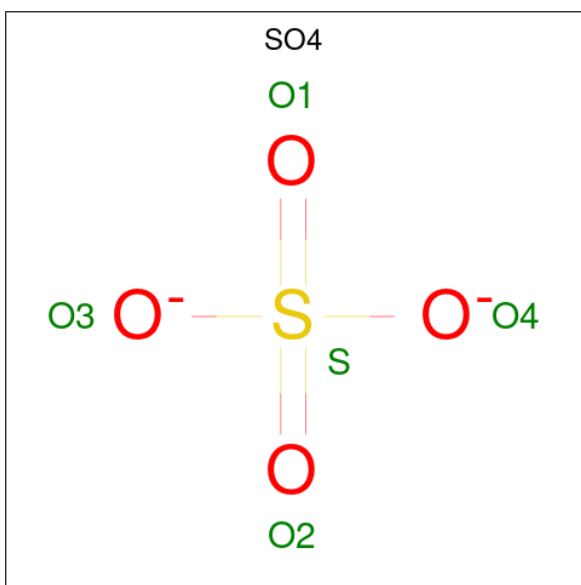
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	C	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0

• Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	56	Total 56	O 56	0	0
3	B	45	Total 45	O 45	0	0
3	C	26	Total 26	O 26	0	0
3	D	20	Total 20	O 20	0	0

3 Residue-property plots ⓘ

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. 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: Genome polyprotein

Chain A:

86%   •   13%

- Molecule 1: Genome polyprotein

Chain B:

89%   •   8%

- Molecule 1: Genome polyprotein

Chain C:

73%   10%   17%

- Molecule 1: Genome polyprotein

Chain D:

75%   8%   17%

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, α, β, γ	91.56Å 91.56Å 316.39Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.91 – 2.62 43.91 – 2.62	Depositor EDS
% Data completeness (in resolution range)	99.7 (43.91-2.62) 99.7 (43.91-2.62)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.18.1_3865	Depositor
R, R_{free}	0.225 , 0.251 0.225 , 0.250	Depositor DCC
R_{free} test set	1207 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	61.9	Xtriage
Anisotropy	0.574	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	4596	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.35% 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: SO4

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	0.25	0/1141	0.50	0/1552
1	B	0.25	0/1212	0.49	0/1649
1	C	0.27	0/1090	0.49	0/1480
1	D	0.25	0/1101	0.51	0/1494
All	All	0.26	0/4544	0.50	0/6175

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 ⓘ

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	1106	0	1076	1	0
1	B	1175	0	1144	1	1
1	C	1056	0	1033	8	0
1	D	1067	0	1045	9	0
2	A	15	0	0	0	0
2	B	15	0	0	0	0
2	C	5	0	0	0	0
2	D	10	0	0	0	0
3	A	56	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	45	0	0	0	0
3	C	26	0	0	0	0
3	D	20	0	0	0	0
All	All	4596	0	4298	19	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 2.

The worst 5 of 19 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:83:GLU:HB3	1:D:84:PRO:HD2	1.84	0.58
1:C:14:LEU:HD11	1:C:29:ILE:HD13	1.87	0.55
1:D:95:ARG:NH2	1:D:97:ARG:HE	2.05	0.55
1:C:82:ALA:HB2	1:C:118:LEU:HD21	1.90	0.53
1:D:83:GLU:HB3	1:D:84:PRO:CD	2.39	0.52

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:B:6:ASP:OD1	1:B:139:GLU:OE2[8_555]	2.17	0.03

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	A	130/152 (86%)	124 (95%)	6 (5%)	0	100	100
1	B	139/152 (91%)	134 (96%)	4 (3%)	1 (1%)	22	41
1	C	124/152 (82%)	118 (95%)	6 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	125/152 (82%)	121 (97%)	4 (3%)	0	100	100
All	All	518/608 (85%)	497 (96%)	20 (4%)	1 (0%)	47	69

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	94	LYS

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	120/138 (87%)	120 (100%)	0	100	100
1	B	129/138 (94%)	129 (100%)	0	100	100
1	C	114/138 (83%)	114 (100%)	0	100	100
1	D	115/138 (83%)	115 (100%)	0	100	100
All	All	478/552 (87%)	478 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

Mol	Chain	Res	Type
1	B	40	HIS

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

5.5   Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6   Ligand geometry ⓘ

9 ligands 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.

5.7   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 ⓘ

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.