



wwPDB X-ray Structure Validation Summary Report

May 16, 2022 – 10:06 am BST

PDB ID : 7QRG
Title : Structure of the post-fusion complex between precursor membrane ectodomain (prM) and envelope ectodomain protein (E) from tick-borne encephalitis virus
Authors : Vaney, M.C.; Rouvinski, A.; Rey, F.A.
Deposited on : 2022-01-11
Resolution : 2.80 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  symbol.

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, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.28.1
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) : 2.28.1

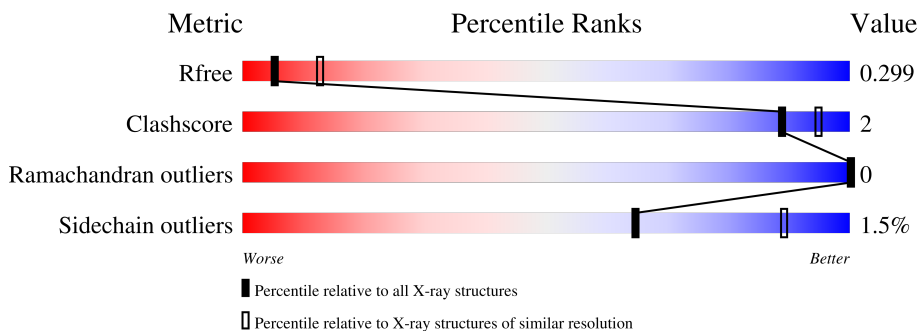
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	443	84% (green), 6% (yellow), 10% (grey)
2	D	138	57% (green), 42% (grey), . (red)

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 3668 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 Envelope protein E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	399	3036	1905	534	576	21	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P14336
A	-2	GLY	-	expression tag	UNP P14336
A	-1	GLY	-	expression tag	UNP P14336
A	0	GLY	-	expression tag	UNP P14336
A	101	ASP	TRP	engineered mutation	UNP P14336
A	401	GLY	-	expression tag	UNP P14336
A	402	PRO	-	expression tag	UNP P14336
A	403	PHE	-	expression tag	UNP P14336
A	404	GLU	-	expression tag	UNP P14336
A	405	ASP	-	expression tag	UNP P14336
A	406	ASP	-	expression tag	UNP P14336
A	407	ASP	-	expression tag	UNP P14336
A	408	ASP	-	expression tag	UNP P14336
A	409	LYS	-	expression tag	UNP P14336
A	410	ALA	-	expression tag	UNP P14336
A	411	GLY	-	expression tag	UNP P14336
A	412	TRP	-	expression tag	UNP P14336
A	413	SER	-	expression tag	UNP P14336
A	414	HIS	-	expression tag	UNP P14336
A	415	PRO	-	expression tag	UNP P14336
A	416	GLN	-	expression tag	UNP P14336
A	417	PHE	-	expression tag	UNP P14336
A	418	GLU	-	expression tag	UNP P14336
A	419	LYS	-	expression tag	UNP P14336
A	420	GLY	-	expression tag	UNP P14336
A	421	GLY	-	expression tag	UNP P14336
A	422	GLY	-	expression tag	UNP P14336

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	423	SER	-	expression tag	UNP P14336
A	424	GLY	-	expression tag	UNP P14336
A	425	GLY	-	expression tag	UNP P14336
A	426	GLY	-	expression tag	UNP P14336
A	427	SER	-	expression tag	UNP P14336
A	428	GLY	-	expression tag	UNP P14336
A	429	GLY	-	expression tag	UNP P14336
A	430	GLY	-	expression tag	UNP P14336
A	431	SER	-	expression tag	UNP P14336
A	432	TRP	-	expression tag	UNP P14336
A	433	SER	-	expression tag	UNP P14336
A	434	HIS	-	expression tag	UNP P14336
A	435	PRO	-	expression tag	UNP P14336
A	436	GLN	-	expression tag	UNP P14336
A	437	PHE	-	expression tag	UNP P14336
A	438	GLU	-	expression tag	UNP P14336
A	439	LYS	-	expression tag	UNP P14336

- Molecule 2 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	80	606	365	105	129	7	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	?	-	ARG	deletion	UNP P14336
D	129	GLY	-	expression tag	UNP P14336
D	130	GLY	-	expression tag	UNP P14336
D	131	GLY	-	expression tag	UNP P14336
D	132	GLY	-	expression tag	UNP P14336
D	133	GLU	-	expression tag	UNP P14336
D	134	ASN	-	expression tag	UNP P14336
D	135	LEU	-	expression tag	UNP P14336
D	136	TYR	-	expression tag	UNP P14336
D	137	PHE	-	expression tag	UNP P14336
D	138	GLN	-	expression tag	UNP P14336

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0

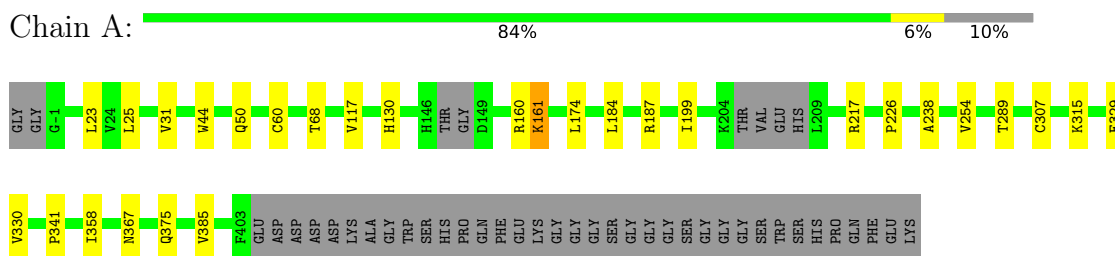
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	13	Total O 13 13	0	0
4	D	1	Total O 1 1	0	0

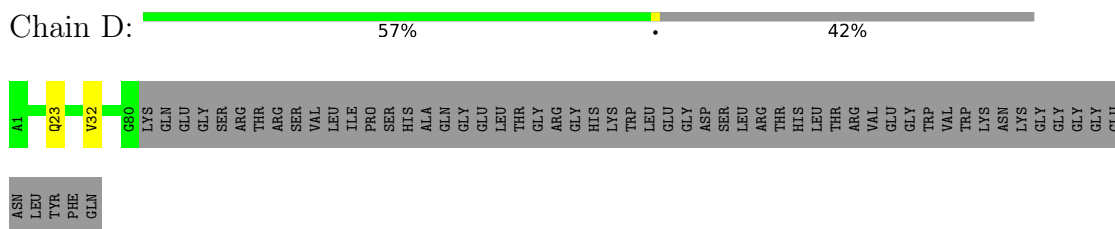
3 Residue-property plots [i](#)

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. 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: Envelope protein E



- Molecule 2: Genome polyprotein



4 Data and refinement statistics

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, α , β , γ	97.94Å 97.94Å 115.16Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.97 – 2.80 48.97 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.2 (48.97-2.80) 98.2 (48.97-2.80)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.25 (at 2.81Å)	Xtrriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.228 , 0.264 0.257 , 0.299	Depositor DCC
R_{free} test set	826 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å ²)	61.5	Xtrriage
Anisotropy	0.477	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.044 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	3668	wwPDB-VP
Average B, all atoms (Å ²)	79.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.18% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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

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	0.34	0/3101	0.56	0/4205
2	D	0.36	0/612	0.57	0/828
All	All	0.35	0/3713	0.56	0/5033

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	3036	0	2988	13	0
2	D	606	0	565	1	0
3	A	12	0	16	0	0
4	A	13	0	0	0	0
4	D	1	0	0	0	0
All	All	3668	0	3569	14	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 14 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:A:60:CYS:HB3	1:A:226:PRO:HG2	1.86	0.57
1:A:315:LYS:HB3	1:A:329:GLU:HB3	1.92	0.51
1:A:226:PRO:HB3	1:A:238:ALA:HA	1.96	0.47
1:A:161:LYS:HD3	1:A:174:LEU:HD22	1.96	0.47
1:A:174:LEU:HD12	1:A:184:LEU:HD12	1.95	0.47

There are no symmetry-related clashes.

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	393/443 (89%)	371 (94%)	22 (6%)	0	100	100
2	D	78/138 (56%)	70 (90%)	8 (10%)	0	100	100
All	All	471/581 (81%)	441 (94%)	30 (6%)	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	327/357 (92%)	321 (98%)	6 (2%)	59	86
2	D	67/114 (59%)	67 (100%)	0	100	100
All	All	394/471 (84%)	388 (98%)	6 (2%)	65	89

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

Mol	Chain	Res	Type
1	A	161	LYS
1	A	217	ARG
1	A	367	ASN
1	A	68	THR
1	A	50	GLN

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

Mol	Chain	Res	Type
1	A	8	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](#)

2 ligands are modelled in this entry.

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	$\# Z > 2$
3	GOL	A	502	-	5,5,5	0.04	0	5,5,5	0.20	0
3	GOL	A	501	-	5,5,5	0.03	0	5,5,5	0.18	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
3	GOL	A	502	-	-	0/4/4/4	-
3	GOL	A	501	-	-	0/4/4/4	-

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 [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

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.