



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2018 – 01:42 pm GMT

PDB ID : 3N41  
Title : Crystal structure of the mature envelope glycoprotein complex (spontaneous cleavage) of Chikungunya virus.  
Authors : Voss, J.; Vaney, M.C.; Duquerroy, S.; Rey, F.A.  
Deposited on : 2010-05-21  
Resolution : 3.01 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

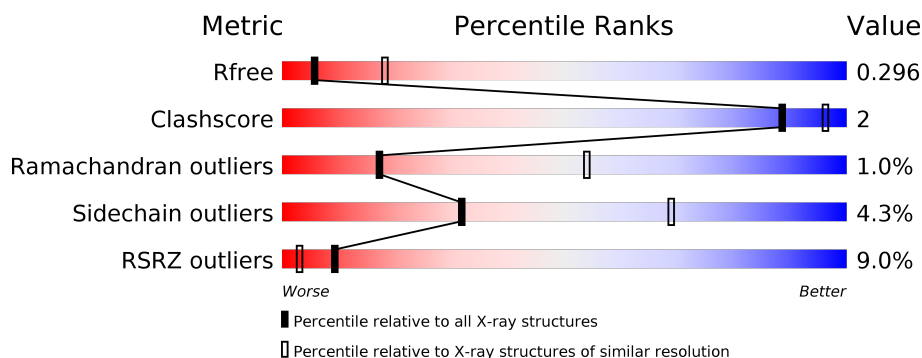
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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))
$R_{free}$	111664	2110 (3.04-3.00)
Clashscore	122126	2436 (3.04-3.00)
Ramachandran outliers	120053	2362 (3.04-3.00)
Sidechain outliers	120020	2365 (3.04-3.00)
RSRZ outliers	108989	2001 (3.04-3.00)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	65	<div> <div>69%</div> <div>12% • 15%</div> </div>
2	B	360	<div> <div>3%</div> <div>85%</div> <div>9% 6%</div> </div>
3	F	473	<div> <div>13%</div> <div>74%</div> <div>6% • 19%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6070 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 E3 envelope glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	55	Total	C	N	O	S	0	0	0
			427	268	67	83	9			

- Molecule 2 is a protein called E2 envelope glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	338	Total	C	N	O	S	0	1	0
			2669	1666	483	500	20			

- Molecule 3 is a protein called E1 envelope glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	382	Total	C	N	O	S	0	1	0
			2909	1838	487	560	24			

There is a discrepancy between the modelled and reference sequences:

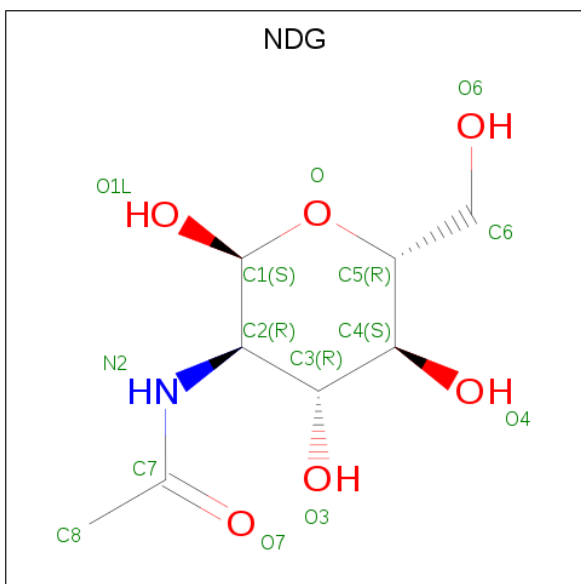
Chain	Residue	Modelled	Actual	Comment	Reference
F	0	GLY	-	expression tag	UNP Q1H8W5

- Molecule 4 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	F	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE (three-letter code: NDG) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	7	Total 7	O 7	0	0
6	F	16	Total 16	O 16	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.24Å 99.61Å 107.10Å 90.00° 102.90° 90.00°	Depositor
Resolution (Å)	50.00 – 3.01 72.07 – 3.01	Depositor EDS
% Data completeness (in resolution range)	98.3 (50.00-3.01) 98.3 (72.07-3.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.82 (at 3.01Å)	Xtriage
Refinement program	BUSTER 2.9.3	Depositor
R, $R_{free}$	0.252 , 0.275 0.272 , 0.296	Depositor DCC
$R_{free}$ test set	1029 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	70.7	Xtriage
Anisotropy	0.197	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 52.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.035 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	6070	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.76% 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

### 5.1 Standard geometry

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

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	0.45	0/438	0.62	0/598
2	B	0.37	0/2741	0.60	0/3731
3	F	0.35	0/2981	0.56	0/4064
All	All	0.37	0/6160	0.58	0/8393

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	427	0	410	6	0
2	B	2669	0	2587	8	0
3	F	2909	0	2818	12	0
4	B	14	0	12	0	0
4	F	14	0	13	0	0
5	B	14	0	13	0	0
6	B	7	0	0	0	0
6	F	16	0	0	0	0
All	All	6070	0	5853	25	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.

All (25) 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:6:VAL:O	1:A:6:VAL:HG12	1.71	0.90
1:A:6:VAL:O	1:A:6:VAL:CG1	2.30	0.79
2:B:198:ARG:HA	2:B:210:LEU:H	1.51	0.73
3:F:225:ALA:H	3:F:230:HIS:HE1	1.38	0.72
3:F:42:LEU:HD11	3:F:266:VAL:HG22	1.85	0.58
3:F:21:ARG:HH12	3:F:284:GLU:HG2	1.70	0.56
3:F:38:LEU:HB2	3:F:268:ALA:HB3	1.90	0.53
2:B:248:LEU:HD13	2:B:251:ARG:HB2	1.93	0.51
1:A:6:VAL:HG13	1:A:17:CYS:SG	2.51	0.49
3:F:60:VAL:HG22	3:F:102:GLN:HG3	1.95	0.49
1:A:6:VAL:CG1	1:A:17:CYS:SG	3.01	0.48
3:F:28:VAL:HG23	3:F:329:ALA:HB1	1.95	0.48
3:F:19:VAL:HB	3:F:27:MET:HB3	1.96	0.48
2:B:52:LEU:HD22	2:B:67:LEU:HD21	1.97	0.47
2:B:184:GLN:HG2	2:B:189:LYS:HB2	1.97	0.46
3:F:30:GLU:HB3	3:F:136:LEU:HB2	1.98	0.46
1:A:7:MET:HE3	1:A:15:PHE:C	2.38	0.43
2:B:184:GLN:HA	2:B:185:SER:HA	1.76	0.42
3:F:59:TYR:HB3	3:F:103:LEU:HB3	2.02	0.42
3:F:147:TYR:H	3:F:152:HIS:HD2	1.68	0.42
3:F:149:ASN:HD22	3:F:149:ASN:H	1.68	0.42
2:B:140:LYS:HB2	2:B:292:PRO:HB2	2.02	0.41
2:B:13:ARG:HE	2:B:236:GLN:HE21	1.69	0.41
1:A:55:LEU:HD13	2:B:8:VAL:HA	2.02	0.41
3:F:7:ILE:HG12	3:F:278:ILE:HD12	2.04	0.40

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	A	53/65 (82%)	50 (94%)	1 (2%)	2 (4%)	3	19
2	B	336/360 (93%)	313 (93%)	19 (6%)	4 (1%)	14	49
3	F	381/473 (80%)	369 (97%)	10 (3%)	2 (0%)	31	70
All	All	770/898 (86%)	732 (95%)	30 (4%)	8 (1%)	17	54

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	121	ILE
3	F	99	GLU
1	A	11	ALA
2	B	72	ASN
2	B	164	THR
3	F	126	THR
2	B	232	HIS
1	A	58	SER

### 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	52/61 (85%)	47 (90%)	5 (10%)	9	33
2	B	300/319 (94%)	287 (96%)	13 (4%)	32	69
3	F	319/378 (84%)	308 (97%)	11 (3%)	40	75
All	All	671/758 (88%)	642 (96%)	29 (4%)	32	69

All (29) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	6	VAL
1	A	7	MET
1	A	9	LEU
1	A	30	GLU
1	A	51	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	6	PHE
2	B	24	GLU
2	B	28	CYS
2	B	80	ARG
2	B	119	ARG
2	B	158	GLN
2	B	181	MET
2	B	192	VAL
2	B	212	THR
2	B	235	TRP
2	B	294	LEU
2	B	307	GLN
2	B	342	LEU
3	F	69	LYS
3	F	110	LYS
3	F	136	LEU
3	F	149	ASN
3	F	157	LYS
3	F	246	GLU
3	F	284	GLU
3	F	327	LYS
3	F	346	VAL
3	F	364	GLU
3	F	370	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	5	ASN
2	B	236	GLN
2	B	282	GLN
3	F	149	ASN
3	F	152	HIS
3	F	230	HIS

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

3 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$
4	NAG	B	1001	2,5	14,14,15	1.40	1 (7%)	17,19,21	1.53	3 (17%)
5	NDG	B	1002	4	14,14,15	1.27	1 (7%)	17,19,21	1.72	6 (35%)
4	NAG	F	1001	3	14,14,15	1.31	2 (14%)	17,19,21	1.18	1 (5%)

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
4	NAG	B	1001	2,5	-	0/6/23/26	0/1/1/1
5	NDG	B	1002	4	-	0/6/23/26	0/1/1/1
4	NAG	F	1001	3	-	0/6/23/26	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	F	1001	NAG	C3-C2	2.15	1.57	1.52
5	B	1002	NDG	C1-C2	2.80	1.56	1.52
4	F	1001	NAG	C1-C2	3.23	1.56	1.52
4	B	1001	NAG	C1-C2	3.67	1.57	1.52

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	1002	NDG	C1-C2-N2	2.06	114.01	110.49
4	B	1001	NAG	C4-C3-C2	2.14	114.16	111.02
5	B	1002	NDG	O-C5-C6	2.28	110.75	107.15
5	B	1002	NDG	C2-N2-C7	2.31	126.31	122.94
5	B	1002	NDG	C3-C4-C5	2.59	114.87	110.24
5	B	1002	NDG	O-C5-C4	3.08	118.32	110.83
4	B	1001	NAG	O4-C4-C3	3.11	117.61	110.34
4	F	1001	NAG	C4-C3-C2	3.18	115.67	111.02
5	B	1002	NDG	C1-O-C5	3.21	116.60	112.19
4	B	1001	NAG	C1-O5-C5	3.34	116.78	112.19

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 ⓘ

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	55/65 (84%)	0.13	0 100 100	63, 81, 92, 101	0
2	B	338/360 (93%)	0.25	10 (2%) 50 22	45, 77, 109, 155	0
3	F	382/473 (80%)	0.79	60 (15%) 2 0	48, 85, 165, 177	0
All	All	775/898 (86%)	0.51	70 (9%) 9 3	45, 81, 163, 177	0

All (70) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	326	GLY	10.7
3	F	327	LYS	7.2
3	F	379	GLU	6.3
3	F	365	PHE	6.1
3	F	360	LEU	6.0
3	F	328	CYS	5.9
2	B	342	LEU	5.7
3	F	351	GLN	4.8
3	F	353	GLN	4.8
2	B	206	SER	4.6
3	F	337	VAL	4.6
3	F	354	ILE	4.5
3	F	352	LEU	4.3
3	F	381	HIS	4.1
3	F	345	GLU	4.1
3	F	348	GLY	4.0
3	F	292	ASP	3.9
3	F	324	LYS	3.6
3	F	356	PHE	3.6
3	F	343	GLU	3.5
3	F	316	ALA	3.5
3	F	358	THR	3.4
3	F	318	ILE	3.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
3	F	333	MET	3.2
3	F	139	GLY	3.2
3	F	325	LYS	3.1
2	B	205	GLY	3.0
3	F	305	ALA	3.0
3	F	331	HIS	3.0
2	B	280	LYS	3.0
2	B	281	ASN	2.9
3	F	357	SER	2.9
3	F	349	ASN	2.9
3	F	362	SER	2.9
3	F	306	CYS	2.9
3	F	320	TYR	2.8
3	F	344	ILE	2.8
3	F	361	ALA	2.8
3	F	377	ALA	2.8
2	B	282	GLN	2.8
3	F	140	ASN	2.8
3	F	370	CYS	2.8
3	F	347	GLU	2.6
3	F	308	HIS	2.5
3	F	363	ALA	2.5
3	F	359	ALA	2.5
3	F	302	GLU	2.5
2	B	223	ASP	2.4
3	F	135	VAL	2.4
2	B	184	GLN	2.4
3	F	27	MET	2.3
3	F	364	GLU	2.3
2	B	146	GLN	2.3
3	F	294	PRO	2.3
2	B	162	ALA	2.2
3	F	313	GLY	2.2
3	F	278	ILE	2.2
3	F	311	ASP	2.2
3	F	373	GLN	2.2
3	F	136	LEU	2.2
3	F	312	PHE	2.1
3	F	144	VAL	2.1
3	F	19	VAL	2.1
3	F	303	VAL	2.1
3	F	161	PHE	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
3	F	1	TYR	2.1
3	F	378	ALA	2.1
3	F	366	ARG	2.0
3	F	371	SER	2.0
3	F	89	TRP	2.0

## 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 carbohydrates in this entry.

## 6.4 Ligands [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
4	NAG	F	1001	14/15	0.60	0.35	115,119,121,121	0
5	NDG	B	1002	14/15	0.82	0.27	85,89,91,91	0
4	NAG	B	1001	14/15	0.88	0.16	78,82,85,85	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.