



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:50 PM GMT

PDB ID : 4WEB  
Title : Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2  
Authors : Khan, A.G.; Whidby, J.; Miller, M.T.; Scarborough, H.; Zatorski, A.V.; Cygan, A.; Price, A.A.; Yost, S.A.; Bohannon, C.D.; Jacob, J.; Grakoui, A.; Marcotrigiano, J.  
Deposited on : 2014-09-09  
Resolution : 2.40 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

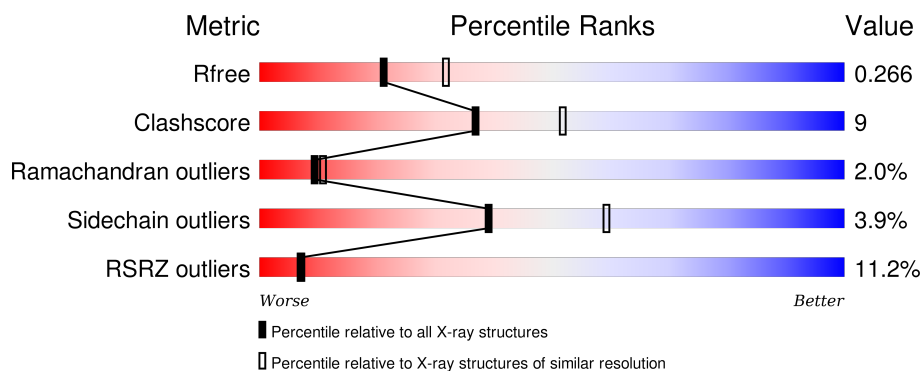
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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))
$R_{free}$	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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	E	217	
2	H	467	
3	L	240	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	NAG	E	701	-	-	-	X
5	ARF	H	501	-	-	-	X
5	ARF	H	502	-	-	-	X

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 4337 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 hepatitis C virus envelope glycoprotein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	123	Total	C	N	O	S	0	0	0
			891	573	144	161	13			

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	450	THR	-	expression tag	UNP Q9QF35
E	451	PRO	-	expression tag	UNP Q9QF35
E	452	VAL	-	expression tag	UNP Q9QF35
E	453	GLY	-	expression tag	UNP Q9QF35
E	454	LEU	-	expression tag	UNP Q9QF35
E	455	ALA	-	expression tag	UNP Q9QF35
E	656	GLY	-	expression tag	UNP Q9QF35
E	657	SER	-	expression tag	UNP Q9QF35
E	658	ALA	-	expression tag	UNP Q9QF35
E	659	SER	-	expression tag	UNP Q9QF35
E	660	GLY	-	expression tag	UNP Q9QF35
E	661	LEU	-	expression tag	UNP Q9QF35
E	662	GLU	-	expression tag	UNP Q9QF35
E	663	VAL	-	expression tag	UNP Q9QF35
E	664	LEU	-	expression tag	UNP Q9QF35
E	665	PHE	-	expression tag	UNP Q9QF35
E	666	GLN	-	expression tag	UNP Q9QF35

- Molecule 2 is a protein called Mouse Fab Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	216	Total	C	N	O	S	0	0	0
			1616	1015	264	332	5			

- Molecule 3 is a protein called Mouse Fab Light Chain.

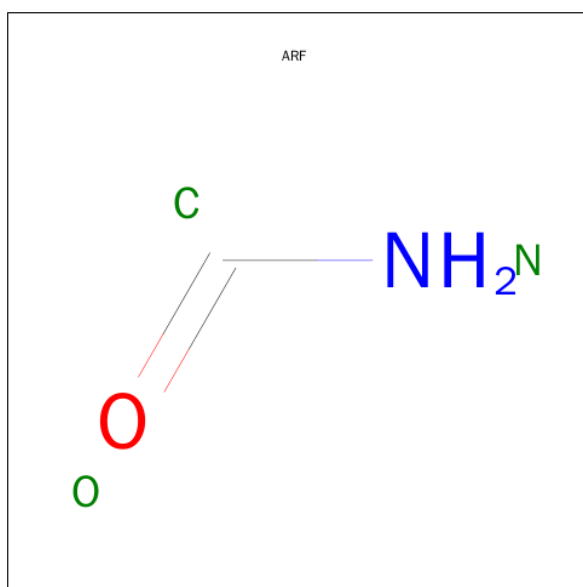
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	L	217	Total	C	N	O	S	0	0	0
			1648	1025	273	342	8			

- Molecule 4 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

- Molecule 5 is FORMAMIDE (three-letter code: ARF) (formula:  $CH_3NO$ ).

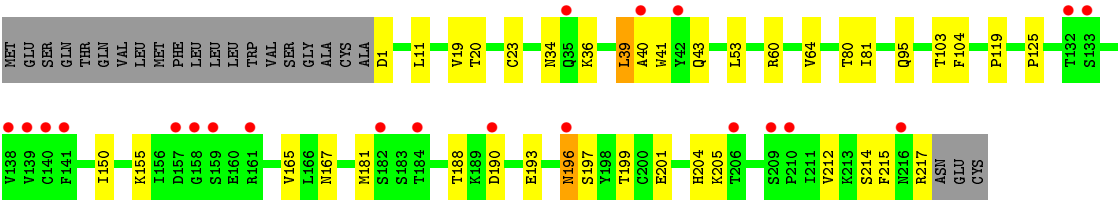


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	E	1	Total	C	N	O	0	0
			3	1	1	1		
5	H	1	Total	C	N	O	0	0
			3	1	1	1		
5	H	1	Total	C	N	O	0	0
			3	1	1	1		
5	H	1	Total	C	N	O	0	0
			3	1	1	1		
5	L	1	Total	C	N	O	0	0
			3	1	1	1		
5	L	1	Total	C	N	O	0	0
			3	1	1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	E	12	Total	O	0	0
			12	12		
6	H	61	Total	O	0	0
			61	61		
6	L	63	Total	O	0	0
			63	63		







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.96Å 194.57Å 37.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.69 – 2.40 29.91 – 2.40	Depositor EDS
% Data completeness (in resolution range)	94.2 (24.69-2.40) 94.2 (29.91-2.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.83 (at 2.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1690)	Depositor
R, $R_{free}$	0.199 , 0.259 0.211 , 0.266	Depositor DCC
$R_{free}$ test set	1998 reflections (8.95%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.9	Xtriage
Anisotropy	0.338	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 54.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 24344 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4337	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.00% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NAG, ARF

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	E	0.46	1/920 (0.1%)	0.72	2/1269 (0.2%)
2	H	0.48	0/1655	0.66	0/2268
3	L	0.46	0/1685	0.62	0/2295
All	All	0.47	1/4260 (0.0%)	0.66	2/5832 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	492	PRO	N-CD	5.02	1.54	1.47

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	643	ARG	NE-CZ-NH2	-5.81	117.39	120.30
1	E	493	PRO	CA-N-CD	-5.21	104.20	111.50

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	E	891	0	812	26	0
2	H	1616	0	1561	34	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	L	1648	0	1535	19	0
4	E	28	0	26	1	0
5	E	3	0	3	1	0
5	H	9	0	9	0	0
5	L	6	0	6	0	0
6	E	12	0	0	2	0
6	H	61	0	0	0	0
6	L	63	0	0	1	0
All	All	4337	0	3952	77	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 9.

The worst 5 of 77 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:H:124:PRO:HB3	2:H:150:TYR:HB3	1.70	0.74
1:E:499:VAL:O	1:E:539:PHE:HB3	1.91	0.69
1:E:501:ALA:HB2	1:E:539:PHE:HE2	1.59	0.68
2:H:55:ASN:HB3	2:H:57:HIS:H	1.60	0.67
1:E:501:ALA:HB2	1:E:539:PHE:CE2	2.30	0.66

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	E	117/217 (54%)	97 (83%)	14 (12%)	6 (5%)	2	1
2	H	212/467 (45%)	196 (92%)	12 (6%)	4 (2%)	10	12
3	L	215/240 (90%)	205 (95%)	9 (4%)	1 (0%)	34	48

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	544/924 (59%)	498 (92%)	35 (6%)	11 (2%)	9	11

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	502	LYS
1	E	540	LEU
2	H	177	SER
2	H	176	GLN
2	H	195	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	E	96/186 (52%)	92 (96%)	4 (4%)	36	56
2	H	185/420 (44%)	179 (97%)	6 (3%)	46	68
3	L	186/214 (87%)	178 (96%)	8 (4%)	35	55
All	All	467/820 (57%)	449 (96%)	18 (4%)	39	59

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	201	ASN
2	H	213	LYS
3	L	165	VAL
2	H	188	VAL
2	H	192	THR

Some 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

8 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	E	701	1	14,14,15	0.33	0	15,19,21	0.32	0
4	NAG	E	702	1	14,14,15	0.39	0	15,19,21	0.30	0
5	ARF	E	703	-	2,2,2	1.12	0	1,1,1	0.67	0
5	ARF	H	501	-	2,2,2	1.20	0	1,1,1	0.64	0
5	ARF	H	502	-	2,2,2	1.18	0	1,1,1	0.91	0
5	ARF	H	503	-	2,2,2	1.26	0	1,1,1	0.46	0
5	ARF	L	301	-	2,2,2	1.12	0	1,1,1	0.74	0
5	ARF	L	302	-	2,2,2	1.36	0	1,1,1	0.41	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
4	NAG	E	701	1	-	0/6/23/26	0/1/1/1
4	NAG	E	702	1	-	0/6/23/26	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	ARF	E	703	-	-	0/0/0/0	0/0/0/0
5	ARF	H	501	-	-	0/0/0/0	0/0/0/0
5	ARF	H	502	-	-	0/0/0/0	0/0/0/0
5	ARF	H	503	-	-	0/0/0/0	0/0/0/0
5	ARF	L	301	-	-	0/0/0/0	0/0/0/0
5	ARF	L	302	-	-	0/0/0/0	0/0/0/0

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.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	E	701	NAG	1	0
5	E	703	ARF	1	0

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	E	123/217 (56%)	1.03	24 (19%) <b>1</b> <b>1</b>	34, 70, 97, 112	0
2	H	216/467 (46%)	0.15	17 (7%) <b>15</b> <b>15</b>	24, 40, 75, 102	0
3	L	217/240 (90%)	0.32	21 (9%) <b>10</b> <b>9</b>	24, 41, 77, 86	0
All	All	556/924 (60%)	0.41	62 (11%) <b>7</b> <b>7</b>	24, 45, 86, 112	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	538	VAL	6.9
1	E	491	TYR	6.6
1	E	501	ALA	5.2
1	E	596	THR	5.2
1	E	539	PHE	5.2

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NAG	E	701	14/15	0.80	0.40	3.12	91,94,101,102	0
5	ARF	H	502	3/3	0.84	0.16	3.09	36,36,36,44	0
5	ARF	H	501	3/3	0.97	0.17	2.15	32,32,35,37	0
5	ARF	H	503	3/3	0.93	0.23	1.36	32,32,36,41	0
5	ARF	E	703	3/3	0.96	0.16	1.25	40,40,47,50	0
5	ARF	L	302	3/3	0.92	0.17	0.59	33,33,36,42	0
4	NAG	E	702	14/15	0.67	0.48	-	83,92,99,103	0
5	ARF	L	301	3/3	0.85	0.14	-	52,52,53,54	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.