



# wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Feb 17, 2018 – 02:39 am GMT

PDB ID : 5UPW  
EMDB ID: : EMD-8595  
Title : CryoEM Structure Refinement by Integrating NMR Chemical Shifts with  
Molecular Dynamics Simulations  
Authors : Perilla, J.R.  
Deposited on : 2017-02-04  
Resolution : 5.00 Å(reported)  
Based on PDB ID : 4XFX

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report  
for a publicly released PDB/EMDB 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/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30686

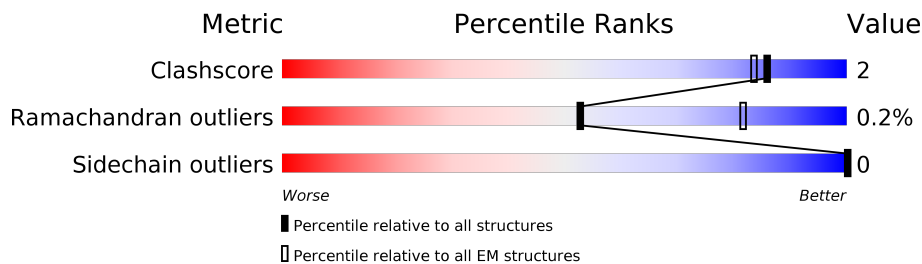
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 5.00 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)	EM structures (#Entries)
Clashscore	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	221	
1	B	221	
1	C	221	
1	D	221	
1	E	221	
1	F	221	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 10350 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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.

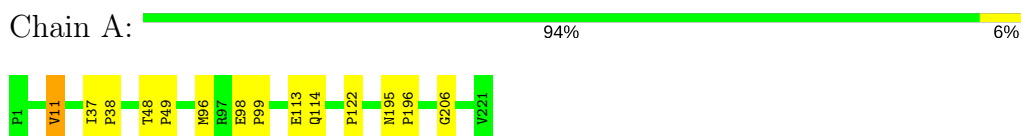
- Molecule 1 is a protein called Gag polyprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	221	Total 1725	1088	301	323	13	0	0
1	B	221	Total 1725	1088	301	323	13	0	0
1	C	221	Total 1725	1088	301	323	13	0	0
1	D	221	Total 1725	1088	301	323	13	0	0
1	E	221	Total 1725	1088	301	323	13	0	0
1	F	221	Total 1725	1088	301	323	13	0	0

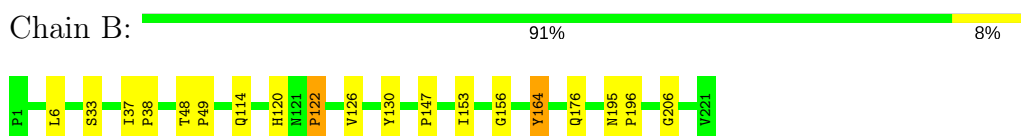
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

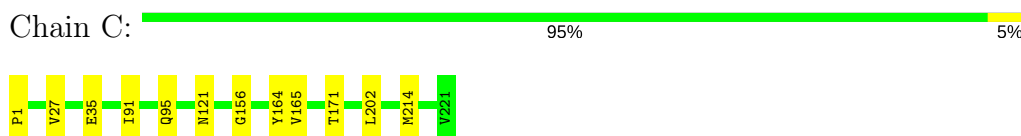
- Molecule 1: Gag polyprotein



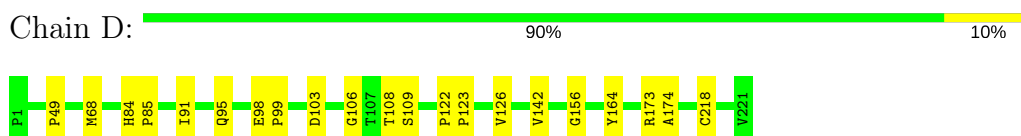
- Molecule 1: Gag polyprotein



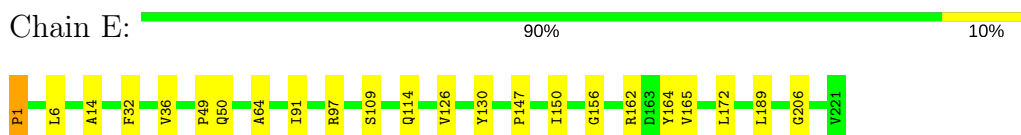
- Molecule 1: Gag polyprotein



- Molecule 1: Gag polyprotein



- Molecule 1: Gag polyprotein



- Molecule 1: Gag polyprotein





## 4 Experimental information

Property	Value	Source
Reconstruction method	HELICAL	Depositor
Imposed symmetry	HELICAL, twist=-31.13°, rise=6.94 Å, axial sym=C1	Depositor
Number of segments used	38452	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{Å}^2$ )	41	Depositor
Minimum defocus (nm)	700	Depositor
Maximum defocus (nm)	2200	Depositor
Magnification	31000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 5 Model quality i

### 5.1 Standard geometry i

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  >2	RMSZ	# Z  >2
1	A	1.03	1/1764 (0.1%)	0.93	1/2398 (0.0%)
1	B	1.03	1/1764 (0.1%)	0.93	2/2398 (0.1%)
1	C	1.04	1/1764 (0.1%)	0.94	3/2398 (0.1%)
1	D	1.04	0/1764	0.93	2/2398 (0.1%)
1	E	1.04	0/1764	0.96	3/2398 (0.1%)
1	F	1.03	0/1764	0.94	1/2398 (0.0%)
All	All	1.03	3/10584 (0.0%)	0.94	12/14388 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	3
1	C	0	2
1	D	0	1
1	E	0	2
1	F	0	2
All	All	0	11

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	122	PRO	CA-C	-5.27	1.42	1.52
1	B	122	PRO	CA-C	-5.23	1.42	1.52
1	C	1	PRO	N-CA	5.13	1.55	1.47

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	164	TYR	CB-CG-CD2	-7.97	116.22	121.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	130	TYR	CB-CG-CD1	-7.03	116.78	121.00
1	D	173	ARG	NE-CZ-NH2	-6.83	116.88	120.30
1	B	164	TYR	CB-CG-CD1	-6.77	116.94	121.00
1	E	1	PRO	CA-N-CD	-5.89	103.25	111.50

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	206	GLY	Mainchain
1	B	156	GLY	Mainchain
1	B	206	GLY	Mainchain
1	B	33	SER	Mainchain
1	C	121	ASN	Peptide

## 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	1725	0	1725	7	0
1	B	1725	0	1725	8	0
1	C	1725	0	1725	5	0
1	D	1725	0	1727	10	0
1	E	1725	0	1725	13	0
1	F	1725	0	1725	9	0
All	All	10350	0	10352	46	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 46 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:C:35:GLU:OE1	1:C:35:GLU:N	2.22	0.71
1:A:11:VAL:HG21	1:B:6:LEU:HA	1.83	0.59

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:6:LEU:HD11	1:F:9:GLN:HB3	1.87	0.56
1:C:165:VAL:HG12	1:E:64:ALA:HB2	1.91	0.53
1:B:37:ILE:HB	1:B:38:PRO:CD	2.40	0.52

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 PDB entries followed by that with respect to all EM entries.

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	219/221 (99%)	208 (95%)	11 (5%)	0	100	100
1	B	219/221 (99%)	209 (95%)	9 (4%)	1 (0%)	31	73
1	C	219/221 (99%)	210 (96%)	9 (4%)	0	100	100
1	D	219/221 (99%)	213 (97%)	6 (3%)	0	100	100
1	E	219/221 (99%)	205 (94%)	13 (6%)	1 (0%)	31	73
1	F	219/221 (99%)	210 (96%)	8 (4%)	1 (0%)	31	73
All	All	1314/1326 (99%)	1255 (96%)	56 (4%)	3 (0%)	53	84

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	147	PRO
1	F	147	PRO
1	E	147	PRO

### 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 PDB entries followed by that with respect to all EM entries.

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	188/188 (100%)	188 (100%)	0	100	100
1	B	188/188 (100%)	188 (100%)	0	100	100
1	C	188/188 (100%)	188 (100%)	0	100	100
1	D	188/188 (100%)	188 (100%)	0	100	100
1	E	188/188 (100%)	188 (100%)	0	100	100
1	F	188/188 (100%)	188 (100%)	0	100	100
All	All	1128/1128 (100%)	1128 (100%)	0	100	100

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

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

### 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 carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

### 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.