wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:47 PM BST

PDB ID	:	3J4F
EMDB ID:	:	EMD-5582
Title	:	Structure of HIV-1 capsid protein by cryo-EM
Authors	:	Zhao, G.; Perilla, J.R.; Meng, X.; Schulten, K.; Zhang, P.
Deposited on	:	2013-07-11
Resolution	:	8.60 Å(reported)
Based on PDB ID	:	2KOD, 3H47

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

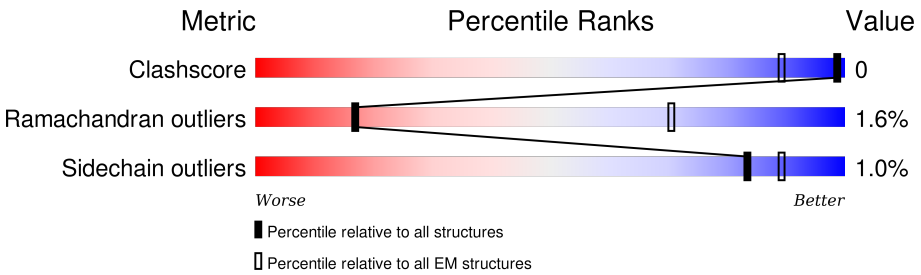
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 8.60 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	A	231	95% 5%
1	B	231	95% • •
1	C	231	94% 6%
1	D	231	94% 6% •
1	E	231	94% 5% •
1	F	231	94% 6%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 10799 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 capsid protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	231	1800	1134	317	336	13	0	0
1	B	231	1800	1134	317	336	13	0	0
1	C	231	1800	1134	317	336	13	0	0
1	D	231	1799	1133	317	336	13	0	0
1	E	231	1800	1134	317	336	13	0	0
1	F	231	1800	1134	317	336	13	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	92	GLU	ALA	ENGINEERED MUTATION	UNP Q79791
B	92	GLU	ALA	ENGINEERED MUTATION	UNP Q79791
C	92	GLU	ALA	ENGINEERED MUTATION	UNP Q79791
D	92	GLU	ALA	ENGINEERED MUTATION	UNP Q79791
E	92	GLU	ALA	ENGINEERED MUTATION	UNP Q79791
F	92	GLU	ALA	ENGINEERED MUTATION	UNP Q79791

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 errors 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: capsid protein

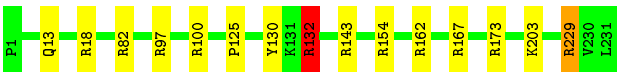
Chain A: 95% 5%

- Molecule 1: capsid protein

Chain B: 95%

- Molecule 1: capsid protein

Chain C: 94% 6%



- Molecule 1: capsid protein

Chain D: 94% 6%

- Molecule 1: capsid protein

Chain E: 94% 5%

- Molecule 1: capsid protein

Chain F: 94% 6%

P1
I15
R82
V86
R97
R100
H120
P125
Y130
R143
P160
F161
R162
D163
Y164
R167
R173
A174
E175
I231

4 Experimental information (i)

Property	Value	Source
Reconstruction method	HELICAL	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	3210	Depositor
Resolution determination method	FSC at 0.5 cut-off	Depositor
CTF correction method	each filament	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	200	Depositor
Electron dose $(e^-/\text{Å}^2)$	15	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	59000	Depositor
Image detector	Kodak SO163	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	0.66	0/1841	1.05	11/2500 (0.4%)
1	B	0.65	0/1841	1.05	10/2500 (0.4%)
1	C	0.66	0/1841	1.07	12/2500 (0.5%)
1	D	0.66	0/1838	1.04	9/2494 (0.4%)
1	E	0.66	0/1841	1.01	6/2500 (0.2%)
1	F	0.65	0/1841	1.04	9/2500 (0.4%)
All	All	0.66	0/11043	1.04	57/14994 (0.4%)

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	B	0	4
1	C	0	5
1	D	0	2
1	E	0	2
1	F	0	1
All	All	0	14

There are no bond length outliers.

The worst 5 of 57 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	82	ARG	NE-CZ-NH1	9.84	125.22	120.30
1	B	162	ARG	NE-CZ-NH1	8.96	124.78	120.30
1	F	82	ARG	NE-CZ-NH1	8.91	124.75	120.30
1	B	162	ARG	NE-CZ-NH2	-8.48	116.06	120.30
1	C	82	ARG	NE-CZ-NH1	8.46	124.53	120.30

There are no chirality outliers.

5 of 14 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	132	ARG	Sidechain
1	B	164	TYR	Sidechain
1	B	82	ARG	Sidechain
1	B	97	ARG	Sidechain
1	C	18	ARG	Sidechain

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	1800	0	1802	0	0
1	B	1800	0	1802	0	0
1	C	1800	0	1802	0	0
1	D	1799	0	1800	3	0
1	E	1800	0	1802	1	0
1	F	1800	0	1802	0	0
All	All	10799	0	10810	4	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 0.

All (4) 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:173:ARG:CA	1:D:174:ALA:N	2.47	0.78
1:D:173:ARG:O	1:D:174:ALA:N	2.24	0.70
1:D:173:ARG:CA	1:D:173:ARG:O	2.41	0.69
1:E:224:PRO:HB3	1:E:226:HIS:CE1	2.56	0.41

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	229/231 (99%)	218 (95%)	8 (4%)	3 (1%)	15	60
1	B	229/231 (99%)	212 (93%)	14 (6%)	3 (1%)	15	60
1	C	229/231 (99%)	219 (96%)	9 (4%)	1 (0%)	39	80
1	D	227/231 (98%)	218 (96%)	6 (3%)	3 (1%)	15	60
1	E	229/231 (99%)	215 (94%)	8 (4%)	6 (3%)	7	45
1	F	229/231 (99%)	217 (95%)	6 (3%)	6 (3%)	7	45
All	All	1372/1386 (99%)	1299 (95%)	51 (4%)	22 (2%)	17	56

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	9	GLN
1	B	226	HIS
1	E	88	ALA
1	E	149	SER
1	F	86	VAL

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	195/195 (100%)	194 (100%)	1 (0%)	92	96
1	B	195/195 (100%)	194 (100%)	1 (0%)	92	96
1	C	195/195 (100%)	192 (98%)	3 (2%)	72	88

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	195/195 (100%)	193 (99%)	2 (1%)	82	92
1	E	195/195 (100%)	191 (98%)	4 (2%)	61	84
1	F	195/195 (100%)	194 (100%)	1 (0%)	92	96
All	All	1170/1170 (100%)	1158 (99%)	12 (1%)	83	92

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	154	ARG
1	D	163	ASP
1	E	158	LYS
1	C	203	LYS
1	E	18	ARG

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

Mol	Chain	Res	Type
1	E	226	HIS

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 (i)

There are no chain breaks in this entry.