



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Mar 2, 2017 – 11:44 am GMT

PDB ID : 3JCA
EMDB ID: : EMD-6441
Title : Core model of the Mouse Mammary Tumor Virus intasome
Authors : Lyumkis, D.L.; Ballandras-Colas, A.; Brown, M.; Cook, N.J.; Dewdney, T.G.;
Demeler, B.; Cherepanov, P.; Engelman, A.N.
Deposited on : 2015-11-24
Resolution : 4.80 Å(reported)
Based on PDB ID : 5D7U, 5CZ2, 5CZ1

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
A user guide is available at
<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : recalc29047

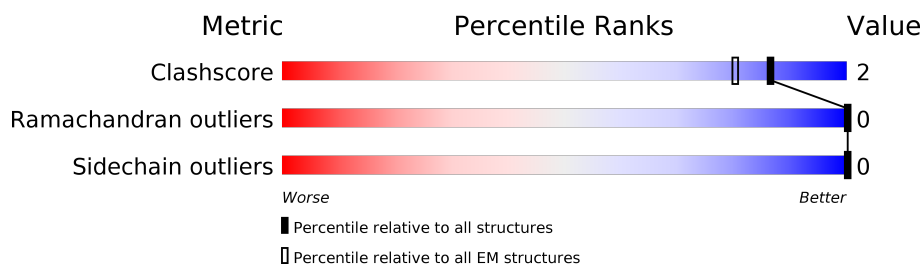
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.80 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	265	94% 5% .
1	B	265	91% 5% 5%
1	E	265	95% . .
1	F	265	94% . 5%
2	C	49	98% .
2	D	49	98% .
2	G	49	98% .
2	H	49	96% .
3	I	22	68% 23% 5% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	K	22	<div><div></div><div>73%23%5%</div></div>
4	J	20	<div><div></div><div>70%25%5%</div></div>
4	L	20	<div><div></div><div>75%20%5%</div></div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11466 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 Integrase.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	262	Total	C	N	O	S	0	0
			2096	1340	383	365	8		
1	B	253	Total	C	N	O	S	0	0
			2031	1299	370	354	8		
1	E	262	Total	C	N	O	S	0	0
			2096	1340	383	365	8		
1	F	253	Total	C	N	O	S	0	0
			2031	1299	370	354	8		

- Molecule 2 is a protein called Integrase.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	49	Total	C	N	O	S	0	0
			393	259	64	67	3		
2	D	49	Total	C	N	O	S	0	0
			393	259	64	67	3		
2	G	49	Total	C	N	O	S	0	0
			393	259	64	67	3		
2	H	49	Total	C	N	O	S	0	0
			393	259	64	67	3		

- Molecule 3 is a DNA chain called 5'-D(*AP*AP*TP*GP*CP*CP*GP*CP*AP*GP*TP*C
P*GP*GP*CP*CP*GP*AP*CP*CP*TP*G)-3'.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	I	21	Total	C	N	O	P	0	0
			425	202	80	123	20		
3	K	21	Total	C	N	O	P	0	0
			425	202	80	123	20		

- Molecule 4 is a DNA chain called 5'-D(*CP*AP*GP*GP*TP*CP*GP*GP*CP*CP*GP*A
P*CP*TP*GP*CP*GP*GP*CP*A)-3'.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	J	19	Total	C	N	O	P	0	0
			393	184	77	113	19		
4	L	19	Total	C	N	O	P	0	0
			393	184	77	113	19		

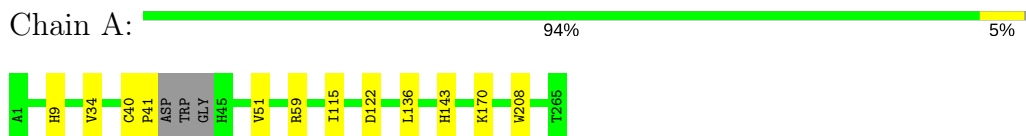
- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
5	B	1	Total	Zn	0
			1	1	
5	A	1	Total	Zn	0
			1	1	
5	F	1	Total	Zn	0
			1	1	
5	E	1	Total	Zn	0
			1	1	

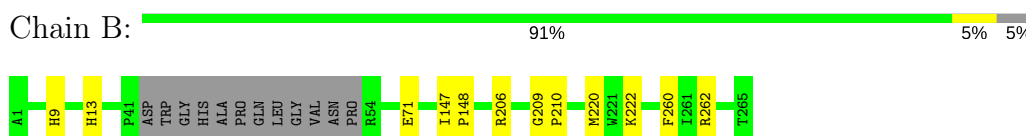
3 Residue-property plots [i](#)

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.

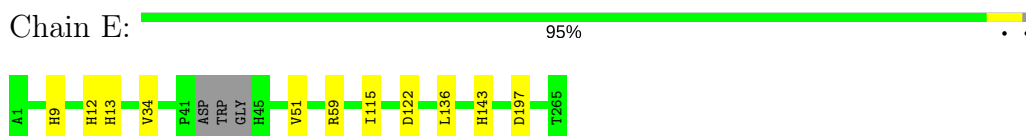
- Molecule 1: Integrase



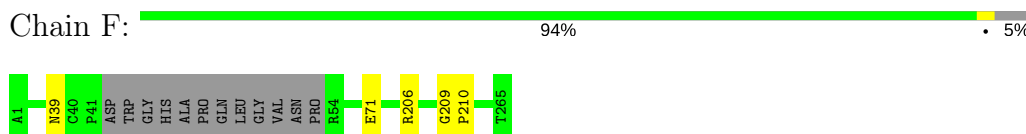
- Molecule 1: Integrase



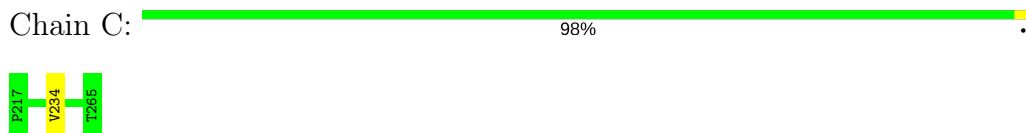
- Molecule 1: Integrase



- Molecule 1: Integrase

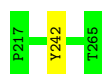


- Molecule 2: Integrase



- Molecule 2: Integrase





- Molecule 2: Integrase

Chain G: 98%



- Molecule 2: Integrase

Chain H: 96%



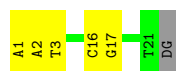
- Molecule 3: 5'-D(*AP*AP*TP*GP*CP*CP*GP*CP*AP*GP*TP*CP*GP*GP*CP*CP*GP*A
P*CP*CP*TP*G)-3'

Chain I: 68% 23% 5% 5%



- Molecule 3: 5'-D(*AP*AP*TP*GP*CP*CP*GP*CP*AP*GP*TP*CP*GP*GP*CP*CP*GP*A
P*CP*CP*TP*G)-3'

Chain K: 73% 23% 5%



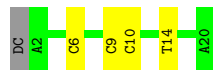
- Molecule 4: 5'-D(*CP*AP*GP*GP*TP*CP*GP*GP*CP*CP*GP*AP*CP*TP*GP*CP*GP*G
P*CP*A)-3'

Chain J: 70% 25% 5%



- Molecule 4: 5'-D(*CP*AP*GP*GP*TP*CP*GP*GP*CP*CP*GP*AP*CP*TP*GP*CP*GP*G
P*CP*A)-3'

Chain L: 75% 20% 5%



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	30307	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	each particle	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	40	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	38167	Depositor
Image detector	GATAN K2 (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.70	0/2160	0.69	1/2938 (0.0%)
1	B	0.69	0/2092	0.67	1/2843 (0.0%)
1	E	0.69	0/2160	0.69	1/2938 (0.0%)
1	F	0.69	0/2092	0.67	1/2843 (0.0%)
2	C	0.70	0/408	0.77	0/556
2	D	0.74	0/408	0.72	0/556
2	G	0.71	0/408	0.77	0/556
2	H	0.74	0/408	0.72	0/556
3	I	0.81	0/476	1.17	2/732 (0.3%)
3	K	0.79	0/476	1.13	1/732 (0.1%)
4	J	0.85	0/441	1.26	3/679 (0.4%)
4	L	0.86	0/441	1.27	3/679 (0.4%)
All	All	0.72	0/11970	0.80	13/16608 (0.1%)

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	L	10	DC	O4'-C1'-N1	6.66	112.66	108.00
1	E	59	ARG	NE-CZ-NH2	-6.15	117.23	120.30
1	A	59	ARG	NE-CZ-NH2	-6.10	117.25	120.30
1	B	206	ARG	NE-CZ-NH2	-5.81	117.39	120.30
4	J	10	DC	O4'-C1'-N1	5.74	112.02	108.00

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	2096	0	2064	10	0
1	B	2031	0	2001	6	0
1	E	2096	0	2064	9	0
1	F	2031	0	2001	3	0
2	C	393	0	386	1	0
2	D	393	0	386	1	0
2	G	393	0	386	1	0
2	H	393	0	386	2	0
3	I	425	0	236	7	0
3	K	425	0	236	7	0
4	J	393	0	212	3	0
4	L	393	0	212	2	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
5	E	1	0	0	0	0
5	F	1	0	0	0	0
All	All	11466	0	10570	50	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 50 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:9:HIS:CE1	1:A:34:VAL:HG12	2.11	0.86
3:K:1:DA:N3	3:K:1:DA:H3'	1.99	0.76
1:E:9:HIS:CE1	1:E:34:VAL:HG12	2.21	0.75
3:I:1:DA:H3'	3:I:1:DA:N3	2.02	0.75
3:K:1:DA:H4'	3:K:2:DA:OP1	1.89	0.72

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 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	258/265 (97%)	246 (95%)	12 (5%)	0	100	100
1	B	249/265 (94%)	231 (93%)	18 (7%)	0	100	100
1	E	258/265 (97%)	243 (94%)	15 (6%)	0	100	100
1	F	249/265 (94%)	232 (93%)	17 (7%)	0	100	100
2	C	47/49 (96%)	39 (83%)	8 (17%)	0	100	100
2	D	47/49 (96%)	43 (92%)	4 (8%)	0	100	100
2	G	47/49 (96%)	39 (83%)	8 (17%)	0	100	100
2	H	47/49 (96%)	42 (89%)	5 (11%)	0	100	100
All	All	1202/1256 (96%)	1115 (93%)	87 (7%)	0	100	100

There are no Ramachandran outliers to report.

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 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	221/223 (99%)	221 (100%)	0	100	100
1	B	214/223 (96%)	214 (100%)	0	100	100
1	E	221/223 (99%)	221 (100%)	0	100	100
1	F	214/223 (96%)	214 (100%)	0	100	100
2	C	42/42 (100%)	42 (100%)	0	100	100
2	D	42/42 (100%)	42 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	G	42/42 (100%)	42 (100%)	0	100	100
2	H	42/42 (100%)	42 (100%)	0	100	100
All	All	1038/1060 (98%)	1038 (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](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

There are no chain breaks in this entry.