



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

Jul 24, 2017 – 02:30 AM EDT

PDB ID : 5M0Q  
EMDB ID: : EMD-4138  
Title : Cryo-EM reconstruction of the maedi-visna virus (MVV) intasome  
Authors : Ballandras-Colas, A.; Maskell, D.; Pye, V.E.; Locke, J.; Swuec, S.; Kotecha, A.; Costa, A.; Cherepanov, P.  
Deposited on : unknown  
Resolution : 4.91 Å(reported)

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

<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) : rb-20029824

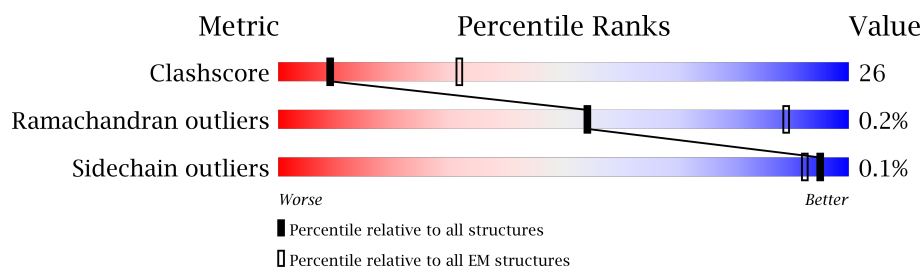
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 4.91 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  $\geq 3$ , 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	281	46% 49% ..
1	B	281	61% 28% . 9%
1	C	281	73% 20% 7%
1	D	281	58% 18% . 23%
1	E	281	63% 28% . 9%
1	F	281	61% 33% . 5%
1	G	281	55% 18% 27%
1	H	281	53% 24% 23%
1	I	281	42% 53% ..

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	J	281	
1	K	281	
1	L	281	
1	M	281	
1	N	281	
1	O	281	
1	P	281	
2	Q	21	
2	S	21	
3	R	19	
3	T	19	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 33171 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	275	Total	C	N	O	S	0	0
			2201	1400	385	407	9		
1	B	256	Total	C	N	O	S	0	0
			2075	1321	363	383	8		
1	C	261	Total	C	N	O	S	0	0
			2123	1353	371	391	8		
1	D	216	Total	C	N	O	S	0	0
			1729	1110	297	316	6		
1	E	257	Total	C	N	O	S	0	0
			2090	1333	365	384	8		
1	F	266	Total	C	N	O	S	0	0
			2148	1367	378	395	8		
1	G	205	Total	C	N	O	S	0	0
			1657	1051	284	314	8		
1	H	217	Total	C	N	O	S	0	0
			1761	1129	308	318	6		
1	I	275	Total	C	N	O	S	0	0
			2201	1400	385	407	9		
1	J	256	Total	C	N	O	S	0	0
			2075	1321	363	383	8		
1	K	261	Total	C	N	O	S	0	0
			2123	1353	371	391	8		
1	L	216	Total	C	N	O	S	0	0
			1729	1110	297	316	6		
1	M	256	Total	C	N	O	S	0	0
			2076	1322	363	383	8		
1	N	264	Total	C	N	O	S	0	0
			2131	1355	376	392	8		
1	O	205	Total	C	N	O	S	0	0
			1657	1051	284	314	8		
1	P	217	Total	C	N	O	S	0	0
			1761	1129	308	318	6		

- Molecule 2 is a DNA chain called vDNA, non-transferred strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	Q	21	Total 428	C 203	N 79	O 126	P 20	0	0
2	S	21	Total 428	C 203	N 79	O 126	P 20	0	0

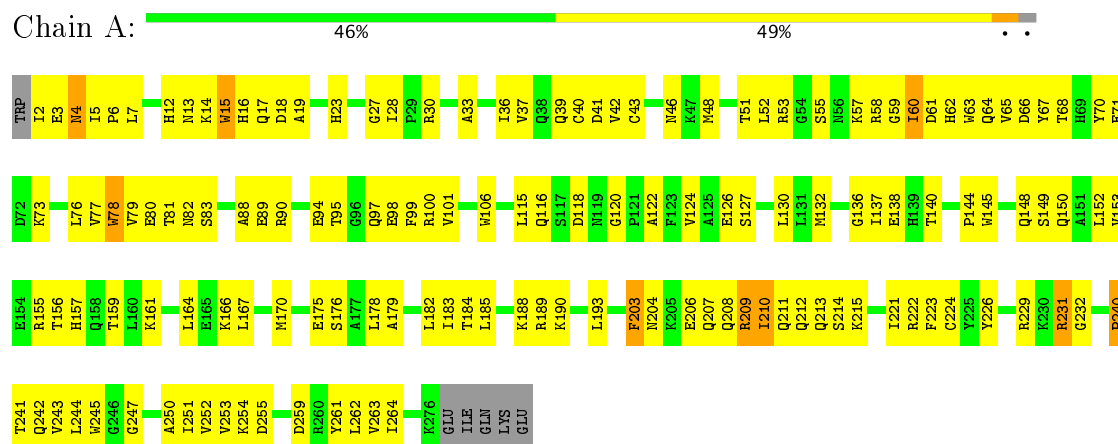
- Molecule 3 is a DNA chain called vDNA, transferred strand.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	R	19	Total 389	C 183	N 75	O 112	P 19	0	0
3	T	19	Total 389	C 183	N 75	O 112	P 19	0	0

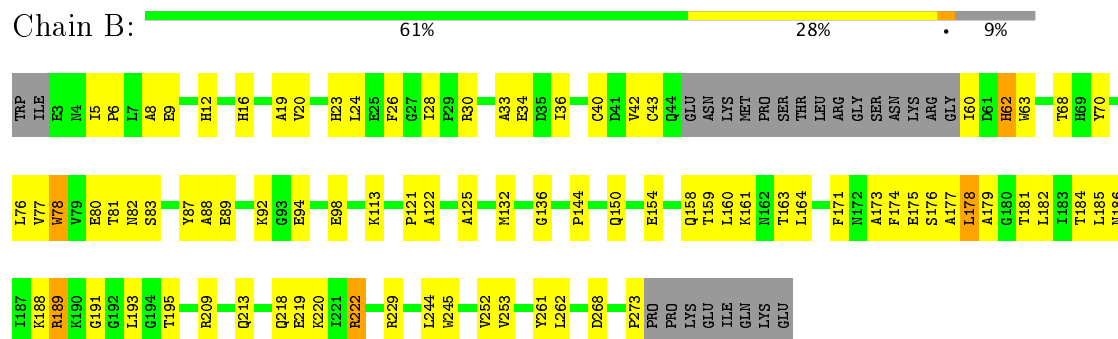
### 3 Residue-property plots

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.

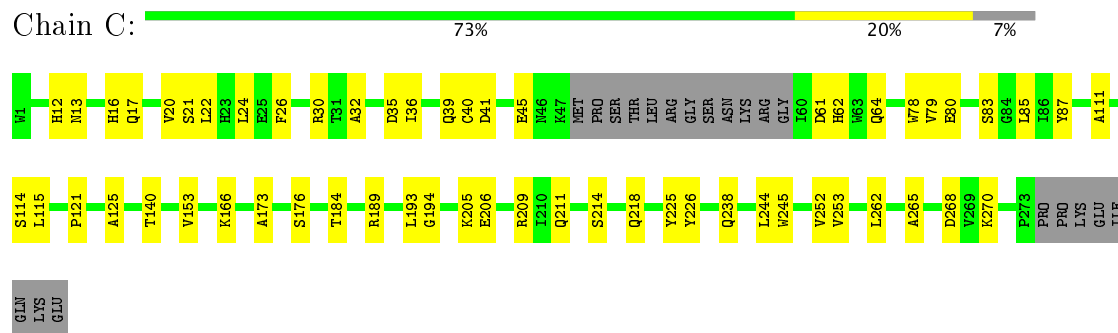
#### • Molecule 1: integrase



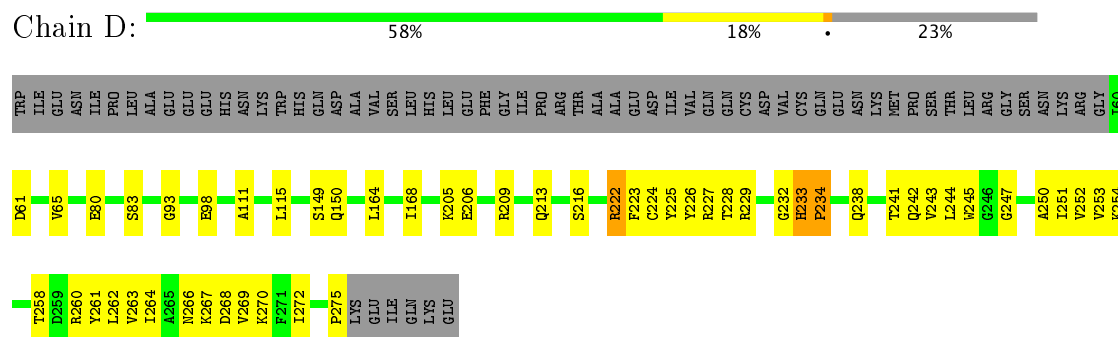
#### • Molecule 1: integrase



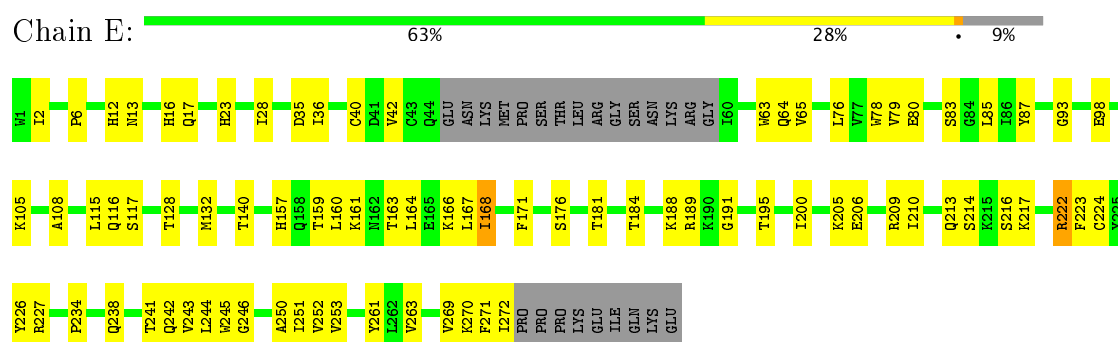
#### • Molecule 1: integrase



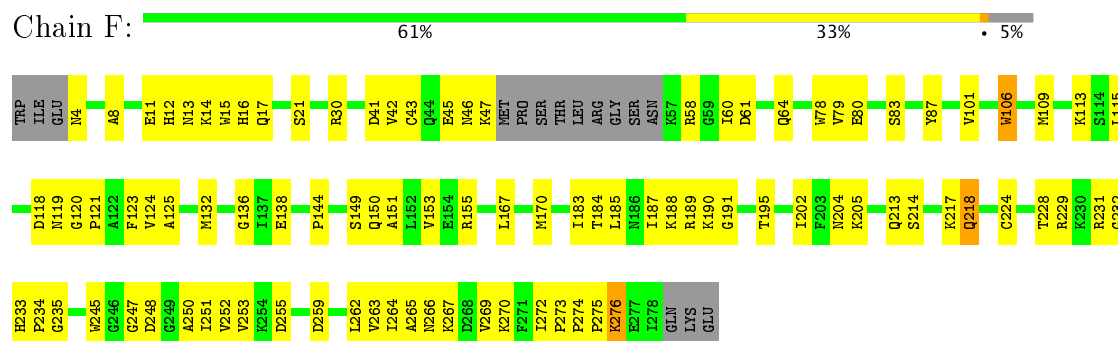
- Molecule 1: integrase



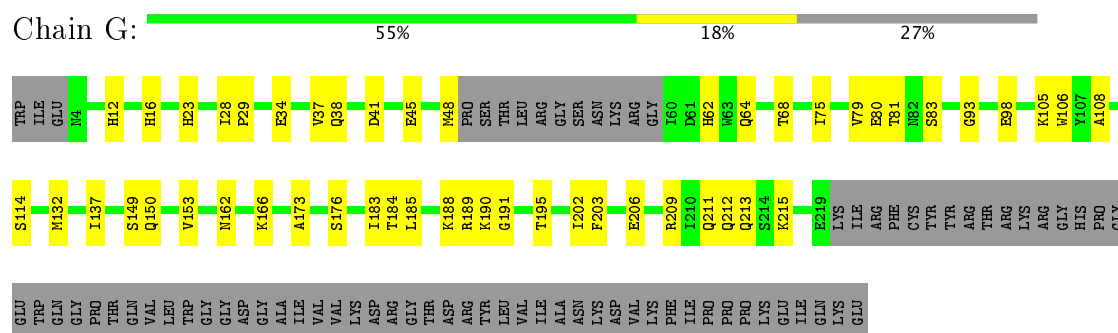
- Molecule 1: integrase



- Molecule 1: integrase

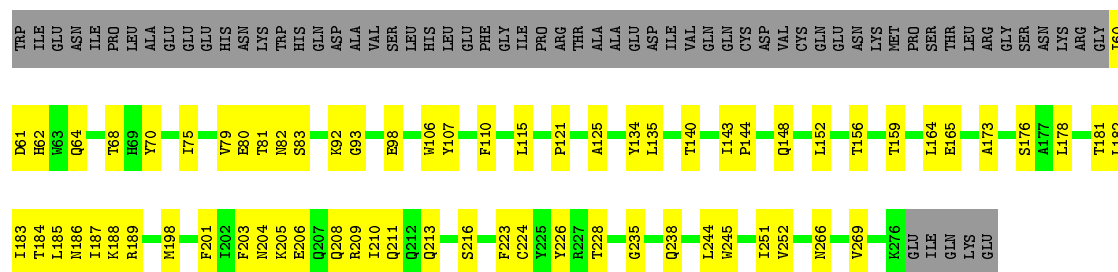


- Molecule 1: integrase



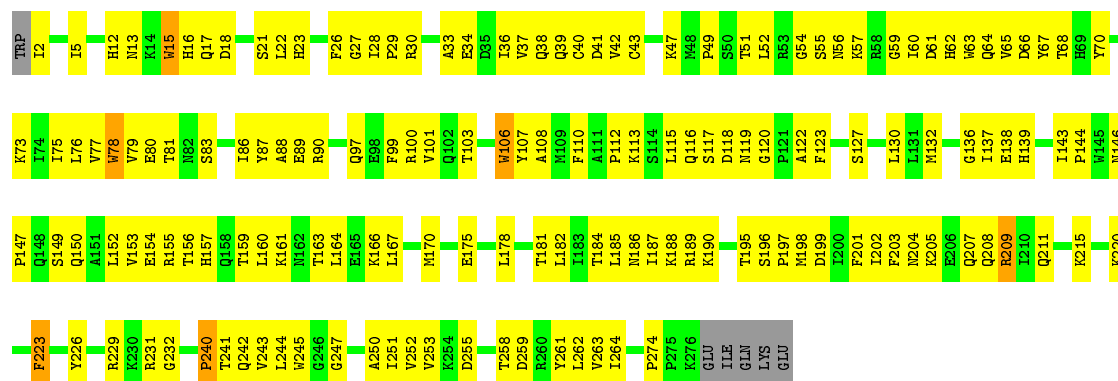
- Molecule 1: integrase

Chain H: 



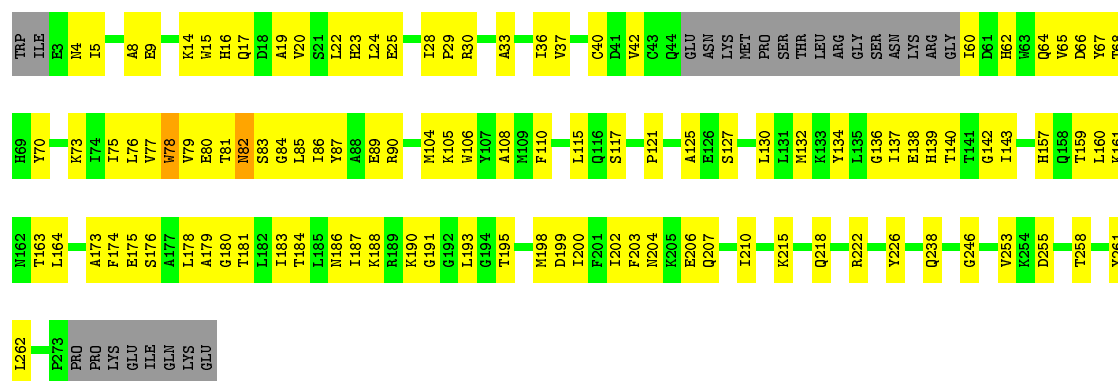
- Molecule 1: integrase

Chain I: 



- Molecule 1: integrase

Chain J: 

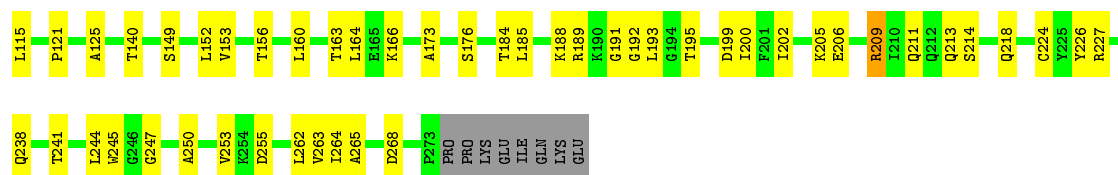


- Molecule 1: integrase

Chain K: 







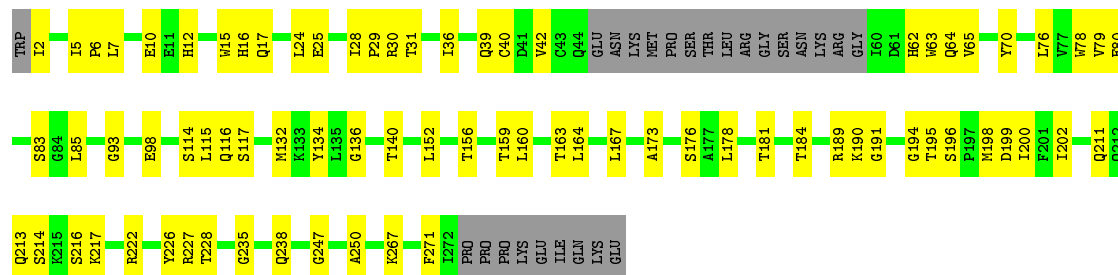
- Molecule 1: integrase

Chain L: 59% 16% 23%



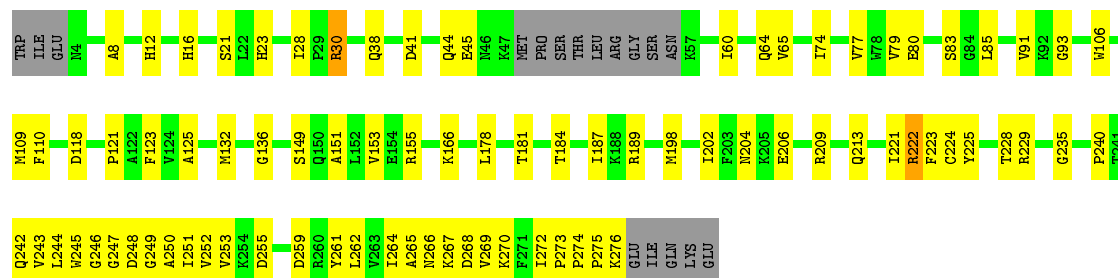
- Molecule 1: integrase

Chain M: 64% 27% 9%



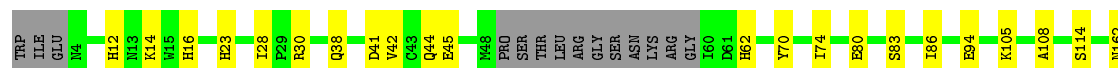
- Molecule 1: integrase

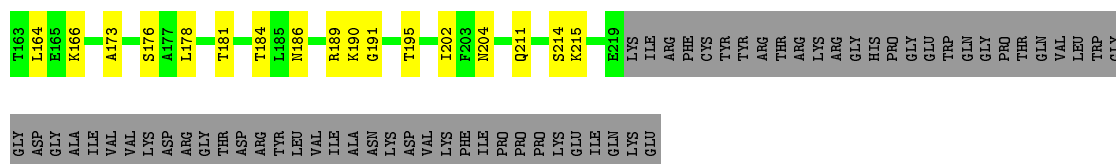
Chain N: 64% 29% 6%



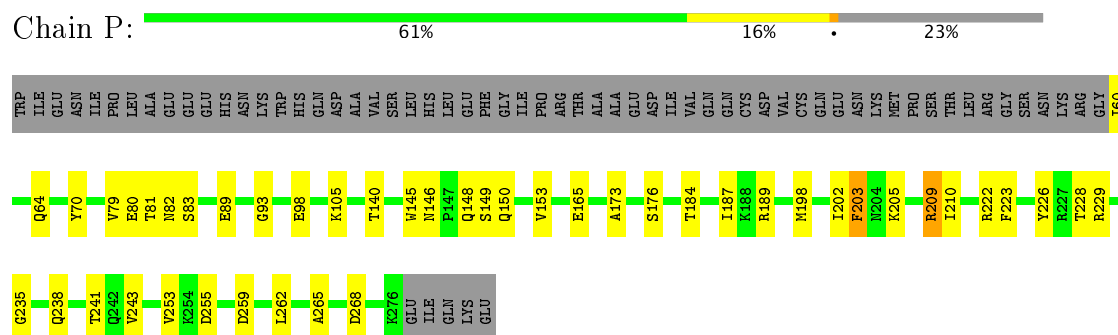
- Molecule 1: integrase

Chain O: 59% 14% 27%





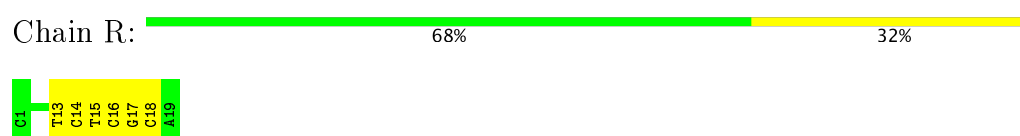
- Molecule 1: integrase



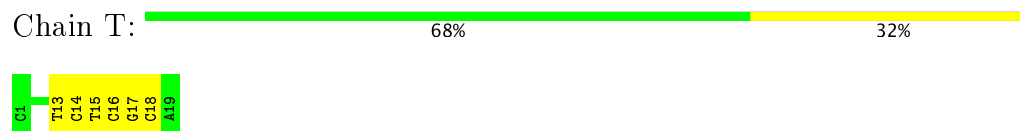
- Molecule 2: vDNA, non-transferred strand



- Molecule 2: vDNA, non-transferred strand



- Molecule 3: vDNA, transferred strand



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C2	Depositor
Number of particles used	94283	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 ( $e^-/\text{\AA}^2$ )	1	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

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	0.91	6/2254 (0.3%)	0.94	3/3053 (0.1%)
1	B	0.75	4/2124 (0.2%)	0.82	3/2874 (0.1%)
1	C	0.54	0/2174	0.70	0/2942
1	D	0.65	3/1772 (0.2%)	0.73	0/2401
1	E	0.69	2/2140 (0.1%)	0.76	0/2896
1	F	0.78	3/2199 (0.1%)	0.81	2/2975 (0.1%)
1	G	0.55	1/1693 (0.1%)	0.67	0/2290
1	H	0.59	0/1805	0.72	0/2441
1	I	0.89	5/2254 (0.2%)	0.93	2/3053 (0.1%)
1	J	0.73	3/2124 (0.1%)	0.84	2/2874 (0.1%)
1	K	0.58	2/2174 (0.1%)	0.69	1/2942 (0.0%)
1	L	0.62	1/1772 (0.1%)	0.68	0/2401
1	M	0.62	0/2124	0.73	0/2873
1	N	0.71	2/2182 (0.1%)	0.75	1/2953 (0.0%)
1	O	0.48	0/1693	0.62	0/2290
1	P	0.63	2/1805 (0.1%)	0.75	1/2441 (0.0%)
2	Q	2.08	14/479 (2.9%)	1.43	7/738 (0.9%)
2	S	2.08	13/479 (2.7%)	1.43	7/738 (0.9%)
3	R	2.08	11/436 (2.5%)	1.29	4/670 (0.6%)
3	T	2.08	10/436 (2.3%)	1.29	4/670 (0.6%)
All	All	0.83	82/34119 (0.2%)	0.82	37/46515 (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	B	0	1
1	L	0	1
All	All	0	2

The worst 5 of 82 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	T	15	DT	C5-C7	-9.45	1.44	1.50
3	R	15	DT	C5-C7	-9.30	1.44	1.50
3	T	13	DT	C5-C7	-7.36	1.45	1.50
1	B	78	TRP	CB-CG	-7.30	1.37	1.50
1	J	78	TRP	CB-CG	-7.28	1.37	1.50

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	P	203	PHE	N-CA-CB	-7.01	97.98	110.60
3	T	15	DT	N3-C4-O4	6.96	124.08	119.90
2	Q	7	DA	O4'-C4'-C3'	-6.93	101.73	104.50
3	R	15	DT	N3-C4-O4	6.92	124.06	119.90
2	S	7	DA	O4'-C4'-C3'	-6.81	101.78	104.50

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	62	HIS	Peptide
1	L	239	GLY	Peptide

## 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	2201	0	2164	346	0
1	B	2075	0	2046	73	0
1	C	2123	0	2095	40	0
1	D	1729	0	1708	114	0
1	E	2090	0	2063	124	0
1	F	2148	0	2119	137	0
1	G	1657	0	1626	53	0
1	H	1761	0	1764	106	0
1	I	2201	0	2165	334	0
1	J	2075	0	2046	153	0
1	K	2123	0	2095	57	0
1	L	1729	0	1708	87	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	M	2076	0	2050	86	0
1	N	2131	0	2099	105	0
1	O	1657	0	1626	28	0
1	P	1761	0	1764	63	0
2	Q	428	0	237	14	0
2	S	428	0	237	44	0
3	R	389	0	212	0	0
3	T	389	0	212	0	0
All	All	33171	0	32036	1669	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 26.

The worst 5 of 1669 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:L:227:ARG:CD	1:L:234:PRO:HB2	1.25	1.58
1:L:227:ARG:CD	1:L:234:PRO:CB	1.91	1.49
1:D:227:ARG:CD	1:D:234:PRO:HB2	1.45	1.45
1:E:223:PHE:CZ	1:E:272:ILE:CG2	1.99	1.45
1:L:227:ARG:HD3	1:L:234:PRO:CB	1.46	1.45

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	273/281 (97%)	248 (91%)	23 (8%)	2 (1%)	25	68
1	B	252/281 (90%)	238 (94%)	14 (6%)	0	100	100
1	C	257/281 (92%)	251 (98%)	6 (2%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	214/281 (76%)	207 (97%)	5 (2%)	2 (1%)	20	63
1	E	253/281 (90%)	249 (98%)	4 (2%)	0	100	100
1	F	262/281 (93%)	257 (98%)	5 (2%)	0	100	100
1	G	201/281 (72%)	198 (98%)	3 (2%)	0	100	100
1	H	215/281 (76%)	213 (99%)	2 (1%)	0	100	100
1	I	273/281 (97%)	257 (94%)	14 (5%)	2 (1%)	25	68
1	J	252/281 (90%)	245 (97%)	7 (3%)	0	100	100
1	K	257/281 (92%)	254 (99%)	3 (1%)	0	100	100
1	L	214/281 (76%)	203 (95%)	9 (4%)	2 (1%)	20	63
1	M	252/281 (90%)	244 (97%)	8 (3%)	0	100	100
1	N	260/281 (92%)	257 (99%)	3 (1%)	0	100	100
1	O	201/281 (72%)	200 (100%)	1 (0%)	0	100	100
1	P	215/281 (76%)	211 (98%)	4 (2%)	0	100	100
All	All	3851/4496 (86%)	3732 (97%)	111 (3%)	8 (0%)	54	85

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	233	HIS
1	L	233	HIS
1	I	119	ASN
1	I	240	PRO
1	A	4	ASN

### 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	235/246 (96%)	234 (100%)	1 (0%)	93	95
1	B	223/246 (91%)	223 (100%)	0	100	100
1	C	228/246 (93%)	228 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	183/246 (74%)	183 (100%)	0	100	100
1	E	224/246 (91%)	223 (100%)	1 (0%)	93	95
1	F	230/246 (94%)	230 (100%)	0	100	100
1	G	181/246 (74%)	181 (100%)	0	100	100
1	H	189/246 (77%)	189 (100%)	0	100	100
1	I	235/246 (96%)	234 (100%)	1 (0%)	93	95
1	J	223/246 (91%)	222 (100%)	1 (0%)	93	95
1	K	228/246 (93%)	228 (100%)	0	100	100
1	L	183/246 (74%)	183 (100%)	0	100	100
1	M	223/246 (91%)	223 (100%)	0	100	100
1	N	228/246 (93%)	228 (100%)	0	100	100
1	O	181/246 (74%)	181 (100%)	0	100	100
1	P	189/246 (77%)	189 (100%)	0	100	100
All	All	3383/3936 (86%)	3379 (100%)	4 (0%)	95	97

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	210	ILE
1	E	168	ILE
1	I	223	PHE
1	J	82	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	I	150	GLN
1	M	16	HIS
1	N	213	GLN

### 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](#)

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.