



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 04:43 AM UTC

PDB ID : 1KSX / pdb_00001ksx
Title : Crystal Structures of Two Intermediates in the Assembly of the Papillomavirus Replication Initiation Complex
Authors : Enemark, E.J.; Stenlund, A.; Joshua-Tor, L.
Deposited on : 2002-01-14
Resolution : 3.20 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0
Xtriage (Phenix) : **NOT EXECUTED**
EDS : **NOT EXECUTED**
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

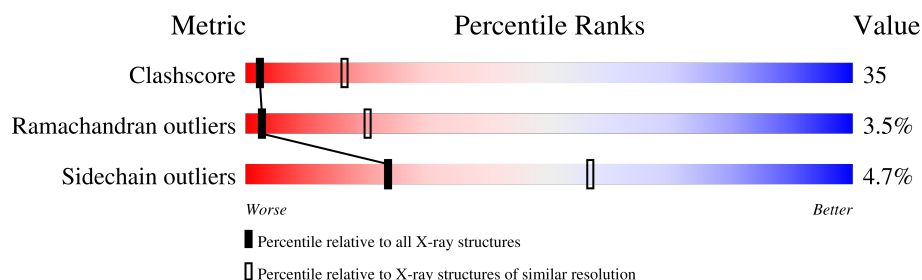
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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))
Clashscore	190562	1573 (3.20-3.20)
Ramachandran outliers	187476	1548 (3.20-3.20)
Sidechain outliers	187428	1547 (3.20-3.20)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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 $\leq 5\%$.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	21	33% 62% 5%
1	G	21	33% 62% 5%
1	K	21	33% 62% 5%
1	O	21	33% 62% 5%
2	A	148	48% 46% ..
2	B	148	39% 51% 7% ..
2	E	148	50% 44% ..
2	F	148	41% 50% 7% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	148	48%46%
2	J	148	39%51%7%
2	M	148	49%45%
2	N	148	39%52%7%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11044 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 DNA chain called E1 Recognition Sequence.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	21	Total	C	N	O	P	0	0	0
			428	208	77	123	20			
1	G	21	Total	C	N	O	P	0	0	0
			428	208	77	123	20			
1	K	21	Total	C	N	O	P	0	0	0
			428	208	77	123	20			
1	O	21	Total	C	N	O	P	0	0	0
			428	208	77	123	20			

- Molecule 2 is a protein called REPLICATION PROTEIN E1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	B	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	E	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	F	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	I	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	J	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	M	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			
2	N	145	Total	C	N	O	S	7	0	0
			1164	754	200	201	9			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	156	GLY	-	cloning artifact	UNP P03116

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	157	SER	-	cloning artifact	UNP P03116
A	158	ARG	-	cloning artifact	UNP P03116
B	156	GLY	-	cloning artifact	UNP P03116
B	157	SER	-	cloning artifact	UNP P03116
B	158	ARG	-	cloning artifact	UNP P03116
E	156	GLY	-	cloning artifact	UNP P03116
E	157	SER	-	cloning artifact	UNP P03116
E	158	ARG	-	cloning artifact	UNP P03116
F	156	GLY	-	cloning artifact	UNP P03116
F	157	SER	-	cloning artifact	UNP P03116
F	158	ARG	-	cloning artifact	UNP P03116
I	156	GLY	-	cloning artifact	UNP P03116
I	157	SER	-	cloning artifact	UNP P03116
I	158	ARG	-	cloning artifact	UNP P03116
J	156	GLY	-	cloning artifact	UNP P03116
J	157	SER	-	cloning artifact	UNP P03116
J	158	ARG	-	cloning artifact	UNP P03116
M	156	GLY	-	cloning artifact	UNP P03116
M	157	SER	-	cloning artifact	UNP P03116
M	158	ARG	-	cloning artifact	UNP P03116
N	156	GLY	-	cloning artifact	UNP P03116
N	157	SER	-	cloning artifact	UNP P03116
N	158	ARG	-	cloning artifact	UNP P03116

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	O	1	Total O 1 1	0	0
3	A	2	Total O 2 2	0	0
3	B	3	Total O 3 3	0	0
3	E	2	Total O 2 2	0	0
3	F	3	Total O 3 3	0	0
3	I	1	Total O 1 1	0	0
3	J	3	Total O 3 3	0	0
3	M	2	Total O 2 2	0	0

Continued on next page...

Continued from previous page...

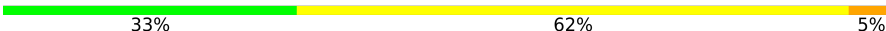
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	N	3	Total	O	0	0
			3	3		

3 Residue-property plots

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

Note EDS was not executed.

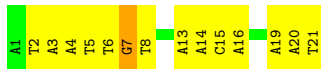
• Molecule 1: E1 Recognition Sequence

Chain C: 



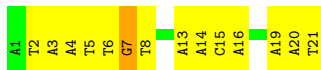
• Molecule 1: E1 Recognition Sequence

Chain G: 

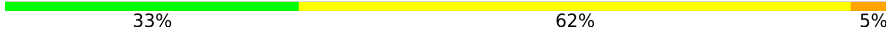


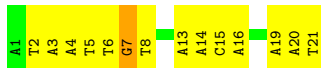
• Molecule 1: E1 Recognition Sequence

Chain K: 



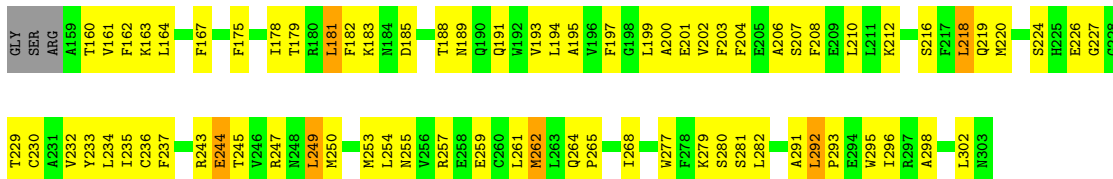
• Molecule 1: E1 Recognition Sequence

Chain O: 



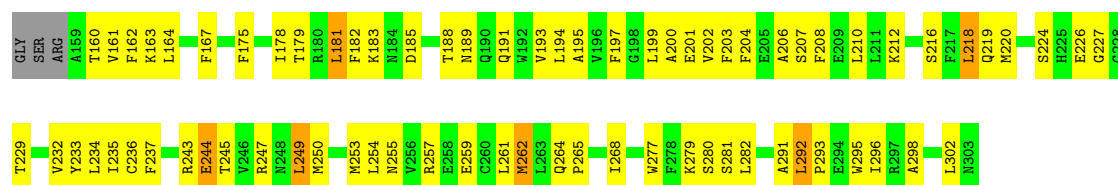
• Molecule 2: REPLICATION PROTEIN E1

Chain A: 

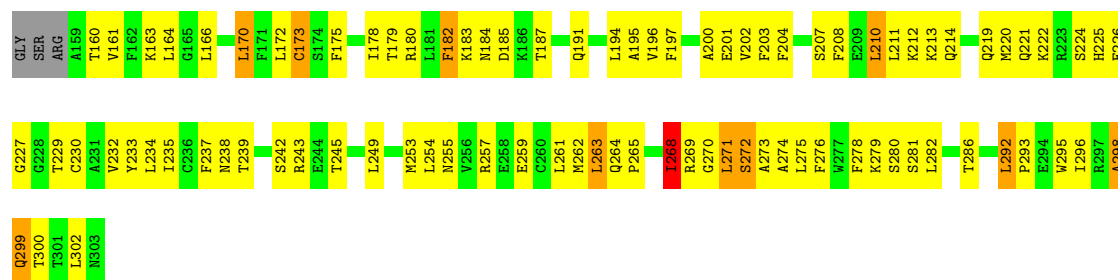




• Molecule 2: REPLICATION PROTEIN E1



• Molecule 2: REPLICATION PROTEIN E1



4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	84.25Å 103.65Å 125.00Å 90.00° 99.53° 90.00°	Depositor
Resolution (Å)	42.62 – 3.20	Depositor
% Data completeness (in resolution range)	90.5 (42.62-3.20)	Depositor
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.263 , 0.285	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	11044	wwPDB-VP
Average B, all atoms (Å ²)	59.0	wwPDB-VP

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	$\# Z > 5$	RMSZ	$\# Z > 5$
1	C	0.44	0/480	0.90	1/739 (0.1%)
1	G	0.44	0/480	0.90	1/739 (0.1%)
1	K	0.44	0/480	0.90	1/739 (0.1%)
1	O	0.44	0/480	0.90	1/739 (0.1%)
2	A	0.47	0/1190	0.92	3/1602 (0.2%)
2	B	0.46	0/1190	1.04	6/1602 (0.4%)
2	E	0.47	0/1190	0.92	3/1602 (0.2%)
2	F	0.46	0/1190	1.04	6/1602 (0.4%)
2	I	0.47	0/1190	0.92	3/1602 (0.2%)
2	J	0.46	0/1190	1.04	6/1602 (0.4%)
2	M	0.47	0/1190	0.92	3/1602 (0.2%)
2	N	0.46	0/1190	1.04	6/1602 (0.4%)
All	All	0.46	0/11440	0.97	40/15772 (0.3%)

There are no bond length outliers.

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	J	292	LEU	CA-C-N	9.28	129.36	119.90
2	J	292	LEU	C-N-CA	9.28	129.36	119.90
2	B	292	LEU	CA-C-N	9.26	129.34	119.90
2	B	292	LEU	C-N-CA	9.26	129.34	119.90
2	F	292	LEU	CA-C-N	9.24	129.33	119.90

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	C	428	0	241	37	0
1	G	428	0	241	59	0
1	K	428	0	241	39	0
1	O	428	0	241	41	0
2	A	1164	0	1189	70	0
2	B	1164	0	1189	87	0
2	E	1164	0	1189	68	0
2	F	1164	0	1189	86	0
2	I	1164	0	1189	92	0
2	J	1164	0	1189	88	0
2	M	1164	0	1189	70	0
2	N	1164	0	1189	92	0
3	A	2	0	0	0	0
3	B	3	0	0	0	0
3	E	2	0	0	0	0
3	F	3	0	0	0	0
3	I	1	0	0	0	0
3	J	3	0	0	0	0
3	M	2	0	0	0	0
3	N	3	0	0	0	0
3	O	1	0	0	0	0
All	All	11044	0	10476	757	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 35.

The worst 5 of 757 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:G:21:DT:H71	2:I:162:PHE:CE2	1.40	1.56
1:G:21:DT:C7	2:I:162:PHE:CZ	2.07	1.38
1:G:21:DT:H73	2:I:162:PHE:CZ	1.64	1.30
1:G:21:DT:C7	2:I:162:PHE:CE2	2.16	1.26
1:O:3:DA:H2''	1:O:4:DA:C8	1.95	1.02

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	
2	A	143/148 (97%)	119 (83%)	21 (15%)	3 (2%)	5	31
2	B	143/148 (97%)	121 (85%)	15 (10%)	7 (5%)	1	13
2	E	143/148 (97%)	119 (83%)	21 (15%)	3 (2%)	5	31
2	F	143/148 (97%)	121 (85%)	15 (10%)	7 (5%)	1	13
2	I	143/148 (97%)	119 (83%)	21 (15%)	3 (2%)	5	31
2	J	143/148 (97%)	121 (85%)	15 (10%)	7 (5%)	1	13
2	M	143/148 (97%)	119 (83%)	21 (15%)	3 (2%)	5	31
2	N	143/148 (97%)	121 (85%)	15 (10%)	7 (5%)	1	13
All	All	1144/1184 (97%)	960 (84%)	144 (13%)	40 (4%)	3	20

5 of 40 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	271	LEU
2	B	299	GLN
2	F	271	LEU
2	F	299	GLN
2	J	271	LEU

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 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	
2	A	127/129 (98%)	120 (94%)	7 (6%)	19	52

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	127/129 (98%)	122 (96%)	5 (4%)	28	62
2	E	127/129 (98%)	120 (94%)	7 (6%)	19	52
2	F	127/129 (98%)	122 (96%)	5 (4%)	28	62
2	I	127/129 (98%)	120 (94%)	7 (6%)	19	52
2	J	127/129 (98%)	122 (96%)	5 (4%)	28	62
2	M	127/129 (98%)	120 (94%)	7 (6%)	19	52
2	N	127/129 (98%)	122 (96%)	5 (4%)	28	62
All	All	1016/1032 (98%)	968 (95%)	48 (5%)	23	57

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	249	LEU
2	J	268	ILE
2	I	262	MET
2	J	210	LEU
2	M	202	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
2	J	191	GLN
2	J	238	ASN
2	N	238	ASN
2	J	225	HIS
2	J	255	ASN

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 [i](#)

There are no oligosaccharides 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.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.