wwPDB X-ray Structure Validation Summary Report (i)

Aug 6, 2023 – 02:05 AM EDT

PDB ID	:	1KSY
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	:	3.05 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 (i) 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 (i)) were used in the production of this report:

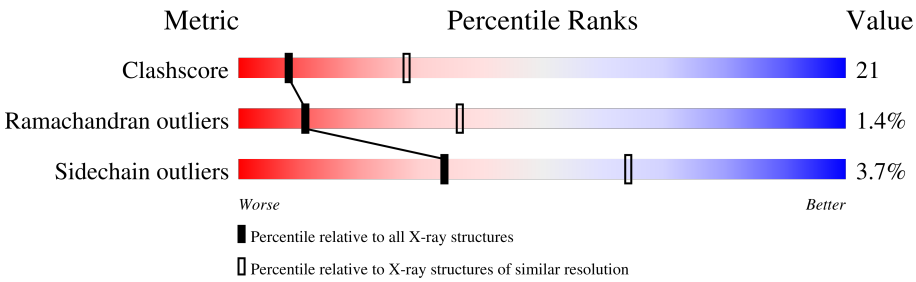
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.05 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	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)

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 <=5%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	D	21	57% 43%
1	E	21	52% 48%
2	F	21	52% 48%
3	A	154	56% 35% • • 6%
3	B	154	57% 36% • •
3	C	154	55% 36% • • 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4795 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, Strand 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	D	21	427	208	74	125	20	0	0	0
1	E	21	427	208	74	125	20	0	0	0

- Molecule 2 is a DNA chain called E1 Recognition Sequence, Strand 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	F	21	428	208	77	123	20	0	0	0

- Molecule 3 is a protein called REPLICATION PROTEIN E1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	145	1163	754	200	200	9	6	0	0
3	B	148	1184	765	206	204	9	6	0	0
3	C	145	1163	754	200	200	9	6	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	156	GLY	-	cloning artifact	UNP P03116
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
C	156	GLY	-	cloning artifact	UNP P03116
C	157	SER	-	cloning artifact	UNP P03116

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	158	ARG	-	cloning artifact	UNP P03116

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O 1 1	0	0
4	B	1	Total O 1 1	0	0
4	C	1	Total O 1 1	0	0

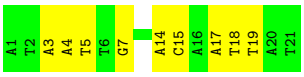
3 Residue-property plots (i)

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.

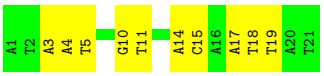
• Molecule 1: E1 Recognition Sequence, Strand 1

Chain D:



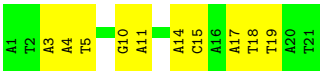
• Molecule 1: E1 Recognition Sequence, Strand 1

Chain E:



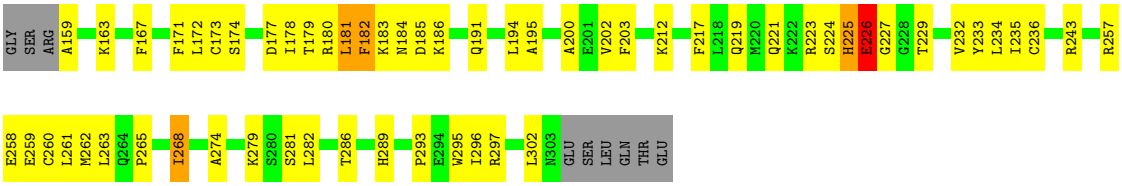
• Molecule 2: E1 Recognition Sequence, Strand 2

Chain F:



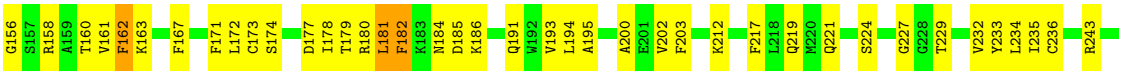
• Molecule 3: REPLICATION PROTEIN E1

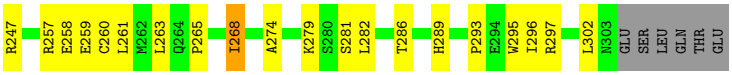
Chain A:



• Molecule 3: REPLICATION PROTEIN E1

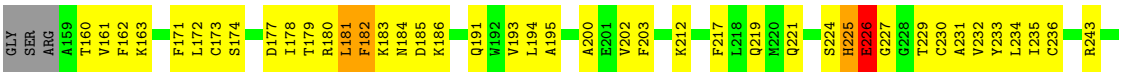
Chain B:

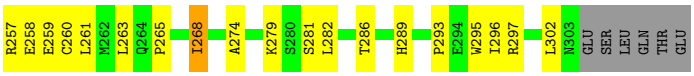




• Molecule 3: REPLICATION PROTEIN E1

Chain C:





4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	149.11Å 110.60Å 75.08Å 90.00° 116.85° 90.00°	Depositor
Resolution (Å)	35.53 – 3.05	Depositor
% Data completeness (in resolution range)	91.8 (35.53-3.05)	Depositor
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.277 , 0.316	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4795	wwPDB-VP
Average B, all atoms (Å^2)	68.0	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	D	0.30	0/478	0.75	0/736
1	E	0.26	0/478	0.72	0/736
2	F	0.29	0/480	0.75	0/739
3	A	0.29	0/1189	0.51	0/1602
3	B	0.29	0/1210	0.51	0/1629
3	C	0.29	0/1189	0.51	0/1602
All	All	0.29	0/5024	0.59	0/7044

There are no bond length outliers.

There are no bond angle outliers.

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	D	427	0	242	8	0
1	E	427	0	242	7	0
2	F	428	0	241	9	0
3	A	1163	0	1189	59	0
3	B	1184	0	1210	55	0
3	C	1163	0	1189	58	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
All	All	4795	0	4313	191	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 21.

The worst 5 of 191 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:4:DA:H2"	1:D:5:DT:H5'	1.59	0.85
2:F:4:DA:H2"	2:F:5:DT:H5'	1.58	0.83
3:A:268:ILE:HA	3:A:295:TRP:CE3	2.14	0.82
3:C:268:ILE:HA	3:C:295:TRP:CE3	2.15	0.82
1:E:4:DA:H2"	1:E:5:DT:H5'	1.59	0.82

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	
3	A	143/154 (93%)	119 (83%)	22 (15%)	2 (1%)	11	36
3	B	146/154 (95%)	126 (86%)	18 (12%)	2 (1%)	11	36
3	C	143/154 (93%)	125 (87%)	16 (11%)	2 (1%)	11	36
All	All	432/462 (94%)	370 (86%)	56 (13%)	6 (1%)	11	36

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	226	GLU
3	A	226	GLU
3	A	274	ALA
3	B	274	ALA
3	C	274	ALA

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	
3	A	127/135 (94%)	122 (96%)	5 (4%)	32	63
3	B	129/135 (96%)	125 (97%)	4 (3%)	40	68
3	C	127/135 (94%)	122 (96%)	5 (4%)	32	63
All	All	383/405 (95%)	369 (96%)	14 (4%)	34	64

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	182	PHE
3	B	268	ILE
3	C	268	ILE
3	C	225	HIS
3	C	226	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	238	ASN
3	C	225	HIS
3	B	238	ASN
3	B	219	GLN
3	C	219	GLN

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 monosaccharides 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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.