Full wwPDB X-ray Structure Validation Report (i)

Jan 30, 2024 – 07:48 PM EST

PDB ID	:	1JJ4
Title	:	Human papillomavirus type 18 E2 DNA-binding domain bound to its DNA target
Authors	:	Kim, S.-S.; Tam, J.K.; Wang, A.F.; Hegde, R.S.
Deposited on	:	2001-07-03
Resolution	:	2.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

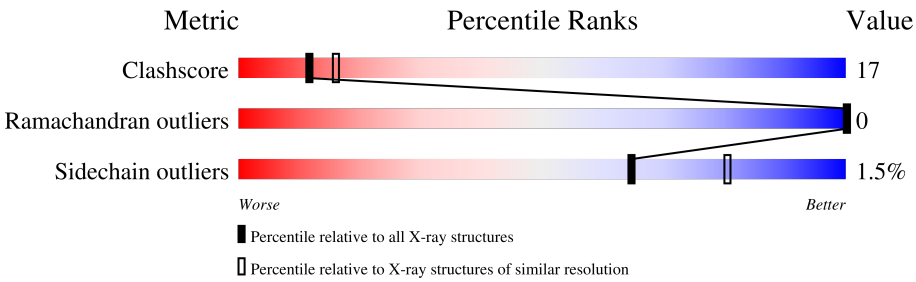
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.40 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	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)

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	C	16	44% 56%
1	D	16	56% 44%
2	A	83	54% 35% 11%
2	B	83	66% 24% • 8%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2035 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 E2 binding site DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	16	325	156	60	94	15	0	0	0
1	D	16	325	156	60	94	15	0	0	0

- Molecule 2 is a protein called Regulatory protein E2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	74	592	375	109	105	3	0	0	0
2	B	76	603	384	109	107	3	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	283	GLY	-	cloning artifact	UNP P06790
A	284	SER	-	cloning artifact	UNP P06790
A	285	HIS	-	cloning artifact	UNP P06790
A	286	MET	-	cloning artifact	UNP P06790
B	283	GLY	-	cloning artifact	UNP P06790
B	284	SER	-	cloning artifact	UNP P06790
B	285	HIS	-	cloning artifact	UNP P06790
B	286	MET	-	cloning artifact	UNP P06790

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	C	49	49	49	0	0
3	D	40	40	40	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	51	51	51	0	0
			Total	O		
3	B	50	50	50	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: E2 binding site DNA

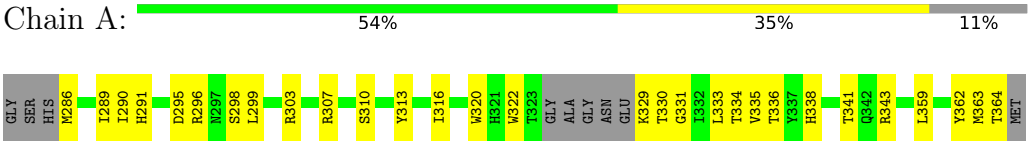
Chain C:

• Molecule 1: E2 binding site DNA

Chain D:

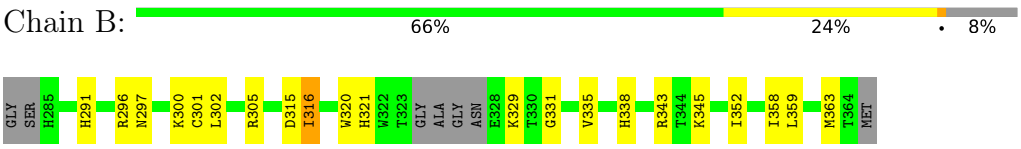
• Molecule 2: Regulatory protein E2

Chain A:



• Molecule 2: Regulatory protein E2

Chain B:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	69.88Å 82.19Å 96.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.65 – 2.40	Depositor
% Data completeness (in resolution range)	83.3 (25.65-2.40)	Depositor
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.210 , 0.287	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2035	wwPDB-VP
Average B, all atoms (Å^2)	25.0	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	0.35	0/364	0.69	0/560
1	D	0.33	0/364	0.73	0/560
2	A	0.34	0/605	0.58	0/818
2	B	0.33	0/616	0.59	0/835
All	All	0.34	0/1949	0.64	0/2773

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	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	615	DT	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	C	325	0	182	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	325	0	182	9	0
2	A	592	0	583	30	0
2	B	603	0	588	22	0
3	A	51	0	0	5	0
3	B	50	0	0	7	0
3	C	49	0	0	1	0
3	D	40	0	0	2	0
All	All	2035	0	1535	59	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 17.

All (59) 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:904:DA:H2"	1:D:905:DC:H5'	1.53	0.90
2:A:363:MET:HG2	2:A:364:THR:HG23	1.57	0.86
1:C:605:DC:H41	2:B:297:ASN:HD21	1.24	0.85
2:B:329:LYS:HA	3:B:211:HOH:O	1.76	0.84
2:B:316:ILE:HG23	2:B:335:VAL:HG22	1.66	0.77
2:B:302:LEU:HA	2:B:305:ARG:NH1	2.05	0.71
2:A:289:ILE:N	2:A:343:ARG:HH21	1.91	0.68
2:A:290:ILE:HG12	2:A:335:VAL:HB	1.77	0.66
2:A:289:ILE:HG22	2:A:336:THR:HA	1.77	0.65
2:A:310:SER:HA	2:A:313:TYR:CE2	2.33	0.63
1:D:912:DC:H2"	1:D:913:DG:C8	2.32	0.63
2:B:321:HIS:HB2	3:B:160:HOH:O	1.99	0.62
2:A:295:ASP:HB3	2:A:298:SER:HB3	1.81	0.62
2:A:341:THR:HG23	3:A:8:HOH:O	1.99	0.62
2:A:286:MET:HB2	2:A:362:TYR:HB3	1.82	0.62
1:D:904:DA:H2"	1:D:905:DC:C5'	2.28	0.61
2:B:301:CYS:HA	3:B:29:HOH:O	2.01	0.61
2:A:364:THR:HG22	2:B:291:HIS:CE1	2.38	0.59
2:B:358:ILE:O	2:B:359:LEU:HD23	2.03	0.58
1:D:912:DC:H2'	2:B:300:LYS:HD2	1.86	0.58
1:C:612:DC:H2"	1:C:613:DG:C8	2.41	0.56
2:A:289:ILE:H	2:A:343:ARG:HH21	1.52	0.56
1:D:904:DA:H1'	1:D:905:DC:H5"	1.87	0.56
2:A:289:ILE:C	2:A:289:ILE:HD12	2.28	0.54
2:B:338:HIS:HB2	3:B:213:HOH:O	2.08	0.54
2:A:364:THR:HG21	3:A:191:HOH:O	2.07	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:907:DG:H2"	3:D:151:HOH:O	2.08	0.52
2:A:313:TYR:CE1	2:A:316:ILE:HG13	2.45	0.52
2:A:295:ASP:HB3	2:A:298:SER:CB	2.40	0.52
1:D:906:DC:H2"	1:D:907:DG:C8	2.44	0.51
2:A:364:THR:HG22	2:B:291:HIS:HE1	1.75	0.51
2:A:296:ARG:HG2	2:A:331:GLY:HA3	1.93	0.51
1:C:616:DT:H73	3:C:187:HOH:O	2.11	0.50
1:C:611:DT:C2	1:C:612:DC:C5	3.00	0.49
2:B:345:LYS:HD3	3:B:177:HOH:O	2.13	0.49
2:B:343:ARG:HD2	3:B:56:HOH:O	2.13	0.48
2:A:320:TRP:HB3	2:B:320:TRP:HB3	1.95	0.48
2:A:299:LEU:HD13	2:A:333:LEU:HG	1.96	0.47
2:B:296:ARG:HG3	2:B:331:GLY:HA3	1.97	0.47
2:A:338:HIS:HE1	3:A:186:HOH:O	1.98	0.46
1:D:916:DT:H71	3:D:46:HOH:O	2.16	0.45
2:A:363:MET:HA	2:B:363:MET:HA	1.99	0.45
2:A:303:ARG:O	2:A:307:ARG:HG3	2.17	0.44
2:A:363:MET:HG2	2:A:364:THR:N	2.31	0.44
2:A:322:TRP:HB3	2:B:315:ASP:OD1	2.18	0.44
1:C:602:DC:H2'	2:B:305:ARG:NH2	2.32	0.44
2:A:289:ILE:HG22	2:A:336:THR:HG22	2.00	0.43
1:C:607:DG:H2"	1:C:608:DA:OP2	2.19	0.43
2:A:290:ILE:HA	2:A:359:LEU:O	2.19	0.42
2:B:315:ASP:HB3	3:B:234:HOH:O	2.18	0.42
2:B:302:LEU:HD12	2:B:305:ARG:NH1	2.35	0.42
2:A:299:LEU:HB3	2:A:333:LEU:HG	2.02	0.42
2:A:291:HIS:CD2	2:A:334:THR:HG23	2.55	0.41
2:A:329:LYS:HA	3:A:195:HOH:O	2.20	0.41
2:B:352:ILE:HG13	2:B:358:ILE:HD11	2.02	0.41
1:D:904:DA:C2'	1:D:905:DC:C5'	2.97	0.41
2:A:329:LYS:N	3:A:204:HOH:O	2.53	0.41
1:C:605:DC:N4	2:B:297:ASN:HD21	2.06	0.41
2:A:295:ASP:O	2:A:299:LEU:HG	2.21	0.40

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	70/83 (84%)	66 (94%)	4 (6%)	0	100	100
2	B	72/83 (87%)	69 (96%)	3 (4%)	0	100	100
All	All	142/166 (86%)	135 (95%)	7 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	65/75 (87%)	64 (98%)	1 (2%)	65	80
2	B	65/75 (87%)	64 (98%)	1 (2%)	65	80
All	All	130/150 (87%)	128 (98%)	2 (2%)	65	80

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	330	THR
2	B	316	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
2	A	291	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	A	312	HIS
2	A	338	HIS
2	A	348	ASN
2	B	291	HIS
2	B	297	ASN

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.