



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 28, 2024 – 03:07 PM EST

PDB ID : 1DTO  
Title : CRYSTAL STRUCTURE OF THE COMPLETE TRANSACTIVATION DOMAIN OF E2 PROTEIN FROM THE HUMAN PAPILLOMAVIRUS TYPE 16  
Authors : Antson, A.A.; Burns, J.E.; Moroz, O.V.; Scott, D.J.; Sanders, C.M.; Bronstein, I.B.; Dodson, G.G.; Wilson, K.S.; Maitland, N.  
Deposited on : 2000-01-13  
Resolution : 1.90 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](mailto: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 : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

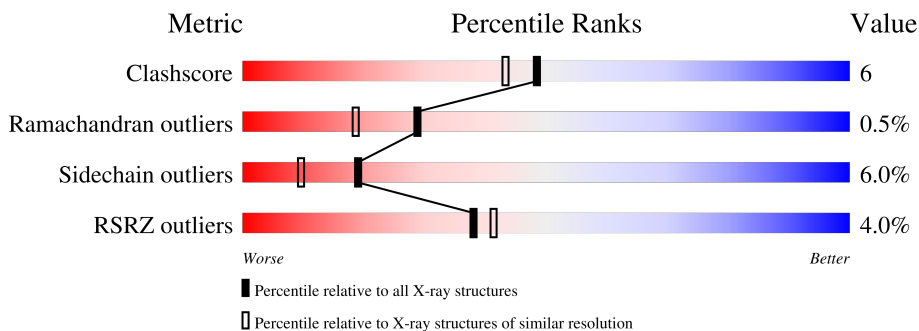
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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  $\geq 3$ , 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	221	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1861 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 protein called REGULATORY PROTEIN E2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	200	1650	1052	274	314	10	30	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	expression tag	UNP P03120
A	-18	GLY	-	expression tag	UNP P03120
A	-17	SER	-	expression tag	UNP P03120
A	-16	SER	-	expression tag	UNP P03120
A	-15	HIS	-	expression tag	UNP P03120
A	-14	HIS	-	expression tag	UNP P03120
A	-13	HIS	-	expression tag	UNP P03120
A	-12	HIS	-	expression tag	UNP P03120
A	-11	HIS	-	expression tag	UNP P03120
A	-10	HIS	-	expression tag	UNP P03120
A	-9	SER	-	expression tag	UNP P03120
A	-8	SER	-	expression tag	UNP P03120
A	-7	GLY	-	expression tag	UNP P03120
A	-6	LEU	-	expression tag	UNP P03120
A	-5	VAL	-	expression tag	UNP P03120
A	-4	PRO	-	expression tag	UNP P03120
A	-3	ARG	-	expression tag	UNP P03120
A	-2	GLY	-	expression tag	UNP P03120
A	-1	SER	-	expression tag	UNP P03120
A	0	HIS	-	expression tag	UNP P03120

- Molecule 2 is water.

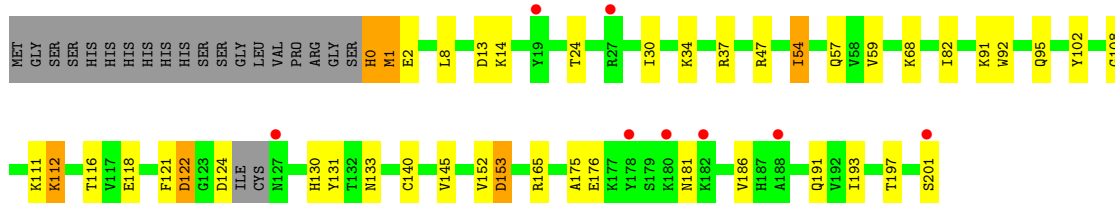
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	211	Total 211	O 211	0	0

### 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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). 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: REGULATORY PROTEIN E2

Chain A: 



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.68Å 54.68Å 155.73Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 1.90 27.34 – 1.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (30.00-1.90) 98.6 (27.34-1.90)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.68 (at 1.91Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.232 , 0.305 0.219 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	33.2	Xtrriage
Anisotropy	0.542	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 54.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.039 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1861	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.68% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 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	A	0.76	3/1691 (0.2%)	1.27	19/2294 (0.8%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	181	ASN	CA-CB	-8.89	1.30	1.53
1	A	13	ASP	CB-CG	-6.39	1.38	1.51
1	A	153	ASP	CA-CB	6.03	1.67	1.53

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	47	ARG	CD-NE-CZ	8.38	135.34	123.60
1	A	124	ASP	CB-CG-OD2	-7.67	111.40	118.30
1	A	1	MET	CA-CB-CG	7.62	126.25	113.30
1	A	153	ASP	CA-CB-CG	-7.41	97.09	113.40
1	A	153	ASP	CB-CG-OD2	7.00	124.60	118.30
1	A	131	TYR	CB-CG-CD1	-6.85	116.89	121.00
1	A	153	ASP	CB-CG-OD1	-6.71	112.26	118.30
1	A	47	ARG	NE-CZ-NH1	6.57	123.59	120.30
1	A	176	GLU	CG-CD-OE2	-6.34	105.62	118.30
1	A	176	GLU	CG-CD-OE1	6.28	130.85	118.30
1	A	47	ARG	NE-CZ-NH2	-6.04	117.28	120.30
1	A	131	TYR	CB-CG-CD2	6.02	124.61	121.00
1	A	37	ARG	NE-CZ-NH1	5.87	123.24	120.30
1	A	112	LYS	CA-CB-CG	5.87	126.32	113.40
1	A	152	VAL	C-N-CA	5.71	135.97	121.70
1	A	152	VAL	CG1-CB-CG2	-5.61	101.92	110.90
1	A	0	HIS	CA-CB-CG	-5.60	104.08	113.60
1	A	165	ARG	NE-CZ-NH2	-5.38	117.61	120.30
1	A	37	ARG	NE-CZ-NH2	-5.26	117.67	120.30

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	1650	0	1581	19	0
2	A	211	0	0	2	1
All	All	1861	0	1581	19	1

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 6.

All (19) 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:82:ILE:CD1	1:A:145:VAL:HG21	2.25	0.66
1:A:191:GLN:HB3	2:A:360:HOH:O	2.03	0.58
1:A:30:ILE:HG22	1:A:34:LYS:HE3	1.84	0.58
1:A:121:PHE:O	1:A:122:ASP:HB2	2.10	0.51
1:A:8:LEU:HD22	1:A:54:ILE:HD13	1.94	0.48
1:A:54:ILE:HG22	1:A:59:VAL:HG22	1.96	0.47
1:A:24:THR:HA	1:A:92:TRP:O	2.15	0.46
1:A:82:ILE:HD11	1:A:145:VAL:HG21	1.97	0.45
1:A:116:THR:HG21	1:A:130:HIS:CD2	2.53	0.44
1:A:121:PHE:CE2	1:A:175:ALA:HB2	2.53	0.44
1:A:92:TRP:CZ3	1:A:112:LYS:HE2	2.52	0.44
1:A:0:HIS:O	1:A:1:MET:C	2.55	0.43
1:A:68:LYS:NZ	2:A:318:HOH:O	2.52	0.42
1:A:82:ILE:HD11	1:A:140:CYS:SG	2.59	0.42
1:A:186:VAL:HB	1:A:193:ILE:HB	2.03	0.41
1:A:57:GLN:HE21	1:A:57:GLN:HB2	1.59	0.41
1:A:82:ILE:HD13	1:A:145:VAL:HG11	2.02	0.41
1:A:102:TYR:CE2	1:A:108:GLY:HA2	2.57	0.40
1:A:30:ILE:O	1:A:34:LYS:HG3	2.22	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:251:HOH:O	2:A:395:HOH:O[6_665]	2.19	0.01

### 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
1	A	196/221 (89%)	191 (97%)	4 (2%)	1 (0%)	29 18

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	122	ASP

#### 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
1	A	182/200 (91%)	171 (94%)	11 (6%)	19 9

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	GLU
1	A	14	LYS
1	A	54	ILE
1	A	91	LYS
1	A	95	GLN

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	111	LYS
1	A	118	GLU
1	A	133	ASN
1	A	153	ASP
1	A	197	THR
1	A	201	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	89	ASN
1	A	95	GLN
1	A	130	HIS
1	A	133	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](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	200/221 (90%)	0.06	8 (4%) 38 41	23, 34, 54, 73	9 (4%)

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	201	SER	5.7
1	A	19	TYR	4.8
1	A	182	LYS	3.1
1	A	127	ASN	3.0
1	A	27	ARG	2.5
1	A	180	LYS	2.4
1	A	188	ALA	2.2
1	A	178	TYR	2.2

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.