



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 05:10 am BST

PDB ID : 5W1N  
Title : Crystal structure of the human astrovirus 2 Oxford serotype capsid protein spike at 1.35-Å resolution  
Authors : Bogdanoff, W.A.; Perez, E.I.; DuBois, R.M.  
Deposited on : 2017-06-04  
Resolution : 1.35 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 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	:	2.11
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)	:	2.11

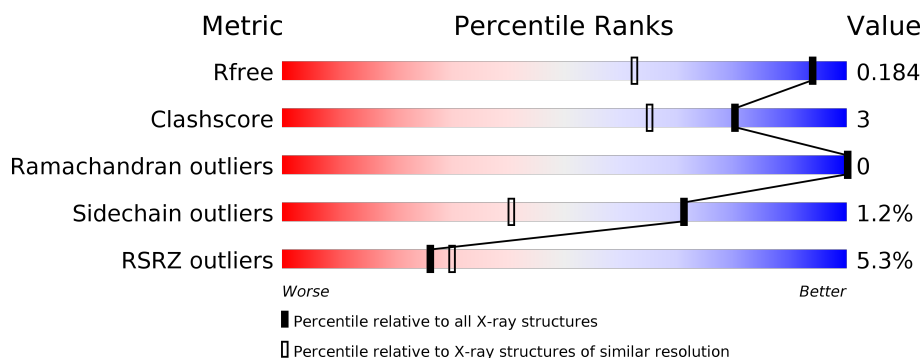
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.35 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))
$R_{free}$	130704	1385 (1.36-1.32)
Clashscore	141614	1417 (1.36-1.32)
Ramachandran outliers	138981	1397 (1.36-1.32)
Sidechain outliers	138945	1397 (1.36-1.32)
RSRZ outliers	127900	1369 (1.36-1.32)

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 on 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	243	<div> <div>3%</div> <div> <div></div> <div>83%</div> <div>6%</div> <div>11%</div> </div> </div>
1	B	243	<div> <div>4%</div> <div> <div></div> <div>86%</div> <div>•</div> <div>11%</div> </div> </div>
1	C	243	<div> <div>7%</div> <div> <div></div> <div>80%</div> <div>7%</div> <div>12%</div> </div> </div>
1	D	243	<div> <div>5%</div> <div> <div></div> <div>86%</div> <div>•</div> <div>11%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13630 atoms, of which 6185 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 Capsid polypeptide VP90.

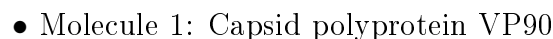
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	217	Total	C	H	N	O	S	0	4	0
			3317	1132	1551	297	327	10			
1	B	216	Total	C	H	N	O	S	0	2	0
			3311	1121	1574	285	321	10			
1	C	213	Total	C	H	N	O	S	0	1	0
			3227	1106	1504	288	319	10			
1	D	217	Total	C	H	N	O	S	0	1	0
			3311	1124	1556	295	326	10			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	109	Total	O	0	0
			109	109		
2	B	117	Total	O	0	0
			117	117		
2	C	109	Total	O	0	0
			109	109		
2	D	129	Total	O	0	0
			129	129		



- Molecule 1: Capsid polyprotein VP90



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.63Å 71.94Å 92.81Å 90.00° 111.22° 90.00°	Depositor
Resolution (Å)	43.97 – 1.35 43.97 – 1.35	Depositor EDS
% Data completeness (in resolution range)	96.0 (43.97-1.35) 96.0 (43.97-1.35)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.85 (at 1.35Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.156 , 0.184 0.157 , 0.184	Depositor DCC
$R_{free}$ test set	8967 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	10.3	Xtriage
Anisotropy	0.599	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.62 , 67.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.032 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	13630	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.92% 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

### 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	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.38	0/1826	0.63	0/2484
1	B	0.40	0/1790	0.62	0/2433
1	C	0.37	0/1771	0.63	0/2407
1	D	0.38	0/1806	0.61	0/2459
All	All	0.38	0/7193	0.62	0/9783

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

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	1766	1551	1741	9	0
1	B	1737	1574	1727	10	0
1	C	1723	1504	1698	17	0
1	D	1755	1556	1727	5	0
2	A	109	0	0	0	0
2	B	117	0	0	2	0
2	C	109	0	0	0	0
2	D	129	0	0	0	0
All	All	7445	6185	6893	38	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 3.

All (38) 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:B:448:ARG:HE	1:C:569:VAL:HG22	1.30	0.93
1:A:433:VAL:HG12	1:A:491:TRP:CE3	2.17	0.80
1:B:605:GLU:OE2	2:B:701:HOH:O	2.02	0.76
1:B:445:SER:OG	1:B:448:ARG:HG3	1.86	0.75
1:B:448:ARG:HE	1:C:569:VAL:CG2	1.98	0.75
1:D:472:HIS:HE1	1:D:474:LEU:HD23	1.54	0.73
1:D:472:HIS:CE1	1:D:474:LEU:HD23	2.26	0.70
1:B:433:VAL:HG12	1:B:491:TRP:CE3	2.27	0.70
1:C:533:ASN:O	1:C:533:ASN:ND2	2.26	0.69
1:C:441:MET:HE2	1:C:583:ILE:HG13	1.81	0.63
1:A:433:VAL:CG1	1:A:491:TRP:CZ3	2.81	0.62
1:D:600:GLU:HG2	1:D:601:ASN:N	2.16	0.60
1:A:433:VAL:HG11	1:A:491:TRP:CZ3	2.37	0.59
1:B:433:VAL:CG1	1:B:491:TRP:CZ3	2.86	0.58
1:C:433:VAL:CG2	1:C:489:VAL:HG13	2.33	0.58
1:C:472:HIS:CE1	1:C:474:LEU:HG	2.42	0.55
1:D:433:VAL:CG2	1:D:489:VAL:HG13	2.37	0.54
1:B:448:ARG:NE	1:C:569:VAL:HG22	2.12	0.54
1:B:433:VAL:HG12	1:B:491:TRP:CZ3	2.43	0.52
1:C:605:GLU:HG2	1:C:606:GLY:N	2.23	0.52
1:A:433:VAL:HG12	1:A:491:TRP:CZ3	2.44	0.52
1:A:493:THR:OG1	1:A:495:ASP:OD1	2.22	0.52
1:C:508:MET:HG3	1:C:523:MET:HE1	1.95	0.47
1:B:433:VAL:HG11	1:B:491:TRP:CZ3	2.50	0.47
1:D:525:GLN:HG3	1:D:607:TRP:CZ3	2.52	0.44
1:C:583:ILE:CD1	1:C:585:PHE:CZ	3.00	0.44
1:C:441:MET:HE2	1:C:583:ILE:CG1	2.46	0.44
1:B:497:GLN:NE2	2:B:702:HOH:O	2.51	0.43
1:C:583:ILE:HD11	1:C:585:PHE:CZ	2.53	0.43
1:C:533:ASN:O	1:C:533:ASN:CG	2.57	0.43
1:C:583:ILE:HG21	1:C:583:ILE:HD13	1.78	0.43
1:C:523:MET:HE3	1:C:523:MET:HB3	1.79	0.42
1:C:525:GLN:HG3	1:C:607:TRP:CZ3	2.55	0.42
1:C:500:ARG:HB3	1:C:500:ARG:CZ	2.50	0.41
1:A:445[A]:SER:OG	1:A:448:ARG:NE	2.48	0.41
1:A:489:VAL:HG21	1:A:542:LEU:HD21	2.02	0.40
1:A:487:LYS:HG2	1:A:488:THR:HG23	2.03	0.40
1:A:525:GLN:HG3	1:A:607:TRP:CZ3	2.56	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	217/243 (89%)	211 (97%)	6 (3%)	0	100	100
1	B	211/243 (87%)	207 (98%)	4 (2%)	0	100	100
1	C	208/243 (86%)	204 (98%)	4 (2%)	0	100	100
1	D	216/243 (89%)	210 (97%)	6 (3%)	0	100	100
All	All	852/972 (88%)	832 (98%)	20 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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 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	200/216 (93%)	196 (98%)	4 (2%)	55	20
1	B	195/216 (90%)	193 (99%)	2 (1%)	76	47
1	C	194/216 (90%)	190 (98%)	4 (2%)	53	18
1	D	198/216 (92%)	198 (100%)	0	100	100
All	All	787/864 (91%)	777 (99%)	10 (1%)	71	36

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	523[A]	MET
1	A	523[B]	MET

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	534	ARG
1	A	543	ILE
1	B	448	ARG
1	B	543	ILE
1	C	464	ASN
1	C	500	ARG
1	C	534	ARG
1	C	610	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	497	GLN
1	C	533	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 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.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

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	217/243 (89%)	0.48	7 (3%) 47 54	6, 14, 23, 33	0
1	B	216/243 (88%)	0.51	9 (4%) 36 42	6, 13, 26, 39	0
1	C	213/243 (87%)	0.57	18 (8%) 10 12	7, 14, 31, 40	0
1	D	217/243 (89%)	0.54	12 (5%) 25 28	6, 14, 25, 35	0
All	All	863/972 (88%)	0.52	46 (5%) 26 30	6, 14, 27, 40	0

All (46) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	461	THR	5.5
1	B	653	VAL	5.4
1	D	429	GLY	4.4
1	C	460	GLY	4.1
1	B	603	LEU	4.0
1	D	462	ASN	3.4
1	C	463	PRO	3.3
1	D	645	THR	3.3
1	C	601	ASN	3.3
1	A	532	ASN	3.2
1	C	567	ASP	3.2
1	C	534	ARG	3.1
1	C	500	ARG	3.1
1	C	473	ASP	3.1
1	B	605	GLU	3.1
1	C	474	LEU	3.0
1	D	430	GLU	3.0
1	D	500	ARG	2.9
1	B	529	SER	2.9
1	D	621	TYR	2.8
1	C	644	ASN	2.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	533	ASN	2.8
1	A	433	VAL	2.7
1	D	531	THR	2.7
1	D	600	GLU	2.6
1	C	602	ASN	2.6
1	C	462	ASN	2.6
1	A	463	PRO	2.5
1	B	462	ASN	2.5
1	B	463	PRO	2.5
1	C	569	VAL	2.5
1	D	463	PRO	2.4
1	C	621	TYR	2.4
1	C	605	GLU	2.3
1	A	605	GLU	2.3
1	A	621	TYR	2.3
1	C	464	ASN	2.2
1	B	460	GLY	2.2
1	D	506	ALA	2.2
1	D	474	LEU	2.1
1	A	464	ASN	2.1
1	A	460	GLY	2.0
1	C	600	GLU	2.0
1	C	430	GLU	2.0
1	B	498	PRO	2.0
1	B	476	HIS	2.0

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

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

## 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.