



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 09:02 am BST

PDB ID : 5KOU  
Title : Crystal structure of the human astrovirus 2 capsid protein spike domain at 1.87-Å resolution  
Authors : Bogdanoff, W.A.; DuBois, R.M.  
Deposited on : 2016-07-01  
Resolution : 1.87 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](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

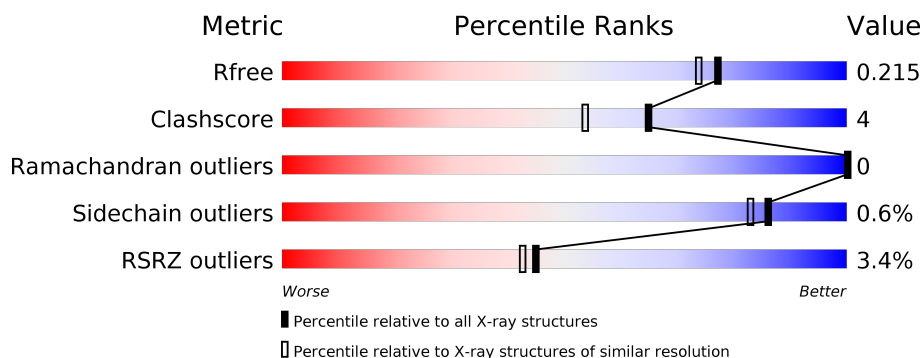
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.87 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	242	<div> <div>82%</div> <div>7%</div> <div>11%</div> </div>
1	B	242	<div>5%</div> <div>83%</div> <div>10%</div> <div>7%</div>
1	C	242	<div>6%</div> <div>77%</div> <div>13%</div> <div>10%</div>
1	D	242	<div>%</div> <div>79%</div> <div>11%</div> <div>10%</div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7459 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 Capsid polypeptide VP25.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	216	Total	C	N	O	S	0	4	0
			1761	1132	290	329	10			
1	B	225	Total	C	N	O	S	0	3	0
			1818	1167	299	342	10			
1	C	219	Total	C	N	O	S	0	3	0
			1781	1141	296	334	10			
1	D	217	Total	C	N	O	S	0	3	0
			1764	1133	292	330	9			

There are 104 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	427	MET	-	initiating methionine	UNP Q82446
A	428	GLY	-	expression tag	UNP Q82446
A	645	ALA	-	expression tag	UNP Q82446
A	646	ALA	-	expression tag	UNP Q82446
A	647	ALA	-	expression tag	UNP Q82446
A	648	GLU	-	expression tag	UNP Q82446
A	649	LEU	-	expression tag	UNP Q82446
A	650	ALA	-	expression tag	UNP Q82446
A	651	LEU	-	expression tag	UNP Q82446
A	652	VAL	-	expression tag	UNP Q82446
A	653	PRO	-	expression tag	UNP Q82446
A	654	ARG	-	expression tag	UNP Q82446
A	655	GLY	-	expression tag	UNP Q82446
A	656	SER	-	expression tag	UNP Q82446
A	657	SER	-	expression tag	UNP Q82446
A	658	ALA	-	expression tag	UNP Q82446
A	659	HIS	-	expression tag	UNP Q82446
A	660	HIS	-	expression tag	UNP Q82446
A	661	HIS	-	expression tag	UNP Q82446
A	662	HIS	-	expression tag	UNP Q82446
A	663	HIS	-	expression tag	UNP Q82446

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	664	HIS	-	expression tag	UNP Q82446
A	665	HIS	-	expression tag	UNP Q82446
A	666	HIS	-	expression tag	UNP Q82446
A	667	HIS	-	expression tag	UNP Q82446
A	668	HIS	-	expression tag	UNP Q82446
B	427	MET	-	initiating methionine	UNP Q82446
B	428	GLY	-	expression tag	UNP Q82446
B	645	ALA	-	expression tag	UNP Q82446
B	646	ALA	-	expression tag	UNP Q82446
B	647	ALA	-	expression tag	UNP Q82446
B	648	GLU	-	expression tag	UNP Q82446
B	649	LEU	-	expression tag	UNP Q82446
B	650	ALA	-	expression tag	UNP Q82446
B	651	LEU	-	expression tag	UNP Q82446
B	652	VAL	-	expression tag	UNP Q82446
B	653	PRO	-	expression tag	UNP Q82446
B	654	ARG	-	expression tag	UNP Q82446
B	655	GLY	-	expression tag	UNP Q82446
B	656	SER	-	expression tag	UNP Q82446
B	657	SER	-	expression tag	UNP Q82446
B	658	ALA	-	expression tag	UNP Q82446
B	659	HIS	-	expression tag	UNP Q82446
B	660	HIS	-	expression tag	UNP Q82446
B	661	HIS	-	expression tag	UNP Q82446
B	662	HIS	-	expression tag	UNP Q82446
B	663	HIS	-	expression tag	UNP Q82446
B	664	HIS	-	expression tag	UNP Q82446
B	665	HIS	-	expression tag	UNP Q82446
B	666	HIS	-	expression tag	UNP Q82446
B	667	HIS	-	expression tag	UNP Q82446
B	668	HIS	-	expression tag	UNP Q82446
C	427	MET	-	initiating methionine	UNP Q82446
C	428	GLY	-	expression tag	UNP Q82446
C	645	ALA	-	expression tag	UNP Q82446
C	646	ALA	-	expression tag	UNP Q82446
C	647	ALA	-	expression tag	UNP Q82446
C	648	GLU	-	expression tag	UNP Q82446
C	649	LEU	-	expression tag	UNP Q82446
C	650	ALA	-	expression tag	UNP Q82446
C	651	LEU	-	expression tag	UNP Q82446
C	652	VAL	-	expression tag	UNP Q82446
C	653	PRO	-	expression tag	UNP Q82446

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	654	ARG	-	expression tag	UNP Q82446
C	655	GLY	-	expression tag	UNP Q82446
C	656	SER	-	expression tag	UNP Q82446
C	657	SER	-	expression tag	UNP Q82446
C	658	ALA	-	expression tag	UNP Q82446
C	659	HIS	-	expression tag	UNP Q82446
C	660	HIS	-	expression tag	UNP Q82446
C	661	HIS	-	expression tag	UNP Q82446
C	662	HIS	-	expression tag	UNP Q82446
C	663	HIS	-	expression tag	UNP Q82446
C	664	HIS	-	expression tag	UNP Q82446
C	665	HIS	-	expression tag	UNP Q82446
C	666	HIS	-	expression tag	UNP Q82446
C	667	HIS	-	expression tag	UNP Q82446
C	668	HIS	-	expression tag	UNP Q82446
D	427	MET	-	initiating methionine	UNP Q82446
D	428	GLY	-	expression tag	UNP Q82446
D	645	ALA	-	expression tag	UNP Q82446
D	646	ALA	-	expression tag	UNP Q82446
D	647	ALA	-	expression tag	UNP Q82446
D	648	GLU	-	expression tag	UNP Q82446
D	649	LEU	-	expression tag	UNP Q82446
D	650	ALA	-	expression tag	UNP Q82446
D	651	LEU	-	expression tag	UNP Q82446
D	652	VAL	-	expression tag	UNP Q82446
D	653	PRO	-	expression tag	UNP Q82446
D	654	ARG	-	expression tag	UNP Q82446
D	655	GLY	-	expression tag	UNP Q82446
D	656	SER	-	expression tag	UNP Q82446
D	657	SER	-	expression tag	UNP Q82446
D	658	ALA	-	expression tag	UNP Q82446
D	659	HIS	-	expression tag	UNP Q82446
D	660	HIS	-	expression tag	UNP Q82446
D	661	HIS	-	expression tag	UNP Q82446
D	662	HIS	-	expression tag	UNP Q82446
D	663	HIS	-	expression tag	UNP Q82446
D	664	HIS	-	expression tag	UNP Q82446
D	665	HIS	-	expression tag	UNP Q82446
D	666	HIS	-	expression tag	UNP Q82446
D	667	HIS	-	expression tag	UNP Q82446
D	668	HIS	-	expression tag	UNP Q82446


- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	85	Total 85	O 85	0	0
2	B	108	Total 108	O 108	0	0
2	C	76	Total 76	O 76	0	0
2	D	66	Total 66	O 66	0	0

### 3 Residue-property plots [i](#)


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

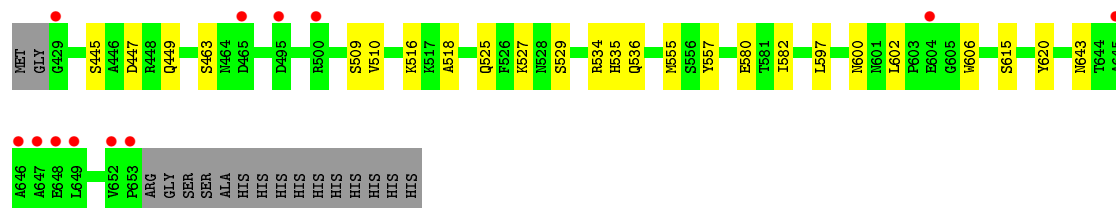
#### • Molecule 1: Capsid polyprotein VP25

Chain A: 




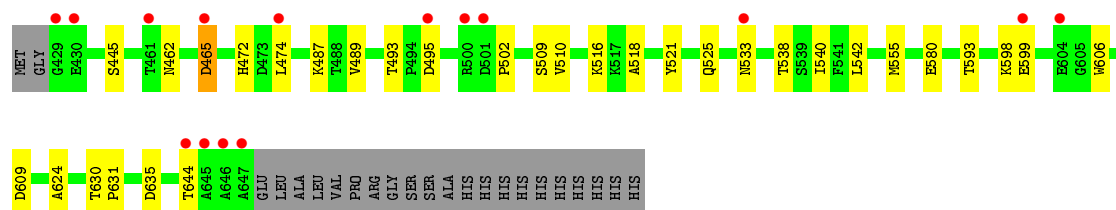
#### • Molecule 1: Capsid polyprotein VP25

Chain B: 




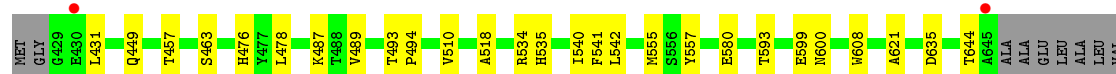
#### • Molecule 1: Capsid polyprotein VP25

Chain C: 



#### • Molecule 1: Capsid polyprotein VP25

Chain D: 



PRO  
ARG  
GLY  
SER  
SER  
ALA  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.07Å 67.72Å 90.53Å 76.53° 80.52° 86.29°	Depositor
Resolution (Å)	47.58 – 1.87 47.58 – 1.87	Depositor EDS
% Data completeness (in resolution range)	96.3 (47.58-1.87) 92.9 (47.58-1.87)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.76 (at 1.87Å)	Xtriage
Refinement program	PHENIX (1.10_2152: ???)	Depositor
R, $R_{free}$	0.193 , 0.215 0.193 , 0.215	Depositor DCC
$R_{free}$ test set	3991 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.3	Xtriage
Anisotropy	0.495	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 42.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7459	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.88% 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	0.29	0/1820	0.51	0/2479
1	B	0.30	0/1875	0.52	0/2555
1	C	0.29	0/1831	0.52	0/2494
1	D	0.29	0/1820	0.52	0/2480
All	All	0.29	0/7346	0.52	0/10008

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	A	1761	0	1741	10	0
1	B	1818	0	1793	13	0
1	C	1781	0	1752	21	0
1	D	1764	0	1743	15	0
2	A	85	0	0	1	0
2	B	108	0	0	0	0
2	C	76	0	0	0	0
2	D	66	0	0	0	0
All	All	7459	0	7029	58	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 4.

The worst 5 of 58 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:C:493:THR:HG22	1:C:495:ASP:H	1.26	1.00
1:A:537:ILE:HD11	1:A:602:LEU:HD11	1.57	0.86
1:D:487:LYS:NZ	1:D:635:ASP:OD1	2.15	0.79
1:B:536:GLN:HE22	1:B:620:TYR:H	1.40	0.67
1:B:555:MET:HE1	1:B:582:ILE:HD11	1.79	0.64

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	
1	A	218/242 (90%)	212 (97%)	6 (3%)	0	100	100
1	B	226/242 (93%)	218 (96%)	8 (4%)	0	100	100
1	C	220/242 (91%)	214 (97%)	6 (3%)	0	100	100
1	D	218/242 (90%)	211 (97%)	7 (3%)	0	100	100
All	All	882/968 (91%)	855 (97%)	27 (3%)	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	
1	A	201/216 (93%)	201 (100%)	0	100	100
1	B	205/216 (95%)	202 (98%)	3 (2%)	65	53
1	C	200/216 (93%)	198 (99%)	2 (1%)	76	69
1	D	200/216 (93%)	200 (100%)	0	100	100
All	All	806/864 (93%)	801 (99%)	5 (1%)	86	83

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	463	SER
1	B	615	SER
1	B	643	ASN
1	C	465	ASP
1	C	533	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	536	GLN
1	D	497	GLN
1	C	533	ASN
1	B	466	ASN
1	C	600	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 ⓘ

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 [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	216/242 (89%)	0.01	1 (0%) 91 91	14, 25, 46, 72	0
1	B	225/242 (92%)	0.13	12 (5%) 26 25	12, 23, 46, 74	0
1	C	219/242 (90%)	0.28	15 (6%) 17 16	14, 27, 61, 75	0
1	D	217/242 (89%)	-0.01	2 (0%) 84 84	13, 26, 49, 64	0
All	All	877/968 (90%)	0.11	30 (3%) 45 42	12, 25, 52, 75	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	645	ALA	6.8
1	B	647	ALA	6.3
1	C	646	ALA	6.1
1	B	652	VAL	4.7
1	C	429	GLY	4.7

### 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 carbohydrates 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.