



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 16, 2024 – 11:59 AM EDT

PDB ID : 4ZQ9  
Title : X-ray structure of AAV-2 OBD bound to AAVS1 site 3:1  
Authors : Musayev, F.N.; Zarate-Perez, F.; Escalante, C.R.  
Deposited on : 2015-05-08  
Resolution : 2.60 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 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	:	2.37.1
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)	:	2.37.1

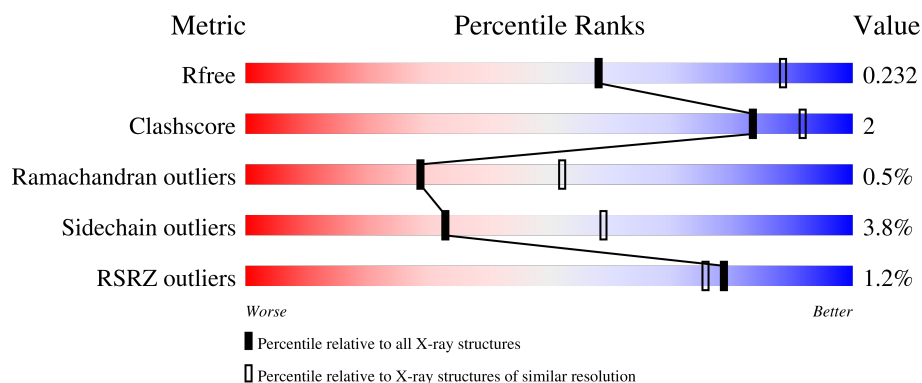
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	211	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>78%</span> <span>8%</span> <span>12%</span> </div> </div>
1	B	211	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green, yellow, orange, red, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>81%</span> <span>7%</span> <span>10%</span> </div> </div>
1	C	211	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red, orange, yellow, green, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>81%</span> <span>7%</span> <span>11%</span> </div> </div>
2	D	21	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green, yellow, orange, red, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>100%</span> </div> </div>
3	E	21	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green, yellow, orange, red, grey);"></div> <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <span>86%</span> <span>14%</span> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5496 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 Protein Rep68.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	185	Total	C	N	O	S	0	0	0
			1515	980	255	275	5			
1	B	189	Total	C	N	O	S	0	0	0
			1539	995	259	280	5			
1	C	187	Total	C	N	O	S	0	0	0
			1527	987	259	276	5			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P03132
A	-1	SER	-	expression tag	UNP P03132
A	0	HIS	-	expression tag	UNP P03132
A	17	GLU	GLY	conflict	UNP P03132
A	151	SER	CYS	engineered mutation	UNP P03132
A	156	PHE	TYR	engineered mutation	UNP P03132
B	-2	GLY	-	expression tag	UNP P03132
B	-1	SER	-	expression tag	UNP P03132
B	0	HIS	-	expression tag	UNP P03132
B	17	GLU	GLY	conflict	UNP P03132
B	151	SER	CYS	engineered mutation	UNP P03132
B	156	PHE	TYR	engineered mutation	UNP P03132
C	-2	GLY	-	expression tag	UNP P03132
C	-1	SER	-	expression tag	UNP P03132
C	0	HIS	-	expression tag	UNP P03132
C	17	GLU	GLY	conflict	UNP P03132
C	151	SER	CYS	engineered mutation	UNP P03132
C	156	PHE	TYR	engineered mutation	UNP P03132

- Molecule 2 is a DNA chain called DNA (5'-D(\*GP\*CP\*GP\*CP\*TP\*CP\*GP\*CP\*TP\*CP\*GP\*CP\*TP\*CP\*GP\*CP\*TP\*CP\*GP\*CP\*TP\*GP\*GP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	21	Total	C	N	O	P	0	0	0
			424	201	75	128	20			

- Molecule 3 is a DNA chain called DNA (5'-D(\*CP\*GP\*CP\*CP\*CP\*AP\*GP\*CP\*GP\*AP\*GP\*CP\*GP\*AP\*GP\*CP\*GP\*AP\*GP\*CP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	21	Total	C	N	O	P	0	0	0
			431	202	89	120	20			

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

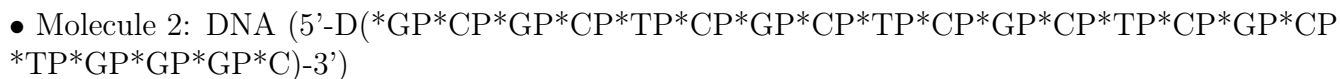
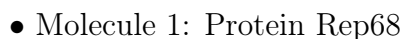
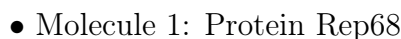
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mn	0	0
			1	1		
4	C	1	Total	Mn	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	14	Total	O	0	0
			14	14		
5	B	17	Total	O	0	0
			17	17		
5	C	15	Total	O	0	0
			15	15		
5	D	4	Total	O	0	0
			4	4		
5	E	8	Total	O	0	0
			8	8		



- Molecule 1: Protein Rep68



- Molecule 3: DNA (5'-D(\*CP\*GP\*CP\*CP\*CP\*AP\*GP\*CP\*GP\*AP\*GP\*CP\*GP\*AP\*GP\*CP\*GP\*AP\*GP\*CP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.10Å 137.25Å 79.25Å 90.00° 111.72° 90.00°	Depositor
Resolution (Å)	30.00 – 2.60 33.02 – 2.60	Depositor EDS
% Data completeness (in resolution range)	85.0 (30.00-2.60) 85.0 (33.02-2.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.26 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.205 , 0.230 0.214 , 0.232	Depositor DCC
$R_{free}$ test set	1949 reflections (5.28%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.7	Xtriage
Anisotropy	0.163	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 32.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	5496	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.99% 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

Bond lengths and bond angles in the following residue types are not validated in this section: MN

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	0.72	0/1554	0.90	5/2108 (0.2%)
1	B	0.73	0/1578	0.84	2/2141 (0.1%)
1	C	0.70	0/1567	0.87	5/2127 (0.2%)
2	D	0.43	0/473	0.89	0/728
3	E	0.47	0/485	0.87	0/747
All	All	0.68	0/5657	0.87	12/7851 (0.2%)

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	B	0	1

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	122	ARG	NE-CZ-NH1	10.07	125.34	120.30
1	A	27	VAL	CA-CB-CG2	6.99	121.38	110.90
1	B	103	MET	CG-SD-CE	-6.94	89.10	100.20
1	C	122	ARG	CD-NE-CZ	6.48	132.68	123.60
1	A	24	ASP	CB-CG-OD1	6.18	123.86	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	85	GLY	Peptide

## 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	1515	0	1507	9	0
1	B	1539	0	1530	11	0
1	C	1527	0	1515	5	0
2	D	424	0	237	0	0
3	E	431	0	233	2	1
4	A	1	0	0	0	0
4	C	1	0	0	0	0
5	A	14	0	0	0	0
5	B	17	0	0	0	0
5	C	15	0	0	0	0
5	D	4	0	0	0	0
5	E	8	0	0	0	0
All	All	5496	0	5022	25	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 2.

The worst 5 of 25 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:14:ASP:O	1:A:25:SER:OG	1.91	0.88
1:B:85:GLY:HA3	1:B:86:GLU:HG3	1.81	0.62
1:B:103:MET:O	1:B:103:MET:HE3	2.03	0.58
1:A:22:ILE:HG22	1:A:23:SER:OG	2.05	0.56
1:A:23:SER:HB3	1:A:24:ASP:HA	1.90	0.53

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)
3:E:22:DC:O5'	3:E:42:DG:O3'[1_655]	2.11	0.09

## 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	181/211 (86%)	178 (98%)	3 (2%)	0	100	100
1	B	185/211 (88%)	177 (96%)	5 (3%)	3 (2%)	9	19
1	C	183/211 (87%)	179 (98%)	4 (2%)	0	100	100
All	All	549/633 (87%)	534 (97%)	12 (2%)	3 (0%)	29	52

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	86	GLU
1	B	14	ASP
1	B	15	LEU

### 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	166/190 (87%)	158 (95%)	8 (5%)	25	49
1	B	168/190 (88%)	161 (96%)	7 (4%)	30	55
1	C	167/190 (88%)	163 (98%)	4 (2%)	49	74

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	501/570 (88%)	482 (96%)	19 (4%)	33 59

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	194	THR
1	C	122	ARG
1	C	195	HIS
1	C	104	VAL
1	B	15	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	92	HIS

### 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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	185/211 (87%)	-0.27	2 (1%) 80 78	16, 37, 70, 107	0
1	B	189/211 (89%)	-0.36	1 (0%) 91 89	17, 34, 64, 102	0
1	C	187/211 (88%)	-0.30	4 (2%) 63 58	15, 37, 68, 88	0
2	D	21/21 (100%)	-0.79	0 100 100	21, 33, 53, 68	0
3	E	21/21 (100%)	-0.73	0 100 100	18, 27, 63, 74	0
All	All	603/675 (89%)	-0.34	7 (1%) 79 76	15, 36, 68, 107	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	22	ILE	4.3
1	A	26	PHE	2.8
1	C	194	THR	2.6
1	C	196	VAL	2.2
1	C	195	HIS	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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
4	MN	A	301	1/1	0.85	0.13	99,99,99,99	0
4	MN	C	301	1/1	0.89	0.10	111,111,111,111	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.