



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 09:14 am BST

PDB ID : 1S9H  
Title : Crystal Structure of Adeno-associated virus Type 2 Rep40  
Authors : James, J.A.; Escalante, C.R.; Yoon-Robarts, M.; Edwards, T.A.; Linden, R.M.; Aggarwal, A.K.  
Deposited on : 2004-02-04  
Resolution : 2.40 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  
Xtrriage (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) : 2.11

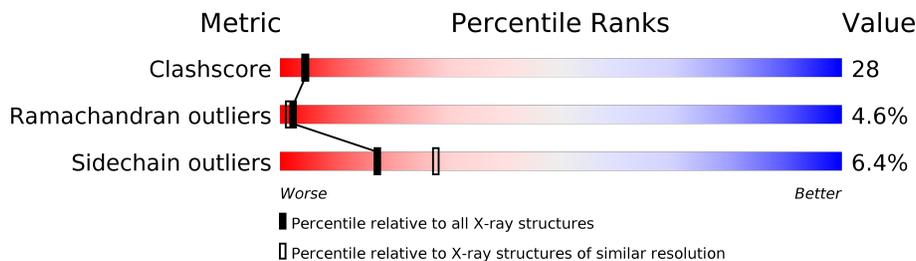
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 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 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  $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	268	
1	B	268	
1	C	268	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 5963 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 Rep 40 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	268	Total 2080	C 1329	N 350	O 391	S 10	0	0	0
1	B	267	Total 2064	C 1319	N 346	O 389	S 10	0	0	0
1	C	199	Total 1462	C 947	N 240	O 270	S 5	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	223	ALA	-	CLONING ARTIFACT	GB 2906019
A	224	GLY	-	CLONING ARTIFACT	GB 2906019
B	223	ALA	-	CLONING ARTIFACT	GB 2906019
B	224	GLY	-	CLONING ARTIFACT	GB 2906019
C	223	ALA	-	CLONING ARTIFACT	GB 2906019
C	224	GLY	-	CLONING ARTIFACT	GB 2906019

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	166	Total 166	O 166	0	0
2	B	117	Total 117	O 117	0	0
2	C	74	Total 74	O 74	0	0

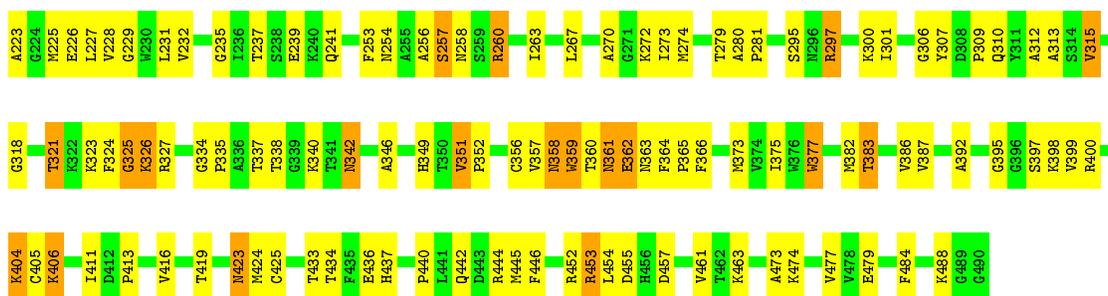
### 3 Residue-property plots [i](#)

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. 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.

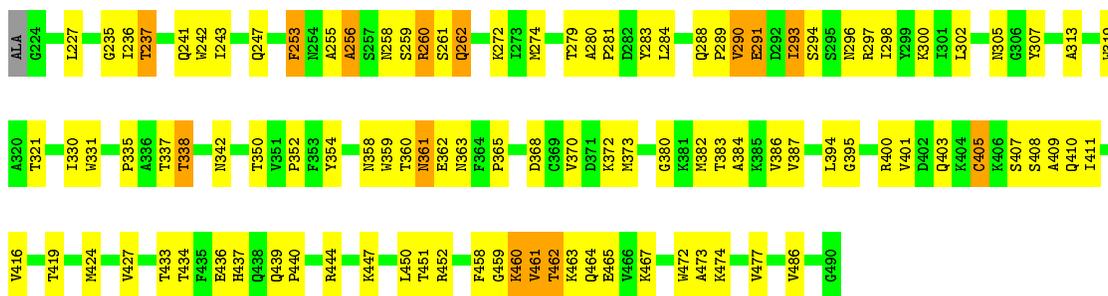
- Molecule 1: Rep 40 protein

Chain A: 



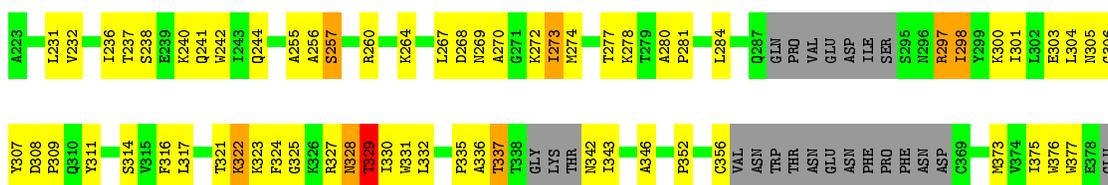
- Molecule 1: Rep 40 protein

Chain B: 



- Molecule 1: Rep 40 protein

Chain C: 



GLY  
LYS  
MET  
THR  
ALA  
LYS  
VAL  
VAL  
GLU  
SER  
ALA  
LYS  
ALA  
ILE  
LEU  
GLY  
GLY  
SER  
LYS  
VAL  
ARG  
VAL  
ASP  
GLN  
LYS  
CYS  
LYS  
SER  
SER  
A409  
Q410  
I411  
D412  
P413  
T414  
P415  
V416  
I417  
V418  
T422  
N423  
MET  
CYS  
ALA  
VAL  
ILE  
ASP  
GLY  
ASN  
SER  
THR  
THR  
PHE  
GLU  
HIS  
GLN  
GLN  
PRO  
I441

Q442  
D443  
R444  
M445  
L450  
V461  
T462  
K463  
Q464  
E465  
V466  
K467  
W472  
D475  
H476  
V477  
V478  
E483  
F484  
Y485  
V486  
K487  
K488  
G489  
G490

## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	126.29Å 126.29Å 97.89Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.97 – 2.40	Depositor
% Data completeness (in resolution range)	85.5 (19.97-2.40)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.06	Depositor
Refinement program	CNS	Depositor
R, $R_{free}$	0.227 , 0.293	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5963	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	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	A	0.43	0/2130	0.68	0/2890
1	B	0.39	0/2113	0.61	0/2868
1	C	0.34	0/1494	0.60	1/2033 (0.0%)
All	All	0.40	0/5737	0.63	1/7791 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	335	PRO	N-CA-CB	5.45	109.83	103.30

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	2080	0	2003	112	0
1	B	2064	0	1969	102	0
1	C	1462	0	1314	99	0
2	A	166	0	0	12	0
2	B	117	0	0	2	0
2	C	74	0	0	3	0
All	All	5963	0	5286	310	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 28.

The worst 5 of 310 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:434:THR:HG22	1:A:436:GLU:H	1.20	1.05
1:B:293:ILE:HD12	1:B:293:ILE:H	1.21	1.01
1:C:238:SER:OG	1:C:241:GLN:HG2	1.65	0.95
1:A:453:ARG:HE	1:A:453:ARG:H	1.17	0.88
1:C:307:TYR:O	1:C:309:PRO:HD3	1.72	0.88

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	266/268 (99%)	243 (91%)	16 (6%)	7 (3%)	5	5
1	B	265/268 (99%)	238 (90%)	17 (6%)	10 (4%)	3	2
1	C	187/268 (70%)	145 (78%)	26 (14%)	16 (9%)	1	0
All	All	718/804 (89%)	626 (87%)	59 (8%)	33 (5%)	2	1

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	257	SER
1	A	325	GLY
1	A	326	LYS
1	A	406	LYS
1	C	298	ILE

### 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	217/229 (95%)	199 (92%)	18 (8%)	11	17
1	B	212/229 (93%)	202 (95%)	10 (5%)	26	42
1	C	132/229 (58%)	124 (94%)	8 (6%)	18	30
All	All	561/687 (82%)	525 (94%)	36 (6%)	17	28

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	453	ARG
1	B	293	ILE
1	C	463	LYS
1	B	237	THR
1	B	338	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	423	ASN
1	A	476	HIS
1	B	464	GLN
1	A	361	ASN
1	B	482	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 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

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.