



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 04:32 am BST

PDB ID : 4AR2  
Title : Dodecahedron formed of penton base protein from adenovirus Ad3  
Authors : Burmeister, W.P.; Szolajska, E.; Zochowska, M.; Nerlo, B.; Andreev, I.; Schoehn, G.; Andrieu, J.-P.; Fender, P.; Naskalska, A.; Zubieta, C.; Cusack, S.; Chroboczek, J.  
Deposited on : 2012-04-20  
Resolution : 3.80 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

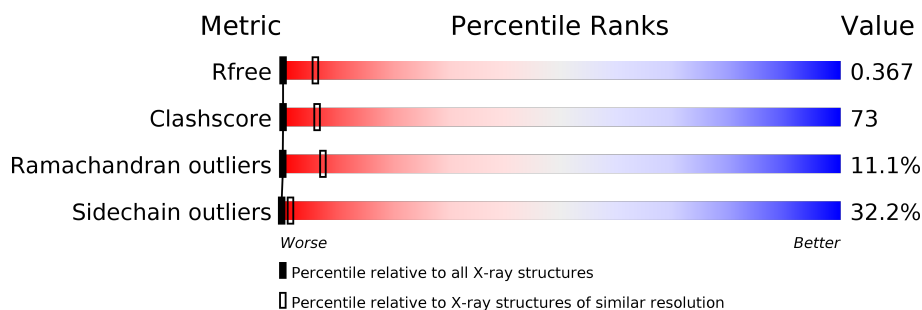
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)

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 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	509	
2	P	21	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3757 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 L2 PROTEIN III (PENTON BASE).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	455	Total	C	N	O	S	0	0	0
			3673	2324	629	706	14			

- Molecule 2 is a protein called FIBER PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	10	Total	C	N	O	S	0	0	0
			83	53	11	18	1			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	21	CYS	-	expression tag	UNP P03275

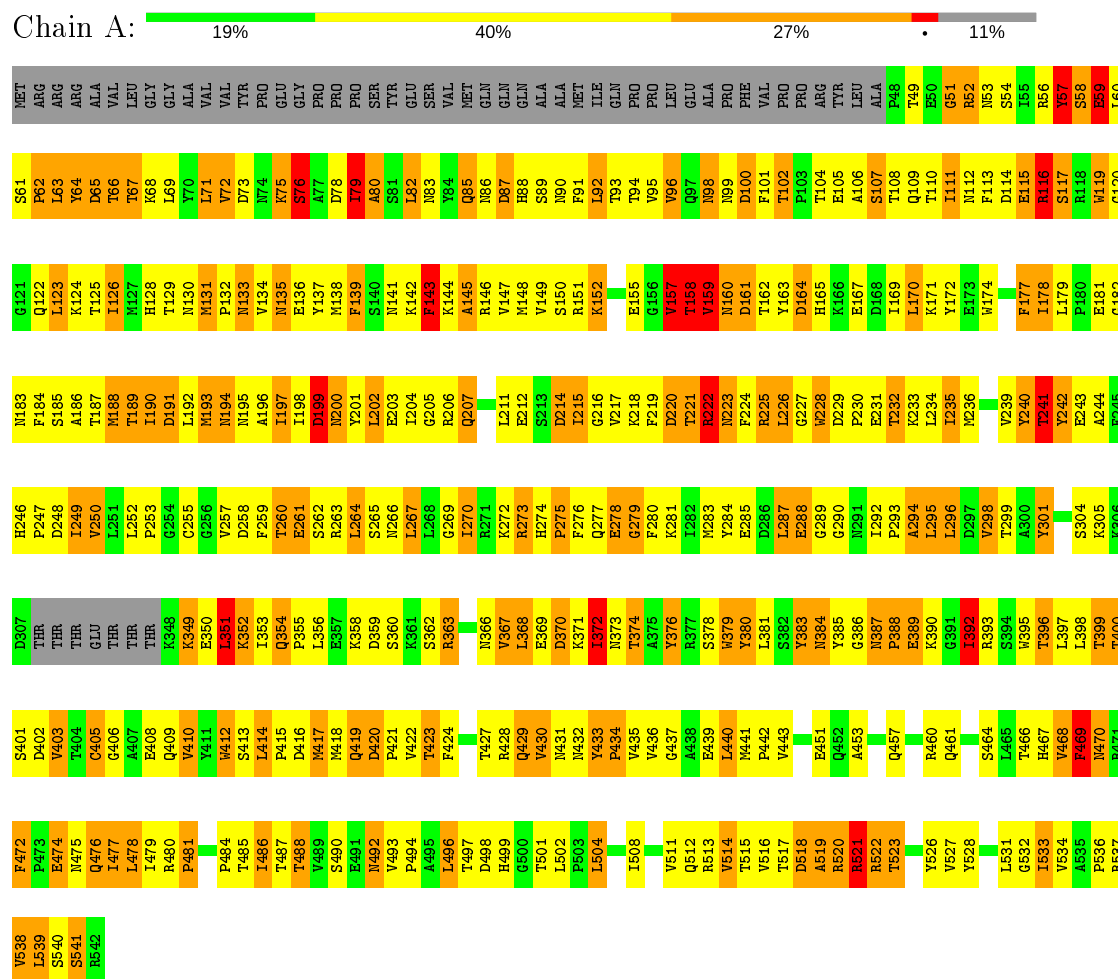
- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		

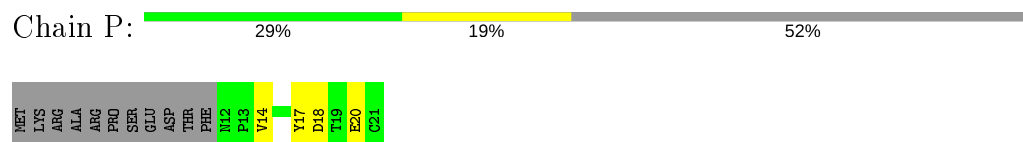
### 3 Residue-property plots

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.

#### • Molecule 1: L2 PROTEIN III (PENTON BASE)



#### • Molecule 2: FIBER PROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	346.04Å 348.55Å 372.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	258.20 – 3.80 29.41 – 3.80	Depositor EDS
% Data completeness (in resolution range)	56.8 (258.20-3.80) 56.9 (29.41-3.80)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.15 (at 3.75Å)	Xtriage
Refinement program	REFMAC 5.6.0114	Depositor
R, $R_{free}$	0.280 , 0.278 0.371 , 0.367	Depositor DCC
$R_{free}$ test set	5040 reflections (2.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	101.2	Xtriage
Anisotropy	0.277	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 17.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.065 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.73	EDS
Total number of atoms	3757	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.98% 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: CA

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	0.67	2/3758 (0.1%)	1.06	24/5103 (0.5%)
2	P	0.53	0/86	0.71	0/119
All	All	0.67	2/3844 (0.1%)	1.06	24/5222 (0.5%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	79	ILE	CA-CB	5.76	1.68	1.54
1	A	80	ALA	N-CA	-5.26	1.35	1.46

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	157	VAL	N-CA-C	8.49	133.92	111.00
1	A	279	GLY	N-CA-C	-8.15	92.72	113.10
1	A	469	PHE	N-CA-C	-8.12	89.08	111.00
1	A	155	GLU	N-CA-C	7.64	131.63	111.00
1	A	387	ASN	N-CA-C	7.21	130.46	111.00

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	3673	0	3600	538	0
2	P	83	0	68	3	0
3	A	1	0	0	0	0
All	All	3757	0	3668	539	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 73.

The worst 5 of 539 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:304:SER:CB	1:A:351:LEU:HB3	1.65	1.24
1:A:131:MET:H	1:A:492:ASN:ND2	1.46	1.12
1:A:431:ASN:O	1:A:521:ARG:HD3	1.52	1.09
1:A:304:SER:HB3	1:A:351:LEU:HB3	1.12	1.08
1:A:114:ASP:CG	1:A:115:GLU:H	1.49	1.08

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	451/509 (89%)	320 (71%)	81 (18%)	50 (11%)	0	7
2	P	8/21 (38%)	5 (62%)	2 (25%)	1 (12%)	0	6
All	All	459/530 (87%)	325 (71%)	83 (18%)	51 (11%)	0	7

5 of 51 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	49	THR
1	A	57	TYR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	59	GLU
1	A	66	THR
1	A	79	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	412/457 (90%)	276 (67%)	136 (33%)	0	1
2	P	10/20 (50%)	10 (100%)	0	100	100
All	All	422/477 (88%)	286 (68%)	136 (32%)	0	2

5 of 136 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	228	TRP
1	A	285	GLU
1	A	497	THR
1	A	231	GLU
1	A	249	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	200	ASN
1	A	354	GLN
1	A	499	HIS
1	A	207	GLN
1	A	223	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](#)

Of 1 ligands modelled in this entry, 1 is 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.