



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 05:08 PM UTC

PDB ID : 2YPF / pdb\_00002ypf  
Title : Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism  
Authors : Stella, S.; Molina, R.; Yefimenko, I.; Prieto, J.; Silva, G.H.; Bertonati, C.; Juillerat, A.; Duchateau, P.; Montoya, G.  
Deposited on : 2012-10-30  
Resolution : 2.55 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 : 4-5-2 with Phenix2.0  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.49

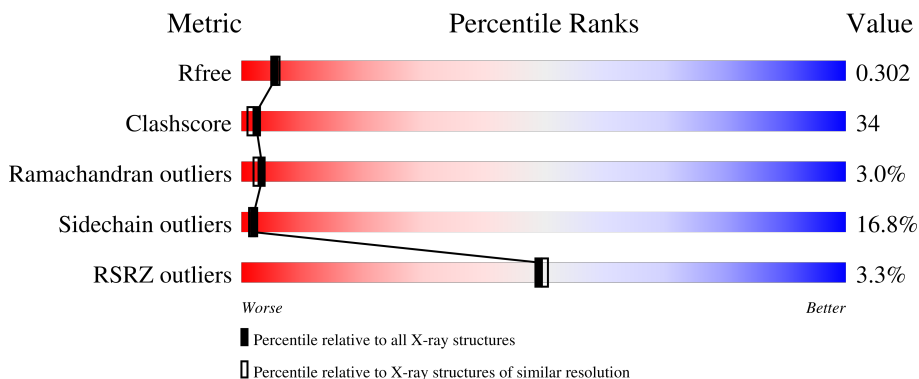
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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))
$R_{free}$	180053	1091 (2.54-2.54)
Clashscore	190562	1120 (2.54-2.54)
Ramachandran outliers	187476	1106 (2.54-2.54)
Sidechain outliers	187428	1106 (2.54-2.54)
RSRZ outliers	180081	1091 (2.54-2.54)

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	758	
2	B	22	
3	C	21	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 5822 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 AVRBS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	675	4772	2988	871	897	16	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	22	439	214	74	130	21	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	21	436	210	81	125	20	0	0	0

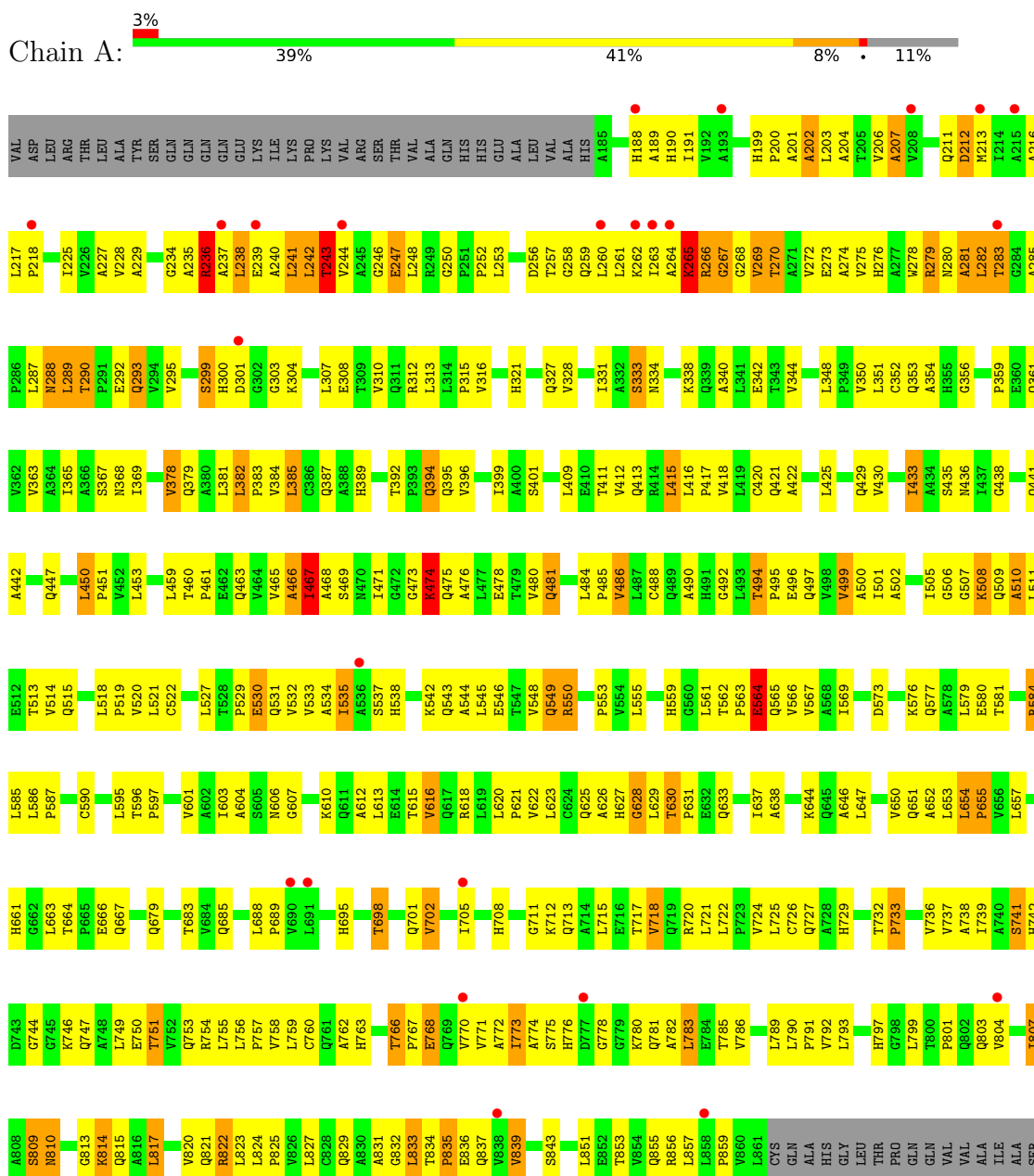
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	139	Total 139	O 139	0	0
4	B	17	Total 17	O 17	0	0
4	C	19	Total 19	O 19	0	0

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

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: AVRBS3



ASN  
GLY  
GLY  
GLY  
ARG  
PRO  
ALA  
LEU  
GLU  
SER  
ILE  
VAL  
ALA  
GLN  
LEU  
SER  
ARG  
PRO  
ASP  
PRO  
GLY  
SER  
SER  
ALA  
ALA  
LEU  
GLU  
HIS  
HIS  
HIS  
HIS  
HIS

- Molecule 2: 5'-D(\*TP\*TP\*TP\*AP\*TP\*AP\*TP\*AP\*AP\*AP\*CP\*CP\*TP\*AP \*AP\*CP\*CP\*C  
P\*TP\*CP\*TP\*AP)-3'

Chain B:  59% 32% 9%

T-2  
T-1  
T0  
A1  
T2  
A3  
A11  
C15  
T18  
A19

- Molecule 3: 5'-D(\*TP\*AP\*GP\*AP\*GP\*GP\*GP\*TP\*TP\*AP\*GP\*GP\*TP\*TP \*TP\*AP\*TP\*  
AP\*TP\*AP\*AP)-3'

Chain C:  43% 48% 10%

T-19  
A-18  
G-17  
A-16  
T-12  
T-11  
A-10  
G-9  
A-2  
T-1  
A0  
A1

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	151.11Å 100.25Å 61.37Å 90.00° 102.55° 90.00°	Depositor
Resolution (Å)	32.29 – 2.55 32.29 – 2.55	Depositor EDS
% Data completeness (in resolution range)	93.3 (32.29-2.55) 93.3 (32.29-2.55)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.70 (at 2.54Å)	Xtrriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.236 , 0.290 0.246 , 0.302	Depositor DCC
$R_{free}$ test set	1475 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	67.8	Xtrriage
Anisotropy	0.041	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 69.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	5822	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.46% 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	0.91	3/4831 (0.1%)	1.27	34/6609 (0.5%)
2	B	0.62	1/490 (0.2%)	1.10	2/752 (0.3%)
3	C	0.41	0/490	1.08	3/757 (0.4%)
All	All	0.85	4/5811 (0.1%)	1.24	39/8118 (0.5%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	236	ARG	CA-C	5.99	1.60	1.52
1	A	499	VAL	CA-CB	-5.75	1.48	1.54
2	B	3	DA	O3'-P	-5.64	1.52	1.61
1	A	471	ILE	C-O	-5.29	1.18	1.24

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	C	-1	DT	C2'-C3'-O3'	-7.91	99.63	111.50
1	A	269	VAL	CB-CA-C	-7.19	102.67	111.88
1	A	200	PRO	N-CA-CB	7.13	110.92	103.15
1	A	218	PRO	N-CA-CB	7.08	110.69	103.25
1	A	212	ASP	N-CA-C	6.92	117.72	108.38

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	4772	0	4877	362	0
2	B	439	0	252	10	0
3	C	436	0	241	12	0
4	A	139	0	0	35	0
4	B	17	0	0	0	0
4	C	19	0	0	4	0
All	All	5822	0	5370	376	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 34.

The worst 5 of 376 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:266:ARG:NH1	1:A:267:GLY:H	1.41	1.17
1:A:278:TRP:HB3	1:A:282:LEU:HD21	1.19	1.17
1:A:244:VAL:HB	4:A:2028:HOH:O	1.41	1.17
1:A:604:ALA:HB2	4:A:2104:HOH:O	1.42	1.15
1:A:289:LEU:HB3	1:A:293:GLN:NE2	1.63	1.13

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	670/758 (88%)	551 (82%)	99 (15%)	20 (3%)	<b>3</b> <b>2</b>

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	202	ALA
1	A	235	ALA

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	236	ARG
1	A	288	ASN
1	A	810	ASN

### 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	483/584 (83%)	402 (83%)	81 (17%)	<b>2</b> <b>2</b>

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	724	VAL
1	A	807	ILE
1	A	729	HIS
1	A	759	LEU
1	A	822	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	515	GLN
1	A	810	ASN
1	A	593	HIS
1	A	776	HIS
1	A	559	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 oligosaccharides 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	195:LEU	C	197:SER	N	9.14
1	A	193:ALA	C	195:LEU	N	4.52

## 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	675/758 (89%)	0.33	24 (3%) 46 47	34, 71, 112, 164	0
2	B	22/22 (100%)	-0.69	0 100 100	32, 41, 70, 147	0
3	C	21/21 (100%)	-0.16	0 100 100	34, 55, 99, 196	0
All	All	718/801 (89%)	0.28	24 (3%) 49 50	32, 70, 115, 196	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	804	VAL	4.8
1	A	263	ILE	3.9
1	A	301	ASP	3.2
1	A	858	LEU	3.0
1	A	215	ALA	3.0

### 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 oligosaccharides 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.