



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 15, 2023 – 08:17 PM JST

PDB ID : 6JW0  
Title : Universal RVD R\* accommodates cytosine via water-mediated interactions  
Authors : Liu, L.; Yi, C.  
Deposited on : 2019-04-18  
Resolution : 2.20 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 : 2.36  
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) : 2.36

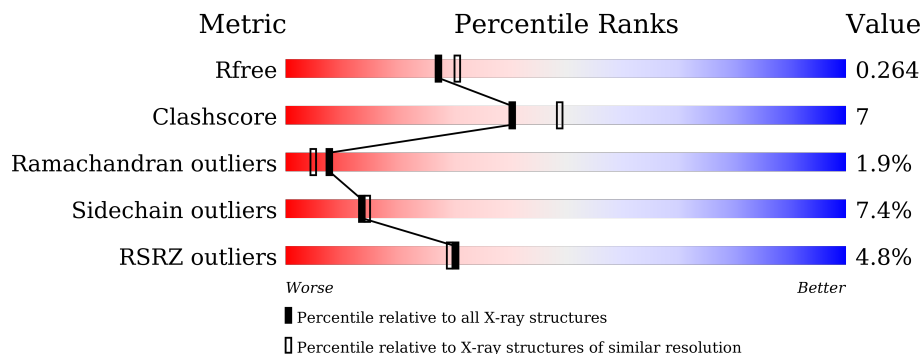
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	498	 3% (poor fit), 84% (0-1 outliers), 13% (2-3 outliers), .. (not modelled)
1	B	498	 7% (poor fit), 81% (0-1 outliers), 14% (2-3 outliers), .. (not modelled)
2	C	17	 6% (poor fit), 88% (0-1 outliers), 12% (2-3 outliers), .. (not modelled)
2	I	17	 6% (poor fit), 76% (0-1 outliers), 24% (2-3 outliers), .. (not modelled)
3	D	17	 6% (poor fit), 71% (0-1 outliers), 29% (2-3 outliers), .. (not modelled)
3	J	17	 12% (poor fit), 71% (0-1 outliers), 29% (2-3 outliers), .. (not modelled)

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 8739 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 TAL effector.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	496	Total 3606	C 2253	N 675	O 666	S 12	0	0	0
1	B	496	Total 3606	C 2253	N 675	O 666	S 12	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(\*TP\*GP\*TP\*CP\*CP\*CP\*TP\*TP\*CP\*GP\*CP\*GP\*TP\*CP\*TP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	I	17	Total 335	C 162	N 50	O 107	P 16	0	0	0
2	C	17	Total 335	C 162	N 50	O 107	P 16	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	J	17	Total 355	C 167	N 79	O 93	P 16	0	0	0
3	D	17	Total 355	C 167	N 79	O 93	P 16	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	50	Total 50 O 50	0	0
4	I	19	Total 19 O 19	0	0
4	J	3	Total 3 O 3	0	0

*Continued on next page...*

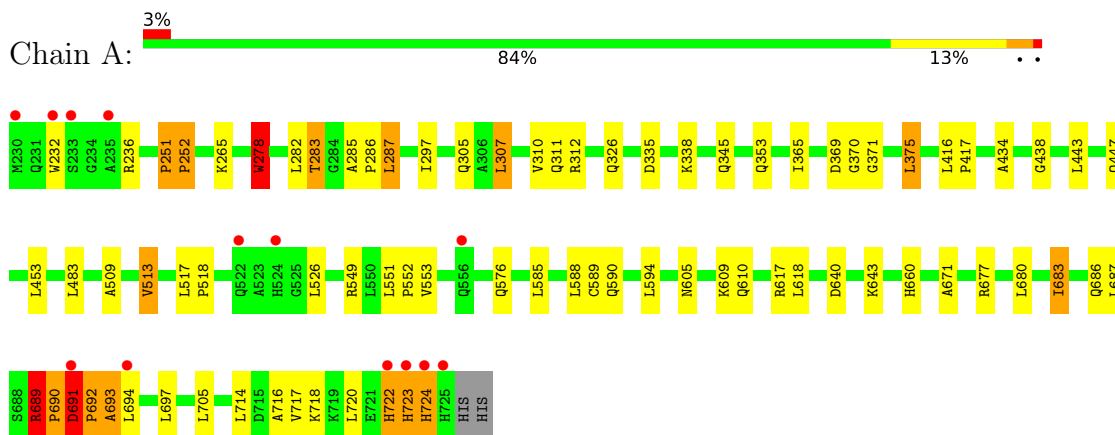
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
4	B	48	Total O 48 48	0	0
4	C	21	Total O 21 21	0	0
4	D	6	Total O 6 6	0	0

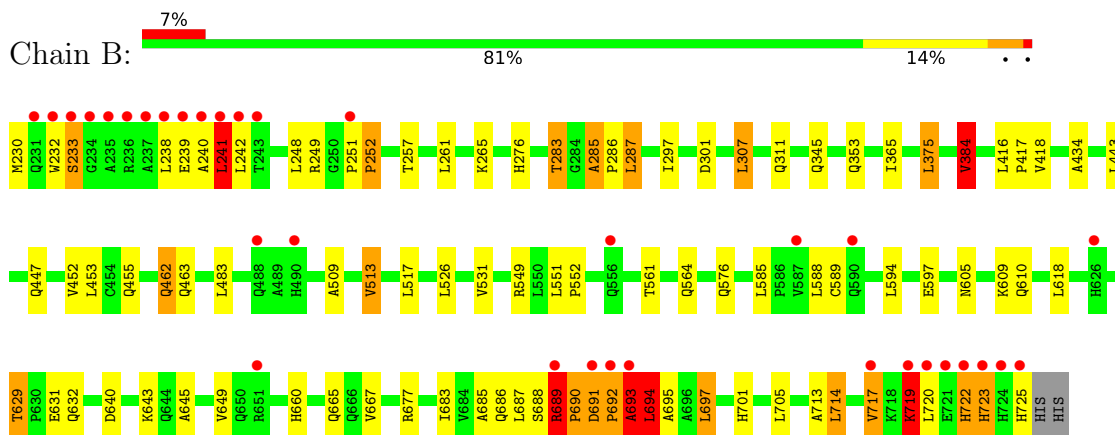
### 3 Residue-property plots [i](#)

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.

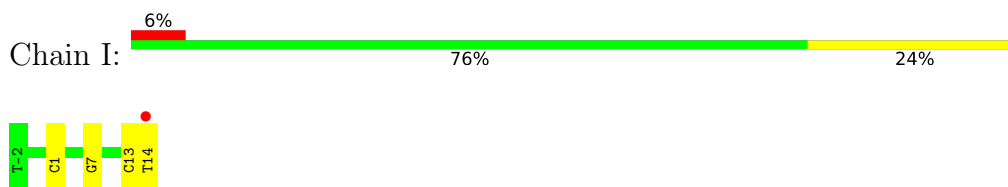
- Molecule 1: TAL effector



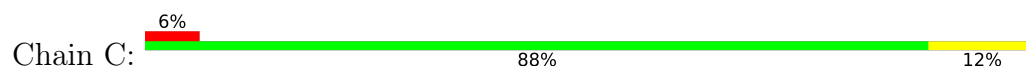
- Molecule 1: TAL effector



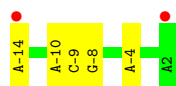
- Molecule 2: DNA (5'-D(\*TP\*GP\*TP\*CP\*CP\*CP\*TP\*TP\*CP\*GP\*CP\*GP\*TP\*CP\*TP\*CP\*T)-3')



- Molecule 2: DNA (5'-D(\*TP\*GP\*TP\*CP\*CP\*CP\*TP\*TP\*CP\*GP\*CP\*GP\*TP\*CP\*TP\*CP\*T)-3')



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



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



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.84Å 87.40Å 88.14Å 90.00° 104.12° 90.00°	Depositor
Resolution (Å)	85.47 – 2.20 32.28 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.0 (85.47-2.20) 97.1 (32.28-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.21 (at 2.20Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.222 , 0.263 0.227 , 0.264	Depositor DCC
$R_{free}$ test set	3193 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	32.7	Xtrriage
Anisotropy	0.434	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 38.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.018 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	8739	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.37% 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	0.69	1/3661 (0.0%)	0.85	8/5001 (0.2%)
1	B	0.65	0/3661	0.87	8/5001 (0.2%)
2	C	0.67	0/371	1.06	0/569
2	I	0.66	0/371	0.97	2/569 (0.4%)
3	D	0.56	0/402	0.84	1/620 (0.2%)
3	J	0.64	0/402	0.86	0/620
All	All	0.67	1/8868 (0.0%)	0.87	19/12380 (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	A	0	8
1	B	0	4
All	All	0	12

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	278	TRP	CB-CG	-5.33	1.40	1.50

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	549	ARG	NE-CZ-NH2	-11.49	114.56	120.30
1	B	549	ARG	NE-CZ-NH1	10.60	125.60	120.30
1	A	549	ARG	NE-CZ-NH1	-9.65	115.47	120.30
1	A	549	ARG	NE-CZ-NH2	9.48	125.04	120.30
1	B	384	VAL	CB-CA-C	-7.74	96.69	111.40

There are no chirality outliers.



5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	232	TRP	Peptide
1	A	236	ARG	Peptide
1	A	251	PRO	Mainchain,Peptide
1	A	285	ALA	Mainchain,Peptide
1	A	689	ARG	Peptide

## 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	3606	0	3740	45	0
1	B	3606	0	3740	64	0
2	C	335	0	193	9	0
2	I	335	0	193	9	0
3	D	355	0	189	4	0
3	J	355	0	189	4	0
4	A	50	0	0	5	0
4	B	48	0	0	2	0
4	C	21	0	0	0	0
4	D	6	0	0	0	0
4	I	19	0	0	0	0
4	J	3	0	0	0	0
All	All	8739	0	8244	121	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 7.

The worst 5 of 121 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:720:LEU:O	2:C:14:DT:C4	2.06	1.08
1:A:723:HIS:HB3	3:D:-14:DA:C6	1.90	1.05
1:B:418:VAL:HG23	4:B:806:HOH:O	1.54	1.05
2:I:13:DC:C2'	2:I:14:DT:H5''	1.91	0.99
1:A:724:HIS:HB3	2:C:14:DT:C4	1.96	0.99

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	494/498 (99%)	471 (95%)	15 (3%)	8 (2%)	9	7
1	B	494/498 (99%)	467 (94%)	16 (3%)	11 (2%)	6	4
All	All	988/996 (99%)	938 (95%)	31 (3%)	19 (2%)	8	5

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	691	ASP
1	A	693	ALA
1	B	252	PRO
1	B	285	ALA
1	B	693	ALA

### 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	380/384 (99%)	356 (94%)	24 (6%)	18	20
1	B	380/384 (99%)	348 (92%)	32 (8%)	11	11
All	All	760/768 (99%)	704 (93%)	56 (7%)	13	14

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	345	GLN
1	B	722	HIS
1	B	462	GLN
1	B	719	LYS
1	B	694	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	699	ASN
1	B	723	HIS
1	A	699	ASN
1	A	724	HIS
1	B	564	GLN

### 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](#)

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	496/498 (99%)	-0.04	13 (2%) 56 53	26, 41, 75, 139	0
1	B	496/498 (99%)	0.51	33 (6%) 17 16	27, 44, 92, 239	0
2	C	17/17 (100%)	-0.17	1 (5%) 22 21	30, 35, 67, 141	0
2	I	17/17 (100%)	-0.55	1 (5%) 22 21	28, 32, 92, 150	0
3	D	17/17 (100%)	0.00	1 (5%) 22 21	37, 47, 78, 137	0
3	J	17/17 (100%)	-0.25	2 (11%) 4 4	32, 44, 90, 112	0
All	All	1060/1064 (99%)	0.20	51 (4%) 30 29	26, 43, 88, 239	0

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	724	HIS	14.7
1	B	235	ALA	14.1
1	B	232	TRP	14.0
1	B	239	GLU	13.7
1	B	234	GLY	13.5

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.