



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 25, 2024 – 03:53 pm GMT

PDB ID : 8OK1  
Title : WD repeat containing protein 5 (WDR5)- N225A mutant  
Authors : Wolf, E.; Boergel, A.  
Deposited on : 2023-03-26  
Resolution : 1.38 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

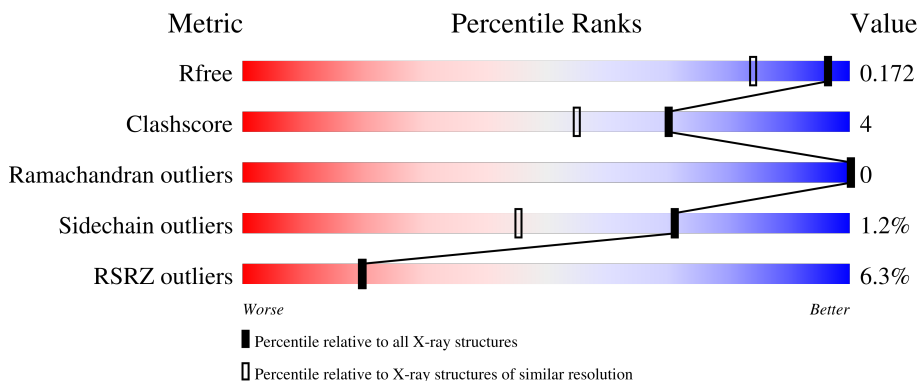
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.38 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	2907 (1.40-1.36)
Clashscore	141614	3037 (1.40-1.36)
Ramachandran outliers	138981	2970 (1.40-1.36)
Sidechain outliers	138945	2969 (1.40-1.36)
RSRZ outliers	127900	2846 (1.40-1.36)

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	555	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 2594 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 Glutathione S-transferase class-mu 26 kDa isozyme, WD repeat-containing protein 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	302	2349	1505	388	447	9	0	2	0

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-220	MET	-	initiating methionine	UNP P08515
A	-219	LYS	-	expression tag	UNP P08515
A	-218	HIS	-	expression tag	UNP P08515
A	-217	HIS	-	expression tag	UNP P08515
A	-216	HIS	-	expression tag	UNP P08515
A	-215	HIS	-	expression tag	UNP P08515
A	-214	HIS	-	expression tag	UNP P08515
A	-213	HIS	-	expression tag	UNP P08515
A	-212	ASN	-	expression tag	UNP P08515
A	-211	THR	-	expression tag	UNP P08515
A	-210	SER	-	expression tag	UNP P08515
A	-209	SER	-	expression tag	UNP P08515
A	-208	ASN	-	expression tag	UNP P08515
A	-207	SER	-	expression tag	UNP P08515
A	12	SER	-	linker	UNP P08515
A	13	ASP	-	linker	UNP P08515
A	14	LEU	-	linker	UNP P08515
A	15	GLU	-	linker	UNP P08515
A	16	VAL	-	linker	UNP P08515
A	17	LEU	-	linker	UNP P08515
A	18	PHE	-	linker	UNP P08515
A	19	GLN	-	linker	UNP P08515
A	20	GLY	-	linker	UNP P08515
A	21	PRO	-	linker	UNP P08515
A	22	SER	-	linker	UNP P08515
A	23	ALA	-	linker	UNP P08515

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	THR	-	linker	UNP P08515
A	25	GLN	-	linker	UNP P08515
A	26	SER	-	linker	UNP P08515
A	27	LYS	-	linker	UNP P08515
A	28	PRO	-	linker	UNP P08515
A	29	THR	-	linker	UNP P08515
A	30	PRO	-	linker	UNP P08515
A	31	VAL	-	linker	UNP P08515
A	225	ALA	ASN	conflict	UNP P61965

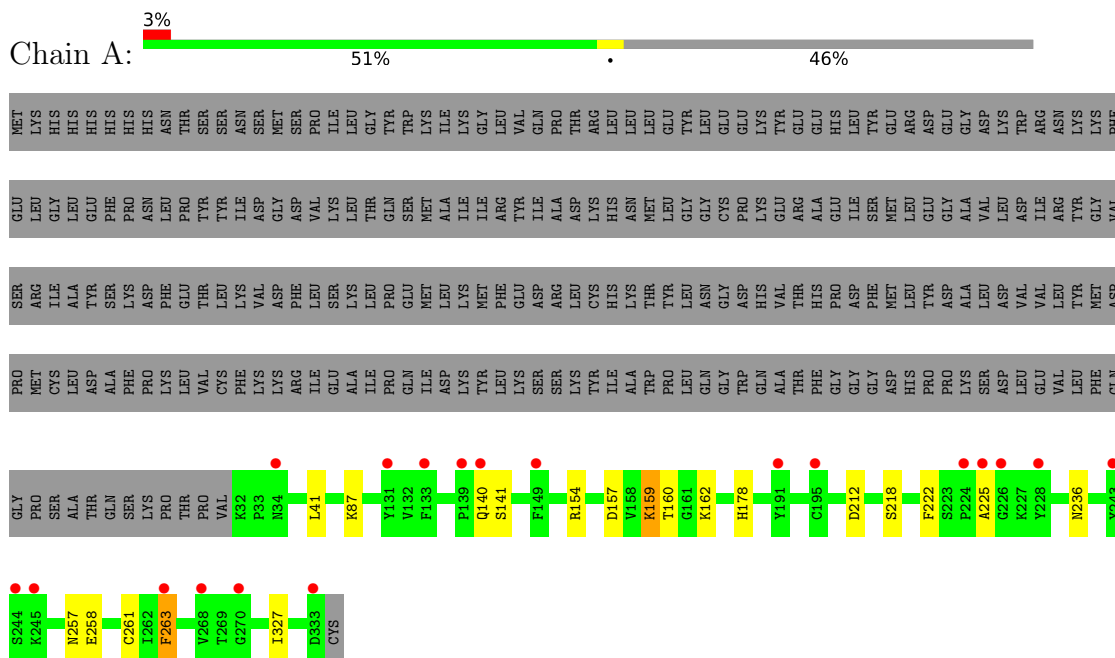
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	245	Total O 245 245	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: Glutathione S-transferase class-mu 26 kDa isozyme, WD repeat-containing protein 5



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.50Å 98.13Å 80.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.67 – 1.38 48.67 – 1.38	Depositor EDS
% Data completeness (in resolution range)	99.2 (48.67-1.38) 99.1 (48.67-1.38)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.66 (at 1.38Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.194 , 0.218 0.148 , 0.172	Depositor DCC
$R_{free}$ test set	3047 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.8	Xtrriage
Anisotropy	0.163	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 38.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2594	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.88% 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	0.76	0/2407	0.88	2/3267 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	154	ARG	NE-CZ-NH2	-8.21	116.19	120.30
1	A	141	SER	CB-CA-C	5.91	121.32	110.10

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	2349	0	2316	17	1
2	A	245	0	0	7	1
All	All	2594	0	2316	17	2

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 4.

All (17) 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:140:GLN:O	2:A:401:HOH:O	1.82	0.97
1:A:236:ASN:HD22	1:A:257:ASN:H	1.20	0.88
1:A:178:HIS:HD2	1:A:222:PHE:H	1.23	0.86
1:A:157:ASP:OD2	1:A:159:LYS:HD3	1.79	0.82
1:A:263:PHE:HZ	2:A:599:HOH:O	1.63	0.81
1:A:263:PHE:CZ	2:A:599:HOH:O	2.39	0.74
1:A:178:HIS:CD2	1:A:222:PHE:H	2.08	0.72
1:A:157:ASP:OD2	1:A:159:LYS:CD	2.41	0.69
1:A:258:GLU:OE1	2:A:402:HOH:O	2.16	0.61
1:A:236:ASN:HD21	1:A:258:GLU:HG3	1.68	0.58
1:A:225:ALA:HB1	2:A:607:HOH:O	2.02	0.58
1:A:236:ASN:ND2	1:A:257:ASN:H	1.96	0.58
1:A:159:LYS:HD3	1:A:160:THR:HG23	1.87	0.57
1:A:41:LEU:HB2	1:A:327:ILE:HB	1.94	0.49
1:A:263:PHE:CE2	2:A:449:HOH:O	2.66	0.48
1:A:218:SER:HB2	1:A:261:CYS:HA	1.97	0.47
1:A:263:PHE:HE1	2:A:594:HOH:O	2.03	0.40

All (2) 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)
2:A:401:HOH:O	2:A:401:HOH:O[4_555]	1.57	0.63
1:A:87:LYS:NZ	1:A:212:ASP:OD1[8_545]	2.07	0.13

## 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	302/555 (54%)	285 (94%)	17 (6%)	0	<a href="#">100</a> <a href="#">100</a>

There are no Ramachandran outliers to report.



### 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	262/490 (54%)	259 (99%)	3 (1%)	<span style="border: 1px solid black; padding: 2px;">73</span> <span style="border: 1px solid black; padding: 2px;">48</span>

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	159	LYS
1	A	162	LYS
1	A	263	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	100	ASN
1	A	178	HIS
1	A	236	ASN
1	A	314	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 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	302/555 (54%)	0.53	19 (6%) 20 20	11, 16, 27, 44	0

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	225	ALA	6.3
1	A	333	ASP	5.2
1	A	263	PHE	4.7
1	A	226	GLY	4.4
1	A	139	PRO	3.2
1	A	34	ASN	3.2
1	A	244	SER	2.9
1	A	195	CYS	2.8
1	A	140	GLN	2.6
1	A	268	VAL	2.5
1	A	191	TYR	2.4
1	A	228	TYR	2.4
1	A	245	LYS	2.3
1	A	270	GLY	2.3
1	A	131	TYR	2.2
1	A	243	TYR	2.2
1	A	133[A]	PHE	2.2
1	A	224	PRO	2.2
1	A	149[A]	PHE	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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.