



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 04:24 PM GMT

PDB ID : 4ERZ  
Title : X-ray structure of WDR5-MLL4 Win motif peptide binary complex  
Authors : Dharmarajan, V.; Lee, J.-H.; Patel, A.; Skalnik, D.G.; Cosgrove, M.S.  
Deposited on : 2012-04-21  
Resolution : 1.75 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  
<http://wwpdb.org/validation/2016/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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

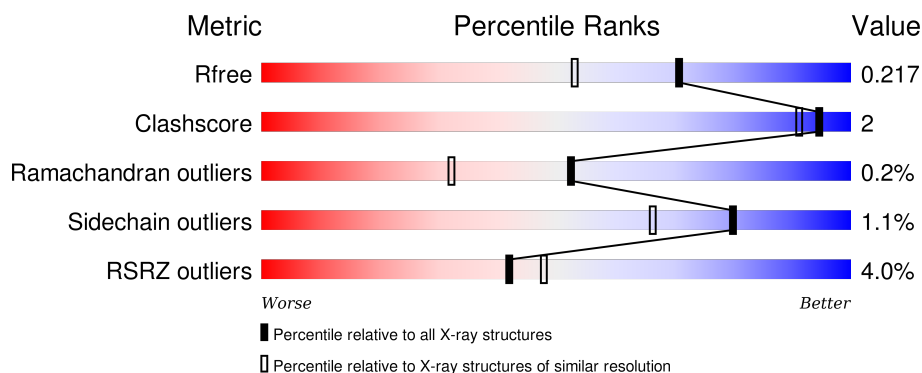
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.75 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))
$R_{free}$	91344	1609 (1.76-1.76)
Clashscore	102246	1730 (1.76-1.76)
Ramachandran outliers	100387	1711 (1.76-1.76)
Sidechain outliers	100360	1711 (1.76-1.76)
RSRZ outliers	91569	1610 (1.76-1.76)

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. 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	312	<div> <div>4%</div> <div>92%</div> <div>5%</div> </div>
1	B	312	<div> <div>3%</div> <div>93%</div> <div>5%</div> </div>
1	C	312	<div> <div>4%</div> <div>94%</div> <div>5%</div> </div>
2	D	14	<div> <div>21%</div> <div>64%</div> <div>21%</div> <div>14%</div> </div>
2	E	14	<div> <div>14%</div> <div>86%</div> <div>14%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	14	<div><div></div><div></div><div></div><div></div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7886 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 WD repeat-containing protein 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	304	Total	C	N	O	S	0	0	0
			2357	1503	393	451	10			
1	B	304	Total	C	N	O	S	0	0	0
			2357	1503	393	451	10			
1	C	304	Total	C	N	O	S	0	0	0
			2357	1503	393	451	10			

- Molecule 2 is a protein called Histone-lysine N-methyltransferase MLL4.

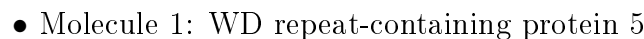
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	0	0	0
			88	54	18	16			
2	E	12	Total	C	N	O	0	0	0
			88	54	18	16			
2	F	12	Total	C	N	O	0	0	0
			88	54	18	16			

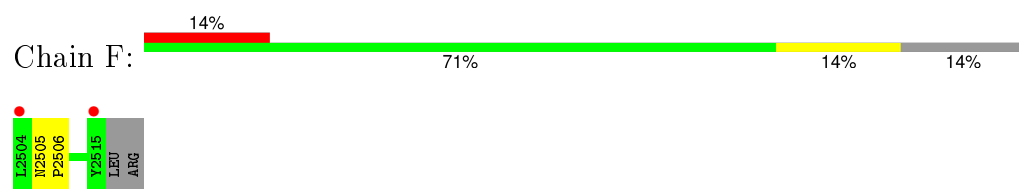
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	165	Total	O	0	0
			165	165		
3	B	186	Total	O	0	0
			186	186		
3	C	180	Total	O	0	0
			180	180		
3	D	6	Total	O	0	0
			6	6		
3	E	8	Total	O	0	0
			8	8		
3	F	6	Total	O	0	0
			6	6		



- Molecule 1: WD repeat-containing protein 5





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	139.03Å 80.29Å 87.63Å 90.00° 90.04° 90.00°	Depositor
Resolution (Å)	27.24 – 1.75 27.24 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.5 (27.24-1.75) 99.1 (27.24-1.75)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.54 (at 1.75Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.3_928)	Depositor
R, $R_{free}$	0.186 , 0.216 0.187 , 0.217	Depositor DCC
$R_{free}$ test set	4841 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	19.0	Xtriage
Anisotropy	0.409	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 44.6	EDS
Estimated twinning fraction	0.025 for -1/2*h+3/2*k,1/2*h+1/2*k,-l 0.024 for -1/2*h-3/2*k,-1/2*h+1/2*k,-l 0.477 for 1/2*h+3/2*k,1/2*h-1/2*k,-l 0.478 for 1/2*h-3/2*k,-1/2*h-1/2*k,-l 0.024 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	1 of 96859 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7886	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.70% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.35	0/2413	0.56	0/3272
1	B	0.35	0/2413	0.57	0/3272
1	C	0.35	0/2413	0.56	0/3272
2	D	0.29	0/90	0.49	0/122
2	E	0.30	0/90	0.48	0/122
2	F	0.28	0/90	0.48	0/122
All	All	0.35	0/7509	0.56	0/10182

There are no bond length outliers.

There are no bond angle outliers.

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	2357	0	2339	9	0
1	B	2357	0	2339	8	0
1	C	2357	0	2339	5	0
2	D	88	0	76	2	0
2	E	88	0	76	0	0
2	F	88	0	76	1	0
3	A	165	0	0	0	0
3	B	186	0	0	1	0
3	C	180	0	0	0	0
3	D	6	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	8	0	0	0	0
3	F	6	0	0	0	0
All	All	7886	0	7245	25	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 2.

All (25) 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:C:245:LYS:HD2	1:C:247:LYS:HD3	1.84	0.58
1:A:32:LYS:HG2	1:A:294:VAL:HB	1.86	0.58
1:A:41:LEU:HB2	1:A:327:ILE:HB	1.90	0.53
1:B:78:LYS:HE3	3:B:450:HOH:O	2.12	0.49
1:B:69:ILE:HD11	1:B:104:SER:HB3	1.93	0.49
1:C:213:ASP:O	1:C:214:ASN:ND2	2.26	0.48
1:B:41:LEU:HB2	1:B:327:ILE:HB	1.96	0.48
1:C:69:ILE:HD11	1:C:104:SER:HB3	1.95	0.47
1:A:69:ILE:HD11	1:A:104:SER:HB3	1.98	0.45
1:B:69:ILE:HB	1:B:83:ILE:HB	1.98	0.45
1:A:242:ASP:HB2	1:A:249:LEU:HD11	1.97	0.45
1:C:41:LEU:HB2	1:C:327:ILE:HB	1.99	0.45
1:C:218:SER:HB2	1:C:261:CYS:HA	1.98	0.44
1:B:218:SER:HB2	1:B:261:CYS:HA	2.00	0.43
2:F:2505:ASN:HA	2:F:2506:PRO:HD3	1.90	0.43
1:A:218:SER:HB2	1:A:261:CYS:HA	2.00	0.42
1:B:310:HIS:HB2	1:B:315:ILE:HB	2.02	0.42
1:B:310:HIS:HB3	1:B:313:GLU:O	2.20	0.42
2:D:2505:ASN:HB3	2:D:2513:GLU:OE1	2.20	0.41
1:B:81:LYS:HB3	1:B:81:LYS:HE3	1.96	0.41
1:A:69:ILE:HB	1:A:83:ILE:HB	2.01	0.41
1:A:172:ASP:HB2	1:A:192:ASP:HB3	2.02	0.41
1:A:315:ILE:HG12	1:A:331:LYS:HG2	2.02	0.41
1:A:38:LYS:HD3	1:A:38:LYS:HA	1.85	0.41
2:D:2505:ASN:HA	2:D:2506:PRO:HD3	1.94	0.40

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	302/312 (97%)	287 (95%)	15 (5%)	0	100	100
1	B	302/312 (97%)	287 (95%)	14 (5%)	1 (0%)	46	25
1	C	302/312 (97%)	286 (95%)	15 (5%)	1 (0%)	46	25
2	D	10/14 (71%)	10 (100%)	0	0	100	100
2	E	10/14 (71%)	10 (100%)	0	0	100	100
2	F	10/14 (71%)	10 (100%)	0	0	100	100
All	All	936/978 (96%)	890 (95%)	44 (5%)	2 (0%)	52	32

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	214	ASN
1	B	214	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	266/273 (97%)	265 (100%)	1 (0%)	93	90
1	B	266/273 (97%)	262 (98%)	4 (2%)	72	55
1	C	266/273 (97%)	262 (98%)	4 (2%)	72	55
2	D	7/10 (70%)	7 (100%)	0	100	100
2	E	7/10 (70%)	7 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	F	7/10 (70%)	7 (100%)	0	100	100
All	All	819/849 (96%)	810 (99%)	9 (1%)	80	66

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	181	ARG
1	B	81	LYS
1	B	149	PHE
1	B	211	ASP
1	B	214	ASN
1	C	149	PHE
1	C	212	ASP
1	C	214	ASN
1	C	247	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	289	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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 ⓘ

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	304/312 (97%)	0.13	11 (3%) 46 52	13, 20, 33, 49	1 (0%)
1	B	304/312 (97%)	0.10	9 (2%) 54 59	13, 20, 33, 47	1 (0%)
1	C	304/312 (97%)	0.12	11 (3%) 46 52	13, 20, 33, 45	1 (0%)
2	D	12/14 (85%)	0.86	3 (25%) 1 1	15, 25, 33, 38	0
2	E	12/14 (85%)	0.84	2 (16%) 2 4	15, 25, 33, 39	0
2	F	12/14 (85%)	0.78	2 (16%) 2 4	15, 25, 33, 40	0
All	All	948/978 (96%)	0.15	38 (4%) 42 48	13, 20, 33, 49	3 (0%)

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	334	CYS	5.6
1	B	334	CYS	4.7
1	A	212	ASP	4.6
1	C	334	CYS	4.6
1	B	31	VAL	4.4
1	B	212	ASP	3.8
2	F	2504	LEU	3.6
2	E	2504	LEU	3.4
1	C	212	ASP	3.3
2	D	2504	LEU	3.2
1	A	31	VAL	3.1
1	B	219	PHE	3.1
1	C	31	VAL	3.1
2	E	2515	TYR	3.0
2	F	2515	TYR	3.0
1	A	219	PHE	2.8
1	C	244	SER	2.8
2	D	2515	TYR	2.7
1	C	213	ASP	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	244	SER	2.6
1	C	219	PHE	2.5
1	C	263	PHE	2.5
1	A	32	LYS	2.5
1	A	263	PHE	2.5
1	A	159	LYS	2.4
1	C	220	VAL	2.4
1	C	159	LYS	2.4
1	A	220	VAL	2.4
1	B	220	VAL	2.4
1	C	247	LYS	2.3
1	B	333	ASP	2.3
1	A	333	ASP	2.2
1	B	213	ASP	2.2
1	C	32	LYS	2.1
1	A	181	ARG	2.1
1	B	181	ARG	2.1
1	B	32	LYS	2.0
2	D	2506	PRO	2.0

## 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 carbohydrates 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.