



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 5, 2020 – 02:08 PM JST

PDB ID : 7BU9  
Title : Crystal Structure of Spindlin1-H3(K4me3-K9me2) complex  
Authors : Zhao, F.; Li, H.  
Deposited on : 2020-04-05  
Resolution : 3.50 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.14.6  
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) : 2.14.6

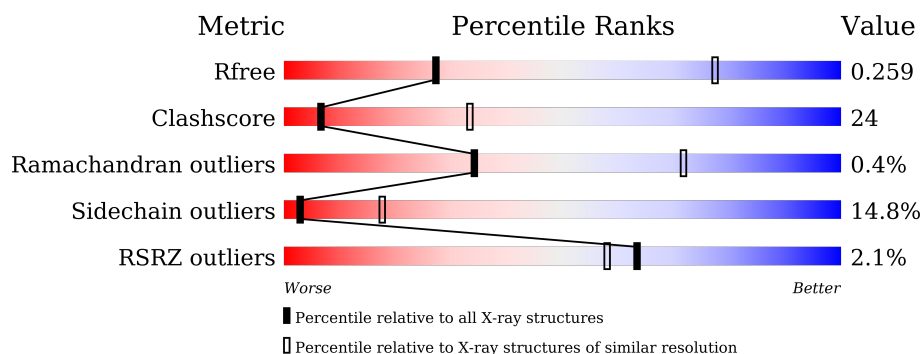
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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.




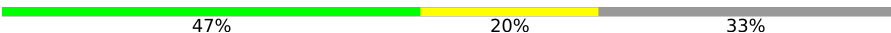
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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  $\leq 5\%$ . 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	220	<div> <div>2%</div> <div>56%</div> <div>31%</div> <div>6%</div> <div>7%</div> </div>
1	C	220	<div> <div>%</div> <div>61%</div> <div>26%</div> <div>6%</div> <div>7%</div> </div>
1	E	220	<div> <div>55%</div> <div>34%</div> <div>5%</div> <div>6%</div> </div>
1	G	220	<div> <div>5%</div> <div>55%</div> <div>27%</div> <div>8%</div> <div>9%</div> </div>
2	B	15	<div> <div>13%</div> <div>47%</div> <div>40%</div> </div>
2	D	15	<div> <div>27%</div> <div>33%</div> <div>7%</div> <div>33%</div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	15	
2	H	15	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	M3L	D	4	-	-	X	-
2	MLY	D	9	-	-	X	-
2	M3L	F	4	-	-	X	-
2	MLY	F	9	-	-	X	-
2	M3L	H	4	-	-	X	-

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6911 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 Spindlin-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	204	Total	C	N	O	S	0	0	0
			1646	1051	273	313	9			
1	C	205	Total	C	N	O	S	0	0	0
			1655	1056	274	316	9			
1	E	206	Total	C	N	O	S	0	0	0
			1662	1060	276	317	9			
1	G	200	Total	C	N	O	S	0	0	0
			1618	1035	269	305	9			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	43	GLY	-	expression tag	UNP Q9Y657
A	44	SER	-	expression tag	UNP Q9Y657
C	43	GLY	-	expression tag	UNP Q9Y657
C	44	SER	-	expression tag	UNP Q9Y657
E	43	GLY	-	expression tag	UNP Q9Y657
E	44	SER	-	expression tag	UNP Q9Y657
G	43	GLY	-	expression tag	UNP Q9Y657
G	44	SER	-	expression tag	UNP Q9Y657

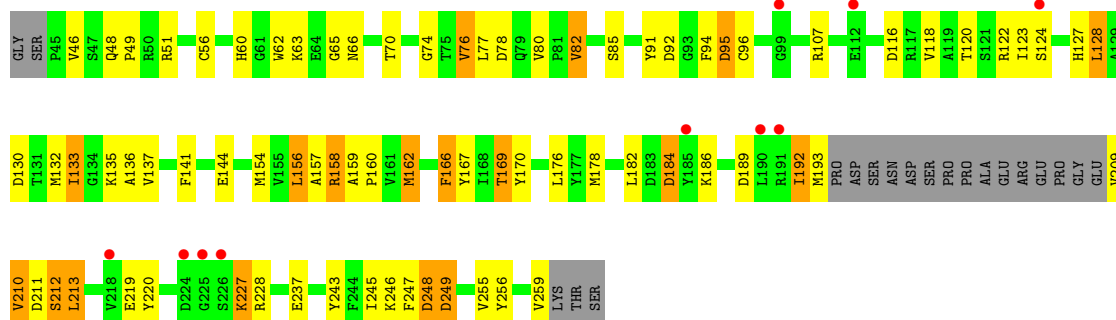
- Molecule 2 is a protein called H3(K4me3-K9me2) peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	9	Total	C	N	O	0	0	0
			78	48	18	12			
2	D	10	Total	C	N	O	0	0	0
			84	51	19	14			
2	F	10	Total	C	N	O	0	0	0
			84	51	19	14			
2	H	10	Total	C	N	O	0	0	0
			84	51	19	14			





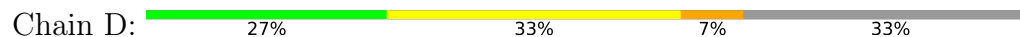
• Molecule 1: Spindlin-1



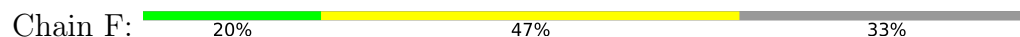
• Molecule 2: H3(K4me3-K9me2) peptide



• Molecule 2: H3(K4me3-K9me2) peptide



• Molecule 2: H3(K4me3-K9me2) peptide



• Molecule 2: H3(K4me3-K9me2) peptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	40.47Å 143.78Å 129.99Å 90.00° 95.86° 90.00°	Depositor
Resolution (Å)	48.07 – 3.50 48.07 – 3.50	Depositor EDS
% Data completeness (in resolution range)	98.7 (48.07-3.50) 84.6 (48.07-3.50)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.41 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.14_3247	Depositor
R, $R_{free}$	0.246 , 0.264 0.248 , 0.259	Depositor DCC
$R_{free}$ test set	876 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	71.6	Xtriage
Anisotropy	1.095	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 60.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.36$ , $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	6911	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.38% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MLY, M3L

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	0.40	0/1685	0.61	0/2283
1	C	0.35	0/1694	0.59	0/2296
1	E	0.35	0/1700	0.57	1/2302 (0.0%)
1	G	0.38	0/1656	0.61	0/2243
2	B	0.20	0/54	0.54	0/71
2	D	0.62	0/60	0.76	0/79
2	F	0.45	0/60	0.75	0/79
2	H	0.57	0/60	0.63	0/79
All	All	0.37	0/6969	0.60	1/9432 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	224	ASP	CB-CG-OD2	5.18	122.97	118.30

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	1646	0	1601	78	0
1	C	1655	0	1606	90	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1662	0	1620	91	0
1	G	1618	0	1581	84	0
2	B	78	0	94	11	0
2	D	84	0	99	26	0
2	F	84	0	100	29	0
2	H	84	0	99	10	0
All	All	6911	0	6800	324	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 24.

The worst 5 of 324 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:141:PHE:CE2	2:D:4:M3L:HM32	1.71	1.23
1:E:72:TRP:CH2	1:E:94:PHE:HE2	1.61	1.18
1:E:207:GLY:HA2	1:G:158:ARG:HG3	1.39	1.03
1:E:72:TRP:HH2	1:E:94:PHE:CE2	1.77	1.00
1:E:72:TRP:CH2	1:E:94:PHE:CE2	2.50	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	200/220 (91%)	178 (89%)	21 (10%)	1 (0%)	29	68
1	C	201/220 (91%)	182 (90%)	19 (10%)	0	100	100
1	E	202/220 (92%)	181 (90%)	20 (10%)	1 (0%)	29	68
1	G	196/220 (89%)	183 (93%)	12 (6%)	1 (0%)	29	68
2	B	6/15 (40%)	6 (100%)	0	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	6/15 (40%)	6 (100%)	0	0	100	100
2	F	6/15 (40%)	6 (100%)	0	0	100	100
2	H	6/15 (40%)	5 (83%)	1 (17%)	0	100	100
All	All	823/940 (88%)	747 (91%)	73 (9%)	3 (0%)	34	72

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	174	PRO
1	E	81	PRO
1	G	160	PRO

### 5.3.2 Protein sidechains ⓘ

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	180/194 (93%)	156 (87%)	24 (13%)	4	21
1	C	181/194 (93%)	158 (87%)	23 (13%)	4	22
1	E	182/194 (94%)	156 (86%)	26 (14%)	3	19
1	G	177/194 (91%)	145 (82%)	32 (18%)	1	9
2	B	5/8 (62%)	3 (60%)	2 (40%)	0	1
2	D	6/8 (75%)	4 (67%)	2 (33%)	0	1
2	F	6/8 (75%)	5 (83%)	1 (17%)	2	12
2	H	6/8 (75%)	6 (100%)	0	100	100
All	All	743/808 (92%)	633 (85%)	110 (15%)	3	17

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	2	ARG
1	E	162	MET
1	G	209	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	47	SER
1	E	108	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

8 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	MLY	D	9	2	9,10,11	0.44	0	6,11,13	0.19	0
2	MLY	B	9	2	9,10,11	0.46	0	6,11,13	0.97	0
2	M3L	H	4	2	10,11,12	0.45	0	9,14,16	0.10	0
2	M3L	F	4	2	10,11,12	0.54	0	9,14,16	0.11	0
2	MLY	H	9	2	9,10,11	0.42	0	6,11,13	0.22	0
2	M3L	D	4	2	10,11,12	0.55	0	9,14,16	0.13	0
2	MLY	F	9	2	9,10,11	0.42	0	6,11,13	0.15	0
2	M3L	B	4	2	10,11,12	0.55	0	9,14,16	0.56	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	MLY	D	9	2	-	4/8/9/11	-
2	MLY	B	9	2	-	1/8/9/11	-

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	M3L	H	4	2	-	4/9/10/12	-
2	M3L	F	4	2	-	5/9/10/12	-
2	MLY	H	9	2	-	4/8/9/11	-
2	M3L	D	4	2	-	6/9/10/12	-
2	MLY	F	9	2	-	4/8/9/11	-
2	M3L	B	4	2	-	0/9/10/12	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 28 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	H	4	M3L	N-CA-CB-CG
2	H	4	M3L	C-CA-CB-CG
2	H	9	MLY	C-CA-CB-CG
2	D	4	M3L	C-CA-CB-CG
2	F	9	MLY	CE-CD-CG-CB

There are no ring outliers.

8 monomers are involved in 65 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	9	MLY	10	0
2	B	9	MLY	6	0
2	H	4	M3L	7	0
2	F	4	M3L	8	0
2	H	9	MLY	2	0
2	D	4	M3L	11	0
2	F	9	MLY	18	0
2	B	4	M3L	3	0

## 5.5 Carbohydrates

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 ⓘ

### 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	204/220 (92%)	0.19	5 (2%) 57 51	30, 79, 128, 176	0
1	C	205/220 (93%)	0.11	3 (1%) 73 68	30, 78, 131, 181	0
1	E	206/220 (93%)	0.13	0 100 100	37, 74, 113, 181	0
1	G	200/220 (90%)	0.23	10 (5%) 28 25	30, 77, 132, 172	0
2	B	7/15 (46%)	-0.24	0 100 100	68, 84, 94, 101	0
2	D	8/15 (53%)	-0.04	0 100 100	79, 88, 108, 110	0
2	F	8/15 (53%)	-0.22	0 100 100	68, 75, 83, 96	0
2	H	8/15 (53%)	0.02	0 100 100	62, 82, 90, 94	0
All	All	846/940 (90%)	0.15	18 (2%) 63 58	30, 78, 129, 181	0

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	123	ILE	5.1
1	G	218	VAL	4.6
1	C	123	ILE	4.0
1	A	220	TYR	3.8
1	G	124	SER	3.6

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	MLY	B	9	11/12	0.73	0.40	74,85,100,103	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	MLY	D	9	11/12	0.89	0.44	96,105,114,117	0
2	M3L	F	4	12/13	0.90	0.48	42,83,90,95	0
2	MLY	H	9	11/12	0.90	0.34	60,69,91,101	0
2	MLY	F	9	11/12	0.92	0.27	62,73,91,92	0
2	M3L	D	4	12/13	0.93	0.31	55,73,97,103	0
2	M3L	H	4	12/13	0.93	0.38	39,67,86,91	0
2	M3L	B	4	12/13	0.94	0.43	42,66,78,81	0

### 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.