



# Full wwPDB X-ray Structure Validation Report

Apr 7, 2014 – 10:33 PM EDT

PDB ID : 4N4I  
Title : Crystal structure of the Bromo-PWWP of the mouse zinc finger MYND-type containing 11 isoform alpha in complex with histone H3.3K36me3  
Authors : Li, Y.; Ren, Y.; Li, H.  
Deposited on : 2013-10-08  
Resolution : 2.00 Å(reported)

This is a full wwPDB 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/ValidationPDFNotes.html>

---

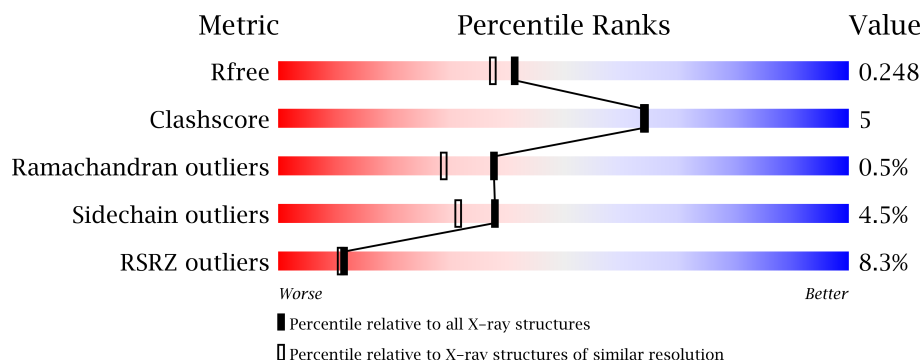
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : **FAILED**  
Xtriage (Phenix) : dev-1439  
EDS : stable22978  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22978

# 1 Overall quality at a glance

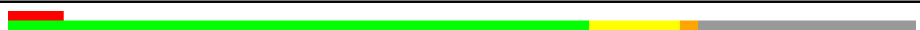

The reported resolution of this entry is 2.00 Å.

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))
$R_{free}$	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (2.00-2.00)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	252	
2	B	24	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
5	PEG	A	403	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 1818 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Zinc finger MYND domain-containing protein 11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	192	1611	1029	290	280	12	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	120	MET	-	EXPRESSION TAG	UNP Q8R5C8
A	121	GLY	-	EXPRESSION TAG	UNP Q8R5C8
A	122	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	123	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	124	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	125	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	126	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	127	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	128	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	129	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	130	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	131	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	132	GLY	-	EXPRESSION TAG	UNP Q8R5C8
A	133	LEU	-	EXPRESSION TAG	UNP Q8R5C8
A	134	VAL	-	EXPRESSION TAG	UNP Q8R5C8
A	135	PRO	-	EXPRESSION TAG	UNP Q8R5C8
A	136	ARG	-	EXPRESSION TAG	UNP Q8R5C8
A	137	GLY	-	EXPRESSION TAG	UNP Q8R5C8
A	138	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	139	HIS	-	EXPRESSION TAG	UNP Q8R5C8
A	140	MET	-	EXPRESSION TAG	UNP Q8R5C8
A	141	ALA	-	EXPRESSION TAG	UNP Q8R5C8
A	142	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	143	MET	-	EXPRESSION TAG	UNP Q8R5C8
A	144	THR	-	EXPRESSION TAG	UNP Q8R5C8
A	145	GLY	-	EXPRESSION TAG	UNP Q8R5C8
A	146	GLY	-	EXPRESSION TAG	UNP Q8R5C8

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	147	GLN	-	EXPRESSION TAG	UNP Q8R5C8
A	148	GLN	-	EXPRESSION TAG	UNP Q8R5C8
A	149	MET	-	EXPRESSION TAG	UNP Q8R5C8
A	150	GLY	-	EXPRESSION TAG	UNP Q8R5C8
A	151	ARG	-	EXPRESSION TAG	UNP Q8R5C8
A	152	GLY	-	EXPRESSION TAG	UNP Q8R5C8
A	153	SER	-	EXPRESSION TAG	UNP Q8R5C8
A	234	ALA	ASP	ENGINEERED MUTATION	UNP Q8R5C8
A	236	ALA	GLU	ENGINEERED MUTATION	UNP Q8R5C8

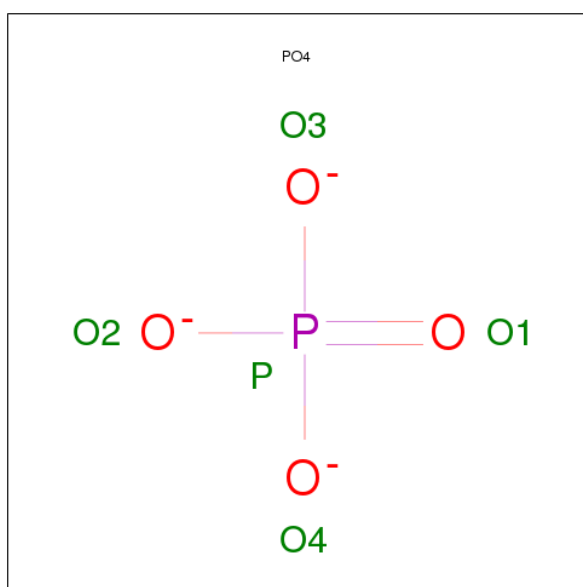
- Molecule 2 is a protein called Peptide from Histone H3.3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	0	0	0
			78	50	15	13			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

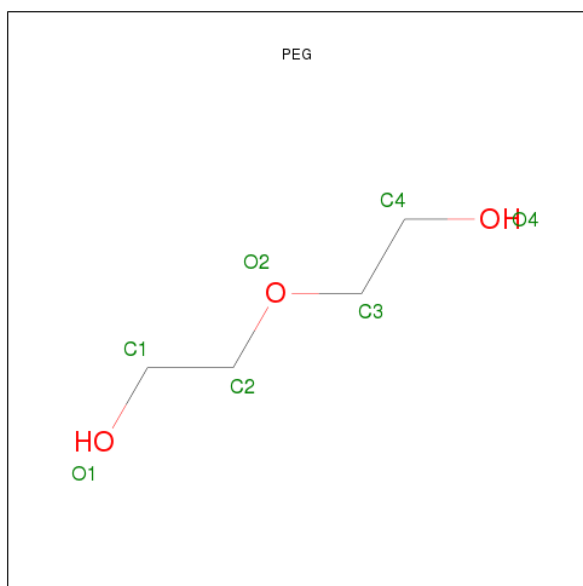
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	P	0	0
			5	4	1		

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 6 is water.

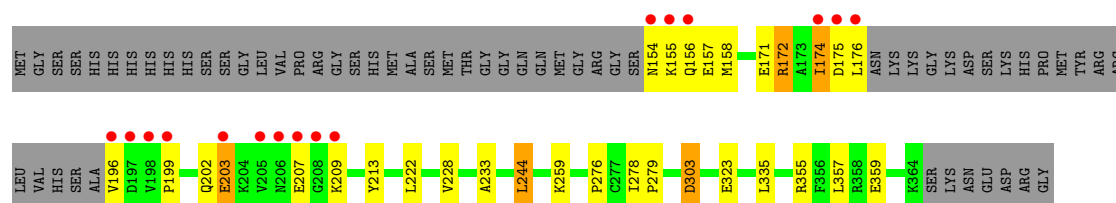
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	105	Total	O	0	0
			105	105		
6	B	11	Total	O	0	0
			11	11		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

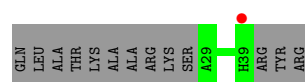
- Molecule 1: Zinc finger MYND domain-containing protein 11

Chain A: 



- Molecule 2: Peptide from Histone H3.3

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.01Å 64.01Å 122.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.45 – 2.00 36.45 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (36.45-2.00) 99.8 (36.45-2.00)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.98 (at 2.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1069)	Depositor
R, $R_{free}$	0.199 , 0.245 0.200 , 0.248	Depositor DCC
$R_{free}$ test set	917 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	31.5	Xtriage
Anisotropy	0.106	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 43.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 17984 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1818	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.27% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, PEG, PO4, 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	0.44	0/1654	0.58	0/2228
2	B	0.34	0/68	0.51	0/92
All	All	0.44	0/1722	0.58	0/2320

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Close contacts

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1611	0	1564	18	0
2	B	78	0	84	0	0
3	A	1	0	0	0	0
4	A	5	0	0	0	0
5	A	7	0	10	0	0
6	A	105	0	0	0	0
6	B	11	0	0	0	0
All	All	1818	0	1658	18	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including



hydrogens) of the entry. The overall clashscore for this entry is 5.

All (18) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:196:VAL:HG23	1:A:222:LEU:HD23	1.70	0.74
1:A:207:GLU:HB3	1:A:209:LYS:HE2	1.78	0.65
1:A:303:ASP:OD2	1:A:303:ASP:N	2.38	0.55
1:A:154:ASN:HB3	1:A:157:GLU:HB2	1.89	0.54
1:A:276:PRO:HB3	1:A:357:LEU:HD13	1.89	0.53
1:A:172:ARG:HB3	1:A:244:LEU:HD11	1.94	0.50
1:A:323:GLU:H	1:A:323:GLU:CD	2.16	0.49
1:A:199:PRO:O	1:A:203:GLU:N	2.44	0.48
1:A:175:ASP:O	1:A:176:LEU:HB2	2.14	0.48
1:A:199:PRO:HA	1:A:202:GLN:HB3	1.94	0.48
1:A:158:MET:HG3	1:A:213:TYR:CE1	2.50	0.47
1:A:228:VAL:CG2	1:A:233:ALA:HA	2.46	0.46
1:A:355:ARG:NH2	1:A:359:GLU:OE2	2.50	0.44
1:A:209:LYS:HA	1:A:209:LYS:HD3	1.65	0.44
1:A:171:GLU:O	1:A:174:ILE:HG13	2.19	0.42
1:A:199:PRO:O	1:A:203:GLU:HB2	2.20	0.42
1:A:196:VAL:CG2	1:A:222:LEU:HD23	2.46	0.42
1:A:278:ILE:HA	1:A:279:PRO:HA	1.89	0.41

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	188/252 (75%)	185 (98%)	2 (1%)	1 (0%)	38	29
2	B	8/24 (33%)	8 (100%)	0	0	100	100
All	All	196/276 (71%)	193 (98%)	2 (1%)	1 (0%)	38	29

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	174	ILE

### 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	171/220 (78%)	163 (95%)	8 (5%)	36	29
2	B	7/17 (41%)	7 (100%)	0	100	100
All	All	178/237 (75%)	170 (96%)	8 (4%)	38	31

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	155	LYS
1	A	156	GLN
1	A	172	ARG
1	A	203	GLU
1	A	244	LEU
1	A	259	LYS
1	A	303	ASP
1	A	335	LEU

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 chains in this entry.

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

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.8 Polymer linkage issues

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	192/252 (76%)	0.58	16 (8%) 11 11	15, 32, 66, 75	0
2	B	11/24 (45%)	0.33	1 (9%) 9 9	23, 27, 40, 46	0
All	All	203/276 (73%)	0.57	17 (8%) 11 10	15, 32, 66, 75	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	155	LYS	8.5
1	A	196	VAL	6.7
1	A	207	GLU	5.6
1	A	156	GLN	4.7
1	A	199	PRO	4.2
1	A	208	GLY	4.1
1	A	154	ASN	3.9
1	A	209	LYS	3.5
1	A	203	GLU	3.3
1	A	176	LEU	3.3
1	A	198	VAL	3.1
1	A	197	ASP	3.1
1	A	205	VAL	2.9
2	B	39	HIS	2.8
1	A	174	ILE	2.7
1	A	175	ASP	2.7
1	A	206	ASN	2.4

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	M3L	B	36	12/13	0.10	-0.48	20,23,27,28	0

### 6.3 Carbohydrates

There are no carbohydrates in this entry.

### 6.4 Ligands

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	PEG	A	403	7/7	0.23	3.84	34,38,43,46	0
3	ZN	A	401	1/1	0.09	-0.51	25,25,25,25	0
4	PO4	A	402	5/5	0.11	-0.76	34,35,42,45	0

### 6.5 Other polymers

There are no such residues in this entry.