



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2020 – 05:13 PM BST

PDB ID : 5R7X  
Title : PanDDA analysis group deposition of ground-state model of Human JMJD1B  
Authors : Snee, M.; Nowak, R.; Johansson, C.; Burgess-Brown, N.A.; Arrowsmith, C.H.;  
Bountra, C.; Edwards, A.M.; Oppermann, U.  
Deposited on : 2020-03-16  
Resolution : 1.44 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  
Xtriage (Phenix) : 1.13  
EDS : 2.13  
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) : 2.13

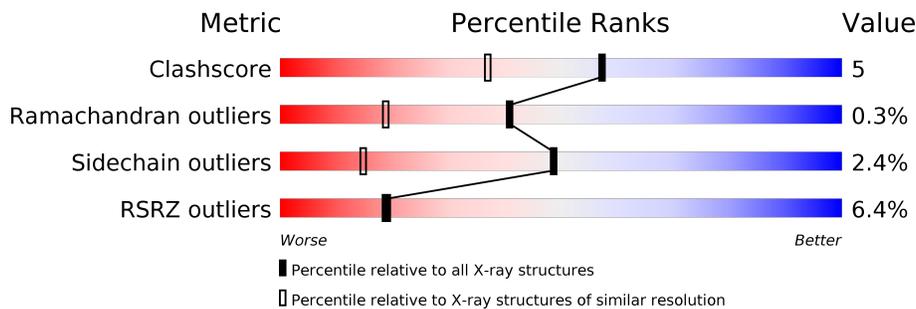
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.44 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))
Clashscore	141614	2086 (1.46-1.42)
Ramachandran outliers	138981	2047 (1.46-1.42)
Sidechain outliers	138945	2047 (1.46-1.42)
RSRZ outliers	127900	1993 (1.46-1.42)

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	372	 5% 83% 8% • 9%
1	B	372	 6% 83% 8% • 8%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6277 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 Lysine-specific demethylase 3B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	340	2765	1751	482	515	17	0	5	0
1	B	344	2798	1774	492	517	15	0	2	0

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1357	MET	-	initiating methionine	UNP Q7LBC6
A	1358	HIS	-	expression tag	UNP Q7LBC6
A	1359	HIS	-	expression tag	UNP Q7LBC6
A	1360	HIS	-	expression tag	UNP Q7LBC6
A	1361	HIS	-	expression tag	UNP Q7LBC6
A	1362	HIS	-	expression tag	UNP Q7LBC6
A	1363	HIS	-	expression tag	UNP Q7LBC6
A	1364	SER	-	expression tag	UNP Q7LBC6
A	1365	SER	-	expression tag	UNP Q7LBC6
A	1366	GLY	-	expression tag	UNP Q7LBC6
A	1367	VAL	-	expression tag	UNP Q7LBC6
A	1368	ASP	-	expression tag	UNP Q7LBC6
A	1369	LEU	-	expression tag	UNP Q7LBC6
A	1370	GLY	-	expression tag	UNP Q7LBC6
A	1371	THR	-	expression tag	UNP Q7LBC6
A	1372	GLU	-	expression tag	UNP Q7LBC6
A	1373	ASN	-	expression tag	UNP Q7LBC6
A	1374	LEU	-	expression tag	UNP Q7LBC6
A	1375	TYR	-	expression tag	UNP Q7LBC6
A	1376	PHE	-	expression tag	UNP Q7LBC6
A	1377	GLN	-	expression tag	UNP Q7LBC6
A	1378	SER	-	expression tag	UNP Q7LBC6
A	1379	MET	-	expression tag	UNP Q7LBC6
A	1601	GLU	GLN	conflict	UNP Q7LBC6
A	1606	HIS	GLY	conflict	UNP Q7LBC6

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1357	MET	-	initiating methionine	UNP Q7LBC6
B	1358	HIS	-	expression tag	UNP Q7LBC6
B	1359	HIS	-	expression tag	UNP Q7LBC6
B	1360	HIS	-	expression tag	UNP Q7LBC6
B	1361	HIS	-	expression tag	UNP Q7LBC6
B	1362	HIS	-	expression tag	UNP Q7LBC6
B	1363	HIS	-	expression tag	UNP Q7LBC6
B	1364	SER	-	expression tag	UNP Q7LBC6
B	1365	SER	-	expression tag	UNP Q7LBC6
B	1366	GLY	-	expression tag	UNP Q7LBC6
B	1367	VAL	-	expression tag	UNP Q7LBC6
B	1368	ASP	-	expression tag	UNP Q7LBC6
B	1369	LEU	-	expression tag	UNP Q7LBC6
B	1370	GLY	-	expression tag	UNP Q7LBC6
B	1371	THR	-	expression tag	UNP Q7LBC6
B	1372	GLU	-	expression tag	UNP Q7LBC6
B	1373	ASN	-	expression tag	UNP Q7LBC6
B	1374	LEU	-	expression tag	UNP Q7LBC6
B	1375	TYR	-	expression tag	UNP Q7LBC6
B	1376	PHE	-	expression tag	UNP Q7LBC6
B	1377	GLN	-	expression tag	UNP Q7LBC6
B	1378	SER	-	expression tag	UNP Q7LBC6
B	1379	MET	-	expression tag	UNP Q7LBC6
B	1601	GLU	GLN	conflict	UNP Q7LBC6
B	1606	HIS	GLY	conflict	UNP Q7LBC6

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Cl 1 1	0	0
2	A	4	Total Cl 4 4	0	0

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Mn 1 1	0	0
3	A	1	Total Mn 1 1	0	0

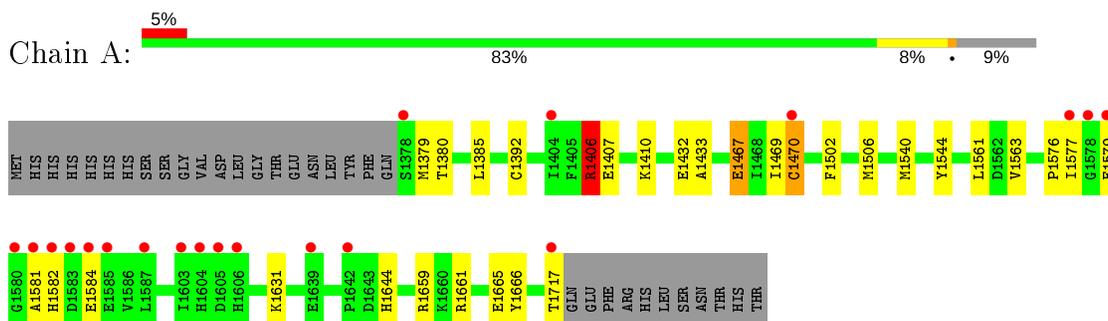
- Molecule 4 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	382	Total 382	O 382	0	0
4	B	325	Total 325	O 325	0	0

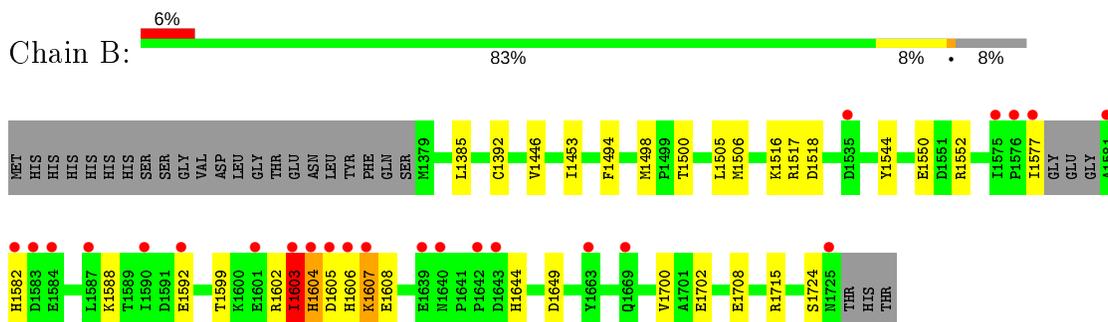
### 3 Residue-property plots [i](#)

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: Lysine-specific demethylase 3B



- Molecule 1: Lysine-specific demethylase 3B



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.81Å 93.79Å 93.25Å 90.00° 107.96° 90.00°	Depositor
Resolution (Å)	47.48 – 1.44 47.44 – 1.44	Depositor EDS
% Data completeness (in resolution range)	98.9 (47.48-1.44) 99.0 (47.44-1.44)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.44 (at 1.44Å)	Xtrriage
Refinement program	REFMAC 5.8.0258	Depositor
R, $R_{free}$	0.187 , 0.207 (Not available) , (Not available)	Depositor DCC
$R_{free}$ test set	8448 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.4	Xtrriage
Anisotropy	0.033	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 45.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.019 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6277	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.53% 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: MN, CL

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	0.94	2/2841 (0.1%)	1.02	7/3854 (0.2%)
1	B	0.89	2/2872 (0.1%)	0.91	1/3895 (0.0%)
All	All	0.91	4/5713 (0.1%)	0.97	8/7749 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1708	GLU	CD-OE2	7.01	1.33	1.25
1	A	1432	GLU	CD-OE1	6.53	1.32	1.25
1	A	1467	GLU	CD-OE2	-5.90	1.19	1.25
1	B	1702	GLU	CD-OE1	-5.33	1.19	1.25

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1540	MET	CG-SD-CE	-8.26	86.98	100.20
1	A	1470	CYS	N-CA-CB	7.76	124.58	110.60
1	A	1406[A]	ARG	CG-CD-NE	-5.53	100.19	111.80
1	A	1406[B]	ARG	CG-CD-NE	-5.53	100.19	111.80
1	A	1544	TYR	CB-CG-CD2	-5.50	117.70	121.00

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	2765	0	2651	23	0
1	B	2798	0	2685	30	0
2	A	4	0	0	0	0
2	B	1	0	0	1	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	382	0	0	10	1
4	B	325	0	0	7	0
All	All	6277	0	5336	53	1

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 5.

The worst 5 of 53 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:B:1516:LYS:HE3	4:B:1936:HOH:O	1.70	0.90
1:A:1385:LEU:HD12	1:A:1392[B]:CYS:SG	2.15	0.85
1:B:1715:ARG:HG3	4:B:2138:HOH:O	1.78	0.83
1:A:1407:GLU:CG	4:A:2207:HOH:O	2.29	0.80
1:A:1407:GLU:HG2	4:A:2207:HOH:O	1.84	0.75

All (1) 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)
4:A:2209:HOH:O	4:A:2262:HOH:O[2_648]	2.04	0.16

## 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	343/372 (92%)	334 (97%)	8 (2%)	1 (0%)	41	19
1	B	342/372 (92%)	332 (97%)	9 (3%)	1 (0%)	41	19
All	All	685/744 (92%)	666 (97%)	17 (2%)	2 (0%)	41	19

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1581	ALA
1	B	1603	ILE

### 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	299/331 (90%)	292 (98%)	7 (2%)	50	16
1	B	302/331 (91%)	294 (97%)	8 (3%)	46	12
All	All	601/662 (91%)	586 (98%)	15 (2%)	49	13

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1717	THR
1	B	1577	ILE
1	B	1605	ASP
1	A	1644	HIS
1	B	1604	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1640	ASN
1	A	1650	GLN
1	B	1656	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](#)

Of 7 ligands modelled in this entry, 7 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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	340/372 (91%)	0.29	20 (5%) 22 22	15, 25, 57, 117	0
1	B	344/372 (92%)	0.09	24 (6%) 16 16	17, 28, 58, 121	0
All	All	684/744 (91%)	0.19	44 (6%) 19 19	15, 27, 58, 121	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1577	ILE	13.0
1	A	1581	ALA	11.3
1	A	1578	GLY	10.2
1	B	1581	ALA	9.7
1	B	1605	ASP	9.1

### 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](#)

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( $\text{\AA}^2$ )	Q<0.9
3	MN	A	1805	1/1	0.92	0.07	43,43,43,43	0
2	CL	A	1804	1/1	0.94	0.06	42,42,42,42	0
2	CL	A	1803	1/1	0.96	0.05	36,36,36,36	0
2	CL	B	1801	1/1	0.97	0.05	41,41,41,41	0
3	MN	B	1802	1/1	0.99	0.03	28,28,28,28	0
2	CL	A	1802	1/1	0.99	0.06	37,37,37,37	0
2	CL	A	1801	1/1	1.00	0.11	21,21,21,21	0

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

There are no such residues in this entry.