



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 10:30 pm GMT

PDB ID : 4B3G  
Title : crystal structure of Ighmbp2 helicase in complex with RNA  
Authors : Lim, S.C.; Song, H.  
Deposited on : 2012-07-24  
Resolution : 2.85 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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

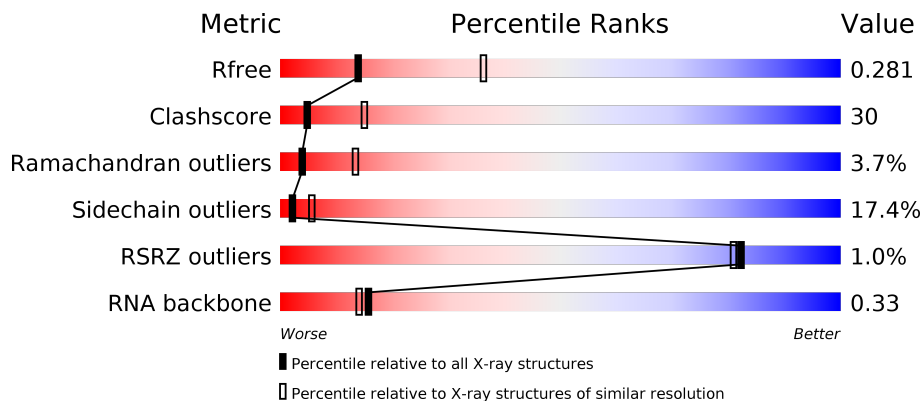
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.85 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))
$R_{free}$	100719	2469 (2.90-2.82)
Clashscore	112137	2749 (2.90-2.82)
Ramachandran outliers	110173	2687 (2.90-2.82)
Sidechain outliers	110143	2690 (2.90-2.82)
RSRZ outliers	101464	2487 (2.90-2.82)
RNA backbone	2435	1033 (3.20-2.52)

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	646	<div> <div>11%</div> <div>43%</div> <div>39%</div> <div>12%</div> <div>5%</div> </div>
1	B	646	<div> <div>49%</div> <div>35%</div> <div>10%</div> <div>6%</div> </div>
2	G	9	<div> <div>44%</div> <div>22%</div> <div>33%</div> </div>
2	H	9	<div> <div>11%</div> <div>11%</div> <div>56%</div> <div>11%</div> <div>11%</div> <div>11%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9979 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 DNA-BINDING PROTEIN SMUBP-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	614	Total	C	N	O	S	0	0	0
			4756	2998	854	889	15			
1	B	608	Total	C	N	O	S	0	0	0
			4704	2964	844	881	15			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	126	GLN	LEU	CONFLICT	UNP P38935
B	126	GLN	LEU	CONFLICT	UNP P38935

- Molecule 2 is a RNA chain called RNA (5'-(AP\*AP\*AP\*AP\*AP\*AP\*AP\*AP\*AP)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	9	Total	C	N	O	P	0	0	0
			198	90	45	54	9			
2	H	8	Total	C	N	O	P	0	0	0
			176	80	40	48	8			

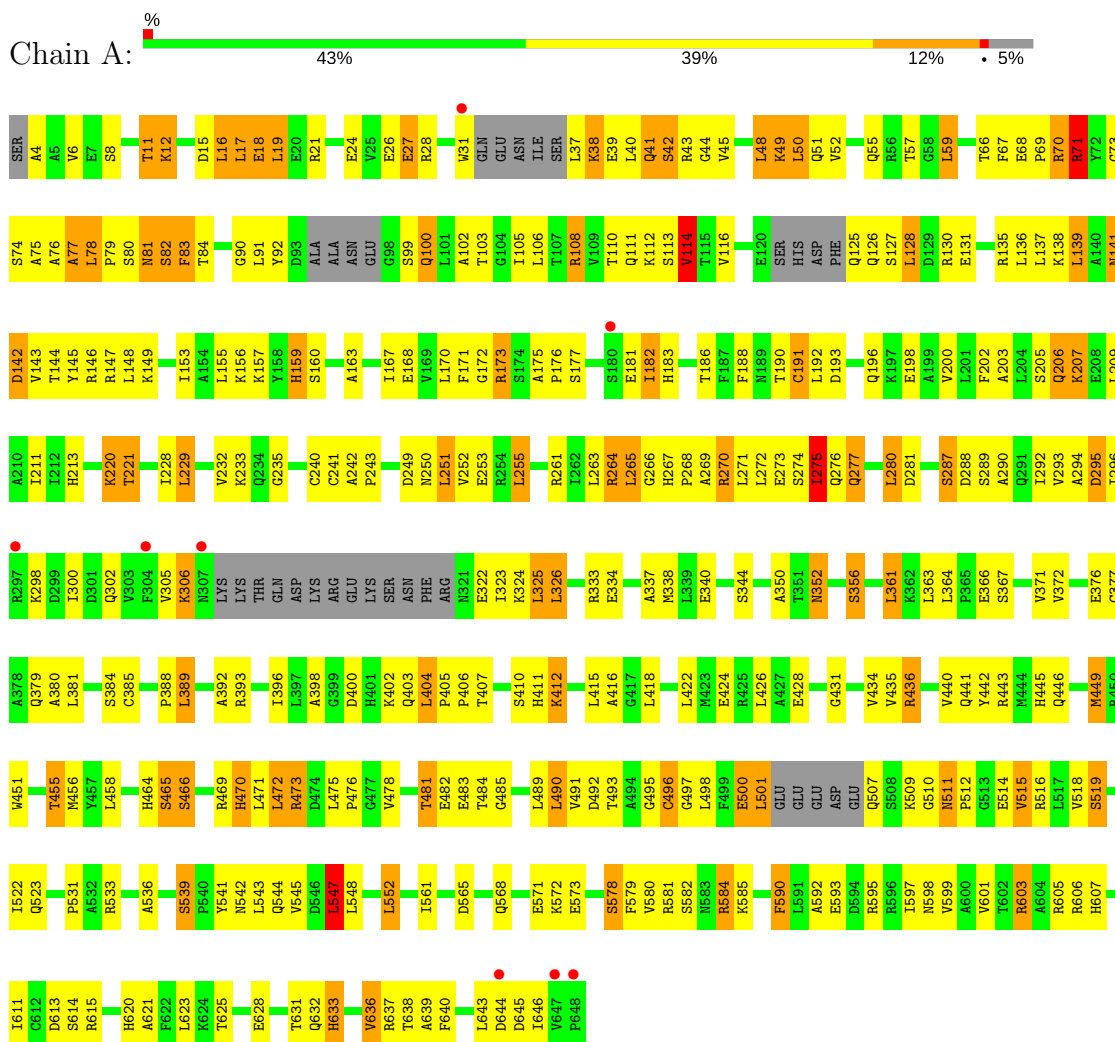
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	67	Total	O	0	0
			67	67		
3	B	69	Total	O	0	0
			69	69		
3	G	5	Total	O	0	0
			5	5		
3	H	4	Total	O	0	0
			4	4		

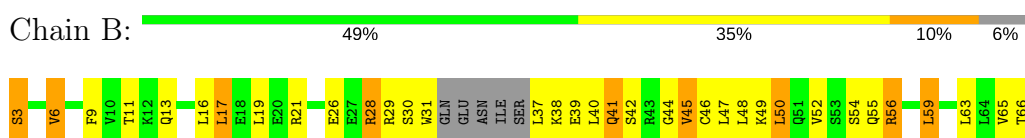
### 3 Residue-property plots

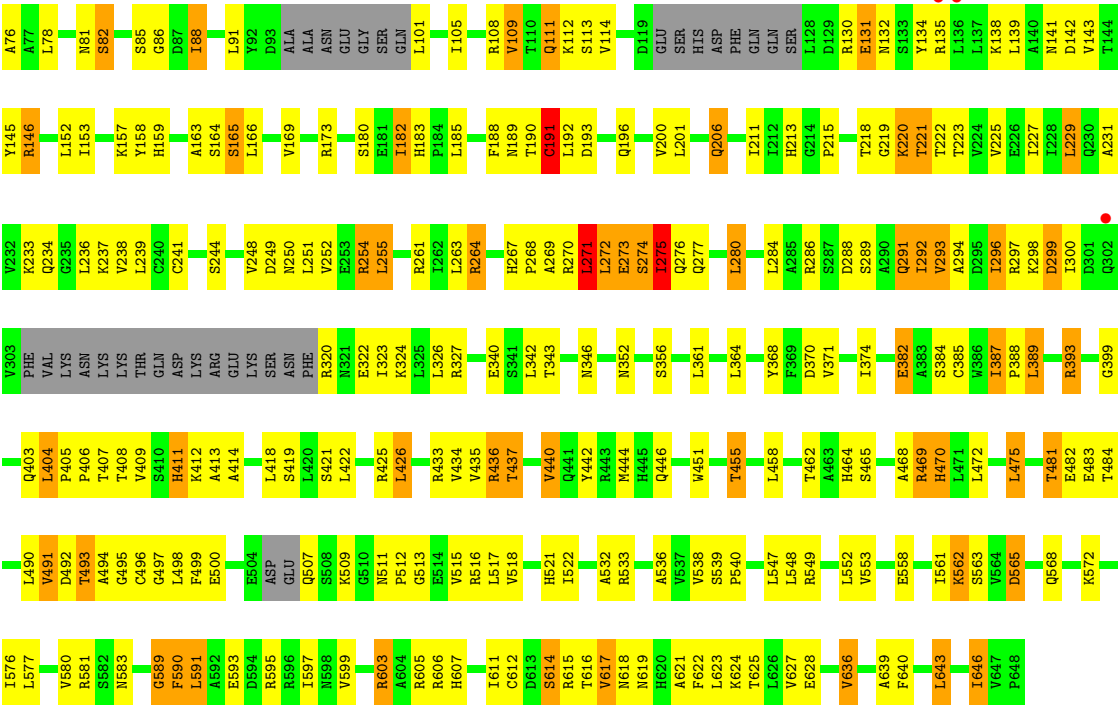
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: DNA-BINDING PROTEIN SMUBP-2



#### • Molecule 1: DNA-BINDING PROTEIN SMUBP-2

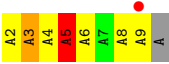
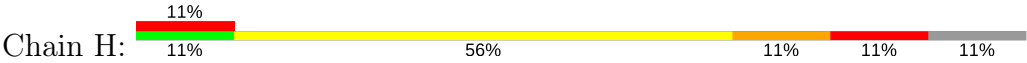




• Molecule 2: RNA (5'-(AP\*AP\*AP\*AP\*AP\*AP\*AP\*AP\*AP)-3')



• Molecule 2: RNA (5'-(AP\*AP\*AP\*AP\*AP\*AP\*AP\*AP\*AP)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.29Å 87.29Å 372.69Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.97 – 2.85 19.97 – 2.85	Depositor EDS
% Data completeness (in resolution range)	97.7 (19.97-2.85) 99.3 (19.97-2.85)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.63 (at 2.83Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.206 , 0.292 0.195 , 0.281	Depositor DCC
$R_{free}$ test set	1974 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	66.5	Xtriage
Anisotropy	0.208	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 46.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	0.075 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9979	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.08% 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

### 5.1 Standard geometry

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	0.45	0/4824	0.66	1/6528 (0.0%)
1	B	0.48	0/4771	0.67	0/6458
2	G	0.68	0/224	1.12	0/347
2	H	0.88	0/199	1.19	1/308 (0.3%)
All	All	0.48	0/10018	0.70	2/13641 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	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	59	LEU	CA-CB-CG	5.67	128.34	115.30
2	H	5	A	C8-N9-C4	-5.34	103.66	105.80

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	271	LEU	Peptide

### 5.2 Too-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 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	4756	0	4878	339	0
1	B	4704	0	4816	257	0
2	G	198	0	100	9	0
2	H	176	0	89	10	0
3	A	67	0	0	15	1
3	B	69	0	0	10	0
3	G	5	0	0	1	0
3	H	4	0	0	0	0
All	All	9979	0	9883	601	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 30.

The worst 5 of 601 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:254:ARG:NH1	1:B:254:ARG:HB3	1.61	1.15
1:B:70:ARG:HH11	1:B:70:ARG:HG2	1.12	1.14
1:B:481:THR:HG22	1:B:484:THR:H	1.15	1.11
1:B:469:ARG:HG2	1:B:469:ARG:HH11	1.12	1.07
1:A:76:ALA:H	1:A:77:ALA:HB2	1.20	1.07

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)
3:A:2058:HOH:O	3:A:2058:HOH:O[4_555]	2.16	0.04

## 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	602/646 (93%)	485 (81%)	87 (14%)	30 (5%)	2	7
1	B	596/646 (92%)	531 (89%)	51 (9%)	14 (2%)	7	24
All	All	1198/1292 (93%)	1016 (85%)	138 (12%)	44 (4%)	4	13

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	159	HIS
1	A	465	SER
1	A	466	SER
1	B	82	SER
1	B	271	LEU

### 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	518/551 (94%)	429 (83%)	89 (17%)	2	6
1	B	511/551 (93%)	421 (82%)	90 (18%)	2	5
All	All	1029/1102 (93%)	850 (83%)	179 (17%)	2	6

5 of 179 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	584	ARG
1	B	59	LEU
1	B	533	ARG
1	A	623	LEU
1	B	6	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	633	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	234	GLN
1	B	619	ASN
1	B	111	GLN
1	B	206	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	G	8/9 (88%)	3 (37%)	0
2	H	7/9 (77%)	5 (71%)	0
All	All	15/18 (83%)	8 (53%)	0

5 of 8 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	G	4	A
2	G	5	A
2	G	6	A
2	H	3	A
2	H	4	A

There are no RNA pucker outliers to report.

## 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 ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	614/646 (95%)	-0.28	8 (1%) 77 76	39, 63, 106, 143	0
1	B	608/646 (94%)	-0.42	3 (0%) 90 90	33, 53, 101, 126	0
2	G	9/9 (100%)	-0.10	0 100 100	58, 69, 107, 111	0
2	H	8/9 (88%)	-0.09	1 (12%) 4 3	49, 62, 86, 103	0
All	All	1239/1310 (94%)	-0.35	12 (0%) 82 81	33, 60, 103, 143	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	648	PRO	4.5
1	A	307	ASN	4.4
1	A	31	TRP	4.1
1	A	647	VAL	3.6
1	A	297	ARG	3.3

### 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.