



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 21, 2017 – 04:04 PM EST

PDB ID : 5LUS  
Title : Structures of DHBN domain of Pelecanus crispus BLM helicase  
Authors : Shi, J.; Chen, W.-F.; Zhang, B.; Fan, S.-H.; Ai, X.; Liu, N.-N.; Rety, S.; Xi, X.-G.  
Deposited on : 2016-09-09  
Resolution : 1.43 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:

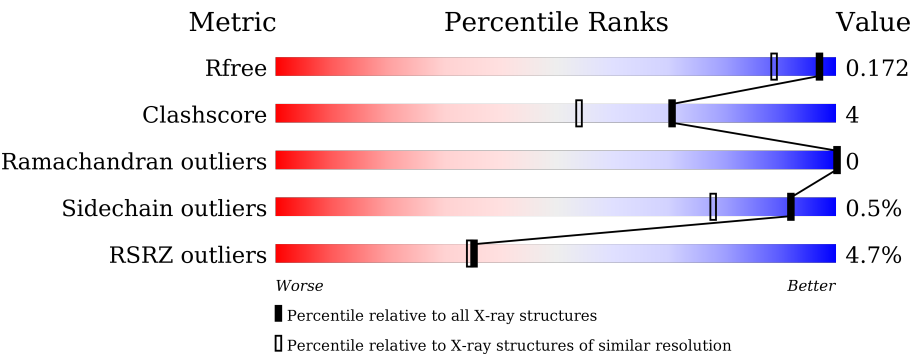
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:  
*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.43 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	1164 (1.46-1.42)
Clashscore	102246	1219 (1.46-1.42)
Ramachandran outliers	100387	1200 (1.46-1.42)
Sidechain outliers	100360	1200 (1.46-1.42)
RSRZ outliers	91569	1166 (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. 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	68	<div><div>4%</div><div><div></div><div>68%</div><div>26%</div></div><div></div></div>
1	B	68	<div><div>4%</div><div><div></div><div>66%</div><div>7%</div><div>26%</div></div><div></div></div>
1	C	68	<div><div></div><div><div></div><div>66%</div><div>7%</div><div>26%</div></div><div></div></div>
1	D	68	<div><div>3%</div><div><div></div><div>71%</div><div>28%</div></div><div></div></div>
1	E	68	<div><div>4%</div><div><div></div><div>69%</div><div>28%</div></div><div></div></div>
1	F	68	<div><div>%</div><div><div></div><div>69%</div><div>29%</div></div><div></div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	68	<div><div><div></div><div></div><div></div></div><div><div>3%</div><div>59%</div><div>12%</div><div>•</div><div>28%</div></div></div>
1	H	68	<div><div><div></div><div></div><div></div></div><div><div>3%</div><div>68%</div><div>•</div><div>28%</div></div></div>
1	I	68	<div><div><div></div><div></div><div></div></div><div><div>6%</div><div>72%</div><div>•</div><div>26%</div></div></div>
1	J	68	<div><div><div></div><div></div><div></div></div><div><div>4%</div><div>66%</div><div>7%</div><div>26%</div></div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8704 atoms, of which 4125 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 BLM helicase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	50	Total 815	C 251	H 417	N 70	O 72	S 5	0	0	0
1	B	50	Total 815	C 251	H 417	N 70	O 72	S 5	0	0	0
1	C	50	Total 816	C 251	H 418	N 70	O 72	S 5	0	0	0
1	D	49	Total 798	C 245	H 410	N 67	O 71	S 5	0	0	0
1	E	49	Total 798	C 245	H 410	N 67	O 71	S 5	0	0	0
1	F	48	Total 774	C 238	H 398	N 66	O 68	S 4	0	0	0
1	G	49	Total 798	C 245	H 410	N 67	O 71	S 5	0	0	0
1	H	49	Total 799	C 245	H 411	N 67	O 71	S 5	0	0	0
1	I	50	Total 815	C 251	H 417	N 70	O 72	S 5	0	0	0
1	J	50	Total 815	C 251	H 417	N 70	O 72	S 5	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	MET	CYS	conflict	UNP A0A091SV96
A	45	MET	CYS	conflict	UNP A0A091SV96
A	68	ASP	-	expression tag	UNP A0A091SV96
B	10	MET	CYS	conflict	UNP A0A091SV96
B	45	MET	CYS	conflict	UNP A0A091SV96
B	68	ASP	-	expression tag	UNP A0A091SV96
C	10	MET	CYS	conflict	UNP A0A091SV96
C	45	MET	CYS	conflict	UNP A0A091SV96
C	68	ASP	-	expression tag	UNP A0A091SV96

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	10	MET	CYS	conflict	UNP A0A091SV96
D	45	MET	CYS	conflict	UNP A0A091SV96
D	68	ASP	-	expression tag	UNP A0A091SV96
E	10	MET	CYS	conflict	UNP A0A091SV96
E	45	MET	CYS	conflict	UNP A0A091SV96
E	68	ASP	-	expression tag	UNP A0A091SV96
F	10	MET	CYS	conflict	UNP A0A091SV96
F	45	MET	CYS	conflict	UNP A0A091SV96
F	68	ASP	-	expression tag	UNP A0A091SV96
G	10	MET	CYS	conflict	UNP A0A091SV96
G	45	MET	CYS	conflict	UNP A0A091SV96
G	68	ASP	-	expression tag	UNP A0A091SV96
H	10	MET	CYS	conflict	UNP A0A091SV96
H	45	MET	CYS	conflict	UNP A0A091SV96
H	68	ASP	-	expression tag	UNP A0A091SV96
I	10	MET	CYS	conflict	UNP A0A091SV96
I	45	MET	CYS	conflict	UNP A0A091SV96
I	68	ASP	-	expression tag	UNP A0A091SV96
J	10	MET	CYS	conflict	UNP A0A091SV96
J	45	MET	CYS	conflict	UNP A0A091SV96
J	68	ASP	-	expression tag	UNP A0A091SV96

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	65	Total O 65 65	0	0
2	B	74	Total O 74 74	0	0
2	C	64	Total O 64 64	0	0
2	D	77	Total O 77 77	0	0
2	E	70	Total O 70 70	0	0
2	F	63	Total O 63 63	0	0
2	G	71	Total O 71 71	0	0
2	H	76	Total O 76 76	0	0
2	I	47	Total O 47 47	0	0

*Continued on next page...*

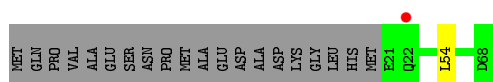
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	J	54	Total	O	0	0
			54	54		

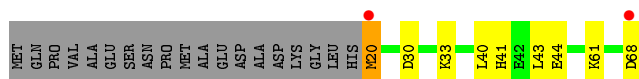


- Molecule 1: BLM helicase

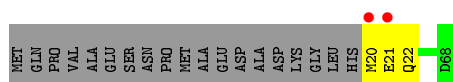




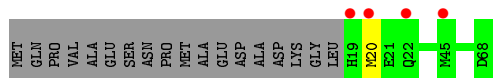
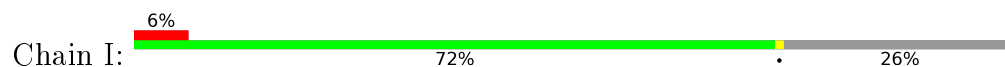
- Molecule 1: BLM helicase



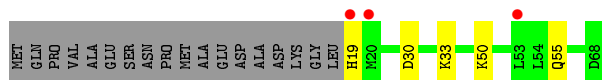
- Molecule 1: BLM helicase



- Molecule 1: BLM helicase



- Molecule 1: BLM helicase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.19Å 72.41Å 79.23Å 90.00° 99.44° 90.00°	Depositor
Resolution (Å)	36.20 – 1.43 78.16 – 1.43	Depositor EDS
% Data completeness (in resolution range)	82.9 (36.20-1.43) 83.0 (78.16-1.43)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.02 (at 1.43Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.157 , 0.172 0.157 , 0.172	Depositor DCC
$R_{free}$ test set	5247 reflections (4.98%)	DCC
Wilson B-factor (Å <sup>2</sup> )	20.1	Xtriage
Anisotropy	0.320	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 49.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	8704	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 7.81% 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](#)

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	0/401	0.69	1/538 (0.2%)
1	B	0.42	0/401	0.63	0/538
1	C	0.44	0/401	0.66	0/538
1	D	0.50	0/390	0.70	0/523
1	E	0.50	0/390	0.70	0/523
1	F	0.50	0/378	0.61	0/508
1	G	0.49	0/390	0.67	0/523
1	H	0.50	0/390	0.66	0/523
1	I	0.36	0/401	0.54	0/538
1	J	0.40	0/401	0.57	0/538
All	All	0.46	0/3943	0.65	1/5290 (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	A	36	ASP	CB-CG-OD1	5.19	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	398	417	417	3	0
1	B	398	417	417	6	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	398	418	417	4	0
1	D	388	410	410	2	0
1	E	388	410	410	3	0
1	F	376	398	397	1	0
1	G	388	410	410	6	0
1	H	388	411	410	5	0
1	I	398	417	417	1	0
1	J	398	417	417	4	0
2	A	65	0	0	2	0
2	B	74	0	0	6	0
2	C	64	0	0	4	0
2	D	77	0	0	1	2
2	E	70	0	0	2	0
2	F	63	0	0	2	0
2	G	71	0	0	2	2
2	H	76	0	0	2	1
2	I	47	0	0	0	0
2	J	54	0	0	4	1
All	All	4579	4125	4122	35	3

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 4.

The worst 5 of 35 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:50:LYS:NZ	2:C:101:HOH:O	1.91	1.03
1:B:19:HIS:N	2:B:102:HOH:O	1.95	0.99
1:B:50:LYS:NZ	2:B:101:HOH:O	1.93	0.97
1:C:45:MET:SD	2:D:106:HOH:O	2.35	0.84
2:C:147:HOH:O	1:D:45:MET:SD	2.36	0.82

All (3) 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)
2:H:173:HOH:O	2:J:149:HOH:O[2_556]	2.01	0.19
2:D:157:HOH:O	2:G:141:HOH:O[2_547]	2.05	0.15
2:D:145:HOH:O	2:G:166:HOH:O[2_547]	2.07	0.13

## 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	48/68 (71%)	48 (100%)	0	0	100	100
1	B	48/68 (71%)	48 (100%)	0	0	100	100
1	C	48/68 (71%)	48 (100%)	0	0	100	100
1	D	47/68 (69%)	47 (100%)	0	0	100	100
1	E	47/68 (69%)	47 (100%)	0	0	100	100
1	F	46/68 (68%)	46 (100%)	0	0	100	100
1	G	47/68 (69%)	47 (100%)	0	0	100	100
1	H	47/68 (69%)	46 (98%)	1 (2%)	0	100	100
1	I	48/68 (71%)	48 (100%)	0	0	100	100
1	J	48/68 (71%)	48 (100%)	0	0	100	100
All	All	474/680 (70%)	473 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

### 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	44/58 (76%)	43 (98%)	1 (2%)	58	19
1	B	44/58 (76%)	44 (100%)	0	100	100
1	C	44/58 (76%)	44 (100%)	0	100	100
1	D	43/58 (74%)	43 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	43/58 (74%)	43 (100%)	0	100	100
1	F	41/58 (71%)	41 (100%)	0	100	100
1	G	43/58 (74%)	42 (98%)	1 (2%)	58	19
1	H	43/58 (74%)	43 (100%)	0	100	100
1	I	44/58 (76%)	44 (100%)	0	100	100
1	J	44/58 (76%)	44 (100%)	0	100	100
All	All	433/580 (75%)	431 (100%)	2 (0%)	92	78

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	36	ASP
1	G	20	MET

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 ⓘ

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 ⓘ

### 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	50/68 (73%)	0.23	3 (6%) 25 25	18, 25, 65, 79	0
1	B	50/68 (73%)	0.37	3 (6%) 25 25	18, 27, 56, 91	0
1	C	50/68 (73%)	0.09	0 100 100	18, 26, 49, 67	0
1	D	49/68 (72%)	0.32	2 (4%) 41 41	17, 23, 53, 93	0
1	E	49/68 (72%)	0.34	3 (6%) 25 24	16, 25, 59, 93	0
1	F	48/68 (70%)	0.15	1 (2%) 67 68	16, 24, 52, 69	0
1	G	49/68 (72%)	0.19	2 (4%) 41 41	16, 23, 60, 67	0
1	H	49/68 (72%)	0.14	2 (4%) 41 41	17, 24, 49, 83	0
1	I	50/68 (73%)	0.46	4 (8%) 15 15	19, 31, 69, 99	0
1	J	50/68 (73%)	0.33	3 (6%) 25 25	20, 32, 67, 74	0
All	All	494/680 (72%)	0.26	23 (4%) 35 34	16, 26, 65, 99	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	19	HIS	9.4
1	D	20	MET	8.0
1	B	19	HIS	6.1
1	E	20	MET	5.7
1	I	20	MET	5.1

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

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.