



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 01:50 PM EST

PDB ID : 5N8R  
Title : Crystal Structure of Drosophila DHX36 helicase in complex with GAG-CACTGC  
Authors : Chen, W.-F.; Rety, S.; Hai-Lei Guo, H.-L.; Wu, W.-Q.; Liu, N.-N.; Liu, Q.-W.; Dai, Y.-X.; Xi, X.-G.  
Deposited on : 2017-02-24  
Resolution : 2.20 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 : rb-20030736  
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) : rb-20030736

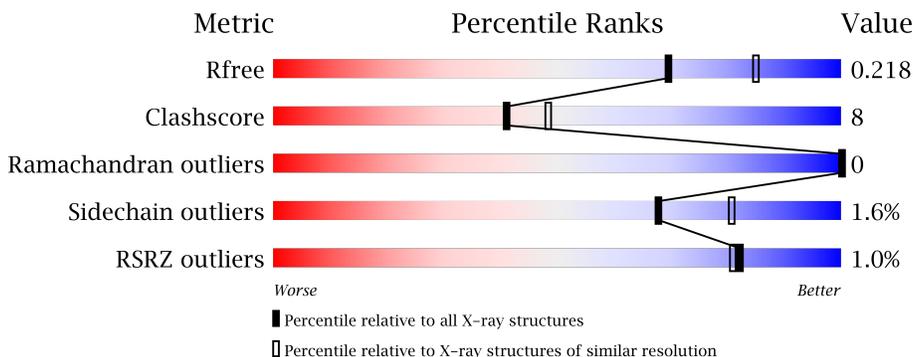
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4002 (2.20-2.20)
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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	944	 74% 15% 10%
1	B	944	 76% 13% 10%
2	C	9	 33% 56% 11%
2	D	9	 33% 67%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 14846 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 CG9323, isoform A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	849	6814	4304	1200	1265	45	0	0	0
1	B	851	6829	4312	1202	1270	45	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	943	VAL	-	expression tag	UNP Q8SWT2
A	944	ASP	-	expression tag	UNP Q8SWT2
B	943	VAL	-	expression tag	UNP Q8SWT2
B	944	ASP	-	expression tag	UNP Q8SWT2

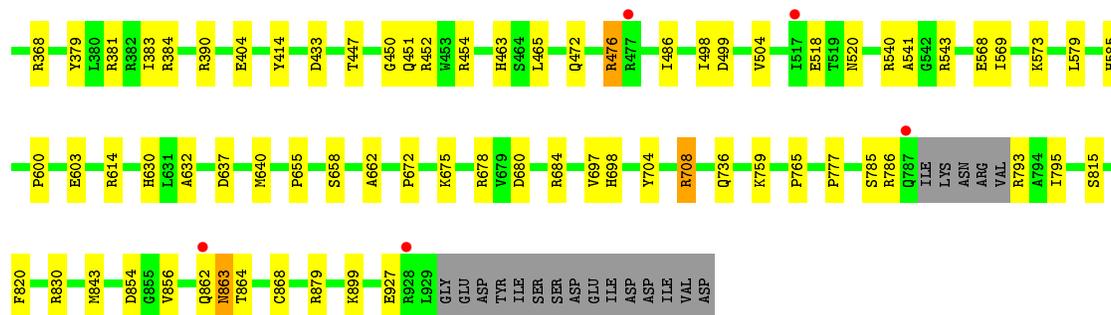
- Molecule 2 is a DNA chain called DNA (5'-D(P\*GP\*AP\*GP\*CP\*AP\*CP\*TP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	9	185	87	36	53	9	0	0	0
2	D	9	185	87	36	53	9	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	377	Total 377	O 377	0	0
3	B	424	Total 424	O 424	0	0
3	C	14	Total 14	O 14	0	0
3	D	18	Total 18	O 18	0	0

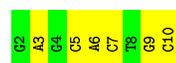




- Molecule 2: DNA (5'-D(P\*GP\*AP\*GP\*CP\*AP\*CP\*TP\*GP\*C)-3')



- Molecule 2: DNA (5'-D(P\*GP\*AP\*GP\*CP\*AP\*CP\*TP\*GP\*C)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	300.87Å 51.48Å 164.39Å 90.00° 114.22° 90.00°	Depositor
Resolution (Å)	149.90 – 2.20 149.92 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.3 (149.90-2.20) 99.3 (149.92-2.20)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 2.20Å)	Xtrriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, $R_{free}$	0.176 , 0.219 0.174 , 0.218	Depositor DCC
$R_{free}$ test set	5861 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	35.4	Xtrriage
Anisotropy	0.323	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 41.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	14846	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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	0.49	5/6933 (0.1%)	0.69	7/9353 (0.1%)
1	B	0.45	0/6950	0.66	1/9378 (0.0%)
2	C	1.00	0/207	0.97	1/317 (0.3%)
2	D	1.12	0/207	1.06	1/317 (0.3%)
All	All	0.50	5/14297 (0.0%)	0.69	10/19365 (0.1%)

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	A	0	2
1	B	0	1
All	All	0	3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	880	GLU	CD-OE1	-7.47	1.17	1.25
1	A	145	LYS	CD-CE	-7.29	1.33	1.51
1	A	684	ARG	CD-NE	-6.18	1.35	1.46
1	A	573	LYS	CD-CE	5.63	1.65	1.51
1	A	684	ARG	CZ-NH2	5.43	1.40	1.33

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	684	ARG	NE-CZ-NH1	-10.40	115.10	120.30
1	A	145	LYS	CD-CE-NZ	-10.01	88.69	111.70
1	A	571	ARG	CG-CD-NE	6.61	125.67	111.80
2	D	6	DA	O4'-C4'-C3'	-6.34	101.97	104.50
1	B	115	ARG	CD-NE-CZ	5.91	131.87	123.60

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	353	PHE	Peptide
1	A	783	ARG	Peptide
1	B	115	ARG	Sidechain

## 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	6814	0	6913	127	0
1	B	6829	0	6918	96	2
2	C	185	0	101	8	0
2	D	185	0	101	7	0
3	A	377	0	0	20	0
3	B	424	0	0	21	0
3	C	14	0	0	0	0
3	D	18	0	0	2	0
All	All	14846	0	14033	228	2

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 8.

The worst 5 of 228 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:A:142:ASN:N	1:A:145:LYS:HZ3	1.53	1.06
1:A:784:LYS:HG2	1:A:796:HIS:HA	1.38	1.05
1:A:476:ARG:NH1	3:A:1004:HOH:O	1.95	0.98
1:A:566:THR:O	1:A:571:ARG:NH1	1.97	0.96
1:A:927:GLU:O	3:A:1001:HOH:O	1.82	0.96

All (2) 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)
1:B:472:GLN:NE2	1:B:603:GLU:OE1[1_565]	2.13	0.07
1:B:476:ARG:NH1	1:B:518:GLU:OE2[1_565]	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	841/944 (89%)	820 (98%)	21 (2%)	0	100	100
1	B	843/944 (89%)	821 (97%)	22 (3%)	0	100	100
All	All	1684/1888 (89%)	1641 (97%)	43 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	757/842 (90%)	749 (99%)	8 (1%)	78	88
1	B	759/842 (90%)	742 (98%)	17 (2%)	57	70
All	All	1516/1684 (90%)	1491 (98%)	25 (2%)	68	81

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	123	ASN
1	B	146	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	863	ASN
1	B	137	ARG
1	B	301	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	94	GLN
1	B	123	ASN
1	B	536	GLN
1	A	758	ASN
1	B	463	HIS

### 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 carbohydrates 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 [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	849/944 (89%)	-0.35	9 (1%) 80 79	23, 41, 75, 108	0
1	B	851/944 (90%)	-0.39	8 (0%) 84 83	24, 39, 70, 97	0
2	C	9/9 (100%)	-0.50	0 100 100	37, 49, 60, 72	0
2	D	9/9 (100%)	-0.45	0 100 100	35, 48, 64, 81	0
All	All	1718/1906 (90%)	-0.37	17 (0%) 82 81	23, 40, 73, 108	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	795	ILE	4.8
1	A	785	SER	3.8
1	B	142	ASN	3.5
1	A	355	LYS	3.2
1	B	517	ILE	3.1

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.