



wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 08:05 am BST

PDB ID : 5N8U
Title : Crystal Structure of Drosophila DHX36 helicase in complex with CTCTCCT
Authors : Chen, W.-F.; Rety, S.; Guo, H.-L.; Wu, W.-Q.; Liu, N.-N.; Liu, Q.-W.; Dai, Y.-X.; Xi, X.-G.
Deposited on : 2017-02-24
Resolution : 2.62 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

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
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

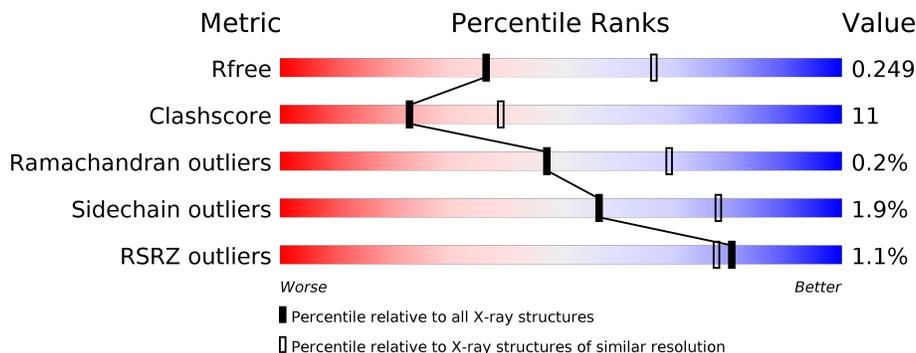
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.62 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}	130704	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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 ≥ 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	944	 % 66% 22% • 10%
1	B	944	 % 68% 20% • 10%
2	C	8	 63% 38%
2	D	8	 63% 38%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 14110 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	849	6814	4304	1200	1265	45	0	0	0

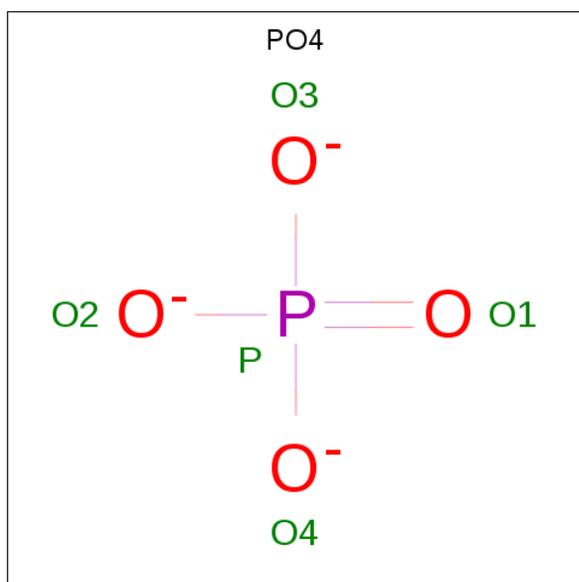
There are 4 discrepancies between the modelled and reference sequences:

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*CP*TP*CP*TP*CP*CP*CP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	8	155	75	21	51	8	0	0	0
2	D	8	155	75	21	51	8	0	0	0

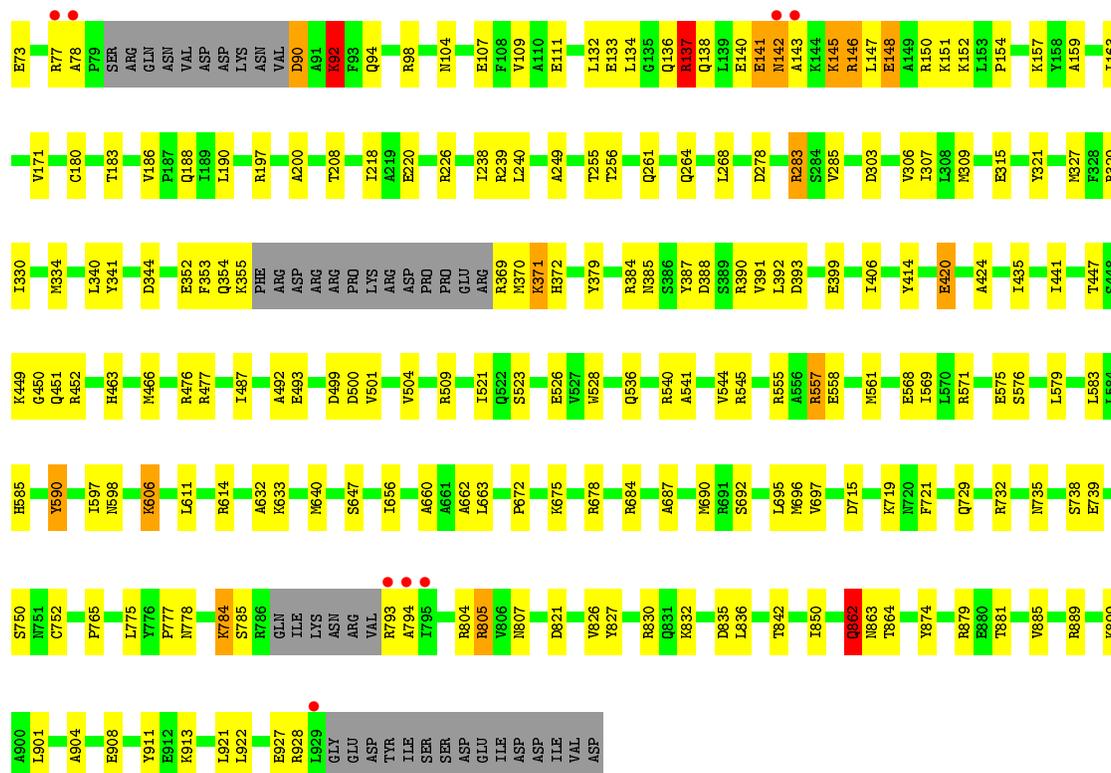
- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



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

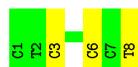
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	77	Total	O	0	0
			77	77		
4	B	80	Total	O	0	0
			80	80		
4	C	1	Total	O	0	0
			1	1		
4	D	4	Total	O	0	0
			4	4		



- Molecule 2: DNA (5'-D(P*CP*TP*CP*TP*CP*CP*CP*T)-3')

Chain C: 63% 38%



- Molecule 2: DNA (5'-D(P*CP*TP*CP*TP*CP*CP*CP*T)-3')

Chain D: 63% 38%



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	302.37Å 51.29Å 164.10Å 90.00° 114.70° 90.00°	Depositor
Resolution (Å)	56.30 – 2.62 56.38 – 2.62	Depositor EDS
% Data completeness (in resolution range)	98.3 (56.30-2.62) 98.3 (56.38-2.62)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.11 (at 2.61Å)	Xtrriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.182 , 0.249 0.183 , 0.249	Depositor DCC
R_{free} test set	3434 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	57.5	Xtrriage
Anisotropy	0.292	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 47.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	14110	wwPDB-VP
Average B, all atoms (Å ²)	66.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.37% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: PO4

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	0.64	14/6933 (0.2%)	0.78	10/9353 (0.1%)
1	B	0.53	5/6933 (0.1%)	0.82	18/9353 (0.2%)
2	C	1.13	0/170	1.08	0/258
2	D	1.21	1/170 (0.6%)	1.08	0/258
All	All	0.61	20/14206 (0.1%)	0.81	28/19222 (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	B	0	5

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	140	GLU	CD-OE1	-13.65	1.10	1.25
1	A	140	GLU	CD-OE2	-13.17	1.11	1.25
1	A	783	ARG	NE-CZ	-9.80	1.20	1.33
1	A	783	ARG	CZ-NH1	-9.61	1.20	1.33
1	A	434	LYS	CD-CE	-9.01	1.28	1.51

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	477	ARG	NE-CZ-NH1	13.37	126.98	120.30
1	B	477	ARG	CG-CD-NE	-10.19	90.40	111.80
1	B	477	ARG	NE-CZ-NH2	-9.92	115.34	120.30
1	B	476	ARG	NE-CZ-NH1	9.70	125.15	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	334	MET	CG-SD-CE	-9.44	85.10	100.20

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	137	ARG	Peptide
1	B	142	ASN	Peptide
1	B	143	ALA	Peptide
1	B	784	LYS	Peptide
1	B	862	GLN	Peptide

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	176	0
1	B	6814	0	6913	141	0
2	C	155	0	92	4	0
2	D	155	0	92	2	0
3	A	5	0	0	0	0
3	B	5	0	0	0	0
4	A	77	0	0	10	0
4	B	80	0	0	9	0
4	C	1	0	0	0	0
4	D	4	0	0	0	0
All	All	14110	0	14010	319	0

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 11.

The worst 5 of 319 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:O	1:A:146:ARG:NH2	1.73	1.21

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:52:ILE:N	4:A:1101:HOH:O	1.86	1.08
1:B:590:TYR:OH	1:B:606:LYS:HD3	1.61	1.01
1:A:515:TYR:OH	1:A:520:ASN:ND2	1.94	0.99
1:B:137:ARG:HA	1:B:140:GLU:OE1	1.69	0.92

There are no symmetry-related clashes.

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%)	814 (97%)	27 (3%)	0	100	100
1	B	841/944 (89%)	819 (97%)	19 (2%)	3 (0%)	34	55
All	All	1682/1888 (89%)	1633 (97%)	46 (3%)	3 (0%)	47	69

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	371	LYS
1	B	862	GLN
1	B	420	GLU

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%)	740 (98%)	17 (2%)	52	74
1	B	757/842 (90%)	745 (98%)	12 (2%)	62	81
All	All	1514/1684 (90%)	1485 (98%)	29 (2%)	57	78

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	689	ASN
1	A	901	LEU
1	B	805	ARG
1	A	783	ARG
1	B	92	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	720	ASN
1	A	729	GLN
1	B	188	GLN
1	A	698	HIS
1	B	209	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	PO4	A	1001	-	4,4,4	1.02	0	6,6,6	0.63	0
3	PO4	B	1001	-	4,4,4	1.00	0	6,6,6	0.77	0

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, 95th 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 ²)	Q<0.9
1	A	849/944 (89%)	-0.13	11 (1%) 77 73	37, 62, 99, 140	0
1	B	849/944 (89%)	-0.17	8 (0%) 84 82	39, 61, 96, 153	0
2	C	8/8 (100%)	0.27	0 100 100	78, 92, 117, 117	0
2	D	8/8 (100%)	0.39	0 100 100	66, 86, 113, 114	0
All	All	1714/1904 (90%)	-0.15	19 (1%) 80 78	37, 62, 99, 153	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	785	SER	5.5
1	A	355	LYS	5.1
1	B	142	ASN	4.4
1	A	795	ILE	4.3
1	A	142	ASN	3.5

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](#)

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, 95th 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(\AA^2)	Q<0.9
3	PO4	A	1001	5/5	0.74	0.32	161,165,170,179	0
3	PO4	B	1001	5/5	0.91	0.23	126,139,143,155	0

6.5 Other polymers [i](#)

There are no such residues in this entry.