wwPDB X-ray Structure Validation Summary Report (i)

Aug 21, 2023 – 10:54 AM EDT

PDB ID	:	2P6N
Title	:	Human DEAD-box RNA helicase DDX41, helicase domain
Authors	:	Karlberg, T.; Ogg, D.; Arrowsmith, C.H.; Berglund, H.; Busam, R.D.; Collins, R.; Dahlgren, L.G.; Edwards, A.; Flodin, S.; Flores, A.; Graslund, S.; Hallberg, B.M.; Hammarstrom, M.; Johansson, I.; Kotenyova, T.; Lehtio, L.; Moche, M.; Nordlund, P.; Nyman, T.; Persson, C.; Sagemark, J.; Stenmark, P.; Sundstrom, M.; Thorsell, A.G.; Van Den Berg, S.; Weigelt, J.; Holmberg-Schiavone, L.; Structural Genomics Consortium (SGC)
Deposited on	:	2007-03-19
Resolution	:	2.60 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

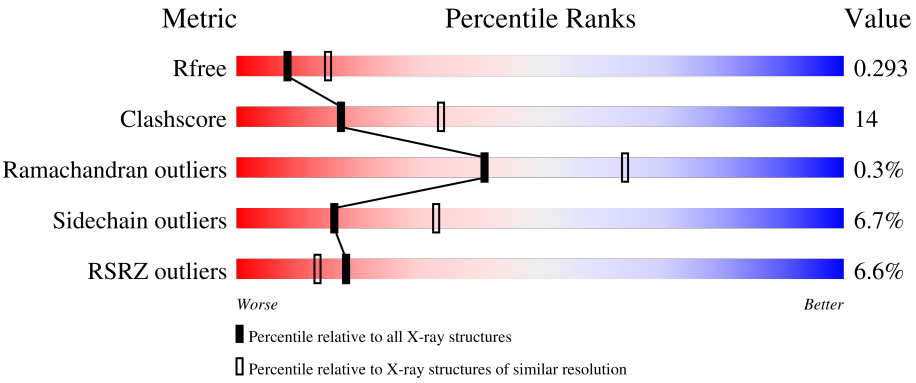
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)

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 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 of 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 <=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	191	6% 59% 25% · 16%
1	B	191	5% 59% 21% · 17%

Validation Pipeline (wwPDB-VP) : 2.35

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2481 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 ATP-dependent RNA helicase DDX41.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	160	Total	C	N	O	S	0	0	0
			1252	802	206	237	7			
1	B	158	Total	C	N	O	S	0	0	0
			1229	788	202	233	6			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	379	MET	-	cloning artifact	UNP Q9UJV9
A	380	HIS	-	cloning artifact	UNP Q9UJV9
A	381	HIS	-	cloning artifact	UNP Q9UJV9
A	382	HIS	-	cloning artifact	UNP Q9UJV9
A	383	HIS	-	cloning artifact	UNP Q9UJV9
A	384	HIS	-	cloning artifact	UNP Q9UJV9
A	385	HIS	-	cloning artifact	UNP Q9UJV9
A	386	SER	-	cloning artifact	UNP Q9UJV9
A	387	SER	-	cloning artifact	UNP Q9UJV9
A	388	GLY	-	cloning artifact	UNP Q9UJV9
A	389	VAL	-	cloning artifact	UNP Q9UJV9
A	390	ASP	-	cloning artifact	UNP Q9UJV9
A	391	LEU	-	cloning artifact	UNP Q9UJV9
A	392	GLY	-	cloning artifact	UNP Q9UJV9
A	393	THR	-	cloning artifact	UNP Q9UJV9
A	394	GLU	-	cloning artifact	UNP Q9UJV9
A	395	ASN	-	cloning artifact	UNP Q9UJV9
A	396	LEU	-	cloning artifact	UNP Q9UJV9
A	397	TYR	-	cloning artifact	UNP Q9UJV9
A	398	PHE	-	cloning artifact	UNP Q9UJV9
A	399	GLN	-	cloning artifact	UNP Q9UJV9
A	400	SER	-	cloning artifact	UNP Q9UJV9
A	401	MET	-	cloning artifact	UNP Q9UJV9
A	525	CYS	ARG	engineered mutation	UNP Q9UJV9
B	379	MET	-	cloning artifact	UNP Q9UJV9

Continued on next page...

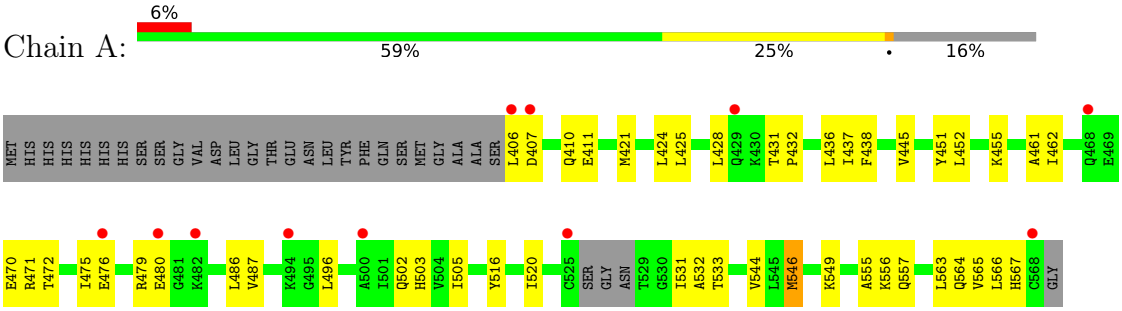
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	380	HIS	-	cloning artifact	UNP Q9UJV9
B	381	HIS	-	cloning artifact	UNP Q9UJV9
B	382	HIS	-	cloning artifact	UNP Q9UJV9
B	383	HIS	-	cloning artifact	UNP Q9UJV9
B	384	HIS	-	cloning artifact	UNP Q9UJV9
B	385	HIS	-	cloning artifact	UNP Q9UJV9
B	386	SER	-	cloning artifact	UNP Q9UJV9
B	387	SER	-	cloning artifact	UNP Q9UJV9
B	388	GLY	-	cloning artifact	UNP Q9UJV9
B	389	VAL	-	cloning artifact	UNP Q9UJV9
B	390	ASP	-	cloning artifact	UNP Q9UJV9
B	391	LEU	-	cloning artifact	UNP Q9UJV9
B	392	GLY	-	cloning artifact	UNP Q9UJV9
B	393	THR	-	cloning artifact	UNP Q9UJV9
B	394	GLU	-	cloning artifact	UNP Q9UJV9
B	395	ASN	-	cloning artifact	UNP Q9UJV9
B	396	LEU	-	cloning artifact	UNP Q9UJV9
B	397	TYR	-	cloning artifact	UNP Q9UJV9
B	398	PHE	-	cloning artifact	UNP Q9UJV9
B	399	GLN	-	cloning artifact	UNP Q9UJV9
B	400	SER	-	cloning artifact	UNP Q9UJV9
B	401	MET	-	cloning artifact	UNP Q9UJV9
B	525	CYS	ARG	engineered mutation	UNP Q9UJV9

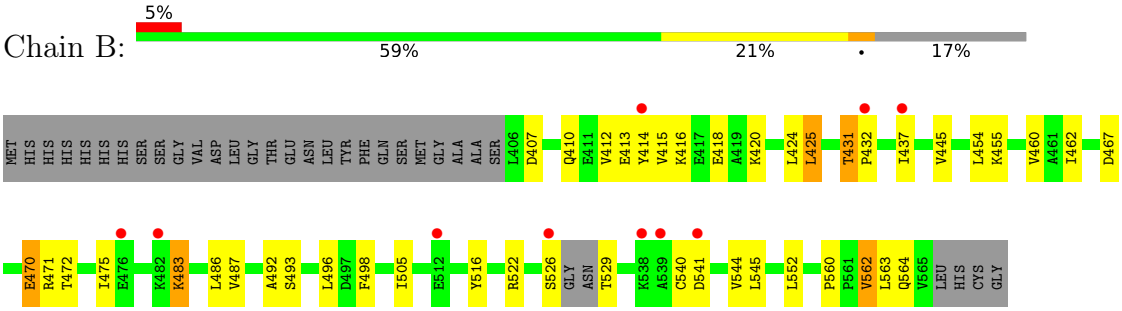
3 Residue-property plots (i)

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: ATP-dependent RNA helicase DDX41



• Molecule 1: ATP-dependent RNA helicase DDX41



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	68.01Å 68.01Å 305.60Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.72 – 2.60 29.72 – 2.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (29.72-2.60) 100.0 (29.72-2.50)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.03	Depositor
$< I/\sigma(I) >$ [1]	7.57 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.3.0032	Depositor
R, R_{free}	0.243 , 0.294 0.247 , 0.293	Depositor DCC
R_{free} test set	776 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	43.1	Xtriage
Anisotropy	0.476	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 70.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	2481	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 49.87 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.0118e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.75	0/1271	0.77	0/1718
1	B	0.72	0/1247	0.76	0/1686
All	All	0.74	0/2518	0.76	0/3404

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1252	0	1284	40	0
1	B	1229	0	1259	32	0
All	All	2481	0	2543	70	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 14.

The worst 5 of 70 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:462:ILE:HG22	1:B:492:ALA:HB1	1.43	0.99
1:A:546:MET:HE2	1:A:566:LEU:HD23	1.45	0.98

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:425:LEU:HD11	1:B:455:LYS:HG2	1.45	0.98
1:A:431:THR:HG22	1:A:432:PRO:O	1.65	0.96
1:B:407:ASP:OD2	1:B:529:THR:HG22	1.75	0.85

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	156/191 (82%)	146 (94%)	10 (6%)	0	100 100
1	B	154/191 (81%)	148 (96%)	5 (3%)	1 (1%)	25 47
All	All	310/382 (81%)	294 (95%)	15 (5%)	1 (0%)	41 64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	562	VAL

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	137/161 (85%)	130 (95%)	7 (5%)	24 46
1	B	133/161 (83%)	122 (92%)	11 (8%)	11 22

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	270/322 (84%)	252 (93%)	18 (7%)	16 33

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	526	SER
1	B	564	GLN
1	B	544	VAL
1	B	425	LEU
1	B	483	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	410	GLN
1	A	567	HIS
1	B	410	GLN
1	B	449	HIS

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 monosaccharides 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^{th} 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	160/191 (83%)	0.47	11 (6%)	16	12	51, 57, 65, 67	0
1	B	158/191 (82%)	0.39	10 (6%)	20	15	51, 57, 64, 66	0
All	All	318/382 (83%)	0.43	21 (6%)	18	13	51, 57, 64, 67	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	525	CYS	11.0
1	A	406	LEU	6.8
1	A	407	ASP	4.4
1	A	500	ALA	4.1
1	B	476	GLU	3.8

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

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.