



wwPDB X-ray Structure Validation Summary Report ⓘ

May 12, 2020 – 11:45 pm BST

PDB ID : 3I8E
Title : Crystal Structure of DDB1 in Complex with the H-Box Motif of WDR42A
Authors : Li, T.; Robert, E.I.; Breugel, P.C.V.; Strubin, M.; Zheng, N.
Deposited on : 2009-07-09
Resolution : 3.40 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
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

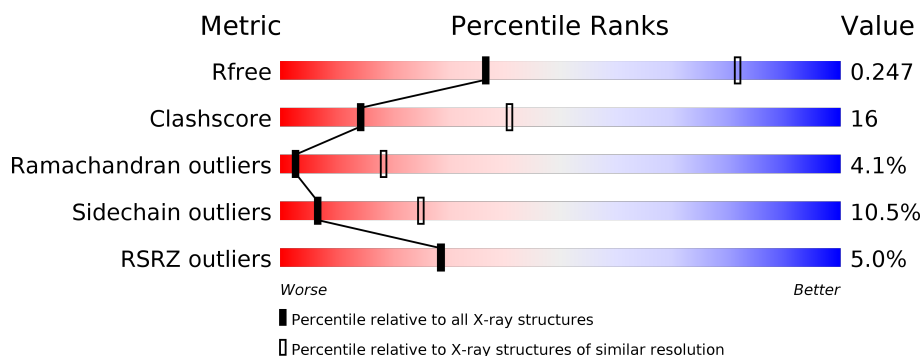
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.40 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	1143	<div> <div>4%</div> <div> <div></div> <div>61%</div> <div>30%</div> <div>6%</div> </div> </div>
1	B	1143	<div> <div>4%</div> <div> <div></div> <div>61%</div> <div>30%</div> <div>6%</div> </div> </div>
2	C	13	<div> <div>23%</div> <div> <div></div> <div>77%</div> <div>23%</div> </div> </div>
2	D	13	<div> <div>69%</div> <div> <div>46%</div> <div>46%</div> <div>8%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 17652 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 damage-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1114	Total	C	N	O	S	0	0	0
			8726	5529	1472	1677	48			
1	B	1114	Total	C	N	O	S	0	0	0
			8726	5529	1472	1677	48			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q16531
A	-1	SER	-	EXPRESSION TAG	UNP Q16531
A	0	HIS	-	EXPRESSION TAG	UNP Q16531
A	422	TYR	ASP	SEE REMARK 999	UNP Q16531
A	898	ASP	GLU	SEE REMARK 999	UNP Q16531
A	899	VAL	LEU	SEE REMARK 999	UNP Q16531
B	-2	GLY	-	EXPRESSION TAG	UNP Q16531
B	-1	SER	-	EXPRESSION TAG	UNP Q16531
B	0	HIS	-	EXPRESSION TAG	UNP Q16531
B	422	TYR	ASP	SEE REMARK 999	UNP Q16531
B	898	ASP	GLU	SEE REMARK 999	UNP Q16531
B	899	VAL	LEU	SEE REMARK 999	UNP Q16531

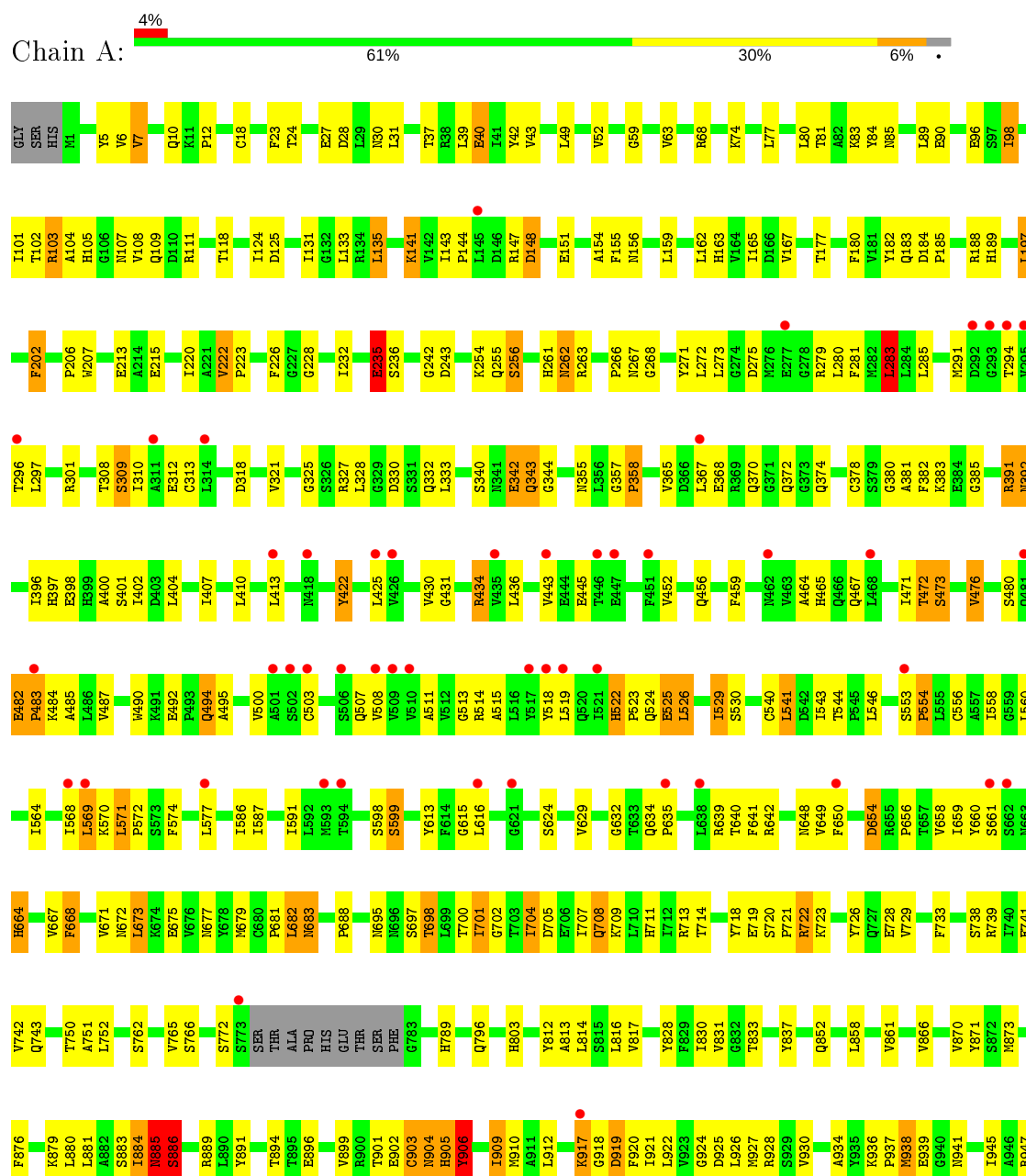
- Molecule 2 is a protein called WD repeat-containing protein 42A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	13	Total	C	N	O	0	0	0
			100	61	20	19			
2	D	13	Total	C	N	O	0	0	0
			100	61	20	19			

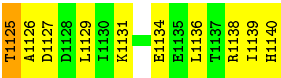
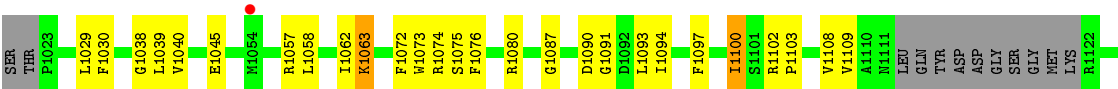
3 Residue-property plots

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.

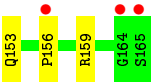
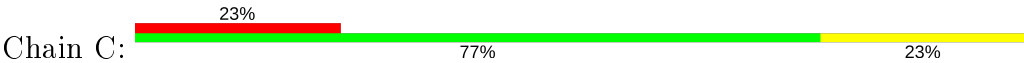
- Molecule 1: DNA damage-binding protein 1



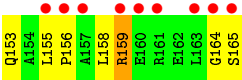




• Molecule 2: WD repeat-containing protein 42A



• Molecule 2: WD repeat-containing protein 42A



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	64.14Å 133.81Å 184.00Å 90.00° 90.45° 90.00°	Depositor
Resolution (Å)	49.15 – 3.40 49.12 – 3.18	Depositor EDS
% Data completeness (in resolution range)	87.3 (49.15-3.40) 81.7 (49.12-3.18)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.22 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.260 , 0.338 0.249 , 0.247	Depositor DCC
R_{free} test set	2192 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å ²)	71.6	Xtriage
Anisotropy	0.097	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 29.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.407 for h,-k,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	17652	wwPDB-VP
Average B, all atoms (Å ²)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.27% 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

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	1/8885 (0.0%)	0.65	2/12034 (0.0%)
1	B	0.46	2/8885 (0.0%)	0.67	4/12034 (0.0%)
2	C	0.44	0/100	0.63	0/133
2	D	0.39	0/100	0.62	0/133
All	All	0.45	3/17970 (0.0%)	0.66	6/24334 (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	A	0	1
1	B	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	367	LEU	C-O	7.37	1.37	1.23
1	B	1013	VAL	CA-C	-7.04	1.34	1.52
1	A	342	GLU	C-N	7.03	1.50	1.34

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	1013	VAL	CB-CA-C	-16.00	81.00	111.40
1	B	1014	MET	N-CA-CB	-14.77	84.01	110.60
1	B	1013	VAL	N-CA-C	13.43	147.25	111.00
1	A	992	LEU	CA-CB-CG	5.10	127.03	115.30
1	B	162	LEU	CA-CB-CG	5.05	126.92	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	886	SER	Mainchain
1	B	367	LEU	Mainchain

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	8726	0	8706	272	0
1	B	8726	0	8706	271	0
2	C	100	0	103	7	0
2	D	100	0	103	8	0
All	All	17652	0	17618	547	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 16.

The worst 5 of 547 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:1014:MET:SD	1:B:1015:GLN:HG2	1.79	1.23
1:B:1014:MET:CG	1:B:1015:GLN:H	1.58	1.12
1:B:1014:MET:SD	1:B:1015:GLN:N	2.30	1.04
1:B:706:GLU:O	1:B:707:ILE:HG13	1.57	1.04
1:A:1063:LYS:H	1:A:1063:LYS:HD3	1.18	1.03

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	1106/1143 (97%)	921 (83%)	134 (12%)	51 (5%)	2	15
1	B	1106/1143 (97%)	906 (82%)	161 (15%)	39 (4%)	3	21
2	C	11/13 (85%)	9 (82%)	2 (18%)	0	100	100
2	D	11/13 (85%)	10 (91%)	0	1 (9%)	1	4
All	All	2234/2312 (97%)	1846 (83%)	297 (13%)	91 (4%)	3	18

5 of 91 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	514	ARG
1	A	599	SER
1	A	683	ASN
1	A	886	SER
1	A	904	ASN

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	977/1001 (98%)	882 (90%)	95 (10%)	8	28
1	B	977/1001 (98%)	866 (89%)	111 (11%)	5	21
2	C	10/10 (100%)	10 (100%)	0	100	100
2	D	10/10 (100%)	9 (90%)	1 (10%)	7	27
All	All	1974/2022 (98%)	1767 (90%)	207 (10%)	7	25

5 of 207 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1093	LEU
1	B	174	GLN
1	B	955	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	27	GLU
1	B	103	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1077	HIS
1	B	267	ASN
1	B	950	ASN
1	B	4	ASN
1	B	30	ASN

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, 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	1114/1143 (97%)	0.14	51 (4%) 32 32	39, 68, 118, 134	0
1	B	1114/1143 (97%)	0.14	49 (4%) 34 34	42, 69, 120, 137	0
2	C	13/13 (100%)	1.02	3 (23%) 0 1	81, 84, 86, 86	0
2	D	13/13 (100%)	3.50	9 (69%) 0 0	100, 103, 104, 104	0
All	All	2254/2312 (97%)	0.17	112 (4%) 28 29	39, 68, 120, 137	0

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	508	VAL	11.6
1	A	293	GLY	10.4
1	B	508	VAL	10.1
2	D	165	SER	9.2
2	D	164	GLY	7.5

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.