



# Full wwPDB X-ray Structure Validation Report ⓘ

May 25, 2020 – 09:18 am BST

PDB ID : 6HTU  
Title : Structure of hStau1 dsRBD3-4 in complex with ARF1 RNA  
Authors : Emmerich, C.; Lazzaretti, D.; Bandholz-Cajamarca, L.; Bono, F.  
Deposited on : 2018-10-04  
Resolution : 2.89 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

<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  
Xtrriage (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

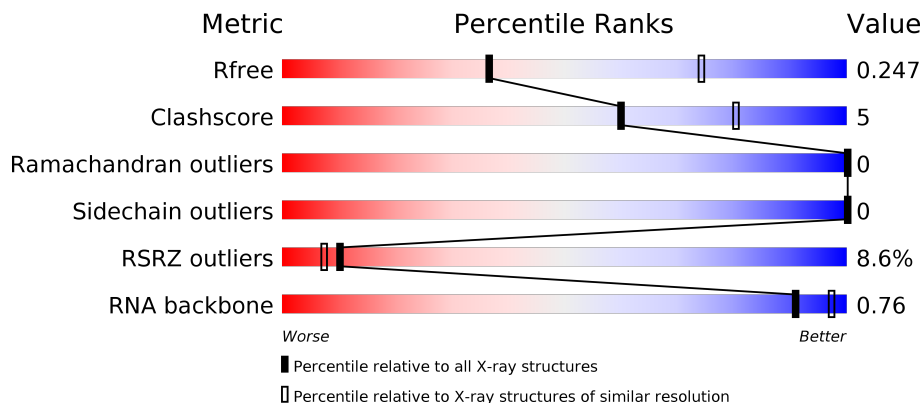
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.89 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.



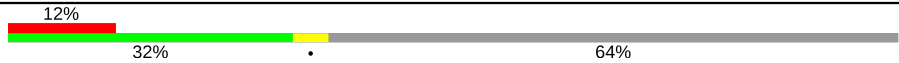
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2691 (2.90-2.86)
Clashscore	141614	2947 (2.90-2.86)
Ramachandran outliers	138981	2868 (2.90-2.86)
Sidechain outliers	138945	2871 (2.90-2.86)
RSRZ outliers	127900	2629 (2.90-2.86)
RNA backbone	3102	1121 (3.16-2.60)

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  $\geq 3$ , 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	D	19	 95% 5%
2	F	19	 79% 21%
3	A	182	 36% 6% 58%
3	B	182	 % 37% 5% 58%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	C	182	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into three segments: a red segment on the left labeled '12%', a green segment in the middle labeled '32%', and a grey segment on the right labeled '64%'. A small black dot is located at the end of the green segment.</p>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4226 atoms, of which 1906 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 RNA chain called RNA (19-MER).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
1	D	19	611	181	208	74	130	18	0	0	0

- Molecule 2 is a RNA chain called RNA (19-MER).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	F	19	606	180	205	69	134	18	0	0	0

- Molecule 3 is a protein called Double-stranded RNA-binding protein Staufen homolog 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	A	76	1119	354	567	95	101	2	0	0	0
3	B	76	1121	355	565	94	105	2	0	0	0
3	C	65	766	251	361	76	76	2	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	179	GLY	-	expression tag	UNP O95793
A	180	HIS	-	expression tag	UNP O95793
A	181	MET	-	expression tag	UNP O95793
A	359	ARG	ALA	conflict	UNP O95793
B	179	GLY	-	expression tag	UNP O95793
B	180	HIS	-	expression tag	UNP O95793
B	181	MET	-	expression tag	UNP O95793
B	359	ARG	ALA	conflict	UNP O95793
C	179	GLY	-	expression tag	UNP O95793
C	180	HIS	-	expression tag	UNP O95793

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	181	MET	-	expression tag	UNP O95793
C	359	ARG	ALA	conflict	UNP O95793

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O 1 1	0	0
4	B	2	Total O 2 2	0	0

### 3 Residue-property plots [i](#)

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: RNA (19-MER)

Chain D: 95% 5%



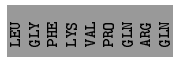
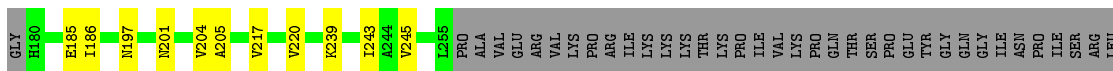
- Molecule 2: RNA (19-MER)

Chain F: 79% 21%



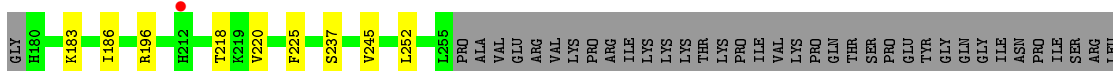
- Molecule 3: Double-stranded RNA-binding protein Staufen homolog 1

Chain A: 36% 6% 58%

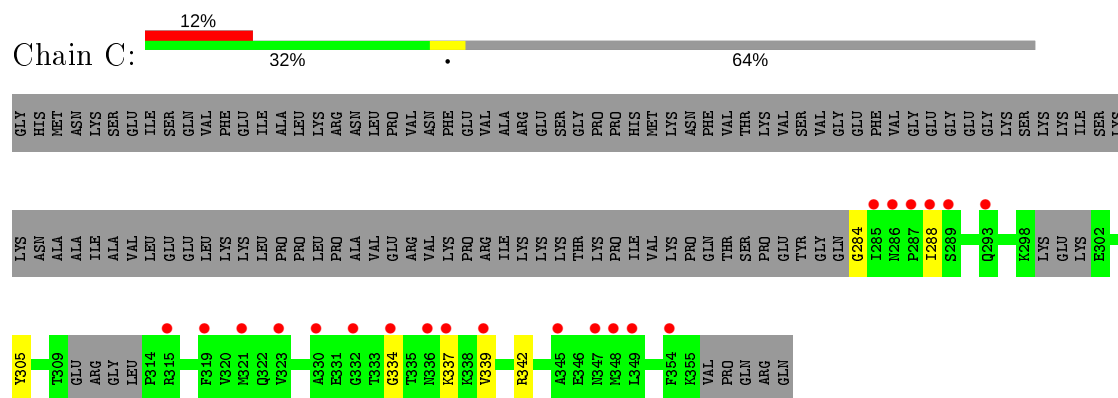


- Molecule 3: Double-stranded RNA-binding protein Staufen homolog 1

Chain B: 37% 5% 58% 1%



- Molecule 3: Double-stranded RNA-binding protein Staufen homolog 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.89Å 105.89Å 169.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.35 – 2.89 47.35 – 2.89	Depositor EDS
% Data completeness (in resolution range)	99.1 (47.35-2.89) 99.4 (47.35-2.89)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.13 (at 2.91Å)	Xtrriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.217 , 0.240 0.224 , 0.247	Depositor DCC
$R_{free}$ test set	1109 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	90.2	Xtrriage
Anisotropy	0.206	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 64.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4226	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	109.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.53% 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	D	0.24	0/450	0.73	0/700
2	F	0.21	0/447	0.72	0/695
3	A	0.28	0/561	0.44	0/757
3	B	0.38	0/565	0.46	0/763
3	C	0.26	0/408	0.45	0/553
All	All	0.29	0/2431	0.58	0/3468

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 [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	D	403	208	208	1	0
2	F	401	205	205	3	0
3	A	552	567	552	8	0
3	B	556	565	549	6	0
3	C	405	361	315	6	0
4	A	1	0	0	0	0
4	B	2	0	0	0	0
All	All	2320	1906	1829	19	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 5.

All (19) 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)
3:B:220:VAL:HG13	3:B:245:VAL:HG21	1.85	0.59
3:B:186:ILE:HD12	3:B:186:ILE:H	1.72	0.54
3:A:186:ILE:H	3:A:186:ILE:HD12	1.75	0.51
3:C:334:GLY:HA3	3:C:339:VAL:HG23	1.93	0.49
3:A:205:ALA:HB3	3:A:217:VAL:HG23	1.94	0.49
3:B:196:ARG:HH12	3:B:252:LEU:HB3	1.77	0.48
3:A:186:ILE:HD11	3:A:239:LYS:HA	1.95	0.48
3:A:185:GLU:HG2	3:A:243:ILE:HG12	1.96	0.47
3:B:218:THR:HG23	3:B:237:SER:OG	2.15	0.46
2:F:201:U:H2'	2:F:202:U:C6	2.53	0.44
3:A:201:ASN:OD1	3:C:342:ARG:NH2	2.44	0.43
3:C:288:ILE:HD11	3:C:305:TYR:CE2	2.53	0.43
3:A:197:ASN:HA	3:C:284:GLY:HA3	2.01	0.43
3:B:225:PHE:CZ	3:B:252:LEU:HD11	2.55	0.42
3:A:204:VAL:HG23	3:A:204:VAL:O	2.20	0.42
1:D:81:C:OP1	3:C:337:LYS:HB2	2.20	0.41
2:F:205:U:O2'	3:B:183:LYS:HE2	2.20	0.41
3:A:220:VAL:HG13	3:A:245:VAL:HG21	2.02	0.41
2:F:199:A:H5'	3:C:288:ILE:CG2	2.51	0.41

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	
3	A	74/182 (41%)	69 (93%)	5 (7%)	0	100	100
3	B	74/182 (41%)	72 (97%)	2 (3%)	0	100	100
3	C	59/182 (32%)	58 (98%)	1 (2%)	0	100	100
All	All	207/546 (38%)	199 (96%)	8 (4%)	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	
3	A	57/157 (36%)	57 (100%)	0	100	100
3	B	58/157 (37%)	58 (100%)	0	100	100
3	C	25/157 (16%)	25 (100%)	0	100	100
All	All	140/471 (30%)	140 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	D	18/19 (94%)	0	0
2	F	18/19 (94%)	0	0
All	All	36/38 (94%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

## 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

### 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	D	19/19 (100%)	-0.02	0	100   100	65, 73, 91, 92	0
2	F	19/19 (100%)	0.05	0	100   100	71, 79, 84, 87	0
3	A	76/182 (41%)	0.09	0	100   100	69, 99, 171, 193	0
3	B	76/182 (41%)	0.06	1 (1%)	77   77	72, 104, 169, 197	0
3	C	65/182 (35%)	1.40	21 (32%)	0   0	119, 171, 229, 243	0
All	All	255/584 (43%)	0.40	22 (8%)	10   7	65, 108, 204, 243	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	319	PHE	5.6
3	C	288	ILE	4.1
3	C	336	ASN	3.9
3	C	293	GLN	3.6
3	C	349	LEU	3.3
3	C	332	GLY	3.0
3	B	212	HIS	3.0
3	C	339	VAL	2.9
3	C	321	MET	2.9
3	C	287	PRO	2.8
3	C	323	VAL	2.8
3	C	347	ASN	2.7
3	C	286	ASN	2.6
3	C	354	PHE	2.6
3	C	330	ALA	2.5
3	C	285	ILE	2.4
3	C	289	SER	2.4
3	C	337	LYS	2.3
3	C	315	ARG	2.3
3	C	345	ALA	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
3	C	348	MET	2.2
3	C	334	GLY	2.0

## 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.