



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 13, 2022 – 12:36 PM JST

PDB ID : 7F3J  
Title : Crystal structure of human YBX2 CSD in complex with m5C RNA in space group P1  
Authors : Zhang, Y.; Huang, Y.  
Deposited on : 2021-06-16  
Resolution : 1.95 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

<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:

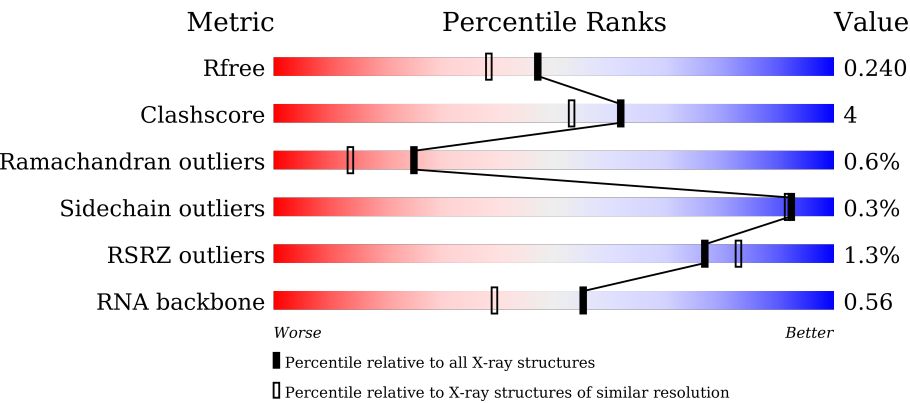
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

# 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 1.95 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 <sub>free</sub>	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)
RNA backbone	3102	1124 (2.50-1.42)

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  $\geq 3$ , 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	A	93	<div><div></div><div>81%15%.</div></div>
1	C	93	<div><div></div><div>91%. .</div></div>
1	E	93	<div><div>2%</div><div>85%12%.</div></div>
1	G	93	<div><div>3%</div><div>90%6%.</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	7	<div><div></div><div></div><div></div><div>43%</div><div>29%</div><div>29%</div></div>
2	D	7	<div><div></div><div></div><div></div><div>57%</div><div>14%</div><div>29%</div></div>
2	F	7	<div><div></div><div></div><div></div><div>43%</div><div>14%</div><div>14%</div><div>29%</div></div>
2	H	7	<div><div></div><div></div><div></div><div>57%</div><div>14%</div><div>29%</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3530 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 Y-box-binding protein 2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	89	Total	C	N	O	0	0	0
			692	433	129	130			
1	C	89	Total	C	N	O	0	0	0
			692	433	129	130			
1	E	90	Total	C	N	O	0	0	0
			699	438	130	131			
1	G	90	Total	C	N	O	0	0	0
			699	438	130	131			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	92	THR	ILE	engineered mutation	UNP Q9Y2T7
A	93	LYS	GLN	engineered mutation	UNP Q9Y2T7
C	92	THR	ILE	engineered mutation	UNP Q9Y2T7
C	93	LYS	GLN	engineered mutation	UNP Q9Y2T7
E	92	THR	ILE	engineered mutation	UNP Q9Y2T7
E	93	LYS	GLN	engineered mutation	UNP Q9Y2T7
G	92	THR	ILE	engineered mutation	UNP Q9Y2T7
G	93	LYS	GLN	engineered mutation	UNP Q9Y2T7

- Molecule 2 is a RNA chain called RNA (5'-R(\*UP\*CP\*AP\*UP\*(5MC))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	5	Total	C	N	O	P	0	0	0
			100	47	15	34	4			
2	D	5	Total	C	N	O	P	0	0	0
			100	47	15	34	4			
2	F	5	Total	C	N	O	P	0	0	0
			100	47	15	34	4			
2	H	5	Total	C	N	O	P	0	0	0
			100	47	15	34	4			


- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	85	Total 85	O 85	0	0
3	B	18	Total 18	O 18	0	0
3	C	94	Total 94	O 94	0	0
3	D	10	Total 10	O 10	0	0
3	E	71	Total 71	O 71	0	0
3	F	5	Total 5	O 5	0	0
3	G	59	Total 59	O 59	0	0
3	H	6	Total 6	O 6	0	0

### 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: Y-box-binding protein 2

Chain A: 




- Molecule 1: Y-box-binding protein 2

Chain C: 

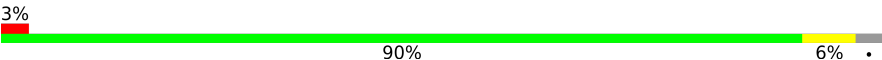


- Molecule 1: Y-box-binding protein 2

Chain E: 



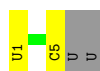
- Molecule 1: Y-box-binding protein 2

Chain G: 



- Molecule 2: RNA (5'-R(\*UP\*CP\*AP\*UP\*(5MC))-3')

Chain B: 



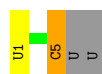
- Molecule 2: RNA (5'-R(\*UP\*CP\*AP\*UP\*(5MC))-3')

Chain D:  57% 14% 29%



- Molecule 2: RNA (5'-R(\*UP\*CP\*AP\*UP\*(5MC))-3')

Chain F:  43% 14% 14% 29%



- Molecule 2: RNA (5'-R(\*UP\*CP\*AP\*UP\*(5MC))-3')

Chain H:  57% 14% 29%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.62Å 46.36Å 57.70Å 69.26° 85.60° 90.02°	Depositor
Resolution (Å)	24.74 – 1.95 29.26 – 1.93	Depositor EDS
% Data completeness (in resolution range)	93.3 (24.74-1.95) 47.5 (29.26-1.93)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.11 (at 1.92Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.188 , 0.229 0.188 , 0.240	Depositor DCC
$R_{free}$ test set	1053 reflections (6.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.9	Xtriage
Anisotropy	0.069	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 43.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.36$	Xtriage
Estimated twinning fraction	0.024 for -h,k,k-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3530	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.66% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 5MC

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	0.41	0/705	0.63	0/951
1	C	0.47	0/705	0.63	0/951
1	E	0.39	0/713	0.68	0/962
1	G	0.39	0/713	0.69	0/962
2	B	0.49	0/87	0.88	0/133
2	D	0.49	0/87	0.97	0/133
2	F	0.43	0/87	0.91	0/133
2	H	0.44	0/87	0.86	0/133
All	All	0.42	0/3184	0.69	0/4358

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	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	151	GLY	Peptide

## 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	692	0	686	12	0
1	C	692	0	686	2	0
1	E	699	0	694	8	0
1	G	699	0	694	4	0
2	B	100	0	57	3	0
2	D	100	0	57	0	0
2	F	100	0	56	4	0
2	H	100	0	57	0	0
3	A	85	0	0	4	0
3	B	18	0	0	0	0
3	C	94	0	0	1	0
3	D	10	0	0	0	0
3	E	71	0	0	0	1
3	F	5	0	0	0	0
3	G	59	0	0	1	1
3	H	6	0	0	0	0
All	All	3530	0	2987	27	1

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 4.

The worst 5 of 27 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:G:123:GLN:OE1	3:G:201:HOH:O	1.95	0.84
1:C:89:VAL:N	3:C:201:HOH:O	2.18	0.75
1:C:104:ARG:HG3	1:E:124:THR:HG21	1.73	0.71
1:A:172:ARG:HH11	1:A:172:ARG:HG2	1.56	0.70
1:A:172:ARG:HD3	3:A:277:HOH:O	1.92	0.69

All (1) 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)
3:E:257:HOH:O	3:G:230:HOH:O[1_655]	1.99	0.21

## 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	87/93 (94%)	85 (98%)	2 (2%)	0	100	100
1	C	87/93 (94%)	84 (97%)	2 (2%)	1 (1%)	14	5
1	E	88/93 (95%)	86 (98%)	1 (1%)	1 (1%)	14	5
1	G	88/93 (95%)	87 (99%)	1 (1%)	0	100	100
All	All	350/372 (94%)	342 (98%)	6 (2%)	2 (1%)	25	14

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	152	GLU
1	E	152	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	73/76 (96%)	73 (100%)	0	100	100
1	C	73/76 (96%)	73 (100%)	0	100	100
1	E	74/76 (97%)	73 (99%)	1 (1%)	67	62

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	G	74/76 (97%)	74 (100%)	0	100	100
All	All	294/304 (97%)	293 (100%)	1 (0%)	92	92

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	104	ARG

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

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	4/7 (57%)	0	0
2	D	4/7 (57%)	0	0
2	F	4/7 (57%)	0	0
2	H	4/7 (57%)	0	0
All	All	16/28 (57%)	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 ⓘ

4 non-standard protein/DNA/RNA residues 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
2	5MC	B	5	2	15,22,23	2.02	5 (33%)	19,32,35	1.35	2 (10%)
2	5MC	H	5	2	15,22,23	2.17	5 (33%)	19,32,35	1.13	2 (10%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	5MC	F	5	2	15,22,23	3.25	5 (33%)	19,32,35	3.28	6 (31%)
2	5MC	D	5	2	15,22,23	2.04	5 (33%)	19,32,35	1.19	2 (10%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	5MC	B	5	2	-	0/5/25/26	0/2/2/2
2	5MC	H	5	2	-	2/5/25/26	0/2/2/2
2	5MC	F	5	2	-	2/5/25/26	0/2/2/2
2	5MC	D	5	2	-	0/5/25/26	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	5	5MC	C5-C4	7.05	1.52	1.41
2	F	5	5MC	C4-N4	6.91	1.51	1.34
2	F	5	5MC	C4-N3	4.68	1.41	1.35
2	H	5	5MC	C4-N3	4.52	1.41	1.35
2	H	5	5MC	C2-N3	4.30	1.46	1.38

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	5	5MC	C4-N3-C2	9.68	127.71	116.02
2	F	5	5MC	C5-C4-N3	-7.38	109.60	121.26
2	F	5	5MC	C6-N1-C1'	4.65	129.68	119.24
2	B	5	5MC	C4-N3-C2	3.96	120.80	116.02
2	D	5	5MC	C4-N3-C2	3.95	120.79	116.02

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	5	5MC	O4'-C1'-N1-C6
2	F	5	5MC	C2'-C1'-N1-C6
2	H	5	5MC	O4'-C1'-N1-C6
2	H	5	5MC	C2'-C1'-N1-C6

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	5	5MC	3	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides 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	89/93 (95%)	-0.60	0 100 100	10, 15, 27, 41	0
1	C	89/93 (95%)	-0.61	0 100 100	10, 16, 31, 49	0
1	E	90/93 (96%)	-0.26	2 (2%) 62 70	13, 20, 42, 54	0
1	G	90/93 (96%)	-0.16	3 (3%) 46 56	11, 20, 40, 56	0
2	B	4/7 (57%)	-0.60	0 100 100	18, 20, 24, 37	0
2	D	4/7 (57%)	-0.91	0 100 100	17, 20, 20, 33	0
2	F	4/7 (57%)	-0.33	0 100 100	24, 26, 29, 45	0
2	H	4/7 (57%)	-0.28	0 100 100	26, 29, 30, 44	0
All	All	374/400 (93%)	-0.41	5 (1%) 77 83	10, 18, 39, 56	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	151	GLY	4.0
1	G	150	GLU	3.2
1	E	90	LEU	3.1
1	E	152	GLU	2.2
1	G	152	GLU	2.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
2	5MC	F	5	21/22	0.91	0.10	18,29,40,49	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	5MC	H	5	21/22	0.92	0.11	19,25,38,46	0
2	5MC	B	5	21/22	0.95	0.07	12,17,24,26	0
2	5MC	D	5	21/22	0.97	0.08	11,16,23,24	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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.