



# wwPDB X-ray Structure Validation Summary Report

Oct 9, 2023 – 12:51 AM EDT

PDB ID : 7JM4  
Title : IRF Transcription Factor  
Authors : Sundararaj, S.; Williams, S.J.; Casarotto, M.G.  
Deposited on : 2020-07-31  
Resolution : 2.95 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35.1  
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) : 2.35.1

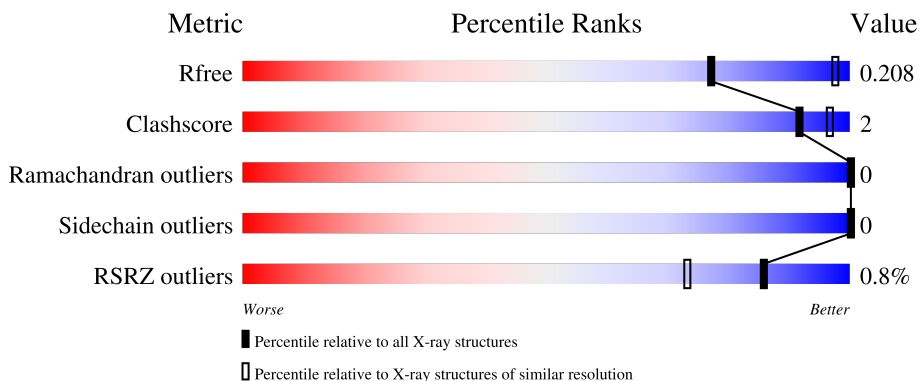
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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.



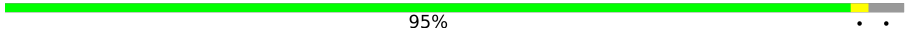
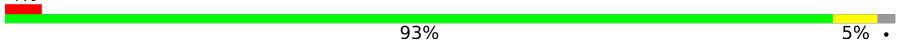
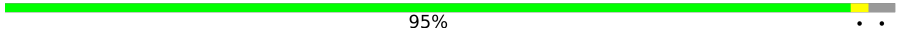
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	C	19	 89% 11%
1	D	19	 100%
2	E	19	 89% 11%
2	F	19	 79% 21%
3	A	111	 96% ..

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	B	111	 95% . .
3	G	111	 4% 93% 5% . .
3	H	111	 95% . .

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9445 atoms, of which 4241 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 DNA chain called Interferon-Stimulated Response Elements.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
1	D	19	603	185	211	82	106	19	0	0	0
1	C	19	601	185	212	82	104	18	0	0	0

- Molecule 2 is a DNA chain called Interferon-Stimulated Response Elements.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	E	19	606	186	219	60	122	19	0	0	0
2	F	19	606	186	219	60	122	19	0	0	0

- Molecule 3 is a protein called Interferon regulatory factor 4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	B	107	1749	584	844	158	162	1	0	0	0
3	A	109	1779	590	862	161	165	1	0	0	0
3	G	109	1715	590	798	161	165	1	0	0	0
3	H	108	1786	585	876	160	164	1	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	19	GLY	-	expression tag	UNP Q15306
B	20	SER	-	expression tag	UNP Q15306
A	19	GLY	-	expression tag	UNP Q15306
A	20	SER	-	expression tag	UNP Q15306

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	19	GLY	-	expression tag	UNP Q15306
G	20	SER	-	expression tag	UNP Q15306
H	19	GLY	-	expression tag	UNP Q15306
H	20	SER	-	expression tag	UNP Q15306

### 3 Residue-property plots


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: Interferon-Stimulated Response Elements

Chain D:  100%

There are no outlier residues recorded for this chain.

- Molecule 1: Interferon-Stimulated Response Elements

Chain C:  89% 11%




- Molecule 2: Interferon-Stimulated Response Elements

Chain E:  89% 11%



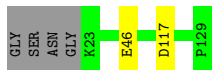
- Molecule 2: Interferon-Stimulated Response Elements

Chain F:  79% 21%



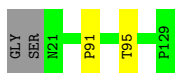
- Molecule 3: Interferon regulatory factor 4

Chain B:  95%

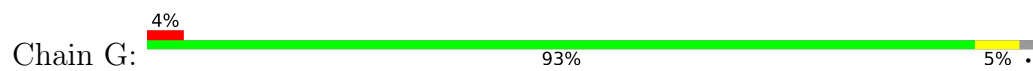


- Molecule 3: Interferon regulatory factor 4

Chain A:  96%



- Molecule 3: Interferon regulatory factor 4



- Molecule 3: Interferon regulatory factor 4



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	117.80Å 117.80Å 154.58Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.44 – 2.95 48.44 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.44-2.95) 99.5 (48.44-2.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.39 (at 2.96Å)	Xtrriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, $R_{free}$	0.185 , 0.205 0.187 , 0.208	Depositor DCC
$R_{free}$ test set	1320 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	76.7	Xtrriage
Anisotropy	0.141	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 29.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.026 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9445	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	83.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.21% 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	C	0.51	0/439	0.78	0/675
1	D	0.52	0/442	0.76	0/679
2	E	0.54	0/430	1.06	0/662
2	F	0.53	0/430	1.02	0/662
3	A	0.26	0/943	0.45	0/1273
3	B	0.26	0/931	0.42	0/1257
3	G	0.25	0/943	0.41	0/1273
3	H	0.26	0/935	0.45	0/1261
All	All	0.36	0/5493	0.64	0/7742

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	C	389	212	212	1	0
1	D	392	211	211	0	0
2	E	387	219	219	1	0
2	F	387	219	219	4	0
3	A	917	862	902	1	0
3	B	905	844	893	2	0
3	G	917	798	902	4	0
3	H	910	876	895	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	5204	4241	4453	13	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 2.

The worst 5 of 13 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:A:91:PRO:O	3:A:95:THR:HG23	2.01	0.59
1:C:1:DC:H2'	1:C:2:DA:C8	2.42	0.53
2:F:1:DG:OP2	2:F:2:DC:OP2	2.29	0.51
2:F:2:DC:H2''	2:F:3:DT:OP2	2.13	0.48
3:G:108:GLU:OE2	3:G:126:ARG:HD2	2.14	0.48

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	107/111 (96%)	103 (96%)	4 (4%)	0	100	100
3	B	105/111 (95%)	100 (95%)	5 (5%)	0	100	100
3	G	107/111 (96%)	101 (94%)	6 (6%)	0	100	100
3	H	106/111 (96%)	101 (95%)	5 (5%)	0	100	100
All	All	425/444 (96%)	405 (95%)	20 (5%)	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	97/98 (99%)	97 (100%)	0	100	100
3	B	96/98 (98%)	96 (100%)	0	100	100
3	G	97/98 (99%)	97 (100%)	0	100	100
3	H	96/98 (98%)	96 (100%)	0	100	100
All	All	386/392 (98%)	386 (100%)	0	100	100

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

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

### 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 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

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	C	19/19 (100%)	-0.23	0	100	100	56, 73, 106, 129	0
1	D	19/19 (100%)	-0.22	0	100	100	53, 64, 120, 129	0
2	E	19/19 (100%)	-0.22	0	100	100	52, 71, 130, 139	0
2	F	19/19 (100%)	-0.14	0	100	100	59, 76, 131, 141	0
3	A	109/111 (98%)	0.09	0	100	100	50, 78, 103, 132	0
3	B	107/111 (96%)	0.27	0	100	100	43, 64, 104, 133	0
3	G	109/111 (98%)	0.37	4 (3%)	41	27	47, 75, 114, 137	0
3	H	108/111 (97%)	0.19	0	100	100	53, 74, 104, 128	0
All	All	509/520 (97%)	0.17	4 (0%)	86	73	43, 74, 112, 141	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	G	60	GLN	2.7
3	G	119	SER	2.6
3	G	58	GLY	2.3
3	G	116	LEU	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 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.