Full wwPDB X-ray Structure Validation Report ⓘ

Apr 8, 2024 – 10:55 pm BST

PDB ID	:	8OJ2
Title	:	Crystal structure of the DNA binding domain of M. polymorpha Auxin Response Factor 2 (MpARF2) in complex with protomor-like sequence IR7
Authors	:	Crespo, I.; Weijers, D.; Boer, D.R.
Deposited on	:	2023-03-23
Resolution	:	2.56 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*
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:

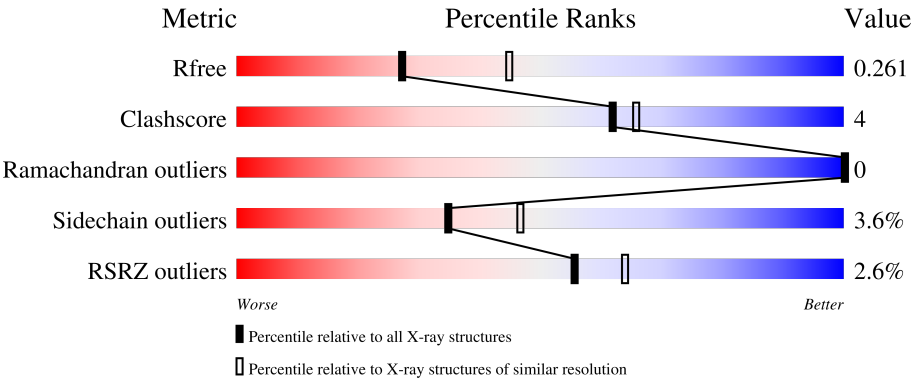
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.56 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	1279 (2.58-2.54)
Clashscore	141614	1327 (2.58-2.54)
Ramachandran outliers	138981	1312 (2.58-2.54)
Sidechain outliers	138945	1312 (2.58-2.54)
RSRZ outliers	127900	1269 (2.58-2.54)

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	366	2% 69% 10% • 21%
2	B	24	58% 33% 8%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2796 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 Auxin response factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	289	2262	1435	404	412	11	0	0	0

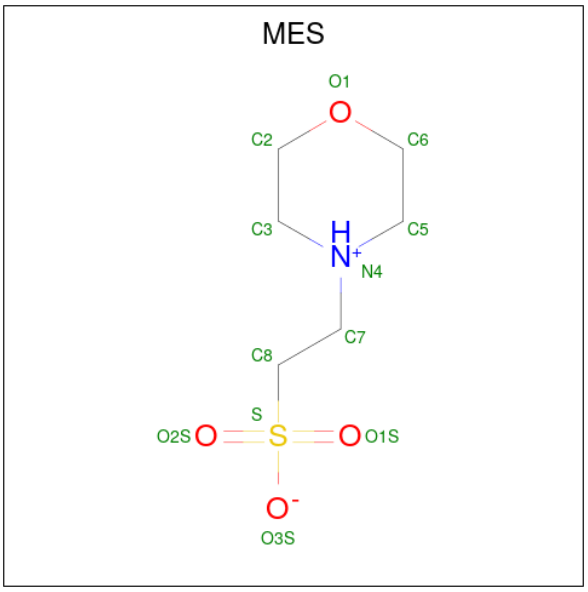
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	396	GLY	-	expression tag	UNP A0A0G3FH20
A	397	ASN	-	expression tag	UNP A0A0G3FH20
A	398	SER	-	expression tag	UNP A0A0G3FH20
A	399	TYR	-	expression tag	UNP A0A0G3FH20
A	400	SER	-	expression tag	UNP A0A0G3FH20
A	401	GLN	-	expression tag	UNP A0A0G3FH20
A	402	SER	-	expression tag	UNP A0A0G3FH20
A	403	MET	-	expression tag	UNP A0A0G3FH20

- Molecule 2 is a DNA chain called IR7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	21	493	235	89	145	24	0	4	0

- Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C N O S 12 6 1 4 1	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Cl 1 1	0	0

- Molecule 5 is water.

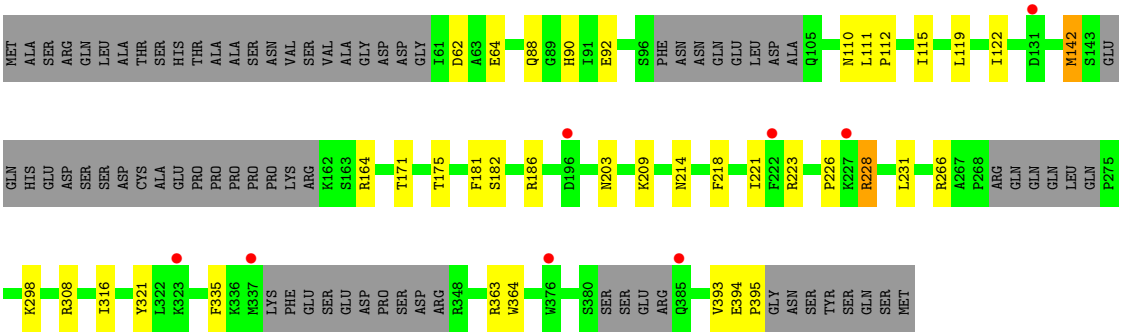
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	26	Total O 26 26	0	0
5	B	2	Total O 2 2	0	0

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: Auxin response factor

Chain A:



• Molecule 2: IR7

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	79.62Å 79.68Å 146.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	56.32 – 2.56 56.32 – 2.56	Depositor EDS
% Data completeness (in resolution range)	82.7 (56.32-2.56) 82.7 (56.32-2.56)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.72 (at 2.55Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.220 , 0.266 0.214 , 0.261	Depositor DCC
R_{free} test set	616 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	76.7	Xtriage
Anisotropy	0.006	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 50.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.007 for k,h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	2796	wwPDB-VP
Average B, all atoms (Å^2)	87.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.13% of the height of the origin peak. No significant pseudotranslation is detected.*

[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

Bond lengths and bond angles in the following residue types are not validated in this section: CL, MES

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	0.27	0/2320	0.53	1/3152 (0.0%)
2	B	1.08	2/570 (0.4%)	1.08	4/871 (0.5%)
All	All	0.53	2/2890 (0.1%)	0.69	5/4023 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	12[B]	DT	O3'-P	20.09	1.85	1.61
2	B	10[B]	DT	O3'-P	7.59	1.70	1.61

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	10[B]	DT	P-O3'-C3'	8.37	129.75	119.70
2	B	12[B]	DT	O3'-P-O5'	-6.40	91.83	104.00
1	A	62	ASP	CB-CG-OD2	6.03	123.72	118.30
2	B	10[B]	DT	OP1-P-O3'	5.96	118.31	105.20
2	B	12[B]	DT	OP1-P-O3'	5.17	116.58	105.20

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	2262	0	2233	17	0
2	B	493	0	266	4	0
3	A	12	0	12	1	0
4	A	1	0	0	0	0
5	A	26	0	0	0	0
5	B	2	0	0	0	0
All	All	2796	0	2511	21	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 4.

All (21) 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)
2:B:17:DG:H2"	2:B:18:DA:H8	1.68	0.58
1:A:181:PHE:HB3	1:A:231:LEU:HB2	1.90	0.55
2:B:17:DG:H2"	2:B:18:DA:C8	2.43	0.54
1:A:110:ASN:HA	3:A:501:MES:H62	1.89	0.52
1:A:111:LEU:HD23	1:A:364:TRP:HH2	1.73	0.52
2:B:11[A]:DA:H2"	2:B:12[A]:DA:C8	2.45	0.52
1:A:112:PRO:HG2	1:A:115:ILE:HG12	1.92	0.51
1:A:335:PHE:CD1	1:A:393:VAL:HB	2.48	0.49
1:A:111:LEU:HD23	1:A:364:TRP:CH2	2.49	0.48
1:A:266:ARG:O	1:A:308:ARG:NH1	2.46	0.48
1:A:88:GLN:O	1:A:92:GLU:HG3	2.15	0.47
1:A:226:PRO:O	1:A:228:ARG:HD2	2.14	0.47
1:A:119:LEU:HD21	1:A:122:ILE:HD11	1.96	0.47
1:A:218:PHE:HB3	1:A:231:LEU:HD22	1.97	0.46
1:A:209:LYS:HA	1:A:214:ASN:O	2.17	0.45
1:A:316:ILE:HD11	1:A:321:TYR:HB2	2.00	0.44
1:A:142:MET:O	1:A:298:LYS:NZ	2.41	0.44
1:A:171:THR:O	1:A:175:THR:HG23	2.19	0.43
1:A:221:ILE:HD13	1:A:223:ARG:HG2	2.01	0.42
2:B:20:DA:H2"	2:B:21:DA:C8	2.55	0.42
1:A:394:GLU:HA	1:A:395:PRO:HD3	1.96	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
1	A	277/366 (76%)	267 (96%)	10 (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
1	A	248/316 (78%)	239 (96%)	9 (4%)	35 47

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	64	GLU
1	A	90	HIS
1	A	142	MET
1	A	164	ARG
1	A	182	SER
1	A	186	ARG
1	A	203	ASN
1	A	228	ARG
1	A	363	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	88	GLN
1	A	90	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

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	$\#\|Z\| > 2$
3	MES	A	501	-	12,12,12	2.27	1 (8%)	14,16,16	2.22	8 (57%)

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
3	MES	A	501	-	-	5/6/14/14	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	501	MES	C8-S	-7.58	1.66	1.77

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	501	MES	C2-C3-N4	-3.55	104.71	110.10
3	A	501	MES	C5-N4-C3	3.54	116.79	108.83
3	A	501	MES	C7-N4-C3	3.20	119.41	111.23
3	A	501	MES	O2S-S-C8	2.79	110.27	106.92
3	A	501	MES	C7-N4-C5	2.57	117.81	111.23
3	A	501	MES	C6-C5-N4	-2.54	106.26	110.10
3	A	501	MES	O1S-S-C8	2.05	109.39	106.92
3	A	501	MES	O3S-S-C8	2.02	109.04	105.77

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	501	MES	C8-C7-N4-C3
3	A	501	MES	C7-C8-S-O2S
3	A	501	MES	C7-C8-S-O3S
3	A	501	MES	C8-C7-N4-C5
3	A	501	MES	C7-C8-S-O1S

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	MES	1	0

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^{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	289/366 (78%)	0.47	8 (2%) 53 62	47, 80, 119, 157	0
2	B	24/24 (100%)	-0.62	0 100 100	87, 116, 133, 137	6 (25%)
All	All	313/390 (80%)	0.38	8 (2%) 56 64	47, 82, 128, 157	6 (1%)

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	222	PHE	4.6
1	A	227	LYS	3.2
1	A	385	GLN	3.2
1	A	196	ASP	3.1
1	A	131	ASP	2.5
1	A	323	LYS	2.2
1	A	337	MET	2.2
1	A	376	TRP	2.2

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

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^{th} 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^2)	Q<0.9
4	CL	A	502	1/1	0.55	0.21	105,105,105,105	0
3	MES	A	501	12/12	0.95	0.15	64,75,97,106	0

6.5 Other polymers (i)

There are no such residues in this entry.