wwPDB X-ray Structure Validation Summary Report (i)

Mar 24, 2022 – 12:28 pm GMT

PDB ID	:	4YJ0
Title	:	Crystal structure of the DM domain of human DMRT1 bound to 25mer target DNA
Authors	:	Murphy, M.W.; Lee, J.K.; Rojo, S.; Gearhart, M.D.; Kurahashi, K.; Banerjee, S.; Loeuille, G.; Bashamboo, A.; McElreavey, K.; Zarkower, D.; Aihara, H.; Bardwell, V.J.
Deposited on	:	2015-03-02
Resolution	:	3.81 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

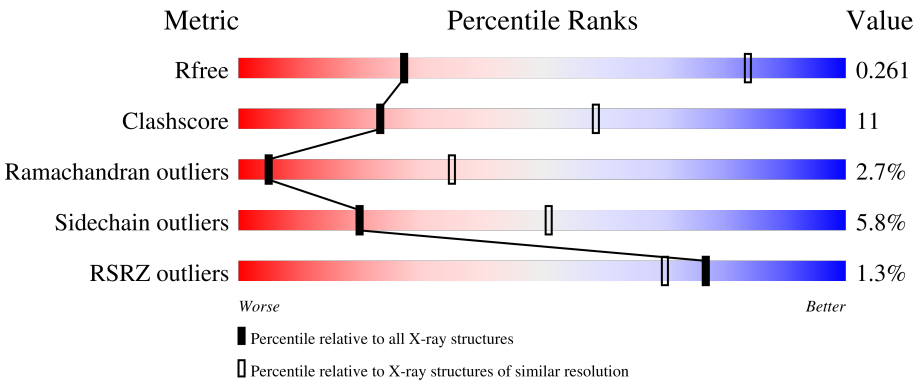
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.27
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 3.81 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	1231 (4.04-3.60)
Clashscore	141614	1031 (4.02-3.62)
Ramachandran outliers	138981	1261 (4.04-3.60)
Sidechain outliers	138945	1255 (4.04-3.60)
RSRZ outliers	127900	1139 (4.04-3.60)

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 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	70	% 70% 17% • 11%
1	B	70	% 67% 23% 10%
1	C	70	% 54% 29% 6% • 10%
2	D	25	52% 48%
3	E	25	52% 48%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2534 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 Doublesex- and mab-3-related transcription factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	62	497	302	106	81	8	0	0	0
1	B	63	507	307	108	84	8	0	0	0
1	C	63	505	308	108	81	8	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	67	SER	-	expression tag	UNP Q9Y5R6
A	68	LYS	-	expression tag	UNP Q9Y5R6
A	69	LYS	-	expression tag	UNP Q9Y5R6
A	132	GLY	-	expression tag	UNP Q9Y5R6
A	133	ILE	-	expression tag	UNP Q9Y5R6
A	134	SER	-	expression tag	UNP Q9Y5R6
A	135	HIS	-	expression tag	UNP Q9Y5R6
A	136	PRO	-	expression tag	UNP Q9Y5R6
B	67	SER	-	expression tag	UNP Q9Y5R6
B	68	LYS	-	expression tag	UNP Q9Y5R6
B	69	LYS	-	expression tag	UNP Q9Y5R6
B	132	GLY	-	expression tag	UNP Q9Y5R6
B	133	ILE	-	expression tag	UNP Q9Y5R6
B	134	SER	-	expression tag	UNP Q9Y5R6
B	135	HIS	-	expression tag	UNP Q9Y5R6
B	136	PRO	-	expression tag	UNP Q9Y5R6
C	67	SER	-	expression tag	UNP Q9Y5R6
C	68	LYS	-	expression tag	UNP Q9Y5R6
C	69	LYS	-	expression tag	UNP Q9Y5R6
C	132	GLY	-	expression tag	UNP Q9Y5R6
C	133	ILE	-	expression tag	UNP Q9Y5R6
C	134	SER	-	expression tag	UNP Q9Y5R6
C	135	HIS	-	expression tag	UNP Q9Y5R6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	136	PRO	-	expression tag	UNP Q9Y5R6

- Molecule 2 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	25	511	246	90	151	24	0	0	0

- Molecule 3 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	E	25	508	244	95	145	24	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

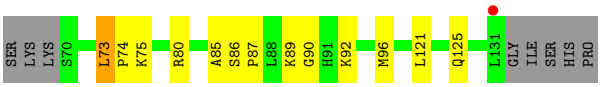
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
4	A	2	2	2	0	0
4	B	2	2	2	0	0
4	C	2	2	2	0	0

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: Doublesex- and mab-3-related transcription factor 1

Chain A:

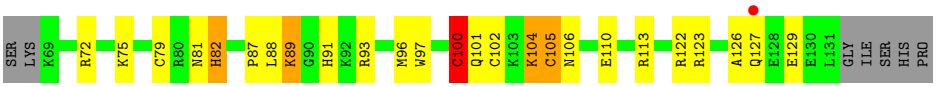


• Molecule 1: Doublesex- and mab-3-related transcription factor 1

Chain B:

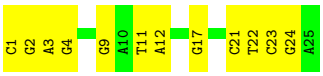
• Molecule 1: Doublesex- and mab-3-related transcription factor 1

Chain C:



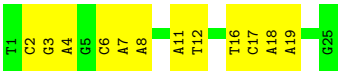
• Molecule 2: DNA (25-MER)

Chain D:



• Molecule 3: DNA (25-MER)

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	83.19Å 138.93Å 141.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.73 – 3.81 49.90 – 3.81	Depositor EDS
% Data completeness (in resolution range)	98.6 (34.73-3.81) 98.5 (49.90-3.81)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	29.81 (at 3.77Å)	Xtriage
Refinement program	PHENIX dev_1801	Depositor
R, R_{free}	0.229 , 0.262 0.231 , 0.261	Depositor DCC
R_{free} test set	809 reflections (9.90%)	wwPDB-VP
Wilson B-factor (Å^2)	128.7	Xtriage
Anisotropy	0.050	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.049 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	2534	wwPDB-VP
Average B, all atoms (Å^2)	132.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.06% 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 (i)

5.1 Standard geometry (i)

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

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	0.25	0/505	0.64	1/674 (0.1%)
1	B	0.22	0/515	0.46	0/686
1	C	1.06	2/513 (0.4%)	0.63	2/684 (0.3%)
2	D	0.52	0/572	0.95	0/882
3	E	0.55	0/570	0.92	0/877
All	All	0.60	2/2675 (0.1%)	0.77	3/3803 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	104	LYS	C-N	22.65	1.86	1.34
1	C	105	CYS	C-N	-6.11	1.20	1.34

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	73	LEU	CA-CB-CG	7.31	132.12	115.30
1	C	104	LYS	C-N-CA	-6.15	106.32	121.70
1	C	104	LYS	O-C-N	5.86	132.08	122.70

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	A	497	0	497	11	0
1	B	507	0	503	13	0
1	C	505	0	503	21	0
2	D	511	0	286	13	0
3	E	508	0	283	10	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
All	All	2534	0	2072	52	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 11.

The worst 5 of 52 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:C:104:LYS:C	1:C:105:CYS:N	1.86	1.27
1:C:100:CYS:SG	1:C:101:GLN:N	2.30	1.03
1:A:73:LEU:HD12	1:A:74:PRO:HD2	1.71	0.72
1:A:125:GLN:HE22	1:B:104:LYS:HZ3	1.41	0.69
1:A:73:LEU:HD13	3:E:19:DA:H4'	1.76	0.66

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	60/70 (86%)	54 (90%)	3 (5%)	3 (5%)	2 23
1	B	61/70 (87%)	53 (87%)	7 (12%)	1 (2%)	9 44
1	C	61/70 (87%)	55 (90%)	5 (8%)	1 (2%)	9 44
All	All	182/210 (87%)	162 (89%)	15 (8%)	5 (3%)	5 35

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	90	GLY
1	C	100	CYS
1	B	100	CYS
1	A	85	ALA
1	A	86	SER

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	51/60 (85%)	48 (94%)	3 (6%)	19	51
1	B	52/60 (87%)	50 (96%)	2 (4%)	33	60
1	C	51/60 (85%)	47 (92%)	4 (8%)	12	42
All	All	154/180 (86%)	145 (94%)	9 (6%)	20	51

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	96	MET
1	C	100	CYS
1	B	73	LEU
1	B	125	GLN
1	C	82	HIS

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

Mol	Chain	Res	Type
1	C	118	GLN
1	C	127	GLN

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)

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	C	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	104:LYS	C	105:CYS	N	1.86
1	C	105:CYS	C	106:ASN	N	1.20

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	62/70 (88%)	0.03	1 (1%)	72	64	105, 130, 188, 197	0
1	B	63/70 (90%)	0.17	1 (1%)	72	64	96, 123, 170, 187	0
1	C	63/70 (90%)	0.01	1 (1%)	72	64	99, 130, 168, 185	0
2	D	25/25 (100%)	-0.77	0	100	100	98, 122, 179, 182	0
3	E	25/25 (100%)	-0.76	0	100	100	92, 128, 181, 190	0
All	All	238/260 (91%)	-0.11	3 (1%)	77	70	92, 128, 180, 197	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	131	LEU	2.7
1	C	127	GLN	2.6
1	B	94	PHE	2.3

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 (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, 95th 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	ZN	C	201	1/1	0.78	0.10	142,142,142,142	0
4	ZN	B	202	1/1	0.93	0.07	124,124,124,124	0
4	ZN	A	201	1/1	0.94	0.16	111,111,111,111	0
4	ZN	B	201	1/1	0.98	0.13	99,99,99,99	0
4	ZN	C	202	1/1	0.98	0.07	125,125,125,125	0
4	ZN	A	202	1/1	0.99	0.06	119,119,119,119	0

6.5 Other polymers (i)

There are no such residues in this entry.