



Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2020 – 03:55 am BST

PDB ID : 5YHU
Title : Crystal structure of the DNA-binding domain of human myelin-gene regulatory factor
Authors : Chen, B.; Zhu, Y.; Ye, S.; Zhang, R.
Deposited on : 2017-09-30
Resolution : 1.85 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 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 : 2.11
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

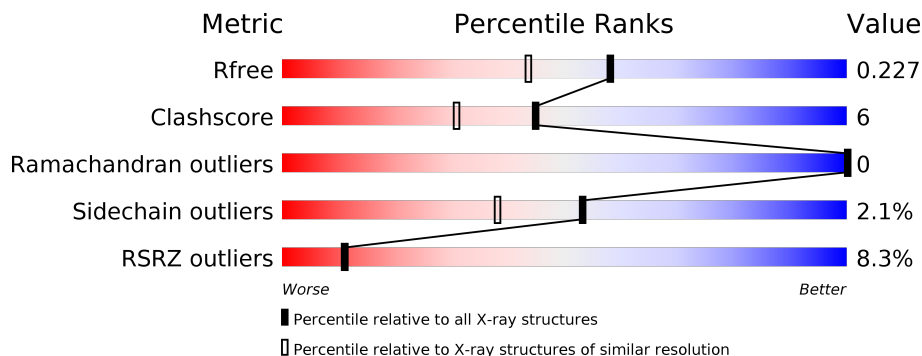
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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 ≥ 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	255	<div style="display: flex; align-items: center;"> <div style="width: 3%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 70%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 5%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 22%; height: 10px; background-color: grey;"></div> </div> <p style="font-size: small; margin-top: 5px;">3% 70% 5% 25%</p>
1	B	255	<div style="display: flex; align-items: center;"> <div style="width: 9%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 58%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 12%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 21%; height: 10px; background-color: grey;"></div> </div> <p style="font-size: small; margin-top: 5px;">9% 58% 12% 29%</p>

2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 3304 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 Myelin regulatory factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	191	1548	986	272	284	6	0	0	0
1	B	182	1476	947	255	268	6	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	315	GLY	-	expression tag	UNP Q9Y2G1
A	316	PRO	-	expression tag	UNP Q9Y2G1
A	317	MET	-	expression tag	UNP Q9Y2G1
A	318	ALA	-	expression tag	UNP Q9Y2G1
A	319	ASP	-	expression tag	UNP Q9Y2G1
A	320	ILE	-	expression tag	UNP Q9Y2G1
A	321	GLY	-	expression tag	UNP Q9Y2G1
A	322	SER	-	expression tag	UNP Q9Y2G1
B	315	GLY	-	expression tag	UNP Q9Y2G1
B	316	PRO	-	expression tag	UNP Q9Y2G1
B	317	MET	-	expression tag	UNP Q9Y2G1
B	318	ALA	-	expression tag	UNP Q9Y2G1
B	319	ASP	-	expression tag	UNP Q9Y2G1
B	320	ILE	-	expression tag	UNP Q9Y2G1
B	321	GLY	-	expression tag	UNP Q9Y2G1
B	322	SER	-	expression tag	UNP Q9Y2G1

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	244	Total 244	O 244	0	0
2	B	36	Total 36	O 36	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α , β , γ	105.12Å 105.12Å 297.41Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	34.18 – 1.85 43.53 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.3 (34.18-1.85) 92.2 (43.53-1.85)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.54 (at 1.86Å)	Xtrriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.204 , 0.228 0.204 , 0.227	Depositor DCC
R_{free} test set	2746 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	28.7	Xtrriage
Anisotropy	0.460	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 49.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	3304	wwPDB-VP
Average B, all atoms (Å ²)	54.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.16% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	A	0.55	1/1586 (0.1%)	0.65	0/2150
1	B	0.32	0/1513	0.54	0/2052
All	All	0.45	1/3099 (0.0%)	0.60	0/4202

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	387	CYS	CB-SG	-5.44	1.73	1.81

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	A	1548	0	1538	10	0
1	B	1476	0	1467	29	0
2	A	244	0	0	7	1
2	B	36	0	0	1	0
All	All	3304	0	3005	39	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 6.

All (39) 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:B:521:ARG:NH1	2:B:601:HOH:O	1.97	0.97
1:B:484:GLY:H	1:B:485:LYS:HZ3	1.31	0.76
1:B:476:THR:HG23	1:B:520:VAL:HG12	1.70	0.73
1:B:476:THR:OG1	1:B:479:ASN:OD1	2.09	0.70
1:B:404:GLU:HG3	1:B:459:PRO:HG3	1.75	0.69
1:B:388:GLN:HB2	1:B:390:LYS:HE3	1.75	0.68
1:A:444:ASP:OD2	2:A:603:HOH:O	2.13	0.67
1:B:392:HIS:HB3	1:B:471:HIS:NE2	2.11	0.66
1:B:484:GLY:H	1:B:485:LYS:NZ	1.95	0.64
1:A:442:GLN:HG2	2:A:602:HOH:O	1.97	0.64
1:A:469:ARG:NH1	2:A:601:HOH:O	2.01	0.61
1:B:388:GLN:HG2	1:B:390:LYS:HG2	1.84	0.60
1:B:390:LYS:HG3	1:B:391:ASN:OD1	2.03	0.59
1:A:374:LYS:NZ	2:A:611:HOH:O	2.38	0.57
1:B:477:ALA:HB3	1:B:491:ARG:HG3	1.85	0.56
1:A:343:LYS:NZ	2:A:604:HOH:O	2.18	0.55
1:A:390:LYS:HD2	1:A:391:ASN:OD1	2.07	0.55
1:B:390:LYS:HG2	1:B:391:ASN:H	1.71	0.54
1:A:522:ALA:O	2:A:605:HOH:O	2.19	0.52
1:B:440:GLN:HG3	1:B:470:LEU:HD13	1.92	0.52
1:B:363:LEU:HD22	1:B:403:GLY:HA3	1.92	0.51
1:B:473:SER:OG	1:B:474:GLU:HG3	2.11	0.50
1:A:513:GLN:OE1	2:A:606:HOH:O	2.20	0.49
1:B:433:ASN:OD1	1:B:433:ASN:N	2.45	0.48
1:B:494:MET:HB3	1:B:514:ILE:HD11	1.97	0.46
1:B:481:ARG:N	1:B:481:ARG:HD2	2.31	0.46
1:B:440:GLN:O	1:B:469:ARG:NE	2.48	0.44
1:B:483:LYS:N	1:B:483:LYS:HE2	2.33	0.43
1:B:390:LYS:CG	1:B:391:ASN:N	2.82	0.43
1:B:388:GLN:HG3	1:B:390:LYS:HD2	2.00	0.43
1:B:485:LYS:HD3	1:B:485:LYS:N	2.33	0.43
1:B:336:ASP:OD1	1:B:338:ASN:ND2	2.49	0.42
1:B:343:LYS:HE3	1:B:516:GLU:OE2	2.20	0.42
1:B:392:HIS:HB3	1:B:471:HIS:CE1	2.54	0.41
1:B:388:GLN:CG	1:B:390:LYS:HG2	2.50	0.41
1:B:388:GLN:OE1	1:B:388:GLN:N	2.53	0.41
1:A:482:LYS:HD3	1:A:482:LYS:HA	1.94	0.41
1:B:468:GLY:C	1:B:470:LEU:HD22	2.40	0.41
1:A:349:GLN:HA	1:A:352:TRP:CE2	2.56	0.40

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)
2:A:763:HOH:O	2:A:830:HOH:O[2_555]	2.17	0.03

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	189/255 (74%)	185 (98%)	4 (2%)	0	100	100
1	B	178/255 (70%)	172 (97%)	6 (3%)	0	100	100
All	All	367/510 (72%)	357 (97%)	10 (3%)	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	171/225 (76%)	170 (99%)	1 (1%)	86	83
1	B	162/225 (72%)	156 (96%)	6 (4%)	34	17
All	All	333/450 (74%)	326 (98%)	7 (2%)	53	38

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	390	LYS
1	B	379	SER
1	B	390	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	450	PHE
1	B	469	ARG
1	B	481	ARG
1	B	483	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	434	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 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, 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	191/255 (74%)	0.06	8 (4%) 36 34	21, 34, 60, 80	0
1	B	182/255 (71%)	0.82	23 (12%) 3 3	49, 70, 111, 136	0
All	All	373/510 (73%)	0.43	31 (8%) 11 11	21, 54, 102, 136	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	485	LYS	5.3
1	B	334	TYR	4.9
1	B	488	PRO	4.1
1	B	450	PHE	3.6
1	B	522	ALA	3.5
1	B	391	ASN	3.4
1	B	480	MET	3.4
1	B	484	GLY	3.4
1	B	505	ASN	3.3
1	B	390	LYS	3.3
1	A	484	GLY	3.2
1	B	481	ARG	3.1
1	B	451	ASN	3.0
1	A	483	LYS	2.9
1	B	333	SER	2.9
1	B	339	TYR	2.7
1	A	485	LYS	2.7
1	B	483	LYS	2.6
1	B	521	ARG	2.6
1	B	392	HIS	2.6
1	B	506	GLN	2.5
1	B	388	GLN	2.5
1	A	443	SER	2.5
1	B	389	LYS	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	486	PRO	2.3
1	A	445	ARG	2.2
1	A	514	ILE	2.2
1	A	344	TRP	2.2
1	B	473	SER	2.1
1	B	342	ILE	2.1
1	A	343	LYS	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.