Full wwPDB X-ray Structure Validation Report ⓘ

Mar 3, 2025 – 08:02 AM EST

PDB ID	:	9MX8
Title	:	Crystal structure of the DNA binding domain of FLI1 in complex with a DNA containing three contiguous GGAA sites
Authors	:	Hou, C.; Tsodikov, O.V.
Deposited on	:	2025-01-17
Resolution	:	3.15 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:

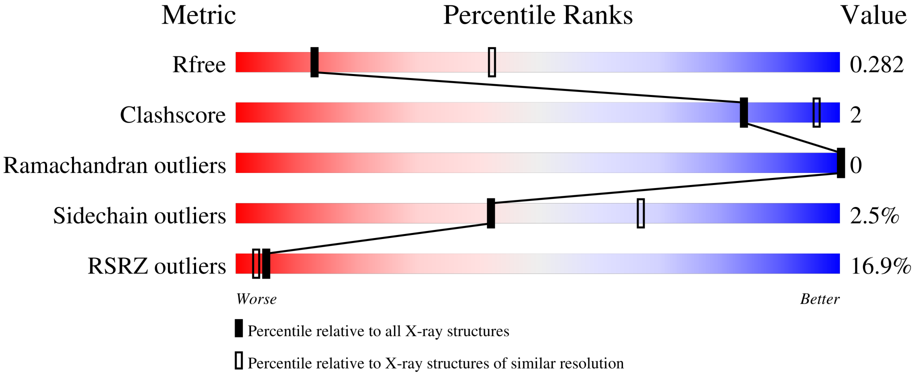
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.41.4

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

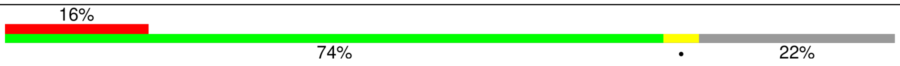
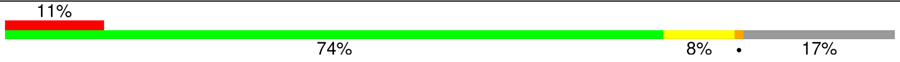
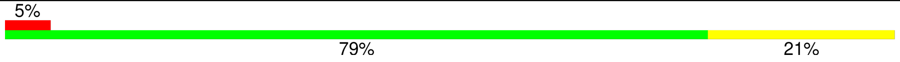
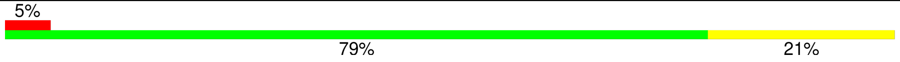
The reported resolution of this entry is 3.15 Å.

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}	164625	2168 (3.20-3.12)
Clashscore	180529	2333 (3.20-3.12)
Ramachandran outliers	177936	2266 (3.20-3.12)
Sidechain outliers	177891	2265 (3.20-3.12)
RSRZ outliers	164620	2169 (3.20-3.12)

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	121	
1	C	121	
1	D	121	
2	E	19	
3	F	19	

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 3190 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 Friend leukemia integration 1 transcription factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	106	Total 841	C 536	N 152	O 149	S 4	0	0	0
1	C	94	Total 768	C 489	N 138	O 137	S 4	0	0	0
1	D	100	Total 806	C 511	N 144	O 147	S 4	0	0	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	255	GLY	-	expression tag	UNP Q01543
A	256	PRO	-	expression tag	UNP Q01543
A	257	HIS	-	expression tag	UNP Q01543
A	258	MET	-	expression tag	UNP Q01543
A	362	ALA	PHE	engineered mutation	UNP Q01543
C	255	GLY	-	expression tag	UNP Q01543
C	256	PRO	-	expression tag	UNP Q01543
C	257	HIS	-	expression tag	UNP Q01543
C	258	MET	-	expression tag	UNP Q01543
C	362	ALA	PHE	engineered mutation	UNP Q01543
D	255	GLY	-	expression tag	UNP Q01543
D	256	PRO	-	expression tag	UNP Q01543
D	257	HIS	-	expression tag	UNP Q01543
D	258	MET	-	expression tag	UNP Q01543
D	362	ALA	PHE	engineered mutation	UNP Q01543

- Molecule 2 is a DNA chain called DNA (5'-D(P*GP*AP*CP*CP*GP*GP*AP*AP*GP*GP*AP*AP*GP*GP*AP*AP*GP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	19	Total 403	C 188	N 88	O 108	P 19	0	0	0

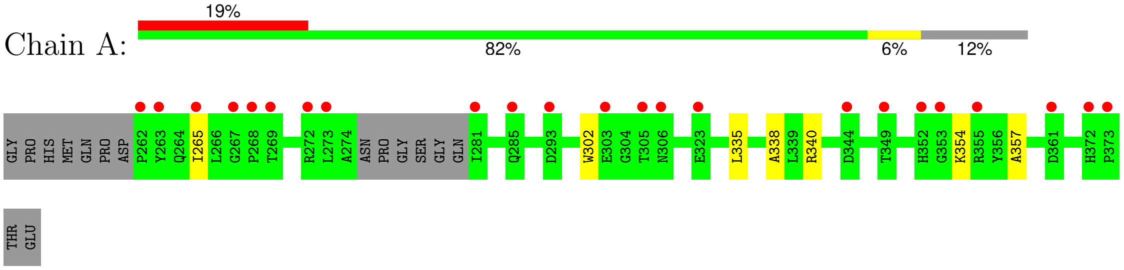
- Molecule 3 is a DNA chain called DNA (5'-D(*CP*AP*CP*TP*TP*CP*CP*TP*TP*CP*CP*TP*TP*CP*CP*GP*GP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	F	19	372	181	56	117	18	0	0	0

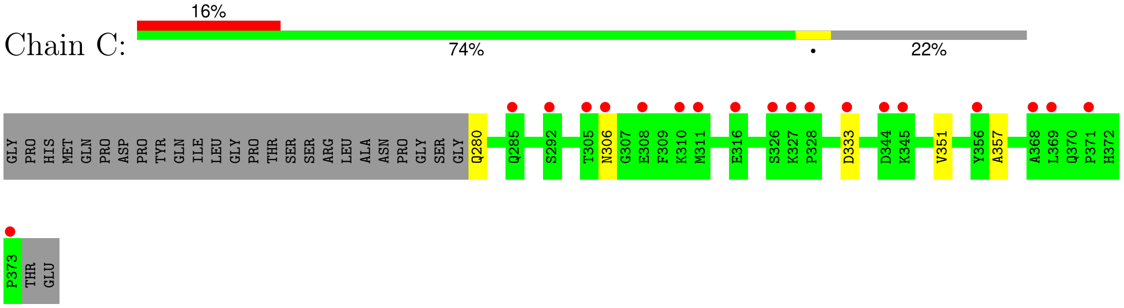
3 Residue-property plots ⓘ

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.

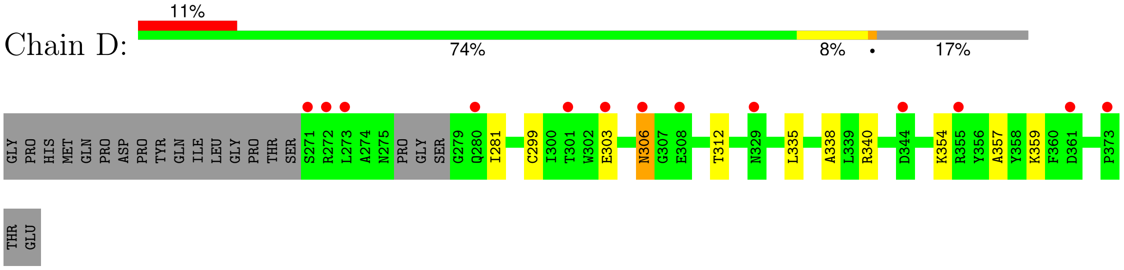
- Molecule 1: Friend leukemia integration 1 transcription factor



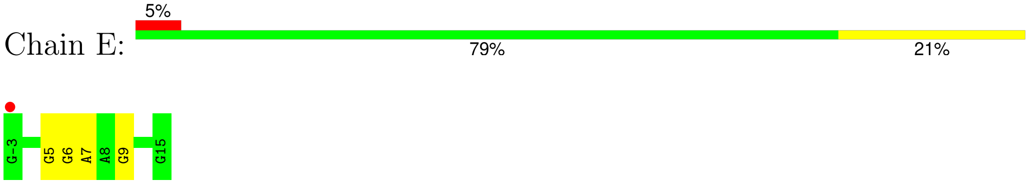
- Molecule 1: Friend leukemia integration 1 transcription factor



- Molecule 1: Friend leukemia integration 1 transcription factor

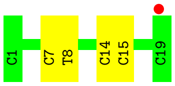


- Molecule 2: DNA (5'-D(P*GP*AP*CP*CP*GP*GP*AP*AP*GP*GP*AP*AP*GP*GP*AP*AP*GP*TP*G)-3')



- Molecule 3: DNA (5’-D(*CP*AP*CP*TP*TP*CP*CP*TP*TP*CP*CP*TP*TP*CP*CP*GP*GP*TP*C)-3’)

Chain F:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	91.54Å 91.54Å 180.73Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.67 – 3.15 39.67 – 3.15	Depositor EDS
% Data completeness (in resolution range)	65.2 (39.67-3.15) 65.2 (39.67-3.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	8.79 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.250 , 0.278 0.253 , 0.282	Depositor DCC
R_{free} test set	553 reflections (5.36%)	wwPDB-VP
Wilson B-factor (Å^2)	26.6	Xtriage
Anisotropy	0.881	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 57.1	EDS
L-test for twinning[2]	$< \|L\| >$ = 0.48, $< L^2 >$ = 0.31	Xtriage
Estimated twinning fraction	0.054 for -h,-k,l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	3190	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.50% 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)

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	0.64	0/863	0.71	0/1163
1	C	0.63	0/789	0.71	0/1065
1	D	0.62	0/826	0.70	0/1113
2	E	0.29	0/456	0.73	0/704
3	F	0.33	0/412	0.82	0/632
All	All	0.57	0/3346	0.73	0/4677

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	A	841	0	807	4	0
1	C	768	0	739	1	0
1	D	806	0	767	5	0
2	E	403	0	211	3	0
3	F	372	0	215	2	0
All	All	3190	0	2739	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.

All (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)
1:A:354:LYS:HB3	1:A:357:ALA:HB3	1.85	0.59
3:F:7:DC:H2'	3:F:8:DT:C6	2.39	0.57
1:C:351:VAL:HB	1:C:357:ALA:HB1	1.93	0.49
1:A:340:ARG:HD2	2:E:9:DG:N7	2.27	0.49
1:D:299:CYS:O	1:D:312:THR:N	2.45	0.49
1:D:306:ASN:HB3	1:D:359:LYS:HE3	1.97	0.45
3:F:14:DC:H2'	3:F:15:DC:O4'	2.18	0.44
1:A:265:ILE:HG21	1:A:302:TRP:HH2	1.82	0.43
1:D:335:LEU:O	1:D:338:ALA:HB3	2.19	0.43
1:D:354:LYS:HB3	1:D:357:ALA:CB	2.51	0.41
1:D:340:ARG:NH2	2:E:5:DG:N7	2.69	0.41
2:E:6:DG:C6	2:E:7:DA:C6	3.08	0.41
1:A:335:LEU:O	1:A:338:ALA:HB3	2.21	0.41

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	102/121 (84%)	92 (90%)	10 (10%)	0	100	100
1	C	92/121 (76%)	85 (92%)	7 (8%)	0	100	100
1	D	96/121 (79%)	91 (95%)	5 (5%)	0	100	100
All	All	290/363 (80%)	268 (92%)	22 (8%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	84/103 (82%)	84 (100%)	0	100	100
1	C	78/103 (76%)	75 (96%)	3 (4%)	28	56
1	D	81/103 (79%)	78 (96%)	3 (4%)	29	57
All	All	243/309 (79%)	237 (98%)	6 (2%)	42	67

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	280	GLN
1	C	306	ASN
1	C	333	ASP
1	D	281	ILE
1	D	303	GLU
1	D	306	ASN

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	285	GLN
1	D	285	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

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

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	106/121 (87%)	1.25	23 (21%)	3	2	29, 67, 117, 135	0
1	C	94/121 (77%)	1.19	19 (20%)	3	2	34, 73, 99, 143	0
1	D	100/121 (82%)	0.99	13 (13%)	9	6	29, 62, 94, 121	0
2	E	19/19 (100%)	-0.10	1 (5%)	33	21	26, 33, 46, 60	0
3	F	19/19 (100%)	0.13	1 (5%)	33	21	27, 35, 57, 67	0
All	All	338/401 (84%)	1.02	57 (16%)	5	3	26, 65, 107, 143	0

All (57) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	271	SER	4.8
1	D	306	ASN	4.5
1	A	262	PRO	4.2
1	A	353	GLY	4.0
1	A	305	THR	3.8
1	D	272	ARG	3.8
1	A	361	ASP	3.7
1	C	333	ASP	3.7
1	C	327	LYS	3.6
1	A	272	ARG	3.6
1	D	280	GLN	3.5
1	C	305	THR	3.4
1	C	306	ASN	3.3
1	A	263	TYR	3.2
1	C	345	LYS	3.1
1	A	344	ASP	3.1
1	A	273	LEU	3.0
1	C	369	LEU	3.0
1	A	267	GLY	3.0
1	A	268	PRO	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	344	ASP	2.9
1	A	323	GLU	2.9
1	A	349	THR	2.8
1	A	373	PRO	2.7
1	A	293	ASP	2.7
1	D	308	GLU	2.7
1	A	303	GLU	2.7
1	D	355	ARG	2.7
1	C	316	GLU	2.6
1	D	344	ASP	2.6
1	A	355	ARG	2.6
1	D	303	GLU	2.5
1	A	265	ILE	2.5
1	C	285	GLN	2.5
1	C	368	ALA	2.5
1	A	306	ASN	2.4
1	C	328	PRO	2.4
3	F	19	DC	2.4
1	C	292	SER	2.3
1	D	373	PRO	2.3
1	C	308	GLU	2.3
2	E	-3	DG	2.3
1	D	301	THR	2.3
1	C	311	MET	2.3
1	D	329	ASN	2.2
1	C	373	PRO	2.2
1	A	352	HIS	2.2
1	C	356	TYR	2.2
1	D	361	ASP	2.2
1	A	269	THR	2.2
1	C	310	LYS	2.1
1	C	371	PRO	2.1
1	A	285	GLN	2.1
1	D	273	LEU	2.0
1	A	281	ILE	2.0
1	C	326	SER	2.0
1	A	372	HIS	2.0

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.