wwPDB X-ray Structure Validation Summary Report (i)

Mar 3, 2025 – 08:02 AM EST

PDB ID	:	9MWY
Title	:	Crystal structure of the DNA binding domain of FLI1 in complex with a DNA containing four contiguous GGAA sites
Authors	:	Hou, C.; Tsodikov, O.V.
Deposited on	:	2025-01-17
Resolution	:	3.28 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

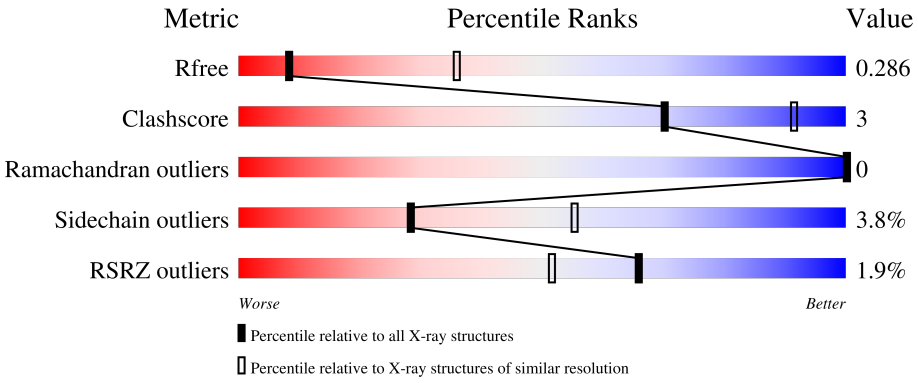
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.28 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	1214 (3.30-3.26)
Clashscore	180529	1265 (3.30-3.26)
Ramachandran outliers	177936	1264 (3.30-3.26)
Sidechain outliers	177891	1263 (3.30-3.26)
RSRZ outliers	164620	1215 (3.30-3.26)

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	145	54% 11% 34%
1	C	145	60% 7% 33%
1	D	145	61% 8% 32%
1	L	145	59% 38%
2	G	25	64% 36%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	25	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4115 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	96	Total	C	N	O	S	0	0	0
			756	477	134	141	4			
1	D	99	Total	C	N	O	S	0	0	0
			811	514	147	146	4			
1	C	97	Total	C	N	O	S	0	0	0
			797	507	144	142	4			
1	L	90	Total	C	N	O	S	0	0	0
			733	466	130	133	4			

There are 20 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
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
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
L	255	GLY	-	expression tag	UNP Q01543
L	256	PRO	-	expression tag	UNP Q01543
L	257	HIS	-	expression tag	UNP Q01543
L	258	MET	-	expression tag	UNP Q01543
L	362	ALA	PHE	engineered mutation	UNP Q01543

- Molecule 2 is a DNA chain called 25-mer DNA containing four contiguous GGAA sites, bottom strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	25	Total	C	N	O	P	0	0	0
			526	247	116	139	24			

- Molecule 3 is a DNA chain called 25-mer DNA containing four contiguous GGAA sites, top strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	25	Total	C	N	O	P	0	0	0
			490	237	74	155	24			

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

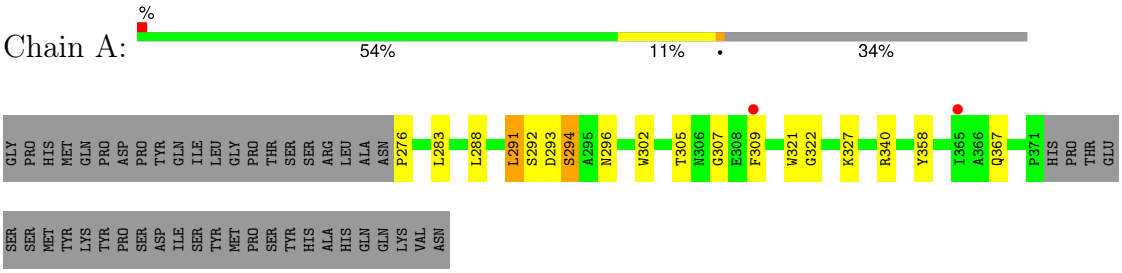
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Na	0	0
			1	1		
4	C	1	Total	Na	0	0
			1	1		

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.

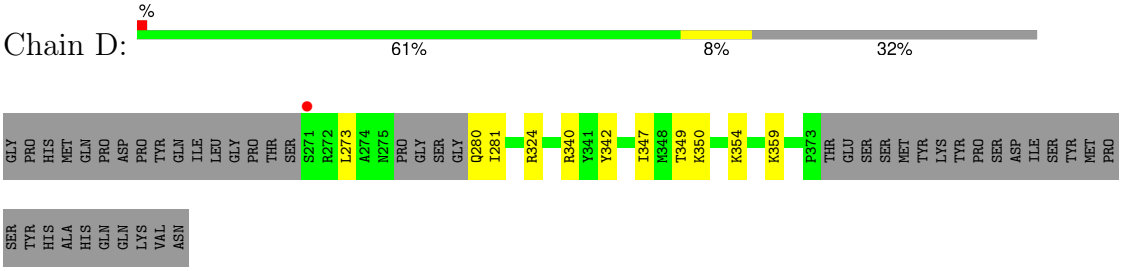
• Molecule 1: Friend leukemia integration 1 transcription factor

Chain A:



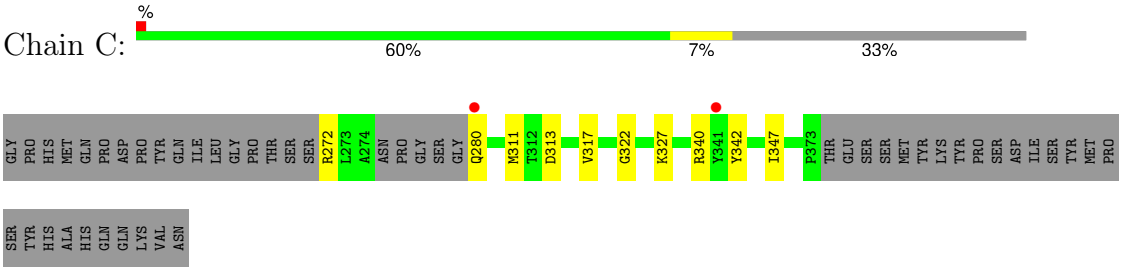
• Molecule 1: Friend leukemia integration 1 transcription factor

Chain D:



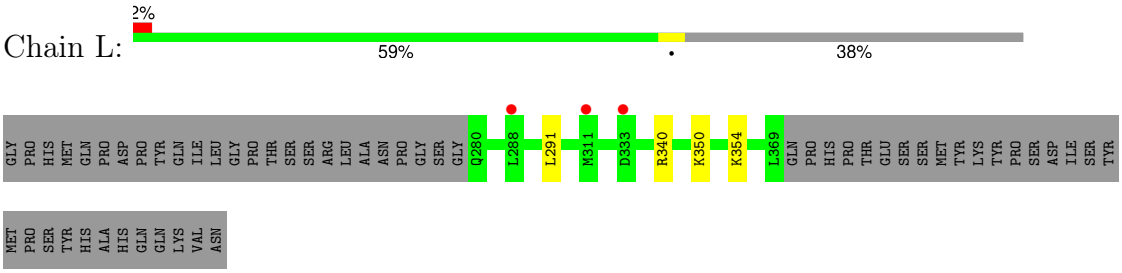
• Molecule 1: Friend leukemia integration 1 transcription factor

Chain C:



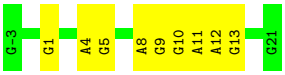
• Molecule 1: Friend leukemia integration 1 transcription factor

Chain L:



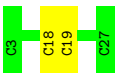
• Molecule 2: 25-mer DNA containing four contiguous GGAA sites, bottom strand

Chain G:



• Molecule 3: 25-mer DNA containing four contiguous GGAA sites, top strand

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	78.20Å 78.20Å 318.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 3.28 40.00 – 3.28	Depositor EDS
% Data completeness (in resolution range)	46.9 (40.00-3.28) 46.9 (40.00-3.28)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.16 (at 3.25Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.214 , 0.281 0.220 , 0.286	Depositor DCC
R_{free} test set	782 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	54.4	Xtriage
Anisotropy	0.542	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 93.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	4115	wwPDB-VP
Average B, all atoms (Å^2)	106.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.46% 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: NA

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	0.66	0/774	0.75	1/1043 (0.1%)
1	C	0.64	0/817	0.72	0/1101
1	D	0.64	0/831	0.70	0/1120
1	L	0.66	0/751	0.70	0/1012
2	G	0.25	0/596	0.72	0/921
3	H	0.27	0/543	0.80	0/833
All	All	0.57	0/4312	0.73	1/6030 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	276	PRO	N-CA-CB	5.53	109.94	103.30

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	756	0	713	9	0
1	C	797	0	772	4	0
1	D	811	0	783	5	0
1	L	733	0	700	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	G	526	0	276	9	0
3	H	490	0	282	1	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
All	All	4115	0	3526	25	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 3.

The worst 5 of 25 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:340:ARG:NH2	2:G:13:DG:N7	2.42	0.67
1:A:288:LEU:HD23	1:A:291:LEU:HD11	1.83	0.60
1:D:340:ARG:NH2	2:G:5:DG:N7	2.53	0.56
1:A:302:TRP:CE3	1:A:307:GLY:HA2	2.43	0.54
1:L:350:LYS:NZ	1:L:354:LYS:O	2.41	0.53

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	94/145 (65%)	82 (87%)	12 (13%)	0	100	100
1	C	93/145 (64%)	84 (90%)	9 (10%)	0	100	100
1	D	95/145 (66%)	88 (93%)	7 (7%)	0	100	100
1	L	88/145 (61%)	81 (92%)	7 (8%)	0	100	100
All	All	370/580 (64%)	335 (90%)	35 (10%)	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	75/126 (60%)	69 (92%)	6 (8%)	10 32
1	C	81/126 (64%)	78 (96%)	3 (4%)	29 56
1	D	83/126 (66%)	81 (98%)	2 (2%)	44 67
1	L	73/126 (58%)	72 (99%)	1 (1%)	62 78
All	All	312/504 (62%)	300 (96%)	12 (4%)	28 55

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	280	GLN
1	C	272	ARG
1	L	291	LEU
1	C	280	GLN
1	A	294	SER

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

Mol	Chain	Res	Type
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 (i)

Of 2 ligands modelled in this entry, 2 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)

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	96/145 (66%)	0.04	2 (2%)	63	49	38, 130, 173, 310	0
1	C	97/145 (66%)	-0.03	2 (2%)	63	49	61, 95, 135, 222	0
1	D	99/145 (68%)	-0.21	1 (1%)	79	67	53, 89, 140, 173	0
1	L	90/145 (62%)	-0.08	3 (3%)	49	36	81, 129, 168, 176	0
2	G	25/25 (100%)	-0.73	0	100	100	56, 81, 119, 134	0
3	H	25/25 (100%)	-0.62	0	100	100	55, 93, 113, 123	0
All	All	432/630 (68%)	-0.14	8 (1%)	66	51	38, 110, 164, 310	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	365	ILE	3.6
1	D	271	SER	3.0
1	L	333	ASP	2.4
1	C	280	GLN	2.4
1	L	288	LEU	2.4

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	NA	A	401	1/1	0.95	0.07	18,18,18,18	0
4	NA	C	401	1/1	0.96	0.16	18,18,18,18	0

6.5 Other polymers

There are no such residues in this entry.