wwPDB X-ray Structure Validation Summary Report (i)

Nov 11, 2026 – 06:08 pm GMT

PDB ID	:	31CY / pdb_000031cy
Title	:	Transcription factors FLI1 and ERG bound to specific DNA
Authors	:	Morgunova, E.; Popov, A.; Taipale, J.
Deposited on	:	2026-05-26
Resolution	:	3.29 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 (i)) were used in the production of this report:

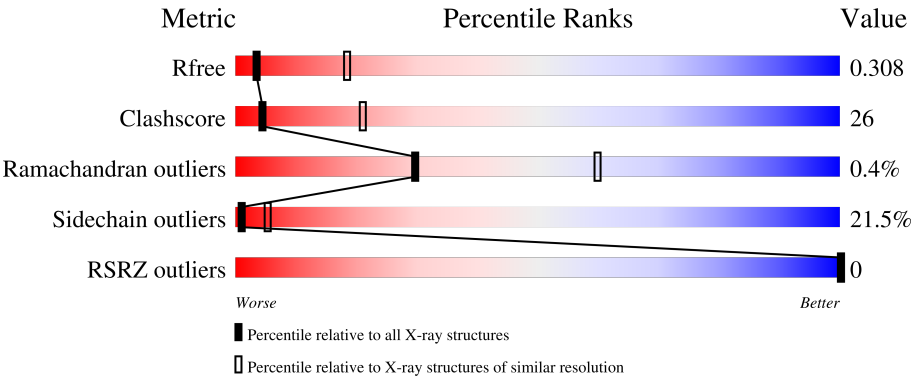
MolProbity	:	4-5-2 with Phenix2.2
Xtriage (Phenix)	:	2.2
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.51

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 3.29 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}	180053	1169 (3.32-3.28)
Clashscore	190562	1209 (3.32-3.28)
Ramachandran outliers	187476	1188 (3.32-3.28)
Sidechain outliers	187428	1187 (3.32-3.28)
RSRZ outliers	180081	1169 (3.32-3.28)

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	93	48% 47% .
1	G	93	42% 48% 9% .
2	B	17	24% 53% 24%
3	C	17	18% 76% 6%
3	M	17	29% 59% 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	92	41% 47% 11%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3369 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 Transcriptional regulator ERG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	93	Total	C	N	O	S	0	0	0
			764	488	134	138	4			
1	G	93	Total	C	N	O	S	0	0	0
			774	492	136	142	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	17	Total	C	N	O	P	0	0	0
			347	165	66	99	17			

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*AP*CP*AP*GP*GP*AP*TP*TP*TP *CP*CP*GP*GP*TP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	17	Total	C	N	O	P	0	0	0
			350	167	61	105	17			
3	M	17	Total	C	N	O	P	0	0	0
			350	167	61	105	17			

- Molecule 4 is a protein called Friend leukemia integration 1 transcription factor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	92	Total	C	N	O	S	0	0	0
			760	484	133	139	4			

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	3	Total	O	0	0
			3	3		

Continued on next page...

Continued from previous page...

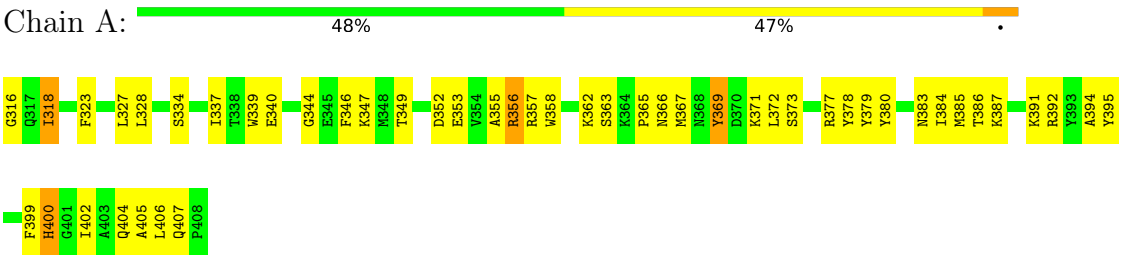
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	4	Total O 4 4	0	0
5	C	8	Total O 8 8	0	0
5	D	5	Total O 5 5	0	0
5	G	4	Total O 4 4	0	0

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.

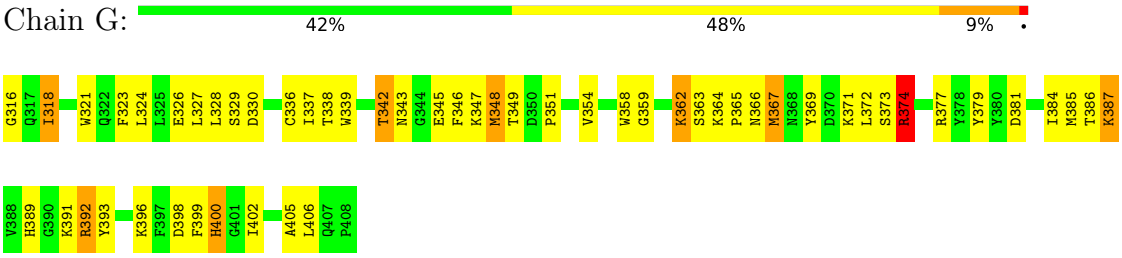
• Molecule 1: Transcriptional regulator ERG

Chain A:



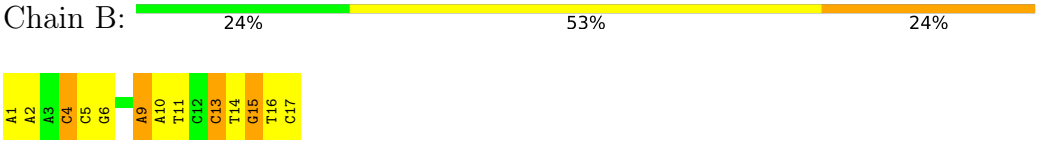
• Molecule 1: Transcriptional regulator ERG

Chain G:



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

Chain B:



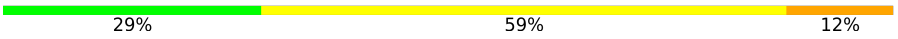
• Molecule 3: DNA (5'-D(P*GP*AP*CP*AP*GP*GP*AP*TP*TP*TP*CP*CP*GP*GP*TP*TP*T)-3')

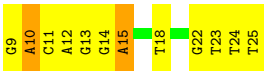
Chain C:



• Molecule 3: DNA (5'-D(P*GP*AP*CP*AP*GP*GP*AP*TP*TP*TP*CP*CP*GP*GP*TP*TP*T)-3')

Chain M:

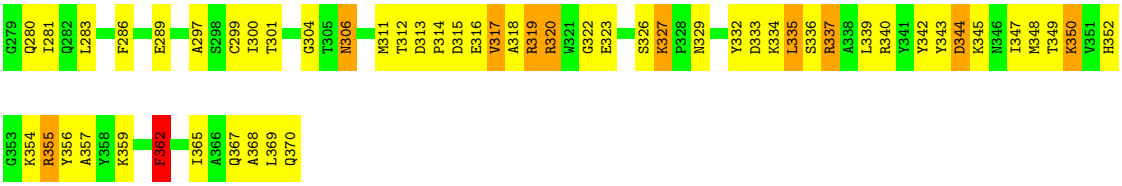




• Molecule 4: Friend leukemia integration 1 transcription factor

Chain D:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	55.92Å 265.45Å 105.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.63 – 3.29 48.63 – 3.29	Depositor EDS
% Data completeness (in resolution range)	98.9 (48.63-3.29) 98.9 (48.63-3.29)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 3.33Å)	Xtriage
Refinement program	REFMAC 5.8.0419	Depositor
R, R_{free}	0.234 , 0.306 0.233 , 0.308	Depositor DCC
R_{free} test set	644 reflections (5.20%)	wwPDB-VP
Wilson B-factor (Å^2)	168.3	Xtriage
Anisotropy	0.461	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 303.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3369	wwPDB-VP
Average B, all atoms (Å^2)	232.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.10% 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

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	0.51	0/784	1.22	2/1056 (0.2%)
1	G	0.50	0/794	1.19	3/1068 (0.3%)
2	B	0.77	0/389	1.42	5/597 (0.8%)
3	C	0.82	0/391	1.34	4/602 (0.7%)
3	M	0.90	0/391	1.51	5/602 (0.8%)
4	D	0.55	0/779	1.39	6/1049 (0.6%)
All	All	0.64	0/3528	1.33	25/4974 (0.5%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	G	0	1

There are no bond length outliers.

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	M	10	DA	O5'-C5'-C4'	11.98	118.99	109.40
4	D	362	PHE	CA-CB-CG	-8.62	105.17	113.80
3	M	9	DG	O4'-C1'-N9	7.91	113.54	108.00
3	C	15	DA	O4'-C1'-N9	6.96	112.88	108.00
1	G	330	ASP	CB-CA-C	6.38	120.06	109.53

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	374	ARG	Sidechain

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	764	0	730	26	0
1	G	774	0	748	40	0
2	B	347	0	191	23	0
3	C	350	0	194	22	0
3	M	350	0	194	29	3
4	D	760	0	725	40	0
5	A	3	0	0	0	0
5	B	4	0	0	1	0
5	C	8	0	0	3	0
5	D	5	0	0	1	0
5	G	4	0	0	1	0
All	All	3369	0	2782	161	3

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 26.

The worst 5 of 161 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:G:374:ARG:HH11	1:G:374:ARG:CA	1.72	1.03
2:B:13:DC:H2"	2:B:14:DT:O5'	1.67	0.94
3:C:11:DC:C5'	5:C:103:HOH:O	2.17	0.92
1:G:374:ARG:HH11	1:G:374:ARG:HA	1.32	0.91
3:C:9:DG:H4'	3:C:10:DA:OP1	1.75	0.86

All (3) 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)
3:M:10:DA:O5'	3:M:10:DA:O5'[6_345]	1.05	1.15
3:M:10:DA:OP2	3:M:10:DA:OP2[6_345]	2.15	0.05
3:M:10:DA:P	3:M:10:DA:O5'[6_345]	2.19	0.01

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	91/93 (98%)	85 (93%)	6 (7%)	0	100 100
1	G	91/93 (98%)	86 (94%)	5 (6%)	0	100 100
4	D	90/92 (98%)	79 (88%)	10 (11%)	1 (1%)	11 39
All	All	272/278 (98%)	250 (92%)	21 (8%)	1 (0%)	30 60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	317	VAL

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	78/82 (95%)	64 (82%)	14 (18%)	2 8
1	G	82/82 (100%)	63 (77%)	19 (23%)	1 4
4	D	77/79 (98%)	59 (77%)	18 (23%)	1 4
All	All	237/243 (98%)	186 (78%)	51 (22%)	1 5

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	362	PHE
1	G	342	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	G	393	TYR
4	D	367	GLN
1	G	318	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	383	ASN
1	A	404	GLN
1	A	407	GLN
1	G	383	ASN
1	G	404	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 oligosaccharides 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, 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	93/93 (100%)	-0.60	0	100	100	167, 211, 264, 293	2 (2%)
1	G	93/93 (100%)	-0.62	0	100	100	165, 252, 328, 395	0
2	B	17/17 (100%)	-1.13	0	100	100	178, 194, 274, 281	0
3	C	17/17 (100%)	-1.04	0	100	100	158, 211, 268, 308	0
3	M	17/17 (100%)	-0.98	0	100	100	221, 300, 512, 539	0
4	D	92/92 (100%)	-0.73	0	100	100	135, 194, 262, 296	1 (1%)
All	All	329/329 (100%)	-0.71	0	100	100	135, 219, 308, 539	3 (0%)

There are no RSRZ outliers to report.

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 oligosaccharides 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.