



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 16, 2020 – 06:27 PM EST

PDB ID : 6VGD  
Title : Crystal structure of the DNA binding domain (DBD) of human FLI1 and the complex of the DBD of human Runx2 with core binding factor beta (Cbfb), in complex with 16mer DNA CAGAGGATGTGGCTTC  
Authors : Hou, C.; Tsodikov, O.V.  
Deposited on : 2020-01-07  
Resolution : 4.20 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](mailto: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	:	2.14.6
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)	:	2.14.6

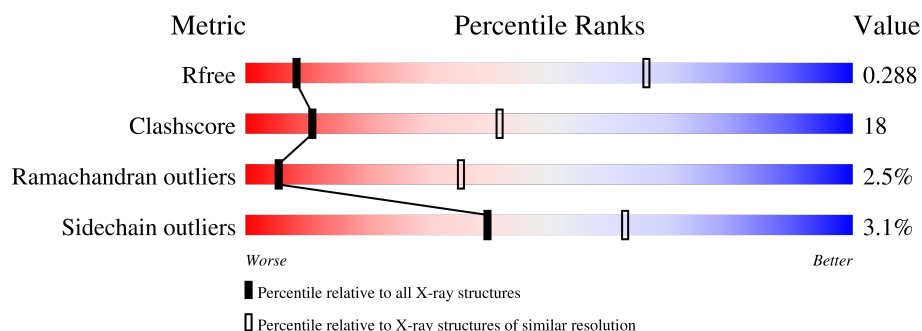
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.20 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	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	104	
2	B	16	
3	C	16	
4	D	177	
5	G	156	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3359 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	98	Total	C	N	O	S	0	0	0
			804	512	143	144	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	272	GLY	-	expression tag	UNP Q01543
A	273	PRO	-	expression tag	UNP Q01543
A	274	HIS	-	expression tag	UNP Q01543
A	275	MET	-	expression tag	UNP Q01543

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	16	Total	C	N	O	P	0	0	0
			332	157	62	97	16			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			324	154	59	95	16			

- Molecule 4 is a protein called Runt-related transcription factor 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	117	Total	C	N	O	S	0	0	0
			901	567	164	166	4			

- Molecule 5 is a protein called Core-binding factor subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	G	121	Total	C	N	O	S	0	0	0
			998	622	181	189	6			

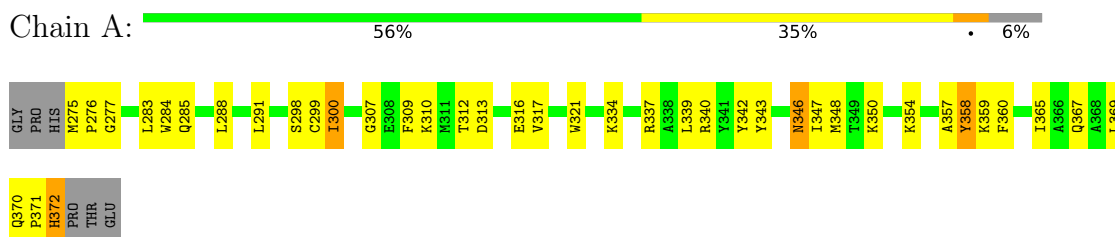
There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-13	MET	-	initiating methionine	UNP Q13951
G	-12	GLY	-	expression tag	UNP Q13951
G	-11	SER	-	expression tag	UNP Q13951
G	-10	SER	-	expression tag	UNP Q13951
G	-9	HIS	-	expression tag	UNP Q13951
G	-8	HIS	-	expression tag	UNP Q13951
G	-7	HIS	-	expression tag	UNP Q13951
G	-6	HIS	-	expression tag	UNP Q13951
G	-5	HIS	-	expression tag	UNP Q13951
G	-4	HIS	-	expression tag	UNP Q13951
G	-3	SER	-	expression tag	UNP Q13951
G	-2	GLN	-	expression tag	UNP Q13951
G	-1	ASP	-	expression tag	UNP Q13951
G	0	PRO	-	expression tag	UNP Q13951

### 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. 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. 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



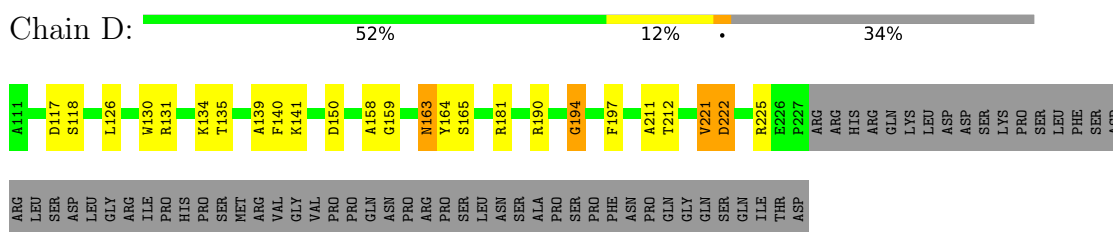
- Molecule 2: DNA (5'-D(P\*CP\*AP\*GP\*AP\*GP\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*TP\*TP\*C)-3')



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

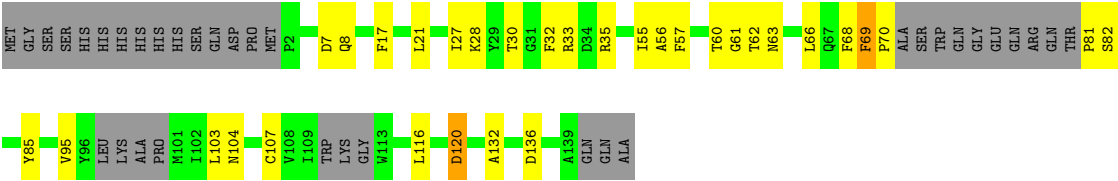


- Molecule 4: Runt-related transcription factor 2



- Molecule 5: Core-binding factor subunit beta





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	104.51Å 104.51Å 322.04Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	35.00 – 4.20 46.99 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.7 (35.00-4.20) 99.4 (46.99-3.25)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.24 (at 3.25Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.295 , 0.293 0.297 , 0.288	Depositor DCC
$R_{free}$ test set	869 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	80.2	Xtriage
Anisotropy	0.252	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.15 , 100.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.82	EDS
Total number of atoms	3359	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	306.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.79% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.65	0/826	0.85	0/1112
2	B	0.39	0/372	0.82	0/573
3	C	0.40	0/362	0.91	0/555
4	D	0.63	0/920	0.80	0/1253
5	G	0.65	0/1014	0.69	0/1355
All	All	0.60	0/3494	0.80	0/4848

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	804	0	777	41	0
2	B	332	0	181	13	0
3	C	324	0	180	20	0
4	D	901	0	905	23	0
5	G	998	0	956	38	0
All	All	3359	0	2999	110	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 18.

The worst 5 of 110 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:334:LYS:HA	3:C:11:DT:H72	1.22	1.13
1:A:275:MET:HG3	1:A:276:PRO:HD2	1.40	1.03
5:G:69:PHE:HB3	5:G:70:PRO:HD2	1.38	1.03
5:G:28:LYS:HG2	5:G:120:ASP:HB2	1.45	0.94
1:A:334:LYS:HA	3:C:11:DT:C7	2.00	0.92

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	96/104 (92%)	79 (82%)	15 (16%)	2 (2%)	7	39
4	D	115/177 (65%)	96 (84%)	14 (12%)	5 (4%)	2	25
5	G	113/156 (72%)	100 (88%)	12 (11%)	1 (1%)	17	56
All	All	324/437 (74%)	275 (85%)	41 (13%)	8 (2%)	5	35

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	135	THR
4	D	163	ASN
4	D	139	ALA
4	D	150	ASP
5	G	69	PHE

### 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	83/89 (93%)	80 (96%)	3 (4%)	35	60
4	D	100/156 (64%)	97 (97%)	3 (3%)	41	63
5	G	106/135 (78%)	103 (97%)	3 (3%)	43	65
All	All	289/380 (76%)	280 (97%)	9 (3%)	40	62

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

Mol	Chain	Res	Type
4	D	221	VAL
5	G	120	ASP
5	G	7	ASP
1	A	372	HIS
4	D	222	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
5	G	8	GLN
5	G	134	GLN
5	G	63	ASN
1	A	367	GLN
5	G	67	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 monosaccharides 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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.