



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:19 PM GMT

PDB ID : 4EFJ  
Title : Crystal structure of I-GzeII LAGLIDADG homing endonuclease in complex with DNA target site  
Authors : Kulshina, N.  
Deposited on : 2012-03-29  
Resolution : 2.80 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  
<http://wwpdb.org/validation/2016/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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

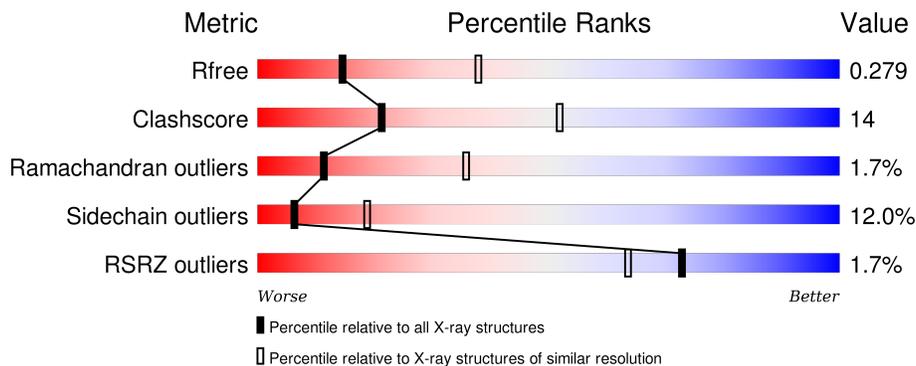
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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. 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	C	27	 33% 59% 7%
1	E	27	 41% 44% 15%
2	D	27	 4% 56% 33% 7% •
2	F	27	 48% 33% 15% •
3	A	300	 3% 63% 29% 5% •

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	B	300	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into three segments: a green segment on the left labeled '66%', a yellow segment in the middle labeled '25%', and a red segment on the right labeled '5%'. A small red square is positioned at the start of the bar, and a small black dot is at the end. A '%' symbol is located above the bar.</p>

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 6846 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 DNA chain called DNA target site top strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	27	565	267	111	160	27	0	0	0
1	E	27	562	267	111	158	26	0	0	0

- Molecule 2 is a DNA chain called DNA target site bottom strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	26	523	251	88	158	26	0	0	0
2	F	26	520	251	88	156	25	0	0	0

- Molecule 3 is a protein called LAGLIDADG endonuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	291	2329	1496	390	434	9	0	0	0
3	B	291	2314	1489	387	429	9	0	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Ca 1 1	0	0
4	A	2	Total Ca 2 2	0	0
4	C	1	Total Ca 1 1	0	0

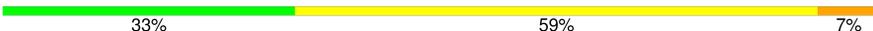
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	4	Total O 4 4	0	0
5	D	3	Total O 3 3	0	0
5	E	4	Total O 4 4	0	0
5	F	6	Total O 6 6	0	0
5	A	9	Total O 9 9	0	0
5	B	3	Total O 3 3	0	0

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: DNA target site top strand

Chain C: 



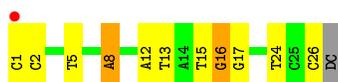
- Molecule 1: DNA target site top strand

Chain E: 



- Molecule 2: DNA target site bottom strand

Chain D: 



- Molecule 2: DNA target site bottom strand

Chain F: 



- Molecule 3: LAGLIDADG endonuclease

Chain A: 





• Molecule 3: LAGLIDADG endonuclease



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	125.83Å 61.18Å 128.88Å 90.00° 116.30° 90.00°	Depositor
Resolution (Å)	38.51 – 2.80 38.51 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.0 (38.51-2.80) 95.0 (38.51-2.80)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.51 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.183 , 0.278 0.182 , 0.279	Depositor DCC
$R_{free}$ test set	1126 reflections (5.69%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.1	Xtriage
Anisotropy	0.385	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 44.5	EDS
Estimated twinning fraction	0.000 for h,-k,-h-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	3 of 20852 reflections (0.014%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6846	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.52 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 6.7384e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	C	0.49	0/636	1.24	11/982 (1.1%)
1	E	0.59	1/633 (0.2%)	1.28	14/978 (1.4%)
2	D	0.57	1/583 (0.2%)	1.23	7/895 (0.8%)
2	F	0.53	0/580	1.31	15/891 (1.7%)
3	A	0.86	0/2383	0.93	4/3215 (0.1%)
3	B	0.81	0/2368	0.87	1/3196 (0.0%)
All	All	0.75	2/7183 (0.0%)	1.05	52/10157 (0.5%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	15	DT	O3'-P	-5.68	1.54	1.61
1	E	17	DG	O3'-P	-5.63	1.54	1.61

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	3	DT	P-O3'-C3'	13.59	136.01	119.70
2	D	16	DG	O5'-P-OP2	-9.86	96.83	105.70
1	C	13	DT	P-O3'-C3'	9.26	130.81	119.70
1	E	15	DT	OP1-P-O3'	-8.14	87.28	105.20
1	C	10	DC	P-O3'-C3'	7.89	129.16	119.70

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	C	565	0	304	6	0
1	E	562	0	305	12	0
2	D	523	0	295	17	0
2	F	520	0	296	7	0
3	A	2329	0	2264	90	0
3	B	2314	0	2244	67	0
4	A	2	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
5	A	9	0	0	9	0
5	B	3	0	0	0	0
5	C	4	0	0	0	0
5	D	3	0	0	1	0
5	E	4	0	0	3	0
5	F	6	0	0	0	0
All	All	6846	0	5708	178	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 14.

The worst 5 of 178 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)
3:A:196:ARG:HG2	3:A:232:VAL:HG13	1.40	1.02
3:B:196:ARG:HG2	3:B:232:VAL:HG13	1.50	0.94
2:D:17:DG:H3'	3:B:23:MET:CE	2.01	0.90
2:D:17:DG:H3'	3:B:23:MET:HE3	1.56	0.87
3:B:134:ALA:HB3	3:B:152:ILE:HG23	1.61	0.81

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	
3	A	289/300 (96%)	255 (88%)	29 (10%)	5 (2%)	11	36
3	B	289/300 (96%)	255 (88%)	29 (10%)	5 (2%)	11	36
All	All	578/600 (96%)	510 (88%)	58 (10%)	10 (2%)	11	36

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	120	GLU
3	A	123	ALA
3	B	122	GLY
3	B	161	ASN
3	B	226	ASN

### 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	
3	A	256/279 (92%)	232 (91%)	24 (9%)	11	31
3	B	252/279 (90%)	215 (85%)	37 (15%)	4	11
All	All	508/558 (91%)	447 (88%)	61 (12%)	6	19

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	45	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	B	93	LEU
3	B	219	PHE
3	B	47	ASP
3	B	63	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	205	GLN
3	B	59	ASN
3	B	161	ASN
3	A	166	HIS
3	B	92	ASN

### 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](#)

Of 4 ligands modelled in this entry, 4 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

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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	C	27/27 (100%)	-0.27	0 <a href="#">100</a> <a href="#">100</a>	29, 52, 84, 130	0
1	E	27/27 (100%)	-0.25	0 <a href="#">100</a> <a href="#">100</a>	29, 53, 70, 95	0
2	D	26/27 (96%)	-0.42	1 (3%) <a href="#">44</a> <a href="#">32</a>	32, 49, 79, 107	0
2	F	26/27 (96%)	-0.36	0 <a href="#">100</a> <a href="#">100</a>	28, 47, 76, 94	0
3	A	291/300 (97%)	-0.20	9 (3%) <a href="#">52</a> <a href="#">40</a>	19, 38, 77, 130	0
3	B	291/300 (97%)	-0.25	2 (0%) <a href="#">89</a> <a href="#">84</a>	21, 41, 72, 95	0
All	All	688/708 (97%)	-0.24	12 (1%) <a href="#">73</a> <a href="#">63</a>	19, 41, 76, 130	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	120	GLU	7.0
3	A	121	VAL	7.0
3	B	123	ALA	6.1
3	A	119	TYR	4.3
3	A	29	ASN	3.8

### 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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	CA	B	401	1/1	0.96	0.17	0.52	37,37,37,37	0
4	CA	A	401	1/1	0.93	0.17	0.46	27,27,27,27	0
4	CA	A	402	1/1	0.99	0.15	-0.08	31,31,31,31	0
4	CA	C	101	1/1	0.97	0.11	-1.63	41,41,41,41	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.