



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:39 AM GMT

PDB ID : 3MLN
Title : DNA binding domain of Early B-cell Factor 1 (Ebf1) bound to DNA (crystal form II)
Authors : Treiber, N.; Treiber, T.; Zocher, G.; Grosschedl, R.
Deposited on : 2010-04-17
Resolution : 2.40 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
<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

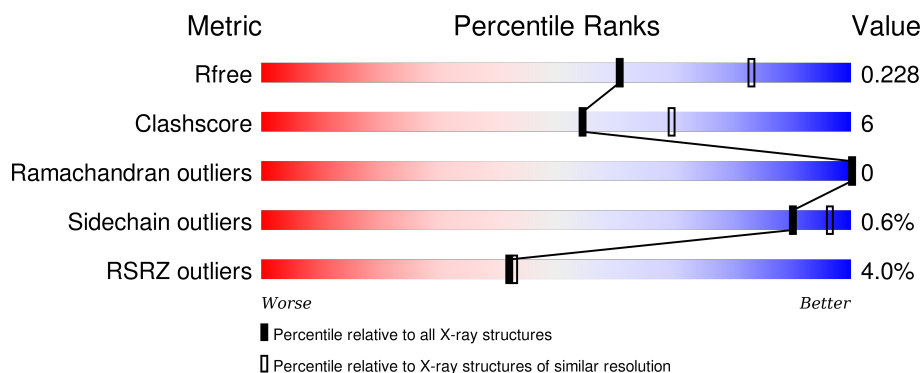
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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 ≥ 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	A	224	 2% 78% 14% 8%
1	B	224	 4% 87% 7% • 5%
1	E	224	 5% 64% 8% 27%
2	C	22	 59% 41%
2	D	22	 64% 36%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6075 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 Transcription factor COE1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	206	Total	C	N	O	S	0	7	0
			1690	1056	309	312	13			
1	B	213	Total	C	N	O	S	0	3	0
			1726	1080	313	320	13			
1	E	164	Total	C	N	O	S	0	2	0
			1316	840	236	232	8			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	242	GLU	-	EXPRESSION TAG	UNP Q07802
A	243	ASN	-	EXPRESSION TAG	UNP Q07802
A	244	LEU	-	EXPRESSION TAG	UNP Q07802
A	245	TYR	-	EXPRESSION TAG	UNP Q07802
A	246	PHE	-	EXPRESSION TAG	UNP Q07802
A	247	GLN	-	EXPRESSION TAG	UNP Q07802
B	242	GLU	-	EXPRESSION TAG	UNP Q07802
B	243	ASN	-	EXPRESSION TAG	UNP Q07802
B	244	LEU	-	EXPRESSION TAG	UNP Q07802
B	245	TYR	-	EXPRESSION TAG	UNP Q07802
B	246	PHE	-	EXPRESSION TAG	UNP Q07802
B	247	GLN	-	EXPRESSION TAG	UNP Q07802
E	242	GLU	-	EXPRESSION TAG	UNP Q07802
E	243	ASN	-	EXPRESSION TAG	UNP Q07802
E	244	LEU	-	EXPRESSION TAG	UNP Q07802
E	245	TYR	-	EXPRESSION TAG	UNP Q07802
E	246	PHE	-	EXPRESSION TAG	UNP Q07802
E	247	GLN	-	EXPRESSION TAG	UNP Q07802

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	22	Total 448	C 216	N 81	O 130	P 21	0	0	0
2	D	22	Total 448	C 216	N 81	O 130	P 21	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total 1	Zn 1	0	0
3	A	1	Total 1	Zn 1	0	0

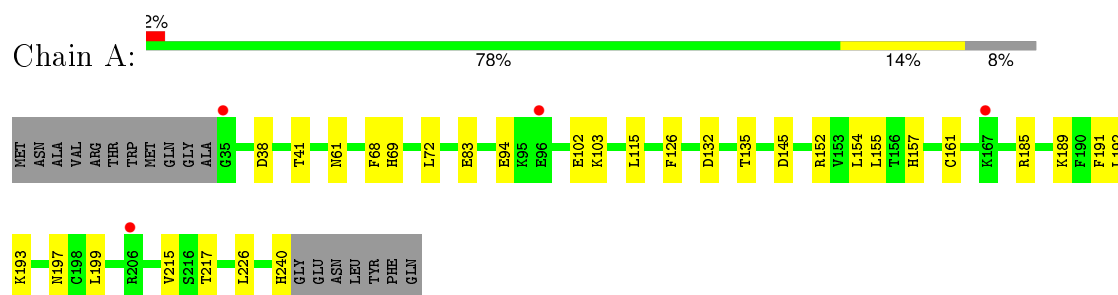
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	163	Total 163	O 163	0	0
4	B	115	Total 115	O 115	0	0
4	E	99	Total 99	O 99	0	0
4	C	34	Total 34	O 34	0	0
4	D	34	Total 34	O 34	0	0

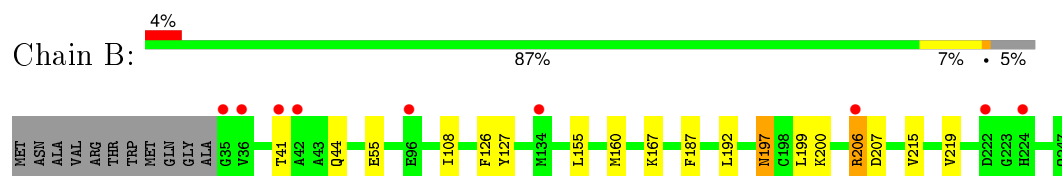
3 Residue-property plots [i](#)

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.

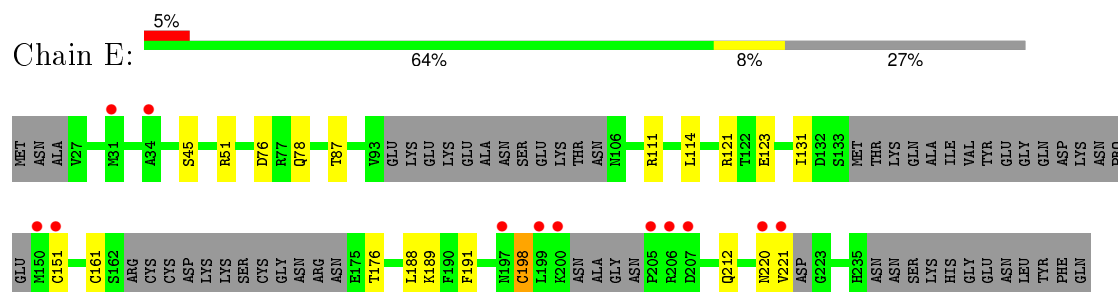
- Molecule 1: Transcription factor COE1



- Molecule 1: Transcription factor COE1



- Molecule 1: Transcription factor COE1



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	71.20Å 100.70Å 72.00Å 90.00° 101.20° 90.00°	Depositor
Resolution (Å)	29.31 – 2.40 29.31 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (29.31-2.40) 99.2 (29.31-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.57 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.171 , 0.219 0.185 , 0.228	Depositor DCC
R_{free} test set	1945 reflections (5.26%)	DCC
Wilson B-factor (Å ²)	33.1	Xtriage
Anisotropy	0.462	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 33.5	EDS
Estimated twinning fraction	0.019 for l,-k,h	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	2 of 38953 reflections (0.005%)	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6075	wwPDB-VP
Average B, all atoms (Å ²)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.28% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: ZN

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	0.54	0/1742	0.64	0/2344
1	B	0.51	0/1768	0.62	0/2380
1	E	0.54	1/1347 (0.1%)	0.66	0/1816
2	C	0.89	0/502	1.60	10/773 (1.3%)
2	D	0.96	0/502	1.45	6/773 (0.8%)
All	All	0.62	1/5861 (0.0%)	0.88	16/8086 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	198	CYS	CB-SG	-5.06	1.73	1.81

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	8	DC	O4'-C1'-N1	10.61	115.43	108.00
2	C	6	DT	O4'-C4'-C3'	-8.87	100.68	106.00
2	C	13	DG	O4'-C1'-N9	-7.97	102.42	108.00
2	C	12	DT	O4'-C1'-N1	-7.27	102.91	108.00
2	D	15	DG	O4'-C1'-N9	6.52	112.57	108.00

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	1690	0	1676	25	0
1	B	1726	0	1695	21	0
1	E	1316	0	1286	26	0
2	C	448	0	251	1	0
2	D	448	0	251	2	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	163	0	0	1	0
4	B	115	0	0	1	0
4	C	34	0	0	0	0
4	D	34	0	0	1	0
4	E	99	0	0	3	0
All	All	6075	0	5159	68	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 6.

The worst 5 of 68 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:154:LEU:HD22	1:A:192:LEU:CD2	1.75	1.16
1:A:154:LEU:HD22	1:A:192:LEU:HD21	1.11	1.07
1:E:151:CYS:SG	1:E:198:CYS:HB2	2.02	0.99
1:E:151:CYS:SG	1:E:198:CYS:CB	2.59	0.90
1:B:108:ILE:HD11	1:B:192:LEU:HD11	1.54	0.88

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	211/224 (94%)	208 (99%)	3 (1%)	0	100	100
1	B	214/224 (96%)	207 (97%)	7 (3%)	0	100	100
1	E	154/224 (69%)	150 (97%)	4 (3%)	0	100	100
All	All	579/672 (86%)	565 (98%)	14 (2%)	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	191/198 (96%)	191 (100%)	0	100	100
1	B	193/198 (98%)	190 (98%)	3 (2%)	70	86
1	E	143/198 (72%)	143 (100%)	0	100	100
All	All	527/594 (89%)	524 (99%)	3 (1%)	90	96

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	197	ASN
1	B	206	ARG
1	B	207	ASP

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

Mol	Chain	Res	Type
1	A	240	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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 ⓘ

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, 95th 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	206/224 (91%)	-0.18	4 (1%) 70 69	35, 45, 62, 72	0
1	B	213/224 (95%)	0.06	9 (4%) 40 41	41, 53, 78, 85	0
1	E	164/224 (73%)	0.19	12 (7%) 18 18	36, 50, 86, 116	0
2	C	22/22 (100%)	-0.30	0 100 100	43, 47, 69, 83	0
2	D	22/22 (100%)	-0.39	0 100 100	37, 46, 68, 83	0
All	All	627/716 (87%)	-0.01	25 (3%) 42 43	35, 50, 78, 116	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	35	GLY	5.6
1	B	35	GLY	4.4
1	E	31	MET	4.3
1	B	206	ARG	3.7
1	B	224	HIS	3.6

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, 95th 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(\AA^2)	Q<0.9
3	ZN	A	501	1/1	1.00	0.06	-2.49	46,46,46,46	0
3	ZN	B	501	1/1	0.99	0.06	-3.36	42,42,42,42	0

6.5 Other polymers [i](#)

There are no such residues in this entry.