



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 08:13 pm BST

PDB ID : 3RI4  
Title : Ets1 cooperative binding to widely separated sites on promoter DNA  
Authors : Babayeva, N.D.; Mino, K.; Tahirov, T.H.  
Deposited on : 2011-04-12  
Resolution : 3.00 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 : 2.11  
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) : 2.11

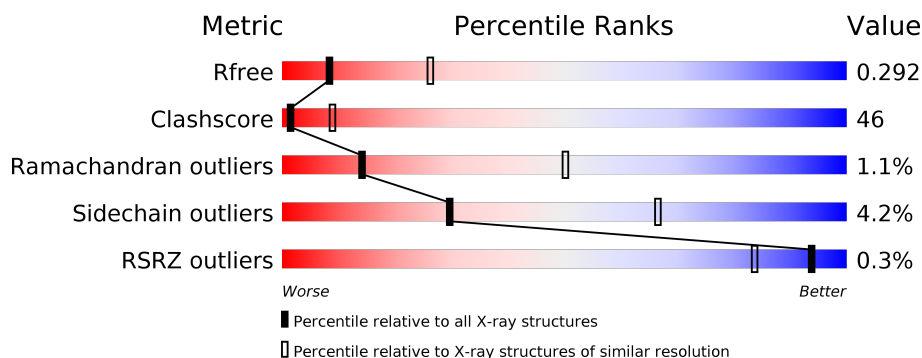
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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  $\leq 5\%$ . 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	163	
1	D	163	
2	B	16	
2	E	16	
3	C	16	
3	F	16	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3567 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 Protein C-ets-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	136	Total	C	N	O	S	0	0	0
			1117	720	193	200	4			
1	D	136	Total	C	N	O	S	0	0	0
			1117	720	193	200	4			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	279	MET	-	INITIATING METHIONINE	UNP P14921
A	288	TYR	SER	CONFLICT	UNP P14921
D	279	MET	-	INITIATING METHIONINE	UNP P14921
D	288	TYR	SER	CONFLICT	UNP P14921

- Molecule 2 is a DNA chain called TCR alpha promoter DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	16	Total	C	N	O	P	0	0	0
			321	154	59	93	15			
2	E	16	Total	C	N	O	P	0	0	0
			321	154	59	93	15			

- Molecule 3 is a DNA chain called TCR alpha promoter DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			329	157	62	95	15			
3	F	16	Total	C	N	O	P	0	0	0
			329	157	62	95	15			

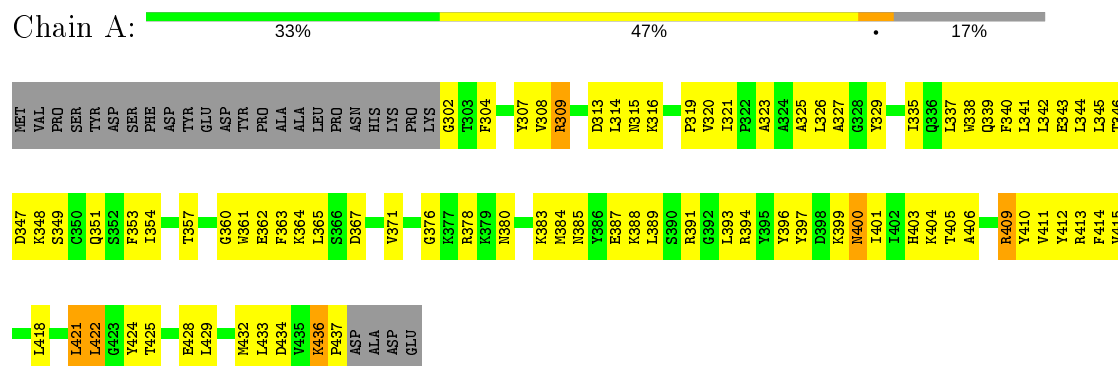
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	6	Total 6	O 6	0	0
4	B	6	Total 6	O 6	0	0
4	C	3	Total 3	O 3	0	0
4	D	8	Total 8	O 8	0	0
4	E	6	Total 6	O 6	0	0
4	F	4	Total 4	O 4	0	0

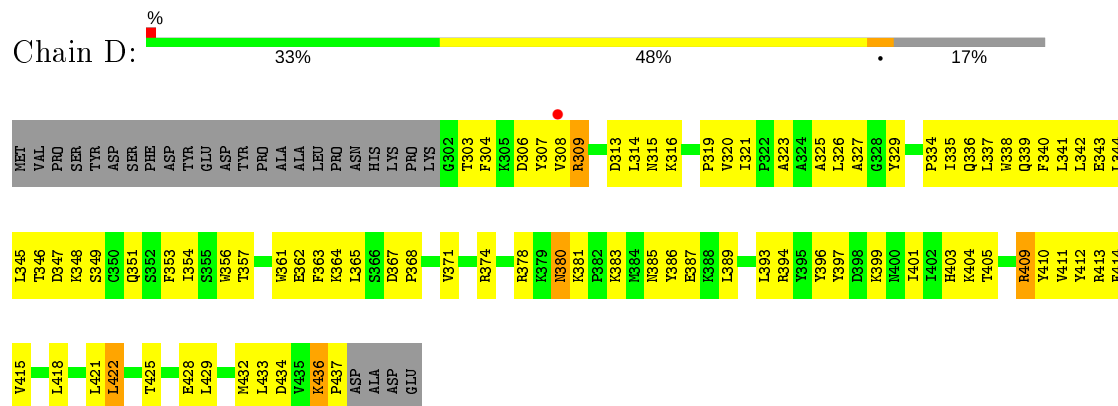
### 3 Residue-property plots [i](#)

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

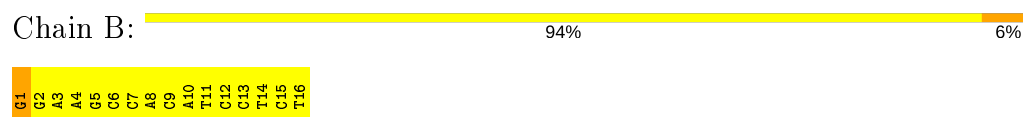
#### • Molecule 1: Protein C-ets-1



#### • Molecule 1: Protein C-ets-1



#### • Molecule 2: TCR alpha promoter DNA

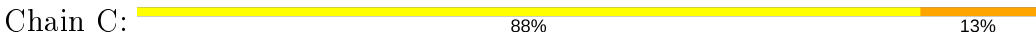


#### • Molecule 2: TCR alpha promoter DNA



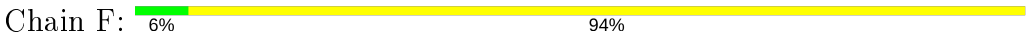
G201
G202
A203
A204
G205
C206
C207
A208
C209
A210
T211
C212
C213
T214
C215
T216

- Molecule 3: TCR alpha promoter DNA



G101
A102
G103
A104
G105
G106
A107
T108
G109
T110
G111
G112
C113
T114
T115
C116

- Molecule 3: TCR alpha promoter DNA



C301
A302
G303
A304
G305
G306
A307
T308
G309
T310
G311
G312
C313
T314
T315
C316

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.18Å 98.12Å 53.58Å 90.00° 109.67° 90.00°	Depositor
Resolution (Å)	29.84 – 3.00 29.84 – 2.99	Depositor EDS
% Data completeness (in resolution range)	94.5 (29.84-3.00) 93.1 (29.84-2.99)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	23.37 (at 3.00Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.230 , 0.284 0.239 , 0.292	Depositor DCC
$R_{free}$ test set	500 reflections (5.34%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.9	Xtriage
Anisotropy	0.291	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 32.1	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	0.89	EDS
Total number of atoms	3567	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.16% 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	0.44	0/1147	0.64	0/1548
1	D	0.48	0/1147	0.67	0/1548
2	B	0.57	0/359	0.83	0/551
2	E	0.57	0/359	0.78	0/551
3	C	0.62	0/369	0.83	1/569 (0.2%)
3	F	0.59	0/369	0.84	0/569
All	All	0.51	0/3750	0.73	1/5336 (0.0%)

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
2	B	0	1
2	E	0	1
3	C	0	1
3	F	0	1
All	All	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	104	DA	O4'-C1'-N9	-5.07	104.45	108.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	1	DG	Sidechain
3	C	103	DG	Sidechain

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Group
2	E	201	DG	Sidechain
3	F	303	DG	Sidechain

## 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	1117	0	1112	104	0
1	D	1117	0	1112	103	0
2	B	321	0	181	31	0
2	E	321	0	181	27	0
3	C	329	0	182	28	0
3	F	329	0	182	30	0
4	A	6	0	0	3	0
4	B	6	0	0	5	0
4	C	3	0	0	1	0
4	D	8	0	0	2	0
4	E	6	0	0	0	0
4	F	4	0	0	1	0
All	All	3567	0	2950	300	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 46.

The worst 5 of 300 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:F:312:DG:H1'	3:F:313:DC:H5'	1.30	1.08
3:F:304:DA:H2''	3:F:305:DG:H5''	1.07	1.03
1:A:436:LYS:HB3	1:A:437:PRO:HD3	1.43	0.96
1:D:319:PRO:HB3	1:D:347:ASP:HB2	1.45	0.96
1:D:436:LYS:HB3	1:D:437:PRO:HD3	1.44	0.96

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	134/163 (82%)	124 (92%)	8 (6%)	2 (2%)	10	42
1	D	134/163 (82%)	124 (92%)	9 (7%)	1 (1%)	22	60
All	All	268/326 (82%)	248 (92%)	17 (6%)	3 (1%)	14	50

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	400	ASN
1	D	436	LYS
1	A	436	LYS

### 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	118/142 (83%)	113 (96%)	5 (4%)	30	66
1	D	118/142 (83%)	113 (96%)	5 (4%)	30	66
All	All	236/284 (83%)	226 (96%)	10 (4%)	30	66

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	422	LEU
1	D	309	ARG
1	D	380	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	421	LEU
1	D	313	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	351	GLN
1	A	419	GLN
1	D	336	GLN
1	D	351	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 carbohydrates 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 [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	A	136/163 (83%)	-0.17	0 100 100	9, 57, 87, 94	0
1	D	136/163 (83%)	-0.20	1 (0%) 87 69	12, 51, 82, 89	0
2	B	16/16 (100%)	-0.51	0 100 100	15, 45, 61, 64	0
2	E	16/16 (100%)	-0.58	0 100 100	13, 45, 62, 77	0
3	C	16/16 (100%)	-0.64	0 100 100	19, 36, 53, 57	0
3	F	16/16 (100%)	-0.62	0 100 100	19, 38, 59, 65	0
All	All	336/390 (86%)	-0.26	1 (0%) 94 84	9, 53, 85, 94	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	308	VAL	2.1

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.