



# wwPDB X-ray Structure Validation Summary Report

Aug 17, 2022 – 03:31 PM EDT

PDB ID : 3WTV  
Title : Crystal structure of the complex comprised of ETS1(V170G), RUNX1, CBF-BETA, and the tcralpha gene enhancer DNA  
Authors : Shiina, M.; Hamada, K.; Ogata, K.  
Deposited on : 2014-04-21  
Resolution : 2.70 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 types of validation reports are described at <http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.29  
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) : 2.29

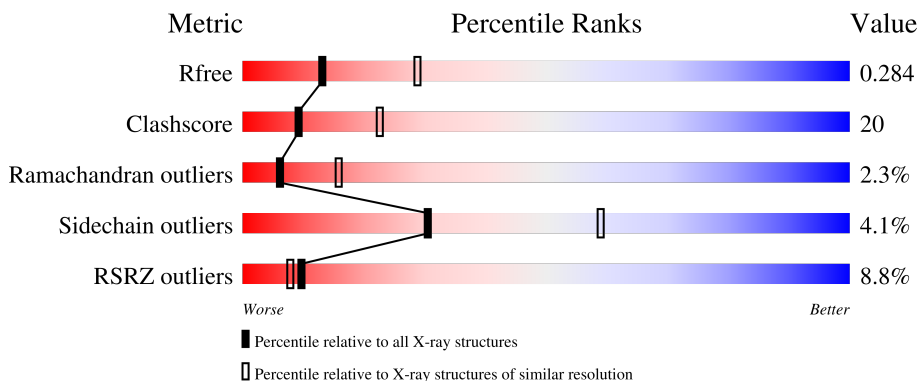
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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.



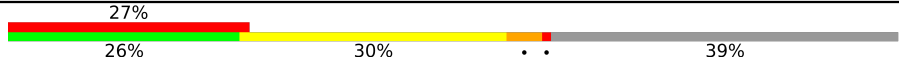
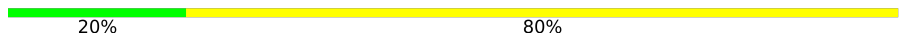
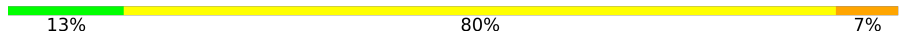


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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  $\geq 3$ , 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	204	
1	F	204	
2	B	142	
2	G	142	
3	C	166	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	H	166	
4	D	15	
4	I	15	
5	E	15	
5	J	15	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7002 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 Runt-related transcription factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	118	911	571	169	167	4	0	0	0
1	F	118	911	571	169	167	4	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	94	LYS	LEU	engineered mutation	UNP Q03347
A	170	GLY	VAL	engineered mutation	UNP Q03347
F	94	LYS	LEU	engineered mutation	UNP Q03347
F	170	GLY	VAL	engineered mutation	UNP Q03347

- Molecule 2 is a protein called Core-binding factor subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	130	1071	671	195	199	6	0	0	0
2	G	129	1062	666	193	197	6	0	0	0

- Molecule 3 is a protein called Protein C-ets-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	118	967	627	165	171	4	0	0	0
3	H	101	853	553	147	149	4	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(\*GP\*AP\*AP\*GP\*CP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			
4	I	15	Total	C	N	O	P	0	0	0
			299	144	54	87	14			

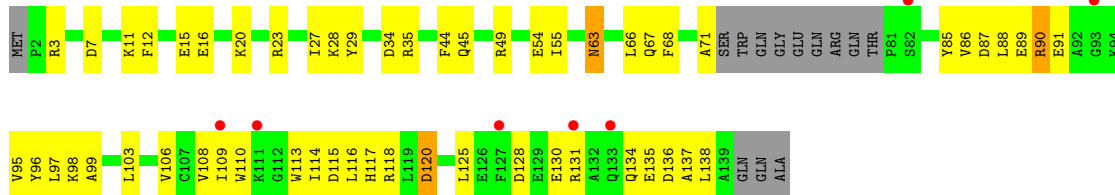
- Molecule 5 is a DNA chain called DNA (5'-D(\*AP\*GP\*AP\*GP\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*TP\*TP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			
5	J	15	Total	C	N	O	P	0	0	0
			310	148	59	89	14			

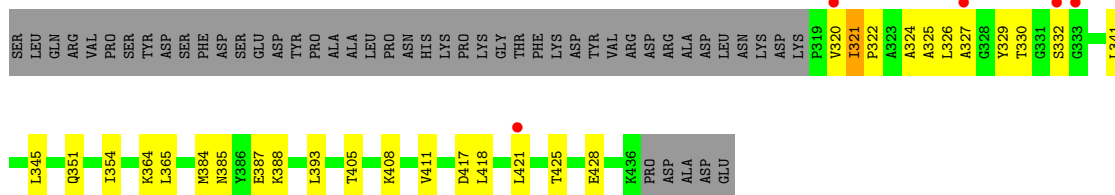
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	3	Total	O	0	0
			3	3		
6	F	4	Total	O	0	0
			4	4		
6	H	1	Total	O	0	0
			1	1		
6	J	1	Total	O	0	0
			1	1		

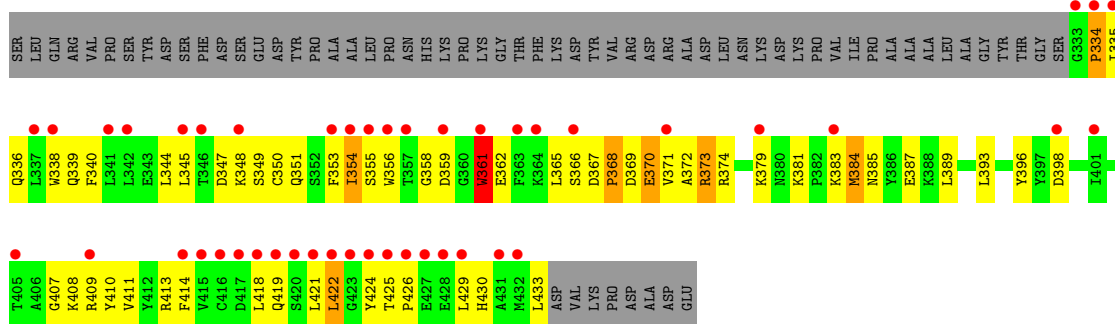
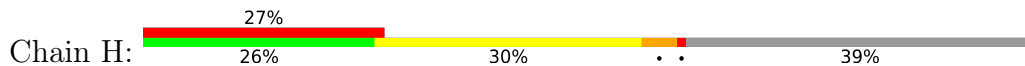




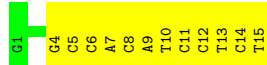
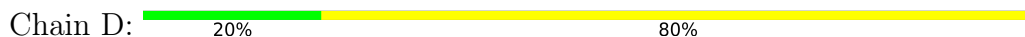
• Molecule 3: Protein C-ets-1



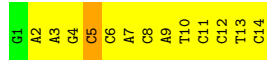
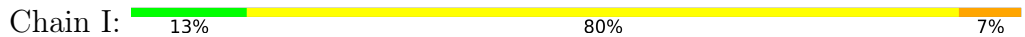
• Molecule 3: Protein C-ets-1




• Molecule 4: DNA (5'-D(\*GP\*AP\*AP\*GP\*CP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*T)-3')



• Molecule 4: DNA (5'-D(\*GP\*AP\*AP\*GP\*CP\*CP\*AP\*CP\*AP\*TP\*CP\*CP\*TP\*CP\*T)-3')



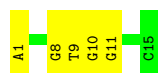
- Molecule 5: DNA (5'-D(\*AP\*GP\*AP\*GP\*GP\*AP\*TP\*GP\*TP\*GP\*GP\*CP\*TP\*TP\*C)-3')

Chain E:  80% 20%



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

Chain J:  67% 33%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.22Å 102.25Å 194.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.26 – 2.70 45.26 – 2.70	Depositor EDS
% Data completeness (in resolution range)	96.4 (45.26-2.70) 96.5 (45.26-2.70)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.62 (at 2.69Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.242 , 0.284 0.242 , 0.284	Depositor DCC
$R_{free}$ test set	4299 reflections (10.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	76.5	Xtrriage
Anisotropy	0.365	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 57.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7002	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	85.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.44% 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	0.42	0/930	0.58	0/1263
1	F	0.48	0/930	0.65	0/1263
2	B	0.34	0/1093	0.49	0/1466
2	G	0.37	0/1084	0.50	0/1454
3	C	0.44	0/994	0.56	0/1342
3	H	0.32	0/877	0.45	0/1181
4	D	0.54	0/334	0.87	0/512
4	I	0.57	0/334	0.83	0/512
5	E	0.50	0/348	0.81	0/537
5	J	0.55	0/348	0.85	0/537
All	All	0.43	0/7272	0.62	0/10067

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
4	D	0	2
4	I	0	2
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	4	DG	Sidechain
4	D	5	DC	Sidechain
4	I	4	DG	Sidechain
4	I	5	DC	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	911	0	916	41	0
1	F	911	0	916	26	0
2	B	1071	0	1034	45	0
2	G	1062	0	1026	47	0
3	C	967	0	966	23	0
3	H	853	0	847	57	0
4	D	299	0	170	13	0
4	I	299	0	170	20	0
5	E	310	0	171	3	0
5	J	310	0	171	6	0
6	C	3	0	0	0	0
6	F	4	0	0	0	0
6	H	1	0	0	0	0
6	J	1	0	0	0	0
All	All	7002	0	6387	266	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 20.

The worst 5 of 266 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)
4:I:2:DA:H2''	4:I:3:DA:H5'	1.31	1.10
4:I:9:DA:H2''	4:I:10:DT:H5'	1.36	1.03
4:I:11:DC:H2''	4:I:12:DC:H5'	1.39	1.02
3:H:424:TYR:HB2	3:H:429:LEU:HG	1.53	0.91
1:A:82:ASN:HD21	1:A:118:ARG:HH11	1.19	0.90

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	116/204 (57%)	109 (94%)	6 (5%)	1 (1%)	17	40
1	F	116/204 (57%)	109 (94%)	6 (5%)	1 (1%)	17	40
2	B	126/142 (89%)	98 (78%)	26 (21%)	2 (2%)	9	24
2	G	125/142 (88%)	103 (82%)	19 (15%)	3 (2%)	6	15
3	C	116/166 (70%)	112 (97%)	4 (3%)	0	100	100
3	H	99/166 (60%)	68 (69%)	22 (22%)	9 (9%)	1	0
All	All	698/1024 (68%)	599 (86%)	83 (12%)	16 (2%)	6	16

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	G	23	ARG
3	H	354	ILE
3	H	359	ASP
3	H	361	TRP
2	B	70	PRO

### 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	99/178 (56%)	94 (95%)	5 (5%)	24	50
1	F	99/178 (56%)	95 (96%)	4 (4%)	31	60
2	B	113/123 (92%)	108 (96%)	5 (4%)	28	56

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	G	112/123 (91%)	107 (96%)	5 (4%)	27	55
3	C	102/145 (70%)	100 (98%)	2 (2%)	55	81
3	H	91/145 (63%)	87 (96%)	4 (4%)	28	56
All	All	616/892 (69%)	591 (96%)	25 (4%)	30	59

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	96	ASP
2	G	63	ASN
3	H	398	ASP
2	G	7	ASP
2	G	90	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	112	ASN
2	G	8	GLN
3	H	419	GLN
1	F	163	HIS
2	G	46	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 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 [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	118/204 (57%)	0.32	1 (0%) 86 87	50, 72, 95, 110	0
1	F	118/204 (57%)	0.25	1 (0%) 86 87	47, 62, 82, 103	0
2	B	130/142 (91%)	0.80	9 (6%) 16 15	77, 101, 128, 135	0
2	G	129/142 (90%)	0.67	7 (5%) 25 24	63, 88, 115, 120	0
3	C	118/166 (71%)	0.41	5 (4%) 36 35	46, 69, 100, 111	0
3	H	101/166 (60%)	1.95	45 (44%) 0 0	71, 131, 173, 179	0
4	D	15/15 (100%)	-0.17	0 100 100	57, 60, 76, 78	0
4	I	15/15 (100%)	-0.14	0 100 100	57, 69, 84, 85	0
5	E	15/15 (100%)	-0.22	0 100 100	50, 61, 73, 77	0
5	J	15/15 (100%)	-0.05	0 100 100	54, 66, 79, 87	0
All	All	774/1084 (71%)	0.64	68 (8%) 10 8	46, 80, 143, 179	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	H	426	PRO	6.2
3	H	421	LEU	5.8
3	H	346	THR	5.5
3	H	423	GLY	5.5
3	H	416	CYS	4.9

### 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 monosaccharides 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.