



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 6, 2023 – 05:21 am GMT

PDB ID : 1HJB  
Title : CRYSTAL STRUCTURE OF RUNX-1/AML1/CBFALPHA RUNT DOMAIN AND C/EBPBETA BZIP HOMODIMER BOUND TO A DNA FRAGMENT FROM THE CSF-1R PROMOTER  
Authors : Tahirov, T.H.; Ogata, K.  
Deposited on : 2001-01-11  
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 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 : 2.36  
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) : 2.36

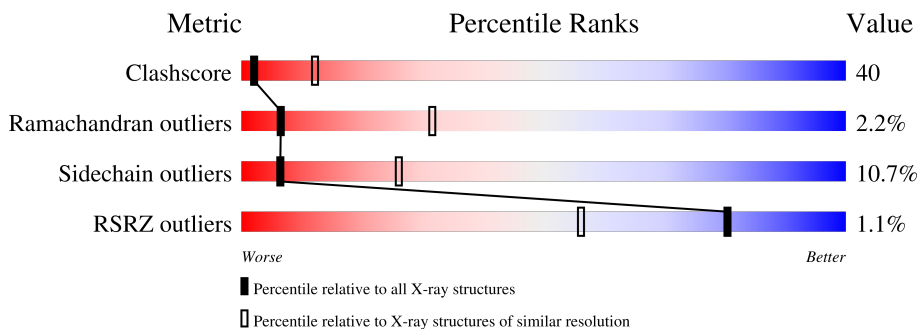
# 1 Overall quality at a glance

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.





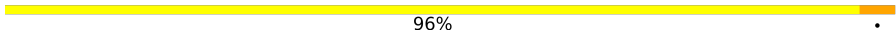
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 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	87	
1	B	87	
1	D	87	
1	E	87	
2	C	123	
2	F	123	
3	G	26	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	I	26	 12% 88%
4	H	26	 8% 92%
4	J	26	 96% .

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 6270 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 CCAAT/ENHANCER BINDING PROTEIN BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	66	Total 560	C 341	N 114	O 104	S 1	0	0	0
1	B	67	Total 568	C 347	N 116	O 104	S 1	0	0	0
1	D	68	Total 579	C 354	N 117	O 107	S 1	0	0	0
1	E	68	Total 575	C 352	N 117	O 105	S 1	0	0	0

- Molecule 2 is a protein called RUNT-RELATED TRANSCRIPTION FACTOR 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	120	Total 934	C 586	N 175	O 169	S 4	0	0	0
2	F	120	Total 934	C 586	N 175	O 169	S 4	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	G	26	Total 532	C 255	N 96	O 156	P 25	0	0	0
3	I	26	Total 532	C 255	N 96	O 156	P 25	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	H	26	Total 528	C 253	N 98	O 152	P 25	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	J	26	528	253	98	152	25	0	0	0



PRO  
LEU  
LEU  
ALA  
SER  
SER  
GLY  
HIS  
CYS

- Molecule 2: RUNT-RELATED TRANSCRIPTION FACTOR 1

Chain C: 42% 51%

G60 E61 L62 V63 R64 T65 D66 F70 S73 V74 L75 F76 I77 H78 W79 R80 C81 T84 L85 P86 I87 A88 F89 R90 V91 V92 V97 P98 D99 L102 V105 M106 M109 D110 E111 M112 E116 L117 R118 M119 K125 M126 Q127 V128 A129 R130 F131 M132 D133 L134 R135

F136 V137 G138 R139 S140 G141 R142 G143 K144 S145 F146 T147 L148 I149 I150 T151 V152 F153 T161 Y162 H163 I166 K167 I168 T169 V170 D171 G172 P173 R174 E175 P176 R177 R178 H179 ARG GLN LYS

- Molecule 2: RUNT-RELATED TRANSCRIPTION FACTOR 1

Chain F: 38% 51% 8%

G60 E61 L62 V63 R64 T65 D66 S67 P68 N69 F70 L71 C72 S73 V74 H78 W79 R80 C81 N82 R83 T84 L85 P86 I87 A88 F89 V90 V91 V92 T104 V105 M109 D110 E111 M112 E116 L117 R118 M119 A120 T121 M124 K125 M126 R130 F131 M132 D133 L134 R135 F136 V137 G138

R139 S140 G141 F146 T147 L148 I149 T150 V151 V152 M165 P166 Q167 H168 V169 V170 D171 G172 P173 R174 E175 P176 R177 R178 H179 ARG GLN LYS

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

Chain G: 8% 88%

G1 A2 A3 G4 A5 T6 T7 T8 C9 C10 A11 A12 A13 C14 T15 T16 C16 T17 G18 T19 G20 G21 T22 T23 G24 C25 G26

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

Chain I: 12% 88%

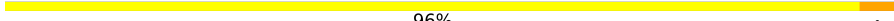
G1 A2 A3 G4 A5 T6 T7 T8 C9 C10 A11 A12 A13 C14 T15 T16 C16 T17 G18 T19 G20 G21 T22 T23 G24 C25 G26

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

Chain H: 8% 92%

G1 C2 G3 C4 A5 A6 C7 C8 A9 C10 A11 A12 A13 G14 T15 G18 G19 A20 A21 A22 T23 C24 T25 T26

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

Chain J:  96%

C1	C2	G3	C4	A5	A6	C7	C8	A9	C10	A11	G12	A13	G14	T15	T16	T17	G18	G19	A20	A21	A22	T23	C24	T25	T26
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	102.17Å 109.27Å 127.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.00 30.03 – 3.00	Depositor EDS
% Data completeness (in resolution range)	89.6 (30.00-3.00) 89.7 (30.03-3.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.58 (at 3.00Å)	Xtrriage
Refinement program	CNS 0.9	Depositor
R, $R_{free}$	0.244 , 0.313 0.258 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	60.0	Xtrriage
Anisotropy	0.378	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.15 , 33.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.23$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6270	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.92% 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	0.37	0/563	0.59	0/746
1	B	0.36	0/571	0.61	0/757
1	D	0.35	0/583	0.54	0/774
1	E	0.38	0/579	0.61	0/769
2	C	0.39	0/954	0.71	0/1297
2	F	0.42	0/954	0.73	0/1297
3	G	0.50	0/596	0.77	0/919
3	I	0.48	0/596	0.79	0/919
4	H	0.48	0/592	0.77	0/911
4	J	0.51	0/592	0.82	0/911
All	All	0.43	0/6580	0.71	0/9300

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
3	G	0	1
4	J	0	3
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
3	G	18	DG	Sidechain
4	J	1	DC	Sidechain
4	J	2	DC	Sidechain
4	J	6	DA	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	560	0	583	49	0
1	B	568	0	596	62	0
1	D	579	0	607	34	0
1	E	575	0	603	60	0
2	C	934	0	940	61	0
2	F	934	0	940	72	1
3	G	532	0	296	34	0
3	I	532	0	296	48	0
4	H	528	0	294	58	0
4	J	528	0	294	47	0
All	All	6270	0	5449	471	1

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 40.

The worst 5 of 471 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:G:9:DC:H2''	3:G:10:DC:H5''	1.25	1.14
1:E:295:ARG:HH11	1:E:295:ARG:HB2	1.07	1.12
4:H:20:DA:H2''	4:H:21:DA:H5''	1.32	1.10
4:J:20:DA:H2''	4:J:21:DA:H5''	1.33	1.06
3:I:24:DG:H2''	3:I:25:DC:H5'	1.34	1.05

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:111:GLU:CB	2:F:111:GLU:CB[2_575]	2.13	0.07

## 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	64/87 (74%)	50 (78%)	12 (19%)	2 (3%)	4	23
1	B	65/87 (75%)	57 (88%)	7 (11%)	1 (2%)	10	42
1	D	66/87 (76%)	55 (83%)	10 (15%)	1 (2%)	10	42
1	E	66/87 (76%)	56 (85%)	8 (12%)	2 (3%)	4	24
2	C	118/123 (96%)	107 (91%)	9 (8%)	2 (2%)	9	39
2	F	118/123 (96%)	105 (89%)	10 (8%)	3 (2%)	5	28
All	All	497/594 (84%)	430 (86%)	56 (11%)	11 (2%)	6	31

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	272	ASP
2	F	82	ASN
2	C	61	GLU
2	F	120	ALA
2	C	111	GLU

### 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	62/81 (76%)	57 (92%)	5 (8%)	11	40
1	B	63/81 (78%)	55 (87%)	8 (13%)	4	19
1	D	65/81 (80%)	53 (82%)	12 (18%)	1	8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	64/81 (79%)	57 (89%)	7 (11%)	6	25
2	C	102/105 (97%)	95 (93%)	7 (7%)	15	48
2	F	102/105 (97%)	92 (90%)	10 (10%)	8	30
All	All	458/534 (86%)	409 (89%)	49 (11%)	6	26

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	324	LEU
1	E	307	THR
1	D	329	ASN
1	E	274	TYR
1	E	319	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	319	GLN
2	F	69	ASN
2	F	155	ASN
1	E	310	ASN
1	B	333	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 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	66/87 (75%)	-0.55	3 (4%) 33 12	27, 81, 163, 183	0
1	B	67/87 (77%)	-0.49	2 (2%) 50 22	29, 78, 147, 166	0
1	D	68/87 (78%)	-0.63	0 100 100	35, 87, 161, 193	0
1	E	68/87 (78%)	-0.81	0 100 100	27, 73, 121, 154	0
2	C	120/123 (97%)	-0.68	1 (0%) 86 65	35, 73, 122, 170	0
2	F	120/123 (97%)	-0.70	1 (0%) 86 65	28, 65, 116, 173	0
3	G	26/26 (100%)	-1.06	0 100 100	33, 55, 77, 89	0
3	I	26/26 (100%)	-1.10	0 100 100	32, 53, 83, 102	0
4	H	26/26 (100%)	-1.10	0 100 100	35, 57, 69, 77	0
4	J	26/26 (100%)	-1.10	0 100 100	38, 53, 78, 95	0
All	All	613/698 (87%)	-0.73	7 (1%) 80 56	27, 69, 132, 193	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	270	HIS	5.5
2	F	111	GLU	4.9
1	A	269	LYS	3.7
1	A	268	ASP	2.8
1	B	321	SER	2.4

### 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.