



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 17, 2024 – 02:21 AM EDT

PDB ID : 5HOI  
Title : Crystal structure of the carboxy-terminal domain of yeast Ctf4 bound to Tof2.  
Authors : Simon, A.C.; Pellegrini, L.  
Deposited on : 2016-01-19  
Resolution : 3.30 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 : 2.37.1  
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) : 2.37.1

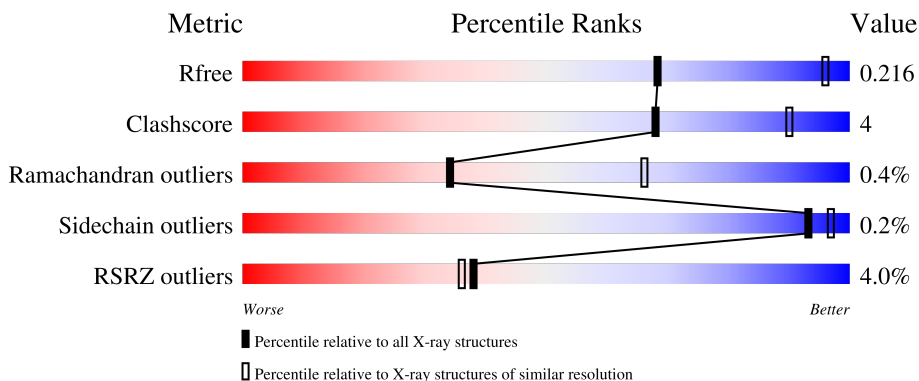
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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	478	
1	B	478	
1	C	478	
2	D	21	
2	E	21	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	21	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a red segment (38%), a green segment (33%), a yellow segment (5%), and a grey segment (62%). The percentages are labeled above and below the bar.</p>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 9650 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 DNA polymerase alpha-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	425	Total 3436	C 2205	N 572	O 644	S 15	0	2	0
1	B	432	Total 3481	C 2233	N 578	O 654	S 16	0	1	0
1	C	296	Total 2405	C 1562	N 392	O 440	S 11	0	1	0

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	450	MET	-	initiating methionine	UNP Q01454
A	451	GLY	-	expression tag	UNP Q01454
A	452	SER	-	expression tag	UNP Q01454
A	453	SER	-	expression tag	UNP Q01454
A	454	HIS	-	expression tag	UNP Q01454
A	455	HIS	-	expression tag	UNP Q01454
A	456	HIS	-	expression tag	UNP Q01454
A	457	HIS	-	expression tag	UNP Q01454
A	458	HIS	-	expression tag	UNP Q01454
A	459	HIS	-	expression tag	UNP Q01454
A	460	SER	-	expression tag	UNP Q01454
A	461	GLN	-	expression tag	UNP Q01454
A	462	ASP	-	expression tag	UNP Q01454
A	463	PRO	-	expression tag	UNP Q01454
A	464	GLU	-	expression tag	UNP Q01454
A	465	ASN	-	expression tag	UNP Q01454
A	466	LEU	-	expression tag	UNP Q01454
A	467	TYR	-	expression tag	UNP Q01454
A	468	PHE	-	expression tag	UNP Q01454
A	469	GLN	-	expression tag	UNP Q01454
A	470	GLY	-	expression tag	UNP Q01454
A	471	THR	-	expression tag	UNP Q01454
B	450	MET	-	initiating methionine	UNP Q01454

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	451	GLY	-	expression tag	UNP Q01454
B	452	SER	-	expression tag	UNP Q01454
B	453	SER	-	expression tag	UNP Q01454
B	454	HIS	-	expression tag	UNP Q01454
B	455	HIS	-	expression tag	UNP Q01454
B	456	HIS	-	expression tag	UNP Q01454
B	457	HIS	-	expression tag	UNP Q01454
B	458	HIS	-	expression tag	UNP Q01454
B	459	HIS	-	expression tag	UNP Q01454
B	460	SER	-	expression tag	UNP Q01454
B	461	GLN	-	expression tag	UNP Q01454
B	462	ASP	-	expression tag	UNP Q01454
B	463	PRO	-	expression tag	UNP Q01454
B	464	GLU	-	expression tag	UNP Q01454
B	465	ASN	-	expression tag	UNP Q01454
B	466	LEU	-	expression tag	UNP Q01454
B	467	TYR	-	expression tag	UNP Q01454
B	468	PHE	-	expression tag	UNP Q01454
B	469	GLN	-	expression tag	UNP Q01454
B	470	GLY	-	expression tag	UNP Q01454
B	471	THR	-	expression tag	UNP Q01454
C	450	MET	-	initiating methionine	UNP Q01454
C	451	GLY	-	expression tag	UNP Q01454
C	452	SER	-	expression tag	UNP Q01454
C	453	SER	-	expression tag	UNP Q01454
C	454	HIS	-	expression tag	UNP Q01454
C	455	HIS	-	expression tag	UNP Q01454
C	456	HIS	-	expression tag	UNP Q01454
C	457	HIS	-	expression tag	UNP Q01454
C	458	HIS	-	expression tag	UNP Q01454
C	459	HIS	-	expression tag	UNP Q01454
C	460	SER	-	expression tag	UNP Q01454
C	461	GLN	-	expression tag	UNP Q01454
C	462	ASP	-	expression tag	UNP Q01454
C	463	PRO	-	expression tag	UNP Q01454
C	464	GLU	-	expression tag	UNP Q01454
C	465	ASN	-	expression tag	UNP Q01454
C	466	LEU	-	expression tag	UNP Q01454
C	467	TYR	-	expression tag	UNP Q01454
C	468	PHE	-	expression tag	UNP Q01454
C	469	GLN	-	expression tag	UNP Q01454
C	470	GLY	-	expression tag	UNP Q01454

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	471	THR	-	expression tag	UNP Q01454

- Molecule 2 is a protein called Topoisomerase 1-associated factor 2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	0	0	0
			104	65	22	17			
2	E	12	Total	C	N	O	0	0	0
			104	65	22	17			
2	F	8	Total	C	N	O	0	0	0
			69	44	13	12			

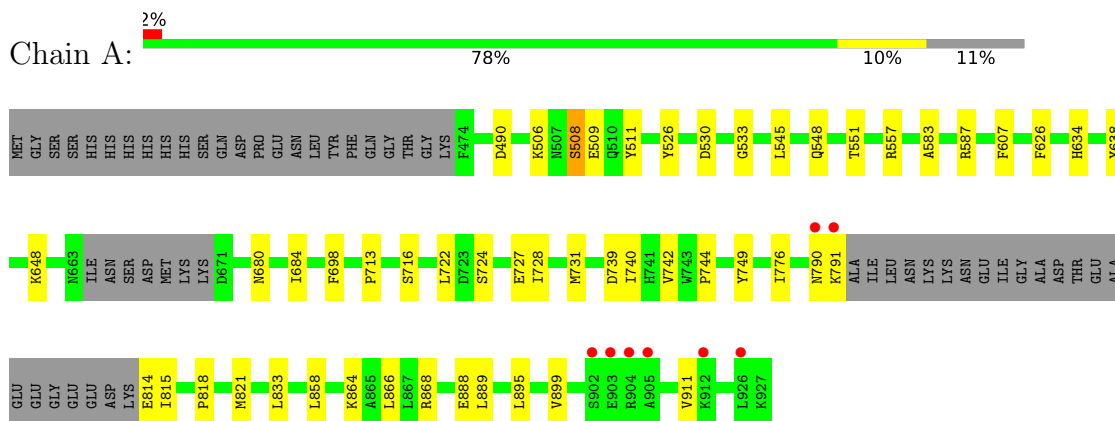
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	19	Total	O	0	0
			19	19		
3	B	22	Total	O	0	0
			22	22		
3	C	10	Total	O	0	0
			10	10		

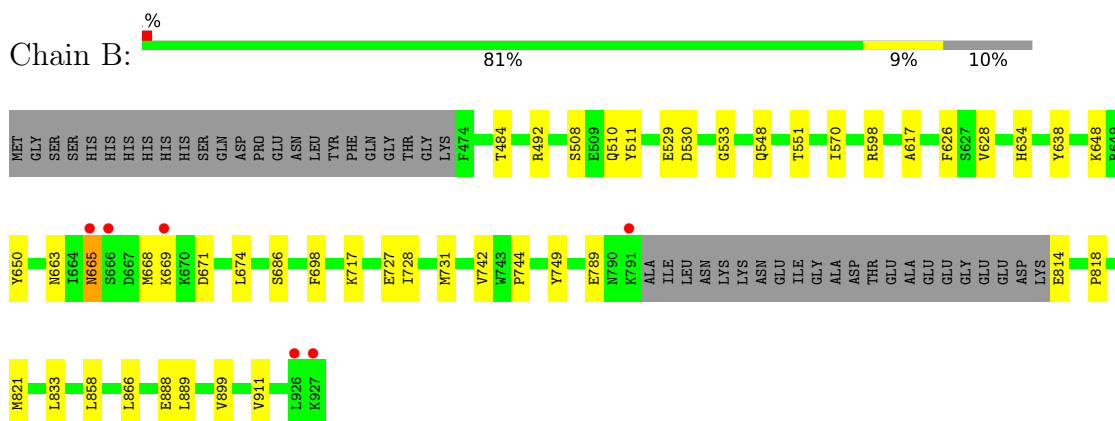
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

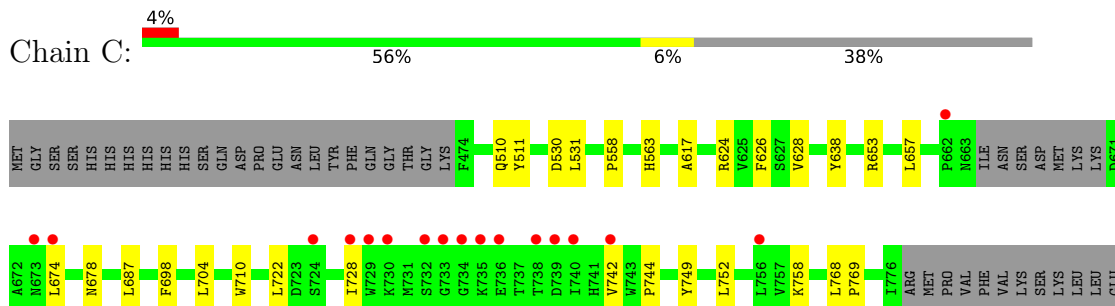
- Molecule 1: DNA polymerase alpha-binding protein



- Molecule 1: DNA polymerase alpha-binding protein



- Molecule 1: DNA polymerase alpha-binding protein







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.58Å 99.55Å 218.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.99 – 3.30 48.99 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.3 (48.99-3.30) 99.2 (48.99-3.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.84 (at 3.33Å)	Xtrriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, $R_{free}$	0.179 , 0.224 0.174 , 0.216	Depositor DCC
$R_{free}$ test set	1471 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	82.1	Xtrriage
Anisotropy	0.319	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 57.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9650	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	79.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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	0.24	0/3521	0.41	0/4767
1	B	0.24	0/3567	0.40	0/4828
1	C	0.25	0/2481	0.43	0/3370
2	D	0.21	0/103	0.36	0/135
2	E	0.21	0/103	0.37	0/135
2	F	0.20	0/68	0.33	0/89
All	All	0.24	0/9843	0.41	0/13324

There are no bond length outliers.

There are no bond angle outliers.

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	3436	0	3382	32	0
1	B	3481	0	3431	28	0
1	C	2405	0	2326	16	0
2	D	104	0	120	2	0
2	E	104	0	120	4	0
2	F	69	0	79	1	0
3	A	19	0	0	0	0
3	B	22	0	0	0	0
3	C	10	0	0	0	0
All	All	9650	0	9458	73	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 4.

The worst 5 of 73 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:B:899:VAL:HG13	1:B:911:VAL:HG13	1.75	0.68
1:C:511:TYR:HB2	1:C:530:ASP:HB3	1.81	0.63
1:A:533:GLY:O	1:A:548:GLN:NE2	2.35	0.60
1:C:704:LEU:HD22	1:C:752:LEU:HD13	1.83	0.59
1:C:728:ILE:HD11	1:C:742:VAL:HG23	1.86	0.57

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	421/478 (88%)	404 (96%)	15 (4%)	2 (0%)	29	61
1	B	429/478 (90%)	412 (96%)	15 (4%)	2 (0%)	29	61
1	C	293/478 (61%)	274 (94%)	18 (6%)	1 (0%)	41	71
2	D	10/21 (48%)	9 (90%)	1 (10%)	0	100	100
2	E	10/21 (48%)	10 (100%)	0	0	100	100
2	F	6/21 (29%)	6 (100%)	0	0	100	100
All	All	1169/1497 (78%)	1115 (95%)	49 (4%)	5 (0%)	34	66

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	508	SER
1	A	749	TYR
1	B	749	TYR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	749	TYR
1	B	508	SER

### 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	379/422 (90%)	378 (100%)	1 (0%)	92	96
1	B	385/422 (91%)	384 (100%)	1 (0%)	92	96
1	C	267/422 (63%)	267 (100%)	0	100	100
2	D	12/20 (60%)	12 (100%)	0	100	100
2	E	12/20 (60%)	12 (100%)	0	100	100
2	F	8/20 (40%)	8 (100%)	0	100	100
All	All	1063/1326 (80%)	1061 (100%)	2 (0%)	93	97

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	739	ASP
1	B	665	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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	425/478 (88%)	0.10	8 (1%) 66 65	45, 68, 113, 139	0
1	B	432/478 (90%)	0.03	6 (1%) 75 75	46, 72, 117, 158	0
1	C	296/478 (61%)	0.19	17 (5%) 23 23	50, 78, 130, 171	0
2	D	12/21 (57%)	1.83	5 (41%) 0 0	115, 124, 146, 154	0
2	E	12/21 (57%)	0.90	3 (25%) 0 0	111, 121, 144, 153	0
2	F	8/21 (38%)	3.55	8 (100%) 0 0	143, 152, 160, 168	0
All	All	1185/1497 (79%)	0.15	47 (3%) 38 36	45, 73, 128, 171	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	504	ILE	5.2
2	F	509	ARG	4.0
2	D	512	ASN	4.0
1	C	738	THR	3.9
1	C	673	ASN	3.7

### 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.