wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2016 – 03:30 PM EDT

PDB ID	:	5HOG
Title	:	Crystal structure of the carboxy-terminal domain of yeast Ctf4 bound to Dna2.
Authors	:	Simon, A.C.; Pellegrini, L.
Deposited on	:	2016-01-19
Resolution	:	3.09 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

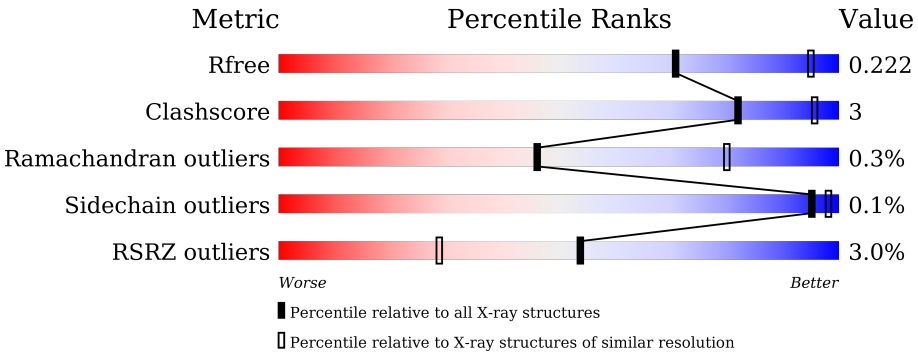
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 3.09 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	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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 <=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	79% 9% 11%
1	B	478	% 83% 8% 10%
1	C	478	4% 55% 6% 38%
2	D	17	35% 71% 6% 24%
2	E	17	41% 76% 6% 18%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9598 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
1	A	425	Total	C	N	O	S	0	1	0
			3425	2199	568	643	15			
1	B	432	Total	C	N	O	S	0	1	0
			3481	2233	578	654	16			
1	C	296	Total	C	N	O	S	0	1	0
			2405	1562	392	440	11			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	450	MET	HIS	conflict	UNP Q01454
A	451	GLY	ASN	conflict	UNP Q01454
A	452	SER	GLU	conflict	UNP Q01454
A	453	SER	HIS	conflict	UNP Q01454
A	454	HIS	SER	conflict	UNP Q01454
A	455	HIS	TYR	conflict	UNP Q01454
A	456	HIS	SER	conflict	UNP Q01454
A	457	HIS	ARG	conflict	UNP Q01454
A	458	HIS	VAL	conflict	UNP Q01454
A	460	SER	LYS	conflict	UNP Q01454
A	461	GLN	THR	conflict	UNP Q01454
A	462	ASP	HIS	conflict	UNP Q01454
A	463	PRO	SER	conflict	UNP Q01454
A	464	GLU	PHE	conflict	UNP Q01454
A	465	ASN	PRO	conflict	UNP Q01454
A	466	LEU	ILE	conflict	UNP Q01454
A	467	TYR	SER	conflict	UNP Q01454
A	468	PHE	LEU	conflict	UNP Q01454
A	469	GLN	ALA	conflict	UNP Q01454
A	470	GLY	ASN	conflict	UNP Q01454
B	450	MET	HIS	conflict	UNP Q01454
B	451	GLY	ASN	conflict	UNP Q01454
B	452	SER	GLU	conflict	UNP Q01454

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	453	SER	HIS	conflict	UNP Q01454
B	454	HIS	SER	conflict	UNP Q01454
B	455	HIS	TYR	conflict	UNP Q01454
B	456	HIS	SER	conflict	UNP Q01454
B	457	HIS	ARG	conflict	UNP Q01454
B	458	HIS	VAL	conflict	UNP Q01454
B	460	SER	LYS	conflict	UNP Q01454
B	461	GLN	THR	conflict	UNP Q01454
B	462	ASP	HIS	conflict	UNP Q01454
B	463	PRO	SER	conflict	UNP Q01454
B	464	GLU	PHE	conflict	UNP Q01454
B	465	ASN	PRO	conflict	UNP Q01454
B	466	LEU	ILE	conflict	UNP Q01454
B	467	TYR	SER	conflict	UNP Q01454
B	468	PHE	LEU	conflict	UNP Q01454
B	469	GLN	ALA	conflict	UNP Q01454
B	470	GLY	ASN	conflict	UNP Q01454
C	450	MET	HIS	conflict	UNP Q01454
C	451	GLY	ASN	conflict	UNP Q01454
C	452	SER	GLU	conflict	UNP Q01454
C	453	SER	HIS	conflict	UNP Q01454
C	454	HIS	SER	conflict	UNP Q01454
C	455	HIS	TYR	conflict	UNP Q01454
C	456	HIS	SER	conflict	UNP Q01454
C	457	HIS	ARG	conflict	UNP Q01454
C	458	HIS	VAL	conflict	UNP Q01454
C	460	SER	LYS	conflict	UNP Q01454
C	461	GLN	THR	conflict	UNP Q01454
C	462	ASP	HIS	conflict	UNP Q01454
C	463	PRO	SER	conflict	UNP Q01454
C	464	GLU	PHE	conflict	UNP Q01454
C	465	ASN	PRO	conflict	UNP Q01454
C	466	LEU	ILE	conflict	UNP Q01454
C	467	TYR	SER	conflict	UNP Q01454
C	468	PHE	LEU	conflict	UNP Q01454
C	469	GLN	ALA	conflict	UNP Q01454
C	470	GLY	ASN	conflict	UNP Q01454

- Molecule 2 is a protein called Dna2p.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	13	Total	C	N	O	0	0	0
			106	64	17	25			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	14	Total	C	N	O	0	0	0
			110	66	18	26			

- Molecule 3 is water.

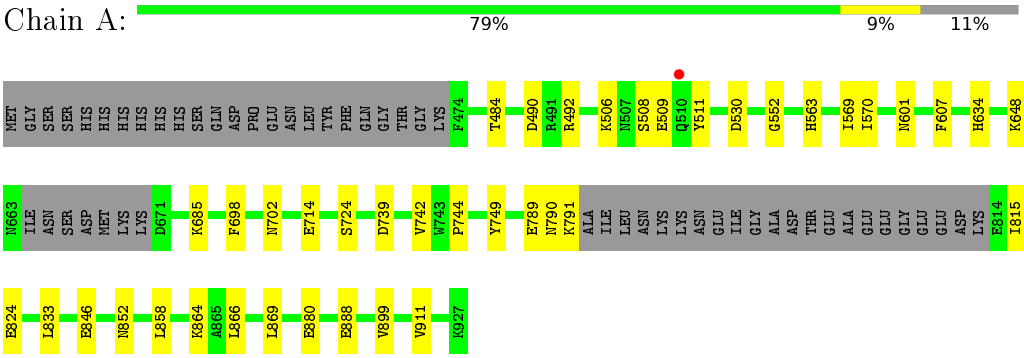
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	29	Total	O	0	0
			29	29		
3	B	29	Total	O	0	0
			29	29		
3	C	13	Total	O	0	0
			13	13		

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.

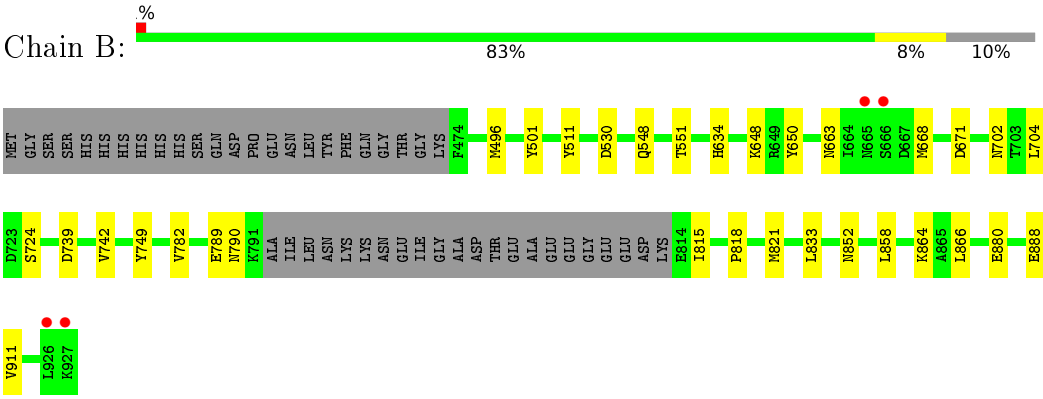
- Molecule 1: DNA polymerase alpha-binding protein

Chain A:



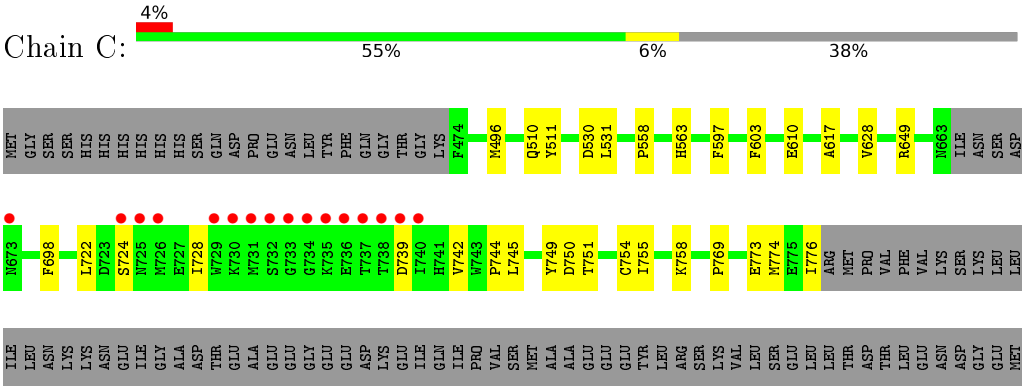
- Molecule 1: DNA polymerase alpha-binding protein

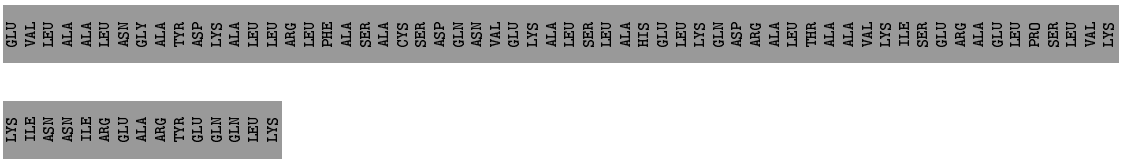
Chain B:



- Molecule 1: DNA polymerase alpha-binding protein

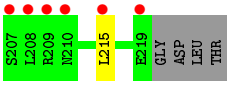
Chain C:





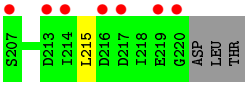
- Molecule 2: Dna2p

Chain D:



- Molecule 2: Dna2p

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	88.68Å 99.55Å 218.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.98 – 3.09 48.98 – 3.09	Depositor EDS
% Data completeness (in resolution range)	99.7 (48.98-3.09) 99.7 (48.98-3.09)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.14 (at 3.07Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.180 , 0.226 0.175 , 0.222	Depositor DCC
R_{free} test set	1817 reflections (5.04%)	DCC
Wilson B-factor (Å^2)	70.7	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9598	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.96% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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	0.25	0/3510	0.42	0/4752
1	B	0.25	0/3567	0.41	0/4828
1	C	0.27	0/2481	0.44	0/3370
2	D	0.20	0/105	0.35	0/141
2	E	0.20	0/109	0.33	0/146
All	All	0.25	0/9772	0.42	0/13237

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

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	3425	0	3369	27	0
1	B	3481	0	3431	22	0
1	C	2405	0	2326	17	0
2	D	106	0	100	1	0
2	E	110	0	103	1	0
3	A	29	0	0	0	0
3	B	29	0	0	0	0
3	C	13	0	0	0	0
All	All	9598	0	9329	60	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 3.

The worst 5 of 60 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:789:GLU:HG3	1:B:818:PRO:HG3	1.72	0.72
1:A:789:GLU:HG3	1:A:818:PRO:HG3	1.73	0.71
1:B:899:VAL:HG13	1:B:911:VAL:HG13	1.77	0.65
1:A:790:ASN:HD21	1:A:815:ILE:HG23	1.67	0.60
1:B:511:TYR:HB2	1:B:530:ASP:HB3	1.85	0.58

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	420/478 (88%)	401 (96%)	18 (4%)	1 (0%)	52 84
1	B	429/478 (90%)	413 (96%)	15 (4%)	1 (0%)	52 84
1	C	293/478 (61%)	275 (94%)	17 (6%)	1 (0%)	46 80
2	D	11/17 (65%)	11 (100%)	0	0	100 100
2	E	12/17 (71%)	11 (92%)	1 (8%)	0	100 100
All	All	1165/1468 (79%)	1111 (95%)	51 (4%)	3 (0%)	46 80

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	749	TYR
1	C	749	TYR
1	A	749	TYR

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	378/422 (90%)	378 (100%)	0	100 100
1	B	385/422 (91%)	384 (100%)	1 (0%)	94 97
1	C	267/422 (63%)	267 (100%)	0	100 100
2	D	13/16 (81%)	13 (100%)	0	100 100
2	E	13/16 (81%)	13 (100%)	0	100 100
All	All	1056/1298 (81%)	1055 (100%)	1 (0%)	95 98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	852	ASN

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	634	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

There are no ligands in this entry.

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

6.1 Protein, DNA and RNA chains

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^{th} 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	425/478 (88%)	-0.16	1 (0%) 95 91	38, 60, 103, 128	0
1	B	432/478 (90%)	-0.13	4 (0%) 85 72	40, 61, 108, 140	0
1	C	296/478 (61%)	-0.00	17 (5%) 27 11	46, 69, 126, 162	0
2	D	13/17 (76%)	2.22	6 (46%) 0 0	120, 130, 144, 145	0
2	E	14/17 (82%)	2.06	7 (50%) 0 0	121, 135, 145, 148	0
All	All	1180/1468 (80%)	-0.05	35 (2%) 54 29	38, 63, 120, 162	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	207	SER	4.8
1	C	733	GLY	4.4
2	D	208	LEU	4.1
2	E	207	SER	3.8
1	B	666	SER	3.6

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.