



wwPDB X-ray Structure Validation Summary Report

Jun 24, 2024 – 04:46 PM EDT

PDB ID : 6YPC
Title : Crystal structure of the kinetochore subunits H/I/K/T/W penta-complex from *S. cerevisiae* at 2.9 angstroms
Authors : Bellini, D.; Zhang, Z.; Barford, D.
Deposited on : 2020-04-15
Resolution : 2.90 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

<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) : 1.20.1
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

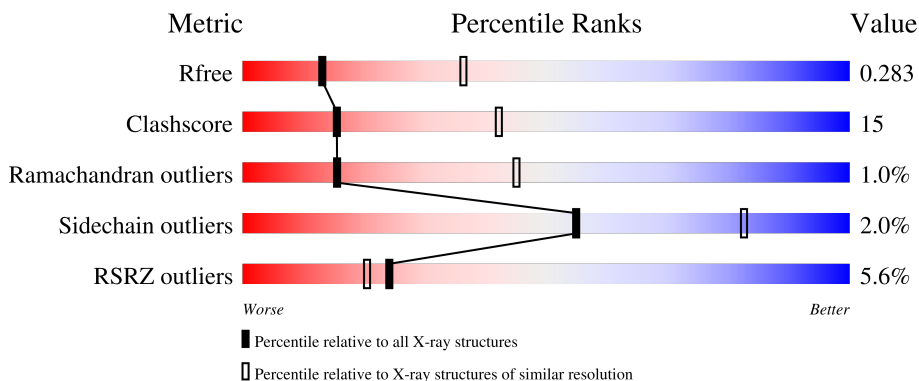
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 ≥ 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	K	110	 5% 71% 18% 10%
2	H	46	 7% 63% 13% 24%
3	T	367	 3% 14% 11% 75%
4	W	89	 6% 37% 40% 22%
5	I	251	 2% 65% 30%

2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 4341 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 Inner kinetochore subunit MCM22.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	K	99	809	515	138	154	2	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	130	MET	-	initiating methionine	UNP P47167

- Molecule 2 is a protein called Inner kinetochore subunit MCM16.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	35	276	173	44	58	1	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	136	MET	-	initiating methionine	UNP Q12262

- Molecule 3 is a protein called Inner kinetochore subunit CNN1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	T	92	761	488	125	144	4	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
T	362	GLU	-	expression tag	UNP P43618
T	363	ASN	-	expression tag	UNP P43618
T	364	LEU	-	expression tag	UNP P43618

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
T	365	TYR	-	expression tag	UNP P43618
T	366	PHE	-	expression tag	UNP P43618
T	367	GLN	-	expression tag	UNP P43618

- Molecule 4 is a protein called Inner kinetochore subunit WIP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	W	69	550	348	96	104	2	0	0	0

- Molecule 5 is a protein called Inner kinetochore subunit CTF3.

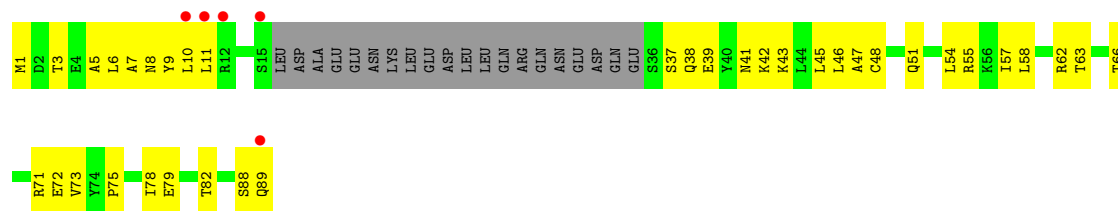
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	I	240	1938	1260	332	339	7	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

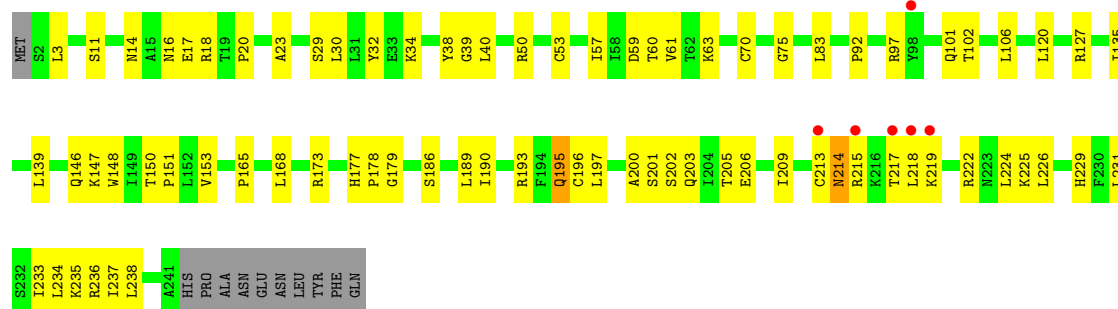
Chain	Residue	Modelled	Actual	Comment	Reference
I	246	GLU	-	expression tag	UNP Q12748
I	247	ASN	-	expression tag	UNP Q12748
I	248	LEU	-	expression tag	UNP Q12748
I	249	TYR	-	expression tag	UNP Q12748
I	250	PHE	-	expression tag	UNP Q12748
I	251	GLN	-	expression tag	UNP Q12748

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	K	2	Total	O	0	0
			2	2		
6	T	2	Total	O	0	0
			2	2		
6	I	3	Total	O	0	0
			3	3		



● Molecule 5: Inner kinetochore subunit CTF3



4 Data and refinement statistics i

Property	Value	Source
Space group	I 41 2 2	Depositor
Cell constants a, b, c, α , β , γ	132.58Å 132.58Å 241.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.85 – 2.90 68.85 – 2.90	Depositor EDS
% Data completeness (in resolution range)	80.9 (68.85-2.90) 80.9 (68.85-2.90)	Depositor EDS
R_{merge}	0.37	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.88 (at 2.91Å)	Xtrriage
Refinement program	REFMAC 1.17.1_3660, PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.228 , 0.283 0.239 , 0.283	Depositor DCC
R_{free} test set	1027 reflections (5.22%)	wwPDB-VP
Wilson B-factor (Å ²)	78.5	Xtrriage
Anisotropy	0.049	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 41.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	4341	wwPDB-VP
Average B, all atoms (Å ²)	72.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.60% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	K	0.50	0/820	0.69	0/1108
2	H	0.54	0/278	0.61	0/375
3	T	0.45	0/772	0.63	0/1040
4	W	0.45	0/556	0.64	0/748
5	I	0.51	0/1985	0.71	0/2704
All	All	0.49	0/4411	0.68	0/5975

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	W	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	W	71	ARG	Peptide

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	K	809	0	828	15	0
2	H	276	0	274	6	0
3	T	761	0	781	39	0
4	W	550	0	559	33	0
5	I	1938	0	2019	63	1
6	I	3	0	0	0	0
6	K	2	0	0	0	0
6	T	2	0	0	0	0
All	All	4341	0	4461	130	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 15.

The worst 5 of 130 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:T:312:LEU:HA	3:T:315:LYS:HB2	1.52	0.90
5:I:225:LYS:C	5:I:226:LEU:HD12	1.94	0.87
5:I:150:THR:HG23	5:I:151:PRO:HD3	1.58	0.84
5:I:165:PRO:HA	5:I:168:LEU:HD12	1.60	0.83
4:W:3:THR:H	5:I:179:GLY:HA2	1.44	0.83

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)
5:I:14:ASN:OD1	5:I:50:ARG:NH2[10_555]	2.07	0.13

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	K	95/110 (86%)	86 (90%)	8 (8%)	1 (1%)	14	42
2	H	33/46 (72%)	30 (91%)	3 (9%)	0	100	100
3	T	90/367 (24%)	82 (91%)	8 (9%)	0	100	100
4	W	65/89 (73%)	60 (92%)	5 (8%)	0	100	100
5	I	238/251 (95%)	219 (92%)	15 (6%)	4 (2%)	9	31
All	All	521/863 (60%)	477 (92%)	39 (8%)	5 (1%)	15	45

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	K	169	GLU
5	I	201	SER
5	I	214	ASN
5	I	196	CYS
5	I	200	ALA

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	K	93/103 (90%)	89 (96%)	4 (4%)	29	62
2	H	31/42 (74%)	31 (100%)	0	100	100
3	T	89/345 (26%)	87 (98%)	2 (2%)	52	81
4	W	57/76 (75%)	56 (98%)	1 (2%)	59	85
5	I	223/233 (96%)	220 (99%)	3 (1%)	69	90
All	All	493/799 (62%)	483 (98%)	10 (2%)	55	82

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

Mol	Chain	Res	Type
5	I	11	SER
5	I	97	ARG
5	I	195	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	K	199	SER
3	T	283	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	K	195	ASN
3	T	335	ASN
4	W	13	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, 95th 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	K	99/110 (90%)	0.33	6 (6%) 21 17	41, 62, 118, 131	0
2	H	35/46 (76%)	0.64	3 (8%) 10 8	38, 59, 108, 110	0
3	T	92/367 (25%)	0.52	10 (10%) 5 4	49, 83, 130, 137	0
4	W	69/89 (77%)	0.36	5 (7%) 15 11	68, 79, 111, 121	0
5	I	240/251 (95%)	0.08	6 (2%) 57 55	40, 63, 94, 120	0
All	All	535/863 (61%)	0.28	30 (5%) 24 20	38, 69, 115, 137	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	166	ASP	10.7
1	K	167	ASN	8.6
1	K	168	ASP	6.7
3	T	269	SER	6.6
2	H	181	ILE	6.2

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.