W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

Full wwPDB X-ray Structure Validation Report ⓘ

May 31, 2022 – 06:18 pm BST

PDB ID	:	7PB4
Title	:	Cenp-HIK 3-protein complex
Authors	:	Bellini, D.; Yatskevich, S.; Muir, W.K.; Barford, D.
Deposited on	:	2021-07-30
Resolution	:	2.49 Å(reported)

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	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 2.49 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	Percentile Ranks	Value
Rfree		0.288
Clashscore		9
Ramachandran outliers		0.9%
Sidechain outliers		1.9%

Worse   *Better*

▊ Percentile relative to all X-ray structures

❏ Percentile relative to X-ray structures of similar resolution
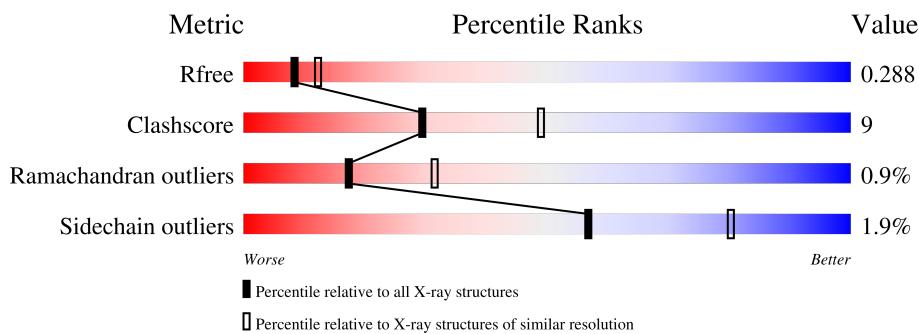
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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 <=5%

Mol	Chain	Length	Quality of chain
1	H	49	78% 18% •
2	I	227	78% 15% •• 5%
3	K	105	78% 9% 13%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 2864 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 Centromere protein H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	47	Total	C	N	O	S	0	0	0
			380	244	63	71	2			

- Molecule 2 is a protein called Centromere protein I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	215	Total	C	N	O	S	0	0	0
			1717	1121	288	297	11			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	63	MET	-	initiating methionine	UNP Q92674
I	282	SER	-	expression tag	UNP Q92674
I	283	PRO	-	expression tag	UNP Q92674
I	284	LEU	-	expression tag	UNP Q92674
I	285	GLU	-	expression tag	UNP Q92674
I	286	VAL	-	expression tag	UNP Q92674
I	287	GLN	-	expression tag	UNP Q92674
I	288	PHE	-	expression tag	UNP Q92674
I	289	GLY	-	expression tag	UNP Q92674

- Molecule 3 is a protein called Centromere protein K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	K	91	Total	C	N	O	S	0	0	0
			767	502	127	136	2			

3 Residue-property plots ⓘ

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. 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. 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: Centromere protein H

Chain H:
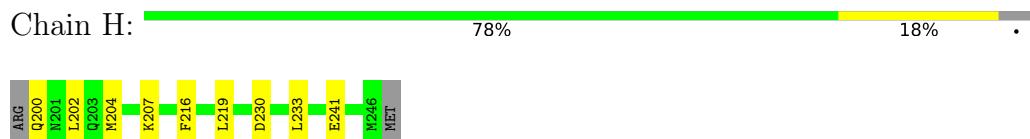
• Molecule 2: Centromere protein I

Chain I:
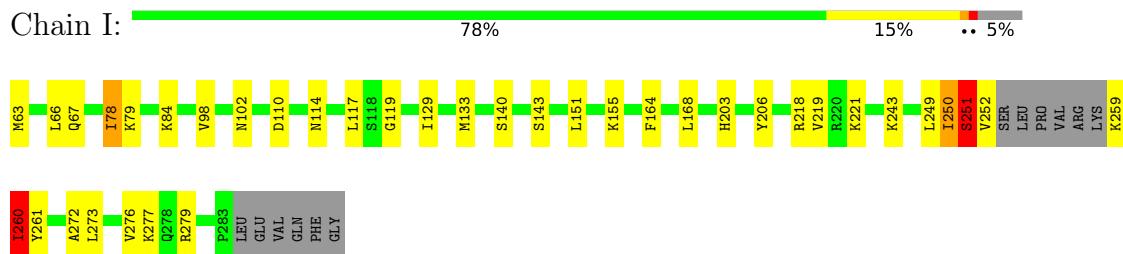
• Molecule 3: Centromere protein K

Chain K:
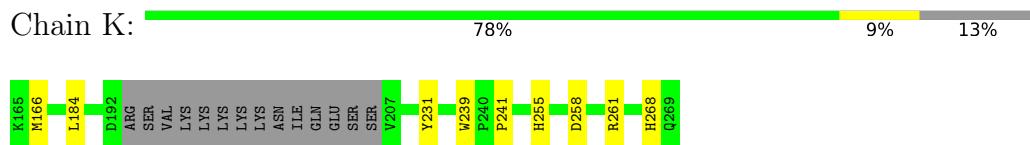
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	44.62Å 56.12Å 176.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.37 – 2.49 47.37 – 2.49	Depositor EDS
% Data completeness (in resolution range)	72.4 (47.37-2.49) 72.4 (47.37-2.49)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.14 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.222 , 0.283 0.230 , 0.288	Depositor DCC
R_{free} test set	595 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	43.5	Xtriage
Anisotropy	0.116	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	2864	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.48% 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	H	0.23	0/384	0.35	0/519
2	I	0.33	0/1752	0.42	0/2361
3	K	0.30	0/788	0.39	0/1068
All	All	0.31	0/2924	0.40	0/3948

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	H	380	0	396	6	0
2	I	1717	0	1786	47	0
3	K	767	0	770	6	0
All	All	2864	0	2952	55	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 9.

All (55) 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)
2:I:219:VAL:CG1	2:I:250:ILE:HG21	1.57	1.33

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:I:219:VAL:HG11	2:I:250:ILE:CG2	1.64	1.27
2:I:243:LYS:CG	2:I:251:SER:O	1.82	1.25
2:I:243:LYS:HG3	2:I:251:SER:O	1.09	1.23
2:I:259:LYS:N	2:I:260:ILE:HA	1.49	1.15
2:I:249:LEU:O	2:I:250:ILE:HG22	1.50	1.12
2:I:219:VAL:HG11	2:I:250:ILE:HG21	1.13	1.07
2:I:250:ILE:HD11	2:I:252:VAL:O	1.64	0.97
2:I:259:LYS:N	2:I:260:ILE:CA	2.29	0.96
2:I:206:TYR:CE2	2:I:259:LYS:N	2.37	0.93
2:I:243:LYS:HB2	2:I:252:VAL:C	1.97	0.83
2:I:249:LEU:O	2:I:250:ILE:CG2	2.26	0.83
2:I:259:LYS:HD2	2:I:260:ILE:HD12	1.67	0.76
1:H:219:LEU:HD13	3:K:184:LEU:HD21	1.68	0.74
2:I:249:LEU:C	2:I:250:ILE:CG2	2.55	0.74
2:I:219:VAL:HG13	2:I:250:ILE:HG21	1.66	0.72
2:I:250:ILE:CD1	2:I:252:VAL:O	2.38	0.71
2:I:219:VAL:CG1	2:I:250:ILE:CG2	2.39	0.69
2:I:63:MET:HE1	2:I:67:GLN:HB3	1.78	0.65
2:I:259:LYS:HD2	2:I:260:ILE:CD1	2.26	0.65
2:I:102:ASN:HA	2:I:273:LEU:HD22	1.81	0.62
1:H:241:GLU:OE1	2:I:218:ARG:NH2	2.33	0.62
2:I:79:LYS:HB2	2:I:84:LYS:HE3	1.83	0.60
2:I:63:MET:CE	2:I:67:GLN:HB3	2.32	0.59
2:I:243:LYS:HG2	2:I:251:SER:O	1.97	0.59
2:I:259:LYS:N	2:I:260:ILE:HG13	2.20	0.56
1:H:207:LYS:HD2	3:K:268:HIS:HB2	1.89	0.54
2:I:249:LEU:C	2:I:250:ILE:HG23	2.28	0.54
2:I:219:VAL:HG11	2:I:250:ILE:HG23	1.77	0.53
2:I:259:LYS:HA	2:I:260:ILE:HD12	1.91	0.51
2:I:110:ASP:O	2:I:114:ASN:ND2	2.37	0.51
2:I:78:ILE:HG13	2:I:84:LYS:HD3	1.93	0.49
2:I:140:SER:O	2:I:143:SER:OG	2.26	0.48
1:H:230:ASP:HB3	1:H:233:LEU:HB2	1.94	0.48
2:I:129:ILE:O	2:I:133:MET:HG2	2.13	0.48
2:I:119:GLY:HA3	2:I:155:LYS:NZ	2.28	0.47
2:I:276:VAL:HG12	2:I:279:ARG:NH2	2.30	0.47
2:I:206:TYR:HE2	2:I:259:LYS:N	2.07	0.46
2:I:243:LYS:CB	2:I:251:SER:O	2.59	0.46
2:I:260:ILE:HG22	2:I:261:TYR:N	2.33	0.44
2:I:272:ALA:O	2:I:276:VAL:HG13	2.17	0.44
3:K:239:TRP:CD1	3:K:241:PRO:HD2	2.53	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:I:260:ILE:CG2	2:I:261:TYR:N	2.81	0.43
2:I:151:LEU:HD11	2:I:164:PHE:CD2	2.53	0.43
2:I:66:LEU:HD22	2:I:98:VAL:HG11	2.00	0.43
2:I:117:LEU:HD11	2:I:164:PHE:CD1	2.54	0.43
2:I:251:SER:OG	2:I:252:VAL:N	2.51	0.43
2:I:273:LEU:HB3	2:I:277:LYS:NZ	2.34	0.42
3:K:231:TYR:HB3	3:K:261:ARG:HB3	2.02	0.42
2:I:221:LYS:HD3	2:I:221:LYS:HA	1.87	0.41
1:H:202:LEU:HD22	3:K:166:MET:HG2	2.02	0.41
2:I:259:LYS:HD2	2:I:259:LYS:HA	1.84	0.41
2:I:273:LEU:O	2:I:276:VAL:HG22	2.20	0.41
1:H:200:GLN:O	1:H:204:MET:HG2	2.20	0.40
3:K:255:HIS:HB3	3:K:258:ASP:O	2.21	0.40

There are no symmetry-related clashes.

5.3　Torsion angles　ⓘ

5.3.1　Protein backbone　ⓘ

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	H	45/49 (92%)	43 (96%)	2 (4%)	0	100	100
2	I	211/227 (93%)	200 (95%)	8 (4%)	3 (1%)	11	20
3	K	87/105 (83%)	87 (100%)	0	0	100	100
All	All	343/381 (90%)	330 (96%)	10 (3%)	3 (1%)	17	31

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	I	251	SER
2	I	260	ILE
2	I	78	ILE

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	H	44/46 (96%)	43 (98%)	1 (2%)	50	76
2	I	187/199 (94%)	182 (97%)	5 (3%)	44	71
3	K	87/101 (86%)	87 (100%)	0	100	100
All	All	318/346 (92%)	312 (98%)	6 (2%)	57	80

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	216	PHE
2	I	168	LEU
2	I	203	HIS
2	I	250	ILE
2	I	251	SER
2	I	260	ILE

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

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 monosaccharides 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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.