



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 7, 2017 – 10:52 PM EDT

PDB ID : 4F9A  
Title : Human CDC7 kinase in complex with DBF4 and nucleotide  
Authors : Hughes, S.; Cherepanov, P.  
Deposited on : unknown  
Resolution : 2.17 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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-20029824

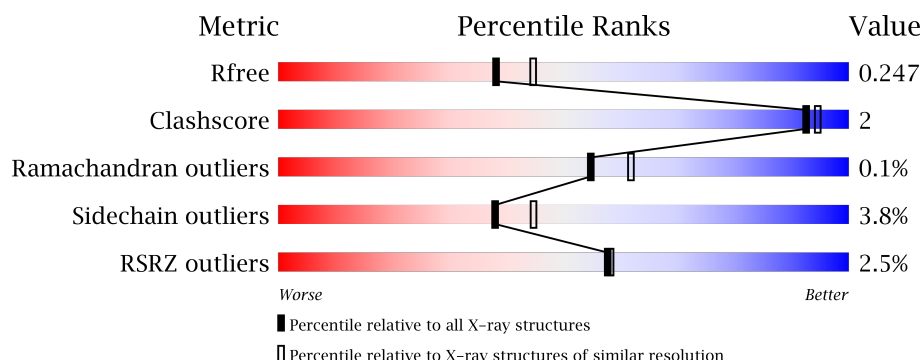
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.17 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))
$R_{free}$	100719	5526 (2.20-2.16)
Clashscore	112137	6386 (2.20-2.16)
Ramachandran outliers	110173	6282 (2.20-2.16)
Sidechain outliers	110143	6282 (2.20-2.16)
RSRZ outliers	101464	5562 (2.20-2.16)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	361	<div> <div>0.1%</div> <div> <div></div> <div>82%</div> <div>5%</div> <div>13%</div> </div> </div>
1	C	361	<div> <div>2%</div> <div> <div></div> <div>80%</div> <div>8%</div> <div>12%</div> </div> </div>
2	B	144	<div> <div>0.1%</div> <div> <div></div> <div>56%</div> <div>5%</div> <div>38%</div> </div> </div>
2	D	144	<div> <div>5%</div> <div> <div></div> <div>56%</div> <div>7%</div> <div>38%</div> </div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6643 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 Cell division cycle 7-related protein kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	315	Total	C	N	O	S	0	0	0
			2502	1605	429	457	11			
1	C	317	Total	C	N	O	S	0	0	0
			2513	1613	429	460	11			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	36	MET	-	EXPRESSION TAG	UNP O00311
C	36	MET	-	EXPRESSION TAG	UNP O00311

- Molecule 2 is a protein called Protein DBF4 homolog A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	90	Total	C	N	O	S	0	0	0
			752	487	118	141	6			
2	D	90	Total	C	N	O	S	0	0	0
			752	487	118	141	6			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	207	GLY	-	EXPRESSION TAG	UNP Q9UBU7
B	208	PRO	-	EXPRESSION TAG	UNP Q9UBU7
B	209	GLY	-	EXPRESSION TAG	UNP Q9UBU7
D	207	GLY	-	EXPRESSION TAG	UNP Q9UBU7
D	208	PRO	-	EXPRESSION TAG	UNP Q9UBU7
D	209	GLY	-	EXPRESSION TAG	UNP Q9UBU7

- Molecule 3 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: C<sub>10</sub>H<sub>15</sub>N<sub>5</sub>O<sub>10</sub>P<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0
			27	10	5	10	2	
3	C	1	Total	C	N	O	P	0
			27	10	5	10	2	

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mg	0	0
			1	1		
4	C	1	Total	Mg	0	0
			1	1		

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Zn	0	0
			1	1		
5	D	1	Total	Zn	0	0
			1	1		

- Molecule 6 is water.

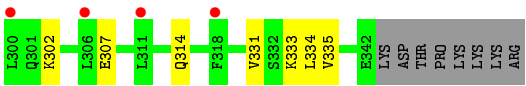
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	43	Total	O	0	0
			43	43		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	6	Total 6	O 6	0	0
6	C	17	Total 17	O 17	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.08 Å 67.04 Å 237.05 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	58.35 – 2.17 58.35 – 2.17	Depositor EDS
% Data completeness (in resolution range)	99.8 (58.35-2.17) 99.8 (58.35-2.17)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 2.16 Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.205 , 0.250 0.203 , 0.247	Depositor DCC
$R_{free}$ test set	2674 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.2	Xtriage
Anisotropy	0.669	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 44.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6643	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.00% 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG, ADP

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	0.60	1/2551 (0.0%)	0.69	1/3443 (0.0%)
1	C	0.53	2/2562 (0.1%)	0.63	0/3459
2	B	0.51	0/771	0.60	0/1042
2	D	0.45	0/771	0.59	0/1042
All	All	0.54	3/6655 (0.0%)	0.64	1/8986 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	398	TRP	CD2-CE2	5.51	1.48	1.41
1	C	535	TRP	CD2-CE2	5.23	1.47	1.41
1	A	398	TRP	CD2-CE2	5.11	1.47	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	464	ARG	NE-CZ-NH1	-5.01	117.79	120.30

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	2502	0	2541	7	0
1	C	2513	0	2552	13	0
2	B	752	0	712	4	0
2	D	752	0	712	3	0
3	A	27	0	12	0	0
3	C	27	0	12	0	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
5	B	1	0	0	0	0
5	D	1	0	0	0	0
6	A	43	0	0	0	0
6	B	6	0	0	0	0
6	C	17	0	0	0	0
All	All	6643	0	6541	23	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 2.

All (23) 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:A:220:GLN:HA	1:A:220:GLN:OE1	1.85	0.75
1:A:109:THR:HG21	2:B:333:LYS:HD3	1.77	0.66
1:C:145:ILE:HD13	1:C:149:LEU:HD11	1.82	0.61
1:A:216:SER:HB2	1:A:219:GLN:HB2	1.89	0.54
1:C:145:ILE:O	1:C:149:LEU:HG	2.08	0.54
1:C:123:CYS:HB3	2:D:331:VAL:HG22	1.92	0.52
1:C:121:LYS:HG2	1:C:135:PRO:HD3	1.94	0.50
1:A:176:ARG:HG2	1:A:395:ILE:HD11	1.93	0.49
1:C:187:ARG:O	1:C:190:LYS:HD3	2.14	0.48
1:C:155:ARG:HG2	1:C:568:PHE:CZ	2.49	0.47
1:A:429:THR:CG2	2:B:242:ILE:HG13	2.44	0.46
1:A:155:ARG:NH2	1:A:570:LYS:O	2.46	0.45
1:C:176:ARG:HG2	1:C:395:ILE:HD11	1.99	0.44
1:C:180:PRO:HD3	1:C:403:ILE:HG12	1.99	0.44
1:C:413:PRO:O	2:D:251:SER:HB2	2.18	0.43
1:C:52:PRO:O	1:C:55:SER:HB2	2.17	0.43
1:C:64:ILE:HD13	1:C:136:TYR:HE1	1.83	0.43
1:C:52:PRO:HD2	1:C:126:LYS:HD2	2.01	0.43
1:C:421:LEU:HA	1:C:421:LEU:HD23	1.89	0.42
2:B:299:CYS:SG	2:B:314:GLN:HG2	2.60	0.42
2:D:333:LYS:HG2	2:D:333:LYS:O	2.18	0.42

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:232:PHE:CD1	2:B:232:PHE:N	2.87	0.42
1:A:570:LYS:HA	1:A:571:ASP:HA	1.85	0.40

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	309/361 (86%)	302 (98%)	7 (2%)	0	100	100
1	C	311/361 (86%)	294 (94%)	16 (5%)	1 (0%)	44	47
2	B	86/144 (60%)	84 (98%)	2 (2%)	0	100	100
2	D	86/144 (60%)	83 (96%)	3 (4%)	0	100	100
All	All	792/1010 (78%)	763 (96%)	28 (4%)	1 (0%)	55	62

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	204	HIS

### 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	271/313 (87%)	263 (97%)	8 (3%)	46	55

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	272/313 (87%)	266 (98%)	6 (2%)	57	68
2	B	86/136 (63%)	80 (93%)	6 (7%)	18	17
2	D	86/136 (63%)	79 (92%)	7 (8%)	14	12
All	All	715/898 (80%)	688 (96%)	27 (4%)	38	45

All (27) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	41	LYS
1	A	68	THR
1	A	81	VAL
1	A	143	LEU
1	A	144	ASP
1	A	146	LEU
1	A	220	GLN
1	A	466	ARG
2	B	227	GLN
2	B	228	LEU
2	B	232	PHE
2	B	304	GLU
2	B	307	GLU
2	B	333	LYS
1	C	42	LYS
1	C	81	VAL
1	C	188	ARG
1	C	214	VAL
1	C	215	GLN
1	C	537	GLU
2	D	236	LEU
2	D	237	THR
2	D	302	LYS
2	D	307	GLU
2	D	314	GLN
2	D	334	LEU
2	D	335	VAL

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

Mol	Chain	Res	Type
2	B	238	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	317	ASN
1	C	127	ASN

### 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 ⓘ

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	ADP	A	601	4	25,29,29	1.01	2 (8%)	24,45,45	1.53	4 (16%)
3	ADP	C	601	4	25,29,29	1.16	2 (8%)	24,45,45	1.68	3 (12%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	ADP	A	601	4	-	0/12/32/32	0/3/3/3
3	ADP	C	601	4	-	0/12/32/32	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	601	ADP	C2-N3	2.17	1.35	1.32
3	A	601	ADP	C5-C4	2.46	1.46	1.40
3	C	601	ADP	O4'-C1'	2.59	1.44	1.41
3	C	601	ADP	C5-C4	3.17	1.47	1.40

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	601	ADP	N3-C2-N1	-6.17	123.48	128.86
3	A	601	ADP	N3-C2-N1	-5.34	124.21	128.86
3	C	601	ADP	C4-C5-N7	-2.43	107.06	109.41
3	A	601	ADP	C4-C5-N7	-2.14	107.34	109.41
3	A	601	ADP	N6-C6-N1	2.08	122.88	118.77
3	A	601	ADP	O3B-PB-O2B	2.26	116.74	107.61
3	C	601	ADP	O3B-PB-O1B	2.67	120.94	110.50

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 ⓘ

### 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<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	315/361 (87%)	0.04	4 (1%) 77 77	24, 40, 67, 90	0
1	C	317/361 (87%)	0.04	7 (2%) 62 62	29, 43, 79, 108	0
2	B	90/144 (62%)	0.18	2 (2%) 62 62	27, 54, 75, 80	0
2	D	90/144 (62%)	0.54	7 (7%) 14 14	37, 64, 98, 105	0
All	All	812/1010 (80%)	0.11	20 (2%) 58 58	24, 45, 82, 108	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	204	HIS	5.0
1	C	69	PHE	4.6
2	D	294	GLY	4.4
1	C	213	PHE	3.8
1	A	571	ASP	3.4
2	D	311	LEU	3.1
2	B	300	LEU	3.1
1	A	218	ALA	3.0
2	D	300	LEU	2.7
1	C	37	LEU	2.7
2	B	302	LYS	2.5
1	C	99	ILE	2.3
2	D	306	LEU	2.3
1	A	84	GLU	2.3
2	D	318	PHE	2.3
2	D	228	LEU	2.2
2	D	230	ARG	2.2
1	C	217	GLU	2.1
1	A	69	PHE	2.1
1	C	220	GLN	2.0

## 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 carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ADP	A	601	27/27	0.98	0.10	-0.45	29,35,43,50	0
3	ADP	C	601	27/27	0.97	0.10	-0.70	31,38,48,56	0
5	ZN	D	401	1/1	0.96	0.07	-2.03	85,85,85,85	0
4	MG	C	602	1/1	0.90	0.06	-	44,44,44,44	0
4	MG	A	602	1/1	0.96	0.06	-	33,33,33,33	0
5	ZN	B	401	1/1	0.99	0.13	-	52,52,52,52	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.