



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 4, 2017 – 10:04 am GMT

PDB ID : 5IOL
Title : Crystal structure of Nucleoside Diphosphate Kinase from *Schistosoma mansoni*
Authors : Torini, J.R.S.; Romanello, L.; Bird, L.E.; Nettleship, J.E.; Owens, R.J.; Aller, P.; DeMarco, R.; Brandao-Neto, J.; Pereira, H.M.
Deposited on : 2016-03-08
Resolution : 1.74 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 ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	recalc29047
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)	:	recalc29047

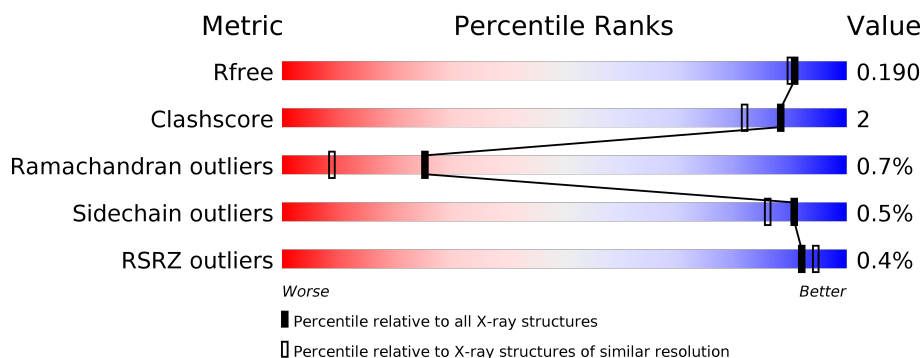
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.74 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	2694 (1.76-1.72)
Clashscore	112137	2854 (1.76-1.72)
Ramachandran outliers	110173	2824 (1.76-1.72)
Sidechain outliers	110143	2824 (1.76-1.72)
RSRZ outliers	101464	2705 (1.76-1.72)

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 $\leq 5\%$. 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	150	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 1%, green 98%);"></div> <div style="display: flex; justify-content: space-between; width: 90%; margin: 2px auto;"> % 93% 7% </div> </div>
1	B	150	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 1%, green 98%);"></div> <div style="display: flex; justify-content: space-between; width: 90%; margin: 2px auto;"> % 91% 8% </div> </div>
1	C	150	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 99%, yellow 1%);"></div> <div style="display: flex; justify-content: space-between; width: 90%; margin: 2px auto;"> % 93% 6% </div> </div>
1	D	150	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 1%, green 98%);"></div> <div style="display: flex; justify-content: space-between; width: 90%; margin: 2px auto;"> % 90% 10% </div> </div>
1	E	150	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 1%, green 98%);"></div> <div style="display: flex; justify-content: space-between; width: 90%; margin: 2px auto;"> % 93% 7% </div> </div>
1	F	150	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 99%, yellow 1%);"></div> <div style="display: flex; justify-content: space-between; width: 90%; margin: 2px auto;"> % 91% 9% </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	150	<div><div></div><div>95%</div><div></div></div> <div><div></div><div></div><div></div></div>
1	H	150	<div>%<div><div></div><div>97%</div><div></div></div><div><div></div><div></div><div></div></div></div>
1	I	150	<div><div></div><div>97%</div><div></div></div> <div><div></div><div></div><div></div></div>
1	J	150	<div><div></div><div>94%</div><div>5%</div></div> <div><div></div><div></div><div></div></div>
1	K	150	<div><div></div><div>94%</div><div>5%</div></div> <div><div></div><div></div><div></div></div>
1	L	150	<div>%<div><div></div><div>97%</div><div></div></div><div><div></div><div></div><div></div></div></div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 16239 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 Nucleoside diphosphate kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	149	Total	C	N	O	S	0	2	0
			1189	752	209	219	9			
1	B	149	Total	C	N	O	S	0	1	0
			1183	748	211	215	9			
1	C	149	Total	C	N	O	S	0	0	0
			1179	745	209	216	9			
1	D	150	Total	C	N	O	S	0	0	0
			1183	747	210	217	9			
1	E	149	Total	C	N	O	S	0	0	0
			1179	745	209	216	9			
1	F	150	Total	C	N	O	S	0	1	0
			1191	752	213	217	9			
1	G	149	Total	C	N	O	S	0	0	0
			1179	745	209	216	9			
1	H	150	Total	C	N	O	S	0	0	0
			1183	747	210	217	9			
1	I	149	Total	C	N	O	S	0	0	0
			1179	745	209	216	9			
1	J	149	Total	C	N	O	S	0	1	0
			1187	750	212	216	9			
1	K	149	Total	C	N	O	S	0	0	0
			1179	745	209	216	9			
1	L	150	Total	C	N	O	S	0	0	0
			1183	747	210	217	9			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP G4VJY9
A	1	PRO	-	expression tag	UNP G4VJY9
B	0	GLY	-	expression tag	UNP G4VJY9
B	1	PRO	-	expression tag	UNP G4VJY9
C	0	GLY	-	expression tag	UNP G4VJY9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	PRO	-	expression tag	UNP G4VJY9
D	0	GLY	-	expression tag	UNP G4VJY9
D	1	PRO	-	expression tag	UNP G4VJY9
E	0	GLY	-	expression tag	UNP G4VJY9
E	1	PRO	-	expression tag	UNP G4VJY9
F	0	GLY	-	expression tag	UNP G4VJY9
F	1	PRO	-	expression tag	UNP G4VJY9
G	0	GLY	-	expression tag	UNP G4VJY9
G	1	PRO	-	expression tag	UNP G4VJY9
H	0	GLY	-	expression tag	UNP G4VJY9
H	1	PRO	-	expression tag	UNP G4VJY9
I	0	GLY	-	expression tag	UNP G4VJY9
I	1	PRO	-	expression tag	UNP G4VJY9
J	0	GLY	-	expression tag	UNP G4VJY9
J	1	PRO	-	expression tag	UNP G4VJY9
K	0	GLY	-	expression tag	UNP G4VJY9
K	1	PRO	-	expression tag	UNP G4VJY9
L	0	GLY	-	expression tag	UNP G4VJY9
L	1	PRO	-	expression tag	UNP G4VJY9

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	135	Total O 135 135	0	0
2	B	153	Total O 153 153	0	0
2	C	176	Total O 176 176	0	0
2	D	169	Total O 169 169	0	0
2	E	172	Total O 172 172	0	0
2	F	182	Total O 182 182	0	0
2	G	187	Total O 187 187	0	0
2	H	178	Total O 178 178	0	0
2	I	172	Total O 172 172	0	0
2	J	204	Total O 204 204	0	0

Continued on next page...

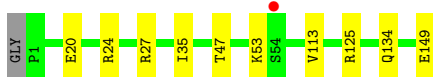
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	K	162	Total 162	O 162	0	0
2	L	155	Total 155	O 155	0	0

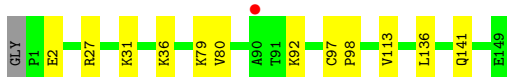
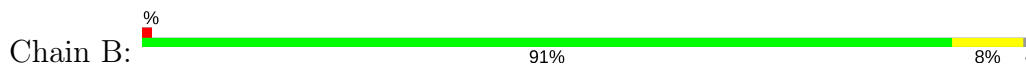
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Nucleoside diphosphate kinase



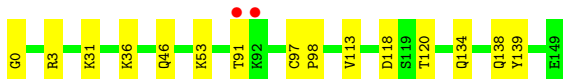
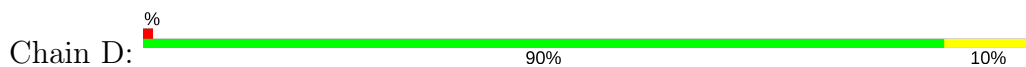
- Molecule 1: Nucleoside diphosphate kinase



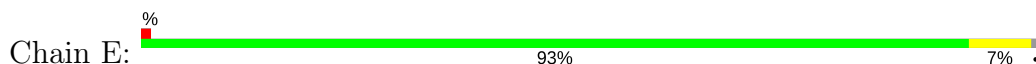
- Molecule 1: Nucleoside diphosphate kinase



- Molecule 1: Nucleoside diphosphate kinase

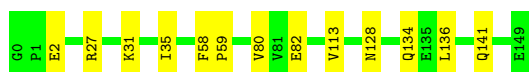


- Molecule 1: Nucleoside diphosphate kinase



- Molecule 1: Nucleoside diphosphate kinase





- Molecule 1: Nucleoside diphosphate kinase

Chain G: 95% ..



- Molecule 1: Nucleoside diphosphate kinase

Chain H: 97% .



- Molecule 1: Nucleoside diphosphate kinase

Chain I: 97% ..



- Molecule 1: Nucleoside diphosphate kinase

Chain J: 94% 5% .



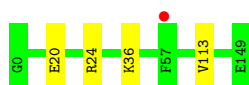
- Molecule 1: Nucleoside diphosphate kinase

Chain K: 94% 5% .



- Molecule 1: Nucleoside diphosphate kinase

Chain L: 97% .



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	211.76Å 82.27Å 122.37Å 90.00° 104.89° 90.00°	Depositor
Resolution (Å)	59.13 – 1.74 59.13 – 1.74	Depositor EDS
% Data completeness (in resolution range)	98.2 (59.13-1.74) 98.2 (59.13-1.74)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.12 (at 1.74Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.163 , 0.191 0.162 , 0.190	Depositor DCC
R_{free} test set	10261 reflections (5.03%)	DCC
Wilson B-factor (Å ²)	22.0	Xtriage
Anisotropy	0.345	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 43.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	16239	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.81% 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	A	0.28	0/1222	0.47	0/1649
1	B	0.38	0/1213	0.49	0/1636
1	C	0.29	0/1206	0.46	0/1627
1	D	0.29	0/1210	0.48	0/1633
1	E	0.28	0/1206	0.47	0/1627
1	F	0.29	0/1221	0.47	0/1647
1	G	0.29	0/1206	0.47	0/1627
1	H	0.30	0/1210	0.48	0/1633
1	I	0.29	0/1206	0.49	0/1627
1	J	0.31	0/1217	0.50	0/1641
1	K	0.29	0/1206	0.48	0/1627
1	L	0.29	0/1210	0.48	0/1633
All	All	0.30	0/14533	0.48	0/19607

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	1189	0	1185	6	0
1	B	1183	0	1179	8	0
1	C	1179	0	1172	6	0
1	D	1183	0	1175	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1179	0	1172	9	0
1	F	1191	0	1188	8	0
1	G	1179	0	1172	4	0
1	H	1183	0	1175	4	0
1	I	1179	0	1172	2	0
1	J	1187	0	1185	6	0
1	K	1179	0	1172	6	0
1	L	1183	0	1175	2	0
2	A	135	0	0	2	0
2	B	153	0	0	5	0
2	C	176	0	0	4	0
2	D	169	0	0	8	0
2	E	172	0	0	5	0
2	F	182	0	0	4	0
2	G	187	0	0	1	0
2	H	178	0	0	2	0
2	I	172	0	0	1	0
2	J	204	0	0	3	0
2	K	162	0	0	2	0
2	L	155	0	0	1	0
All	All	16239	0	14122	65	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.

The worst 5 of 65 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:D:46:GLN:NE2	2:D:202:HOH:O	2.06	0.88
1:D:98:PRO:O	2:D:201:HOH:O	1.94	0.84
1:D:97:CYS:SG	2:L:342:HOH:O	2.37	0.81
1:A:53:LYS:NZ	2:A:201:HOH:O	2.17	0.78
1:E:92:LYS:HA	1:E:109:VAL:HG22	1.69	0.74

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	149/150 (99%)	145 (97%)	3 (2%)	1 (1%)	25	9
1	B	148/150 (99%)	144 (97%)	3 (2%)	1 (1%)	25	9
1	C	147/150 (98%)	143 (97%)	3 (2%)	1 (1%)	25	9
1	D	148/150 (99%)	144 (97%)	3 (2%)	1 (1%)	25	9
1	E	147/150 (98%)	143 (97%)	3 (2%)	1 (1%)	25	9
1	F	149/150 (99%)	145 (97%)	3 (2%)	1 (1%)	25	9
1	G	147/150 (98%)	144 (98%)	2 (1%)	1 (1%)	25	9
1	H	148/150 (99%)	145 (98%)	2 (1%)	1 (1%)	25	9
1	I	147/150 (98%)	143 (97%)	3 (2%)	1 (1%)	25	9
1	J	148/150 (99%)	144 (97%)	3 (2%)	1 (1%)	25	9
1	K	147/150 (98%)	143 (97%)	3 (2%)	1 (1%)	25	9
1	L	148/150 (99%)	144 (97%)	3 (2%)	1 (1%)	25	9
All	All	1773/1800 (98%)	1727 (97%)	34 (2%)	12 (1%)	25	9

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	I	113	VAL
1	B	113	VAL
1	C	113	VAL
1	L	113	VAL
1	D	113	VAL

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	131/129 (102%)	130 (99%)	1 (1%)	85	75
1	B	129/129 (100%)	128 (99%)	1 (1%)	85	75
1	C	129/129 (100%)	127 (98%)	2 (2%)	68	49
1	D	129/129 (100%)	127 (98%)	2 (2%)	68	49
1	E	129/129 (100%)	129 (100%)	0	100	100
1	F	130/129 (101%)	129 (99%)	1 (1%)	85	75
1	G	129/129 (100%)	129 (100%)	0	100	100
1	H	129/129 (100%)	129 (100%)	0	100	100
1	I	129/129 (100%)	129 (100%)	0	100	100
1	J	130/129 (101%)	130 (100%)	0	100	100
1	K	129/129 (100%)	128 (99%)	1 (1%)	85	75
1	L	129/129 (100%)	129 (100%)	0	100	100
All	All	1552/1548 (100%)	1544 (100%)	8 (0%)	91	86

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	149	GLU
1	K	149	GLU
1	D	134	GLN
1	C	78	ARG
1	D	118	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	L	39	HIS

5.3.3 RNA ⓘ

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 carbohydrates 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 ⓘ

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, 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	A	149/150 (99%)	-0.31	1 (0%) 87 91	17, 25, 45, 53	0
1	B	149/150 (99%)	-0.33	1 (0%) 87 91	15, 24, 44, 58	0
1	C	149/150 (99%)	-0.30	0 100 100	16, 25, 40, 47	0
1	D	150/150 (100%)	-0.33	2 (1%) 77 83	14, 22, 43, 49	0
1	E	149/150 (99%)	-0.30	1 (0%) 87 91	16, 25, 40, 51	0
1	F	150/150 (100%)	-0.22	0 100 100	16, 23, 40, 58	0
1	G	149/150 (99%)	-0.29	0 100 100	14, 21, 35, 52	0
1	H	150/150 (100%)	-0.17	2 (1%) 77 83	13, 21, 40, 62	0
1	I	149/150 (99%)	-0.40	0 100 100	14, 20, 35, 47	0
1	J	149/150 (99%)	-0.21	0 100 100	13, 19, 35, 46	0
1	K	149/150 (99%)	-0.38	0 100 100	17, 23, 42, 53	0
1	L	150/150 (100%)	-0.34	1 (0%) 87 91	14, 21, 43, 54	0
All	All	1792/1800 (99%)	-0.30	8 (0%) 92 94	13, 22, 42, 62	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	0	GLY	3.5
1	B	90	ALA	2.6
1	H	1	PRO	2.5
1	D	91	THR	2.3
1	L	57	PHE	2.3

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 [i](#)

There are no carbohydrates 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.