



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 03:01 PM BST

PDB ID : 1J5K  
Title : COMPLEX OF THE KH3 DOMAIN OF HNRNP K WITH A SINGLE\_STRANDED 10MER DNA OLIGONUCLEOTIDE  
Authors : Clore, G.M.; Braddock, D.T.  
Deposited on : 2002-05-13

This is a Full wwPDB NMR 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/NMRValidationReportHelp>  
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:

Cyrange : Kirchner and Güntert (2011)  
NmrClust : Kelley et al. (1996)  
MolProbity : 4.02b-467  
Mogul : unknown  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : rb-20027457  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027457

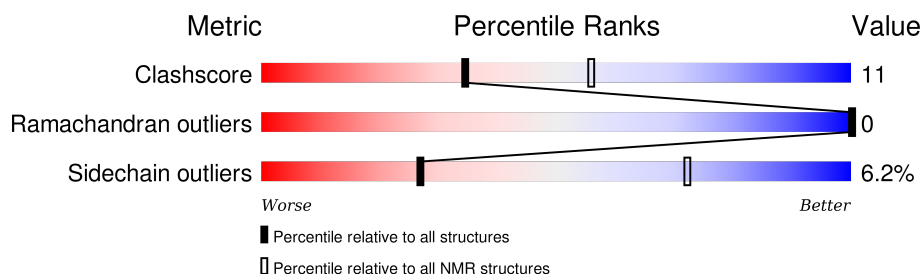
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment was not calculated.

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)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	B	10	<div> <div>20%</div> <div>30%</div> <div>50%</div> </div>
2	A	89	<div> <div>66%</div> <div>19%</div> <div>15%</div> </div>

## 2 Ensemble composition and analysis ⓘ

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1321 atoms, of which 650 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called 5'-D(\*AP\*TP\*AP\*T\*TP\*CP\*CP\*CP\*TP\*C)-3'.

Mol	Chain	Residues	Atoms						Trace
1	B	5	Total	C	H	N	O	P	0
			151	47	57	13	30	4	

- Molecule 2 is a protein called Heterogeneous nuclear ribonucleoprotein K.

Mol	Chain	Residues	Atoms					Trace
2	A	76	Total	C	H	N	O	0
			1170	357	593	103	117	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P61978
A	2	SER	-	CLONING ARTIFACT	UNP P61978
A	3	HIS	-	CLONING ARTIFACT	UNP P61978
A	4	MET	-	CLONING ARTIFACT	UNP P61978

## 4 Residue-property plots

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: 5'-D(\*AP\*TP\*AP\*T\*TP\*CP\*CP\*CP\*TP\*C)-3'

Chain B: 

DA	DT	DA	DT	T105	C106	C107	C108	T109	DC
----	----	----	----	------	------	------	------	------	----

- Molecule 2: Heterogeneous nuclear ribonucleoprotein K

Chain A: 

GLY	SER	HIS	MET	SER	TYR	GLY	ASP	LEU	G10	K22	D23	L24	S27	I28	I29	G30	K31	G32	G33	R40	G44	K48	I49	D50	S56	E57	D58	D68	Q69	V81	Y84	S85	GLY	LYS	PHE	PHE
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *SIMULATED ANNEALING*.

Of the 125 calculated structures, 1 were deposited, based on the following criterion: *REGULARIZED MEAN STRUCTURE*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR NIH	refinement	( <a href="http://NMR.CIT.NIH.GOV/XPLOR_NIH">HTTP://NMR.CIT.NIH.GOV/XPLOR_NIH</a> )
XPLOR NIH	structure solution	

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

## 6 Model quality [i](#)

### 6.1 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 (average) 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	B	1.33	0/103 (0.0%)	2.26	9/156 (5.8%)
2	A	1.05	0/582 (0.0%)	0.90	0/784 (0.0%)
All	All	1.09	0/685 (0.0%)	1.23	9/940 (1.0%)

There are no bond-length outliers.

All angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	105	DT	C4'-C3'-C2'	8.87	111.08	103.10
1	B	106	DC	C4'-C3'-C2'	7.58	109.92	103.10
1	B	109	DT	C4'-C3'-C2'	7.53	109.88	103.10
1	B	108	DC	C4'-C3'-C2'	7.38	109.74	103.10
1	B	107	DC	C4'-C3'-C2'	6.87	109.28	103.10
1	B	105	DT	C4-C5-C6	5.50	121.30	118.00
1	B	109	DT	C6-C5-C7	-5.18	119.79	122.90
1	B	109	DT	C4-C5-C6	5.14	121.09	118.00
1	B	105	DT	C6-C5-C7	-5.08	119.85	122.90

There are no chirality outliers.

There are no planarity outliers.

### 6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	B	94	57	59	8
2	A	577	593	592	11

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	671	650	651	14

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 11.

All clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
1:B:109:DT:H73	1:B:109:DT:OP2	0.62	1.94
2:A:24:LEU:CD2	2:A:84:TYR:CD2	0.55	2.90
2:A:57:GLU:O	2:A:58:ASP:OD1	0.52	2.28
1:B:109:DT:H5'	2:A:40:ARG:NH1	0.48	2.23
1:B:106:DC:O2	2:A:29:ILE:HG21	0.48	2.08
1:B:109:DT:O5'	1:B:109:DT:H6	0.47	1.91
1:B:109:DT:OP2	1:B:109:DT:C7	0.45	2.63
1:B:109:DT:O2	2:A:48:LYS:CD	0.44	2.66
1:B:108:DC:O3'	2:A:40:ARG:NH1	0.42	2.41
2:A:81:VAL:O	2:A:81:VAL:HG12	0.42	2.14
1:B:106:DC:O2	2:A:29:ILE:CG2	0.41	2.68
2:A:44:GLY:O	2:A:69:GLN:NE2	0.41	2.54
2:A:22:LYS:O	2:A:24:LEU:N	0.41	2.54
2:A:31:LYS:C	2:A:33:GLY:H	0.40	2.19

## 6.3 Torsion angles

### 6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	
2	A	74/89 (83%)	70 (95%)	4 (5%)	0 (0%)	100	100
All	All	74/89 (83%)	70 (95%)	4 (5%)	0 (0%)	100	100

There are no Ramachandran outliers.

### 6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	
2	A	64/74 (86%)	60 (94%)	4 (6%)	27	72
All	All	64/74 (86%)	60 (94%)	4 (6%)	27	72

All 4 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
2	A	27	SER
2	A	68	ASP
2	A	50	ASP
2	A	56	SER

### 6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 6.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.6 Ligand geometry [i](#)

There are no ligands in this entry.

### 6.7 Other polymers [i](#)

There are no such molecules in this entry.

## 6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

## 7 Chemical shift validation

No chemical shift data were provided