



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2017 – 07:18 pm GMT

PDB ID : 5HN2  
Title : Base Pairing and Structure Insights into the 5-Formylcytosine in RNA Duplex  
Authors : Luo, Z.P.; Sheng, J.  
Deposited on : 2016-01-18  
Resolution : 1.50 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 : trunk28620  
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) : recalc28949

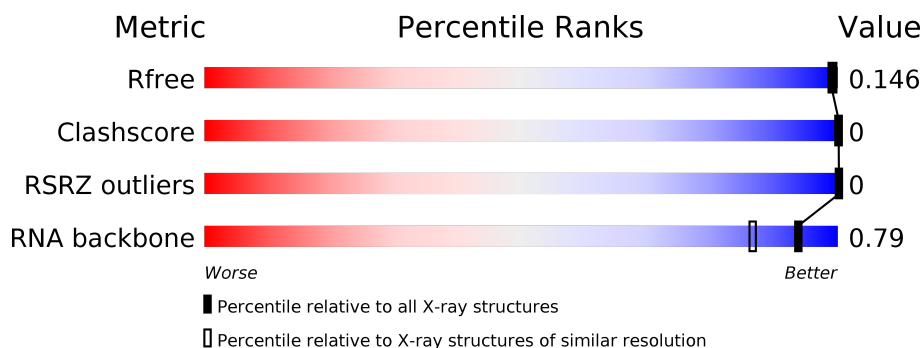
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.50 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	2279 (1.50-1.50)
Clashscore	112137	2503 (1.50-1.50)
RSRZ outliers	101464	2305 (1.50-1.50)
RNA backbone	2435	1043 (2.60-0.62)

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	8	<div> <div></div> <div>75%</div> <div>25%</div> </div>
1	B	8	<div> <div></div> <div>75%</div> <div>25%</div> </div>
1	C	8	<div> <div></div> <div>63%</div> <div>38%</div> </div>
1	D	8	<div> <div></div> <div>75%</div> <div>25%</div> </div>
1	E	8	<div> <div></div> <div>88%</div> <div>13%</div> </div>
1	F	8	<div> <div></div> <div>88%</div> <div>13%</div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 1212 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 RNA chain called RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	8	Total	C	N	O	P	0	0	0
			169	77	30	55	7			
1	B	8	Total	C	N	O	P	0	1	0
			171	78	30	56	7			
1	C	8	Total	C	N	O	P	0	0	0
			169	77	30	55	7			
1	D	8	Total	C	N	O	P	0	0	0
			169	77	30	55	7			
1	E	8	Total	C	N	O	P	0	0	0
			169	77	30	55	7			
1	F	8	Total	C	N	O	P	0	0	0
			169	77	30	55	7			

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	F	1	Total	Na	0	0
			1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	31	Total	O	0	0
			31	31		
3	B	28	Total	O	0	0
			28	28		
3	C	24	Total	O	0	0
			24	24		
3	D	23	Total	O	0	0
			23	23		
3	E	49	Total	O	0	0
			49	49		

*Continued on next page...*

*Continued from previous page...*

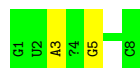
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	F	40	Total	O	0	0
			40	40		

### 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: RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3')

Chain A: 



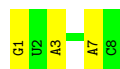
- Molecule 1: RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3')

Chain B: 



- Molecule 1: RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3')

Chain C: 




- Molecule 1: RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3')

Chain D: 




- Molecule 1: RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3')

Chain E: 



- Molecule 1: RNA (5'-R(\*GP\*UP\*AP\*(OFC)P\*GP\*UP\*AP\*C)-3')

Chain F: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.34Å 42.34Å 58.65Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	22.90 – 1.50 22.90 – 1.50	Depositor EDS
% Data completeness (in resolution range)	98.1 (22.90-1.50) 98.1 (22.90-1.50)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.82 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, $R_{free}$	0.123 , 0.140 0.130 , 0.146	Depositor DCC
$R_{free}$ test set	896 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	17.5	Xtriage
Anisotropy	0.256	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 49.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.006 for -h,-k,l 0.056 for h,-h-k,-l 0.030 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	1212	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 30.72 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.2477e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: NA, RSQ

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	1.92	2/163 (1.2%)	1.69	3/250 (1.2%)
1	B	1.44	0/186	1.37	2/287 (0.7%)
1	C	1.57	1/163 (0.6%)	1.35	1/250 (0.4%)
1	D	1.61	1/163 (0.6%)	1.36	2/250 (0.8%)
1	E	1.62	1/163 (0.6%)	1.27	0/250
1	F	1.51	0/163	1.39	2/250 (0.8%)
All	All	1.61	5/1001 (0.5%)	1.41	10/1537 (0.7%)

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
1	C	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	3	A	P-OP2	5.92	1.59	1.49
1	E	7	A	C3'-O3'	5.34	1.49	1.42
1	C	7	A	P-OP2	-5.25	1.40	1.49
1	A	5	G	C3'-C2'	5.21	1.58	1.52
1	D	7	A	P-O5'	-5.18	1.54	1.59

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	3	A	O5'-P-OP2	-15.44	91.80	105.70
1	B	6	U	O5'-P-OP2	-8.58	97.97	105.70

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	5	G	O5'-P-OP2	-7.17	99.25	105.70
1	C	3	A	O5'-P-OP2	-6.79	99.59	105.70
1	B	7	A	O5'-P-OP2	-6.77	99.61	105.70
1	F	7	A	O5'-P-OP2	-6.58	99.78	105.70
1	A	5	G	C3'-C2'-C1'	-5.79	96.86	101.50
1	A	3	A	OP1-P-OP2	5.62	128.02	119.60
1	D	7	A	OP1-P-OP2	5.39	127.68	119.60
1	F	7	A	C2'-C3'-O3'	5.28	122.15	113.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	1	G	Sidechain

## 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	169	0	88	0	0
1	B	171	0	84	0	0
1	C	169	0	88	0	0
1	D	169	0	88	0	0
1	E	169	0	88	0	0
1	F	169	0	88	0	0
2	F	1	0	0	0	0
3	A	31	0	0	0	0
3	B	28	0	0	0	0
3	C	24	0	0	0	0
3	D	23	0	0	0	0
3	E	49	0	0	0	0
3	F	40	0	0	0	0
All	All	1212	0	524	0	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 0.

There are no clashes within the asymmetric unit.

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

There are no protein molecules in this entry.

### 5.3.2 Protein sidechains [i](#)

There are no protein molecules in this entry.

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	6/8 (75%)	0	0
1	B	6/8 (75%)	0	0
1	C	6/8 (75%)	0	0
1	D	6/8 (75%)	0	0
1	E	6/8 (75%)	0	0
1	F	6/8 (75%)	0	0
All	All	36/48 (75%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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

6 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\# Z  > 2$
1	RSQ	A	4	1	15,23,24	1.10	1 (6%)	17,33,36	2.60	4 (23%)
1	RSQ	B	4	1	15,23,24	1.63	2 (13%)	17,33,36	2.01	4 (23%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	RSQ	C	4	1	15,23,24	1.34	1 (6%)	17,33,36	2.11	5 (29%)
1	RSQ	D	4	1	15,23,24	1.36	2 (13%)	17,33,36	1.32	2 (11%)
1	RSQ	E	4	1	15,23,24	2.03	5 (33%)	17,33,36	2.29	3 (17%)
1	RSQ	F	4	1	15,23,24	1.27	2 (13%)	17,33,36	1.48	4 (23%)

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
1	RSQ	A	4	1	-	0/5/27/28	0/2/2/2
1	RSQ	B	4	1	-	0/5/27/28	0/2/2/2
1	RSQ	C	4	1	-	0/5/27/28	0/2/2/2
1	RSQ	D	4	1	-	0/5/27/28	0/2/2/2
1	RSQ	E	4	1	-	0/5/27/28	0/2/2/2
1	RSQ	F	4	1	-	0/5/27/28	0/2/2/2

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	4	RSQ	O4'-C1'	-4.00	1.35	1.41
1	B	4	RSQ	C2'-C1'	-3.69	1.47	1.53
1	E	4	RSQ	O5'-C5'	-3.25	1.40	1.44
1	E	4	RSQ	O4'-C4'	-2.88	1.38	1.45
1	F	4	RSQ	C6-C5	-2.75	1.34	1.39
1	D	4	RSQ	O5'-C5'	-2.64	1.41	1.44
1	D	4	RSQ	C6-C5	-2.40	1.35	1.39
1	E	4	RSQ	C5'-C4'	-2.24	1.44	1.51
1	E	4	RSQ	C2'-C1'	-2.18	1.50	1.53
1	F	4	RSQ	O4'-C1'	2.11	1.44	1.41
1	A	4	RSQ	O3'-C3'	2.11	1.47	1.43
1	B	4	RSQ	O4'-C1'	3.33	1.45	1.41
1	E	4	RSQ	O4'-C1'	4.74	1.47	1.41

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	4	RSQ	C5-C4-N4	-9.19	114.43	122.79
1	B	4	RSQ	C5-C4-N4	-6.02	117.31	122.79
1	E	4	RSQ	C5-C4-N4	-5.94	117.39	122.79

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	4	RSQ	C2'-C3'-C4'	-3.48	95.84	102.62
1	F	4	RSQ	C5-C4-N4	-3.40	119.70	122.79
1	E	4	RSQ	C5-C4-N3	-3.34	116.59	121.11
1	C	4	RSQ	N4-C4-N3	-2.76	112.93	117.00
1	D	4	RSQ	O4'-C4'-C3'	-2.46	100.28	105.17
1	A	4	RSQ	O4'-C4'-C3'	-2.35	100.50	105.17
1	B	4	RSQ	O4'-C4'-C3'	-2.25	100.70	105.17
1	D	4	RSQ	C5-C4-N3	-2.22	118.10	121.11
1	F	4	RSQ	O4'-C4'-C3'	-2.14	100.90	105.17
1	A	4	RSQ	N4-C4-N3	2.04	120.02	117.00
1	F	4	RSQ	C6-C5-C4	2.40	117.06	114.68
1	B	4	RSQ	C5-C4-N3	2.41	124.37	121.11
1	C	4	RSQ	C4'-O4'-C1'	2.49	112.42	109.77
1	F	4	RSQ	O4'-C1'-N1	2.49	113.08	108.08
1	B	4	RSQ	C6-C5-C4	2.62	117.28	114.68
1	A	4	RSQ	C5-C4-N3	3.27	125.55	121.11
1	C	4	RSQ	C5-C4-N4	3.80	126.25	122.79
1	C	4	RSQ	C6-C5-C4	5.05	119.69	114.68
1	E	4	RSQ	N4-C4-N3	5.97	125.82	117.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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	7/8 (87%)	-1.15	0 100 100	18, 18, 24, 24	0
1	B	7/8 (87%)	-1.16	0 100 100	18, 19, 23, 25	0
1	C	7/8 (87%)	-1.08	0 100 100	20, 21, 27, 29	0
1	D	7/8 (87%)	-1.01	0 100 100	20, 22, 28, 30	0
1	E	7/8 (87%)	-1.07	0 100 100	17, 19, 25, 26	0
1	F	7/8 (87%)	-1.02	0 100 100	18, 21, 28, 31	0
All	All	42/48 (87%)	-1.08	0 100 100	17, 22, 28, 31	0

There are no RSRZ outliers to report.

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

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(Å <sup>2</sup> )	Q<0.9
1	RSQ	A	4	22/23	0.99	0.04	-	14,18,21,25	0
1	RSQ	B	4	22/23	0.99	0.05	-	14,16,20,20	0
1	RSQ	E	4	22/23	0.98	0.05	-	17,22,25,29	0
1	RSQ	C	4	22/23	0.98	0.04	-	18,24,32,41	0
1	RSQ	F	4	22/23	0.99	0.05	-	16,18,22,23	0
1	RSQ	D	4	22/23	0.99	0.05	-	15,18,22,23	0

### 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
2	NA	F	101	1/1	0.98	0.16	-	30,30,30,30	0

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

There are no such residues in this entry.