



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 7, 2023 – 05:44 AM EDT

PDB ID : 4GSI  
Title : DNA Holliday junction stabilized by fluorine halogen bond. F2J construct of related reference.  
Authors : Ho, P.S.; Carter, M.  
Deposited on : 2012-08-27  
Resolution : 2.38 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.35

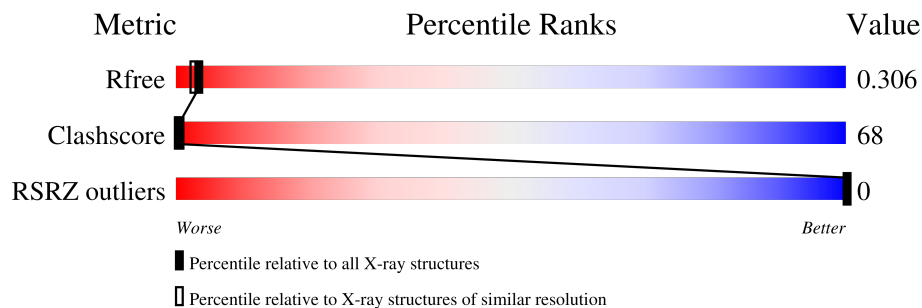
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.38 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.



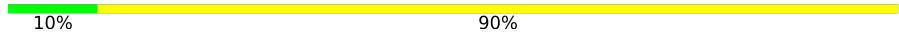
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	10	90% 10%
1	C	10	10% 80% 10%
1	F	10	20% 70% 10%
1	H	10	20% 70% 10%
2	B	10	20% 80%
2	D	10	20% 80%
2	E	10	100%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	10	 10% 90%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
1	UFP	A	7[A]	-	-	X	-

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 1716 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 DNA chain called DNA (5'-D(\*CP\*CP\*GP\*GP\*TP\*AP\*(UFP)P\*CP\*GP\*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	F	N	O	P			
1	A	10	203	96	1	38	59	9	0	10	0
1	C	10	203	96	1	38	59	9	0	10	0
1	F	10	203	96	1	38	59	9	0	10	0
1	H	10	203	96	1	38	59	9	0	10	0

- Molecule 2 is a DNA chain called DNA (5'-D(\*CP\*CP\*GP\*AP\*TP\*AP\*CP\*CP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	10	201	96	39	57	9	0	10	0
2	D	10	201	96	39	57	9	0	10	0
2	E	10	201	96	39	57	9	0	10	0
2	G	10	201	96	39	57	9	0	10	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	8	Total 8 8	0	0
3	B	10	Total 10 10	0	0
3	C	10	Total 10 10	0	0

*Continued on next page...*

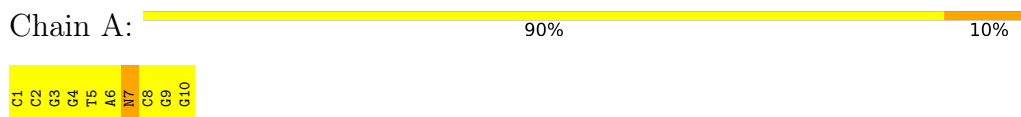
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
3	D	8	Total O 8 8	0	0
3	E	17	Total O 17 17	0	0
3	F	22	Total O 22 22	0	0
3	G	17	Total O 17 17	0	0
3	H	8	Total O 8 8	0	0

### 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 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: DNA (5'-D(\*CP\*CP\*GP\*GP\*TP\*AP\*(UFP)P\*CP\*GP\*G)-3')



- Molecule 1: DNA (5'-D(\*CP\*CP\*GP\*GP\*TP\*AP\*(UFP)P\*CP\*GP\*G)-3')



- Molecule 1: DNA (5'-D(\*CP\*CP\*GP\*GP\*TP\*AP\*(UFP)P\*CP\*GP\*G)-3')



- Molecule 1: DNA (5'-D(\*CP\*CP\*GP\*GP\*TP\*AP\*(UFP)P\*CP\*GP\*G)-3')



- Molecule 2: DNA (5'-D(\*CP\*CP\*GP\*AP\*TP\*AP\*CP\*CP\*GP\*G)-3')



- Molecule 2: DNA (5'-D(\*CP\*CP\*GP\*AP\*TP\*AP\*CP\*CP\*GP\*G)-3')




C31  
C32  
G33  
A34  
T35  
A36  
C37  
C38  
G39  
G40

- Molecule 2: DNA (5'-D(\*CP\*CP\*GP\*AP\*TP\*AP\*CP\*CP\*GP\*G)-3')

Chain E:  100%

C201  
C202  
G203  
A204  
T205  
A206  
C207  
C208  
G209  
G210

- Molecule 2: DNA (5'-D(\*CP\*CP\*GP\*AP\*TP\*AP\*CP\*CP\*GP\*G)-3')

Chain G:  10%  90%

C221  
G222  
G223  
A224  
T225  
A226  
C227  
C228  
G229  
G230

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	65.21Å 23.92Å 77.45Å 90.00° 114.80° 90.00°	Depositor
Resolution (Å)	32.61 – 2.38 32.61 – 2.30	Depositor EDS
% Data completeness (in resolution range)	74.2 (32.61-2.38) 70.1 (32.61-2.30)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	18.29 (at 2.31Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.225 , 0.291 0.224 , 0.306	Depositor DCC
$R_{free}$ test set	190 reflections (5.39%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.6	Xtrriage
Anisotropy	1.166	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 79.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.33$ , $\langle L^2 \rangle = 0.16$	Xtrriage
Estimated twinning fraction	0.108 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	1716	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	15.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 29.27 % 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 1.5961e-03.*

<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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: UFP

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	0.22	0/204	0.61	0/311
1	C	0.23	0/204	0.65	0/311
1	F	0.23	0/204	0.58	0/311
1	H	0.25	0/204	0.62	0/311
2	B	0.23	0/225	0.58	0/345
2	D	0.25	0/225	0.62	0/345
2	E	0.22	0/225	0.62	0/345
2	G	0.24	0/225	0.65	0/345
All	All	0.24	0/1716	0.62	0/2624

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	203	0	105	34	0
1	C	203	0	107	21	0
1	F	203	0	106	30	0
1	H	203	0	108	11	0
2	B	201	0	110	33	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	201	0	108	12	0
2	E	201	0	110	29	0
2	G	201	0	110	20	0
3	A	8	0	0	0	0
3	B	10	0	0	8	0
3	C	10	0	0	4	0
3	D	8	0	0	3	0
3	E	17	0	0	7	0
3	F	22	0	0	18	0
3	G	17	0	0	4	0
3	H	8	0	0	1	0
All	All	1716	0	864	165	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 68.

The worst 5 of 165 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:4[A]:DG:C8	1:A:5[A]:DT:H71	1.73	1.24
1:A:4[A]:DG:N7	1:A:5[A]:DT:H73	1.58	1.18
1:A:4[A]:DG:C8	1:A:5[A]:DT:C7	2.27	1.16
2:E:203[B]:DG:H2''	2:E:204[B]:DA:C8	1.81	1.15
1:A:3[A]:DG:H2''	1:A:4[A]:DG:C8	1.87	1.08

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 ⓘ

There are no RNA molecules in this entry.

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

4 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	UFP	C	27[A]	1,2	18,21,22	0.81	1 (5%)	26,30,33	0.97	2 (7%)
1	UFP	A	7[A]	1,2	18,21,22	0.85	1 (5%)	26,30,33	1.02	2 (7%)
1	UFP	F	217[B]	1	18,21,22	0.80	1 (5%)	26,30,33	1.01	3 (11%)
1	UFP	H	237[B]	1	18,21,22	0.82	1 (5%)	26,30,33	1.01	2 (7%)

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	UFP	C	27[A]	1,2	-	0/7/21/22	0/2/2/2
1	UFP	A	7[A]	1,2	-	0/7/21/22	0/2/2/2
1	UFP	F	217[B]	1	-	4/7/21/22	0/2/2/2
1	UFP	H	237[B]	1	-	4/7/21/22	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	7[A]	UFP	F5-C5	-3.16	1.30	1.35
1	H	237[B]	UFP	F5-C5	-3.10	1.30	1.35
1	F	217[B]	UFP	F5-C5	-3.08	1.30	1.35
1	C	27[A]	UFP	F5-C5	-3.08	1.30	1.35

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	7[A]	UFP	F5-C5-C6	3.21	123.85	120.97
1	F	217[B]	UFP	F5-C5-C6	3.06	123.71	120.97
1	H	237[B]	UFP	F5-C5-C6	3.03	123.69	120.97
1	H	237[B]	UFP	C6-C5-C4	-2.98	119.82	122.60
1	F	217[B]	UFP	C6-C5-C4	-2.96	119.84	122.60

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	F	217[B]	UFP	C2'-C1'-N1-C6
1	H	237[B]	UFP	C2'-C1'-N1-C6
1	F	217[B]	UFP	C2'-C1'-N1-C2
1	H	237[B]	UFP	C2'-C1'-N1-C2
1	F	217[B]	UFP	O4'-C1'-N1-C6

There are no ring outliers.

4 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	27[A]	UFP	4	0
1	A	7[A]	UFP	7	0
1	F	217[B]	UFP	2	0
1	H	237[B]	UFP	2	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	9/10 (90%)	-0.23	0 100 100	5, 16, 29, 32	9 (100%)
1	C	9/10 (90%)	-0.26	0 100 100	6, 15, 20, 21	9 (100%)
1	F	9/10 (90%)	-0.33	0 100 100	4, 10, 30, 34	9 (100%)
1	H	9/10 (90%)	-0.28	0 100 100	5, 9, 29, 30	9 (100%)
2	B	10/10 (100%)	-0.32	0 100 100	5, 14, 30, 34	10 (100%)
2	D	10/10 (100%)	-0.26	0 100 100	5, 9, 29, 30	10 (100%)
2	E	10/10 (100%)	-0.23	0 100 100	6, 16, 29, 32	10 (100%)
2	G	10/10 (100%)	-0.32	0 100 100	6, 12, 19, 21	10 (100%)
All	All	76/80 (95%)	-0.28	0 100 100	4, 15, 30, 34	76 (100%)

There are no RSRZ outliers to report.

### 6.2 Non-standard residues in protein, DNA, RNA chains [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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
1	UFP	H	237[B]	20/21	0.91	0.16	0,10,18,19	20
1	UFP	C	27[A]	20/21	0.92	0.14	7,12,15,16	20
1	UFP	F	217[B]	20/21	0.92	0.15	22,28,30,31	20
1	UFP	A	7[A]	20/21	0.92	0.14	0,4,14,15	20

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.