



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 11, 2017 – 07:13 AM EST

PDB ID : 5I6T  
Title : Crystal structure of color device state C  
Authors : Hao, Y.; Birktoft, J.; Seeman, N.C.  
Deposited on : 2016-02-16  
Resolution : 5.28 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

<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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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-20028442

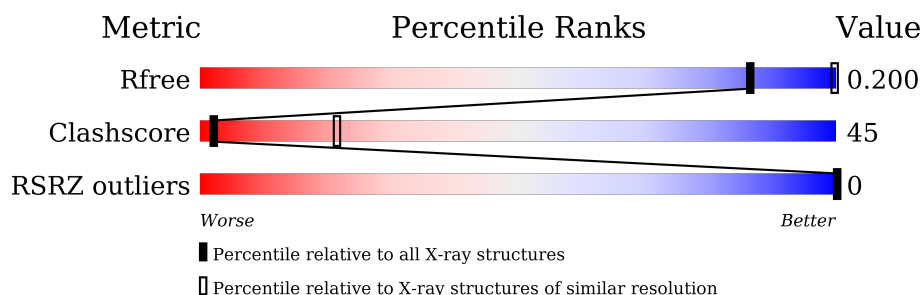
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 5.28 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))
$R_{free}$	91344	1148 (6.92-3.60)
Clashscore	102246	1005 (6.86-3.68)
RSRZ outliers	91569	1149 (6.92-3.60)

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	21	<div> <div>62%</div> <div>38%</div> </div>
2	B	26	<div> <div>23%</div> <div>73%</div> <div>.</div> </div>
3	C	14	<div> <div>7%</div> <div>79%</div> <div>14%</div> </div>
4	D	14	<div> <div>21%</div> <div>71%</div> <div>7%</div> </div>
5	E	14	<div> <div>7%</div> <div>86%</div> <div>7%</div> </div>
6	F	21	<div> <div>90%</div> <div>10%</div> </div>
7	G	21	<div> <div>19%</div> <div>76%</div> <div>5%</div> </div>
8	H	5	<div> <div>100%</div> </div>

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 2767 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(\*GP\*TP\*CP\*CP\*TP\*AP\*CP\*CP\*TP\*GP\*GP\*CP\*AP\*GP\*GP\*AP\*CP\*GP\*AP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	21	Total	C	N	O	P	0	0	0
			426	203	79	124	20			

- Molecule 2 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	26	Total	C	N	O	P	0	0	0
			520	249	93	153	25			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	14	Total	C	N	O	P	0	0	0
			290	139	53	85	13			

- Molecule 4 is a DNA chain called DNA (5'-D(\*AP\*CP\*AP\*GP\*TP\*CP\*GP\*TP\*GP\*GP\*TP\*AP\*TP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	14	Total	C	N	O	P	0	0	0
			285	137	52	83	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	14	Total	C	N	O	P	0	0	0
			287	137	52	85	13			

- Molecule 6 is a DNA chain called DNA (5'-D(\*CP\*AP\*GP\*AP\*TP\*AP\*CP\*CP\*TP\*GP\*

AP\*TP\*CP\*GP\*GP\*AP\*CP\*TP\*AP\*CP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	21	Total	C	N	O	P	0	0	0
			427	204	81	122	20			

- Molecule 7 is a DNA chain called DNA (5'-D(\*GP\*AP\*GP\*CP\*GP\*AP\*CP\*CP\*TP\*GP\*TP\*AP\*CP\*GP\*GP\*AP\*CP\*AP\*TP\*CP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	21	Total	C	N	O	P	0	0	0
			429	204	84	121	20			

- Molecule 8 is a DNA chain called DNA (5'-D(P\*TP\*AP\*GP\*AP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	5	Total	C	N	O	P	0	0	0
			103	49	20	29	5			

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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(\*GP\*TP\*CP\*CP\*TP\*AP\*CP\*CP\*TP\*GP\*GP\*CP\*AP\*GP\*GP\*AP\*CP\*GP\*AP\*CP\*T)-3')

Chain A: 



- Molecule 2: DNA (26-MER)

Chain B: 



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

Chain C: 



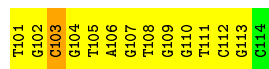
- Molecule 4: DNA (5'-D(\*AP\*CP\*AP\*GP\*TP\*CP\*GP\*TP\*GP\*GP\*TP\*AP\*TP\*C)-3')

Chain D: 

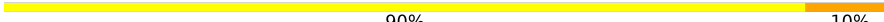


- Molecule 5: DNA (5'-D(\*TP\*GP\*CP\*GP\*TP\*AP\*GP\*TP\*GP\*GP\*TP\*CP\*GP\*C)-3')

Chain E: 




- Molecule 6: DNA (5'-D(\*CP\*AP\*GP\*AP\*TP\*AP\*CP\*CP\*TP\*GP\*AP\*TP\*CP\*GP\*GP\*AP\*CP\*TP\*AP\*CP\*G)-3')

Chain F:  90% 10%

C201  
A202  
G203  
A204  
T205  
A206  
C207  
C208  
T209  
G210  
A211  
T212  
C213  
G214  
G215  
A216  
C217  
T218  
A219  
C220  
G221

- Molecule 7: DNA (5'-D(\*GP\*AP\*GP\*CP\*GP\*AP\*CP\*CP\*TP\*GP\*TP\*AP\*CP\*GP\*GP\*AP\*CP\*AP\*TP\*CP\*A)-3')

Chain G:  19% 76% 5%

G301  
A302  
G303  
C304  
G305  
A306  
C307  
C308  
T309  
G310  
T311  
A312  
C313  
G314  
G315  
A316  
C317  
A318  
T319  
C320  
A321

- Molecule 8: DNA (5'-D(P\*TP\*AP\*GP\*AP\*C)-3')

Chain H:  100%

There are no outlier residues recorded for this chain.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.06Å 69.41Å 69.86Å 98.73° 97.58° 99.27°	Depositor
Resolution (Å)	41.74 – 5.28 41.74 – 5.28	Depositor EDS
% Data completeness (in resolution range)	98.0 (41.74-5.28) 85.8 (41.74-5.28)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.61 (at 5.38Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.155 , 0.203 0.153 , 0.200	Depositor DCC
$R_{free}$ test set	394 reflections (9.76%)	DCC
Wilson B-factor (Å <sup>2</sup> )	249.2	Xtriage
Anisotropy	0.186	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.08 , 108.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.098 for k,l,h 0.098 for l,h,k 0.050 for -k,-h,-l 0.049 for -l,-k,-h 0.059 for -h,-l,-k	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	2767	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	358.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.32% 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 [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	1.35	3/477 (0.6%)	1.30	5/734 (0.7%)
2	B	1.10	0/581	1.17	1/892 (0.1%)
3	C	1.10	0/325	1.24	2/502 (0.4%)
4	D	1.17	1/319 (0.3%)	1.23	1/491 (0.2%)
5	E	0.95	0/321	1.12	1/495 (0.2%)
6	F	1.22	2/479 (0.4%)	1.21	1/737 (0.1%)
7	G	1.08	1/482 (0.2%)	1.06	0/742
8	H	1.04	0/115	0.89	0/175
All	All	1.15	7/3099 (0.2%)	1.18	11/4768 (0.2%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	113	DA	C3'-O3'	-7.42	1.34	1.44
6	F	212	DT	C3'-O3'	6.85	1.52	1.44
6	F	212	DT	C1'-N1	6.57	1.57	1.49
1	A	104	DC	C3'-O3'	-6.06	1.36	1.44
4	D	108	DT	C1'-N1	5.89	1.56	1.49

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	112	DC	O5'-P-OP2	-9.97	96.72	105.70
4	D	108	DT	O4'-C4'-C3'	-8.18	101.09	106.00
2	B	110	DA	O4'-C1'-N9	6.77	112.74	108.00
1	A	114	DG	O5'-P-OP1	6.73	118.78	110.70
6	F	206	DA	O4'-C1'-N9	6.32	112.42	108.00

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	426	0	237	37	0
2	B	520	0	293	40	0
3	C	290	0	161	20	0
4	D	285	0	160	26	0
5	E	287	0	160	17	0
6	F	427	0	237	39	0
7	G	429	0	236	28	0
8	H	103	0	57	0	0
All	All	2767	0	1541	172	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 45.

The worst 5 of 172 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:104:DC:N4	3:C:113:DG:O6	1.96	0.98
1:A:103:DC:N4	3:C:114:DG:O6	2.06	0.87
5:E:110:DG:N2	7:G:308:DC:N3	2.36	0.74
1:A:108:DC:O2	3:C:110:DG:N2	2.19	0.74
2:B:111:DC:H2'	2:B:112:DC:C2	2.27	0.70

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](#)

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, 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	21/21 (100%)	-1.38	0 100 100	235, 292, 367, 383	0
2	B	26/26 (100%)	-1.11	0 100 100	259, 287, 726, 835	0
3	C	14/14 (100%)	-1.42	0 100 100	283, 359, 421, 424	0
4	D	14/14 (100%)	-1.53	0 100 100	273, 340, 398, 399	0
5	E	14/14 (100%)	-1.52	0 100 100	294, 339, 388, 406	0
6	F	21/21 (100%)	-1.43	0 100 100	264, 333, 349, 357	0
7	G	21/21 (100%)	-1.40	0 100 100	307, 325, 380, 391	0
8	H	5/5 (100%)	-0.29	0 100 100	714, 843, 870, 877	0
All	All	136/136 (100%)	-1.33	0 100 100	235, 329, 714, 877	0

There are no RSRZ outliers to report.

### 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 ⓘ

There are no carbohydrates 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.