



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:03 PM GMT

PDB ID : 4NM6  
Title : Crystal structure of TET2-DNA complex  
Authors : Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Wang, P.; Xu, Y.  
Deposited on : 2013-11-14  
Resolution : 2.03 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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

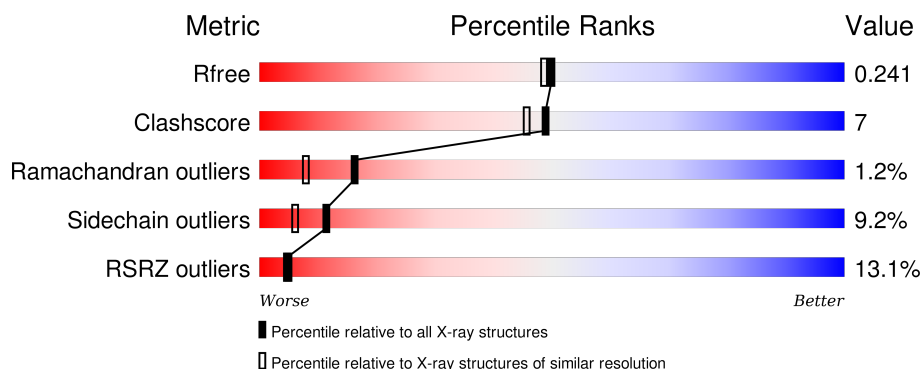
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.03 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	7858 (2.04-2.00)
Clashscore	102246	9060 (2.04-2.00)
Ramachandran outliers	100387	8952 (2.04-2.00)
Sidechain outliers	100360	8951 (2.04-2.00)
RSRZ outliers	91569	7873 (2.04-2.00)

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	463	<div> <div>11%</div> <div> <div></div> <div>71%</div> <div>13%</div> <div>• •</div> <div>12%</div> </div> </div>
2	B	12	<div> <div>33%</div> <div>42%</div> <div>17%</div> <div>8%</div> </div>
2	C	12	<div> <div>25%</div> <div>17%</div> <div>58%</div> <div>8%</div> <div>17%</div> </div>

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
3	ZN	A	2003	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3781 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 Methylcytosine dioxygenase TET2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	407	Total	C	N	O	S	0	1	0
			3205	2003	576	602	24			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1126	GLY	-	EXPRESSION TAG	UNP Q6N021
A	1127	GLY	-	EXPRESSION TAG	UNP Q6N021
A	1128	SER	-	EXPRESSION TAG	UNP Q6N021
A	1829	GLY	-	LINKER	UNP Q6N021
A	1830	GLY	-	LINKER	UNP Q6N021
A	1831	GLY	-	LINKER	UNP Q6N021
A	1832	GLY	-	LINKER	UNP Q6N021
A	1833	SER	-	LINKER	UNP Q6N021
A	1834	GLY	-	LINKER	UNP Q6N021
A	1835	GLY	-	LINKER	UNP Q6N021
A	1836	GLY	-	LINKER	UNP Q6N021
A	1837	GLY	-	LINKER	UNP Q6N021
A	1838	SER	-	LINKER	UNP Q6N021
A	1839	GLY	-	LINKER	UNP Q6N021
A	1840	GLY	-	LINKER	UNP Q6N021
A	1841	GLY	-	LINKER	UNP Q6N021
A	1842	GLY	-	LINKER	UNP Q6N021
A	1843	SER	-	LINKER	UNP Q6N021

- Molecule 2 is a DNA chain called 5'-D(\*AP\*CP\*CP\*AP\*CP\*(5CM)P\*GP\*GP\*TP\*GP\*GP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	P	0	0	0
			224	107	44	63	10			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	P	0	0	0
			207	98	38	61	10			

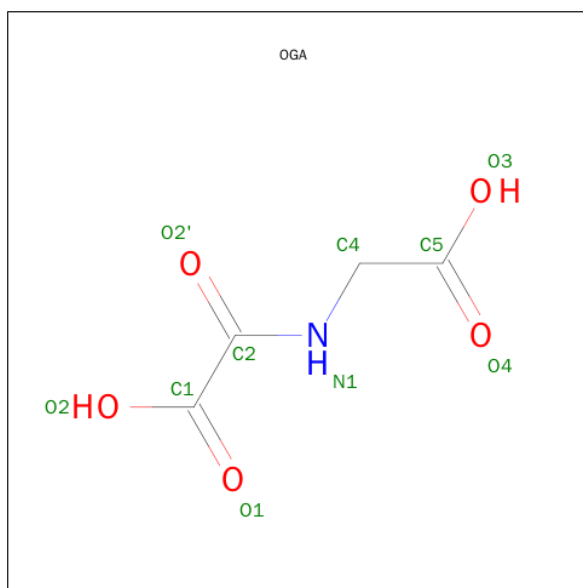
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total	Zn	0	0
			3	3		

- Molecule 4 is FE (II) ION (three-letter code: FE2) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Fe	0	0
			1	1		

- Molecule 5 is N-OXALYLGLYCINE (three-letter code: OGA) (formula: C<sub>4</sub>H<sub>5</sub>NO<sub>5</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			10	4	1	5		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	127	Total	O	0	0
			127	127		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	4	Total	O	0	0
			4	4		



- Molecule 1: Methylcytosine dioxygenase TET2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	48.30 Å 88.19 Å 262.96 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.80 – 2.03 41.80 – 2.03	Depositor EDS
% Data completeness (in resolution range)	93.8 (41.80-2.03) 93.8 (41.80-2.03)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.86 (at 2.03 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.202 , 0.243 0.213 , 0.241	Depositor DCC
$R_{free}$ test set	1732 reflections (5.23%)	DCC
Wilson B-factor (Å <sup>2</sup> )	42.6	Xtriage
Anisotropy	0.452	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 58.5	EDS
Estimated twinning fraction	0.043 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.055 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 34839 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3781	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.24% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: OGA, ZN, FE2, 5CM

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	0.39	0/3272	0.64	5/4408 (0.1%)
2	B	0.58	0/228	1.27	2/348 (0.6%)
2	C	0.63	0/208	1.39	2/317 (0.6%)
All	All	0.42	0/3708	0.76	9/5073 (0.2%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	B	7	DG	O4'-C1'-N9	-7.48	102.77	108.00
1	A	1229	LEU	CA-CB-CG	6.63	130.54	115.30
2	C	3	DC	O4'-C1'-N1	6.62	112.64	108.00
2	B	10	DG	O4'-C1'-N9	6.00	112.20	108.00
1	A	1371	VAL	CB-CA-C	-5.95	100.09	111.40

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	3205	0	3173	38	0
2	B	224	0	126	8	0
2	C	207	0	115	12	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	3	0	0	0	0
4	A	1	0	0	0	0
5	A	10	0	4	1	0
6	A	127	0	0	3	0
6	B	4	0	0	0	0
All	All	3781	0	3418	51	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 7.

The worst 5 of 51 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)
2:B:8:DG:H1	2:C:5:DC:H42	1.26	0.82
2:B:2:DC:O2	2:C:11:DG:N2	2.14	0.77
1:A:1277:ASP:HB3	1:A:1280:THR:HG22	1.67	0.76
1:A:1262:ARG:HB2	2:B:5:DC:H5'	1.68	0.76
1:A:1447:LEU:HB3	1:A:1448:SER:HB2	1.73	0.70

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	402/463 (87%)	385 (96%)	12 (3%)	5 (1%)	<b>16</b> <b>8</b>

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1448	SER
1	A	1411	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	1861	GLY
1	A	1462	LYS
1	A	1410	PRO

### 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	349/385 (91%)	317 (91%)	32 (9%)	11 6

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1362	LEU
1	A	1415	LEU
1	A	1899	LEU
1	A	1394	LEU
1	A	1418	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

2 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	$\# Z  > 2$
2	5CM	B	6	2	13,21,22	1.84	4 (30%)	17,30,33	1.31	2 (11%)
2	5CM	C	6	2	13,21,22	1.82	6 (46%)	17,30,33	1.56	5 (29%)

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
2	5CM	B	6	2	-	0/3/21/22	0/2/2/2
2	5CM	C	6	2	-	0/3/21/22	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	6	5CM	C2'-C3'	-2.09	1.47	1.52
2	B	6	5CM	C2'-C3'	-2.06	1.47	1.52
2	C	6	5CM	C3'-C4'	-2.01	1.47	1.53
2	C	6	5CM	O4'-C1'	2.20	1.47	1.42
2	B	6	5CM	O4'-C1'	2.31	1.47	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	6	5CM	C5-C4-N4	-2.36	118.72	122.20
2	C	6	5CM	C5A-C5-C6	2.11	122.86	118.62
2	C	6	5CM	O5'-C5'-C4'	2.12	116.90	109.12
2	C	6	5CM	N4-C4-N3	2.39	120.41	116.95
2	B	6	5CM	O5'-C5'-C4'	2.68	118.93	109.12

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	6	5CM	3	0

## 5.5 Carbohydrates

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 for Mogul analysis.

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	$\# Z  > 2$
5	OGA	A	2005	4	3,9,9	2.57	1 (33%)	3,11,11	1.14	0

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
5	OGA	A	2005	4	-	0/3/9/9	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	2005	OGA	C2-N1	4.33	1.43	1.33

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	2005	OGA	1	0

## 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	407/463 (87%)	0.63	53 (13%) 5 5	37, 54, 97, 124	0
2	B	10/12 (83%)	0.52	0 100 100	72, 85, 92, 117	0
2	C	9/12 (75%)	1.30	3 (33%) 0 1	79, 88, 134, 139	0
All	All	426/487 (87%)	0.64	56 (13%) 5 5	37, 55, 101, 139	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1917	TRP	7.0
1	A	1922	ALA	6.6
1	A	1919	ALA	6.1
1	A	1218	GLY	4.7
1	A	1463	THR	4.6

### 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. 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
2	5CM	C	6	20/21	0.91	0.26	-	92,99,120,121	0
2	5CM	B	6	20/21	0.84	0.23	-	55,96,105,111	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(Å <sup>2</sup> )	Q<0.9
3	ZN	A	2003	1/1	1.00	0.17	3.65	52,52,52,52	0
3	ZN	A	2001	1/1	0.99	0.12	1.99	51,51,51,51	0
5	OGA	A	2005	10/10	0.95	0.18	-0.52	40,53,65,66	0
3	ZN	A	2002	1/1	0.99	0.04	-2.45	77,77,77,77	0
4	FE2	A	2004	1/1	1.00	0.20	-	43,43,43,43	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.