



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 15, 2022 – 01:33 pm BST

PDB ID : 7NE3  
Title : Human TET2 in complex with favourable DNA substrate.  
Authors : Rafalski, D.; Bochtler, M.  
Deposited on : 2021-02-03  
Resolution : 2.26 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 : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.29  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.29

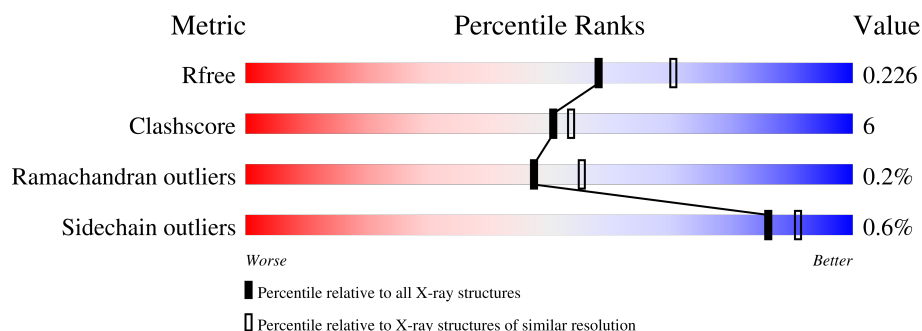
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.26 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	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)

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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	463	
2	B	12	
2	C	12	

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 3969 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	416	Total	C	N	O	S	0	2	0
			3201	1997	570	611	23			

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 DNA (5'-D(\*AP\*CP\*AP\*GP\*GP\*(5CM)P\*GP\*CP\*CP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	P	0	0	0
			227	108	43	65	11			

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	12	Total	C	N	O	P	0	0	0
			247	118	45	72	12			

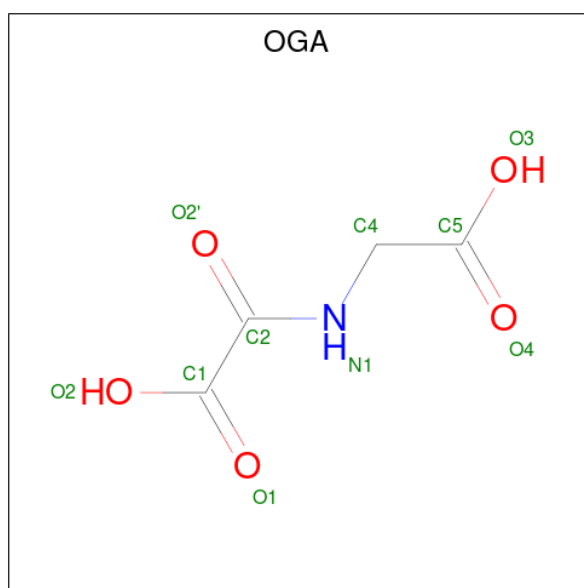
- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

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>).



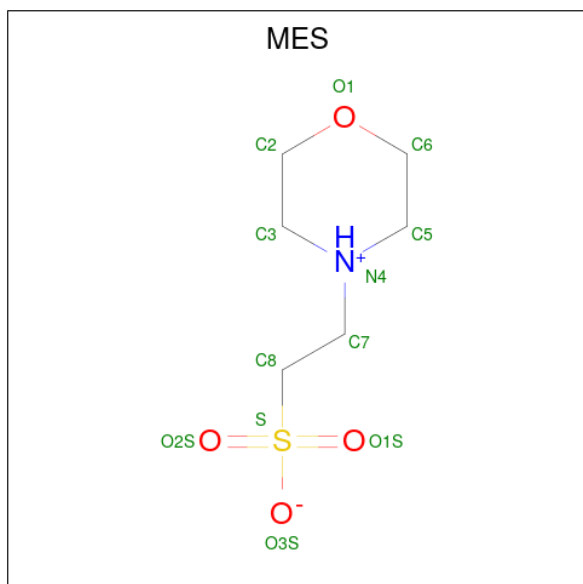
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			10	4	1	5		

- Molecule 6 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			4	2	2		
6	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 7 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula:  $C_6H_{13}NO_4S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	242	Total 242	O 242	0	0
8	B	14	Total 14	O 14	0	0
8	C	4	Total 4	O 4	0	0



- 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, $\alpha$ , $\beta$ , $\gamma$	47.99Å 87.71Å 260.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.86 – 2.26 43.86 – 2.26	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.86-2.26) 99.9 (43.86-2.26)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.59 (at 2.27Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, $R_{free}$	0.175 , 0.226 0.175 , 0.226	Depositor DCC
$R_{free}$ test set	1322 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.5	Xtriage
Anisotropy	0.320	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.032 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.053 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3969	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 4.45% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: OGA, MES, FE2, ZN, EDO, 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	0.40	0/3268	0.55	0/4421
2	B	0.67	0/231	0.88	0/352
2	C	0.75	0/253	0.98	0/386
All	All	0.46	0/3752	0.62	0/5159

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	3201	0	3064	32	0
2	B	227	0	126	6	0
2	C	247	0	138	6	0
3	A	3	0	0	0	0
4	A	1	0	0	0	0
5	A	10	0	3	0	0
6	A	8	0	12	1	0
7	A	12	0	12	0	0
8	A	242	0	0	7	0
8	B	14	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	C	4	0	0	0	0
All	All	3969	0	3355	40	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 6.

The worst 5 of 40 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:C:6:5CM:O4'	2:C:6:5CM:C1'	1.63	1.22
1:A:1211:CYS:HB3	8:A:2324:HOH:O	1.89	0.72
1:A:1262:ARG:HB2	2:B:5:DA:H5'	1.82	0.61
2:C:10:DT:H2''	2:C:11:DG:C8	2.37	0.59
1:A:1369:SER:HB2	1:A:1903:GLN:HG3	1.84	0.59

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	412/463 (89%)	400 (97%)	11 (3%)	1 (0%)	47 55

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1861	GLY

### 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	336/385 (87%)	334 (99%)	2 (1%)	86 91

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1284	SER
1	A	1843	SER

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

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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	17,21,22	5.17	13 (76%)	24,30,33	1.39	5 (20%)
2	5CM	C	6	2	17,21,22	5.40	13 (76%)	24,30,33	1.49	5 (20%)

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	-	4/7/21/22	0/2/2/2
2	5CM	C	6	2	-	5/7/21/22	0/2/2/2

The worst 5 of 26 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	6	5CM	O4'-C1'	9.54	1.63	1.42
2	C	6	5CM	C6-C5	9.36	1.50	1.34
2	B	6	5CM	O4'-C1'	8.87	1.62	1.42
2	B	6	5CM	C6-C5	8.68	1.48	1.34
2	C	6	5CM	C2-N3	7.54	1.51	1.36

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	6	5CM	C5-C6-N1	-3.00	120.25	123.34
2	C	6	5CM	C5-C6-N1	-2.65	120.61	123.34
2	C	6	5CM	O4'-C1'-N1	2.44	112.22	107.86
2	C	6	5CM	O4'-C4'-C3'	-2.39	100.09	105.67
2	B	6	5CM	C1'-N1-C2	2.24	121.67	117.74

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	6	5CM	C2'-C1'-N1-C2
2	C	6	5CM	C3'-C4'-C5'-O5'
2	B	6	5CM	O4'-C1'-N1-C6
2	B	6	5CM	C2'-C1'-N1-C6
2	B	6	5CM	O4'-C1'-N1-C2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	6	5CM	1	0
2	C	6	5CM	1	0

## 5.5 Carbohydrates

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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
7	MES	A	2008	-	12,12,12	2.01	1 (8%)	14,16,16	2.46	7 (50%)
6	EDO	A	2006	-	3,3,3	0.54	0	2,2,2	0.13	0
6	EDO	A	2007	-	3,3,3	0.50	0	2,2,2	0.31	0
5	OGA	A	2005	4	9,9,9	2.86	2 (22%)	10,11,11	1.99	5 (50%)

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
7	MES	A	2008	-	-	3/6/14/14	0/1/1/1
6	EDO	A	2006	-	-	1/1/1/1	-
6	EDO	A	2007	-	-	1/1/1/1	-
5	OGA	A	2005	4	-	0/8/9/9	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	2005	OGA	C2-N1	7.41	1.46	1.33
7	A	2008	MES	C8-S	-6.38	1.68	1.77
5	A	2005	OGA	O2'-C2	-2.76	1.18	1.23

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	2008	MES	C5-N4-C3	5.08	120.26	108.83
7	A	2008	MES	O2S-S-C8	3.98	111.71	106.92
5	A	2005	OGA	O2'-C2-N1	-3.35	117.16	123.30
7	A	2008	MES	C7-N4-C3	3.00	118.91	111.23

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	2008	MES	C7-N4-C5	2.96	118.80	111.23

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	A	2008	MES	C7-C8-S-O1S
7	A	2008	MES	C7-C8-S-O3S
6	A	2007	EDO	O1-C1-C2-O2
7	A	2008	MES	C7-C8-S-O2S
6	A	2006	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	2007	EDO	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.