



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 12:30 pm GMT

PDB ID : 1MUS  
Title : crystal structure of Tn5 transposase complexed with resolved outside end DNA  
Authors : Holden, H.M.; Thoden, J.B.; Steiniger-White, M.; Reznikoff, W.S.; Lovell, S.; Rayment, I.  
Deposited on : 2002-09-24  
Resolution : 1.90 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 : 1.7.2 (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

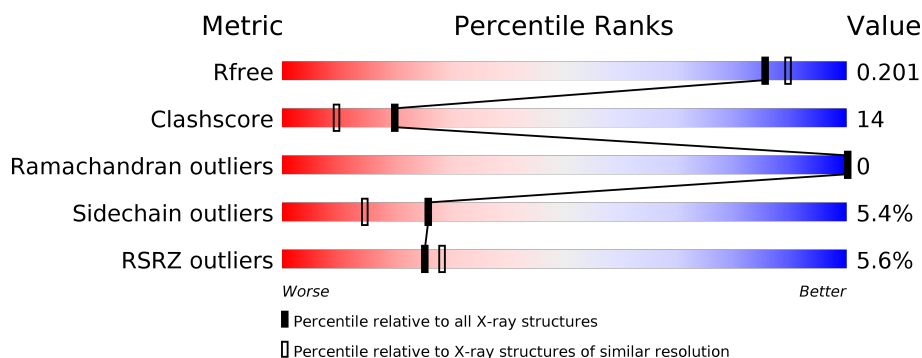
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	5047 (1.90-1.90)
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)
RSRZ outliers	101464	5100 (1.90-1.90)

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	B	20	<div> <div>5%</div> <div>15% 40% 45%</div> </div>
2	C	20	<div> <div>10% 50% 40%</div> </div>
3	A	477	<div> <div>6%</div> <div>66% 25% . . .</div> </div>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 5060 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 transferred strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	20	Total	C	N	O	P	0	0	0
			413	198	78	118	19			

- Molecule 2 is a DNA chain called DNA non-transferred strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	20	Total	C	N	O	P	0	0	0
			405	194	70	121	20			

- Molecule 3 is a protein called Tn5 transposase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	458	Total	C	N	O	S	0	4	0
			3611	2276	664	658	13			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	54	LYS	GLU	ENGINEERED	UNP Q46731
A	56	ALA	MET	ENGINEERED	UNP Q46731
A	119	LYS	ASP	ENGINEERED	UNP Q46731
A	120	ALA	LYS	ENGINEERED	UNP Q46731
A	345	LYS	GLU	ENGINEERED	UNP Q46731
A	372	PRO	LEU	ENGINEERED	UNP Q46731
A	477	GLY	-	CLONING ARTIFACT	UNP Q46731

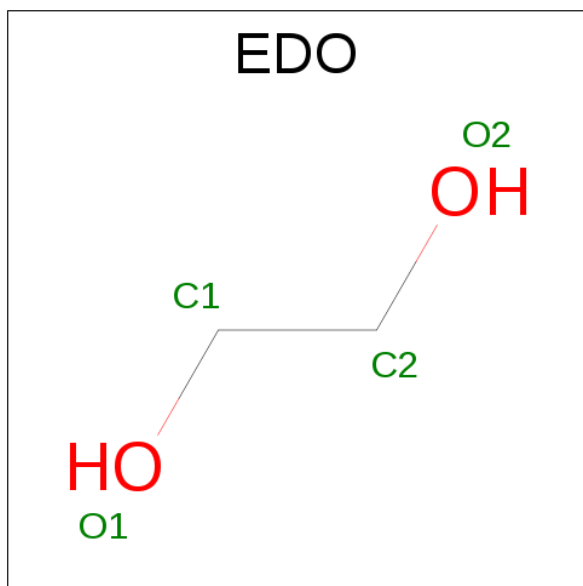
- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Mn	0	0
			2	2		

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Mg	0	0
			2	2		

- 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		
6	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	494	Total	O	0	0
			494	494		
7	B	71	Total	O	0	0
			71	71		
7	C	50	Total	O	0	0
			50	50		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	112.70Å 112.70Å 235.90Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 1.90 25.42 – 1.90	Depositor EDS
% Data completeness (in resolution range)	85.5 (30.00-1.90) 85.7 (25.42-1.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.97 (at 1.91Å)	Xtriage
Refinement program	TNT	Depositor
R, $R_{free}$	0.191 , 0.249 0.205 , 0.201	Depositor DCC
$R_{free}$ test set	5942 reflections (10.94%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.4	Xtriage
Anisotropy	0.305	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 89.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5060	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.79% 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: MG, MN, EDO

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	B	1.34	1/464 (0.2%)	2.06	22/716 (3.1%)
2	C	1.52	4/452 (0.9%)	2.49	31/692 (4.5%)
3	A	0.97	23/3698 (0.6%)	1.45	50/4985 (1.0%)
All	All	1.08	28/4614 (0.6%)	1.67	103/6393 (1.6%)

The worst 5 of 28 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	DC	OP3-P	-7.92	1.51	1.61
3	A	344	GLU	CD-OE2	7.07	1.33	1.25
3	A	219	GLU	CD-OE2	6.54	1.32	1.25
3	A	454	GLU	CD-OE2	6.44	1.32	1.25
3	A	304	GLU	CD-OE2	6.43	1.32	1.25

The worst 5 of 103 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	16	DA	O4'-C1'-N9	-15.99	96.81	108.00
2	C	15	DC	C2-N1-C1'	-13.11	104.38	118.80
2	C	20	DC	C2-N1-C1'	-12.97	104.53	118.80
3	A	324	ARG	NE-CZ-NH1	12.06	126.33	120.30
3	A	27	ARG	NE-CZ-NH2	-11.65	114.48	120.30

There are no chirality outliers.

There are no planarity outliers.

## 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	B	413	0	227	23	0
2	C	405	0	227	14	0
3	A	3611	0	3661	88	1
4	A	2	0	0	0	0
5	A	2	0	0	0	0
6	A	12	0	18	3	0
7	A	494	0	0	18	0
7	B	71	0	0	2	0
7	C	50	0	0	1	0
All	All	5060	0	4133	122	1

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 14.

The worst 5 of 122 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:9:DT:H2''	2:C:10:DA:H5''	1.40	1.01
1:B:4:DT:H6	1:B:4:DT:H5''	1.25	0.98
1:B:4:DT:C6	1:B:4:DT:H5''	2.07	0.89
3:A:115:GLY:HA3	3:A:119:LYS:HD2	1.56	0.88
3:A:274:GLN:NE2	3:A:274:GLN:H	1.77	0.81

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:470:MET:SD	3:A:470:MET:SD[12_555]	2.13	0.07



## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	
3	A	458/477 (96%)	444 (97%)	14 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains

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	
3	A	375/388 (97%)	353 (94%)	22 (6%)	23	12

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	229	LYS
3	A	252	LYS
3	A	415[B]	ARG
3	A	240	SER
3	A	250	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	57	GLN
3	A	81	GLN
3	A	274	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	A	403	GLN

### 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](#)

Of 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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	$\# Z  > 2$
6	EDO	A	482	-	3,3,3	0.38	0	2,2,2	0.29	0
6	EDO	A	483	-	3,3,3	0.66	0	2,2,2	0.53	0
6	EDO	A	484	-	3,3,3	0.54	0	2,2,2	0.34	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
6	EDO	A	482	-	-	0/1/1/1	0/0/0/0
6	EDO	A	483	-	-	0/1/1/1	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	EDO	A	484	-	-	0/1/1/1	0/0/0/0

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.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	482	EDO	1	0
6	A	484	EDO	2	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	B	20/20 (100%)	-0.26	1 (5%) 30 33	23, 30, 65, 71	0
2	C	20/20 (100%)	-0.36	0 100 100	20, 36, 55, 71	0
3	A	458/477 (96%)	0.14	27 (5%) 23 26	16, 28, 64, 99	0
All	All	498/517 (96%)	0.10	28 (5%) 25 28	16, 28, 64, 99	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	291	LYS	6.0
3	A	294	THR	5.1
3	A	251	GLY	4.7
3	A	473	GLY	3.7
3	A	250	ARG	3.7

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 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
6	EDO	A	483	4/4	0.96	0.10	1.84	19,25,31,31	0
6	EDO	A	482	4/4	0.94	0.12	-0.39	45,66,66,72	0
6	EDO	A	484	4/4	0.92	0.12	-0.42	24,34,47,55	0
4	MN	A	478	1/1	0.99	0.03	-3.42	21,21,21,21	0
4	MN	A	479	1/1	0.99	0.03	-3.58	22,22,22,22	0
5	MG	A	480	1/1	0.98	0.05	-	15,15,15,15	1
5	MG	A	481	1/1	0.98	0.07	-	28,28,28,28	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.