



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 20, 2020 – 11:58 PM BST

PDB ID : 6DWY  
Title : Hermes transposase deletion dimer complex with (C/G) DNA and Ca<sup>2+</sup>  
Authors : Dyda, F.; Hickman, A.B.  
Deposited on : 2018-06-28  
Resolution : 3.20 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
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)	:	2.13.1

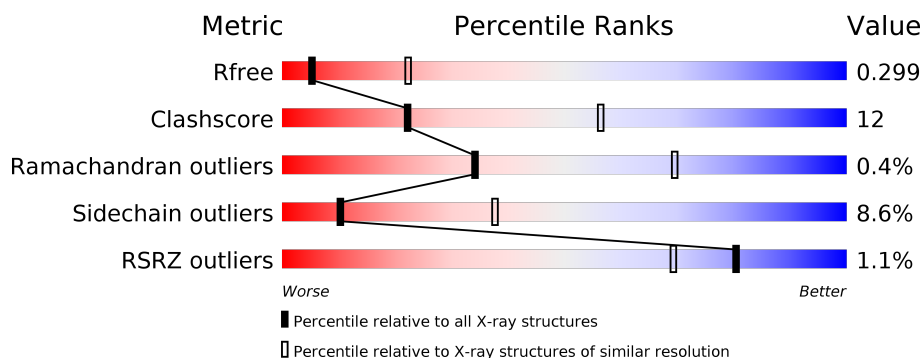
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 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	517	<div> <div></div> <div> <div></div> <div>64%</div> <div>26%</div> <div>7%</div> </div> </div>
2	B	16	<div> <div>13%</div> <div>88%</div> </div>
3	C	26	<div> <div>4%</div> <div>31%</div> <div>65%</div> <div></div> </div>
4	D	8	<div> <div>50%</div> <div>38%</div> <div>13%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4867 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 Hermes transposase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	483	Total	C	N	O	S	0	0	0
			3888	2489	658	722	19			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	76	GLY	-	expression tag	UNP Q25438
A	77	SER	-	expression tag	UNP Q25438
A	78	HIS	-	expression tag	UNP Q25438
A	79	MET	-	expression tag	UNP Q25438
A	128	GLY	LYS	conflict	UNP Q25438
A	?	-	ASP	deletion	UNP Q25438
A	?	-	ILE	deletion	UNP Q25438
A	?	-	SER	deletion	UNP Q25438
A	?	-	THR	deletion	UNP Q25438
A	?	-	THR	deletion	UNP Q25438
A	?	-	SER	deletion	UNP Q25438
A	?	-	PHE	deletion	UNP Q25438
A	?	-	PHE	deletion	UNP Q25438
A	?	-	PHE	deletion	UNP Q25438
A	?	-	PRO	deletion	UNP Q25438
A	?	-	GLN	deletion	UNP Q25438
A	?	-	LEU	deletion	UNP Q25438
A	?	-	THR	deletion	UNP Q25438
A	?	-	GLN	deletion	UNP Q25438
A	?	-	ASN	deletion	UNP Q25438
A	?	-	ASN	deletion	UNP Q25438
A	?	-	SER	deletion	UNP Q25438
A	?	-	ARG	deletion	UNP Q25438
A	?	-	GLU	deletion	UNP Q25438
A	?	-	PRO	deletion	UNP Q25438
A	519	SER	CYS	engineered mutation	UNP Q25438

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	16	Total	C	N	O	P	0	0	0
			330	157	74	84	15			

- Molecule 3 is a DNA chain called DNA (26-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	25	Total	C	N	O	P	0	0	0
			506	244	80	158	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	7	Total	C	N	O	P	0	0	0
			142	68	28	40	6			

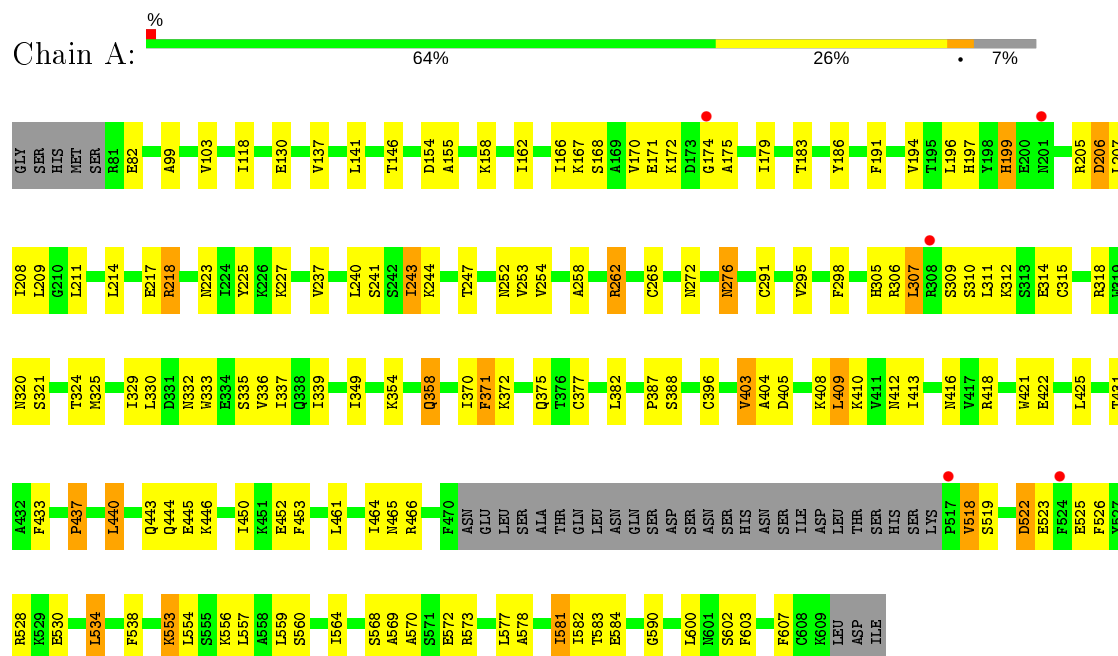
- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Ca	0	0
			1	1		

### 3 Residue-property plots [i](#)

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: Hermes transposase



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

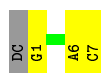


- Molecule 3: DNA (26-MER)



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

Chain D:  50% 38% 13%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	121.38Å 135.35Å 103.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.55 – 3.20 29.55 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.7 (29.55-3.20) 99.7 (29.55-3.10)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.05 (at 3.11Å)	Xtriage
Refinement program	PHENIX (1.10 _2155: ???)	Depositor
R, $R_{free}$	0.259 , 0.295 0.263 , 0.299	Depositor DCC
$R_{free}$ test set	943 reflections (6.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	78.3	Xtriage
Anisotropy	0.767	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.22 , 20.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4867	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	95.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.93% 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:  
CA

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	0.35	1/3962 (0.0%)	0.46	0/5343
2	B	0.65	0/374	0.76	0/575
3	C	0.73	0/563	1.09	0/868
4	D	0.73	0/159	0.89	0/244
All	All	0.45	1/5058 (0.0%)	0.62	0/7030

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	437	PRO	C-N	12.98	1.58	1.34

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	241	SER	Peptide
1	A	590	GLY	Peptide



## 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	A	3888	0	3957	81	0
2	B	330	0	178	9	0
3	C	506	0	288	18	0
4	D	142	0	80	4	0
5	A	1	0	0	0	0
All	All	4867	0	4503	108	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 12.

The worst 5 of 108 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:519:SER:HB3	1:A:522:ASP:HB2	1.51	0.90
1:A:254:VAL:O	1:A:262:ARG:NH1	2.16	0.78
1:A:525:GLU:HG3	1:A:528:ARG:HH21	1.52	0.74
1:A:199:HIS:HD2	1:A:559:LEU:HB3	1.56	0.70
1:A:312:LYS:HG3	1:A:324:THR:HG22	1.74	0.69

There are no symmetry-related clashes.

## 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
1	A	479/517 (93%)	441 (92%)	36 (8%)	2 (0%)	34 69

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	518	VAL
1	A	314	GLU

### 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	
1	A	442/475 (93%)	404 (91%)	38 (9%)	10	38

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	307	LEU
1	A	371	PHE
1	A	553	LYS
1	A	315	CYS
1	A	403	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	197	HIS
1	A	358	GLN
1	A	362	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

## 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	483/517 (93%)	-0.03	5 (1%) 82 72	64, 93, 136, 151	0
2	B	16/16 (100%)	-0.20	0 100 100	53, 83, 108, 111	0
3	C	25/26 (96%)	-0.15	1 (4%) 38 25	68, 89, 104, 134	0
4	D	7/8 (87%)	-0.56	0 100 100	71, 78, 102, 103	0
All	All	531/567 (93%)	-0.04	6 (1%) 80 69	53, 93, 135, 151	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	524	PHE	3.2
1	A	517	PRO	2.9
1	A	308	ARG	2.8
1	A	174	GLY	2.8
1	A	201	ASN	2.3

### 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 monosaccharides 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. 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( $\text{\AA}^2$ )	Q<0.9
5	CA	A	701	1/1	0.93	0.34	68,68,68,68	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.