



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 4, 2026 – 11:22 PM UTC

PDB ID : 4E7L / pdb\_00004e7l  
Title : PFV integrase Strand Transfer Complex (STC-Mn\*) following reaction in crystallo, at 3.0 Å resolution.  
Authors : Maertens, G.N.; Cherepanov, P.  
Deposited on : 2012-03-17  
Resolution : 3.00 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 : 4-5-2 with Phenix2.0  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.49

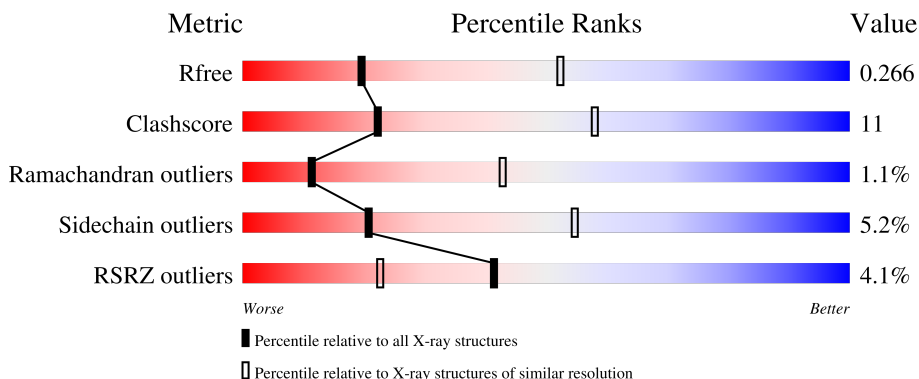
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	180053	2672 (3.00-3.00)
Clashscore	190562	2977 (3.00-3.00)
Ramachandran outliers	187476	2877 (3.00-3.00)
Sidechain outliers	187428	2880 (3.00-3.00)
RSRZ outliers	180081	2671 (3.00-3.00)

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 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  $\leq 5\%$ . 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	395	 2% 69% 23% 8%
1	B	395	 3% 26% 12% 59%
2	C	19	 5% 53% 42% 5%
3	D	17	 6% 29% 59% 12%
4	t	13	 23% 31% 46%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	T	17	

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
8	SO4	B	401	-	-	X	-

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 5236 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 Pro-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	365	Total 2887	C 1853	N 508	O 522	S 4	0	0	0
1	B	163	Total 1275	C 836	N 206	O 232	S 1	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P14350
A	-1	PRO	-	expression tag	UNP P14350
A	0	GLY	-	expression tag	UNP P14350
A	217	SER	GLY	variant	UNP P14350
A	218	GLY	SER	variant	UNP P14350
B	-2	GLY	-	expression tag	UNP P14350
B	-1	PRO	-	expression tag	UNP P14350
B	0	GLY	-	expression tag	UNP P14350
B	217	SER	GLY	variant	UNP P14350
B	218	GLY	SER	variant	UNP P14350

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	19	Total 387	C 187	N 68	O 114	P 18	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	D	17	Total 345	C 166	N 65	O 98	P 16	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	t	7	124	58	23	37	6	0	0	1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	T	11	209	97	38	63	11	0	0	0

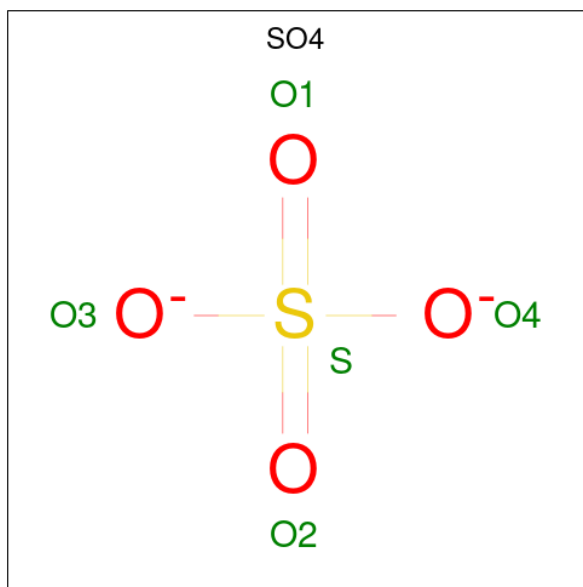
- Molecule 6 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
6	A	1	1	1	0	0

- Molecule 7 is MANGANESE (II) ION (CCD ID: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
7	A	2	2	2	0	0
7	B	1	1	1	0	0

- Molecule 8 is SULFATE ION (CCD ID: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	B	1	Total	O	S	0	0
			5	4	1		



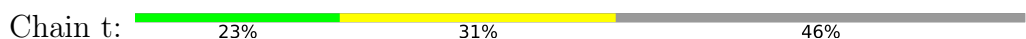
- Molecule 2: DNA (5'-D(\*AP\*TP\*TP\*GP\*TP\*CP\*AP\*TP\*GP\*GP\*AP\*AP\*TP\*TP\*TP\*CP\*GP\*CP\*A)-3')



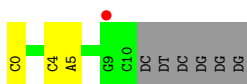
- Molecule 3: DNA (5'-D(\*TP\*GP\*CP\*GP\*AP\*AP\*AP\*TP\*TP\*CP\*CP\*AP\*TP\*GP\*AP\*CP\*A)-3')



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



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	159.08Å 159.08Å 126.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.95 – 3.00 37.95 – 3.00	Depositor EDS
% Data completeness (in resolution range)	98.5 (37.95-3.00) 98.2 (37.95-3.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.12 (at 3.01Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.228 , 0.264 0.226 , 0.266	Depositor DCC
$R_{free}$ test set	1625 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	63.9	Xtrriage
Anisotropy	0.350	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 43.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	5236	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.53% 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: MN, ZN, SO4

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	0.48	0/2965	0.86	2/4049 (0.0%)
1	B	0.52	0/1316	0.97	5/1804 (0.3%)
2	C	0.40	0/433	1.11	2/667 (0.3%)
3	D	0.42	0/387	1.23	3/595 (0.5%)
4	t	0.27	0/138	0.86	0/212
5	T	0.33	0/233	0.97	0/358
All	All	0.47	0/5472	0.95	12/7685 (0.2%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	17	DA	O4'-C1'-N9	6.71	118.46	108.40
1	B	249	VAL	CB-CA-C	-6.68	103.33	111.88
2	C	19	DA	P-O5'-C5'	-6.37	110.44	120.00
1	B	238	ARG	CA-C-N	6.37	126.06	119.56
1	B	238	ARG	C-N-CA	6.37	126.06	119.56

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	2887	0	2918	55	1
1	B	1275	0	1260	40	0
2	C	387	0	218	7	0
3	D	345	0	192	6	1
4	t	124	0	68	5	0
5	T	209	0	112	3	0
6	A	1	0	0	0	0
7	A	2	0	0	0	0
7	B	1	0	0	0	0
8	B	5	0	0	2	0
All	All	5236	0	4768	103	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 11.

The worst 5 of 103 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:B:222:ARG:NH1	8:B:401:SO4:O4	2.14	0.81
1:A:69:ARG:NH1	1:A:70:GLU:OE2	2.20	0.73
1:A:180:LYS:NZ	1:B:275:ASN:OD1	2.22	0.72
1:A:174:THR:HB	1:A:178:ILE:HD13	1.72	0.71
1:B:227:ILE:HG12	1:B:253:LEU:HD13	1.77	0.66

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)
1:A:311:SER:OG	3:D:13:DT:OP1[8_554]	2.16	0.04

## 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	363/395 (92%)	338 (93%)	22 (6%)	3 (1%)	16	50
1	B	161/395 (41%)	150 (93%)	8 (5%)	3 (2%)	6	30
All	All	524/790 (66%)	488 (93%)	30 (6%)	6 (1%)	11	43

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	216	SER
1	A	328	ALA
1	B	259	PRO
1	A	22	LYS
1	B	135	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	322/354 (91%)	310 (96%)	12 (4%)	30	64
1	B	141/354 (40%)	129 (92%)	12 (8%)	10	36
All	All	463/708 (65%)	439 (95%)	24 (5%)	21	55

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	232	THR
1	B	248	VAL
1	B	240	THR
1	B	258	SER
1	A	215	GLN

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

Mol	Chain	Res	Type
1	A	255	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	280	ASN
1	A	324	GLN
1	A	360	ASN

### 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 oligosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

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	$\# Z  > 2$
8	SO4	B	401	-	4,4,4	0.24	0	6,6,6	0.20	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.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	B	401	SO4	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	A	365/395 (92%)	0.00	9 (2%) 58 35	37, 58, 90, 110	0
1	B	163/395 (41%)	0.42	12 (7%) 20 10	38, 62, 98, 122	0
2	C	19/19 (100%)	0.08	1 (5%) 32 16	41, 61, 80, 97	0
3	D	17/17 (100%)	0.03	1 (5%) 28 14	41, 54, 80, 109	0
4	t	7/13 (53%)	0.41	0 100 100	50, 63, 108, 122	0
5	T	11/17 (64%)	0.37	1 (9%) 15 8	59, 65, 120, 142	0
All	All	582/856 (67%)	0.13	24 (4%) 41 23	37, 59, 95, 142	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	278	PHE	3.8
1	B	257	TYR	3.7
1	B	259	PRO	3.3
2	C	1	DA	3.2
1	B	213	HIS	3.1

### 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 oligosaccharides 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(Å <sup>2</sup> )	Q<0.9
8	SO4	B	401	5/5	0.89	0.14	97,98,111,122	0
7	MN	A	402	1/1	0.96	0.05	67,67,67,67	0
7	MN	A	403	1/1	0.97	0.12	66,66,66,66	0
7	MN	B	402	1/1	0.99	0.10	53,53,53,53	0
6	ZN	A	401	1/1	1.00	0.02	56,56,56,56	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.