



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 09:25 PM UTC

PDB ID : 5THG / pdb\_00005thg  
Title : Engineered variant of I-OnuI meganuclease targeting the HIV CCR5 gene;  
harbors 43 point mutations relative to wild-type I-OnuI  
Authors : Hallinan, J.P.; Stoddard, B.L.  
Deposited on : 2016-09-29  
Resolution : 3.11 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

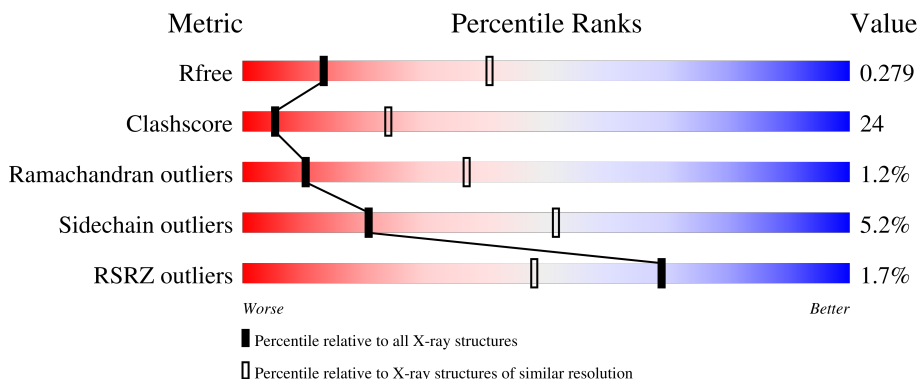
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.11 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.



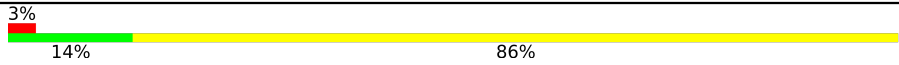
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	180053	1456 (3.10-3.10)
Clashscore	190562	1539 (3.10-3.10)
Ramachandran outliers	187476	1467 (3.10-3.10)
Sidechain outliers	187428	1467 (3.10-3.10)
RSRZ outliers	180081	1456 (3.10-3.10)

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	304	
1	D	304	
2	B	29	
2	E	29	
3	C	29	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	29	 3% 14% 86%

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 6772 atoms, of which 16 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 I-OnuI\_e-hCCR5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	294	Total 2261	C 1455	N 390	O 408	S 8	0	0	0
1	D	291	Total 2149	C 1382	N 369	O 390	S 8	0	0	0

- Molecule 2 is a DNA chain called DNA (29-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	28	Total 565	C 270	N 96	O 171	P 28	0	0	0
2	E	28	Total 565	C 270	N 96	O 171	P 28	0	0	0

- Molecule 3 is a DNA chain called DNA (29-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	28	Total 583	C 275	N 115	O 165	P 28	0	0	0
3	F	29	Total 605	C 285	N 120	O 171	P 29	0	0	0

- Molecule 4 is CALCIUM ION (CCD ID: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total 2 Ca 2	0	0
4	D	2	Total 2 Ca 2	0	0

- Molecule 5 is GLYCEROL (CCD ID: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).

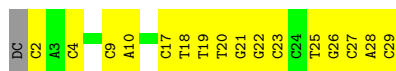


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	
5	B	1	Total	C	H	O	0	0
			14	3	8	3		
5	E	1	Total	C	H	O	0	0
			14	3	8	3		

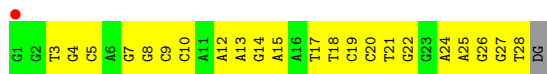
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	2	Total	O	0	0
			2	2		
6	B	2	Total	O	0	0
			2	2		
6	C	1	Total	O	0	0
			1	1		
6	D	1	Total	O	0	0
			1	1		
6	E	4	Total	O	0	0
			4	4		
6	F	2	Total	O	0	0
			2	2		

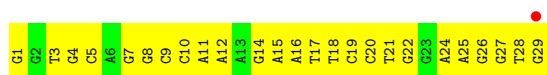




- Molecule 3: DNA (29-MER)



- Molecule 3: DNA (29-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.96Å 112.35Å 122.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.77 – 3.11 47.77 – 3.11	Depositor EDS
% Data completeness (in resolution range)	98.7 (47.77-3.11) 90.8 (47.77-3.11)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.11 (at 3.12Å)	Xtrriage
Refinement program	PHENIX (1.10_2155)	Depositor
R, $R_{free}$	0.221 , 0.279 0.223 , 0.279	Depositor DCC
$R_{free}$ test set	1948 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.5	Xtrriage
Anisotropy	0.832	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 59.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6772	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	55.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 79.62 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.3637e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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, GOL

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	0.22	0/2304	0.38	0/3115
1	D	0.20	0/2191	0.35	0/2976
2	B	0.40	0/630	0.57	0/968
2	E	0.34	0/630	0.52	0/968
3	C	0.35	0/656	0.46	0/1012
3	F	0.33	0/681	0.44	0/1051
All	All	0.27	0/7092	0.42	0/10090

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	2261	0	2178	97	0
1	D	2149	0	1952	97	1
2	B	565	0	317	36	0
2	E	565	0	317	31	0
3	C	583	0	314	31	0
3	F	605	0	325	39	1
4	A	2	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	2	0	0	0	0
5	B	6	8	8	0	0
5	E	6	8	8	1	0
6	A	2	0	0	0	0
6	B	2	0	0	1	0
6	C	1	0	0	0	0
6	D	1	0	0	0	0
6	E	4	0	0	2	0
6	F	2	0	0	0	0
All	All	6756	16	5419	293	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 24.

The worst 5 of 293 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:E:21:DG:OP1	6:E:201:HOH:O	1.83	0.96
2:E:22:DG:N1	3:F:9:DC:N3	2.16	0.93
5:E:101:GOL:HO1	5:E:101:GOL:HO3	1.16	0.92
2:E:22:DG:N2	3:F:9:DC:O2	2.04	0.90
1:A:232:GLY:H	1:A:233:SER:HA	1.38	0.88

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:D:257:THR:O	3:F:1:DG:N2[2_485]	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	288/304 (95%)	256 (89%)	27 (9%)	5 (2%)	7	30
1	D	283/304 (93%)	259 (92%)	22 (8%)	2 (1%)	18	49
All	All	571/608 (94%)	515 (90%)	49 (9%)	7 (1%)	10	37

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	8	GLU
1	A	231	LYS
1	A	194	VAL
1	D	7	ARG
1	D	302	VAL

### 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	227/274 (83%)	216 (95%)	11 (5%)	23	54
1	D	199/274 (73%)	188 (94%)	11 (6%)	19	50
All	All	426/548 (78%)	404 (95%)	22 (5%)	21	51

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

Mol	Chain	Res	Type
1	D	50	HIS
1	D	191	SER
1	D	80	SER
1	D	199	ARG
1	A	193	LYS

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

Mol	Chain	Res	Type
1	A	51	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	93	HIS
1	A	256	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 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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	GOL	B	101	-	5,5,5	0.35	0	5,5,5	0.35	0
5	GOL	E	101	-	5,5,5	0.41	0	5,5,5	0.36	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	GOL	B	101	-	-	0/4/4/4	-
5	GOL	E	101	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	E	101	GOL	O1-C1-C2-C3
5	E	101	GOL	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	E	101	GOL	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	294/304 (96%)	-0.13	4 (1%) 73 53	26, 44, 91, 110	0
1	D	291/304 (95%)	0.01	6 (2%) 63 42	32, 54, 98, 127	0
2	B	28/29 (96%)	-0.09	0 100 100	33, 47, 75, 94	0
2	E	28/29 (96%)	0.11	0 100 100	48, 60, 90, 113	0
3	C	28/29 (96%)	0.09	1 (3%) 46 26	37, 49, 73, 80	0
3	F	29/29 (100%)	0.13	1 (3%) 48 28	47, 60, 94, 95	0
All	All	698/724 (96%)	-0.04	12 (1%) 69 48	26, 50, 94, 127	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	32	THR	3.2
1	A	157	GLU	2.8
1	A	230	SER	2.5
1	D	36	SER	2.4
1	D	228	ASN	2.4

### 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
5	GOL	E	101	6/6	0.83	0.12	58,69,80,80	0
5	GOL	B	101	6/6	0.89	0.09	46,59,76,79	0
4	CA	D	402	1/1	0.94	0.13	65,65,65,65	0
4	CA	D	401	1/1	0.96	0.04	56,56,56,56	0
4	CA	A	401	1/1	0.96	0.04	50,50,50,50	0
4	CA	A	402	1/1	0.99	0.05	36,36,36,36	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.